ONTAPとエンタープライズアプリケーション

Enterprise applications

NetApp
May 09, 2024

目次

ONTAPとエンタープライズアプリケーション

Hyper-V

導入ガイドラインとストレージのベストプラクティス

概要

Microsoft Windows Serverは、ネットワーク、セキュリティ、仮想化、プライベートクラウド、ハイブリッドクラウド、 仮想デスクトップインフラ、アクセス保護、情報保護、Webサービス、アプリケーションプラットフォームインフラ、 その他多数。

このドキュメントは、以前に公開されていたテクニカルレポート**_TR-4568：『NetApp Deployment Guidelines and Storage Best Practices for Windows Server_』**の内容を置き換えます。

NetApp ONTAP®管理ソフトウェアは、**NetApp**ストレージコントローラ上で動作します。複数の形式で使用できます。

- ファイル、オブジェクト、ブロックの各プロトコルをサポートするユニファイドアーキテクチャ。これにより、ストレージコントローラがNASデバイスとSANデバイスの両方、およびオブジェクトストアとして機能できるようになります。
- ブロックプロトコルのみに重点を置き、接続ホストに対称アクティブ/アクティブマルチパスを追加することでI/O再開時間（IORT）を最適化するオールSANアレイ（ASA）
- ソフトウェア定義型のユニファイドアーキテクチャ
  - VMware vSphereまたはKVMで実行されるONTAP Select
  - クラウドネイティブインスタンスとして実行されるCloud Volumes ONTAP
- ハイパースケールクラウドプロバイダが提供するファーストパーティ製品
  - NetApp ONTAP 対応の Amazon FSX
  - Azure NetApp Files の特長
  - Google Cloud NetAppボリューム

ONTAPは、NetApp Snapshot（R）テクノロジ、クローニング、重複排除、シンプロビジョニング、シンレプリケーションなどのNetApp Storage Efficiency機能を提供 圧縮、バーチャルストレージ階層化など、パフォーマンスと効率性が向上しています。

Windows ServerとONTAPを併用すれば、大規模な環境でも運用でき、データセンターの統合やプライベートクラウドやハイブリッドクラウドの導入に大きな価値をもたらすことができます。また、この組み合わせにより、システムを停止することなくワークロードを効率的に実行でき、シームレスな拡張性

対象読者

本ドキュメントは、Windows Server向けのNetAppストレージソリューションを設計するシステムアーキテクトおよびストレージアーキテクトを対象としています。

本ドキュメントでは、次のことを前提としています。

- NetAppのハードウェアおよびソフトウェアソリューションに関する一般的な知識がある。を参照してくだ

さい "システムアドミニストレーションガイド（クラスタ管理）" を参照してください。

- iSCSI、FC、ファイルアクセスプロトコルSMB / CIFSなどのブロックアクセスプロトコルに関する一般的な知識がある読者。を参照してください "clustered Data ONTAPのSAN管理" を参照してください。を参照してください "NAS管理" を参照してください。
- 読者には、Windows Server OSおよびHyper-Vに関する一般的な知識があります。

テスト済みでサポートされているSANおよびNAS構成のマトリックスについては、定期的に更新される完全なマトリックスを参照してください。 "Interoperability Matrix Tool （ IMT ）" NetApp Support Site上。IMTを使用すると、特定の環境でサポートされている製品や機能のバージョンを確認できます。NetApp IMTには、NetAppでサポートされる構成と互換性のある製品コンポーネントとバージョンが定義されています。サポートの可否は、お客様の実際のインストール環境が公表されている仕様に従っているかどうかによって異なります。

NetAppストレージおよびWindows Server環境

に記載されているように "概要"NetAppストレージコントローラは、ファイル、ブロック、オブジェクトの各プロトコルをサポートする真のユニファイドアーキテクチャを提供します。これには、SMB / CIFS、NFS、NVMe/TCP、NVMe/FC、iSCSI、 FC（FCP）とS3が統合され、クライアントとホストのアクセスが統合されます。同じストレージコントローラで、NFSやSMB / CIFSと同様にSAN LUNとファイルサービスという形式のブロックストレージサービスを同時に提供できます。ONTAPは、iSCSIおよびFCPとの対称アクティブ/アクティブマルチパスによってホストアクセスを最適化するオールSANアレイ（ASA）としても利用できますが、Unified ONTAPシステムでは非対称アクティブ/アクティブマルチパスが使用されます。どちらのモードでも、ONTAPはNVMe over Fabrics（NVMe-oF）マルチパス管理にANAを使用します。

ONTAPソフトウェアを実行するNetAppストレージコントローラは、Windows Server環境で次のワークロードをサポートします。

- 継続的可用性を備えたSMB 3.0共有でホストされるVM
- iSCSIまたはFCで実行されているCluster Shared Volume（CSV；クラスタ共有ボリューム）LUNでホストされているVM
- SMB 3.0共有上のSQL Serverデータベース
- NVMe-oF、iSCSI、FC上のSQL Serverデータベース
- その他のアプリケーションワークロード

さらに、NetApp重複排除、NetApp FlexClone（R）コピー、NetApp Snapshotテクノロジ、シンプロビジョニング、圧縮、 また、ストレージ階層化は、Windows Serverで実行されるワークロードに大きな価値をもたらします。

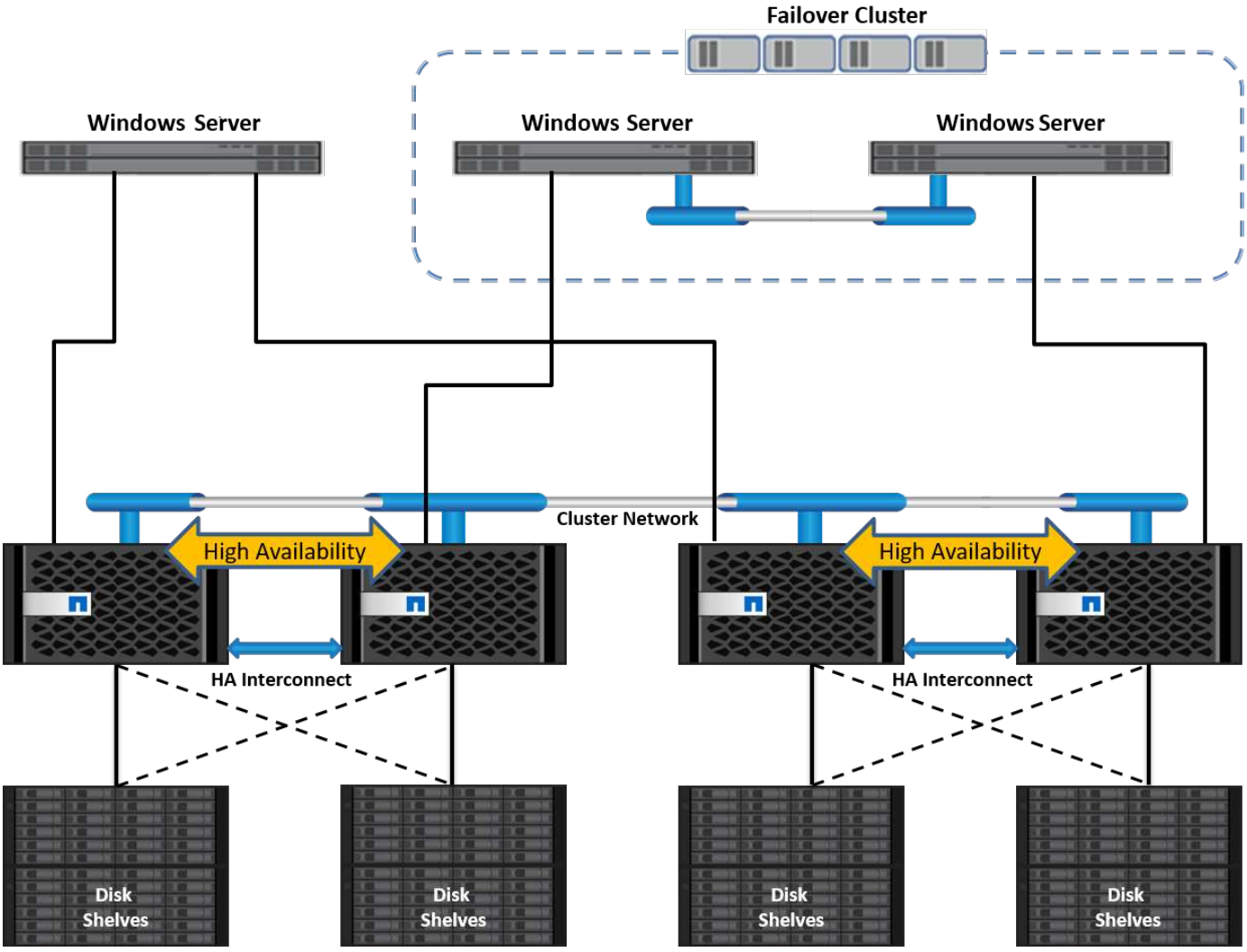
ONTAPデータ管理

ONTAPは、NetAppストレージコントローラ上で実行される管理ソフトウェアです。NetAppストレージコントローラはノードと呼ばれ、プロセッサ、RAM、NVRAMを搭載したハードウェアデバイスです。ノードは、SATA、SAS、SSDのディスクドライブ、またはそれらのドライブの組み合わせに接続できます。

複数のノードが1つのクラスタシステムに集約されます。クラスタ内のノードは相互に継続的に通信し、クラ

スタのアクティビティを調整します。2つの10Gbイーサネットスイッチで構成される専用のクラスタネットワークへの冗長パスを使用することで、ノード間でデータを透過的に移動することもできます。クラスタ内のノードが相互にテイクオーバーして、フェイルオーバー時の高可用性を実現できます。クラスタは、ノード単位ではなくクラスタ全体が1つの単位として管理され、データは1つ以上のStorage Virtual Machine（SVM）から提供されます。クラスタからデータを提供するには、少なくとも1つのSVMが必要です。

クラスタの基本単位はノードで、ノードはハイアベイラビリティ（HA）ペアの一部としてクラスタに追加されます。HAペアは、（専用のクラスタネットワークから分離された）HAインターコネクトを介して相互に通信し、HAペアのディスクへの接続を冗長化することで、高可用性を実現します。シェルフにはHAペアのどちらかのメンバーに属するディスクが含まれている場合もありますが、HAペア間でディスクが共有されることはありません。次の図は、Windows Server環境におけるNetAppストレージの導入を示しています。

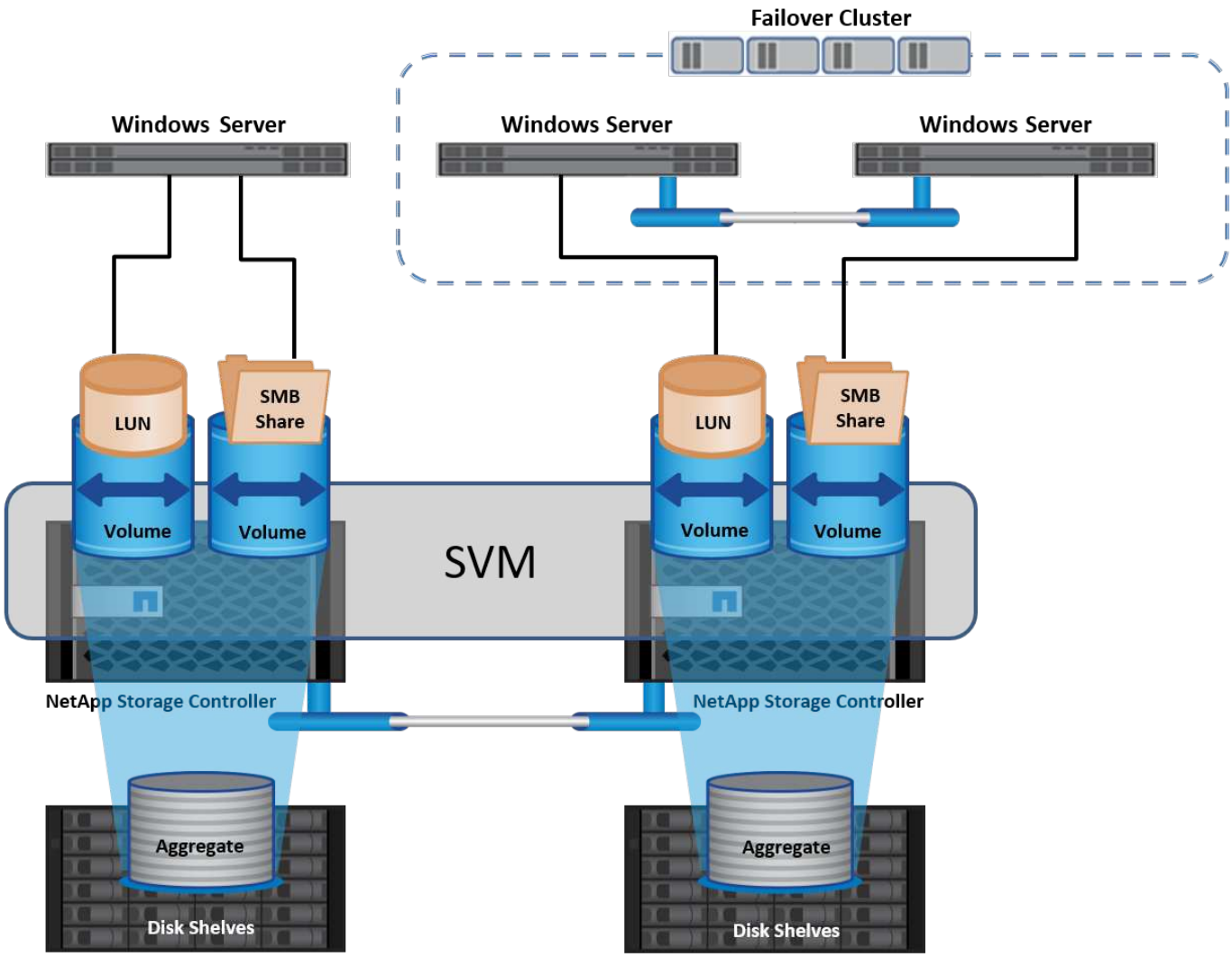


Storage Virtual Machine

ONTAP SVMは、1つ以上の論理インターフェイス（LIF）からLUNやNASネームスペースへのデータアクセスを提供する論理ストレージサーバです。したがって、SVMはストレージセグメント化の基本単位であり、ONTAPでセキュアマルチテナンシーを実現します。各SVMは、物理アグリゲートからプロビジョニングされた専用のストレージボリュームと、物理イーサネットネットワークまたはFCターゲットポートに割り当てられた論理インターフェイス（LIF）で構成されます。

論理ディスク（LUN）またはCIFS共有は、SVMのボリューム内に作成され、Windowsホストおよびクラスタにマッピングされてストレージスペースを提供します（次の図を参照）。SVMはノードに依存せず、クラス

タベースです。クラスタ内の任意の場所のボリュームやネットワークポートなどの物理リソースを使用できます。



Windows Server用のNetAppストレージのプロビジョニング

ストレージは、SAN環境とNAS環境の両方でWindows Serverにプロビジョニングできます。SAN環境では、ストレージはNetApp上のLUNのディスクとしてブロックストレージとして提供されます。NAS環境では、ストレージはファイルストレージとしてNetAppボリューム上のCIFS/SMB共有として提供されます。これらのディスクと共有は、次のようにWindows Serverに適用できます。

- アプリケーションワークロード用のWindows Serverホスト用ストレージ
- ナノサーバおよびコンテナ向けストレージ
- VMを格納するための個 々 のHyper-Vホストのストレージ
- VMを格納するCSV形式のHyper-Vクラスタ用共有ストレージ
- SQL Serverデータベース用のストレージ

NetAppストレージの管理

Windows Server 2016からNetAppストレージに接続、構成、および管理するには、次のいずれかの方法を使用します。

- セキュアシェル**(SSH)**。 Windows Server上の任意のSSHクライアントを使用して、NetApp CLIコマンドを実行します。
- * System Manager。*ネットアップのGUIベースの管理機能製品です。
- * NetApp PowerShell Toolkit。*これは、カスタムスクリプトおよびワークフローを自動化および実装するためのNetApp PowerShell Toolkitです。

NetApp PowerShellツールキット

NetApp PowerShell Toolkit（PSTK）は、NetApp ONTAPをエンドツーエンドで自動化し、ストレージ管理を可能にするPowerShellモジュールです。ONTAPモジュールには2、000を超えるコマンドレットが含まれており、FAS、NetApp All Flash FAS（AFF）、コモディティハードウェア、クラウドリソースの管理に役立ちます。

覚えておくべきこと

- NetAppでは、Windows Serverストレージスペースはサポートされていません。ストレージスペースはJBOD（単なるディスクの束）にのみ使用され、どのタイプのRAID（直接接続ストレージ[DAS]またはSAN）でも機能しません。
- Windows Serverのクラスタ化されたストレージプールは、ONTAPではサポートされていません。
- NetAppは、Windows SAN環境でのゲストクラスタリング用に共有仮想ハードディスクフォーマット（VHDX）をサポートしています。
- Windows Serverでは、iSCSI LUNまたはFC LUNを使用したストレージプールの作成はサポートされていません。

さらに読みます

- NetApp PowerShell Toolkitの詳細については、 "NetApp Support Site"。
- NetApp PowerShell Toolkitのベストプラクティスについては、を参照してください。 "TR-4475：『NetApp PowerShell Toolkit Best Practices Guide』"。

ネットワークのベストプラクティス

イーサネットネットワークは、次のグループに大きく分けることができます。

- VMのクライアントネットワーク
- 1つ以上のストレージネットワーク（ストレージシステムに接続するiSCSIまたはSMB）
- クラスタ通信ネットワーク（クラスタのノード間のハートビートおよびその他の通信）
- 管理ネットワーク（システムの監視とトラブルシューティング用）
- 移行ネットワーク（ホストのライブマイグレーション用）
- VMレプリケーション（Hyper-Vレプリカ）

ベストプラクティス

- NetAppでは、ネットワークの分離とパフォーマンスを確保するために、上記の機能ごとに専用の物理ポートを用意することを推奨しています。
- 上記のネットワーク要件（ストレージ要件を除く）ごとに、複数の物理ネットワークポートを集約して負

荷を分散したり、フォールトトレランスを実現できます。

- NetAppでは、VM内のゲストストレージ接続用に、Hyper-Vホスト上に専用の仮想スイッチを作成することを推奨しています。
- Hyper-VホストとゲストiSCSIのデータパスで別 々 の物理ポートと仮想スイッチを使用して、ゲストとホスト間のセキュアな分離を実現します。
- NetAppでは、iSCSI NICのNICチーミングを避けることを推奨しています。
- NetAppでは、ストレージ用にホストに設定されたONTAP Multipath Input/Output（MPIO；マルチパス入出力）を使用することを推奨しています。
- ゲストiSCSIイニシエータを使用する場合は、ゲストVM内でMPIOを使用することを推奨しますNetApp。パススルーディスクを使用する場合は、ゲスト内でMPIOの使用を避ける必要があります。この場合、ホストにMPIOをインストールすれば十分です。
- NetAppでは、ストレージネットワークに割り当てられた仮想スイッチにQoSポリシーを適用しないことを推奨しています。
- NetAppでは、物理NICで自動プライベートIPアドレッシング（APIPA）を使用しないことを推奨しています。これは、APIPAがルーティングされず、DNSに登録されていないためです。
- NetAppでは、CSV、iSCSI、ライブマイグレーションの各ネットワークでジャンボフレームを有効にして、スループットを向上させ、CPUサイクルを短縮することを推奨しています。
- NetAppでは、Hyper-V仮想スイッチ用に管理オペレーティングシステムがこのネットワークアダプタを共有できるようにするオプションをオフにして、VM専用のネットワークを作成することを推奨しています。
- NetAppでは、ライブマイグレーション用に冗長なネットワークパス（複数のスイッチ）を作成し、耐障害性とQoSを確保することを推奨しています。

SAN環境でのプロビジョニング

ONTAP SVMは、ブロックプロトコルiSCSIおよびFCをサポートしています。ブロックプロトコルiSCSIまたはFCを使用してSVMを作成すると、SVMにはiSCSI Qualified Name（IQN）またはFC Worldwide Name（WWN）がそれぞれ取得されます。この識別子は、NetAppブロックストレージにアクセスするホストにSCSIターゲットを提供します。

Windows ServerでのNetApp LUNのプロビジョニング

前提条件

Windows ServerのSAN環境でNetAppストレージを使用するには、次の要件があります。

- NetAppクラスタには、1つ以上のNetAppストレージコントローラが設定されています。
- NetAppクラスタまたはストレージコントローラに有効なiSCSIライセンスがある。
- iSCSIポートまたはFCポートが設定されていることを確認します。
- FCゾーニングはFCスイッチでFC用に実行されます。
- アグリゲートが少なくとも1つ作成されている。
- SVMには、iSCSIまたはファイバチャネルを使用してデータを提供するすべてのストレージコントローラ上のイーサネットネットワークまたはファイバチャネルファブリックごとに1つのLIFが必要です。

導入

1. ブロックプロトコルiSCSIまたはFCを有効にして、新しいSVMを作成します。新しいSVMは次のいずれかの方法で作成できます。
   - NetAppストレージのCLIコマンド
   - ONTAP システムマネージャ
   - NetApp PowerShellツールキット
2. iSCSIプロトコル/ FCプロトコルを設定
3. SVMに各クラスタノードのLIFを割り当てます。
4. SVMでiSCSIサービス/ FCサービスを開始します。

   。
5. SVM LIFを使用して、iSCSIポートセットやFCポートセットを作成します。
6. 作成したポートセットを使用して、Windows用のiSCSIイニシエータ/ FCイニシエータグループを作成します。
7. イニシエータグループにイニシエータを追加します。イニシエータは、iSCSIのIQNとFCのWWPNです。Windows Serverからクエリを実行するには、PowerShellコマンドレットGet-InitiatorPortを実行します。

```
# Get the IQN for iSCSI
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'iSCSI'} | Select-
Object -Property NodeAddress
```

```
# Get the WWPN for FC
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'Fibre Channel'} |
Select-Object -Property PortAddress
```

```
# While adding initiator to the initiator group in case of FC, make sure
to provide the initiator(PortAddress) in the standard WWPN format
```

Windows Server上のiSCSIのIQNは、iSCSIイニシエータプロパティの構成でも確認できます。

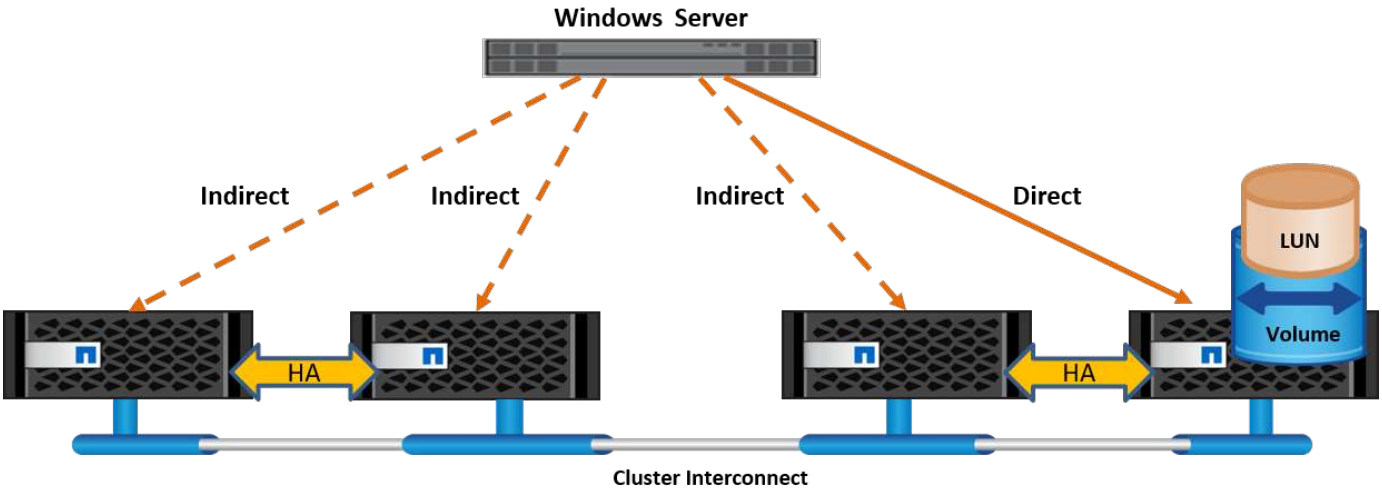
- LUN作成ウィザードを使用してLUNを作成し、作成したイニシエータグループに関連付けます。

ホスト統合

Windows Serverでは、Asymmetrical Logical Unit Access（ALUA；非対称論理ユニットアクセス）拡張MPIOを使用して、LUNへの直接パスと間接パスを決定します。SVMが所有するすべてのLIFはLUNの読み取り/書き込み要求を受け入れますが、そのLUNの元となるディスクを実際に所有しているクラスタノードは常に1つだけです。これにより、次の図に示すように、LUNへの使用可能なパスが直接パスと間接パスの2種類に分けられます。

LUNの直接パスとは、SVMのLIFとアクセス対象のLUNが同じノードにあるパスです。物理ターゲットポートからディスクに移動する場合、クラスタネットワークを経由する必要はありません。

間接パスは、SVMのLIFとアクセス対象のLUNが別 々 のノードにあるデータパスです。物理ターゲットポートからディスクに移動するには、データがクラスタネットワークを経由する必要があります。



MPIO

NetApp ONTAPは、ストレージコントローラからWindows Serverへの複数のパスが存在できる高可用性ストレージを提供します。マルチパスは、サーバからストレージアレイへの複数のデータパスを確立する機能です。マルチパスは、ハードウェア障害（ケーブルの切断、スイッチおよびHost Bus Adapter（HBA；ホストバスアダプタ）の障害など）から保護します。また、複数の接続を集約したパフォーマンスを使用することで、より高いパフォーマンス制限を実現できます。一方のパスまたは接続が使用できなくなると、マルチパスソフトウェアは自動的に他の使用可能なパスのいずれかに負荷を移します。MPIO機能は、ストレージへの複数の物理パスをデータアクセスに使用する単一の論理パスとして組み合わせて、ストレージの耐障害性と負荷分散を実現します。この機能を使用するには、Windows ServerでMPIO機能を有効にする必要があります。

MPIOを有効にする

Windows ServerでMPIOを有効にするには、次の手順を実行します。

1. 管理者グループのメンバーとしてWindows Serverにログインします。
2. Server Managerを起動します。
3. [管理]セクションで、[ロールと機能の追加]をクリックします。
4. [Select Features]ページで、[Multipath I/O]を選択します。

MPIOの設定

iSCSIプロトコルを使用する場合は、MPIOプロパティでiSCSIデバイスにマルチパスサポートを適用するようにWindows Serverに指示する必要があります。

Windows ServerでMPIOを設定するには、次の手順を実行します。

1. 管理者グループのメンバーとしてWindows Serverにログオンします。
2. Server Managerを起動します。

3. [ツール]セクションで、[MPIO]をクリックします。
4. [Discover Multi-Paths]の[MPIO Properties]で、[Add Support for iSCSI Devices]を選択し、[Add]をクリックします。コンピュータの再起動を求めるプロンプトが表示されます。
5. Windows Serverをリブートして、[MPIOのプロパティ]の[MPIOデバイス]セクションにMPIOデバイスが表示されることを確認します。

iSCSIを設定

Windows ServerでiSCSIブロックストレージを検出するには、次の手順を実行します。

1. 管理者グループのメンバーとしてWindows Serverにログオンします。
2. Server Managerを起動します。
3. [Tools]セクションで、[iSCSI Initiator]をクリックします。
4. [Discovery]タブで、[Discover Portal]をクリックします。
5. SANプロトコル用のNetAppストレージ用に作成したSVMに関連付けられているLIFのIPアドレスを指定します。[詳細設定]をクリックし、[全般]タブで情報を設定して、[OK]をクリックします。
6. iSCSIイニシエータによってiSCSIターゲットが自動的に検出され、[ターゲット]タブに一覧表示されます。
7. [Discovered Targets]でiSCSIターゲットを選択します。[Connect]をクリックして[Connect to Target]ウィンドウを開きます。
8. Windows ServerホストからNetAppストレージクラスタ上のターゲットiSCSI LIFへのセッションを複数作成する必要があります。これには、次の手順を実行します。
9. [Connect to Target]ウィンドウで、[Enable MPIO]を選択し、[Advanced]をクリックします。
10. [詳細設定]の[全般]タブで、ローカルアダプタをMicrosoft iSCSIイニシエータとして選択し、[イニシエータIP]と[ターゲットポータルIP]を選択します。
11. また、2番目のパスを使用して接続する必要があります。そのため、手順5から手順8を繰り返しますが、今回は2番目のパスとして[Initiator IP]と[Target Portal IP]を選択します。
12. [iSCSI Properties]メインウィンドウの[Discovered Targets]でiSCSIターゲットを選択し、[Properties]をクリックします。
13. [プロパティ]ウィンドウに、複数のセッションが検出されたことが表示されます。セッションを選択して[Devices]をクリックし、MPIOをクリックしてロードバランシングポリシーを設定します。デバイスに設定されているすべてのパスが表示され、すべてのロードバランシングポリシーがサポートされます。通常、NetAppではサブセットを使用したラウンドロビンを推奨しています。この設定は、ALUAが有効なアレイのデフォルトです。ラウンドロビンは、ALUAをサポートしないアクティブ/アクティブアレイのデフォルトです。

ブロックストレージを検出

Windows ServerでiSCSIまたはFCブロックストレージを検出するには、次の手順を実行します。

1. サーバーマネージャの[ツール]セクションで[コンピュータの管理]をクリックします。
2. [コンピュータの管理]で、[ストレージのディスクの管理]セクションをクリックし、[その他の操作]と[ディスクの再スキャン]をクリックします。これにより、raw iSCSI LUNが表示されます。
3. 検出されたLUNをクリックしてオンラインにします。次に、MBRまたはGPTパーティションを使用してディスクを初期化を選択します。ボリュームサイズとドライブ文字を指定して新しいシンプルボリュームを

作成し、FAT、FAT32、NTFS、またはResilient File System（ReFS）を使用してフォーマットします。

ベストプラクティス

- NetAppでは、LUNをホストするボリュームでシンプロビジョニングを有効にすることを推奨しています。
- マルチパスの問題を回避するために、NetAppでは、特定のLUNに対するすべての10Gbセッションまたはすべての1Gbセッションのいずれかを使用することを推奨しています。
- NetAppでは、ストレージシステムでALUAが有効になっていることを確認することを推奨しています。ONTAPでは、ALUAがデフォルトで有効になっています。
- NetApp LUNのマッピング先のWindows Serverホストで、ファイアウォールの設定で、インバウンドの場合はiSCSIサービス（TCP-IN）、アウトバウンドの場合はiSCSIサービス（TCP-OUT）を有効にします。これらの設定により、Hyper-VホストおよびNetAppコントローラとの間でiSCSIトラフィックが送受信されます。

Nanoサーバでの**NetApp LUN**のプロビジョニング

前提条件

前のセクションで説明した前提条件に加えて、ストレージロールをNano Server側から有効にする必要があります。たとえば、Nano Serverは-Storageオプションを使用して導入する必要があります。Nano Serverを展開するには、「"Nano Serverを展開します。""

導入

ナノサーバでNetApp LUNをプロビジョニングするには、次の手順を実行します。

1. 「"Nanoサーバーへの接続". 」
2. iSCSIを設定するには、Nano Serverで次のPowerShellコマンドレットを実行します。

```
# Start iSCSI service, if it is not already running
Start-Service msiscsi
```

```
# Create a new iSCSI target portal
New-IscsiTargetPortal â€"TargetPortalAddress <SVM LIF>
```

```
# View the available iSCSI targets and their node address
Get-IscsiTarget
```

```
# Connect to iSCSI target
Connect-IscsiTarget -NodeAddress <NodeAddress>
```

```
# NodeAddress is retrived in above cmdlet Get-IscsiTarget
# OR
Get-IscsiTarget | Connect-IscsiTarget
```

```
# View the established iSCSI session
Get-IscsiSession
```

```
# Note the InitiatorNodeAddress retrieved in the above cmdlet Get-
IscsiSession. This is the IQN for Nano server and this needs to be added
in the Initiator group on NetApp Storage
```

```
# Rescan the disks
Update-HostStorageCache
```

3. イニシエータグループにイニシエータを追加します。

```
Add the InitiatorNodeAddress retrieved from the cmdlet Get-IscsiSession
to the Initiator Group on NetApp Controller
```

4. MPIOを設定します。

```
# Enable MPIO Feature
Enable-WindowsOptionalFeature -Online -FeatureName MultipathIo
```

```
# Get the Network adapters and their IPs
Get-NetIPAddress â€"AddressFamily IPv4 â€"PrefixOrigin <Dhcp or Manual>
```

```
# Create one MPIO-enabled iSCSI connection per network adapter
Connect-IscsiTarget -NodeAddress <NodeAddress> -IsPersistent $True â
€"IsMultipathEnabled $True â€"InitiatorPortalAddress <IP Address of
ethernet adapter>
```

```
# NodeAddress is retrieved from the cmdlet Get-IscsiTarget
# IPs are retrieved in above cmdlet Get-NetIPAddress
```

```
# View the connections
Get-IscsiConnection
```

5. ブロックストレージを検出

```
# Rescan disks
Update-HostStorageCache
```

```
# Get details of disks
Get-Disk
```

```
# Initialize disk
Initialize-Disk -Number <DiskNumber> -PartitionStyle <GPT or MBR>
```

```
# DiskNumber is retrived in the above cmdlet Get-Disk
# Bring the disk online
Set-Disk -Number <DiskNumber> -IsOffline $false
```

```
# Create a volume with maximum size and default drive letter
New-Partition -DiskNumber <DiskNumber> -UseMaximumSize
-AssignDriveLetter
```

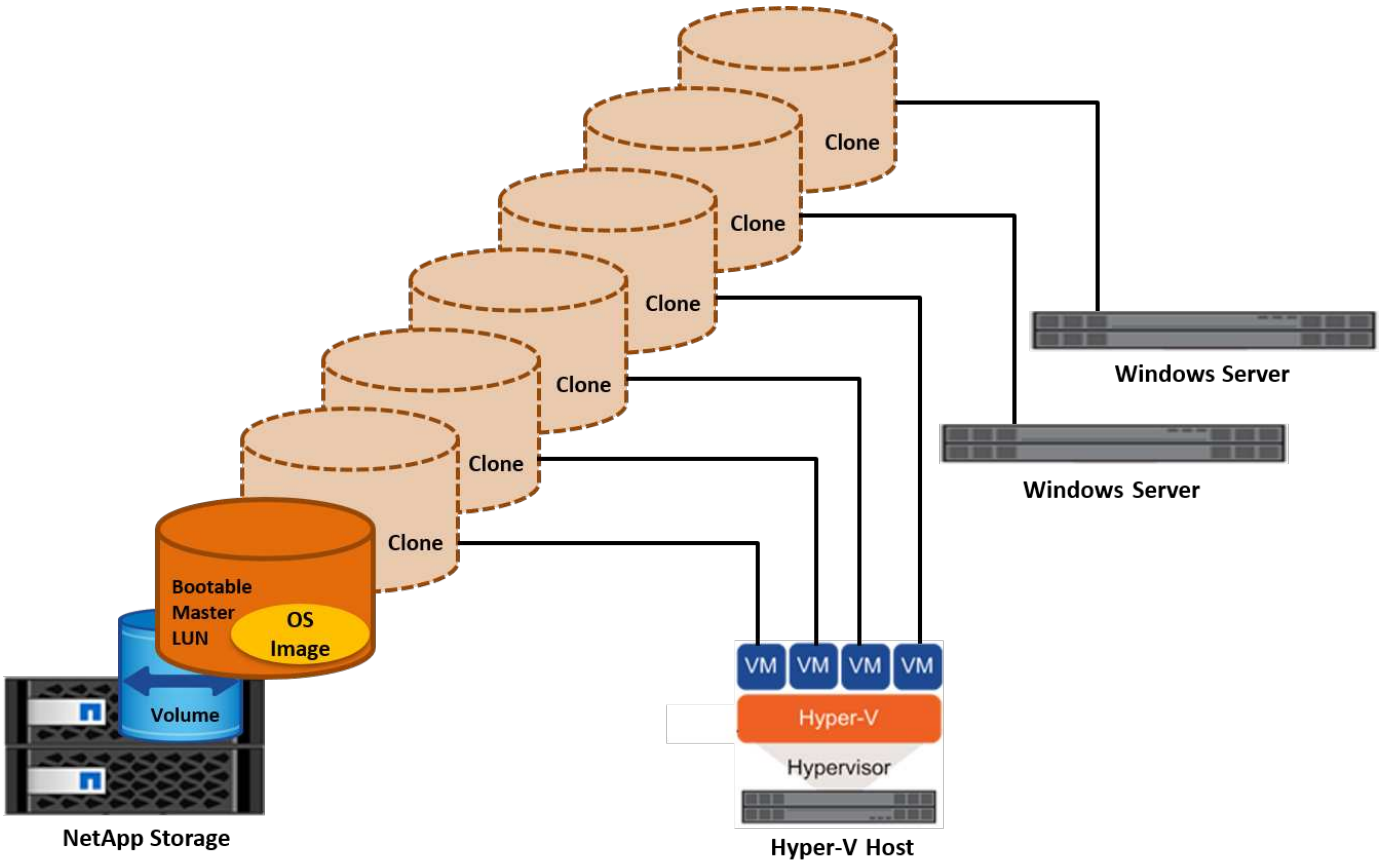
```
# To choose the size and drive letter use -Size and -DriveLetter
parameters
# Format the volume
Format-Volume -DriveLetter <DriveLetter> -FileSystem <FAT32 or NTFS or
REFS>
```

SANからのブート

物理ホスト（サーバ）またはHyper-V VMは、内蔵ハードディスクではなくNetApp LUNから直接Windows Server OSをブートできます。SANブートのアプローチでは、ブート元のOSイメージは、物理ホストまたはVMに接続されたNetApp LUNに格納されます。物理ホストの場合、物理ホストのHBAは、NetApp LUNをブートに使用するように設定されます。VMの場合、NetApp LUNはブート用のパススルーディスクとして接続されます。

NetApp FlexCloneのアプローチ

NetApp FlexCloneテクノロジを使用すると、次の図に示すように、OSイメージを含むブートLUNのクローンを瞬時に作成し、サーバやVMに接続して、クリーンなOSイメージを迅速に提供できます。



物理ホストの**SAN**からのブート

前提条件

- 物理ホスト（サーバ）に適切なiSCSI HBAまたはFC HBAが搭載されている。
- Windows Serverをサポートしているサーバに適したHBAデバイスドライバをダウンロードしておきます。
- サーバーにWindows Server ISOイメージを挿入するのに適したCD/DVDドライブまたは仮想メディアがあり、HBAデバイスドライバがダウンロードされている。
- NetApp iSCSIまたはFC LUNは、NetAppストレージコントローラ上にプロビジョニングされます。

導入

物理ホストに対してSANからのブートを設定するには、次の手順を実行します。

1. サーバHBAでBootBIOSを有効にします
2. iSCSI HBAの場合は、ブートBIOS設定でイニシエータIP、iSCSIノード名、およびアダプタのブートモードを設定します。
3. NetAppストレージコントローラでiSCSIまたはFCのイニシエータグループを作成する場合は、サーバHBAイニシエータをグループに追加します。サーバのHBAイニシエータは、FC HBAのWWPNまたはiSCSI

HBAのiSCSIノード名です。

4. NetAppストレージコントローラにLUN ID 0のLUNを作成し、前の手順で作成したイニシエータグループに関連付けます。このLUNはブートLUNとして機能します。
5. HBAをブートLUNへの単一のパスに制限します。Windows ServerをブートLUNにインストールしたあとにパスを追加して、マルチパス機能を利用できます。
6. HBAのBootBIOSユーティリティを使用して、LUNをブートデバイスとして設定します。
7. ホストをリブートし、ホストBIOSユーティリティを起動します。
8. ブートLUNがブート順序の最初のデバイスになるようにホストBIOSを設定します。
9. Windows Server ISOから、インストールセットアップを起動します。
10. 「Where do you want to install Windows？」というメッセージが表示されたら、インストール画面の下部にある「Load Driver（ドライバのロード）」をクリックして、「Select Driver to Install（インストールするドライバの選択）」ページを起動します。前の手順でダウンロードしたHBAデバイスドライバのパスを入力し、ドライバのインストールを完了します。
11. これで、前の手順で作成したブートLUNがWindowsのインストールページに表示されるようになります。ブートLUNにWindows Serverをインストールするブート LUNを選択し、インストールを完了します。

仮想マシンの**SAN**からのブート

VMに対してSANからのブートを設定するには、次の手順を実行します。

導入

1. NetAppストレージコントローラでiSCSIまたはFCのイニシエータグループを作成する場合は、Hyper-VサーバのIQN（iSCSIの場合）またはWWN（FCの場合）をコントローラに追加します。
2. NetAppストレージコントローラでLUNまたはLUNクローンを作成し、前の手順で作成したイニシエータグループに関連付けます。これらのLUNは、VMのブートLUNとして機能します。
3. Hyper-Vサーバ上のLUNを検出してオンラインにし、初期化します。
4. LUNをオフラインにします。
5. [Connect Virtual Hard Disk]ページで、[Attach a Virtual Hard Disk]オプションを使用してVMを作成します。
6. LUNをVMにパススルーディスクとして追加します。
   a. VM設定を開きます。
   b. [IDE Controller 0]をクリックし、[Hard Drive]を選択して、[Add]をクリックします。[IDE Controller 0]を選択すると、このディスクがVMの最初の起動デバイスになります。
   c. [Hard Disk]オプションで[Physical Hard Disk]を選択し、リストからパススルーディスクとしてディスクを選択します。ディスクは、前の手順で設定したLUNです。
7. パススルーディスクにWindows Serverをインストールします。

ベストプラクティス

- LUNがオフラインであることを確認します。そうしないと、ディスクをVMにパススルーディスクとして追加できません。
- LUNが複数存在する場合は、ディスク管理でLUNのディスク番号をメモしておいてください。VMのリス

トにはディスク番号が記載されているため、この処理は必須です。また、VMのパススルーディスクとしてのディスクの選択は、このディスク番号に基づいて行われます。

- NetAppでは、iSCSI NICのNICチーミングを避けることを推奨しています。
- NetAppでは、ストレージ用にホストに設定されたONTAP MPIOを使用することを推奨しています。

SMB環境でのプロビジョニング

ONTAPは、SMB3プロトコルを使用して、Hyper-V仮想マシン用に耐障害性とパフォーマンスに優れたNASストレージを提供します。

CIFSプロトコルを使用してSVMを作成すると、Windows Active Directoryドメインに属するSVM上でCIFSサーバが実行されます。SMB共有をホームディレクトリに使用したり、Hyper-VおよびSQL Serverのワークロードをホストしたりできます。ONTAPでは、SMB 3.0の次の機能がサポートされます。

- 永続的ハンドル（継続的可用性を備えたファイル共有）
- カンシフロトコル
- クラスタクライアントフェイルオーバー
- スケールアウト対応
- ODX
- リモオトVSS

Windows ServerでのSMB共有のプロビジョニング

前提条件

Windows ServerのNAS環境でNetAppストレージを使用するには、次の要件があります。

- ONTAPクラスタに有効なCIFSライセンスが必要です。
- アグリゲートが少なくとも1つ作成されている。
- データ論理インターフェイス（LIF）が1つ作成され、そのデータLIFをCIFS用に設定する必要があります。
- DNSが設定したWindows Active Directoryドメインサーバとドメイン管理者のクレデンシャルがある。
- NetAppクラスタ内の各ノードは、Windowsドメインコントローラと時刻が同期されます。

Active Directoryドメインコントローラ

NetAppストレージコントローラは、Windowsサーバと同様にActive Directoryに参加して、Active Directory内で動作することができます。SVMの作成時に、ドメイン名とネームサーバの詳細を指定してDNSを設定できます。SVMは、Windows Serverと同様の方法で、DNSにActive Directory / Lightweight Directory Access Protocol（LDAP）サーバを照会することで、Active Directoryドメインコントローラの検索を試みます。

CIFSのセットアップが正しく機能するためには、NetAppストレージコントローラとWindowsドメインコントローラの時刻が同期されている必要があります。NetAppでは、WindowsドメインコントローラとNetAppストレージコントローラの間の時間差を5分以内にすることを推奨しています。ONTAPクラスタを外部の時間ソースと同期するには、ネットワークタイムプロトコル（NTP）サーバを設定することを推奨します。WindowsドメインコントローラをNTPサーバとして設定するには、ONTAPクラスタで次のコマンドを実行します。

```
$domainControllerIP = "<input IP Address of windows domain controller>"
cluster::> system services ntp server create â€"server $domainControllerIP
```

導入

1. 新しいSVMを作成してNASプロトコルCIFSを有効にします。新しいSVMは次のいずれかの方法で作成できます。
   - NetApp ONTAPノCLIコマント
   - System Manager の略
   - NetApp PowerShellツールキット
2. CIFSプロトコルの設定
   a. CIFSサーバ名を指定します。
   b. CIFSサーバを追加するActive Directoryを指定します。CIFSサーバをActive Directoryに追加するには、ドメイン管理者のクレデンシャルが必要です。
3. SVMに各クラスタノードのLIFを割り当てます。
4. SVMでCIFSサービスを開始します。
5. アグリゲートからNTFSセキュリティ形式のボリュームを作成します。
6. ボリュームにqtreeを作成します（オプション）。
7. Windows Serverからアクセスできるように、ボリュームまたはqtreeディレクトリに対応する共有を作成します。共有をHyper-Vストレージに使用する場合は、共有の作成時にHyper-Vの継続的可用性を有効にするを選択します。これにより、ファイル共有の高可用性が実現します。
8. 作成した共有を編集し、共有へのアクセスに必要に応じて権限を変更します。SMB共有にアクセスするすべてのサーバのコンピュータアカウントにアクセスを許可するように、SMB共有の権限を設定する必要があります。

ホスト統合

NASプロトコルCIFSは、ONTAPに標準で統合されています。したがって、Windows Serverは、NetApp ONTAP上のデータにアクセスするために追加のクライアントソフトウェアを必要としません。NetAppストレージコントローラは、ネットワーク上でネイティブファイルサーバとして認識され、Microsoft Active Directory認証をサポートします。

Windows Serverで作成したCIFS共有を検出するには、次の手順を実行します。

1. 管理者グループのメンバーとしてWindows Serverにログインします。
2. run.exeに移動し、共有にアクセスするために作成したCIFS共有の完全パスを入力します。
3. 共有をWindows Serverに永続的にマッピングするには、[This PC]を右クリックし、[Map Network Drive]をクリックして、CIFS共有のパスを指定します。
4. 一部のCIFS管理タスクは、Microsoft管理コンソール（MMC）を使用して実行できます。これらのタスクを実行する前に、MMCメニューコマンドを使用してMMCをNetApp ONTAPストレージに接続する必要があります。
   a. Windows ServerでMMCを開くには、サーバーマネージャの[ツール]セクションで[コンピュータの管理]をクリックします。

b. [その他の操作]をクリックして[別のコンピュータに接続]をクリックすると、[コンピュータの選択]ダイアログが開きます。

c. CIFSサーバの名前またはCIFSサーバに接続するSVM LIFのIPアドレスを入力します。

d. [システムツール]と[共有フォルダ]を展開して、開いているファイル、セッション、および共有を表示および管理します。

ベストプラクティス

- NetAppでは、ボリュームがあるノードから別のノードに移動されたときやノードで障害が発生したときにダウンタイムが発生しないことを確認するために、ファイル共有でcontinuous availabilityオプションを有効にすることを推奨しています。
- Hyper-V over SMB環境用にVMをプロビジョニングする場合はNetApp、ストレージシステムでコピーオフロードを有効にすることを推奨します。これにより、VMのプロビジョニング時間が短縮されます。
- ストレージクラスタでSQL Server、Hyper-V、CIFSサーバなどの複数のSMBワークロードをホストするNetApp場合は、別 々 のアグリゲートにある別 々 のSVMで異なるSMBワークロードをホストすることを推奨します。この構成は、各ワークロードに固有のストレージネットワークとボリュームレイアウトが必要になるため、有益です。
- NetAppでは、Hyper-VホストとNetApp ONTAPストレージを10GBのネットワーク（使用可能な場合）で接続することを推奨しています。1GBのネットワーク接続の場合、NetAppでは、複数の1GBポートで構成されるインターフェイスグループを作成することを推奨します。
- NetAppでは、あるSMB 3.0共有から別の共有にVMを移行する際に、移行時間を短縮するために、ストレージシステムでCIFSコピーオフロード機能を有効にすることを推奨しています。

覚えておくべきこと

- SMB環境用のボリュームをプロビジョニングする場合は、ボリュームをNTFSセキュリティ形式で作成する必要があります。
- クラスタ内のノードの時間設定は、それに応じて設定する必要があります。NetApp CIFSサーバがWindows Active Directoryドメインに参加している必要がある場合は、NTPを使用します。
- 永続的ハンドルは、HAペアのノード間でのみ機能します。
- 監視プロトコルは、HAペアのノード間でのみ機能します。
- 継続的可用性を備えたファイル共有は、Hyper-VおよびSQL Serverワークロードでのみサポートされます。
- SMBマルチチャネルはONTAP 9.4以降でサポートされます。
- RDMAはサポートされません。
- Refsはサポートされていません。

Nanoサーバーでの**SMB**共有のプロビジョニング

Nano Serverでは、NetAppストレージコントローラ上のCIFS共有上のデータにアクセスするために、追加のクライアントソフトウェアは必要ありません。

Nano ServerからCIFS共有にファイルをコピーするには、リモートサーバで次のコマンドレットを実行します。

```
$ip = "<input IP Address of the Nano Server>"
```

```
# Create a New PS Session to the Nano Server
$session = New-PSSession -ComputerName $ip -Credential ~\Administrator
```

```
 Copy-Item -FromSession $s -Path C:\Windows\Logs\DISM\dism.log
-Destination \\cifsshare
```

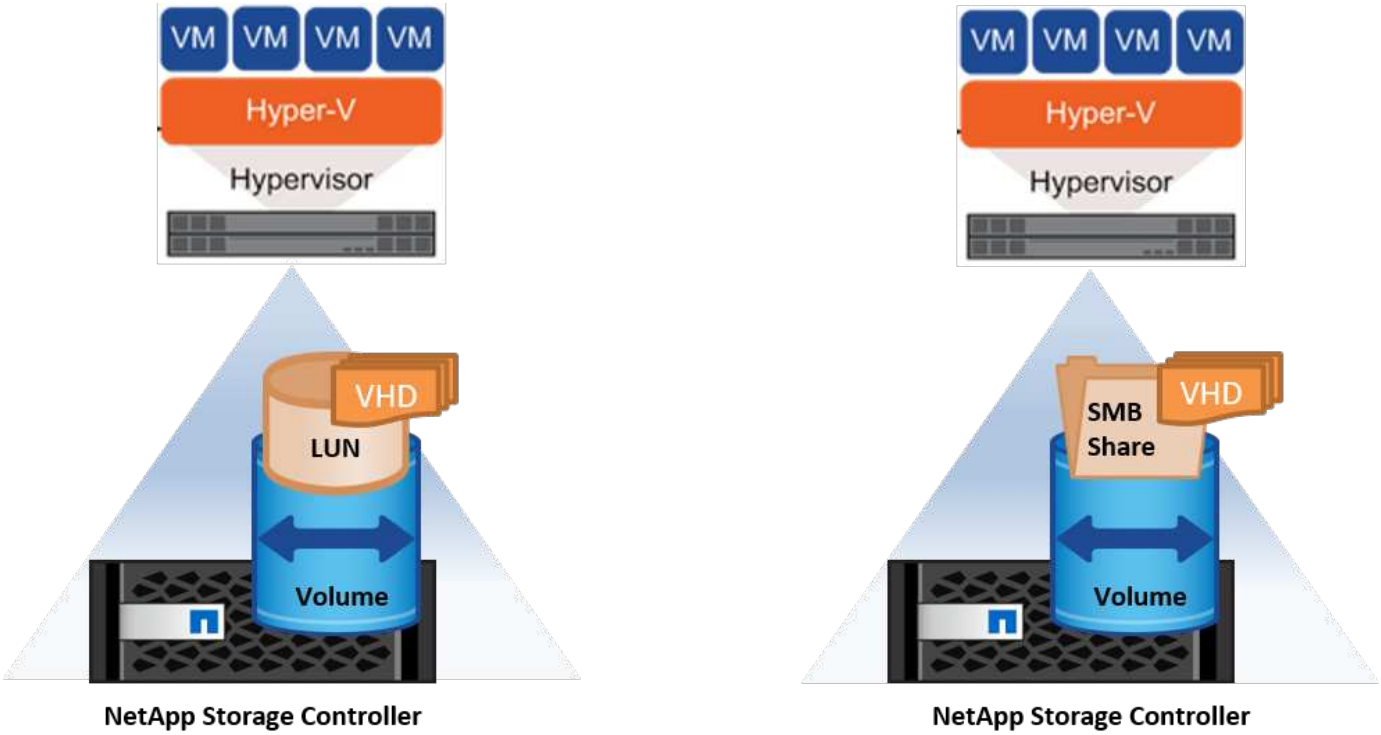
* `cifsshare` は、NetAppストレージコントローラ上のCIFS共有です。
* Nano Serverにファイルをコピーするには、次のコマンドレットを実行します。

+
Copy-Item -ToSession $s -Path \\cifsshare\<file> -Destination C:\

フォルダの内容全体をコピーするには、フォルダ名を指定し、コマンドレットの末尾にある-Recurseパラメータを使用します。

NetApp上のHyper-Vストレージインフラ

Hyper-Vストレージインフラは、ONTAPストレージシステムでホストできます。Hyper-VでVMファイルとそのディスクを格納するためのストレージは、次の図に示すように、NetApp LUNまたはNetApp CIFS共有を使用して提供できます。



NetApp LUN上のHyper-Vストレージ

- Hyper-VサーバマシンでNetApp LUNをプロビジョニングします。詳細については、「"SAN環境でのプロビジョニング". 」
- [Server Manager]の[Tools]セクションから[Hyper-V Manager]を開きます。
- Hyper-Vサーバを選択し、[Hyper-V Settings]をクリックします。
- VMとそのディスクをLUNとして格納するデフォルトのフォルダを指定します。これにより、Hyper-Vストレージのデフォルトパスが LUNとして設定されます。VMのパスを明示的に指定する場合は、VMの作成時に指定できます。

NetApp CIFS上のHyper-Vストレージ

このセクションに記載されている手順を開始する前に、"SMB環境でのプロビジョニング"」。 NetApp CIFS共有でHyper-Vストレージを設定するには、次の手順を実行します。

1. [Server Manager]の[Tools]セクションから[Hyper-V Manager]を開きます。
2. Hyper-Vサーバを選択し、[Hyper-V Settings]をクリックします。
3. VMとそのディスクをCIFS共有として格納するデフォルトのフォルダを指定します。これにより、Hyper-VストレージのCIFS共有としてデフォルトパスが設定されます。VMのパスを明示的に指定する場合は、VMの作成時に指定できます。

Hyper-Vの各VMには、物理ホストに提供されたNetApp LUNとCIFS共有を提供できます。この手順は、任意の物理ホストの場合と同じです。VMにストレージをプロビジョニングするには、次の方法を使用します。

- VM内のFCイニシエータを使用したストレージLUNの追加
- VM内のiSCSIイニシエータを使用したストレージLUNの追加
- VMへのパススルー物理ディスクの追加
- ホストからVMへのVHD / VHDXの追加

ベストプラクティス

- VMとそのデータがNetAppストレージに格納される場合、NetAppでは、NetApp重複排除をボリュームレベルで定期的に実行することを推奨しています。これにより、同一のVMがCSV共有またはSMB共有でホストされている場合、スペースが大幅に削減されます。重複排除はストレージコントローラ上で実行され、ホストシステムとVMのパフォーマンスには影響しません。
- Hyper-VでiSCSI LUNを使用する場合は、 `iSCSI Service (TCP-In) for Inbound` および `iSCSI Service (TCP-Out) for Outbound` Hyper-Vホストのファイアウォール設定。これにより、Hyper-VホストとNetAppコントローラとの間でiSCSIトラフィックが送受信されます。
- NetAppでは、[Allow Management Operating System to Share this Network Adapter for the Hyper-V virtual switch]オプションをオフにすることを推奨しています。これにより、VM専用のネットワークが作成されます。

覚えておくべきこと

- 仮想ファイバチャネルを使用してVMをプロビジョニングするには、NポートID仮想化が有効なFC HBAが必要です。最大4つのFCポートがサポートされます。
- ホストシステムに複数のFCポートが設定されており、VMに提供されている場合は、マルチパスを有効に

するためにMPIOをVMにインストールする必要があります。

- パススルーディスクではMPIOがサポートされないため、そのホストでMPIOが使用されている場合、パススルーディスクをホストにプロビジョニングすることはできません。
- VHD / VHDxファイルに使用するディスクは、割り当てに64Kのフォーマットを使用する必要があります。

さらに読みます

- FC HBAの詳細については、を参照してください。 "NetApp Interoperability Matrix を参照してください"。
- 仮想ファイバチャネルの詳細については、Microsoftの "Hyper-V仮想ファイバチャネルの概要" ページ

オフロードデータ転送

Microsoft ODX（コピーオフロード）を使用すると、ホストコンピュータを介さずに、ストレージデバイス内または互換性があるストレージデバイス間でデータを直接転送できます。NetApp ONTAPは、CIFSプロトコルとSANプロトコルの両方でODX機能をサポートしています。ODXを使用すると、コピーが同じボリューム内にある場合にパフォーマンスが向上したり、クライアントでのCPUとメモリの使用率が低下したり、ネットワークI/O帯域幅の使用率が低下したりする可能性があります。

ODXを使用すると、SMB共有内、LUN内、およびSMB共有とLUN（同じボリューム内の場合）間でファイルをコピーする処理が高速かつ効率的になります。この方法は、OS（VHD / VHDX）のゴールデンイメージの複数のコピーが同じボリューム内に必要な場合に役立ちます。コピーが同じボリューム内にある場合、同じゴールデンイメージの複数のコピーを作成する時間が大幅に短縮されます。ODXは、VMストレージを移動するためのHyper-Vストレージのライブマイグレーションでも適用されます。

複数のボリューム間でコピーを行う場合は、ホストベースのコピーに比べてパフォーマンスが大幅に向上することはありません。

CIFSでODX機能を有効にするには、NetAppストレージコントローラで次のCLIコマンドを実行します。

1. CIFS用のODXを有効にします。
   #権限レベルをdiagnosticに設定する
   cluster：：> set -privilege diagnostic

```
#enable the odx feature
cluster::> vserver cifs options modify -vserver <vserver_name> -copy
-offload-enabled true
```

```
#return to admin privilege level
cluster::> set privilege admin
```

2. SANでODX機能を有効にするには、NetAppストレージコントローラで次のCLIコマンドを実行します。
   #権限レベルをdiagnosticに設定する
   cluster：：> set -privilege diagnostic

```
#enable the odx feature
cluster::> copy-offload modify -vserver <vserver_name> -scsi enabled
```

```
#return to admin privilege level
cluster::> set privilege admin
```

覚えておくべきこと

- CIFSの場合、ODXを使用できるのは、クライアントとストレージサーバの両方でSMB 3.0およびODX機能がサポートされている場合だけです。
- SAN環境でODXを使用できるのは、クライアントとストレージサーバの両方でODX機能がサポートされている場合のみです。

さらに読みます

ODXの詳細については、を参照してください。 "Microsoftリモートコピーのパフォーマンスの向上" および "Microsoftオフロードデータ転送" 。

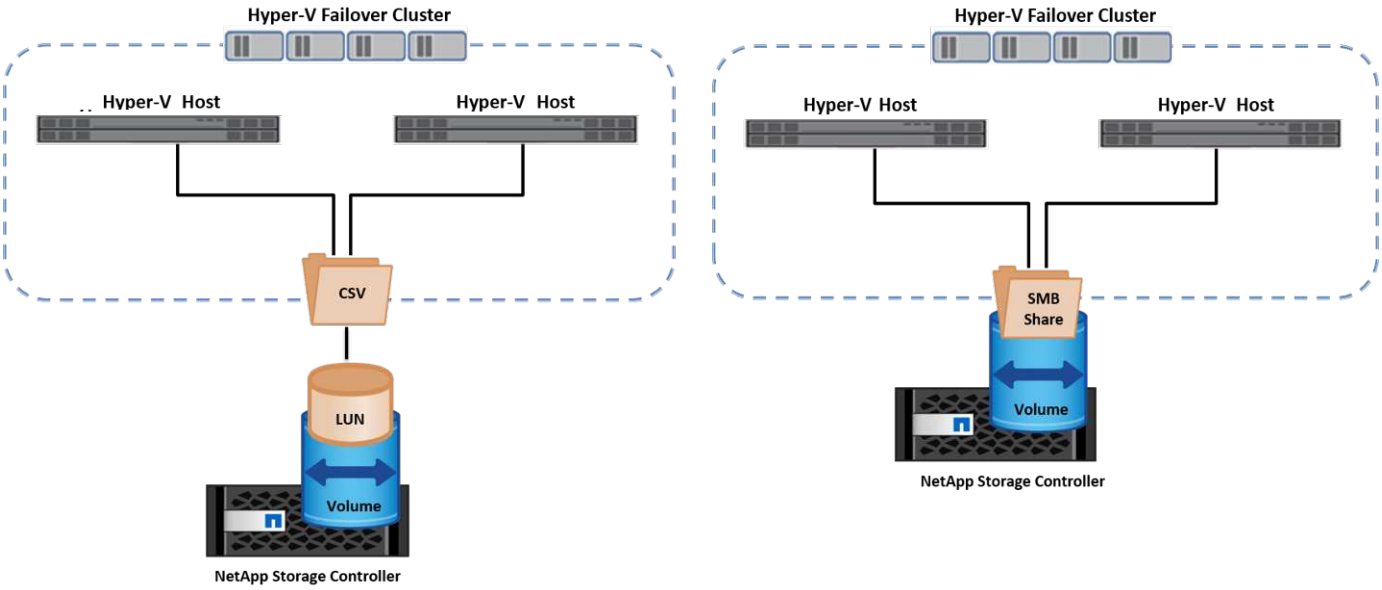
Hyper-Vクラスタリング：仮想マシンの高可用性と拡張性

フェイルオーバークラスタは、Hyper-Vサーバに対して高可用性と拡張性を提供します。フェイルオーバークラスタは、VMの可用性と拡張性を高めるために連携する独立したHyper-Vサーバのグループです。

Hyper-Vクラスタサーバ（ノード）は、物理ネットワークとクラスタソフトウェアによって接続されます。これらのノードは共有ストレージを使用して、構成、仮想ハードディスク（VHD）ファイル、Snapshotコピーなどの VMファイルを格納します。共有ストレージには、図6に示すように、NetApp SMB/CIFS共有またはNetApp LUN上のCSVを使用できます。この共有ストレージは、一貫性のある分散されたネームスペースを提供し、クラスタ内のすべてのノードから同時にアクセスできます。したがって、クラスタ内の1つのノードに障害が発生すると、もう一方のノードがフェイルオーバーと呼ばれるプロセスによってサービスを提供します。フェイルオーバークラスタは、フェイルオーバークラスタマネージャスナップインおよびフェイルオーバークラスタリングWindows PowerShellコマンドレットを使用して管理できます。

クラスタ共有ボリューム

CSVを使用すると、NTFSまたはReFSボリュームとしてプロビジョニングされた同じNetApp LUNへの読み取り/書き込みアクセスを、フェイルオーバークラスタ内の複数のノードで同時に実行できます。CSVを使用すると、クラスタ化されたロールは、ドライブ所有権を変更したり、ボリュームをディスマウントおよび再マウントしたりすることなく、ノード間で迅速にフェイルオーバーできます。CSVを使用すると、フェイルオーバークラスタ内の多数のLUNを簡単に管理できます。CSVは、NTFSまたはReFS上に階層化された汎用クラスタファイルシステムを提供します。



ベストプラクティス

- NetAppでは、内部クラスタ通信とCSVトラフィックが同じネットワークを経由しないように、iSCSIネットワークでクラスタ通信をオフにすることを推奨しています。
- NetAppでは、耐障害性とQoSを確保するために冗長なネットワークパス（複数のスイッチ）を使用することを推奨しています

覚えておくべきこと

- CSVに使用するディスクは、NTFSまたはReFSでパーティショニングする必要があります。FATまたはFAT32でフォーマットされたディスクはCSVに使用できません。
- CSVに使用するディスクの割り当てには64Kのフォーマットを使用する必要があります。

さらに読みます

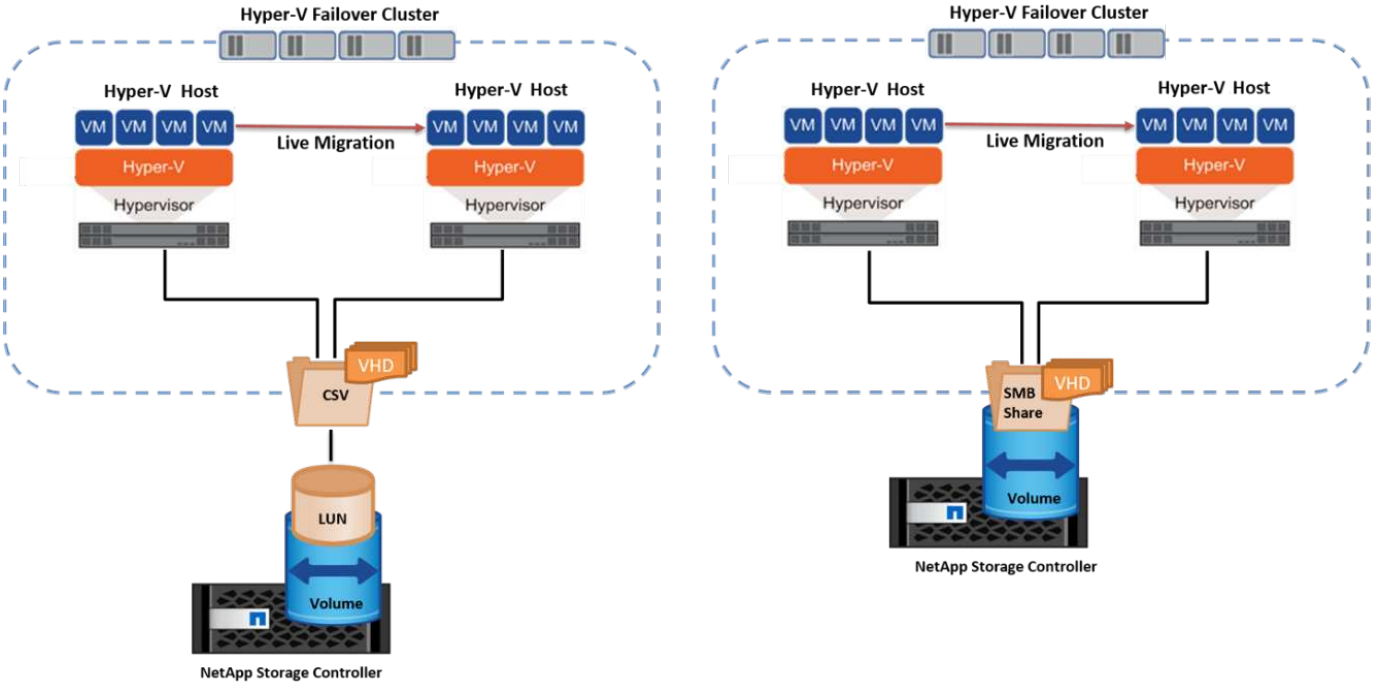
Hyper-Vクラスタの導入については、「付録B： "Hyper-Vクラスタの導入"。

Hyper-Vライブマイグレーション：VMの移行

VMの有効期間中に、Windowsクラスタ上の別のホストにVMを移動しなければならない場合があります。この処理は、ホストのシステムリソースが不足している場合や、メンテナンスのためにホストのリブートが必要な場合に必要になることがあります。同様に、VMを別のLUNまたはSMB共有に移動しなければならない場合があります。これは、現在のLUNまたは共有でスペースが不足しているか、パフォーマンスが想定よりも低い場合に必要になることがあります。Hyper-Vライブマイグレーションでは、実行中のVMを物理Hyper-Vサーバ間で移動します。VMの可用性には影響しません。フェイルオーバークラスタの一部であるHyper-Vサーバ間、またはどのクラスタにも属さない独立したHyper-Vサーバ間で、VMをライブマイグレーションできます。

クラスタ環境でのライブマイグレーション

VMは、クラスタのノード間でシームレスに移動できます。クラスタ内のすべてのノードが同じストレージを共有し、VMとそのディスクにアクセスできるため、VMの移行は瞬時に完了します。次の図に、クラスタ環境でのライブマイグレーションを示します。



ベストプラクティス

- ライブマイグレーショントラフィック専用のポートを用意します。
- 移行中のネットワーク関連の問題を回避するために、専用のホストライブマイグレーションネットワークを用意します。

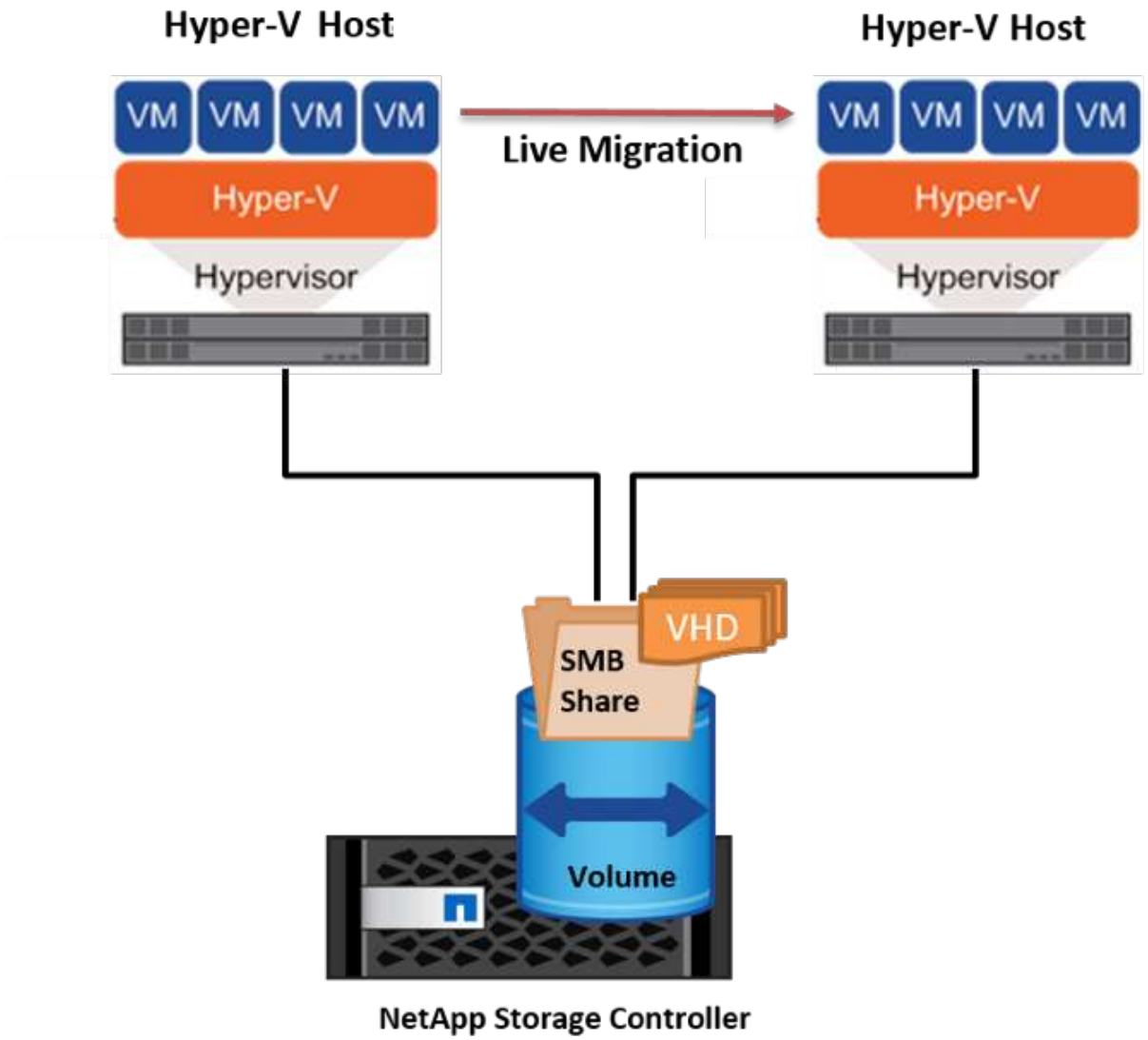
さらに読みます

クラスタ環境へのライブマイグレーションの導入については、を参照してください。 "付録C：クラスタ環境へのHyper-Vライブマイグレーションの導入"。

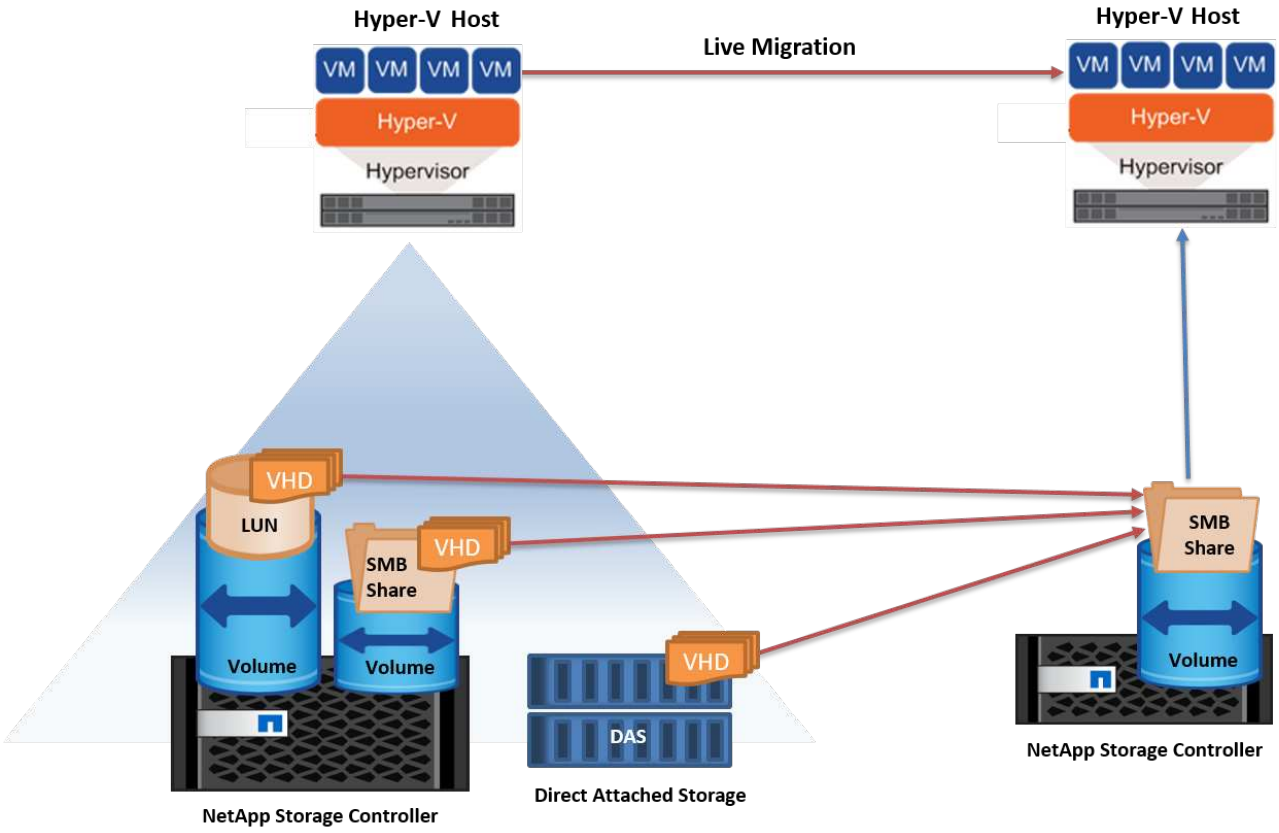
クラスタ環境外でのライブマイグレーション

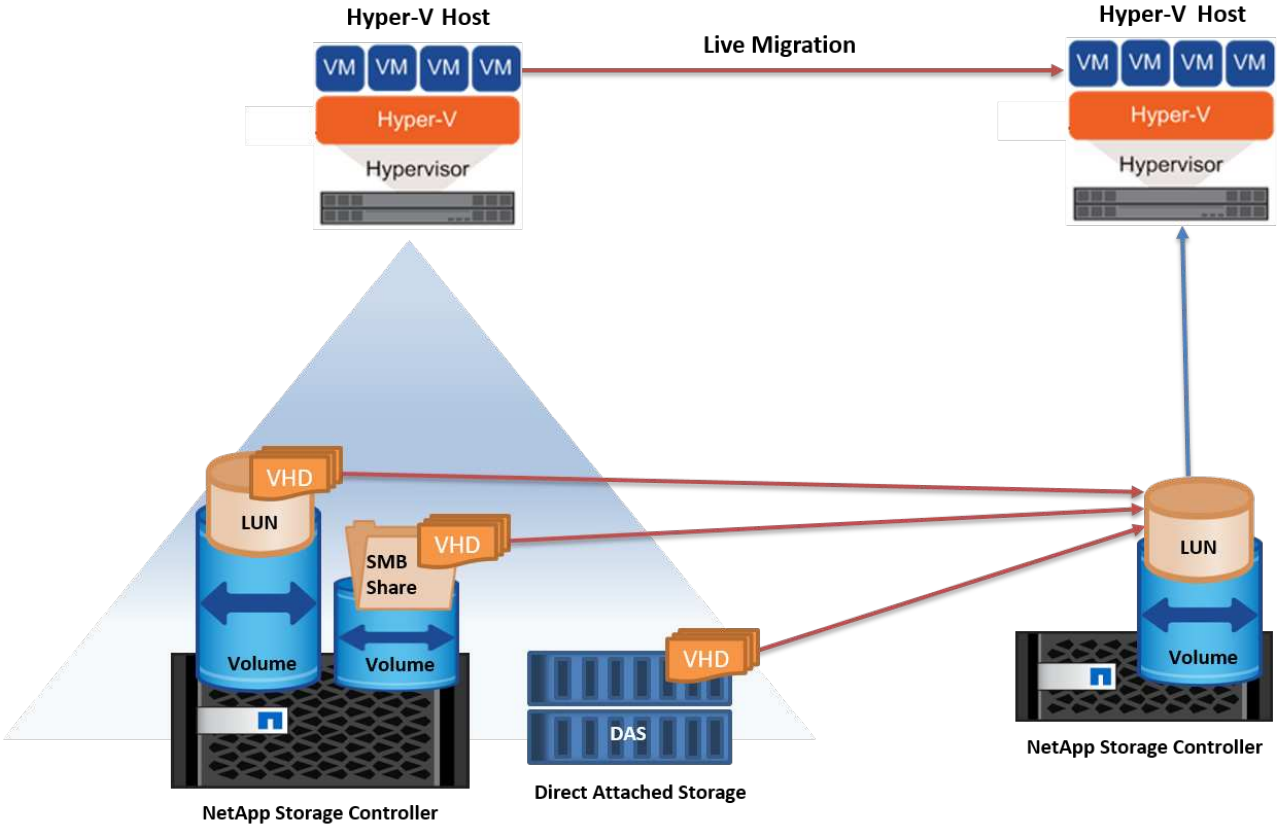
VMは、クラスタ化されておらず、独立した2台のHyper-Vサーバ間でライブマイグレーションできます。このプロセスでは、シェアードナッシングまたはシェアードナッシングライブマイグレーションを使用できます。

- 共有ライブマイグレーションでは、VMはSMB共有に格納されます。したがって、VMをライブマイグレーションする場合、次の図に示すように、VMのストレージは中央のSMB共有に残り、もう一方のノードから即座にアクセスできます。



- シェアードナッシングライブマイグレーションでは、各Hyper-Vサーバに独自のローカルストレージ（SMB共有、LUN、DAS）があり、VMのストレージはHyper-Vサーバに対してローカルになります。VMをライブマイグレーションすると、VMのストレージがクライアントネットワーク経由でデスティネーションサーバにミラーリングされ、その後VMが移行されます。DAS、LUN、またはSMB / CIFS共有に格納されているVMは、次の図に示すように、もう一方のHyper-Vサーバ上のSMB / CIFS共有に移動できます。2番目の図に示すように、LUNに移動することもできます。





さらに読みます

クラスタ環境外へのライブマイグレーションの導入については、を参照してください。 "付録D：クラスタ環

境以外にHyper-Vライブマイグレーションを導入する"。

Hyper-Vストレージのライブマイグレーション

VMの有効期間中に、VMストレージ（VHD / VHDX）を別のLUNまたはSMB共有に移動しなければならない場合があります。これは、現在のLUNまたは共有でスペースが不足しているか、パフォーマンスが想定よりも低い場合に必要になることがあります。

VMを現在ホストしているLUNまたは共有は、スペース不足、転用、またはパフォーマンスの低下を招く可能性があります。このような状況では、ダウンタイムを発生させずに、別のボリューム、アグリゲート、またはクラスタ上の別のLUNや共有にVMを移動できます。ストレージシステムにコピーオフロード機能がある場合は、この処理の方が高速です。NetAppストレージシステムは、CIFSおよびSAN環境ではデフォルトでコピーオフロードが有効になります。

ODX機能は、リモートサーバ上にある2つのディレクトリ間でファイル全体またはサブファイルのコピーを実行します。コピーは、サーバ間（ソースファイルとデスティネーションファイルが同じサーバ上にある場合は同じサーバ）でデータをコピーすることによって作成されます。コピーは、クライアントがソースからデータを読み取ったり、デスティネーションに書き込んだりすることなく作成されます。このプロセスにより、クライアントまたはサーバのプロセッサとメモリの使用量が削減され、ネットワークI/O帯域幅が最小限に抑えられます。同じボリューム内にある場合は、より高速にコピーできます。複数のボリューム間でコピーを行う場合は、ホストベースのコピーに比べてパフォーマンスが大幅に向上することはありません。ホストでコピー処理を開始する前に、ストレージシステムにコピーオフロードが設定されていることを確認してください。

VMストレージのライブマイグレーションをホストから開始すると、ソースとデスティネーションが特定され、コピーアクティビティがストレージシステムにオフロードされます。このアクティビティはストレージシステムによって実行されるため、ホストのCPU、メモリ、またはネットワークの使用量はごくわずかです。

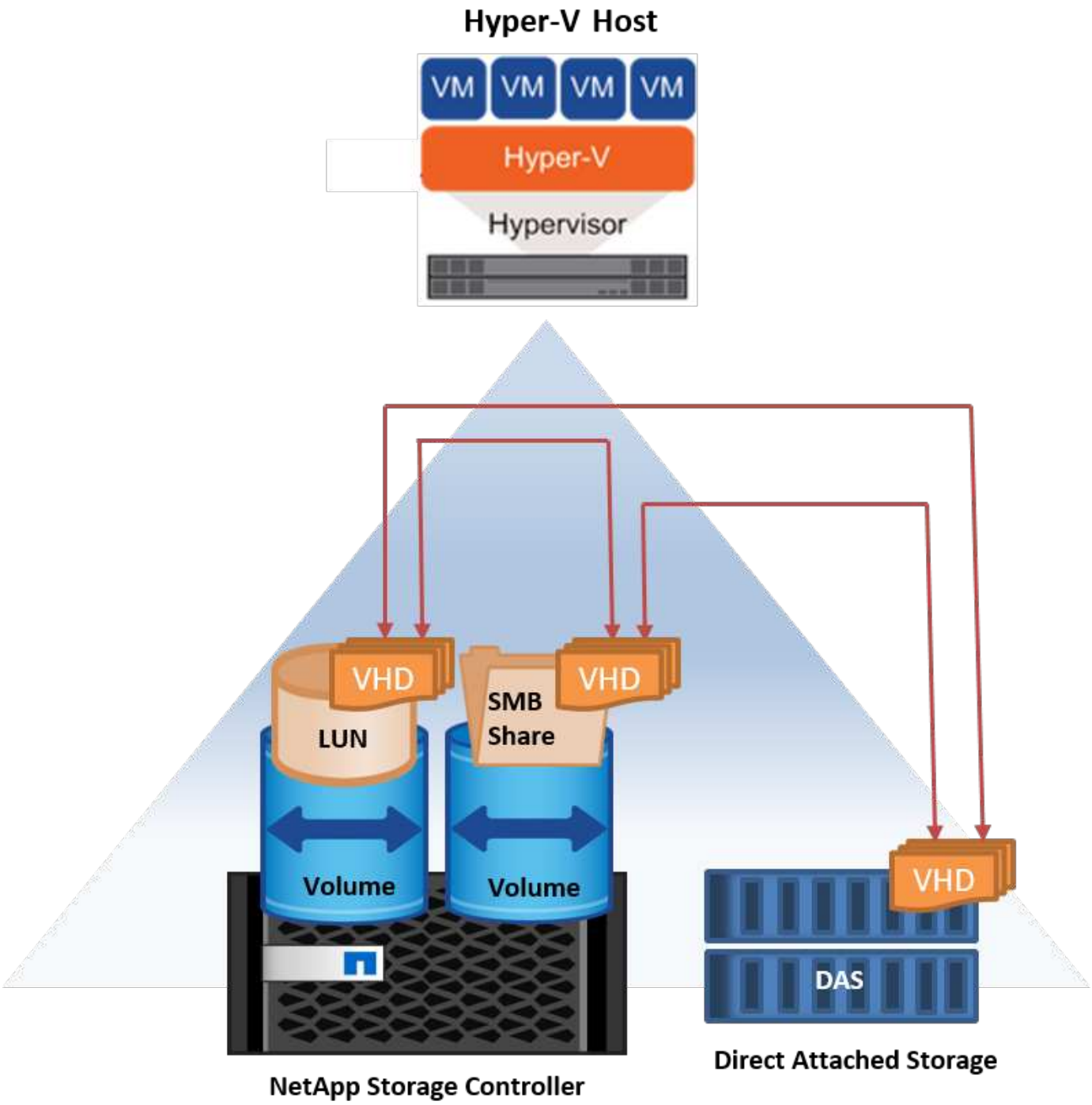
NetAppストレージコントローラでは、次のようなODXシナリオがサポートされます。

- * IntraSVM。*データは同じSVMに所有されます。
- *ボリューム内、イントラノード。*ソースとデスティネーションのファイルまたはLUNは同じボリューム内に存在します。コピーはFlexCloneファイルテクノロジを使用して実行されるため、リモートコピーのパフォーマンスがさらに向上します。
- *ボリューム間、イントラノード。*ソースとデスティネーションのファイルまたはLUNは、同じノード上の異なるボリュームにあります。
- *ボリューム間、ノード間。*ソースとデスティネーションのファイルまたはLUNは、異なるノード上にある異なるボリュームにあります。
- * InterSVM。*データは別 々 のSVMに所有されています。
- *ボリューム間、イントラノード。*ソースとデスティネーションのファイルまたはLUNは、同じノード上の異なるボリュームにあります。
- *ボリューム間、ノード間。*ソースとデスティネーションのファイルまたはLUNは、異なるノード上の異なるボリュームにあります。
- クラスタ間。 ONTAP 9.0以降では、SAN環境でのクラスタ間LUN転送でもODXがサポートされます。クラスタ間ODXはSANプロトコルでのみサポートされ、SMBではサポートされません。

移行が完了したら、VMを保持する新しいボリュームを反映するようにバックアップポリシーとレプリケーションポリシーを再設定する必要があります。以前に作成されたバックアップは使用できません。

VMストレージ（VHD / VHDX）は、次のストレージタイプ間で移行できます。

- DASとSMB共有
- DASとLUN
- SMB共有とLUN
- LUNカン
- SMBキヨウユウカン



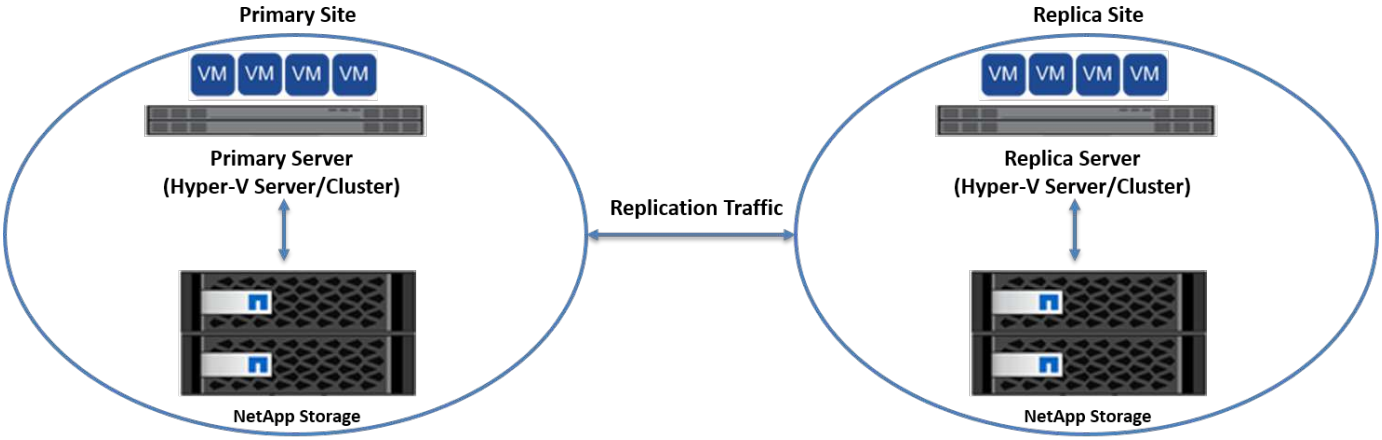
さらに読みます

ストレージライブマイグレーションの導入については、を参照してください。 "付録E：Hyper-Vストレージラ

イブマイグレーションの導入"。

Hyper-Vレプリカ：仮想マシンのディザスタリカバリ

Hyper-Vレプリカは、プライマリサイトからセカンダリサイトのレプリカVMにHyper-V VMをレプリケートし、VMのディザスタリカバリを非同期で提供します。VMをホストするプライマリサイトのHyper-Vサーバをプライマリサーバと呼び、レプリケートされたVMを受け取るセカンダリサイトのHyper-Vサーバをレプリカサーバと呼びます。次の図に、Hyper-Vレプリカのシナリオ例を示します。Hyper-Vレプリカは、フェイルオーバークラスタの一部であるHyper-Vサーバ間、またはどのクラスタにも属さない独立したHyper-Vサーバ間で、VMに使用できます。



レプリケーション

プライマリサーバ上のVMに対してHyper-Vレプリカが有効になると、最初のレプリケーションではレプリカサーバ上に同一のVMが作成されます。最初のレプリケーション後、Hyper-VレプリカはVMのVHDのログファイルを保持します。ログファイルは、レプリケーション頻度に応じてレプリカVHDに対して逆の順序で再生されます。このログと逆の順序を使用することで、最新の変更が非同期で保存され、レプリケートされます。想定される頻度でレプリケーションが実行されない場合は、アラートが発行されます。

拡張レプリケーション

Hyper-Vレプリカは、セカンダリレプリカサーバをディザスタリカバリ用に構成できる拡張レプリケーションをサポートしています。セカンダリレプリカサーバは、レプリカサーバがレプリカVM上の変更を受信するように構成できます。拡張レプリケーションシナリオでは、プライマリサーバ上のプライマリVMの変更がレプリカサーバにレプリケートされます。その後'変更内容が拡張レプリカ・サーバに複製されますプライマリサーバとレプリカサーバの両方がダウンした場合にのみ、VMを拡張レプリカサーバにフェイルオーバーできます。

フェイルオーバー

フェイルオーバーは自動ではなく、手動で実行する必要があります。フェイルオーバーには、次の3種類があります。

- *フェイルオーバーのテスト。*このタイプは、レプリカVMがレプリカサーバで正常に起動し、レプリカVMで開始されることを確認するために使用されます。このプロセスでは、フェイルオーバー時にテストVMの複製が作成され、通常の本番レプリケーションには影響しません。
- *計画的フェイルオーバー。*このタイプは、計画的停止または予期される停止中にVMをフェイルオーバーするために使用されます。このプロセスはプライマリVMで開始されます。計画的フェイルオーバーを実行する前に、プライマリサーバでこのプロセスをオフにする必要があります。マシンがフェイルオーバー

すると、Hyper-V Replicaはレプリカサーバ上のレプリカVMを起動します。

- *計画外フェイルオーバー。*このタイプは、予期しない停止が発生した場合に使用されます。このプロセスはレプリカVMで開始され、プライマリマシンに障害が発生した場合にのみ使用する必要があります。

リカバリ

VMのレプリケーションを設定するときに、リカバリポイントの数を指定できます。リカバリポイントは、レプリケートされたマシンからデータをリカバリできる時点を表します。

さらに読みます

- Hyper-Vレプリカをクラスタ環境外に導入する方法については、「"クラスタ環境外にHyper-Vレプリカを導入する". 」
- クラスタ環境へのHyper-Vレプリカの導入については、「"クラスタ環境へのHyper-Vレプリカの導入". 」

ストレージ効率

ONTAPは、Microsoft Hyper-Vをはじめとする仮想環境向けに、業界をリードするStorage Efficiencyテクノロジを提供します。NetAppでは、ストレージ容量削減保証プログラムも提供しています。

NetApp重複排除

NetApp重複排除は、ストレージボリュームレベルで重複ブロックを削除することで機能します。論理コピーの数に関係なく、物理コピーは1つだけ保存されます。そのため、重複排除機能を使用すると、そのブロックのコピーが多数あるという錯覚が生じます。重複排除は、ボリューム全体の4KBブロックレベルで重複データブロックを自動的に削除します。このプロセスでは、ディスクへの物理的な書き込み回数が減るため、ストレージが再利用されてスペースが確保され、パフォーマンスが削減される可能性があります。Hyper-V環境では、重複排除機能によってスペースを70%以上削減できます。

シンプロビジョニング

シンプロビジョニングは、ストレージを事前に割り当てる必要がないため、効率的にストレージをプロビジョニングできます。つまり、シンプロビジョニングを使用してボリュームまたはLUNを作成した場合、ストレージシステム上のスペースは使用されません。スペースは、データがLUNまたはボリュームに書き込まれ、データの格納に必要なスペースだけが使用されるまで未使用のままです。NetAppでは、ボリュームでシンプロビジョニングを有効にし、LUNリザベーションを無効にすることを推奨します。

Quality of Service の略

clustered ONTAPのストレージQoSを使用すると、ストレージオブジェクトをグループ化し、グループにスループットの制限を設定できます。ストレージQoSを使用すると、ワークロードに対するスループットを制限したり、ワークロードのパフォーマンスを監視したりできます。これにより、ストレージ管理者は、組織、アプリケーション、ビジネスユニット、本番環境や開発環境ごとにワークロードを分離できます。

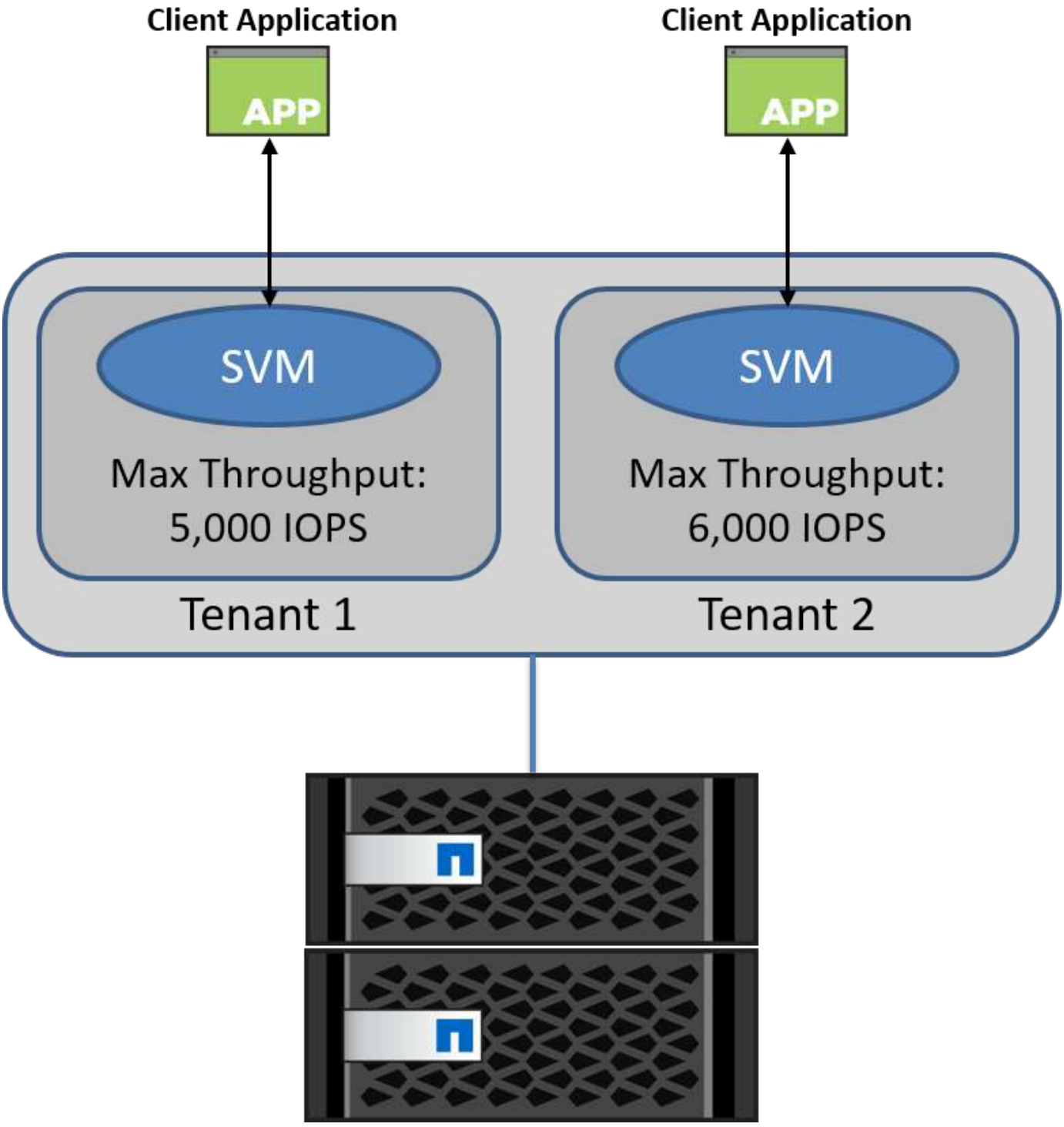
エンタープライズ環境では、ストレージQoSによって次のことが実現されます。

- ユーザのワークロード間での影響を防止
- IT-as-a-Service（ITaaS）環境で満たす必要がある特定の応答時間がある重要なアプリケーションを保護します。

- テナント間での影響を防止
- 新しいテナントを1つずつ追加することで、パフォーマンスの低下を回避できます。

QoSを使用すると、SVM、フレキシブルボリューム、LUN、またはファイルに送信するI/Oの量を制限できます。I/Oは処理数または物理スループットによって制限される場合があります。

次の図は、最大スループット制限を適用する独自のQoSポリシーが設定されたSVMを示しています。



SVMに独自のQoSポリシーを設定し、ポリシーグループを監視するには、ONTAPクラスタで次のコマンドを実行します。

```
# create a new policy group pg1 with a maximum throughput of 5,000 IOPS
cluster::> qos policy-group create pg1 -vserver vs1 -max-throughput
5000iops
```

```
# create a new policy group pg2 without a maximum throughput
cluster::> qos policy-group create pg2 -vserver vs2
```

```
# monitor policy group performance
cluster::> qos statistics performance show
```

```
# monitor workload performance
cluster::> qos statistics workload performance show
```

セキュリティ

ONTAPは、Windowsオペレーティングシステムにセキュアなストレージシステムを提供します。

Windows Defenderアンチウイルス

Windows Defenderは、Windows Serverにインストールされ、デフォルトで有効になっているマルウェア対策ソフトウェアです。このソフトウェアは、既知のマルウェアからWindows Serverを積極的に保護し、Windows Updateを介して定期的にマルウェア対策の定義を更新することができます。NetApp LUNおよびSMB共有は、Windows Defenderを使用してスキャンできます。

さらに読みます

詳細については、を参照してください。 "Windows Defenderの概要"。

BitLocker

BitLockerドライブ暗号化は、Windows Server 2012から引き継がれたデータ保護機能です。この暗号化により、物理ディスク、LUN、およびCSVが保護されます。

ベストプラクティス

BitLockerを有効にする前に、CSVをメンテナンスモードにする必要があります。そのため、NetAppでは、ダウンタイムを回避するために、CSV上にVMを作成する前に、BitLockerベースのセキュリティに関する決定を行うことを推奨しています。

Nanoサーバーの導入

Microsoft Windows Nano Serverの導入について説明します。

導入

Nano ServerをHyper-Vホストとして導入するには、次の手順を実行します。

1. 管理者グループのメンバーとしてWindows Serverにログインします。
2. Windows Server ISOの\NanoServerフォルダからNanoServerImageGeneratorフォルダをローカルハードドライブにコピーします。
3. Nano Server VHD/VHDXを作成するには、次の手順を実行します。
   a. Windows PowerShellを管理者として起動し、ローカルハードドライブ上のコピーされたNanoServerImageGeneratorフォルダに移動して、次のコマンドレットを実行します。

```
Set-ExecutionPolicy RemoteSigned
Import-Module .\NanoServerImageGenerator -Verbose
```

   b. 次のPowerShellコマンドレットを実行して、Nano Server用のVHDをHyper-Vホストとして作成します。このコマンドを実行すると、新しいVHDの管理者パスワードを入力するように求められます。

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath <"input the path to the root of the contents of Windows
Server 2016 ISO"> -TargetPath <"input the path, including the
filename and extension where the resulting VHD/VHDX will be created">
-ComputerName <"input the name of the nano server computer you are
about to create"> -Compute
```

. . 次の例では、フェイルオーバークラスタリングが有効なHyper-Vホスト機能を持つNano Server VHDを作成します。この例では、f:\にマウントされたISOからNano Server VHDを作成します。新しく作成したVHDは、コマンドレットの実行元のフォルダ内のNanoServerという名前のフォルダに配置されます。コンピュータ名はNanoServerで、作成されたVHDにはWindows Serverの標準エディションが含まれています。

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath f:\ -TargetPath .\NanoServer.vhd -ComputerName NanoServer
-Compute -Clustering
```

. . コマンドレットNew-NanoServerImageを使用して、
IPアドレス、サブネットマスク、デフォルトゲートウェイ、DNSサーバ、ドメイン名を設定するパラメータを設定します。 など。

4. VMまたは物理ホストでVHDを使用して、Nano ServerをHyper-Vホストとして導入します。
   a. VMに導入する場合は、Hyper-V Managerで新しいVMを作成し、手順3で作成したVHDを使用します。
   b. 物理ホストに導入する場合は、VHDを物理コンピュータにコピーし、この新しいVHDから起動するように構成します。まず、VHDをマウントし、bcdboot e:\windows（VHDがE:\の下にマウントされている場所）を実行し、VHDをアンマウントし、物理コンピュータを再起動して、Nano Serverを起動します。

5. Nano Serverをドメインに参加させる（オプション）：

   a. ドメイン内の任意のコンピュータにログインし、次のPowerShellコマンドレットを実行してデータブロブを作成します。

```
$domain = "<input the domain to which the Nano Server is to be
joined>"
$nanoserver = "<input name of the Nano Server>"
```

```
 djoin.exe /provision /domain $domain /machine $nanoserver /savefile
C:\temp\odjblob /reuse
```

.. リモートマシンで次のPowerShellコマンドレットを実行して、odjblob ファイルをNano Serverにコピーします。

```
$nanoserver = "<input name of the Nano Server>"
$nanouname = ""<input username of the Nano Server>"
$nanopwd = ""<input password of the Nano Server>"
```

```
$filePath = 'c:\temp\odjblob'
$fileContents = Get-Content -Path $filePath -Encoding Unicode
```

```
$securenanopwd = ConvertTo-SecureString -AsPlainText -Force $nanopwd
$nanosecurecred = new-object management.automation.pscredential
$nanouname, $securenanopwd
```

```
Invoke-Command -VMName $nanoserver -Credential $nanosecurecred
-ArgumentList @($filePath,$fileContents) -ScriptBlock \{
    param($filePath,$data)
    New-Item -ItemType directory -Path c:\temp
    Set-Content -Path $filePath -Value $data -Encoding Unicode
    cd C:\temp
    djoin /requestodj /loadfile c:\temp\odjblob /windowspath
c:\windows /localos
}
```

   b. Nano Serverを再起動します。

Nanoサーバーへの接続

PowerShellを使用してNano Serverにリモート接続するには、次の手順を実行します。

1. リモートサーバーで次のコマンドレットを実行して、Nano Serverをリモートコンピュータ上の信頼できるホストとして追加します。

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts "<input IP Address of the Nano Server>"
```

.
環境が安全で、すべてのホストを信頼できるホストとしてサーバに追加する場合は、次のコマンドを実行します。

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts *
```

.
リモートサーバで次のコマンドレットを実行して、リモートセッションを開始します。プロンプトが表示されたら、`Nano Server`のパスワードを入力します。

```
Enter-PSSession -ComputerName "<input IP Address of the Nano Server>" -Credential ~\Administrator
```

リモートWindows ServerからGUI管理ツールを使用してNano Serverにリモート接続するには、次のコマンドを実行します。

1. 管理者グループのメンバーとしてWindows Serverにログインします。
2. Server Managerの起動。
3. Server ManagerからNano Serverをリモートで管理するには、All Servers（すべてのサーバー）を右クリックし、Add Servers（サーバーの追加）をクリックして、Nano Serverの情報を入力して追加します。これで、サーバーリストにNano Serverが表示されます。Nano Serverを選択し、右クリックして、提供されたさまざまなオプションで管理を開始します。
4. Nano Server上のサービスをリモートで管理するには、次の手順を実行します。
   a. サーバーマネージャのツールセクションからサービスを開きます。
   b. [サービス（ローカル）]を右クリックします。
   c. [Connect to Server]をクリックします。
   d. Nano Serverのサービスを表示および管理するためのNano Serverの詳細情報を提供します。
5. Nano ServerでHyper-Vの役割が有効になっている場合は、次の手順を実行してHyper-V Managerからリモートで管理します。
   a. [Server Manager]の[Tools]セクションから[Hyper-V Manager]を開きます。
   b. [Hyper-V Manager]を右クリックします。
   c. [Connect to Server]をクリックし、Nano Serverの詳細を入力します。Nano ServerをHyper-Vサーバとして管理し、その上にVMを作成および管理できるようになりました。
6. Nano Serverでフェールオーバークラスタリングロールが有効になっている場合は、次の手順を実行してフェールオーバークラスタマネージャからリモートで管理します。

a. Server ManagerのToolsセクションからFailover Cluster Managerを開きます。

b. Nano Serverを使用してクラスタリング関連の操作を実行します。

Hyper-Vクラスタの導入

この付録では、Hyper-Vクラスタの導入について説明します。

前提条件

- 2台以上のHyper-Vサーバが相互に接続されている。
- 各Hyper-Vサーバに少なくとも1つの仮想スイッチが設定されている。
- フェイルオーバークラスタ機能は、各Hyper-Vサーバで有効になっています。
- SMB共有（CSV）は、Hyper-VクラスタリングのためにVMとそのディスクを格納する共有ストレージとして使用されます。
- 異なるクラスタ間でストレージを共有しないでください。クラスタごとにCSV / CIFS共有を1つだけ設定する必要があります。
- SMB共有を共有ストレージとして使用する場合は、SMB共有に対する権限を設定して、クラスタ内のすべてのHyper-Vサーバのコンピュータアカウントにアクセスを許可する必要があります。

導入

1. いずれかのWindows Hyper-Vサーバに管理者グループのメンバーとしてログインします。
2. Server Managerの起動。
3. [Tools]セクションで、[Failover][Cluster Manager]をクリックします。
4. [Actions]メニューから[Create Cluster]をクリックします。
5. このクラスタに含まれるHyper-Vサーバの詳細を指定します。
6. クラスタ構成を検証クラスタ構成の検証を求められたら[Yes]を選択し、Hyper-Vサーバがクラスタに参加するための前提条件を満たしているかどうかを検証するために必要なテストを選択します。
7. 検証に成功すると、クラスタ作成ウィザードが開始されます。ウィザードで、新しいクラスタのクラスタ名とクラスタのIPアドレスを入力します。次に、Hyper-Vサーバ用の新しいフェイルオーバークラスタが作成されます。
8. フェイルオーバークラスタマネージャで、新しく作成したクラスタをクリックして管理します。
9. クラスタで使用する共有ストレージを定義します。SMB共有またはCSVのいずれかです。
10. SMB共有を共有ストレージとして使用する場合、特別な手順は必要ありません。
    - NetAppストレージコントローラにCIFS共有を設定します。これを行うには、「"SMB環境でのプロビジョニング""。
11. CSVを共有ストレージとして使用するには、次の手順を実行します。
    a. NetAppストレージコントローラでLUNを設定します。これを行うには、「SAN環境でのプロビジョニング」を参照してください。
    b. フェイルオーバークラスタ内のすべてのHyper-VサーバがNetApp LUNを認識できることを確認します。フェイルオーバークラスタに含まれるすべてのHyper-Vサーバに対してこの処理を実行するには、それぞれのイニシエータがNetAppストレージのイニシエータグループに追加されていることを確認し

ます。また、LUNが検出され、MPIOが有効になっていることを確認してください。

c. クラスタ内のいずれかのHyper-Vサーバで、次の手順を実行します。

   i. LUNをオンラインにし、ディスクを初期化し、新しいシンプルボリュームを作成し、NTFSまたはReFSを使用してフォーマットします。

   ii. フェイルオーバークラスタマネージャで、クラスタを展開し、ストレージを展開し、ディスクを右クリックして、ディスクの追加をクリックします。追加すると、クラスタへのディスクの追加ウィザードが開き、LUNがディスクとして表示されます。[OK]をクリックしてLUNをディスクとして追加します。

   iii. これで、LUNの名前がClustered Diskになり、[Disks]に[Available Storage]と表示されます。

d. LUN（クラスタディスク）を右クリックし、[Add to Cluster Shared Volumes]をクリックします。これで、LUNがCSVとして表示されます。

e. CSVは、ローカルの場所C：\ClusterStorage\にあるフェイルオーバークラスタのすべてのHyper-Vサーバから同時に認識およびアクセスできます。

12. 高可用性VMを作成します。

    a. フェイルオーバークラスタマネージャで、前の手順で作成したクラスタを選択して展開します。

    b. [Roles]をクリックし、[Actions]で[Virtual Machines]をクリックします。[New Virtual Machine]をクリックします。

    c. VMを配置するクラスタからノードを選択します。

    d. [Virtual Machine Creation]ウィザードで、VMとそのディスクを格納するパスとして共有ストレージ（SMB共有またはCSV）を指定します。

    e. Hyper-V Managerを使用して、VMとそのディスクをHyper-Vサーバ用に格納するためのデフォルトパスとして共有ストレージ（SMB共有またはCSV）を設定します。

13. 計画的フェイルオーバーをテストする。ライブマイグレーション、クイックマイグレーション、またはストレージマイグレーション（移動）を使用して、VMを別のノードに移動します。レビュー "クラスタ環境でのライブマイグレーション" 詳細：

14. 計画外フェイルオーバーをテストVMを所有するサーバでクラスタサービスを停止します。

クラスタ環境へのHyper-Vライブマイグレーションの導入

この付録では、クラスタ環境へのライブマイグレーションの導入について説明します。

前提条件

ライブマイグレーションを導入するには、共有ストレージを使用するフェイルオーバークラスタでHyper-Vサーバを構成する必要があります。レビュー "Hyper-Vクラスタの導入" 詳細：

導入

クラスタ環境でライブマイグレーションを使用するには、次の手順を実行します。

1. フェイルオーバークラスタマネージャで、クラスタを選択して展開します。クラスタが表示されない場合は、[Failover][Cluster Manager]をクリックし、[Connect to Cluster]をクリックしてクラスタ名を指定します。

2. [Roles]をクリックします。クラスタで使用可能なすべてのVMが表示されます。

3. VMを右クリックし、[Move]をクリックします。これにより、次の3つのオプションが提供されます。
    - *ライブマイグレーション。*手動でノードを選択することも、クラスタが最適なノードを選択できるようにすることもできます。ライブマイグレーションでは、クラスタはVMで使用されているメモリを現在のノードから別のノードにコピーします。そのため、VMを別のノードに移行すると、VMに必要なメモリと状態の情報がVMにすでに用意されています。この移行方法はほぼ瞬時ですが、一度に移行できるVMは1つだけです。
    - *クイック移行。*ノードを手動で選択することも、クラスタが最適なノードを選択できるようにすることもできます。クイックマイグレーションでは、クラスタはVMで使用されているメモリをストレージ内のディスクにコピーします。そのため、VMを別のノードに移行すると、VMに必要なメモリと状態の情報を、もう一方のノードがディスクからすばやく読み取ることができます。クイックマイグレーションでは、複数のVMを同時に移行できます。
    - *仮想マシンストレージの移行。*この方法では、仮想マシンストレージの移動ウィザードを使用します。このウィザードでは、VMディスクと他のファイルを選択して、別の場所（CSV共有またはSMB共有）に移動できます。

クラスタ環境外への**Hyper-V**ライブマイグレーションの導入

このセクションでは、クラスタ環境外でのHyper-Vライブマイグレーションの導入について説明します。

前提条件

- 独立したストレージまたは共有SMBストレージを備えたスタンドアロンのHyper-Vサーバ。
- ソースサーバとデスティネーションサーバの両方にインストールされているHyper-Vの役割。
- 両方のHyper-Vサーバが同じドメインに属しているか、相互に信頼されているドメインに属しています。

導入

非クラスタ環境でライブマイグレーションを実行するには、ソースとデスティネーションのHyper-Vサーバがライブマイグレーション処理を送受信できるように設定します。両方のHyper-Vサーバで、次の手順を実行します。

1. [Server Manager]の[Tools]セクションから[Hyper-V Manager]を開きます。
2. [Actions]で[Hyper-V Settings]をクリックします。
3. [Live Migrations]をクリックし、[Enable Incoming and Outgoing Live Migrations]を選択します。
4. 使用可能なネットワーク上でライブマイグレーショントラフィックを許可するか、特定のネットワーク上でのみ許可するかを選択します。
5. 必要に応じて、Live MigrationsのAdvancedセクションから認証プロトコルとパフォーマンスオプションを設定できます。
6. CredSSPを認証プロトコルとして使用している場合は、VMを移動する前に、デスティネーションHyper-VサーバからソースHyper-Vサーバにログインしてください。
7. Kerberosを認証プロトコルとして使用する場合は、制約付き委任を設定します。そのためには、Active Directoryドメインコントローラへのアクセスが必要です。委譲を設定するには、次の手順を実行します。
    a. Active Directoryドメインコントローラに管理者としてログインします。
    b. Server Managerを起動します。

c. [ツール]セクションで、[Active Directoryユーザーとコンピュータ]をクリックします。

d. ドメインを展開し、[コンピュータ]をクリックします。

e. リストからソースHyper-Vサーバを選択して右クリックし、[Properties]をクリックします。

f. [委任]タブで、[このコンピュータを信頼して指定されたサービスのみに委任する]を選択します。

g. [Kerberosのみを使用]を選択します。

h. [追加]をクリックします。[サービスの追加]ウィザードが開きます。

i. [サービスの追加]で[ユーザーとコンピュータ]をクリックすると、[ユーザーまたはコンピュータの選択]が開きます。

j. デスティネーションHyper-Vサーバ名を指定し、[OK]をクリックします。

- VMストレージを移動するには、[CIFS]を選択します。
- VMを移動するには、[Microsoft Virtual System Migration]サービスを選択します。

k. [委任]タブで、[OK]をクリックします。

l. [Computers]フォルダで、リストから移行先のHyper-Vサーバを選択し、この処理を繰り返します。[Select Users or Computers]で、ソースHyper-Vサーバ名を指定します。

8. VMを移動します。

a. Hyper-V Managerを開きます。

b. VMを右クリックし、[Move]をクリックします。

c. [Move the Virtual Machine]を選択します。

d. VMのデスティネーションHyper-Vサーバを指定します。

e. 移動オプションを選択します。[Shared Live Migration]で、[Move Only the Virtual Machine]を選択します。シェアードナッシングライブマイグレーションでは、設定に基づいて他の2つのオプションのいずれかを選択します。

f. 必要に応じて、デスティネーションHyper-Vサーバ上のVMの場所を指定します。

g. 概要を確認し、[OK]をクリックしてVMを移動します。

Hyper-Vストレージのライブマイグレーションの導入

Hyper-Vストレージのライブマイグレーションの設定方法

前提条件

- 独立したストレージ（DASまたはLUN）またはSMBストレージ（ローカルまたは他のHyper-Vサーバ間で共有）を備えたスタンドアロンのHyper-Vサーバが必要です。
- Hyper-Vサーバがライブマイグレーション用に設定されている必要があります。の導入に関するセクションを確認します。 "クラスタ環境外でのライブマイグレーション"。

導入

1. Hyper-V Managerを開きます。
2. VMを右クリックし、[Move]をクリックします。

3. [仮想マシンのストレージの移動]を選択します。
4. 設定に基づいてストレージの移動オプションを選択します。
5. VMの項目の新しい場所を指定します。
6. 概要を確認し、[OK]をクリックしてVMのストレージを移動します。

クラスタ環境外への**Hyper-V**レプリカの導入

この付録では、クラスタ環境の外部にHyper-Vレプリカを導入する方法について説明します。

前提条件

- プライマリサーバおよびレプリカサーバとして機能する、同じまたは別の地理的な場所にスタンドアロンのHyper-Vサーバが必要です。
- 別々のサイトを使用する場合は、プライマリサーバとレプリカサーバ間の通信を許可するように各サイトのファイアウォールを設定する必要があります。
- レプリカサーバには、レプリケートされたワークロードを格納するための十分なスペースが必要です。

導入

1. レプリカサーバを構成します。
   a. インバウンドファイアウォールルールでレプリケーショントラフィックの受信を許可するには、次のPowerShellコマンドレットを実行します。

```
 Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP Listener
(TCP-In)"
.. [Server Manager]の[Tools]セクションから[Hyper-V Manager]を開きます。
.. [Actions]から[Hyper-V Settings]をクリックします。
.. [Replication Configuration]をクリックし、[Enable this computer as a
Replica Server]を選択します。
.. [Authentication and Ports]セクションで、認証方法とポートを選択します。
.. [Authorization and Storage]セクションで、レプリケートされた
VMとファイルを格納する場所を指定します。
```

2. プライマリサーバ上のVMのVMレプリケーションを有効にします。VMのレプリケーションは、Hyper-Vサーバ全体ではなく、VM単位で有効になります。
   a. Hyper-V Managerで、VMを右クリックして[Enable Replication]をクリックし、[Enable Replication]ウィザードを開きます。
   b. VMをレプリケートするレプリカサーバの名前を指定します。
   c. レプリカサーバでレプリケーショントラフィックを受信するように構成された認証タイプとレプリカサーバポートを指定します。
   d. レプリケートするVHDを選択します。
   e. 変更がレプリカサーバに送信される頻度（期間）を選択します。

f. リカバリポイントを構成して、レプリカサーバ上で保持するリカバリポイントの数を指定します。

g. [Initial Replication Method]を選択して、VMデータの初期コピーをレプリカサーバに転送する方法を指定します。

h. 概要を確認し、[Finish]をクリックします。

i. このプロセスでは、レプリカサーバ上にVMレプリカが作成されます。

レプリケーション

1. テストフェイルオーバーを実行して、レプリカVMがレプリカサーバ上で正常に機能することを確認します。このテストでは、レプリカサーバに一時VMを作成します。
   a. レプリカサーバにログインします。
   b. Hyper-V Managerで、レプリカVMを右クリックし、[Replication]をクリックして、[Test Failover]をクリックします。
   c. 使用するリカバリポイントを選択します。
   d. このプロセスでは、-Testが追加された同じ名前のVMが作成されます。
   e. VMを検証して、すべてが正常に動作することを確認します。
   f. フェイルオーバー後、レプリカテストVMに対して[Stop Test Failover]を選択すると、レプリカテストVMは削除されます。
2. 計画的フェイルオーバーを実行して、プライマリVMの最新の変更をレプリカVMにレプリケートします。
   a. プライマリサーバにログインします。
   b. フェイルオーバーするVMをオフにします。
   c. Hyper-V Managerで、オフになっているVMを右クリックし、[Replication]をクリックして、[Planned Failover]をクリックします。
   d. [Failover]をクリックして、最新のVM変更をレプリカサーバに転送します。
3. プライマリVMに障害が発生した場合は、計画外フェイルオーバーを実行します。
   a. レプリカサーバにログインします。
   b. Hyper-V Managerで、レプリカVMを右クリックし、[Replication]をクリックして、[Failover]をクリックします。
   c. 使用するリカバリポイントを選択します。
   d. [Failover]をクリックしてVMをフェイルオーバーします。

クラスタ環境への**Hyper-V**レプリカの導入

Windows Serverフェイルオーバークラスタを使用してHyper-Vレプリカを導入および構成する方法について説明します。

前提条件

- Hyper-Vクラスタを同じ場所または別の地理的な場所に配置し、プライマリクラスタとレプリカクラスタとして機能させる必要があります。レビュー "Hyper-Vクラスタの導入" 詳細：
- 別 々 のサイトを使用する場合は、プライマリクラスタとレプリカクラスタ間の通信を許可するように各

サイトのファイアウォールを設定する必要があります。

- レプリカクラスタには、レプリケートされたワークロードを格納できるだけの十分なスペースが必要です。

導入

1. クラスタのすべてのノードでファイアウォールルールを有効にします。プライマリクラスタとレプリカクラスタの両方のすべてのノードで、管理者権限で次のPowerShellコマンドレットを実行します。

```
# For Kerberos authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP
Listener (TCP-In)"}}
```

```
# For Certificate authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica
HTTPS Listener (TCP-In)"}}
```

2. レプリカクラスタを構成します。
   a. レプリカクラスタとして使用されるクラスタへの接続ポイントとして使用するNetBIOS名とIPアドレスを使用して、Hyper-Vレプリカブローカーを設定します。
      i. フェイルオーバークラスタマネージャを開きます。
      ii. クラスタを展開し、[Roles]をクリックして、[Actions]ペインで[Configure Role]をクリックします。
      iii. [Select Role]ページで[Hyper-V Replica Broker]を選択します。
      iv. クラスタ（クライアントアクセスポイント）への接続ポイントとして使用するNetBIOS名とIPアドレスを指定します。
      v. このプロセスでは、Hyper-Vレプリカブローカーの役割が作成されます。正常にオンラインになったことを確認します。
   b. レプリケーション設定を構成します。
      i. 前の手順で作成したレプリカブローカーを右クリックし、[Replication Settings]をクリックします。
      ii. [このクラスタをレプリカサーバとして有効にする]を選択します。
      iii. [Authentication and Ports]セクションで、認証方法とポートを選択します。
      iv. [Authorization and storage]セクションで、このクラスタへのVMのレプリケートを許可するサーバを選択します。また、レプリケートされたVMが格納されるデフォルトの場所を指定します。

レプリケーション

レプリケーションは、で説明したプロセスと似ています。 "クラスタ環境外のレプリカ"。

追加情報の参照先

Microsoft WindowsおよびHyper-Vに関するその他のリソース

- ONTAP の概念
  https://docs.netapp.com/us-en/ontap/concepts/introducing-data-management-software-concept.html
- 最新SANのベストプラクティス
  https://www.netapp.com/media/10680-tr4080.pdf
- NetAppオールSANアレイデータの可用性とNetApp ASAとの整合性
  https://www.netapp.com/pdf.html?item=/media/85671-tr-4968.pdf
- SMBのドキュメント
  https://docs.netapp.com/us-en/ontap/smb-admin/index.html
- Nano Server入門+
  https://technet.microsoft.com/library/mt126167.aspx
- Windows Server上のHyper-Vの新機能+
  https://technet.microsoft.com/windows-server-docs/compute/hyper-v/what-s-new-in-hyper-v-on-windows

Microsoft SQL Server の場合

ONTAP上のMicrosoft SQL Server

ONTAPは解決策、Microsoft SQL Serverデータベースにエンタープライズクラスのセキュリティとパフォーマンスを提供すると同時に、環境を管理するためのワールドクラスのツールも提供します。

このドキュメントは、以前に公開されたテクニカルレポート_TR-4590：『Best Practice Guide for Microsoft SQL Server with ONTAP』の内容を置き換えます。

NetAppは、読者が以下について実用的な知識を持っていることを前提としています。

- ONTAP ソフトウェア
- バックアップソフトウェアとしてのNetApp SnapCenterには、次のものが含まれます。
  - SnapCenter Plug-in for Microsoft Windows の略
  - SQL Server向けSnapCenterプラグイン
- Microsoft SQL Serverのアーキテクチャと管理

このベストプラクティスセクションの範囲は、NetAppがストレージインフラに推奨する設計原則と推奨される標準に基づいた技術設計に限定されます。エンドツーエンドの実装は範囲外です。

NetApp製品間での設定の互換性については、を参照してください。 "ネットアップの Interoperability Matrix Tool （ IMT ）"。

Microsoft SQL Serverのワークロード

SQL Serverを導入する前に、SQL Serverインスタンスがサポートするアプリケーションのデータベースワークロード要件を理解しておく必要があります。容量、パフォーマンス、可用性に関する要件はアプリケーションごとに異なるため、各データベースはこれらの要件を最適にサポートするように設計する必要があります。多くの組織では、アプリケーション要件を使用してSLAを定義し、データベースを複数の管理階層に分類しています。SQL Serverのワークロードは次のように記述できます。

- OLTPデータベースは、多くの場合、組織で最も重要なデータベースでもあります。これらのデータベースは通常、顧客向けのアプリケーションをバックアップし、企業の中核業務に不可欠であると考えられています。ミッションクリティカルなOLTPデータベースとサポート対象のアプリケーションには、高レベルのパフォーマンスが必要で、パフォーマンスの低下や可用性の影響を受けやすいSLAが設定されていることがよくあります。また、Always-Onフェイルオーバークラスタ&Always-On可用性グループの候補になることもあります。これらのタイプのデータベースのI/O構成は、通常、ランダムリードが75$_{90\%、書き込みが25}$10%という特徴があります。
- 意思決定支援システム(DSS)データベースは、データウェアハウスとも呼ばれます。これらのデータベースは、ビジネスの分析に依存している多くの組織でミッションクリティカルです。これらのデータベースは、クエリの実行時にCPU利用率やディスクからの読み取り処理の影響を受けます。多くの組織では、DSSデータベースは月、四半期、年末に最も重要なデータベースです。このワークロードは、一般に100%の読み取りI/O構成です。

データベース設定

Microsoft SQL ServerのCPU構成

システムパフォーマンスを向上させるには、SQL Serverの設定とサーバ構成を変更して、適切な数のプロセッサを実行に使用する必要があります。

ハイパースレッディング

ハイパースレッディングはインテル独自の同時マルチスレッド(SMT)実装であり、x86マイクロプロセッサ上で実行される計算（マルチタスク）の並列化を改善します。

ハイパースレッディングを使用するハードウェアでは、論理ハイパースレッディングCPUを物理CPUとしてオペレーティングシステムに認識させることができます。SQL Serverは、オペレーティングシステムが提供する物理CPUを認識し、ハイパースレッドプロセッサを使用できます。これにより、並列化が促進され、パフォーマンスが向上します。

ここで注意すべき点は、SQL Serverの各バージョンには、使用できるコンピューティング能力に独自の制限があることです。詳細については、「SQL Serverのエディション別のコンピューティング容量制限」を参照してください。

SQL Serverのライセンスには2つのオプションがあります。1つ目はサーバ+クライアントアクセスライセンス（CAL）モデルと呼ばれ、2つ目はプロセッサごとのコアモデルです。SQL Serverで利用可能なすべての製品機能には、サーバ+ CAL戦略でアクセスできますが、ソケットあたりのCPUコア数はハードウェアで20に制限されています。ソケットあたり20個以上のCPUコアを搭載したサーバ用のSQL Server Enterprise Edition+CALがある場合でも、アプリケーションはそのインスタンスですべてのコアを一度に使用することはできません。

次の図は、起動後のSQL Serverログメッセージを示しています。これは、コア制限の適用を示しています。

ログエントリは、**SQL Server**の起動後に使用されているコアの数を示します。

```
2017-01-11 07:16:30.71 Server      Microsoft SQL Server 2016
(RTM) - 13.0.1601.5 (X64)
     Apr 29 2016 23:23:58
     Copyright (c) Microsoft Corporation
     Enterprise Edition (64-bit) on Windows Server 2016
Datacenter 6.3 <X64> (Build 14393: )

2017-01-11 07:16:30.71 Server      UTC adjustment: -8:00
2017-01-11 07:16:30.71 Server      (c) Microsoft Corporation.
2017-01-11 07:16:30.71 Server      All rights reserved.
2017-01-11 07:16:30.71 Server      Server process ID is 10176.
2017-01-11 07:16:30.71 Server      System Manufacturer:
'FUJITSU', System Model: 'PRIMERGY RX2540 M1'.
2017-01-11 07:16:30.71 Server      Authentication mode is MIXED.
2017-01-11 07:16:30.71 Server      Logging SQL Server messages
in file 'C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG'.
2017-01-11 07:16:30.71 Server      The service account is 'SEA-
TM\FUJIA2R30$'. This is an informational message; no user action
is required.
2017-01-11 07:16:30.71 Server      Registry startup parameters:
     -d C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\master.mdf
     -e C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG
     -l C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\mastlog.ldf
     -T 3502
     -T 834
2017-01-11 07:16:30.71 Server      Command Line Startup
Parameters:
     -s "MSSQLSERVER"
2017-01-11 07:16:30.72 Server      SQL Server detected 2 sockets
with 18 cores per socket and 36 logical processors per socket,
72 total logical processors; using 40 logical processors based
on SQL Server licensing. This is an informational message; no
user action is required.
2017-01-11 07:16:30.72 Server      SQL Server is starting at
```

したがって、すべてのCPUを使用するには、プロセッサ単位のコアライセンスを使用する必要があります。SQL Serverのライセンスの詳細については、を参照してください。 "SQL Server 2022：最新データプラットフォームを実現"。

CPUアフィニティ

パフォーマンスの問題が発生しない限り、プロセッサアフィニティのデフォルトを変更する必要はありませんが、その内容と動作を理解する価値はあります。

SQL Serverは、次の2つのオプションでプロセッサアフィニティをサポートします。

- CPUアフィニティマスク
- アフィニティI/Oマスク

SQL Serverは、オペレーティングシステムで使用可能なすべてのCPUを使用します（プロセッサ単位のコアライセンスが選択されている場合）。すべてのCPUにスケジューラを作成し、特定のワークロードでリソースを最大限に活用します。マルチタスクを実行する場合、オペレーティングシステムやサーバー上のその他のアプリケーションは、プロセススレッドをプロセッサ間で切り替えることができます。SQL Serverはリソースを大量に消費するアプリケーションであるため、この状況が発生するとパフォーマンスに影響する可能性があります。影響を最小限に抑えるには、SQL Serverのすべての負荷が事前に選択されたプロセッサグループに送られるようにプロセッサを構成します。これは、CPUアフィニティマスクを使用することによって実現されます。

アフィニティI/Oマスクオプションは、SQL ServerディスクI/OをCPUのサブセットにバインドします。SQL Server OLTP環境では、この拡張により、I/O処理を実行するSQL Serverスレッドのパフォーマンスが向上します。

並列処理の最大次数**(MAXDOP)**

デフォルトでは、プロセッサ単位のコアライセンスが選択されている場合、SQL Serverはクエリの実行中に使用可能なすべてのCPUを使用します。

これは大規模なクエリには役立ちますが、原因のパフォーマンスが低下し、同時実行数が制限される可能性があります。1つのCPUソケット内の物理コアの数に並列処理を制限する方法が適しています。たとえば、ソケットあたり12コアの2つの物理CPUソケットを持つサーバでは、ハイパースレッディングに関係なく、MAXDOPを12に設定する必要があります。MAXDOPでは、使用するCPUを制限したり、指定したりすることはできません。代わりに、単一のバッチクエリで使用できるCPUの数を制限します。

* NetAppでは、データウェアハウスなどのDSSでは、MAXDOP 50から開始し、必要に応じてチューニングアップまたはチューニングダウンを検討することを推奨しています。変更を加えるときは、必ずアプリケーション内の重要なクエリを測定してください。

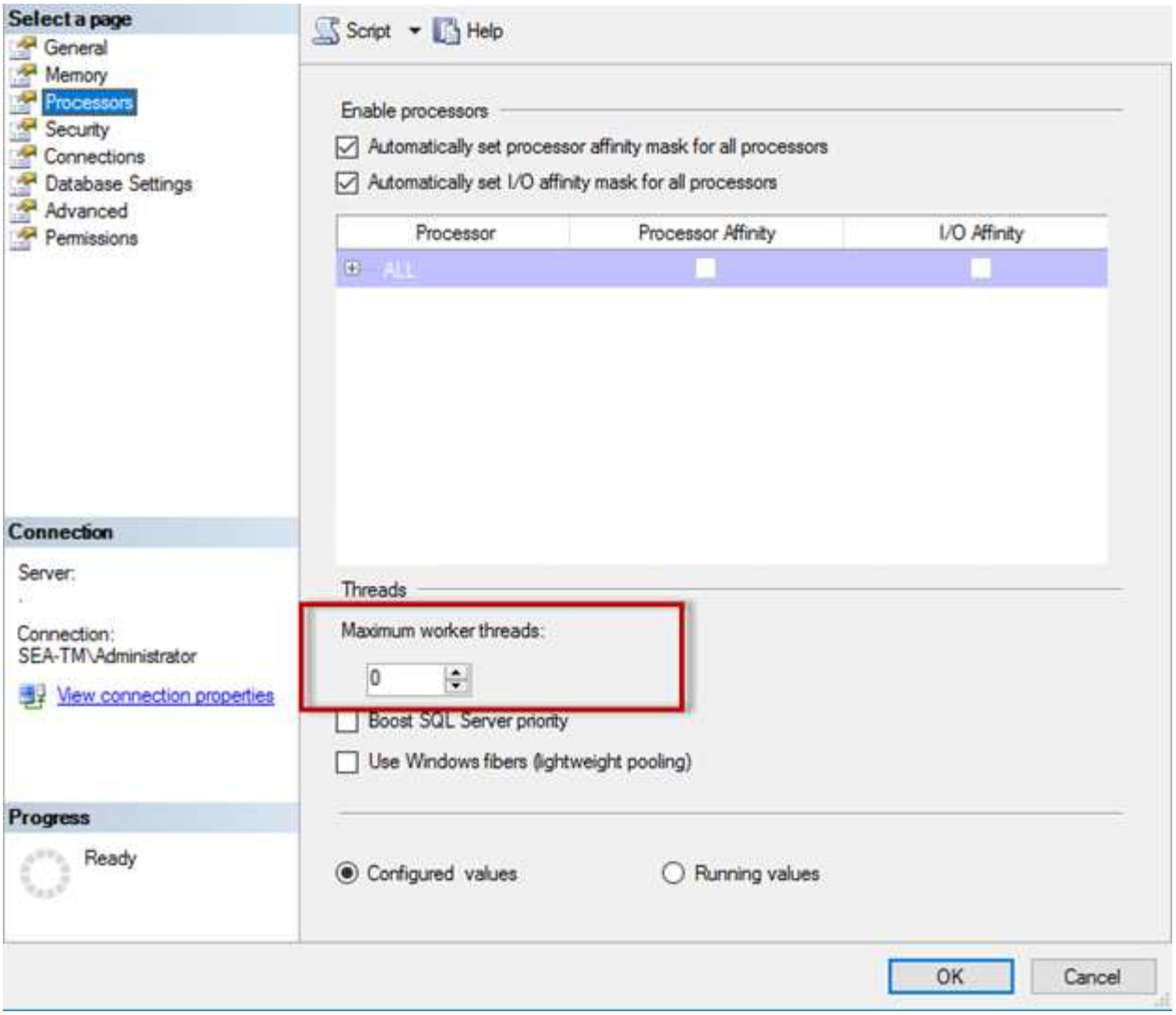
ワーカースレッドの最大数

最大ワーカースレッド数オプションは、多数のクライアントがSQL Serverに接続されている場合にパフォーマンスを最適化するのに役立ちます。

通常、クエリ要求ごとに個別のオペレーティングシステムスレッドが作成されます。SQL Serverへの同時接続が数百もの場合、クエリ要求ごとに1つのスレッドが大量のシステムリソースを消費します。最大ワーカースレッド数オプションを使用すると、SQL Serverでワーカースレッドのプールを作成して多数のクエリ要求を処理できるようになるため、パフォーマンスが向上します。

デフォルト値は0で、SQL Serverは起動時にワーカースレッド数を自動的に設定できます。これはほとんどのシステムで機能します。ワーカースレッドの最大数は高度なオプションであり、経験豊富なデータベース管理者（DBA）の支援なしに変更しないでください。

より多くのワーカースレッドを使用するようにSQL Serverを設定する必要があるのはいつですか？各スケジューラの平均ワークキューの長さが1を超える場合は、負荷がCPUに制限されていないか、その他の重い待機時間が発生している場合にのみ、システムにスレッドを追加することでメリットが得られます。これらのいずれかが発生している場合、スレッドを追加しても、他のシステムボトルネックを待つことになるため、効果はありません。最大ワーカースレッド数の詳細については、を参照してください。 "max worker threadsサーバ設定オプションの設定"。



SQL Server Management Studioを使用した最大ワーカースレッド数の設定。

```
The following example shows how to configure the max work threads option
using T-SQL.

EXEC sp_configure 'show advanced options', 1;
GO
RECONFIGURE ;
GO
EXEC sp_configure 'max worker threads', 900 ;
GO
RECONFIGURE;
GO
```

Microsoft SQL Serverのメモリ構成

次のセクションでは、データベースのパフォーマンスを最適化するためのSQL Serverメ

モリ設定の構成について説明します。

最大サーバメモリ

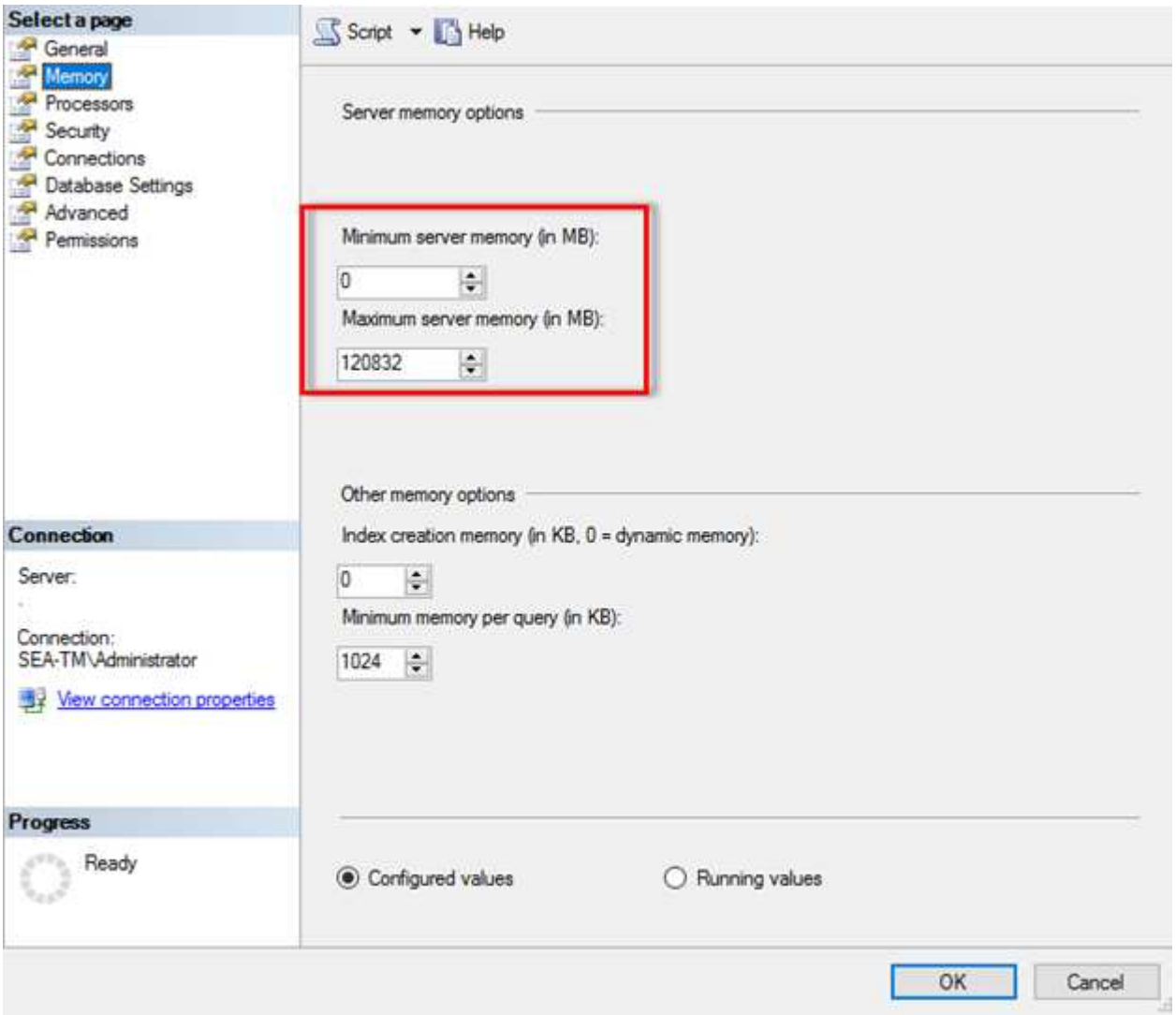
max server memoryオプションは、SQL Serverインスタンスで使用できるメモリの最大容量を設定します。

通常、SQL Serverが実行されている同じサーバで複数のアプリケーションが実行されていて、これらのアプリケーションが正常に機能するのに十分なメモリを確保したい場合に使用されます。

アプリケーションによっては、起動時に使用可能なメモリのみを使用し、必要に応じて要求しないものもあります。ここで、最大サーバメモリ設定が有効になります。

複数のSQL Serverインスタンスを持つSQL Serverクラスタでは、各インスタンスがリソースを競合する可能性があります。SQL Serverインスタンスごとにメモリ制限を設定すると、各インスタンスのパフォーマンスを最大限に高めることができます。

* NetAppでは、パフォーマンスの問題を回避するために、オペレーティングシステム用に少なくとも4GBから6GBのRAMを残しておくことを推奨しています。



SQL Server Management Studioを使用したサーバの最小メモリと最大メモリの調整

SQL Server Management Studioを使用してサーバの最小メモリまたは最大メモリを調整するには、SQL Serverサービスを再起動する必要があります。次のコードを使用して、Transact SQL（T-SQL）を使用してサーバメモリを調整できます。

```
EXECUTE sp_configure 'show advanced options', 1
GO
EXECUTE sp_configure 'min server memory (MB)', 2048
GO
EXEC sp_configure 'max server memory (MB)', 120832
GO
RECONFIGURE WITH OVERRIDE
```

不均一なメモリアクセス

NUMA（Nonuniform Memory Access）は、プロセッサバスの負荷を増やすことなくプロセッサ速度を向上させるメモリアクセス最適化方法です。

SQL Serverがインストールされているサーバで**NUMA**が構成されている場合、SQL ServerはNUMAを認識し、NUMAハードウェアで優れたパフォーマンスを発揮するため、追加の構成は必要ありません。

インデックス作成メモリ

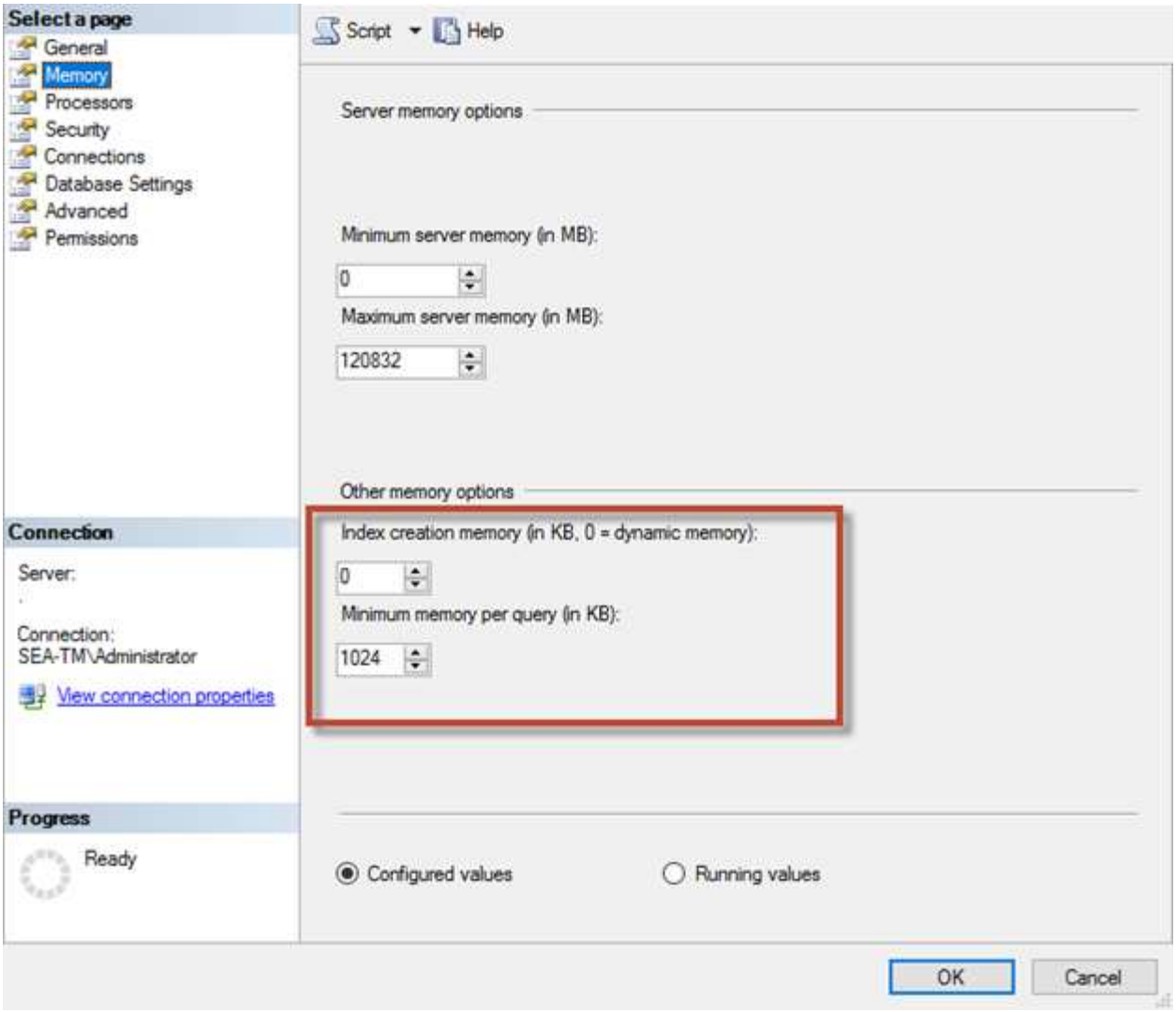
index create memoryオプションは、通常は変更しないもう1つの高度なオプションです。

インデックスを作成するために最初に割り当てられるRAMの最大容量を制御します。このオプションのデフォルト値は0です。これは、SQL Serverによって自動的に管理されることを意味します。ただし、インデックスの作成に問題がある場合は、このオプションの値を増やすことを検討してください。

クエリあたりの最小メモリ

クエリを実行すると、SQL Serverは効率的に実行するために最適なメモリ容量を割り当てようとします。

デフォルトでは、クエリごとの最小メモリ設定では、実行するクエリごとに>=が1024KBに割り当てられます。SQL Serverがインデックス作成処理に割り当てられるメモリ容量を動的に管理できるようにするには、この設定をデフォルト値の0のままにしておくことを推奨します。ただし、SQL ServerのRAM容量が効率的に実行するために必要な容量を超えている場合は、この設定を大きくすると、一部のクエリのパフォーマンスが向上することがあります。したがって、SQL Server、その他のアプリケーション、またはオペレーティングシステムで使用されていないサーバ上のメモリが使用可能であれば、この設定を大きくすることでSQL Serverの全体的なパフォーマンスを向上させることができます。空きメモリがない場合、この設定を増やすと、全体的なパフォーマンスが低下する可能性があります。



バッファプール拡張

バッファプール拡張機能を使用すると、NVRAM拡張機能とデータベースエンジンのバッファプールをシームレスに統合して、I/Oスループットを大幅に向上させることができます。

バッファプール拡張機能は、すべてのSQL Serverエディションで使用できるわけではありません。64ビットのSQL Server Standard、Business Intelligence、およびEnterpriseエディションでのみ使用できます。

バッファプール拡張機能は、不揮発性ストレージ（通常はSSD）を使用してバッファプールキャッシュを拡張します。この拡張機能により、バッファプールはより大規模なデータベースワーキングセットに対応できるようになり、RAMとSSD間のI/Oのページングが強制され、小さなランダムI/OがメカニカルディスクからSSDに効果的にオフロードされます。SSDのレイテンシが低く、ランダムI/Oのパフォーマンスが向上するため、バッファプールを拡張することでI/Oスループットが大幅に向上します。

バッファプール拡張機能には、次の利点があります。

- ランダムI/Oスループットの向上
- I/Oレイテンシの低減
- トランザクションスループットの向上
- ハイブリッドバッファプールの拡張による読み取りパフォーマンスの向上

- 既存および将来の低コストメモリを活用できるキャッシュアーキテクチャ

- NetAppでは、バッファプール拡張を次のように設定することを推奨しています。
- バッファプール拡張ターゲットディスクとして使用できるように、SSDベースのLUN（NetApp AFFなど）がSQL Serverホストに提供されていることを確認します。
- 拡張ファイルのサイズは、バッファプールと同じかそれよりも大きくする必要があります。

次に、バッファプール拡張を32GBに設定するT-SQLコマンドの例を示します。

```
USE master
GO
ALTER SERVER CONFIGURATION
SET BUFFER POOL EXTENSION ON
  (FILENAME = 'P:\BUFFER POOL EXTENSION\SQLServerCache.BUFFER POOL
EXTENSION', SIZE = 32 GB);
GO
```

Microsoft SQL Server共有インスタンスと専用インスタンス

複数のSQL Serverは、サーバごとに1つのインスタンスとして構成することも、複数のインスタンスとして構成することもできます。通常、適切な決定は、サーバを本番用と開発用のどちらに使用するか、インスタンスがビジネスの運用やパフォーマンスの目標にとって重要であると判断されるかなどの要因によって決まります。

共有インスタンスの構成は、最初は簡単に設定できますが、リソースが分割されたりロックされたりする問題が発生し、共有SQL Serverインスタンスでデータベースがホストされている他のアプリケーションでパフォーマンスの問題が発生する可能性があります。

どのインスタンスがルート原因であるかを把握する必要があるため、パフォーマンスの問題のトラブルシューティングは複雑になります。この質問は、オペレーティングシステムライセンスとSQL Serverライセンスのコストと比較して検討されます。アプリケーションのパフォーマンスを最優先する場合は、専用インスタンスを使用することを推奨します。

Microsoftでは、SQL Serverのライセンスは、インスタンス単位ではなく、コア単位でサーバレベルで付与されます。このため、データベース管理者は、ライセンスコストを削減するために、サーバで処理できる数のSQL Serverインスタンスをインストールしようとします。これは、後で大きなパフォーマンスの問題につながる可能性があります。

* NetAppでは、最適なパフォーマンスを得るために、可能な限り専用のSQL Serverインスタンスを選択することを推奨しています。

ストレージ構成

Microsoft SQL Serverストレージに関する考慮事項

ONTAPストレージソリューションとMicrosoft SQL Serverを組み合わせることで、今日の最も要求の厳しいアプリケーション要件を満たすエンタープライズレベルのデータベースストレージ設計を作成できます。

両方のテクノロジを最適化するには、SQL ServerのI/Oパターンと特性を理解することが重要です。SQL Serverデータベース用の適切に設計されたストレージレイアウトは、SQL Serverのパフォーマンスと SQL Serverインフラの管理をサポートします。また、ストレージレイアウトを適切に配置すれば、初期導入を成功させ、ビジネスの成長に合わせて環境をスムーズに拡張できます。

データストレージ設計

SnapCenter を使用してバックアップを実行しないSQL Serverデータベースについては、データファイルとログファイルを別 々 のドライブに配置することを推奨します。データを同時に更新して要求するアプリケーションでは、ログファイルに書き込み負荷がかかり、（アプリケーションによっては）データファイルの読み取り/書き込み負荷が高くなります。データを取得する場合、ログファイルは必要ありません。そのため、データの要求は、そのドライブに配置されたデータファイルから満たすことができます。

新しいデータベースを作成するときは、データとログ用に別 々 のドライブを指定することを推奨します。データベース作成後にファイルを移動するには、データベースをオフラインにする必要があります。Microsoftのその他の推奨事項については、 "データファイルとログファイルを別 々 のドライブに配置"。

アグリゲート

アグリゲートは、NetAppストレージ構成の最下位レベルのストレージコンテナです。一部のレガシードキュメントはインターネット上に存在し、異なるドライブセットにIOを分離することを推奨しています。これはONTAPでは推奨されません。NetAppは、データファイルとトランザクションログファイルを分離した共有アグリゲートと専用アグリゲートを使用して、さまざまなI/Oワークロードの特性評価テストを実施してきました。このテストでは、1つの大規模アグリゲートに複数のRAIDグループとドライブを配置することで、ストレージのパフォーマンスが最適化され、向上し、管理者が管理しやすくなることがわかりました。その理由は次の2つです。

- 1つの大きなアグリゲートで、すべてのドライブのI/O機能をすべてのファイルで使用できます。
- 1 つの大きなアグリゲートで、最も効率的なディスクスペースを使用できます。

高可用性（HA）を実現するには、SQL Server Always On可用性グループのセカンダリ同期レプリカを、アグリゲート内の別のStorage Virtual Machine（SVM）に配置します。ディザスタリカバリを目的とした場合は、DRサイト内の別のストレージクラスタの一部であるアグリゲートに非同期レプリカを配置し、NetApp SnapMirrorテクノロジを使用してコンテンツをレプリケートします。NetAppでは、ストレージのパフォーマンスを最適化するために、アグリゲートに利用可能な空きスペースを少なくとも10%確保することを推奨しています。

個のボリューム

NetApp FlexVolボリュームはアグリゲート内に作成され、格納されます。ONTAPボリュームがLUNではないため、この用語を使用すると混乱が生じることがあります。 ONTAPボリュームはデータの管理コンテナです。ボリュームには、ファイル、LUN、さらにはS3オブジェクトが含まれている可能性があります。ボリュームはスペースを消費せず、格納されたデータの管理にのみ使用されます。

データベースボリュームの設計を作成する前に、SQL ServerのI/Oパターンと特性がワークロードやバックアップとリカバリの要件に応じてどのように変わるかを理解しておくことが重要です。フレキシブルボリュームについては、NetAppに関する次の推奨事項を参照してください。

- ホスト間でのボリュームの共有は避けてください。たとえば、1つのボリュームに2つのLUNを作成し、各LUNを別のホストで共有することは可能ですが、管理が複雑になる可能性があるため、この方法は避けてください。
- ドライブレターではなくNTFSマウントポイントを使用して、Windowsのドライブレターの制限（26文字）を超えます。ボリュームマウントポイントを使用する場合は、ボリュームラベルにマウントポイントと同じ名前を付けることを一般的に推奨します。
- 必要に応じて、スペース不足が発生しないようにボリュームのオートサイズポリシーを設定します。17 ONTAPを使用したMicrosoft SQL Serverのベストプラクティスガイド©2022 NetApp、Inc. 無断転載を禁じます。
- SQL ServerをSMB共有にインストールする場合は、フォルダを作成するためにSMB/CIFSボリュームでUnicodeが有効になっていることを確認してください。
- 運用面からの監視を容易にするために、ボリュームのスナップショット予約値をゼロに設定します。
- Snapshotスケジュールと保持ポリシーを無効にします。代わりに、SnapCenterを使用してSQL Serverデータボリュームのスナップショットコピーを調整します。
- SQL Serverシステムデータベースを専用ボリュームに配置します。
- tempdbは、特にI/O負荷の高いDBCC CHECKDB処理のために、SQL Serverが一時的なワークスペースとして使用するシステムデータベースです。したがって、このデータベースは、独立したスピンドルセットを持つ専用ボリュームに配置します。ボリューム数が課題となる大規模な環境では、慎重に計画を立てたあと、tempdbを少数のボリュームに統合し、他のシステムデータベースと同じボリュームに格納できます。tempdbのデータ保護は、SQL Serverを再起動するたびにこのデータベースが再作成されるため、優先度の高いものではありません。
- ランダムな読み取り/書き込みワークロードであるため、ユーザデータファイル（.mdf）を別 々 のボリュームに配置します。トランザクションログバックアップは、データベースバックアップよりも頻繁に作成するのが一般的です。そのため、トランザクションログファイル（.ldf）をデータファイルとは別のボリュームまたはVMDKに配置して、それぞれに個別のバックアップスケジュールを作成できるようにします。この分離により、ログファイルのシーケンシャルライトI/Oがデータファイルのランダムリード/ライトI/Oから分離され、SQL Serverのパフォーマンスが大幅に向上します。

LUN

- ユーザデータベースファイルとログバックアップを格納するログディレクトリが別 々 のボリュームにあることを確認して、SnapVaultテクノロジでSnapshotが使用されている場合に保持ポリシーによって上書きされないようにしてください。
- SQL Serverデータベースが、フルテキスト検索関連ファイルなど、データベース以外のファイルを持つLUNとは別のLUN上に存在することを確認します。
- データベースのセカンダリファイルを（ファイルグループの一部として）別 々 のボリュームに配置すると、SQL Serverデータベースのパフォーマンスが向上します。この分離は、データベースの.mdfファイルがLUNを他の.mdfファイルと共有していない場合にのみ有効です。
- DiskManagerなどのツールを使用してLUNを作成する場合は、LUNをフォーマットするときに、パーティションの割り当て単位サイズが64Kに設定されていることを確認してください。
- を参照してください "最新SANに対するONTAPのベストプラクティスに基づくMicrosoft Windowsとネイ

ティブMPIO" WindowsのマルチパスサポートをMPIOプロパティのiSCSIデバイスに適用するには、次の手順を実行します。

Microsoft SQL Serverデータベースファイルおよびファイルグループ

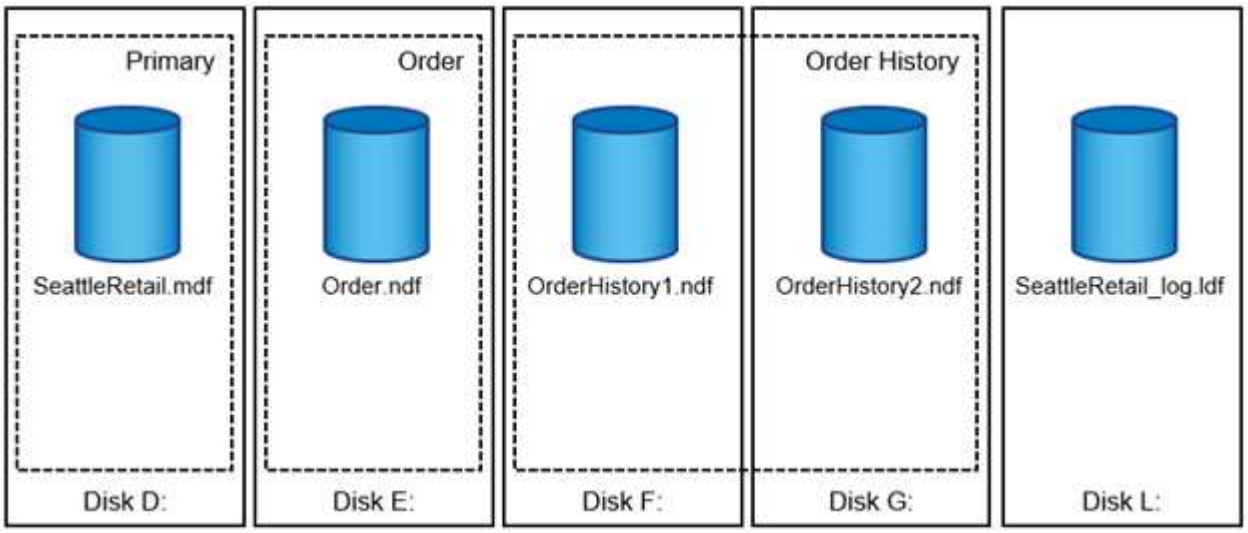
初期導入段階では、SQL ServerデータベースファイルをONTAPに適切に配置することが重要です。これにより、パフォーマンス、スペース管理、バックアップとリストアの最適な時間が確保され、ビジネス要件に合わせて設定できます。

理論的には、SQL Server（64ビット）ではインスタンスあたり32、767個のデータベースと524、272TBのデータベースサイズがサポートされますが、通常のインストールでは複数のデータベースが使用されます。ただし、SQL Serverで処理できるデータベースの数は、負荷とハードウェアによって異なります。SQL Serverインスタンスでは、数十、数百、場合によっては数千の小規模データベースをホストしていることも珍しくありません。

各データベースは、1つ以上のデータファイルと1つ以上のトランザクションログファイルで構成されます。トランザクションログには、データベーストランザクションに関する情報と、各セッションで行われたすべてのデータ変更が格納されます。データが変更されるたびに、SQL Serverはトランザクションログに十分な情報を格納して、アクションを元に戻す（ロールバックする）か、やり直す（再生する）かを指定します。SQL Serverトランザクションログは、データの整合性と堅牢性に関するSQL Serverの評価に不可欠な要素です。トランザクションログは、SQL Serverの不可分性、整合性、分離、耐久性（ACID）機能に不可欠です。SQL Serverは、データページが変更されるとすぐにトランザクションログに書き込みます。すべてのData Manipulation Language（DML）ステートメント（SELECT、INSERT、UPDATE、DELETEなど）は完全なトランザクションであり、トランザクションログによってセットベースの操作全体が確実に実行され、トランザクションの不可分性が保証されます。

各データベースには1つのプライマリデータファイルがあり、デフォルトでは.mdf拡張子が付いています。また、各データベースにセカンダリデータベースファイルを含めることもできます。これらのファイルには、デフォルトで.ndf拡張子が付いています。

すべてのデータベースファイルはファイルグループにグループ化されます。ファイルグループは論理ユニットであり、データベース管理を簡素化します。論理オブジェクトの配置と物理データベースファイルを分離できます。データベースオブジェクトテーブルを作成するときは、基になるデータファイルの設定を気にすることなく、配置するファイルグループを指定します。



ファイルグループ内に複数のデータファイルを配置できるため、複数のストレージデバイスに負荷を分散し

て、システムのI/Oパフォーマンスを向上させることができます。一方、SQL Serverはトランザクションログにシーケンシャルに書き込むため、トランザクションログには複数のファイルを使用するメリットはありません。

ファイルグループ内の論理オブジェクトの配置と物理データベースファイルの配置を分離することで、データベースファイルのレイアウトを微調整し、ストレージサブシステムを最大限に活用できます。たとえば、異なる顧客に自社製品を導入している独立系ソフトウェアベンダー（ISV）は、基盤となるI/O構成と導入段階で予想されるデータ量に基づいてデータベースファイルの数を調整できます。これらの変更は、データベースファイルではなくファイルグループにデータベースオブジェクトを配置するアプリケーション開発者には透過的です。

* NetAppでは、システムオブジェクト以外にプライマリファイルグループを使用しないことを推奨しています。ユーザオブジェクト用に別のファイルグループまたはファイルグループのセットを作成すると、特に大規模なデータベースの場合、データベースの管理とディザスタリカバリが容易になります。

データベースを作成するとき、または既存のデータベースに新しいファイルを追加するときに、初期ファイルサイズと自動拡張パラメータを指定できます。SQL Serverでは、Proportional Fill Algorithmを使用して、データを書き込むデータファイルを選択します。ファイルで使用可能な空きスペースに比例してデータ量が書き込まれます。ファイル内の空きスペースが多いほど、処理する書き込み数も多くなります。

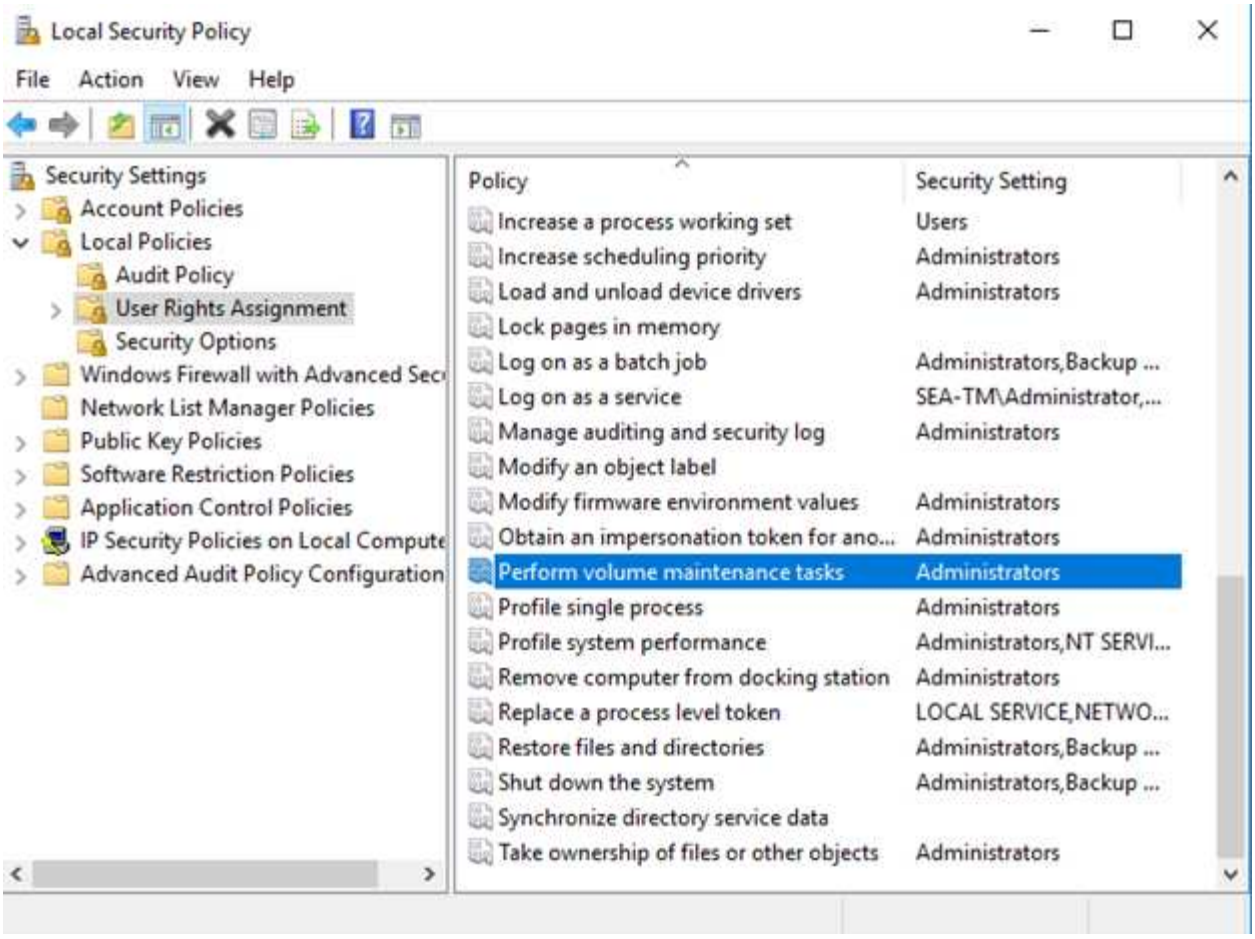
* NetAppでは、1つのファイルグループ内のすべてのファイルに同じ初期サイズと自動拡張パラメータを設定し、拡張サイズをパーセンテージではなくメガバイト単位で定義することを推奨しています*。これにより、Proportional Fill Algorithmは、データファイル間で書き込みアクティビティのバランスを均等に調整できます。

SQL Serverは、ファイルを拡張するたびに、新しく割り当てられたスペースをゼロでいっぱいにします。このプロセスは、対応するファイルへの書き込みが必要なすべてのセッションをブロックします。トランザクションログが増加した場合は、トランザクションログレコードを生成します。

SQL Serverは常にトランザクションログをゼロにし、その動作を変更することはできません。ただし、インスタントファイルの初期化を有効または無効にすることで、データファイルを初期化するかどうかを制御できます。インスタントファイルの初期化を有効にすると、データファイルの増加を高速化し、データベースの作成やリストアに必要な時間を短縮できます。

インスタントファイルの初期化には、わずかなセキュリティリスクが伴います。このオプションを有効にすると、データファイルの未割り当て部分に、以前に削除されたOSファイルの情報を含めることができます。データベース管理者はこのようなデータを調べることができます。

インスタントファイルの初期化を有効にするには、「ボリュームメンテナンスタスクの実行」とも呼ばれるSA_MANAGE_VOLUME_name権限をSQL Serverスタートアップアカウントに追加します。これは、次の図に示すように、ローカルセキュリティポリシー管理アプリケーション（secpol.msc）で実行できます。「Perform volume maintenance task」権限のプロパティを開き、SQL Serverスタートアップアカウントをユーザのリストに追加します。



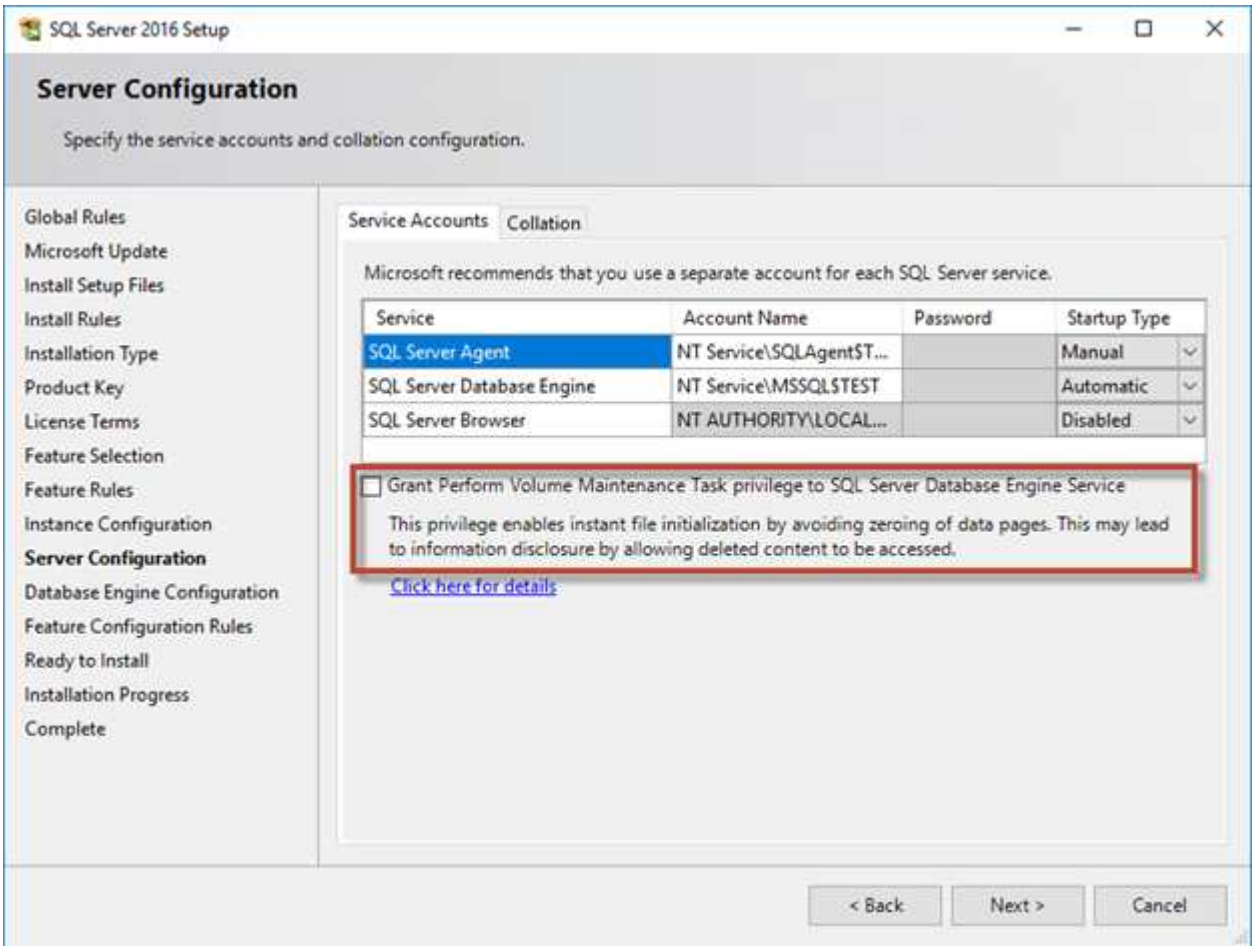
権限が有効になっているかどうかを確認するには、次の例のコードを使用します。このコードは、SQL Serverがエラーログに追加情報を書き込み、小さなデータベースを作成し、ログの内容を読み取るように強制する2つのトレースフラグを設定します。

```
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO
```

インスタントファイルの初期化が有効になっていない場合、次の例に示すように、SQL Serverのエラーログには、LDFログファイルの初期化に加えてMDFデータファイルが初期化されていることが示されます。インスタントファイルの初期化を有効にすると、ログファイルの初期化のみが表示されます。

	LogDate	ProcessInfo	Text
365	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 flush delta counts.
366	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 logging active xact info.
367	2017-02-09 08:10:07.750	spid53	Ckpt dbid 3 phase 1 ended (8)
368	2017-02-09 08:10:07.750	spid53	About to log Checkpoint end.
369	2017-02-09 08:10:07.880	spid53	Ckpt dbid 3 complete
370	2017-02-09 08:10:08.130	spid53	Starting up database 'DelMe'.
371	2017-02-09 08:10:08.150	spid53	FixupLogTail(progress) zeroing C:\Program Files\Micros
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQ
373	2017-02-09 08:10:08.170	spid53	Zeroing completed on C:\Program Files\Microsoft SQL
374	2017-02-09 08:10:08.710	spid53	Ckpt dbid 6 started
375	2017-02-09 08:10:08.710	spid53	About to log Checkpoint begin.

ボリュームメンテナンスタスクはSQL Server 2016では簡単に実行でき、インストールプロセス中にオプションとして提供されます。この図は、SQL Serverデータベースエンジンサービスにボリュームメンテナンスタスクを実行する権限を付与するオプションを示しています。



データベースファイルのサイズを制御するもう1つの重要なデータベースオプションは、自動縮小です。このオプションを有効にすると、SQL Serverはデータベースファイルを定期的に縮小してサイズを縮小し、オペレーティングシステムにスペースを解放します。この処理はリソースを大量に消費するため、新しいデータがシステムに入ってくるとデータベースファイルが再び拡張されるため、あまり有用ではありません。データベースで自動縮小を有効にしないでください。

Microsoft SQL Serverログディレクトリ

ログディレクトリは、トランザクションログバックアップデータをホストレベルで格納するためにSQL Serverで指定します。SnapCenterを使用してログファイルをバックアップする場合は、SnapCenterで使用される各SQL Serverホストに、ログバックアップを実行するようにホストログディレクトリを設定する必要があります。SnapCenter にはデータベースリポジトリがあるため、バックアップ、リストア、クローニングの処理に関連するメタデータは中央のデータベースリポジトリに格納されます。

ホストログディレクトリのサイズは、次のように計算されます。
ホストログディレクトリのサイズ=（（最大DB LDFサイズx日次ログ変更率%）x（Snapshot保持率）÷（1 LUNオーバーヘッドスペース%）
ホストログディレクトリのサイジング式では、LUNオーバーヘッドスペースが10%であることを前提としています。

ログディレクトリは専用のボリュームまたはLUNに配置します。ホストログディレクトリのデータ量は、バックアップのサイズとバックアップを保持する日数によって異なります。SnapCenterでは、SQL Serverホストごとに1つのホストログディレクトリのみが許可されます。ホストログディレクトリは、SnapCenter -→ホスト-→プラグインの設定で設定できます。

- NetAppでは、ホストログディレクトリに次のことを推奨しています*。
- ホストログディレクトリが、バックアップSnapshotデータを破損する可能性のある他のタイプのデータと共有されていないことを確認してください。
- マウントポイントをホストするLUNにユーザデータベースまたはシステムデータベースを配置しないでください。
- SnapCenterによるトランザクション・ログのコピー先となる専用のFlexVolボリューム上に、ホスト・ログ・ディレクトリを作成します。
- SnapCenterウィザードを使用してデータベースをNetAppストレージに移行し、データベースを有効な場所に格納できるようにすることで、SnapCenterのバックアップおよびリストア処理を正常に実行できます。移行プロセスはシステムの停止を伴うため、移行の実行中にデータベースを原因でオフラインにする可能性があることに注意してください。
- SQL Serverのフェイルオーバークラスタインスタンス（FCI）では、次の条件が満たされている必要があります。
  - フェイルオーバークラスタインスタンスを使用している場合は、ホストログディレクトリLUNがSnapCenter、バックアップ対象のSQL Serverインスタンスと同じクラスタグループ内のクラスタディスクリソースである必要があります。
  - フェイルオーバークラスタインスタンスを使用している場合は、SQL Serverインスタンスに関連付けられたクラスタグループに割り当てられた物理ディスククラスタリソースである共有LUNにユーザデータベースを配置する必要があります。

Microsoft SQL Server tempdbファイル

tempdbデータベースは大量に利用できます。ONTAPへのユーザデータベースファイルの最適な配置に加えて、tempdbデータファイルを変更して割り当ての競合を軽減

ページ競合は、SQL Serverが新しいオブジェクトを割り当てるために特別なシステムページに書き込む必要がある場合に、グローバル割り当てマップ（GAM）、共有グローバル割り当てマップ（SGAM）、またはペ

ージ空きスペース（PFS）ページで発生する可能性があります。ラッチは、メモリ内のこれらのページを保護（ロック）します。ビジー状態のSQL Serverインスタンスでは、tempdbのシステムページでラッチを取得するのに時間がかかることがあります。その結果、クエリの実行時間が長くなり、ラッチ競合と呼ばれます。tempdbデータファイルを作成する場合は、次のベストプラクティスを参照してください。

- 8コア以下の場合：tempdbデータファイル=コア数
- 8コアを超える場合：8個のtempdbデータファイル

次のスクリプト例は、8つのtempdbファイルを作成し、tempdbをマウントポイントに移動することで、tempdbを変更します。 `C:\MSSQL\tempdb` （SQL Server 2012以降）。

```
use master

go

-- Change logical tempdb file name first since SQL Server shipped with
logical file name called tempdev

alter database tempdb modify file (name = 'tempdev', newname =
'tempdev01');



-- Change location of tempdev01 and log file

alter database tempdb modify file (name = 'tempdev01', filename =
'C:\MSSQL\tempdb\tempdev01.mdf');

alter database tempdb modify file (name = 'templog', filename =
'C:\MSSQL\tempdb\templog.ldf');

GO

-- Assign proper size for tempdev01

ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'tempdev01', SIZE = 10GB );

ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'templog', SIZE = 10GB );

GO

-- Add more tempdb files

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev02', FILENAME =
N'C:\MSSQL\tempdb\tempdev02.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev03', FILENAME =
```

```
N'C:\MSSQL\tempdb\tempdev03.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev04', FILENAME =
N'C:\MSSQL\tempdb\tempdev04.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev05', FILENAME =
N'C:\MSSQL\tempdb\tempdev05.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev06', FILENAME =
N'C:\MSSQL\tempdb\tempdev06.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev07', FILENAME =
N'C:\MSSQL\tempdb\tempdev07.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev08', FILENAME =
N'C:\MSSQL\tempdb\tempdev08.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

GO
```

SQL Server 2016以降では、インストール時にオペレーティングシステムが認識できるCPUコアの数が自動的に検出され、その数に基づいて、最適なパフォーマンスを実現するために必要なtempdbファイルの数が計算および設定されます。

Microsoft SQL ServerとStorage Efficiency

ONTAPのStorage Efficiency機能は、消費するストレージスペースが最小限に抑えられ、システム全体のパフォーマンスにほとんど影響しないようにSQL Serverデータを格納、管理できるように最適化されています。

Storage Efficiencyは、RAID、プロビジョニング（全体的なレイアウトと利用率）、ミラーリング、その他のデータ保護テクノロジを組み合わせたものです。Snapshot、シンプロビジョニング、クローニングなどのNetAppテクノロジは、インフラ内の既存のストレージを最適化し、将来のストレージ支出を先送りまたは回避します。これらのテクノロジを組み合わせて使用するほど、削減効果が大きくなります。

圧縮、コンパクション、重複排除などのスペース効率化機能は、特定の量の物理ストレージに収まる論理データの量を増やすように設計されています。その結果、コストと管理オーバーヘッドが削減されます。

圧縮とは、大まかに言って、データのパターンを検出してスペースを削減する方法でエンコードする数学的プロセスです。一方、重複排除機能は、実際に繰り返されるデータブロックを検出し、不要なコピーを削除します。コンパクションを使用すると、複数の論理ブロックのデータをメディア上の同じ物理ブロックで共有できます。

Storage Efficiencyとフラクショナルリザベーションの連動については、シンプロビジョニングに関する以下のセクションを参照してください。

圧縮

オールフラッシュストレージシステムが登場する以前は、アレイベースの圧縮の価値は限られていまし

た。I/O負荷の高いワークロードのほとんどでは、許容可能なパフォーマンスを提供するために非常に多数のスピンドルが必要だったためです。ストレージシステムには、ドライブ数が多いことの副作用として、必要以上の容量が常に搭載されていました。この状況は、ソリッドステートストレージの登場によって変化しました。優れたパフォーマンスを得るためだけにドライブを過剰にオーバープロビジョニングする必要はもうありません。ストレージシステムのドライブスペースは、実際の容量ニーズに合わせて調整できます。

ソリッドステートドライブ（SSD）ではIOPSが向上するため、ほとんどの場合、回転式ドライブに比べてコストを削減できますが、圧縮を使用すると、ソリッドステートメディアの実効容量を増やすことで、さらに削減効果を高めることができます。

データを圧縮する方法はいくつかあります。多くのデータベースには独自の圧縮機能が搭載されていますが、お客様の環境ではこのような圧縮機能はほとんど見られません。その理由は、通常、圧縮データを*変更*するとパフォーマンスが低下することに加え、一部のアプリケーションではデータベースレベルの圧縮のライセンスコストが高くなることにあります。最後に、データベース処理のパフォーマンスが全体的に低下します。実際のデータベース作業ではなく、データの圧縮と解凍を実行するCPUに、高いCPU単位のライセンスコストを支払うことはほとんど意味がありません。より適切な方法は、圧縮処理をストレージシステムにオフロードすることです。

適応圧縮

アダプティブ圧縮は、レイテンシがマイクロ秒単位で測定されるオールフラッシュ環境であっても、パフォーマンスに影響を及ぼさないエンタープライズワークロードで徹底的にテストされています。一部のお客様から、圧縮機能によってデータがキャッシュ内で圧縮されたままになるため、パフォーマンスが向上したとの報告もあります。これは、コントローラで使用可能なキャッシュ容量が実質的に増加するためです。

ONTAPは物理ブロックを4KB単位で管理します。アダプティブ圧縮では、デフォルトの圧縮ブロックサイズである8KBが使用されます。つまり、データは8KB単位で圧縮されます。これは、リレーショナルデータベースで最もよく使用される8KBのブロックサイズに一致します。圧縮アルゴリズムは、より多くのデータが1つの単位として圧縮されるので、より効率的になります。圧縮ブロックサイズが32KBの場合、8KBの圧縮ブロックユニットよりもスペース効率に優れています。つまり、デフォルトの8KBのブロックサイズを使用するアダプティブ圧縮の場合、削減率はわずかに低くなりますが、圧縮ブロックサイズを小さくすることには大きなメリットがあります。データベースワークロードには、大量の上書きアクティビティが含まれています。32KBの圧縮されたデータブロックの8KBを上書きするには、32KBの論理データ全体を読み取って解凍し、必要な8KB領域を更新してから再度圧縮し、32KB全体をドライブに書き込む必要があります。この処理はストレージシステムでは非常にコストがかかります。このため、圧縮ブロックサイズの大きい競合ストレージアレイでも、データベースワークロードのパフォーマンスが大幅に低下します。

適応圧縮で使用されるブロックサイズは、最大32KBまで拡張できます。これによりストレージ効率が向上する可能性があります。このようなデータがアレイに大量に格納されている場合は、トランザクションログやバックアップファイルなどの静止ファイルについて検討する必要があります。状況によっては、適応圧縮のブロックサイズをそれに合わせて増やすことで、16KBまたは32KBのブロックサイズを使用するアクティブデータベースでもメリットが得られる場合があります。この方法がお客様のワークロードに適しているかどうかについては、NetAppまたはパートナーの担当者にお問い合わせください。

ストリーミングバックアップデスティネーションでは、重複排除と一緒に8KBを超える圧縮ブロックサイズを使用しないでください。これは、バックアップデータへのわずかな変更が32KBの圧縮ウィンドウに影響するためです。ウィンドウが移動すると、圧縮されたデータはファイル全体で異なります。重複排除は圧縮後に実行されます。つまり、重複排除エンジンは、圧縮された各バックアップを別 々 に認識します。ストリーミングバックアップの重複排除が必要な場合は、8KBのブロックアダプティブ圧縮のみを使用します。アダプティブ圧縮を使用することを推奨します。アダプティブ圧縮はブロックサイズが小さく、重複排除による効率化の妨げにならないためです。同様の理由から、ホスト側の圧縮も重複排除による効率化の妨げになります。

圧縮のアライメント

データベース環境でアダプティブ圧縮を使用する場合は、圧縮ブロックのアライメントについて考慮する必要があります。これは、非常に特定のブロックでランダムオーバーライトが発生するデータについてのみ考慮する必要があります。このアプローチは、ファイルシステム全体のアライメントと概念的に似ています。ファイルシステムの開始は4Kデバイスの境界に合わせて調整する必要があり、ファイルシステムのブロックサイズは4Kの倍数でなければなりません。

たとえば、ファイルへの8KBの書き込みは、ファイルシステム自体の8KBの境界にアライメントされている場合にのみ圧縮されます。これは、ファイルの最初の8KB、ファイルの2番目の8KBなどに配置する必要があることを意味します。アライメントを正しく行う最も簡単な方法は、正しいLUNタイプを使用することです。作成するパーティションには、デバイスの先頭から8Kの倍数のオフセットを設定し、データベースのブロックサイズの倍数のファイルシステムのブロックサイズを使用する必要があります。

バックアップやトランザクションログなどのデータは、複数のブロックにまたがるシーケンシャル書き込み処理であり、すべて圧縮されます。したがって、アライメントを考慮する必要はありません。問題となるI/Oパターンは、ファイルのランダムオーバーライトだけです。

データコンパクション

データコンパクションは、圧縮効率を向上させるテクノロジです。前述したように、アダプティブ圧縮では4KBのWAFLブロックに8KBのI/Oが格納されるため、削減率は最大でも2：1です。ブロックサイズが大きい圧縮方式では、効率性が向上します。ただし、小さなブロックの上書きが発生するデータには適していません。32KBのデータユニットを解凍して8KB部分を更新し、再度圧縮してからドライブにライトバックすると、オーバーヘッドが発生します。

データコンパクションでは、複数の論理ブロックを物理ブロック内に格納できます。たとえば、テキストブロックや部分的にフルブロックなど、圧縮率の高いデータを含むデータベースは、8KBから1KBに圧縮できます。コンパクションを使用しない場合、この1KBのデータが4KBブロック全体を占有します。インラインデータコンパクションでは、圧縮された1KBのデータを、他の圧縮データと一緒にわずか1KBの物理スペースに格納できます。これは圧縮テクノロジではありません。ドライブのスペースをより効率的に割り当てる方法なので、検出できるほどのパフォーマンスへの影響はありません。

得られる削減効果の程度はさまざまです。すでに圧縮または暗号化されているデータは、通常それ以上圧縮することはできないため、コンパクションによるメリットはありません。一方、初期化されたばかりのデータファイルで、ブロックメタデータとゼロブロックしか含まれていない場合は、最大80：1まで圧縮できます。

温度に基づくストレージ効率

ONTAP 9.8以降で利用できるTemperature Sensitive Storage Efficiency（TSSE）は、ブロックアクセスのヒートマップを利用してアクセス頻度の低いブロックを特定し、圧縮して効率を高めます。

重複排除

重複排除とは、データセットから重複するブロックサイズを削除することです。たとえば、同じ4KBブロックが10個のファイルに存在する場合、重複排除機能は、10個のファイルすべてのうち、その4KBブロックを同じ4KBの物理ブロックにリダイレクトします。その結果、そのデータの効率が10分の1に向上します。

VMwareゲストブートLUNなどのデータは、同じオペレーティングシステムファイルの複数のコピーで構成されるため、通常は重複排除が非常に効果的です。100:1以上の効率が観測されている。

一部のデータに重複データが含まれていません。たとえば、Oracleブロックには、データベースに対してグローバルに一意のヘッダーと、ほぼ一意のトレーラが含まれています。そのため、Oracleデータベースの重複排除によって1%以上の削減効果が得られることはほとんどありません。MS SQLデータベースでの重複排除はやや優れていますが、ブロックレベルでの固有のメタデータは依然として制限されています。

16KBでブロックサイズが大きいデータベースでは、最大15%のスペース削減効果が確認されたケースがいくつかあります。各ブロックの最初の4KBにはグローバルに一意なヘッダーが含まれ、最後の4KBブロックにはほぼ一意のトレーラが含まれます。内部ブロックは重複排除の対象となりますが、実際には、初期化されたデータの重複排除にほぼ完全に起因しています。

競合するアレイの多くは、データベースが複数回コピーされていると仮定して、データベースの重複排除機能があると主張しています。この点では、NetAppの重複排除も使用できますが、ONTAPにはNetApp FlexCloneテクノロジというより優れたオプションがあります。最終的な結果は同じで、基盤となる物理ブロックの大部分を共有するデータベースのコピーが複数作成されます。FlexCloneを使用すると、時間をかけてデータベースファイルをコピーしてから重複を排除するよりも、はるかに効率的です。重複は最初から作成されないため、実際には重複排除ではなく重複排除です。

効率性とシンプロビジョニング

効率化機能はシンプロビジョニングの一形態です。たとえば、100GBのボリュームを使用している100GBのLUNを50GBに圧縮するとします。ボリュームが100GBのままなので、実際の削減はまだ実現されていません。削減されたスペースをシステムの他の場所で使用できるように、まずボリュームのサイズを縮小する必要があります。100GBのLUNにあとから変更した結果、データの圧縮率が低下すると、LUNのサイズが大きくなり、ボリュームがいっぱいになる可能性があります。

シンプロビジョニングは、管理を簡易化しながら、使用可能な容量を大幅に改善し、コストを削減できるため、強く推奨されます。これは、単純なデータベース環境では、多くの場合、空のスペース、多数のボリュームやLUN、圧縮可能なデータが含まれているためです。シックプロビジョニングでは、ボリュームとLUNのストレージにスペースがリザーブされます。これは、100%フルになり、100%圧縮不可能なデータが含まれる場合に限られます。これは起こりそうもないことですシンプロビジョニングを使用すると、スペースを他の場所で再生して使用できます。また、容量の管理は、多数の小さいボリュームやLUNではなく、ストレージシステム自体に基づいて行うことができます。

一部のお客様は、特定のワークロードにシックプロビジョニングを使用するか、一般的には確立された運用と調達の手法に基づいてシックプロビジョニングを使用します。

*注意：*ボリュームがシックプロビジョニングされている場合は、ボリュームの圧縮解除や重複排除の削除など、そのボリュームのすべての効率化機能を完全に無効にするように注意する必要があります。 `sis undo` コマンドを実行しますボリュームが `volume efficiency show` 出力。有効になっている場合、ボリュームはまだ部分的に効率化機能用に設定されています。その結果、オーバーライトギャランティの動作が異なります。そのため、設定で原因が見落とされてボリュームのスペースが予期せず不足し、データベースI/Oエラーが発生する可能性が高くなります。

効率化のベストプラクティス

NetAppの推奨事項は次のとおりです。

AFFのデフォルト

オールフラッシュAFFシステムで実行されているONTAPで作成されたボリュームは、すべてのインライン効率化機能が有効になった状態でシンプロビジョニングされます。一般にデータベースには重複排除機能はなく、圧縮不可能なデータも含まれている可能性がありますが、デフォルト設定はほとんどすべてのワークロードに適しています。ONTAPは、あらゆる種類のデータとI/Oパターンを効率的に処理するように設計されており、削減効果があるかどうかは関係ありません。デフォルトは、理由が完全に理解されていて、逸脱するメリットがある場合にのみ変更する必要があります。

一般的な推奨事項

- ボリュームやLUNがシンプロビジョニングされていない場合は、すべての効率化設定を無効にする必要があります。これらの機能を使用しても削減は得られず、シックプロビジョニングとスペース効率化が有効になっていると、スペース不足エラーなどの予期しない動作が原因に発生する可能性があるためです。
- バックアップやデータベーストランザクションログなどでデータが上書きされない場合は、クーリング期間を短くしてTSSEを有効にすることで、効率を高めることができます。
- アプリケーションレベルで圧縮がすでに有効になっているファイルが暗号化されている場合など、一部のファイルには圧縮不可能なデータが大量に含まれていることがあります。上記のいずれかに該当する場合は、圧縮可能なデータを含む他のボリュームでより効率的に処理できるように、圧縮を無効にすることを検討してください。
- データベースバックアップでは、32KBの圧縮機能と重複排除機能の両方を使用しないでください。を参照してください [適応圧縮] を参照してください。

データベース圧縮

SQL Server自体には、データを圧縮して効率的に管理する機能もあります。SQL Serverでは現在、行圧縮とページ圧縮の2種類のデータ圧縮がサポートされています。

行圧縮を使用すると、データストレージ形式が変更されます。たとえば、整数と小数を、ネイティブの固定長形式ではなく可変長形式に変更します。また、空白スペースを排除することで、固定長の文字列を可変長形式に変更します。ページ圧縮では、行圧縮と他の2つの圧縮方式（プレフィックス圧縮とディクショナリ圧縮）が実装されます。ページ圧縮の詳細については、 "ページ圧縮の実装"。

データ圧縮は現在、SQL Server 2008以降のEnterprise、Developer、およびEvaluationエディションでサポートされています。圧縮はデータベース自体で実行できますが、SQL Server環境ではほとんど実行されません。

SQL Serverデータファイルのスペース管理の推奨事項は次のとおりです。

- SQL Server環境でシンプロビジョニングを使用すると、スペース利用率を向上し、スペースギャランティ機能を使用する場合に必要なストレージ全体を削減できます。
- ストレージ管理者が監視する必要があるのはアグリゲート内のスペース使用量だけであるため、一般的な構成では自動拡張を使用します。
- バックアップから単一ボリュームへのデータベースのリストアなど、同じデータのコピーがボリュームに複数含まれていることがわかっている場合を除き、SQL Serverデータファイルを含むボリュームで重複排除を有効にしないことを推奨します。

スペース再生

スペース再生は、LUN内の未使用スペースをリカバリするために定期的に開始できます。SnapCenterでは、次のPowerShellコマンドを使用してスペース再生を開始できます。

```
Invoke-SdHostVolumeSpaceReclaim -Path drive_path
```

スペース再生を実行する必要がある場合は、最初にホストのサイクルを消費するため、アクティビティが少ない時間帯にこのプロセスを実行する必要があります。

NetApp管理ソフトウェアによるMicrosoft SQL Serverデータ保護

データベースのバックアップは、ビジネス要件に基づいて計画します。ONTAPのNetApp Snapshotテクノロジを組み合わせ、Microsoft SQL Server APIを活用することで、ユーザデータベースのサイズに関係なく、アプリケーションと整合性のあるバックアップを迅速に作成できます。より高度なデータ管理やスケールアウトデータ管理の要件に対応するために、NetAppはSnapCenterを提供しています。

SnapCenter

SnapCenterは、エンタープライズアプリケーション向けのNetAppデータ保護ソフトウェアです。SnapCenter Plug-in for SQL Serverや、SnapCenter Plug-in for Microsoft Windowsで管理されるOS処理を使用して、SQL Serverデータベースを迅速かつ簡単に保護できます。

SQL Serverインスタンスは、スタンドアロンセットアップ、フェイルオーバークラスタインスタンス、または常時稼働の可用性グループにすることができます。その結果、データベースの保護、クローニング、リストアをプライマリコピーまたはセカンダリコピーから単一コンソールで実行できます。SnapCenterでは、SQL Serverデータベースをオンプレミス、クラウド、ハイブリッド構成の両方で管理できます。また、開発やレポート作成のために、元のホストまたは代替ホストに数分でデータベースコピーを作成することもできます。

* NetAppでは* SnapCenterを使用してSnapshotコピーを作成することを推奨しています。以下に示すT-SQL方式も機能しますが、SnapCenterでは、バックアップ、リストア、クローニングのプロセスを完全に自動化できます。また、検出を実行して、正しいSnapshotが作成されていることを確認します。事前設定は必要ありません。
な…何だ?
また、SQL Serverでは、作成時にSnapshotに正しいデータが存在するように、OSとストレージの間で調整を行う必要があります。ほとんどの場合、これを行う唯一の安全な方法は、SnapCenterまたはT-SQLを使用することです。この追加の調整なしで作成されたSnapshotは、確実にリカバリできない可能性があります。

SQL Server Plug-in for SnapCenterの詳細については、を参照してください。 "TR-4714：『Best Practice Guide for SQL Server using NetApp SnapCenter』"。

T-SQLスナップショットを使用したデータベースの保護

SQL Server 2022では、MicrosoftがT-SQLスナップショットを導入しました。これにより、バックアップ処理のスクリプト作成と自動化を行うことができます。フルサイズのコピーを実行する代わりに、Snapshot用に

データベースを準備できます。データベースのバックアップ準備が完了したら、ONTAP REST APIを使用してSnapshotを作成できます。

次に、バックアップワークフローの例を示します。

1. ALTERコマンドを使用してデータベースをフリーズします。これにより、基盤となるストレージ上で整合性のあるSnapshotを作成するためのデータベースが準備されます。フリーズ後、backupコマンドを使用してデータベースをフリーズ解除し、スナップショットを記録できます。
2. 新しいbackup groupコマンドとbackup serverコマンドを使用して、ストレージボリューム上の複数のデータベースのスナップショットを同時に実行します。
3. フルバックアップまたはCOPY_ONLYフルバックアップを実行します。これらのバックアップもmsdbに記録されます。
4. スナップショットフルバックアップ後に通常のストリーミング方式で作成されたログバックアップを使用して、ポイントインタイムリカバリを実行します。必要に応じて、ストリーミング差分バックアップもサポートされます。

詳細については、を参照してください "T-SQLスナップショットについて知るためのMicrosoftのドキュメント"。

ONTAPを使用したMicrosoft SQL Serverディザスタリカバリ

エンタープライズデータベースやアプリケーションインフラでは、自然災害や予期しないビジネスの中断からダウンタイムを最小限に抑えて保護するために、レプリケーションが必要になることがよくあります。

SQL Server Always-On可用性グループレプリケーション機能は優れたオプションであり、NetAppはデータ保護とAlways-Onを統合するオプションを提供します。ただし、ONTAPレプリケーションテクノロジを検討する必要がある場合もあります。MetroClusterやSnapMirrorなどのONTAPレプリケーションオプションを使用すると、パフォーマンスへの影響を最小限に抑えながら拡張性が向上し、SQL以外のデータを保護できます。また、一般にインフラ全体のレプリケーションとDR解決策が提供されます。

SnapMirror非同期

SnapMirrorテクノロジは、LANおよびWAN経由でデータをレプリケートするための、高速で柔軟な非同期エンタープライズ解決策を提供します。最初のミラーリングの作成後は、変更されたデータブロックのみがデスティネーションに転送されるため、必要なネットワーク帯域幅が大幅に削減されます。

SnapMirror for SQL Serverの推奨事項は次のとおりです。

- CIFSを使用する場合は、デスティネーションSVMがソースSVMと同じActive Directoryドメインのメンバーである必要があります。これにより、NASファイルに格納されているアクセス制御リスト（ACL）が災害からのリカバリ時に破損しないようになります。
- ソースボリューム名と同じデスティネーションボリューム名を使用する必要はありませんが、デスティネーションボリュームをデスティネーションにマウントするプロセスを管理しやすくすることができます。CIFSを使用する場合は、デスティネーションNASネームスペースをソースネームスペースとパスおよびディレクトリ構造で同一にする必要があります。
- 整合性を確保するために、コントローラからのSnapMirror更新のスケジュールを設定しないでください。代わりに、フルバックアップまたはログバックアップの完了後にSnapCenterからのSnapMirror更新を有効にしてSnapMirrorを更新します。

- SQL Serverデータを含むボリュームをクラスタ内の複数のノードに分散して、すべてのクラスタノードでSnapMirrorレプリケーションアクティビティを共有できるようにします。この分散により、ノードリソースの使用が最適化されます。

SnapMirrorの詳細については、を参照してください。 "TR-4015：『SnapMirrorの設定およびベストプラクティスガイド- ONTAP 9』"。

ONTAP上のMicrosoft SQL Serverのセキュリティ保護

SQL Serverデータベース環境のセキュリティ保護は、データベース自体の管理にとどまらない多次元的な取り組みです。ONTAPには、データベースインフラのストレージを保護するために設計された独自の機能がいくつか用意されています。

Snapshot コピー

ストレージスナップショットは、ターゲットデータのポイントインタイムレプリカです。ONTAPには、さまざまなポリシーを設定し、ボリュームあたり最大1024個のSnapshotを格納する機能が実装されています。ONTAPのSnapshotはスペース効率に優れています。スペースが消費されるのは、元のデータセットが変更されたときだけです。また、読み取り専用です。Snapshotは削除できますが、変更することはできません。

場合によっては、ONTAPで直接Snapshotのスケジュールを設定できます。また、スナップショットを作成する前に、SnapCenterなどのソフトウェアでアプリケーションやOSの処理をオーケストレーションしなければならない場合もあります。ワークロードに最適なアプローチにかかわらず、アグレッシブなスナップショット戦略を使用すると、ブートLUNからミッションクリティカルなデータベースまで、すべてのバックアップに頻繁かつ簡単にアクセスできるため、データのセキュリティを確保できます。

注：ONTAPフレキシブルボリューム、つまり、ボリュームはLUNと同義ではありません。ボリュームは、ファイルやLUNなどのデータ用の管理コンテナです。たとえば、データベースを8 LUNのストライプセットに配置し、すべてのLUNを1つのボリュームに格納するとします。

スナップショットの詳細については、 "こちらをご覧ください。"

スナップショットの改ざん防止

ONTAP 9.12.1以降では、Snapshotは読み取り専用であるだけでなく、偶発的または意図的な削除からも保護できます。この機能は改ざん防止スナップショットと呼ばれます。保持期間は、Snapshotポリシーを使用して設定および適用できます。作成されたスナップショットは、有効期限に達するまで削除できません。管理またはサポートセンターのオーバーライドはありません。

これにより、侵入者、悪意のある内部者、さらにはランサムウェア攻撃さえも、バックアップが原因でONTAPシステム自体にアクセスできたとしても、バックアップを侵害することはできません。頻繁なSnapshotスケジュールと組み合わせることで、非常に強力なデータ保護と非常に低いRPOを実現できます。

改ざん防止スナップショットの詳細については、 "こちらをご覧ください。"

SnapMirror レプリケーション

Snapshotはリモートシステムにレプリケートすることもできます。これには改ざん防止Snapshotも含まれます。このSnapshotでは、リモートシステムに保持期間が適用され、適用されます。その結果、ローカ

ルSnapshotと同じデータ保護のメリットが得られますが、データは2つ目のストレージアレイに配置されます。これにより、元のアレイを破棄してもバックアップが損なわれることはありません。

2つ目のシステムでは、管理セキュリティのための新しいオプションも開きます。たとえば、NetAppのお客様によっては、プライマリストレージシステムとセカンダリストレージシステムの認証クレデンシャルを分離している場合があります。1人の管理ユーザが両方のシステムにアクセスできることはないため、悪意のある管理者がデータのすべてのコピーを削除することはできません。

SnapMirrorの詳細については、 "こちらをご覧ください。"

Storage Virtual Machine

新しく構成したONTAPストレージシステムは、新しくプロビジョニングしたVMware ESXサーバと似ています。これは、仮想マシンが作成されるまで、どちらもユーザをサポートできないためです。ONTAPでは、ストレージ管理の最も基本的な単位となるStorage Virtual Machine（SVM）を作成します。各SVMには、独自のストレージリソース、プロトコル構成、IPアドレス、FCP WWNがあります。 これがONTAPマルチテナンシーの基盤です。

たとえば、重要な本番環境のワークロード用に1つのSVMを設定し、開発アクティビティ用にもう1つのSVMを別のネットワークセグメントに設定できます。これにより、本番用SVMへのアクセスを特定の管理者に制限し、開発者は開発用SVM内のストレージリソースをより広範に制御できるようになります。また、特に重要な目に見えるデータを格納するために、財務チームや人事チームに3つ目のSVMを用意しなければならない場合もあります。

SVMの詳細については、 "こちらをご覧ください。"

カンリRBAC

ONTAPには、管理者ログイン用の強力なロールベースアクセス制御（RBAC）が用意されています。クラスタへのフルアクセスが必要な管理者もいれば、特定のSVMへのアクセスのみが必要な管理者もいます。高度なヘルプデスク担当者は、ボリュームのサイズを増やす機能を必要とする場合があります。その結果、管理者ユーザに、職務を遂行するために必要なアクセス権を付与するだけでなく、それ以上のアクセス権を付与することができます。さらに、さまざまなベンダーのPKIを使用してこれらのログインを保護し、sshキーのみへのアクセスを制限し、失敗したログイン試行のロックアウトを強制できます。

管理アクセス制御の詳細については、 "こちらをご覧ください。"

マルチファクタ認証

ONTAPおよびその他の一部のNetApp製品では、さまざまな方法を使用した多要素認証（MFA）がサポートされるようになりました。その結果、ユーザー名/パスワードだけでは、FOBやスマートフォンアプリなどの2番目の要因からのデータがないセキュリティスレッドではありません。

詳細については、をクリックしてください "こちらをご覧ください。"

API RBAC

自動化にはAPI呼び出しが必要ですが、すべてのツールにフル管理アクセスが必要となるわけではありません。自動化システムを保護するために、APIレベルでRBACを使用することもできます。自動化ユーザアカウントは、必要なAPI呼び出しだけに制限できます。たとえば、監視ソフトウェアには変更アクセスは必要なく、読み取りアクセスのみが必要です。ストレージをプロビジョニングするワークフローでは、ストレージを削除する機能は必要ありません。

詳細については、c here.

マルチ管理者認証（**MAV**）

特定のアクティビティを承認するために、それぞれ独自のクレデンシャルを持つ2人の異なる管理者を要求することで、多要素認証をさらに進めることができます。これには、ログイン権限の変更、診断コマンドの実行、データの削除が含まれます。

マルチ管理者認証（MAV）の詳細については、 "こちらをご覧ください"

MySQL

ONTAPノMySQLテタヘス

MariaDBやPercona MySQLを含むMySQLとその変種は、世界で最も人気のあるデータベースです。

ONTAPとMySQLデータベースに関するこのドキュメントは、以前に公開されていた_TR-4722：『MySQL database on ONTAP best practices』に代わるものです。_

ONTAPは文字通りデータベース向けに設計されているため、ONTAPはMySQLデータベースに最適なプラットフォームです。データベースワークロードのニーズに対応するために、高度なQuality of Service（QoS；サービス品質）機能や基本的なFlexClone機能に対するランダムI/Oレイテンシの最適化など、多数の機能が特別に開発されました。

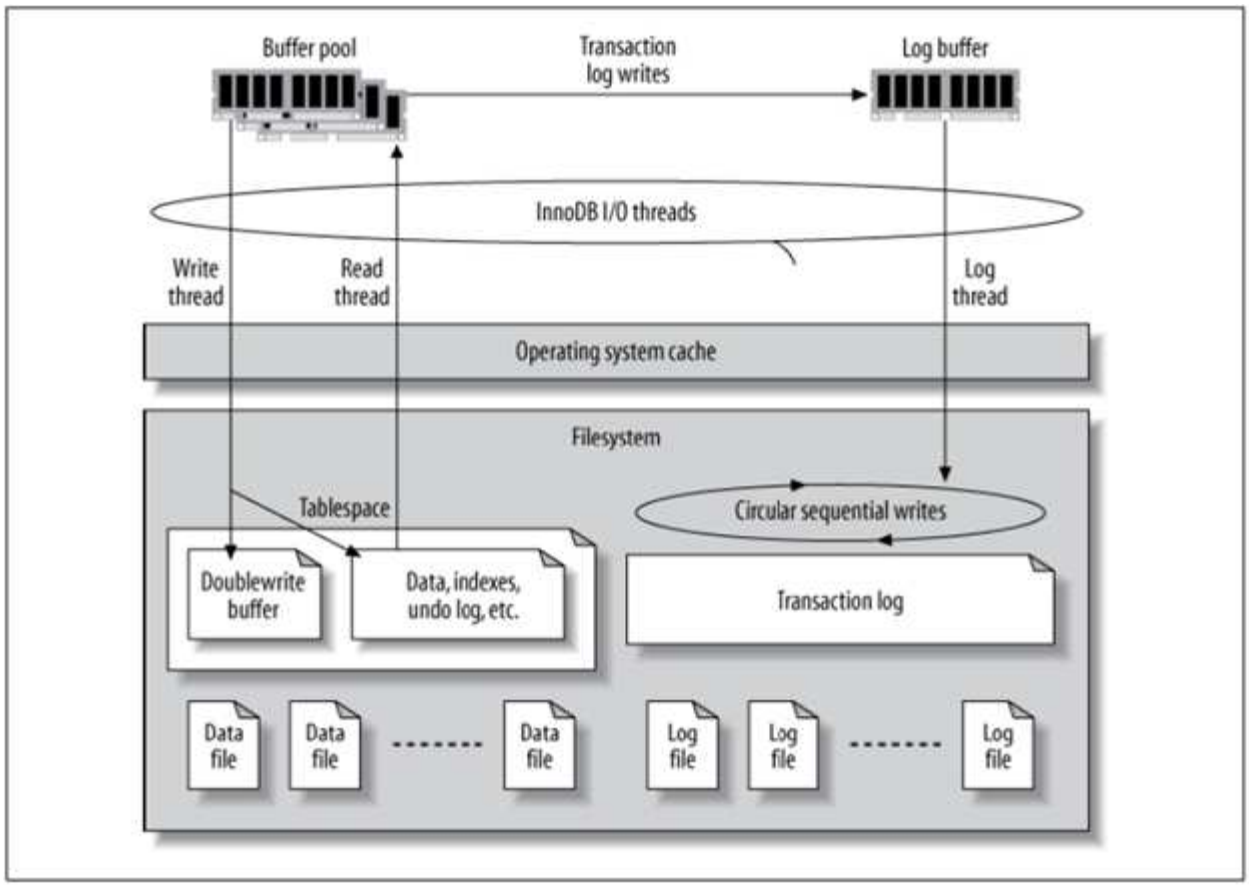
無停止アップグレード（ストレージの交換など）などの追加機能により、重要なデータベースの可用性を維持できます。また、MetroClusterを使用して大規模な環境で瞬時にディザスタリカバリを実行したり、SnapMirrorアクティブ同期を使用してデータベースを選択したりすることもできます。

最も重要なことは、ONTAPが卓越したパフォーマンスを提供し、お客様固有のニーズに合わせて解決策をサイジングできることです。ネットアップのハイエンドシステムは100万超のIOPSをマイクロ秒単位のレイテンシで提供できますが、必要なIOPSが10万だけの場合は、同じストレージオペレーティングシステムを実行する小型のコントローラでストレージ解決策を適切にサイジングできます。

データベース設定

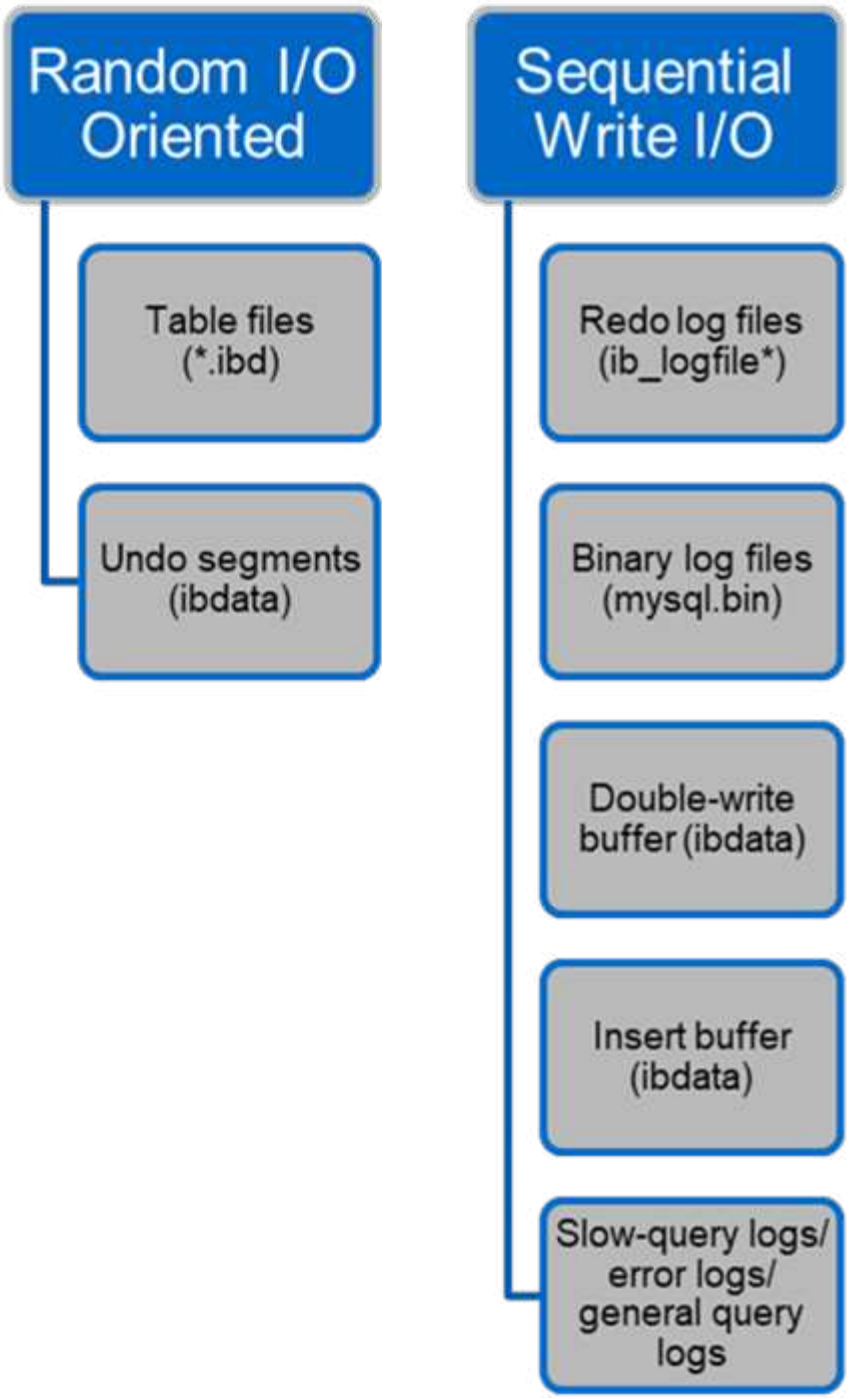
MySQLとInnoDB

InnoDBはストレージとMySQLサーバの中間層として機能し、データをドライブに格納します。



MySQL I/Oは次の2つのタイプに分類されます。

- ランダムファイルI/O
- シーケンシャルファイルI/O



データファイルはランダムに読み取りおよび上書きされるため、IOPSが高くなります。そのため、SSDストレージを推奨します。

REDOログファイルとバイナリログファイルはトランザクションログです。それらはシーケンシャルに書き込まれるので、書き込みキャッシュを備えたHDDで優れたパフォーマンスを得ることができます。リカバリ時にシーケンシャルリードが発生しますが、ログファイルのサイズは通常データファイルより小さく、シーケンシャルリードはランダムリード（データファイルで発生）よりも高速であるため、パフォーマンスの問題が発生することはほとんどありません。

ダブル書き込みバッファはInnoDBの特別な機能です。InnoDBは、最初にフラッシュされたページをダブルライトバッファに書き込み、次にページをデータファイル上の正しい位置に書き込みます。このプロセスによ

り、ページの破損が防止されます。二重書き込みバッファがない場合、ドライブへの書き込みプロセス中に電源障害が発生すると、ページが破損する可能性があります。ダブル書き込みバッファへの書き込みはシーケンシャルであるため、HDD向けに高度に最適化されています。リカバリ時にシーケンシャルリードが発生します。

ONTAP NVRAMはすでに書き込み保護を提供しているため、ダブル書き込みバッファは必要ありません。MySQLにはパラメータがあります。 `skip_innodb_doublewrite`、ダブルライトバッファをディセーブルにします。この機能により、パフォーマンスが大幅に向上します。

挿入バッファはInnoDBの特別な機能でもあります。一意でないセカンダリインデックスブロックがメモリ内にない場合、InnoDBはエントリを挿入バッファに挿入して、ランダムなI/O操作を回避します。定期的に、挿入バッファはデータベース内のセカンダリインデックスツリーにマージされます。挿入バッファは、I/O要求を同じブロックにマージすることでI/O処理数を削減します。ランダムI/O処理はシーケンシャルです。また、インサートバッファはHDD用に高度に最適化されています。シーケンシャルライトと読み取りは、どちらも通常運用時に発生します。

元に戻すセグメントは、ランダムI/Oです。Multi-Version Concurrency（MVCC）を保証するために、InnoDBは元に戻すセグメントに古いイメージを登録する必要があります。元に戻すセグメントから以前の画像を読み取るには、ランダムな読み取りが必要です。繰り返し実行可能な読み取りで長時間のトランザクション（mysqldump—単一トランザクションなど）を実行したり、長時間のクエリを実行したりすると、ランダムリードが発生する可能性があります。したがって、この場合、元に戻すセグメントをSSDに保存する方が適しています。短いトランザクションまたはクエリのみを実行する場合、ランダムリードは問題ではありません。

- NetAppでは、InnoDBのI/O特性を考慮して、以下のストレージ設計レイアウトを推奨しています。
- MySQLのランダムI/OおよびシーケンシャルI/O指向ファイルを1つのボリュームに格納
- 純粋にシーケンシャルなI/O指向のMySQLファイルを格納するための別のボリューム

このレイアウトは、データ保護のポリシーと戦略の設計にも役立ちます。

MySQLセツテイハラメエタ

NetAppでは、最適なパフォーマンスを得るために、いくつかの重要なMySQL構成パラメータを推奨しています。

パラメータ	値
innodb_log_file_size	2億5、600万
innodb_flush_log_at_trx_commit	2.
innodb_doublewrite	0
innodb_flush_method	fsync
innodb_buffer_pool_size	11g
innodb_io_capacity	8、192です
innodb_buffer_pool_instances	8
innodb_lru_scan_depth	8、192です
open_file_limit	65535

このセクションで説明するパラメータを設定するには、MySQL構成ファイル（my.cnf）でパラメータを変更する必要があります。NetAppのベストプラクティスは、社内で実施したテストの結果です。

innodb_log_file_size

InnoDBログファイルのサイズに適したサイズを選択することは、書き込み処理およびサーバクラッシュ後の適切なリカバリ時間を確保するために重要です。

ファイルに記録されるトランザクションの数が非常に多いため、書き込み処理ではログファイルのサイズが重要になります。レコードが変更されても、変更はすぐにはテーブルスペースに書き込まれません。代わりに、変更はログファイルの最後に記録され、ページはダーティとしてマークされます。InnoDBはログを使用してランダムI/Oをシーケンシャル I/Oに変換します。

ログがいっぱいになると、ダーティページが順番にテーブルスペースに書き込まれ、ログファイルのスペースが解放されます。たとえば、トランザクションの途中でサーバがクラッシュし、書き込み処理のみがログファイルに記録されるとします。サーバを再び稼働させるには、ログファイルに記録された変更が再生されるリカバリフェーズを実行する必要があります。ログファイル内のエントリ数が多いほど、サーバのリカバリにかかる時間が長くなります。

この例では、ログファイルのサイズがリカバリ時間と書き込みパフォーマンスの両方に影響します。ログファイルのサイズに適切な数を選択する場合は、リカバリ時間と書き込みパフォーマンスのバランスを取ります。通常、128Mから512Mの間の値は良い値です。

innodb_flush_log_at_trx_commit

データに変更があっても、変更はすぐにはストレージに書き込まれません。

代わりに、データはログバッファに記録されます。これは、InnoDBがログファイルに記録されるバッファ変更に割り当てるメモリの一部です。InnoDBは、トランザクションがコミットされたとき、バッファがいっぱいになったとき、または1秒に1回のイベントが発生したときに、バッファをログファイルにフラッシュします。このプロセスを制御する構成変数は、Innodb_flush_log_at_trx_commitです。値オプションには次のものがあります。

- 設定時 `innodb_flush_log_trx_at_commit=0`では、（InnoDBバッファプール内の）変更されたデータがログファイル（ib_logfile）に書き込まれ、ログファイルが1秒ごとにフラッシュされます（ストレージへの書き込み）。ただし、トランザクションがコミットされても何も実行されません。停電またはシステムクラッシュが発生した場合、フラッシュされていないデータはログファイルまたはドライブに書き込まれないため、リカバリできません。
- 設定時 `innodb_flush_log_trx_commit=1`InnoDBはログバッファをトランザクションログに書き込み、トランザクションごとに永続的なストレージにフラッシュします。たとえば、すべてのトランザクションコミットについて、InnoDBはログに書き込み、その後ストレージに書き込みます。ストレージの速度が遅いと、パフォーマンスが低下します。たとえば、1秒あたりのInnoDBトランザクション数が減少します。
- 設定時 `innodb_flush_log_trx_commit=2`InnoDBはコミットのたびにログバッファをログファイルに書き込みますが、ストレージにデータを書き込むことはありません。InnoDBは、1秒に1回データをフラッシュします。停電やシステムクラッシュが発生した場合でも、オプション2のデータはログファイルに保存され、リカバリ可能です。

パフォーマンスが主な目標である場合は、値を2に設定します。InnoDBは、トランザクションコミットごとではなく、1秒に1回ドライブに書き込みを行うため、パフォーマンスが大幅に向上します。停電やクラッシュが発生した場合は、トランザクションログからデータをリカバリできます。

データの安全性が主な目的である場合は、トランザクションコミットごとにInnoDBがドライブにフラッシュされるように、値を1に設定します。ただし、パフォーマンスに影響する可能性があります。

* NetAppでは、パフォーマンスを向上させるために、innodb_flush_log_trx_commit値を2に設定することを推奨しています。

innodb_doublewrite

いつ `innodb_doublewrite` イネーブル(デフォルト)では、InnoDBはすべてのデータを2回格納します。最初にダブルライトバッファに格納し、次に実際のデータファイルに格納します。

このパラメータは、次のコマンドで無効にできます。 `--skip-innodb_doublewrite` ベンチマークの場合や、データの整合性や障害の可能性よりも最高のパフォーマンスに関心がある場合。InnoDBでは、ダブルライトと呼ばれるファイルフラッシュ技術が使用されています。InnoDBは、ページをデータファイルに書き込む前に、ダブル書き込みバッファと呼ばれる連続領域にページを書き込みます。書き込みとダブル書き込みバッファへのフラッシュが完了すると、InnoDBはページをデータファイル内の適切な位置に書き込みます。ページの書き込み中にオペレーティングシステムまたはmysqldプロセスがクラッシュした場合、InnoDBは後でクラッシュリカバリ中にダブルライトバッファからページの適切なコピーを見つけることができます。

* NetAppでは、ダブル書き込みバッファを無効にすることを推奨しています。ONTAP NVRAMは同じ機能を果たします。ダブルバッファリングは不必要にパフォーマンスを低下させます。

innodb_buffer_pool_size

InnoDBバッファプールは、チューニングアクティビティの中で最も重要な部分です。

InnoDBは、インデックスのキャッシュとデータのローイング、アダプティブハッシュインデックス、挿入バッファ、および内部で使用される他の多くのデータ構造をバッファプールに大きく依存しています。また、データへの変更もバッファに格納されるため、書き込み処理をストレージに対してすぐに実行する必要はありません。これにより、パフォーマンスが向上します。バッファプールはInnoDBの不可欠な部分であり、それに応じてサイズを調整する必要があります。バッファプールサイズを設定するときは、次の点を考慮してください。

- 専用のInnoDB専用マシンの場合は、バッファプールサイズを使用可能なRAMの80%以上に設定します。
- MySQL専用サーバーでない場合は、サイズをRAMの50%に設定します。

innodb_flush_method

innodb_flush_methodパラメータは、InnoDBがログファイルとデータファイルを開いてフラッシュする方法を指定します。

最適化

InnoDB最適化では、このパラメータを設定すると、必要に応じてデータベースのパフォーマンスが調整されます。

次のオプションは、InnoDBを使用してファイルをフラッシュするためのものです。

- `fsync`。InnoDBでは、 `fsync()` データファイルとログファイルの両方をフラッシュするシステムコール。このオプションはデフォルト設定です。
- `O_DSYNC`。InnoDBでは、 `O_DSYNC` ログファイルを開いてフラッシュし、データファイルをフラッシュするにはfsync()を使用します。InnoDBが使用しない `O_DSYNC` 直接、なぜなら、多くの種類のUNIXで問題があったからです。
- `O_DIRECT`。InnoDBでは、 `O_DIRECT` オプション（または `directio()` （Solarisの場合）データファイルを開くには、次のコマンドを使用します。 `fsync()` データファイルとログファイルの両方をフラッシュします。このオプションは、一部のGNU/Linuxバージョン、FreeBSD、Solarisで利用できます。
- `O_DIRECT_NO_FSYNC`。InnoDBでは、 `O_DIRECT` オプションはI/Oのフラッシュ時に使用されますが、 `fsync()` 後でシステムコール。このオプションは、一部のタイプのファイルシステム（XFSなど）には適していません。ファイルシステムに `fsync()` システムコール（たとえば、すべてのファイルメタデータを保持する場合）は、 `O_DIRECT` 代わりにオプション。

観察

NetAppラボテストでは、 `fsync` NFSとSANではdefaultオプションが使用されており、 `O_DIRECT`。flushメソッドを使用している場合 `O_DIRECT` ONTAPでは、クライアントが4096ブロックの境界で大量のシングルバイト書き込みをシリアル方式で書き込みました。この書き込みにより、ネットワーク経由のレイテンシが増加し、パフォーマンスが低下します。

innodb_io_capacity

InnoDBプラグインでは、MySQL 5.7からInnoDB_io_capacityという新しいパラメータが追加されました。

InnoDBが実行する最大IOPS（ダーティページのフラッシュレートと挿入バッファ[ibuf]バッチサイズを含む）を制御します。innodb_io_capacityパラメータは、バッファプールからのページのフラッシュや変更バッファからのデータのマージなど、InnoDBバックグラウンドタスクによるIOPSの上限を設定します。

innodb_io_capacityパラメータに、システムが1秒あたりに実行できるI/O処理のおおよその数を設定します。理想的には、設定をできるだけ低くしてくださいが、バックグラウンドアクティビティが遅くなるように低くしないでください。この設定が高すぎると、データがバッファプールから削除され、キャッシュするにはバッファの挿入が早すぎて大きなメリットが得られません。

* NetAppでは*この設定をNFSで使用する場合は、IOPSのテスト結果（SysBench / Fio）を分析し、それに応じてパラメータを設定することを推奨します。InnoDBバッファプールで必要以上に変更されたページやダーティなページが表示されない場合は、フラッシュとパージにできるだけ小さい値を使用してください。

ワークロードに対して低い値では不十分であることを証明した場合を除き、20、000以上などの極端な値は使用しないでください。

innodb_io_capacityパラメータは、フラッシュ速度と関連するI/Oを調整します。

このパラメータまたはinnodb_io_capacity_maxパラメータを設定すると、パフォーマンスに深刻な影響を与える可能性があります。このパラメータが高すぎて、早期フラッシュでI/O処理が無駄になります。

innodb_lru_scan_depth

。 `innodb_lru_scan_depth` パラメータは、InnoDBバッファプールのフラッシュ操作のアルゴリズムとヒューリスティックに影響を与えます。

このパラメータは、I/O負荷の高いワークロードを調整するパフォーマンスのエキスパートが主に関心を持っています。このパラメータは、バッファプールインスタンスごとに、Least Recently Used（LRU）ページリスト内でページクリーナースレッドがスキャンを続行し、フラッシュするダーティページを探す距離を指定します。このバックグラウンド処理は1秒に1回実行されます。

値を上下に調整して、空きページ数を最小限に抑えることができます。必要以上の値を設定しないでください。スキャンを実行すると、パフォーマンスが大幅に低下する可能性があります。また、バッファプールインスタンスの数を変更する場合は、このパラメータを調整することを検討してください。 `innodb_lru_scan_depth * innodb_buffer_pool_instances` ページクリーナースレッドが毎秒実行する作業量を定義します。

デフォルトよりも小さい設定は、ほとんどのワークロードに適しています。一般的なワークロードでスペアのI/O容量がある場合にのみ、値を増やすことを検討してください。逆に、大量の書き込みが発生するワークロードでI/O容量が飽和状態になった場合は、特にバッファプールが大きい場合は値を小さくしてください。

open_file_limits

。 `open_file_limits` パラメータは、オペレーティングシステムがmysqldを開くことを許可するファイルの数を決定します。

実行時のこのパラメータの値は、システムで許可されている実際の値であり、サーバの起動時に指定した値とは異なる場合があります。MySQLが開いているファイルの数を変更できないシステムでは、値は0です。効果的な `open_files_limit` 値は、システムの起動時に指定された値（存在する場合）と `max_connections` および `table_open_cache` 次の式を使用します。

- 10以上 `max_connections` [+] (`table_open_cache` ×2）
- `max_connections` X 5
- オペレーティングシステムの制限（正の場合）
- オペレーティングシステムの制限が無限大の場合： `open_files_limit` 起動時に値が指定されます。指定されていない場合は5,000

サーバは、これら4つの値の最大値を使用してファイル記述子の数を取得しようとします。これだけ多くのディスクリプタを取得できない場合、サーバはシステムが許可する数だけのディスクリプタを取得しようとします。

ホストの設定

MySQLのコンテナ化

MySQLデータベースのコンテナ化はますます普及しています。

低レベルのコンテナ管理は、ほとんどの場合Dockerを使用して実行されます。OpenShiftやKubernetesなどのコンテナ管理プラットフォームを使用すると、大規模なコンテナ環境の管理がさらに簡単になります。コンテナ化のメリットとしては、ハイパーバイザーのライセンスが不要なため、コストの削減が挙げられます。ま

た、コンテナを使用すると、基盤となる同じカーネルとオペレーティングシステムを共有しながら、複数のデータベースを互いに分離して実行できます。コンテナはマイクロ秒単位でプロビジョニングできます。

NetAppは、ストレージの高度な管理機能を提供するAstra Tridentを提供しています。たとえば、Astra Tridentを使用すると、Kubernetesで作成されたコンテナは、適切な階層にストレージを自動的にプロビジョニングしたり、エクスポートポリシーを適用したり、Snapshotポリシーを設定したり、コンテナを別の階層にクローニングしたりできます。追加情報の場合は、を参照してください "Astra Trident のドキュメント"。

MySQLとNFSv3のスロットテーブル

LinuxでのNFSv3のパフォーマンスは、 `tcp_max_slot_table_entries`。

TCPスロットテーブルは、NFSv3でホストバスアダプタ（HBA）のキュー深度に相当します。一度に未処理となることのできるNFS処理の数を制御します。デフォルト値は通常16ですが、最適なパフォーマンスを得るには小さすぎます。逆に、新しいLinuxカーネルでTCPスロットテーブルの上限をNFSサーバが要求でいっぱいになるレベルに自動的に引き上げることができるため、問題が発生します。

パフォーマンスを最適化し、パフォーマンスの問題を回避するには、TCPスロットテーブルを制御するカーネルパラメータを調整します。

を実行します `sysctl -a | grep tcp.*.slot_table` コマンドを実行し、次のパラメータを確認します。

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

すべてのLinuxシステムに `sunrpc.tcp_slot_table_entries`ただし、次のようなものがあります。 `sunrpc.tcp_max_slot_table_entries`。どちらも128に設定する必要があります。

注意
これらのパラメータを設定しないと、パフォーマンスに大きく影響する可能性があります。Linux OSが十分なI/Oを発行していないためにパフォーマンスが制限される場合もあります。一方では、Linux OSが問題で処理できる以上のI/Oを試行すると、I/Oレイテンシが増加します。

I/OスケジューラとMySQL

Linuxカーネルでは、ブロックデバイスへのI/Oのスケジュール方法を低レベルで制御できます。

Linuxのさまざまなディストリビューションでのデフォルト設定は大きく異なります。MySQLでは、 `NOOP` または `deadline` Linuxでネイティブの非同期I/O（AIO）を使用するI/Oスケジューラ。一般に、NetAppのお客様や社内テストでは、NoOpsの方が効果的です。

MySQLのInnoDBストレージエンジンは、Linux上の非同期I/Oサブシステム（ネイティブAIO）を使用して、データファイルページの先読み要求と書き込み要求を実行します。この動作は、 `innodb_use_native_aio` 設定オプション。デフォルトで有効になっています。ネイティブの一体型I/Oでは、I/Oスケジューラのタイプ がI/Oパフォーマンスに大きく影響します。ベンチマークを実施して、ワークロードと環境に最適な結果を提供するI/Oスケジューラを特定します。

I/Oスケジューラの設定手順については、該当するLinuxおよびMySQLのドキュメントを参照してください。

MySQLファアイルキシユツシ

MySQLサーバを実行するにはファイル記述子が必要であり、デフォルト値では不十分です。

これらを使用して、新しい接続を開いたり、テーブルをキャッシュに格納したり、複雑なクエリを解決するための一時テーブルを作成したり、永続的なテーブルにアクセスしたりします。mysqldが必要なときに新しいファイルを開くことができない場合、正常に機能しなくなる可能性があります。この問題の一般的な現象は、エラー24「開いているファイルが多すぎます」です。 mysqldが同時に開くことができるファイル記述子の数は、 `open_files_limit` 構成ファイルに設定されたオプション (`/etc/my.cnf`）。でも...。 `open_files_limit` オペレーティングシステムの制限によっても異なります。この依存関係により、変数の設定がより複雑になります。

MySQLカセツテイテキナイ `open_files_limit` オプションがで指定されている値よりも大きい `ulimit 'open files'`。したがって、必要に応じてMySQLがファイルを開くことができるように、これらの制限をオペレーティングシステムレベルで明示的に設定する必要があります。Linuxでファイル制限を確認するには、次の2つの方法があります。

- 。 `ulimit` コマンドを実行すると、許可またはロックされているパラメータの詳細な概要がすばやく表示されます。このコマンドを実行して行った変更は永続的ではなく、システムのリブート後に消去されます。
- に対する変更 `/etc/security/limit.conf` ファイルは永続的であり、システムの再起動の影響を受けません。

ユーザmysqlのハードリミットとソフトリミットの両方を変更してください。構成からの抜粋を次に示します。

```
mysql hard nofile 65535
mysql soft nofile 65353
```

並行して、 `my.cnf` 開いているファイル制限を完全に使用するには。

ストレージ構成

MySQLとNFS

MySQLのドキュメントでは、NAS環境にはNFSv4を使用することを推奨しています。

ONTAP NFS転送サイズ

ONTAPでは、デフォルトでNFS IOサイズが64Kに制限されます。MySQLデータベースでランダムIOを使用する場合、ブロックサイズは64Kの最大値よりもはるかに小さくなります。ラージブロックIOは通常並列化されるため、64Kの最大値も制限ではありません。

一部のワークロードでは、最大64Kに制限があります。特に、フルテーブルスキャンバックアップ処理などのシングルスレッド処理は、データベースで実行するI/Oが少なくても大容量であれば、より高速かつ効率的に実行されます。データベースワークロードを使用するONTAPに最適なIO処理サイズは256Kです。以下のオペ

レーティングシステムのNFSマウントオプションは、それに応じて64Kから256Kに更新されています。

特定のONTAP SVMの最大転送サイズは、次のように変更できます。

```
Cluster01::> set advanced

Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.

Do you want to continue? {y|n}: y

Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
```

ONTAPで許容される最大転送サイズを、現在マウントされているNFSファイルシステムのrsize/wsizeの値より小さくしないでください。これにより、一部のオペレーティングシステムでハングしたり、データが破損したりする可能性があります。たとえば、NFSクライアントのrsize / wsizeが65536に設定されている場合は、クライアント自体が制限されているため、ONTAPの最大転送サイズを65536~1048576の間で調整しても効果はありません。最大転送サイズを65536未満に縮小すると、可用性やデータが損傷する可能性があります。

- NetAppの推奨事項*

次のNFSv4 fstab（/etc/fstab）設定を設定します。

```
nfs4 rw,
hard,nointr,bg,vers=4,proto=tcp,noatime,rsize=262144,wsize=262144
```

NFSv3を使用する一般的な問題は、停電後にロックされたInnoDBログファイルでした。時間または切り替えログファイルを使用して、この問題を解決しました。ただし、NFSv4にはロック処理があり、開いているファイルや委譲が追跡されます。

MySQLとSAN

通常の2ボリュームモデルを使用してSANでMySQLを構成するには、2つのオプションがあります。

I/Oと容量の要件が単一のLUNファイルシステムの制限内であれば、小規模なデータベースを標準LUNのペアに配置できます。たとえば、約2、000 IOPSのランダムIOPSを必要とするデータベースを、1つのLUN上の単一のファイルシステムでホストできます。同様に、サイズが100GBしかないデータベースでも、1つのLUNに収まります。管理上の問題は発生しません。

大規模なデータベースには複数のLUNが必要です。たとえば、10万IOPSを必要とするデータベースには、少なくとも8つのLUNが必要です。ドライブへのSCSIチャネルの数が不十分なため、1つのLUNがボトルネックになります。同じように、10TBのデータベースを1つの10TB LUNで管理するのは困難です。論理ボリュームマネージャは、複数のLUNのパフォーマンス機能と容量機能を結合して、パフォーマンスと管理性を向上させるように設計されています。

どちらの場合も、ONTAPボリュームのペアで十分です。単純な設定では、ログLUNと同様に、データファイルLUNも専用ボリュームに配置されます。論理ボリュームマネージャ構成の場合、データファイルボリュームグループ内のすべてのLUNは専用ボリュームに配置され、ログボリュームグループのLUNは2つ目の専用ボリュームに配置されます。

- NetAppでは* SANへのMySQL導入には2つのファイルシステムを使用することを推奨しています。
- 最初のファイルシステムには、表領域、データ、インデックスを含むすべてのMySQLデータが格納されます。
- 2番目のファイルシステムには、すべてのログ（バイナリログ、低速ログ、トランザクションログ）が格納されます。

この方法でデータを分離する理由には、次のようなものがあります。

- データファイルとログファイルのI/Oパターンは異なります。これらを分離すると、QoS制御でより多くのオプションを使用できるようになります。
- Snapshotテクノロジを最適に使用するには、データファイルを個別にリストアできる必要があります。データファイルとログファイルを混在させると、データファイルのリストアが妨げられます。
- NetApp SnapMirrorテクノロジを使用すると、シンプルでRPOの低いディザスタリカバリ機能をデータベースに提供できますが、データファイルとログには異なるレプリケーションスケジュールが必要です。

ONTAPのすべての機能を必要に応じて使用解決策できるように、この基本的な2ボリュームレイアウトを使用して、将来のニーズにも対応します。

- NetAppでは、次の機能により、ドライブをext4ファイルシステムでフォーマットすることを推奨しています。
- JFS（ジャーナルファイルシステム）で使用されるブロック管理機能の拡張アプローチと、XFS（拡張ファイルシステム）の遅延割り当て機能。
- ext4は、最大1エクスビバイト（2^{60}バイト）のファイルシステムと最大16テビバイト（$16*2$40バイト）のファイルシステムを許可します。一方、ext3ファイルシステムでサポートされる最大ファイルシステムサイズは16TB、最大ファイルサイズは2TBです。
- ext4ファイルシステムでは、複数ブロック割り当て(mballoc)は、ext3のようにファイルを1つずつ割り当てるのではなく、1回の操作で複数のブロックを割り当てます。この構成により、ブロックアロケータを数回呼び出すオーバーヘッドが削減され、メモリの割り当てが最適化されます。
- XFSは多くのLinuxディストリビューションではデフォルトですが、メタデータの管理方法が異なり、一部のMySQL構成には適していません。

- NetAppでは、mkfsユーティリティで4kブロックサイズオプションを使用して、既存のブロックLUNサイズに合わせることを推奨*しています。

```
mkfs.ext4 -b 4096
```

NetApp LUNは4KBの物理ブロックにデータを格納するため、512バイトの論理ブロックが8個生成されます。

同じブロックサイズを設定しないと、I/Oは物理ブロックと正しくアライメントされず、RAIDグループ内の2つの異なるドライブに書き込みが行われてレイテンシが発生する可能性があります。

スムーズな読み取り/書き込み処理を実現するためには、I/Oのアライメントが重要です。ただし、物理ブロックの先頭以外の論理ブロックからI/Oが開始されると、I/Oはミスアライメントされます。I/O処理がアライメントされるのは、I/O処理が論理ブロック（物理ブロック内の最初の論理ブロック）で開始されたときだけです。

Oracle データベース

ONTAP上のOracleデータベース

ONTAPはOracleデータベース向けに設計されています。ONTAPは数十年にわたり、リレーショナルデータベースI/O固有の要求に合わせて最適化されてきました。また、Oracleデータベースのニーズに対応するために、さらにはOracle Inc自体の要求にも対応するために、複数のONTAP機能が開発されました。

> 本ドキュメントは、これまでに公開されていたテクニカルレポート_TR-3633：『Oracle databases on ONTAP』、TR-4591：『Oracle data protection：Backup、recovery、replication』、TR-4592：『Oracle on MetroCluster』、TR-4534：『Migration of Oracle Databases to NetApp Storage Systems_

ONTAPがデータベース環境にもたらすさまざまな価値に加えて、データベースのサイズ、パフォーマンス要件、データ保護のニーズなど、ユーザのさまざまな要件もあります。NetAppストレージの既知の導入には、VMware ESXで実行される約6、000個のデータベースの仮想環境から、現在996TBのシングルインスタンスデータウェアハウスまで、あらゆるものが含まれます。そのため、NetAppストレージ上にOracleデータベースを設定する際の明確なベストプラクティスはほとんどありません。

NetAppストレージでOracleデータベースを運用するための要件には、2つの方法があります。まず、明確なベストプラクティスが存在する場合は、具体的に説明します。大まかに、Oracleストレージソリューションの設計者が、それぞれのビジネス要件に基づいて対処しなければならない、設計上のさまざまな考慮事項について説明します。

ONTAP構成

RAIDデータベースとOracleデータベース

RAIDとは、冗長性を使用してドライブの損失からデータを保護することです。

Oracleデータベースやその他のエンタープライズアプリケーションに使用するNetAppストレージの構成では、RAIDレベルに関して疑問が生じることがあります。ストレージアレイ構成に関する従来のOracleのベストプラクティスの多くには、RAIDミラーリングの使用や特定のタイプのRAIDの回避に関する警告が含まれています。これらは有効なポイントを上げますが、これらのソースは、RAID 4およびONTAPで使用されているNetApp RAID DPおよびRAID-TECテクノロジには適用されません。

RAID 4、RAID 5、RAID 6、RAID DP、およびRAID-TECはいずれもパリティを使用して、ドライブ障害によってデータが失われないようにします。これらのRAIDオプションはミラーリングよりもストレージ利用率がはるかに優れていますが、ほとんどのRAID実装には書き込み処理に影響する欠点があります。他のRAID実装で書き込み操作が完了すると、パリティデータを再生成するために複数のドライブ読み取りが必要になる場合があります。これは、一般にRAIDペナルティと呼ばれるプロセスです。

ただし、ONTAPではこのようなRAIDペナルティは発生しません。これは、NetApp WAFL（Write Anywhere File Layout）とRAIDレイヤが統合されているためです。書き込み処理はRAMで結合され、パリティ生成を含む完全なRAIDストライプとして準備されます。ONTAPは書き込みを完了するために読み取りを実行する必要がないため、ONTAPとWAFLはRAIDペナルティを回避できます。レイテンシが重要な処理（Redoロギングなど）のパフォーマンスが妨げられることはありません。また、データファイルのランダム書き込みでは、パリティの再生成が必要になるためにRAIDのペナルティが発生することもありません。

統計的信頼性に関しては、RAID DPでさえRAIDミラーリングよりも優れた保護を提供します。主な問題は、RAIDのリビルド中にドライブに要求が発生することです。ミラーリングされたRAIDセットでは、RAIDセット内のパートナーへのリビルド中にドライブ障害によるデータ損失のリスクが、RAID DPセット内の三重ドライブ障害のリスクよりもはるかに高くなります。

Oracleデータベースとストレージ容量の管理

データベースやその他のエンタープライズアプリケーションを、予測性と管理性に優れたハイパフォーマンスのエンタープライズストレージで管理するには、データとメタデータを管理するためにドライブ上の空きスペースがいくらか必要です。必要な空きスペースの量は、使用するドライブのタイプやビジネスプロセスによって異なります。

空きスペースとは実際のデータに使用されていないスペースのことで、アグリゲート自体の未割り当てスペースやコンスティチュエントボリューム内の未使用スペースを含みます。シンプロビジョニングも考慮する必要があります。たとえば、あるボリュームに含まれている1TBのLUNのうち、実際のデータに使用されているのは50%だけであるとします。シンプロビジョニング環境では、500GBのスペースが消費されているように見えます。ただし、フルプロビジョニング環境では、1TBの全容量が使用中と表示されます。500GBの未割り当てスペースは非表示になります。このスペースは実際のデータには使用されていないため、空きスペースの合計の計算に含める必要があります。

エンタープライズアプリケーションに使用するストレージシステムに関するNetAppの推奨事項は次のとおりです。

SSDアグリゲート（**AFF**システムを含む）

* NetAppでは*最低10%の空き容量を推奨しています。これには、アグリゲートまたはボリューム内の空きスペース、フルプロビジョニングのために割り当てられているが実際のデータには使用されていない空きスペースなど、すべての未使用スペースが含まれます。論理スペースは重要ではありません。問題は、データストレージに実際に使用できる物理的な空きスペースの量です。

推奨される10%の空きスペースは非常に控えめな値です。SSDアグリゲートでは、パフォーマンスに影響を与えることなく、さらに高い利用率でワークロードをサポートできます。ただし、アグリゲートの使用率が高くなると、使用率を注意深く監視しないと、スペースが不足するリスクも高まります。さらに、容量99%でシステムを実行している場合はパフォーマンスが低下することはありませんが、ハードウェアの追加発注時にシステムが完全にいっぱいにならないように管理作業が必要になる可能性があり、追加のドライブの調達と取り付けに時間がかかることがあります。

HDDアグリゲート（**Flash Pool**アグリゲートを含む）

* NetAppでは*回転式ドライブを使用する場合は、最低15%の空き容量を確保することを推奨しています。これには、アグリゲートまたはボリューム内の空きスペース、フルプロビジョニングのために割り当てられているが実際のデータには使用されていない空きスペースなど、すべての未使用スペースが含まれます。言論の自由が10%に近づくと、パフォーマンスに影響が出ます。

Oracleデータベースと**Storage Virtual Machine**

Oracleデータベースのストレージ管理をStorage Virtual Machine（SVM）で一元化

SVMは、ONTAP CLIではSVMと呼ばれ、ストレージの基本的な機能ユニットであり、VMware ESXサーバ上のゲストと比較すると便利です。

ESXを最初にインストールした時点では、ゲストOSのホストやエンドユーザアプリケーションのサポートなど、事前に設定された機能はありません。仮想マシン（VM）が定義されるまでは空のコンテナです。ONTAPも同様です。ONTAPを最初にインストールした時点では、SVMが作成されるまでデータを提供する機能はありません。データサービスはSVMの特性によって定義されます。

ストレージアーキテクチャの他の要素と同様に、SVMと論理インターフェイス（LIF）の設計に最適なオプションは、拡張要件とビジネスニーズによって大きく異なります。

SVM

ONTAP用にSVMをプロビジョニングする公式のベストプラクティスはありません。適切なアプローチは、管理要件とセキュリティ要件によって異なります。

ほとんどのお客様は、プライマリSVMを1つ運用して日常的な要件のほとんどに対応しつつ、特殊なニーズに対応するSVMを少数作成しています。たとえば、次のようなものを作成できます。

- スペシャリストチームが管理する重要なビジネスデータベース用のSVM
- 開発グループ用のSVM。独自のストレージを個別に管理できるように、完全な管理権限が与えられています。
- 人事や財務報告のデータなど、機密性の高いビジネスデータを格納するSVM。管理チームを限定する必要がある

マルチテナント環境では、各テナントのデータに専用のSVMを割り当てることができます。クラスタ、HAペア、およびノードあたりのSVMとLIFの数の上限は、使用するプロトコル、ノードモデル、およびONTAPのバージョンによって異なります。 を参照してください "NetApp Hardware Universe の略" これらの限界のために。

ONTAP QoSによるOracleデータベースのパフォーマンス管理

複数のOracleデータベースを安全かつ効率的に管理するには、効果的なQoS戦略が必要です。その理由は、最新のストレージシステムのパフォーマンス機能が絶えず向上していることです。

特に、オールフラッシュストレージの採用が増えたことで、ワークロードの統合が実現しました。回転式メディアに依存するストレージアレイでは、古い回転式ドライブテクノロジではIOPS機能が制限されているため、I/O負荷の高いワークロードの数が限られていました。1つまたは2つの高アクティブデータベースでは、ストレージコントローラが制限に達するずっと前に基盤となるドライブがいっぱいになります。これは変更されました。SSDドライブの数が比較的少ないパフォーマンス機能は、最も強力なストレージコントローラでさえ飽和状態になる可能性があります。つまり、回転式メディアのレイテンシが急激に低下することなく、コントローラのすべての機能を活用できます。

参考例として、シンプルな2ノードHA AFF A800システムでは、レイテンシが1ミリ秒を超える前に最大100万IOPSのランダムIOPSを処理できます。このレベルに達すると予想される単一のワークロードはほとんどありません。このAFF A800システムアレイをフルに活用するには、複数のワークロードをホストする必要がありますが、予測可能性を確保しながら安全にこれを行うには、QoS制御が必要です。

ONTAPには、IOPSと帯域幅の2種類のサービス品質（QoS）があります。QoS制御は、SVM、ボリューム、LUN、およびファイルに適用できます。

IOPS QoS

IOPS QoS制御は、特定のリソースの合計IOPSに基づいていることは明らかですが、IOPS QoSには直感的でない側面がいくつかあります。当初、一部のお客様は、IOPSのしきい値に達したときにレイテンシが明らかに上昇したことに困惑していました。レイテンシの増加は、IOPSが制限されることによる当然の結果です。論理的には、トークンシステムと同様に機能します。たとえば、データファイルが格納されている特定のボリュームに10、000 IOPSの制限がある場合、受信した各I/Oは、処理を続行するために最初にトークンを受信する必要があります。1秒間に10Kを超えるトークンが消費されない限り、遅延は発生しません。IO処理がトークンの受信を待機する必要がある場合、この待機は追加のレイテンシとして表示されます。ワークロードがQoS制限に押し上げられにくくなると、各IOが処理されるまでキューで待機する時間が長くなり、レイテンシが高くなります。

データベースのトランザクション/ REDOログデータにQoS制御を適用する場合は注意が必要です。通常、Redoログに必要なパフォーマンスはデータファイルよりもはるかに低くなりますが、Redoログのアクティビティはバースト性が高くなります。IOは短時間のパルスで発生し、平均REDO IOレベルに適したQoS制限が実際の要件に対して低すぎる可能性があります。その結果、Redoログのバーストごとに QoSが適用されるため、パフォーマンスが大幅に制限される可能性があります。一般に、REDOログとアーカイブログはQoSによって制限されるべきではありません。

帯域幅QoS

すべてのI/Oサイズが同じではありません。たとえば、あるデータベースが大量の小さなブロック読み取りを実行していて、IOPSのしきい値に達しているとします。 一方、データベースがテーブルのフルスキャン処理を実行している場合もあります。この処理では、大量のブロック読み取りが非常に少なく、大量の帯域幅を消費しますが、IOPSは比較的低くなります。

同様に、VMware環境ではブート時に非常に多くのランダムIOPSが発生する可能性がありますが、外部バックアップ時に実行されるI/Oは少なくても大きくなります。

パフォーマンスを効果的に管理するために、IOPSまたは帯域幅のQoS制限、あるいはその両方が必要になる場合があります。

最小/保証されたQoS

多くのお客様が、QoS保証付きの解決策を求めていますが、これは見た目よりも達成が難しく、無駄になる可能性があります。たとえば、10個のデータベースを10K IOPS保証で配置する場合、10個のデータベースすべてが同時に10K IOPSで実行され、合計で100Kになるようにシステムをサイジングする必要があります。

QoS管理を最小限に抑えるには、重要なワークロードを保護するのが最適です。たとえば、最大IOPSが500Kで、本番ワークロードと開発ワークロードが混在しているONTAPコントローラについて考えてみましょう。特定のデータベースがコントローラを独占しないように、開発ワークロードに最大QoSポリシーを適用する必要があります。次に、最小限のQoSポリシーを本番環境のワークロードに適用して、必要なIOPSを必要なときに常に利用できるようにします。

アダプティブ QoS

アダプティブQoSとは、ONTAPの機能のことで、ストレージオブジェクトの容量に基づいてQoS制限が設定されます。通常、データベースのサイズとそのパフォーマンス要件との間にリンクがないため、データベースで使用されることはほとんどありません。大規模なデータベースはほとんど不活性になる可能性がありますが、小規模なデータベースはIOPS負荷が最も高くなる可能性があります。

アダプティブQoSは、仮想化データストアで非常に役立ちます。このようなデータセットのIOPS要件は、デ

ータベースの合計サイズに相関する傾向があるためです。1TBのVMDKファイルを格納した新しいデータストアでは、2TBデータストアの約半分のパフォーマンスが必要になる可能性があります。アダプティブQoSを使用すると、データストアにデータが入力されたときに、QoS制限を自動的に増やすことができます。

Oracleデータベースと**ONTAP**の効率化機能

ONTAPのスペース効率化機能は、Oracleデータベース向けに最適化されています。ほとんどの場合、すべての効率化機能を有効にした状態でデフォルトのままにすることを推奨します。

圧縮、コンパクション、重複排除などのスペース効率化機能は、特定の量の物理ストレージに収まる論理データの量を増やすように設計されています。その結果、コストと管理オーバーヘッドが削減されます。

圧縮とは、大まかに言って、データのパターンを検出してスペースを削減する方法でエンコードする数学的プロセスです。一方、重複排除機能は、実際に繰り返されるデータブロックを検出し、不要なコピーを削除します。コンパクションを使用すると、複数の論理ブロックのデータをメディア上の同じ物理ブロックで共有できます。

Storage Efficiencyとフラクショナルリザベーションの連動については、シンプロビジョニングに関する以下のセクションを参照してください。

圧縮

オールフラッシュストレージシステムが登場する以前は、アレイベースの圧縮の価値は限られていました。I/O負荷の高いワークロードのほとんどでは、許容可能なパフォーマンスを提供するために非常に多数のスピンドルが必要だったためです。ストレージシステムには、ドライブ数が多いことの副作用として、必要以上の容量が常に搭載されていました。この状況は、ソリッドステートストレージの登場によって変化しました。優れたパフォーマンスを得るためだけにドライブを過剰にオーバープロビジョニングする必要はもうありません。ストレージシステムのドライブスペースは、実際の容量ニーズに合わせて調整できます。

ソリッドステートドライブ（SSD）ではIOPSが向上するため、ほとんどの場合、回転式ドライブに比べてコストを削減できますが、圧縮を使用すると、ソリッドステートメディアの実効容量を増やすことで、さらに削減効果を高めることができます。

データを圧縮する方法はいくつかあります。多くのデータベースには独自の圧縮機能が搭載されていますが、お客様の環境ではこのような圧縮機能はほとんど見られません。その理由は、通常、圧縮データを*変更*するとパフォーマンスが低下することに加え、一部のアプリケーションではデータベースレベルの圧縮のライセンスコストが高くなることにあります。最後に、データベース処理のパフォーマンスが全体的に低下します。実際のデータベース作業ではなく、データの圧縮と解凍を実行するCPUに、高いCPU単位のライセンスコストを支払うことはほとんど意味がありません。より適切な方法は、圧縮処理をストレージシステムにオフロードすることです。

適応圧縮

アダプティブ圧縮は、レイテンシがマイクロ秒単位で測定されるオールフラッシュ環境であっても、パフォーマンスに影響を及ぼさないエンタープライズワークロードで徹底的にテストされています。一部のお客様から、圧縮機能によってデータがキャッシュ内で圧縮されたままになるため、パフォーマンスが向上したとの報告もあります。これは、コントローラで使用可能なキャッシュ容量が実質的に増加するためです。

ONTAPは物理ブロックを4KB単位で管理します。アダプティブ圧縮では、デフォルトの圧縮ブロックサイズである8KBが使用されます。つまり、データは8KB単位で圧縮されます。これは、リレーショナルデータベースで最もよく使用される8KBのブロックサイズに一致します。圧縮アルゴリズムは、より多くのデータが1つ

の単位として圧縮されるので、より効率的になります。圧縮ブロックサイズが32KBの場合、8KBの圧縮ブロックユニットよりもスペース効率に優れています。つまり、デフォルトの8KBのブロックサイズを使用するアダプティブ圧縮の場合、削減率はわずかに低くなりますが、圧縮ブロックサイズを小さくすることには大きなメリットがあります。データベースワークロードには、大量の上書きアクティビティが含まれています。32KBの圧縮されたデータブロックの8KBを上書きするには、32KBの論理データ全体を読み取って解凍し、必要な8KB領域を更新してから再度圧縮し、32KB全体をドライブに書き込む必要があります。この処理はストレージシステムでは非常にコストがかかります。このため、圧縮ブロックサイズの大きい競合ストレージアレイでも、データベースワークロードのパフォーマンスが大幅に低下します。

適応圧縮で使用されるブロックサイズは、最大32KBまで拡張できます。これによりストレージ効率が向上する可能性があります。このようなデータがアレイに大量に格納されている場合は、トランザクションログやバックアップファイルなどの静止ファイルについて検討する必要があります。状況によっては、適応圧縮のブロックサイズをそれに合わせて増やすことで、16KBまたは32KBのブロックサイズを使用するアクティブデータベースでもメリットが得られる場合があります。この方法がお客様のワークロードに適しているかどうかについては、NetAppまたはパートナーの担当者にお問い合わせください。

ストリーミングバックアップデスティネーションでは、重複排除と一緒に8KBを超える圧縮ブロックサイズを使用しないでください。これは、バックアップデータへのわずかな変更が32KBの圧縮ウィンドウに影響するためです。ウィンドウが移動すると、圧縮されたデータはファイル全体で異なります。重複排除は圧縮後に実行されます。つまり、重複排除エンジンは、圧縮された各バックアップを別 々 に認識します。ストリーミングバックアップの重複排除が必要な場合は、8KBのブロックアダプティブ圧縮のみを使用します。アダプティブ圧縮を使用することを推奨します。アダプティブ圧縮はブロックサイズが小さく、重複排除による効率化の妨げにならないためです。同様の理由から、ホスト側の圧縮も重複排除による効率化の妨げになります。

圧縮のアライメント

データベース環境でアダプティブ圧縮を使用する場合は、圧縮ブロックのアライメントについて考慮する必要があります。これは、非常に特定のブロックでランダムオーバーライトが発生するデータについてのみ考慮する必要があります。このアプローチは、ファイルシステム全体のアライメントと概念的に似ています。ファイルシステムの開始は4Kデバイスの境界に合わせて調整する必要があり、ファイルシステムのブロックサイズは4Kの倍数でなければなりません。

たとえば、ファイルへの8KBの書き込みは、ファイルシステム自体の8KBの境界にアライメントされている場合にのみ圧縮されます。これは、ファイルの最初の8KB、ファイルの2番目の8KBなどに配置する必要があることを意味します。アライメントを正しく行う最も簡単な方法は、正しいLUNタイプを使用することです。作成するパーティションには、デバイスの先頭から8Kの倍数のオフセットを設定し、データベースのブロックサイズの倍数のファイルシステムのブロックサイズを使用する必要があります。

バックアップやトランザクションログなどのデータは、複数のブロックにまたがるシーケンシャル書き込み処理であり、すべて圧縮されます。したがって、アライメントを考慮する必要はありません。問題となるI/Oパターンは、ファイルのランダムオーバーライトだけです。

データコンパクション

データコンパクションは、圧縮効率を向上させるテクノロジです。前述したように、アダプティブ圧縮では4KBのWAFLブロックに8KBのI/Oが格納されるため、削減率は最大でも2：1です。ブロックサイズが大きい圧縮方式では、効率性が向上します。ただし、小さなブロックの上書きが発生するデータには適していません。32KBのデータユニットを解凍して8KB部分を更新し、再度圧縮してからドライブにライトバックすると、オーバーヘッドが発生します。

データコンパクションでは、複数の論理ブロックを物理ブロック内に格納できます。たとえば、テキストブロックや部分的にフルブロックなど、圧縮率の高いデータを含むデータベースは、8KBから1KBに圧縮できます。コンパクションを使用しない場合、この1KBのデータが4KBブロック全体を占有します。インラインデータコンパクションでは、圧縮された1KBのデータを、他の圧縮データと一緒にわずか1KBの物理スペースに格納できます。これは圧縮テクノロジではありません。ドライブのスペースをより効率的に割り当てる方法なので、検出できるほどのパフォーマンスへの影響はありません。

得られる削減効果の程度はさまざまです。すでに圧縮または暗号化されているデータは、通常それ以上圧縮することはできないため、コンパクションによるメリットはありません。一方、初期化されたばかりのデータファイルで、ブロックメタデータとゼロブロックしか含まれていない場合は、最大80：1まで圧縮できます。

温度に基づくストレージ効率

ONTAP 9.8以降で利用できるTemperature Sensitive Storage Efficiency（TSSE）は、ブロックアクセスのヒートマップを利用してアクセス頻度の低いブロックを特定し、圧縮して効率を高めます。

重複排除

重複排除とは、データセットから重複するブロックサイズを削除することです。たとえば、同じ4KBブロックが10個のファイルに存在する場合、重複排除機能は、10個のファイルすべてのうち、その4KBブロックを同じ4KBの物理ブロックにリダイレクトします。その結果、そのデータの効率が10分の1に向上します。

VMwareゲストブートLUNなどのデータは、同じオペレーティングシステムファイルの複数のコピーで構成されるため、通常は重複排除が非常に効果的です。100:1以上の効率が観測されている。

一部のデータに重複データが含まれていません。たとえば、Oracleブロックには、データベースに対してグローバルに一意のヘッダーと、ほぼ一意のトレーラが含まれています。そのため、Oracleデータベースの重複排除によって1%以上の削減効果が得られることはほとんどありません。MS SQLデータベースでの重複排除はやや優れていますが、ブロックレベルでの固有のメタデータは依然として制限されています。

16KBでブロックサイズが大きいデータベースでは、最大15%のスペース削減効果が確認されたケースがいくつかあります。各ブロックの最初の4KBにはグローバルに一意なヘッダーが含まれ、最後の4KBブロックにはほぼ一意のトレーラが含まれます。内部ブロックは重複排除の対象となりますが、実際には、初期化されたデータの重複排除にほぼ完全に起因しています。

競合するアレイの多くは、データベースが複数回コピーされていると仮定して、データベースの重複排除機能があると主張しています。この点では、NetAppの重複排除も使用できますが、ONTAPにはNetApp FlexCloneテクノロジというより優れたオプションがあります。最終的な結果は同じで、基盤となる物理ブロックの大部分を共有するデータベースのコピーが複数作成されます。FlexCloneを使用すると、時間をかけてデータベースファイルをコピーしてから重複を排除するよりも、はるかに効率的です。重複は最初から作成されないため、実際には重複排除ではなく重複排除です。

効率性とシンプロビジョニング

効率化機能はシンプロビジョニングの一形態です。たとえば、100GBのボリュームを使用している100GBのLUNを50GBに圧縮するとします。ボリュームが100GBのままなので、実際の削減はまだ実現されていません。削減されたスペースをシステムの他の場所で使用できるように、まずボリュームのサイズを縮小する必要があります。100GBのLUNにあとから変更した結果、データの圧縮率が低下すると、LUNのサイズが大きくなり、ボリュームがいっぱいになる可能性があります。

シンプロビジョニングは、管理を簡易化しながら、使用可能な容量を大幅に改善し、コストを削減できるため、強く推奨されます。これは、単純なデータベース環境では、多くの場合、空のスペース、多数のボリュームやLUN、圧縮可能なデータが含まれているためです。シックプロビジョニングでは、ボリュームとLUNのス

トレージにスペースがリザーブされます。これは、100%フルになり、100%圧縮不可能なデータが含まれる場合に限られます。これは起こりそうもないことですシンプロビジョニングを使用すると、スペースを他の場所で再生して使用できます。また、容量の管理は、多数の小さいボリュームやLUNではなく、ストレージシステム自体に基づいて行うことができます。

一部のお客様は、特定のワークロードにシックプロビジョニングを使用するか、一般的には確立された運用と調達の手法に基づいてシックプロビジョニングを使用します。

*注意：*ボリュームがシックプロビジョニングされている場合は、ボリュームの圧縮解除や重複排除の削除など、そのボリュームのすべての効率化機能を完全に無効にするように注意する必要があります。 `sis undo` コマンドを実行しますボリュームが `volume efficiency show` 出力。有効になっている場合、ボリュームはまだ部分的に効率化機能用に設定されています。その結果、オーバーライトギャランティの動作が異なります。そのため、設定で原因が見落とされてボリュームのスペースが予期せず不足し、データベースI/Oエラーが発生する可能性が高くなります。

効率化のベストプラクティス

NetAppの推奨事項は次のとおりです。

AFFのデフォルト

オールフラッシュAFFシステムで実行されているONTAPで作成されたボリュームは、すべてのインライン効率化機能が有効になった状態でシンプロビジョニングされます。一般にデータベースには重複排除機能はなく、圧縮不可能なデータも含まれている可能性がありますが、デフォルト設定はほとんどすべてのワークロードに適しています。ONTAPは、あらゆる種類のデータとI/Oパターンを効率的に処理するように設計されており、削減効果があるかどうかは関係ありません。デフォルトは、理由が完全に理解されていて、逸脱するメリットがある場合にのみ変更する必要があります。

一般的な推奨事項

- ボリュームやLUNがシンプロビジョニングされていない場合は、すべての効率化設定を無効にする必要があります。これらの機能を使用しても削減は得られず、シックプロビジョニングとスペース効率化が有効になっていると、スペース不足エラーなどの予期しない動作が原因に発生する可能性があるためです。
- バックアップやデータベーストランザクションログなどでデータが上書きされない場合は、クーリング期間を短くしてTSSEを有効にすることで、効率を高めることができます。
- アプリケーションレベルで圧縮がすでに有効になっているファイルが暗号化されている場合など、一部のファイルには圧縮不可能なデータが大量に含まれていることがあります。上記のいずれかに該当する場合は、圧縮可能なデータを含む他のボリュームでより効率的に処理できるように、圧縮を無効にすることを検討してください。
- データベースバックアップでは、32KBの圧縮機能と重複排除機能の両方を使用しないでください。を参照してください [適応圧縮] を参照してください。

Oracleデータベースのシンプロビジョニング

Oracleデータベースのシンプロビジョニングでは、ストレージシステムに物理的に使用可能なスペースよりも多くのスペースが設定されるため、慎重な計画が必要になります。適切に行うと、大幅なコスト削減と管理性の向上につながるため、努力する価値は非常に高くなります。

シンプロビジョニングにはさまざまな形式があり、ONTAPがエンタープライズアプリケーション環境に提供

する多くの機能に欠かせない機能です。シンプロビジョニングも効率化テクノロジと密接に関連しています。効率化機能を使用すると、ストレージシステムに実際に存在するよりも多くの論理データを格納できます。

Snapshotを使用する場合、シンプロビジョニングが必要になります。たとえば、NetAppストレージ上の一般的な10TBのデータベースには、約30日間のSnapshotが含まれています。この構成では、アクティブファイルシステムに表示されるデータは約10TB、Snapshot専用のデータは300TBになります。合計310TBのストレージは、通常、約12$_{15TB}$のスペースに配置されます。アクティブデータベースは$_{10TB}$を消費しますが、残りの$_{300TB}$のデータは元のデータに加えられた変更のみが格納されるため、$_{2TB}$5TBのスペースしか必要としません。

クローニングもシンプロビジョニングの一例です。NetAppの主要なお客様が、80TBのデータベースのクローンを40個作成し、開発用に使用しました。これらのクローンを使用している40人の開発者全員がすべてのデータファイルのすべてのブロックを上書きした場合、3.2PBを超えるストレージが必要になります。実際には書き替え率は低く、変更のみがドライブに格納されるため、必要なスペースは合計で40TBに近くなります。

スペース管理

データの変更率が予期せず増加する可能性があるため、アプリケーション環境のシンプロビジョニングには注意が必要です。たとえば、データベーステーブルのインデックスを再作成したり、VMwareゲストに大規模なパッチを適用したりすると、Snapshotによるスペース消費が急増します。バックアップの配置が間違っていると、非常に短時間で大量のデータが書き込まれる可能性があります。最後に、ファイルシステムの空きスペースが予期せず不足した場合、一部のアプリケーションのリカバリが困難になることがあります。

幸いなことに、これらのリスクは慎重に構成することで対処できます。 `volume-autogrow` および `snapshot-autodelete` ポリシー：名前からわかるように、これらのオプションを使用すると、Snapshotによって消費されるスペースを自動的にクリアしたり、追加データに対応するためにボリュームを拡張したりするポリシーを作成できます。多くのオプションが用意されており、ニーズはお客様によって異なります。

を参照してください "論理ストレージ管理に関する文書" を参照してください。

フラクショナルリザベーション

フラクショナルリザーブは、ボリューム内でのスペース効率に関するLUNの動作です。オプション `fractional-reserve` が100%に設定されている場合、ボリュームのスペースを使い切ることなく、任意のデータパターンでボリューム内のすべてのデータを100%書き替えることができます。

たとえば、1TBのボリュームに配置された単一の250GB LUNにデータベースが格納されているとします。Snapshotを作成すると、ただちにボリュームに250GBのスペースが追加でリザーブされ、何らかの理由でボリュームのスペースが不足することはありません。データベースボリュームのすべてのバイトの上書きが必要になることはほとんどないため、フラクショナルリザーブの使用は一般に無駄です。決して発生しないイベントのためにスペースを予約する理由はありません。ただし、ストレージシステムのスペース消費を監視できず、スペースが不足しないように保証する必要がある場合は、Snapshotの使用に100%のフラクショナルリザベーションが必要になります。

圧縮と重複排除

圧縮と重複排除はどちらもシンプロビジョニングの形式です。たとえば、50TBのデータ容量を30TBに圧縮すると、20TBが削減されます。圧縮によってメリットが得られるようにするには、20TBの一部を他のデータに使用するか、50TB未満のストレージシステムを購入する必要があります。その結果、ストレージシステムで実際に使用可能な量よりも多くのデータが格納されます。データの観点から見ると、ドライブでは30TBしか占有していないにもかかわらず、50TBのデータがあります。

データセットの圧縮率は常に変化し、実際のスペースの消費量が増加する可能性があります。このように消費量が増加するため、他の形式のシンプロビジョニングと同様に、監視と使用の観点から圧縮を管理する必要が

あります。 `volume-autogrow` および `snapshot-autodelete`。

圧縮機能と重複排除機能の詳細については、efficiency.htmlのリンクを参照してください。

圧縮とフラクショナルリザベーション

圧縮はシンプロビジョニングの一形態です。フラクショナルリザベーションは圧縮の使用に影響します。スペースはSnapshotの作成前にリザーブされる点に注意してください。通常、フラクショナルリザーブが重要になるのはSnapshotが存在する場合のみです。Snapshotがない場合、フラクショナルリザーブは重要ではありません。これは、圧縮の場合には当てはまりません。圧縮が有効なボリュームでLUNを作成すると、ONTAPではSnapshotに対応するためのスペースが確保されます。この動作は設定時に混乱を招く可能性がありますが、これは想定される動作です。

たとえば、10GBのボリュームに5GBのLUNが格納され、2.5GBに圧縮されてSnapshotが作成されていないとします。次の2つのシナリオを検討します。

- フラクショナルリザーブ= 100では7.5GBが使用されます。
- フラクショナルリザーブ= 0の場合、2.5GBの使用率が得られます。

最初のシナリオでは、現在のデータ用に2.5GBのスペースが使用され、スナップショットの使用を想定してソースの100%の切り替えに使用される5GBのスペースが使用されます。2番目のシナリオでは、追加のスペースは確保されません。

この状況は混乱しているように見えるかもしれませんが、実際に遭遇することはほとんどありません。圧縮はシンプロビジョニングを意味し、LUN環境でのシンプロビジョニングにはフラクショナルリザベーションが必要です。圧縮されたデータは圧縮不可能なデータで上書きされる可能性があります。つまり、圧縮によって削減効果が得られるように、ボリュームをシンプロビジョニングする必要があります。

- NetAppでは*次の予約構成を推奨しています。
- 設定 `fractional-reserve` 基本的な容量監視と `volume-autogrow` および `snapshot-autodelete`。
- 設定 `fractional-reserve` 監視機能がない場合、または何らかの状況でスペースを使い切ることができない場合は、100にします。

空きスペースと**LVM**スペースの割り当て

ファイルシステム環境でアクティブなLUNをシンプロビジョニングした場合、データが削除されるにつれて効率が低下する可能性があります。削除されたデータがゼロで上書きされない限り（ "ASMRU" または、スペースがTRIM/UNMAPスペース再生で解放されると、「消去された」データは、ファイルシステム内の割り当てられていないスペースをますます占有します。さらに、データファイルは作成時にフルサイズに初期化されるため、アクティブなLUNのシンプロビジョニングは多くのデータベース環境ではあまり使用されません。

LVMの構成を慎重に計画すると、効率が向上し、ストレージのプロビジョニングやLUNのサイズ変更の必要性を最小限に抑えることができます。Veritas VxVMやOracle ASMなどのLVMを使用すると、基盤となるLUNが複数のエクステントに分割され、必要な場合にのみ使用されます。たとえば、最初は2TBのサイズだったデータセットが、やがて10TBに拡張される可能性がある場合、このデータセットを、シンプロビジョニングされた10TBのLUNがLVMディスクグループにまとめられて配置できます。作成時に消費されるスペースは2TBにすぎず、データ量の増大に対応するためにエクステントが割り当てられた場合にのみ追加のスペースが必要になります。このプロセスは、スペースが監視されているかぎり安全です。

Oracleデータベースと**ONTAP**コントローラのフェイルオーバー/スイッチオーバー

Oracleデータベースの処理が中断されないようにするには、ストレージのテイクオーバーとスイッチオーバーの機能を理解しておく必要があります。また、テイクオーバー処理やスイッチオーバー処理で使用される引数を正しく使用しないと、データの整合性に影響する可能性があります。

- 通常の状態では、特定のコントローラへの書き込みは、パートナーに同期ミラーリングされます。NetApp MetroCluster環境では、書き込みはリモートコントローラにもミラーリングされます。書き込みがすべての場所の不揮発性メディアに格納されるまで、ホストアプリケーションに確認応答は返されません。
- 書き込みデータを格納するメディアは、不揮発性メモリ（NVMEM）と呼ばれます。不揮発性ランダムアクセスメモリ（NVRAM）と呼ばれることもあります。機能はジャーナルですが、書き込みキャッシュとみなすことができます。通常の処理でNVMEMのデータが読み取られることはなく、ソフトウェアまたはハードウェアに障害が発生した場合のデータ保護にのみ使用されます。ドライブにデータが書き込まれると、NVMEMではなくシステムのRAMからデータが転送されます。
- テイクオーバー処理では、ハイアベイラビリティ（HA）ペアの一方のノードがパートナーの処理をテイクオーバーします。スイッチオーバーは基本的に同じですが、IT環境 MetroCluster構成ではリモートノードがローカルノードの機能を引き継ぎます。

定期的なメンテナンス作業中は、ネットワークパスの変更によって一時的に運用が停止する可能性がある場合を除き、ストレージのテイクオーバーやスイッチオーバーは透過的に行われる必要があります。ただし、ネットワークの設定は複雑なため、エラーが発生しやすいため、NetAppでは、ストレージシステムを本番環境に移行する前に、テイクオーバーとスイッチオーバーの処理を徹底的にテストすることを強く推奨します。これは、すべてのネットワークパスが正しく設定されていることを確認する唯一の方法です。SAN環境では、コマンドの出力を慎重に確認します。 `sanlun lun show -p` 想定されるすべてのプライマリパスとセカンダリパスが使用可能であることを確認します。

テイクオーバーやスイッチオーバーを強制的に実行する場合は注意が必要です。これらのオプションを使用してストレージ構成を強制的に変更すると、ドライブを所有するコントローラの状態が無視され、代わりのノードが強制的にドライブの制御を引き継ぐことになります。テイクオーバーの強制が正しく行われないと、データの損失や破損が発生する可能性があります。これは、強制的なテイクオーバーやスイッチオーバーによってNVMEMの内容が破棄される可能性があるためです。テイクオーバーまたはスイッチオーバーの完了後にそのデータが失われると、データベースから見ると、ドライブに格納されているデータが少し古い状態に戻る可能性があります。

通常のHAペアを使用した強制テイクオーバーが必要になることはほとんどありません。ほぼすべての障害シナリオでは、ノードがシャットダウンし、パートナーに通知して自動フェイルオーバーが実行されます。一部のエッジケース（ローリング障害など）では、ノード間のインターコネクトが失われたあとに一方のコントローラが失われた場合など、強制テイクオーバーが必要になります。この場合、コントローラで障害が発生する前にノード間のミラーリングが失われるため、稼働しているコントローラには実行中の書き込みのコピーが存在しなくなります。その後、テイクオーバーを強制的に実行する必要があります。つまり、データが失われる可能性があります。

同じ論理環境A MetroClusterスイッチオーバー。通常の状態では、スイッチオーバーはほぼ透過的に実行されます。ただし、災害が発生すると、サバイバーサイトとディザスタサイトの間の接続が失われる可能性があります。サバイバーサイトから見ると、問題はサイト間の接続の中断にすぎず、元のサイトが引き続きデータを処理している可能性があります。ノードがプライマリコントローラの状態を検証できない場合は、強制スイッチオーバーのみが実行できます。

- NetAppでは、次の注意事項を遵守することを推奨しています。
- テイクオーバーやスイッチオーバーを誤って強制的に実行しないように十分注意してください。通常は強制は必要ありません。強制的に変更すると、原因のデータが失われる可能性があります。
- テイクオーバーやスイッチオーバーの強制実行が必要な場合は、アプリケーションがシャットダウンされ、すべてのファイルシステムがディスマウントされ、論理ボリュームマネージャ（LVM）ボリュームグループが分離されていることを確認してください。ASMディスクグループをアンマウントする必要があります。
- MetroClusterの強制スイッチオーバーが発生した場合は、障害が発生したノードを残りのすべてのストレージリソースからフェンシングします。詳細については、該当するバージョンのONTAPの『MetroCluster管理およびディザスタリカバリガイド』を参照してください。

MetroClusterと複数のアグリゲート

MetroClusterは同期レプリケーションテクノロジで、接続が中断されると非同期モードに切り替わります。これはお客様からの最も一般的な要求です。同期レプリケーションが保証されていると、サイト接続が中断されるとデータベースI/Oが完全に停止し、データベースのサービスが停止します。

MetroClusterを使用すると、接続がリストアされたあとにアグリゲートが迅速に再同期されます。他のストレージテクノロジとは異なり、MetroClusterでは、サイト障害後に完全な再ミラーリングを行う必要はありません。変更された差分のみを出荷する必要があります。

複数のアグリゲートにまたがるデータセットでは、ローリングディザスタシナリオでデータリカバリ手順を追加する必要があるというわずかなリスクがあります。具体的には、（a）サイト間の接続が中断された場合、（b）接続がリストアされた場合、（c）アグリゲートが同期されている状態と同期されていない状態になった場合、 そして（d）プライマリサイトが失われると、サバイバーサイトが作成され、アグリゲートが相互に同期されなくなります。この場合、データセットの一部が相互に同期され、アプリケーション、データベース、またはデータストアをリカバリなしで起動することはできません。データセットが複数のアグリゲートにまたがっている場合、NetAppでは、Snapshotベースのバックアップと多数のツールのいずれかを活用して、このような異常な状況で迅速にリカバリできるかどうかを検証することを強く推奨しています。

データベース設定

Oracleデータベースのブロックサイズ

ONTAPでは内部的に可変ブロックサイズが使用されるため、Oracleデータベースには任意のブロックサイズを設定できます。ただし、ファイルシステムのブロックサイズがパフォーマンスに影響することがあり、場合によってはRedoブロックサイズを大きくすることでパフォーマンスが向上することがあります。

データファイルのブロックサイズ

OSによっては、ファイルシステムのブロックサイズを選択できます。Oracleデータファイルをサポートするファイルシステムでは、圧縮を使用する場合にブロックサイズを8KBにする必要があります。圧縮が不要な場合は、8KBまたは4KBのブロックサイズを使用できます。

512バイトのブロックを使用するファイルシステムにデータファイルが配置されていると、ファイルのミスアライメントが発生する可能性があります。LUNとファイルシステムはNetAppの推奨事項に基づいて適切にア

ライメントされていても、ファイルI/Oはミスアライメントされます。このようなミスアライメントが発生すると、原因で重大なパフォーマンス問題が発生します。

Redoログをサポートするファイルシステムでは、Redoブロックサイズの倍数のブロックサイズを使用する必要があります。そのためには、通常、RedoログファイルシステムとRedoログ自体の両方で512バイトのブロックサイズを使用する必要があります。

Redoブロックサイズ

Redo率が非常に高い場合は、少ない処理で効率よくI/Oを実行できるため、4KBのブロックサイズの方がパフォーマンスが向上する可能性があります。Redo率が50MBpsを超える場合は、4KBのブロックサイズをテストすることを検討してください。

一部のお客様では、ブロックサイズが512バイトのRedoログをブロックサイズが4KBのファイルシステムで使用し、非常に小さなトランザクションが大量に発生するという問題が報告されています。このパフォーマンスの問題は、複数の512バイトの変更を4KBの単一のファイルシステムブロックに適用する際のオーバーヘッドが原因で発生していましたが、ファイルシステムのブロックサイズを512バイトに変更することで解決されました。

* NetAppでは、関連するカスタマーサポートまたはプロフェッショナルサービス部門から指示があった場合、または正式な製品ドキュメントに基づく場合を除き、Redoブロックサイズを変更しないことを推奨しています*。

Oracleデータベースパラメータ：db_file_multiblock_read_count

。 `db_file_multiblock_read_count` パラメータは、シーケンシャルI/OでOracleが単一の処理として読み取るOracleデータベースブロックの最大数を制御します。

ただし、このパラメータは、すべての読み取り処理でOracleが読み取るブロック数には影響しません。また、ランダムI/Oにも影響しません。影響を受けるのはシーケンシャルI/Oのブロックサイズだけです。

Oracleでは、このパラメータを未設定のままにしておくことを推奨しています。これにより、データベースソフトウェアは自動的に最適な値を設定できます。つまり、このパラメータは通常、I/Oサイズが1MBになる値に設定されます。たとえば、8KBブロックの1MB読み取りでは128ブロックを読み取る必要があるため、このパラメータのデフォルト値は128になります。

顧客サイトのNetAppで発生したデータベースパフォーマンスの問題のほとんどには、このパラメータの設定が誤っています。Oracleバージョン8および9では、この値を変更する正当な理由があります。そのため、パラメータが `init.ora` ファイル：データベースがOracle 10以降にアップグレードされたため。従来の設定である8または16をデフォルト値の128と比較すると、シーケンシャルI/Oのパフォーマンスが大幅に低下します。

* NetApp推奨*設定 `db_file_multiblock_read_count` パラメータがに存在していてはなりません。 `init.ora` ファイル。NetAppでは、このパラメータを変更することでパフォーマンスが向上するという状況は発生していませんが、シーケンシャルI/Oのスループットに明らかな影響を及ぼすケースは少なくありません。

Oracleデータベースパラメータ:filesystemio_options

Oracle初期化パラメータ `filesystemio_options` 非同期I/Oとダイレクト I/Oの使用を制御します。

一般的な考え方に反して、非同期I/Oとダイレクト I/Oは相互に排他的ではありません。NetAppでは、お客様の環境でこのパラメータの設定ミスが頻繁に発生し、この設定ミスが多くのパフォーマンス問題の直接的な原因となっていることを確認しています。

非同期I/Oとは、Oracle I/O処理を並行処理できることを意味します。さまざまなOSで非同期I/Oが使用可能になる前は、ユーザが多数のdbwriterプロセスを設定し、サーバプロセスの設定を変更していました。非同期I/Oでは、OS自体がデータベースソフトウェアに代わって効率的かつ並列的にI/Oを実行します。このプロセスによってデータがリスクにさらされることはなく、OracleのRedoロギングなどの重要な処理も同期的に実行されます。

ダイレクトI/OはOSのバッファキャッシュをバイパスします。UNIXシステムのI/Oは、通常、OSのバッファキャッシュを通過します。これは内部キャッシュを持たないアプリケーションでは便利ですが、OracleはSGA内に独自のバッファキャッシュを備えています。ほとんどの場合、OSのバッファキャッシュを使用するよりも、ダイレクトI/Oを有効にしてサーバRAMをSGAに割り当てる方が適しています。Oracle SGAはメモリをより効率的に使用します。さらに、I/OがOSバッファを通過すると、追加の処理が発生し、レイテンシが増加します。レイテンシの増加は、低レイテンシが重要な要件である書き込みI/Oの負荷が高い場合に特に顕著です。

オフションは `filesystemio_options` 次のとおりです。

- * async。* OracleはI/O要求をOSに送信して処理します。このプロセスにより、OracleはI/Oの完了を待たずに他の処理を実行できるため、I/Oの並列化が促進されます。
- * directio.* Oracleは、ホストOSキャッシュを介してI/Oをルーティングするのではなく、物理ファイルに対して直接I/Oを実行します。
- なし。 Oracleは同期I/Oとバッファ I/Oを使用します。この構成では、共有サーバプロセスと専用サーバプロセスの選択、およびdbwriterの数がより重要になります。
- * SETALL。* Oracleは非同期I/Oとダイレクト I/Oの両方を使用します。ほとんどすべての場合、 `setall` 最適です。

。 `filesystemio_options` パラメータは、DNFS環境とASM環境では効果がありません。DNFSまたはASMを使用すると、自動的に非同期I/Oとダイレクト I/Oの両方が使用されます。

一部のお客様では、特に以前のRed Hat Enterprise Linux 4（RHEL4）リリースで、過去に非同期I/Oの問題が発生していました。インターネット上のいくつかの時代遅れのアドバイスは、時代遅れの情報のために非同期IOを避けることを提案しています。非同期I/Oは、現在のすべてのOSで安定しています。OSの既知のバグがない限り、無効にする理由はありません。

データベースでバッファ I/Oが使用されている場合は、ダイレクト I/Oに切り替えてもSGAサイズの変更が必要になることがあります。バッファ I/Oを無効にすると、ホストOSキャッシュがデータベースに提供するパフォーマンス上のメリットがなくなります。RAMをSGAに再度追加すると、この問題が解決します。最終的には、I/Oパフォーマンスの向上につながります。

RAMはOSのバッファキャッシュよりもOracle SGAに使用する方がほとんどですが、最適な値を特定できない場合もあります。たとえば、断続的にアクティブになるOracleインスタンスが多数あるデータベースサーバでは、SGAサイズが非常に小さいバッファ I/Oを使用することを推奨します。この方法では、実行中のすべてのデータベースインスタンスが、空いているOSのRAMを柔軟に使用できます。これは非常にまれな状況ですが、一部のお客様のサイトで確認されています。

* NetApp推奨*設定 `filesystemio_options` 終了： `setall`ただし、状況によっては、ホストのバッファキャッシュが失われた場合にOracle SGAの拡張が必要になることがあります。

Oracle Real Application Clusters（RAC）タイムアウト

Oracle RACは、クラスタの健常性を監視する複数のタイプの内部ハートビートプロセスを備えたクラスタウェア製品です。

の情報 "MissCount" セクションには、ネットワーク・ストレージを使用するOracle RAC環境に関する重要な情報が記載されています。多くの場合、RACクラスタがネットワーク・パスの変更やストレージのフェイルオーバー/スイッチオーバー操作に耐えられるように、デフォルトのOracle RAC設定を変更する必要があります。

ディスクタイムアウト

プライマリストレージ関連のRACパラメータはです。 `disktimeout`。このパラメータは、投票ファイルI/Oが完了しなければならないしきい値を制御します。状況に応じて `disktimeout` パラメータの値を超えると、そのRACノードがクラスタから削除されます。このパラメータのデフォルトは200です。ストレージのテイクオーバーとギブバックの標準的な手順では、この値で十分です。

テイクオーバーやギブバックには多くの要素が影響するため、NetAppでは、RAC構成を本番環境に導入する前に徹底的にテストすることを強く推奨します。ストレージフェイルオーバーの完了に必要な時間に加えて、Link Aggregation Control Protocol（LACP；リンクアグリゲーション制御プロトコル）の変更が伝播されるまでの時間も長くなります。また、SANマルチパスソフトウェアはI/Oタイムアウトを検出し、代替パスで再試行する必要があります。データベースが非常にアクティブな場合は、投票ディスクI/Oが処理される前に、大量のI/Oをキューに入れて再試行する必要があります。

ストレージのテイクオーバーやギブバックを実際に実行できない場合は、データベースサーバでケーブルを取り外すテストを実行して影響をシミュレートできます。

- NetAppの推奨事項*：
- を終了します。 `disktimeout` パラメータを指定します。デフォルト値は200です。
- RAC構成は常に十分にテストしてください。

MissCount

。 `misscount` パラメータは通常、RACノード間のネットワークハートビートにのみ影響します。デフォルトは30秒です。Gridバイナリがストレージアレイ上にある場合やOSのブートドライブがローカルでない場合は、このパラメータが重要になることがあります。これには、ブートドライブがFC SANに配置されたホスト、NFSブートOS、およびVMDKファイルなどの仮想データストアに配置されたブートドライブが含まれます。

ブートドライブへのアクセスがストレージのテイクオーバーやギブバックによって中断された場合、Gridバイナリの場所またはOS全体が一時的に停止する可能性があります。ONTAPがストレージ処理を完了するのに必要な時間、およびOSがパスを変更してI/Oを再開するのに必要な時間が、 `misscount` しきい値。そのため、ブートLUNまたはGridバイナリへの接続がリストアされたあと、ノードはただちに削除されます。ほとんどの場合、削除とその後のリブートは実行されますが、リブートの理由を示すログメッセージは表示されません。すべての構成に影響するわけではないので、RAC環境内のSANブート、NFSブート、またはデータストアベースのホストをテストして、ブートドライブへの通信が中断してもRACが安定した状態になるようにします。

ローカルでないブートドライブまたはローカルでないファイルシステムをホストしている場合 `grid` バイナリ、 `misscount` 一致するように変更する必要があります `disktimeout`。このパラメータを変更した場合は、さらにテストを行い、ノードのフェイルオーバー時間など、RACの動作への影響を特定します。

- NetAppの推奨事項*：
- そのままにします。 `misscount` 次のいずれかの条件が適用されない場合は、デフォルト値の30のパラメータを使用します。
  - `grid` バイナリが、ネットワークに接続されたドライブ（NFS、iSCSI、FC、データストアベースのドライブを含む）に配置されている。
  - OSがSANブートである。
- このような場合は、ネットワークの中断がOSへのアクセスに影響するか、 `GRID_HOME` ファイルシステム：このような中断によって原因Oracle RACデーモンが停止し、 `misscount`-ベースのタイムアウトおよび削除。タイムアウトのデフォルトは27秒です。これは `misscount` マイナス `reboottime`。このような場合、 `misscount` 200にして一致させる `disktimeout`。

ホストの設定

IBM AIXを使用するOracleデータベース

ONTAPを使用したIBM AIX上のOracleデータベースの構成に関するトピック。

同時I/O

IBM AIXで最適なパフォーマンスを実現するには、同時I/Oを使用する必要があります。AIXはシリアル化されたアトミックなI/Oを実行するため、大量のオーバーヘッドが発生するため、同時I/Oがないとパフォーマンスが制限される可能性があります。

従来のNetAppでは、 `cio` マウントオプション：ファイルシステムで強制的に同時I/Oを使用しますが、このプロセスには欠点があるため不要になりました。AIX 5.2とOracle 10gR1が導入されて以降、AIX上のOracleでは、ファイルシステム全体で同時I/Oを強制的に実行するのではなく、個 々 のファイルを開いて同時IOを実行できるようになりました。

同時I/Oを有効にする最適な方法は、 `init.ora` パラメータ `filesystemio_options` 終了： `setall`。これにより、Oracleが特定のファイルを開いて同時I/Oで使用できるようになります。

を使用します `cio` マウントオプションを指定すると、同時I/Oが強制的に使用されるため、悪影響が生じる可能性があります。たとえば、同時I/Oを強制するとファイルシステムの先読みが無効になり、Oracleデータベースソフトウェアの外部で発生するI/O（ファイルのコピーやテープバックアップの実行など）のパフォーマンスが低下する可能性があります。さらに、Oracle GoldenGateやSAP BR * Toolsなどの製品は、 `cio` 特定のバージョンのOracleでのマウントオプション。

- NetAppの推奨事項*：
- を使用しないでください `cio` ファイルシステムレベルのマウントオプション。代わりに、を使用して同時I/Oを有効にします。 `filesystemio_options=setall`。
- 使用するのは、 `cio` マウントオプションは次のように設定できない場合に実行します： `filesystemio_options=setall`。

AIX NFSのマウントオプション

次の表に、OracleシングルインスタンスデータベースのAIX NFSマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144`
制御ファイル データファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,intr`

次の表に、RACのAIX NFSマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144`
制御ファイル データファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac`
`CRS/Voting`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac`
専用 `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144`
共有 `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr`

シングルインスタンスとRACマウントオプションの主な違いは、 `noac` をマウントオプションに移動します。このオプションを使用するとホストOSのキャッシングが無効になるため、データの状態について、RACクラスタ内のすべてのインスタンスが一貫した情報を認識できるようになります。

ただし、 `cio` マウントオプションと `init.ora` パラメータ `filesystemio_options=setall` ホストのキャッシングを無効にした場合と同じ効果がありますが、引き続きを使用する必要があります。 `noac`。 `noac` 共有の場合は必須です `ORACLE_HOME` OracleパスワードファイルやOracleパスワードファイルなどのファイルの整合性を維持するための導入 `spfile` パラメータファイル。RACクラスタ内の各インスタンスに専用の`ORACLE_HOME`の場合、このパラメータは必要ありません。

AIX JFS / JFS2のマウントオプション

次の表に、AIX JFS / JFS2のマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	デフォルト値です
制御ファイル データファイル REDO ログ	デフォルト値です
ORACLE_HOME を参照してください	デフォルト値です

AIXを使用する前に `hdisk` データベースを含むあらゆる環境のデバイスで、パラメータをチェックします。 `queue_depth`。このパラメータはHBAのキュー深度ではなく、個 々 ののSCSIキュー深度に関連します。 `hdisk device. Depending on how the LUNs are configured, the value for` `queue_depth` パフォーマンスを向上させるには低すぎる可能性があります。テストでは、最適値は64であることが示されています。

HP-UXを使用したOracleデータベース

ONTAPを使用したHP-UX上のOracleデータベースの設定に関するトピック。

HP-UX NFSのマウントオプション

次の表に、単一インスタンスのHP-UX NFSマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`
制御ファイル データファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144,forc edirectio, nointr,suid`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`

次の表に、RACのHP-UX NFSマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, noac,suid`
制御ファイル データファイル REDO ログ	`rw, bg,hard, [vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,forcedirectio,suid`
CRS /投票	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac, forcedirectio,suid`

ファイルタイプ	マウントオプション
専用 ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,suid
共有 ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,suid

シングルインスタンスとRACマウントオプションの主な違いは、 noac および forcedirectio をマウントオプションに移動します。このオプションを使用するとホストOSのキャッシングが無効になるため、データの状態について、RACクラスタ内のすべてのインスタンスが一貫した情報を認識できるようになります。ただし、 init.ora パラメータ filesystemio_options=setall ホストのキャッシングを無効にした場合と同じ効果がありますが、引き続きを使用する必要があります。 noac および forcedirectio。

理由 noac 共有の場合は必須です ORACLE_HOME を導入すると、Oracleパスワードファイルやspfileなどのファイルの整合性が維持されます。RACクラスタ内の各インスタンスに専用の ORACLE_HOME、このパラメータは必須ではありません。

HP-UX VxFSマウントオプション

Oracleバイナリをホストするファイルシステムには、次のマウントオプションを使用します。

```
delaylog,nodatainlog
```

データファイル、Redoログ、アーカイブログ、制御ファイルが格納されているファイルシステムで、HP-UXのバージョンが同時I/Oをサポートしていない場合は、次のマウントオプションを使用します。

```
nodatainlog,mincache=direct,convosync=direct
```

同時I/Oがサポートされている場合（VxFS 5.0.1以降、またはServiceGuard Storage Management Suiteを使用している場合）は、データファイル、Redoログ、アーカイブログ、および制御ファイルを格納しているファイルシステムで次のマウントオプションを使用します。

```
delaylog,cio
```

パラメータ db_file_multiblock_read_count VxFS環境では特に重要です。Oracleでは、特に指示がないかぎり、Oracle 10g R1以降ではこのパラメータを未設定のままにすることを推奨しています。Oracleの8KBブロックサイズの場合、デフォルトは128です。このパラメータの値が強制的に16以下になる場合は、 convosync=direct シーケンシャルI/Oのパフォーマンスが低下する可能性があるため、マウントオプションを使用します。この手順は、パフォーマンスの他の側面に損傷を与えるため、 db_file_multiblock_read_count デフォルト値から変更する必要があります。

Linuxを使用したOracleデータベース

Linux OSに固有の設定に関するトピック。

Linux NFSv3 TCPスロットテーブル

TCPスロットテーブルは、NFSv3でホストバスアダプタ（HBA）のキュー深度に相当します。一度に未処理となることのできるNFS処理の数を制御します。デフォルト値は通常16ですが、最適なパフォーマンスを得るには小さすぎます。逆に、新しいLinuxカーネルでTCPスロットテーブルの上限をNFSサーバが要求でいっぱいになるレベルに自動的に引き上げることができるため、問題が発生します。

パフォーマンスを最適化し、パフォーマンスの問題を回避するには、TCPスロットテーブルを制御するカーネルパラメータを調整します。

を実行します `sysctl -a | grep tcp.*.slot_table` コマンドを実行し、次のパラメータを確認します。

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

すべてのLinuxシステムに `sunrpc.tcp_slot_table_entries`ただし、次のようなものがあります。 `sunrpc.tcp_max_slot_table_entries`。どちらも128に設定する必要があります。

注意
これらのパラメータを設定しないと、パフォーマンスに大きく影響する可能性があります。Linux OSが十分なI/Oを発行していないためにパフォーマンスが制限される場合もあります。一方では、Linux OSが問題で処理できる以上のI/Oを試行すると、I/Oレイテンシが増加します。

Linux NFSのマウントオプション

次の表に、単一インスタンスのLinux NFSのマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144`
制御ファイル データファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr`
ORACLE_HOME を参照してください	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr`

次の表に、RACのLinux NFSマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,actimeo=0
制御ファイル データ・ファイル REDO ログ	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,actimeo=0
CRS /投票	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,actimeo=0
専用 ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144
共有 ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,actimeo=0

シングルインスタンスとRACマウントオプションの主な違いは、 actimeo=0 をマウントオプションに移動します。このオプションを使用するとホストOSのキャッシングが無効になるため、データの状態について、RACクラスタ内のすべてのインスタンスが一貫した情報を認識できるようになります。ただし、 init.ora パラメータ filesystemio_options=setall ホストのキャッシングを無効にした場合と同じ効果がありますが、引き続きを使用する必要があります。 actimeo=0。

理由 actimeo=0 共有の場合は必須です ORACLE_HOME を導入すると、Oracleパスワードファイルやspfileなどのファイルの整合性が維持されます。RACクラスタ内の各インスタンスに専用の `ORACLE_HOME`の場合、このパラメータは必要ありません。

一般に、データベース以外のファイルは、シングルインスタンスのデータファイルと同じオプションを使用してマウントします。ただしアプリケーションによっては要件が異なる場合があります。マウントオプションを使用しない noac および actimeo=0 これらのオプションは、ファイルシステムレベルの先読みとバッファリングを無効にするため、可能であれば可能です。これにより、原因抽出、変換、ロードなどのプロセスで重大なパフォーマンスの問題が発生する可能性があります。

ACCESSとGETATTR

一部のお客様は、ACCESSやGETATTRなどのIOPSがワークロードを占有する可能性が非常に高いことを指摘しています。極端なケースでは、読み取りや書き込みなどの処理が全体の10%にまで低下することがあります。これは、を含むデータベースでは正常に動作します。 actimeo=0 および / または noac Linuxの場合：これらのオプションは、Linux OSを原因して、ストレージシステムからファイルメタデータを定期的にリロードします。ACCESSやGETATTRなどの処理は影響力の低い処理で、データベース環境ではONTAPキャッシュから処理されます。読み取りや書き込みなど、ストレージシステムに真の需要を生み出す純粋なIOPSとみなすべきではありません。ただし、特にRAC環境では、これらのIOPSにはある程度の負荷がかかります。この状況に対処するには、DNFSを有効にして、OSのバッファキャッシュをバイパスし、不要なメタデータ処理を回避します。

Linux Direct NFS

もう1つのマウントオプション（ nosharecache`は、（a）DNFSが有効で、（b）1つのソースボリュームが1つのサーバ（c）に複数回マウントされ、NFSマウントがネストされている場合に必要です。この構成は、主にSAPアプリケーションをサポートしている環境で見られます。たとえば、NetAppシステム上の1つのボリュームに、次の場所にディレクトリを配置できます。 `/vol/oracle/base 1秒前に

/vol/oracle/home。状況 /vol/oracle/base はにマウントされます。 /oracle および /vol/oracle/home はにマウントされます。 `/oracle/home`を指定すると、同じソースからのNFSマウントがネストされます。

OSは、次の事実を検出できます。 /oracle および /oracle/home 同じボリューム（同じソースファイルシステム）に配置します。その後、OSは同じデバイスハンドルを使用してデータにアクセスします。これにより、OSキャッシングなどの特定の処理の使用が改善されますが、DNFSの妨げになります。DNFSが次のようなファイルにアクセスする必要がある場合 spfile、オン /oracle/home`は、誤ってデータへの間違ったパスを使用しようとする可能性があります。その結果、I/O処理が失敗します。これらの構成では、`nosharecache マウントオプションは、ソースFlexVolボリュームをそのホスト上の別のNFSファイルシステムと共有する任意のNFSファイルシステムに適用されます。これにより、Linux OSはそのファイルシステムに独立したデバイスハンドルを割り当てるようになります。

Linux Direct NFSとOracle RAC

Linux OS上のOracle RACでは、ノード間の一貫性を維持するためにRACで必要となるダイレクトI/Oを強制的に実行する方法がLinuxにないため、DNFSを使用するとパフォーマンスが特別に向上します。Linuxを回避策として使用するには、 actimeo=0 マウントオプション。OSキャッシュからファイルデータがただちに期限切れになります。このオプションを使用すると、Linux NFSクライアントは属性データを定期的に再読み取りするため、レイテンシが低下し、ストレージコントローラの負荷が増加します。

DNFSを有効にすると、ホストNFSクライアントがバイパスされ、この被害を回避できます。DNFSを有効にしたところ、RACクラスタのパフォーマンスが大幅に向上し、（特に他のIOPSに関して）ONTAPの負荷が大幅に減少したという報告が複数のお客様から寄せられています。

Linux Direct NFSとoranfstabファイル

マルチパスオプションを指定してLinuxでDNFSを使用する場合は、複数のサブネットを使用する必要があります。他のOSでは、を使用して複数のDNFSチャネルを確立できます。 LOCAL および DONTROUTE 単一のサブネット上に複数のDNFSチャネルを設定するためのオプション。ただし、これはLinuxでは正しく機能せず、予期しないパフォーマンスの問題が発生する可能性があります。Linuxでは、DNFSトラフィックに使用するNICをそれぞれ別 々 のサブネットに配置する必要があります。

I/Oスケジューラ

Linuxカーネルでは、ブロックデバイスへのI/Oのスケジュール方法を低レベルで制御できます。デフォルト値はLinuxのディストリビューションによって大きく異なります。テストでは、通常はDeadlineが最良の結果を提供することが示されていますが、場合によってはNOOPがわずかに改善されています。パフォーマンスの違いはごくわずかですが、データベース構成から最大限のパフォーマンスを引き出す必要がある場合は、両方のオプションをテストしてください。CFQは多くの構成でデフォルトであり、データベースワークロードのパフォーマンスに重大な問題があることが実証されています。

I/Oスケジューラの設定手順については、該当するLinuxベンダーのドキュメントを参照してください。

マルチパス

一部のお客様では、マルチパスデーモンがシステムで実行されていなかったために、ネットワーク停止中にクラッシュが発生しました。最近のバージョンのLinuxでは、OSとマルチパスデーモンのインストールプロセスによって、これらのOSがこの問題に対して脆弱なままになる可能性があります。パッケージは正しくインストールされていますが、再起動後の自動起動が設定されていません。

たとえば、RHEL5.5のマルチパスデーモンのデフォルトは次のようになります。

```
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:off   3:off   4:off   5:off   6:off
```

これを修正するには、次のコマンドを使用します。

```
[root@host1 iscsi]# chkconfig multipathd on
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:on    3:on    4:on    5:on    6:off
```

ASMミラーリング

ASM ミラーリングでは、 ASM が問題を認識して代替の障害グループに切り替えるために、 Linux マルチパス設定の変更が必要になる場合があります。ONTAP 上のほとんどの ASM 構成では、外部冗長性が使用されます。つまり、データ保護は外部アレイによって提供され、 ASM はデータをミラーリングしません。一部のサイトでは、通常の冗長性を備えた ASM を使用して、通常は異なるサイト間で双方向ミラーリングを提供しています。

に表示されるLinux設定 "NetApp Host Utilitiesのマニュアル" I/Oが無期限にキューイングされるマルチパスパラメータを指定します。つまり、アクティブなパスがないLUNデバイス上のI/Oは、I/Oが完了するまで待機します。これは、SANパスの変更が完了するまで、FCスイッチがリブートするまで、またはストレージシステムがフェイルオーバーを完了するまで、Linuxホストが必要な時間だけ待機するために、通常は推奨されます。

この無制限のキューイング動作により、ASMミラーリングで問題が発生します。ASMは、代替LUNでI/Oを再試行するためにI/O障害を受信する必要があるためです。

Linuxで次のパラメータを設定します。 `multipath.conf` ASMミラーリングで使用されるASM LUNのファイル：

```
polling_interval 5
no_path_retry 24
```

これらの設定により、ASMデバイスに120秒のタイムアウトが作成されます。タイムアウトは、 `polling_interval` * `no_path_retry` 秒として。状況によっては正確な値の調整が必要になる場合がありますが、ほとんどの場合は120秒のタイムアウトで十分です。具体的には、コントローラのテイクオーバーまたはギブバックが120秒以内に実行され、I/Oエラーが発生しないようにしてください。この場合、障害グループはオフラインになります。

A下限 `no_path_retry` この値を指定すると、ASMが代替障害グループに切り替えるのに必要な時間を短縮できますが、これにより、コントローラのテイクオーバーなどのメンテナンス作業中に不要なフェイルオーバーが発生するリスクも高まります。ASMミラーリングの状態を注意深く監視することで、このリスクを軽減できます。不要なフェイルオーバーが発生した場合、再同期が比較的短時間で実行されると、ミラーを迅速に再同期できます。追加情報については、使用しているOracleソフトウェアのバージョンに対応するASM高速ミラー再同期に関するOracleのマニュアルを参照してください。

Linuxのxfs、ext3、ext4のマウントオプション

* NetAppでは*デフォルトのマウントオプションを使用することを推奨しています。

ASMLib/AFD（ASM Filter Driver）を使用するOracleデータベース

AFDとASMlibを使用するLinux OSに固有の設定トピック

ASMlibブロックサイズ

ASMlibは、オプションのASM管理ライブラリおよび関連ユーティリティです。その主な価値は、LUNまたはNFSベースのファイルにASMリソースとして人間が判読可能なラベルを付けることです。

ASMlibの最近のバージョンでは、Logical Blocks Per Physical Block Exponent（LBPPBE）というLUNパラメータが検出されています。最近まで、この値はONTAP SCSIターゲットによって報告されていませんでした。4KBのブロックサイズが推奨されることを示す値が返されるようになりました。これはブロックサイズの定義ではありませんが、LBPPBEを使用するアプリケーションにとって、特定のサイズのI/Oがより効率的に処理される可能性があることを示唆しています。ただし、ASMlibはLBPPBEをブロックサイズとして解釈し、ASMデバイスの作成時にASMヘッダーを永続的にスタンプします。

このプロセスは、さまざまな方法でアップグレードや移行で原因の問題を引き起こす可能性があります。すべては、同じASMディスクグループにブロックサイズの異なるASMlibデバイスを混在させることができないことが原因です。

たとえば、古いアレイでは通常、LBPPBE値が0と報告されているか、この値がまったく報告されていませんでした。ASMlibはこれを512バイトのブロックサイズと解釈します。新しいアレイは、4KBのブロックサイズと解釈されます。512バイトと4KBのデバイスを同じASMディスクグループに混在させることはできません。これにより、2つのアレイのLUNを使用してASMディスクグループのサイズを拡張したり、ASMを移行ツールとして活用したりすることができなくなります。それ以外の場合、RMANでは、512バイトのブロックサイズのASMディスクグループと4KBのブロックサイズのASMディスクグループの間でファイルを複製できないことがあります。

推奨される解決策は、ASMlibにパッチを適用することです。OracleのバグIDは13999609で、パッチはoracleasm-support-2.1.8-1以降に存在します。このパッチを適用すると、ユーザーはパラメータを設定できます。 `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` 終了： `true` を参照してください `/etc/sysconfig/oracleasm` 構成ファイルこれにより、ASMlibはLBPPBEパラメータを使用できなくなります。つまり、新しいアレイ上のLUNが512バイトのブロックデバイスとして認識されるようになります。

このオプションを使用しても、以前にASMlibによってスタンプされたLUNのブロックサイズは変更されません。たとえば、ブロック数が512バイトのASMディスクグループを、ブロック数が4KBと報告される新しいストレージシステムに移行する必要がある場合は、オプション `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` 新しいLUNがASMlibでスタンプされる前に設定する必要があります。 デバイスがoracleasmによってすでにスタンプされている場合は、新しいブロックサイズで再スタンプする前に再フォーマットする必要があります。まず、デバイスの設定を解除します。 `oracleasm deletedisk`をクリックし、デバイスの最初の`1GB`を消去します。 `dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`。最後に、デバイスが以前にパーティション分割されていた場合は、 `kpartx` 古いパーティションを削除するか、単にOSを再起動するためのコマンド。

ASMlibにパッチを適用できない場合は、ASMlibを構成から削除できます。この変更はシステムの停止を伴うため、ASMディスクのスタンプを解除し、 `asm_diskstring` パラメータが正しく設定されている。ただし、この変更ではデータの移行は必要ありません。

ASMフィルタドライブ（AFD）のブロックサイズ

AFDは、ASMlibに代わるオプションのASM管理ライブラリです。ストレージの観点から見ると、ASMlibはASMlibに非常に似ていますが、Oracle以外のI/Oをブロックして、データが破損する可能性のあるユーザーまたはアプリケーションのエラーの可能性を減らすなどの追加機能が含まれています。

デバイスのブロックサイズ

ASMlibと同様に、AFDもLUNパラメータLogical Blocks Per Physical Block Exponent（LBPPBE）を読み取り、デフォルトでは論理ブロックサイズではなく物理ブロックサイズを使用します。

ASMデバイスがすでに512バイトのブロックデバイスとしてフォーマットされている既存の構成にAFDを追加すると、問題が発生する可能性があります。AFDドライバはLUNを4Kデバイスとして認識し、ASMラベルと物理デバイスの不一致が原因でアクセスできなくなります。同様に、512バイトと4KBのデバイスを同じASMディスクグループに混在させることはできないため、移行も影響を受けます。これにより、2つのアレイのLUNを使用してASMディスクグループのサイズを拡張したり、ASMを移行ツールとして活用したりすることができなくなります。それ以外の場合、RMANでは、512バイトのブロックサイズのASMディスクグループと4KBのブロックサイズのASMディスクグループの間でファイルを複製できないことがあります。

解決策はシンプルです- AFDには、論理ブロックサイズと物理ブロックサイズのどちらを使用するかを制御するパラメータが含まれています。これは、システム上のすべてのデバイスに影響を与えるグローバルパラメータです。AFDで強制的に論理ブロックサイズを使用するには、 `options oracleafd oracleafd_use_logical_block_size=1` を参照してください `/etc/modprobe.d/oracleafd.conf` ファイル。

マルチパステンソウサイス

最近のLinuxカーネルの変更では、マルチパスデバイスに送信されるI/Oサイズ制限が適用されますが、AFDではこれらの制限が適用されません。その後I/Oが拒否され、LUNパスがオフラインになります。その結果、Oracle Gridのインストール、ASMの設定、データベースの作成ができなくなります。

解決策では、ONTAP LUNのmultipath.confファイルに最大転送長を手動で指定します。

```
devices {
             device {
                 vendor "NETAPP"
                 product "LUN.*"
                 max_sectors_kb 4096
         }
     }
```

現在問題が存在しない場合でも、AFDを使用して将来のLinuxアップグレードで予期せず原因の問題が発生しないようにする場合は、このパラメータを設定する必要があります。

Microsoft Windowsを使用したOracleデータベース

ONTAPを使用したMicrosoft Windows上のOracleデータベースの構成に関するトピック

NFS

Oracleでは、Direct NFSクライアントでのMicrosoft Windowsの使用がサポートされています。この機能は、複数の環境にわたるファイルの表示、ボリュームの動的なサイズ変更、安価なIPプロトコルの活用など、NFSの管理上のメリットをもたらします。DNFSを使用してMicrosoft Windowsにデータベースをインストールおよび設定する方法については、Oracleの公式ドキュメントを参照してください。特別なベストプラクティスはありません。

SAN

圧縮効率を最適化するには、NTFSファイルシステムで8K以上の割り当て単位を使用するようにしてください。一般にデフォルトである4Kの割り当て単位を使用すると、圧縮効率が低下します。

Solarisを使用したOracleデータベース

Solaris OSに固有の構成に関するトピック

Solaris NFSのマウントオプション

次の表に、単一インスタンスのSolaris NFSのマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1], roto=tcp, timeo=600,rsize=262144,wsize=262144`
制御ファイル データファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,llock,suid`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`

の使用 `llock` ストレージシステムでロックを取得および解放する際のレイテンシを排除することで、お客様の環境のパフォーマンスが劇的に向上することが実証されています。多数のサーバが同じファイルシステムをマウントするように構成され、Oracleがこれらのデータベースをマウントするよう構成されている環境では、このオプションの使用に注意してください。これは非常に珍しい構成ですが、少数のお客様によって使用されています。インスタンスが誤って2回目に開始された場合、Oracleは外部サーバ上のロックファイルを検出できないため、データが破損する可能性があります。NFSロックは、NFSバージョン3のように保護を提供するものではなく、推奨されるだけです。

なぜなら、 `llock` および `forcedirectio` パラメータは相互に排他的です。次のことが重要です。 `filesystemio_options=setall` は、 `init.ora` ファイルを作成して `directio` を使用します。このパラメータを指定しないと、ホストOSのバッファキャッシュが使用され、パフォーマンスが低下する可能性があります。

次の表に、Solaris NFSのマウントオプションを示します。

ファイルタイプ	マウントオプション
ADRホーム	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,noac`
制御ファイル データ・ファイル REDO ログ	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,forcedirectio`
CRS /投票	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,forcedirectio`
専用 `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,suid`
共有 `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsize=262144,nointr,noac,suid`

シングルインスタンスとRACマウントオプションの主な違いは、 `noac` および `forcedirectio` をマウントオプションに移動します。このオプションを使用するとホストOSのキャッシングが無効になるため、データの状態について、RACクラスタ内のすべてのインスタンスが一貫した情報を認識できるようになります。ただし、 `init.ora` パラメータ `filesystemio_options=setall` ホストのキャッシングを無効にした場合と同じ効果がありますが、引き続きを使用する必要があります。 `noac` および `forcedirectio`。

理由 `actimeo=0` 共有の場合は必須です `ORACLE_HOME` を導入すると、Oracleパスワードファイルやspfileなどのファイルの整合性が維持されます。RACクラスタ内の各インスタンスに専用の `ORACLE_HOME`、このパラメータは必須ではありません。

Solaris UFSのマウントオプション

NetAppでは、ロギングマウントオプションを使用して、Solarisホストがクラッシュしたり**FC**接続が中断したりした場合にデータの整合性が維持されるようにすることを強く推奨しています。ロギングマウントオプションを使用すると、Snapshotバックアップのユーザビリティも維持されます。

Solaris ZFS

最適なパフォーマンスを実現するには、Solaris ZFSをインストールして慎重に設定する必要があります。

mvector

Solaris 11では、大規模なI/O処理の処理方法が変更され、SANストレージアレイのパフォーマンスに重大な問題が発生する可能性がありました。この問題の詳細については、NetAppバグレポート630173「Solaris 11 ZFSパフォーマンスの回帰」を参照してください。"The解決策is to change a OS parameter called `zfs_mvector_max_size`。

rootとして次のコマンドを実行します。

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t131072" |mdb -kw
```

この変更によって予期しない問題が発生した場合は、次のコマンドをrootとして実行することで簡単に元に戻すことができます。

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t1048576" |mdb -kw
```

カーネル

信頼性の高いZFSパフォーマンスを実現するには、LUNのアライメントの問題に対してSolarisカーネルにパッチを適用する必要があります。この修正は、Solaris 10のパッチ147440-19とSolaris 11のSRU 10.5で導入されました。ZFSではSolaris 10以降のみを使用してください。

LUN構成

LUNを設定するには、次の手順を実行します。

1. タイプがのLUNを作成します。 `solaris`。
2. で指定された適切なHost Utility Kit（HUK）をインストールします。 "ネットアップの Interoperability Matrix Tool （ IMT ）"。
3. HUKに記載されている手順に正確に従ってください。基本的な手順は以下のとおりですが、 "最新のドキュメント" を参照してください手順。
   a. を実行します `host_config` を更新するユーティリティ `sd.conf/sdd.conf` ファイル。これにより、SCSIドライブがONTAP LUNを正しく検出できるようになります。
   b. の指示に従ってください。 `host_config` Multipath Input/Output（MPIO；マルチパス入出力）を有効にするユーティリティ。
   c. リブートします。この手順は、システム全体で変更が認識されるようにするために必要です。
4. LUNをパーティショニングし、適切にアライメントされていることを確認します。アライメントを直接テストして確認する方法については、「付録B：WAFLアライメントの検証」を参照してください。

zpools

zpoolは、の手順を実行したあとに作成する必要があります。 "LUNの設定" が実行されます。手順を正しく実行しないと、I/Oのアライメントが原因でパフォーマンスが大幅に低下する可能性があります。ONTAPのパフォーマンスを最適化するには、I/Oがドライブの4Kの境界にアライメントされている必要があります。zpoolに作成されるファイルシステムは、というパラメータで制御される実効ブロックサイズを使用します。 `ashift`（コマンドを実行すると表示できます） `zdb -C`。

の値 `ashift` デフォルトは9です。これは2^9、つまり512バイトを意味します。最適なパフォーマンスを実現するには、 `ashift` 値は12（2^12=4K）である必要があります。この値はzpoolの作成時に設定され、変更することはできません。つまり、 `ashift` 12以外の場合は、新しく作成したzpoolにデータをコピーして移行する必要があります。

zpoolを作成したら、の値を確認します。 `ashift` 次に進む前に、値が12以外の場合は、LUNが正しく検出されていません。zpoolを削除し、関連するHost Utilitiesのドキュメントに記載された手順をすべて正しく実行したことを確認してから、zpoolを再作成します。

zpoolとSolaris LDOM

Solaris LDOMには、I/Oアライメントが正しいことを確認するための追加の要件があります。LUNは4Kデバイ

スとして適切に検出されますが、LDOM上の仮想vdskデバイスはI/Oドメインの設定を継承しません。このLUNに基づくvdskは、デフォルトで512バイトブロックに戻ります。

追加の構成ファイルが必要です。まず、追加の設定オプションを有効にするために、個 々 のLDOMにOracleのバグ15824910のパッチを適用する必要があります。このパッチは、現在使用されているすべてのバージョンのSolarisに移植されています。LDOMにパッチを適用すると、適切にアライメントされた新しいLUNを設定できるようになります。手順は次のとおりです。

1. 新しいzpoolで使用するLUNを特定します。この例では、c2d1デバイスです。

```
[root@LDOM1 ~]# echo | format
Searching for disks...done
AVAILABLE DISK SELECTIONS:
  0. c2d0 <Unknown-Unknown-0001-100.00GB>
     /virtual-devices@100/channel-devices@200/disk@0
  1. c2d1 <SUN-ZFS Storage 7330-1.0 cyl 1623 alt 2 hd 254 sec 254>
     /virtual-devices@100/channel-devices@200/disk@1
```

2. ZFSプールに使用するデバイスのVDCインスタンスを取得します。

```
[root@LDOM1 ~]#  cat /etc/path_to_inst
#
# Caution! This file contains critical kernel state
#
"/fcoe" 0 "fcoe"
"/iscsi" 0 "iscsi"
"/pseudo" 0 "pseudo"
"/scsi_vhci" 0 "scsi_vhci"
"/options" 0 "options"
"/virtual-devices@100" 0 "vnex"
"/virtual-devices@100/channel-devices@200" 0 "cnex"
"/virtual-devices@100/channel-devices@200/disk@0" 0 "vdc"
"/virtual-devices@100/channel-devices@200/pciv-communication@0" 0 "vpci"
"/virtual-devices@100/channel-devices@200/network@0" 0 "vnet"
"/virtual-devices@100/channel-devices@200/network@1" 1 "vnet"
"/virtual-devices@100/channel-devices@200/network@2" 2 "vnet"
"/virtual-devices@100/channel-devices@200/network@3" 3 "vnet"
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc" << We want
this one
```

3. 編集 `/platform/sun4v/kernel/drv/vdc.conf`：

```
block-size-list="1:4096";
```

つまり、デバイスインスタンス1には4096のブロックサイズが割り当てられます。

追加の例として、vdskインスタンス1~6を4Kブロックサイズに設定する必要があり、`/etc/path_to_inst` 読み取り値は次のとおりです。

```
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc"
"/virtual-devices@100/channel-devices@200/disk@2" 2 "vdc"
"/virtual-devices@100/channel-devices@200/disk@3" 3 "vdc"
"/virtual-devices@100/channel-devices@200/disk@4" 4 "vdc"
"/virtual-devices@100/channel-devices@200/disk@5" 5 "vdc"
"/virtual-devices@100/channel-devices@200/disk@6" 6 "vdc"
```

4. 決勝戦 `vdc.conf` ファイルには以下が含まれている必要があります

```
block-size-list="1:8192","2:8192","3:8192","4:8192","5:8192","6:8192";
```

注意
vdc.confを設定してvdskを作成したら、LDOMをリブートする必要があります。この手順は避けられません。ブロックサイズの変更はリブート後にのみ有効になります。zpoolの設定に進み、前述のようにashiftが12に正しく設定されていることを確認します。

ZFSインテントログ（ZIL）

通常'ZFSインテントログ(ZIL)を別のデバイスに配置する理由はありませんログはメインプールとスペースを共有できます。ZILを別 々 に使用する主な用途は、最新のストレージアレイには書き込みキャッシュ機能がない物理ドライブを使用する場合です。

ロバイアス

を設定します `logbias` Oracleデータをホストするzfsファイルシステムのパラメータ。

```
zfs set logbias=throughput <filesystem>
```

このパラメータを使用すると、全体的な書き込みレベルが低下します。デフォルトでは、書き込まれたデータはまずZILにコミットされ、次にメインのストレージプールにコミットされます。このアプローチは、SSDベースのZILデバイスとメインストレージプール用の回転式メディアを含む、プレーンドライブ構成を使用する構成に適しています。これは、利用可能な最も低レイテンシのメディア上の単一のI/Oトランザクションでコミットを実行できるためです。

独自のキャッシュ機能を備えた最新のストレージアレイを使用する場合は、通常、このアプローチは必要ありません。まれに、レイテンシの影響を受けやすい大量のランダム書き込みで構成されるワークロードのように、単一のトランザクションで書き込みをログにコミットした方が望ましい場合があります。ログに記録されたデータは最終的にメインのストレージプールに書き込まれ、書き込みアクティビティが2倍になるため、ライトアンプリフィケーションという結果になります。

ダイレクトI/O

Oracle製品を含む多くのアプリケーションでは、ダイレクトI/Oを有効にすることでホストのバッファキャッシュをバイパスできます。ZFSファイルシステムでは、この方法は想定どおりに機能しません。ホストのバッファキャッシュはバイパスされますが、ZFS自体はデータのキャッシュを継続します。I/Oがストレージシステムに到達しているかどうか、またはI/OがOS内にローカルにキャッシュされているかどうかを予測することが困難であるため、FIOやSIOなどのツールを使用してパフォーマンステストを実行すると、誤った結果になる可能性があります。また、このような総合的なテストを使用してZFSと他のファイルシステムのパフォーマンスを比較することも非常に困難になります。実際のユーザワークロードでは、ファイルシステムのパフォーマンスにほとんど違いはありません。

複数の**zpool**

ZFSベースのデータのSnapshotベースのバックアップ、リストア、クローニング、アーカイブは、zpoolレベルで実行する必要があり、通常は複数のzpoolが必要です。zpoolはLVMディスクグループに似ており、同じルールを使用して設定する必要があります。たとえば、データベースのレイアウトには、データファイルが配置されているのが最適です。 `zpool1` およびにあるアーカイブログ、制御ファイル、REDOログ `zpool2`。このアプローチでは、データベースがホットバックアップモードに設定された標準のホットバックアップに続いて、 `zpool1`。次に、データベースがホットバックアップモードから削除され、ログアーカイブが強制的に実行され、 `zpool2` が作成されます。リストア処理では、zfsファイルシステムをアンマウントし、zpoolを完全にオフラインにしてから、SnapRestoreのリストア処理を実行する必要があります。その後、zpoolをオンラインに戻してデータベースをリカバリできます。

ファイルシステムオプション

Oracleパラメータ `filesystemio_options` ZFSでは動作が異なります。状況 `setall` または `directio` を使用します。書き込み処理は同期でOSのバッファキャッシュをバイパスしますが、読み取りはZFSによってバッファされます。この場合、I/OがZFSキャッシュによって代行受信されて処理されることがあるため、ストレージのレイテンシと総I/Oが想定よりも低くなるため、パフォーマンス分析が困難になります。

ネットワーク構成：

Oracleデータベース向けの論理インターフェイス設計

Oracleデータベースにはストレージへのアクセスが必要です。Logical Interface（LIF；論理インターフェイス）は、Storage Virtual Machine（SVM）をネットワークに接続し、さらにデータベースに接続するネットワーク配管です。各データベースワークロードに十分な帯域幅を確保し、フェイルオーバーによってストレージサービスが失われないようにするには、LIFを適切に設計する必要があります。

このセクションでは、LIFの主な設計原則の概要を説明します。より包括的なドキュメントについては、"ONTAPネットワーク管理に関するドキュメント"。データベースアーキテクチャの他の要素と同様に、Storage Virtual Machine（SVM、CLIではVserver）と論理インターフェイス（LIF）の設計に最適なオプションは、拡張要件とビジネスニーズに大きく依存します。

LIFの戦略を立てる際は、主に次の点を考慮してください。

- *パフォーマンス。*ネットワーク帯域幅は十分か。
- *耐障害性。*設計に単一点障害はありますか？
- *管理性。*ネットワークを無停止で拡張できますか？

これらのトピックは、ホストからスイッチ、ストレージシステムまで、エンドツーエンドの解決策に適用されます。

LIFタイプ

LIFには複数のタイプがあります。 "LIFタイプに関するONTAPのドキュメント" このトピックのより包括的な情報を提供しますが、機能的にはLIFを次のグループに分類できます。

- *クラスタおよびノードの管理LIF。*ストレージクラスタの管理に使用するLIF。
- * SVM管理LIF。* REST APIまたはONTAPI（ZAPIとも呼ばれます）を使用してSVMへのアクセスを許可するインターフェイス。Snapshotの作成やボリュームのサイズ変更などの機能に使用できます。SnapManager for Oracle（SMO）などの製品では、SVM管理LIFにアクセスする必要があります。
- データ**LIF**。 FC、iSCSI、NVMe/FC、NVMe/TCP、NFS、 またはSMB / CIFSデータ。

> ファイアウォールポリシーを `data` 終了： `mgmt` または、HTTP、HTTPS、SSHを許可する別のポリシー。この変更により、NFSデータLIFと別の管理LIFの両方にアクセスするように各ホストを設定する必要がなくなるため、ネットワーク設定が簡易化されます。iSCSIトラフィックと管理トラフィックの両方にIPプロトコルを使用しているにもかかわらず、インターフェイスを設定することはできません。iSCSI環境では、個別の管理LIFが必要です。

SAN LIFの設計

SAN環境でのLIFの設計は、マルチパスという1つの理由で比較的簡単です。最新のSAN実装では、クライアントは複数の独立したネットワークパス経由でデータにアクセスし、アクセスに最適なパス（複数可）を選択できます。その結果、SANクライアントは使用可能な最適なパス間でI/Oの負荷を自動的に分散するため、パフォーマンスに関してはLIFの設計は容易に対処できます。

あるパスが使用できなくなった場合、クライアントは自動的に別のパスを選択します。その結果、設計がシンプルになるため、一般にSAN LIFの管理性が向上します。だからといって、SAN環境の方が常に簡単に管理できるわけではありません。SANストレージには、NFSよりもはるかに複雑な要素が多数あるからです。単純に、SAN LIFの設計が容易であることを意味します。

パフォーマンス

SAN環境でLIFのパフォーマンスを考慮する際に最も重要な考慮事項は、帯域幅です。たとえば、2ノードONTAP AFFクラスタの各ノードに16Gb FCポートを2つ搭載すると、各ノードとの間で最大32Gbの帯域幅を確保できます。

耐障害性

AFFストレージシステムでは、SAN LIFはフェイルオーバーしません。コントローラのフェイルオーバーが原因でSAN LIFに障害が発生すると、クライアントのマルチパスソフトウェアがパスの損失を検出し、I/Oを別のLIFにリダイレクトします。ASAストレージシステムでは、LIFは短時間でフェイルオーバーされますが、もう一方のコントローラにすでにアクティブなパスがあるためIOが中断されることはありません。フェイルオーバープロセスは、定義されたすべてのポートでホストアクセスをリストアするために実行されます。

管理性

NFS環境では、クラスタ内でのボリュームの再配置にLIFの移行が伴うことが多いため、LIFの移行ははるかに一般的なタスクです。SAN環境でHAペア内でボリュームを再配置しても、LIFを移行する必要はありません。ボリュームの移動が完了すると、ONTAPはパスの変更をSANに通知し、SANクライアントは自動的に再最適

化します。SANを使用したLIFの移行は、主に物理ハードウェアの大幅な変更に関連しています。たとえば、コントローラの無停止アップグレードが必要な場合は、SAN LIFを新しいハードウェアに移行します。FCポートで障害が検出された場合は、LIFを未使用のポートに移行できます。

設計上の推奨事項

NetAppの推奨事項は次のとおりです。

- 必要以上の数のパスを作成しないでください。パスの数が多すぎると管理全体が複雑になり、一部のホストでのパスのフェイルオーバーで原因の問題が発生する可能性があります。さらに、一部のホストでは、SANブートなどの構成でパスが予期せず制限されます。
- ごく少数の構成では、LUNへのパスが4つ以上必要です。LUNにパスをアドバタイズするノードが3つ以上あると、LUNを所有するノードとそのHAパートナーに障害が発生した場合、LUNをホストしているアグリゲートにアクセスできなくなるため、その価値には制限があります。このような状況では、プライマリHAペア以外のノードにパスを作成しても役に立ちません。
- 参照可能なLUNパスの数はFCゾーンに含めるポートを選択することで管理できますが、一般には、ターゲットとなるポイントをすべてFCゾーンに含め、LUNの可視性をONTAPレベルで制御する方が簡単です。
- ONTAP 8.3以降では、選択的LUNマッピング（SLM）機能がデフォルトです。SLMを使用すると、新しいLUNはすべて、基盤となるアグリゲートを所有するノードとノードのHAパートナーから自動的に通知されます。これにより、ポートのアクセス性を制限するためにポートセットを作成したりゾーニングを設定したりする必要がなくなります。各LUNは、最適なパフォーマンスと耐障害性の両方を実現するために必要な最小限のノードで利用できます。
- LUNを2台のコントローラの外部に移行する必要がある場合は、 `lun mapping add-reporting-nodes` コマンドを実行して、新しいノードでLUNがアドバタイズされるようにします。これにより、LUNの移行用にLUNへの追加のSANパスが作成されます。ただし、新しいパスを使用するには、ホストで検出処理を実行する必要があります。
- 間接トラフィックを過度に気にしないでください。I/Oが大量に発生する環境ではレイテンシがマイクロ秒単位で重要になるため、間接トラフィックは避けることを推奨しますが、一般的なワークロードではパフォーマンスに目に見える影響はごくわずかです。

NFS LIFの設計

NFSでは、SANプロトコルとは異なり、データへの複数のパスを定義する機能に制限があります。NFSv4に対するParallel NFS（pNFS）拡張ではこの制限に対応していますが、イーサネットの速度が100GB以上に達しているため、パスを追加する価値があることはほとんどありません。

パフォーマンスと耐障害性

SAN LIFのパフォーマンスを測定することは、主にすべてのプライマリパスの合計帯域幅を計算することですが、NFS LIFのパフォーマンスを判断するには、正確なネットワーク構成を詳しく調べる必要があります。たとえば、2つの10Gbポートを物理ポートとして構成することも、Link Aggregation Control Protocol（LACP）インターフェイスグループとして構成することもできます。インターフェイスグループとして設定されている場合は、複数のロードバランシングポリシーを使用できます。ロードバランシングポリシーの動作は、トラフィックがスイッチングされるかルーティングされるかによって異なります。最後に、Oracle Direct NFS（dNFS）は、現時点ではどのOS NFSクライアントにも存在しないロードバランシング設定を提供します。

SANプロトコルとは異なり、NFSファイルシステムにはプロトコルレイヤでの耐障害性が必要です。たとえば、LUNは常にマルチパスを有効にして設定されるため、ストレージシステムではFCプロトコルを使用する複数の冗長チャネルを使用できます。一方NFSファイルシステムは、物理レイヤでのみ保護できる単一のTCP/IPチャネルの可用性に依存します。このような理由から、ポートフェイルオーバーやLACPポートアグリゲーションなどのオプションが用意されています。

NFS環境では、パフォーマンスと耐障害性の両方がネットワークプロトコルレイヤで提供されます。その結果、両方のトピックが絡み合っており、一緒に議論する必要があります。

ポートグループへの**LIF**のバインド

LIFをポートグループにバインドするには、LIFのIPアドレスを物理ポートのグループに関連付けます。物理ポートを1つに集約する主な方法はLACPです。LACPのフォールトトレランス機能は非常に簡単です。LACPグループ内の各ポートは監視され、障害が発生した場合はポートグループから削除されます。ただし、パフォーマンスに関してLACPがどのように機能するかについては、多くの誤解があります。

- LACPでは、エンドポイントと一致するようにスイッチで設定する必要はありません。たとえば、ONTAPにIPベースのロードバランシングを設定し、スイッチにMACベースのロードバランシングを使用することができます。
- LACP接続を使用する各エンドポイントは、パケット送信ポートを個別に選択できますが、受信に使用するポートは選択できません。これは、ONTAPから特定の宛先へのトラフィックが特定のポートに結び付けられ、リターントラフィックが別のインターフェイスに到達する可能性があることを意味します。ただし、これは原因の問題ではありません。
- LACPでは、常にトラフィックが均等に分散されるわけではありません。多数のNFSクライアントを含む大規模な環境では、通常はLACPアグリゲーションのすべてのポートが均等に使用されます。ただし、環境内の1つのNFSファイルシステムの帯域幅は、アグリゲーション全体ではなく、1つのポートの帯域幅に制限されます。
- ONTAPではロビンベースのLACPポリシーを使用できますが、スイッチからホストへの接続には対応していません。たとえば、ホストで4ポートのLACPトランクを、ONTAPで4ポートのLACPトランクを使用する構成でも、ファイルシステムの読み取りには1つのポートしか使用できません。ONTAPは4つのポートすべてを介してデータを送信できますが、4つのポートすべてを介してスイッチからホストに送信するスイッチテクノロジは現在使用できません。使用されるのは1つだけです。

多数のデータベースホストで構成される大規模な環境で最も一般的なアプローチは、IPロードバランシングを使用して、適切な数の10Gb（またはそれよりも高速）インターフェイスでLACPアグリゲートを構築する方法です。このアプローチにより、ONTAPはクライアントが十分に存在する限り、すべてのポートを均等に使用できます。LACPトランキングでは負荷が動的に再分散されないため、構成内のクライアント数が少なくなるとロードバランシングが機能しません。

接続が確立されると、特定の方向のトラフィックは1つのポートにのみ配置されます。たとえば、あるデータベースがNFSファイルシステムに対してテーブルのフルスキャンを実行し、接続に4ポートのLACPトランクを使用している場合、データの読み取りには1枚のネットワークインターフェイスカード（NIC）のみが使用されます。このような環境にデータベースサーバが3台しかない場合は、3台すべてが同じポートから読み取りを行い、他の3つのポートはアイドル状態になる可能性があります。

物理ポートへの**LIF**のバインド

物理ポートにLIFをバインドすると、ネットワーク構成をきめ細かく制御できるようになります。これは、ONTAPシステム上の特定のIPアドレスは、一度に1つのネットワークポートにのみ関連付けられるためです。フェイルオーバーグループとフェイルオーバーポリシーを設定することで耐障害性が実現します。

フェイルオーバーポリシーとフェイルオーバーグループ

ネットワーク停止時のLIFの動作は、フェイルオーバーポリシーとフェイルオーバーグループによって制御されます。設定オプションは、ONTAPのバージョンによって変更されました。を参照してください "フェイルオーバーグループとポリシーに関するONTAPのネットワーク管理に関するドキュメント" を参照して、導入するONTAPのバージョンの詳細を確認してください。

ONTAP 8.3以降では、ブロードキャストドメインに基づいてLIFのフェイルオーバーを管理できます。そのため、特定のサブネットにアクセスできるすべてのポートを管理者が定義し、ONTAPが適切なフェイルオーバーLIFを選択できるようにすることができます。このアプローチは一部のお客様にも使用できますが、予測性がないため、高速ストレージネットワーク環境では制限があります。たとえば、ファイルシステムへの日常的なアクセスに使用する1Gbポートと、データファイルI/Oに使用する10Gbポートの両方を環境に含めることができます。両方のタイプのポートが同じブロードキャストドメインにあると、LIFのフェイルオーバーによって、データファイルI/Oが10Gbポートから1Gbポートに移動される可能性があります。

要約すると、次の方法を検討してください。

1. ユーザ定義のフェイルオーバーグループを設定します。
2. フェイルオーバーグループにストレージフェイルオーバー（SFO）パートナーコントローラのポートを含め、ストレージフェイルオーバー時にLIFがアグリゲートに従って移動するようにします。これにより、間接トラフィックの作成が回避されます。
3. パフォーマンス特性が元のLIFと一致するフェイルオーバーポートを使用します。たとえば、1つの物理10Gbポート上のLIFには、1つの10Gbポートを含むフェイルオーバーグループを含める必要があります。4ポートLACP LIFは、別の4ポートLACP LIFにフェイルオーバーする必要があります。これらのポートは、ブロードキャストドメインに定義されているポートのサブセットになります。
4. SFOパートナーのみにフェイルオーバーポリシーを設定します。これにより、フェイルオーバー時にLIFがアグリゲートに従うようになります。

自動リバート

を設定します `auto-revert` 必要に応じてパラメータを指定する。ほとんどのお客様は、このパラメータを `true` LIFをホームポートにリバートします。ただし、場合によっては、想定外のフェイルオーバーを調査してからLIFをホームポートに戻すように、このパラメータを「false」に設定することもできます。

LIFとボリュームの比率

よくある誤解の1つは、ボリュームとNFS LIFの間には1：1の関係が必要であるということです。この構成は、ボリュームをクラスタ内の任意の場所に移動する際に必要ですが、インターコネクトトラフィックが増えることはありません。ただし、この構成は必須要件ではありません。クラスタ間トラフィックは考慮する必要がありますが、クラスタ間トラフィックが存在するだけでは問題は発生しません。ONTAP用に作成された公開済みのベンチマークの多くには、主に間接I/Oが含まれています。

たとえば、パフォーマンスが重視されるデータベースの数が比較的少なく、合計で40個のボリュームしか必要としないデータベースプロジェクトの場合、ボリューム対LIFの戦略は1：1で、必要なIPアドレスは40個です。これにより、すべてのボリュームを関連付けられたLIFと一緒にクラスタ内の任意の場所に移動でき、トラフィックは常に直接送信されるため、レイテンシのすべてのソースをマイクロ秒レベルでも最小限に抑えることができます。

反対の例として、大規模なホスト環境では、お客様とLIFが1：1の関係にある場合、より簡単に管理できます。時間が経つにつれて、ボリュームを別のノードに移行しなければならない場合があり、間接トラフィックが原因になることがあります。ただし、インターコネクトスイッチのネットワークポートが飽和状態になっていないかぎり、パフォーマンスへの影響は検出されません。懸念がある場合は、ノードを追加して新しいLIFを設定し、次回のメンテナンス時間にホストを更新して、構成から間接トラフィックを取り除くことができます。

Oracleデータベース用の**TCP/IP**およびイーサネット構成

Oracle on ONTAPをご利用のお客様の多くは、NFS、iSCSI、NVMe/TCPのネットワーク

プロトコルであるイーサネットを使用しており、特にクラウドを使用しています。

ホストOSの設定

ほとんどのアプリケーションベンダーのドキュメントには、アプリケーションが最適に動作することを確認するためのTCPおよびイーサネットの設定が含まれています。これらの設定は通常、IPベースのストレージパフォーマンスを最適化するのに十分です。

イーサネットフロー制御

このテクノロジを使用すると、クライアントは送信者にデータ転送を一時的に停止するように要求できます。これは通常、受信側が受信データを十分に迅速に処理できないために行われます。一時期、送信者に送信の中止を要求しても、バッファがいっぱいになったために受信者がパケットを破棄するよりも、中断が少なくて済みました。現在OSで使用されているTCPスタックでは、これは当てはまりません。実際、フロー制御は解決するよりも多くの問題を引き起こします。

近年、イーサネットフロー制御に起因するパフォーマンスの問題が増加しています。これは、イーサネットフロー制御が物理レイヤで動作するためです。ネットワーク構成で、任意のホストOSからストレージシステムへのイーサネットフロー制御要求の送信が許可されていると、接続されているすべてのクライアントのI/Oが一時停止します。1台のストレージコントローラで対応するクライアントの数が増えているため、1台以上のクライアントがフロー制御要求を送信する可能性が高くなります。この問題は、OSの仮想化が広範に行われているお客様のサイトで頻繁に発生しています。

NetAppシステム上のNICは、フロー制御要求を受信しないでください。この結果を得る方法は、ネットワークスイッチの製造元によって異なります。ほとんどの場合、イーサネットスイッチのフロー制御は次のように設定できます。 `receive desired` または `receive on`これは、フロー制御要求がストレージコントローラに転送されないことを意味します。それ以外の場合は、ストレージコントローラのネットワーク接続でフロー制御の無効化が許可されないことがあります。このような場合は、ホストサーバ自体またはホストサーバが接続されているスイッチポートのNIC設定に変更して、フロー制御要求を送信しないようにクライアントを設定する必要があります。

* NetAppでは* NetAppストレージコントローラがイーサネットフロー制御パケットを受信しないようにすることを推奨しています。これは通常、コントローラが接続されているスイッチポートを設定することで実行できますが、一部のスイッチハードウェアには制限があり、代わりにクライアント側の変更が必要になる場合があります。

MTUサイス

ジャンボフレームを使用すると、CPUとネットワークのオーバーヘッドが軽減され、1Gbネットワークのパフォーマンスがある程度向上することが示されていますが、通常はそれほど大きなメリットはありません。

* NetAppでは、可能な限りジャンボフレームを実装することを推奨しています。これは、パフォーマンス上のメリットを実現し解決策、将来のニーズにも対応するためです。

10Gbネットワークではジャンボフレームの使用がほぼ必須です。これは、ほとんどの10Gb環境では、ジャンボフレームを使用しないと10Gbに達する前に1秒あたりのパケット数が制限されるためです。ジャンボフレームを使用すると、OS、サーバ、NIC、およびストレージシステムで処理できるパケットの数は少なくても大きいため、TCP / IP処理の効率が向上します。パフォーマンスの向上はNICによって異なりますが、大幅に向上します。

ジャンボフレームの実装では、接続されているすべてのデバイスでジャンボフレームがサポートされている必要があり、MTUサイズがエンドツーエンドで同じである必要があるという誤った考えがよくあります。代わ

りに、2つのネットワークエンドポイントは、接続を確立するときに、相互に許容可能な最大フレームサイズをネゴシエートします。一般的な環境では、ネットワークスイッチのMTUサイズは9216、NetAppコントローラは9000、クライアントは9000と1514が混在するように設定されています。MTU 9000をサポートできるクライアントはジャンボフレームを使用でき、1514しかサポートできないクライアントは低い値をネゴシエートできます。

完全にスイッチが接続された環境では、この構成に問題が生じることはほとんどありません。ただし、ルーティングされた環境では、中間ルータが強制的にジャンボフレームをフラグメント化しないように注意してください。

- NetAppでは*次の設定を推奨しています。
- ジャンボフレームの使用を推奨しますが、1Gbイーサネット（GbE）の場合は必須ではありません。
- 10GbE以上の速度で最大のパフォーマンスを実現するには、ジャンボフレームが必要です。

TCPパラメータ

TCPタイムスタンプ、選択的確認応答（SACK）、TCPウィンドウスケーリングの3つの設定が誤って設定されることがよくあります。インターネット上の古いドキュメントの多くは、パフォーマンスを向上させるために、これらのパラメータの1つまたは複数を無効にすることを推奨しています。CPU能力がはるかに低く、TCP処理のオーバーヘッドを可能な限り削減できるというメリットが何年も前にあったこの推奨事項には、いくつかのメリットがありました。

ただし、最新のOSでは、これらのTCP機能のいずれかを無効にしても、通常は検出できるメリットはなく、パフォーマンスも低下する可能性があります。特に仮想ネットワーク環境では、パケット損失やネットワーク品質の変化を効率的に処理するためにこれらの機能が必要になるため、パフォーマンスが低下する可能性があります。

* NetAppでは、ホストでTCPタイムスタンプ、SACK、TCPウィンドウスケーリングを有効にすることを推奨しています。現在のOSでは、これら3つのパラメータはすべてデフォルトでオンにする必要があります。

Oracleテタヘスヨウノ**FC**ノセツテイ

Oracleデータベース用にFC SANを構成する主な目的は、日常的なSANのベストプラクティスに従うことです。

これには、ホストとストレージシステムの間のSANに十分な帯域幅があることを確認したり、必要なすべてのデバイス間にすべてのSANパスが存在することを確認したり、FCスイッチベンダーが必要とするFCポート設定を使用してISLの競合を回避したりするなど、一般的な計画方法が含まれます。 SANファブリックを適切に監視します。

ゾーニング

FCゾーンに複数のイニシエータを含めることはできません。このような配置は最初は機能しているように見えるかもしれませんが、最終的にはイニシエータ間のクロストークがパフォーマンスと安定性の妨げになります。

マルチターゲットゾーンは一般に安全とみなされますが、まれに、ベンダーが異なるFCターゲットポートの

動作が問題を引き起こすことがあります。たとえば、NetAppとネットアップ以外のストレージアレイのターゲットポートを同じゾーンに配置することは避けてください。また、NetAppストレージシステムとテープデバイスを同じゾーンに配置すると、原因の問題が発生する可能性がさらに高くなります。

Oracleデータベースと直接接続のONTAP接続

ストレージ管理者は、構成からネットワークスイッチを削除してインフラを簡易化したいと考える場合があります。これは一部のシナリオでサポートされます。

iSCSIとNVMe/TCP

iSCSIまたはNVMe/TCPを使用するホストは、ストレージシステムに直接接続して正常に動作することができます。その理由はパス設定です。2つの異なるストレージコントローラに直接接続すると、データフローが2つの独立したパスになります。パス、ポート、またはコントローラが失われても、他のパスの使用が妨げられることはありません。

NFS

直接接続されたNFSストレージも使用できますが、フェイルオーバーには大きな制限があります。スクリプト作成にはお客様の責任が伴います。

直接接続されたNFSストレージで無停止フェイルオーバーが複雑になるのは、ローカルOSで発生するルーティングが原因です。たとえば、ホストのIPアドレスが192.168.1.1/24で、IPアドレスが192.168.1.50/24のONTAPコントローラに直接接続されているとします。フェールオーバー中、192.168.1.50アドレスはもう一方のコントローラにフェールオーバーでき、ホストが使用できるようになりますが、ホストはそのアドレスの存在をどのように検出しますか。元の192.168.1.1アドレスは、動作中のシステムに接続されていないホストNICに残っています。192.168.1.50宛てのトラフィックは、動作不能なネットワークポートに引き続き送信されます。

2番目のOS NICは19に設定できます。 2.168.1.2およびは、192.168.1.50経由でフェールオーバーされたアドレスと通信できますが、ローカルルーティングテーブルのデフォルトでは、192.168.1.0/24サブネットと通信するために1つの*および1つの*アドレスのみを使用することになります。システム管理者は、失敗したネットワーク接続を検出し、ローカルルーティングテーブルを変更したり、インターフェイスをアップ/ダウンしたりするスクリプトフレームワークを作成できます。正確な手順は、使用しているOSによって異なります。

実際にはNetAppを使用していますが、通常はフェイルオーバー中のIO一時停止が許容されるワークロードのみが対象です。ハードマウントを使用する場合は、一時停止中にIOエラーが発生しないようにしてください。ホスト上のNIC間でIPアドレスを移動するためのフェイルバックまたは手動操作によって、サービスが復元されるまでIOはハングします。

FC直接接続

FCプロトコルを使用してホストをONTAPストレージシステムに直接接続することはできません。その理由はNPIVの使用です。FCネットワークへのONTAP FCポートを識別するWWNは、NPIVと呼ばれる仮想化タイプを使用します。ONTAPシステムに接続されているすべてのデバイスがNPIV WWNを認識できる必要があります。現在、NPIVターゲットをサポートできるホストにインストールできるHBAを提供しているHBAベンダーはありません。

ストレージ構成

FC SAN

Oracleデータベース**I/O**の**LUN**アライメント

LUNアライメントとは、基盤となるファイルシステムのレイアウトに合わせてI/Oを最適化することです。

ONTAPシステムでは、ストレージは4KB単位で編成されます。データベースまたはファイルシステムの8KBブロックは、4KBブロック2個に正確にマッピングする必要があります。LUNの構成エラーによってアライメントがいずれかの方向に1KBずれた場合、8KBの各ブロックは、4KBのストレージブロックが2つではなく3つに配置されます。このようにすると、原因によってレイテンシが増加し、ストレージシステム内で実行される原因の追加I/Oが発生します。

アライメントはLVMアーキテクチャにも影響します。論理ボリュームグループ内の物理ボリュームがドライブデバイス全体に定義されている場合（パーティションは作成されません）、LUN上の最初の4KBブロックがストレージシステム上の最初の4KBブロックとアライメントされます。これは正しいアライメントです。パーティションで問題が発生するのは、OSがLUNを使用する開始場所が変わるためです。オフセットが4KB単位でずれているかぎり、LUNはアライメントされます。

Linux環境では、ドライブデバイス全体に論理ボリュームグループを構築します。パーティションが必要な場合は、次のコマンドを実行してアライメントを確認します。 `fdisk -u` 各パーティションの開始が8の倍数であることを確認します。つまり、パーティションは8の倍数の512バイトセクター（4KB）から開始されます。

圧縮ブロックのアライメントに関するセクションも参照してください。 "効率性"。8KBの圧縮ブロックの境界でアライメントされたレイアウトも、4KBの境界でアライメントされます。

ミスアライメントノケイコク

データベースのRedo /トランザクションログでは通常、アライメントされていないI/Oが生成されるため、ONTAPでLUNがミスアライメントされているという警告が原因で誤って表示される可能性があります。

ロギングは、さまざまなサイズの書き込みでログファイルのシーケンシャルライトを実行します。4KBの境界にアライメントされないログ書き込み処理では、次のログ書き込み処理でブロックが完了するため、通常は原因のパフォーマンスの問題は発生しません。その結果、一部の4KBブロックが2つの別 々 の処理で書き込まれていても、ONTAPはほぼすべての書き込みを完全な4KBブロックとして処理できます。

次のようなユーティリティを使用してアライメントを確認します。 `sio` または `dd` 定義されたブロックサイズでI/Oを生成できます。ストレージシステムのI/Oアライメント統計は、 `stats` コマンドを実行しますを参照してください "WAFLアライメントの検証" を参照してください。

Solaris環境ではアライメントがより複雑になります。を参照してください "ONTAP SAN ホスト構成" を参照してください。

注意
Solaris x86環境では、ほとんどの構成に複数のパーティションレイヤがあるため、適切なアライメントにさらに注意してください。Solaris x86パーティションスライスは通常、標準のマスターブートレコードパーティションテーブルの上に存在します。

Oracleデータベースの**LUN**のサイジングと数

Oracleデータベースのパフォーマンスと管理性を最適化するには、最適なLUNサイズと

使用するLUNの数を選択することが重要です。

LUNはONTAP上の仮想オブジェクトで、ホストしているアグリゲートのすべてのドライブにわたって配置されます。そのため、LUNはどのサイズを選択してもアグリゲートの潜在的なパフォーマンスを最大限に引き出すため、サイズによるLUNのパフォーマンスへの影響はありません。

便宜上、特定のサイズのLUNを使用したい場合があります。たとえば、データベースを2つの1TB LUNで構成されるLVMまたはOracle ASMディスクグループ上に構築する場合、そのディスクグループは1TB単位で拡張する必要があります。8個の500GB LUNでディスクグループを構築し、ディスクグループの増分単位を小さくできるようにすることを推奨します。

汎用性に優れた標準LUNサイズを設定すると、管理が複雑になる可能性があるため、推奨されません。たとえば、標準サイズの100GBのLUNは、1TB~2TBのデータベースまたはデータストアの場合に適していますが、サイズが20TBのデータベースまたはデータストアには200個のLUNが必要です。つまり、サーバのリブート時間が長くなり、さまざまなUIで管理するオブジェクトが増え、SnapCenterなどの製品は多くのオブジェクトに対して検出を実行する必要があります。LUNのサイズを大きくすることで、このような問題を回避できます。

- LUNの数は、サイズよりも重要です。
- LUNのサイズは、主に必要なLUN数によって決まります。
- 必要以上の数のLUNを作成することは避けてください。

LUN数

LUNのサイズとは異なり、LUNの数はパフォーマンスに影響します。アプリケーションのパフォーマンスは、多くの場合、SCSIレイヤを介して並列I/Oを実行できるかどうかに左右されます。その結果、2つのLUNの方が単一のLUNよりもパフォーマンスが向上します。Veritas VxVM、Linux LVM2、Oracle ASMなどのLVMを使用すると、並列処理を強化する最も簡単な方法です。

NetAppのお客様は、LUNの数を16個以上に増やすことによるメリットはほとんどありませんが、ランダムI/Oが非常に大きい100% SSD環境のテストでは、最大64個のLUNがさらに向上していることが実証されています。

> - NetAppの推奨事項*：
>
> 一般に、あらゆるデータベースワークロードのI/Oニーズに対応するには、4~16個のLUNで十分です。LUNを4つ未満にすると、ホストのSCSI実装の制限が原因でパフォーマンスが制限される可能性があります。

OracleデータベースのLUN配置

ONTAPボリューム内でのデータベースLUNの最適な配置は、主に、さまざまなONTAP機能の使用方法によって異なります。

個のボリューム

ONTAPを初めて導入するお客様と混同される共通点の1つは、FlexVol（一般に単に「ボリューム」と呼ばれる）を使用することです。

ボリュームがLUNではありません。これらの用語は、クラウドプロバイダを含む他の多くのベンダー製品と同義語として使用されています。ONTAPボリュームは、単なる管理コンテナです。単独でデータを提供するこ

とも、スペースを占有することもありません。ファイルまたはLUN用のコンテナであり、特に大規模環境で管理性を向上および簡易化するために用意されています。

ボリュームと**LUN**

関連するLUNは通常、1つのボリュームに同じ場所に配置されます。たとえば、10個のLUNが必要なデータベースでは、通常、10個のLUNすべてが同じボリュームに配置されます。

- LUNとボリュームの比率を1：1（ボリュームごとに1つのLUN）にすることは、正式なベストプラクティスでは*ありません。
- 代わりに、ボリュームをワークロードまたはデータセットのコンテナとみなす必要があります。各ボリュームにLUNを1つだけ配置することも、多数配置することもできます。適切な回答は、管理要件によって異なります。
- LUNを不要な数のボリュームに分散させると、Snapshot処理などの処理でオーバーヘッドやスケジュールに関する追加の問題が発生したり、UIに表示されるオブジェクトの数が多すぎたり、LUNの制限に達する前にプラットフォームのボリューム制限に達したりする可能性があります。

ボリューム、**LUN**、**Snapshot**

Snapshotポリシーとスケジュールは、LUNではなくボリュームに配置されます。10個のLUNで構成されるデータセットでは、これらのLUNが同じボリュームに同じ場所にある場合、Snapshotポリシーは1つだけで済みます。

さらに、1つのボリューム内の特定のデータセットに関連するすべてのLUNを同じ場所に配置することで、アトミックなスナップショット操作が可能になります。たとえば、10個のLUNにあるデータベースや、10個のOSで構成されるVMwareベースのアプリケーション環境を、基盤となるLUNがすべて1つのボリュームに配置されている場合は、1つの整合性のあるオブジェクトとして保護できます。Snapshotが別のボリュームに配置されている場合は、同時にスケジュールされていても、Snapshotが100%同期されている場合とそうでない場合があります。

場合によっては、リカバリ要件のために、関連する一連のLUNを2つのボリュームに分割しなければならないことがあります。たとえば、データベースにデータファイル用のLUNが4つ、ログ用のLUNが2つあるとします。この場合は、4つのLUNを含むデータファイルボリュームと2つのLUNを含むログボリュームが最適なオプションです。その理由は独立した回復可能性です。たとえば、データファイルボリュームを選択して以前の状態にリストアすると、4つのLUNすべてがSnapshotの状態にリバートされ、重要なデータを含むログボリュームには影響はありません。

ボリューム、**LUN**、**SnapMirror**

SnapMirrorのポリシーや処理は、Snapshotの処理と同様に、LUNではなくボリュームに対して実行されます。

関連するLUNを1つのボリュームに同じ場所に配置すると、1つのSnapMirror関係を作成し、1回の更新ですべてのデータを更新できます。スナップショットと同様に、更新もアトミックな操作になります。SnapMirrorデスティネーションには、ソースLUNの単一のポイントインタイムレプリカが保証されます。LUNが複数のボリュームに分散している場合は、レプリカ間で整合性がとれている場合とそうでない場合があります。

ボリューム、**LUN**、**QoS**

QoSは個 々 のLUNに選択して適用できますが、通常はボリュームレベルで設定する方が簡単です。たとえば、特定のESXサーバのゲストが使用するすべてのLUNを1つのボリュームに配置し、ONTAPアダプティ

ブQoSポリシーを適用できます。その結果、すべての環境がTBあたりのIOPS制限を自己拡張できるようになります。

同様に、データベースに100K IOPSが必要で、10個のLUNを使用している場合は、LUNごとに1つずつ10K IOPSの制限を個別に10個設定するよりも、1つのボリュームに100K IOPSの制限を1つ設定する方が簡単です。

マルチボリュームレイアウト

複数のボリュームにLUNを分散すると効果的な場合があります。主な理由は、コントローラのストライピングです。たとえば、HAストレージシステムで単一のデータベースをホストし、各コントローラの処理能力とキャッシュ能力をフルに発揮する必要があるとします。この場合、一般的な設計では、LUNの半分をコントローラ1の1つのボリュームに配置し、残りの半分をコントローラ2の1つのボリュームに配置します。

同様に、コントローラストライピングをロードバランシングに使用することもできます。10個のLUNからなる100個のデータベースをホストするHAシステムは、2台のコントローラそれぞれで5個のLUNのボリュームを各データベースに格納するように設計できます。その結果、追加のデータベースがプロビジョニングされるたびに、各コントローラの対称的なロードが保証されます。

ただし、これらの例では、ボリュームとLUNの比率が1：1である必要はありません。その目標は'関連するLUNをボリューム内に共存させることで'管理性を最適化することです

たとえばコンテナ化では、LUNとボリュームの比率を1：1にすることが理にかなっています。コンテナ化では、各LUNは実際には単一のワークロードに相当し、それぞれを個別に管理する必要があります。このような場合、1:1の比率が最適な場合があります。

Oracleデータベースの**LUN**のサイズ変更と**LVM**ベースのサイズ変更

SANベースのファイルシステムが容量の上限に達した場合は、次の2つの方法で使用可能なスペースを増やすことができます。

- LUNのサイズを拡張する
- 既存のボリュームグループにLUNを追加し、それに含まれる論理ボリュームを拡張する

LUNのサイズ変更は容量を拡張するためのオプションですが、一般にはOracle ASMなどのLVMを使用することを推奨します。LVMが存在する主な理由の1つは、LUNのサイズ変更を回避することです。LVMでは、複数のLUNが1つの仮想ストレージプールにボンディングされます。このプールから切り分けられた論理ボリュームはLVMで管理されるため、サイズを簡単に変更できます。もう1つの利点は、特定の論理ボリュームを使用可能なすべてのLUNに分散することで、特定のドライブ上のホットスポットを回避できることです。透過的な移行は、通常、ボリュームマネージャを使用して論理ボリュームの基盤となるエクステントを新しいLUNに再配置することで実行できます。

Oracleデータベースでの**LVM**ストライピング

LVMストライピングとは、複数のLUNにデータを分散することです。その結果、多くのデータベースのパフォーマンスが大幅に向上します。

フラッシュドライブが登場する以前は、回転式ドライブのパフォーマンス上の制限を克服するためにストライピングが使用されていました。たとえば、OSが1MBの読み取り操作を実行する必要がある場合、1つのドライブからその1MBのデータを読み取るには、1MBがゆっくり転送されるため、多くのドライブヘッドのシークと読み取りが必要になります。この1MBのデータが8つのLUNにストライピングされている場合、OSは8つの128K読み取り処理を並行して問題できるため、1MB転送の完了に必要な時間が短縮されます。

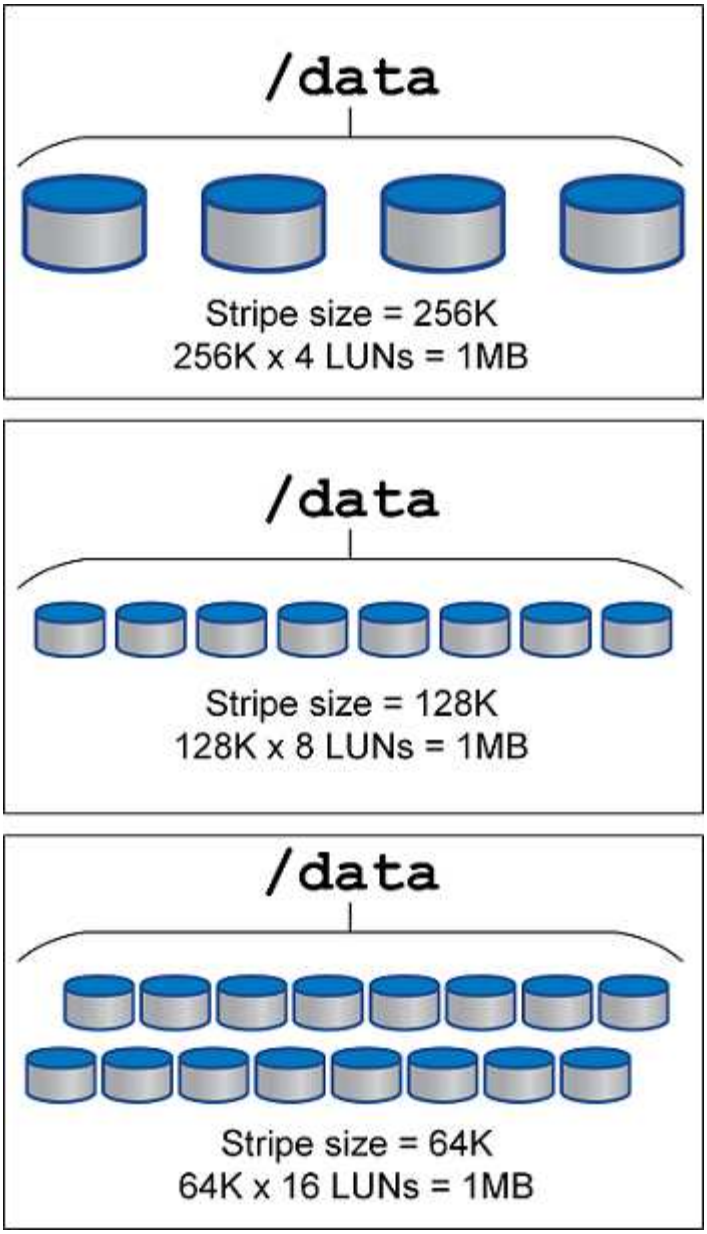
回転式ドライブを使用したストライピングは、I/Oパターンを事前に把握しておく必要があったため、より困難でした。ストライピングが実際のI/Oパターンに合わせて正しく調整されていない場合、ストライピングされた構成ではパフォーマンスが低下する可能性があります。Oracleデータベース、特にオールフラッシュ構成では、ストライピングは設定がはるかに簡単で、パフォーマンスが劇的に向上することが実証されています。

デフォルトではOracle ASMなどの論理ボリュームマネージャがストライプされますが、ネイティブOS LVMはストライプされません。その中には、複数のLUNを連結されたデバイスとして結合するものもあります。そのため、データファイルは1つのLUNデバイスにしか存在しません。これにより、ホットスポットが発生します。他のLVM実装では、デフォルトで分散エクステントが使用されます。これはストライピングに似ていますが、粗いです。ボリュームグループ内のLUNはエクステントと呼ばれる大きな部分にスライスされ、通常は数メガバイト単位で測定され、論理ボリュームがそれらのエクステントに分散されます。その結果、ファイルに対するランダムI/OはLUN間で適切に分散されますが、シーケンシャルI/O処理はそれほど効率的ではありません。

高いパフォーマンスを必要とするアプリケーションI/Oは、ほとんどの場合（a）基本ブロックサイズの単位または（b）1メガバイトのいずれかです。

ストライピング構成の主な目的は、シングルファイルI/Oを1つのユニットとして実行し、マルチブロックI/O（サイズは1MB）をストライピングされたボリューム内のすべてのLUNで均等に並列化できるようにすることです。つまり、ストライプ・サイズはデータベース・ブロック・サイズより小さくすることはできず、ストライプ・サイズにLUN数を掛けたサイズは1MBにする必要があります。

次の図に、ストライプサイズと幅の調整に使用できる3つのオプションを示します。LUNの数は、前述のパフォーマンス要件を満たすように選択されますが、いずれの場合も、1つのストライプ内の総データ量は1MBです。



NFS

Oracleテタヘスヨウノ**NFS**ノセツテイ

NetAppは30年以上にわたってエンタープライズクラスのNFSストレージを提供してきましたが、クラウドベースのインフラへの移行が進むにつれ、その使用がますます増えています。その理由は、シンプルさです。

NFSプロトコルには、要件が異なる複数のバージョンが含まれています。ONTAPを使用したNFSの完全な概要構成については、を参照してください。 "TR-4067『NFS on ONTAP Best Practices』"。次のセクションでは、より重要な要件と一般的なユーザーエラーについて説明します。

NFSハアシヨン

オペレーティングシステムNFSクライアントがNetAppでサポートされている必要があります。

- NFSv3は、NFSv3標準に準拠したOSでサポートされます。
- Oracle dNFSクライアントではNFSv3がサポートされています。
- NFSv4は、NFSv4標準に準拠するすべてのOSでサポートされます。
- NFSv4.1およびNFSv4.2では、特定のOSのサポートが必要です。を参照してください "NetApp IMT" サポートされているOSの場合。
- NFSv4.1でのOracle dNFSのサポートには、Oracle 12.2.0.2以降が必要です。

。 "NetAppのサポートマトリックス" NFSv3およびNFSv4の場合、特定のオペレーティングシステムは含まれません。RFCに準拠するすべてのOSが一般的にサポートされています。オンラインのIMTでNFSv3またはNFSv4のサポートを検索する場合は、該当するOSが表示されないため、特定のOSを選択しないでください。すべてのOSは、一般ポリシーで暗黙的にサポートされています。

Linux NFSv3 TCPスロットテーブル

TCPスロットテーブルは、NFSv3でホストバスアダプタ（HBA）のキュー深度に相当します。一度に未処理となることのできるNFS処理の数を制御します。デフォルト値は通常16ですが、最適なパフォーマンスを得るには小さすぎます。逆に、新しいLinuxカーネルでTCPスロットテーブルの上限をNFSサーバが要求でいっぱいになるレベルに自動的に引き上げることができるため、問題が発生します。

パフォーマンスを最適化し、パフォーマンスの問題を回避するには、TCPスロットテーブルを制御するカーネルパラメータを調整します。

を実行します `sysctl -a | grep tcp.*.slot_table` コマンドを実行し、次のパラメータを確認します。

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

すべてのLinuxシステムに `sunrpc.tcp_slot_table_entries`ただし、次のようなものがあります。 `sunrpc.tcp_max_slot_table_entries`。どちらも128に設定する必要があります。

注意
これらのパラメータを設定しないと、パフォーマンスに大きく影響する可能性があります。Linux OSが十分なI/Oを発行していないためにパフォーマンスが制限される場合もあります。一方では、Linux OSが問題で処理できる以上のI/Oを試行すると、I/Oレイテンシが増加します。

ADRとNFS

一部のお客様から、内のデータに対する過剰なI/Oが原因でパフォーマンスの問題が発生すると報告されています。 `ADR` 場所。通常、この問題は、大量のパフォーマンスデータが蓄積されるまで発生しません。過剰なI/Oの理由は不明ですが、Oracleプロセスがターゲットディレクトリを繰り返しスキャンして変更を求めたことが原因と考えられます。

の取り外し `noac` および / または `actimeo=0` マウントオプションを使用すると、ホストOSのキャッシュを実行してストレージI/Oレベルを削減できます。

* NetApp推奨*配置しないこと ADR フアイルシステムノテエタ noac または actimeo=0 パフォーマンスの問題が発生する可能性があるからです。分離 ADR データを別のマウントポイントに格納（必要な場合）

nfs-rootonlyおよびmount-rootonly

ONTAPには、という名前のNFSオプションがあります。 nfs-rootonly これにより、サーバが高ポートからのNFSトラフィック接続を受け入れるかどうかが制御されます。セキュリティ対策として、1024未満の送信元ポートを使用してTCP/IP接続を開くことができるのはrootユーザだけです。これは、このようなポートは通常、ユーザプロセスではなくOS用に予約されているためです。この制限により、NFSトラフィックが実際のオペレーティングシステムNFSクライアントからのものであり、NFSクライアントをエミュレートする悪意のあるプロセスではないことを確認できます。Oracle dNFSクライアントはユーザスペースドライバですが、このプロセスはrootとして実行されるため、通常は nfs-rootonly。接続は低ポートから行われます。

。 mount-rootonly オプションは環境NFSv3のみです。1024より大きいポートからRPCマウント呼び出しを受け入れるかどうかを制御します。dNFSを使用すると、クライアントは再びルートとして実行されるため、1024未満のポートを開くことができます。このパラメータは効果がありません。

NFSバージョン4.0以降でdNFSを使用して接続を開いているプロセスは、rootとして実行されないため、1024以上のポートが必要です。。 nfs-rootonly dNFSが接続を完了するには、パラメータをdisabledに設定する必要があります。

状況 nfs-rootonly が有効になっている場合、マウントフェーズでdNFS接続を開いているときにハングします。sqlplusの出力は次のようになります。

```
SQL>startup
ORACLE instance started.
Total System Global Area 4294963272 bytes
Fixed Size                  8904776 bytes
Variable Size             822083584 bytes
Database Buffers         3456106496 bytes
Redo Buffers                7868416 bytes
```

パラメータは次のように変更できます。

```
Cluster01::> nfs server modify -nfs-rootonly disabled
```

まれに、nfs-rootonlyとmount-rootonlyの両方をdisabledに変更しなければならないことがあります。サーバが非常に多くのTCP接続を管理している場合、1024未満のポートが使用できない可能性があり、OSはより高いポートを使用するように強制されます。接続を完了するには、これら2つのONTAPパラメータを変更する必要があります。

NFSエクスポートポリシー：superuserとsetuid

OracleバイナリがNFS共有に配置されている場合は、エクスポートポリシーにsuperuser権限とsetuid権限を含める必要があります。

ユーザホームディレクトリなどの汎用ファイルサービスに使用される共有NFSエクスポートでは、通常、root

ユーザが引き下げられます。これは、ファイルシステムをマウントしたホスト上のrootユーザからの要求が、より低い権限を持つ別のユーザとして再マッピングされることを意味します。これは、特定のサーバ上のrootユーザが共有サーバ上のデータにアクセスできないようにすることで、データを保護するのに役立ちます。setuidビットは、共有環境ではセキュリティリスクになることもあります。setuidビットを使用すると、コマンドを呼び出すユーザとは別のユーザとしてプロセスを実行できます。たとえば、rootがsetuidビットを持つシェルスクリプトはrootとして実行されます。そのシェルスクリプトが他のユーザによって変更される可能性がある場合、root以外のユーザはスクリプトを更新することでrootとしてコマンドを問題できます。

Oracleバイナリには、rootが所有するsetuidビットを使用するファイルが含まれます。OracleバイナリがNFS共有にインストールされている場合は、エクスポートポリシーに適切なsuperuser権限とsetuid権限が含まれている必要があります。次の例では、ルールに `allow-suid` 許可します `superuser` （root）システム認証を使用したNFSクライアントのアクセス。

```
Cluster01::> export-policy rule show -vserver vserver1 -policyname orabin
-fields allow-suid,superuser
vserver   policyname ruleindex superuser allow-suid
--------- ---------- --------- --------- ----------
vserver1  orabin     1         sys       true
```

NFSv4 / 4.1構成

ほとんどのアプリケーションで、NFSv3とNFSv4の違いはほとんどありません。通常、アプリケーションI/Oは非常に単純なI/Oであり、NFSv4の高度な機能の一部からあまりメリットが得られません。上位バージョンのNFSは、データベースストレージから見ると「アップグレード」ではなく、機能を追加したNFSのバージョンとみなすべきです。たとえば、Kerberosプライバシーモード（krb5p）のエンドツーエンドのセキュリティが必要な場合は、NFSv4が必要です。

* NetAppでは* NFSv4の機能が必要な場合はNFSv4.1を使用することを推奨します。NFSv4.1では、一部のエッジにおける耐障害性を向上させるために、NFSv4プロトコルの機能がいくつか強化されています。

NFSv4への切り替えは、マウントオプションを単にvers=3からvers=4.1に変更するよりも複雑です。ONTAPを使用したNFSv4設定の詳細（OSの設定に関するガイダンスなど）については、を参照してください。 "TR-4067『NFS on ONTAP』のベストプラクティス"。このTRの以降のセクションでは、NFSv4を使用するための基本的な要件の一部について説明します。

NFSv4ドメイン

NFSv4 / 4.1の設定について詳しくは本ドキュメントでは説明していませんが、よく発生する問題の1つとして、ドメインマッピングの不一致があります。sysadminから見ると、NFSファイルシステムは正常に動作しているように見えますが、アプリケーションからは特定のファイルに対する権限やsetuidに関するエラーが報告されます。場合によっては、管理者は、アプリケーションバイナリのアクセス許可が破損していると誤って判断し、実際の問題がドメイン名であったときにchownまたはchmodコマンドを実行しました。

ONTAP SVMでNFSv4ドメイン名が設定されます。

```
Cluster01::> nfs server show -fields v4-id-domain
vserver   v4-id-domain
--------- ------------
vserver1  my.lab
```

ホストのNFSv4ドメイン名は、 `/etc/idmap.cfg`

```
[root@host1 etc]# head /etc/idmapd.conf
[General]
#Verbosity = 0
# The following should be set to the local NFSv4 domain name
# The default is the host's DNS domain name.
Domain = my.lab
```

ドメイン名が一致している必要があります。マッピングされていない場合は、次のようなマッピングエラーがに表示されます。 `/var/log/messages`：

```
Apr 12 11:43:08 host1 nfsidmap[16298]: nss_getpwnam: name 'root@my.lab'
does not map into domain 'default.com'
```

アプリケーションバイナリ（Oracleデータベースバイナリなど）には、rootが所有するsetuidビットのファイルが含まれています。つまり、NFSv4ドメイン名が一致していないとOracleの起動に失敗し、という名前のファイルの所有権または権限に関する警告が表示されます。 `oradism`をクリックします。 `$ORACLE_HOME/bin` ディレクトリ。次のように表示されます。

```
[root@host1 etc]# ls -l /orabin/product/19.3.0.0/dbhome_1/bin/oradism
-rwsr-x--- 1 root oinstall 147848 Apr 17  2019
/orabin/product/19.3.0.0/dbhome_1/bin/oradism
```

所有権がnobodyのファイルが表示される場合は、NFSv4ドメインのマッピングに問題がある可能性があります。

```
[root@host1 bin]# ls -l oradism
-rwsr-x--- 1 nobody oinstall 147848 Apr 17  2019 oradism
```

これを修正するには、 `/etc/idmap.cfg` ファイルをONTAPのv4-id-domain設定に対して作成し、整合性を確保します。設定されていない場合は、必要な変更を行い、 `nfsidmap -c`をクリックし、変更が反映されるまでしばらく待ちます。これで、ファイル所有権が`root`として正しく認識されます。ユーザがを実行しようとした場合 `chown root` NFSドメインの設定が修正される前に、このファイルで次のコマンドを実行する必要があります。 `chown root` をもう一度クリックします

Oracle directNFS

Oracleデータベースでは、NFSを2つの方法で使用できます。

まず、オペレーティングシステムの一部であるネイティブのNFSクライアントを使用してマウントされたファイルシステムを使用できます。これはカーネルNFS（kNFS）と呼ばれることもあります。NFSファイルシステムは、NFSファイルシステムを使用する他のアプリケーションとまったく同じように、Oracleデータベースによってマウントされ、使用されます。

2つ目の方法はOracle Direct NFS（dNFS）です。これは、OracleデータベースソフトウェアにNFS標準を実装したものです。DBAによるOracleデータベースの設定や管理方法は変更されません。ストレージシステム自体に正しい設定があるかぎり、DBAチームやエンドユーザはdNFSを透過的に使用できなければなりません。

dNFS機能を有効にしたデータベースでは、通常のNFSファイルシステムが引き続きマウントされています。データベースが開くと、Oracleデータベースは一連のTCP/IPセッションを開き、NFS操作を直接実行します。

Direct NFS

OracleのDirect NFSの主なメリットは'ホストのNFSクライアントをバイパスして'NFSサーバ上で直接NFSファイル操作を実行することですこれを有効にするには、Oracle Disk Manager（ODM）ライブラリを変更する必要があります。このプロセスの手順については、Oracleのマニュアルを参照してください。

dNFSを使用すると、I/Oが最も効率的な方法で実行されるため、I/Oパフォーマンスが大幅に向上し、ホストとストレージシステムの負荷が軽減されます。

さらに、Oracle dNFSには、ネットワークインターフェイスのマルチパスとフォールトトレランス用の*オプション*が含まれています。たとえば、2つの10Gbインターフェイスをバインドして、20Gbの帯域幅を提供できます。一方のインターフェイスで障害が発生すると、もう一方のインターフェイスでI/Oが再試行されます。全体的な処理はFCマルチパスとほぼ同じです。マルチパスは、1Gbイーサネットが最も一般的な標準であった数年前には一般的でした。ほとんどのOracleワークロードには10Gb NICで十分ですが、必要に応じて10Gb NICをボンディングできます。

dNFSを使用する場合は、Oracleのドキュメント1495104.1に記載されているパッチをすべてインストールしておくことが重要です。パッチをインストールできない場合は、環境を評価して、そのドキュメントに記載されているバグが原因の問題ではないことを確認する必要があります。必要なパッチをインストールできないためにdNFSを使用できない場合があります。

dNFSでラウンドロビン方式の名前解決（DNS、DDNS、NISなど）を使用しないでください。これには、ONTAPで使用できるDNSロードバランシング機能も含まれます。dNFSを使用するOracleデータベースがあるホスト名をIPアドレスに解決した場合、以降の検索でもホスト名が変更されないようにする必要があります。その結果、Oracleデータベースがクラッシュし、データが破損する可能性があります。

Direct NFSとホストファイルシステムへのアクセス

アプリケーションやユーザのアクティビティが、ホストにマウントされた参照可能なファイルシステムに依存している場合、dNFSを使用すると原因の問題が発生することがあります。これは、dNFSクライアントがホストOSの帯域外でファイルシステムにアクセスするためです。dNFSクライアントは、OSが認識されていなくてもファイルの作成、削除、および変更を行うことができます。

シングルインスタンスデータベースのマウントオプションを使用すると、ファイルおよびディレクトリの属性のキャッシュが有効になり、ディレクトリの内容もキャッシュされます。そのため、dNFSでファイルが作成される可能性があり、OSがディレクトリの内容を再読み取りしてファイルがユーザに表示されるまでに少し時間がかかります。これは通常は問題になりませんが、まれに、SAP BR * Toolsなどのユーティリティで問題

が発生することがあります。この場合は、マウントオプションをOracle RACの推奨事項に変更して、問題に対処してください。この変更により、ホストのキャッシュがすべて無効になります。

マウントオプションを変更するのは、（a）dNFSが使用されていて、（b）問題がファイルが参照可能になるまでの遅延が原因で発生した場合のみにしてください。dNFSを使用していない場合は、シングルインスタンスデータベースでOracle RACマウントオプションを使用すると、パフォーマンスが低下します。

次の注を参照： `nosharecache` インチ "Linux NFSのマウントオプション" 通常とは異なる結果を生成する可能性があるLinux固有のdNFS問題の場合。

Oracleデータベースと**NFS**のリースとロック

NFSv3はステートレスです。つまり、NFSサーバ（ONTAP）は、どのファイルシステムがマウントされているのか、誰がどのロックが実際に有効であるのかを追跡しません。

ONTAPにはマウントの試行を記録する機能がいくつかあります。そのため、どのクライアントがデータにアクセスしている可能性があるかを把握できます。また、アドバイザリロックが存在する可能性がありますが、その情報が100%完了する保証はありません。NFSクライアントの状態の追跡はNFSv3標準には含まれていないため、この処理を完了できません。

NFSv4のステートフル

一方、NFSv4はステートフルです。NFSv4サーバは、どのクライアントが使用しているファイルシステム、どのファイルが存在するか、どのファイルやファイル領域がロックされているかなどを追跡します。 つまり、状態のデータを最新の状態に保つためには、NFSv4サーバ間で定期的な通信が必要です。

NFSサーバによって管理されている最も重要な状態は、NFSv4ロックとNFSv4リースであり、これらは非常に密接に関連しています。それぞれがそれ自体でどのように機能し、それらが互いにどのように関連しているかを理解する必要があります。

NFSv4ロック

NFSv3では、ロックは推奨されます。NFSクライアントは、「ロックされた」ファイルを変更または削除できます。NFSv3のロックは自動的に期限切れになるわけではなく、削除する必要があります。これは問題を引き起こします。たとえば、クラスタ化されたアプリケーションでNFSv3ロックを作成していて、いずれかのノードで障害が発生した場合は、どうすればよいですか？残りのノードでアプリケーションをコーディングしてロックを解除できますが、それが安全であることをどのようにして確認できますか？「failed」ノードは動作しているが、クラスタの残りのノードと通信していない可能性があります。

NFSv4では、ロックの期間に制限があります。ロックを保持しているクライアントがNFSv4サーバにチェックインし続けるかぎり、他のクライアントはこれらのロックを取得できません。クライアントがNFSv4へのチェックインに失敗すると、最終的にサーバによってロックが取り消され、他のクライアントはロックを要求および取得できます。

NFSv4リース

NFSv4ロックはNFSv4リースに関連付けられます。NFSv4クライアントがNFSv4サーバとの接続を確立すると、リースを取得します。クライアントがロックを取得した場合（ロックにはさまざまな種類があります）、ロックはリースに関連付けられます。

このリースには定義済みのタイムアウトがあります。デフォルトでは、ONTAPはタイムアウト値を30秒に設定します。

```
Cluster01::*> nfs server show -vserver vserver1 -fields v4-lease-seconds

vserver   v4-lease-seconds
--------- ----------------
vserver1  30
```

つまり、NFSv4クライアントはリースを更新するために、30秒ごとにNFSv4サーバにチェックインする必要があります。

リースはすべてのアクティビティによって自動的に更新されるため、クライアントが作業を行っている場合は追加操作を実行する必要はありません。アプリケーションが静かになり、実際の作業を行っていない場合は、代わりに一種のキープアライブ操作(シーケンスと呼ばれる)を実行する必要があります。それは基本的に「私はまだここにいる、私のリースを更新してください」と言っているだけです。

```
 *Question:* What happens if you lose network connectivity for 31 seconds?
```
NFSv3はステートレスです。クライアントからの通信を期待していません。NFSv4はステートフルであり、リース期間が経過するとリースが期限切れになり、ロックが取り消され、ロックされたファイルが他のクライアントから利用可能になります。

NFSv3では、ネットワークケーブルを移動したり、ネットワークスイッチをリブートしたり、設定を変更したりすることができ、問題が発生しないように十分に確認することができます。アプリケーションは通常、ネットワーク接続が再び機能するのを辛抱強く待つだけです。

NFSv4では、作業を完了するまでに30秒かかります（ONTAP内でそのパラメータの値を増やした場合を除く）。それを超えると、リースはタイムアウトになります。通常、この結果、アプリケーションがクラッシュします。

たとえば、Oracleデータベースを使用していて、リースタイムアウトを超えるネットワーク接続（「ネットワークパーティション」と呼ばれることもあります）が失われると、データベースがクラッシュします。

これが発生した場合のOracleアラート・ログの出力例を次に示します

```
2022-10-11T15:52:55.206231-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00202: control file: '/redo0/NTAP/ctrl/control01.ctl'
ORA-27072: File I/O error
Linux-x86_64 Error: 5: Input/output error
Additional information: 4
Additional information: 1
Additional information: 4294967295
2022-10-11T15:52:59.842508-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00206: error in writing (block 3, # blocks 1) of control file
ORA-00202: control file: '/redo1/NTAP/ctrl/control02.ctl'
ORA-27061: waiting for async I/Os failed
```

syslogを確認すると、次のエラーのいくつかが表示されます。

```
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
```

ログメッセージは通常、アプリケーションがフリーズする以外に、問題の最初の兆候です。通常、ネットワークの停止中は何も表示されません。これは、NFSファイルシステムにアクセスしようとするプロセスとOS自体がブロックされるためです。

エラーは、ネットワークが再び動作可能になると表示されます。上記の例では、接続が再確立されると、OSはロックの再取得を試みましたが、遅すぎました。リースが期限切れになり、ロックが削除されました。その結果、エラーがOracleレイヤまで伝播し、アラートログにメッセージが記録されます。これらのパターンは、データベースのバージョンと構成によって異なる場合があります。

要約すると、NFSv3はネットワークの中断は許容されますが、NFSv4はより機密性が高く、リース期間が定義されます。

30秒のタイムアウトが許容されない場合はどうなりますか。スイッチが再起動されたり、ケーブルが再配置されたりする動的に変化するネットワークを管理していて、その結果、時 々 ネットワークが中断される場合はどうなりますか。リース期間を延長することもできますが、その場合はNFSv4猶予期間の説明が必要です。

NFSv4猶予期間

NFSv3サーバをリブートすると、ほぼ瞬時にIOを処理できるようになります。それはクライアントの状態を維持することではありませんでした。そのため、ONTAPのテイクオーバー処理はほぼ瞬時に実行されることがよくあります。コントローラがデータの提供を開始する準備ができた時点で、ネットワークにARPを送信し、トポロジの変更を通知します。通常、クライアントはこれをほぼ瞬時に検出し、データの流れを再開します。

ただし、NFSv4では一時停止が発生します。これは、NFSv4がどのように機能するかの一部にすぎません。

NFSv4サーバは、リース、ロック、および誰がどのデータを使用しているかを追跡する必要があります。NFSサーバがパニック状態になってリブートされた場合、または一時的に電力が失われた場合、またはメンテナンス作業中に再起動された場合は、リース/ロックなどのクライアント情報が失われます。サーバは、処理を再開する前に、どのクライアントがどのデータを使用しているかを把握する必要があります。ここで猶予期間が入ります。

NFSv4サーバの電源が突然再投入された場合。再起動すると、IOを再開しようとするクライアントは、基本的に「リース/ロック情報が失われました。ロックを再登録しますか？" これが猶予期間の始まりですONTAPではデフォルトで45秒です。

```
Cluster01::> nfs server show -vserver vserver1 -fields v4-grace-seconds

vserver   v4-grace-seconds
--------- ----------------
vserver1  45
```

その結果、再起動後、すべてのクライアントがリースとロックを再要求する間、コントローラはIOを一時停止します。猶予期間が終了すると、サーバはIO処理を再開します。

リースタイムアウトと猶予期間

猶予期間とリース期間が接続されます。前述したように、デフォルトのリースタイムアウトは30秒です。つまり、NFSv4クライアントは少なくとも30秒ごとにサーバにチェックインする必要があります。そうしないと、リースとロックが失われます。この猶予期間はNFSサーバがリース/ロックデータを再構築できるようにするためのもので、デフォルトは45秒です。ONTAPでは、猶予期間をリース期間より15秒長くする必要があります。これにより、リースを30秒以上更新するように設計されたNFSクライアント環境では、再起動後にサーバにチェックインできます。猶予期間を45秒に設定することで、少なくとも30秒ごとにリースを更新することを期待するすべてのクライアントが確実に更新する機会を得ることができます。

30秒のタイムアウトが許容されない場合は、リース期間を延長することもできます。60秒のネットワーク停止に耐えるためにリースタイムアウトを60秒に延長する場合は、猶予期間を少なくとも75秒に延長する必要があります。ONTAPでは、リース期間より15秒長くする必要があります。つまり、コントローラフェイルオーバー中にIOが一時停止する時間が長くなります。

これは通常は問題ではありません。一般的なユーザはONTAPコントローラを年に1~2回更新するだけで、ハードウェア障害による計画外フェイルオーバーは非常にまれです。また、ネットワークに60秒のネットワーク停止が発生する可能性があり、リースタイムアウトを60秒にする必要がある場合は、まれにストレージシステムのフェイルオーバーに異議を唱えず、75秒の一時停止も発生する可能性があります。ネットワークが60秒以上頻繁に一時停止していることをすでに認識しています。

Oracleデータベースでの**NFS**キャッシュ

次のマウントオプションが存在すると、ホストのキャッシュが無効になります。

```
cio, actimeo=0, noac, forcedirectio
```

これらの設定は、ソフトウェアのインストール、パッチ適用、およびバックアップ/リストアの処理速度に重大な悪影響を及ぼす可能性があります。場合によっては、特にクラスタ化されたアプリケーションでは、クラスタ内のすべてのノードにキャッシュの一貫性を提供する必要があるため、これらのオプションが必然的に必要になることがあります。それ以外の場合、顧客はこれらのパラメータを誤って使用し、結果は不要な性能の損傷です。

多くのお客様は、アプリケーションバイナリのインストール時やパッチ適用時に、これらのマウントオプションを一時的に削除しています。インストールまたはパッチ適用プロセス中にターゲットディレクトリを他のプロセスがアクティブに使用していないことをユーザーが確認した場合は、この削除を安全に実行できます。

ONTAPでは、デフォルトでNFS I/Oサイズが64Kに制限されています。

ほとんどのアプリケーションとデータベースでランダムI/Oを実行すると、ブロックサイズがはるかに小さくなり、最大64Kよりもはるかに小さくなります。ラージブロックI/Oは通常並列処理されるため、最大64Kも最大帯域幅の確保に制限されるわけではありません。

一部のワークロードでは、最大64Kに制限があります。特に、バックアップ/リカバリ処理やデータベースのフルテーブルスキャンなどのシングルスレッド処理は、実行回数が少なくても大容量のI/Oを実行できるのであれば、より高速かつ効率的に実行できます。ONTAPに最適なI/O処理サイズは256Kです。

特定のONTAP SVMの最大転送サイズは、次のように変更できます。

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

注意
ONTAPで許容される最大転送サイズを、現在マウントされているNFSファイルシステムのrsize/wsizeの値より小さくしないでください。これにより、一部のオペレーティングシステムでハングしたり、データが破損したりする可能性があります。たとえば、NFSクライアントのrsize / wsizeが65536に設定されている場合は、クライアント自体が制限されているため、ONTAPの最大転送サイズを65536~1048576の間で調整しても効果はありません。最大転送サイズを65536未満に縮小すると、可用性やデータが損傷する可能性があります。

Oracleデータベースと**NVFAIL**

NVFailは、重大なフェイルオーバーシナリオの際に整合性を確保するONTAPの機能です。

データベースは大規模な内部キャッシュを保持するため、ストレージフェイルオーバー時に破損の影響を受けやすくなります。構成全体の健全性に関係なく、重大なイベントによってONTAPフェイルオーバーの強制またはMetroClusterスイッチオーバーの強制が必要になった場合は、以前に確認された変更が実質的に破棄されることがあります。ストレージアレイの内容が時間を遡るようになり、データベースキャッシュの状態がディスク上のデータの状態を反映しなくなります。この不整合により、データが破損します。

キャッシュはアプリケーションレイヤまたはサーバレイヤで実行できます。たとえば、プライマリサイトとリモートサイトの両方でアクティブなサーバを使用するOracle Real Application Cluster（RAC）構成では、Oracle SGA内のデータがキャッシュされます。強制スイッチオーバー処理によってデータが失われると、SGAに格納されているブロックがディスク上のブロックと一致しない可能性があるため、データベースが破損するリスクがあります。

キャッシュの使用は、OSファイルシステムレイヤではあまり明白ではありません。マウントされたNFSファイルシステムのブロックは、OSにキャッシュされる場合があります。または、プライマリサイトにあるLUN

に基づくクラスタ化されたファイルシステムをリモートサイトのサーバにマウントして、データをキャッシュすることもできます。このような状況でNVRAMの障害、強制テイクオーバー、強制スイッチオーバーが発生すると、ファイルシステムが破損する可能性があります。

ONTAPは、NVFAILとその関連設定を使用して、このシナリオからデータベースとオペレーティングシステムを保護します。

ASM再生ユーティリティとONTAPゼロブロック検出

インライン圧縮が有効な場合、ONTAPはファイルまたはLUNに書き込まれた初期化済みブロックを効率的に削除します。Oracle ASM Reclamation Utility（ASRU）などのユーティリティは、未使用のASMエクステントにゼロを書き込むことで機能します。

これにより、DBAはデータが削除されたあとにストレージアレイのスペースを再生できます。ONTAPはゼロをインターセプトし、LUNからスペースの割り当てを解除します。ストレージシステム内にデータが書き込まれていないため、再生プロセスは非常に高速です。

データベースに関しては、ASMディスクグループには0が含まれているため、LUNのこれらの領域を読み取ると0のストリームが生成されますが、ONTAPはドライブに0を格納しません。代わりに、メタデータが単純に変更され、LUNの初期化された領域がデータの空として内部的にマークされます。

同様の理由から、ゼロのブロックは実際にはストレージアレイ内で書き込みとして処理されないため、初期化されたデータを使用したパフォーマンステストは無効です。

ASRUを使用する場合は、Oracleが推奨するすべてのパッチがインストールされていることを確認してください。

Oracleデータベースの仮想化

VMware、Oracle OLVM、KVMを使用したデータベースの仮想化は、最もミッションクリティカルなデータベースでさえ仮想化を選択したNetAppのお客様にとって、ますます一般的な選択肢となっています。

サポート性

Oracleによる仮想化のサポートポリシーについては、多くの誤解があります。特にVMware製品については、誤解が生じています。Oracle OUTRIGHTが仮想化をサポートしていないという話は珍しくありません。この概念は正しくないため、仮想化によるメリットを得る機会を逃してしまいます。実際の要件についてはOracleのドキュメントID 249212.1で説明されており、お客様が懸念事項と考えることはほとんどありません。

仮想化されたサーバで問題が発生し、これまでOracleサポートがその問題を認識していなかった場合は、物理ハードウェアで問題を再現するようにお客様に依頼されることがあります。最新バージョンの製品を使用しているOracleのお客様は、サポート性の問題が発生する可能性があるため、仮想化の使用を望まないかもしれませんが、一般提供されているOracle製品バージョンを使用しているお客様にとって、このような状況は現実のものではありません。

ストレージ提供

データベースの仮想化を検討しているお客様は、ビジネスニーズに基づいてストレージに関する意思決定を行

う必要があります。これはすべてのIT意思決定に一般的に当てはまりますが、要件のサイズと範囲が大幅に異なるため、データベースプロジェクトでは特に重要です。

ストレージプレゼンテーションには、次の3つの基本的なオプションがあります。

- ハイパーバイザーデータストア上の仮想LUN
- iSCSI LUNをハイパーバイザーではなくVMのiSCSIイニシエータで管理
- （NFSベースのデータストアからではなく）VMによってマウントされたNFSファイルシステム
- 直接デバイスマッピング。VMware RDMはお客様から嫌われていますが、多くの場合、物理デバイスはKVMやOLVM仮想化と同様に直接マッピングされています。

パフォーマンス

仮想ゲストにストレージを提供する方法は、通常、パフォーマンスに影響しません。ホストOS、仮想ネットワークドライバ、ハイパーバイザーデータストアの実装はいずれも高度に最適化されており、基本的なベストプラクティスに従うかぎり、ハイパーバイザーとストレージシステムの間で使用可能なFCまたはIPネットワーク帯域幅をすべて消費できます。場合によっては、あるストレージプレゼンテーションのアプローチを使用した方が、別のアプローチよりもわずかに簡単に最適なパフォーマンスを得ることができますが、最終的な結果は同等である必要があります。

管理性

仮想ゲストにストレージをどのように提供するかを決定する際の重要な要素は、管理性です。正しい方法も間違った方法もありません。最適なアプローチは、IT運用のニーズ、スキル、好みによって異なります。

考慮すべき要素は次のとおりです。

- 透過性。 VMがファイルシステムを管理する場合、データベース管理者やシステム管理者は、データのファイルシステムのソースを簡単に特定できます。ファイルシステムとLUNへのアクセス方法は、物理サーバと同じです。
- 一貫性。 VMが自身のファイルシステムを所有している場合、ハイパーバイザーレイヤを使用するかどうかは管理性に影響します。プロビジョニング、監視、データ保護などの手順は、仮想環境と非仮想環境の両方を含め、資産全体で同じです。

  一方、完全に仮想化されたデータセンターでは、前述のように、プロビジョニング、保護、モンターリング、データ保護に同じ手順を使用できる一貫性という同じ根拠に基づいて、フットプリント全体でデータストアベースのストレージを使用することを推奨します。

- 安定性とトラブルシューティング。 VMが自身のファイルシステムを所有している場合、ストレージスタック全体がVM上に存在するため、優れた安定したパフォーマンスが提供され、問題のトラブルシューティングが簡単になります。ハイパーバイザーの役割は、FCフレームまたはIPフレームを転送することだけです。構成にデータストアが含まれていると、タイムアウト、パラメータ、ログファイル、および潜在的なバグが新たに発生するため、構成が複雑になります。
- モビリティ。 VMが自身のファイルシステムを所有している場合、Oracle環境を移動するプロセスははるかにシンプルになります。ファイルシステムは、仮想ゲストと非仮想ゲストの間で簡単に移動できます。
- *ベンダーロックイン。*データをデータストアに配置すると、別のハイパーバイザーを使用したり、仮想環境からデータを取り出すことが完全に困難になります。
- *スナップショットの有効化。*仮想環境での従来のバックアップ手順は、帯域幅が比較的限られているため、問題になる可能性があります。たとえば、多くの仮想データベースで日常的に必要とされるパフォー

マンスには4ポート10GbEトランクで十分ですが、RMANなどのバックアップ製品を使用してバックアップを実行するには、データのフルサイズのコピーをストリーミングする必要がある場合は、このようなトランクでは不十分です。その結果、統合が進む仮想環境では、ストレージスナップショットを使用してバックアップを実行する必要が生じています。これにより、バックアップウィンドウで帯域幅とCPUの要件をサポートするためだけにハイパーバイザー構成を過剰に構築する必要がなくなります。

ゲスト所有のファイルシステムを使用すると、保護を必要とするストレージオブジェクトをより簡単にターゲットにできるため、Snapshotベースのバックアップとリストアを簡単に活用できる場合があります。しかし、データストアやスナップショットとうまく統合できる仮想化データ保護製品の数はますます増えています。仮想化されたホストにストレージを提供する方法を決定する前に、バックアップ戦略を十分に検討する必要があります。

準仮想化ドライバ

最適なパフォーマンスを実現するには、準仮想化ネットワークドライバを使用することが重要です。データストアを使用する場合は、準仮想化SCSIドライバが必要です。準仮想化デバイスドライバを使用すると、エミュレートされたドライバとは対照的に、ゲストをハイパーバイザーにより深く統合できます。エミュレートされたドライバでは、ハイパーバイザーは物理ハードウェアの動作を模倣するためにより多くのCPU時間を消費します。

RAMのオーバーコミット

RAMのオーバーコミットとは、物理ハードウェア上に存在するよりも多くの仮想RAMをさまざまなホストに設定することを意味します。原因で予期しないパフォーマンスの問題が発生する可能性があります。データベースを仮想化する場合、Oracle SGAの基盤となるブロックがハイパーバイザーによってストレージにスワップアウトされないようにする必要があります。このように設定すると、パフォーマンスが非常に不安定になります。

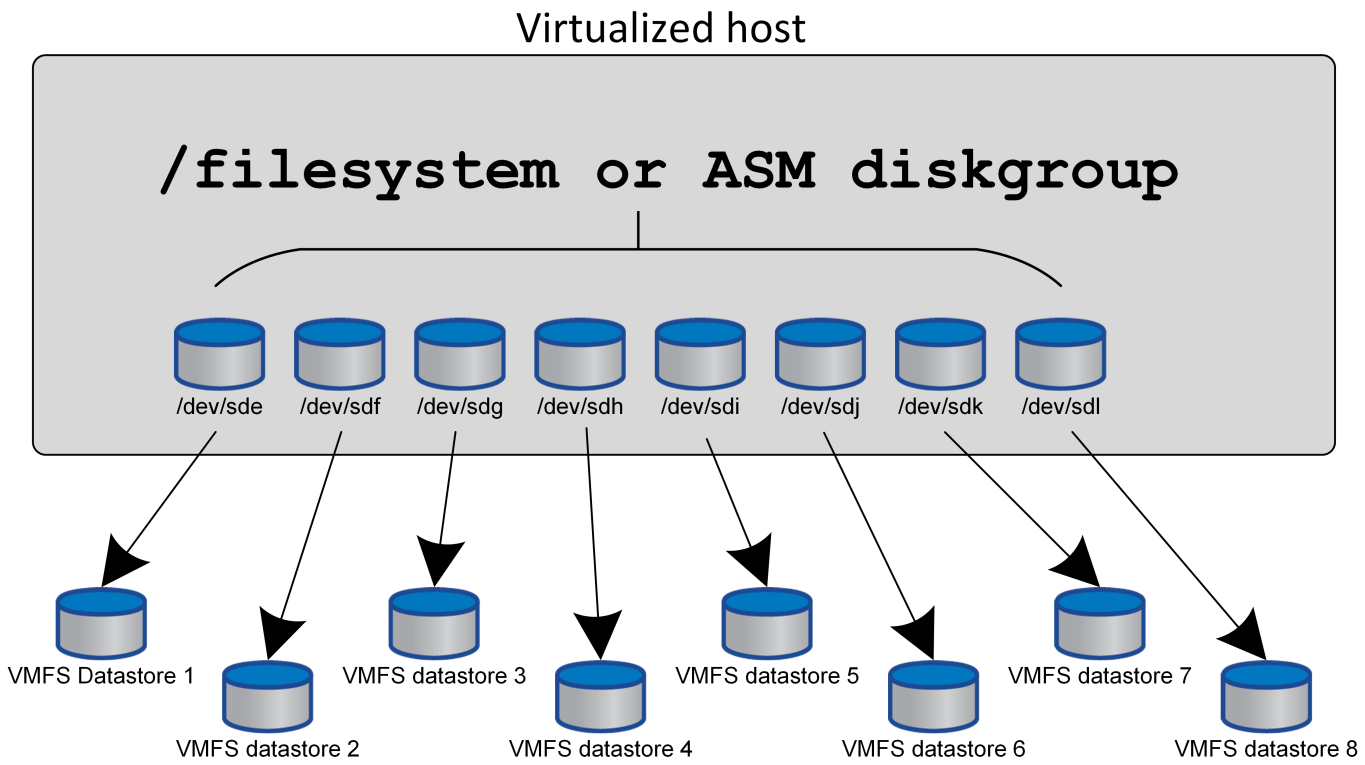
データストアのストライピング

データストアでデータベースを使用する場合、パフォーマンスストライピングに関して考慮すべき重要な要素が1つあります。

VMFSなどのデータストアテクノロジは複数のLUNにまたがることができますが、ストライプデバイスではありません。LUNが連結されます。その結果、LUNのホットスポットが発生する可能性があります。たとえば、一般的なOracleデータベースに8 LUNのASMディスクグループがあるとします。8つの仮想化されたLUNはすべて8 LUNのVMFSデータストアにプロビジョニングできますが、データがどのLUNに格納されるかは保証されません。この構成では、8つの仮想LUNすべてがVMFSデータストア内の1つのLUNを占有するようになります。これがパフォーマンスのボトルネックになります。

通常、ストライピングが必要です。KVMなどの一部のハイパーバイザーでは、次の説明に従ってLVMストライピングを使用してデータストアを構築できます。 "こちらをご覧ください"。VMwareの場合、アーキテクチャは少し異なります。各仮想LUNを別 々 のVMFSデータストアに配置する必要があります。

例：



このアプローチの主な推進力はONTAPではありません。これは、1つのVMまたはハイパーバイザーLUNが並行して処理できる処理数に固有の制限があるためです。1つのONTAP LUNでサポートできるIOPSは、通常、ホストが要求できるIOPSよりもはるかに多くなります。単一LUNのパフォーマンス制限は、ほとんどの場合、ホストOSが原因です。そのため、ほとんどのデータベースでは、パフォーマンスのニーズを満たすために4~8個のLUNが必要になります。

VMwareアーキテクチャでは、データストアやLUNパスの最大数がこのアプローチで発生しないように、アーキテクチャを慎重に計画する必要があります。また、すべてのデータベースに固有のVMFSデータストアセットを用意する必要はありません。主に必要なのは、各ホストに、仮想化されたLUNからストレージシステム自体のバックエンドLUNへの4 ~ 8個のIOパスのクリーンなセットがあることを確認することです。まれに、より多くのデータストアが本当に極端なパフォーマンス要求に対して有益な場合もありますが、一般に、全データベースの95%に対して4~8個のLUNで十分です。8個のLUNを含む1つのONTAPボリュームでは、一般的なOS / ONTAP /ネットワーク構成で、Oracleブロックのランダム IOPSを最大25、000個サポートできます。

階層化

Oracle Database FabricPoolの階層化の概要

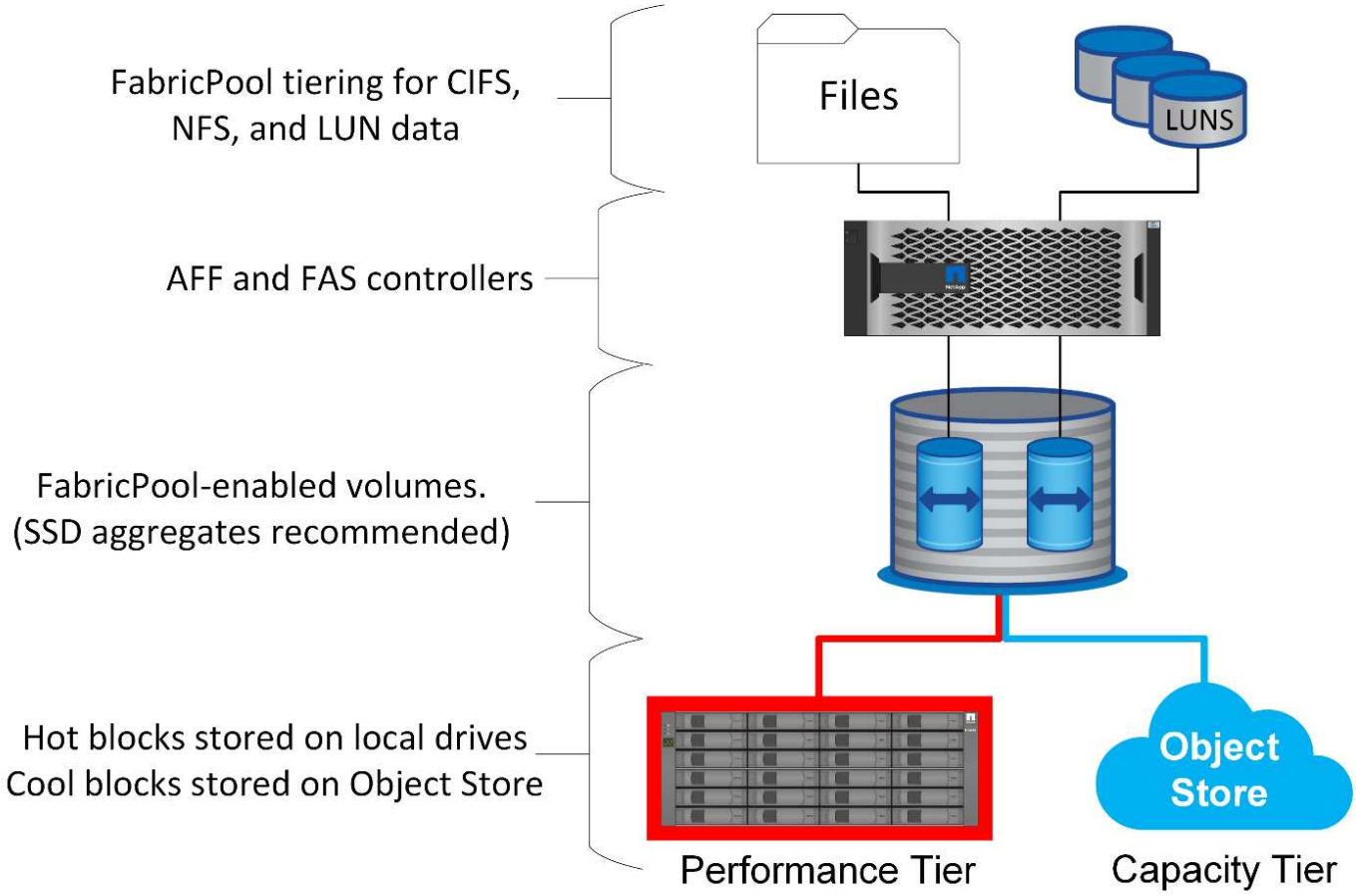
FabricPoolの階層化がOracleやその他のデータベースに与える影響を理解するには、低レベルのFabricPoolアーキテクチャについて理解しておく必要があります。

アーキテクチャ

FabricPoolは、ブロックをホットまたはクールに分類し、最も適切なストレージ階層に配置する階層化テクノロジです。ほとんどの場合、パフォーマンス階層はSSDストレージに配置され、ホットデータブロックをホストします。大容量階層はオブジェクトストアに配置され、クールデータブロックをホストします。サポートされるオブジェクトストレージには、NetApp StorageGRID、ONTAP S3、Microsoft Azure Blobストレージ、Alibaba Cloud Object Storageサービス、IBM Cloud Object Storage、Google Cloudストレージ、Amazon

AWS S3などがあります。

ブロックをホットまたはクールに分類する方法を制御する複数の階層化ポリシーを使用できます。ポリシーはボリューム単位で設定し、必要に応じて変更できます。パフォーマンス階層と大容量階層の間で移動されるのはデータブロックのみです。LUNとファイルシステムの構造を定義するメタデータは、常にパフォーマンス階層に残ります。そのため、管理はONTAPに一元化されます。ファイルとLUNは、他のONTAP構成に格納されているデータと変わりません。NetApp AFFコントローラまたはFASコントローラが定義されたポリシーを適用して、データを適切な階層に移動します。



オブジェクトストレージプロバイダ

オブジェクトストレージプロトコルでは、単純なHTTP要求またはHTTPS要求を使用して大量のデータオブジェクトを格納します。ONTAPからのデータアクセスは要求の迅速な処理に依存するため、オブジェクトストレージへのアクセスは信頼できるものでなければなりません。オプションには、Amazon S3の[Standard and InFrequent Access]オプション、Microsoft Azure Hot and Cool Blob Storage、IBM Cloud、Google Cloudなどがあります。Amazon GlacierやAmazon Archiveなどのアーカイブオプションはサポートされていません。データの読み出しに要する時間がホストのオペレーティングシステムやアプリケーションの許容範囲を超える可能性があるためです。

NetApp StorageGRIDもサポートされており、最適なエンタープライズクラスの解決策です。ハイパフォーマンス、拡張性、セキュリティに優れたオブジェクトストレージシステムであり、FabricPoolデータだけでなく、エンタープライズアプリケーション環境に組み込まれる可能性が高まるその他のオブジェクトストレージアプリケーションにも、地理的な冗長性を提供します。

また、StorageGRIDは、多くのパブリッククラウドプロバイダがサービスからデータを読み返す際に出力料金を課す必要がないため、コストを削減できます。

データとメタデータ

ここで「data」という用語は、メタデータではなく実際のデータブロックを環境することに注意してください。データブロックのみが階層化され、メタデータはパフォーマンス階層に残ります。また、あるブロックのステータス（hotまたはcool）が影響を受けるのは、実際のデータブロックを読み取った場合のみです。ファイルの名前、タイムスタンプ、所有権のメタデータを読み取っても、基盤となるデータブロックの場所には影響しません。

バックアップ

FabricPoolはストレージの設置面積を大幅に削減できますが、それだけではバックアップ解決策ではありません。NetApp WAFLメタデータは常に高パフォーマンス階層に残ります。大容量階層にはWAFLメタデータが含まれていないため、大容量階層のデータを使用して新しい環境を作成することはできません。

ただし、FabricPoolはバックアップ戦略の一部になる可能性があります。たとえば、FabricPoolにはNetApp SnapMirrorレプリケーションテクノロジを設定できます。ミラーの各半分は、オブジェクトストレージターゲットに独自に接続できます。その結果、データの2つの独立したコピーが作成されます。プライマリコピーは、パフォーマンス階層のブロックと大容量階層内の関連するブロックで構成され、レプリカはパフォーマンスブロックと容量ブロックの2つ目のセットです。

階層化ポリシー

Oracle Database FabricPool階層化ポリシー

ONTAPでは4つのポリシーを使用して、高パフォーマンス階層にあるOracleデータを大容量階層に再配置する方法を制御できます。

Snapshotのみ

。 `snapshot-only` `tiering-policy` アクティブファイルシステムと共有されていないブロックにのみ適用されます。基本的には、データベースバックアップの階層化につながります。Snapshotが作成されてそのブロックが上書きされると、ブロックが階層化の候補となり、その結果、Snapshot内にのみ存在するブロックが作成されます。Aの前の遅延 `snapshot-only` ブロックは冷却されていると見なされ、によって制御されます。 `tiering-minimum-cooling-days` ボリュームの設定。ONTAP 9.8で指定できる範囲は2～183日です。

多くのデータセットは変更率が低いため、このポリシーによる削減効果は最小限に抑えられます。たとえば、ONTAPで観察される一般的なデータベースの変更率は、週あたり5%未満です。データベースのアーカイブログは大量のスペースを占有することがありますが、通常はアクティブファイルシステムに引き続き存在するため、このポリシーでは階層化の対象になりません。

自動

。 `auto` 階層化ポリシーは、階層化をSnapshot固有のブロックだけでなく、アクティブなファイルシステム内のブロックにも拡張します。ブロックが冷却されるまでの遅延は、 `tiering-minimum-cooling-days` ボリュームの設定。ONTAP 9.8で指定できる範囲は2～183日です。

このアプローチでは、では使用できない階層化オプションが有効になります。 `snapshot-only` ポリシー：たとえば、データ保護ポリシーで特定のログファイルを90日間保持する必要がある場合があります。クーリング期間を3日に設定すると、3日を超過した古いログファイルがパフォーマンスレイヤから階層化されます。この操作により、パフォーマンス階層のかなりのスペースが解放されると同時に、90日間分のすべてのデータを表示して管理することができます。

なし

。 none 階層化ポリシーを使用すると、追加のブロックがストレージレイヤから階層化されなくなりますが、大容量階層のデータは読み取りが行われるまで大容量階層に残ります。その後ブロックが読み取られると、元に戻されてパフォーマンス階層に配置されます。

を使用する主な理由は、 none 階層化ポリシーはブロックが階層化されないようにするためのものですが、時間の経過とともにポリシーを変更すると便利です。たとえば、あるデータセットが大容量レイヤに階層化されているとしますが、完全なパフォーマンス機能が予期せず必要になったとします。このポリシーを変更すると、追加の階層化が不要になり、I/Oの増加に伴って読み取られたブロックがパフォーマンス階層に残るようにすることができます。

すべて

。 all 階層化ポリシーで置き換えられる backup ONTAP 9.6以降のポリシー。。 backup データ保護ボリューム（SnapMirrorまたはNetApp SnapVaultのデスティネーション）にのみ適用されるポリシー。。 all ポリシーの機能は同じですが、データ保護ボリュームに限定されません。

このポリシーでは、ブロックはすぐにクールとみなされ、すぐに容量レイヤに階層化できるようになります。

このポリシーは、長期的なバックアップに特に適しています。Hierarchical Storage Management（HSM；階層型ストレージ管理）の一種としても使用できます。以前は、ファイルシステム上でファイル自体を認識したまま、ファイルのデータブロックをテープに階層化するためにHSMが一般的に使用されていました。FabricPoolボリューム all ポリシーを使用すると、表示および管理可能なにファイルを格納できますが、ローカルストレージ階層のスペースはほとんど消費しません。

Oracleデータベースと**FabricPool**の読み出しポリシー

階層化ポリシーは、どのOracleデータベースブロックをパフォーマンス階層から大容量階層に階層化するかを制御します。読み出しポリシーは、階層化されたブロックが読み取られたときの処理を制御します。

デフォルト

すべてのFabricPoolボリュームの初期設定は `default`これは、動作が「cloud-retrieval-policy」によって制御されることを意味します。'正確な動作は、使用する階層化ポリシーによって異なります。

- auto–ランダムリードデータのみを取得
- snapshot-only–すべてのシーケンシャルまたはランダムリード・データを取得
- none–すべてのシーケンシャルまたはランダムリード・データを取得
- all–大容量階層からデータを取得しない

オンリード

設定 cloud-retrieval-policy をオンリードに設定するとデフォルトの動作が無効になるため、階層化されたデータが読み取られた場合、そのデータはパフォーマンス階層に返されます。

たとえば、ボリュームは、 auto 階層化ポリシーとほとんどのブロックが階層化されます。

ビジネスニーズの予期しない変化によって、特定のレポートを作成するために一部のデータを繰り返しスキャンする必要がある場合は、 cloud-retrieval-policy 終了： on-read シーケンシャルデータとランダム

リードデータの両方を含む、読み取りされるすべてのデータがパフォーマンス階層に返されます。これにより、ボリュームに対するシーケンシャルI/Oのパフォーマンスが向上します。

プロモート

昇格ポリシーの動作は階層化ポリシーによって異なります。階層化ポリシーがの場合 `auto`をクリックし、`cloud-retrieval-policy `to `promote 次回の階層化スキャンで大容量階層のすべてのブロックを戻します。

階層化ポリシーがの場合 `snapshot-only`を指定すると、アクティブファイルシステムに関連付けられているブロックのみが返されます。通常、これは効果がありません。これは、 `snapshot-only ポリシーは、Snapshotにのみ関連付けられたブロックになります。アクティブファイルシステムに階層化されたブロックはありません。

ただし、ボリュームSnapRestoreまたはSnapshotからのファイルクローン操作によってボリューム上のデータがリストアされた場合、Snapshotにのみ関連付けられていたために階層化されたブロックの一部がアクティブファイルシステムで必要になることがあります。一時的に `cloud-retrieval-policy` ポリシーの宛先 `promote` ローカルで必要なすべてのブロックを迅速に取得できます。

なし

大容量階層からブロックを取得しないでください。

階層化戦略

Oracleデータベースのファイル**FabricPool**のフル階層化

FabricPool階層化はブロックレベルで動作しますが、ファイルレベルの階層化に使用できる場合もあります。

多くのアプリケーションデータセットは日付別に整理されており、そのようなデータが古くなるにつれてアクセスされる可能性はますます低くなっています。たとえば、銀行が5年間の顧客明細書を含むPDFファイルのリポジトリを持っていても、最近の数か月のみがアクティブになっているとします。FabricPoolを使用して、古いデータファイルを大容量階層に再配置できます。クーリング期間を14日間にすると、直近の14日間のPDFファイルがパフォーマンス階層に残ります。さらに、少なくとも14日ごとに読み取られたファイルはホットのままであるため、パフォーマンス階層に残ります。

ポリシー

ファイルベースの階層化アプローチを実装するには、ファイルが書き込まれ、その後変更されないようにする必要があります。。 `tiering-minimum-cooling-days` 必要なファイルが高パフォーマンス階層に残るように、ポリシーを十分に高く設定する必要があります。たとえば、最新の60日分のデータが必要なデータセットで、最適なパフォーマンス保証が設定されているとします。 `tiering-minimum-cooling-days` 60までの期間。ファイルアクセスパターンに基づいても同様の結果が得られます。たとえば、最新の90日間のデータが必要で、アプリケーションがその90日間のデータにアクセスしている場合、データは高パフォーマンス階層に残ります。を設定する `tiering-minimum-cooling-days` 2までの期間では、データの使用頻度が低下した後、迅速な階層化が行われます。

。 `auto` これらのブロックの階層化を推進するにはポリシーが必要です。これは、 `auto` ポリシーは、アクティブファイルシステム内のブロックに影響します。

データにアクセスすると、ヒートマップデータがリセットされます。ウィルススキャン、インデックス作成、さらにはソースファイルを読み取るバックアップ処理も行われるため、 `tiering-minimum-cooling-days` しきい値に達していません。

Oracleの部分ファイル**FabricPool**階層化

FabricPoolはブロックレベルで機能するため、変更の可能性があるファイルは部分的にオブジェクトストレージに階層化し、部分的にパフォーマンス階層に残すことができます。

これはデータベースで一般的です。アクセス頻度の低いブロックが含まれていることがわかっているデータベースも、FabricPool階層化の候補になります。たとえば、サプライチェーン管理データベースには履歴情報が含まれている可能性があります。この情報は、必要に応じて利用できなければなりませんが、通常の運用中はアクセスできません。FabricPoolを使用すると、非アクティブなブロックを選択的に再配置できます。

たとえば、FabricPoolで実行されているデータファイルの場合、 `tiering-minimum-cooling-days` 90日の期間には、過去90日間にアクセスされたブロックがパフォーマンス階層に保持されます。ただし、90日間アクセスされなかったデータはすべて大容量階層に再配置されます。それ以外の場合は、通常のアプリケーションアクティビティで正しいブロックが正しい階層に保持されます。たとえば、データベースが通常、過去60日間のデータを定期的に処理するために使用されている場合、 `tiering-minimum-cooling-days` 期間を設定できるのは、アプリケーションの自然なアクティビティによって、ブロックが早期に再配置されないようにするためです。

。 `auto` データベースには注意してポリシーを使用する必要があります。多くのデータベースには、四半期末のプロセスやインデックスの再作成などの定期的なアクティビティがあります。これらの操作の期間が `tiering-minimum-cooling-days` パフォーマンスに問題が生じる可能性があります。たとえば、四半期末の処理で1TBのデータをまだ使用していない状態で処理する必要がある場合、そのデータは大容量階層に配置される可能性があります。大容量階層からの読み取りは非常に高速であることが多く、原因のパフォーマンスに問題はない可能性がありますが、正確な結果はオブジェクトストアの設定によって異なります。

ポリシー

。 `tiering-minimum-cooling-days` パフォーマンス階層で必要になる可能性のあるファイルを保持できるように、ポリシーを十分な高さに設定する必要があります。たとえば、最新の60日分のデータが必要でパフォーマンスが最適なデータベースでは、 `tiering-minimum-cooling-days` 60日までの期間。同様の結果は、ファイルのアクセスパターンに基づいても達成できます。たとえば、最新の90日間のデータが必要で、アプリケーションがその90日間のデータにアクセスしている場合、データは高パフォーマンス階層に残ります。を設定します `tiering-minimum-cooling-days` データの使用頻度が低下した場合は、2日間の期間でデータが階層化されます。

。 `auto` これらのブロックの階層化を推進するにはポリシーが必要です。これは、 `auto` ポリシーは、アクティブファイルシステム内のブロックに影響します。

データにアクセスすると、ヒートマップデータがリセットされます。そのため、データベースのテーブル全体がスキャンされ、ソースファイルを読み取るバックアップアクティビティも行われるため、 `tiering-minimum-cooling-days` しきい値に達していません。

Oracleデータベースのアーカイブログの階層化

FabricPoolの最も重要な用途は、データベーストランザクションログなどの既知のコールドデータの効率化です。

ほとんどのリレーショナルデータベースは、ポイントインタイムリカバリを実現するためにトランザクションログアーカイブモードで動作します。データベースへの変更は、トランザクションログに変更を記録することによってコミットされ、トランザクションログは上書きされずに保持されます。そのため、大量のアーカイブトランザクションログを保持しなければならない場合があります。同様の例は、保持する必要があるデータを生成する他の多くのアプリケーションワークフローにも存在しますが、アクセスされることはほとんどありません。

FabricPoolは、階層化が統合された単一の解決策を提供することで、これらの問題を解決します。ファイルは通常の場所に保存されてアクセス可能な状態に維持されますが'プライマリ・アレイのスペースはほとんど消費されません

ポリシー

を使用します `tiering-minimum-cooling-days` 数日間のポリシーを設定すると、最近作成されたファイル（短期的に必要になる可能性が高いファイル）のブロックが高パフォーマンス階層に保持されます。その後、古いファイルのデータブロックが大容量階層に移動されます。

。 `auto` ログが削除されたか、プライマリファイルシステムに引き続き存在しているかに関係なく、クーリングしきい値に達したときに、プロンプト階層化を適用します。必要となる可能性があるすべてのログをアクティブファイルシステムの1つの場所に格納することも、管理を簡易化します。リストアが必要なファイルを特定するためにSnapshotを検索する必要はありません。

Microsoft SQL Serverなどの一部のアプリケーションでは、バックアップ処理中にトランザクションログファイルが切り捨てられ、ログがアクティブファイルシステムに記録されなくなります。容量は、 `snapshot-only` 階層化ポリシー `auto` アクティブファイルシステムにはログデータが冷却されることはほとんどないため、ログデータにはポリシーは役立ちません。

Oracleと**FabricPool**スナップショットの階層化

FabricPoolの初期リリースでは、バックアップのユースケースを対象としていました。階層化できるブロックのタイプは、アクティブファイルシステム内のデータに関連付けられなくなったブロックだけです。そのため、大容量階層に移動できるのはSnapshotデータブロックだけです。これは、パフォーマンスに影響を与えないようにする必要がある場合に、最も安全な階層化オプションの1つです。

ポリシー**-**ローカル**Snapshot**

アクセス頻度の低いSnapshotブロックを大容量階層に階層化する方法は2つあります。まず、 `snapshot-only` ポリシーはSnapshotブロックのみを対象としています。ただし、 `auto` ポリシーには、 `snapshot-only` ブロックの場合は、アクティブファイルシステムのブロックも階層化されます。これは望ましくない可能性があります。

。 `tiering-minimum-cooling-days` この値は、リストア時に必要となる可能性のあるデータを高パフォーマンス階層で使用できるようにする期間に設定する必要があります。たとえば、重要な本番環境データベースのリストアシナリオのほとんどには、過去数日間のある時点のリストアポイントが含まれます。セッテイ `tiering-minimum-cooling-days` 値を3に設定すると、ファイルをリストアしたときにパフォーマンスがすぐに最大になるようにファイルが作成されます。アクティブファイル内のすべてのブロックは、大容量階層からリカバリすることなく高速ストレージに残ります。

ポリシー**-**レプリケートされた**Snapshot**

リカバリのみに使用されるSnapMirrorまたはSnapVaultでレプリケートされるSnapshotには、一般

にFabricPoolを使用する必要があります。 `all` ポリシー：このポリシーでは、メタデータはレプリケートされますが、すべてのデータブロックがただちに大容量階層に送信されるため、パフォーマンスが最大限に向上します。ほとんどのリカバリプロセスではシーケンシャルI/Oが発生しますが、これは本質的に効率的です。オブジェクトストアのデスティネーションからのリカバリ時間を評価する必要がありますが、適切に設計されたアーキテクチャでは、このリカバリプロセスにローカルデータからのリカバリよりも大幅に時間がかかる必要はありません。

レプリケートされたデータをクローニングにも使用する場合は、 `auto` ポリシーはより適切であり、 `tiering-minimum-cooling-days` クローニング環境で定期的に使用されることが期待されるデータを含む価値。たとえば、データベースのアクティブなワーキングセットには、過去3日間に読み書きされたデータが含まれている場合がありますが、さらに6カ月分の履歴データが含まれている場合もあります。その場合は、 `auto` SnapMirrorデスティネーションでポリシーを設定すると、作業セットを高パフォーマンス階層で使用できるようになります。

Oracleデータベースバックアップの階層化

従来のアプリケーションバックアップには、Oracle Recovery Managerなどの製品が含まれています。Oracle Recovery Managerは、元のデータベースの場所以外にファイルベースのバックアップを作成します。

```
 `tiering-minimum-cooling-days` policy of a few days preserves the most
recent backups, and therefore the backups most likely to be required for
an urgent recovery situation, on the performance tier. The data blocks of
the older files are then moved to the capacity tier.
```

。 `auto`
ポリシーは、バックアップデータに最も適したポリシーです。これにより、ファイルが削除されたか、プライマリファイルシステムに引き続き存在しているかに関係なく、クーリングしきい値に達したときに迅速に階層化されます。必要となる可能性があるすべてのファイルをアクティブファイルシステムの1つの場所に格納することも、管理を簡易化します。リストアが必要なファイルを特定するために`Snapshot`を検索する必要はありません。

。 `snapshot-only` ポリシーは機能するように設定できますが、アクティブファイルシステムに存在しなくなった環境ブロックのみが対象となります。そのため、データを階層化するには、まずNFS共有またはSMB共有上のファイルを削除する必要があります。

LUNからファイルを削除するとファイル参照がファイルシステムのメタデータから削除されるだけなので、このポリシーはLUN設定の場合はさらに効率的ではありません。LUN上の実際のブロックは、上書きされるまでそのまま維持されます。このような状況では、ファイルが削除されてブロックが上書きされて階層化の候補になるまでに長時間の遅延が発生する可能性があります。の移動にはいくつかの利点があります。 `snapshot-only` ブロックは大容量階層に移動しますが、全体的にはバックアップデータのFabricPool管理が最適なのは、 `auto` ポリシー：

> このアプローチは、バックアップに必要なスペースをより効率的に管理するのに役立ちますが、FabricPool自体はバックアップテクノロジではありません。バックアップファイルをオブジェクトストアに階層化すると、ファイルは元のストレージシステムに引き続き表示されるため、管理が簡易化されますが、オブジェクトストアデスティネーションのデータブロックは元のストレージシステムに依存します。ソースボリュームが失われると、オブジェクトストアのデータを使用できなくなります。

Oracleデータベースとオブジェクトストアへのアクセスの中断

FabricPoolでデータセットを階層化すると、プライマリストレージアレイとオブジェクトストア階層の間に依存関係が生じます。オブジェクトストレージには、さまざまなレベルの可用性を提供するオプションが多数あります。プライマリストレージアレイとオブジェクトストレージ階層の間の接続が失われた場合の影響を理解することが重要です。

ONTAPに対して実行するI/Oで大容量階層のデータが必要になり、ONTAPが大容量階層に到達してブロックを読み出すことができない場合、最終的にI/Oはタイムアウトします。このタイムアウトの影響は、使用するプロトコルによって異なります。NFS環境では、ONTAPはプロトコルに応じてEJUKEBOXまたはEDELAYのいずれかの応答で応答します。一部の古いオペレーティングシステムではエラーと解釈される場合がありますが、現在のオペレーティングシステムとOracle Direct NFSクライアントの現在のパッチレベルでは、これを再試行可能なエラーとして扱い、I/Oの完了を待ち続けます。

環境SAN環境のタイムアウトを短縮します。オブジェクトストア環境のブロックが必要で、2分間アクセスできない場合は、読み取りエラーがホストに返されます。ONTAPボリュームとLUNはオンラインのままですが、ホストOSからファイルシステムにエラー状態のフラグが設定されることがあります。

オブジェクトストレージの接続の問題 `snapshot-only` バックアップデータのみが階層化されるため、ポリシーはそれほど重要ではありません。通信に問題があると、データのリカバリに時間がかかりますが、それ以外の場合はアクティブに使用されているデータに影響。 `auto` および `all` ポリシーを使用すると、アクティブなLUNからコールドデータを階層化できます。つまり、オブジェクトストアデータの読み出し中にエラーが発生すると、データベースの可用性に影響する可能性があります。これらのポリシーを使用したSAN環境は、高可用性を実現するように設計されたエンタープライズクラスのオブジェクトストレージとネットワーク接続でのみ使用してください。NetApp StorageGRIDは優れたオプションです。

Oracleのデータ保護

ONTAPによるOracleデータ保護

NetAppは、最もミッションクリティカルなデータがデータベースに含まれていることを認識しています。

企業はデータへのアクセスなしでは業務を遂行できず、場合によってはデータによってビジネスが決まることもあります。このようなデータは保護する必要がありますが、データ保護では、使用可能なバックアップを確保するだけでなく、バックアップを安全に保管するだけでなく、迅速かつ確実に実行することも重要です。

データ保護のもう1つの側面は、データリカバリです。データにアクセスできなくなると企業は影響を受け、データがリストアされるまで操作できなくなる可能性があります。このプロセスは高速で信頼性が必要です。最後に、ほとんどのデータベースを災害から保護する必要があります。つまり、データベースのレプリカを維持する必要があります。レプリカは十分に最新である必要があります。また'レプリカを完全に動作可能なデータベースにするには'迅速かつ簡単に行う必要があります

> このドキュメントは、以前に公開されていたテクニカルレポート_TR-4591：『Oracle data protection：Backup、recovery、and replication』に代わるものです。_

計画

適切なエンタープライズデータ保護アーキテクチャは、さまざまなイベントにおけるデータの保持、リカバリ

性、耐障害性に関するビジネス要件に依存します。

たとえば、対象となるアプリケーション、データベース、重要なデータセットの数を考えてみましょう。管理するオブジェクトが少ないため、一般的なSLAへの準拠を保証する単一データセットのバックアップ戦略の構築は非常に簡単です。データセットの数が増えるにつれて監視が複雑になり、バックアップの失敗に対処するために、管理者がますます多くの時間を費やすことになる可能性があります。環境がクラウドに到達し、サービスプロバイダが拡張するにつれて、まったく異なるアプローチが必要になります。

データセットのサイズも戦略に影響します。たとえば、データセットが非常に小さいため、100GBのデータベースのバックアップとリカバリには多くのオプションがあります。従来のツールを使用してバックアップメディアからデータをコピーするだけで、リカバリに十分なRTOが得られます。通常、100TBのデータベースでは、RTOによって複数日の停止が許容される場合を除き、まったく異なる戦略が必要になります。その場合は、従来のコピーベースのバックアップおよびリカバリの手順で十分かもしれません。

最後に、バックアップとリカバリのプロセス自体以外にもさまざまな要因があります。たとえば、重要な本番環境のアクティビティをサポートしているデータベースがあり、熟練したDBAだけがリカバリを実行するまれなイベントになっているとしますか。あるいは、データベースは、リカバリが頻繁に発生し、ジェネラリストのITチームが管理する大規模な開発環境に含まれていますか。

Oracleデータベースの**RTO**、**RPO**、**SLA**計画

ONTAPを使用すると、Oracleデータベースのデータ保護戦略をビジネス要件に簡単にカスタマイズできます。

これらの要件には、リカバリの速度、許容される最大データ損失、バックアップの保持ニーズなどの要因が含まれます。データ保護計画では、データの保持とリストアに関するさまざまな規制要件も考慮する必要があります。最後に、さまざまなデータリカバリシナリオを検討する必要があります。たとえば、ユーザやアプリケーションのエラーに起因する一般的で予測可能なリカバリから、サイト全体の損失を含むディザスタリカバリのシナリオまで、さまざまなシナリオを検討する必要があります。

データ保護ポリシーとリカバリポリシーのわずかな変更は、ストレージ、バックアップ、リカバリのアーキテクチャ全体に大きな影響を与える可能性があります。データ保護アーキテクチャが複雑にならないように、設計作業を開始する前に標準を定義して文書化することが重要です。不要な機能や保護レベルは、不要なコストや管理オーバーヘッドにつながります。また、最初に見落とされた要件は、プロジェクトを間違った方向に進めたり、直前の設計変更を必要としたりする可能性があります。

目標復旧時間

Recovery Time Objective（RTO；目標復旧時間）は、サービスのリカバリに許容される最大時間を定義します。たとえば、人事データベースのRTOが24時間になる可能性があります。これは、営業日中にこのデータにアクセスできなくなることは非常に不便ですが、ビジネスを継続できるためです。一方、銀行の総勘定元帳をサポートするデータベースでは、数分または数秒でRTOを測定できます。RTOをゼロにすることはできません。これは、実際のサービス停止と、ネットワークパケットの損失などの日常的なイベントを区別する方法が必要であるためです。ただし、一般的な要件はRTOがほぼゼロです。

目標復旧時点

Recovery Point Objective（RPO；目標復旧時点）は、最大許容データ損失を定義します。多くの場合、RPOはSnapshotまたはSnapMirror更新の頻度によって決まります。

場合によっては、RPOをより積極的に設定し、特定のデータをより頻繁に選択的に保護することができます。データベースのコンテキストでは、通常、RPOは、特定の状況で失われる可能性のあるログデータの量です。製品のバグやユーザエラーによってデータベースが破損した一般的なリカバリシナリオでは、RPOは

ゼロ、つまりデータ損失がないはずです。リカバリ手順では、データベースファイルの以前のコピーをリストアし、ログファイルを再生して、データベースを希望する時点の状態にします。この処理に必要なログファイルは元の場所にすでに存在している必要があります。

通常とは異なる状況では、ログデータが失われる可能性があります。たとえば、偶発的または悪意のある `rm -rf *` データベースファイルのすべてのデータが削除される可能性があります。唯一の方法は、ログファイルを含むバックアップからリストアすることであり、一部のデータは必然的に失われます。従来のバックアップ環境でRPOを向上させる唯一の方法は、ログデータのバックアップを繰り返し実行することです。しかし、データが絶えず移動し、バックアップシステムを継続的に実行されるサービスとして維持することが困難であるため、これには限界があります。高度なストレージシステムのメリットの1つは、偶発的または悪意のあるファイルの破損からデータを保護し、データを移動せずにRPOを向上できることです。

ディザスタリカバリ

ディザスタリカバリには、物理的な災害が発生した場合にサービスをリカバリするために必要なITアーキテクチャ、ポリシー、および手順が含まれます。これには、洪水、火災、または悪意または過失の意図を持って行動する人が含まれます。

ディザスタリカバリは、単なるリカバリ手順ではありません。これは、さまざまなリスクを特定し、データリカバリとサービス継続性の要件を定義し、適切なアーキテクチャと関連手順を提供する完全なプロセスです。

データ保護の要件を確立するには、一般的なRPOとRTOの要件と、ディザスタリカバリに必要なRPOとRTOの要件を区別することが重要です。一部のアプリケーション環境では、比較的通常のユーザエラーからデータセンターの破壊に至るまで、データ損失の状況に対して、RPOゼロとRTOほぼゼロを達成する必要があります。ただし、これらの高レベルの保護にはコストと管理上の影響があります。

一般に、ディザスタ以外のデータリカバリの要件は、次の2つの理由で厳しいものにする必要があります。まず、データに損害を与えるアプリケーションのバグやユーザエラーは、ほぼ避けられないほど予測可能です。2つ目は、ストレージシステムが破損していないかぎり、RPOをゼロにしてRTOを短縮できるバックアップ戦略を設計することです。容易に修復できる重大なリスクに対処しない理由はありません。そのため、ローカルリカバリのRPOとRTOの目標を積極的に設定する必要があります。

ディザスタリカバリのRTOとRPOの要件は、災害が発生する可能性や、関連するデータの損失やビジネスの中断がもたらす影響によって大きく異なります。RPOとRTOの要件は、一般的な原則ではなく、実際のビジネスニーズに基づいている必要があります。論理的および物理的な複数の災害シナリオを考慮する必要があります。

論理的災害

論理的災害には、ユーザによるデータ破損、アプリケーションやOSのバグ、ソフトウェアの誤動作などがあります。論理的災害には、ウイルスやワームによる外部からの悪意のある攻撃や、アプリケーションの脆弱性を悪用した悪意のある攻撃も含まれます。この場合、物理インフラは破損していませんが、基盤となるデータは無効になります。

ランサムウェアと呼ばれる論理災害のタイプはますます一般的になりつつあり、攻撃ベクトルを使用してデータを暗号化します。暗号化はデータを損傷することはありませんが、サードパーティに支払いが行われるまで使用できなくなります。ランサムウェアのハッキングを特に標的にされる企業は、ますます増えています。この脅威に対して、NetAppには改ざん防止スナップショットが用意されており、ストレージ管理者であっても、設定された有効期限までに保護されたデータを変更することはできません。

物理的災害

物理的災害には、インフラストラクチャのコンポーネントの障害がその冗長性機能を超え、データの損失やサービスの長期的な損失につながることが含まれます。たとえば、RAID保護ではディスクドライブの冗長性が

提供され、HBAを使用することでFCポートとFCケーブルの冗長性が提供されます。このようなコンポーネントのハードウェア障害は予測可能であり、可用性には影響しません。

エンタープライズ環境では、通常、サイト全体のインフラストラクチャを冗長コンポーネントで保護し、予測可能な唯一の物理的災害シナリオがサイトの完全な損失である時点まで保護することができます。ディザスタリカバリ計画は、サイト間レプリケーションによって異なります。

同期および非同期のデータ保護

理想的な環境では、地理的に分散したサイト間ですべてのデータを同期的にレプリケートできます。このようなレプリケーションは、次のようないくつかの理由により、必ずしも実現可能ではありません。

- 同期レプリケーションでは、アプリケーションやデータベースの処理を続行する前にすべての変更を両方の場所にレプリケートする必要があるため、書き込みレイテンシが避けられません。このようなパフォーマンスへの影響が許容できない場合があり、同期ミラーリングの使用が除外されます。
- 100% SSDストレージの採用が増加しているため、期待されるパフォーマンスには数十万IOPSと1ミリ秒未満のレイテンシが含まれているため、書き込みレイテンシの増加に気付く可能性が高くなります。100% SSDを使用するメリットを最大限に引き出すには、ディザスタリカバリ戦略を見直す必要があります。
- データセットはバイト単位で増え続けているため、同期レプリケーションを維持するのに十分な帯域幅を確保するという課題が生じています。
- データセットも複雑化し、大規模な同期レプリケーションの管理が困難になっています。
- クラウドベースの戦略では、多くの場合、レプリケーションの距離とレイテンシが長くなり、同期ミラーリングの使用がさらに困難になります。

NetAppは、最も厳しいデータリカバリ要件に対応する同期レプリケーションと、パフォーマンスと柔軟性の向上を可能にする非同期ソリューションの両方を含むソリューションを提供しています。さらに、NetAppテクノロジは、Oracle DataGuardなどの多くのサードパーティ製レプリケーションソリューションとシームレスに統合されます。

保持時間

データ保護戦略の最後の側面は、データの保持期間です。データの保持期間は大きく異なる場合があります。

- 一般的な要件は、プライマリサイトに夜間バックアップを14日間、セカンダリサイトにバックアップを90日間保存することです。
- 多くのお客様が'異なるメディアに保存された四半期ごとのスタンドアロンアーカイブを作成しています
- 定期的に更新されるデータベースでは、履歴データは不要であり、バックアップは数日間だけ保持する必要があります。
- 規制要件によっては、任意のトランザクションを365日以内にリカバリできることが求められる場合があります。

ONTAPによるOracleデータベースの可用性

ONTAPは、Oracleデータベースの可用性を最大限に高めるように設計されています。概要of ONTAPの高可用性機能は、本ドキュメントでは扱いません。ただし、データ保護と同様に、データベースインフラを設計する際には、この機能の基本的な理解が重要です。

HA ペア

ハイアベイラビリティの基本単位はHAペアです。各ペアには、NVRAMへのデータのレプリケーションをサポートするための冗長リンクが含まれています。NVRAMは書き込みキャッシュではありません。コントローラ内部のRAMは書き込みキャッシュとして機能します。NVRAMの目的は、予期しないシステム障害から保護するためにデータを一時的にジャーナルすることです。この点では、データベースのREDOログに似ています。

NVRAMとデータベースのRedoログはどちらもデータを迅速に格納するために使用されるため、データに対する変更をできるだけ迅速にコミットできます。ドライブ（データファイル）上の永続的データの更新は、ONTAPとほとんどのデータベースプラットフォームの両方でチェックポイントと呼ばれるプロセスが実行されるまで行われません。通常運用時は、NVRAMデータもデータベースのREDOログも読み取られません。

コントローラで突然障害が発生した場合、ドライブにまだ書き込まれていない保留中の変更がNVRAMに保存されている可能性があります。パートナーコントローラが障害を検出してドライブを制御し、NVRAMに保存されている必要な変更を適用します。

テイクオーバーとギブバック

テイクオーバーとギブバックは、HAペアのノード間でストレージリソースの責任を移すプロセスです。テイクオーバーとギブバックには次の2つの側面があります。

- ドライブへのアクセスを許可するネットワーク接続の管理
- ドライブ自体の管理

CIFSおよびNFSトラフィックをサポートするネットワークインターフェイスには、ホームロケーションとフェイルオーバーロケーションの両方が設定されます。テイクオーバーでは、ネットワークインターフェイスを元の場所と同じサブネットにある物理インターフェイス上の一時的なホームに移動します。ギブバックでは、ネットワークインターフェイスを元の場所に戻します。必要に応じて、正確な動作を調整できます。

iSCSIやFCなどのSANブロックプロトコルをサポートしているネットワークインターフェイスは、テイクオーバーやギブバックの実行時に再配置されません。代わりに、完全なHAペアを含むパスを使用してLUNをプロビジョニングする必要があります。これにより、プライマリパスとセカンダリパスが作成されます。

> 大規模なクラスタ内のノード間でデータを再配置できるように、追加のコントローラへの追加のパスを設定することもできますが、これはHAプロセスの一部ではありません。

テイクオーバーとギブバックの2つ目の側面は、ディスク所有権の移行です。具体的なプロセスは、テイクオーバー/ギブバックの理由や実行したコマンドラインオプションなど、複数の要因によって異なります。目標は、できるだけ効率的に操作を実行することです。全体的なプロセスには数分かかるように見えるかもしれませんが、ドライブの所有権がノードからノードに移行される実際の瞬間は、通常数秒で測定できます。

テイクオーバー時間

テイクオーバー処理やギブバック処理の実行中にホストI/Oが短時間中断されますが、正しく設定された環境ではアプリケーションが停止することはありません。I/Oが遅延する実際の移行プロセスは通常数秒で測定されますが、ホストがデータパスの変更を認識してI/O処理を再送信するために、さらに時間がかかる場合があります。

中断の内容はプロトコルによって異なります。

- NFSおよびCIFSトラフィックをサポートするネットワークインターフェイスは、新しい物理的な場所への

移行後に、ネットワークに対してAddress Resolution Protocol（ARP；アドレス解決プロトコル）要求を発行します。これにより、ネットワークスイッチはメディアアクセス制御（MAC）アドレステーブルを更新し、I/Oの処理を再開します。計画的なテイクオーバーとギブバックの停止は、通常数秒で測定され、多くの場合は検出されません。ネットワークによっては、ネットワークパスの変更を完全に認識するのに時間がかかる場合があります。また、OSによっては、再試行が必要な大量のI/Oが短時間にキューイングされる場合があります。これにより、I/Oの再開に必要な時間が長くなる可能性があります。

- SANプロトコルをサポートするネットワークインターフェイスが新しい場所に移行されない。ホストOSが使用中のパスを変更する必要があります。ホストで検出されるI/Oの一時停止は、複数の要因によって異なります。ストレージシステムの観点から見ると、I/Oを処理できない時間はわずか数秒です。ただし、ホストOSによっては、I/Oがタイムアウトしてから再試行されるまでにさらに時間がかかる場合があります。新しいOSではパスの変更をより迅速に認識できますが、古いOSでは通常、変更を認識するのに最大30秒かかります。

次の表に、ストレージシステムがアプリケーション環境にデータを提供できない場合の想定テイクオーバー時間を示します。どのアプリケーション環境にもエラーは発生しません。テイクオーバーはI/O処理の一時停止として表示されます。

	NFS	AFF	ASA
計画的テイクオーバー	15秒	6～10秒	2～3秒
計画外のテイクオーバー	30秒	6～10秒	2～3秒

チェックサムと**Oracle**データベースの整合性

ONTAPとそのサポートされているプロトコルには、保存データとネットワーク経由で転送されるデータの両方を含む、Oracleデータベースの整合性を保護する複数の機能が含まれています。

ONTAPでの論理データ保護は、次の3つの重要な要件で構成されます。

- データを破損から保護する必要があります。
- データはドライブ障害から保護する必要があります。
- データへの変更は損失から保護する必要があります。

この3つのニーズについては、以降のセクションで説明します。

ネットワークの破損:チェックサム

最も基本的なデータ保護レベルはチェックサムです。チェックサムは、データと一緒に格納される特別なエラー検出コードです。ネットワーク転送中のデータの破損は、チェックサムを使用して検出されます。場合によっては、複数のチェックサムを使用します。

たとえば、FCフレームには巡回冗長検査（CRC）と呼ばれるチェックサム形式が含まれており、転送中にペイロードが破損していないことを確認できます。送信機は、データのデータとCRCの両方を送信します。FCフレームの受信側は、受信したデータのCRCを再計算して、送信されたCRCと一致することを確認します。新しく計算されたCRCがフレームに接続されたCRCと一致しない場合、データは破損し、FCフレームは破棄または拒否されます。iSCSI I/O処理には、TCP/IPおよびイーサネットレイヤでのチェックサムが含まれます。また、保護を強化するために、SCSIレイヤでオプションのCRC保護を含めることもできます。ワイヤ上のビットの破損はTCPレイヤまたはIPレイヤによって検出され、パケットが再送信されます。FCと同様に、SCSI CRCでエラーが発生すると、処理が破棄または拒否されます。

ドライブの破損：チェックサム

チェックサムは、ドライブに格納されているデータの整合性を検証するためにも使用されます。ドライブに書き込まれたデータブロックは、元のデータに関連付けられた予測不可能な数を生成するチェックサム機能で格納されます。ドライブからデータが読み取られると、チェックサムが再計算され、保存されているチェックサムと比較されます。一致しない場合は、データが破損しているため、RAIDレイヤでリカバリする必要があります。

データ破損：失われた書き込み

検出するのが最も困難な種類の破損の1つは、書き込みの紛失または置き忘れです。書き込みが確認応答されたら、正しい場所にあるメディアに書き込む必要があります。インプレースデータの破損は、データとともに保存されたシンプルなチェックサムを使用することで、比較的簡単に検出できます。ただし、書き込みが失われただけの場合は、以前のバージョンのデータが残っている可能性があり、チェックサムが正しいことになります。書き込みが間違った物理的な場所に配置された場合、書き込みによって他のデータが破壊されても、関連するチェックサムは保存データに対して再び有効になります。

この課題に対する解決策は次のとおりです。

- 書き込み処理には、書き込みが予想される場所を示すメタデータが含まれている必要があります。
- 書き込み処理には、何らかのバージョン識別子が含まれている必要があります。

ONTAPがブロックを書き込むときは、そのブロックが属する場所のデータも含まれます。後続の読み取りでブロックが識別されていても、メタデータにブロックが456の場所で見つかったときに123の場所に属していることが示されている場合、書き込みは誤って配置されています。

完全に失われた書き込みを検出することは、より困難です。説明は非常に複雑ですが、基本的にONTAPは、書き込み処理によってドライブ上の2つの場所が更新されるようにメタデータを格納します。書き込みが失われると、その後のデータおよび関連するメタデータの読み取りで、2つの異なるバージョンIDが表示されます。これは、ドライブによる書き込みが完了しなかったことを示します。

書き込みの破損が失われたり置き忘れられたりすることは非常にまれですが、ドライブが増え続け、データセットがエクサバイト規模になると、リスクが増大します。データベースワークロードをサポートするストレージシステムには、Lost Write検出機能を含める必要があります。

ドライブ障害：**RAID**、**RAID DP**、**RAID-TEC**

ドライブ上のデータブロックが破損していることが検出された場合、またはドライブ全体で障害が発生して完全に使用できなくなった場合は、データを再構成する必要があります。これは、ONTAPでパリティドライブを使用して行われます。データが複数のデータドライブにストライピングされ、パリティデータが生成されます。これは元のデータとは別に保存されます。

ONTAPは元 々 RAID 4を使用していました。RAID 4は、データドライブのグループごとにパリティドライブを1本使用します。その結果、グループ内のいずれかのドライブで障害が発生してもデータが失われることはありませんでした。パリティドライブで障害が発生してもデータは破損しておらず、新しいパリティドライブを構築できました。1本のデータドライブで障害が発生した場合は、残りのドライブをパリティドライブと一緒に使用して失われたデータを再生成します。

ドライブが小さい場合、2本のドライブで同時に障害が発生する可能性はほとんどありませんでした。ドライブ容量の増大に伴い、ドライブ障害発生後のデータの再構築に必要な時間も増加しています。これにより、2つ目のドライブ障害が発生してデータが失われる時間が長くなりました。また、再構築プロセスでは、稼働しているドライブに多くのI/Oが追加で作成されます。ドライブが古くなると、負荷が増えて2つ目のドライブ障害が発生するリスクも高まります。最後に、RAID 4を継続して使用することでデータ損失のリスクが増加し

なかったとしても、データ損失の影響はより深刻になります。RAIDグループで障害が発生した場合に失われるデータが多いほど、データのリカバリにかかる時間が長くなり、ビジネスの中断が長くなります。

これらの問題により、NetAppはRAID 6の一種であるNetApp RAID DP技術を開発した。この解決策にはパリティドライブが2本含まれているため、RAIDグループ内の2本のドライブで障害が発生してもデータが失われることはありません。ドライブのサイズは拡大を続けており、その結果、NetAppは3つ目のパリティドライブを導入するNetApp RAID-TECテクノロジを開発しました。

一部の履歴データベースのベストプラクティスでは、ストライプミラーリングとも呼ばれるRAID-10の使用を推奨しています。2本のディスクで障害が発生するシナリオが複数あるのに対し、RAID DPでは何も発生しないため、RAID DPよりもデータ保護が劣ります。

また、パフォーマンス上の懸念から、RAID-4 / 5 / 6よりもRAID-10が推奨されることを示す履歴データベースのベストプラクティスもいくつかあります。これらの推奨事項は、RAIDペナルティを意味する場合があります。これらの推奨事項は一般的に正しいものですが、ONTAP内でのRAIDの実装には適用されません。パフォーマンスの問題はパリティ再生に関連しています。従来のRAID実装では、データベースによって実行されるルーチンのランダムライトを処理するには、パリティデータを再生成して書き込みを完了するために、複数のディスク読み取りが必要です。ペナルティは、書き込み処理の実行に必要な追加の読み取りIOPSとして定義されます。

書き込みはメモリでステージングされ、パリティが生成されてから単一のRAIDストライプとしてディスクに書き込まれるため、ONTAPではRAIDペナルティは発生しません。書き込み処理を完了するための読み取りは必要ありません。

要約すると、RAID DPとRAID-TECは、RAID 10と比較して使用可能な容量がはるかに多く、ドライブ障害に対する保護が強化され、パフォーマンスが低下することはありません。

ハードウェア障害からの保護**:NVRAM**

データベースワークロードを処理するストレージアレイでは、書き込み処理をできるだけ迅速に処理する必要があります。さらに、電源障害などの予期しないイベントから書き込み処理を損失から保護する必要があります。つまり、書き込み処理は少なくとも2つの場所に安全に格納する必要があります。

AFFシステムとFASシステムは、これらの要件を満たすためにNVRAMを利用しています。書き込みプロセスは次のように機能します。

1. インバウンド書き込みデータはRAMに格納されます。
2. ディスク上のデータに加えなければならない変更は、ローカルノードとパートナーノードの両方のNVRAMに記録されます。NVRAMは書き込みキャッシュではなく、データベースのRedoログに似たジャーナルです。通常の条件下では、読み取りは行われません。I/O処理中に電源障害が発生した場合など、リカバリにのみ使用されます。
3. その後、書き込みがホストに確認応答されます。

この段階の書き込みプロセスはアプリケーションの観点からは完了しており、データは2つの異なる場所に格納されるため、損失から保護されます。最終的に変更はディスクに書き込まれますが、書き込みが確認されたあとに実行されるためレイテンシに影響しないため、このプロセスはアプリケーションの観点からはアウトオブバンドです。このプロセスもデータベースロギングに似ています。データベースに対する変更はできるだけ早くREDOログに記録され、変更がコミットされたことが確認されます。データファイルの更新はかなり遅れて行われ、処理速度に直接影響することはありません。

コントローラで障害が発生すると、パートナーコントローラが必要なディスクの所有権を取得し、ログに記録されたデータをNVRAMに再生して、障害発生時に転送中だったI/O処理をリカバリします。

ハードウェア障害からの保護：**NVFAIL**

前述したように、書き込みの確認応答は、少なくとも1台の他のコントローラでローカルのNVRAMとNVRAMに記録されるまで返されません。このアプローチにより、ハードウェア障害や停電が発生しても、転送中のI/Oが失われることはありません。ローカルのNVRAMに障害が発生したり、HAパートナーへの接続に障害が発生したりすると、この実行中のデータはミラーリングされなくなります。

ローカルNVRAMからエラーが報告されると、ノードはシャットダウンします。このシャットダウンにより、HAパートナーコントローラにフェイルオーバーします。障害が発生したコントローラが書き込み処理を確認していないため、データが失われることはありません。

データが同期されていない場合、ONTAPは、強制的にフェイルオーバーを実行しない限り、フェイルオーバーを許可しません。この方法で条件を変更すると、元のコントローラにデータが残っている可能性があり、データ損失が許容されることが確認されます。

データベースはディスク上のデータの大規模な内部キャッシュを保持しているため、フェイルオーバーが強制された場合、データベースが破損する可能性が特に高くなります。強制的なフェイルオーバーが発生した場合、以前に承認された変更は事実上破棄されます。ストレージアレイの内容は実質的に時間を逆方向に移動し、データベースキャッシュの状態はディスク上のデータの状態を反映しなくなります。

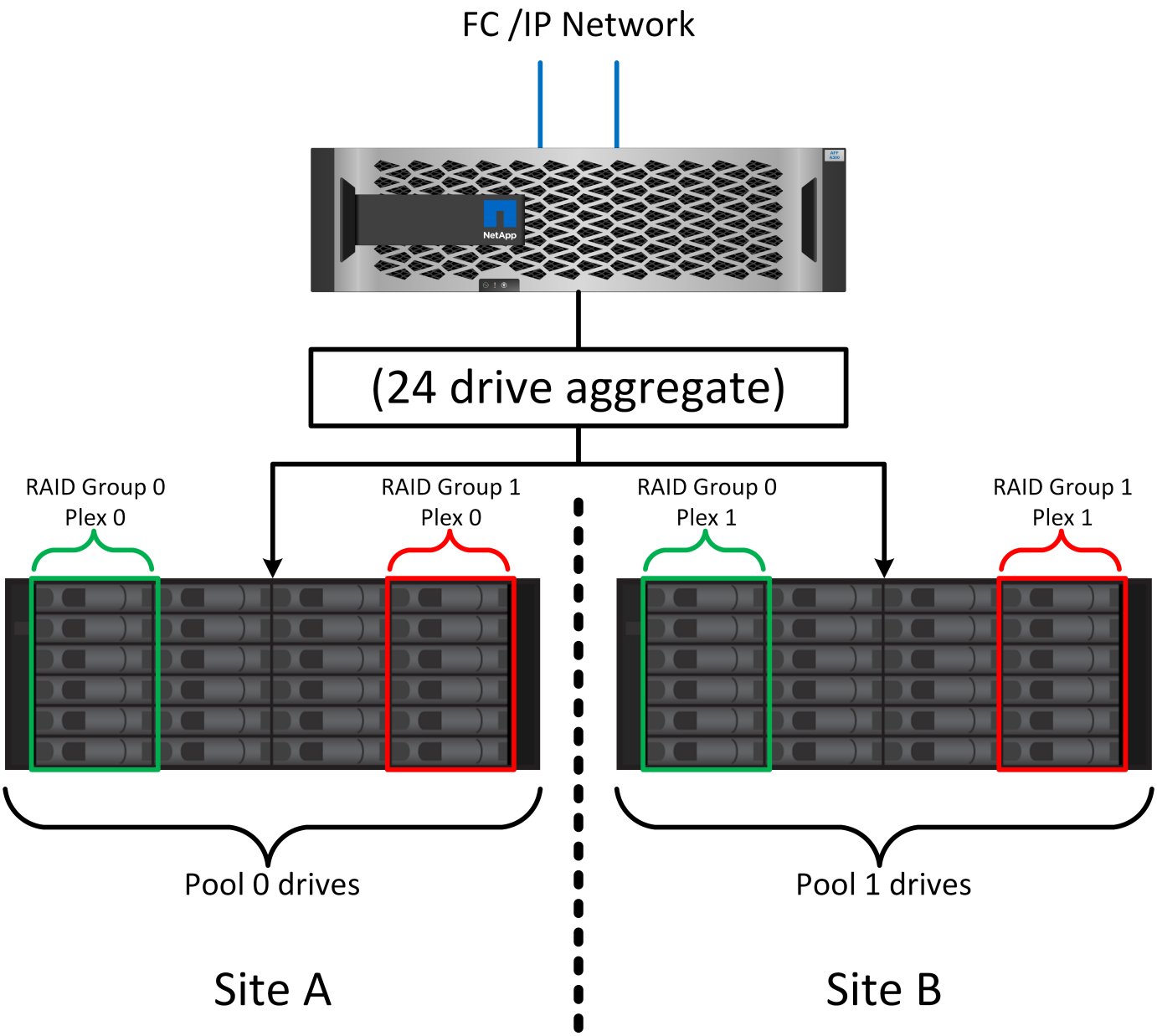
この状況からデータを保護するために、ONTAPでは、NVRAMの障害に対する特別な保護をボリュームに設定できます。この保護メカニズムがトリガーされると、ボリュームがNVFAILという状態になります。この状態では、古いデータを使用しないように原因Aアプリケーションをシャットダウンするl/Oエラーが発生します。確認済みの書き込みがストレージアレイに存在する必要があるため、データは失われません。

次の手順では、管理者がホストを完全にシャットダウンしてから、LUNとボリュームを手動で再度オンラインに戻します。これらの手順にはいくつかの作業が含まれる可能性がありますが、このアプローチはデータの整合性を確保するための最も安全な方法です。すべてのデータがこの保護を必要とするわけではありません。そのため、NVFAILの動作はボリューム単位で設定できます。

サイトおよびシェルフ障害からの保護：**SyncMirror**とプレックス

SyncMirrorは、RAID DPやRAID-TECを強化するミラーリングテクノロジですが、これに代わるものではありません。2つの独立したRAIDグループの内容をミラーリングします。論理構成は次のとおりです。

- ドライブは、場所に基づいて2つのプールに構成されます。1つのプールはサイトAのすべてのドライブで構成され、2つ目のプールはサイトBのすべてのドライブで構成されます。
- 次に、アグリゲートと呼ばれる共通のストレージプールが、RAIDグループのミラーセットに基づいて作成されます。各サイトから同じ数のドライブが引き出されます。たとえば、20ドライブのSyncMirrorアグリゲートは、サイトAの10本のドライブとサイトBの10本のドライブで構成されます。
- 特定のサイトのドライブセットは、ミラーリングを使用することなく、1つ以上の完全に冗長化されたRAID-DPまたはRAID-TECグループとして自動的に構成されます。これにより、サイトが失われても継続的なデータ保護が実現します。



上の図は、SyncMirror構成の例を示しています。24ドライブのアグリゲートをコントローラに作成しました。このアグリゲートは、サイトAで割り当てられたシェルフの12本のドライブと、サイトBで割り当てられたシェルフの12本のドライブで構成されています。ドライブは2つのミラーRAIDグループにグループ化されました。RAIDグループ0には、サイトAの6ドライブプレックスが含まれており、サイトBの6ドライブプレックスにミラーリングされています。同様に、RAIDグループ1にはサイトAの6ドライブプレックスが含まれており、サイトBの6ドライブプレックスにミラーリングされています。

SyncMirrorは通常、MetroClusterシステムにリモートミラーリングを提供するために使用され、各サイトにデータのコピーが1つずつ配置されます。場合によっては、1つのシステムで追加レベルの冗長性を提供するために使用されます。特に、シェルフレベルの冗長性を提供します。ドライブシェルフにはすでにデュアル電源装置とコントローラが搭載されており、全体的には板金をほとんど使用していませんが、場合によっては追加の保護が保証されることがあります。たとえば、あるNetAppのお客様は、自動車テストで使用するモバイルリアルタイム分析プラットフォームにSyncMirrorを導入しています。システムは、独立したUPSシステムからの独立した電源供給によって供給される2つの物理ラックに分割されました。

==チェックサム

チェックサムのトピックは、Oracle RMANのストリーミングバックアップをSnapshotベースのバックアップに移行することに慣れているDBAにとって特に関心があります。RMANの機能の1つは、バックアップ処理中に整合性チェックを実行することです。この機能には何らかの価値がありますが、その最大のメリットは、データベースが最新のストレージアレイで使用されていないことです。Oracleデータベースに物理ドライブが使用されている場合、ドライブの使用年数が経つと最終的にはほぼ確実に破損します。この問題は、真のストレージアレイではアレイベースのチェックサムによって解決されます。

実際のストレージアレイでは、複数のレベルでチェックサムを使用してデータの整合性が保護されます。IPベースのネットワークでデータが破損した場合、Transmission Control Protocol（TCP）レイヤはパケットデータを拒否し、再送信を要求します。FCプロトコルには、カプセル化されたSCSIデータと同様にチェックサムが含まれます。アレイに配置されたONTAPは、RAIDとチェックサムによる保護を備えています。破損は発生する可能性がありますが、ほとんどのエンタープライズアレイと同様に検出されて修正されます。通常、ドライブ全体に障害が発生してRAIDのリビルドが要求され、データベースの整合性は影響を受けません。ONTAPがチェックサムエラーを検出することもあります。これは、ドライブ上のデータが破損していることを意味します。ドライブが故障し、RAIDのリビルドが開始されます。繰り返しになりますが、データの整合性には影響はありません。

OracleのデータファイルとRedoログのアーキテクチャも、極度の状況下でも可能な限り最高レベルのデータ整合性を提供するように設計されています。最も基本的なレベルでは、Oracleのブロックにはチェックサムが含まれており、ほぼすべてのI/Oについて基本的な論理チェックが実行されます。Oracleがクラッシュしたり表領域がオフラインになったりしていない場合、データはそのまま維持されます。データ整合性チェックの程度は調整可能で、書き込みを確認するようにOracleを設定することもできます。その結果、クラッシュや障害のほぼすべてのシナリオをリカバリでき、非常にまれにリカバリ不能な状況が発生した場合は、破損がすぐに検出されます。

Oracleデータベースを使用しているNetAppのお客様のほとんどは'スナップショット・ベースのバックアップに移行すると'RMANなどのバックアップ製品の使用を中止しますRMANを使用してSnapCenterでブロックレベルのリカバリを実行できるオプションはまだあります。ただし、日常的には、RMAN、NetBackup、およびその他の製品は、月次または四半期ごとのアーカイブコピーの作成にのみ使用されます。

お客様の中には、 `dbv` 既存のデータベースの整合性チェックを定期的に実行します。NetAppでは、不必要なI/O負荷が発生するため、この方法は推奨されません。前述したように、データベースに以前に問題が発生していなかった場合、 `dbv` 問題の検出はほぼゼロです。このユーティリティは、ネットワークおよびストレージシステムに非常に高いシーケンシャルI/O負荷を生成します。Oracleの既知のバグにさらされるなど、破損が存在すると信じる理由がないかぎり、 `dbv`。

バックアップとリカバリの基本

Oracleデータベースと**Snapshot**ベースのバックアップ

ONTAPでのOracleデータベースのデータ保護の基盤となるのが、NetAppのSnapshotテクノロジです。

主な値は次のとおりです。

- *簡易性。*スナップショットは、特定の時点におけるデータのコンテナの内容の読み取り専用コピーです。
- 効率性。 Snapshotは作成時にスペースを必要としません。スペースが消費されるのは、データが変更されたときだけです。
- *管理性。*スナップショットをベースにしたバックアップ戦略は、ストレージOSに標準で組み込まれているため、構成と管理が容易です。ストレージシステムの電源がオンになっていれば、バックアップを作成できます。

- *拡張性。*ファイルとLUNの単一コンテナの最大1024個のバックアップを保持できます。複雑なデータセットの場合、データの複数のコンテナを、整合性のある単一のSnapshotセットで保護できます。
- ボリュームに1024個のSnapshotが含まれているかどうかに関係なく、パフォーマンスに影響はありません。

多くのストレージベンダーがSnapshotテクノロジを提供していますが、ONTAP内のSnapshotテクノロジは他に類を見ないものであり、エンタープライズアプリケーションやデータベース環境に次のような大きなメリットをもたらします。

- Snapshotコピーは、基盤となるWrite-Anywhere File Layout（WAFL）の一部です。アドオンや外部テクノロジではありません。これにより、ストレージシステムがバックアップシステムであるため、管理が簡易化されます。
- Snapshotコピーはパフォーマンスには影響しません。ただし、Snapshotに大量のデータが格納され、基盤となるストレージシステムがいっぱいになる場合など、一部のエッジケースを除きます。
- 「整合グループ」という用語は、整合性のあるデータの集合として管理されるストレージオブジェクトをグループ化したものを指す場合によく使用されます。特定のONTAPボリュームのSnapshotが整合グループのバックアップを構成します。

また、ONTAPスナップショットは、競合するテクノロジよりも拡張性に優れています。パフォーマンスに影響を与えることなく、5、50、500個のスナップショットを保存できます。ボリュームに現在許可されているSnapshotの最大数は1024です。Snapshotの保持期間を延長する必要がある場合は、Snapshotを追加のボリュームにカスケードするオプションがあります。

そのため、ONTAPでホストされているデータセットの保護はシンプルで拡張性に優れています。バックアップはデータの移動を必要としないため、ネットワーク転送速度、多数のテープドライブ、ディスクステージング領域の制限ではなく、ビジネスのニーズに合わせてバックアップ戦略を調整できます。

Snapshotはバックアップですか？

データ保護戦略としてSnapshotを使用する場合によく寄せられる質問の1つは、「実際の」データとSnapshotデータが同じドライブに配置されていることです。これらのドライブが失われると、プライマリデータとバックアップの両方が失われます。

これは有効な問題です。ローカルSnapshotは、日 々 のバックアップとリカバリのニーズに使用され、その点でSnapshotはバックアップです。NetApp環境のすべてのリカバリシナリオの99%近くが、最も厳しいRTO要件を満たすためにSnapshotを使用しています。

ただし、ローカルSnapshotが唯一のバックアップ戦略であるべきではありません。そのため、NetAppは、SnapMirrorやSnapVaultレプリケーションなどのテクノロジを提供し、独立したドライブセットにSnapshotを迅速かつ効率的にレプリケートします。スナップショットとスナップショットレプリケーションを使用して適切に設計された解決策では、テープの使用を最小限に抑えて四半期ごとのアーカイブを作成することも、完全に排除することもできます。

Snapshotベースのバックアップ

ONTAP Snapshotコピーを使用してデータを保護する方法は多数ありますが、Snapshotは、レプリケーション、ディザスタリカバリ、クローニングなど、ONTAPの他の多くの機能の基盤となります。Snapshotテクノロジの完全な概要については本ドキュメントでは説明しませんが、ここでは概要について説明します。

データセットのスナップショットを作成するには、主に次の2つの方法があります。

- crash-consistentバックアップ

- アプリケーションと整合性のあるバックアップ

データセットのcrash-consistentバックアップとは、ある時点におけるデータセット構造全体のキャプチャです。データセットが単一のNetApp FlexVolボリュームに格納されている場合は、Snapshotはいつでも作成できるため、このプロセスは簡単です。データセットが複数のボリュームにまたがっている場合は、整合性グループ（CG）Snapshotを作成する必要があります。CG Snapshotを作成するには、NetApp SnapCenterソフトウェア、ONTAPのネイティブ整合グループ機能、ユーザが管理するスクリプトなど、いくつかのオプションがあります。

crash-consistentバックアップは、主にpoint-of-the-backupリカバリで十分な場合に使用します。よりきめ細かなリカバリが必要な場合は、通常、アプリケーションと整合性のあるバックアップが必要です。

「application-consistent」の「consistent」という言葉は、しばしば誤った名義である。たとえば、Oracleデータベースをバックアップモードにすることをアプリケーション整合性バックアップと呼びますが、データの整合性が確保されたり休止されたりすることはありません。バックアップ中もデータは変化し続けます。一方、ほとんどのMySQLおよびMicrosoft SQL Serverのバックアップでは、バックアップを実行する前にデータが休止されます。VMwareは、特定のファイルの整合性を確保する場合としない場合があります。

整合グループ

「コンシステンシグループ」とは、ストレージアレイが複数のストレージリソースを単一のイメージとして管理できることを指します。たとえば、データベースが10個のLUNで構成されているとします。アレイは、これらの10個のLUNを一貫した方法でバックアップ、リストア、およびレプリケートできる必要があります。バックアップ時点でLUNのイメージに一貫性がなかった場合は、リストアを実行できません。これらの10個のLUNをレプリケートするには、すべてのレプリカが相互に完全に同期されている必要があります。

ONTAPのボリュームとアグリゲートのアーキテクチャでは、整合性は常に基本的な機能であるため、ONTAPについて説明する際に「整合グループ」という用語はあまり使用されません。他の多くのストレージアレイは、LUNまたはファイルシステムを個別のユニットとして管理します。その後、データ保護を目的とした「整合グループ」として設定することもできますが、これは追加の設定手順です。

ONTAPは、常に一貫性のあるローカルイメージとレプリケートされたデータイメージをキャプチャすることができました。ONTAPシステム上のさまざまなボリュームは、通常、正式には整合グループと呼ばれませんが、それが整合グループです。このボリュームのSnapshotは整合グループのイメージであり、そのSnapshotのリストアは整合グループのリストアです。SnapMirrorとSnapVaultはどちらも整合グループのレプリケーションを提供します。

整合性グループの**Snapshot**

整合グループSnapshot（cg-snapshots）は、ONTAPの基本的なSnapshotテクノロジを拡張したものです。標準のSnapshot処理では、1つのボリューム内のすべてのデータの整合性のあるイメージが作成されますが、複数のボリューム間、さらには複数のストレージシステム間で整合性のある一連のSnapshotを作成する必要がある場合があります。その結果、1つのボリュームのSnapshotと同じ方法で使用できる一連のSnapshotが作成されます。ローカルデータのリカバリに使用することも、ディザスタリカバリの目的でレプリケートすることも、単一の一貫したユニットとしてクローニングすることもできます。

cg-snapshotsの最大の用途は、12台のコントローラにまたがる約1PBのデータベース環境です。このシステムで作成されたcg-snapshotは、バックアップ、リカバリ、クローニングに使用されています。

ほとんどの場合、データセットが複数のボリュームにまたがっており、書き込み順序を維持する必要がある場合、選択した管理ソフトウェアによってcg-snapshotが自動的に使用されます。このような場合、cg-snapshotsの技術的な詳細を理解する必要はありません。ただし、複雑なデータ保護要件によっては、データ保護とレプリケーションのプロセスを詳細に管理しなければならない場合があります。ワークフローの自動化や、cg-snapshot APIの呼び出しにカスタムスクリプトを使用することもできます。最適なオプションとcg-

snapshotの役割を理解するには、テクノロジの詳細な説明が必要です。

一連のcg-snapshotsの作成は、次の2つの手順で行います。

1. すべてのターゲットボリュームで書き込みフェンシングを確立します。
2. フェンシングされた状態のボリュームのSnapshotを作成します。

書き込みフェンシングは順番に確立されます。つまり、フェンシングプロセスが複数のボリュームにまたがって設定されている間は、最初のボリュームで書き込みI/Oがフリーズされ、以降に表示されるボリュームにコミットされ続けます。これは、最初は書き込み順序を維持するための要件に違反しているように見えるかもしれませんが、環境ホストで非同期的に実行され、他の書き込みには依存しません。

たとえば、データベースでは大量の非同期データファイル更新が問題され、OSがI/Oの順序を変更して、独自のスケジューラ設定に従って完了できる場合があります。アプリケーションとオペレーティングシステムが書き込み順序を保持する要件をすでにリリースしているため、このタイプのI/Oの順序は保証できません。

カウンタの例として、ほとんどのデータベースロギングアクティビティは同期です。I/Oが確認応答され、書き込み順序を維持する必要があるまで、データベースはログへの以降の書き込みを続行しません。ログI/Oがフェンシングされたボリュームに到達した場合、そのことは確認されず、アプリケーションはそれ以降の書き込みをブロックします。同様に、ファイルシステムのメタデータI/Oは通常同期です。たとえば、ファイル削除処理が失われることはありません。xfsファイルシステムを使用するオペレーティングシステムがファイルを削除し、xfsファイルシステムのメタデータを更新して、フェンシングされたボリュームにあるファイルへの参照を削除するI/Oを実行すると、ファイルシステムのアクティビティが一時停止します。これにより、cg-snapshot処理中のファイルシステムの整合性が保証されます。

ターゲットボリューム間で書き込みフェンシングを設定すると、それらのボリュームでSnapshotを作成できるようになります。ボリュームの状態は従属書き込みの観点からフリーズされるため、Snapshotを正確に同時に作成する必要はありません。cg-snapshotを作成するアプリケーションの欠陥を防ぐために、初期の書き込みフェンシングには設定可能なタイムアウトが含まれています。このタイムアウトでは、ONTAPが自動的にフェンシングを解除し、定義された秒数後に書き込み処理を再開します。タイムアウト時間の経過前にすべてのSnapshotが作成された場合、作成される一連のSnapshotは有効な整合グループになります。

従属書き込み順序

技術的な観点から見ると、整合性グループの鍵となるのは、書き込み順序（特に従属書き込み順序）を維持することです。たとえば、10個のLUNに書き込むデータベースは、すべてのLUNに同時に書き込みます。多くの書き込みは非同期で発行されます。つまり、書き込みが完了する順序は重要ではなく、実際の書き込み順序はオペレーティングシステムやネットワークの動作によって異なります。

データベースが追加の書き込みを続行するには、一部の書き込み処理がディスク上に存在している必要があります。このような重要な書き込み処理は、依存書き込みと呼ばれます。以降の書き込みI/Oは、これらの書き込みがディスクに存在するかどうかに左右されます。これら10個のLUNのスナップショット、リカバリ、またはレプリケーションでは、従属書き込み順序が保証されていることを確認する必要があります。ファイルシステムの更新も、書き込み順序に依存した書き込みの例です。ファイルシステムの変更の順序を維持する必要があります。そうしないと、ファイルシステム全体が破損する可能性があります。

戦略

Snapshotベースのバックアップには、主に次の2つの方法があります。

- crash-consistentバックアップ
- Snapshotで保護されたホットバックアップ

データベースのcrash-consistentバックアップとは、データファイル、REDOログ、制御ファイルなど、データベース構造全体をある時点でキャプチャすることです。データベースが単一のNetApp FlexVolボリュームに格納されている場合は、Snapshotはいつでも作成できるため、このプロセスは簡単です。データベースが複数のボリュームにまたがっている場合は、整合性グループ（CG）Snapshotを作成する必要があります。CG Snapshotを作成するには、NetApp SnapCenterソフトウェア、ONTAPのネイティブ整合グループ機能、ユーザが管理するスクリプトなど、いくつかのオプションがあります。

crash-consistent Snapshotバックアップは、主にポイントオブザバックアップリカバリで十分な場合に使用されます。状況によってはアーカイブログを適用できますが、よりきめ細かなポイントインタイムリカバリが必要な場合は、オンラインバックアップを推奨します。

Snapshotベースのオンラインバックアップの基本的な手順は次のとおりです。

1. データベースを `backup` モード（Mode）：
2. データファイルをホストしているすべてのボリュームのSnapshotを作成します。
3. 終了します `backup` モード（Mode）：
4. コマンドを実行します `alter system archive log current` ログのアーカイブを強制的に実行します。
5. アーカイブログをホストするすべてのボリュームのSnapshotを作成します。

この手順により、バックアップモードのデータファイルと、バックアップモード中に生成された重要なアーカイブログを含む一連のSnapshotが作成されます。データベースのリカバリには、次の2つの要件があります。制御ファイルなどのファイルも便宜上保護する必要がありますが、絶対に必要なのはデータファイルとアーカイブログの保護だけです。

戦略はお客様によって大きく異なる可能性がありますが、これらの戦略のほとんどは、最終的には以下に概説されているのと同じ原則に基づいています。

Snapshotベースのリカバリ

Oracleデータベースのボリュームレイアウトを設計する際には、ボリュームベースNetApp SnapRestore（VBSR）テクノロジを使用するかどうかを最初に決定します。

ボリュームベースのSnapRestoreを使用すると、ボリュームをある時点の状態にほぼ瞬時にリバートできます。VBSRはボリューム上のすべてのデータがリバートされるため、すべてのユースケースに適しているとは限りません。たとえば、データファイル、Redoログ、アーカイブログを含むデータベース全体が1つのボリュームに格納されている場合、このボリュームをVBSRでリストアすると、新しいアーカイブログとRedoデータが破棄されるためデータが失われます。

リストアにVBSRは必要ありません。データベースの多くは、ファイルベースのSingle-File SnapRestore（SFSR）を使用するか、Snapshotからアクティブファイルシステムにファイルをコピーして戻すだけでリストアできます。

VBSRは、データベースが非常に大規模な場合やできるだけ迅速にリカバリする必要がある場合に推奨されます。また、VBSRを使用するにはデータファイルを分離する必要があります。NFS環境では、特定のデータベースのデータファイルを、他の種類のファイルの影響を受けない専用ボリュームに格納する必要があります。SAN環境では、データファイルを専用のFlexVolボリューム上の専用LUNに格納する必要があります。ボリュームマネージャを使用する場合は（Oracle Automatic Storage Management[ASM]を含む）、ディスクグループもデータファイル専用にする必要があります。

この方法でデータファイルを分離すると、他のファイルシステムに影響を与えることなく、データファイルを

以前の状態にリバートできます。

Snapshot リザーブ

SAN環境内のOracleデータを含むボリュームごとに、 `percent-snapshot-space` LUN環境でSnapshot用にスペースをリザーブしても役に立たないため、ゼロに設定する必要があります。フラクショナルリザーブを100に設定すると、LUNを含むボリュームのSnapshotでは、すべてのデータの書き替えを100%吸収するために、Snapshotリザーブを除くボリューム内に十分な空きスペースが必要になります。フラクショナルリザーブの値を小さい値に設定すると、それに応じて必要な空きスペースは少なくなりますが、Snapshotリザーブは常に除外されます。これは、LUN環境のスナップショット予約スペースが無駄になることを意味します。

NFS環境には2つのオプションがあります。

- を設定します `percent-snapshot-space` 予想されるSnapshotスペース消費量に基づきます。
- を設定します `percent-snapshot-space` アクティブなスペース使用量とSnapshotスペース使用量をまとめてゼロにして管理できます。

最初のオプションでは、 `percent-snapshot-space` は、ゼロ以外の値（通常は約20%）に設定されます。このスペースはユーザーには表示されません。ただし、この値によって利用率が制限されるわけではありません。リザーブが20%のデータベースで30%の入れ替えが発生した場合、スナップショット領域は20%リザーブの範囲を超えて拡張され、リザーブされていないスペースを占有する可能性があります。

リザーブを20%などの値に設定する主な利点は、一部のスペースが常にスナップショットに使用可能であることを確認することです。たとえば、1TBのボリュームに20%のリザーブが設定されている場合、データベース管理者（DBA）が格納できるのは800GBのデータのみです。この構成では、Snapshot用に少なくとも200GBのスペースが保証されます。

いつ `percent-snapshot-space` がゼロに設定されている場合、ボリューム内のすべてのスペースをエンドユーザが使用できるため、可視性が向上します。データベース管理者は、Snapshotを利用する1TBのボリュームが表示された場合、この1TBのスペースはアクティブデータとSnapshotの書き替えの間で共有されることを理解しておく必要があります。

エンドユーザ間では、オプション1とオプション2の間に明確な優先順位はありません。

ONTAPとサードパーティのスナップショット

Oracle Doc ID 604683.1には、サードパーティ製スナップショットのサポート要件と、バックアップおよびリストア処理に使用できる複数のオプションが説明されています。

サードパーティベンダーは、会社のスナップショットが次の要件に準拠していることを保証する必要があります。

- スナップショットは、Oracleが推奨するリストアおよびリカバリ処理と統合する必要があります。
- スナップショットは、スナップショットの時点でデータベースクラッシュ整合性がある必要があります。
- スナップショット内のファイルごとに書き込み順序が保持されます。

ONTAPおよびNetAppのOracle管理製品は、これらの要件に準拠しています。

SnapRestoreによるOracleデータベースの高速リカバリ

NetApp SnapRestoreテクノロジは、SnapshotからのONTAPでのデータの高速リストア

を実現します。

重要なデータセットが使用できないと、重要なビジネス処理が停止します。テープが破損する可能性があり、ディスク・ベースのバックアップからリストアする場合でも、ネットワーク上での転送に時間がかかることがあります。SnapRestoreでは、データセットをほぼ瞬時にリストアできるため、このような問題を回避できます。ペタバイト規模のデータベースでも、わずか数分で完全にリストアできます。

SnapRestoreには、ファイル/ LUNベースとボリュームベースの2つの形式があります。

- 個 々 のファイルやLUNは、2TBのLUNでも4KBのファイルでも、数秒でリストアできます。
- ファイルやLUNのコンテナは、10GBでも100TBのデータでも、数秒でリストアできます。

「ファイルまたはLUNのコンテナ」とは、通常はFlexVolボリュームを指します。たとえば、1つのボリューム内に1つのLVMディスクグループを構成する10個のLUNを配置したり、1つのボリュームに1、000ユーザのNFSホームディレクトリを格納したりできます。個 々 のファイルまたはLUNに対してリストア処理を実行する代わりに、ボリューム全体を単一の処理としてリストアできます。このプロセスは、FlexGroupやONTAP整合グループなど、複数のボリュームを含むスケールアウトコンテナとも連携します。

SnapRestoreがこれほど迅速かつ効率的に機能するのは、Snapshotの性質によるものです。Snapshotは本質的には、特定の時点におけるボリュームの内容を読み取り専用で並行して表示する機能です。アクティブブロックは変更可能な実際のブロックですが、Snapshotは、Snapshot作成時のファイルおよびLUNを構成するブロックの状態を読み取り専用で表示します。

ONTAPでは、スナップショットデータへの読み取り専用アクセスのみが許可されますが、SnapRestoreを使用してデータを再アクティブ化できます。スナップショットはデータの読み取り/書き込みビューとして再度有効になり、データは以前の状態に戻ります。SnapRestoreは、ボリュームレベルまたはファイルレベルで動作できます。この技術は基本的に同じで、動作に若干の違いがあります。

ボリューム**SnapRestore**

ボリュームベースのSnapRestoreは、データのボリューム全体を以前の状態に戻します。この処理ではデータの移動は必要ありません。つまり、API処理やCLI処理の処理には数秒かかることがありますが、リストアプロセスは基本的に瞬時に完了します。1GBのデータをリストアするのは、1PBのデータをリストアするのと同じくらい複雑で時間のかかる作業ではありません。この機能は、多くの企業のお客様がONTAPストレージシステムに移行する主な理由です。大規模なデータセットでも数秒でRTOを達成できます。

ボリュームベースSnapRestoreの欠点の1つは、ボリューム内の変更が時間の経過とともに累積されることが原因です。したがって、各Snapshotとアクティブなファイルデータは、その時点までの変更に依存します。ボリュームを以前の状態にリバートすると、データに対する以降の変更がすべて破棄されます。ただし、これには以降に作成されたスナップショットが含まれることはあまり明白ではありません。これは必ずしも望ましいとは限りません。

たとえば、データ保持のSLAで夜間バックアップを30日間指定するとします。ボリュームSnapRestoreを使用して5日前に作成されたSnapshotにデータセットをリストアすると、過去5日間に作成されたSnapshotがすべて破棄され、SLAに違反します。

この制限に対処するために、いくつかのオプションが用意されています。

1. ボリューム全体のSnapRestoreを実行するのではなく、以前のSnapshotからデータをコピーできます。この方法は、データセットが小さい場合に最も適しています。
2. Snapshotはリストアではなくクローニングできます。このアプローチの制限事項は、ソーススナップショットがクローンの依存関係であることです。したがって、クローンも削除されるか、独立したボリューム

にスプリットされないかぎり、削除することはできません。

3. ファイルベースのSnapRestoreの使用。

File SnapRestore

ファイルベースのSnapRestoreは、Snapshotベースのより詳細なリストアプロセスです。ボリューム全体の状態をリバートする代わりに、個 々 のファイルまたはLUNの状態がリバートされます。スナップショットを削除する必要はありません。また、この操作によって以前のスナップショットへの依存関係が作成されることもありません。ファイルまたはLUNがアクティブボリュームですぐに使用可能になります。

ファイルまたはLUNのSnapRestoreリストア中にデータを移動する必要はありません。ただし、ファイルまたはLUNの基盤となるブロックがSnapshotとアクティブボリュームの両方に存在するようになったことを反映するには、一部の内部メタデータの更新が必要になります。パフォーマンスへの影響はありませんが、この処理が完了するまでSnapshotの作成はブロックされます。処理速度は約5GBps（18TB/時）です。これは、リストアするファイルの合計サイズに基づきます。

Oracleデータベースのオンラインバックアップ

バックアップモードでOracleデータベースを保護およびリカバリするには、2セットのデータが必要です。これはOracleの唯一のバックアップ・オプションではなく'最も一般的なバックアップ・オプションであることに注意してください

- バックアップモードでのデータファイルのSnapshot
- データファイルがバックアップモードのときに作成されたアーカイブログ

コミットされたすべてのトランザクションを含む完全なリカバリが必要な場合は、3つ目の項目が必要です。

- 最新のREDOログのセット

オンラインバックアップのリカバリを促進する方法はいくつかあります。多くのお客様は、ONTAP CLIを使用してSnapshotをリストアし、次にOracle RMANまたはsqlplusを使用してリカバリを完了します。これは、データベースをリストアする可能性と頻度が非常に低く、すべてのリストア手順が熟練したデータベース管理者によって処理される大規模な本番環境では特に顕著です。完全な自動化を実現するために、NetApp SnapCenterなどのソリューションには、コマンドラインインターフェイスとグラフィカルインターフェイスの両方を備えたOracleプラグインが含まれています。

一部の大規模なお客様では、スケジュールされたSnapshotに備えて特定の時間にデータベースをバックアップモードにするように、ホストで基本的なスクリプトを設定することで、よりシンプルなアプローチを採用しています。たとえば、次のコマンドをスケジュールします。 `alter database begin backup` 23時58分、`alter database end backup` 00:02に実行し、午前0時にストレージシステム上でSnapshotの直接スケジュールを設定します。その結果、外部のソフトウェアやライセンスを必要としない、シンプルで拡張性に優れたバックアップ戦略が実現します。

データレイアウト

最もシンプルなレイアウトは、データファイルを1つ以上の専用ボリュームに分離する方法です。これらのファイルは、他のファイルタイプによって汚染されていない必要があります。これは、重要なREDOログ、制御ファイル、またはアーカイブログを削除することなく、SnapRestore処理によってデータファイルボリュームを迅速にリストアできるようにするためです。

SANでは、専用ボリューム内でのデータファイルの分離についても同様の要件があります。Microsoft Windowsなどのオペレーティングシステムでは、1つのボリュームに複数のデータファイルLUNが含まれ、そ

れぞれにNTFSファイルシステムが配置される場合があります。他のオペレーティング・システムでは'通常'論理ボリューム・マネージャが使用されますたとえば、Oracle ASMでは、ASMディスクグループのLUNを1つのボリュームに限定し、1つのボリュームとしてバックアップおよびリストアできるようにするのが最も簡単なオプションです。パフォーマンスまたは容量管理のために追加のボリュームが必要な場合は、新しいボリュームに追加のディスクグループを作成すると、管理が簡単になります。

これらのガイドラインに従うと、整合性グループSnapshotを実行する必要なく、ストレージシステム上で直接Snapshotをスケジュールできます。これは、Oracleのバックアップではデータファイルを同時にバックアップする必要がないためです。オンラインバックアップ手順は、データファイルが数時間にわたってテープにゆっくりとストリーミングされても、継続的に更新されるように設計されています。

ASMディスクグループを複数のボリュームに分散して使用すると、複雑な状況が発生します。このような場合は、cg-snapshotを実行して、すべてのコンスティチュエントボリュームでASMメタデータの整合性を確保する必要があります。

注意： ASMが `spfile` および `passwd` データファイルをホストしているディスクグループにファイルがありません。これにより、データファイルのみを選択してリストアすることができなくなります。

ローカルリカバリ手順—NFS

この手順は、手動で実行することも、SnapCenterなどのアプリケーションを使用して実行することもできます。基本的な手順は次のとおりです。

1. データベースをシャットダウンします。
2. 目的のリストアポイントの直前に、データファイルボリュームをSnapshotにリカバリします。
3. アーカイブログを目的のポイントまで再生します。
4. 完全なリカバリが必要な場合は、現在のREDOログを再生します。

この手順では、目的のアーカイブログがアクティブファイルシステムにまだ存在していることを前提としています。サポートされていない場合は、アーカイブログをリストアする必要があります。リストアされていない場合は、RMAN / sqlplusをsnapshotディレクトリ内のデータに転送できます。

また、小規模なデータベースの場合は、エンドユーザがデータファイルを `.snapshot` 自動化ツールやストレージ管理者の支援がないディレクトリで、 `snaprestore` コマンドを実行します

ローカルリカバリ手順—SAN

この手順は、手動で実行することも、SnapCenterなどのアプリケーションを使用して実行することもできます。基本的な手順は次のとおりです。

1. データベースをシャットダウンします。
2. データファイルをホストしているディスクグループを休止します。手順は、選択した論理ボリュームマネージャによって異なります。ASMでは、このプロセスでディスクグループをディスマウントする必要があります。Linuxでは、ファイルシステムをディスマウントし、論理ボリュームとボリュームグループを非アクティブ化する必要があります。目的は、リストア対象のターゲットボリュームグループに対するすべての更新を停止することです。
3. 目的のリストアポイントの直前に、データファイルディスクグループをSnapshotにリストアします。
4. 新しくリストアしたディスクグループを再アクティブ化します。
5. アーカイブログを目的のポイントまで再生します。

6. 完全なリカバリが必要な場合は、すべてのREDOログを再生します。

この手順では、目的のアーカイブログがアクティブファイルシステムにまだ存在していることを前提としています。サポートされていない場合は、アーカイブログLUNをオフラインにしてリストアを実行し、アーカイブログをリストアする必要があります。この例では、アーカイブログを専用ボリュームに分割すると便利です。アーカイブログがRedoログとボリュームグループを共有している場合は、LUNのセット全体をリストアする前にRedoログを他の場所にコピーする必要があります。この手順により、最終的に記録されたトランザクションの損失を防ぐことができます。

Oracle Databaseストレージの**Snapshot**による最適化されたバックアップ

Oracle 12cがリリースされた時点では、データベースをホットバックアップモードにする必要がないため、Snapshotベースのバックアップとリカバリはさらにシンプルになりました。そのため、Snapshotベースのバックアップをストレージシステム上で直接スケジュール設定しても、完全なリカバリやポイントインタイムリカバリを引き続き実行できます。

データベース管理者にとってはホットバックアップリカバリの手順の方がなじみがありますが、データベースがホットバックアップモードのときに作成されなかったSnapshotを使用することは以前から可能でした。Oracle 10gおよび11gでは、データベースの整合性を維持するために、リカバリ時に手動で追加の手順を実行する必要がありました。Oracle 12cでは、 `sqlplus` および `rman` ホットバックアップモードではないデータファイルバックアップでアーカイブログを再生するための追加ロジックが含まれています。

前述したように、スナップショットベースのホットバックアップをリカバリするには、次の2セットのデータが必要です。

- バックアップモードで作成されたデータファイルのSnapshot
- データファイルがホットバックアップモードのときに生成されたアーカイブログ

リカバリ中、データベースはデータファイルからメタデータを読み取り、リカバリに必要なアーカイブログを選択します。

ストレージSnapshotを最適化したリカバリでは、同じ結果を達成するために必要なデータセットがわずかに異なります。

- データファイルのSnapshot、およびSnapshotが作成された時刻を識別する方法
- 最新のデータファイルチェックポイントの時刻からSnapshotの正確な時刻までのログをアーカイブします。

リカバリ中、データベースはデータファイルからメタデータを読み取り、必要な最も古いアーカイブログを特定します。フルリカバリまたはポイントインタイムリカバリを実行できます。ポイントインタイムリカバリを実行する場合は、データファイルのSnapshotの時刻を把握することが重要です。指定したリカバリポイントは、Snapshotの作成時刻以降である必要があります。NetAppでは、クロックの変動を考慮して、スナップショット時間に少なくとも数分を追加することを推奨しています。

詳細については、Oracle 12cの各種ドキュメントで「Recovery Using Storage Snapshot Optimization」のトピックを参照してください。また、Oracleサードパーティ製スナップショットのサポートについては、OracleのドキュメントID Doc ID 604683.1を参照してください。

データレイアウト

最も簡単なレイアウトは、データファイルを1つ以上の専用ボリュームに分離する方法です。これらのファイルは、他のファイルタイプによって汚染されていない必要があります。これは、重要なREDOログ、制御ファイル、またはアーカイブログを削除することなく、SnapRestore処理でデータファイルボリュームを迅速にリストアできるようにするためです。

SANでは、専用ボリューム内でのデータファイルの分離についても同様の要件があります。Microsoft Windowsなどのオペレーティングシステムでは、1つのボリュームに複数のデータファイルLUNが含まれ、それぞれにNTFSファイルシステムが配置される場合があります。他のオペレーティング・システムでは'通常'論理ボリューム・マネージャも使用されますたとえば、Oracle ASMでは、ディスクグループを1つのボリュームに限定し、1つのボリュームとしてバックアップおよびリストアできるようにするのが最も簡単なオプションです。パフォーマンスまたは容量管理のために追加のボリュームが必要な場合は、新しいボリュームに追加のディスクグループを作成すると、管理が容易になります。

これらのガイドラインに従うと、整合性グループSnapshotを実行することなく、ONTAPで直接Snapshotをスケジュールできます。これは、Snapshotで最適化されたバックアップでは、データファイルを同時にバックアップする必要がないためです。

ASMディスクグループが複数のボリュームに分散されている場合は、複雑な問題が発生します。このような場合は、cg-snapshotを実行して、すべてのコンスティチュエントボリュームでASMメタデータの整合性を確保する必要があります。

[注] ASM spfileファイルとpasswdファイルが、データファイルをホストしているディスクグループにないことを確認します。これにより、データファイルのみを選択してリストアすることができなくなります。

ローカルリカバリ手順**—NFS**

この手順は、手動で実行することも、SnapCenterなどのアプリケーションを使用して実行することもできます。基本的な手順は次のとおりです。

1. データベースをシャットダウンします。
2. 目的のリストアポイントの直前に、データファイルボリュームをSnapshotにリカバリします。
3. アーカイブログを目的のポイントまで再生します。

この手順では、目的のアーカイブログがアクティブファイルシステムにまだ存在していることを前提としています。サポートされていない場合は、アーカイブログをリストアする必要があります。または、 `rman` または `sqlplus` のデータに転送できます。 `.snapshot` ディレクトリ。

また、小規模なデータベースの場合は、エンドユーザがデータファイルを `.snapshot` SnapRestoreコマンドを実行するための自動化ツールやストレージ管理者の支援がないディレクトリ。

ローカルリカバリ手順**—SAN**

この手順は、手動で実行することも、SnapCenterなどのアプリケーションを使用して実行することもできます。基本的な手順は次のとおりです。

1. データベースをシャットダウンします。
2. データファイルをホストしているディスクグループを休止します。手順は、選択した論理ボリュームマネージャによって異なります。ASMでは、このプロセスでディスクグループをディスマウントする必要があります。Linuxでは、ファイルシステムをディスマウントし、論理ボリュームとボリュームグループを非アクティブ化する必要があります。目的は、リストア対象のターゲットボリュームグループに対するすべ

ての更新を停止することです。

3. 目的のリストアポイントの直前に、データファイルディスクグループをSnapshotにリストアします。
4. 新しくリストアしたディスクグループを再アクティブ化します。
5. アーカイブログを目的のポイントまで再生します。

この手順では、目的のアーカイブログがアクティブファイルシステムにまだ存在していることを前提としています。サポートされていない場合は、アーカイブログLUNをオフラインにしてリストアを実行し、アーカイブログをリストアする必要があります。この例では、アーカイブログを専用ボリュームに分割すると便利です。アーカイブログがRedoログとボリュームグループを共有している場合は、記録された最終的なトランザクションが失われないように、LUNセット全体のリストア前にRedoログを別の場所にコピーする必要があります。

フルリカバリの例

データファイルが破損または破壊されており、完全なリカバリが必要であると仮定します。そのための手順は次のとおりです。

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover automatic;
Media recovery complete.
SQL> alter database open;
Database altered.
SQL>
```

ポイントインタイムリカバリの例

リカバリ手順全体は1つのコマンドで実行できます。 `recover automatic`。

ポイントインタイムリカバリが必要な場合は、Snapshotのタイムスタンプがわかっている必要があり、次のように特定できます。

```
Cluster01::> snapshot show -vserver vserver1 -volume NTAP_oradata -fields
create-time
vserver   volume        snapshot   create-time
--------  ------------  ---------  -------------------------
vserver1  NTAP_oradata  my-backup  Thu Mar 09 10:10:06 2017
```

Snapshotの作成時間は3月9日と10：10：06と表示されます。安全のために、Snapshotの時刻に1分が追加されます。

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover database until time '09-MAR-2017 10:44:15' snapshot time '09-
MAR-2017 10:11:00';
```

リカバリが開始されました。スナップショット時間は記録された時間の1分後の10：11：00、目標復旧時間は10：44と指定されています。次に、sqlplusは目的のリカバリ時間（10：44）に到達するために必要なアーカイブログを要求します。

```
ORA-00279: change 551760 generated at 03/09/2017 05:06:07 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_31_930813377.dbf
ORA-00280: change 551760 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 552566 generated at 03/09/2017 05:08:09 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_32_930813377.dbf
ORA-00280: change 552566 for thread 1 is in sequence #32
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 553045 generated at 03/09/2017 05:10:12 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_33_930813377.dbf
ORA-00280: change 553045 for thread 1 is in sequence #33
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 753229 generated at 03/09/2017 05:15:58 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_34_930813377.dbf
ORA-00280: change 753229 for thread 1 is in sequence #34
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
Log applied.
Media recovery complete.
SQL> alter database open resetlogs;
Database altered.
SQL>
```

Snapshotを使用してデータベースを完全にリカバリするには、 `recover automatic` コマンドには特定のライセンスは不要ですが、を使用してポイントインタイムリカバリを実行できます。 `snapshot time` Oracle Advanced Compressionのライセンスが必要です。

Oracleデータベースの管理と自動化のためのツール

Oracleデータベース環境におけるONTAPの主な価値は、瞬時のSnapshotコピー、シンプルなSnapMirrorレプリケーション、効率的なFlexCloneボリュームの作成など、ONTAPのコアテクノロジにあります。

これらのコア機能をONTAPに直接簡単に設定して要件を満たす場合もありますが、より複雑なニーズにはオーケストレーションレイヤが必要です。

SnapCenter

SnapCenterは、NetAppの主力データ保護製品です。データベースバックアップの実行方法という点ではSnapManager製品に似ていますが、NetAppストレージシステム上のデータ保護管理を単一コンソールで管理できるように一から構築されています。

SnapCenterには、Snapshotベースのバックアップとリストア、SnapMirrorとSnapVaultのレプリケーションな

ど、大企業の大規模な運用に必要な基本機能が含まれています。これらの高度な機能には、拡張されたロールベースアクセス制御（RBAC）機能、サードパーティのオーケストレーション製品と統合するためのRESTful API、データベースホスト上のSnapCenterプラグインの無停止での一元管理、クラウド規模環境向けに設計されたユーザインターフェイスなどがあります。

REST

ONTAPには、豊富なRESTful APIセットも含まれています。これにより、サードパーティベンダーは、ONTAPとの緊密な統合により、データ保護やその他の管理アプリケーションを作成できます。さらに、独自の自動化ワークフローやユーティリティを作成したいお客様も、RESTful APIを簡単に利用できます。

Oracleのディザスタリカバリ

ONTAPによるOracleデータベースのディザスタリカバリ

ディザスタリカバリとは、火災によってストレージシステムやサイト全体が破壊されるなど、重大な災害が発生した場合にデータサービスをリストアすることです。

> このドキュメントは、以前に公開されたテクニカルレポート_TR-4591：『Oracle Data Protection_and_TR-4592：Oracle on MetroCluster』を差し替えます。_

ディザスタリカバリは、もちろんSnapMirrorを使用してデータを単純にレプリケーションすることで実現できます。多くのお客様は、ミラーされたレプリカを1時間に何度も更新します。

ほとんどのお客様にとって、DRに必要なのはデータのリモートコピーだけではなく、そのデータを迅速に利用できることです。NetAppは、このニーズに対応する2つのテクノロジを提供します。MetroClusterとSnapMirrorのアクティブ同期です。

MetroClusterとは、低レベルの同期ミラーリングストレージと多数の追加機能を含むハードウェア構成のONTAPのことです。MetroClusterなどの統合ソリューションは、今日の複雑なスケールアウトデータベース、アプリケーション、仮想化インフラストラクチャを簡素化します。複数の外部データ保護製品や戦略を、1つのシンプルな中央集中型ストレージアレイに置き換えます。また、単一のクラスタストレージシステム内に、バックアップ、リカバリ、ディザスタリカバリ、高可用性（HA）が統合されています。

SnapMirrorアクティブ同期はSnapMirror Synchronousに基づいています。MetroClusterでは、各ONTAPコントローラがドライブデータをリモートサイトにレプリケートします。SnapMirrorアクティブ同期を使用すると、基本的には2つの異なるONTAPシステムでLUNデータの独立したコピーを維持しながら、このLUNの単一インスタンスを提供できます。ホストの観点からは、単一のLUNエンティティです。

SnapMirrorアクティブ同期とMetroClusterの内部的な動作は大きく異なりますが、ホストにとってはほぼ同じ結果になります。主な違いは粒度です。同期レプリケートするワークロードのみを選択する場合は、SnapMirrorアクティブ同期が適しています。環境全体やデータセンターをレプリケートする必要がある場合は、MetroClusterをお勧めします。また、SnapMirrorアクティブ同期は現在SAN専用ですが、MetroClusterはSAN、NFS、SMBなどのマルチプロトコルです。

MetroCluster

MetroCluster物理アーキテクチャとOracleデータベース

MetroCluster環境でのOracleデータベースの動作を理解するには、MetroClusterシステム

の物理設計についてある程度の説明が必要です。

このドキュメントは、以前に公開されていたテクニカルレポート（TR-4592：『Oracle on MetroCluster』）に代わるものです。_

MetroClusterは3種類の構成で使用できます。

- IPセツソクノHAヘア
- FCセツソクノHAヘア
- シングルコントローラ、FC接続

[注]「接続」という用語は、サイト間レプリケーションに使用されるクラスタ接続を指します。ホストプロトコルを指しているわけではありません。MetroCluster構成では、クラスタ間通信に使用される接続の種類に関係なく、すべてのホスト側プロトコルが通常どおりサポートされます。

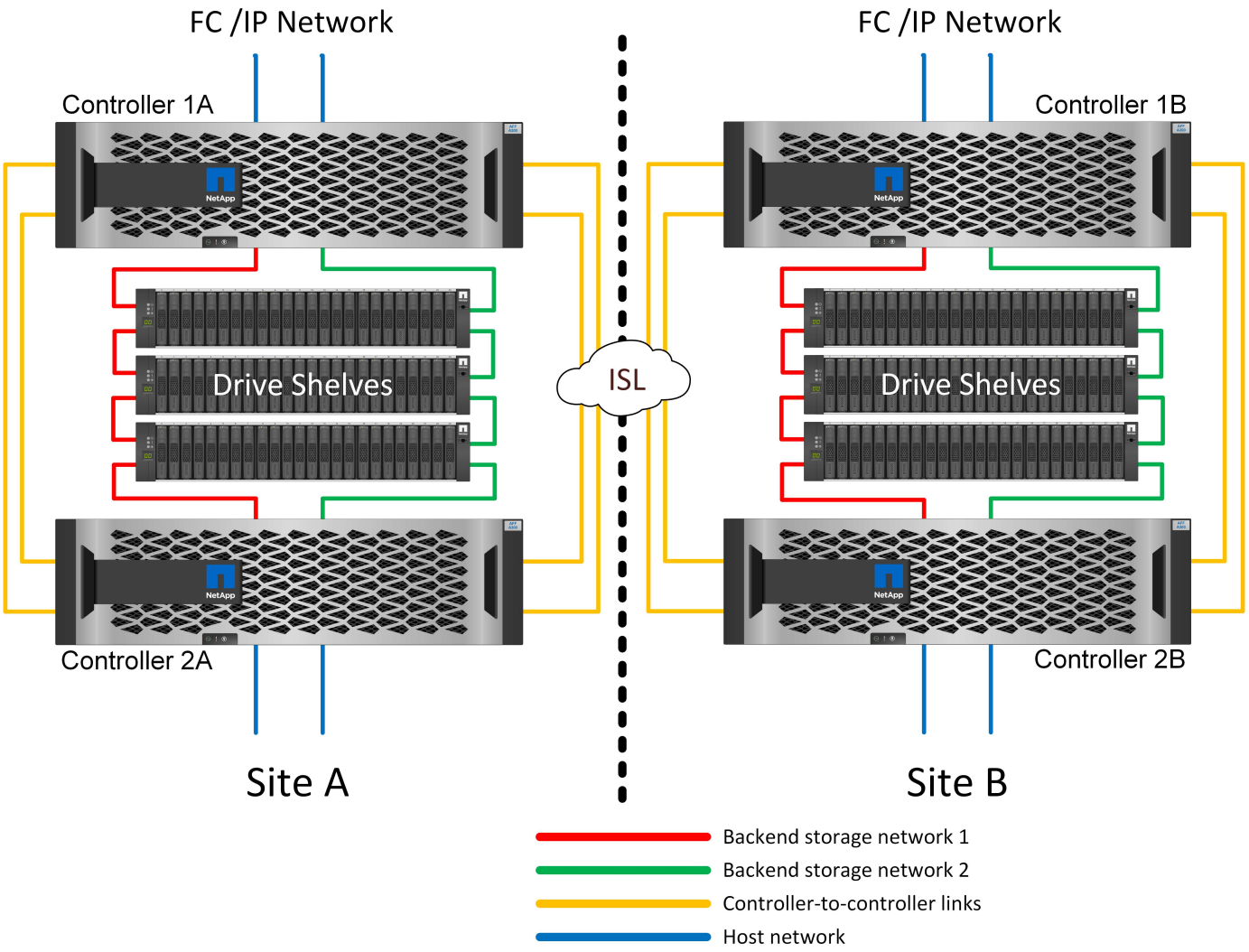
MetroCluster IP の略

HAペアMetroCluster IP構成では、サイトごとに2ノードまたは4ノードを使用します。この設定オプションを使用すると、2ノードオプションに比べて複雑さとコストが増加しますが、サイト内の冗長性という重要なメリットがあります。単純なコントローラ障害では、WAN経由のデータアクセスは必要ありません。データアクセスは、代替ローカルコントローラを介してローカルのままです。

ほとんどのお客様は、インフラストラクチャの要件がシンプルであるため、IP接続を選択しています。これまでは、ダークファイバやFCスイッチを使用した場合、サイト間での高速接続のプロビジョニングは一般的に容易でしたが、今日では、高速で低レイテンシのIP回線がより容易に利用可能になっています。

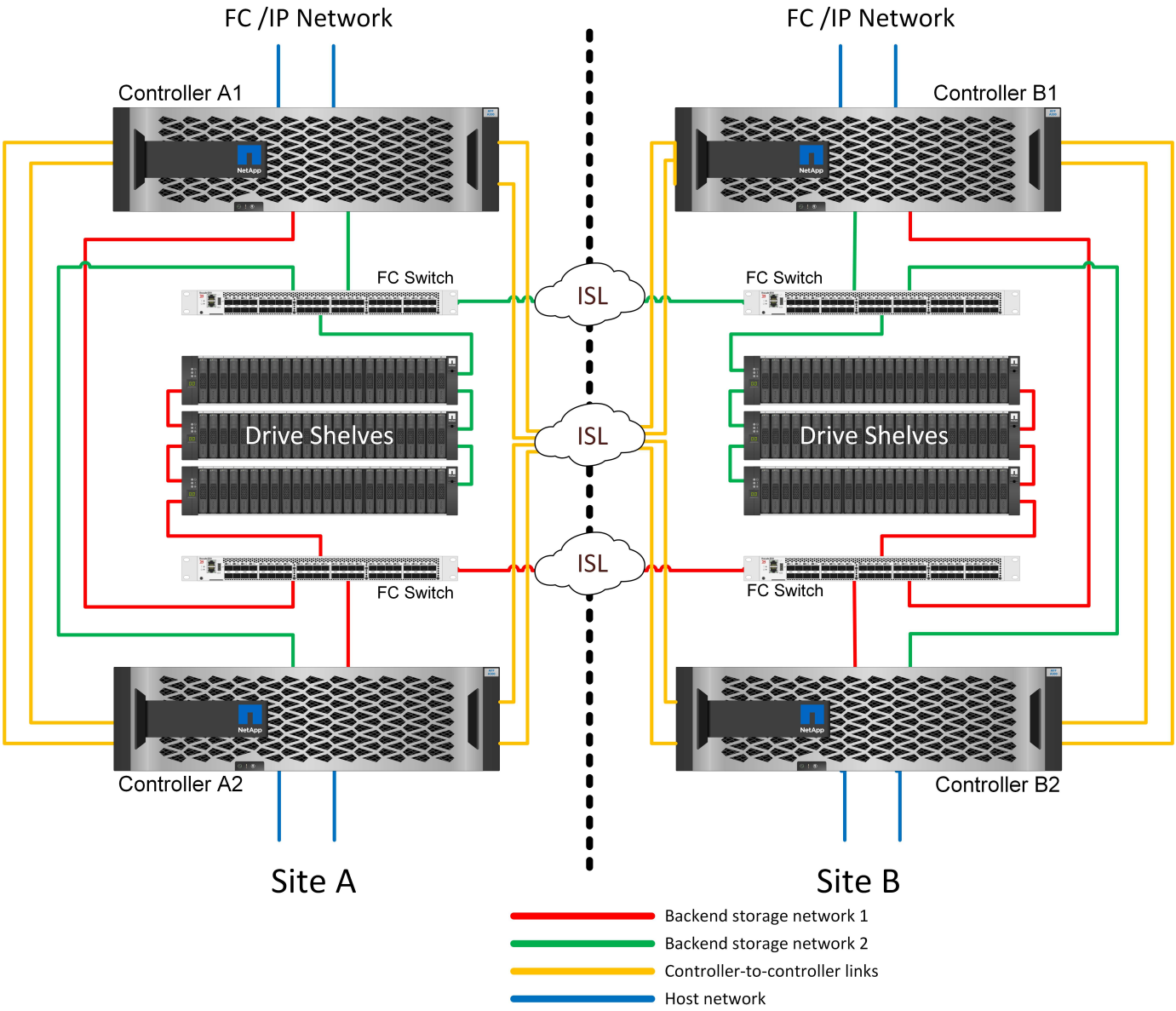
サイト間接続はコントローラのみであるため、アーキテクチャもシンプルです。FC SAN接続MetroClusterでは、コントローラが反対側サイトのドライブに直接書き込むため、追加のSAN接続、スイッチ、およびブリッジが必要になります。一方、IP構成のコントローラは、コントローラを介して反対側のドライブに書き込みます。

追加情報については、ONTAPの公式ドキュメントを参照してください。 "MetroCluster IP 解決策のアーキテクチャと設計"。



HAペアFC SAN接続MetroCluster

HAペアMetroCluster FC構成では、サイトごとに2ノードまたは4ノードを使用します。この設定オプションを使用すると、2ノードオプションに比べて複雑さとコストが増加しますが、サイト内の冗長性という重要なメリットがあります。単純なコントローラ障害では、WAN経由のデータアクセスは必要ありません。データアクセスは、代替ローカルコントローラを介してローカルのままです。

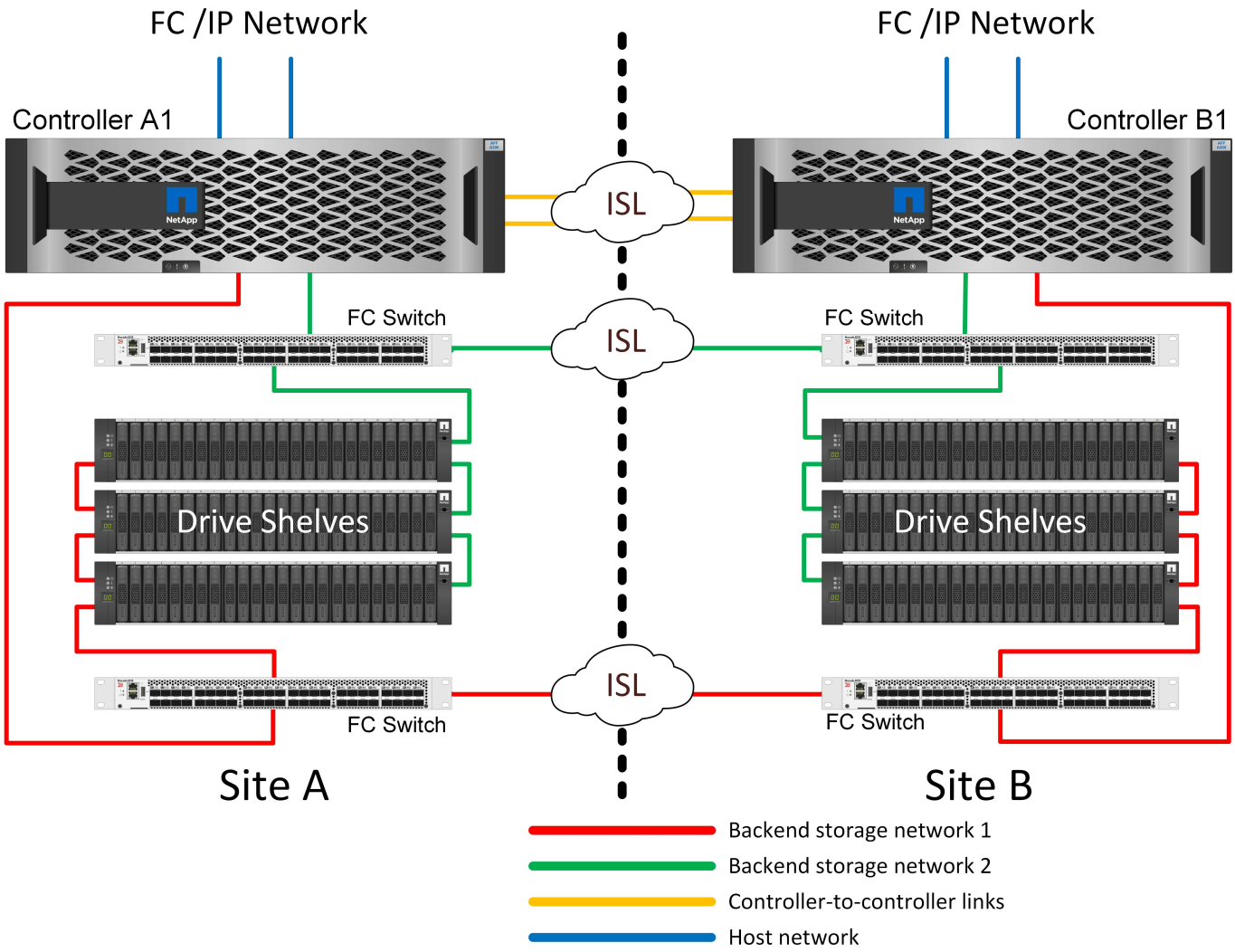


一部のマルチサイトインフラは、アクティブ/アクティブ運用向けに設計されたものではなく、プライマリサイトやディザスタリカバリサイトとして使用されます。この場合、一般にHAペアMetroClusterオプションが推奨される理由は次のとおりです。

- 2ノードMetroClusterクラスタはHAシステムですが、コントローラに予期しない障害が発生した場合や計画的メンテナンスを行う場合は、反対側のサイトでデータサービスをオンラインにする必要があります。サイト間のネットワーク接続が必要な帯域幅をサポートできない場合は、パフォーマンスが低下します。唯一の選択肢は、さまざまなホストOSと関連サービスを代替サイトにフェイルオーバーすることです。HAペアMetroClusterクラスタでは、コントローラが停止すると同じサイト内で単純なフェイルオーバーが発生するため、この問題は解消されます。
- 一部のネットワークトポロジは、サイト間アクセス用に設計されていませんが、異なるサブネットまたは分離されたFC SANを使用します。この場合、代替コントローラが反対側のサイトのサーバにデータを提供できないため、2ノードMetroClusterクラスタはHAシステムとして機能しなくなります。完全な冗長性を実現するには、HAペアMetroClusterオプションが必要です。
- 2サイトインフラを単一の高可用性インフラとみなす場合は、2ノードMetroCluster構成が適しています。ただし、サイト障害後もシステムが長時間機能しなければならない場合は、HAペアが推奨されます。HAペアは、単一サイト内でHAを提供し続けるためです。

2ノードFC SAN接続MetroCluster

2ノードMetroCluster構成では、サイトごとに1つのノードのみが使用されます。設定とメンテナンスが必要なコンポーネントが少ないため、HAペアオプションよりもシンプルな設計になっています。また、ケーブル配線やFCスイッチの点でインフラストラクチャの必要性も軽減されています。最後に、コストを削減します。



この設計の明らかな影響は、1つのサイトでコントローラに障害が発生した場合、反対側のサイトからデータを利用できることです。この制限は必ずしも問題ではありません。多くの企業は、本質的に単一のインフラとして機能する、拡張された高速で低レイテンシのネットワークを使用したマルチサイトデータセンター運用を行っています。このような場合は、2ノードバージョンのMetroClusterが推奨されます。2ノードシステムは現在、複数のサービスプロバイダでペタバイト規模で使用されています。

MetroClusterの耐障害性機能

MetroCluster 解決策 には単一点障害（Single Point of Failure）はありません。

- 各コントローラに、ローカルサイトのドライブシェルフへの独立したパスが2つあります。
- 各コントローラに、リモートサイトのドライブシェルフへの独立したパスが2つあります。
- 各コントローラには、反対側のサイトのコントローラへの独立したパスが2つあります。
- HAペア構成では、各コントローラからローカルパートナーへのパスが2つあります。

つまり、構成内のコンポーネントを1つでも削除しても、MetroClusterのデータ提供機能を損なうことはありません。2つのオプションの耐障害性の違いは、サイト障害後もHAペアバージョンが全体的なHAストレージシステムになる点だけです。

MetroCluster論理アーキテクチャと**Oracle**データベース

MetroCluster環境でOracleデータベースがどのように動作するかを理解するAlsopでは、MetroClusterシステムの論理機能について説明する必要があります。

サイト障害からの保護：**NVRAM**と**MetroCluster**

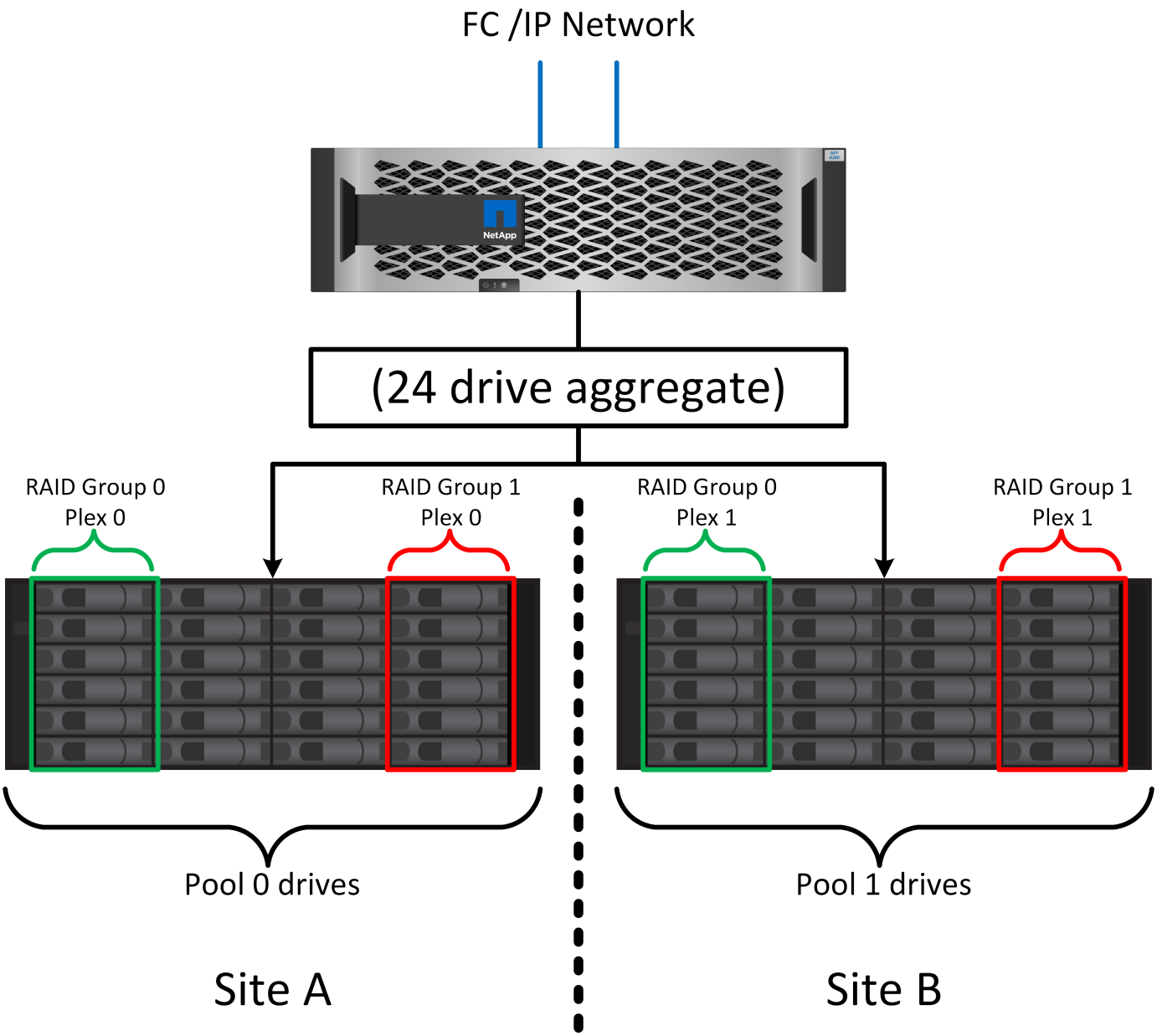
MetroClusterは、次の方法でNVRAMデータ保護を拡張します。

- 2ノード構成では、NVRAMデータがスイッチ間リンク（ISL）を使用してリモートパートナーにレプリケートされます。
- HAペア構成では、NVRAMデータがローカルパートナーとリモートパートナーの両方にレプリケートされます。
- 書き込みは、すべてのパートナーにレプリケートされるまで確認応答されません。このアーキテクチャは、NVRAMデータをリモートパートナーにレプリケートすることで、転送中のI/Oをサイト障害から保護します。このプロセスは、ドライブレベルのデータレプリケーションには関係ありません。アグリゲートを所有するコントローラは、アグリゲート内の両方のプレックスに書き込むことでデータレプリケーションを実行しますが、サイトが失われた場合でも転送中のI/Oの損失からデータを保護する必要があります。レプリケートされたNVRAMデータは、障害が発生したコントローラをパートナーコントローラがテイクオーバーする必要がある場合にのみ使用されます。

サイトおよびシェルフ障害からの保護：**SyncMirror**とプレックス

SyncMirrorは、RAID DPやRAID-TECを強化するミラーリングテクノロジですが、これに代わるものではありません。2つの独立したRAIDグループの内容をミラーリングします。論理構成は次のとおりです。

1. ドライブは、場所に基づいて2つのプールに構成されます。1つのプールはサイトAのすべてのドライブで構成され、2つ目のプールはサイトBのすべてのドライブで構成されます。
2. 次に、アグリゲートと呼ばれる共通のストレージプールが、RAIDグループのミラーセットに基づいて作成されます。各サイトから同じ数のドライブが引き出されます。たとえば、20ドライブのSyncMirrorアグリゲートは、サイトAの10本のドライブとサイトBの10本のドライブで構成されます。
3. サイト上の各ドライブセットは、ミラーリングを使用せずに、完全に冗長化された1つ以上のRAID DPグループまたはRAID-TECグループとして自動的に構成されます。ミラーリングの下でRAIDを使用することで、サイトが失われた場合でもデータを保護できます。



上の図は、SyncMirror構成の例を示しています。24ドライブのアグリゲートをコントローラに作成しました。このアグリゲートは、サイトAで割り当てられたシェルフの12本のドライブと、サイトBで割り当てられたシェルフの12本のドライブで構成されています。ドライブは2つのミラーRAIDグループにグループ化されました。RAIDグループ0には、サイトAの6ドライブのプレックスが含まれており、サイトBの6ドライブのプレックスにミラーリングされています。同様に、RAIDグループ1にはサイトAの6ドライブのプレックスが含まれており、サイトBの6ドライブのプレックスにミラーリングされています。

SyncMirrorは通常、MetroClusterシステムにリモートミラーリングを提供するために使用され、各サイトにデータのコピーが1つずつ配置されます。場合によっては、1つのシステムで追加レベルの冗長性を提供するために使用されます。特に、シェルフレベルの冗長性を提供します。ドライブシェルフにはすでにデュアル電源装置とコントローラが搭載されており、全体的には板金をほとんど使用していませんが、場合によっては追加の保護が保証されることがあります。たとえば、あるNetAppのお客様は、自動車テストで使用するモバイルリアルタイム分析プラットフォームにSyncMirrorを導入しています。システムは、独立した電源供給と独立したUPSシステムを備えた2つの物理ラックに分かれていました。

前述したように、書き込みの確認応答は、少なくとも1台の他のコントローラでローカルのNVRAMとNVRAMに記録されるまで返されません。このアプローチにより、ハードウェア障害や停電が発生しても、転送中のI/Oが失われることはありません。ローカルのNVRAMに障害が発生したり、他のノードへの接続に障害が発生したりすると、データはミラーリングされなくなります。

ローカルNVRAMからエラーが報告されると、ノードはシャットダウンします。このシャットダウンにより、HAペアが使用されている場合はパートナーコントローラにフェイルオーバーされます。MetroClusterでは、動作は選択した全体的な設定によって異なりますが、リモートノートに自動的にフェイルオーバーされる場合があります。いずれの場合も、障害が発生したコントローラが書き込み処理を認識していないため、データは失われません。

リモートノードへのNVRAMレプリケーションがブロックされるサイト間接続障害は、より複雑な状況です。書き込みがリモートノードにレプリケートされなくなるため、コントローラで重大なエラーが発生した場合にデータが失われる可能性があります。さらに重要なことは、このような状況で別のノードにフェイルオーバーしようとするとデータが失われることです。

制御要素は、NVRAMが同期されているかどうかです。NVRAMが同期されていれば、ノード間のフェイルオーバーを安全に実行でき、データ損失のリスクはありません。MetroCluster構成では、NVRAMと基盤となるアグリゲートのプレックスが同期されていれば、データ損失のリスクなしにスイッチオーバーを実行できます。

データが同期されていない場合、ONTAPは、フェイルオーバーまたはスイッチオーバーを強制的に実行しないかぎり、フェイルオーバーまたはスイッチオーバーを許可しません。この方法で条件を変更すると、元のコントローラにデータが残っている可能性があり、データ損失が許容されることが確認されます。

データベースやその他のアプリケーションは、ディスク上のデータのより大きな内部キャッシュを保持するため、フェイルオーバーやスイッチオーバーを強制的に実行した場合に特に破損の影響を受けやすくなります。強制的なフェイルオーバーまたはスイッチオーバーが発生した場合、以前に確認済みの変更は事実上破棄されます。ストレージアレイの内容は実質的に時間を逆方向にジャンプし、キャッシュの状態はディスク上のデータの状態を反映しなくなります。

この状況を回避するために、ONTAPでは、NVRAMの障害に対する特別な保護をボリュームに設定できます。この保護メカニズムがトリガーされると、ボリュームがNVFAILという状態になります。この状態になると、原因アプリケーションがクラッシュするI/Oエラーが発生します。このクラッシュにより、古いデータを使用しないようにアプリケーションがシャットダウンされます。コミットされたトランザクションデータがログに含まれている必要があるため、データが失われないようにしてください。次の手順では、管理者がホストを完全にシャットダウンしてから、LUNとボリュームを手動で再度オンラインに戻します。これらの手順にはいくつかの作業が含まれる可能性がありますが、このアプローチはデータの整合性を確保するための最も安全な方法です。すべてのデータがこの保護を必要とするわけではありません。そのため、NVFAILの動作はボリューム単位で設定できます。

HAペアと**MetroCluster**

MetroClusterには、2ノードとHAペアの2つの構成があります。2ノード構成の動作は、NVRAMに関してはHAペアと同じです。突然の障害が発生した場合、パートナーノードはNVRAMデータを再生してドライブの整合性を確保し、確認済みの書き込みが失われていないことを確認できます。

HAペア構成では、ローカルパートナーノードにもNVRAMがレプリケートされます。MetroClusterを使用しないスタンドアロンHAペアの場合と同様に、単純なコントローラ障害ではパートナーノードでNVRAMが再生されます。サイト全体が突然失われた場合、リモートサイトには、ドライブの整合性を確保してデータの提供を開始するために必要なNVRAMも用意されています。

MetroClusterの重要な側面の1つは、通常の運用状態ではリモートノードがパートナーデータにアクセスできないことです。各サイトは本質的に、反対のサイトのパーソナリティを想定できる独立したシステムとして機能します。このプロセスはスイッチオーバーと呼ばれ、計画的スイッチオーバーでは、サイトの処理が無停止で反対側のサイトに移行されます。また、サイトが失われ、ディザスタリカバリの一環として手動または自動のスイッチオーバーが必要になる計画外の状況も含まれます。

スイッチオーバーとスイッチバック

スイッチオーバーとスイッチバックという用語は、MetroCluster構成のリモートコントローラ間でボリュームを移行するプロセスを指します。このプロセスでは、リモートノードのみが環境されます。4ボリューム構成でMetroClusterを使用する場合のローカルノードのフェイルオーバーは、前述したテイクオーバーとギブバックのプロセスと同じです。

計画的スイッチオーバーとスイッチバック

計画的スイッチオーバーまたはスイッチバックは、ノード間のテイクオーバーやギブバックと似ています。このプロセスには複数の手順があり、数分かかるように見える場合もありますが、実際には、ストレージリソースとネットワークリソースを複数のフェーズで正常に移行します。完全なコマンドの実行に必要な時間よりもはるかに短時間で制御転送が行われる瞬間。

テイクオーバー/ギブバックとスイッチオーバー/スイッチバックの主な違いは、FC SAN接続への影響です。ローカルのテイクオーバー/ギブバックでは、ローカルノードへのFCパスがすべて失われ、ホストのネイティブMPIOを使用して使用可能な代替パスに切り替えます。ポートは再配置されません。スイッチオーバーとスイッチバックでは、コントローラの仮想FCターゲットポートがもう一方のサイトに移行します。一時的にSAN上に存在しなくなり、代わりのコントローラに再表示されます。

SyncMirrorタイムアウト

SyncMirrorは、シェルフ障害から保護するONTAPのミラーリングテクノロジです。シェルフが離れた場所に配置されている場合は、リモートデータ保護が実現します。

SyncMirrorは汎用同期ミラーリングを提供しません。その結果、可用性が向上します。一部のストレージシステムでは、一定のオールオアナッシングミラーリング（Dominoモードと呼ばれることもあります）を使用します。リモートサイトへの接続が失われるとすべての書き込みアクティビティが停止する必要があるため、この形式のミラーリングはアプリケーションで制限されます。そうしないと、書き込みは一方のサイトに存在し、もう一方のサイトには存在しません。通常、このような環境では、サイト間の接続が短時間（30秒など）以上切断された場合にLUNがオフラインになるように構成されます。

この動作は、一部の環境に適しています。ただし、ほとんどのアプリケーションには、通常の動作条件下で保証された同期レプリケーションを提供しながら、レプリケーションを一時停止できる解決策が必要です。サイト間の接続が完全に失われると、多くの場合、災害が近い状況とみなされます。通常、このような環境は、接続が修復されるか、データを保護するために環境をシャットダウンする正式な決定が下されるまで、オンラインのままでデータを提供します。リモートレプリケーションの障害のみが原因でアプリケーションを自動的にシャットダウンする必要があるのは珍しいことです。

SyncMirrorは、タイムアウトの柔軟性を備えた同期ミラーリングの要件に対応しています。リモートコントローラやプレックスへの接続が失われると、30秒のタイマーがカウントダウンを開始します。カウンタが0に達すると、ローカルデータを使用して書き込みI/O処理が再開されます。データのリモートコピーは使用可能ですが、接続が回復するまで時間内に凍結されます。再同期では、アグリゲートレベルのSnapshotを使用してシステムをできるだけ迅速に同期モードに戻します。

特に、多くの場合、この種の汎用的なオールオアナッシングDominoモードレプリケーションは、アプリケーションレイヤでより適切に実装されています。たとえば、Oracle DataGuardには最大保護モードが用意され

ており、どのような状況でも長時間のインスタンスレプリケーションが保証されます。設定可能なタイムアウトを超えてレプリケーションリンクに障害が発生すると、データベースはシャットダウンします。

ファブリック接続**MetroCluster**による自動無人スイッチオーバー

Automatic Unattended Switchover（AUSO；自動無人スイッチオーバー）は、クロスサイトHAの形式を提供するファブリック接続MetroClusterの機能です。前述したように、MetroClusterには2つのタイプ（各サイトに1台のコントローラを配置する場合と、各サイトに1台のHAペアを配置する場合）があります。HAオプションの主な利点は、コントローラの計画的シャットダウンと計画外シャットダウンのどちらでもすべてのI/Oをローカルで処理できることです。シングルノードオプションのメリットは、コスト、複雑さ、インフラの削減です。

AUSOの主な価値は、ファブリック接続MetroClusterシステムのHA機能を向上させることです。各サイトが反対側のサイトの健常性を監視し、データを提供するノードがなくなると、AUSOによって迅速なスイッチオーバーが実行されます。このアプローチは、可用性の点でHAペアに近い構成になるため、サイトごとにノードが1つだけのMetroCluster構成で特に役立ちます。

AUSOでは、HAペアレベルで包括的な監視を行うことはできません。HAペアには、ノード間の直接通信用の2本の冗長な物理ケーブルが含まれているため、きわめて高い可用性を実現できます。さらに、HAペアの両方のノードが冗長ループ上の同じディスクセットにアクセスできるため、1つのノードが別のノードの健常性を監視するための別のルートが提供されます。

MetroClusterクラスタは複数のサイトにまたがって存在し、ノード間の通信とディスクアクセスの両方がサイト間ネットワーク接続に依存します。クラスタの残りの部分のハートビートを監視する機能には制限があります。AUSOは、ネットワークの問題が原因で、もう一方のサイトが使用できない状況ではなく、実際にダウンしている状況を区別する必要があります。

その結果、HAペアのコントローラで、システムパニックなどの特定の理由で発生したコントローラ障害が検出された場合、テイクオーバーが要求されることがあります。また、接続が完全に失われた場合（ハートビートの損失とも呼ばれます）、テイクオーバーを促すこともあります。

MetroClusterシステムで自動スイッチオーバーを安全に実行できるのは、元のサイトで特定の障害が検出された場合のみです。また、ストレージシステムの所有権を取得するコントローラは、ディスクとNVRAMのデータが同期されていることを保証できる必要があります。コントローラは、ソースサイトとの通信が失われて稼働している可能性があるため、スイッチオーバーの安全性を保証できません。スイッチオーバーを自動化するためのその他のオプションについては、次のセクションのMetroCluster Tiebreaker（MCTB）解決策に関する情報を参照してください。

ファブリック接続**MetroCluster**を使用した**MetroCluster Tiebreaker**

。 "NetApp MetroCluster Tiebreaker" ソフトウェアを第3のサイトで実行して、MetroCluster環境の健全性を監視し、通知を送信し、必要に応じて災害時にスイッチオーバーを強制的に実行できます。Tiebreakerの完全な概要は、 "NetApp Support Site"ただし、MetroCluster Tiebreakerの主な目的はサイトの損失を検出することです。また、サイトの損失と接続の損失を区別する必要があります。たとえば、Tiebreakerがプライマリサイトに到達できなかったためにスイッチオーバーが発生しないようにします。そのため、Tiebreakerはリモートサイトがプライマリサイトに接続する能力も監視します。

AUSOによる自動スイッチオーバーもMCTBと互換性があります。AUSOは、特定の障害イベントを検出し、NVRAMとSyncMirrorのプレックスが同期されている場合にのみスイッチオーバーを実行するように設計されているため、非常に迅速に対応します。

一方、Tiebreakerはリモートに配置されているため、サイトの停止を宣言する前にタイマーが経過するのを待つ必要があります。Tiebreakerは最終的にAUSOの対象となるコントローラ障害を検出しますが、一般的にはAUSOがスイッチオーバーを開始しており、Tiebreakerが機能する前にスイッチオーバーを完了している可

能性があります。Tiebreakerから送信される2つ目のswitchoverコマンドは拒否されます。

注意： MCTBソフトウェアは、強制的なスイッチオーバー時にNVRAMが同期されていること、またはプレックスが同期されていることを確認しません。メンテナンス作業中に自動スイッチオーバーが設定されている場合は無効にして、NVRAMまたはSyncMirrorプレックスの同期が失われるようにしてください。

また、MCTBは、次の一連のイベントにつながるローリングディザスタに対応できない場合があります。

1. サイト間の接続が30秒以上中断されます。
2. SyncMirrorレプリケーションがタイムアウトし、プライマリサイトで処理が続行されるため、リモートレプリカは古くなります。
3. プライマリサイトが失われます。その結果、プライマリサイトにレプリケートされていない変更が存在します。その場合、次のようないくつかの理由でスイッチオーバーが望ましくない可能性があります。
   - 重要なデータはプライマリサイトに存在し、最終的にリカバリ可能になる可能性があります。スイッチオーバーによってアプリケーションの動作が継続されると、重要なデータは実質的に破棄されます。
   - サバイバーサイトのアプリケーションで、サイト障害時にプライマリサイトのストレージリソースを使用していた場合、データがキャッシュされている可能性があります。スイッチオーバーでは、キャッシュと一致しない古いバージョンのデータが生成されます。
   - サバイバーサイトのオペレーティングシステムで、サイト障害時にプライマリサイトのストレージリソースを使用していた場合、キャッシュデータがある可能性があります。スイッチオーバーでは、キャッシュと一致しない古いバージョンのデータが生成されます。最も安全な方法は、Tiebreakerがサイト障害を検出した場合にアラートを送信するように設定し、スイッチオーバーを強制的に実行するかどうかを決定することです。キャッシュされたデータを消去するには、アプリケーションやオペレーティングシステムのシャットダウンが必要になる場合があります。さらに、NVFAIL設定を使用して保護を強化し、フェイルオーバープロセスを合理化することもできます。

MetroCluster IPを使用した**ONTAP**メディエーター

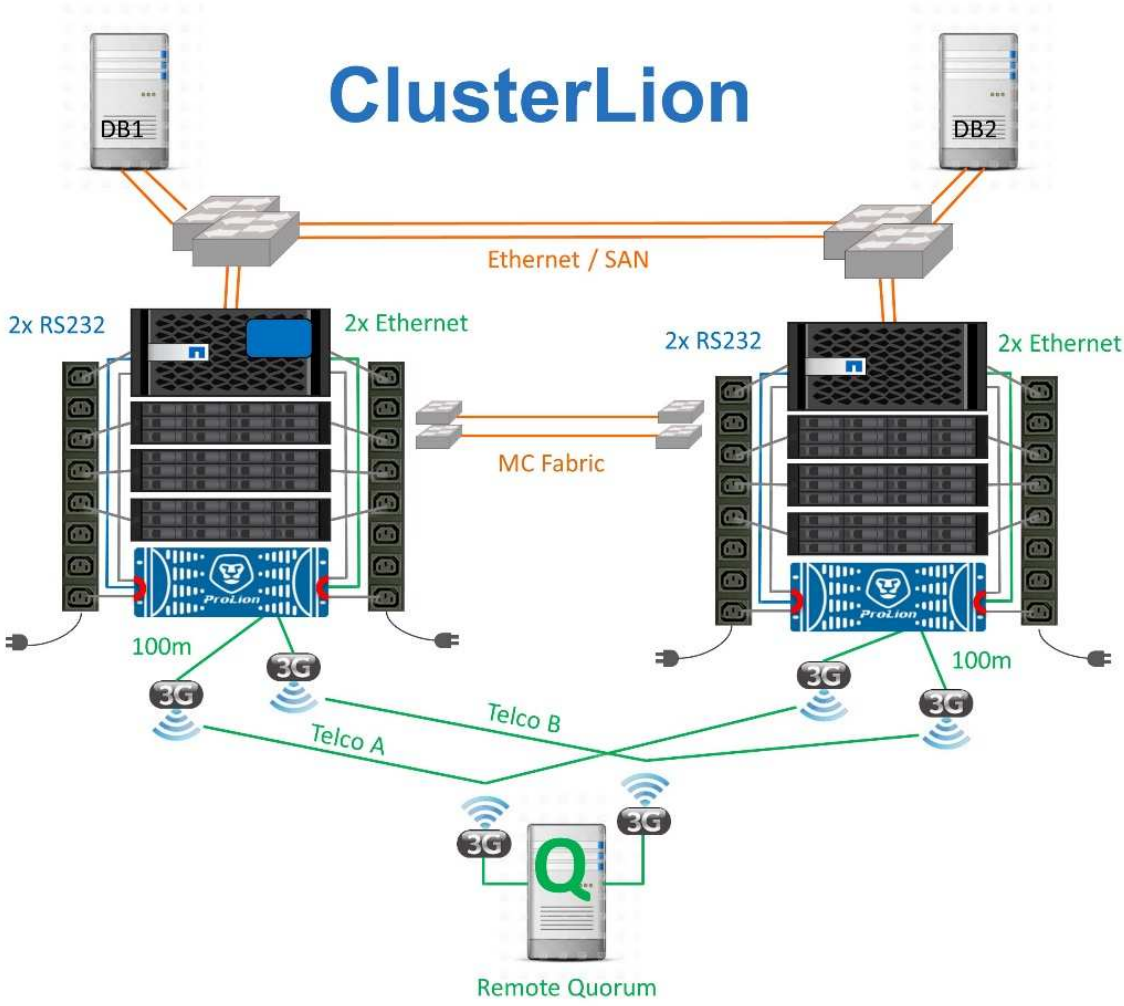
ONTAPメディエーターは、MetroCluster IPおよびその他の特定のONTAPソリューションで使用されます。これは、前述のMetroCluster Tiebreakerソフトウェアと同様に従来のTiebreakerサービスとして機能しますが、自動無人スイッチオーバーの実行という重要な機能も備えています。

ファブリック接続MetroClusterは、反対側のサイトのストレージデバイスに直接アクセスできます。これにより、一方のMetroClusterコントローラがドライブからハートビートデータを読み取ることで、他のコントローラの健常性を監視できます。これにより、一方のコントローラがもう一方のコントローラの障害を認識し、スイッチオーバーを実行できるようになります。

一方、MetroCluster IPアーキテクチャでは、すべてのI/Oがコントローラとコントローラの接続を介して排他的にルーティングされるため、リモートサイトのストレージデバイスに直接アクセスすることはありません。これにより、コントローラで障害を検出してスイッチオーバーを実行する機能が制限されます。そのため、サイトの損失を検出して自動的にスイッチオーバーを実行するためには、ONTAPメディエーターがTiebreakerデバイスとして必要になります。

ClusterLionを使用した**3**番目の仮想サイト

ClusterLionは、仮想の第3サイトとして機能する高度なMetroCluster監視アプライアンスです。このアプローチにより、完全に自動化されたスイッチオーバー機能により、MetroClusterを2サイト構成で安全に導入できます。さらに、ClusterLionでは、追加のネットワークレベル監視を実行し、スイッチオーバー後の処理を実行できます。完全なドキュメントはProLionから入手できます。



- ClusterLionアプライアンスは、直接接続されたイーサネットケーブルとシリアルケーブルでコントローラの健常性を監視します。
- 2つのアプライアンスは、冗長3Gワイヤレス接続で相互に接続されています。
- ONTAPコントローラへの電源は、内部リレーを介して配線されます。サイト障害が発生すると、内部UPSシステムを搭載したClusterLionによって電源接続が切断されてからスイッチオーバーが実行されます。このプロセスにより、スプリットブレイン状態が発生しないようにします。
- ClusterLionは、30秒のSyncMirrorタイムアウト内にスイッチオーバーを実行するか、まったく実行しません。
- ClusterLionでは、NVRAMプレックスとSyncMirrorプレックスの状態が同期されていないかぎり、スイッチオーバーは実行されません。
- ClusterLionでは、MetroClusterが完全に同期されている場合にのみスイッチオーバーが実行されるため、NVFAILは必要ありません。この構成では、計画外スイッチオーバーが発生しても、拡張Oracle RACなどのサイトスパニング環境をオンラインのまま維持できます。
- ファブリック接続MetroClusterとMetroCluster IPの両方をサポート

Oracleデータベースと**SyncMirror**

MetroClusterシステムを使用したOracleデータ保護の基盤となるのは、最大パフォーマンスのスケールアウト同期ミラーリングテクノロジであるSyncMirrorです。

SyncMirrorによるデータ保護

最も簡単な意味では、同期レプリケーションとは、変更がミラーされたストレージの両側に対して確認応答される前に行われなければならないことを意味します。たとえば、データベースがログを書き込んでいる場合やVMwareゲストにパッチを適用している場合は、書き込みが失われることはありません。プロトコルレベルでは、両方のサイトの不揮発性メディアにコミットされるまで、ストレージシステムは書き込みを確認応答しないでください。その場合にのみ、データ損失のリスクなしに作業を安全に進めることができます。

同期レプリケーションテクノロジの使用は、同期レプリケーション解決策を設計および管理するための最初のステップです。最も重要な考慮事項は、計画的および計画外のさまざまな障害シナリオで何が発生するかを理解することです。すべての同期レプリケーションソリューションが同じ機能を提供するわけではありません。Recovery Point Objective（RPO；目標復旧時点）がゼロ（つまりデータ損失ゼロ）の解決策が必要な場合は、すべての障害シナリオを考慮する必要があります。特に、サイト間の接続が失われてレプリケーションが不可能になった場合、どのような結果が予想されますか。

SyncMirrorデータの可用性

MetroClusterレプリケーションは、同期モードに効率的に切り替えられるように設計されたNetApp SyncMirrorテクノロジに基づいています。この機能は、同期レプリケーションを必要とする一方で、データサービスに高可用性も必要とするお客様の要件を満たします。たとえば、リモートサイトへの接続が切断されている場合は、通常、ストレージシステムをレプリケートされていない状態で運用し続けることを推奨します。

多くの同期レプリケーションソリューションは、同期モードでしか動作できません。このタイプのall-or-nothingレプリケーションは、Dominoモードと呼ばれることがあります。このようなストレージシステムでは、データのローカルコピーとリモートコピーが非同期になるのではなく、データの提供が停止します。レプリケーションが強制的に解除された場合、再同期には非常に時間がかかり、ミラーリングの再確立中にデータが完全に失われる可能性があります。

リモートサイトに到達できない場合にSyncMirrorを同期モードからシームレスに切り替えることができるだけでなく、接続がリストアされたときにRPO=0状態に迅速に再同期することもできます。再同期中にリモートサイトにある古いデータコピーを使用可能な状態で保持することもできるため、データのローカルコピーとリモートコピーを常に維持できます。

Dominoモードが必要な場合、NetAppはSnapMirror Synchronous（SM-S）を提供します。Oracle DataGuardや、ホスト側のディスクミラーリングのタイムアウト延長など、アプリケーションレベルのオプションもあります。追加情報とオプションについては、担当のNetAppまたはパートナーアカウントチームにお問い合わせください。

MetroClusterによるOracleデータベースのフェイルオーバー

```
 Metrocluster is an ONTAP feature that can protect your Oracle databases
with RPO=0 synchronous mirroring across sites, and it scales up to support
hundreds of databases on a single MetroCluster system. It's also simple to
use. The use of MetroCluster does not necessarily add to or change any
best practices for operating a enterprise applications and databases.
```

通常のベストプラクティスも引き続き適用され、必要なデータ保護がRPO=0の場合はMetroClusterで対応します。しかし、ほとんどのお客様は、RPO=0のデータ保護だけでなく、災害時のRTOを向上させ、サイトメンテナンス作業の一環として透過的なフェイルオーバーを実現するためにMetroClusterを使用しています。

事前設定された**OS**を使用したフェイルオーバー

SyncMirrorはディザスタリカバリサイトにデータの同期コピーを提供しますが、そのデータを利用できるようにするには、オペレーティングシステムと関連するアプリケーションが必要です。基本的な自動化により、環境全体のフェイルオーバー時間を大幅に短縮できます。Oracle RAC、Veritas Cluster Server（VCS）、VMware HAなどのClusterware製品は、サイト間でクラスタを作成するためによく使用されます。多くの場合、フェイルオーバープロセスは単純なスクリプトで実行できます。

プライマリノードが失われた場合、代替サイトでアプリケーションをオンラインにするようにクラスタウェア（またはスクリプト）が設定されます。1つは、アプリケーションを構成するNFSリソースまたはSANリソース用に事前設定されたスタンバイサーバを作成する方法です。プライマリサイトに障害が発生すると、クラスタウェアまたはスクリプト化された代替サイトが次のような一連の処理を実行します。

1. MetroClusterスイッチオーバーの強制実行
2. FC LUNの検出の実行（SANのみ）
3. ファイルシステムのマウント
4. アプリケーションの起動

このアプローチの主な要件は、リモートサイトでOSを実行することです。アプリケーションバイナリを使用して事前に設定する必要があります。つまり、パッチ適用などのタスクをプライマリサイトとスタンバイサイトで実行する必要があります。また、災害が発生した場合は、アプリケーションバイナリをリモートサイトにミラーリングしてマウントすることもできます。

実際のアクティベーション手順は簡単です。LUN検出などのコマンドでは、FCポートあたりのコマンド数が少なくて済みます。ファイル・システムのマウントは' `mount` コマンドを実行し、データベースとASMの両方をCLIで1つのコマンドで起動および停止できます。スイッチオーバーの前にディザスタリカバリサイトでボリュームとファイルシステムを使用していない場合は、 `dr-force- nvfail` ボリューム：

仮想**OS**によるフェイルオーバー

データベース環境のフェイルオーバーを拡張して、オペレーティングシステム自体を含めることができます。理論的には、このフェイルオーバーはブートLUNで実行できますが、ほとんどの場合、仮想OSで実行されます。手順の手順は次のようになります。

1. MetroClusterスイッチオーバーの強制実行
2. データベースサーバ仮想マシンをホストするデータストアのマウント
3. 仮想マシンの起動
4. データベースを手動で起動するか、またはデータベースを自動的に起動するように仮想マシンを設定する

たとえば、ESXクラスタが複数のサイトにまたがっているとします。災害が発生した場合は、スイッチオーバー後にディザスタリカバリサイトで仮想マシンをオンラインにすることができます。災害発生時に仮想データベースサーバをホストするデータストアが使用されていないかぎり、 `dr-force- nvfail` 関連付けられているボリューム。

Oracleデータベース、**MetroCluster**、**NVFAIL**

NVFailはONTAPの一般的なデータ整合性機能で、データベースを使用してデータ整合性を最大限に保護するように設計されています。

このセクションでは、基本的なONTAP NVFAILについて説明し、MetroCluster固有のトピックを扱います。

MetroClusterでは、少なくとも1台の他のコントローラのローカルNVRAMとNVRAMに書き込みが記録されるまで、書き込み確認は行われません。このアプローチにより、ハードウェア障害や停電が発生しても、転送中のI/Oが失われることはありません。ローカルのNVRAMに障害が発生したり、他のノードへの接続に障害が発生したりすると、データはミラーリングされなくなります。

ローカルNVRAMからエラーが報告されると、ノードはシャットダウンします。このシャットダウンにより、HAペアが使用されている場合はパートナーコントローラにフェイルオーバーされます。MetroClusterでは、動作は選択した全体的な設定によって異なりますが、リモートノートに自動的にフェイルオーバーされる場合があります。いずれの場合も、障害が発生したコントローラが書き込み処理を認識していないため、データは失われません。

リモートノードへのNVRAMレプリケーションがブロックされるサイト間接続障害は、より複雑な状況です。書き込みがリモートノードにレプリケートされなくなるため、コントローラで重大なエラーが発生した場合にデータが失われる可能性があります。さらに重要なことは、このような状況で別のノードにフェイルオーバーしようとするとデータが失われることです。

制御要素は、NVRAMが同期されているかどうかです。NVRAMが同期されていれば、ノード間のフェイルオーバーを安全に実行でき、データ損失のリスクはありません。MetroCluster構成では、NVRAMと基盤となるアグリゲートのプレックスが同期されていれば、データ損失のリスクなしにスイッチオーバーを安全に実行できます。

データが同期されていない場合、ONTAPは、フェイルオーバーまたはスイッチオーバーを強制的に実行しないかぎり、フェイルオーバーまたはスイッチオーバーを許可しません。この方法で条件を変更すると、元のコントローラにデータが残っている可能性があり、データ損失が許容されることが確認されます。

データベースは、ディスク上のデータのより大きな内部キャッシュを保持するため、フェイルオーバーやスイッチオーバーを強制的に実行した場合、データベースが破損する可能性が特に高くなります。強制的なフェイルオーバーまたはスイッチオーバーが発生した場合、以前に確認済みの変更は事実上破棄されます。ストレージアレイの内容は実質的に時間を逆方向に移動し、データベースキャッシュの状態はディスク上のデータの状態を反映しなくなります。

この状況からアプリケーションを保護するために、ONTAPでは、NVRAMの障害に対する特別な保護をボリュームに設定できます。この保護メカニズムがトリガーされると、ボリュームがNVFAILという状態になります。この状態になると、古いデータを使用しないように原因アプリケーションをシャットダウンするI/Oエラーが発生します。確認済みの書き込みはストレージシステムに残っているため、データが失われることはありません。データベースの場合は、コミットされたトランザクションデータがログに含まれている必要があります。

次の手順では、管理者がホストを完全にシャットダウンしてから、LUNとボリュームを手動で再度オンラインに戻します。これらの手順にはいくつかの作業が含まれる可能性がありますが、このアプローチはデータの整合性を確保するための最も安全な方法です。すべてのデータがこの保護を必要とするわけではありません。そのため、NVFAILの動作はボリューム単位で設定できます。

手動強制**NVFAIL**

サイト間に分散されているアプリケーションクラスタ（VMware、Oracle RACなど）でスイッチオーバーを強制的に実行する最も安全なオプションは、 `-force-nvfail-all` コマンドラインで。このオプションは、キャッシュされたすべてのデータが確実にフラッシュされるようにするための緊急措置として使用できます。災害が発生したサイトにもともと配置されていたストレージリソースをホストが使用している場合、I/Oエラーまたは古いファイルハンドルのいずれかを受信します。（`ESTALE`）エラー。Oracleデータベースがクラッシ

ュし、ファイルシステムが完全にオフラインになるか、読み取り専用モードに切り替わります。

スイッチオーバーの完了後、 `in-nvfailed-state` フラグをクリアし、LUNをオンラインにする必要があります。このアクティビティが完了したら、データベースを再起動できます。これらのタスクを自動化してRTOを短縮できます。

dr-force-nvfail

一般的な安全対策として、 `dr-force-nvfail` 通常運用時にリモートサイトからアクセスされる可能性があるすべてのボリューム（フェイルオーバー前に使用されるアクティビティ）にフラグを付けます。この設定により、選択したリモートボリュームが `in-nvfailed-state` スイッチオーバー中。スイッチオーバーの完了後、 `in-nvfailed-state` フラグをクリアし、LUNをオンラインにする必要があります。これらのアクティビティが完了したら、アプリケーションを再起動できます。これらのタスクを自動化してRTOを短縮できます。

結果は、 `-force-nvfail-all` 手動スイッチオーバーのフラグ。ただし、影響を受けるボリュームの数は、古いキャッシュを使用するアプリケーションまたはオペレーティングシステムから保護する必要があるボリュームだけに制限される場合があります。

を使用しない環境には、次の2つの重要な要件があります。 `dr-force-nvfail` アプリケーションボリューム：

- 強制スイッチオーバーは、プライマリサイトの障害から30秒以内に実行する必要があります。
- メンテナンスタスクの実行中や、SyncMirrorプレックスやNVRAMレプリケーションが同期されていないその他の状況では、スイッチオーバーを実行しないでください。最初の要件を満たすには、Tiebreakerソフトウェアを使用します。Tiebreakerソフトウェアは、サイト障害から30秒以内にスイッチオーバーを実行するように設定されています。これは、サイト障害が検出されてから30秒以内にスイッチオーバーを実行する必要があるという意味ではありません。これは、サイトが動作していることが確認されてから30秒が経過した場合に強制的にスイッチオーバーを実行しても安全ではないことを意味します。

2つ目の要件は、MetroCluster構成が同期されていないことが判明した場合に、自動スイッチオーバー機能をすべて無効にすることで部分的に満たすことができます。NVRAMレプリケーションとSyncMirrorプレックスの健常性を監視できるTiebreaker解決策を使用することを推奨します。クラスタが完全に同期されていない場合、Tiebreakerはスイッチオーバーをトリガーしません。

NetApp MCTBソフトウェアは同期ステータスを監視できないため、何らかの理由でMetroClusterが同期されていない場合は無効にする必要があります。ClusterLionにはNVRAM監視機能とプレックス監視機能が搭載されており、MetroClusterシステムが完全に同期されていることが確認されないかぎり、スイッチオーバーをトリガーしないように設定できます。

MetroCluster上のOracle単一インスタンス

前述したように、MetroClusterシステムが存在しても、データベースの運用に関するベストプラクティスが必ずしも追加されたり変更されたりするわけではありません。お客様のMetroClusterシステムで現在実行されているデータベースの大部分はシングルインスタンスであり、Oracle on ONTAPドキュメントに記載されている推奨事項に従っています。

事前設定されたOSを使用したフェイルオーバー

SyncMirrorはディザスタリカバリサイトにデータの同期コピーを提供しますが、そのデータを利用できるようにするには、オペレーティングシステムと関連するアプリケーションが必要です。基本的な自動化により、環

境全体のフェイルオーバー時間を大幅に短縮できます。Veritas Cluster Server（VCS）などのClusterware製品は、サイト間でクラスタを作成するためによく使用されます。多くの場合、フェイルオーバープロセスは単純なスクリプトで実行できます。

プライマリノードが失われた場合、代替サイトでデータベースをオンラインにするようにクラスタウェア（またはスクリプト）が設定されます。1つは、データベースを構成するNFSリソースまたはSANリソース用に事前設定されたスタンバイサーバを作成する方法です。プライマリサイトに障害が発生すると、クラスタウェアまたはスクリプト化された代替サイトが次のような一連の処理を実行します。

1. MetroClusterスイッチオーバーの強制実行
2. FC LUNの検出の実行（SANのみ）
3. ファイルシステムのマウント、ASMディスクグループのマウント
4. データベースの起動

このアプローチの主な要件は、リモートサイトでOSを実行することです。Oracleバイナリを使用して事前に設定する必要があります。つまり、Oracleのパッチ適用などのタスクをプライマリサイトとスタンバイサイトで実行する必要があります。また、災害が発生した場合は、Oracleバイナリをリモートサイトにミラーリングしてマウントすることもできます。

実際のアクティベーション手順は簡単です。LUN検出などのコマンドでは、FCポートあたりのコマンド数が少なくて済みます。ファイル・システムのマウントは' `mount` コマンドを実行し、データベースとASMの両方をCLIで1つのコマンドで起動および停止できます。スイッチオーバーの前にディザスタリカバリサイトでボリュームとファイルシステムを使用していない場合は、 `dr-force- nvfail` ボリューム：

仮想**OS**によるフェイルオーバー

データベース環境のフェイルオーバーを拡張して、オペレーティングシステム自体を含めることができます。理論的には、このフェイルオーバーはブートLUNで実行できますが、ほとんどの場合、仮想OSで実行されます。手順の手順は次のようになります。

1. MetroClusterスイッチオーバーの強制実行
2. データベースサーバ仮想マシンをホストするデータストアのマウント
3. 仮想マシンの起動
4. データベースを手動で起動するか、データベースを自動的に起動するように仮想マシンを設定します。たとえば、ESXクラスタが複数のサイトにまたがっている場合があります。災害が発生した場合は、スイッチオーバー後にディザスタリカバリサイトで仮想マシンをオンラインにすることができます。災害発生時に仮想データベースサーバをホストするデータストアが使用されていないかぎり、 `dr-force- nvfail` 関連付けられているボリューム。

MetroCluster上の拡張**Oracle RAC**

多くのお客様が、Oracle RACクラスタを複数のサイトにまたがって構成し、完全なアクティブ/アクティブ構成を実現することで、RTOを最適化しています。Oracle RACのクォーラム管理を含める必要があるため、設計全体が複雑になります。また、データは両方のサイトからアクセスされるため、強制的スイッチオーバーによって古いデータコピーが使用される可能性があります。

データのコピーは両方のサイトに存在しますが、データを提供できるのはアグリゲートを現在所有しているコントローラだけです。そのため、拡張RACクラスタでは、リモートのノードがサイト間接続でI/Oを実行する

必要があります。その結果、I/Oレイテンシが増加しますが、このレイテンシは一般的には問題になりません。RACインターコネクトネットワークは複数のサイトにまたがって拡張する必要があるため、とにかく高速で低レイテンシのネットワークが必要です。レイテンシが増加して原因に問題が発生した場合は、クラスタをアクティブ/パッシブで運用できます。I/O負荷の高い処理は、アグリゲートを所有するコントローラに対してローカルなRACノードに対して実行する必要があります。リモートノードは、より軽いI/O処理を実行するか、純粋にウォームスタンバイサーバとして使用されます。

アクティブ/アクティブの拡張RACが必要な場合は、MetroClusterではなくASMミラーリングを検討する必要があります。ASMミラーリングでは、データの特定のレプリカを優先することができます。したがって、すべての読み取りがローカルに行われる拡張RACクラスタを構築できます。読み取りI/Oがサイトを経由することはないため、レイテンシは最小限に抑えられます。すべての書き込みアクティビティは引き続きサイト間接続を転送する必要がありますが、このようなトラフィックは同期ミラーリング解決策では回避できません。

> 仮想ブートディスクを含むブートLUNがOracle RACで使用されている場合は、 `misscount` パラメータの変更が必要になる場合があります。RACタイムアウトパラメータの詳細については、を参照してください。 "ONTAPを使用したOracle RAC"。

2サイト構成

2サイトの拡張RAC構成では、すべてではないが多くの災害シナリオに無停止で対応できるアクティブ/アクティブデータベースサービスを提供できます。

RAC投票ファイル

MetroClusterに拡張RACを導入する場合は、クォーラム管理を最初に検討する必要があります。Oracle RACには、クォーラムを管理するための2つのメカニズム（ディスクハートビートとネットワークハートビート）があります。ディスクハートビートは、投票ファイルを使用してストレージアクセスを監視します。単一サイトのRAC構成では、基盤となるストレージシステムがHA機能を提供していれば、単一の投票リソースで十分です。

以前のバージョンのOracleでは、投票ファイルは物理ストレージデバイスに配置されていましたが、現在のバージョンのOracleでは、投票ファイルはASMディスクグループに格納されていました。

> Oracle RACはNFSでサポートされています。グリッドのインストールプロセスでは、一連のASMプロセスが作成され、グリッドファイルに使用されるNFSの場所がASMディスクグループとして提供されます。このプロセスはエンドユーザに対してほぼ透過的であり、インストール完了後にASMを継続的に管理する必要はありません。

2サイト構成の最初の要件は、無停止のディザスタリカバリプロセスを保証する方法で、各サイトが常に半数以上の投票ファイルにアクセスできるようにすることです。このタスクは、投票ファイルがASMディスクグループに格納される前は簡単でしたが、今日の管理者はASM冗長性の基本原則を理解する必要があります。

ASMディスクグループには3つの冗長性オプションがあります。 `external`、 `normal`および `high`。つまり、ミラーリングされていない、ミラーリングされている、3方向ミラーリングされているということです。という新しいオプションがあります。 `Flex` 利用可能ですが、めったに使用されません。冗長デバイスの冗長性レベルと配置によって、障害が発生した場合の動作が制御されます。例：

- 投票ファイルをに配置する `diskgroup` を使用 `external` 冗長性リソースを使用すると、サイト間接続が失われた場合に一方のサイトの削除が保証されます。
- 投票ファイルをに配置する `diskgroup` を使用 `normal` 各サイトにASMディスクが1つしかない冗長性を確保すると、どちらのサイトにもマジョリティクォーラムが存在しないためにサイト間接続が失われた場合に、両方のサイトでノードが削除されます。

- 投票ファイルをに配置する `diskgroup` を使用 `high` 一方のサイトに2本のディスクを配置し、もう一方のサイトに1本のディスクを配置する冗長性により、両方のサイトが動作していて相互にアクセスできる場合にアクティブ/アクティブ処理が可能になります。ただし、シングルディスクサイトがネットワークから分離されている場合、そのサイトは削除されます。

RACネットワークハートビート

Oracle RACネットワークハートビートは、クラスタインターコネクト経由でノードに到達できるかどうかを監視します。クラスタに残すには、あるノードが他のノードの半数以上にアクセスできる必要があります。この要件により、2サイトアーキテクチャのRACノード数は次のように選択されます。

- サイトごとに同じ数のノードを配置すると、ネットワーク接続が失われた場合に1つのサイトが削除されます。
- 一方のサイトにN個のノードを配置し、もう一方のサイトにN+1個のノードを配置すると、サイト間接続が失われてネットワーククォーラムに残っているノードの数が多くなり、削除するノードの数が少なくなります。

Oracle 12cR2より前のバージョンでは、サイト障害時にどの側で削除するかを制御することは不可能でした。各サイトのノード数が同じ場合、削除はマスターノード（通常は最初にブートするRACノード）によって制御されます。

Oracle 12cR2では、ノードの重み付け機能が導入されています。この機能により、管理者はOracleによるスプリットブレイン状態の解決方法をより細かく制御できます。簡単な例として、次のコマンドはRAC内の特定のノードの優先順位を設定します。

```
[root@host-a ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.
```

Oracle High-Availability Servicesを再起動すると、構成は次のようになります。

```
[root@host-a lib]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
```

ノード `host-a` が重要なサーバとして指定されました。2つのRACノードが分離されている場合は、 `host-a` 生き残って `host-b` 削除されます。

詳細については、Oracleのホワイトペーパー『Oracle Clusterware 12c Release 2 Technical Overview』を参照してください。」

12cR2より前のバージョンのOracle RACでは、CRSログを確認することでマスターノードを特定できます。

```
[root@host-a ~]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
 [root@host-a ~]# grep -i 'master node' /grid/diag/crs/host-
a/crs/trace/crsd.trc
2017-05-04 04:46:12.261525 :   CRSSE:2130671360: {1:16377:2} Master Change
Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:01:24.979716 :   CRSSE:2031576832: {1:13237:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
2017-05-04 05:11:22.995707 :   CRSSE:2031576832: {1:13237:221} Master
Change Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:28:25.797860 :   CRSSE:3336529664: {1:8557:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
```

このログは、マスターノードが 2 ノード host-a ID： 1。これはつまり host-a はマスターノードではありません。マスターノードのIDは、コマンドで確認できます。 olsnodes -n。

```
[root@host-a ~]# /grid/bin/olsnodes -n
host-a  1
host-b  2
```

IDがのノード 2 はです host-b`をクリックします。これはマスターノードです。各サイトに同じ数のノードがある構成では、 `host-b 2つのセットが何らかの理由でネットワーク接続を失った場合に存続するサイトです。

マスターノードを識別するログエントリがシステムから期限切れになる可能性があります。この場合、Oracle Cluster Registry（OCR）バックアップのタイムスタンプを使用できます。

```
[root@host-a ~]#  /grid/bin/ocrconfig -showbackup
host-b     2017/05/05 05:39:53     /grid/cdata/host-cluster/backup00.ocr
0
host-b     2017/05/05 01:39:53     /grid/cdata/host-cluster/backup01.ocr
0
host-b     2017/05/04 21:39:52     /grid/cdata/host-cluster/backup02.ocr
0
host-a     2017/05/04 02:05:36     /grid/cdata/host-cluster/day.ocr     0
host-a     2017/04/22 02:05:17     /grid/cdata/host-cluster/week.ocr     0
```

次の例では、マスターノードが host-b。また、マスターノードの変更も示します。 host-a 終了： host-b 5月4日の2時5分から21時39分までの間。マスターノードを識別するこの方法は、前回のOCRバックアップ以降にマスターノードが変更されている可能性があるため、CRSログもチェックされている場合にのみ使用で

きます。この変更が発生した場合は、OCRログに表示されます。

ほとんどのお客様は、環境全体と各サイトで同数のRACノードにサービスを提供する投票ディスクグループを1つ選択しています。ディスクグループは、データベースが格納されているサイトに配置する必要があります。接続が失われると、リモートサイトが削除されます。リモートサイトにはクォーラムがなくなり、データベースファイルにもアクセスできなくなりますが、ローカルサイトは通常どおり稼働し続けます。接続が回復したら、リモートインスタンスを再びオンラインにすることができます。

災害が発生した場合は、サバイバーサイトでデータベースファイルと投票ディスクグループをオンラインにするためにスイッチオーバーが必要です。災害によってAUSOでスイッチオーバーがトリガーされた場合、クラスタが同期されていてストレージリソースが正常にオンラインになるため、NVFAILはトリガーされません。AUSOは非常に高速な操作であり、 `disktimeout` 有効期限が切れます。

サイトが2つしかないため、自動化された外部タイブレークソフトウェアを使用することは不可能であり、強制スイッチオーバーは手動で行う必要があります。

3サイト構成

3つのサイトで拡張RACクラスタを構築する方がはるかに簡単です。MetroClusterシステムの各半分をホストする2つのサイトもデータベースワークロードをサポートし、3つ目のサイトはデータベースとMetroClusterシステムの両方のTiebreakerとして機能します。Oracle Tiebreakerの構成は、第3のサイトに投票に使用するASMディスクグループのメンバーを配置するだけで簡単に構成できます。また、RACクラスタに奇数のノードを配置するために、第3のサイトに運用インスタンスを配置することもできます。

> 拡張RAC構成でNFSを使用する場合の重要な情報については、「クォーラム障害グループ」に関するOracleのドキュメントを参照してください。要するに、クォーラムリソースをホストする3番目のサイトへの接続が失われても、プライマリOracleサーバまたはOracle RACプロセスが停止しないように、NFSマウントオプションを変更してsoftオプションを含める必要がある場合があります。

SnapMirrorアクティブ同期

SnapMirrorアクティブ同期ありのOracleデータベース

SnapMirrorアクティブ同期により、個 々 のOracleデータベースおよびアプリケーション環境に対して、選択的なRPO=0同期ミラーリングが可能になります。

SnapMirrorアクティブ同期は、SAN向けに強化されたSnapMirror機能です。これにより、ホストは、LUNをホストしているシステムとレプリカをホストしているシステムの両方からLUNにアクセスできます。

SnapMirror Active SyncとSnapMirror Syncはレプリケーションエンジンを共有しますが、SnapMirror Active Syncには、エンタープライズアプリケーションに対する透過的なアプリケーションフェイルオーバーやフェイルバックなどの追加機能が含まれています。

実際には、個 々 のワークロードに対して選択的かつきめ細かなRPO=0の同期レプリケーションを有効にすることで、MetroClusterのきめ細かなバージョンと同様に機能します。下位レベルのパスの動作はMetroClusterとは大きく異なりますが、ホスト側から見た結果はほぼ同じです。

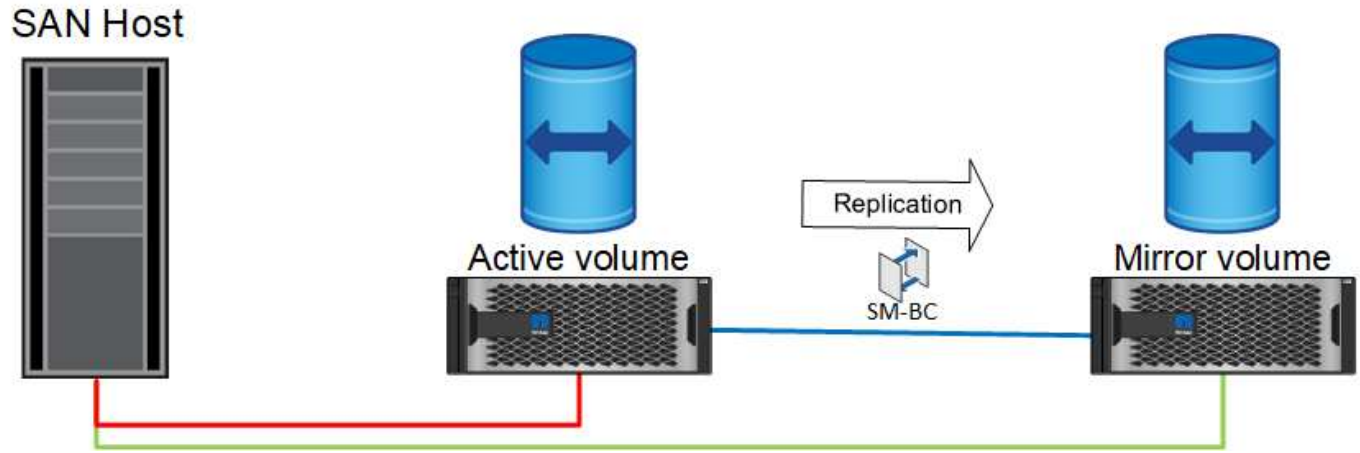
パスアクセス

SnapMirrorをアクティブに同期すると、プライマリとリモートの両方のストレージアレイから、ホストオペレーティングシステムからストレージデバイスを認識できるようになります。パスは、ストレージシステムとホ

ストの間の最適なパスを特定するための業界標準プロトコルであるAsymmetric Logical Unit Access（ALUA；非対称論理ユニットアクセス）を通じて管理されます。

I/Oへのアクセスに最も短いデバイスパスはアクティブ/最適パスとみなされ、残りのパスはアクティブ/非最適パスとみなされます。

SnapMirrorアクティブ同期関係は、異なるクラスタにあるSVMのペア間で確立されます。どちらのSVMもデータを提供できますが、ALUAは、LUNが配置されているドライブの所有権を現在持っているSVMを優先的に使用します。リモートSVMへのIOは、SnapMirrorのアクティブな同期インターコネクトを使用して経由でプロキシされます。



同期レプリケーション

通常の運用では、1つの例外を除き、リモートコピーは常にRPO=0の同期レプリカです。データをレプリケートできない場合、SnapMirrorのアクティブな同期により、データをレプリケートしてIOの提供を再開する必要がなくなります。このオプションは、レプリケーションリンクの損失がほぼ災害になると考えているお客様や、データをレプリケートできないときに業務の停止を望まないお客様に適しています。

ストレージハードウェア

他のストレージディザスタリカバリソリューションとは異なり、SnapMirrorアクティブ同期は非対称プラットフォームの柔軟性を提供します。各サイトのハードウェアが同一である必要はありません。この機能を使用すると、SnapMirrorアクティブ同期をサポートするために使用するハードウェアのサイズを適正化できます。リモートストレージシステムは、本番環境のワークロードを完全にサポートする必要がある場合はプライマリサイトと同一にすることができますが、災害によってI/Oが減少した場合は、リモートサイトの小規模システムよりも対費用効果が高くなります。

ONTAPメディエーター

ONTAPメディエーターは、NetAppサポートからダウンロードするソフトウェアアプリケーションです。Mediatorは、プライマリサイトとリモートサイトの両方のストレージクラスタのフェイルオーバー処理を自動化します。オンプレミスまたはクラウドでホストされた小規模な仮想マシン（VM）に導入できます。設定後は、両方のサイトのフェイルオーバーシナリオを監視するための第3のサイトとして機能します。

SnapMirror Active Syncを使用した**Oracle**データベースのフェイルオーバー

SnapMirrorのアクティブな同期でOracleデータベースをホストする主な理由は、計画的ストレージイベントと計画外ストレージイベントの発生時に透過的なフェイルオーバー

を実現するためです。

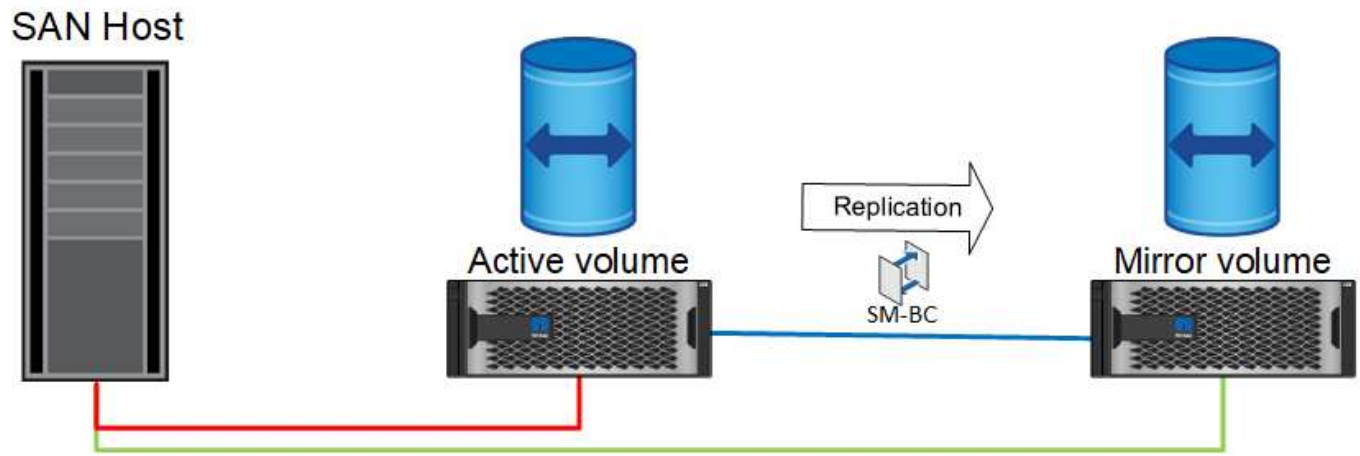
SnapMirror Active Syncでは、計画的フェイルオーバーと計画外フェイルオーバーの2種類のストレージフェイルオーバー処理がサポートされます。どちらの処理も多少異なります。計画的フェイルオーバーは、リモートサイトへの迅速なスイッチオーバーのために管理者が手動で開始し、計画外フェイルオーバーは3番目のサイトのメディエーターによって自動的に開始されます。計画的フェイルオーバーの主な目的は、パッチ適用とアップグレードの差分実行、ディザスタリカバリテストの実行、または完全なアクティブ同期機能を実証するためのサイト間の運用の切り替えという正式なポリシーの採用です。

次の図は、通常、フェイルオーバー、フェイルバックの各処理中に何が発生するかを示しています。わかりやすいように、レプリケートされたLUNを表しています。実際のSnapMirrorアクティブ同期構成では、レプリケーションはボリュームに基づいて行われます。各ボリュームには1つ以上のLUNが含まれていますが、わかりやすくするためにボリュームレイヤは削除されています。

通常運用時

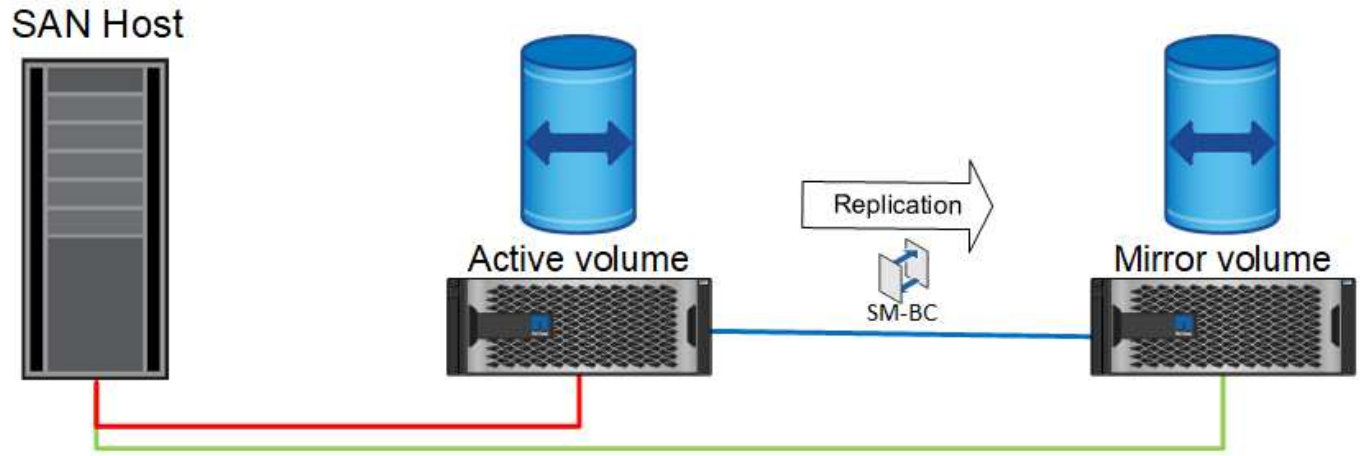
通常の操作では、ローカルレプリカまたはリモートレプリカからLUNにアクセスできます。赤い線はALUAによってアドバタイズされる最適パスを示し、そのパスにIOが優先的に送信されることになります。

緑の線はアクティブなパスですが、そのパスのIOをSnapMirrorのアクティブな同期パス経由で渡す必要があるため、レイテンシが高くなります。追加のレイテンシは、SnapMirrorアクティブ同期に使用されるサイト間のインターコネクトの速度によって異なります。



失敗

計画的フェイルオーバーまたは計画外フェイルオーバーのためにアクティブなミラーコピーを使用できなくなった場合は、明らかに使用できなくなります。ただし、リモートシステムには同期レプリカがあり、リモートサイトへのSANパスはすでに存在します。リモートシステムは、そのLUNのIOを処理できます。

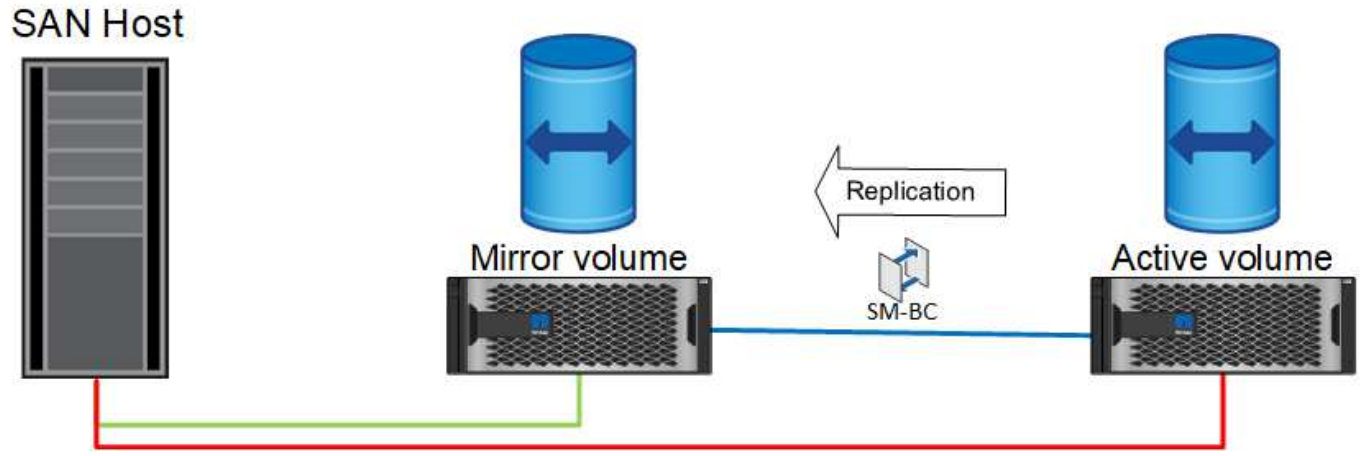


フェイルオーバー

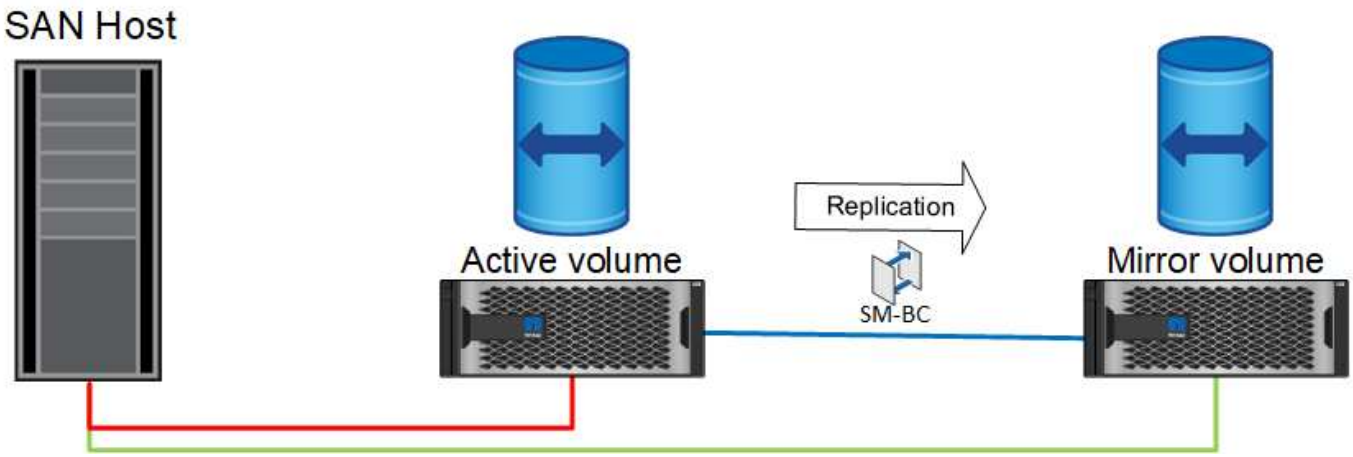
フェイルオーバーを実行すると、リモートコピーがアクティブコピーになります。パスが[Active]から[Active]/[Optimized]に変更され、IOは引き続きデータ損失なしで処理されます。

修復する

ソースシステムが稼働状態に戻ると、SnapMirrorアクティブ同期はレプリケーションを再同期できますが、逆方向に実行されます。現在の構成は、アクティブミラーサイトが反転されている点を除き、基本的には開始点と同じです。



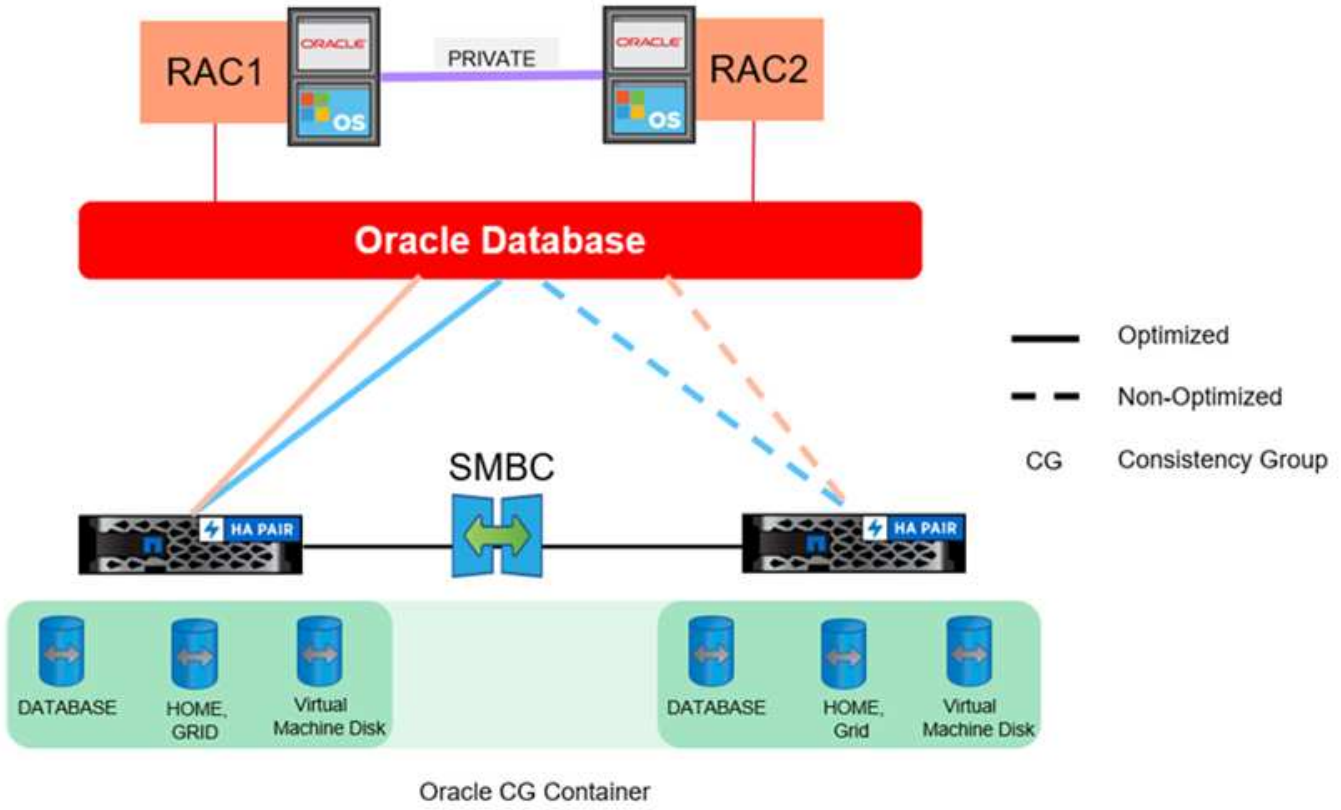
必要に応じて、管理者はフェイルバックを実行し、LUNのアクティブコピーを元のコントローラに戻すことができます。



シングルインスタンスの**Oracle**データベースと**SnapMirror**アクティブ同期

次の図は、Oracleデータベースのプライマリストレージクラスタとリモートストレージクラスタの両方からストレージデバイスをゾーニングまたは接続するシンプルな導入モデルを示しています。

Oracleはプライマリにのみ設定されます。このモデルは、ストレージ側で災害が発生した場合にシームレスなストレージフェイルオーバーに対応し、アプリケーションのダウンタイムを発生させることなくデータの損失をゼロにします。ただし、このモデルでは、サイト障害時のデータベース環境の高可用性を実現できません。このタイプのアーキテクチャは、データ損失ゼロの解決策でストレージサービスの高可用性を実現したいが、データベースクラスタが完全に失われた場合には手動での作業が必要である場合に便利です。



このアプローチでは、Oracleのライセンスコストも削減できます。リモートサイトでOracleデータベースノードを事前に設定するには、ほとんどのOracleライセンス契約に基づいてすべてのコアがライセンスされている必要があります。Oracleデータベースサーバのインストールと、稼働しているデータコピーのマウントにかかる時間が原因で遅延が発生しても問題ない場合は、コスト効率に優れた設計にすることができます。

Oracle RACと**SnapMirror**のアクティブな同期

SnapMirror Active Syncを使用すると、ロードバランシングや個 々 のアプリケーションのフェイルオーバーなど、データセットのレプリケーションをきめ細かく制御できます。アーキテクチャ全体は拡張RACクラスタのように見えますが、一部のデータベースは特定のサイト専用で、全体の負荷は分散されます。

たとえば、6つのデータベースを個別にホストするOracle RACクラスタを構築できます。3つのデータベースのストレージは主にサイトAでホストされ、残りの3つのデータベースのストレージはサイトBでホストされます。この構成により、クロスサイトトラフィックが最小限に抑えられ、可能な限り最高のパフォーマンスが保証されます。また、ストレージシステムに対してローカルなデータベースインスタンスをアクティブパスで使用するようにアプリケーションを設定します。これにより、RACインターコネクトトラフィックが最小限に抑えられます。最後に、この全体的な設計により、すべてのコンピューティングリソースが均等に使用されます。ワークロードの変化に応じて、データベースをサイト間で選択的にフェイルバックして、ロードが均一になるようにすることができます。

きめ細かさを除けば、SnapMirror Active Syncareを使用するOracle RACの基本原則とオプションは"MetroCluster上のOracle RAC"

Oracleデータベースと**SnapMirror**のアクティブな同期が失敗した場合

SnapMirror Active Sync（SM-AS）で障害が発生した場合は、それぞれ結果が異なりま

す。

シナリオ（ **Scenario** ）	結果
レプリケーションリンク障害	Mediatorはこのスプリットブレインシナリオを認識し、マスターコピーを保持するノードでI/Oを再開します。サイト間の接続がオンラインに戻ると、代替サイトで自動再同期が実行されます。
プライマリサイトストレージの障害	自動計画外フェイルオーバーはMediatorによって開始されます。 I/Oの中断はありません。
リモートサイトのストレージ障害	I/Oの中断はありません。ネットワークが原因で一時的に停止し、同期レプリケーションが中断され、マスターがI/O処理を継続する正当な所有者であることが確認されます（コンセンサス）。そのため、数秒間I/Oが一時停止してから、I/Oが再開されます。 サイトがオンラインの場合は自動再同期が実行されます。
MediatorまたはMediatorとストレージアレイの間のリンクの停止	I/Oは継続してリモートクラスタとの同期が維持されますが、Mediatorがないと、計画外フェイルオーバーや自動フェイルバックは実行できません。
HAクラスタ内の一方のストレージコントローラの停止	HAクラスタのパートナーノードでテイクオーバー（NDO）が試行されます。テイクオーバーに失敗すると、Mediatorはストレージの両方のノードが停止していることを認識し、リモートクラスタへの自動計画外フェイルオーバーを実行します。
ディスクノテイシ	IOは、連続して3つのディスク障害が発生しても継続されます。これはRAID-TECの一部です。
一般的な環境でサイト全体が停止する	障害が発生したサイトのサーバは、明らかに使用できなくなります。クラスタリングをサポートするアプリケーションは、両方のサイトで実行し、代替サイトで処理を継続するように設定できます。ただし、ほとんどのアプリケーションでは、SM-ASでメディエーターが必要な場合と同様の第3のサイトTiebreakerが必要です。 アプリケーションレベルのクラスタがない場合は、サバイバーサイトでアプリケーションを起動する必要があります。これは可用性に影響しますが、RPO=0は維持されます。データが失われることはありません。

Oracleデータベースの移行

ONTAPストレージシステムへのOracleデータベースの移行

新しいストレージプラットフォームの機能を活用するには、必ず新しいストレージシステムにデータを配置する必要があるという、避けられない要件が1つあります。ONTAP

を使用すると、ONTAPからONTAPへの移行とアップグレード、外部LUNのインポート、ホストオペレーティングシステムまたはOracleデータベースソフトウェアを直接使用する手順など、移行プロセスを簡単に実行できます。

以前に公開されていたテクニカルレポート_TR-4534：『Migration of Oracle Databases to NetApp Storage Systems_

新しいデータベースプロジェクトの場合、データベースとアプリケーション環境が適切に構築されているため、これは問題になりません。ただし、移行には、ビジネスの中断、移行の完了に必要な時間、必要なスキルセット、リスクの最小化という特別な課題が伴います。

スクリプト

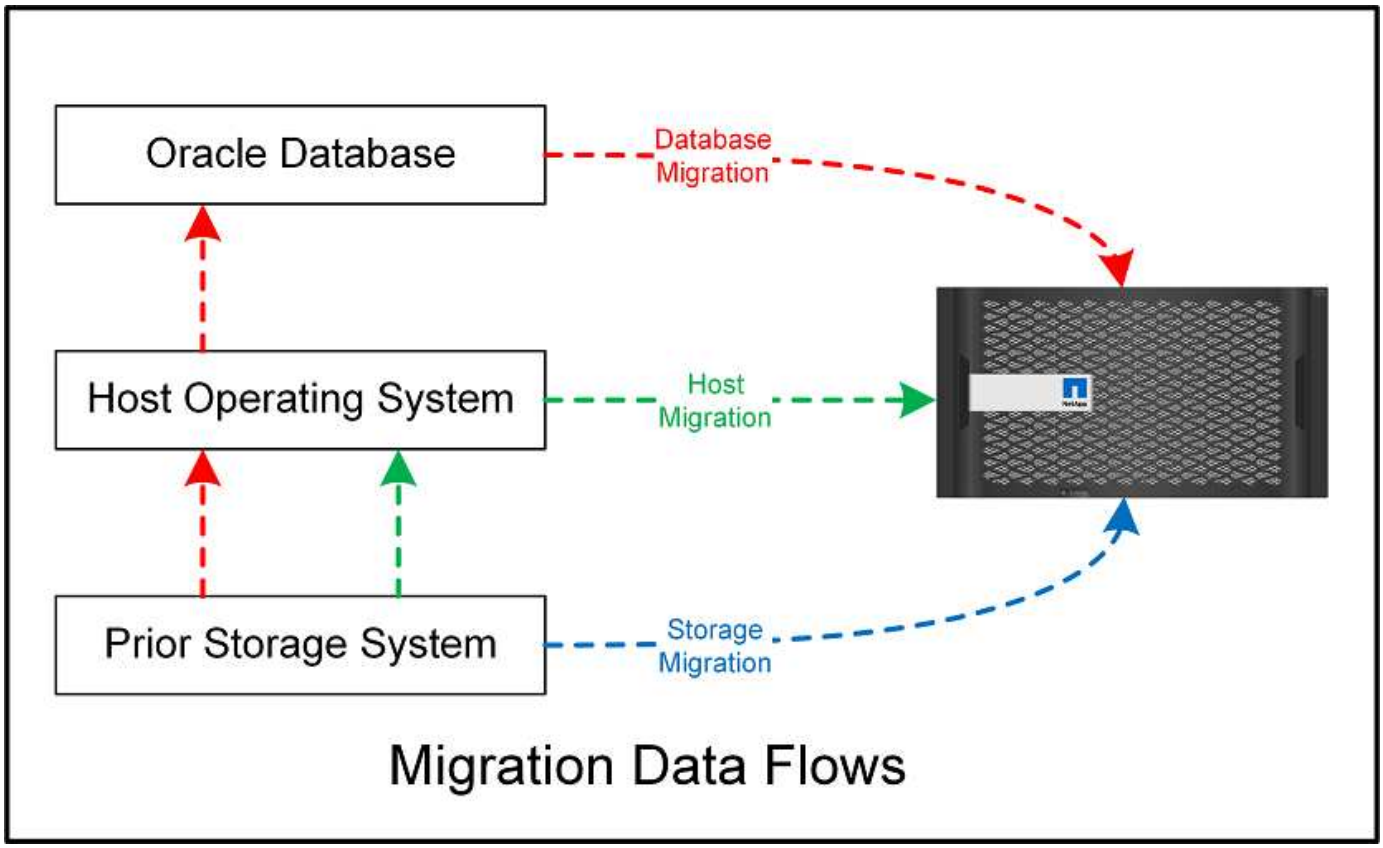
このドキュメントには、サンプルスクリプトが記載されています。これらのスクリプトは、ユーザによるミスの可能性を減らすために、移行のさまざまな側面を自動化するサンプル方法を提供します。スクリプトを使用すると、移行を担当するITスタッフの全体的な要求を軽減し、プロセス全体をスピードアップできます。これらのスクリプトはすべて、NetAppプロフェッショナルサービスとNetAppパートナーが実施した実際の移行プロジェクトを基に作成されています。このドキュメント全体で使用例が示されています。

Oracleデータベースの移行計画

Oracleのデータ移行は、データベース、ホスト、ストレージアレイの3つのレベルのいずれかで実行できます。

違いは、解決策全体のどのコンポーネント（データベース、ホストオペレーティングシステム、ストレージシステム）がデータの移動を担当しているかです。

次の図は、移行レベルとデータフローの例を示しています。データベースレベルの移行では、元のストレージシステムからホストレイヤとデータベースレイヤを経由して新しい環境にデータが移動されます。ホストレベルの移行も同様ですが、データはアプリケーションレイヤを通過せず、代わりにホストプロセスを使用して新しい場所に書き込まれます。最後に、ストレージレベルの移行では、NetApp FASシステムなどのアレイがデータ移動を行います。



データベースレベルの移行とは、通常、スタンバイデータベースを介してOracleのログ配布を使用して、Oracleレイヤでの移行を完了することを指します。ホストレベルの移行は、ホストオペレーティングシステム構成の標準機能を使用して実行されます。この構成には、cp、tar、およびOracle Recovery Manager（RMAN）などのコマンドを使用するか、論理ボリュームマネージャ（LVM）を使用してファイルシステムの基盤となるバイトを再配置するファイルコピー処理が含まれます。Oracle Automatic Storage Management（ASM）は、データベースアプリケーションのレベル以下で実行されるため、ホストレベルの機能に分類されます。ASMは、ホスト上の通常の論理ボリュームマネージャの代わりに使用されます。最後に、データをストレージアレイレベル（オペレーティングシステムのレベル以下）で移行できます。

計画に関する考慮事項

移行に最適な方法は、移行する環境の規模、ダウンタイムの回避の必要性、移行の実行に必要な全体的な作業など、さまざまな要因の組み合わせによって異なります。大規模なデータベースの移行には明らかに多くの時間と労力が必要ですが、そのような移行の複雑さは最小限です。小規模なデータベースは迅速に移行できますが、数千ものデータベースを移行する必要がある場合は、規模の大きさによって複雑な作業が発生する可能性があります。最後に、データベースの規模が大きいほど、ビジネスクリティカルである可能性が高くなります。そのため、バックアウトパスを維持しながらダウンタイムを最小限に抑える必要があります。

ここでは、移行戦略を計画する際の考慮事項について説明します。

データサイズ

移動するデータベースのサイズは移行計画に明らかに影響しますが、サイズがカットオーバー時間に必ずしも影響するとは限りません。大量のデータを移行する必要がある場合、主に帯域幅を考慮する必要があります。コピー処理は通常、効率的なシーケンシャルI/Oを使用して実行されます。控えめな見積もりでは、コピー処理に使用できるネットワーク帯域幅の使用率が50%であると想定しています。たとえば、8GBのFCポートでは理論上約800MBpsの転送が可能です。使用率が50%であれば、約400Mbpsの速度でデータベースをコピーできます。そのため、この速度では約7時間で10TBのデータベースをコピーできます。

長距離の移行では、通常、ログ配布プロセスなど、より創造的なアプローチが必要になります。詳細については、を参照してください。 "データファイルのオンライン移動"。長距離IPネットワークでは、LANやSANの速度に近い場所で帯域幅が使用されることはほとんどありません。あるケースでは、アーカイブログの生成頻度が非常に高い220TBデータベースの長距離移行をNetAppが支援しました。データ転送のために選択されたアプローチは、可能な限り最大の帯域幅を提供するため、テープの毎日の出荷でした。

データベース数

多くの場合、大量のデータを移動する際の問題はデータサイズではなく、データベースをサポートする構成の複雑さです。50TBのデータベースを移行する必要があるというだけでは、十分な情報ではありません。1つの50TBのミッションクリティカルなデータベース、4、000個のレガシーデータベースの集まり、または本番環境と非本番環境の混在環境などが考えられます。場合によっては、ほとんどのデータがソースデータベースのクローンで構成されています。これらのクローンは簡単に再作成できるため、マイグレートする必要はありません。特に、新しいアーキテクチャでNetApp FlexCloneボリュームを利用するように設計されている場合は、クローンを簡単に再作成できるためです。

移行を計画するには、対象となるデータベースの数と、データベースの優先順位を決定する方法を理解しておく必要があります。データベースの数が増えるにつれて、優先される移行オプションはスタック内で徐 々 に低くなる傾向があります。たとえば、RMANを使用して単一のデータベースのコピーを簡単に実行でき、短時間の停止が発生する可能性があります。これはホストレベルのレプリケーションです。

データベースが50個ある場合は、RMANコピーを受信する新しいファイルシステム構造を設定してデータを所定の場所に移動することを避けた方が簡単な場合があります。このプロセスは、ホストベースのLVM移行を利用して、古いLUNから新しいLUNにデータを再配置することで実行できます。これにより、データベース管理者（DBA）チームからOSチームに責任が移され、データベースに関して透過的にデータが移行されます。ファイルシステム構成は変更されません。

最後に、200台のサーバにまたがる500個のデータベースを移行する必要がある場合は、ONTAP Foreign LUN Import（FLI）機能などのストレージベースのオプションを使用して、LUNの直接移行を実行できます。

リアーキテクチャノヨウケン

通常、データベースファイルのレイアウトを変更して新しいストレージアレイの機能を利用する必要がありますが、必ずしもそうであるとは限りません。たとえば、EFシリーズオールフラッシュアレイの機能は、主にSANのパフォーマンスと信頼性を重視しています。ほとんどの場合、データレイアウトに関する特別な考慮事項なしに、データベースをEFシリーズアレイに移行できます。要件は、高いIOPS、低レイテンシ、堅牢な信頼性だけです。RAID構成やDynamic Disk Poolsなどの要素に関連するベストプラクティスはありますが、EFシリーズのプロジェクトでこれらの機能を活用するためにストレージアーキテクチャ全体を大幅に変更しなければならないことはほとんどありません。

これとは対照的に、ONTAPへの移行では、最終的な構成で最大限の価値を実現するために、データベースレイアウトをより慎重に検討する必要があります。ONTAP自体は、特定のアーキテクチャの作業がなくても、データベース環境に多くの機能を提供します。最も重要なのは、現在のハードウェアが寿命に達したときに、システムを停止せずに新しいハードウェアに移行できることです。一般的には、ONTAPへの移行は最後に実行する必要があります。後続のハードウェアはインプレースでアップグレードされ、データは無停止で新しいメディアに移行されます。

いくつかの計画では、さらに多くの利点が利用可能です。スナップショットの使用には、最も重要な考慮事項があります。Snapshotは、バックアップ、リストア、クローニング処理をほぼ瞬時に実行するための基盤です。スナップショットの機能の一例として、最大の用途は、6台のコントローラ上の約250個のLUNで996TBの単一データベースを実行している場合です。このデータベースのバックアップは2分で完了し、リストアは2分で完了し、クローニングは15分で完了します。その他のメリットとしては、ワークロードの変化に応じてクラスタ内でデータを移動できる機能や、サービス品質（QoS）制御を適用して、マルチデータベース環境

で安定した優れたパフォーマンスを提供できる機能などがあります。

QoS制御、データ再配置、Snapshot、クローニングなどのテクノロジは、ほぼすべての構成で機能します。しかし、利益を最大化するためには、一般的にいくつかの考えが必要です。場合によっては、新しいストレージアレイへの投資を最大限に活用するために、データベースストレージのレイアウトの設計変更が必要になることがあります。ホストベースまたはストレージベースの移行では元のデータレイアウトがレプリケートされるため、このような設計の変更は移行戦略に影響する可能性があります。移行を完了してONTAP向けに最適化されたデータレイアウトを提供するには、追加の手順が必要になる場合があります。に示す手順 "Oracle移行手順の概要" また、データベースを移行するだけでなく、最小限の労力で最適な最終レイアウトに移行する方法についても説明します。

カットオーバー時間

カットオーバー中のサービス停止の許容最大値を決定する必要があります。移行プロセス全体がシステム停止を引き起こすと想定するのはよくある間違いです。サービスの中断が発生する前に多くのタスクを完了できます。また、多くのオプションを使用すると、システム停止やシステム停止を伴わずに移行を完了できます。カットオーバーの時間は手順によって異なるため、システム停止が避けられない場合でも、許容されるサービス停止の最大値を定義する必要があります手順。

たとえば、10TBのデータベースのコピーには、通常、約7時間かかります。ビジネスニーズが7時間の停止を許容している場合、ファイルのコピーは簡単で安全な移行オプションです。5時間で対応できない場合は、シンプルなログ配布プロセス（ "Oracleのログ配布"最小限の労力でセットアップでき、カットオーバー時間を約15分に短縮できます。この間、データベース管理者はプロセスを完了できます。許容できない時間が15分であった場合は、スクリプトを使用して最終カットオーバープロセスを自動化し、カットオーバー時間をわずか数分に短縮できます。移行はいつでも高速化できますが、そのためには時間と労力がかかります。カットオーバーの所要時間は、ビジネス部門が許容できる範囲で決定する必要があります。

バックアウトパス

完全にリスクのない移行はありません。テクノロジが完全に動作していても、ユーザエラーの可能性は常にあります。選択した移行パスに関連するリスクと、失敗した移行の結果を考慮する必要があります。たとえば、Oracle ASMの透過的オンラインストレージ移行機能は、Oracle ASMの主要機能の1つであり、この方法は、最も信頼性の高い方法の1つです。ただし、この方法ではデータが不可逆的にコピーされています。万一ASMで問題が発生した場合、簡単にバックアウトできるパスはありません。唯一の選択肢は、元の環境をリストアするか、ASMを使用して移行を元のLUNに戻すことです。このリスクは、元のストレージ・システムでスナップショット・タイプのバックアップを実行できる場合には、最小限に抑えることができますが、排除することはできません。

リハーサル

一部の移行手順は、実行前に完全に検証する必要があります。移行とカットオーバープロセスのリハーサルは、ミッションクリティカルなデータベースへの一般的な要求であり、移行を成功させ、ダウンタイムを最小限に抑える必要があります。また、ユーザ受け入れテストは移行後の作業に含まれることが多く、システム全体を本番環境に戻すには、これらのテストが完了する必要があります。

リハーサルが必要な場合は、いくつかのONTAP機能を使用すると、プロセスがはるかに簡単になります。特に、スナップショットを使用すると、テスト環境をリセットして、データベース環境のスペース効率に優れた複数のコピーをすばやく作成できます。

の手順

Oracleデータベースへの移行には、さまざまな手順を使用できます。最適なソリューションは、ビジネスニーズに応じて異なります。

多くの場合、システム管理者とDBAには、物理ボリュームのデータの再配置、ミラーリングとミラーリングの解除、Oracle RMANを使用したデータのコピーなど、独自の方法があります。

これらの手順は、主に、利用可能なオプションの一部に慣れていないITスタッフ向けのガイダンスとして提供されています。さらに、各移行アプローチのタスク、時間要件、スキルセットの要求についても説明します。これにより、NetAppやパートナーのプロフェッショナルサービスやIT管理者などの他の関係者が、各手順の要件をより十分に理解できるようになります。

移行戦略を作成するための単一のベストプラクティスはありません。計画を作成するには、まず可用性オプションを理解してから、ビジネスのニーズに最適な方法を選択する必要があります。次の図は、基本的な考慮事項と一般的な結論を示していますが、すべての状況に当てはまるわけではありません。

たとえば、1つのステップで合計データベースサイズの問題が生成されます。次の手順は、データベースが1TBを超えているかどうかによって異なります。推奨される手順は、一般的なお客様の慣行に基づいた推奨事項です。ほとんどのお客様は、DataGuardを使用して小規模なデータベースをコピーすることはありませんが、場合によってはコピーすることもあります。ほとんどのお客様は、時間がかかるため50TBのデータベースをコピーしようとはしませんが、一部のお客様では、このような処理を実行できるだけの十分なメンテナンス時間がある場合があります。

移行パスが最適な考慮事項のタイプのフローチャートを確認できます。 "こちらをご覧ください"。

データファイルのオンライン移動

Oracle 12cR1以降では、データベースをオンラインにしたままデータファイルを移動できます。さらに、異なる種類のファイルシステム間で動作します。たとえば、データファイルをxfsファイルシステムからASMに再配置できます。個 々 のデータファイルの移動処理が必要になるため、この方法は一般に大規模な環境では使用されませんが、データファイルの数が少ない小規模なデータベースの場合は検討する価値があります。

また、データファイルを移動するだけで、既存のデータベースの一部を移行することもできます。たとえば、アクティブでないデータファイルをコスト効率に優れたストレージ（アイドルブロックをオブジェクトストアに格納できるFabricPoolなど）に再配置できます。

データベースレベルの移行

データベースレベルでの移行とは、データベースがデータを再配置できることを意味します。具体的には、ログ配布を意味します。RMANやASMなどのテクノロジはOracle製品ですが、移行の目的では、ファイルのコピーやボリュームの管理を行うホストレベルで動作します。

ログ配布

データベースレベルの移行の基盤となるのがOracleアーカイブログです。このログには、データベースに対する変更のログが記録されます。ほとんどの場合、アーカイブログはバックアップおよびリカバリ戦略の一部です。リカバリプロセスでは、まずデータベースをリストアし、次に1つ以上のアーカイブログを再生して、データベースを目的の状態に戻します。これと同じ基本テクノロジを使用して、運用の中断をほとんどまたはまったく伴わずに移行を実行できます。さらに重要なのは、このテクノロジにより、元のデータベースに手を加えずに移行できるため、バックアウトパスが維持されます。

移行プロセスは、まずセカンダリサーバへのデータベースバックアップのリストアから始まります。これには

さまざまな方法がありますが、ほとんどのお客様は通常のバックアップアプリケーションを使用してデータファイルをリストアしています。データファイルがリストアされたら、ユーザはログの配布方法を設定します。その目的は、プライマリデータベースで生成されたアーカイブログの一定のフィードを作成し、リストアしたデータベースでそれらのログを再生して、両方を同じ状態に保つことです。カットオーバー時間に達すると、ソースデータベースが完全にシャットダウンされ、最終的なアーカイブログと場合によってはREDOログがコピーされて再生されます。コミットされた最終トランザクションの一部がREDOログに含まれている可能性があるため、REDOログも考慮することが重要です。

これらのログが転送されて再生されると、両方のデータベースの整合性が維持されます。この時点で、ほとんどのお客様はいくつかの基本的なテストを実施します。移行プロセス中にエラーが発生した場合は、ログ再生でエラーが報告されて失敗します。構成が最適であることを確認するために、既知のクエリまたはアプリケーションベースのアクティビティに基づいていくつかのクイックテストを実行することをお勧めします。また、移行したデータベースに元のデータベースが存在するかどうかを確認する前に、最後のテストテーブルを1つ作成してから元のデータベースをシャットダウンすることも一般的です。この手順では、最終的なログ同期中にエラーが発生していないことを確認します。

シンプルなログ配布移行は、元のデータベースに対してアウトオブバンドで構成できるため、ミッションクリティカルなデータベースに特に役立ちます。ソースデータベースの構成の変更は必要ありません。移行環境のリストアと初期設定は、本番環境の運用には影響しません。ログ配布を構成すると、一部のI/O要求が本番サーバに送信されます。ただし、ログ配布ではアーカイブログの単純なシーケンシャル読み取りが行われるため、本番データベースのパフォーマンスへの影響はほとんどありません。

ログ配布は、長距離の変更率の高い移行プロジェクトに特に有用であることが証明されています。たとえば、1つの220TBデータベースを約500マイル離れた場所に移行しました。変更率は非常に高く、セキュリティ上の制約があるため、ネットワーク接続を使用できませんでした。ログ配布はテープと宅配便を使用して実施しました。ソース・データベースのコピーは'最初に以下の手順を使用してリストアされましたログは、カットオーバーの最終セットが配信され、ログがレプリカデータベースに適用されるまで、宅配便によって毎週出荷されました。

Oracle DataGuard

場合によっては、完全なDataGuard環境が保証されることもあります。ログ配布またはスタンバイデータベースの構成をDataGuardと呼ぶのは正しくありません。Oracle DataGuardは、データベースレプリケーションを管理するための包括的なフレームワークですが、レプリケーションテクノロジではありません。移行作業における完全なDataGuard環境の主なメリットは、データベース間で透過的にスイッチオーバーできることです。また、新しい環境とのパフォーマンスやネットワーク接続問題などの問題が検出された場合に、元のデータベースに透過的にスイッチオーバーすることもできます。DataGuard環境を完全に構成するには、データベースレイヤだけでなくアプリケーションも構成して、アプリケーションがプライマリデータベースの場所の変更を検出できるようにする必要があります。一般的に、DataGuardを使用して移行を完了する必要はありませんが、DataGuardに関する豊富な専門知識を社内に持ち、移行作業にすでにDataGuardを利用しているお客様もいます。

再アーキテクチャ

前述したように、ストレージアレイの高度な機能を活用するには、データベースレイアウトの変更が必要になる場合があります。さらに、ASMからNFSファイルシステムへの移行などのストレージプロトコルの変更によって、ファイルシステムのレイアウトが変更される必要があります。

DataGuardなどのログ配布方法の主な利点の1つは、レプリケーション先がソースと一致している必要がないことです。ログ配布アプローチを使用してASMから通常のファイルシステムに（またはその逆に）移行する場合、問題はありません。データファイルの正確なレイアウトを宛先で変更して、Pluggable Database（PDB）テクノロジの使用を最適化したり、特定のファイルに対してQoS制御を選択的に設定したりできます。つまり、ログ配布に基づく移行プロセスを使用すると、データベースストレージレイアウトを簡単かつ安全に最適化できます。

サーバリソース

データベースレベルの移行の制限事項の1つに、2台目のサーバの必要性があります。この2台目のサーバは、次の2つの方法で使用できます。

1. 2番目のサーバは、データベースの永続的な新しいホームとして使用できます。
2. 2番目のサーバーを一時的なステージングサーバーとして使用できます。新しいストレージアレイへのデータ移行が完了してテストされると、LUNまたはNFSファイルシステムがステージングサーバから切断され、元のサーバに再接続されます。

最初のオプションは最も簡単ですが、非常に強力なサーバを必要とする非常に大規模な環境では使用できない可能性があります。2番目のオプションでは、ファイルシステムを元の場所に再配置するための追加作業が必要です。ファイルシステムをステージングサーバからアンマウントして元のサーバに再マウントできるため、NFSをストレージプロトコルとして使用する単純な操作にすることができます。

ブロックベースのファイルシステムでは、FCゾーニングまたはiSCSIイニシエータを更新するために追加の作業が必要です。ほとんどの論理ボリュームマネージャ（ASMを含む）では、元のサーバでLUNが使用可能になると、LUNが自動的に検出されてオンラインになります。ただし、ファイルシステムやLVMの実装によっては、データのエクスポートとインポートにより多くの作業が必要になる場合があります。正確な手順は異なる場合がありますが、通常は、移行を完了し、元のサーバにデータをリホームするためのシンプルで反復可能な手順を確立するのは簡単です。

単一のサーバ環境内でログ配布を設定してデータベースをレプリケートすることは可能ですが、ログを再生するには、新しいインスタンスに別のプロセスSIDを設定する必要があります。異なるSIDを持つ別のプロセスIDセットの下でデータベースを一時的に起動し、後で変更することができます。ただし、管理作業が複雑になり、データベース環境がユーザミスのリスクにさらされる可能性があります。

ホストレベルの移行

ホストレベルでデータを移行するとは、ホストオペレーティングシステムと関連するユーティリティを使用して移行を完了することを意味します。このプロセスには、Oracle RMANやOracle ASMなど、データをコピーするすべてのユーティリティが含まれます。

データコピー

単純なコピー操作の値を過小評価してはなりません。最新のネットワークインフラでは、1秒あたりのギガバイト数でデータを移動できます。ファイルのコピー処理は、効率的なシーケンシャル読み取り/書き込みI/Oに基づいています。ログ配布と比較すると、ホストのコピー処理ではこれ以上のシステム停止は避けられませんが、移行は単なるデータ移動ではありません。通常は、ネットワークへの変更、データベースの再起動時間、移行後のテストが含まれます。

データのコピーに実際に必要な時間はそれほど長くはありません。さらに、コピー処理では、元のデータが変更されないため、保証されたバックアウトパスが維持されます。移行プロセス中に問題が発生した場合は、元のデータを持つ元のファイルシステムを再アクティブ化できます。

プラットフォームの変更

再プラットフォーム化とは、CPUタイプの変更を指します。従来のSolaris、AIX、またはHP-UXプラットフォームからx86 Linuxにデータベースを移行する場合、CPUアーキテクチャの変更により、データを再フォーマットする必要があります。SPARC、IA64、POWER CPUはビッグエンディアンプロセッサとして知られ、x86とx86_64アーキテクチャはリトルエンディアンとして知られている。その結果、Oracleデータファイル内の一部のデータは、使用中のプロセッサによって順序が異なります。

従来、お客様はDataPumpを使用してプラットフォーム間でデータをレプリケートしてきました。データダンプは、ターゲットデータベースでより迅速にインポートできる特別なタイプの論理データエクスポートを作成するユーティリティです。データの論理コピーが作成されるため、DataPumpはプロセッサエンディアンの依存関係を残します。一部のお客様はデータダンプを再プラットフォーム化に使用していますが、Oracle 11gではより高速なオプションが利用できるようになりました。クロスプラットフォームで移動可能な表領域です。このアドバンスにより、テーブルスペースを別のエンディアン形式に変換できます。これは、DataPumpエクスポートよりも優れたパフォーマンスを提供する物理的な変換です。DataPumpエクスポートでは、物理バイトを論理データに変換してから、物理バイトに戻す必要があります。

DataPumpと移動可能な表領域の詳細については、NetAppのドキュメントでは説明していませんが、NetAppでは、新しいCPUアーキテクチャを使用して新しいストレージアレイログに移行する際にお客様をサポートしてきた経験に基づいて、次のような推奨事項がいくつかあります。

- DataPumpを使用している場合は、移行の完了に必要な時間をテスト環境で測定する必要があります。お客様は、移行の完了に必要な時間に驚かれることがあります。このような予期しないダウンタイムが発生すると、原因の停止が発生
- 多くのお客様は、クロスプラットフォームの移動可能な表領域はデータ変換を必要としないと誤って考えています。異なるエンディアンを持つCPUが使用されている場合、RMAN `convert` データファイルに対しては、事前に操作を実行しておく必要があります。これは瞬間的な操作ではありません。場合によっては、異なるデータファイルで複数のスレッドを動作させることで変換処理を高速化することができますが、変換処理を回避することはできません。

論理ボリュームマネージャによる移行

LVMは、1つ以上のLUNのグループを作成し、それらをエクステントと呼ばれる小さな単位に分割することで機能します。次に、エクステントのプールをソースとして使用して、基本的に仮想化された論理ボリュームを作成します。この仮想化レイヤーは、さまざまな方法で価値を提供します。

- 論理ボリュームは、複数のLUNから取得されたエクステントを使用できます。論理ボリューム上に作成されたファイルシステムは、すべてのLUNのパフォーマンス機能をフルに使用できます。また、ボリュームグループ内のすべてのLUNの均等なロードが促進され、より予測可能なパフォーマンスが提供されます。
- 論理ボリュームのサイズは、エクステントを追加したり、場合によっては削除したりすることで変更できます。論理ボリューム上のファイルシステムのサイズ変更は、通常無停止で実行されます。
- 基盤となるエクステントを移動することで、論理ボリュームを無停止で移行できます。

LVMを使用した移行は、エクステントの移動またはエクステントのミラーリング/ミラーリングという2つの方法のいずれかで機能します。LVMの移行では、効率的な大容量ブロックのシーケンシャルI/Oが使用され、パフォーマンスに関する懸念が生じることはほとんどありません。これが問題になった場合は、通常、I/O速度を調整するオプションがあります。これにより、移行の完了に必要な時間が長くなりますが、ホストとストレージシステムのI/O負荷が軽減されます。

ミラーおよびデミラー

AIX LVMなどの一部のボリュームマネージャでは、各エクステントのコピー数を指定したり、各コピーをホストするデバイスを制御したりできます。移行では、既存の論理ボリュームを取得し、基盤となるエクステントを新しいボリュームにミラーリングし、コピーの同期を待ってから、古いコピーをドロップします。バックアウトパスが必要な場合は、ミラーコピーが破棄される前に元のデータのSnapshotを作成できます。または、サーバを短時間シャットダウンして元のLUNをマスクしてから、格納されているミラーコピーを強制的に削除することもできます。これにより、リカバリ可能なデータのコピーが元の場所に保持されます。

エクステントの移行

ほとんどすべてのボリューム・マネージャではエクステントの移行が可能であり'複数のオプションが存在する場合もありますたとえば、一部のボリュームマネージャでは、管理者が特定の論理ボリュームの個 々 のエクステントを古いストレージから新しいストレージに再配置できます。Linux LVM2などのボリュームマネージャは、 `pvmove` コマンド。指定したLUNデバイス上のすべてのエクステントを新しいLUNに再配置します。古いLUNは退避後に削除できます。

運用の主なリスクは、古い未使用のLUNを構成から削除することです。FCゾーニングを変更したり、古いLUNデバイスを削除したりする場合は、十分に注意する必要があります。

Oracle自動ストレージ管理

Oracle ASMは、論理ボリュームマネージャとファイルシステムを組み合わせたものです。大まかに言えば、Oracle ASMはLUNの集まりを受け取り、それらを小さな割り当て単位に分割して、ASMディスクグループと呼ばれる単一のボリュームとして提供します。ASMには、冗長性レベルを設定してディスクグループをミラーリングする機能もあります。ボリュームは、ミラーリングされていない（外部冗長性）、ミラーリングされている（通常の冗長性）、または3方向ミラーリングされている（高冗長性）ことができます。冗長性レベルの設定は作成後に変更できないため、慎重に行う必要があります。

ASMは、ファイルシステム機能も提供します。ファイルシステムはホストから直接認識されませんが、Oracleデータベースではディスクグループ上のファイルやディレクトリを作成、移動、削除できます。また、asmcmdユーティリティを使用して構造体をナビゲートすることもできます。

他のLVM実装と同様に、Oracle ASMは、使用可能なすべてのLUNにわたって各ファイルのI/Oをストライピングおよびロードバランシングすることで、I/Oパフォーマンスを最適化します。次に、基盤となるエクステントを再配置して、ASMディスクグループのサイズ変更と移行の両方を可能にします。Oracle ASMは、リバランシング処理を通じてプロセスを自動化します。新しいLUNがASMディスクグループに追加され、古いLUNが削除されると、エクステントの再配置と、退避したLUNがディスクグループから削除されます。このプロセスは、最も実証された移行方法の1つであり、透過的な移行を提供するASMの信頼性は、ASMの最も重要な機能である可能性があります。

Oracle ASMのミラーリングレベルは固定されているため、mirrorおよびdemirror方式の移行では使用できません。

ストレージレベルの移行

ストレージレベルの移行とは、アプリケーションレベルとオペレーティングシステムレベルの両方を下回るレベルで移行を実行することを意味します。以前は、これはネットワークレベルでLUNをコピーする専用のデバイスを使用することを意味していましたが、現在ではこれらの機能はONTAPに標準で搭載されています。

SnapMirror

NetAppシステム間でのデータベースの移行は、ほとんどの場合、NetApp SnapMirrorデータレプリケーションソフトウェアを使用して実行されます。このプロセスでは、移動するボリュームのミラー関係を設定して同期を許可し、カットオーバー時間を待機します。到着すると、ソースデータベースがシャットダウンされ、最後のミラー更新が1回実行され、ミラーが解除されます。レプリカボリュームは、格納されているNFSファイルシステムディレクトリをマウントするか、格納されているLUNを検出してデータベースを開始することで、使用できる状態になります。

単一のONTAPクラスタ内でのボリュームの再配置は、移動とはみなされず、日常的な作業です。 `volume move` 操作。SnapMirrorは、クラスタ内でデータレプリケーションエンジンとして使用されます。このプロセ

スは完全に自動化されています。LUNマッピングやNFSエクスポート権限など、ボリュームの属性がボリューム自体と一緒に移動された場合に実行する追加の移行手順はありません。再配置では、ホストの処理が中断されません。場合によっては、再配置されたデータに可能な限り効率的にアクセスできるようにネットワークアクセスを更新する必要がありますが、これらのタスクも無停止で実行できます。

Foreign LUN Import（FLI）

FLIは、8.3以降を実行するData ONTAPシステムで既存のLUNを別のストレージアレイから移行できる機能です。手順はシンプルです。ONTAPシステムは、他のSANホストと同様に既存のストレージアレイにゾーニングされます。次に、Data ONTAPが必要な従来型LUNを制御し、基盤となるデータを移行します。また、インポートプロセスでは、データの移動時に新しいボリュームの効率化設定が使用されます。つまり、移動プロセス中にデータをインラインで圧縮したり重複排除したりできます。

Data ONTAP 8.3で初めて実装されたFLIでは、オフライン移行のみが可能でした。これは非常に高速な転送でしたが、移行が完了するまでLUNデータを使用できないことを意味していました。オンライン移行はData ONTAP 8.3.1で導入されました。このような移行では、転送プロセス中にONTAPがLUNデータを提供できるようになるため、システム停止を最小限に抑えることができます。ONTAP経由でLUNを使用するようにホストをゾーニングしている間、システムが短時間停止します。ただし、これらの変更が行われるとすぐに、データに再びアクセスでき、移行プロセス中も引き続きアクセスできます。

コピー処理が完了するまで読み取りI/OはONTAP経由でプロキシされ、書き込みI/Oは外部LUNとONTAP LUNの両方に同期的に書き込まれます。管理者が完全なカットオーバーを実行して外部LUNを解放し、書き込みをレプリケートしなくなるまで、2つのLUNコピーはこの方法で同期されます。

FLIはFCと連携するように設計されていますが、iSCSIに変更する必要がある場合は、移行の完了後に、移行したLUNをiSCSI LUNとして簡単に再マッピングできます。

FLIの機能の1つに、アライメントの自動検出と調整があります。アライメントという用語は、LUNデバイス上のパーティションを指します。パフォーマンスを最適化するには、I/Oが4Kブロックにアライメントされている必要があります。パーティションを4Kの倍数ではないオフセットに配置すると、パフォーマンスが低下します。

アライメントには、パーティションオフセット（ファイルシステムのブロックサイズ）を調整して修正できないもう1つの側面があります。たとえば、ZFSファイルシステムのデフォルトの内部ブロックサイズは512バイトです。AIXを使用しているお客様の中には、ブロックサイズが512バイトまたは1バイトのJFS2ファイルシステムを作成するケースもあります。ファイルシステムは4Kの境界にアライメントされていても、そのファイルシステム内に作成されたファイルはアライメントされず、パフォーマンスが低下します。

このような状況ではFLIを使用しないでください。移行後はデータにアクセスできますが、その結果、ファイルシステムのパフォーマンスが大幅に制限されます。一般的な原則として、ONTAPでランダムオーバーライトワークロードをサポートするファイルシステムでは、4Kブロックサイズを使用する必要があります。これは主に、データベースデータファイルやVDI環境などのワークロードに該当します。ブロックサイズは、関連するホストオペレーティングシステムコマンドを使用して特定できます。

たとえば、AIXでは、ブロックサイズを `lsfs -q`。Linuxの場合、 `xfs_info` および `tune2fs` 次の用途に使用できます。 `xfs` および `ext3/ext4`をクリックします。を使用 `zfs`コマンドは次のようになります。 `zdb -C`。

ブロックサイズを制御するパラメータは次のとおりです。 `ashift` 通常、デフォルト値は9です。これは2^9、つまり512バイトを意味します。最適なパフォーマンスを実現するには、 `ashift` 値は12（2^12=4K）である必要があります。この値はzpoolの作成時に設定され、変更することはできません。つまり、 `ashift` 12以外の場合は、新しく作成したzpoolにデータをコピーして移行する必要があります。

Oracle ASMには基本ブロックサイズはありません。唯一の要件は、ASMディスクを構築するパーティションが適切にアライメントされていることです。

7-Mode Transition Tool

7-Mode Transition Tool（7MTT）は、7-Modeの大規模な構成をONTAPに移行するための自動化ユーティリティです。データベースをご利用のお客様は、ストレージの設置面積全体を移動するのではなく、データベース単位で環境のデータベースを移行することが多いため、他の方法を簡単に見つけることができます。また、多くの場合、データベースは大規模なストレージ環境の一部にすぎません。そのため、データベースは多くの場合個別に移行され、その後7MTTを使用して残りの環境を移動できます。

複雑なデータベース環境に特化したストレージシステムを運用しているお客様は少なくありませんが、かなりの数のお客様がいらっしゃいます。これらの環境には、多数のボリュームやSnapshotのほか、エクスポート権限、LUNイニシエータグループ、ユーザ権限、Lightweight Directory Access Protocolの設定など、さまざまな設定の詳細が含まれている可能性があります。このような場合は、7MTTの自動化機能によって移動が簡易化されます。

7MTTは次の2つのモードのいずれかで動作します。

- コピーベースの移行（**CBT**）。 7MTTとCBTにより、新しい環境の既存の7-ModeシステムからSnapMirrorボリュームがセットアップされます。データの同期が完了すると、7MTTによってカットオーバープロセスがオーケストレーションされます。
- コピーフリーの移行（**CFT**）。 CFTを使用する7MTTは、既存の7-Modeディスクシェルフのインプレース変換に基づいています。データはコピーされず、既存のディスクシェルフは再利用できます。データ保護とStorage Efficiencyの既存の設定は維持されます。

これら2つのオプションの主な違いは、コピーフリーの移行はビッグバンアプローチであり、元の7-Mode HAペアに接続されているすべてのディスクシェルフを新しい環境に再配置する必要がある点です。シェルフのサブセットを移動するオプションはありません。コピーベースのアプローチでは、選択したボリュームを移動できます。また、ディスクシェルフを再ケーブル接続してメタデータを変換する際にも同様の接続が必要になるため、コピーフリーの移行ではカットオーバー時間が長くなる可能性があります。NetAppでは、現場での経験に基づき、ディスクシェルフの再配置と再接続には1時間、メタデータ変換には15分から2時間かかることを推奨しています。

Oracleデータファイルの移行

1つのコマンドで個 々 のOracleデータファイルを移動できます。

たとえば、次のコマンドはデータファイルIOPST.dbfをファイルシステムから移動します。 `/oradata2` ファイルシステムへ `/oradata3`。

```
SQL> alter database move datafile  '/oradata2/NTAP/IOPS002.dbf' to
'/oradata3/NTAP/IOPS002.dbf';
Database altered.
```

この方法でデータファイルを移動すると時間がかかることがありますが、通常はI/Oが十分に発生しないため、日常のデータベースワークロードを妨げることはありません。一方、ASMのリバランシングを使用した移行ははるかに高速ですが、データの移動中にデータベース全体の処理速度が低下するという代償があります。

データファイルの移動に要する時間は、テストデータファイルを作成して移動することで簡単に測定できま

す。操作の経過時間は、V$セッションデータに記録されます。

```
SQL> set linesize 300;
SQL> select elapsed_seconds||':'||message from v$session_longops;
ELAPSED_SECONDS||':'||MESSAGE
------------------------------------------------------------------------
---------------
351:Online data file move: data file 8: 22548578304 out of 22548578304
bytes done
SQL> select bytes / 1024 / 1024 /1024 as GB from dba_data_files where
FILE_ID = 8;
        GB
----------
        21
```

この例では、移動したファイルはデータファイル8です。データファイルのサイズは21GBで、移行に約6分かかりました。必要な時間は、ストレージシステムの機能、ストレージネットワーク、および移行時に発生する全体的なデータベースアクティビティによって異なります。

ログ配布による**Oracle**データベースの移行

ログ配布を使用した移行の目的は、元のデータファイルのコピーを新しい場所に作成し、変更を新しい環境に配布する方法を確立することです。

いったん確立されると、ログの送信と再生を自動化して、レプリカデータベースをソースとほぼ同期した状態に保つことができます。たとえば、（a）最新のログを新しい場所にコピーし、（b）15分ごとに再生するようにcronジョブをスケジュールできます。再生が必要なアーカイブログは15分以内であるため、カットオーバー時のシステム停止は最小限に抑えられます。

次に示す手順は、基本的にはデータベースのクローニング処理です。表示されるロジックは、NetApp SnapManager for Oracle（SMO）およびNetApp SnapCenter Oracleプラグインのエンジンと似ています。一部のお客様は、スクリプトまたはWFAワークフローに表示されている手順をカスタムクローニング処理に使用しています。この手順はSMOやSnapCenterを使用するよりも手動で作成する必要がありますが、スクリプト化も容易で、ONTAP内のデータ管理APIによってプロセスがさらに簡易化されます。

ログ配布-ファイルシステムからファイルシステムへ

この例では、Waffleというデータベースを通常のファイルシステムから別のサーバにある別の通常のファイルシステムに移行する方法を示します。また、SnapMirrorを使用してデータファイルの高速コピーを作成する方法も示していますが、これは手順全体に不可欠な要素ではありません。

データベースバックアップの作成

まず、データベースのバックアップを作成します。具体的には、この手順には、アーカイブログの再生に使用できる一連のデータファイルが必要です。

環境

この例では、ソースデータベースはONTAPシステム上にあります。データベースのバックアップを作成する

最も簡単な方法は、Snapshotを使用する方法です。データベースがホットバックアップモードになるまでの数秒間、 `snapshot create` この処理は、データファイルをホストしているボリュームで実行されます。

```
SQL> alter database begin backup;
Database altered.
```

```
Cluster01::*> snapshot create -vserver vserver1 -volume jfsc1_oradata
hotbackup
Cluster01::*>
```

```
SQL> alter database end backup;
Database altered.
```

その結果、という名前のディスク上のスナップショットが作成されます。 `hotbackup` ホットバックアップモード時のデータファイルのイメージを含むデータファイルを展開します。適切なアーカイブログと組み合わせてデータファイルの整合性を確保すると、このSnapshot内のデータをリストアまたはクローンのベースとして使用できます。この場合、新しいサーバに複製されます。

新しい環境へのリストア

これで、新しい環境でバックアップをリストアする必要があります。これは、Oracle RMAN、NetBackupなどのバックアップアプリケーションからのリストア、ホットバックアップモードに設定されたデータファイルの単純なコピー操作など、さまざまな方法で実行できます。

この例では、SnapMirrorを使用してSnapshotホットバックアップを新しい場所にレプリケートします。

1. Snapshotデータを受信する新しいボリュームを作成します。ミラーリングの初期化 `jfsc1_oradata` 終了： `vol_oradata`。

```
Cluster01::*> volume create -vserver vserver1 -volume vol_oradata
-aggregate data_01 -size 20g -state online -type DP -snapshot-policy
none -policy jfsc3
[Job 833] Job succeeded: Successful
```

```
Cluster01::*> snapmirror initialize -source-path vserver1:jfsc1_oradata
-destination-path vserver1:vol_oradata
Operation is queued: snapmirror initialize of destination
"vserver1:vol_oradata".
Cluster01::*> volume mount -vserver vserver1 -volume vol_oradata
-junction-path /vol_oradata
Cluster01::*>
```

2. SnapMirrorによって同期が完了したことを示す状態が設定されたら、目的のSnapshotに基づいてミラーを更新します。

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields state
source-path             destination-path        state
----------------------- ----------------------- ------------
vserver1:jfsc1_oradata vserver1:vol_oradata SnapMirrored
```

```
Cluster01::*> snapmirror update -destination-path vserver1:vol_oradata
-source-snapshot hotbackup
Operation is queued: snapmirror update of destination
"vserver1:vol_oradata".
```

3. 同期が正常に完了したかどうかは、 `newest-snapshot` フィールドを指定します。

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields newest-snapshot
source-path             destination-path        newest-snapshot
----------------------- ----------------------- ---------------
vserver1:jfsc1_oradata vserver1:vol_oradata hotbackup
```

4. その後、ミラーを壊すことができます。

```
Cluster01::> snapmirror break -destination-path vserver1:vol_oradata
Operation succeeded: snapmirror break for destination
"vserver1:vol_oradata".
Cluster01::>
```

5. 新しいファイルシステムをマウントします。ブロックベースのファイルシステムでは、使用するLVMによって正確な手順が異なります。FCゾーニングまたはiSCSI接続を設定する必要があります。LUNへの接続が確立されたら、Linuxなどのコマンド `pvscan` ASMで検出できるように設定する必要があるボリュームグループまたはLUNを検出する場合に、が必要になることがあります。

   この例では、シンプルなNFSファイルシステムを使用しています。このファイルシステムは直接マウントできます。

```
fas8060-nfs1:/vol_oradata        19922944   1639360   18283584   9%
/oradata
fas8060-nfs1:/vol_logs            9961472       128    9961344   1%
/logs
```

制御ファイル作成テンプレートの作成

次に、制御ファイルテンプレートを作成する必要があります。。 `backup controlfile to trace` コマンド制御ファイルを再作成するためのテキストコマンドを作成します。この機能は、状況によってはバックアップからデータベースをリストアする場合に役立ちます。また、データベースのクローニングなどのタスクを実行するスクリプトでよく使用されます。

1. 移行されたデータベースの制御ファイルを再作成するには、次のコマンドの出力を使用します。

```
SQL> alter database backup controlfile to trace as '/tmp/waffle.ctrl';
Database altered.
```

2. 制御ファイルが作成されたら、ファイルを新しいサーバにコピーします。

```
[oracle@jfsc3 tmp]$ scp oracle@jfsc1:/tmp/waffle.ctrl /tmp/
oracle@jfsc1's password:
waffle.ctrl                                           100% 5199
5.1KB/s   00:00
```

バックアップパラメータファイル

新しい環境ではパラメータファイルも必要です。最も簡単な方法は、現在のspfileまたはpfileからpfileを作成することです。この例では、ソースデータベースでspfileが使用されています。

```
SQL> create pfile='/tmp/waffle.tmp.pfile' from spfile;
File created.
```

oratabエントリの作成

oratabエントリの作成は、oraenvなどのユーティリティが適切に機能するために必要です。oratabエントリを作成するには、次の手順を実行します。

```
WAFFLE:/orabin/product/12.1.0/dbhome_1:N
```

ディレクトリ構造の準備

必要なディレクトリがまだ存在していない場合は、作成する必要があります。作成しないと、データベースの起動手順が失敗します。ディレクトリ構造を準備するには、次の最小要件を満たしている必要があります。

```
[oracle@jfsc3 ~]$ . oraenv
ORACLE_SID = [oracle] ? WAFFLE
The Oracle base has been set to /orabin
[oracle@jfsc3 ~]$ cd $ORACLE_BASE
[oracle@jfsc3 orabin]$ cd admin
[oracle@jfsc3 admin]$ mkdir WAFFLE
[oracle@jfsc3 admin]$ cd WAFFLE
[oracle@jfsc3 WAFFLE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

パラメータファイルの更新

1. パラメータファイルを新しいサーバにコピーするには、次のコマンドを実行します。デフォルトの場所は `$ORACLE_HOME/dbs` ディレクトリ。この場合、pfileは任意の場所に配置できます。これは、移行プロセスの中間ステップとしてのみ使用されます。

```
[oracle@jfsc3 admin]$ scp oracle@jfsc1:/tmp/waffle.tmp.pfile
$ORACLE_HOME/dbs/waffle.tmp.pfile
oracle@jfsc1's password:
waffle.pfile                                          100%  916
0.9KB/s   00:00
```

1. 必要に応じてファイルを編集します。たとえば、アーカイブログの場所が変更された場合は、新しい場所を反映するようにpfileを変更する必要があります。この例では、制御ファイルだけが再配置されています。その一部は、ログファイルシステムとデータファイルシステム間で制御ファイルを分散するためです。

```
[root@jfsc1 tmp]# cat waffle.pfile
WAFFLE.__data_transfer_cache_size=0
WAFFLE.__db_cache_size=507510784
WAFFLE.__java_pool_size=4194304
WAFFLE.__large_pool_size=20971520
WAFFLE.__oracle_base='/orabin'#ORACLE_BASE set from environment
WAFFLE.__pga_aggregate_target=268435456
WAFFLE.__sga_target=805306368
WAFFLE.__shared_io_pool_size=29360128
WAFFLE.__shared_pool_size=234881024
WAFFLE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/WAFFLE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/oradata//WAFFLE/control01.ctl','/oradata//WAFFLE/contr
ol02.ctl'
*.control_files='/oradata/WAFFLE/control01.ctl','/logs/WAFFLE/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='WAFFLE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=WAFFLEXDB)'
*.log_archive_dest_1='LOCATION=/logs/WAFFLE/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

2. 編集が完了したら、このpfileに基づいてspfileを作成します。

```
SQL> create spfile from pfile='waffle.tmp.pfile';
File created.
```

制御ファイルの再作成

前の手順では、 `backup controlfile to trace` が新しいサーバにコピーされました。必要な出力の具体的な部分は、 `controlfile recreation` コマンドを実行しますこの情報は、ファイルのマークされたセクションの下に記載されています。 `Set #1. NORESETLOGS`。次の行から始まります `create controlfile reuse database` 次の単語を含める必要があります。 `noresetlogs`。最後はセミコロン（;）文字です。

1. この手順の例では、ファイルは次のように表示されます。

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/logs/WAFFLE/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/logs/WAFFLE/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/logs/WAFFLE/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

2. このスクリプトを必要に応じて編集し、さまざまなファイルの新しい場所を反映します。たとえば、高I/Oをサポートすると認識されている特定のデータファイルは、ハイパフォーマンスストレージ階層上のファイルシステムにリダイレクトされる可能性があります。また、特定のPDBのデータファイルを専用ボリュームに分離するなど、管理者のみが変更を行う場合もあります。
3. この例では、を使用しています `DATAFILE` スタンザは変更されませんが、REDOログは `/redo` アーカイブログでスペースを共有する代わりに `/logs`。

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
SQL> CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
  2      MAXLOGFILES 16
  3      MAXLOGMEMBERS 3
  4      MAXDATAFILES 100
  5      MAXINSTANCES 8
  6      MAXLOGHISTORY 292
  7  LOGFILE
  8    GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  9    GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
 10    GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
 11  -- STANDBY LOGFILE
 12  DATAFILE
 13    '/oradata/WAFFLE/system01.dbf',
 14    '/oradata/WAFFLE/sysaux01.dbf',
 15    '/oradata/WAFFLE/undotbs01.dbf',
 16    '/oradata/WAFFLE/users01.dbf'
 17  CHARACTER SET WE8MSWIN1252
 18  ;
Control file created.
SQL>
```

ファイルが正しく配置されていない場合やパラメータが正しく設定されていない場合は、修正が必要な項目を示すエラーが生成されます。データベースはマウントされていますが、使用中のデータファイルがホットバックアップモードとしてマークされているため、まだ開いておらず、開くことができません。データベースの整合性を維持するには、まずアーカイブログを適用する必要があります。

初期ログレプリケーション

データファイルの整合性を確保するには、少なくとも1つのログ応答処理が必要です。ログの再生には、さまざまなオプションを使用できます。場合によっては、元のサーバ上の元のアーカイブログの場所をNFS経由で共有し、ログの返信を直接行うことができます。それ以外の場合は、アーカイブログをコピーする必要があります。

例えば、単純な `scp` この処理では、現在のすべてのログを移行元サーバから移行先サーバにコピーできます。

```
[oracle@jfsc3 arch]$ scp jfsc1:/logs/WAFFLE/arch/* ./
oracle@jfsc1's password:
1_22_912662036.dbf                                          100%   47MB
47.0MB/s   00:01
1_23_912662036.dbf                                          100%   40MB
40.4MB/s   00:00
1_24_912662036.dbf                                          100%   45MB
45.4MB/s   00:00
1_25_912662036.dbf                                          100%   41MB
40.9MB/s   00:01
1_26_912662036.dbf                                          100%   39MB
39.4MB/s   00:00
1_27_912662036.dbf                                          100%   39MB
38.7MB/s   00:00
1_28_912662036.dbf                                          100%   40MB
40.1MB/s   00:01
1_29_912662036.dbf                                          100%   17MB
16.9MB/s   00:00
1_30_912662036.dbf                                          100%  636KB
636.0KB/s   00:00
```

初回のログ再生

アーカイブログの場所に保存されたファイルは、コマンドを実行して再生できます。 `recover database until cancel` その後に応答が続きます `AUTO` 使用可能なすべてのログを自動的に再生します。

```
SQL> recover database until cancel;
ORA-00279: change 382713 generated at 05/24/2016 09:00:54 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_23_912662036.dbf
ORA-00280: change 382713 for thread 1 is in sequence #23
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 405712 generated at 05/24/2016 15:01:05 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_24_912662036.dbf
ORA-00280: change 405712 for thread 1 is in sequence #24
ORA-00278: log file '/logs/WAFFLE/arch/1_23_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
ORA-00278: log file '/logs/WAFFLE/arch/1_30_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_31_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

最後のアーカイブログの応答でエラーが報告されますが、これは正常な動作です。ログは次のことを示します。 `sqlplus` 特定のログファイルを探していましたが、見つかりませんでした。ログファイルがまだ存在しない可能性があります。

アーカイブログをコピーする前にソースデータベースをシャットダウンできる場合、この手順は1回だけ実行する必要があります。アーカイブログがコピーされて再生されたら、重要なRedoログをレプリケートするカットオーバープロセスに直接進むことができます。

差分ログのレプリケーションと再生

ほとんどの場合、移行はすぐには実行されません。移行プロセスが完了するまでに数日、場合によっては数週間かかることもあります。つまり、ログをレプリカデータベースに継続的に送信して再生する必要があります。そのため、カットオーバーが完了したら、最小限のデータを転送して再生する必要があります。

これはさまざまな方法でスクリプト化できますが、最も一般的な方法の1つは、一般的なファイルレプリケーションユーティリティであるrsyncを使用することです。このユーティリティを使用する最も安全な方法は、このユーティリティをデーモンとして設定することです。たとえば、などです `rsyncd.conf` 次のファイルは、という名前のリソースを作成する方法を示しています。 `waffle.arch` Oracleユーザクレデンシャルでアクセスされ、次にマッピングされます。 `/logs/WAFFLE/arch`。最も重要なことは、リソースが読み取り専用に設定されていることです。これにより、本番データの読み取りは可能ですが、変更はできません。

```
[root@jfsc1 arch]# cat /etc/rsyncd.conf
[waffle.arch]
   uid=oracle
   gid=dba
   path=/logs/WAFFLE/arch
   read only = true
[root@jfsc1 arch]# rsync --daemon
```

次のコマンドは'新しいサーバのアーカイブログデスティネーションをrsyncリソースと同期します `waffle.arch` 元のサーバ。。 `t` の引数 `rsync - potg` タイムスタンプに基づいてファイルリストが比較され、新しいファイルのみがコピーされます。このプロセスでは、新しいサーバの増分アップデートが提供されます。このコマンドは、cronで定期的に実行するようにスケジュールすることもできます。

```
[oracle@jfsc3 arch]$ rsync -potg --stats --progress jfsc1::waffle.arch/*
/logs/WAFFLE/arch/
1_31_912662036.dbf
      650240 100%  124.02MB/s    0:00:00 (xfer#1, to-check=8/18)
1_32_912662036.dbf
     4873728 100%  110.67MB/s    0:00:00 (xfer#2, to-check=7/18)
1_33_912662036.dbf
     4088832 100%   50.64MB/s    0:00:00 (xfer#3, to-check=6/18)
1_34_912662036.dbf
     8196096 100%   54.66MB/s    0:00:00 (xfer#4, to-check=5/18)
1_35_912662036.dbf
    19376128 100%   57.75MB/s    0:00:00 (xfer#5, to-check=4/18)
1_36_912662036.dbf
       71680 100%  201.15kB/s    0:00:00 (xfer#6, to-check=3/18)
1_37_912662036.dbf
     1144320 100%    3.06MB/s    0:00:00 (xfer#7, to-check=2/18)
1_38_912662036.dbf
    35757568 100%   63.74MB/s    0:00:00 (xfer#8, to-check=1/18)
1_39_912662036.dbf
      984576 100%    1.63MB/s    0:00:00 (xfer#9, to-check=0/18)
Number of files: 18
Number of files transferred: 9
Total file size: 399653376 bytes
Total transferred file size: 75143168 bytes
Literal data: 75143168 bytes
Matched data: 0 bytes
File list size: 474
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 204
Total bytes received: 75153219
sent 204 bytes  received 75153219 bytes  150306846.00 bytes/sec
total size is 399653376  speedup is 5.32
```

ログを受信したら、それらのログを再生する必要があります。上記の例では、sqlplusを使用して手動で `recover database until cancel`、簡単に自動化できるプロセス。この例では、で説明されているスクリプトを使用しています。 "データベースのログを再生"。スクリプトは、リプレイ操作を必要とするデータベースを指定する引数を受け入れます。これにより、同じスクリプトをマルチデータベース移行で使用できます。

```
[oracle@jfsc3 logs]$ ./replay.logs.pl WAFFLE
ORACLE_SID = [WAFFLE] ? The Oracle base remains unchanged with value
/orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu May 26 10:47:16 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 814256 generated at 05/26/2016 04:52:30 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_32_912662036.dbf
ORA-00280: change 814256 for thread 1 is in sequence #32
ORA-00278: log file '/logs/WAFFLE/arch/1_31_912662036.dbf' no longer
needed for
this recovery
ORA-00279: change 814780 generated at 05/26/2016 04:53:04 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_33_912662036.dbf
ORA-00280: change 814780 for thread 1 is in sequence #33
ORA-00278: log file '/logs/WAFFLE/arch/1_32_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
ORA-00278: log file '/logs/WAFFLE/arch/1_39_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

カットオーバー

新しい環境にカットオーバーする準備ができたら、アーカイブログとREDOログの両方を含む最終的な同期を実行する必要があります。元のREDOログの場所が不明な場合は、次のように特定できます。

```
SQL> select member from v$logfile;
MEMBER
--------------------------------------------------------------------------------
------
/logs/WAFFLE/redo/redo01.log
/logs/WAFFLE/redo/redo02.log
/logs/WAFFLE/redo/redo03.log
```

1. ソースデータベースをシャットダウンします。
2. 目的の方法を使用して、新しいサーバでアーカイブログの最終的な同期を1回実行します。
3. ソースREDOログを新しいサーバにコピーする必要があります。この例では、REDOログがの新しいディレクトリに再配置されています。 `/redo`。

   ```
   [oracle@jfsc3 logs]$ scp jfsc1:/logs/WAFFLE/redo/* /redo/
   oracle@jfsc1's password:
   redo01.log
   100%   50MB  50.0MB/s   00:01
   redo02.log
   100%   50MB  50.0MB/s   00:00
   redo03.log
   100%   50MB  50.0MB/s   00:00
   ```

4. この段階で、新しいデータベース環境には、ソースとまったく同じ状態にするために必要なすべてのファイルが含まれています。アーカイブログは最後に1回再生する必要があります。

```
SQL> recover database until cancel;
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

5. 完了したら、Redoログを再生する必要があります。というメッセージが表示されます `Media recovery complete` が返されると、プロセスが成功し、データベースが同期されてオープンできるようになります。

```
SQL> recover database;
Media recovery complete.
SQL> alter database open;
Database altered.
```

ログ配布**- ASM**からファイルシステムへ

この例では、Oracle RMANを使用してデータベースを移行します。ファイルシステムからファイルシステムへのログ配布の前の例と非常によく似ていますが、ASM上のファイルはホストには表示されません。ASMデバイス上にあるデータを移行するには、ASM LUNを再配置するか、Oracle RMANを使用してコピー処理を実行するしかありません。

Oracle ASMからファイルをコピーするにはRMANが必要ですが、RMANを使用できるのはASMに限られません。RMANを使用すると、任意のタイプのストレージから他のタイプのストレージに移行できます。

この例は'pancakeというデータベースをASMストレージから'パスにある別のサーバにある通常のファイルシステムに再配置する例を示しています `/oradata` および `/logs`。

データベースバックアップの作成

最初の手順では、代替サーバに移行するデータベースのバックアップを作成します。ソースではOracle ASMを使用するため、RMANを使用する必要があります。単純なRMANバックアップは、次のように実行できます。この方法で作成されるタグ付きバックアップは、あとでRMANで簡単に識別できるように手順なります。

最初のコマンドは、バックアップ先のタイプと使用する場所を定義します。2番目のコマンドでは、データファイルのみのバックアップが開始されます。

```
RMAN> configure channel device type disk format '/rman/pancake/%U';
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters are successfully stored
RMAN> backup database tag 'ONTAP_MIGRATION';
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=251 device type=DISK
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/PANCAKE/system01.dbf
input datafile file number=00002 name=+ASM0/PANCAKE/sysaux01.dbf
input datafile file number=00003 name=+ASM0/PANCAKE/undotbs101.dbf
input datafile file number=00004 name=+ASM0/PANCAKE/users01.dbf
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1gr6c161_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:03
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1hr6c164_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

バックアップ制御ファイルバックアップセイギョファイル

バックアップ制御ファイルは、手順の後半の工程で `duplicate database` 操作。

```
RMAN> backup current controlfile format '/rman/pancake/ctrl.bkp';
Starting backup at 24-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/ctrl.bkp tag=TAG20160524T032651 comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

バックアップパラメータファイル

新しい環境ではパラメータファイルも必要です。最も簡単な方法は、現在のspfileまたはpfileからpfileを作成することです。この例では、ソースデータベースでspfileが使用されています。

```
RMAN> create pfile='/rman/pancake/pfile' from spfile;
Statement processed
```

ASMファイル名変更スクリプト

データベースを移動すると、制御ファイルに現在定義されている複数のファイルの場所が変更されます。次のスクリプトは、プロセスを簡単にするためにRMANスクリプトを作成します。この例は、データファイルの数が非常に少ないデータベースを示していますが、通常、データベースには数百、場合によっては数千のデータファイルが含まれています。

このスクリプトは、 "ASMからファイルシステム名への変換" 2つのことができます

まず、REDOログの場所を再定義するパラメータを作成します。 `log_file_name_convert`。基本的には交互のフィールドのリストです。最初のフィールドは現在のREDOログの場所で、2番目のフィールドは新しいサーバ上の場所です。その後、パターンが繰り返されます。

2つ目の機能は、データファイルの名前を変更するためのテンプレートを提供することです。スクリプトは、データファイルをループ処理し、名前とファイル番号の情報を取得して、RMANスクリプトとしてフォーマットします。次に、一時ファイルについても同じことが行われます。その結果、必要に応じて編集してファイルを目的の場所にリストアできるシンプルなRMANスクリプトが作成されます。

```
SQL> @/rman/mk.rename.scripts.sql
Parameters for log file conversion:
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/NEW_PATH/redo01.log','+ASM0/PANCAKE/redo02.log',
'/NEW_PATH/redo02.log','+ASM0/PANCAKE/redo03.log', '/NEW_PATH/redo03.log'
rman duplication script:
run
{
set newname for datafile 1 to '+ASM0/PANCAKE/system01.dbf';
set newname for datafile 2 to '+ASM0/PANCAKE/sysaux01.dbf';
set newname for datafile 3 to '+ASM0/PANCAKE/undotbs101.dbf';
set newname for datafile 4 to '+ASM0/PANCAKE/users01.dbf';
set newname for tempfile 1 to '+ASM0/PANCAKE/temp01.dbf';
duplicate target database for standby backup location INSERT_PATH_HERE;
}
PL/SQL procedure successfully completed.
```

この画面の出力をキャプチャします。。 `log_file_name_convert` パラメータは、次のようにpfileに配置されます。RMANデータ・ファイルの名前変更および複製スクリプトを編集して、必要な場所にデータ・ファイルを配置する必要があります。この例では、これらはすべて `/oradata/pancake`。

```
run
{
set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
set newname for datafile 4 to '/oradata/pancake/users.dbf';
set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
duplicate target database for standby backup location '/rman/pancake';
}
```

ディレクトリ構造の準備

スクリプトの実行準備はほぼ完了していますが、最初にディレクトリ構造を設定する必要があります。必要なディレクトリが存在しない場合は、それらのディレクトリを作成する必要があります。存在しないと、データベースの起動手順が失敗します。次の例は、最小要件を示しています。

```
[oracle@jfsc2 ~]$ mkdir /oradata/pancake
[oracle@jfsc2 ~]$ mkdir /logs/pancake
[oracle@jfsc2 ~]$ cd /orabin/admin
[oracle@jfsc2 admin]$ mkdir PANCAKE
[oracle@jfsc2 admin]$ cd PANCAKE
[oracle@jfsc2 PANCAKE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

oratabエントリの作成

次のコマンドは、oraenvなどのユーティリティが正常に動作するために必要です。

```
PANCAKE:/orabin/product/12.1.0/dbhome_1:N
```

パラメータの更新

保存したpfileを更新して、新しいサーバ上のパスの変更を反映する必要があります。データ・ファイル・パスの変更は、RMAN複製スクリプトによって変更されます。ほとんどのデータベースでは、 `control_files` および `log_archive_dest` パラメータ変更が必要な監査ファイルの場所や、次のようなパラメータが存在する場合もあります。 `db_create_file_dest` ASM以外では関連性がない可能性があります。経験豊富なデータベース管理者は、次に進む前に提案された変更を慎重に確認する必要があります。

この例では、制御ファイルの場所、ログのアーカイブ先、 `log_file_name_convert` パラメータ

```
PANCAKE.__data_transfer_cache_size=0
PANCAKE.__db_cache_size=545259520
PANCAKE.__java_pool_size=4194304
PANCAKE.__large_pool_size=25165824
PANCAKE.__oracle_base='/orabin'#ORACLE_BASE set from environment
PANCAKE.__pga_aggregate_target=268435456
PANCAKE.__sga_target=805306368
PANCAKE.__shared_io_pool_size=29360128
PANCAKE.__shared_pool_size=192937984
PANCAKE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/PANCAKE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='+ASM0/PANCAKE/control01.ctl','+ASM0/PANCAKE/control02.ctl
'
*.control_files='/oradata/pancake/control01.ctl','/logs/pancake/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='PANCAKE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=PANCAKEXDB)'
*.log_archive_dest_1='LOCATION=+ASM1'
*.log_archive_dest_1='LOCATION=/logs/pancake'
*.log_archive_format='%t_%s_%r.dbf'
'/logs/path/redo02.log'
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/logs/pancake/redo01.log', '+ASM0/PANCAKE/redo02.log',
'/logs/pancake/redo02.log', '+ASM0/PANCAKE/redo03.log',
'/logs/pancake/redo03.log'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

新しいパラメータが確認されたら、パラメータを有効にする必要があります。複数のオプションがありますが、ほとんどのお客様はテキストpfileに基づいてspfileを作成します。

```
bash-4.1$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0 Production on Fri Jan 8 11:17:40 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> create spfile from pfile='/rman/pancake/pfile';
File created.
```

スタートアップの登録

データベースをレプリケートする前の最後の手順では、データベースプロセスを起動しますが、ファイルはマウントしません。この手順では、spfileの問題が明らかになる可能性があります。状況に応じて `startup nomount` パラメータエラーが原因でコマンドが失敗します。シャットダウンし、pfileテンプレートを修正し、spfileとしてリロードして、再試行するのは簡単です。

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             373296240 bytes
Database Buffers          423624704 bytes
Redo Buffers                5455872 bytes
```

データベースの複製

以前のRMANバックアップを新しい場所にリストアするには、このプロセスの他の手順よりも時間がかかります。データベースID（DBID）を変更したり、ログをリセットしたりせずに、データベースを複製する必要があります。これにより、ログが適用されなくなります。これは、コピーを完全に同期するために必要な手順です。

前の手順で作成したスクリプトを使用して、RMANをauxとしてデータベースに接続し、DUPLICATE DATABASEコマンドを問題します。

```
[oracle@jfsc2 pancake]$ rman auxiliary /
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 03:04:56
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to auxiliary database: PANCAKE (not mounted)
RMAN> run
2> {
3> set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
4> set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
5> set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
6> set newname for datafile 4 to '/oradata/pancake/users.dbf';
7> set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
```

```
8> duplicate target database for standby backup location '/rman/pancake';
9> }
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting Duplicate Db at 24-MAY-16
contents of Memory Script:
{
   restore clone standby controlfile from  '/rman/pancake/ctrl.bkp';
}
executing Memory Script
Starting restore at 24-MAY-16
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
channel ORA_AUX_DISK_1: restoring control file
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:01
output file name=/oradata/pancake/control01.ctl
output file name=/logs/pancake/control02.ctl
Finished restore at 24-MAY-16
contents of Memory Script:
{
   sql clone 'alter database mount standby database';
}
executing Memory Script
sql statement: alter database mount standby database
released channel: ORA_AUX_DISK_1
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
contents of Memory Script:
{
   set newname for tempfile  1 to
 "/oradata/pancake/temp.dbf";
   switch clone tempfile all;
   set newname for datafile  1 to
 "/oradata/pancake/pancake.dbf";
   set newname for datafile  2 to
 "/oradata/pancake/sysaux.dbf";
   set newname for datafile  3 to
 "/oradata/pancake/undotbs1.dbf";
   set newname for datafile  4 to
 "/oradata/pancake/users.dbf";
   restore
   clone database
   ;
```

```
}
executing Memory Script
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/pancake/temp.dbf in control file
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting restore at 24-MAY-16
using channel ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: starting datafile backup set restore
channel ORA_AUX_DISK_1: specifying datafile(s) to restore from backup set
channel ORA_AUX_DISK_1: restoring datafile 00001 to
/oradata/pancake/pancake.dbf
channel ORA_AUX_DISK_1: restoring datafile 00002 to
/oradata/pancake/sysaux.dbf
channel ORA_AUX_DISK_1: restoring datafile 00003 to
/oradata/pancake/undotbs1.dbf
channel ORA_AUX_DISK_1: restoring datafile 00004 to
/oradata/pancake/users.dbf
channel ORA_AUX_DISK_1: reading from backup piece
/rman/pancake/1gr6c161_1_1
channel ORA_AUX_DISK_1: piece handle=/rman/pancake/1gr6c161_1_1
tag=ONTAP_MIGRATION
channel ORA_AUX_DISK_1: restored backup piece 1
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:07
Finished restore at 24-MAY-16
contents of Memory Script:
{
   switch clone datafile all;
}
executing Memory Script
datafile 1 switched to datafile copy
input datafile copy RECID=5 STAMP=912655725 file
name=/oradata/pancake/pancake.dbf
datafile 2 switched to datafile copy
input datafile copy RECID=6 STAMP=912655725 file
name=/oradata/pancake/sysaux.dbf
datafile 3 switched to datafile copy
input datafile copy RECID=7 STAMP=912655725 file
name=/oradata/pancake/undotbs1.dbf
datafile 4 switched to datafile copy
input datafile copy RECID=8 STAMP=912655725 file
name=/oradata/pancake/users.dbf
Finished Duplicate Db at 24-MAY-16
```

初期ログレプリケーション

ソースデータベースから新しい場所に変更を出荷する必要があります。そのためには、いくつかの手順が必要になる場合があります。最も簡単な方法は、ソース・データベースのRMANでアーカイブ・ログを共有ネットワーク接続に書き込む方法です。共有の場所を使用できない場合は、RMANを使用してローカルファイルシステムに書き込み、rcpまたはrsyncを使用してファイルをコピーする方法もあります。

この例では、を使用しています `/rman` ディレクトリは、元のデータベースと移行後のデータベースの両方で使用できるNFS共有です。

ここでの重要な問題の1つは、 `disk format` 条項。バックアップのディスクフォーマットは次のとおりです。 `%h_%e_%a.dbf`これは、スレッド番号、シーケンス番号、およびデータベースのアクティベーションIDの形式を使用する必要があることを意味します。文字は異なりますが、これは `log_archive_format='%t_%s_%r.dbf` パラメータをpfileに指定します。このパラメータは、スレッド番号、シーケンス番号、およびアクティベーションIDの形式でアーカイブログを指定します。最終的に、ソース上のログファイルのバックアップでは、データベースで想定される命名規則が使用されます。これにより、次のような操作が行われます。 `recover database` sqlplusはアーカイブログの名前を正しく予測して再生できるため、はるかにシンプルです。

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/arch/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
released channel: ORA_DISK_1
RMAN> backup as copy archivelog from time 'sysdate-2';
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=373 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=70 STAMP=912658508
output file name=/rman/pancake/logship/1_54_912576125.dbf RECID=123
STAMP=912659482
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=29 STAMP=912654101
output file name=/rman/pancake/logship/1_41_912576125.dbf RECID=124
STAMP=912659483
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=33 STAMP=912654688
output file name=/rman/pancake/logship/1_45_912576125.dbf RECID=152
STAMP=912659514
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=36 STAMP=912654809
output file name=/rman/pancake/logship/1_47_912576125.dbf RECID=153
STAMP=912659515
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

初回のログ再生

アーカイブログの場所に保存されたファイルは、コマンドを実行して再生できます。 `recover database until cancel` その後に応答が続きます `AUTO` 使用可能なすべてのログを自動的に再生します。パラメータファイルは現在、アーカイブログを次の場所に転送しています：`/logs/archive`ただし、これは、RMANを使用してログを保存した場所と一致しません。この場所は、データベースをリカバリする前に、次のように一時的にリダイレクトできます。

```
SQL> alter system set log_archive_dest_1='LOCATION=/rman/pancake/logship'
scope=memory;
System altered.
SQL> recover standby database until cancel;
ORA-00279: change 560224 generated at 05/24/2016 03:25:53 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_49_912576125.dbf
ORA-00280: change 560224 for thread 1 is in sequence #49
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 560353 generated at 05/24/2016 03:29:17 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_50_912576125.dbf
ORA-00280: change 560353 for thread 1 is in sequence #50
ORA-00278: log file '/rman/pancake/logship/1_49_912576125.dbf' no longer
needed
for this recovery
...
ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
ORA-00278: log file '/rman/pancake/logship/1_53_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_54_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

最後のアーカイブログの応答でエラーが報告されますが、これは正常な動作です。エラーは、sqlplusが特定のログファイルを探していたが見つからなかったことを示しています。ログファイルがまだ存在しない可能性があります。

アーカイブログをコピーする前にソースデータベースをシャットダウンできる場合、この手順は1回だけ実行する必要があります。アーカイブログがコピーされて再生されたら、重要なRedoログをレプリケートするカットオーバープロセスに直接進むことができます。

差分ログのレプリケーションと再生

ほとんどの場合、移行はすぐには実行されません。移行プロセスが完了するまでに数日、場合によっては数週間かかることもあります。つまり、ログをレプリカデータベースに継続的に送信して再生する必要があります。これにより、カットオーバーの到着時に最小限のデータの転送と再生が必要になります。

このプロセスは簡単にスクリプト化できます。たとえば、次のコマンドを元のデータベースでスケジュールして、ログ配布に使用される場所が継続的に更新されるようにすることができます。

```
[oracle@jfsc1 pancake]$ cat copylogs.rman
configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
backup as copy archivelog from time 'sysdate-2';
```

```
[oracle@jfsc1 pancake]$ rman target / cmdfile=copylogs.rman
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 04:36:19
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: PANCAKE (DBID=3574534589)
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=369 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:22
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=124 STAMP=912659483
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:23
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_41_912576125.dbf
continuing other job steps, job failed will not be re-run
...
channel ORA_DISK_1: starting archived log copy
```

```
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:55
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:57
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
Recovery Manager complete.
```

ログを受信したら、それらのログを再生する必要があります。上記の例では、sqlplusを使用して手動で`recover database until cancel`をクリックします。これは簡単に自動化できます。この例では、で説明されているスクリプトを使用しています。 "スタンバイデータベースのリプレイログ"。スクリプトは、リプレイ操作を必要とするデータベースを指定する引数を受け取ります。このプロセスでは、同じスクリプトをマルチデータベース移行で使用できます。

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 04:47:10 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 562219 generated at 05/24/2016 04:15:08 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_55_912576125.dbf
ORA-00280: change 562219 for thread 1 is in sequence #55
ORA-00278: log file '/rman/pancake/logship/1_54_912576125.dbf' no longer
needed for this recovery
ORA-00279: change 562370 generated at 05/24/2016 04:19:18 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_56_912576125.dbf
ORA-00280: change 562370 for thread 1 is in sequence #56
ORA-00278: log file '/rman/pancake/logship/1_55_912576125.dbf' no longer
needed for this recovery
...
ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
ORA-00278: log file '/rman/pancake/logship/1_64_912576125.dbf' no longer
needed for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_65_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

カットオーバー

新しい環境にカットオーバーする準備ができたら、最後の同期を1回実行する必要があります。通常のファイルシステムを使用する場合は、元のREDOログがコピーされて再生されるため、移行したデータベースが元のデータベースと完全に同期されていることを簡単に確認できます。ASMでこれを行う良い方法はありません。簡単に再コピーできるのはアーカイブログだけです。データが失われないようにするには、元のデータベースの最終的なシャットダウンを慎重に実行する必要があります。

1. まず、データベースを休止して、変更が行われていないことを確認する必要があります。この休止には、スケジュールされた処理の無効化、リスナーのシャットダウン、アプリケーションのシャットダウンなどが含まれます。
2. この手順を実行すると、ほとんどのDBAはダミーテーブルを作成し、シャットダウンのマーカーとして機能します。
3. ログを強制的にアーカイブし、ダミーテーブルの作成がアーカイブログに記録されるようにします。これを行うには、次のコマンドを実行します。

```
SQL> create table cutovercheck as select * from dba_users;
Table created.
SQL> alter system archive log current;
System altered.
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
```

4. 最後のアーカイブログをコピーするには、次のコマンドを実行します。データベースは使用可能であるが、開いていない必要があります。

```
SQL> startup mount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
Database mounted.
```

5. アーカイブログをコピーするには、次のコマンドを実行します。

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=8 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:24
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:58
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:59:00
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
```

6. 最後に、残りのアーカイブログを新しいサーバで再生します。

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 05:00:53 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed
for thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 563629 generated at 05/24/2016 04:55:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_66_912576125.dbf
ORA-00280: change 563629 for thread 1 is in sequence #66
ORA-00278: log file '/rman/pancake/logship/1_65_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_66_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

7. この段階では、すべてのデータをレプリケートします。データベースをスタンバイデータベースからアクティブ運用データベースに変換してオープンする準備が整いました。

```
SQL> alter database activate standby database;
Database altered.
SQL> alter database open;
Database altered.
```

8. ダミーテーブルの存在を確認してからドロップします。

```
SQL> desc cutovercheck
 Name                                      Null?    Type
 ----------------------------------------- --------
---------------------------
 USERNAME                                  NOT NULL VARCHAR2(128)
 USER_ID                                   NOT NULL NUMBER
 PASSWORD                                           VARCHAR2(4000)
 ACCOUNT_STATUS                            NOT NULL VARCHAR2(32)
 LOCK_DATE                                          DATE
 EXPIRY_DATE                                        DATE
 DEFAULT_TABLESPACE                        NOT NULL VARCHAR2(30)
 TEMPORARY_TABLESPACE                      NOT NULL VARCHAR2(30)
 CREATED                                   NOT NULL DATE
 PROFILE                                   NOT NULL VARCHAR2(128)
 INITIAL_RSRC_CONSUMER_GROUP                        VARCHAR2(128)
 EXTERNAL_NAME                                      VARCHAR2(4000)
 PASSWORD_VERSIONS                                  VARCHAR2(12)
 EDITIONS_ENABLED                                   VARCHAR2(1)
 AUTHENTICATION_TYPE                                VARCHAR2(8)
 PROXY_ONLY_CONNECT                                 VARCHAR2(1)
 COMMON                                             VARCHAR2(3)
 LAST_LOGIN                                         TIMESTAMP(9) WITH
TIME ZONE
 ORACLE_MAINTAINED                                  VARCHAR2(1)
SQL> drop table cutovercheck;
Table dropped.
```

Redoログの無停止移行

REDOログを除き、データベース全体が正しく構成されている場合があります。これはさまざまな理由で発生する可能性がありますが、最も一般的なのはスナップショットに関連しています。SnapManager for Oracle、SnapCenter、NetApp Snap Creatorのストレージ管理フレームワークなどの製品では、データファイルボリュームの状態をリバートする場合にのみ、データベースをほぼ瞬時にリカバリできます。REDOログがデータファイルとスペースを共有している場合は、REDOログが破棄されてデータが失われる可能性があるため、リバートを安全に実行できません。そのため、REDOログを再配置する必要があります。

この手順はシンプルで、無停止で実行できます。

現在のREDOログ設定

1. REDOロググループの数とそれぞれのグループ番号を確認します。

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
--------------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
rows selected.
```

2. Redoログのサイズを入力します。

```
SQL> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
--------------------------------------------------------------------------------
--------
1 524288000
2 524288000
3 524288000
```

新しいログを作成する

1. Redoログごとに、サイズとメンバー数が一致する新しいグループを作成します。

```
SQL> alter database add logfile ('/newredo0/redo01a.log',
'/newredo1/redo01b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo02a.log',
'/newredo1/redo02b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo03a.log',
'/newredo1/redo03b.log') size 500M;
Database altered.
SQL>
```

2. 新しい設定を確認します。

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
--------------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
4 /newredo0/redo01a.log
4 /newredo1/redo01b.log
5 /newredo0/redo02a.log
5 /newredo1/redo02b.log
6 /newredo0/redo03a.log
6 /newredo1/redo03b.log
12 rows selected.
```

古いログを削除

1. 古いログ（グループ1、2、3）を削除します。

```
SQL> alter database drop logfile group 1;
Database altered.
SQL> alter database drop logfile group 2;
Database altered.
SQL> alter database drop logfile group 3;
Database altered.
```

2. アクティブなログをドロップできないエラーが発生した場合は、次のログに切り替えてロックを解除し、グローバルチェックポイントを強制的に実行します。このプロセスの次の例を参照してください。古い場所にあるログファイルグループ2を削除しようとしましたが、このログファイルにアクティブなデータが残っているため拒否されました。

```
SQL> alter database drop logfile group 2;
alter database drop logfile group 2
*
ERROR at line 1:
ORA-01623: log 2 is current log for instance NTAP (thread 1) - cannot
drop
ORA-00312: online log 2 thread 1: '/redo0/NTAP/redo02a.log'
ORA-00312: online log 2 thread 1: '/redo1/NTAP/redo02b.log'
```

3. ログアーカイブの後にチェックポイントを追加すると、ログファイルをドロップできます。

```
SQL> alter system archive log current;
System altered.
SQL> alter system checkpoint;
System altered.
SQL> alter database drop logfile group 2;
Database altered.
```

4. 次に、ファイルシステムからログを削除します。このプロセスは細心の注意を払って実行する必要があります。

Oracleデータベースホストのデータコピー

データベースレベルの移行と同様に、ホストレイヤでの移行では、ストレージベンダーに依存しないアプローチが提供されます。

言い換えれば、いつか「ファイルをコピーするだけ」が最良のオプションです。

このローテクなアプローチは基本的すぎるように思われるかもしれませんが、特別なソフトウェアは必要なく、プロセス中に元のデータに安全に触れることができないため、大きな利点があります。主な制限事項は、ファイルコピーデータの移行はシステムの停止を伴うプロセスであることです。これは、コピー処理を開始する前にデータベースをシャットダウンする必要があるためです。ファイル内の変更を同期する適切な方法はないため、コピーを開始する前にファイルを完全に休止する必要があります。

コピー処理で必要なシャットダウンが望ましくない場合、次に推奨されるホストベースのオプションは論理ボリュームマネージャ（LVM）を利用することです。Oracle ASMを含む多くのLVMオプションは、すべて同様の機能を備えていますが、いくつかの制限事項を考慮する必要があります。ほとんどの場合、移行はダウンタイムやシステム停止なしで完了します。

ファイルシステムからファイルシステムへのコピー

単純なコピー操作の有用性を過小評価してはなりません。この処理はコピープロセス中のダウンタイムを必要としますが、信頼性の高いプロセスであり、オペレーティングシステム、データベース、ストレージシステムに関する特別な専門知識は必要ありません。さらに、元のデータに影響を与えないため、非常に安全です。通常'システム管理者は'ソース・ファイル・システムを読み取り専用としてマウントするように変更してから'サーバを再起動して'現在のデータに損傷を与えないようにしますコピープロセスをスクリプト化して、ユーザーエラーのリスクなしにできるだけ迅速に実行できるようにすることができます。I/Oのタイプはデータの単純なシーケンシャル転送であるため、帯域幅効率に優れています。

次の例は、安全かつ迅速な移行のための1つのオプションを示しています。

環境

移行する環境は次のとおりです。

- 現在のファイルシステム

```
ontap-nfs1:/host1_oradata       52428800  16196928  36231872  31%
/oradata
ontap-nfs1:/host1_logs          49807360    548032  49259328   2% /logs
```

- 新しいファイルシステム

```
ontap-nfs1:/host1_logs_new      49807360       128  49807232   1%
/new/logs
ontap-nfs1:/host1_oradata_new   49807360       128  49807232   1%
/new/oradata
```

概要

データベースは、データベースをシャットダウンしてファイルをコピーするだけで移行できますが、多数のデータベースを移行する必要がある場合や、ダウンタイムを最小限に抑えることが重要な場合は、プロセスを簡単にスクリプト化できます。スクリプトを使用すると、ユーザエラーの可能性も低くなります。

このスクリプトの例では、次の処理が自動化されています。

- データベースのシャットダウン
- 既存のファイルシステムの読み取り専用状態への変換
- ソース・ファイル・システムからターゲット・ファイル・システムへのすべてのデータのコピー（すべてのファイル権限を保持）
- 古いファイルシステムと新しいファイルシステムのアンマウント
- 以前のファイルシステムと同じパスでの新しいファイルシステムの再マウント

手順

1. データベースをシャットダウンします。

```
[root@host1 current]# ./dbshut.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 15:58:48 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
```

2. ファイルシステムを読み取り専用に変換します。スクリプトを使用すると、に示すように、この処理をより迅速に実行できます。 "ファイルシステムを読み取り専用に変換"。

```
[root@host1 current]# ./mk.fs.readonly.pl /oradata
/oradata unmounted
/oradata mounted read-only
[root@host1 current]# ./mk.fs.readonly.pl /logs
/logs unmounted
/logs mounted read-only
```

3. ファイルシステムが読み取り専用になったことを確認します。

```
ontap-nfs1:/host1_oradata on /oradata type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_logs on /logs type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

4. ファイルシステムの内容を `rsync` コマンドを実行します

```
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /oradata/ /new/oradata/
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
```

```
 10737426432 100%  153.50MB/s    0:01:06 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%   12.09MB/s    0:00:01 (xfer#2, to-check=9/13)
...
NTAP/undotbs02.dbf
  1073750016 100%  131.60MB/s    0:00:07 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    3.95MB/s    0:00:01 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  162204017.96 bytes/sec
total size is 18570092218  speedup is 1.00
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /logs/ /new/logs/
sending incremental file list
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%   95.98MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   49.45MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   44.68MB/s    0:00:01 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   68.03MB/s    0:00:00 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
```

```
sent 527098156 bytes  received 278 bytes  95836078.91 bytes/sec
total size is 527032832  speedup is 1.00
```

5. 古いファイルシステムをアンマウントし、コピーしたデータを再配置します。スクリプトを使用すると、に示すように、この処理をより迅速に実行できます。 "ファイルシステムの置き換え"。

```
[root@host1 current]# ./swap.fs.pl /logs,/new/logs
/new/logs unmounted
/logs unmounted
Updated /logs mounted
[root@host1 current]# ./swap.fs.pl /oradata,/new/oradata
/new/oradata unmounted
/oradata unmounted
Updated /oradata mounted
```

6. 新しいファイルシステムが所定の位置にあることを確認します。

```
ontap-nfs1:/host1_logs_new on /logs type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_oradata_new on /oradata type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

7. データベースを起動します。

```
[root@host1 current]# ./dbstart.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 16:10:07 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
```

カットオーバーを完全に自動化

このサンプルスクリプトでは、データベースSIDの引数に続いて、共通区切りのファイルシステムペアを指定します。上記の例では、コマンドは次のように実行されます。

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
```

このサンプルスクリプトを実行すると、次のシーケンスが試行されます。いずれかの手順でエラーが発生すると終了します。

1. データベースをシャットダウンします。
2. 現在のファイルシステムを読み取り専用ステータスに変換します。
3. カンマで区切られた各ファイルシステム引数のペアを使用し、最初のファイルシステムを2番目のファイルシステムに同期します。
4. 以前のファイルシステムをディスマウントします。
5. を更新します `/etc/fstab` ファイルは次のとおりです。
   a. バックアップの作成場所 `/etc/fstab.bak`。
   b. 以前のファイルシステムと新しいファイルシステムの前のエントリをコメントアウトします。
   c. 古いマウントポイントを使用する新しいファイルシステム用の新しいエントリを作成します。
6. ファイルシステムをマウントします。
7. データベースを起動します。

次のテキストは、このスクリプトの実行例を示しています。

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:05:50 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
```

```
sending incremental file list
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%  185.40MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   81.34MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   70.42MB/s    0:00:00 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   47.08MB/s    0:00:01 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
sent 527098156 bytes  received 278 bytes  150599552.57 bytes/sec
total size is 527032832  speedup is 1.00
Succesfully replicated filesystem /logs to /new/logs
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
 10737426432 100%  176.55MB/s    0:00:58 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%    9.48MB/s    0:00:02 (xfer#2, to-check=9/13)
... NTAP/undotbs01.dbf
   309338112 100%   70.76MB/s    0:00:04 (xfer#9, to-check=2/13)
NTAP/undotbs02.dbf
  1073750016 100%  187.65MB/s    0:00:05 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    5.09MB/s    0:00:00 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
```

```
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  177725933.55 bytes/sec
total size is 18570092218  speedup is 1.00
Succesfully replicated filesystem /oradata to /new/oradata
swap 0 /logs /new/logs
/new/logs unmounted
/logs unmounted
Mounted updated /logs
Swapped filesystem /logs for /new/logs
swap 1 /oradata /new/oradata
/new/oradata unmounted
/oradata unmounted
Mounted updated /oradata
Swapped filesystem /oradata for /new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:08:59 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
[root@host1 current]#
```

Oracle ASM spfileとpasswdの移行

ASMを含む移行を完了する際の難しさの1つに、ASM固有のspfileとパスワードファイルがあります。デフォルトでは、これらの重要なメタデータファイルは、最初に定義されたASMディスクグループに作成されます。特定のASMディスクグループを退避して削除する必要がある場合は、そのASMインスタンスを制御するspfileファイルとパスワードファイルを再配置する必要があります。

これらのファイルの再配置が必要になる別のユースケースとして、SnapManager for OracleやSnapCenter Oracleプラグインなどのデータベース管理ソフトウェアを導入する場合があります。これらの製品の機能の1つは、データファイルをホストしているASM LUNの状態をリバートして、データベースを迅速にリストアすることです。そのためには、リストアを実行する前にASMディスクグループをオフラインにする必要があり

ます。特定のデータベースのデータファイルが専用のASMディスクグループに分離されていれば、これは問題になりません。

そのディスクグループにASM spfile/passwdファイルも含まれている場合、ディスクグループをオフラインにするには、ASMインスタンス全体をシャットダウンするしかありません。これはシステムの停止を伴うプロセスであり、spfile/passwdファイルを再配置する必要があります。

環境

1. データベースSID =トースト
2. 現在のデータファイル： `+DATA`
3. 現在のログファイルと制御ファイル `+LOGS`
4. シンシイASMテイスククルウフノセツテイ `+NEWDATA` および `+NEWLOGS`

ASM spfile/passwdファイルの場所

これらのファイルは、システムを停止することなく再配置できます。ただし、安全のために、NetAppでは、ファイルが再配置され、構成が適切に更新されたことを確実に確認できるように、データベース環境をシャットダウンすることを推奨しています。サーバに複数のASMインスタンスが存在する場合は、この手順を繰り返す必要があります。

ASMインスタンスの識別

に記録されたデータに基づいてASMインスタンスを特定します。 `oratab` ファイル。ASMインスタンスは+記号で示されます。

```
-bash-4.1$ cat /etc/oratab | grep '^+'
+ASM:/orabin/grid:N             # line added by Agent
```

このサーバーには+asmというASMインスタンスが1つあります。

すべてのデータベースがシャットダウンされていることを確認する

表示されるSMONプロセスは、使用中のASMインスタンスのSMONだけです。別のSMONプロセスが存在する場合は、データベースが実行中であることを示します。

```
-bash-4.1$ ps -ef | grep smon
oracle     857     1   0 18:26 ?        00:00:00 asm_smon_+ASM
```

SMONプロセスはASMインスタンス自体のみです。これは、他のデータベースが実行されていないことを意味し、データベースの処理を中断するリスクを伴わずに、安全に処理を続行できることを意味します。

ファイルの検索

次のコマンドを使用して、ASM spfileおよびパスワードファイルの現在の場所を特定します。 `spget` および `pwget` コマンド

```
bash-4.1$ asmcmd
ASMCMD> spget
+DATA/spfile.ora
```

```
ASMCMD> pwget --asm
+DATA/orapwasm
```

これらのファイルは両方とも、 `+DATA` ディスクグループ：

ファイルのコピー

次のコマンドを使用して、ファイルを新しいASMディスクグループにコピーします。 `spcopy` および `pwcopy` コマンド新しいディスクグループが最近作成され、現在空の場合は、最初にマウントする必要があります。

```
ASMCMD> mount NEWDATA
```

```
ASMCMD> spcopy +DATA/spfile.ora +NEWDATA/spfile.ora
copying +DATA/spfile.ora -> +NEWDATA/spfilea.ora
```

```
ASMCMD> pwcopy +DATA/orapwasm +NEWDATA/orapwasm
copying +DATA/orapwasm -> +NEWDATA/orapwasm
```

ファイルは次の場所からコピーされました： `+DATA` 終了： `+NEWDATA`。

ASMインスタンスの更新

ASMインスタンスを更新して、場所の変更を反映する必要があります。。 `spset` および `pwset` コマンドは、ASMディスクグループの起動に必要なASMメタデータを更新します。

```
ASMCMD> spset +NEWDATA/spfile.ora
ASMCMD> pwset --asm +NEWDATA/orapwasm
```

更新ファイルを使用した**ASM**のアクティブ化

この時点で、ASMインスタンスは引き続きこれらのファイルの以前の場所を使用します。新しい場所からファイルを強制的に再読み込みし、以前のファイルのロックを解除するには、インスタンスを再起動する必要があります。

```
-bash-4.1$ sqlplus / as sysasm
SQL> shutdown immediate;
ASM diskgroups volume disabled
ASM diskgroups dismounted
ASM instance shutdown
```

```
SQL> startup
ASM instance started
Total System Global Area 1140850688 bytes
Fixed Size                  2933400 bytes
Variable Size            1112751464 bytes
ASM Cache                  25165824 bytes
ORA-15032: not all alterations performed
ORA-15017: diskgroup "NEWDATA" cannot be mounted
ORA-15013: diskgroup "NEWDATA" is already mounted
```

古い**spfile**ファイルとパスワードファイルを削除する

手順が正常に実行されると、以前のファイルはロックされなくなり、削除できるようになります。

```
-bash-4.1$ asmcmd
ASMCMD> rm +DATA/spfile.ora
ASMCMD> rm +DATA/orapwasm
```

Oracle ASMカラASMへノコヒイ

Oracle ASMは、基本的に軽量なボリュームマネージャとファイルシステムを統合したものです。ファイルシステムはすぐには認識されないため、RMANを使用してコピー処理を実行する必要があります。コピーベースの移動プロセスは安全でシンプルですが、システム停止が発生することがあります。システム停止を最小限に抑えることはできますが、完全に排除することはできません。

ASMベースのデータベースを無停止で移行する場合は、ASMの機能を活用して、古いLUNを削除しながらASMエクステントを新しいLUNにリバランシングすることを推奨します。これは一般に安全でノンストップオペレーションですが、バックアウトパスは提供されません。機能またはパフォーマンスの問題が発生した場合、唯一の選択肢はデータをソースに戻すことです。

このリスクを回避するには、データを移動するのではなく、データベースを新しい場所にコピーして、元のデータに変更を加えないようにします。データベースは、稼働を開始する前に新しい場所で完全にテストすることができ、問題が見つかった場合は、元のデータベースをフォールバックオプションとして使用できます。

この手順は、RMANに関連する多数のオプションの1つです。最初のバックアップが作成され、ログ再生によって後で同期される2段階のプロセスが可能になります。このプロセスでは、最初のベースラインコピーの実行中もデータベースの運用を維持し、データを提供できるため、ダウンタイムを最小限に抑えることが推奨されます。

データベースコピー

Oracle RMANは、ASMディスクグループに現在配置されているソースデータベースのレベル0（完全）コピーを作成します。 `+DATA` 次の場所に移動します： `+NEWDATA`。

```
-bash-4.1$ rman target /
Recovery Manager: Release 12.1.0.2.0 - Production on Sun Dec 6 17:40:03
2015
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: TOAST (DBID=2084313411)
RMAN> backup as copy incremental level 0 database format '+NEWDATA' tag
'ONTAP_MIGRATION';
Starting backup at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=302 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
...
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
output file name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
tag=ONTAP_MIGRATION RECID=5 STAMP=897759622
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWDATA/TOAST/BACKUPSET/2015_12_06/nnsnn0_ontap_migration_0.262.89
7759623 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

アーカイブログの強制切り替え

コピーの完全な整合性を確保するために必要なすべてのデータがアーカイブログに含まれていることを確認するには、アーカイブログを強制的に切り替えます。このコマンドを使用しないと、REDOログにキーデータが残っている可能性があります。

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

ソースデータベースのシャットダウン

データベースがシャットダウンされ、アクセスが制限された読み取り専用モードになるため、システムが停止します。ソースデータベースをシャットダウンするには、次のコマンドを実行します。

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                390073456 bytes
Database Buffers             406847488 bytes
Redo Buffers                   5455872 bytes
```

制御ファイルのバックアップ

移行を中止して元のストレージの場所に戻す必要がある場合に備えて、制御ファイルをバックアップする必要があります。バックアップ制御ファイルのコピーは100%必要ではありませんが、データベースファイルの場所を元の場所にリセットする処理が簡単になります。

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=358 device type=DISK
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151206T174753 RECID=6
STAMP=897760073
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

パラメータの更新

現在のspfileには、古いASMディスクグループ内の現在の場所にある制御ファイルへの参照が含まれています。編集する必要があります。これは、中間のpfileバージョンを編集することで簡単に実行できます。

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

pfileの更新

古いASMディスクグループを参照しているすべてのパラメータを更新し、新しいASMディスクグループ名を反映させます。次に、更新されたpfileを保存します。次のことを確認します。 `db_create` パラメータが存在します。

次の例では、 `+DATA` 変更されました `+NEWDATA` 黄色で強調表示されます。主なパラメータは次の2つです。 `db_create` 正しい場所に新しいファイルを作成するパラメータ。

```
*.compatible='12.1.0.2.0'
*.control_files='+NEWLOGS/TOAST/CONTROLFILE/current.258.897683139'
*.db_block_size=8192
*. db_create_file_dest='+NEWDATA'
*. db_create_online_log_dest_1='+NEWLOGS'
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=+NEWLOGS'
*.log_archive_format='%t_%s_%r.dbf'
```

init.oraファイルの更新

ほとんどのASMベースのデータベースでは、 `init.ora` ファイルはにあります `$ORACLE_HOME/dbs` ディレクトリ。ASMディスクグループ上のspfileへのポイントです。このファイルは、新しいASMディスクグループの場所にリダイレクトする必要があります。

```
-bash-4.1$ cd $ORACLE_HOME/dbs
-bash-4.1$ cat initTOAST.ora
SPFILE='+DATA/TOAST/spfileTOAST.ora'
```

このファイルを次のように変更します。

```
SPFILE=+NEWLOGS/TOAST/spfileTOAST.ora
```

パラメータファイルの再作成

これで'編集したpfileのデータをspfileに入力する準備が整いました

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

新しい**spfile**の使用を開始するには'データベースを起動します

データベースを起動して、新しく作成されたspfileが使用されていること、およびシステムパラメータに対するそれ以降の変更が正しく記録されていることを確認します。

```
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                373296240 bytes
Database Buffers             423624704 bytes
Redo Buffers                   5455872 bytes
```

制御ファイルのリストア

RMANによって作成されたバックアップ制御ファイルは、RMANによって、新しいspfileに指定された場所に直接リストアすることもできます。

```
RMAN> restore controlfile from
'+DATA/TOAST/CONTROLFILE/current.258.897683139';
Starting restore at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=417 device type=DISK
channel ORA_DISK_1: copied control file copy
output file name=+NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
Finished restore at 06-DEC-15
```

データベースをマウントし、新しい制御ファイルが使用されていることを確認します。

```
RMAN> alter database mount;
using target database control file instead of recovery catalog
Statement processed
```

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
+NEWLOGS/TOAST/CONTROLFILE/cur
                                                 rent.273.897761061
```

ログ再生

データベースは現在、古い場所にあるデータファイルを使用しています。コピーを使用する前に、コピーを同期する必要があります。最初のコピープロセスで時間が経過し、主にアーカイブログに変更が記録されました。これらの変更は次のように複製されます。

1. アーカイブ・ログを含むRMAN増分バックアップを実行します。

```
RMAN> backup incremental level 1 format '+NEWLOGS' for recover of copy
with tag 'ONTAP_MIGRATION' database;
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=62 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
input datafile file number=00002
name=+DATA/TOAST/DATAFILE/sysaux.260.897683143
input datafile file number=00003
name=+DATA/TOAST/DATAFILE/undotbs1.257.897683145
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/ncsnn1_ontap_migration_0.267.
897762697 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

2. ログを再生します。

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 06-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001
name=+NEWDATA/TOAST/DATAFILE/system.259.897759609
recovering datafile copy file number=00002
name=+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
recovering datafile copy file number=00003
name=+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
recovering datafile copy file number=00004
name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
channel ORA_DISK_1: reading from backup piece
+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.8977626
93
channel ORA_DISK_1: piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

アクティブ化

リストアされた制御ファイルは元の場所にあるデータ・ファイルを参照しており、コピーされたデータ・ファイルのパス情報も含まれています。

1. アクティブなデータファイルを変更するには、 `switch database to copy` コマンドを実行します

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/system.259.897759609"
datafile 2 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615"
datafile 3 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619"
datafile 4 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/users.258.897759623"
```

アクティブなデータファイルがコピーされたデータファイルになりますが、最終的なREDOログに変更が含まれている可能性があります。

2. 残りのログをすべて再生するには、 `recover database` コマンドを実行しますというメッセージが表示

されます `media recovery complete` と表示され、プロセスは成功しました。

```
RMAN> recover database;
Starting recover at 06-DEC-15
using channel ORA_DISK_1
starting media recovery
media recovery complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

このプロセスで変更されるのは、通常のデータファイルの場所だけです。一時データファイルの名前は変更する必要がありますが、一時データファイルであるためコピーする必要はありません。データベースは現在ダウンしているため、一時データファイルにアクティブなデータはありません。

3. 一時データファイルを移動するには、まずその場所を特定します。

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
-------------------------------------------------------------------------
--------
1 +DATA/TOAST/TEMPFILE/temp.263.897683145
```

4. 各データファイルに新しい名前を設定するRMANコマンドを使用して、一時データファイルを移動します。Oracle Managed Files（OMF）では、完全な名前は必要ありません。ASMディスクグループで十分です。データベースが開くと、OMFはASMディスクグループ上の適切な場所にリンクします。ファイルを再配置するには、次のコマンドを実行します。

```
run {
set newname for tempfile 1 to '+NEWDATA';
switch tempfile all;
}
```

```
RMAN> run {
2> set newname for tempfile 1 to '+NEWDATA';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to +NEWDATA in control file
```

Redoログの移行

移行プロセスはほぼ完了していますが、REDOログは元のASMディスクグループに残ります。REDOログは直接再配置できません。代わりに、新しいREDOログセットが作成されて設定に追加され、古いログがドロップされます。

1. REDOロググループの数とそれぞれのグループ番号を確認します。

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
------------------------------------------------------------------------
--------
1 +DATA/TOAST/ONLINELOG/group_1.261.897683139
2 +DATA/TOAST/ONLINELOG/group_2.259.897683139
3 +DATA/TOAST/ONLINELOG/group_3.256.897683139
```

2. Redoログのサイズを入力します。

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Redoログごとに、設定が一致する新しいグループを作成します。OMFを使用しない場合は、フルパスを指定する必要があります。また、この例では、 `db_create_online_log` パラメータ前述のように、このパラメータは+NEWLOGSに設定されています。この設定では、次のコマンドを使用して、ファイルの場所や特定のASMディスクグループを指定することなく、新しいオンラインログを作成できます。

```
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
```

4. データベースを開きます。

```
SQL> alter database open;
Database altered.
```

5. 古いログを削除します。

```
RMAN> alter database drop logfile group 1;
Statement processed
```

6. アクティブなログをドロップできないエラーが発生した場合は、次のログに切り替えてロックを解除し、グローバルチェックポイントを強制的に実行します。以下に例を示します。古い場所にあるログファイルグループ3を削除しようとしましたが、このログファイルにアクティブなデータが残っているため拒否されました。チェックポイントに続くログアーカイブでは、ログファイルを削除できます。

```
RMAN> alter database drop logfile group 3;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 3 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 3 thread 1:
'+LOGS/TOAST/ONLINELOG/group_3.259.897563549'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 3;
Statement processed
```

7. 環境をレビューして、すべてのロケーションベースのパラメータが更新されていることを確認します。

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

8. 次のスクリプトは、このプロセスを簡素化する方法を示しています。

```
[root@host1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
+NEWDATA/TOAST/DATAFILE/system.259.897759609
+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
+NEWDATA/TOAST/DATAFILE/users.258.897759623
TOAST redo logs:
+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123
+NEWLOGS/TOAST/ONLINELOG/group_5.265.897763125
+NEWLOGS/TOAST/ONLINELOG/group_6.264.897763125
TOAST temp datafiles:
+NEWDATA/TOAST/TEMPFILE/temp.260.897763165
TOAST spfile
spfile                                string
+NEWDATA/spfiletoast.ora
TOAST key parameters
control_files +NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
log_archive_dest_1 LOCATION=+NEWLOGS
db_create_file_dest +NEWDATA
db_create_online_log_dest_1 +NEWLOGS
```

9. ASMディスクグループが完全に退避された場合は、次のコマンドを使用してアンマウントできます。 `asmcmd`。ただし、多くの場合、他のデータベースまたはASM spfile/passwdファイルに属するファイルが存在する可能性があります。

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

Oracle ASMからファイルシステムへのコピー

Oracle ASMからファイルシステムへのコピー手順は、ASMからASMへのコピー手順と非常によく似ていますが、利点と制限は似ています。主な違いは、ASMディスクグループではなく可視ファイルシステムを使用する場合の、さまざまなコマンドや設定パラメータの構文です。

データベースコピー

Oracle RMANを使用して、ASMディスクグループに現在配置されているソースデータベースのレベル0（完全）コピーを作成します。 `+DATA` 次の場所に移動します： `/oradata`。

```
RMAN> backup as copy incremental level 0 database format
'/oradata/TOAST/%U' tag 'ONTAP_MIGRATION';
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=377 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-
1_01r5fhjg tag=ONTAP_MIGRATION RECID=1 STAMP=911722099
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-
2_02r5fhjo tag=ONTAP_MIGRATION RECID=2 STAMP=911722106
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt tag=ONTAP_MIGRATION RECID=3 STAMP=911722113
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/oradata/TOAST/cf_D-TOAST_id-2098173325_04r5fhk5
tag=ONTAP_MIGRATION RECID=4 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting datafile copy
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-
4_05r5fhk6 tag=ONTAP_MIGRATION RECID=5 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/oradata/TOAST/06r5fhk7_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
```

アーカイブログの強制切り替え

コピーの完全な整合性を確保するために必要なすべてのデータがアーカイブログに含まれていることを確認するには、アーカイブログの切り替えを強制する必要があります。このコマンドを使用しないと、REDOログにキーデータが残っている可能性があります。アーカイブログを強制的に切り替えるには、次のコマンドを実行

します。

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

ソースデータベースのシャットダウン

データベースがシャットダウンされ、アクセスが制限された読み取り専用モードになるため、システムが停止します。ソースデータベースをシャットダウンするには、次のコマンドを実行します。

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

制御ファイルのバックアップ

移行を中止して元のストレージの場所に戻す必要がある場合に備えて、制御ファイルをバックアップします。バックアップ制御ファイルのコピーは100%必要ではありませんが、データベースファイルの場所を元の場所にリセットする処理が簡単になります。

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 08-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151208T194540 RECID=30
STAMP=897939940
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 08-DEC-15
```

パラメータの更新

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

pfileの更新

古いASMディスクグループを参照するすべてのパラメータは、関連性がなくなったときに更新し、場合によっては削除する必要があります。新しいファイルシステムパスを反映するように更新し、更新されたpfileを保存します。完全なターゲットパスが表示されていることを確認します。これらのパラメータを更新するには、次のコマンドを実行します。

```
*.audit_file_dest='/orabin/admin/TOAST/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/logs/TOAST/arch/control01.ctl','/logs/TOAST/redo/control
02.ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=/logs/TOAST/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

元の**init.ora**ファイルを無効にする

このファイルは、 `$ORACLE_HOME/dbs` ディレクトリとは、通常、ASMディスクグループ上のspfileへのポインタとして機能するpfile内にあります。元のspfileが使用されていないことを確認するには、名前を変更します。ただし、このファイルは移行を中止する必要がある場合に必要になるため、削除しないでください。

```
[oracle@jfsc1 ~]$ cd $ORACLE_HOME/dbs
[oracle@jfsc1 dbs]$ cat initTOAST.ora
SPFILE='+ASM0/TOAST/spfileTOAST.ora'
[oracle@jfsc1 dbs]$ mv initTOAST.ora initTOAST.ora.prev
[oracle@jfsc1 dbs]$
```

パラメータファイルの再作成

これは'spfile再配置の最後の手順です元のspfileは使用されなくなり'中間ファイルを使用してデータベースが現

在起動されています（マウントされていません）このファイルの内容は'次のようにして新しいspfileの場所に書き出すことができます

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

新しい**spfile**の使用を開始するには'データベースを起動します

中間ファイルのロックを解除するには、データベースを起動し、新しいspfileファイルのみを使用してデータベースを起動する必要があります。データベースを起動すると、新しいspfileの場所が正しいことと、そのデータが有効であることも証明されます。

```
RMAN> shutdown immediate;
Oracle instance shut down
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

制御ファイルのリストア

バックアップ制御ファイルがパスに作成されました `/tmp/TOAST.ctrl` 手順の初期段階。新しいspfileでは、制御ファイルの場所を次のように定義します。 `/logfs/TOAST/ctrl/ctrlfile1.ctrl` および `/logfs/TOAST/redo/ctrlfile2.ctrl`。ただし、これらのファイルはまだ存在しません。

1. このコマンドは、spfileに定義されているパスに制御ファイルのデータをリストアします。

   ```
   RMAN> restore controlfile from '/tmp/TOAST.ctrl';
   Starting restore at 13-MAY-16
   using channel ORA_DISK_1
   channel ORA_DISK_1: copied control file copy
   output file name=/logs/TOAST/arch/control01.ctl
   output file name=/logs/TOAST/redo/control02.ctl
   Finished restore at 13-MAY-16
   ```

2. mountコマンドを問題して、制御ファイルが正しく検出され、有効なデータが含まれていることを確認します。

```
RMAN> alter database mount;
Statement processed
released channel: ORA_DISK_1
```

を検証するには `control_files` パラメータを指定して、次のコマンドを実行します。

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
/logs/TOAST/arch/control01.ctl
                                                 ,
/logs/TOAST/redo/control02.c
                                                 tl
```

ログ再生

データベースは現在、古い場所にあるデータファイルを使用しています。コピーを使用する前に、データファイルを同期する必要があります。最初のコピープロセスで時間が経過し、変更は主にアーカイブログに記録されました。これらの変更は、次の2つのステップで複製されます。

1. アーカイブ・ログを含むRMAN増分バックアップを実行します。

```
RMAN>  backup incremental level 1 format '/logs/TOAST/arch/%U' for
recover of copy with tag 'ONTAP_MIGRATION' database;
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=124 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/logs/TOAST/arch/09r5fj8i_1_1 tag=ONTAP_MIGRATION
comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

2. ログを再生します。

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 13-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
recovering datafile copy file number=00002 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
recovering datafile copy file number=00003 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
recovering datafile copy file number=00004 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
channel ORA_DISK_1: reading from backup piece
/logs/TOAST/arch/09r5fj8i_1_1
channel ORA_DISK_1: piece handle=/logs/TOAST/arch/09r5fj8i_1_1
tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

アクティブ化

リストアされた制御ファイルは元の場所にあるデータ・ファイルを参照しており、コピーされたデータ・ファイルのパス情報も含まれています。

1. アクティブなデータファイルを変更するには、 `switch database to copy` コマンドを実行します

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSTEM_FNO-1_01r5fhjg"
datafile 2 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSAUX_FNO-2_02r5fhjo"
datafile 3 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt"
datafile 4 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-USERS_FNO-4_05r5fhk6"
```

2. データファイルの整合性は完全である必要がありますが、オンラインREDOログに記録された残りの変更を再生するには、最後に1つの手順を実行する必要があります。を使用します `recover database` これらの変更を再生し、コピーを元のコピーと100%同一にするコマンド。ただし、コピーはまだ開いていません。

```
RMAN> recover database;
Starting recover at 13-MAY-16
using channel ORA_DISK_1
starting media recovery
archived log for thread 1 with sequence 28 is already on disk as file
+ASM0/TOAST/redo01.log
archived log file name=+ASM0/TOAST/redo01.log thread=1 sequence=28
media recovery complete, elapsed time: 00:00:00
Finished recover at 13-MAY-16
```

一時データファイルの再配置

1. 元のディスクグループでまだ使用されている一時データファイルの場所を特定します。

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
------------------------------------------------------------------------
--------
1 +ASM0/TOAST/temp01.dbf
```

2. データファイルを移動するには、次のコマンドを実行します。一時ファイルが多数ある場合は、テキスト・エディタを使用してRMANコマンドを作成し、それをカットアンドペーストします。

```
RMAN> run {
2> set newname for tempfile 1 to '/oradata/TOAST/temp01.dbf';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/TOAST/temp01.dbf in control file
```

Redoログの移行

移行プロセスはほぼ完了していますが、REDOログは元のASMディスクグループに残ります。REDOログは直接再配置できません。代わりに、新しいREDOログセットが作成され、古いログがドロップされて設定に追加されます。

1. REDOロググループの数とそれぞれのグループ番号を確認します。

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
--------------------------------------------------------------------------------
--------
1 +ASM0/TOAST/redo01.log
2 +ASM0/TOAST/redo02.log
3 +ASM0/TOAST/redo03.log
```

2. Redoログのサイズを入力します。

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
--------------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Redoログごとに、新しいファイルシステムの場所を使用して、現在のRedoロググループと同じサイズを使用して新しいグループを作成します。

```
RMAN> alter database add logfile '/logs/TOAST/redo/log00.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log01.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log02.rdo' size
52428800;
Statement processed
```

4. 以前のストレージにまだ配置されている古いログファイルグループを削除します。

```
RMAN> alter database drop logfile group 4;
Statement processed
RMAN> alter database drop logfile group 5;
Statement processed
RMAN> alter database drop logfile group 6;
Statement processed
```

5. アクティブログのドロップをブロックするエラーが発生した場合は、次のログに強制的に切り替えてロックを解除し、グローバルチェックポイントを強制的に実行します。以下に例を示します。古い場所にある

ログファイルグループ3を削除しようとしましたが、このログファイルにアクティブなデータが残っているため拒否されました。ログをアーカイブしたあとにチェックポイントを追加すると、ログファイルの削除が可能になります。

```
RMAN> alter database drop logfile group 4;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 4 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 4 thread 1:
'+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 4;
Statement processed
```

6. 環境をレビューして、すべてのロケーションベースのパラメータが更新されていることを確認します。

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

7. 次のスクリプトは、このプロセスを簡単にする方法を示しています。

```
[root@jfsc1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
TOAST redo logs:
/logs/TOAST/redo/log00.rdo
/logs/TOAST/redo/log01.rdo
/logs/TOAST/redo/log02.rdo
TOAST temp datafiles:
/oradata/TOAST/temp01.dbf
TOAST spfile
spfile                                  string
/orabin/product/12.1.0/dbhome_
                                                      1/dbs/spfileTOAST.ora
TOAST key parameters
control_files /logs/TOAST/arch/control01.ctl,
/logs/TOAST/redo/control02.ctl
log_archive_dest_1 LOCATION=/logs/TOAST/arch
```

8. ASMディスクグループが完全に退避された場合は、次のコマンドを使用してアンマウントできます。 `asmcmd`。多くの場合、他のデータベースまたはASM spfile/passwdファイルに属するファイルは引き続き存在する可能性があります。

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

データファイルのクリーンアップ手順

Oracle RMANの使用方法によっては、移行プロセスの結果、構文が長いデータファイルや暗号化されたデータファイルが生成されることがあります。この例では、次のファイル形式でバックアップが実行されています：`/oradata/TOAST/%U`。 `%U` RMANが各データ・ファイルにデフォルトの一意の名前を作成する必要があることを示します。結果は次のテキストに示されているものと似ています。データファイルの従来の名前は、名前の中に埋め込まれています。これは、に示すスクリプト化されたアプローチを使用してクリーンアップできます。 "ASM移行クリーンアップ"。

```
[root@jfsc1 current]# ./fixuniquenames.pl TOAST
#sqlplus Commands
shutdown immediate;
startup mount;
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/system.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/sysaux.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt /oradata/TOAST/undotbs1.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
/oradata/TOAST/users.dbf
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSTEM_FNO-1_01r5fhjg' to '/oradata/TOAST/system.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSAUX_FNO-2_02r5fhjo' to '/oradata/TOAST/sysaux.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
UNDOTBS1_FNO-3_03r5fhjt' to '/oradata/TOAST/undotbs1.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
USERS_FNO-4_05r5fhk6' to '/oradata/TOAST/users.dbf';
alter database open;
```

Oracle ASMのリバランシング

前述したように、Oracle ASMディスクグループは、リバランシングプロセスを使用して新しいストレージシステムに透過的に移行できます。つまり、リバランシングプロセスでは、既存のLUNグループに同じサイズのLUNを追加してから、前のLUNを破棄する必要があります。Oracle ASMは、基盤となるデータを最適なレイアウトで新しいストレージに自動的に再配置し、完了すると古いLUNを解放します。

マイグレーションプロセスでは効率的なシーケンシャルI/Oを使用し、通常は原因パフォーマンスの中断は発生しませんが、必要に応じてマイグレーション速度を調整できます。

移行するデータを特定

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
SQL> select group_number||' '||name from v$asm_diskgroup;
1 NEWDATA
```

新しいLUNを作成する

同じサイズの新しいLUNを作成し、必要に応じてユーザとグループのメンバーシップを設定します。LUNはと表示されます。 `CANDIDATE` ディスク：

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
 0 0 /dev/mapper/3600a098038303537762b47594c31586b CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315869 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315858 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c31586a CANDIDATE
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
```

新しいLUNの追加

追加処理と削除処理は同時に実行できますが、新しいLUNを追加する方が2つの手順で簡単に実行できます。まず、新しいLUNをディスクグループに追加します。この手順により、エクステントの半分が現在のASM LUNから新しいLUNに移行されます。

リバランシング電力は、データが転送される速度を示します。数値が大きいほど、データ転送の並列性が高くなります。移行は、効率的なシーケンシャルI/O処理を使用して実行されますが、原因のパフォーマンスに問題が生じることはほとんどありません。ただし、必要に応じて、進行中の移行のリバランシング機能を `alter diskgroup [name] rebalance power [level]` コマンドを実行します一般的な移行では、値5が使用されます。

```
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586b' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315869' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315858' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586a' rebalance power 5;
Diskgroup altered.
```

動作の監視

リバランシング処理は、さまざまな方法で監視および管理できます。この例では、次のコマンドを使用しました。

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

移行が完了しても、リバランシング処理は報告されません。

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

古い**LUN**を削除する

移行は途中で完了しました。環境が健全であることを確認するために、いくつかの基本的なパフォーマンステストを実行することを推奨します。確認後、古いLUNを削除して残りのデータを再配置できます。これによってLUNがすぐに解放されるわけではないことに注意してください。drop処理は、最初にエクステントを再配置してからLUNを解放するようOracle ASMに通知します。

```
sqlplus / as sysasm
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0000 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0001 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0002 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0003 rebalance power 5;
Diskgroup altered.
```

動作の監視

リバランシング処理は、さまざまな方法で監視および管理できます。この例では、次のコマンドを使用しました。

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

移行が完了しても、リバランシング処理は報告されません。

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

古い**LUN**を削除する

ディスクグループから古いLUNを削除する前に、ヘッダーのステータスを最後に確認する必要があります。ASMからLUNを解放すると、LUNの名前は表示されなくなり、ヘッダーステータスが FORMER。これは、これらのLUNをシステムから安全に削除できることを示します。

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NAME||''||GROUP_NUMBER||''||TOTAL_MB||''||PATH||''||HEADER_STATUS
--------------------------------------------------------------------------
------
 0 0 /dev/mapper/3600a098038303537762b47594c315863 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315864 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315861 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315862 FORMER
NEWDATA_0005 1 10240 /dev/mapper/3600a098038303537762b47594c315869 MEMBER
NEWDATA_0007 1 10240 /dev/mapper/3600a098038303537762b47594c31586a MEMBER
NEWDATA_0004 1 10240 /dev/mapper/3600a098038303537762b47594c31586b MEMBER
NEWDATA_0006 1 10240 /dev/mapper/3600a098038303537762b47594c315858 MEMBER
8 rows selected.
```

LVMの移行

ここに示す手順は、LVMベースのボリュームグループ移動の原則を示しています。 datavg。これらの例はLinux LVMを参考にしていますが、原則はAIX、HP-UX、VxVMにも当てはまります。正確なコマンドは異なる場合があります。

1. 現在に含まれているLUNを特定します。 datavg ボリュームグループ：

```
[root@host1 ~]# pvdisplay -C | grep datavg
  /dev/mapper/3600a098038303537762b47594c31582f datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31585a datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c315859 datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31586c datavg lvm2 a--  10.00g
10.00g
```

2. 物理サイズが同じか少し大きい新しいLUNを作成し、物理ボリュームとして定義します。

```
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315864
  Physical volume "/dev/mapper/3600a098038303537762b47594c315864"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315863
  Physical volume "/dev/mapper/3600a098038303537762b47594c315863"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315862
  Physical volume "/dev/mapper/3600a098038303537762b47594c315862"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315861
  Physical volume "/dev/mapper/3600a098038303537762b47594c315861"
successfully created
```

3. 新しいボリュームをボリュームグループに追加します。

```
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315864
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315863
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315862
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315861
  Volume group "datavg" successfully extended
```

4. 問題 `pvmove` コマンドを使用して、現在の各LUNのエクステントを新しいLUNに再配置します。。 `-i [seconds]` 引数は、操作の進行状況を監視します。

```
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31582f
/dev/mapper/3600a098038303537762b47594c315864
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 14.2%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 28.4%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 42.5%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 57.1%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 72.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 87.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31585a
/dev/mapper/3600a098038303537762b47594c315863
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 14.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 29.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 44.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 60.1%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 75.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 90.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c315859
/dev/mapper/3600a098038303537762b47594c315862
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 14.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 29.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 45.5%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 61.1%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 76.6%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 91.7%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31586c
/dev/mapper/3600a098038303537762b47594c315861
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 15.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 30.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 46.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 61.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 77.2%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 92.3%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 100.0%
```

5. このプロセスが完了したら、 `vgreduce` コマンドを実行します成功すると、LUNをシステムから安全に削除できるようになります。

```
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31582f
Removed "/dev/mapper/3600a098038303537762b47594c31582f" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31585a
  Removed "/dev/mapper/3600a098038303537762b47594c31585a" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c315859
  Removed "/dev/mapper/3600a098038303537762b47594c315859" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31586c
  Removed "/dev/mapper/3600a098038303537762b47594c31586c" from volume
group "datavg"
```

ForeignLUNImport

FLIによるOracleの移行計画

FLIを使用してSANリソースを移行する手順については、NetAppを参照してください。 "TR-4380：『SAN Migration Using Foreign LUN Import』"。

データベースとホストの観点からは、特別な手順は必要ありません。FCゾーンが更新されてLUNがONTAPで使用可能になると、LVMはLUNからLVMメタデータを読み取れるようになります。また、ボリュームグループを使用するための準備が整い、それ以上の設定手順は必要ありません。まれに、以前のストレージアレイへの参照がハードコーディングされた構成ファイルが環境に含まれることがあります。例えばLinuxシステムには `/etc/multipath.conf` 特定のデバイスのWWNを参照するルールは、FLIで導入された変更を反映するように更新する必要があります。

> サポートされている構成については、NetApp互換性マトリックスを参照してください。お使いの環境が含まれていない場合は、NetAppの担当者にお問い合わせください。

この例は、Linuxサーバでホストされ ているASM LUNとLVM LUNの両方の移行を示しています。FLIは他のオペレーティングシステムでもサポートされており、ホスト側のコマンドは異なる場合がありますが、原則は同じで、ONTAPの手順も同じです。

LVM LUNの特定

準備の最初の手順は、移行するLUNを特定することです。この例では、2つのSANベースのファイルシステムが `/orabin` および `/backups`。

```
[root@host1 ~]# df -k
Filesystem                          1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root                52403200   8811464  43591736  17% /
devtmpfs                             65882776         0  65882776   0% /dev
...
fas8060-nfs-public:/install      199229440 119368128  79861312  60%
/install
/dev/mapper/sanvg-lvorabin        20961280  12348476   8612804  59%
/orabin
/dev/mapper/sanvg-lvbackups       73364480  62947536  10416944  86%
/backups
```

ボリューム・グループの名前は'（ボリューム・グループ名）-（論理ボリューム名）という形式のデバイス名から抽出できますこの場合、ボリュームグループの名前は sanvg。

。 pvdisplay このボリュームグループをサポートするLUNを特定するには、コマンドを次のように使用します。この例では、 sanvg ボリュームグループ：

```
[root@host1 ~]# pvdisplay -C -o pv_name,pv_size,pv_fmt,vg_name
  PV                                                    PSize   VG
  /dev/mapper/3600a0980383030445424487556574266  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574267  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574268  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574269  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426a  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426b  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426c  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426d  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426e  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426f  10.00g sanvg
  /dev/sda2                                             278.38g rhel
```

ASM LUNの識別

ASM LUNも移行する必要があります。LUNとLUNパスの数をsqlplusからSYSASMユーザとして取得するには、次のコマンドを実行します。

```
SQL> select path||' '||os_mb from v$asm_disk;
PATH||''||OS_MB
--------------------------------------------------------------------------------
------
/dev/oracleasm/disks/ASM0 10240
/dev/oracleasm/disks/ASM9 10240
/dev/oracleasm/disks/ASM8 10240
/dev/oracleasm/disks/ASM7 10240
/dev/oracleasm/disks/ASM6 10240
/dev/oracleasm/disks/ASM5 10240
/dev/oracleasm/disks/ASM4 10240
/dev/oracleasm/disks/ASM1 10240
/dev/oracleasm/disks/ASM3 10240
/dev/oracleasm/disks/ASM2 10240
10 rows selected.
SQL>
```

FCネットワークの変更

現在の環境には、移行するLUNが20個含まれています。現在のSANを更新して、ONTAPが現在のLUNにアクセスできるようにします。データはまだ移行されていませんが、ONTAPは現在のLUNから構成情報を読み取って、そのデータの新しいホームを作成する必要があります。

AFF / FASシステムの少なくとも1つのHBAポートをイニシエータポートとして設定する必要があります。また、ONTAPが外部ストレージアレイ上のLUNにアクセスできるように、FCゾーンを更新する必要があります。一部のストレージアレイでは、特定のLUNにアクセスできるWWNを制限するLUNマスキングが設定されています。その場合は、LUNマスキングも更新して、ONTAP WWNへのアクセスを許可する必要があります。

この手順が完了すると、ONTAPは外部ストレージアレイを `storage array show` コマンドを実行します返されるキーフィールドは、システム上の外部LUNの識別に使用されるプレフィックスです。次の例では、外部アレイ上のLUN `FOREIGN_1` プレフィックスを使用してONTAP内に表示されます。 `FOR-1`。

外部アレイの識別

```
Cluster01::> storage array show -fields name,prefix
name          prefix
------------- ------
FOREIGN_1     FOR-1
Cluster01::>
```

外部**LUN**の識別

LUNを表示するには、 `array-name` に移動します `storage disk show` コマンドを実行します返されるデータは、移行手順中に複数回参照されます。

```
Cluster01::> storage disk show -array-name FOREIGN_1 -fields disk,serial
disk     serial-number
-------- -------------
FOR-1.1  800DT$HuVWBX
FOR-1.2  800DT$HuVWBZ
FOR-1.3  800DT$HuVWBW
FOR-1.4  800DT$HuVWBY
FOR-1.5  800DT$HuVWB/
FOR-1.6  800DT$HuVWBa
FOR-1.7  800DT$HuVWBd
FOR-1.8  800DT$HuVWBb
FOR-1.9  800DT$HuVWBc
FOR-1.10 800DT$HuVWBe
FOR-1.11 800DT$HuVWBf
FOR-1.12 800DT$HuVWBg
FOR-1.13 800DT$HuVWBi
FOR-1.14 800DT$HuVWBh
FOR-1.15 800DT$HuVWBj
FOR-1.16 800DT$HuVWBk
FOR-1.17 800DT$HuVWBm
FOR-1.18 800DT$HuVWBl
FOR-1.19 800DT$HuVWBo
FOR-1.20 800DT$HuVWBn
20 entries were displayed.
Cluster01::>
```

外部アレイ**LUN**をインポート候補として登録

外部LUNは、最初は特定のLUNタイプとして分類されます。データをインポートする前に、LUNを外部としてタグ付けする必要があるため、インポートプロセスの候補になる必要があります。この手順は、シリアル番号を `storage disk modify` 次の例に示すように、コマンドを実行します。このプロセスでは、ONTAP内でLUNのみが外部としてタグ付けされることに注意してください。外部LUN自体にはデータは書き込まれません。

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign true
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign true
Cluster01::*>
```

移行したLUNをホストするボリュームの作成

移行したLUNをホストするにはボリュームが必要です。正確なボリューム構成は、ONTAPの機能を活用する全体的な計画によって異なります。この例では、ASM LUNが1つのボリュームに配置され、LVM LUNが2つ目のボリュームに配置されています。これにより、階層化、Snapshotの作成、QoS制御の設定などの目的で、LUNを独立したグループとして管理できます。

を設定します `snapshot-policy `to `none`。移行プロセスには、大量のデータの入れ替えが含まれる場合があります。そのため、Snapshotに不要なデータがキャプチャされるために誤ってSnapshotを作成すると、スペース消費が大幅に増加する可能性があります。

```
Cluster01::> volume create -volume new_asm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1152] Job succeeded: Successful
Cluster01::> volume create -volume new_lvm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1153] Job succeeded: Successful
Cluster01::>
```

ONTAP LUNの作成

ボリュームを作成したら、新しいLUNを作成する必要があります。通常、LUNを作成する際にはLUNサイズなどの情報を指定する必要がありますが、この場合はforeign-disk引数がコマンドに渡されます。その結果、ONTAPは指定されたシリアル番号から現在のLUN設定データを複製します。また、LUNジオメトリとパーティションテーブルのデータを使用してLUNのアライメントを調整し、最適なパフォーマンスを確立します。

この手順では、外部アレイに対してシリアル番号を相互参照して、正しい外部LUNが正しい新しいLUNに照合されるようにする必要があります。

```
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN0 -ostype
linux -foreign-disk 800DT$HuVWBW
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN1 -ostype
linux -foreign-disk 800DT$HuVWBX
Created a LUN of size 10g (10737418240)
...
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN8 -ostype
linux -foreign-disk 800DT$HuVWBn
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN9 -ostype
linux -foreign-disk 800DT$HuVWBo
Created a LUN of size 10g (10737418240)
```

インポート関係を作成する

LUNは作成されましたが、レプリケーション先としては設定されていません。この手順を実行する前に、LUNをオフラインにする必要があります。この追加手順は、ユーザエラーからデータを保護するように設計されています。ONTAPでオンラインのLUNで移行を実行できると、入力ミスが原因でアクティブなデータが上書きされるリスクがあります。ユーザに最初にLUNをオフラインにするよう強制する追加手順は、正しいターゲットLUNが移行先として使用されていることを確認するのに役立ちます。

```
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN0
Warning: This command will take LUN "/vol/new_asm/LUN0" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN1
Warning: This command will take LUN "/vol/new_asm/LUN1" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
...
Warning: This command will take LUN "/vol/new_lvm/LUN8" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_lvm/LUN9
Warning: This command will take LUN "/vol/new_lvm/LUN9" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
```

LUNがオフラインになったら、外部LUNのシリアル番号を `lun import create` コマンドを実行します

```
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN0
-foreign-disk 800DT$HuVWBW
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN1
-foreign-disk 800DT$HuVWBX
...
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN8
-foreign-disk 800DT$HuVWBn
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN9
-foreign-disk 800DT$HuVWBo
Cluster01::*>
```

すべてのインポート関係が確立されたら、LUNをオンラインに戻すことができます。

```
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN9
Cluster01::*>
```

イニシエータグループの作成

イニシエータグループ（igroup）は、ONTAP LUNマスキングアーキテクチャの一部です。新しく作成したLUNには、ホストに最初にアクセスを許可しないかぎりアクセスできません。そのためには、アクセスを許可するFC WWNまたはiSCSIイニシエータ名をリストするigroupを作成します。このレポートの作成時点では、FLIはFC LUNでのみサポートされていました。ただし、移行後のiSCSIへの変換は簡単です（を参照）。"プロトコル変換"。

この例では、ホストのHBAで使用可能な2つのポートに対応する2つのWWNを含むigroupが作成されます。

```
Cluster01::*> igroup create linuxhost -protocol fcp -ostype linux
-initiator 21:00:00:0e:1e:16:63:50 21:00:00:0e:1e:16:63:51
```

新しい**LUN**をホストにマッピング

igroupの作成後、LUNは定義したigroupにマッピングされます。これらのLUNは、このigroupに含まれるWWNでのみ使用できます。NetAppでは、移動プロセスのこの段階で、ホストがONTAPにゾーニングされていないことを前提としています。これは重要なことです。ホストが外部アレイと新しいONTAPシステムに同時にゾーニングされていると、各アレイで同じシリアル番号のLUNが検出されるリスクがあるためです。マルチパスの誤動作やデータの破損が発生する可能性があります。

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
Cluster01::*>
```

FLIカットオーバーを使用した**Oracle**移行

FCネットワーク設定を変更する必要があるため、Foreign LUN Importの実行中にシステムが一部停止することは避けられません。ただし、システム停止は、データベース環境を再起動してFCゾーニングを更新し、ホストのFC接続を外部LUNからONTAPに切り替えるために必要な時間よりもはるかに長く続く必要はありません。

このプロセスは次のように要約できます。

1. 外部LUN上のすべてのLUNアクティビティを休止します。
2. ホストのFC接続を新しいONTAPシステムにリダイレクトします。
3. インポートプロセスをトリガーします。
4. LUNを再検出します。
5. データベースを再起動します。

移行プロセスが完了するまで待つ必要はありません。特定のLUNの移行を開始すると、そのLUNをONTAPで使用できるようになり、データコピープロセスを続行しながらデータを提供できます。すべての読み取りが外部LUNに渡され、すべての書き込みが両方のアレイに同期的に書き込まれます。コピー処理は非常に高速で、FCトラフィックのリダイレクトによるオーバーヘッドも最小限であるため、パフォーマンスへの影響は一時的で最小限に抑えてください。懸念事項がある場合は、移行プロセスが完了してインポート関係が削除されるまで、環境の再起動を遅らせることができます。

データベースをシャットダウン

この例の環境を休止する最初の手順は、データベースをシャットダウンすることです。

```
[oracle@host1 bin]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base remains unchanged with value /orabin
[oracle@host1 bin]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, Automatic Storage Management, OLAP, Advanced
Analytics
and Real Application Testing options
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
SQL>
```

グリッドサービスをシャットダウン

移行するSANベースのファイルシステムの1つには、Oracle ASMサービスも含まれています。基盤となるLUNを休止するには、ファイルシステムをディスマウントする必要があります。つまり、このファイルシステム上で開いているファイルを含むプロセスをすべて停止する必要があります。

```
[oracle@host1 bin]$ ./crsctl stop has -f
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'host1'
CRS-2673: Attempting to stop 'ora.evmd' on 'host1'
CRS-2673: Attempting to stop 'ora.DATA.dg' on 'host1'
CRS-2673: Attempting to stop 'ora.LISTENER.lsnr' on 'host1'
CRS-2677: Stop of 'ora.DATA.dg' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.asm' on 'host1'
CRS-2677: Stop of 'ora.LISTENER.lsnr' on 'host1' succeeded
CRS-2677: Stop of 'ora.evmd' on 'host1' succeeded
CRS-2677: Stop of 'ora.asm' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.cssd' on 'host1'
CRS-2677: Stop of 'ora.cssd' on 'host1' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'host1' has completed
CRS-4133: Oracle High Availability Services has been stopped.
[oracle@host1 bin]$
```

ファイルシステムのディスマウント

すべてのプロセスがシャットダウンされると、アンマウント処理は成功します。権限が拒否された場合は、ファイルシステムがロックされているプロセスが存在する必要があります。。 `fuser` コマンドは、これらのプロセスを識別するのに役立ちます。

```
[root@host1 ~]# umount /orabin
[root@host1 ~]# umount /backups
```

ボリュームグループの非アクティブ化

特定のボリュームグループ内のすべてのファイルシステムがディスマウントされたら、そのボリュームグループを非アクティブ化できます。

```
[root@host1 ~]# vgchange --activate n sanvg
  0 logical volume(s) in volume group "sanvg" now active
[root@host1 ~]#
```

FCネットワークの変更

FCゾーンを更新して、ホストから外部アレイへのすべてのアクセスを削除し、ONTAPへのアクセスを確立できるようになりました。

インポートプロセスの開始

LUNインポートプロセスを開始するには、 `lun import start` コマンドを実行します

```
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN0
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN1
...
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN8
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN9
Cluster01::lun import*>
```

インポートの進捗状況の監視

インポート操作を監視するには、 `lun import show` コマンドを実行します次の図に示すように、20個すべてのLUNのインポートを実行中です。つまり、データコピー処理がまだ進行中であっても、ONTAPからデータにアクセスできるようになります。

```
Cluster01::lun import*> lun import show -fields path,percent-complete
vserver   foreign-disk path               percent-complete
--------- ------------ ----------------- ----------------
vserver1  800DT$HuVWB/ /vol/new_asm/LUN4 5
vserver1  800DT$HuVWBW /vol/new_asm/LUN0 5
vserver1  800DT$HuVWBX /vol/new_asm/LUN1 6
vserver1  800DT$HuVWBY /vol/new_asm/LUN2 6
vserver1  800DT$HuVWBZ /vol/new_asm/LUN3 5
vserver1  800DT$HuVWBa /vol/new_asm/LUN5 4
vserver1  800DT$HuVWBb /vol/new_asm/LUN6 4
vserver1  800DT$HuVWBc /vol/new_asm/LUN7 4
vserver1  800DT$HuVWBd /vol/new_asm/LUN8 4
vserver1  800DT$HuVWBe /vol/new_asm/LUN9 4
vserver1  800DT$HuVWBf /vol/new_lvm/LUN0 5
vserver1  800DT$HuVWBg /vol/new_lvm/LUN1 4
vserver1  800DT$HuVWBh /vol/new_lvm/LUN2 4
vserver1  800DT$HuVWBi /vol/new_lvm/LUN3 3
vserver1  800DT$HuVWBj /vol/new_lvm/LUN4 3
vserver1  800DT$HuVWBk /vol/new_lvm/LUN5 3
vserver1  800DT$HuVWBl /vol/new_lvm/LUN6 4
vserver1  800DT$HuVWBm /vol/new_lvm/LUN7 3
vserver1  800DT$HuVWBn /vol/new_lvm/LUN8 2
vserver1  800DT$HuVWBo /vol/new_lvm/LUN9 2
20 entries were displayed.
```

オフラインプロセスが必要な場合は、サービスの再検出または再開を `lun import show` コマンドは、すべての移行が正常に完了したことを示します。その後、移行プロセスを完了できます（を参照）。 "Foreign LUN Import—完了"。

オンライン移行が必要な場合は、新しいホーム内のLUNの再検出に進み、サービスを起動します。

SCSIデバイスの変更をスキャン

ほとんどの場合、新しいLUNを再検出する最も簡単なオプションは、ホストを再起動することです。これにより、古いデバイスが自動的に削除され、新しいLUNがすべて適切に検出され、マルチパスデバイスなどの関連デバイスが構築されます。この例では、デモ用の完全オンラインプロセスを示しています。

注意：ホストを再起動する前に、 `/etc/fstab` 移行されたSANリソースについては、コメントアウトされています。これを行わず、LUNアクセスに問題があると、OSがブートしない可能性があります。この状況ではデータが破損することはありません。ただし、レスキューモードまたは同様のモードで起動し、 `/etc/fstab` これにより、OSを起動してトラブルシューティングを有効にすることができます。

この例で使用しているLinuxバージョンのLUNは、 `rescan-scsi-bus.sh` コマンドを実行しますコマンドが成功すると、各LUNパスが出力に表示されます。出力は解釈が難しい場合がありますが、ゾーニングとigroupの設定が正しい場合は、 `NETAPP` ベンダー文字列。

```
[root@host1 /]# rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 0 2 0 0 ...
OLD: Host: scsi0 Channel: 02 Id: 00 Lun: 00
      Vendor: LSI      Model: RAID SAS 6G 0/1  Rev: 2.13
      Type:   Direct-Access                    ANSI SCSI revision: 05
Scanning host 1 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 1 0 0 0 ...
OLD: Host: scsi1 Channel: 00 Id: 00 Lun: 00
      Vendor: Optiarc  Model: DVD RW AD-7760H  Rev: 1.41
      Type:   CD-ROM                           ANSI SCSI revision: 05
Scanning host 2 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 3 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 4 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 5 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 6 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 7 for  all SCSI target IDs, all LUNs
 Scanning for device 7 0 0 10 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 10
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 11 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 11
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 12 ...
...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 18
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 9 0 1 19 ...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 19
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
0 new or changed device(s) found.
0 remapped or resized device(s) found.
0 device(s) removed.
```

マルチハステハイスノカクニン

LUN検出プロセスではマルチパスデバイスの再作成もトリガーされますが、Linuxのマルチパスドライバでは時折問題が発生することがわかっています。の出力 `multipath - ll` 出力が想定どおりに表示されることを確認する必要があります。たとえば、次の出力は、に関連付けられているマルチパスデバイスを示しています。 `NETAPP` ベンダー文字列。各デバイスには4つのパスがあり、2つはプライオリティ50、2つはプライオリティ10です。正確な出力はLinuxのバージョンによって異なりますが、この出力は想定どおりです。

使用するLinuxのバージョンに対応するHost Utilitiesのマニュアルを参照して、 `/etc/multipath.conf` 設定が正しい。

```
[root@host1 /]# multipath -ll
3600a098038303558735d493762504b36 dm-5 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:4  sdat 66:208 active ready running
| `- 9:0:1:4  sdbn 68:16  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:4  sdf  8:80   active ready running
  `- 9:0:0:4  sdz  65:144 active ready running
3600a098038303558735d493762504b2d dm-10 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:8  sdax 67:16  active ready running
| `- 9:0:1:8  sdbr 68:80  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:8  sdj  8:144  active ready running
  `- 9:0:0:8  sdad 65:208 active ready running
...
3600a098038303558735d493762504b37 dm-8 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:5  sdau 66:224 active ready running
| `- 9:0:1:5  sdbo 68:32  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:5  sdg  8:96   active ready running
  `- 9:0:0:5  sdaa 65:160 active ready running
3600a098038303558735d493762504b4b dm-22 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:19 sdbi 67:192 active ready running
| `- 9:0:1:19 sdcc 69:0   active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:19 sdu  65:64  active ready running
  `- 9:0:0:19 sdao 66:128 active ready running
```

LVMボリュームグループの再アクティブ化

LVM LUNが正しく検出されていれば、 `vgchange --activate y` コマンドは成功するはずです。これは、

論理ボリュームマネージャの価値を示す良い例です。ボリュームグループのメタデータはLUN自体に書き込まれるため、LUNのWWNやシリアル番号の変更は重要ではありません。

OSがLUNをスキャンし、LUNに書き込まれている少量のデータが検出され、LUNがLUNに属する物理ボリュームであることがわかりました。 `sanvg volumegroup`。その後、必要なすべてのデバイスを構築しました。必要なのは、ボリュームグループを再アクティブ化することだけです。

```
[root@host1 /]# vgchange --activate y sanvg
  Found duplicate PV fpCzdLTuKfy2xDZjai1NliJh3TjLUBiT: using
/dev/mapper/3600a098038303558735d493762504b46 not /dev/sdp
  Using duplicate PV /dev/mapper/3600a098038303558735d493762504b46 from
subsystem DM, ignoring /dev/sdp
  2 logical volume(s) in volume group "sanvg" now active
```

ファイルシステムの再マウント

ボリューム・グループを再アクティブ化すると'元のデータをすべてそのまま使用してファイル・システムをマウントできます前述したように、バックグループでデータレプリケーションがまだアクティブであっても、ファイルシステムは完全に動作します。

```
[root@host1 /]# mount /orabin
[root@host1 /]# mount /backups
[root@host1 /]# df -k
Filesystem                        1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root              52403200   8837100  43566100  17% /
devtmpfs                           65882776         0  65882776   0% /dev
tmpfs                               6291456        84   6291372   1%
/dev/shm
tmpfs                              65898668      9884  65888784   1% /run
tmpfs                              65898668         0  65898668   0%
/sys/fs/cgroup
/dev/sda1                            505580    224828    280752  45% /boot
fas8060-nfs-public:/install      199229440 119368256  79861184  60%
/install
fas8040-nfs-routable:/snapomatic   9961472     30528   9930944   1%
/snapomatic
tmpfs                              13179736        16  13179720   1%
/run/user/42
tmpfs                              13179736         0  13179736   0%
/run/user/0
/dev/mapper/sanvg-lvorabin         20961280  12357456   8603824  59%
/orabin
/dev/mapper/sanvg-lvbackups        73364480  62947536  10416944  86%
/backups
```

ASMテハイスノサイスキヤン

ASMlibデバイスは、SCSIデバイスが再スキャンされたときに再検出されているはずです。再検出をオンラインで確認するには、ASMlibを再起動してからディスクをスキャンします。

この手順は、ASMlibを使用するASM構成にのみ関連します。

注意：ASMlibを使用しない場合は、 `/dev/mapper` デバイスは自動的に再作成されているはずです。ただし、権限が正しくない可能性があります。ASMlibがない場合は、ASMの基盤となるデバイスに特別な権限を設定する必要があります。これは通常、次のいずれかの特別なエントリによって達成されます。 `/etc/multipath.conf` または `udev` ルール、または両方のルールセットに含まれている可能性があります。ASMデバイスに正しいアクセス許可が設定されていることを確認するには、WWNまたはシリアル番号に関する環境の変更を反映するために、これらのファイルの更新が必要になる場合があります。

この例では、ASMlibを再起動してディスクをスキャンすると、元の環境と同じ10個のASM LUNが表示されます。

```
[root@host1 /]# oracleasm exit
Unmounting ASMlib driver filesystem: /dev/oracleasm
Unloading module "oracleasm": oracleasm
[root@host1 /]# oracleasm init
Loading module "oracleasm": oracleasm
Configuring "oracleasm" to use device physical block size
Mounting ASMlib driver filesystem: /dev/oracleasm
[root@host1 /]# oracleasm scandisks
Reloading disk partitions: done
Cleaning any stale ASM disks...
Scanning system for ASM disks...
Instantiating disk "ASM0"
Instantiating disk "ASM1"
Instantiating disk "ASM2"
Instantiating disk "ASM3"
Instantiating disk "ASM4"
Instantiating disk "ASM5"
Instantiating disk "ASM6"
Instantiating disk "ASM7"
Instantiating disk "ASM8"
Instantiating disk "ASM9"
```

グリッドサービスの再起動

LVMデバイスとASMデバイスがオンラインで使用可能になったので、グリッドサービスを再起動できます。

```
[root@host1 /]# cd /orabin/product/12.1.0/grid/bin
[root@host1 bin]# ./crsctl start has
```

データベースの再起動

グリッドサービスが再起動されたら、データベースを起動できます。ASMサービスが完全に使用可能になるまで数分待ってからデータベースを起動しなければならない場合があります。

```
[root@host1 bin]# su - oracle
[oracle@host1 ~]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base has been set to /orabin
[oracle@host1 ~]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> startup
ORACLE instance started.
Total System Global Area 3221225472 bytes
Fixed Size                  4502416 bytes
Variable Size            1207962736 bytes
Database Buffers         1996488704 bytes
Redo Buffers               12271616 bytes
Database mounted.
Database opened.
SQL>
```

FLIを使用した**Oracle**移行の完了

ホスト側から見ると移行は完了しますが、インポート関係が削除されるまでは外部アレイからI/Oが提供されます。

関係を削除する前に、すべてのLUNの移行プロセスが完了していることを確認する必要があります。

```
Cluster01::*> lun import show -vserver vserver1 -fields foreign-
disk,path,operational-state
vserver   foreign-disk path              operational-state
--------- ------------ ----------------- -----------------
vserver1 800DT$HuVWB/ /vol/new_asm/LUN4 completed
vserver1 800DT$HuVWBW /vol/new_asm/LUN0 completed
vserver1 800DT$HuVWBX /vol/new_asm/LUN1 completed
vserver1 800DT$HuVWBY /vol/new_asm/LUN2 completed
vserver1 800DT$HuVWBZ /vol/new_asm/LUN3 completed
vserver1 800DT$HuVWBa /vol/new_asm/LUN5 completed
vserver1 800DT$HuVWBb /vol/new_asm/LUN6 completed
vserver1 800DT$HuVWBc /vol/new_asm/LUN7 completed
vserver1 800DT$HuVWBd /vol/new_asm/LUN8 completed
vserver1 800DT$HuVWBe /vol/new_asm/LUN9 completed
vserver1 800DT$HuVWBf /vol/new_lvm/LUN0 completed
vserver1 800DT$HuVWBg /vol/new_lvm/LUN1 completed
vserver1 800DT$HuVWBh /vol/new_lvm/LUN2 completed
vserver1 800DT$HuVWBi /vol/new_lvm/LUN3 completed
vserver1 800DT$HuVWBj /vol/new_lvm/LUN4 completed
vserver1 800DT$HuVWBk /vol/new_lvm/LUN5 completed
vserver1 800DT$HuVWBl /vol/new_lvm/LUN6 completed
vserver1 800DT$HuVWBm /vol/new_lvm/LUN7 completed
vserver1 800DT$HuVWBn /vol/new_lvm/LUN8 completed
vserver1 800DT$HuVWBo /vol/new_lvm/LUN9 completed
20 entries were displayed.
```

インポート関係を削除します

移行プロセスが完了したら、移行関係を削除します。I/O処理が完了すると、ONTAP上のドライブからのみI/Oが提供されます。

```
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN9
```

外部**LUN**の登録解除

最後に、ディスクを変更して `is-foreign` 指定。

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign false
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign false
Cluster01::*>
```

FLIプロトコル変換を使用した**Oracle**移行

LUNへのアクセスに使用するプロトコルの変更は、一般的な要件です。

場合によっては、全体的な戦略の一環としてデータをクラウドに移行することもあります。TCP/IPはクラウドのプロトコルであり、FCからiSCSIに変更することで、さまざまなクラウド環境への移行が容易になります。また、IP SANのコスト削減を活用するためにiSCSIが望ましい場合もあります。移行では、一時的な手段として別のプロトコルが使用されることがあります。たとえば、外部アレイとONTAPベースのLUNを同じHBA上に共存させることができない場合は、iSCSI LUNを使用して古いアレイからデータをコピーできます。その後、古いLUNをシステムから削除したあとにFCに変換し直すことができます。

次の手順はFCからiSCSIへの変換を示していますが、全体的な原則はiSCSIからFCへの逆変換に適用されます。

iSCSIイニシエータのインストール

ほとんどのオペレーティングシステムには、デフォルトでソフトウェアiSCSIイニシエータが含まれていますが、含まれていない場合は簡単にインストールできます。

```
[root@host1 /]# yum install -y iscsi-initiator-utils
Loaded plugins: langpacks, product-id, search-disabled-repos,
subscription-
              : manager
Resolving Dependencies
--> Running transaction check
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.el7 will be
updated
--> Processing Dependency: iscsi-initiator-utils = 6.2.0.873-32.el7 for
package: iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7 will be
an update
--> Running transaction check
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.el7 will
be updated
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
```

```
will be an update
--> Finished Dependency Resolution
Dependencies Resolved
================================================================================
===
 Package                         Arch   Version              Repository
Size
================================================================================
===
Updating:
 iscsi-initiator-utils           x86_64 6.2.0.873-32.0.2.el7 ol7_latest 416
k
Updating for dependencies:
 iscsi-initiator-utils-iscsiuio x86_64 6.2.0.873-32.0.2.el7 ol7_latest  84
k
Transaction Summary
================================================================================
===
Upgrade  1 Package (+1 Dependent package)
Total download size: 501 k
Downloading packages:
No Presto metadata available for ol7_latest
(1/2): iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_6 | 416 kB   00:00
(2/2): iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2. |  84 kB   00:00
--------------------------------------------------------------------------------
---
Total                                           2.8 MB/s | 501 kB
00:00Cluster01
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Updating   : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
1/4
  Updating   : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
2/4
  Cleanup    : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
3/4
  Cleanup    : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
rhel-7-server-eus-rpms/7Server/x86_64/productid       | 1.7 kB     00:00
rhel-7-server-rpms/7Server/x86_64/productid           | 1.7 kB     00:00
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
1/4
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
2/4
```

```
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
3/4
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
Updated:
  iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7
Dependency Updated:
  iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
Complete!
[root@host1 /]#
```

iSCSIイニシエータ名の識別

インストールプロセス中に一意のiSCSIイニシエータ名が生成されます。Linuxの場合は、`/etc/iscsi/initiatorname.iscsi` ファイル。この名前は、IP SAN上のホストを識別するために使用されます。

```
[root@host1 /]# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1992-05.com.redhat:497bd66ca0
```

新しいイニシエータグループを作成する

イニシエータグループ（igroup）は、ONTAP LUNマスキングアーキテクチャの一部です。新しく作成したLUNには、ホストに最初にアクセスを許可しないかぎりアクセスできません。そのためには、アクセスが必要なFC WWNまたはiSCSIイニシエータ名のいずれかをリストするigroupを作成します。

この例では、LinuxホストのiSCSIイニシエータを含むigroupを作成しています。

```
Cluster01::*> igroup create -igroup linuxiscsi -protocol iscsi -ostype
linux -initiator iqn.1994-05.com.redhat:497bd66ca0
```

環境をシャットダウンする

LUNプロトコルを変更する前に、LUNを完全に休止する必要があります。変換するLUNのいずれかのデータベースをシャットダウンし、ファイルシステムをディスマウントし、ボリュームグループを非アクティブ化する必要があります。ASMを使用する場合は、ASMディスクグループがディスマウントされていることを確認し、すべてのグリッドサービスをシャットダウンします。

FCネットワークからの**LUN**のマッピング解除

LUNが完全に休止されたら、元のFC igroupからマッピングを削除します。

```
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
```

IPネットワークへの**LUN**の再マッピング

新しいiSCSIベースのイニシエータグループに各LUNへのアクセスを許可します。

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxiscsi
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxiscsi
Cluster01::*>
```

iSCSIターゲットの検出

iSCSI検出には2つのフェーズがあります。1つ目はターゲットの検出です。これは、LUNの検出とは異なります。。 `iscsiadm` 次のコマンドは、 `-p argument` およびには、iSCSIサービスを提供するすべてのIPアドレスとポートのリストが格納されます。この場合、デフォルトポート3260にiSCSIサービスを持つIPアドレスが4つあります。

> いずれかのターゲットIPアドレスに到達できない場合、このコマンドは完了までに数分かかることがあります。

```
[root@host1 ~]# iscsiadm -m discovery -t st -p fas8060-iscsi-public1
10.63.147.197:3260,1033 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
10.63.147.198:3260,1034 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.203:3260,1030 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.202:3260,1029 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
```

iSCSI LUNの検出

iSCSIターゲットが検出されたら、iSCSIサービスを再起動して使用可能なiSCSI LUNを検出し、マルチパスやASMlibデバイスなどの関連デバイスを構築します。

```
[root@host1 ~]# service iscsi restart
Redirecting to /bin/systemctl restart  iscsi.service
```

環境の再起動

ボリュームグループの再アクティブ化、ファイルシステムの再マウント、RACサービスの再起動などを実行して、環境を再起動します。予防措置としてNetApp、変換プロセスの完了後にサーバを再起動して、すべての構成ファイルが正しいことと古いデバイスがすべて削除されることを確認することをお勧めします。

注意：ホストを再起動する前に、 `/etc/fstab` 移行されたSANリソースについては、コメントアウトされています。この手順を実行せず、LUNアクセスに問題があると、OSがブートしない可能性があります。この問題はデータに損傷を与えません。ただし、レスキューモードまたは同様のモードで起動して修正するのは非常に不便な場合があります。 `/etc/fstab` OSを起動してトラブルシューティング作業を開始できるようにします。

Oracle移行手順のサンプルスクリプト

ここで紹介するスクリプトは、さまざまなOSおよびデータベースタスクのスクリプト作成方法の例として提供されています。それらはそのまま供給されます。特定の手順のサポートが必要な場合は、NetAppまたはNetAppリセラーにお問い合わせください。

データベースのシャットダウン

次のPerlスクリプトは、Oracle SIDの引数を1つ指定してデータベースをシャットダウンします。Oracleユーザまたはrootとして実行できます。

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
77 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF4
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
`;};
print @out;
if ("@out" =~ /ORACLE instance shut down/) {
print "$oraclesid shut down\n";
exit 0;}
elsif ("@out" =~ /Connected to an idle instance/) {
print "$oraclesid already shut down\n";
exit 0;}
else {
print "$oraclesid failed to shut down\n";
exit 1;}
```

データベースの起動

次のPerlスクリプトは、Oracle SIDの引数を1つ指定してデータベースをシャットダウンします。Oracleユーザまたはrootとして実行できます。

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
'
`;}
else {
@out=`. oraenv << EOF3
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
`;};
print @out;
if ("@out" =~ /Database opened/) {
print "$oraclesid started\n";
exit 0;}
elsif ("@out" =~ /cannot start already-running ORACLE/) {
print "$oraclesid already started\n";
exit 1;}
else {
78 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
print "$oraclesid failed to start\n";
exit 1;}
```

ファイルシステムを読み取り専用に変換

次のスクリプトは、ファイルシステム引数を取り、ディスマウントして読み取り専用として再マウントしようとします。これは、データをレプリケートするためにファイルシステムの可用性を維持しつつ、偶発的な破損から保護する必要がある移行プロセスで役立ちます。

```
#! /usr/bin/perl
use strict;
#use warnings;
my $filesystem=$ARGV[0];
my @out=`umount '$filesystem'`;
if ($? == 0) {
 print "$filesystem unmounted\n";
 @out = `mount -o ro '$filesystem'`;
 if ($? == 0) {
 print "$filesystem mounted read-only\n";
 exit 0;}}
else {
 print "Unable to unmount $filesystem\n";
 exit 1;}
print @out;
```

ファイルシステムの交換

次のスクリプト例は、あるファイルシステムを別のファイルシステムに置き換えるために使用します。`/etc/fstab`ファイルを編集するので、rootとして実行する必要があります。古いファイルシステムと新しいファイルシステムの単一のカンマ区切り引数を受け入れます。

1. ファイルシステムを交換するには、次のスクリプトを実行します。

```
#! /usr/bin/perl
use strict;
#use warnings;
my $oldfs;
my $newfs;
my @oldfstab;
my @newfstab;
my $source;
my $mountpoint;
my $leftover;
my $oldfstabentry='';
my $newfstabentry='';
my $migratedfstabentry='';
($oldfs, $newfs) = split (',',$ARGV[0]);
open(my $filehandle, '<', '/etc/fstab') or die "Could not open
/etc/fstab\n";
while (my $line = <$filehandle>) {
 chomp $line;
 ($source, $mountpoint, $leftover) = split(/[ , ]/,$line, 3);
 if ($mountpoint eq $oldfs) {
 $oldfstabentry = "#Removed by swap script $source $oldfs $leftover";}
```

```
 elsif ($mountpoint eq $newfs) {
 $newfstabentry = "#Removed by swap script $source $newfs $leftover";
 $migratedfstabentry = "$source $oldfs $leftover";}
 else {
 push (@newfstab, "$line\n")}}
79 Migration of Oracle Databases to NetApp Storage Systems © 2021
NetApp, Inc. All rights reserved
push (@newfstab, "$oldfstabentry\n");
push (@newfstab, "$newfstabentry\n");
push (@newfstab, "$migratedfstabentry\n");
close($filehandle);
if ($oldfstabentry eq ''){
 die "Could not find $oldfs in /etc/fstab\n";}
if ($newfstabentry eq ''){
 die "Could not find $newfs in /etc/fstab\n";}
my @out=`umount '$newfs'`;
if ($? == 0) {
 print "$newfs unmounted\n";}
else {
 print "Unable to unmount $newfs\n";
 exit 1;}
@out=`umount '$oldfs'`;
if ($? == 0) {
 print "$oldfs unmounted\n";}
else {
 print "Unable to unmount $oldfs\n";
 exit 1;}
system("cp /etc/fstab /etc/fstab.bak");
open ($filehandle, ">", '/etc/fstab') or die "Could not open /etc/fstab
for writing\n";
for my $line (@newfstab) {
 print $filehandle $line;}
close($filehandle);
@out=`mount '$oldfs'`;
if ($? == 0) {
 print "Mounted updated $oldfs\n";
 exit 0;}
else{
 print "Unable to mount updated $oldfs\n";
 exit 1;}
exit 0;
```

このスクリプトの使用例として、 `/oradata` の移行先 `/neworadata` および `/logs` の移行先 `/newlogs`。このタスクを実行する最も簡単な方法の1つは、単純なファイルコピー操作を使用して、新しいデバイスを元のマウントポイントに再配置することです。

2. 古いファイルシステムと新しいファイルシステムが /etc/fstab ファイルは次のとおりです。

```
cluster01:/vol_oradata /oradata nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
cluster01:/vol_logs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /newlogs nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
```

3. このスクリプトを実行すると、現在のファイルシステムがアンマウントされ、新しいファイルシステムに置き換えられます。

```
[root@jfsc3 scripts]# ./swap.fs.pl /oradata,/neworadata
/neworadata unmounted
/oradata unmounted
Mounted updated /oradata
[root@jfsc3 scripts]# ./swap.fs.pl /logs,/newlogs
/newlogs unmounted
/logs unmounted
Mounted updated /logs
```

4. このスクリプトでは、 /etc/fstab 必要に応じてファイルを作成この例では、次の変更が含まれています。

```
#Removed by swap script cluster01:/vol_oradata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_logs /logs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_newlogs /newlogs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0
0
```

データベース移行の自動化

この例では、シャットダウン、起動、およびファイルシステム置換スクリプトを使用して移行を完全に自動化する方法を示します。

```
#! /usr/bin/perl
use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my @oldfs;
my @newfs;
my $x=1;
while ($x < scalar(@ARGV)) {
   ($oldfs[$x-1], $newfs[$x-1]) = split (',',$ARGV[$x]);
   $x+=1;}
my @out=`./dbshut.pl '$oraclesid'`;
print @out;
if ($? ne 0) {
   print "Failed to shut down database\n";
   exit 0;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`./mk.fs.readonly.pl '$oldfs[$x]'`;
   if ($? ne 0) {
      print "Failed to make filesystem $oldfs[$x] readonly\n";
      exit 0;}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`rsync -rlpogt --stats --progress --exclude='.snapshot'
'$oldfs[$x]/' '/$newfs[$x]/'`;
   print @out;
   if ($? ne 0) {
      print "Failed to copy filesystem $oldfs[$x] to $newfs[$x]\n";
      exit 0;}
   else {
      print "Succesfully replicated filesystem $oldfs[$x] to
$newfs[$x]\n";}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   print "swap $x $oldfs[$x] $newfs[$x]\n";
   my @out=`./swap.fs.pl '$oldfs[$x],$newfs[$x]'`;
   print @out;
   if ($? ne 0) {
      print "Failed to swap filesystem $oldfs[$x] for $newfs[$x]\n";
      exit 1;}
   else {
      print "Swapped filesystem $oldfs[$x] for $newfs[$x]\n";}
   $x+=1;}
my @out=`./dbstart.pl '$oraclesid'`;
```

```
print @out;
```

ファイルの場所を表示する

このスクリプトは、多数の重要なデータベースパラメータを収集し、読みやすい形式で出力します。このスクリプトは、データレイアウトを確認する場合に役立ちます。また、他の用途に合わせてスクリプトを変更することもできます。

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;};
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print "$oraclesid datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
```

```
@out=dosql('select member from v\\\\\$logfile;');
print "$oraclesid redo logs:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name from v\\\\\$tempfile;');
print "$oraclesid temp datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('show parameter spfile;');
print "$oraclesid spfile\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name||\' \'||value from v\\\\\$parameter where
isdefault=\'FALSE\';');
print "$oraclesid key parameters\n";
for $line (@out) {
        chomp($line);
        if ($line =~ /control_files/) {print "$line\n";}
        if ($line =~ /db_create/) {print "$line\n";}
        if ($line =~ /db_file_name_convert/) {print "$line\n";}
        if ($line =~ /log_archive_dest/) {print "$line\n";}}
        if ($line =~ /log_file_name_convert/) {print "$line\n";}
        if ($line =~ /pdb_file_name_convert/) {print "$line\n";}
        if ($line =~ /spfile/) {print "$line\n";}
print "\n";
```

ASM移行のクリーンアップ

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
```

```
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;}
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print @out;
print "shutdown immediate;\n";
print "startup mount;\n";
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "host mv $line $1$newname\n";}}
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "alter database rename file '$line' to
'$1$newname';\n";}}
```

```
print "alter database open;\n";
print "\n";
```

ASMからファイルシステム名への変換

```
set serveroutput on;
set wrap off;
declare
    cursor df is select file#, name from v$datafile;
    cursor tf is select file#, name from v$tempfile;
    cursor lf is select member from v$logfile;
    firstline boolean := true;
begin
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('Parameters for log file conversion:');
    dbms_output.put_line(CHR(13));
    dbms_output.put('*.log_file_name_convert = ');
    for lfrec in lf loop
        if (firstline = true) then
            dbms_output.put('''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*/','') || '''');
        else
            dbms_output.put(',''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*/','') || '''');
        end if;
        firstline:=false;
    end loop;
    dbms_output.put_line(CHR(13));
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('rman duplication script:');
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('run');
    dbms_output.put_line('{');
    for dfrec in df loop
        dbms_output.put_line('set newname for datafile ' ||
            dfrec.file# || ' to ''' || dfrec.name ||''';');
    end loop;
    for tfrec in tf loop
        dbms_output.put_line('set newname for tempfile ' ||
            tfrec.file# || ' to ''' || tfrec.name ||''';');
    end loop;
    dbms_output.put_line('duplicate target database for standby backup
location INSERT_PATH_HERE;');
    dbms_output.put_line('}');
end;
/
```

データベースでログを再生

このスクリプトは、マウントモードのデータベースに対してOracle SIDの引数を1つ指定し、現在使用可能なすべてのアーカイブログを再生します。

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
84 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
`;
}
print @out;
```

スタンバイデータベースでログを再生

このスクリプトは、スタンバイデータベース用に設計されている点を除き、上記のスクリプトと同じです。

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
`;
}
print @out;
```

その他の注意事項

Oracleデータベースのパフォーマンス最適化とベンチマーク手順

データベースストレージのパフォーマンスを正確にテストすることは、非常に複雑な課題です。次の問題について理解しておく必要があります。

- IOPSとスループット
- フォアグラウンドI/O処理とバックグラウンドI/O処理の違い
- データベースへのレイテンシの影響
- ストレージのパフォーマンスにも影響する多数のOSとネットワーク設定

また、ストレージデータベース以外のタスクについても考慮する必要があります。ストレージパフォーマンスがパフォーマンスの制限要因ではなくなったため、ストレージパフォーマンスを最適化しても有益なメリットは得られなくなります。

現在、データベースユーザの大半がオールフラッシュアレイを選択していることから、新たな考慮事項がいくつか生まれています。たとえば、2ノードのAFF A900システムでパフォーマンスをテストする場合を考えてみましょう。

- 読み取り/書き込み比率が80対20のA900ノードでは、レイテンシが150 µ sマークを超える前に、100万を超えるランダムデータベースIOPSを達成できます。これは、ほとんどのデータベースで現在必要とされているパフォーマンスをはるかに超えているため、予想される改善を予測することは困難です。ストレージがボトルネックになることはほとんどありません。
- ネットワーク帯域幅は、パフォーマンス上の制約の原因としてますます一般的になっています。たとえば、回転式ディスクソリューションはI/Oレイテンシが非常に高いため、データベースパフォーマンスのボトルネックになることがよくあります。オールフラッシュアレイでレイテンシの制限が取り除かれると、多くの場合、その障壁はネットワークに移ります。これは、真のネットワーク接続を可視化することが困難な仮想環境やブレードシステムで特に顕著です。帯域幅の制限のためにストレージシステム自体をフルに活用できない場合、パフォーマンステストが複雑になる可能性があります。
- オールフラッシュアレイのレイテンシが劇的に改善されるため、オールフラッシュアレイと回転式ディスクを搭載したアレイのパフォーマンスを比較することは一般的に不可能です。通常、テスト結果は意味がありません。
- オールフラッシュアレイでピーク時のIOPSパフォーマンスを比較することは、データベースがストレージI/Oの制約を受けないため、あまり有益なテストではありません。たとえば、あるアレイが50万IOPSを維持でき、別のアレイが30万IOPSを維持できるとします。データベースの処理時間の99%がCPU処理に費やされている場合、この違いは現実の世界では無関係です。ワークロードがストレージアレイのすべての機能を利用することはありません。一方、ストレージアレイの能力を最大限に引き出すことが期待される統合プラットフォームでは、ピーク時のIOPS性能が重要になる場合があります。
- どのストレージテストでも、レイテンシとIOPSを常に考慮してください。市場に出回っているストレージアレイの多くは、非常に高いIOPSを謳っていますが、このレベルのIOPSはレイテンシによって役に立たなくなります。オールフラッシュアレイの一般的なターゲットは1ミリ秒です。テストの優れた方法は、可能な最大IOPSを測定することではなく、平均レイテンシが1ミリ秒を超える前にストレージアレイが維持できるIOPSを特定することです。

Oracle自動ワークロードリポジトリとベンチマーク

Oracleのパフォーマンス比較のゴールドスタンダードは、Oracle Automatic Workload Repository（AWR）レポートです。

AWRレポートには複数のタイプがあります。ストレージの観点から見ると、 `awrrpt.sql` コマンドは、特定のデータベースインスタンスを対象としており、レイテンシに基づいてストレージI/Oイベントを内訳表示する詳細なヒストグラムが含まれているため、最も包括的で価値があります。

2つのパフォーマンスアレイを比較するには、各アレイで同じワークロードを実行し、ワークロードを正確に対象とするAWRレポートを作成するのが理想的です。非常に長時間のワークロードの場合は、開始時間と停止時間を含む経過時間を含む単一のAWRレポートを使用できますが、AWRデータを複数のレポートとして分割することを推奨します。たとえば、バッチジョブが午前0時から午前6時まで実行された場合は、午前0時から午前1時、午前1時から午前2時などの1時間のAWRレポートを作成します。

それ以外の場合は、非常に短いクエリを最適化する必要があります。最適なオプションは、クエリの開始時に作成されたAWRスナップショットと、クエリの終了時に作成された2番目のAWRスナップショットに基づくAWRレポートです。データベースサーバは、分析中のクエリのアクティビティを隠すバックグラウンドアクティビティを最小限に抑えるために、それ以外の場合は静かにしておく必要があります。

AWRレポートを使用できない場合は、代わりにOracle Statspackレポートを使用することを推奨します。AWRレポートとほとんど同じI/O統計情報が含まれています。

Oracle AWRとトラブルシューティング

AWRレポートは、パフォーマンスの問題を分析するための最も重要なツールでもあります。

ベンチマークと同様に、パフォーマンスのトラブルシューティングでは、特定のワークロードを正確に測定する必要があります。可能な場合は、パフォーマンスの問題をNetAppサポートセンターに報告するとき、または新しい解決策についてNetAppまたはパートナーアカウントチームと協力するときにAWRデータを提供してください。

AWRデータを提供する場合は、次の要件を考慮してください。

- を実行します `awrrpt.sql` レポートを生成するコマンド。出力はテキストまたはHTMLのいずれかになります。
- Oracle Real Application Clusters（RAC）を使用する場合は、クラスタ内の各インスタンスについてAWRレポートを生成します。
- 問題が発生した特定の時間をターゲットにします。AWRレポートの最大許容経過時間は、通常1時間です。問題が複数時間続く場合、またはバッチジョブなどの複数時間の操作を伴う場合は、分析対象の期間全体をカバーする複数の1時間のAWRレポートを提供します。
- 可能であれば、AWRスナップショット間隔を15分に調整します。この設定では、より詳細な分析を実行できます。これには、次の追加の実行も必要です。 `awrrpt.sql` 15分間隔ごとにレポートを作成します。
- 実行中のクエリが非常に短い場合は、操作の開始時に作成されたAWRスナップショットと、操作の終了時に作成された2つ目のAWRスナップショットに基づいてAWRレポートを提供します。それ以外の場合は、分析中の操作のアクティビティを隠すバックグラウンドアクティビティを最小限に抑えるために、データベースサーバは静かにしておく必要があります。
- パフォーマンスの問題が特定の時間に報告され、他の時間には報告されない場合は、比較のために優れたパフォーマンスを示す追加のAWRデータを提供します。

キャリブレーション_IO

。 `calibrate_io` コマンドは、ストレージシステムのテスト、比較、ベンチマークには使用しないでください。Oracleのドキュメントに記載されているように、この手順はストレージのI/O機能を調整します。

キャリブレーションはベンチマークと同じではありません。このコマンドの目的は、問題I/Oを使用して、データベース処理を調整し、ホストに対して実行されるI/Oのレベルを最適化することで効率を向上させることです。これは、によって実行されるI/Oのタイプが `calibrate_io` 処理が実際のデータベースユーザI/Oを表しているわけではありません。結果は予測不可能であり、再現さえできないこともよくあります。

SLOB2

Silly Little Oracle BenchmarkであるSLOB2は、データベースのパフォーマンス評価に好まれるツールになりました。Kevin Clossonによって開発され、次のサイトで入手できます。 "https://kevinclosson.net/slob/"。インストールと設定には数分かかり、実際のOracleデータベースを使用してユーザ定義の表領域にI/Oパターンを生成します。オールフラッシュアレイをI/Oで飽和状態にすることができる数少ないテストオプションの1つです。生成されるI/Oのレベルをはるかに低くして、IOPSは低くてもレイテンシの影響を受けやすいストレージワークロードをシミュレートする場合にも役立ちます。

スイングベンチ

Swingbenchはデータベースのパフォーマンスをテストするのに役立ちますが、ストレージに負荷がかかるような方法でSwingbenchを使用することは非常に困難です。NetAppでは、Swingbenchによるテストで、AFF

アレイに多大な負荷をかけるのに十分なI/Oが生成されたことはありません。一部のケースでは、Order Entry Test（OET）を使用してレイテンシの観点からストレージを評価できます。これは、データベースに特定のクエリに対する既知のレイテンシの依存関係がある場合に役立ちます。オールフラッシュアレイの潜在的なレイテンシを実現できるように、ホストとネットワークを適切に設定する必要があります。

HammerDB

HammerDBは、TPC-CやTPC-Hのベンチマークなどをシミュレートするデータベーステストツールです。テストを適切に実行するために十分な大きさのデータセットを構築するには、多くの時間がかかることがありますが、OLTPアプリケーションやデータウェアハウスアプリケーションのパフォーマンスを評価するための効果的なツールになる可能性があります。

オリオン

Oracle OrionツールはOracle 9で一般的に使用されていましたが、さまざまなホストオペレーティングシステムの変更に対応するためにメンテナンスが行われていません。OSやストレージ構成との互換性がないため、Oracle 10やOracle 11で使用されることはほとんどありません。

Oracleはこのツールを書き直し、Oracle 12cにデフォルトでインストールされます。この製品は改良され、実際のOracleデータベースと同じ呼び出しの多くを使用しますが、コードパスやI/O動作はOracleで使用されているものとまったく同じではありません。たとえば、ほとんどのOracle I/Oは同期的に実行されます。つまり、I/O処理はフォアグラウンドで完了するため、I/Oが完了するまでデータベースは停止します。ストレージシステムをランダムI/Oでフラッディングするだけでは、実際のOracle I/Oが再現されるわけではなく、ストレージアレイを比較したり、構成変更の影響を測定したりする直接的な方法もありません。

とはいえ、特定のホスト/ネットワーク/ストレージ構成の最大パフォーマンスの一般的な測定や、ストレージシステムの健全性の測定など、Orionのユースケースもあります。綿密なテストを実施すれば、Orionの使用可能なテストを考案して、ストレージアレイを比較したり、構成変更の影響を評価したりすることができます。ただし、パラメータにIOPS、スループット、レイテンシを考慮し、現実的なワークロードを忠実にレプリケートしようとする必要があります。

古い**NFSv3**ロックと**Oracle**データベース

Oracleデータベースサーバがクラッシュすると、再起動時に古いNFSロックで問題が発生する可能性があります。この問題は、サーバの名前解決を注意深く設定することで回避できます。

この問題は、ロックの作成とロックの解除で使用される名前解決方法がわずかに異なるために発生します。Network Lock Manager（NLM；ネットワークロックマネージャ）とNFSクライアントの2つのプロセスが関係しています。NLMでは、 `uname n` ホスト名を確認するには、 `rpc.statd` プロセスの用途 `gethostbyname()`。OSが古いロックを適切に解除するには、これらのホスト名が一致している必要があります。たとえば、ホストで所有されているロックが検索されているとします。 `dbserver5`が、ロックはホストによって次のように登録されています。 `dbserver5.mydomain.org`。状況 `gethostbyname()` と同じ値を返さない `uname –a`をクリックすると、ロック解除プロセスが成功しませんでした。

次のサンプルスクリプトは、名前解決が完全に一貫しているかどうかを検証します。

```
#! /usr/bin/perl
$uname=`uname -n`;
chomp($uname);
($name, $aliases, $addrtype, $length, @addrs) = gethostbyname $uname;
print "uname -n yields: $uname\n";
print "gethostbyname yields: $name\n";
```

状況 `gethostbyname` 一致しません `uname`古いロックが使用されている可能性があります。たとえば、次の結果は潜在的な問題を示しています。

```
uname -n yields: dbserver5
gethostbyname yields: dbserver5.mydomain.org
```

解決策は通常、に表示されるホストの順序を変更することによって検出されます。 `/etc/hosts`。たとえば、hostsファイルに次のエントリが含まれているとします。

```
10.156.110.201  dbserver5.mydomain.org dbserver5 loghost
```

この問題を解決するには、完全修飾ドメイン名と短いホスト名の表示順序を変更します。

```
10.156.110.201  dbserver5 dbserver5.mydomain.org loghost
```

`gethostbyname()` では、shortを返します。 `dbserver5` ホスト名（の出力に一致） `uname`。したがって、ロックはサーバクラッシュ後に自動的にクリアされます。

OracleデータベースのWAFLアライメント検証

優れたパフォーマンスを実現するには、WAFLを正しくアライメントすることが重要です。ONTAPはブロックを4KB単位で管理しますが、すべての処理がONTAPで4KB単位で実行されるわけではありません。実際、ONTAPはさまざまなサイズのブロック処理に対応しますが、基盤となる計算処理はWAFLによって4KB単位で管理されます。

「アライメント」という用語は、Oracle I/Oがこれらの4KBユニットにどのように対応するかを意味します。パフォーマンスを最適化するには、ドライブ上の2つの4KB WAFL物理ブロックにOracleの8KBブロックが配置されている必要があります。1つのブロックが2KBずれて配置されると、このブロックは1つの4KBブロックの半分、別の4KBブロック全体、3つ目の4KBブロックの半分に配置されます。このように配置すると、パフォーマンスが低下します。

NASファイルシステムでは、アライメントは問題になりません。Oracleデータファイルは、Oracleブロックのサイズに基づいてファイルの先頭にアライメントされます。したがって、8KB、16KB、32KBのブロックサイズは常にアライメントされます。すべてのブロック処理は、ファイルの先頭から4KB単位でオフセットされます。

一方、LUNの開始位置には何らかのドライバヘッダーやファイルシステムのメタデータが含まれているため、

オフセットが作成されます。最新のOSでは、アライメントが問題になることはほとんどありません。最新のOSは、標準の4KBセクターを使用する物理ドライブ向けに設計されており、パフォーマンスを最適化するためにI/Oを4KBの境界にアライメントする必要があるためです。

ただし、いくつかの例外があります。4KB I/O用に最適化されていない古いOSからデータベースが移行された場合や、パーティション作成時のユーザエラーによって4KB単位以外のオフセットが発生した場合があります。

以下はLinux固有の例ですが、手順はどのOSにも適用できます。

アライメント済み

次の例は、パーティションが1つの単一のLUNでアライメントチェックを示しています。

まず、ドライブで使用可能なすべてのパーティションを使用するパーティションを作成します。

```
[root@host0 iscsi]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
disklabel
Building a new DOS disklabel with disk identifier 0xb97f94c1.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.
The device presents a logical sector size that is smaller than
the physical sector size. Aligning to a physical sector (or optimal
I/O) size boundary is recommended, or performance may be impacted.
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-10240, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-10240, default 10240):
Using default value 10240
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
[root@host0 iscsi]#
```

アライメントは、次のコマンドを使用して数学的にチェックできます。

```
[root@host0 iscsi]# fdisk -u -l /dev/sdb
Disk /dev/sdb: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 4096 bytes
I/O size (minimum/optimal): 4096 bytes / 65536 bytes
Disk identifier: 0xb97f94c1
   Device Boot      Start         End      Blocks   Id  System
/dev/sdb1              32    20971519    10485744   83  Linux
```

出力は、単位が512バイトで、パーティションの開始が32ユニットであることを示しています。これは32 x 512 = 16、834バイトで、これは4KBのWAFLブロックの倍数です。このパーティションは正しくアライメントされています。

アライメントが正しいことを確認するには、次の手順を実行します。

1. LUNのUniversally Unique Identifier（UUID）を特定します。

```
FAS8040SAP::> lun show -v /vol/jfs_luns/lun0
              Vserver Name: jfs
                  LUN UUID: ed95d953-1560-4f74-9006-85b352f58fcd
                    Mapped: mapped`                 `
```

2. ONTAPコントローラでノードシェルを開始します。

```
FAS8040SAP::> node run -node FAS8040SAP-02
Type 'exit' or 'Ctrl-D' to return to the CLI
FAS8040SAP-02> set advanced
set not found.  Type '?' for a list of commands
FAS8040SAP-02> priv set advanced
Warning: These advanced commands are potentially dangerous; use
         them only when directed to do so by NetApp
         personnel.
```

3. 最初の手順で特定したターゲットUUIDで統計収集を開始します。

```
FAS8040SAP-02*> stats start lun:ed95d953-1560-4f74-9006-85b352f58fcd
Stats identifier name is 'Ind0xffffff08b9536188'
FAS8040SAP-02*>
```

4. I/Oを実行します。次のツールを使用することが重要です。 `iflag` I/Oが同期でバッファされていないことを確認する引数。

このコマンドには十分注意してください。の反転 `if` および `of` 引数はデータを破棄します。

```
[root@host0 iscsi]# dd if=/dev/sdb1 of=/dev/null iflag=dsync count=1000
bs=4096
1000+0 records in
1000+0 records out
4096000 bytes (4.1 MB) copied, 0.0186706 s, 219 MB/s
```

5. 統計を停止し、アライメントのヒストグラムを表示します。すべてのI/Oが `.0` Bucket。4KBのブロック境界にアライメントされたI/Oを示します。

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff08b9536188
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:186%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
```

ミスアライメント状態です

次の例は、ミスアライメントI/Oを示しています。

1. 4KBの境界にアライメントされないパーティションを作成します。最新のOSでは、これはデフォルトの動作ではありません。

```
[root@host0 iscsi]# fdisk -u /dev/sdb
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First sector (32-20971519, default 32): 33
Last sector, +sectors or +size{K,M,G} (33-20971519, default 20971519):
Using default value 20971519
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
```

2. パーティションは、デフォルトの32ではなく33セクターオフセットで作成されています。で説明されている手順を繰り返します。 "アライメント済み"。ヒストグラムは次のように表示されます。

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:136%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_partial_blocks:31%
```

ミスアライメントは明らかです。I/Oの大部分は* *.1 バケット。想定されるオフセットに一致します。パーティションが作成されたときに、最適化されたデフォルトよりも512バイト先のデバイスに移動されました。これは、ヒストグラムが512バイトオフセットされることを意味します。

また、も参照してください `read_partial_blocks` 統計がゼロ以外の場合は、実行されたI/Oが4KBブロック全体を一杯にしなかったことを意味します。

Redoロギング

ここで説明する手順はデータファイルに適用できます。OracleのREDOログとアーカイブログでは、I/Oパターンが異なります。たとえば、Redoロギングでは、単一ファイルを繰り返し上書きします。デフォルトの512バイトのブロックサイズを使用する場合、書き込み統計は次のようになります。

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-4f74-
9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.0:12%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.1:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.3:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.4:13%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.5:6%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.6:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.7:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_partial_blocks:85%
```

I/Oはすべてのヒストグラムバケットに分散されますが、これはパフォーマンス上の問題ではありません。ただし、4KBのブロックサイズを使用すると、Redoロギング率が非常に高くなる場合があります。この場合は、RedoロギングLUNが適切にアライメントされていることを確認することを推奨します。ただし、これは優れたパフォーマンスにとってデータファイルのアライメントほど重要ではありません。

PostgreSQL

ONTAP上のPostgreSQLデータベース

PostgreSQLには、PostgreSQL、PostgreSQL Plus、EDB Postgres Advanced Server（EPAS）などのバリアントが付属しています。PostgreSQLは通常、多層アプリケーションのバックエンドデータベースとして導入されます。一般的なミドルウェアパッケージ(PHP、Java、Python、Tcl/Tk、ODBCなど)でサポートされています。 とJDBC）は、オープンソースのデータベース管理システムでは、歴史的に人気のある選択肢でした。ONTAPは、信頼性、パフォーマンス、効率に優れたデータ管理機能を備えたPostgreSQLデータベースを実行するための優れた選択肢です。

ONTAPおよびPostgreSQLデータベースに関するこのドキュメントは、以前に公開されていた_TR-4770：『PostgreSQL database on ONTAP best practices』に代わるものです。_

データが指数関数的に増加するにつれて、企業にとってデータ管理はより複雑になります。この複雑さにより、ライセンス、運用、サポート、メンテナンスのコストが増大します。全体的なTCOを削減するには、信頼性とパフォーマンスに優れたバックエンドストレージを使用して、商用データベースからオープンソースデータベースに切り替えることを検討してください。

ONTAPは理想的なプラットフォームです。ONTAPは文字通りデータベース向けに設計されているからです。データベースワークロードのニーズに対応するために、高度なQuality of Service（QoS；サービス品質）機能や基本的なFlexClone機能に対するランダムI/Oレイテンシの最適化など、多数の機能が特別に開発されました。

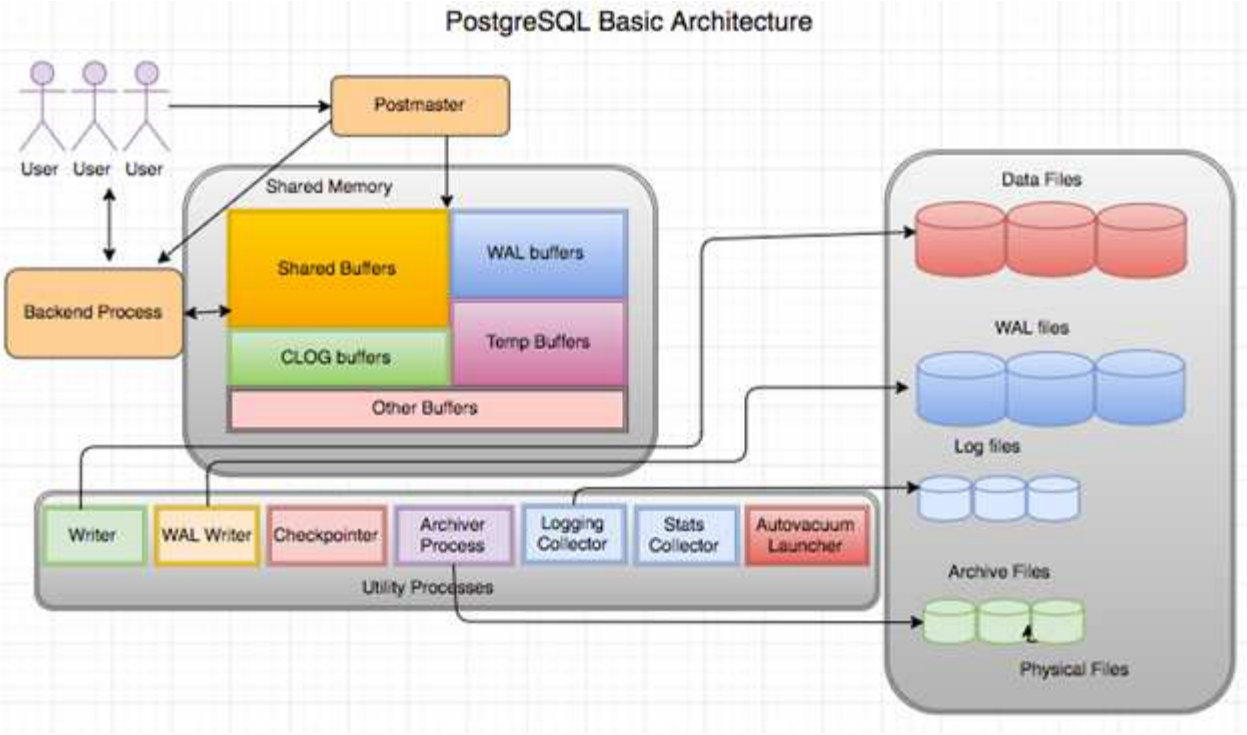
無停止アップグレード（ストレージの交換など）などの追加機能により、重要なデータベースの可用性を維持できます。また、MetroClusterを使用して大規模な環境で瞬時にディザスタリカバリを実行したり、SnapMirrorアクティブ同期を使用してデータベースを選択したりすることもできます。

最も重要なことは、ONTAPが卓越したパフォーマンスを提供し、お客様固有のニーズに合わせて解決策をサイジングできることです。ネットアップのハイエンドシステムは100万超のIOPSをマイクロ秒単位のレイテンシで提供できますが、必要なIOPSが10万だけの場合は、同じストレージオペレーティングシステムを実行する小型のコントローラでストレージ解決策を適切にサイジングできます。

データベース設定

PostgreSQLアーキテクチャ

PostgreSQLは、クライアントとサーバのアーキテクチャに基づいたRDBMSです。PostgreSQLインスタンスはデータベースクラスタと呼ばれ、サーバの集合ではなくデータベースの集合です。



PostgreSQLデータベースには、postmaster、フロントエンド(クライアント)、バックエンドの3つの要素があります。クライアントは、IPプロトコルや接続先データベースなどの情報を含む要求をポストマスターに送信します。postmasterは接続を認証し、さらに通信するためにバックエンドプロセスに渡します。バックエンドプロセスはクエリを実行し、結果を直接フロントエンド（クライアント）に送信します。

PostgreSQLインスタンスは、マルチスレッドモデルではなく、マルチプロセスモデルに基づいています。ジョブごとに複数のプロセスが生成され、各プロセスには独自の機能があります。主なプロセスには、クライアントプロセス、WALライタプロセス、バックグラウンドライタプロセス、およびcheckpointerプロセスが含まれます。

- クライアント（フォアグラウンド）プロセスがPostgreSQLインスタンスに読み取りまたは書き込み要求を送信しても、データを直接ディスクに読み書きすることはありません。最初に、共有バッファとWAL(Write-Ahead Logging)バッファにデータをバッファします。
- WALライタプロセスは、共有バッファとWALバッファの内容を操作してWALログに書き込みます。WALログは通常PostgreSQLのトランザクションログであり、シーケンシャルに書き込まれます。したがって、データベースからの応答時間を短縮するために、PostgreSQLはまずトランザクションログに書き込み、クライアントに確認応答します。
- データベースを整合性のある状態にするために、バックグラウンドライタープロセスは共有バッファにダーティページがないか定期的にチェックします。次に、NetAppボリュームまたはLUNに格納されているデータファイルにデータをフラッシュします。
- checkpointerプロセスも定期的に（バックグラウンドプロセスよりも少ない頻度で）実行され、バッファへの変更を防ぎます。WALライタプロセスに、NetAppディスクに保存されているWALログの末尾にチェックポイントレコードを書き込み、フラッシュするように指示します。また、すべてのダーティページをディスクに書き込み、フラッシュするようにバックグラウンドライタープロセスに通知します。

PostgreSQL初期化パラメータ

新しいデータベースクラスタを作成するには、 `initdb` プログラム。A `initdb` スクリプトは、クラスタを定義するデータファイル、システムテーブル、およびテンプレート

データベース（template0およびtemplate1）を作成します。

テンプレートデータベースはストックデータベースを表します。システムテーブル、標準ビュー、関数、およびデータ型の定義が含まれています。 `pgdata` の引数として機能します。 `initdb` データベースクラスタの場所を指定するスクリプト。

PostgreSQLのすべてのデータベースオブジェクトは、それぞれのOIDによって内部的に管理されます。テーブルとインデックスは、個 々 のOIDによっても管理されます。データベースオブジェクトとそれぞれのOIDとの関係は、オブジェクトのタイプに応じて適切なシステムカタログテーブルに格納されます。たとえば、データベースとヒープテーブルのOIDは、 `pg_database` それぞれ`pg_classと`pg_classです。OIDを確認するには、PostgreSQLクライアントでクエリを発行します。

各データベースには、1GBに制限された個別のテーブルとインデックスファイルがあります。各テーブルには、それぞれサフィックス付きの2つのファイルが関連付けられています。 `_fsm` および `_vm`。これらは、フリースペースマップおよび可視性マップと呼ばれます。これらのファイルには空きスペース容量に関する情報が格納され、テーブルファイルの各ページに表示されます。インデックスには個 々 の空き領域マップのみがあり、可視性マップはありません。

。 `pg_xlog/pg_wal` ディレクトリには、先行書き込みログが格納されます。先行書き込みログは、データベースの信頼性とパフォーマンスを向上させるために使用されます。テーブル内の行を更新するたびに、PostgreSQLは先読みログに変更内容を書き込み、その後実際のデータページに変更内容をディスクに書き込みます。。 `pg_xlog` ディレクトリには通常複数のファイルが含まれていますが、initdbは最初のファイルだけを作成します。必要に応じて追加のファイルが追加されます。各xlogファイルの長さは16MBです。

ONTAPを使用したPostgreSQLデータベースの設定

パフォーマンスを向上させるPostgreSQLのチューニング設定がいくつかあります。

最も一般的に使用されるパラメータは次のとおりです。

- `max_connections = <num>`:一度に持つデータベース接続の最大数。このパラメータを使用して、ディスクへのスワップを制限し、パフォーマンスを強制終了します。アプリケーションの要件に応じて、このパラメータを接続プールの設定に合わせて調整することもできます。
- `shared_buffers = <num>`：データベース・サーバのパフォーマンスを向上させる最も簡単な方法最新のほとんどのハードウェアでは、デフォルトはlowです。導入時に、システム上の使用可能なRAMの約25%に設定されます。このパラメータ設定は、特定のデータベースインスタンスでの動作によって異なります。試行錯誤して値を増減しなければならない場合があります。ただし、この値をHighに設定すると、パフォーマンスが低下する可能性があります。
- `effective_cache_size = <num>`:この値は、PostgreSQLのオプティマイザに、PostgreSQLがデータをキャッシュするために使用できるメモリ量を伝え、インデックスを使用するかどうかを判断するのに役立ちます。値を大きくすると、インデックスを使用する可能性が高くなります。このパラメータは、に割り当てられているメモリの量に設定する必要があります。 `shared_buffers` さらに、使用可能なOSキャッシュの容量も表示されます。多くの場合、この値はシステムメモリ全体の50%を超えています。
- `work_mem = <num>`:このパラメータは、ソート操作およびハッシュテーブルで使用するメモリの量を制御します。アプリケーションで大量のソートを行う場合は、メモリの量を増やす必要があるかもしれませんが、注意が必要です。これはシステム全体のパラメータではなく、操作ごとのパラメータです。複雑なクエリに複数のソート操作が含まれている場合、複数のwork_mem単位のメモリを使用し、複数のバックエンドが同時にこれを実行する可能性があります。このクエリを実行すると'値が大きすぎると'データベース・サーバがスワップされることがよくありますこのオプションは、以前のバージョンのPostgreSQLではsort_memと呼ばれていました。

- `fsync = <boolean> (on or off)`:このパラメータは、トランザクションがコミットされる前にfsync()を使用して、すべてのWALページをディスクに同期するかどうかを決定します。電源をオフにすると、書き込みパフォーマンスが向上し、オンにすると、システムクラッシュ時の破損のリスクからの保護が強化されます。
- `checkpoint_timeout`:チェックポイント・プロセスは'コミットされたデータをディスクにフラッシュしますこれには、ディスク上で多くの読み取り/書き込み処理が含まれます。値は秒単位で設定され、値を小さくするとクラッシュリカバリ時間が短縮されます。値を大きくすると、チェックポイントコールが削減されるため、システムリソースの負荷が軽減されます。アプリケーションの重要度、使用状況、データベースの可用性に応じて、checkpoint_timeoutの値を設定します。
- `commit_delay = <num>` および `commit_siblings = <num>`:これらのオプションを組み合わせて使用すると、一度にコミットする複数のトランザクションを書き出すことで、パフォーマンスを向上させることができます。トランザクションがコミットされた瞬間に複数のcommit_siblingsオブジェクトがアクティブになっている場合、サーバはcommit_delayマイクロ秒を待って、一度に複数のトランザクションをコミットしようとします。
- `max_worker_processes / max_parallel_workers`:プロセスに最適なワーカー数を設定します。max_parallel_workersは、使用可能なCPUの数に対応します。アプリケーションの設計によっては、クエリで並列処理に必要なワーカーの数が少なくて済みます。両方のパラメータの値は同じにし、テスト後に値を調整することをお勧めします。
- `random_page_cost = <num>`:この値は、PostgreSQLが非シーケンシャルディスク読み取りを表示する方法を制御します。値を大きくすると、PostgreSQLはインデックススキャンではなくシーケンシャルスキャンを使用する可能性が高くなります。これは、サーバーに高速ディスクがあることを示します。計画ベースの最適化、真空化、クエリやスキーマの変更に対するインデックス付けなど、他のオプションを評価した後で、この設定を変更してください。
- `effective_io_concurrency = <num>`:このパラメータは、PostgreSQLが同時に実行を試みる同時ディスクI/O処理の数を設定します。この値を大きくすると、個 々 のPostgreSQLセッションが並行して開始しようとするI/O処理の数が増加します。指定できる範囲は1〜1、000です。非同期I/O要求の発行を無効にする場合は0にします。現在、この設定はビットマップヒープスキャンにのみ影響します。ソリッドステートドライブ（SSD）やその他のメモリベースストレージ（NVMe）は、多数の同時要求を処理できることが多いため、数百の数を推奨します。

PostgreSQL設定パラメータの完全なリストについては、PostgreSQLのドキュメントを参照してください。

トースト

TOASTは、特大属性ストレージテクニックを表しています。PostgreSQLは固定のページサイズ（通常は8KB）を使用しており、タプルを複数のページにまたがることはできません。したがって、大きなフィールド値を直接保存することはできません。このサイズを超える行を格納しようとすると、トーストは大きな列のデータを小さな「ピース」に分割してトーストテーブルに格納します。

トーストされた属性の大きな値は'結果セットがクライアントに送信されるときにのみ(選択されている場合)プルアウトされますテーブル自体は非常に小さく、アウトオブラインストレージ（TOAST）を使用しない場合よりも多くの行を共有バッファキャッシュに格納できます。

バキューム

通常のPostgreSQL操作では、更新によって削除または廃止されたタプルはテーブルから物理的に削除されず、VACUUMが実行されるまで存在したままになります。したがって、特に頻繁に更新されるテーブルでは、VACUMを定期的に実行する必要があります。ディスクスペースが使用されているスペースは、ディスクスペースが停止しないように、新しい行で再利用できるように再利用する必要があります。ただし、スペースはオペレーティングシステムに返されません。

ページ内の空き領域は断片化されません。VACUUMはブロック全体を書き換え、残りの行を効率的にパッキングし、1つの連続した空き領域ブロックをページに残します。

一方、VACUUM FULLは、デッドスペースのないまったく新しいバージョンのテーブルファイルを作成することで、テーブルを積極的に圧縮します。この操作により、テーブルのサイズは最小限に抑えられますが、時間がかかることがあります。また、処理が完了するまで、テーブルの新しいコピー用に追加のディスクスペースが必要になります。ルーチンバキュームの目的は、バキュームフルアクティビティを回避することです。このプロセスでは、テーブルが最小サイズに維持されるだけでなく、ディスクスペースの安定した使用量も維持されます。

PostgreSQLテーブルスペース

データベース・クラスタが初期化されると、2つの表領域が自動的に作成されます。

。 `pg_global` 表領域は共有システムカタログに使用されます。。 `pg_default` tablespaceは、template1およびtemplate0データベースのデフォルトのテーブルスペースです。クラスタが初期化されたパーティションまたはボリュームの容量が不足し、拡張できない場合は、別のパーティションに表領域を作成して、システムを再構成できるようになるまで使用できます。

頻繁に使用されるインデックスは、ソリッドステートデバイスのような高速で可用性の高いディスクに配置できます。また、ほとんど使用されない、またはパフォーマンスが重要でないアーカイブデータを格納するテーブルは、SASドライブやSATAドライブなどの低コストで低速なディスクシステムに格納できます。

表領域はデータベースクラスタの一部であり、データファイルの自律的なコレクションとして扱うことはできません。これらは、メインデータディレクトリに含まれるメタデータに依存するため、別のデータベースクラスタに接続したり、個別にバックアップしたりすることはできません。同様に、（ファイル削除やディスク障害などによって）テーブルスペースが失われると、データベースクラスタが読み取り不能になったり、起動できなくなったりすることがあります。RAMディスクのような一時ファイルシステムに表領域を配置すると、クラスタ全体の信頼性が低下します。

作成後、要求元ユーザに十分な権限があれば、任意のデータベースから表領域を使用できます。PostgreSQLは、テーブルスペースの実装を簡素化するためにシンボリックリンクを使用します。PostgreSQLは、 `pg_tablespace` Table（クラスタ全体のテーブル）を作成し、その行に新しいオブジェクト識別子（OID）を割り当てます。最後に、サーバはOIDを使用して、クラスタと指定されたディレクトリの間にシンボリックリンクを作成します。ディレクトリ `$PGDATA/pg_tblspc` クラスタで定義されている組み込み以外の各表領域を参照するシンボリックリンクが含まれます。

ストレージ構成

NFSファイルシステムを使用したPostgreSQLデータベース

PostgreSQLデータベースは、NFSv3またはNFSv4ファイルシステムでホストできます。最適なオプションは、データベース外の要因によって異なります。

たとえば、特定のクラスタ環境ではNFSv4のロック動作が推奨されます。（参照： "こちらをご覧ください" 詳細はこちら）

それ以外の場合は、パフォーマンスも含めて、データベース機能はほぼ同一である必要があります。唯一の要件は、 `hard` マウントオプション。これは、ソフトタイムアウトによって回復不能なIOエラーが発生しないようにするために必要です。

NFSv4がプロトコルとして選択されている場合、NetAppではNFSv4.1の使用を推奨します。NFSv4.1では、NFSv4.0よりも耐障害性が向上するように、NFSv4プロトコルの機能がいくつか拡張されています。

一般的なデータベースワークロードには、次のマウントオプションを使用します。

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=65536,wsize=65536
```

大量のシーケンシャルIOが予想される場合は、次のセクションの説明に従ってNFS転送サイズを増やすことができます。

NFSテンソウサイス

ONTAPでは、デフォルトでNFS I/Oサイズが64Kに制限されています。

ほとんどのアプリケーションとデータベースでランダムI/Oを実行すると、ブロックサイズがはるかに小さくなり、最大64Kよりもはるかに小さくなります。ラージブロックI/Oは通常並列処理されるため、最大64Kも最大帯域幅の確保に制限されるわけではありません。

一部のワークロードでは、最大64Kに制限があります。特に、バックアップ/リカバリ処理やデータベースのフルテーブルスキャンなどのシングルスレッド処理は、実行回数が少なくても大容量のI/Oを実行できるのであれば、より高速かつ効率的に実行できます。ONTAPに最適なI/O処理サイズは256Kです。

特定のONTAP SVMの最大転送サイズは、次のように変更できます。

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

注意
ONTAPで許容される最大転送サイズを、現在マウントされているNFSファイルシステムのrsize/wsizeの値より小さくしないでください。これにより、一部のオペレーティングシステムでハングしたり、データが破損したりする可能性があります。たとえば、NFSクライアントのrsize / wsizeが65536に設定されている場合は、クライアント自体が制限されているため、ONTAPの最大転送サイズを65536~1048576の間で調整しても効果はありません。最大転送サイズを65536未満に縮小すると、可用性やデータが損傷する可能性があります。

転送サイズをONTAPレベルで拡張すると、次のマウントオプションが使用されます。

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=262144,wsize=262144
```

NFSv3 TCPスロットテーブル

LinuxでNFSv3を使用する場合は、TCPスロットテーブルを適切に設定することが重要です。

TCPスロットテーブルは、NFSv3でホストバスアダプタ（HBA）のキュー深度に相当します。一度に未処理となることのできるNFS処理の数を制御します。デフォルト値は通常16ですが、最適なパフォーマンスを得るには小さすぎます。逆に、新しいLinuxカーネルでTCPスロットテーブルの上限をNFSサーバが要求でいっぱいになるレベルに自動的に引き上げることができるため、問題が発生します。

パフォーマンスを最適化し、パフォーマンスの問題を回避するには、TCPスロットテーブルを制御するカーネルパラメータを調整します。

を実行します `sysctl -a | grep tcp.*.slot_table` コマンドを実行し、次のパラメータを確認します。

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

すべてのLinuxシステムに `sunrpc.tcp_slot_table_entries`ただし、次のようなものがあります。 `sunrpc.tcp_max_slot_table_entries`。どちらも128に設定する必要があります。

注意
これらのパラメータを設定しないと、パフォーマンスに大きく影響する可能性があります。Linux OSが十分なI/Oを発行していないためにパフォーマンスが制限される場合もあります。一方では、Linux OSが問題で処理できる以上のI/Oを試行すると、I/Oレイテンシが増加します。

SANファイルシステムを使用するPostgreSQL

SANを使用したPostgreSQLデータベースは、通常xfsファイルシステムでホストされますが、OSベンダーがサポートしていれば他のデータベースも使用できます。

1つのLUNで最大10万IOPSをサポートできますが、I/O負荷の高いデータベースでは、一般にLVMとストライピングを使用する必要があります。

LVMストライピング

フラッシュドライブが登場する以前は、回転式ドライブのパフォーマンス上の制限を克服するためにストライピングが使用されていました。たとえば、OSが1MBの読み取り操作を実行する必要がある場合、1つのドライブからその1MBのデータを読み取るには、1MBがゆっくり転送されるため、多くのドライブヘッドのシークと読み取りが必要になります。この1MBのデータが8つのLUNにストライピングされている場合、OSは8つの128K読み取り処理を並行して問題できるため、1MB転送の完了に必要な時間が短縮されます。

回転式ドライブを使用したストライピングは、I/Oパターンを事前に把握しておく必要があったため、より困難でした。ストライピングが実際のI/Oパターンに合わせて正しく調整されていない場合、ストライピングされた構成ではパフォーマンスが低下する可能性があります。Oracleデータベース、特にオールフラッシュ構成では、ストライピングは設定がはるかに簡単で、パフォーマンスが劇的に向上することが実証されています。

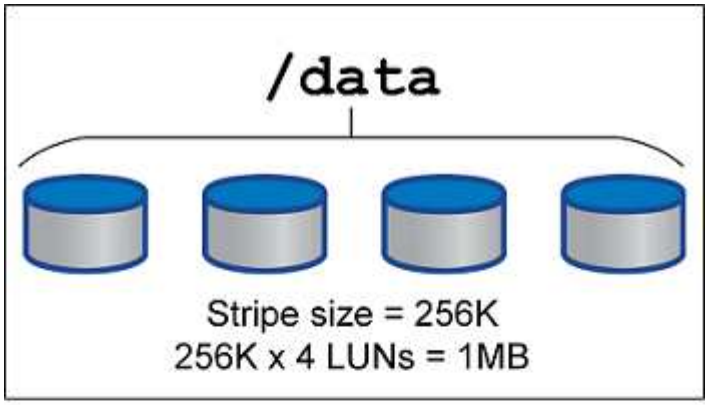
デフォルトではOracle ASMなどの論理ボリュームマネージャがストライプされますが、ネイティブOS LVMはストライプされません。その中には、複数のLUNを連結されたデバイスとして結合するものもあります。その

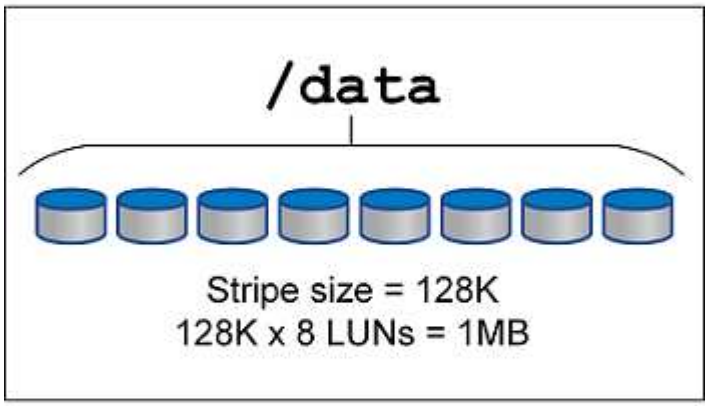
ため、データファイルは1つのLUNデバイスにしか存在しません。これにより、ホットスポットが発生します。他のLVM実装では、デフォルトで分散エクステントが使用されます。これはストライピングに似ていますが、粗いです。ボリュームグループ内のLUNはエクステントと呼ばれる大きな部分にスライスされ、通常は数メガバイト単位で測定され、論理ボリュームがそれらのエクステントに分散されます。その結果、ファイルに対するランダムI/OはLUN間で適切に分散されますが、シーケンシャルI/O処理はそれほど効率的ではありません。

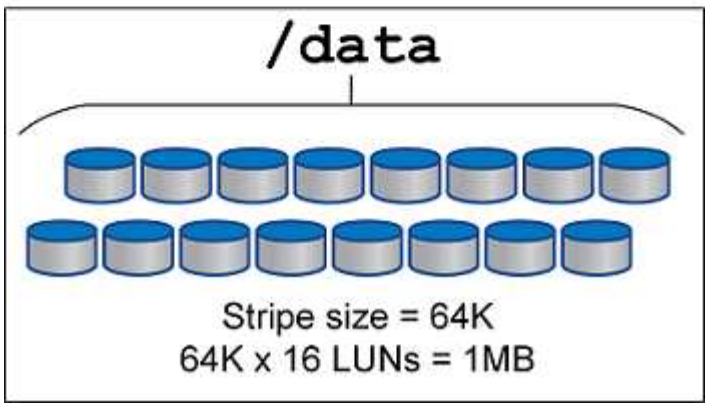
高いパフォーマンスを必要とするアプリケーションI/Oは、ほとんどの場合（a）基本ブロックサイズの単位または（b）1メガバイトのいずれかです。

ストライピング構成の主な目的は、シングルファイルI/Oを1つのユニットとして実行し、マルチブロックI/O（サイズは1MB）をストライピングされたボリューム内のすべてのLUNで均等に並列化できるようにすることです。つまり、ストライプ・サイズはデータベース・ブロック・サイズより小さくすることはできず、ストライプ・サイズにLUN数を掛けたサイズは1MBにする必要があります。

次の図に、ストライプサイズと幅の調整に使用できる3つのオプションを示します。LUNの数は、前述のパフォーマンス要件を満たすように選択されますが、いずれの場合も、1つのストライプ内の総データ量は1MBです。







データ保護

PostgreSQLのデータ保護

ストレージ設計の主な側面の1つは、PostgreSQLボリュームの保護を有効にすることです。お客様は、ダンプアプローチを使用するか、ファイルシステムバックアップを使用して、PostgreSQLデータベースを保護できます。このセクションでは、個 々 のデータベースまたはクラスタ全体をバックアップするさまざまな方法について説明します。

PostgreSQLデータをバックアップするには、次の3つの方法があります。

- SQL Serverダンプ
- ファイルシステムレベルのバックアップ
- 継続的アーカイブ

SQL Serverダンプ方式の背後にある考え方は、SQL Serverコマンドを使用してファイルを生成することです。このコマンドをサーバに返すと、ダンプ時と同じようにデータベースを再作成できます。PostgreSQLはユーティリティプログラムを提供します。 `pg_dump` および `pg_dump_all` 個 々 のバックアップとクラスタレベルのバックアップの作成に使用します。これらのダンプは論理的であり、WAL再生で使用するのに十分な情報が含まれていません。

別のバックアップ戦略として、PostgreSQLがデータベースにデータを保存するために使用するファイルを管理者が直接コピーするファイルシステムレベルのバックアップを使用する方法があります。この方法はオフラインモードで実行されます。データベースまたはクラスタをシャットダウンする必要があります。もう1つの選択肢は、 `pg_basebackup` PostgreSQLデータベースのホットストリーミングバックアップを実行します。

PostgreSQLデータベースとストレージスナップショット

PostgreSQLを使用したスナップショットベースのバックアップでは、フルリカバリまたはポイントインタイムリカバリを提供するために、データファイル、WALファイル、およびアーカイブされたWALファイルのスナップショットを構成する必要があります。

PostgreSQLデータベースの場合、Snapshotを使用した平均バックアップ時間は数秒~数分です。このバックアップ速度は、 `pg_basebackup` その他のファイル・システム・ベースのバックアップ・アプローチ

NetAppストレージ上のSnapshotは、crash-consistentとアプリケーション整合性の両方が可能です。crash-consistent Snapshotはデータベースを休止せずにストレージ上に作成されますが、アプリケーション整合性Snapshotはデータベースがバックアップモードの間に作成されます。NetAppでは、後続のスナップショットが永久増分バックアップとなるため、ストレージの節約とネットワークの効率化が促進されます。

スナップショットは高速で、システムのパフォーマンスに影響を与えないため、他のストリーミングバックアップテクノロジのように1日1回のバックアップを作成するのではなく、1日に複数のスナップショットをスケジュールできます。リストアとリカバリの処理が必要な場合は、次の2つの主な機能によってシステムのダウンタイムが短縮されます。

- NetApp SnapRestoreのデータリカバリテクノロジにより、リストア処理が数秒で実行されます。
- Recovery Point Objective（RPO；目標復旧時点）が頻繁に発生するため、適用するデータベースログの数が減り、フォワードリカバリも高速化されます。

PostgreSQLをバックアップするには、データボリュームが（コンシステンシグループ）WALとアーカイブログと同時に保護されていることを確認する必要があります。Snapshotテクノロジを使用してWALファイルをコピーする場合は、次のコマンドを実行してください： `pg_stop` アーカイブする必要があるすべてのWALエントリをフラッシュします。リストア中にWALエントリをフラッシュする場合は、データベースを停止するか、既存のデータディレクトリをアンマウントまたは削除し、ストレージでSnapRestore操作を実行するだけで済みます。リストアが完了したら、システムをマウントして現在の状態に戻すことができます。ポイントインタイムリカバリの場合は、WALとアーカイブログをリストアすることもできます。次に、PostgreSQLは最も整合性のあるポイントを決定し、自動的にリカバリします。

整合グループはONTAPの機能であり、1つのインスタンスまたは複数の表領域を含むデータベースに複数のボリュームがマウントされている場合に推奨されます。整合性グループSnapshotを使用すると、すべてのボリュームがグループ化されて保護されます。整合グループはONTAP System Managerから効率的に管理できます。また、整合グループをクローニングして、テストや開発用にデータベースのインスタンスコピーを作成することもできます。

コンシステンシグループの詳細については、を参照してください。 "NetApp整合性グループの概要"。

PostgreSQLデータ保護ソフトウェア

PostgreSQLデータベース用のNetApp SnapCenterプラグインをSnapshotおよびNetApp FlexCloneテクノロジと組み合わせると、次のようなメリットがあります。

- 高速なバックアップとリストア：
- スペース効率に優れたクローン：
- 迅速で効果的なディザスタリカバリシステムを構築する能力。

> ⓘ 次のような状況では、Veeam SoftwareやCommvaultなど、ネットアップのプレミアムバックアップパートナーを選択することもできます。
>
> - 異機種混在環境全体でのワークロードの管理
> - バックアップをクラウドまたはテープに保存して長期保持
> - さまざまなバージョンと種類のOSをサポート

PostgreSQL用のSnapCenterプラグインはコミュニティでサポートされているプラグインであり、セットアップとドキュメントはNetAppオートメーションストアで入手できます。SnapCenterを使用すると、データベースのバックアップ、データのクローニング、リストアをリモートで実行できます。

SAP

VMware

ONTAP を使用した VMware vSphere

ONTAP を使用した VMware vSphere

ONTAPは、約20年にわたって業界をリードするVMware vSphere環境向けストレージ解決策であり、コストを削減しながら管理を簡易化する革新的な機能を継続的に追加しています。このドキュメントでは、導入の合理化、リスクの軽減、管理の簡易化を実現するために、最新の製品情報とベストプラクティスを含む ONTAP 解決策 for vSphere について説明します。

以前に公開されていたテクニカルレポート_TR-4597：『VMware vSphere for ONTAP』をこのドキュメントに差し替えます。

ベストプラクティスは、ガイドや互換性リストなどの他のドキュメントを補うものです。ラボテストに基づいて開発されており、ネットアップのエンジニアやお客様は広範な現場経験を積んでいます。すべての環境で機能する唯一のサポート対象となるわけではありませんが、一般に、ほとんどのお客様のニーズを満たす最もシンプルなソリューションです。

本ドキュメントでは、vSphere 7.0以降で実行されるONTAPの最新リリース（9.x）の機能について説明します。を参照してください "NetApp Interoperability Matrix Tool で確認できます" および "VMware Compatibility Guide" 特定のリリースに関する詳細については、を参照してください。

ONTAP for vSphere を選ぶ理由

ONTAPをvSphereのストレージ解決策として選択した理由は数多くあります。たとえば、SANとNASの両方のプロトコルをサポートするユニファイドストレージシステム、スペース効率に優れたSnapshotを使用した堅牢なデータ保護機能、アプリケーションデータの管理に役立つ豊富なツールなどです。ハイパーバイザーとは別のストレージシステムを使用すると、さまざまな機能をオフロードして、 vSphere ホストシステムへの投資を最大限に活用できます。このアプローチにより、ホストリソースをアプリケーションワークロードに集中できるだけでなく、ストレージ運用によるアプリケーションのランダムなパフォーマンスへの影響も回避できます。

vSphere と ONTAP を併用すると、ホストハードウェアと VMware ソフトウェアのコストを削減できます。また、一貫した高パフォーマンスを維持しながら、低コストでデータを保護することもできます。仮想化されたワークロードはモバイル対応であるため、 Storage vMotion を使用して、 VMFS 、 NFS 、または VVOL データストア間で VM を移動するさまざまなアプローチを、すべて同じストレージシステム上で検討できます。

お客様が現在重視している主な要因は次のとおりです。

- * ユニファイド・ストレージ。 * ONTAP ソフトウェアを実行するシステムは、いくつかの重要な方法で統合されています。当初、このアプローチは NAS プロトコルと SAN プロトコルの両方を指していましたが、 ONTAP は業界をリードする SAN プラットフォームであり続けており、 NAS における従来の強みもあります。vSphere 環境では、このアプローチは仮想デスクトップインフラ（ VDI ）向けのユニファイドシステムと仮想サーバインフラ（ VSI ）の組み合わせを意味する場合もあります。ONTAP ソフトウェアを実行するシステムは一般に、従来のエンタープライズアレイに比べて VSI の方が安価ですが、同じシステムで VDI を処理するための高度な Storage Efficiency 機能も備えています。また、 ONTAP は、 SSD から SATA までさまざまなストレージメディアを統合し、クラウドへの拡張を容易にします。パフォーマンスのためにフラッシュアレイを1つ、アーカイブ用にSATAアレイを1つ、クラウド用に別 々 のシステムを

購入する必要はありません。ONTAP は、これらすべてを 1 つにまとめます。

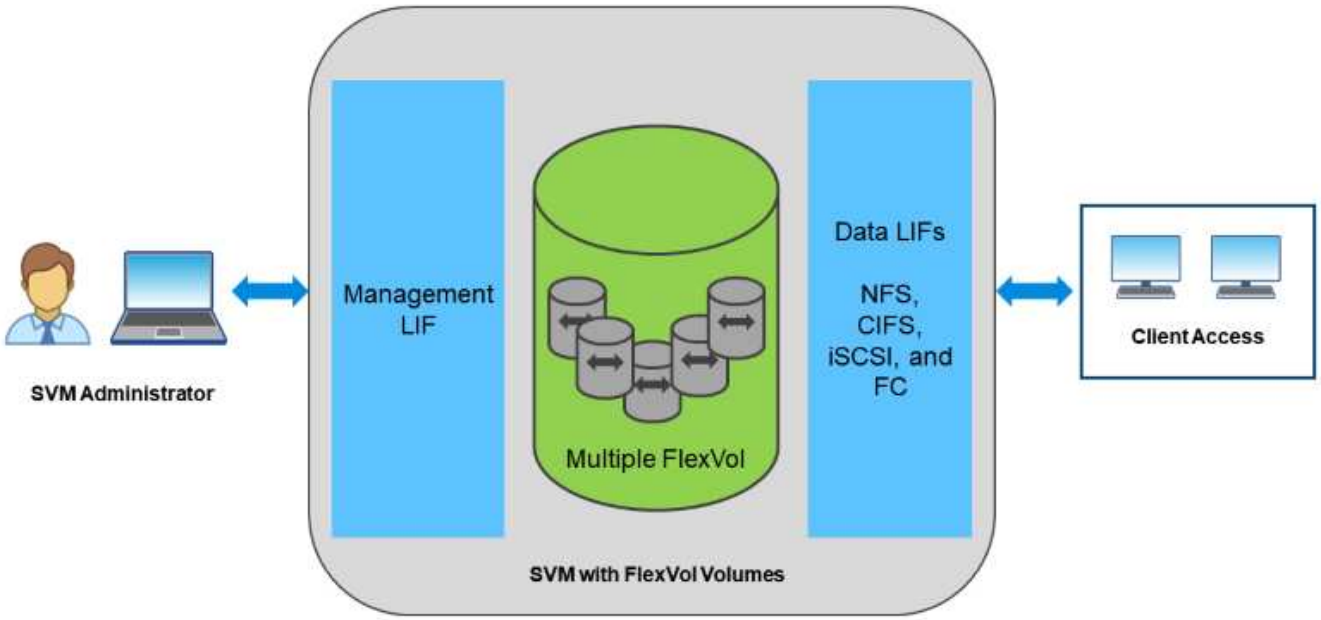
- 仮想ボリュームとストレージポリシーベースの管理。 NetAppは、vSphere Virtual Volume（VVOL）の開発においてVMwareの初期の設計パートナーであり、アーキテクチャに関する情報を提供し、VVOLとVMware vSphere APIs for Storage Awareness（VASA）を早期にサポートしています。このアプローチにより、VMFSでVMストレージをきめ細かく管理できるだけでなく、ストレージポリシーベースの管理によるストレージプロビジョニングの自動化もサポートされました。このアプローチにより、ストレージアーキテクトは、 VM 管理者が簡単に利用できるさまざまな機能を備えたストレージプールを設計できます。ONTAP は VVOL 規模でストレージ業界をリードし、 1 つのクラスタで数十万もの VVol をサポートします。一方、エンタープライズアレイや小規模なフラッシュアレイベンダーは、アレイあたり数千の VVol をサポートします。ネットアップは、 VVOL 3.0 のサポートに向けて、今後追加される機能で、きめ細かな VM 管理の進化も推進しています。
- ストレージ効率。NetAppは本番ワークロードに重複排除機能を初めて提供しましたが、このイノベーションはこの分野の最初のものでも最後のものでもありませんでした。まず、パフォーマンスに影響を与えないスペース効率に優れたデータ保護メカニズムであるSnapshotと、本番環境およびバックアップ用にVMの読み取り/書き込みコピーを瞬時に作成するFlexCloneテクノロジから始まりました。ネットアップは、重複排除、圧縮、ゼロブロック重複排除などのインライン機能を提供し、高価な SSD のストレージを最後まで絞ります。ONTAP は最近、圧縮機能を使用して、より小さな I/O 処理とファイルをディスクブロックに圧縮する機能を追加しました。これらの機能を組み合わせることで、 VSI では最大 5 分の 1 、 VDI では最大 30 分の 1 のコストを削減できました。
- * ハイブリッド・クラウド。 * オンプレミスのプライベート・クラウド、パブリック・クラウド・インフラストラクチャー、または両方の利点を組み合わせたハイブリッド・クラウドのいずれに使用しても、ONTAP ソリューションはデータ管理を合理化し、最適化するためのデータ・ファブリックの構築を支援します。まずハイパフォーマンスのオールフラッシュシステムを導入し、データ保護とクラウドコンピューティングのためにディスクストレージシステムとクラウドストレージシステムのどちらかと組み合わせます。Azure 、 AWS 、 IBM 、 Google のクラウドから選択して、コストを最適化し、ロックインを回避できます。必要に応じて、 OpenStack とコンテナテクノロジの高度なサポートを活用できます。ネットアップ ONTAP では、クラウドベースのバックアップ（ SnapMirror クラウド、 Cloud Backup Service 、 Cloud Sync ）やストレージ階層化 / アーカイブツール（ FabricPool ）も提供しており、運用コストの削減とクラウドの幅広いリーチの活用を支援します。
- * その他。 * NetApp AFF A シリーズアレイの卓越したパフォーマンスを活用して、コストを管理しながら仮想インフラを高速化できます。スケールアウト ONTAP クラスタを使用して、ストレージシステムのメンテナンスからアップグレード、完全な交換まで、完全なノンストップオペレーションを実現します。ネットアップの暗号化機能を追加コストなしで使用して、保存データを保護できます。きめ細かいサービス品質機能により、パフォーマンスがビジネスサービスレベルを満たしていることを確認します。これらはすべて、業界をリードするエンタープライズデータ管理ソフトウェアであるONTAPに付属する幅広い機能の一部です。

ユニファイドストレージ

NetApp ONTAPは、シンプルなソフトウェア定義型アプローチによってストレージを統合し、セキュアで効率的な管理、パフォーマンスの向上、シームレスな拡張性を実現します。このアプローチにより、データ保護が強化され、クラウドリソースを効果的に利用できるようになります。

当初、このユニファイドアプローチでは、1つのストレージシステムでNASとSANの両方のプロトコルをサポートすることが推奨されていましたが、ONTAPは引き続き業界をリードするSAN向けプラットフォームであり、当初からNASで強みを発揮しています。ONTAPでは、S3オブジェクトプロトコルもサポートされるようになりました。S3はデータストアには使用されませんが、ゲスト内アプリケーションに使用できます。S3プロトコルのサポートの詳細については、ONTAPを参照してください。 "S3構成の概要"。

Storage Virtual Machine（SVM）は、ONTAPのセキュアマルチテナンシーの単位です。これは、ONTAPソフトウェアを実行しているシステムへのクライアントアクセスを許可する論理構成要素です。SVM は、論理インターフェイス（ LIF ）を介して複数のデータアクセスプロトコルを使用して同時にデータをやり取りできます。SVM は、 CIFS や NFS などの NAS プロトコルでファイルレベルのデータアクセスを提供し、 iSCSI 、 FC / FCoE 、 NVMe などの SAN プロトコルでブロックレベルのデータアクセスを提供します。SVM は、S3と同様に、SANクライアントとNASクライアントそれぞれに同時にデータを提供できます。



vSphere 環境では、このアプローチは仮想デスクトップインフラ（ VDI ）向けのユニファイドシステムと仮想サーバインフラ（ VSI ）の組み合わせを意味する場合もあります。ONTAP ソフトウェアを実行するシステムは一般に、従来のエンタープライズアレイに比べて VSI の方が安価ですが、同じシステムで VDI を処理するための高度な Storage Efficiency 機能も備えています。また、 ONTAP は、 SSD から SATA までさまざまなストレージメディアを統合し、クラウドへの拡張を容易にします。パフォーマンスのためにフラッシュアレイを1つ、アーカイブ用にSATAアレイを1つ、クラウド用に別 々 のシステムを購入する必要はありません。ONTAP は、これらすべてを 1 つにまとめます。

注： SVM、ユニファイドストレージ、およびクライアントアクセスの詳細については、 "ストレージ仮想化" ONTAP 9 ドキュメントセンターを参照してください。

ONTAP の仮想化ツール

ネットアップでは、 ONTAP および vSphere と組み合わせて使用し、仮想環境を管理できるスタンドアロンのソフトウェアツールをいくつか提供しています。

ONTAP ライセンスには、追加コストなしで次のツールが含まれています。vSphere 環境でこれらのツールがどのように連携するかについては、図 1 を参照してください。

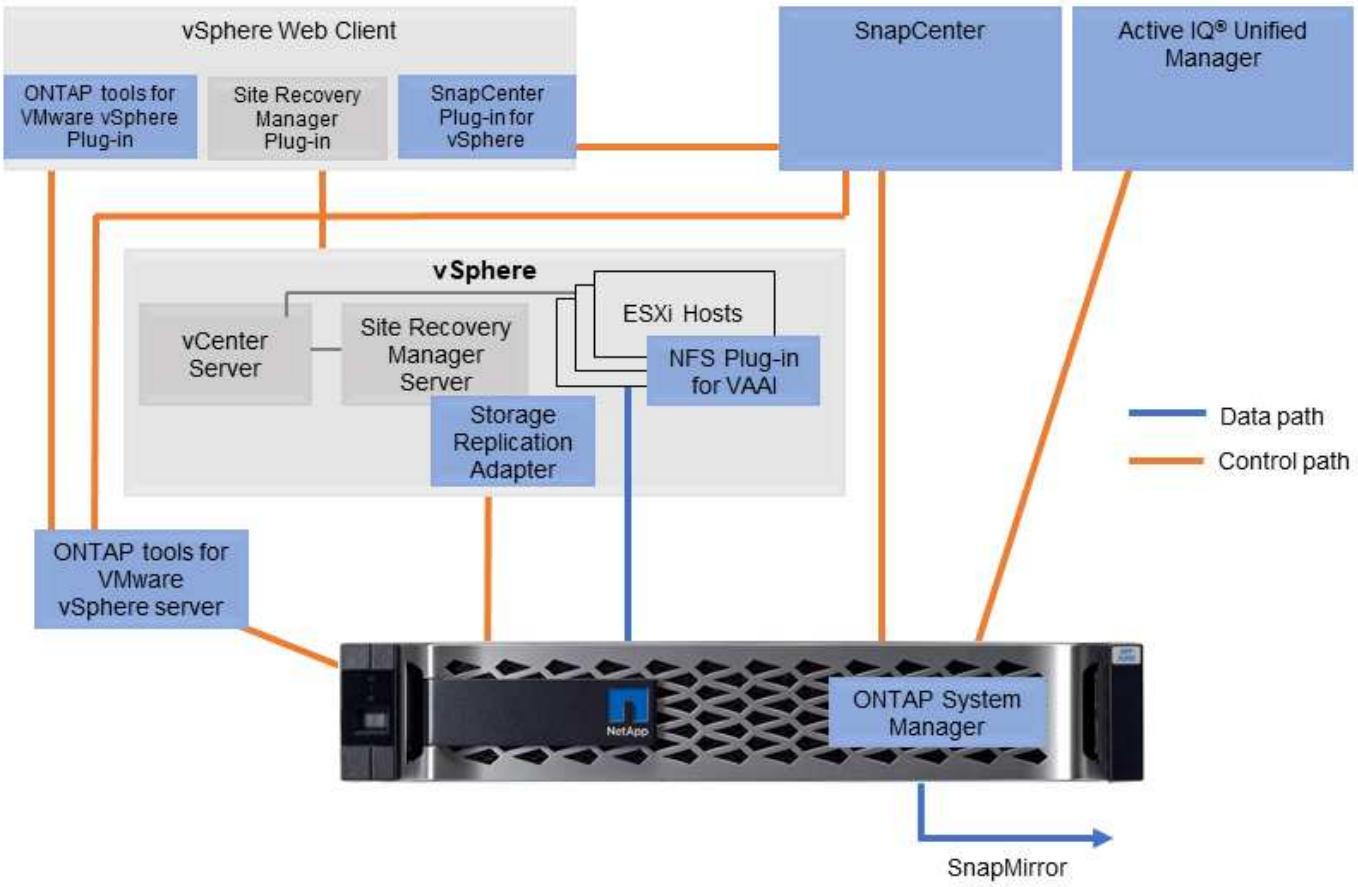
VMware vSphere 用の ONTAP ツール

VMware vSphere 用の ONTAP ツールは、 vSphere とともに ONTAP ストレージを使用するための一連のツー

ルです。vCenter プラグインは、以前 Virtual Storage Console （ VSC ）と呼ばれていたもので、 SAN と NAS のどちらを使用している場合でも、ストレージ管理と効率化機能の簡易化、可用性の向上、ストレージコストと運用オーバーヘッドの削減を実現します。データストアのプロビジョニングのベストプラクティスを使用して、 NFS 環境およびブロックストレージ環境用の ESXi ホスト設定を最適化します。以上のメリットのために、ネットアップでは、 ONTAP ソフトウェアを実行しているシステムで vSphere を使用する際のベストプラクティスとして、これらの ONTAP ツールを使用することを推奨します。サーバアプライアンス、vCenter、VASA Provider、Storage Replication Adapterのユーザインターフェイス拡張機能が含まれています。ONTAP ツールのほぼすべてを、最新の自動化ツールで利用できるシンプルなREST APIを使用して自動化できます。

- * vCenter UIの拡張機能* ONTAP ツールのUI拡張機能は、vCenter UIにホストとストレージを管理するための使いやすいコンテキスト依存メニュー、情報ポートレット、およびネイティブアラート機能を直接組み込み、ワークフローを合理化することで、運用チームやvCenter管理者の業務を簡素化します。
- * VASA Provider for ONTAP 。* VASA Provider for ONTAP は、VMware vStorage APIs for Storage Awareness（VASA）フレームワークをサポートしています。VMware vSphere 用の ONTAP ツールの一部として提供され、導入を容易にする単一の仮想アプライアンスとして提供されます。VASA Provider では、 VM ストレージのプロビジョニングと監視に役立つように vCenter Server と ONTAP を接続します。VMware Virtual Volumes （ VVol ）のサポート、ストレージ機能プロファイルと個々の VM VVol のパフォーマンスの管理、およびプロファイルの容量と準拠状況の監視用アラームが可能になります。
- * Storage Replication Adapter. SRAは、VMware Site Recovery Manager（SRM）と併用して、本番サイトと災害復旧サイト間のデータ複製を管理し、DRレプリカを無停止でテストします。検出、リカバリ、再保護のタスクを自動化します。Windows SRM サーバおよび SRM アプライアンス用の SRA サーバアプライアンスと SRA アダプタの両方が含まれています。

次の図は、 vSphere 用の ONTAP ツールを示しています。



NFS Plug-in for VMware VAAI のこと

NetApp NFS Plug-in for VMware VAAIはESXiホスト向けのプラグインで、ONTAP 上のNFSデータストアでVAAI機能を使用できます。クローン処理、シック仮想ディスクファイルのスペースリザベーション、およびスナップショットオフロードのコピーオフロードをサポートします。コピー処理をストレージにオフロードしても、完了までの時間が必ずしも短縮されるとは限りませんが、ネットワーク帯域幅の要件が軽減され、CPUサイクル、バッファ、キューなどのホストリソースがオフロードされます。VMware vSphere用のONTAP ツールを使用して、ESXiホストまたはサポートされている場合はvSphere Lifecycle Manager（VLCM）にプラグインをインストールできます。

Virtual Volumes （ VVol ）と Storage Policy Based Management （ SPBM ）

ネットアップは、 vSphere Virtual Volumes （ VVol ）の開発において VMware と初期の設計パートナーとして、アーキテクチャに関する情報提供と、 VVol および VMware vSphere APIs for Storage Awareness （ VASA ）のサポートを提供していました。このアプローチにより、VMのきめ細かなストレージ管理がVMFSで実現しただけでなく、Storage Policy Based Management（SPBM）によるストレージプロビジョニングの自動化もサポートされました。

SPBM は、仮想化環境で使用できるストレージサービスと、プロビジョニングされたストレージ要素の間の抽象化レイヤとして機能するフレームワークを、ポリシーを通じて提供します。このアプローチにより、ストレージアーキテクトは、 VM 管理者が簡単に利用できるさまざまな機能を備えたストレージプールを設計できます。仮想マシンのワークロード要件をプロビジョニングされたストレージプールと照合することで、仮想マシンごとまたは仮想ディスクレベルのさまざまな設定をきめ細かく制御できます。

ONTAP は VVol の規模においてストレージ業界をリードし、 1 つのクラスタで数十万もの VVol をサポートします。一方、エンタープライズアレイや小規模なフラッシュアレイベンダーは、アレイあたり数千の VVol をサポートします。また、 VVOL 3.0 をサポートする機能が追加され、 VM のきめ細かな管理が進化しています。

VMware vSphere Virtual Volumes 、 SPBM 、および ONTAP の詳細については、を参照してください "TR-4400 ：『 VMware vSphere Virtual Volumes with ONTAP 』"。

データストアおよびプロトコル

vSphereデータストアとプロトコルの機能の概要

VMware vSphereとONTAP ソフトウェアを実行しているシステム上のデータストアの接続には、次の7つのプロトコルが使用されます。

- FCP
- FCoE
- NVMe/FC
- NVMe/FC
- iSCSI
- NFS v3
- NFS v4.1

FCP、FCoE、NVMe/FC、NVMe/FC、NVMe/FC、NVMe/FC、およびiSCSIはブロックプロトコルで、vSphere Virtual Machine File System（VMFS）を使用して、ONTAP FlexVol ボリュームに含まれるONTAP LUNまたはNVMeネームスペースにVMを格納します。vSphere 7.0 以降では、 VMware は本番環境でのソフトウェア FCoE をサポートしなくなりました。NFS はファイルプロトコルで、 VM をデータストア（ ONTAP ボリューム）に配置し、 VMFS を必要としません。SMB（CIFS）、iSCSI、NVMe/FC、NFSもゲストOSからONTAPに直接使用できます。

次の表に、vSphereがサポートするONTAPの従来のデータストア機能を示します。この情報はVVOLデータストアには該当しませんが、通常は、サポートされているONTAP リリースを使用する環境 vSphere 6.x以降のリリースで使用されます。を参照することもできます "VMwareコウセイノサイダイスウ" 個々の vSphere リリースに固有の制限を確認するため。

機能 / 特徴	FC / FCoE	iSCSI	NVMe-oF	NFS
の形式で入力し	VMFS または raw デバイスマッピング（ RDM ）	VMFS または RDM	VMFS	該当なし
データストアまたは LUN の最大数	ホストあたり1、024個のLUN	サーバあたり1、024個のLUN	サーバごとに256名を指定します	256マウント デフォルトのNFS。MaxVolumes は 8 です。VMware vSphere 用の ONTAP ツールを使用して 256 まで増やす。
データストアの最大サイズ	64TB	64TB	64TB	100TB 以上の FlexVol ボリュームと FlexGroup ボリューム
データストアの最大ファイルサイズ	62TB	62TB	62TB	62TB（ONTAP 9.12.1P2以降使用時）
LUN またはファイルシステムごとのキューの深さの最適値	64 ~ 256	64 ~ 256	自動ネゴシエーション	のNFS.MaxQueueDepthを参照してください "推奨される ESXi ホストとその他の ONTAP 設定"。

次の表に、サポートされる VMware ストレージ関連機能を示します。

容量 / 機能	FC / FCoE	iSCSI	NVMe-oF	NFS
vMotion	はい。	はい。	はい。	はい。
Storage vMotion の機能です	はい。	はい。	はい。	はい。
VMware HA	はい。	はい。	はい。	はい。
ストレージ分散リソーススケジューラ（ SDRS ）	はい。	はい。	はい。	はい。

容量 / 機能	FC / FCoE	iSCSI	NVMe-oF	NFS
VMware vStorage APIs for Data Protection （ VADP ）対応のバックアップソフトウェア	はい。	はい。	はい。	はい。
VM 内の Microsoft Cluster Service （ MSCS ）またはフェイルオーバークラスタリング	はい。	はい *	はい *	サポート対象外
フォールトトレランス	はい。	はい。	はい。	はい。
Site Recovery Manager の略	はい。	はい。	いいえ **	v3のみ**
シンプロビジョニングされた VM （仮想ディスク）	はい。	はい。	はい。	はい。 VAAIを使用しない場合、NFS上のすべてのVMに対してこの設定がデフォルトになります。
VMware 標準マルチパス	はい。	はい。	はい、新しい高性能プラグイン（HPP）を使用して	NFS v4.1セッショントランキングにはONTAP 9.14.1以降が必要

次の表に、サポートされる ONTAP ストレージ管理機能を示します。

機能 / 特徴	FC / FCoE	iSCSI	NVMe-oF	NFS
データ重複排除	アレイ内での容量削減	アレイ内での容量削減	アレイ内での容量削減	データストア内での容量削減
シンプロビジョニング	データストアまたは RDM	データストアまたは RDM	データストア	データストア
データストアのサイズを変更	拡張のみ	拡張のみ	拡張のみ	拡張、自動拡張、縮小
Windows 、 Linux アプリケーション用の SnapCenter プラグイン（ゲスト内）	はい。	はい。	いいえ	はい。
VMware vSphere 用の ONTAP ツールを使用した監視とホストの設定	はい。	はい。	いいえ	はい。

機能 / 特徴	FC / FCoE	iSCSI	NVMe-oF	NFS
VMware vSphere 用の ONTAP ツールを使用したプロビジョニング	はい。	はい。	いいえ	はい。

次の表に、サポートされるバックアップ機能を示します。

機能 / 特徴	FC / FCoE	iSCSI	NVMe-oF	NFS
ONTAPスナップショット	はい。	はい。	はい。	はい。
複製バックアップでサポートされるSRM	はい。	はい。	いいえ **	v3のみ**
Volume SnapMirrorの略	はい。	はい。	はい。	はい。
VMDK イメージアクセス	VADP 対応のバックアップソフトウェア	VADP 対応のバックアップソフトウェア	VADP 対応のバックアップソフトウェア	VADP 対応のバックアップソフトウェア、 vSphere Client 、 vSphere Web Client データストアブラウザ
VMDK のファイルレベルアクセス	VADP 対応のバックアップソフトウェア、 Windows のみ	VADP 対応のバックアップソフトウェア、 Windows のみ	VADP 対応のバックアップソフトウェア、 Windows のみ	VADP 対応のバックアップソフトウェアとサードパーティ製アプリケーション
NDMP の単位	データストア	データストア	データストア	データストアまたはVM

- VMFSデータストア内でマルチライター対応のVMDKを使用するのではなく、Microsoftクラスタにゲスト内iSCSIを使用することを推奨します。このアプローチは Microsoft と VMware によって完全にサポートされており、 ONTAP （オンプレミスまたはクラウドの ONTAP システムへの SnapMirror ）を使用した優れた柔軟性、設定と自動化が容易で、 SnapCenter で保護できます。vSphere 7 で、新しいクラスタ化された VMDK オプションが追加されました。これは、マルチライター対応のVMDKとは異なります。マルチライター対応のVMDKを使用するには、クラスタ化されたVMDKをサポートするFCプロトコルを介して提供されるデータストアが必要です。その他の制限が適用されます。VMwareの詳細 "Windows Server フェールオーバークラスタリングのセットアップ" 設定ガイドラインについては、ドキュメントを参照してください
    - NVMe-oFとNFS v4.1を使用するデータストアには、vSphereレプリケーションが必要です。アレイベースのレプリケーションはSRMではサポートされていません。

ストレージプロトコルを選択

ONTAP ソフトウェアを実行するシステムは、主要なストレージプロトコルをすべてサポートしているため、既存および計画されているネットワークインフラやスタッフのスキルに応じて、お客様は環境に最適なものを選択できます。ネットアップのテストでは、一般に、ほぼ同じ速度の回線で実行されているプロトコル間の違いはほとんど見られませんでした。そのため、物理プロトコルのパフォーマンスよりもネットワークインフラとスタッフの能力に重点を置くことを推奨します。

プロトコルの選択を検討する際には、次の要素が役立ちます。

- * 現在のお客様の環境。 * 一般に、 IT チームはイーサネット IP インフラの管理のスキルを持っていますが、すべてのチームが FC SAN ファブリックの管理のスキルを持っているわけではありません。ただし、ストレージトラフィック用に設計されていない汎用IPネットワークを使用すると、うまく機能しない場合があります。現在利用しているネットワークインフラストラクチャ、計画的な改善点、およびそれらを管理するためのスタッフのスキルと可用性を考慮します。
- * セットアップの容易さ * FC ファブリックの初期構成（追加のスイッチとケーブル配線、ゾーニング、 HBA とファームウェアの相互運用性の検証）に加えて、ブロックプロトコルを使用するには、 LUN の作成とマッピング、ゲスト OS による検出とフォーマットも必要です。作成およびエクスポートされた NFS ボリュームは、 ESXi ホストによってマウントされ、使用可能な状態になります。NFS では、ハードウェアの認定や管理に関する特別なファームウェアはありません。
- * 管理の容易さ。 * SAN プロトコルでは、より多くのスペースが必要な場合、 LUN の拡張、新しいサイズの検出のための再スキャン、ファイルシステムの拡張など、いくつかの手順が必要です。LUN の拡張は可能ですが、 LUN のサイズを縮小することはできず、未使用スペースのリカバリには追加の作業が必要になる場合があります。NFS を使用すると、簡単なサイジングが可能です。このサイズ変更は、ストレージシステムで自動化できます。SAN では、ゲスト OS のトリム / マッピング解除コマンドを使用してスペース再生が可能で、削除されたファイルのスペースをアレイに戻すことができます。NFS データストアでは、このようなスペース再生がより困難になります。
- * ストレージスペースの透過性。 * シンプロビジョニングによって削減効果が即座に現れるため、 NFS 環境では一般にストレージ利用率が見やすくなります。同様に、重複排除とクローニングによる削減効果は、同じデータストア内の他の VM や他のストレージシステムボリュームで即座に利用できます。一般に、 VM の密度は NFS データストア内でも高くなります。管理するデータストアが少ないため、重複排除による削減効果が向上すると同時に管理コストも削減されます。

データストアのレイアウト

ONTAP ストレージシステムは、 VM および仮想ディスク用のデータストアを柔軟に作成できます。を使用する場合、 ONTAP の多くのベストプラクティスが適用されますが vSphere 用のデータストアをプロビジョニングする VSC （を参照） "推奨される ESXi ホストとその他の ONTAP 設定") 、考慮すべきその他のガイドラインを次に示します。

- ONTAP NFS データストアを使用して vSphere を導入することで、高性能でありながら管理が容易な実装を実現でき、ブロックベースのストレージプロトコルでは達成できない VM / データストア比率が提供されます。このアーキテクチャでは、データストア密度を 10 倍に増やすことも可能で、それに伴いデータストアの数は減少します。データストアのサイズを大きくするとストレージ効率が向上し、運用上のメリットが得られますが、ハードウェアリソースのパフォーマンスを最大限に引き出すためには、少なくとも 4 つのデータストア（ FlexVol ボリューム）を使用して 1 つの ONTAP コントローラに VM を格納することを検討してください。また、異なるリカバリポリシーを使用してデータストアを確立することもできます。ビジネスニーズに基づいて、他のバックアップや複製の頻度を高められるものもあります。FlexGroup ボリュームは設計上拡張できるため、複数のデータストアを使用する必要はありません。
- NetAppでは、ほとんどのNFSデータストアにFlexVolボリュームを使用することを推奨しています。ONTAP 9.8以降でFlexGroupは、データストアとしての使用もサポートされており、特定のユースケースでの使用が一般的に推奨されます。qtreeなどのその他のONTAPストレージコンテナは、現在ONTAP Tools for VMware vSphereまたはNetApp SnapCenter Plugin for VMware vSphereでサポートされていないため、一般に推奨されません。とはいえ、1つのボリューム内の複数のqtreeとしてデータストアを導入することは、データストアレベルのクォータやVMファイルクローンのメリットが得られる高度に自動化された環境に役立つ可能性があります。
- FlexVol ボリュームデータストアの適切なサイズは 4~8TB です。このサイズは、パフォーマンス、管理のしやすさ、データ保護のバランスが取れた適切なサイズです。小規模構成から開始して（ 4TB など）、必要に応じてデータストアを拡張します（最大 100TB まで）。小規模なデータストアは、バックアップ

や災害からのリカバリにかかる時間が短く、クラスタ間で迅速に移動できます。使用済みスペースの変化に応じてボリュームを自動的に拡張または縮小するには、 ONTAP のオートサイズを使用することを検討してください。VMware vSphere データストアプロビジョニングウィザードの ONTAP ツールでは、新しいデータストアに対してデフォルトでオートサイズが使用されます。拡張および縮小のしきい値と最大および最小サイズは、 System Manager またはコマンドラインを使用して追加でカスタマイズできます。

- または、 VMFS データストアを、 FC 、 iSCSI または FCoE でアクセスする LUN で構成することもできます。VMFS を使用すると、クラスタ内の各 ESX サーバから同時に従来型の LUN にアクセスすることができます。VMFS データストアは、最大 64TB まで拡張でき、最大 32 個の 2TB LUN （ VMFS 3 ）または単一の 64TB LUN （ VMFS 5 ）で構成できます。ONTAP の最大LUNサイズは、ほとんどのシステムで16TBで、オールSANアレイシステムでは128TBです。したがって、ほとんどの ONTAP システムでは、最大サイズの VMFS 5 データストアを、 4 つの 16TB LUN を使用して作成できます。複数のLUN（ハイエンドのFAS またはAFF システムを使用）を使用する高I/Oワークロードではパフォーマンス上のメリットを得られますが、データストアLUNの作成、管理、保護の複雑さが増し、可用性のリスクが増大することで、このメリットを相殺することができます。ネットアップでは、通常、各データストアに 1 つの大きな LUN を使用し、 16TB を超えるデータストアを追加する必要がある場合にのみスパンすることを推奨しています。NFS と同様に、複数のデータストア（ボリューム）を使用することで、 1 台の ONTAP コントローラのパフォーマンスを最大化することを検討してください。
- 古いゲストオペレーティングシステム（ OS ）では、パフォーマンスとストレージ効率を最大化するために、ストレージシステムとのアライメントが必要でした。しかし、 Microsoft や Linux ディストリビュータ（ Red Hat など）が提供する、ベンダーがサポートする最新の OS では、ファイルシステムのパーティションを仮想環境の基盤となるストレージシステムのブロックにアライメントするように調整する必要はありません。アライメントが必要な古い OS を使用している場合は、ネットアップサポートの技術情報で「 VM のアライメント」に関する記事を検索するか、ネットアップの営業担当者またはパートナー担当者に TR-3747 のコピーを請求してください。
- デフラグユーティリティはゲストOS内では使用しないでください。パフォーマンス上のメリットはなく、ストレージ効率とスナップショット容量の使用にも影響します。また、仮想デスクトップのゲスト OS で検索インデックスを無効にすることを検討してください。
- ONTAP は、革新的な Storage Efficiency 機能で業界をリードし、使用可能なディスクスペースを最大限に活用できるようにしています。AFF システムでは、デフォルトのインライン重複排除機能と圧縮機能により、この効率性がさらに向上しています。データはアグリゲート内のすべてのボリュームにわたって重複排除されるため、類似するオペレーティングシステムやアプリケーションを 1 つのデータストア内にまとめて、最大限の削減効果を得る必要はありません。
- 場合によっては、データストアが不要なこともあります。パフォーマンスと管理性を最大限に高めるためには、データベースや一部のアプリケーションなどの高 I/O アプリケーションにはデータストアを使用しないでください。代わりに、ゲストが管理する NFS や iSCSI ファイルシステムなど、ゲスト所有のファイルシステムや RDM を使用することを検討してください。アプリケーションに関する具体的なガイダンスについては、ご使用のアプリケーションに関するネットアップのテクニカルレポートを参照してください。例： "ONTAP を基盤にした Oracle データベース" 仮想化に関するセクションと役立つ詳細情報が記載されています。
- 第 1 クラスのディスク（または強化された仮想ディスク）を使用すると、 vSphere 6.5 以降を搭載した VM に関係なく、 vCenter で管理されるディスクを使用できます。主に API で管理されますが、 VVol では特に OpenStack ツールや Kubernetes ツールで管理する場合に便利です。ONTAP および VMware vSphere 用の ONTAP ツールでサポートされています。

データストアと **VM** 移行

別のストレージシステム上の既存のデータストアから ONTAP に VM を移行する際は、いくつか注意しておくべきプラクティスがあります。

- Storage vMotion を使用して、仮想マシンの大部分を ONTAP に移動します。このアプローチでは、実行中の VM を停止する必要がなくなるだけでなく、インラインの重複排除や圧縮などの ONTAP の Storage

Efficiency 機能を使用して、移行時にデータを処理できます。vCenter 機能を使用してインベントリリストから複数の VM を選択し、適切なタイミングで移行をスケジュール（ Ctrl キーを押しながら [アクション] をクリック）することを検討します。

- 適切なデスティネーションデータストアへの移行を慎重に計画することもできますが、多くの場合、一括で移行して必要に応じてあとから整理する方が簡単です。Snapshotスケジュールの変更など、データ保護に関する特定のニーズがある場合は、このアプローチを使用して別のデータストアに移行できます。
- ほとんどの VM とそのストレージは、実行中（ホット）に移行できますが、 ISO 、 LUN 、 NFS ボリュームなどの接続されたストレージ（データストア内にない）を別のストレージシステムから移行する場合は、コールドマイグレーションが必要になることがあります。
- より慎重な移行が必要な仮想マシンには、接続されたストレージを使用するデータベースやアプリケーションなどがあります。一般的に、移行を管理するためにアプリケーションのツールを使用することを検討してください。Oracle の場合は、 RMAN や ASM などの Oracle ツールを使用してデータベース・ファイルを移行することを検討してください。を参照してください "TR-4534" を参照してください。同様に、 SQL Server の場合は、 SQL Server Management Studio を使用するか、 SnapManager for SQL Server や SnapCenter などのネットアップのツールを使用することを検討します。

VMware vSphere 用の ONTAP ツール

ONTAP ソフトウェアを実行しているシステムで vSphere を使用する際に最も重要なベストプラクティスは、VMware vSphere プラグイン（旧 Virtual Storage Console ）用の ONTAP ツールをインストールして使用することです。この vCenter プラグインは、 SAN と NAS のどちらを使用している場合でも、ストレージ管理を簡易化し、可用性を向上させ、ストレージコストと運用オーバーヘッドを削減します。データストアのプロビジョニングのベストプラクティスを使用して、マルチパスと HBA タイムアウト（これらは付録 B で説明）用の ESXi ホスト設定を最適化します。vCenterプラグインであるため、vCenterサーバに接続するすべてのvSphere Web Clientで使用できます。

このプラグインは、 vSphere 環境で他の ONTAP ツールを使用する場合にも役立ちます。NFS Plug-in for VMware VAAIをインストールできます。これにより、VMのクローニング処理、シック仮想ディスクファイルのスペースリザベーション、ONTAPスナップショットのオフロードのために、ONTAPへのコピーオフロードが可能になります。

VASA Provider for ONTAP の多くの機能を使用するための管理インターフェイスでもあり、 VVol でのストレージポリシーベースの管理がサポートされています。VMware vSphere 用の ONTAP ツールを登録したら、ストレージ機能プロファイルを作成してストレージにマッピングし、データストアがプロファイルに一定期間にわたって準拠していることを確認します。VASA Provider には、 VVol データストアの作成と管理を行うためのインターフェイスも用意されています。

一般に、 vCenter 内で VMware vSphere インターフェイス用の ONTAP ツールを使用して、従来のデータストアと VVol データストアをプロビジョニングし、ベストプラクティスに従っていることを確認することを推奨します。

一般的なネットワーク

ONTAP ソフトウェアを実行しているシステムで vSphere を使用する場合のネットワーク設定の構成は簡単で、他のネットワーク構成と同様です。考慮すべき点をいくつか挙げます。

- ストレージネットワークのトラフィックを他のネットワークから分離します。専用の VLAN を使用するか、ストレージ用に別個のスイッチを使用することで、別のネットワークを実現できます。ストレージネットワークがアップリンクなどの物理パスを共有している場合は、十分な帯域幅を確保するために QoS または追加のアップリンクポートが必要になることがあります。ホストをストレージに直接接続しないでください。スイッチを使用して冗長パスを確保し、VMware HAが介入なしで機能できるようにします。を参照してください "直接接続ネットワーク" 追加情報 の場合。

- ジャンボフレームは、必要に応じてネットワークでサポートされていれば、特に iSCSI を使用している場合に使用できます。使用する場合は、ストレージと ESXi ホストの間のパスにあるすべてのネットワークデバイスや VLAN で設定が同じであることを確認してください。そうしないと、パフォーマンスや接続の問題が発生する可能性があります。MTU は、 ESXi 仮想スイッチ、 VMkernel ポート、および各 ONTAP ノードの物理ポートまたはインターフェイスグループでも同一の設定にする必要があります。
- ネットワークフロー制御は、 ONTAP クラスタ内のクラスタネットワークポートでのみ無効にすることを推奨します。データトラフィックに使用される残りのネットワークポートについては、推奨されるベストプラクティスはありません。必要に応じて有効または無効にしてください。を参照してください "TR-4182" を参照してください。
- ESXi および ONTAP ストレージアレイをイーサネットストレージネットワークに接続するときは、接続先のイーサネットポートを Rapid Spanning Tree Protocol （ RSTP ；高速スパニングツリープロトコル）のエッジポートとして設定するか、 Cisco の PortFast 機能を使用して設定することを推奨します。ネットアップでは、 Cisco の PortFast 機能を使用していて、 ESXi サーバまたは ONTAP ストレージアレイへの 802.1Q VLAN トランキングが有効になっている環境では、 Spanning-Tree PortFast trunk 機能を有効にすることを推奨します。
- リンクアグリゲーションのベストプラクティスとして次を推奨します。
  - CiscoのVirtual PortChannel（vPC）などのマルチシャーシリンクアグリゲーショングループアプローチを使用して、2つの別 々 のスイッチシャーシ上のポートのリンクアグリゲーションをサポートするスイッチを使用します。
  - LACPが設定されたdvSwitches 5.1以降を使用していない場合、ESXiに接続されているスイッチポートのLACPを無効にします。
  - LACPを使用して、ポートハッシュまたはIPハッシュを使用したダイナミックマルチモードインターフェイスグループを使用するONTAPストレージシステムのリンクアグリゲートを作成します。を参照してください "Network Management の略" を参照してください。
  - ESXiで静的リンクアグリゲーション（EtherChannelなど）と標準vSwitchを使用する場合、またはvSphere Distributed Switchを使用するLACPベースのリンクアグリゲーションを使用する場合は、IPハッシュチーミングポリシーを使用します。リンクアグリゲーションを使用しない場合は、代わりに[Route based on the originating virtual port ID]を使用します。

次の表に、ネットワーク設定項目とその適用先をまとめます。

項目	ESXi	スイッチ	ノード	SVM
IP アドレス	VMkernel	いいえ **	いいえ **	はい。
リンクアグリゲーション	仮想スイッチ	はい。	はい。	いいえ *
VLAN	VMkernel と VM ポートグループ	はい。	はい。	いいえ *
フロー制御	NIC	はい。	はい。	いいえ *
スパニングツリー	いいえ	はい。	いいえ	いいえ
MTU （ジャンボフレーム用）	仮想スイッチと VMkernel ポート（ 9000 ）	○（最大に設定）	○（ 9000 ）	いいえ *
フェイルオーバーグループ	いいえ	いいえ	○（作成）	○（選択）

- SVM LIFは、VLANやMTUなどが設定されたポート、インターフェイスグループ、またはVLANインターフェイスに接続します。ただし、設定の管理はSVMレベルではありません。
  - これらのデバイスには管理用に独自の IP アドレスがありますが、 ESXi ストレージネットワークのコンテキストでは使用されません。

SAN （ FC 、 FCoE 、 NVMe/FC 、 iSCSI ）、 RDM

NetApp ONTAPは、iSCSI、ファイバチャネルプロトコル（FCP、またはFC）、NVMe over Fabrics（NVMe-oF）を使用して、VMware vSphereにエンタープライズクラスのブロックストレージを提供します。vSphereとONTAPを使用してVMストレージにブロックプロトコルを実装する場合のベストプラクティスを次に示します。

vSphere では、ブロックストレージ LUN を 3 通りの方法で使用します。

- VMFS データストアを使用する場合
- raw デバイスマッピング（ RDM ）で使用
- ソフトウェアイニシエータがアクセスおよび制御する LUN として使用 VM ゲスト OS から作成します

VMFS は、共有ストレージプールであるデータストアを提供する、高性能なクラスタファイルシステムです。VMFSデータストアは、FC、iSCSI、FCoEを使用してアクセスするLUN、またはNVMe/FCまたはNVMe/TCPプロトコルを使用してアクセスするNVMeネームスペースで構成できます。VMFSを使用すると、クラスタ内のすべてのESXサーバから同時にストレージにアクセスできます。ONTAP 9.12.1P2以降（およびASAシステムの以前のバージョン）では、一般に最大LUNサイズは128TBです。したがって、単一のLUNを使用して、64TBの最大サイズのVMFS 5または6データストアを作成できます。

vSphere は、ストレージデバイスへの複数のパスを標準でサポートします。この機能はネイティブマルチパス（ NMP ）と呼ばれます。NMP は、サポートされるストレージシステムのストレージタイプを検出し、使用中のストレージシステムの機能をサポートするように NMP スタックを自動的に設定できます。

NMPとONTAPはどちらも、Asymmetric Logical Unit Access（ALUA；非対称論理ユニットアクセス）による最適パスと非最適パスのネゴシエートをサポートします。ONTAP では、アクセス対象の LUN をホストするノード上のターゲットポートを使用する直接データパスが、 ALUA の最適パスとなります。ALUA は、vSphere と ONTAP の両方でデフォルトで有効になっています。NMPはONTAPクラスタをALUAとして認識し、ALUAストレージアレイタイププラグインを使用します。（`VMW_SATP_ALUA`）を入力し、ラウンドロビンパス選択プラグインを選択します。（`VMW_PSP_RR`）。

ESXi 6 は、最大 256 個の LUN と、 LUN への最大 1 、 024 個の合計パスをサポートします。これらの制限を超えるLUNやパスはESXiで認識されません。最大数の LUN を使用した場合、 LUN あたりのパス数は最大 4 つです。大規模な ONTAP クラスタでは、 LUN 数の上限に達する前にパス数の制限に達する可能性があります。この制限に対処するため、 ONTAP では、リリース 8.3 以降の選択的 LUN マップ（ SLM ）がサポートされています。

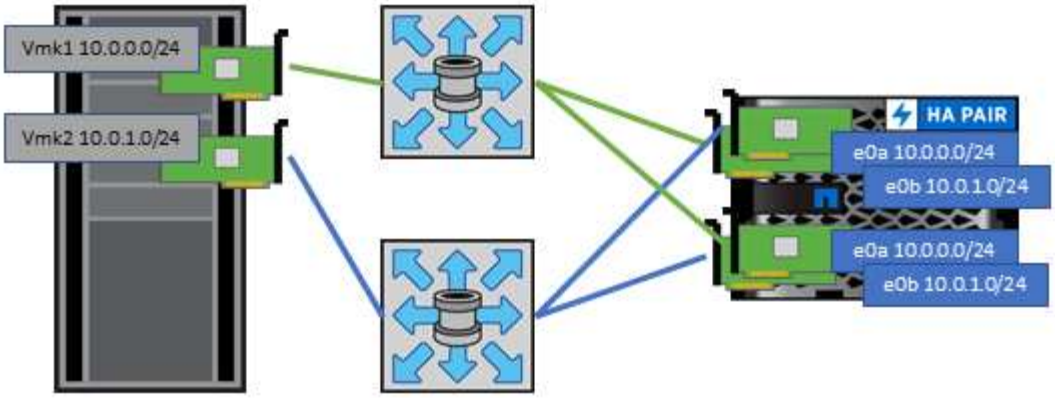
SLM は、特定の LUN へのパスをアドバタイズするノードを制限します。ネットアップのベストプラクティスでは、各 SVM のノードごとに少なくとも 1 つの LIF を配置し、 SLM を使用して、 LUN とその HA パートナーをホストするノードへのアドバタイズパスを制限することを推奨しています。他のパスは存在しますが、デフォルトではアドバタイズされません。SLM 内で、レポートノードの追加引数および削除引数を使用して通知されたパスを変更することができます。8.3 より前のリリースで作成された LUN ではすべてのパスがアドバタイズされるため、ホストしている HA ペアへのパスのみがアドバタイズされるように変更する必要があることに注意してください。SLM の詳細については、のセクション 5.9 を参照してください "TR-4080"。以前のポートセットの方式を使用すると、 LUN の使用可能なパスをさらに削減できます。ポートセットを使用すると、 igroup 内のイニシエータが LUN を認識する際に経由可能なパス数を減らすことができます。

- SLM はデフォルトでは有効になっています。ポートセットを使用しないかぎり、これ以上の設定は必要ありません。
- Data ONTAP 8.3より前のバージョンで作成したLUNの場合、次のコマンドを実行してSLMを手動で適用します。 `lun mapping remove-reporting-nodes` LUNレポートノードを削除し、LUNへのアクセスをLUNの所有者ノードとそのHAパートナーに制限するコマンド。

ブロックプロトコル（ iSCSI 、 FC 、 FCoE ）は、一意の名前に加え、 LUN ID とシリアル番号を使用して LUN にアクセスします。FC と FCoE は Worldwide Name （ WWNN および WWPN ）を使用し、 iSCSI は iSCSI Qualified Name （ IQN ）を使用します。ストレージ内での LUN へのパスはブロックプロトコルにとっては意味がないため、どこにも表示されません。したがって、 LUN のみが含まれるボリュームは内部でマウントする必要がなく、データストアで使用される LUN を含むボリュームのジャンクションパスも必要ありません。ONTAP の NVMe サブシステムも同様に機能します。

考慮すべきその他のベストプラクティス：

- 可用性と移動性を最大限に高めるために、 ONTAP クラスタ内の各ノード上の各 SVM に論理インターフェイス（ LIF ）が作成されていることを確認します。ONTAP SAN では、各ファブリックに対して 1 つずつ、ノードごとに 2 つの物理ポートと LIF を使用することを推奨します。ALUA を使用してパスが解析され、アクティブな最適化（直接）パスとアクティブな非最適化パスが特定されます。ALUA は FC 、 FCoE 、および iSCSI に使用されます。
- iSCSI ネットワークの場合、複数の仮想スイッチがある場合は、 NIC チーミングを使用して、異なるネットワークサブネット上の複数の VMkernel ネットワークインターフェイスを使用します。また、複数の物理スイッチに接続された複数の物理 NIC を使用して、 HA を実現し、スループットを向上させることもできます。次の図に、マルチパス接続の例を示します。ONTAP では、 2 つ以上のスイッチに接続された 2 つ以上のリンクでフェイルオーバーするシングルモードインターフェイスグループを設定するか、 LACP または他のリンクアグリゲーションテクノロジをマルチモードインターフェイスグループと併用して HA を実現し、リンクアグリゲーションのメリットを活かすことができます。
- ESXiでターゲット認証にチャレンジハンドシェイク認証プロトコル（CHAP）が使用されている場合は、CLIを使用してONTAPでもCHAPを設定する必要があります。（`vserver iscsi security create`）またはSystem Managerで（[ストレージ]>[SVM]>[SVM設定]>[プロトコル]>[iSCSI]で[イニシエータセキュリティ]を編集します）。
- LUN と igroup の作成と管理には、 VMware vSphere の ONTAP ツールを使用します。プラグインによってサーバの WWPN が自動的に判別され、適切な igroup が作成されます。また、ベストプラクティスに従って LUN を設定し、正しい igroup にマッピングします。
- RDMは管理が困難になる可能性があるため、使用には注意が必要です。また、前述したように制限されているパスも使用します。ONTAP LUN は両方をサポートします "物理互換モードと仮想互換モード" RDM ：
- vSphere 7.0 での NVMe/FC の使用については、以下を参照してください "ONTAP NVMe/FC Host Configuration Guide" および "TR-4684"次の図に、 vSphere ホストから ONTAP LUN へのマルチパス接続を示します。



NFS

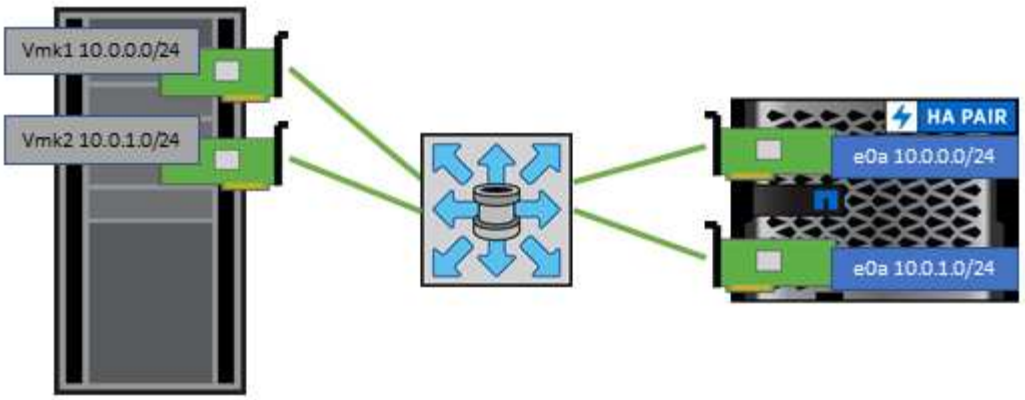
NetApp ONTAPは、とりわけエンタープライズクラスのスケールアウトNASアレイです。ONTAPは、VMware vSphereを強化し、多数のESXiホストからNFS接続データストアに同時にアクセスできるようにします。VMFSファイルシステムの制限をはるかに超えています。vSphereでNFSを使用すると、使いやすさとストレージ効率の可視化のメリットが得られます。詳細については、 "データストア" セクション。

vSphere で ONTAP NFS を使用する際に推奨されるベストプラクティスは次のとおりです。

- ONTAP クラスタ内の各ノードの各 SVM で、 1 つの論理インターフェイス（ LIF ）を使用します。データストアごとの LIF の過去の推奨事項は不要になりました。直接アクセス（同じノード上のLIFとデータストア）を推奨しますが、一般にパフォーマンスへの影響は最小限（マイクロ秒）であるため、間接アクセスについて心配する必要はありません。
- VMware は、 VMware Infrastructure 3 以降で NFSv3 をサポートしています。vSphere 6.0 では NFSv4.1 がサポートされるようになり、 Kerberos セキュリティなどの高度な機能が使用できるようになりました。NFSv3 ではクライアント側のロックが使用され、 NFSv4.1 ではサーバ側のロックが使用されます。ONTAP ボリュームは両方のプロトコルでエクスポートできますが、 ESXi は 1 つのプロトコルでしかマウントできません。この単一プロトコルのマウントにより、他の ESXi ホストが同じデータストアを別のバージョンでマウントすることができるわけではありません。すべてのホストが同じバージョン、つまり同じロック形式を使用するように、マウント時に使用するプロトコルバージョンを指定してください。NFS のバージョンをホスト間で混在させないでください。可能であれば、ホストプロファイルを使用して準拠しているかどうかを確認します
  - NFSv3 と NFSv4.1 間ではデータストアが自動変換されないため、新しい NFSv4.1 データストアを作成し、 Storage vMotion を使用して新しいデータストアに VM を移行します。
  - に記載されている NFS v4.1 と相互運用性に関する表の注を参照してください "NetApp Interoperability Matrix Tool で確認できます" をサポートするには、特定の ESXi パッチレベルが必要です。
  - vSphere 8.0U2以降では、VMwareでNFSv3でのnconnectがサポートされます。nconnectの詳細については、 "NetAppおよびVMwareでのNFSv3 nconnect機能"
- NFS エクスポートポリシーは、 vSphere ホストによるアクセスの制御に使用されます。複数のボリューム（データストア）で 1 つのポリシーを使用できます。NFSv3 では、 ESXi で sys （ UNIX ）セキュリティ形式が使用され、 VM を実行するためにルートマウントオプションが必要となります。ONTAP では、このオプションはスーパーユーザと呼ばれます。スーパーユーザオプションを使用する場合は、匿名ユーザ ID を指定する必要はありません。の値が異なるエクスポートポリシールールに注意してください `-anon` および `-allow-suid` 原因 SVM検出がONTAP ツールで問題を検出できるかどうか。ポリシーの例を次に示します。

- Access Protocol：nfs（nfs3とnfs4の両方を含む）
- クライアント一致仕様： 192.168.42.21
- RO アクセスルール： sys
- RWアクセスルール:sys
- 匿名UIDの形式です
- superuser ： sys

- NetApp NFS Plug-in for VMware VAAIを使用する場合は、プロトコルをに設定する必要があります。 `nfs` ではなく `nfs3` エクスポートポリシールールが作成または変更されたとき。VAAIコピーオフロード機能を使用するには、データプロトコルがNFSv3であっても、NFSv4プロトコルが機能する必要があります。フロトコルノシテイ `nfs` NFSv3とNFSv4の両方のバージョンが含まれます。
- NFS データストアのボリュームは SVM のルートボリュームからジャンクションされるため、 ESXi がデータストアボリュームに移動してマウントするためにはルートボリュームへのアクセス権も必要となります。ルートボリューム、およびデータストアボリュームのジャンクションがネストされているその他のボリュームのエクスポートポリシーには、ESXiサーバに読み取り専用アクセスを許可するルールが含まれている必要があります。VAAIプラグインを使用したルートボリュームのポリシーの例を次に示します。
  - Access Protocol：nfs（nfs3とnfs4の両方を含む）
  - クライアント一致仕様： 192.168.42.21
  - RO アクセスルール： sys
  - RW Access Rule：never（ルートボリュームに最適なセキュリティ）
  - 匿名UIDの形式です
  - superuser：sys（VAAIを使用するルートボリュームの場合も必要）
- VMware vSphere 用の ONTAP ツール（最も重要なベストプラクティス）を使用：
  - VMware vSphere 用の ONTAP ツールを使用してデータストアをプロビジョニングすると、エクスポートポリシーの自動管理が簡易化されます。
  - プラグインを使用してVMwareクラスタ用のデータストアを作成するときは、単一のESXサーバではなくクラスタを選択します。これにより、データストアがクラスタ内のすべてのホストに自動的にマウントされます。
  - プラグインのマウント機能を使用して、既存のデータストアを新しいサーバに適用します。
  - VMware vSphere 用の ONTAP ツールを使用しない場合は、すべてのサーバ、または追加のアクセス制御が必要なサーバクラスタごとに、 1 つのエクスポートポリシーを使用します。
- ONTAP にはフレキシブルボリュームのネームスペース構造が用意されており、ジャンクションを使用してボリュームをツリーにまとめることができますが、このアプローチは vSphere には価値がありません。ストレージのネームスペース階層に関係なく、データストアのルートに各 VM 用のディレクトリが作成されます。そのため、単に SVM のルートボリュームに vSphere のボリュームのジャンクションパスをマウントすることがベストプラクティスです。これは、 VMware vSphere 用の ONTAP ツールでデータストアをプロビジョニングする方法です。ジャンクションパスがネストされていないと、ルートボリューム以外のボリュームに依存しているボリュームがないこと、またボリュームをオフラインにするか破棄するかによって意図的に他のボリュームへのパスに影響が及ぶこともありません。
- NFS データストアの NTFS パーティションのブロックサイズは 4K で十分です。次の図は、 vSphere ホストから ONTAP NFS データストアへの接続を示しています。



次の表に、 NFS のバージョンとサポートされる機能を示します。

vSphere の機能	**NFSv3**	**NFSv4.1**
vMotion と Storage vMotion	はい。	はい。
高可用性	はい。	はい。
フォールトトレランス	はい。	はい。
DRS	はい。	はい。
ホストプロファイル	はい。	はい。
Storage DRS	はい。	いいえ
ストレージ I/O の制御	はい。	いいえ
SRM の場合	はい。	いいえ
仮想ボリューム	はい。	いいえ
ハードウェアアクセラレーション（ VAAI ）	はい。	はい。
Kerberos 認証	いいえ	○（ vSphere 6.5 以降で拡張して、 AES 、 krb5i ）
マルチパスのサポート	いいえ	はい。

FlexGroup ボリューム

VMware vSphereでONTAPボリュームとFlexGroupボリュームを使用すれば、ONTAPクラスタ全体の能力を最大限に活用できるシンプルで拡張性に優れたデータストアを構築できます。

ONTAP 9.8、ONTAP Tools for VMware vSphere 9.8、SnapCenterプラグインfor VMware 4.4リリースに加えて、vSphereでのFlexGroupボリュームベースデータストアのサポートが追加されました。FlexGroupボリュームは大規模なデータストアの作成を簡易化し、必要な分散コンスティチュエントボリュームをONTAPクラスタ全体に自動的に作成して、ONTAPシステムのパフォーマンスを最大限に引き出します。

FlexGroupボリュームに関する詳細情報 "『FlexCache and FlexGroup Volume Technical Report』を参照してください"。

ONTAPクラスタ全体の機能を備えた拡張性に優れた単一のvSphereデータストアが必要な場合や、非常に大規模なクローニングワークロードがあり、新しいFlexGroupクローニングメカニズムのメリットがある場合は、vSphereでFlexGroupボリュームを使用します。

コピーオフロード

ONTAP 9.8では、vSphereワークロードを使用した広範なシステムテストに加えて、FlexGroupデータストア用の新しいコピーオフロードメカニズムが追加されました。この新しいシステムでは、強化されたコピーエンジンを使用して、ソースとデスティネーションの両方へのアクセスを許可しながら、バックグラウンドでコンスティチュエント間でファイルをレプリケートします。このローカルキャッシュを使用して、VMクローンをオンデマンドで迅速にインスタンス化します。

FlexGroup最適化コピーオフロードを有効にする方法については、を参照してください。 "VAAIコピーオフロードを許可するようにONTAP FlexGroupを設定する方法"

VAAIクローニングを使用していても、キャッシュをウォームアップするのに十分なクローンを作成しないと、ホストベースのコピーよりも高速ではない場合があります。その場合は、必要に応じてキャッシュタイムアウトを調整できます。

次のシナリオを考えてみましょう。

- 8つのコンスティチュエントで新しいFlexGroupを作成しました
- 新しいFlexGroupのキャッシュタイムアウトが160分に設定されている

このシナリオでは、ローカルファイルクローンではなく、最初に完了する8つのクローンがフルコピーになります。160秒のタイムアウトが経過する前にそのVMをクローニングすると、各コンスティチュエント内のファイルクローンエンジンがラウンドロビン方式で使用され、コンスティチュエントボリューム間でほぼ瞬時に均等に分散されたコピーが作成されます。

ボリュームが新しいクローンジョブを受信するたびに、タイムアウトがリセットされます。この例のFlexGroup内のコンスティチュエントボリュームがタイムアウトまでにクローン要求を受信しなかった場合、そのVMのキャッシュはクリアされ、ボリュームに再度データを入力する必要があります。また、元のクローンのソースが変更された場合（テンプレートを更新した場合など）、競合を防ぐために各構成要素のローカルキャッシュが無効になります。前述したように、キャッシュは調整可能であり、環境のニーズに合わせて設定できます。

VAAIでFlexGroupを使用する方法の詳細については、次の技術情報アーティクルを参照してください。 "VAAI：FlexGroupボリュームでのキャッシュの仕組みを教えてください。"

FlexGroupキャッシュを十分に活用できないものの、ボリューム間での高速クローニングが必要な環境では、VVOLの使用を検討してください。VVolを使用したボリューム間クローニングは、従来のデータストアよりもはるかに高速で、キャッシュに依存しません。

QoSセツテイ

ONTAP System Managerまたはクラスタシェルを使用してFlexGroupレベルでQoSを設定することはサポートされていますが、VMに対応したりvCenterと統合したりすることはできません。

QoS（最大/最小IOPS）は、vCenter UIまたはREST APIを使用して、個 々 のVMまたはデータストア内のすべてのVMに対して設定できますONTAP。すべての VM に QoS を設定すると、 VM ごとに個別に設定する必要がなくなります。今後は、新規または移行された VM には適用されません。新しい VM に QoS を設定するか、データストア内のすべての VM に QoS を再適用してください。

VMware vSphereでは、NFSデータストアのすべてのIOがホストごとに単一のキューとして扱われるため、1つのVMでのQoS調整が、同じデータストア内の他のVMのパフォーマンスに影響する可能性があることに注意してください。これに対し、VVOLでは、別のデータストアに移行してもQoSポリシーの設定を維持でき、調整しても他のVMのIOに影響しません。

指標

また、ONTAP 9.8では、FlexGroupファイル用のファイルベースのパフォーマンス指標（IOPS、スループット、レイテンシ）が新たに追加され、これらの指標はONTAP tools for VMware vSphereのダッシュボードとVMレポートで確認できるようになりました。VMware vSphere プラグイン用の ONTAP ツールでは、最大 IOPS と最小 IOPS の組み合わせを使用してサービス品質（ QoS ）ルールを設定することもできます。これらは、データストア内のすべての VM に対して個別に設定することも、特定の VM に対して個別に設定することもできます。

ベストプラクティス

- ONTAPツールを使用してFlexGroupデータストアを作成すると、FlexGroupが最適に作成され、vSphere環境に合わせてエクスポートポリシーが設定されます。ただし、ONTAP toolsを使用してFlexGroupボリュームを作成すると、vSphereクラスタ内のすべてのノードが1つのIPアドレスを使用してデータストアをマウントすることがわかります。その結果、ネットワークポートがボトルネックになる可能性があります。この問題を回避するには、データストアをアンマウントし、SVM上のLIF間でロードバランシングを行うラウンドロビンDNS名を使用して標準のvSphereデータストアウィザードを使用して再マウントします。再マウントが完了すると、ONTAP toolsは再びデータストアを管理できるようになります。ONTAP toolsを使用できない場合は、FlexGroupのデフォルト値を使用し、のガイドラインに従ってエクスポートポリシーを作成します。 "データストアとプロトコル- NFS"。
- FlexGroup データストアのサイジングを行う場合、 FlexVol は、より大容量のネームスペースを作成する複数の小さい FlexGroup で構成されることに注意してください。そのため、データストアのサイズは、最大のVMDKファイルのサイズの8倍以上（デフォルトのコンスティチュエントが8つの場合）、さらに10~20%の未使用のヘッドルームを使用して、リバランシングを柔軟に実行できるようにします。たとえば、環境に6TBのVMDKがある場合は、FlexGroupデータストアのサイズを52.8TB（6x8+10%）以上に設定します。
- ONTAP 9.14.1以降では、VMwareとNetAppでNFSv4.1セッショントランキングがサポートされます。特定のバージョンの詳細については、NetApp NFS 4.1のInteroperability Matrixの注意事項を参照してください。NFSv3では、ボリュームへの複数の物理パスはサポートされませんが、vSphere 8.0U2以降ではnconnectがサポートされます。nconnectの詳細については、 "NetAppおよびVMwareでのNFSv3 nconnect機能"。
- コピーオフロードには、 NFS Plug-in for VMware VAAI を使用します。前述したように、クローニングはFlexGroupデータストア内で強化されますが、FlexVolボリュームとFlexGroupボリュームの間でVMをコピーする場合、ONTAPはESXiホストのコピーに比べてパフォーマンス上の大きなメリットはありません。そのため、VAAIとFlexGroupのどちらを使用するかを決定する際は、ワークロードのクローニングを検討してください。コンスティチュエントボリュームの数の変更は、FlexGroupベースのクローニングを最適化する1つの方法です。前述のキャッシュタイムアウトの調整と同様に、
- ONTAP tools for VMware vSphere 9.8以降を使用して、ONTAP指標（ダッシュボードとVMレポート）を使用してFlexGroup VMのパフォーマンスを監視し、個 々 のVMのQoSを管理します。現時点では、これらの指標は ONTAP コマンドや API では使用できません。
- SnapCenter Plug-in for VMware vSphereリリース4.4以降では、プライマリストレージシステム上のFlexGroupデータストアのVMのバックアップとリカバリがサポートされます。SCV 4.6では、FlexGroupベースのデータストアに対するSnapMirrorのサポートが追加されています。アレイベースのスナップショットとレプリケーションを使用することは、データを保護する最も効率的な方法です。

ネットワーク構成：

ONTAP ソフトウェアを実行しているシステムで vSphere を使用する場合のネットワーク設定の構成は簡単で、他のネットワーク構成と同様です。

考慮すべき点をいくつか挙げます。

- ストレージネットワークのトラフィックを他のネットワークから分離します。専用の VLAN を使用するか、ストレージ用に別個のスイッチを使用することで、別のネットワークを実現できます。ストレージネットワークがアップリンクなどの物理パスを共有している場合は、十分な帯域幅を確保するために QoS または追加のアップリンクポートが必要になることがあります。ホストをストレージに直接接続しないでください。スイッチを使用して冗長パスを確保し、VMware HAが介入なしで機能できるようにします。を参照してください "直接接続ネットワーク" 追加情報 の場合。
- ジャンボフレームは、必要に応じてネットワークでサポートされていれば、特に iSCSI を使用している場合に使用できます。使用する場合は、ストレージと ESXi ホストの間のパスにあるすべてのネットワークデバイスや VLAN で設定が同じであることを確認してください。そうしないと、パフォーマンスや接続の問題が発生する可能性があります。MTU は、 ESXi 仮想スイッチ、 VMkernel ポート、および各 ONTAP ノードの物理ポートまたはインターフェイスグループでも同一の設定にする必要があります。
- ネットワークフロー制御は、 ONTAP クラスタ内のクラスタネットワークポートでのみ無効にすることを推奨します。データトラフィックに使用される残りのネットワークポートについては、推奨されるベストプラクティスはありません。必要に応じて有効または無効にする必要があります。を参照してください "TR-4182" を参照してください。
- ESXi および ONTAP ストレージアレイをイーサネットストレージネットワークに接続するときは、接続先のイーサネットポートを Rapid Spanning Tree Protocol （ RSTP ；高速スパニングツリープロトコル）のエッジポートとして設定するか、 Cisco の PortFast 機能を使用して設定することを推奨します。ネットアップでは、 Cisco の PortFast 機能を使用していて、 ESXi サーバまたは ONTAP ストレージアレイへの 802.1Q VLAN トランキングが有効になっている環境では、 Spanning-Tree PortFast trunk 機能を有効にすることを推奨します。
- リンクアグリゲーションのベストプラクティスとして次を推奨します。
  - CiscoのVirtual PortChannel（vPC）などのマルチシャーシリンクアグリゲーショングループアプローチを使用して、2つの別 々 のスイッチシャーシ上のポートのリンクアグリゲーションをサポートするスイッチを使用します。
  - LACPが設定されたdvSwitches 5.1以降を使用していない場合、ESXiに接続されているスイッチポートのLACPを無効にします。
  - LACPを使用して、IPハッシュを持つダイナミックマルチモードインターフェイスグループを持つONTAP ストレージシステムのリンクアグリゲートを作成します。
  - ESXiでIPハッシュチーミングポリシーを使用します。

次の表に、ネットワーク設定項目とその適用先をまとめます。

項目	**ESXi**	スイッチ	ノード	**SVM**
IP アドレス	VMkernel	いいえ **	いいえ **	はい。
リンクアグリゲーション	仮想スイッチ	はい。	はい。	いいえ *
VLAN	VMkernel と VM ポートグループ	はい。	はい。	いいえ *

項目	**ESXi**	スイッチ	ノード	**SVM**
フロー制御	NIC	はい。	はい。	いいえ *
スパニングツリー	いいえ	はい。	いいえ	いいえ
MTU （ジャンボフレーム用）	仮想スイッチと VMkernel ポート（ 9000 ）	○（最大に設定）	○（ 9000 ）	いいえ *
フェイルオーバーグループ	いいえ	いいえ	○（作成）	○（選択）

- SVM LIFは、VLANやMTUなどが設定されたポート、インターフェイスグループ、またはVLANインターフェイスに接続します。ただし、設定の管理はSVMレベルではありません。
  - これらのデバイスには管理用に独自の IP アドレスがありますが、 ESXi ストレージネットワークのコンテキストでは使用されません。

SAN （ FC 、 FCoE 、 NVMe/FC 、 iSCSI ）、 RDM

vSphere では、ブロックストレージ LUN を 3 通りの方法で使用します。

- VMFS データストアを使用する場合
- raw デバイスマッピング（ RDM ）で使用
- ソフトウェアイニシエータがアクセスおよび制御する LUN として使用 VM ゲスト OS から作成します

VMFS は、共有ストレージプールであるデータストアを提供する、高性能なクラスタファイルシステムです。VMFS データストアは、 NVMe/FC プロトコルによってアクセスされる FC 、 iSCSI 、 FCoE 、または NVMe ネームスペースを使用してアクセスする LUN で構成できます。VMFS を使用すると、クラスタ内の各 ESX サーバから同時に従来型の LUN にアクセスすることができます。ONTAP の最大 LUN サイズは通常 16TB であるため、最大サイズの 64TB （このセクションの最初の表を参照）の VMFS 5 データストアは、 4 つの 16TB LUN を使用して作成されます（すべての SAN アレイシステムが最大 VMFS LUN サイズ 64TB をサポート）。ONTAP LUN アーキテクチャでは個々のキュー深度が小さくないため、 ONTAP の VMFS データストアは、比較的簡単な方法で従来のアレイアーキテクチャよりも大規模に拡張できます。

vSphere は、ストレージデバイスへの複数のパスを標準でサポートします。この機能はネイティブマルチパス（ NMP ）と呼ばれます。NMP は、サポートされるストレージシステムのストレージタイプを検出し、使用中のストレージシステムの機能をサポートするように NMP スタックを自動的に設定できます。

NMPとONTAPはどちらも、Asymmetric Logical Unit Access（ALUA；非対称論理ユニットアクセス）による最適パスと非最適パスのネゴシエートをサポートします。ONTAP では、アクセス対象の LUN をホストするノード上のターゲットポートを使用する直接データパスが、 ALUA の最適パスとなります。ALUA は、 vSphere と ONTAP の両方でデフォルトで有効になっています。NMPはONTAPクラスタをALUAとして認識し、ALUAストレージアレイタイププラグインを使用します。（`VMW_SATP_ALUA`）を入力し、ラウンドロビンパス選択プラグインを選択します。（`VMW_PSP_RR`）。

ESXi 6 は、最大 256 個の LUN と、 LUN への最大 1 、 024 個の合計パスをサポートします。これらの制限を超える LUN やパスは、 ESXi で認識されません。最大数の LUN を使用した場合、 LUN あたりのパス数は最大 4 つです。大規模な ONTAP クラスタでは、 LUN 数の上限に達する前にパス数の制限に達する可能性があります。この制限に対処するため、 ONTAP では、リリース 8.3 以降の選択的 LUN マップ（ SLM ）がサポートされています。

SLM は、特定の LUN へのパスをアドバタイズするノードを制限します。ネットアップのベストプラクティス

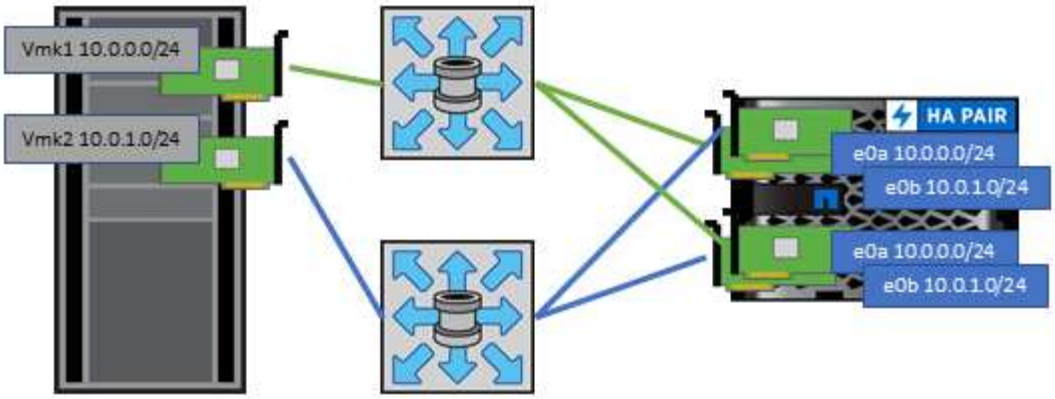
では、各 SVM のノードごとに少なくとも 1 つの LIF を配置し、 SLM を使用して、 LUN とその HA パートナーをホストするノードへのアドバタイズパスを制限することを推奨しています。他のパスは存在しますが、デフォルトではアドバタイズされません。SLM 内で、レポートノードの追加引数および削除引数を使用して通知されたパスを変更することができます。8.3より前のリリースで作成されたLUNではすべてのパスがアドバタイズされるため、ホストしているHAペアへのパスのみがアドバタイズされるように変更する必要があります。SLM の詳細については、のセクション 5.9 を参照してください "TR-4080"。以前のポートセットの方式を使用すると、 LUN の使用可能なパスをさらに削減できます。ポートセットを使用すると、 igroup 内のイニシエータが LUN を認識する際に経由可能なパス数を減らすことができます。

- SLM はデフォルトでは有効になっています。ポートセットを使用しないかぎり、これ以上の設定は必要ありません。
- Data ONTAP 8.3より前のバージョンで作成したLUNの場合、 `lun mapping remove-reporting-nodes` LUNレポートノードを削除し、LUNへのアクセスをLUNの所有者ノードとそのHAパートナーに制限するコマンド。

ブロックプロトコル（ iSCSI 、 FC 、 FCoE ）は、一意の名前に加え、 LUN ID とシリアル番号を使用して LUN にアクセスします。FC と FCoE は Worldwide Name （ WWNN および WWPN ）を使用し、 iSCSI は iSCSI Qualified Name （ IQN ）を使用します。ストレージ内での LUN へのパスはブロックプロトコルにとっては意味がないため、どこにも表示されません。したがって、 LUN のみが含まれるボリュームは内部でマウントする必要がなく、データストアで使用される LUN を含むボリュームのジャンクションパスも必要ありません。ONTAP の NVMe サブシステムも同様に機能します。

考慮すべきその他のベストプラクティス：

- 可用性と移動性を最大限に高めるために、 ONTAP クラスタ内の各ノード上の各 SVM に論理インターフェイス（ LIF ）が作成されていることを確認します。ONTAP SAN では、各ファブリックに対して 1 つずつ、ノードごとに 2 つの物理ポートと LIF を使用することを推奨します。ALUA を使用してパスが解析され、アクティブな最適化（直接）パスとアクティブな非最適化パスが特定されます。ALUA は FC 、 FCoE 、および iSCSI に使用されます。
- iSCSI ネットワークの場合、複数の仮想スイッチがある場合は、 NIC チーミングを使用して、異なるネットワークサブネット上の複数の VMkernel ネットワークインターフェイスを使用します。また、複数の物理スイッチに接続された複数の物理 NIC を使用して、 HA を実現し、スループットを向上させることもできます。次の図に、マルチパス接続の例を示します。ONTAPでは、高可用性とリンクアグリゲーションを実現するために、異なるスイッチへの複数のリンクを含むシングルモードインターフェイスグループを使用するか、マルチモードインターフェイスグループを使用したLACPを使用します。
- ESXiでターゲット認証にチャレンジハンドシェイク認証プロトコル（CHAP）が使用されている場合は、CLIを使用してONTAPでもCHAPを設定する必要があります。 (`vserver iscsi security create`）またはSystem Managerで（[ストレージ]>[SVM]>[SVM設定]>[プロトコル]>[iSCSI]で[イニシエータセキュリティ]を編集します）。
- LUN と igroup の作成と管理には、 VMware vSphere の ONTAP ツールを使用します。プラグインによってサーバの WWPN が自動的に判別され、適切な igroup が作成されます。また、ベストプラクティスに従って LUN を設定し、正しい igroup にマッピングします。
- RDMは管理が困難になる可能性があるため、使用には注意が必要です。また、前述したように制限されているパスも使用します。ONTAP LUN は両方をサポートします "物理互換モードと仮想互換モード" RDM ：
- vSphere 7.0 での NVMe/FC の使用については、以下を参照してください "ONTAP NVMe/FC Host Configuration Guide" および "TR-4684"。次の図は、vSphereホストからONTAP LUNへのマルチパス接続を示しています。



NFS

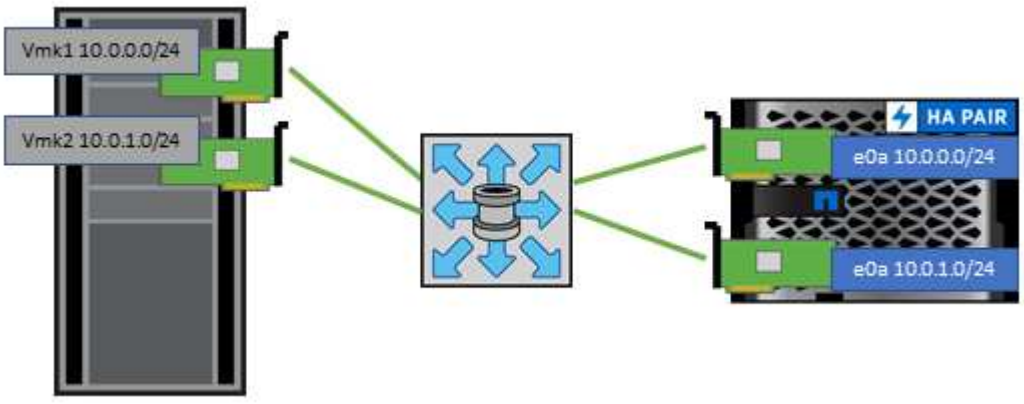
vSphere を使用すると、エンタープライズクラスの NFS アレイを使用して、 ESXi クラスタ内のすべてのノードへのデータストアへの同時アクセスを提供できます。データストアのセクションで説明したように、 vSphere で NFS を使用すると、使いやすさが向上し、ストレージ効率を可視化できるというメリットがあります。

vSphere で ONTAP NFS を使用する際に推奨されるベストプラクティスは次のとおりです。

- ONTAP クラスタ内の各ノードの各 SVM で、 1 つの論理インターフェイス（ LIF ）を使用します。データストアごとの LIF の過去の推奨事項は不要になりました。直接アクセス（LIFとデータストアが同じノード上にある場合）を推奨しますが、一般にパフォーマンスへの影響は最小限（マイクロ秒）であるため、間接アクセスについて心配する必要はありません。
- 現在サポートされているすべてのバージョンのVMware vSphereで、NFS v3とv4.1の両方を使用できます。nconnectの公式サポートは、NFS v3用のvSphere 8.0 Update 2に追加されました。NFS v4.1 のvSphereは、セッショントランキング、Kerberos認証、整合性を維持したKerberos認証を引き続きサポートします。セッショントランキングにはONTAP 9.14.1以降のバージョンが必要であることに注意してください。nconnect機能の詳細と、nconnect機能によってパフォーマンスがどのように向上するかについては、 "NetAppおよびVMwareでのNFSv3 nconnect機能"。

NFSv3とNFSv4.1では、異なるロックメカニズムが使用されていることに注目してください。NFSv3ではクライアント側ロックが使用され、NFSv4.1ではサーバ側ロックが使用されます。ONTAPボリュームは両方のプロトコルでエクスポートできますが、ESXiは1つのプロトコルでしかデータストアをマウントできません。ただしこれは、他のESXiホストが異なるバージョンを使用して同じデータストアをマウントできないという意味ではありません。問題を回避するには、マウント時に使用するプロトコルのバージョンを指定して、すべてのホストで同じバージョン、つまり同じロック形式を使用するようにする必要があります。NFSバージョンをホスト間で混在させないことが重要です。可能であれば、ホストプロファイルを使用して準拠を確認します。データストアは**NFSv3**と**NFSv4.1**の間で自動で変換されないため、新しい**NFSv4.1**データストアを作成し、**Storage vMotion**を使用して新しいデータストアに**VM**を移行します。
NFS v4.1の相互運用性の表を参照してください。 "NetApp Interoperability Matrix Tool で確認できます" をサポートするには、特定の ESXi パッチレベルが必要です。
* NFSエクスポートポリシーは、vSphereホストによるアクセスの制御に使用されます。複数のボリューム（データストア）で 1 つのポリシーを使用できます。NFSv3 では、 ESXi で sys （ UNIX ）セキュリティ形式が使用され、 VM を実行するためにルートマウントオプションが必要となります。ONTAP では、このオプションはスーパーユーザと呼ばれます。スーパーユーザオプションを使用する場合は、匿名ユーザ ID を指定する必要はありません。の値が異なるエクスポートポリシールールに注意してください `-anon` および `-allow -suid` 原因 SVM検出がONTAP ツールで問題を検出できるかどうか。ポリシーの例を次に示します。
アクセスプロトコル：**NFS3**
クライアント一致仕様：192.168.42.21

ROアクセスルール**: sys**
RWアクセスルール: sys
匿名**UID**
スーパーユーザ: sys
* NetApp NFS Plug-in for VMware VAAIを使用する場合、プロトコルは次のように設定する必要があります。 `nfs` エクスポートポリシールールが作成または変更されたとき。VAAIコピーオフロードが機能するためには、次のように指定してNFSv4プロトコルが必要です。 `nfs` NFSv3とNFSv4の両方のバージョンが自動的に含まれます。
* NFSデータストアボリュームはSVMのルートボリュームからジャンクションされるため、ESXiがデータストアボリュームに移動してマウントするには、ルートボリュームへのアクセスも必要です。ルートボリューム、およびデータストアボリュームのジャンクションがネストされているその他のボリュームのエクスポートポリシーには、ESXiサーバに読み取り専用アクセスを許可するルールが含まれている必要があります。VAAIプラグインを使用したルートボリュームのポリシーの例を次に示します。
アクセスプロトコル：**NFS**（**NFS3**と**nfs4**の両方を含む）
クライアント一致仕様：192.168.42.21
ROアクセスルール**: sys**
RW Access Rule：never（ルートボリュームに最適なセキュリティ）
匿名**UID**
Superuser：sys（VAAIを使用するルートボリュームにも必要）
* ONTAP Tools for VMware vSphere（最も重要なベストプラクティス）を使用します。
ONTAP Tools for VMware vSphereを使用すると、エクスポートポリシーの管理が自動的に簡素化されるため、データストアをプロビジョニングできます。
プラグインを使用してVMwareクラスタ用のデータストアを作成する場合は、単一のESXサーバではなくクラスタを選択します。これにより、データストアがクラスタ内のすべてのホストに自動的にマウントされます。既存のデータストアを新しいサーバに適用するには、プラグインマウント機能を使用します。
ONTAP Tools for VMware vSphereを使用しない場合は、すべてのサーバ、または追加のアクセス制御が必要なサーバのクラスタごとに1つのエクスポートポリシーを使用します。
* ONTAPは柔軟なボリュームネームスペース構造を提供し、ジャンクションを使用してボリュームをツリーにまとめることができますが、このアプローチはvSphereには意味がありません。ストレージのネームスペース階層に関係なく、データストアのルートに各 VM 用のディレクトリが作成されます。そのため、単に SVM のルートボリュームに vSphere のボリュームのジャンクションパスをマウントすることがベストプラクティスです。これは、 VMware vSphere 用の ONTAP ツールでデータストアをプロビジョニングする方法です。ジャンクションパスがネストされていないと、ルートボリューム以外のボリュームに依存しているボリュームがないこと、またボリュームをオフラインにするか破棄するかによって意図的に他のボリュームへのパスに影響が及ぶこともありません。
* NFSデータストア上のNTFSパーティションでは、ブロックサイズを4Kに設定しても問題ありません。次の図は、 vSphere ホストから ONTAP NFS データストアへの接続を示しています。



次の表に、 NFS のバージョンとサポートされる機能を示します。

vSphere の機能	NFSv3	NFSv4.1
vMotion と Storage vMotion	はい。	はい。
高可用性	はい。	はい。
フォールトトレランス	はい。	はい。
DRS	はい。	はい。
ホストプロファイル	はい。	はい。
Storage DRS	はい。	いいえ
ストレージ I/O の制御	はい。	いいえ
SRM の場合	はい。	いいえ
仮想ボリューム	はい。	いいえ
ハードウェアアクセラレーション（ VAAI ）	はい。	はい。
Kerberos 認証	いいえ	○（ vSphere 6.5 以降で拡張して、 AES 、 krb5i ）
マルチパスのサポート	いいえ	○（ONTAP 9.14.1）

直接接続ネットワーク

ストレージ管理者は、構成からネットワークスイッチを削除してインフラを簡易化したいと考える場合があります。これは一部のシナリオでサポートされます。

iSCSIとNVMe/TCP

iSCSIまたはNVMe/TCPを使用するホストは、ストレージシステムに直接接続して正常に動作することができます。その理由はパス設定です。2つの異なるストレージコントローラに直接接続すると、データフローが2つの独立したパスになります。パス、ポート、またはコントローラが失われても、他のパスの使用が妨げられることはありません。

NFS

直接接続されたNFSストレージも使用できますが、フェイルオーバーには大きな制限があります。スクリプト作成にはお客様の責任が伴います。

直接接続されたNFSストレージで無停止フェイルオーバーが複雑になるのは、ローカルOSで発生するルーティングが原因です。たとえば、ホストのIPアドレスが192.168.1.1/24で、IPアドレスが192.168.1.50/24のONTAPコントローラに直接接続されているとします。フェールオーバー中、192.168.1.50アドレスはもう一方のコントローラにフェールオーバーでき、ホストが使用できるようになりますが、ホストはそのアドレスの存在をどのように検出しますか。元の192.168.1.1アドレスは、動作中のシステムに接続されていないホストNICに残っています。192.168.1.50宛てのトラフィックは、動作不能なネットワークポートに引き続き送信されます。

2番目のOS NICは19に設定できます。 2.168.1.2およびは、192.168.1.50経由でフェールオーバーされたアドレスと通信できますが、ローカルルーティングテーブルのデフォルトでは、192.168.1.0/24サブネットと通信するために1つの*および1つの*アドレスのみを使用することになります。システム管理者は、失敗したネットワーク接続を検出し、ローカルルーティングテーブルを変更したり、インターフェイスをアップ/ダウンしたりするスクリプトフレームワークを作成できます。正確な手順は、使用しているOSによって異なります。

実際にはNetAppを使用していますが、通常はフェイルオーバー中のIO一時停止が許容されるワークロードのみが対象です。ハードマウントを使用する場合は、一時停止中にIOエラーが発生しないようにしてください。ホスト上のNIC間でIPアドレスを移動するためのフェイルバックまたは手動操作によって、サービスが復元されるまでIOはハングします。

FC直接接続

FCプロトコルを使用してホストをONTAPストレージシステムに直接接続することはできません。その理由はNPIVの使用です。FCネットワークへのONTAP FCポートを識別するWWNは、NPIVと呼ばれる仮想化タイプを使用します。ONTAPシステムに接続されているすべてのデバイスがNPIV WWNを認識できる必要があります。現在、NPIVターゲットをサポートできるホストにインストールできるHBAを提供しているHBAベンダーはありません。

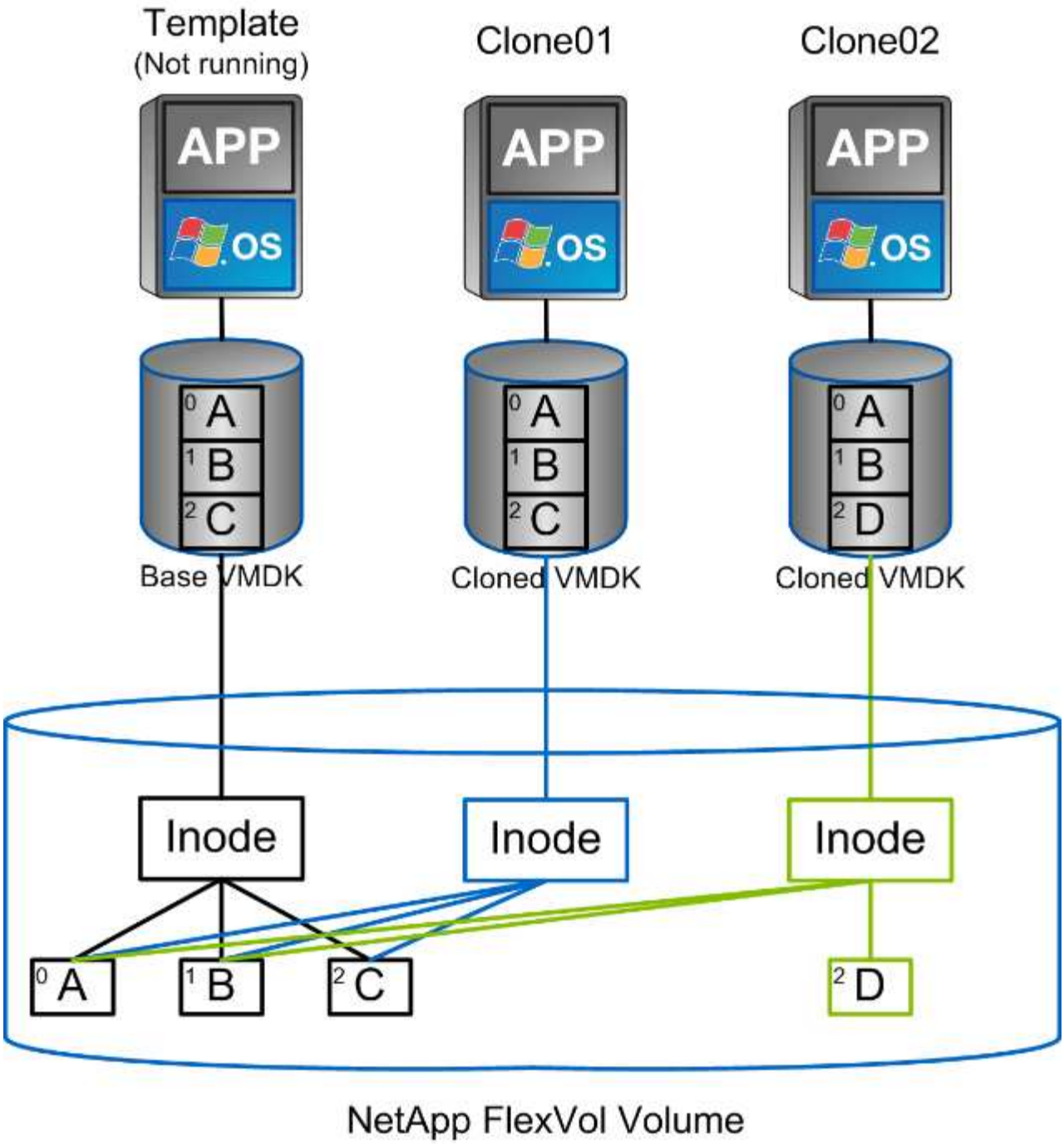
VM とデータストアのクローニング

ストレージオブジェクトをクローニングすると、追加の VM のプロビジョニングやバックアップ / リカバリ処理などの用途に使用できるコピーを簡単に作成できます。

vSphere では、 VM 、仮想ディスク、 VVOL 、またはデータストアをクローニングできます。クローニングされたオブジェクトは、多くの場合、自動化されたプロセスによってさらにカスタマイズできます。vSphereでは、フルコピークローンとリンククローンの両方がサポートされます。リンククローンでは、元のオブジェクトとは別に変更が追跡されます。

リンククローンはスペースを節約するのに適していますが、 vSphere が VM に対して処理する I/O 量が増えるため、その VM のパフォーマンスや場合によってはホスト全体のパフォーマンスに影響します。そのため、NetAppのお客様は、ストレージシステムベースのクローンを使用して、ストレージの効率的な使用とパフォーマンスの向上という2つのメリットを活用することがよくあります。

次の図は、 ONTAP クローニングを示しています。



クローニングは、 ONTAP ソフトウェアを実行するシステムに複数のメカニズムを使用してオフロードできます。通常は、 VM 、 VVol 、データストアのレベルでオフロードします。これには次のものが含まれます。

- NetApp vSphere APIs for Storage Awareness （ VASA ） Provider を使用した VVol のクローニング。vCenterで管理されるVVol Snapshotをサポートするために、ONTAPクローンを使用します。VVol Snapshotの作成や削除によるI/Oへの影響は最小限で、スペース効率に優れています。 VM のクローニングは vCenter を使用して行うこともでき、 1 つのデータストア / ボリューム内かデータストア / ボリューム間かに関係なく、 ONTAP にオフロードされます。
- vSphere APIs – Array Integration （ VAAI ）を使用した vSphere のクローニングと移行：SAN 環境と NAS 環境の両方で、 VM のクローニング処理を ONTAP にオフロードできます（ネットアップでは、 NFS 用の VAAI を有効にするために ESXi プラグインを提供しています）。 vSphere は、 NAS データストア内のコールド（電源オフ） VM にのみオフロードします。一方、ホット VM （クローニングと

Storage vMotion ）の処理も SAN にオフロードされます。ONTAP では、ソース、デスティネーション、インストールされている製品ライセンスに基づいて最も効率的なアプローチを採用しています。この機能は VMware Horizon View でも使用されています。

- SRA （ VMware Site Recovery Manager で使用）。ここでは、クローンを使用して、 DR レプリカのリカバリを無停止でテストします。
- SnapCenter などのネットアップのツールを使用したバックアップとリカバリVM クローンは、バックアップ処理の検証や VM バックアップのマウントに使用され、個々のファイルをコピーできるようにします。

ONTAP オフロードクローニングは、 VMware 、ネットアップ、サードパーティのツールから実行できます。ONTAP にオフロードされたクローンには、いくつかのメリットがあります。ほとんどの場合、スペース効率に優れており、オブジェクトの変更にのみ対応するストレージが必要です。読み取りや書き込みのパフォーマンスには影響しません。また、高速キャッシュでブロックを共有することでパフォーマンスが向上する場合もあります。また、 CPU サイクルとネットワーク I/O も ESXi サーバからオフロードされます。FlexVol を使用する従来のデータストア内でのコピーオフロードは、 FlexClone ライセンスを使用すると高速かつ効率的ですが、 FlexVol 間のコピーの方が低速になる可能性があります。VM テンプレートをクローンのソースとして管理する場合は、スペース効率に優れた高速クローンを作成するために、テンプレートをデータストアボリューム内に配置することを検討してください（フォルダやコンテンツライブラリを使用してテンプレートを整理します）。

ONTAP 内で直接ボリュームまたは LUN をクローニングして、データストアをクローニングすることもできます。NFS データストアの場合は、 FlexClone テクノロジでボリューム全体をクローニングし、 ONTAP からクローンをエクスポートして、別のデータストアとして ESXi にマウントできます。VMFS データストアの場合は、ボリューム内の LUN 、または 1 つ以上の LUN を含むボリューム全体を ONTAP でクローニングできます。VMFS を含む LUN を通常のデータストアとしてマウントして使用するためには、 LUN を ESXi igroup にマッピングし、 ESXi から再署名を受ける必要があります。ただし一部の一時的なユースケースでは、クローニングされた VMFS を再署名なしでマウントすることができます。クローニングしたデータストア内の VM は、個別にクローニングした VM と同様に登録、再設定、およびカスタマイズすることができます。

バックアップや FlexClone 用の SnapRestore など、追加のライセンス機能を使用してクローニングを強化できる場合があります。これらのライセンスは、追加コストなしでライセンスバンドルに含まれていることがよくあります。FlexCloneライセンスは、VVolのクローニング処理や、VVolの管理対象Snapshot（ハイパーバイザーからONTAPにオフロードされる）をサポートするために必要です。FlexClone をデータストア / ボリューム内で使用すると、特定の VAAI ベースのクローンの品質を向上させることもできます（ブロックコピーではなく、スペース効率に優れたコピーが瞬時に作成されます）。 また、 DR レプリカのリカバリをテストする際に SRA で使用され、クローニング処理用に SnapCenter でバックアップコピーを参照して個々のファイルをリストアする際にも使用されます。

データ保護

VM のバックアップと迅速なリカバリは、 ONTAP for vSphere の大きな特長の 1 つです。この機能は、 SnapCenter Plug-in for VMware vSphere を使用して vCenter 内で簡単に管理できます。

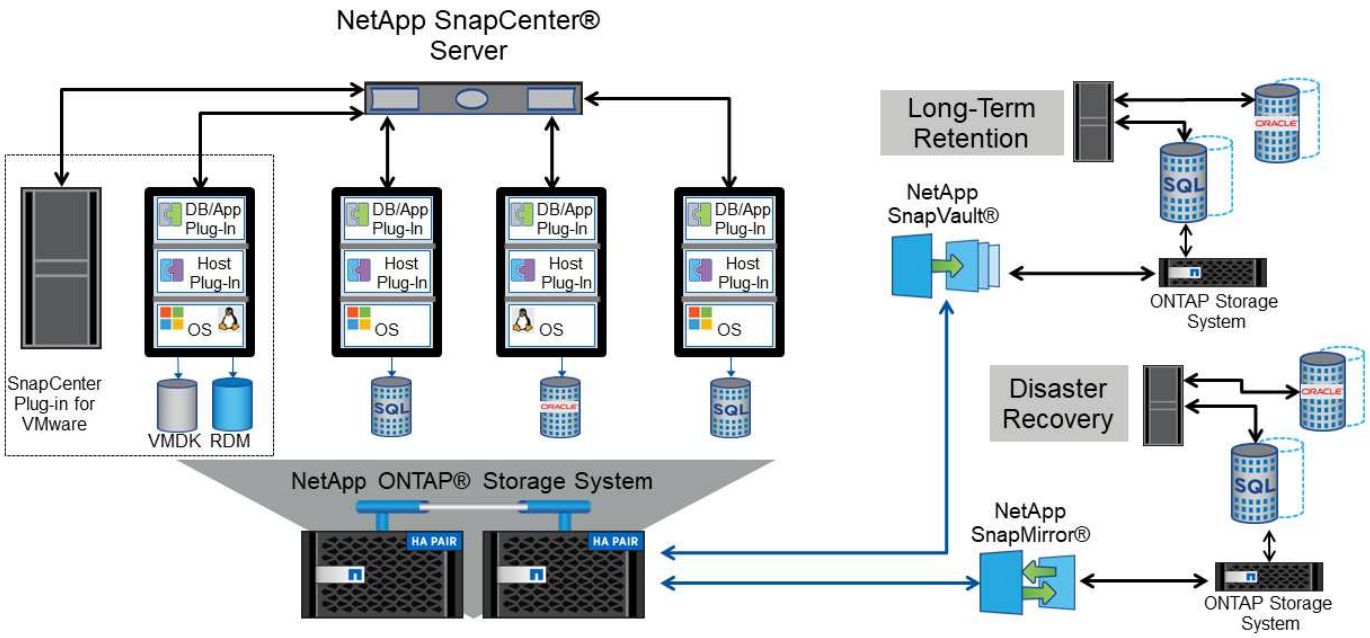
Snapshotを使用すると、パフォーマンスに影響を与えずにVMやデータストアのコピーをすばやく作成でき、SnapMirrorを使用してセカンダリシステムに送信することで、オフサイトでの長期的なデータ保護を実現できます。このアプローチでは、変更された情報のみを格納することで、ストレージスペースとネットワーク帯域幅を最小限に抑えます。

SnapCenter では、複数のジョブに適用可能なバックアップポリシーを作成できます。これらのポリシーでは、スケジュール、保持、レプリケーションなどの機能を定義できます。VMwareスナップショットを作成す

る前にI/Oを休止するハイパーバイザーの機能を活用して、VM整合性スナップショットをオプションで選択できます。ただし、 VMware スナップショットはパフォーマンスへの影響があるため、ゲストファイルシステムを休止する必要がないかぎり、一般には推奨されません。代わりに、スナップショットを使用して一般的な保護を行い、SnapCenterプラグインなどのアプリケーションツールを使用してSQL ServerやOracleなどのトランザクションデータを保護します。これらのスナップショットはVMware（整合性）スナップショットとは異なり、長期的な保護に適しています。 VMware スナップショットはのみです "（推奨）" パフォーマンスやその他の影響があるため、短期的な使用に適しています。

これらのプラグインは、物理環境と仮想環境の両方でデータベースを保護する拡張機能を提供します。vSphere では、これらのプロトコルを使用して、 RDM LUN 、ゲスト OS に直接接続された iSCSI LUN 、 VMFS または NFS データストア上の VMDK ファイルにデータが格納されている SQL Server または Oracle データベースを保護できます。プラグインでは、さまざまなタイプのデータベースバックアップを指定し、オンラインまたはオフラインのバックアップをサポートし、ログファイルとともにデータベースファイルを保護できます。プラグインは、バックアップとリカバリに加えて、開発やテスト目的でのデータベースのクローニングにも対応しています。

次の図は、 SnapCenter の導入例を示しています。



ディザスタリカバリ機能を強化するには、 ONTAP 用 NetApp SRA と VMware Site Recovery Manager の使用を検討してください。DR サイトへのデータストアのレプリケーションをサポートするだけでなく、レプリケートしたデータストアをクローニングすることで DR 環境を無停止でテストすることもできます。SRA に組み込まれている自動化機能を使用すると、災害からのリカバリや、システム停止が解決したあとの本番環境の再保護も簡単に実行できます。

最後に、最高レベルのデータ保護を実現するために、 NetApp MetroCluster を使用した VMware vSphere Metro Storage Cluster （ vMSC ）設定を検討してください。vMSC は、同期レプリケーションとアレイベースのクラスタリングを組み合わせた VMware 認定の解決策です。高可用性クラスタと同じメリットを提供しますが、複数のサイトに分散してサイト障害から保護します。NetApp MetroCluster は、同期レプリケーション向けの対費用効果の高い構成を提供します。ストレージコンポーネントのあらゆる単一障害から透過的にリカバリでき、サイト障害時にコマンド 1 つでリカバリできます。vMSC の詳細については、を参照してください "TR-4128"。

サービス品質（ **QoS** ）

ONTAP ソフトウェアを実行するシステムでは、 ONTAP ストレージ QoS 機能を使用して、ファイル、 LUN 、ボリューム、 SVM 全体などの異なるストレージオブジェクトに対するスループットを MBps や IOPS （ 1 秒あたりの I/O 数）で制限できます。

スループット制限は、他のワークロードに影響しないように、導入前に未知のワークロードやテストワークロードを制御するのに役立ちます。また、 Bully ワークロードが特定された場合に、この 2 つを使用して抑制することもできます。ONTAP 9.2 では SAN オブジェクトに、 ONTAP 9.3 では NAS オブジェクトに一貫したパフォーマンスを提供するために、 IOPS に基づく最小サービスレベルもサポートされています。

NFS データストアの場合は、 QoS ポリシーを FlexVol 全体またはボリューム内の個々の VMDK ファイルに適用できます。ONTAP LUN を使用する VMFS データストアでは、 LUN を含む FlexVol ボリュームには QoS ポリシーを適用できますが、 ONTAP が VMFS ファイルシステムを認識しないため、個々の VMDK ファイルには適用できません。VVol を使用する場合は、ストレージ機能プロファイルと VM ストレージポリシーを使用して、個々の VM に最小 QoS と最大 QoS を設定できます。

オブジェクトに対する QoS の最大スループット制限は、 MBps と IOPS のいずれかまたは両方で設定できます。両方を使用する場合は、最初に到達した制限が ONTAP によって適用されます。ワークロードには複数のオブジェクトを含めることができ、 QoS ポリシーは 1 つ以上のワークロードに適用できます。ポリシーを複数のワークロードに適用した場合は、ポリシーの制限はワークロード全体に適用されます。ネストされたオブジェクトはサポートされません（たとえば、ボリューム内のファイルには個別のポリシーを設定することはできません）。QoS の最小値は IOPS 単位でのみ設定できます。

ONTAP QoS ポリシーの管理とオブジェクトへの適用に現在使用できるツールは次のとおりです。

- ONTAP CLI
- ONTAP システムマネージャ
- OnCommand Workflow Automation のサポートを利用できます
- Active IQ Unified Manager
- NetApp PowerShell Toolkit for ONTAP
- VMware vSphere VASA Provider 用の ONTAP ツール

NFS 上の VMDK に QoS ポリシーを割り当てる場合は、次のガイドラインに注意してください。

- ポリシーは、 `vmname- flat.vmdk` ではなく、実際の仮想ディスクイメージが含まれています。 `vmname.vmdk` （仮想ディスク記述ファイル）または `vmname.vmx` （VM記述ファイル）。
- 仮想スワップファイルなど、他のVMファイルにポリシーを適用しない (`vmname.vswp`）。
- vSphere Web Clientを使用してファイルパスを検索する場合（[Datastore]>[Files]）は、 `- flat.vmdk` および `. vmdk` 1つのファイルが表示されます。このファイルには、 `. vmdk` しかしその大きさは `- flat.vmdk`。追加（ Add ） `-flat` ファイル名に入力して、正しいパスを取得します。

VMFS と RDM 、 ONTAP SVM （ SVM として表示）、 LUN パス、シリアル番号などの LUN に QoS ポリシーを割り当てるには、 ONTAP Tools for VMware vSphere のホームページのストレージシステムメニューから QoS ポリシーを取得します。ストレージシステム（SVM）を選択し、[Related Objects]>[SAN]を選択します。 この方法は、いずれかの ONTAP ツールを使用して QoS を指定する場合に使用します。

VVol ベースの VM には、 VMware vSphere または Virtual Storage Console 7.1 以降の ONTAP ツールを使用して、最大 QoS と最小 QoS を簡単に割り当てることができます。VVolコンテナのストレージ機能プロファ

イルを作成するときは、パフォーマンス機能で最大IOPSと最小IOPSの値を指定し、このSCPをVMのストレージポリシーで参照します。このポリシーは VM を作成するときに使用するか、ポリシーを既存の VM に適用します。

FlexGroup データストアでは、 ONTAP ツールを VMware vSphere 9.8 以降で使用する場合に、 QoS 機能が強化されています。QoS は、データストア内のすべての VM 、または特定の VM に簡単に設定できます。詳細については、本レポートの「 FlexGroup 」セクションを参照してください。

ONTAP の QoS と VMware の SIOC

ONTAP の QoS と VMware vSphere の Storage I/O Control （ SIOC ）は、 vSphere 管理者とストレージ管理者が組み合わせて、 ONTAP ソフトウェアを実行するシステムでホストされる vSphere VM のパフォーマンスを管理できる、相互に補完するテクノロジです。各ツールには、次の表に示すようにそれぞれの長所があります。VMware vCenter と ONTAP ではスコープが異なるため、一部のオブジェクトは一方のシステムで認識および管理でき、もう一方のシステムではできません。

プロパティ（ **Property** ）	**ONTAP QoS**	**VMware SIOC**
アクティブになっている場合	ポリシーは常にアクティブです	競合が発生している（データストアのレイテンシがしきい値を超えている）場合
単位のタイプ	IOPS 、 MBps	IOPS 、共有数
対象となる vCenter またはアプリケーション	複数の vCenter 環境、その他のハイパーバイザーとアプリケーションがあります	単一の vCenter サーバ
VM に QoS を設定？	NFS 上の VMDK のみ	NFS 上または VMFS 上の VMDK です
LUN （ RDM ）で QoS を設定？	はい。	いいえ
LUN （ VMFS ）への QoS の設定	はい。	いいえ
ボリューム（ NFS データストア）への QoS の設定	はい。	いいえ
SVM （テナント）に QoS を設定？	はい。	いいえ
ポリシーベースのアプローチ	はい。ポリシー内のすべてのワークロードで共有することも、ポリシー内の各ワークロードにフルに適用することもできます。	はい。 vSphere 6.5 以降が必要です。
ライセンスが必要です	ONTAP に付属しています	Enterprise Plus

VMware Storage Distributed Resource Scheduler の略

VMware Storage Distributed Resource Scheduler （ SDRS ）は、現在の I/O レイテンシとスペース使用量に基づいて VM をストレージに配置する vSphere の機能です。その後、 VM や VMDK の配置先として最適なデータストアをデータストアクラスタ内から選択し、システムを停止することなくデータストアクラスタ（ポッドとも呼ばれます）内のデータストア間で VM や VMDK を移動します。データストアクラスタは、類似するデータストアをvSphere管理者から見た単一の消費単位に集約したものです。

SDRSとONTAP tools for VMware vSphereを使用する場合は、まずプラグインを使用してデータストアを作成し、vCenterを使用してデータストアクラスタを作成してから、そのデータストアにデータストアを追加する

必要があります。データストアクラスタを作成したら、プロビジョニングウィザードの詳細ページからデータストアクラスタにデータストアを直接追加できます。

SDRS に関するその他の ONTAP のベストプラクティスは、次のとおりです。

- クラスタ内のすべてのデータストアで同じタイプのストレージ（ SAS 、 SATA 、 SSD など）を使用し、すべて VMFS データストアまたは NFS データストアとし、レプリケーションと保護の設定を同じにします。
- デフォルト（手動）モードでは SDRS の使用を検討してください。このアプローチでは、推奨事項を確認し、適用するかどうかを決定できます。VMDK の移行による影響を次に示します。
  - SDRS がデータストア間で VMDK を移動すると、 ONTAP のクローニングや重複排除によるスペース削減効果は失われます。重複排除機能を再実行すれば、削減効果を取り戻すことができます。
  - NetAppでは、VMDKを移動したあとに、移動したVMによってスペースがロックされるため、ソースデータストアでSnapshotを再作成することを推奨しています。
  - 同じアグリゲート上のデータストア間で VMDK を移動してもメリットはほとんどなく、 SDRS はアグリゲートを共有する可能性のある他のワークロードを可視化できません。

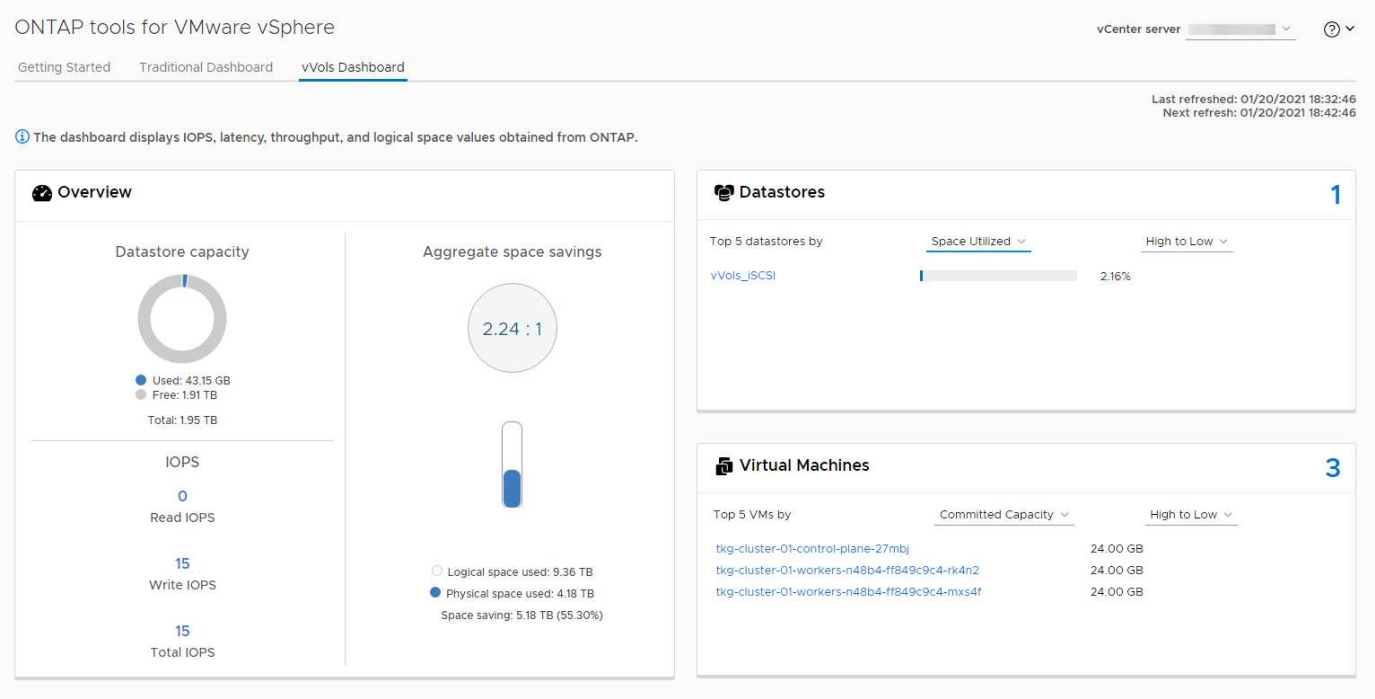
ストレージポリシーベースの管理と**VVOL**

VMware vSphere APIs for Storage Awareness （ VASA ）を使用すると、ストレージ管理者は、明確に定義された機能を使用してデータストアを簡単に設定でき、 VM 管理者は、相互にやり取りすることなく、いつでも VM をプロビジョニングするためのこれらの機能を使用できます。このアプローチを見て、仮想化ストレージの運用を合理化し、単純な作業の多くを回避する方法を確認することをお勧めします。

VASA が導入される前は、 VM 管理者が VM ストレージポリシーを定義することもできましたが、適切なデータストアを特定するには、多くの場合、ドキュメントや命名規則を使用する必要がありました。VASA を使用すると、ストレージ管理者は、パフォーマンス、階層化、暗号化、レプリケーションなど、さまざまなストレージ機能を定義できます。1 つのボリュームまたはボリュームセットの一連の機能を、ストレージ機能プロファイル（ SCP ）と呼びます。

SCPでは、VMのデータVVOLに対して最小または最大のQoSがサポートされます。最小 QoS は AFF システムでのみサポートされます。VMware vSphere 用の ONTAP ツールには、 ONTAP システム上の VVOL の VM の詳細なパフォーマンスと論理容量を表示するダッシュボードがあります。

次の図は、 VMware vSphere 9.8 VVol ダッシュボード用の ONTAP ツールを示しています。



ストレージ機能プロファイルを定義したら、そのプロファイルを使用して要件を定義するストレージポリシーを使用して VM をプロビジョニングできます。vCenter では、 VM ストレージポリシーとデータストアストレージ機能プロファイルのマッピングに基づいて、互換性があるデータストアのリストを選択対象として表示できます。このアプローチは、ストレージポリシーベースの管理と呼ばれます。

VASA は、ストレージを照会して一連のストレージ機能を vCenter に返すためのテクノロジを提供します。VASA ベンダープロバイダは、ストレージシステムの API およびコンストラクトと、 vCenter が認識可能な VMware API との間の変換機能を提供します。ネットアップのVASA Provider for ONTAPは、ONTAP Tools for VMware vSphereアプライアンスVMの一部として提供されます。vCenterプラグインは、VVOLデータストアをプロビジョニングおよび管理するためのインターフェイスと、ストレージ機能プロファイル（SCP）を定義する機能を提供します。

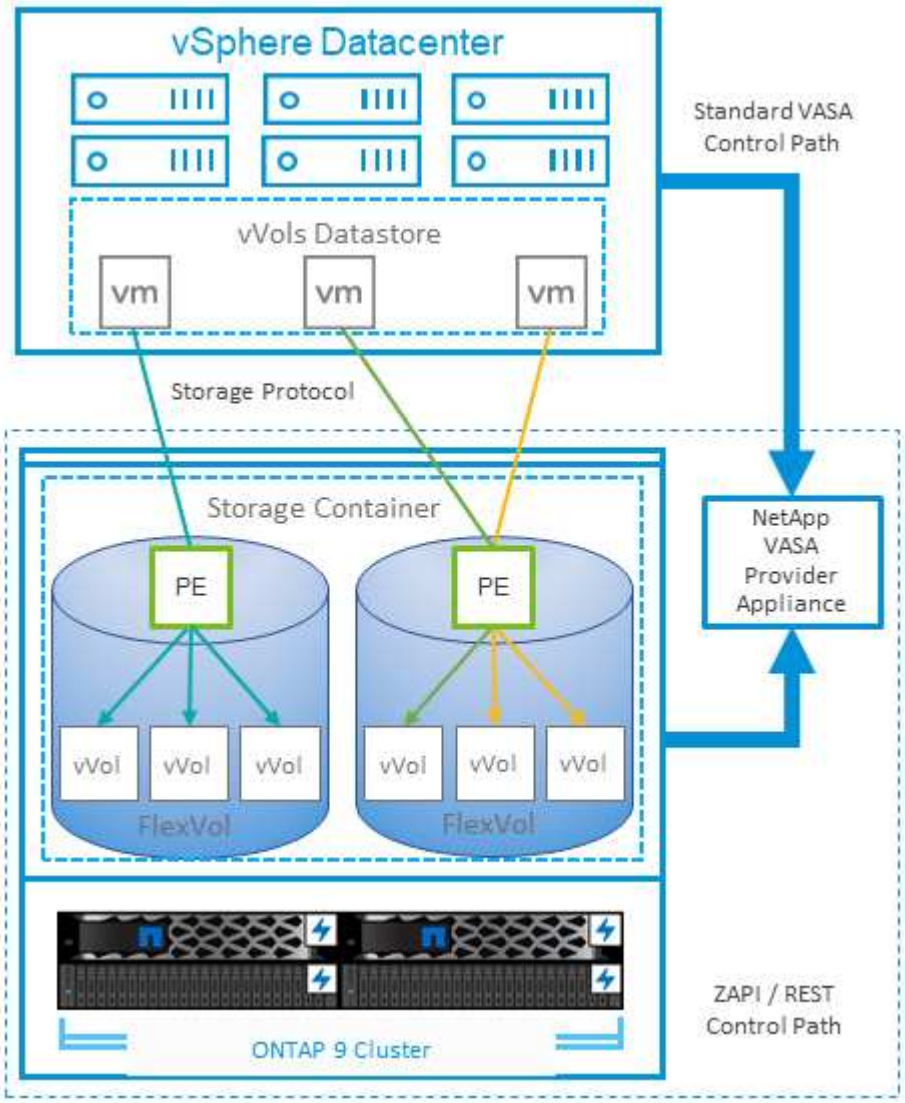
ONTAP は、 VMFS データストアと NFS データストアの両方をサポートしています。SAN データストアで VVOL を使用すると、 VM レベルのきめ細かさなど、 NFS のメリットの一部を活用できます。ここでは考慮すべきベストプラクティスをいくつか示します。また、追加情報はにあります "TR-4400"：

- VVOL データストアは、複数のクラスタノードにある複数の FlexVol で構成できます。ボリュームごとに機能が異なる場合でも、最もシンプルなアプローチは 1 つのデータストアです。SPBM により、互換性のあるボリュームが VM に使用されています。ただし、すべてのボリュームが 1 つの ONTAP SVM に含まれていて、単一のプロトコルでアクセスできる必要があります。各プロトコルでノードごとに 1 つの LIF で十分です。1 つの VVOL データストアで複数の ONTAP リリースを使用することは避けてください。リリースによってストレージ機能が異なる場合があります。
- VVol データストアの作成と管理には、 VMware vSphere プラグインの ONTAP ツールを使用します。データストアとそのプロファイルの管理に加え、必要に応じて、 VVOL にアクセスするためのプロトコルエンドポイントが自動的に作成されます。LUN を使用する場合、 LUN PE は 300 以上の LUN ID を使用してマッピングされます。ESXiホストの詳細なシステム設定を確認する `Disk.MaxLUN` 300を超えるLUN ID番号を許可します（デフォルトは1、024）。そのためには、vCenterでESXiホストを選択し、[Configure]タブで `Disk.MaxLUN` をクリックします。
- VASA Provider 、 vCenter Server （アプライアンスまたは Windows ベース）、または VMware vSphere 用の ONTAP ツールは相互に依存するため、 VVOL データストアにインストールしたり移行したりしないでください。これらのツールは、停電やその他のデータセンターの停止が発生した場合に管理しなくなる

ためです。

- VASA Provider VM を定期的にバックアップします。VASA Providerが格納された従来のデータストアのSnapshotを少なくとも1時間ごとに作成してください。VASA Provider の保護とリカバリの詳細については、こちらを参照してください "こちらの技術情報アーティクル"。

次の図は、 VVOL のコンポーネントを示しています。



クラウドへの移行とバックアップ

ONTAP のもう 1 つの強みは、ハイブリッドクラウドを幅広くサポートすることで、オンプレミスのプライベートクラウドのシステムとパブリッククラウドの機能を統合できることです。vSphere と組み合わせて使用できるネットアップのクラウドソリューションには、次のものがあります。

- * Cloud Volumes。* NetApp Cloud Volumes Service for Amazon Web ServicesまたはGoogle Cloud PlatformとAzure NetApp Files for ANFは、主要なパブリッククラウド環境でハイパフォーマンスなマルチプロトコルマネージドストレージサービスを提供します。VMware Cloud VM ゲストで直接使用できます。
- * Cloud Volumes ONTAP 。 * NetApp Cloud Volumes ONTAP データ管理ソフトウェアは、お客様が選択したクラウド上のデータを管理、保護、柔軟性、効率性で保護します。Cloud Volumes ONTAP は、ONTAPストレージ上に構築されたクラウドネイティブのデータ管理ソフトウェアです。Cloud

Volumes ONTAP インスタンスをオンプレミスの ONTAP システムと一緒に導入、管理する際には、Cloud Manager と組み合わせて使用できます。NASおよびiSCSI SANの高度な機能と、スナップショットやSnapMirrorレプリケーションなどの統合データ管理機能を活用できます。

- * Cloud Backup Service * 。クラウドサービスまたは SnapMirror クラウドを使用して、パブリッククラウドストレージを使用してオンプレミスシステムからデータを保護します。Cloud Sync を使用すると、NAS 、オブジェクトストア、 Cloud Volumes Service ストレージ間でデータを移行し、同期を維持できます。
- * ONTAP * FabricPool は、 FabricPool データの階層化を迅速かつ容易にします。コールドブロックは、パブリッククラウドまたはStorageGRIDのプライベートオブジェクトストアにあるオブジェクトストアに移行でき、ONTAPデータが再度アクセスされると自動的にリコールされます。または、 SnapVault ですでに管理されているデータの第 3 レベルの保護としてオブジェクト階層を使用することもできます。この方法を使用すると、を実行できます "VMのより多くのスナップショットを保存" プライマリおよびセカンダリ ONTAP ストレージシステム。
- * ONTAP Select * 。ネットアップの Software-Defined Storage を使用して、インターネット経由でプライベートクラウドをリモートの施設やオフィスに拡張できます。 ONTAP Select を使用すれば、ブロックサービスやファイルサービスのほか、エンタープライズデータセンターと同じ vSphere データ管理機能をサポートできます。

VM ベースのアプリケーションを設計する際は、将来のクラウドのモビリティを考慮してください。たとえば、アプリケーションファイルとデータファイルを一緒に配置するのではなく、データ用に別の LUN または NFS エクスポートを使用します。これにより、 VM とデータを別々にクラウドサービスに移行できます。

vSphere データの暗号化

現在、保管データを暗号化で保護する必要性はますます高まっています。当初は財務情報や医療情報に重点が置かれていましたが'ファイル'データベース'その他のデータタイプに保存されているかどうかにかかわらず'すべての情報を保護することへの関心が高まっています

ONTAP ソフトウェアを実行するシステムでは、保存データの暗号化を使用してあらゆるデータを簡単に保護できます。NetApp Storage Encryption （ NSE ）は、 ONTAP を備えた自己暗号化ディスクドライブを使用して、 SAN と NAS のデータを保護します。また、 NetApp Volume Encryption と NetApp Aggregate Encryption も、シンプルなソフトウェアベースの手法として、ディスクドライブ上のボリュームを暗号化します。このソフトウェア暗号化では、特別なディスクドライブや外部キー管理ツールは必要ありません。ONTAPのお客様は追加料金なしで利用できます。クライアントやアプリケーションを停止することなくアップグレードして使用を開始でき、オンボードキーマネージャなどの FIPS 140-2 レベル 1 標準で検証されます。

VMware vSphere 上で実行される仮想アプリケーションのデータを保護する方法はいくつかあります。1 つは、 VM 内のソフトウェアをゲスト OS レベルで使用してデータを保護する方法です。別の方法として、vSphere 6.5 などの新しいハイパーバイザーでは VM レベルの暗号化がサポートされるようになりました。ただし、ネットアップのソフトウェア暗号化はシンプルで使いやすく、次のようなメリットがあります。

- * 仮想サーバの CPU には影響しません。 * 仮想サーバ環境によっては、アプリケーションに使用可能なすべての CPU サイクルが必要ですが、ハイパーバイザーレベルの暗号化では最大 5 倍の CPU リソースが必要です。暗号化ソフトウェアがインテルのAES-NI命令セットをサポートして暗号化ワークロードをオフロードしていても（NetAppソフトウェア暗号化がサポートしているように）、古いサーバーと互換性のない新しいCPUが必要なため、このアプローチは実現できない可能性があります。
- * オンボードキーマネージャを含む。 * ネットアップのソフトウェア暗号化機能には、追加料金なしでオンボードキーマネージャが含まれているため、購入や使用が複雑な高可用性キー管理サーバなしで簡単に利用を開始できます。
- * ストレージ効率への影響はありません。 * 重複排除や圧縮などの Storage Efficiency テクノロジは現在広

く使用されており、フラッシュディスクメディアをコスト効率よく使用する上で鍵となります。ただし、一般に、暗号化されたデータは重複排除も圧縮もできません。ネットアップのハードウェアとストレージの暗号化は下位レベルで動作し、他のアプローチとは異なり、業界をリードするネットアップの Storage Efficiency 機能を最大限に活用できます。

- * データストアのきめ細かい暗号化が容易。 * NetApp Volume Encryption を使用すると、各ボリュームに専用の AES 256 ビットキーが設定されます。変更が必要な場合は、 1 つのコマンドで変更できます。このアプローチは、テナントが複数ある場合や、さまざまな部門やアプリケーションに対して個別に暗号化を証明する必要がある場合に適しています。この暗号化はデータストアレベルで管理されるため、個々の VM の管理よりもはるかに簡単です。

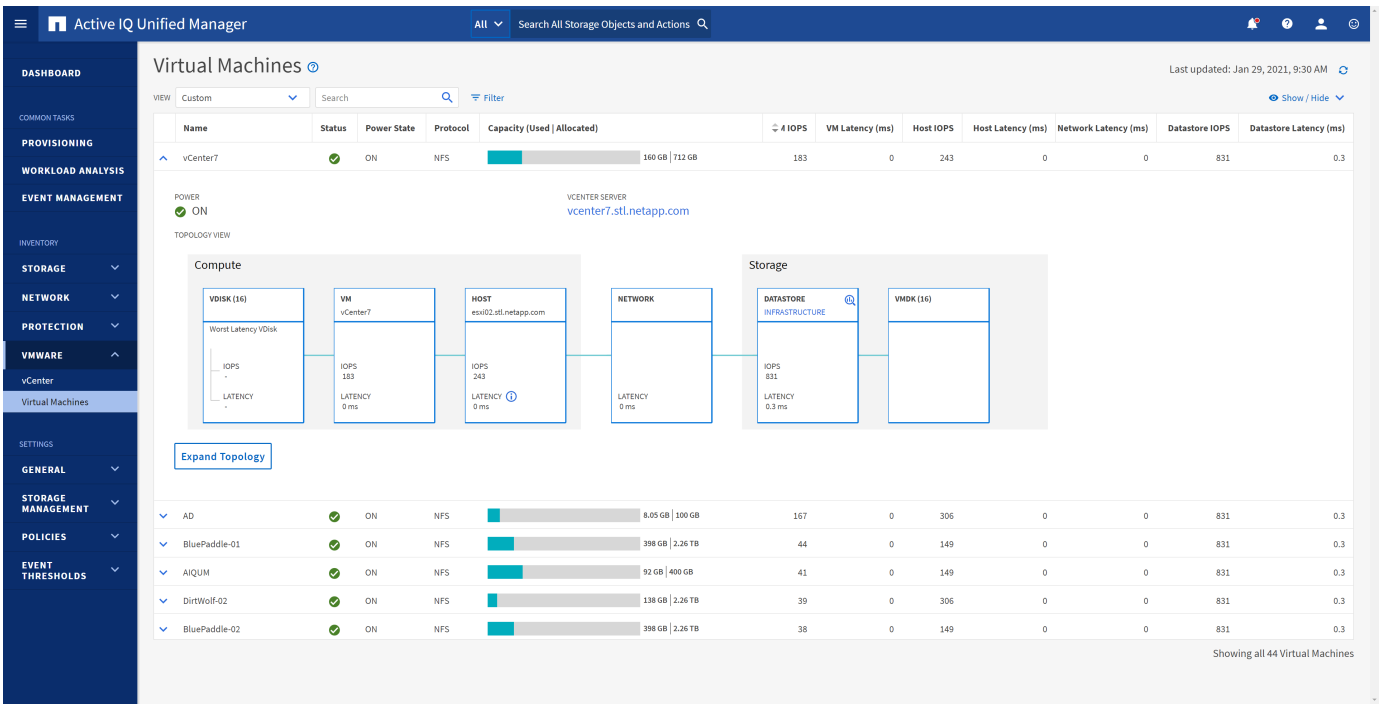
ソフトウェア暗号化を開始するのは簡単です。ライセンスのインストールが完了したら、パスフレーズを指定してオンボードキーマネージャを設定し、新しいボリュームを作成するかストレージ側のボリューム移動を実行して暗号化を有効にします。ネットアップでは、 VMware ツールの今後のリリースで、暗号化機能のサポートをさらに統合する予定です。

Active IQ Unified Manager

Active IQ Unified Manager を使用すると、仮想インフラ内の VM を可視化し、仮想環境内のストレージやパフォーマンスの問題を監視してトラブルシューティングすることができます。

ONTAP の一般的な仮想インフラ環境には、さまざまなコンポーネントがコンピューティングレイヤ、ネットワークレイヤ、ストレージレイヤに分散して配置されています。VM アプリケーションのパフォーマンス低下は、各レイヤのさまざまなコンポーネントでレイテンシが生じていることが原因である可能性があります。

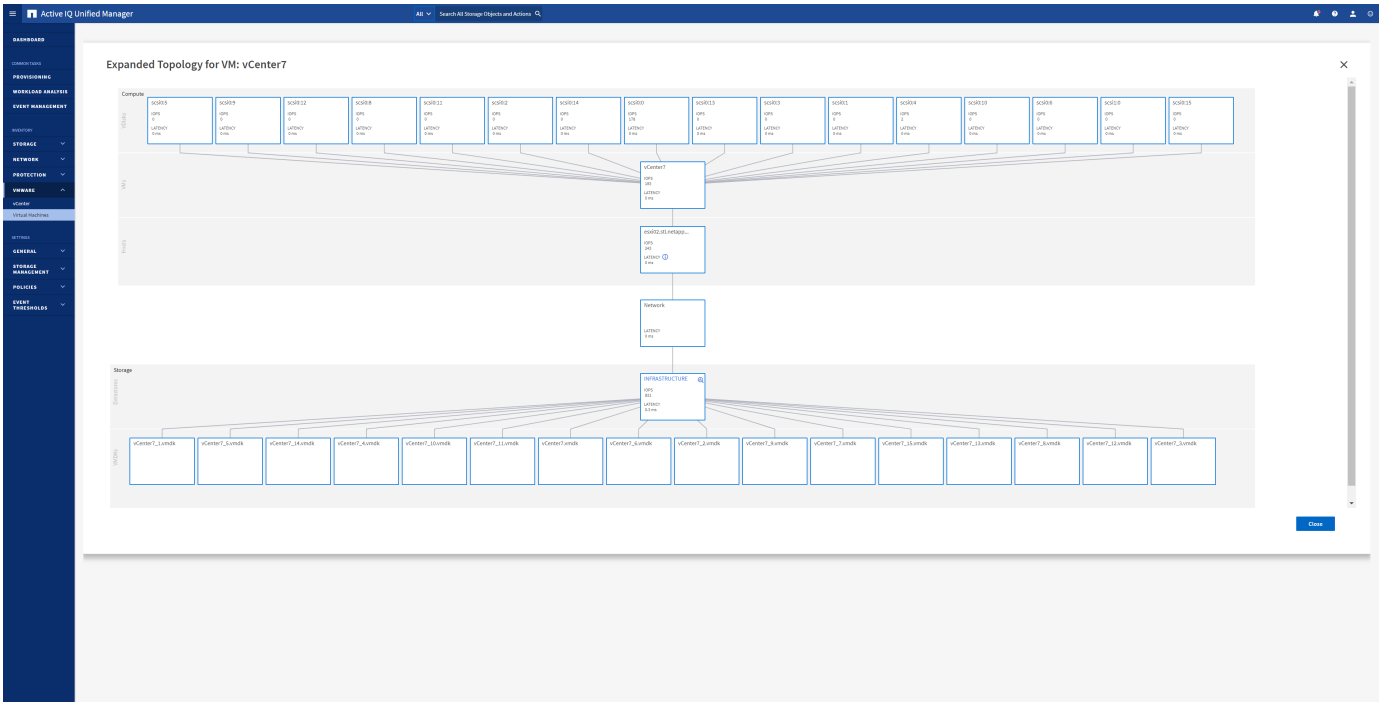
次のスクリーンショットは、 Active IQ Unified Manager の仮想マシンビューを示しています。



ビュー"]

Unified Manager のトポロジビューには、仮想環境の基盤となるサブシステムが表示され、コンピューティングノード、ネットワーク、またはストレージでレイテンシ問題が発生したかどうかが確認されます。また、修復手順を実行して基盤となる問題に対応するために、パフォーマンス低下の原因となっているオブジェクトが強調表示されます。

次のスクリーンショットは、 AIQUM の拡張トポロジを示しています。



ストレージポリシーベースの管理と**VVOL**

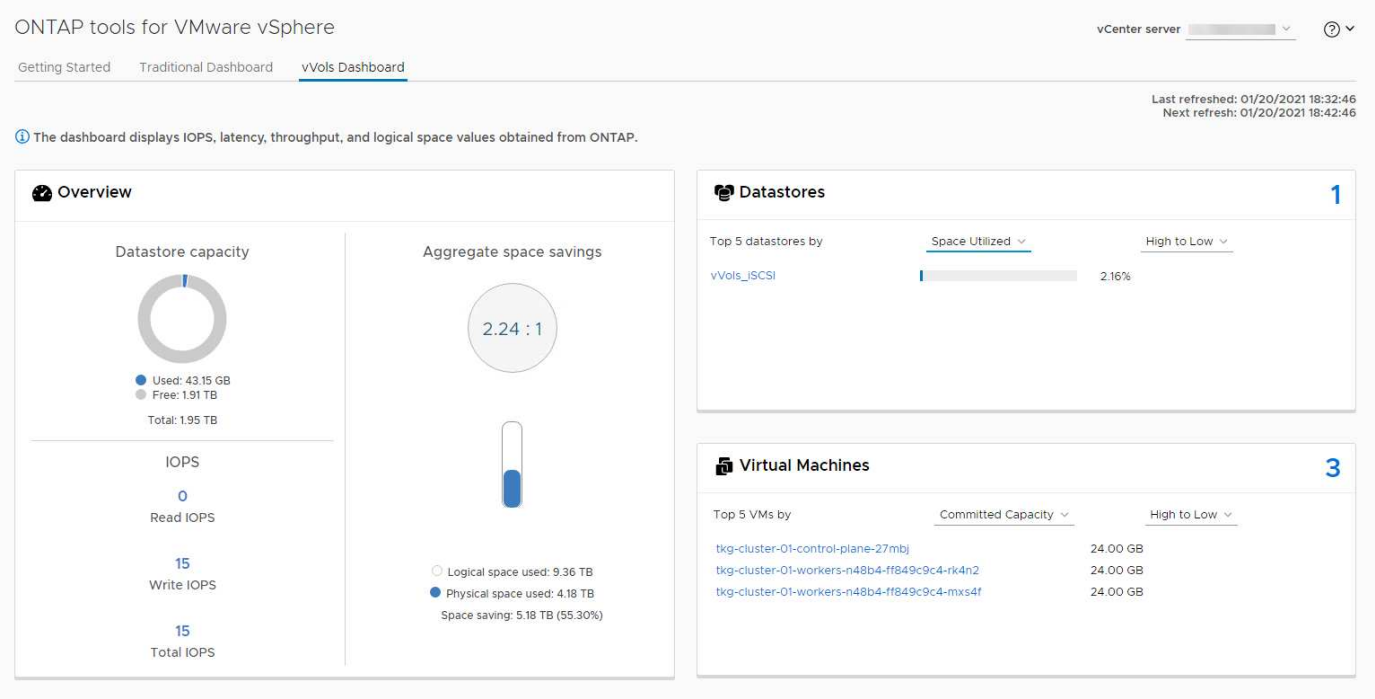
VMware vSphere APIs for Storage Awareness （ VASA ）を使用すると、ストレージ管理者は、明確に定義された機能を使用してデータストアを簡単に設定でき、 VM 管理者は、相互にやり取りすることなく、いつでも VM をプロビジョニングするためのこれらの機能を使用できます。

このアプローチを見て、仮想化ストレージの運用を合理化し、単純な作業の多くを回避する方法を確認することをお勧めします。

VASA が導入される前は、 VM 管理者が VM ストレージポリシーを定義することもできましたが、適切なデータストアを特定するには、多くの場合、ドキュメントや命名規則を使用する必要がありました。VASA を使用すると、ストレージ管理者は、パフォーマンス、階層化、暗号化、レプリケーションなど、さまざまなストレージ機能を定義できます。1 つのボリュームまたはボリュームセットの一連の機能を、ストレージ機能プロファイル（ SCP ）と呼びます。

SCPでは、VMのデータVVOLに対して最小または最大のQoSがサポートされます。最小 QoS は AFF システムでのみサポートされます。VMware vSphere 用の ONTAP ツールには、 ONTAP システム上の VVOL の VM の詳細なパフォーマンスと論理容量を表示するダッシュボードがあります。

次の図は、 VMware vSphere 9.8 VVol ダッシュボード用の ONTAP ツールを示しています。



ストレージ機能プロファイルを定義したら、そのプロファイルを使用して要件を定義するストレージポリシーを使用して VM をプロビジョニングできます。vCenter では、 VM ストレージポリシーとデータストアストレージ機能プロファイルのマッピングに基づいて、互換性があるデータストアのリストを選択対象として表示できます。このアプローチは、ストレージポリシーベースの管理と呼ばれます。

VASA は、ストレージを照会して一連のストレージ機能を vCenter に返すためのテクノロジを提供します。VASA ベンダープロバイダは、ストレージシステムの API およびコンストラクトと、 vCenter が認識可能な VMware API との間の変換機能を提供します。ネットアップのVASA Provider for ONTAPは、ONTAP Tools for VMware vSphereアプライアンスVMの一部として提供されます。vCenterプラグインは、VVOLデータストアをプロビジョニングおよび管理するためのインターフェイスと、ストレージ機能プロファイル（SCP）を定義する機能を提供します。

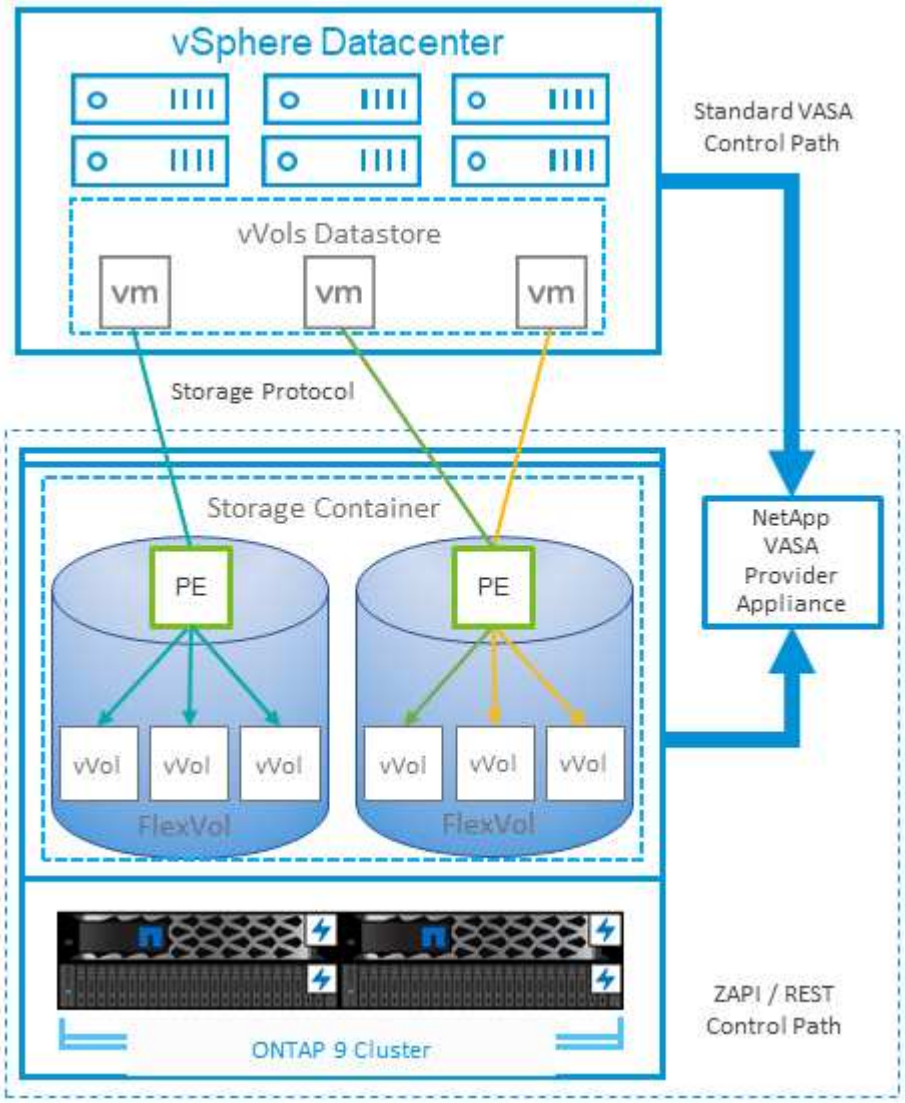
ONTAP は、 VMFS データストアと NFS データストアの両方をサポートしています。SAN データストアで VVOL を使用すると、 VM レベルのきめ細かさなど、 NFS のメリットの一部を活用できます。ここでは考慮すべきベストプラクティスをいくつか示します。また、追加情報はにあります "TR-4400"：

- VVOL データストアは、複数のクラスタノードにある複数の FlexVol で構成できます。ボリュームごとに機能が異なる場合でも、最もシンプルなアプローチは 1 つのデータストアです。SPBM により、互換性のあるボリュームが VM に使用されています。ただし、すべてのボリュームが 1 つの ONTAP SVM に含まれていて、単一のプロトコルでアクセスできる必要があります。各プロトコルでノードごとに 1 つの LIF で十分です。1 つの VVOL データストアで複数の ONTAP リリースを使用することは避けてください。リリースによってストレージ機能が異なる場合があります。
- VVol データストアの作成と管理には、 VMware vSphere プラグインの ONTAP ツールを使用します。データストアとそのプロファイルの管理に加え、必要に応じて、 VVOL にアクセスするためのプロトコルエンドポイントが自動的に作成されます。LUN を使用する場合、 LUN PE は 300 以上の LUN ID を使用してマッピングされます。ESXiホストの詳細なシステム設定を確認する `Disk.MaxLUN` 300を超えるLUN ID番号を許可します（デフォルトは1、024）。そのためには、vCenterでESXiホストを選択し、[Configure]タブで `Disk.MaxLUN` をクリックします。
- VASA Provider 、 vCenter Server （アプライアンスまたは Windows ベース）、または VMware vSphere 用の ONTAP ツールは相互に依存するため、 VVOL データストアにインストールしたり移行したりしないでください。これらのツールは、停電やその他のデータセンターの停止が発生した場合に管理しなくなる

ためです。

- VASA Provider VM を定期的にバックアップします。VASA Providerが格納された従来のデータストアのSnapshotを少なくとも1時間ごとに作成してください。VASA Provider の保護とリカバリの詳細については、こちらを参照してください "こちらの技術情報アーティクル"。

次の図は、 VVOL のコンポーネントを示しています。



VMware Storage Distributed Resource Scheduler の略

VMware Storage Distributed Resource Scheduler （ SDRS ）は、現在の I/O レイテンシとスペース使用量に基づいて VM をストレージに配置する vSphere の機能です。

その後、 VM や VMDK の配置先として最適なデータストアをデータストアクラスタ内から選択し、システムを停止することなくデータストアクラスタ（ポッドとも呼ばれます）内のデータストア間で VM や VMDK を移動します。データストアクラスタは、類似するデータストアをvSphere管理者から見た単一の消費単位に集約したものです。

SDRSとONTAP tools for VMware vSphereを使用する場合は、まずプラグインを使用してデータストアを作成し、vCenterを使用してデータストアクラスタを作成してから、そのデータストアにデータストアを追加する

必要があります。データストアクラスタを作成したら、プロビジョニングウィザードの詳細ページからデータストアクラスタにデータストアを直接追加できます。

SDRS に関するその他の ONTAP のベストプラクティスは、次のとおりです。

- クラスタ内のすべてのデータストアで同じタイプのストレージ（ SAS 、 SATA 、 SSD など）を使用し、すべて VMFS データストアまたは NFS データストアとし、レプリケーションと保護の設定を同じにします。
- デフォルト（手動）モードでは SDRS の使用を検討してください。このアプローチでは、推奨事項を確認し、適用するかどうかを決定できます。VMDK の移行による影響を次に示します。
  - SDRS がデータストア間で VMDK を移動すると、 ONTAP のクローニングや重複排除によるスペース削減効果は失われます。重複排除機能を再実行すれば、削減効果を取り戻すことができます。
  - NetAppでは、VMDKを移動したあとに、移動したVMによってスペースがロックされるため、ソースデータストアでSnapshotを再作成することを推奨しています。
  - 同じアグリゲート上のデータストア間で VMDK を移動してもメリットはほとんどなく、 SDRS はアグリゲートを共有する可能性のある他のワークロードを可視化できません。

推奨される **ESXi** ホストとその他の **ONTAP** 設定

NetAppは、NFSプロトコルとブロックプロトコルの両方に最適なESXiホスト設定を開発しました。また、NetAppとVMwareの内部テストに基づいて、ONTAPで適切に動作するようにマルチパスとHBAタイムアウトを設定するための具体的なガイダンスも提供されます。

これらの値は、ONTAP tools for VMware vSphereを使用して簡単に設定できます。[Summary]ダッシュボードで、[Host Systems]ポートレットの[Edit Settings]をクリックするか、vCenterでホストを右クリックし、ONTAP tools]>[Set Recommended Values]に移動します。

ここでは、9.8~9.13リリースで推奨されるホスト設定を示します。

ホスト設定	ネットアップが推奨する値	再起動が必要です
* ESXi Advanced Configuration *		
VMFS3.HardwareAcceleratedLocking	デフォルトのまま（1）	いいえ
VMFS3.EnableBlockDelete の 2 つのオプションがあります	デフォルト（0）のままにしますが、必要に応じて変更できます。詳細については、を参照してください "VMware KB 2007427"	いいえ
VMFS3.EnableVMFS6Unmap	デフォルトのまま（1） 詳細については、を参照してください "VMware vSphere API：アレイ統合（VAAI）"	いいえ
* NFS設定*		
Net.TcpipHeapSize の場合	vSphere 6.0 以降： 32 に設定 他のすべてのNFS設定の場合は、30に設定されます	はい。

Net.TcpipHeapMax	vSphere 6.Xのほとんどのリリースでは512 MBに設定されています。6.5U3、6.7U3、7.0以降の場合は、1024MBに設定します。	はい。
NFS.MaxVolumes の場合	vSphere 6.0以降：256に設定その他のNFS構成はすべて64に設定されます。	いいえ
NFS41.MaxVolumes	vSphere 6.0 以降では、 256 に設定されます。	いいえ
NFS.MaxQueueDepth^1 ^	vSphere 6.0以降では、128に設定されます	はい。
NFS.HeartbeatMaxFailures の略	すべてのNFS設定について、10に設定されます	いいえ
nfs.HeartbeatFrequency	すべてのNFS構成で12に設定	いいえ
nfs.HeartbeatTimeout	すべてのNFS構成で5に設定されます。	いいえ
SunRPC.MaxConnPerIP	vSphere 7.0 以降では 128 に設定されます。	いいえ
* FC / FCoE 設定 *		
パス選択ポリシー	FC パスの ALUA を使用する場合は、 RR （ラウンドロビン）に設定されます。それ以外の構成では、すべて FIXED に設定されます。 この値を RR に設定すると、最適化されたすべてのアクティブなパスで負荷を分散できます。 FIXED は、 ALUA に対応していない従来の構成用の値で、プロキシ I/O を防止できますつまり、Data ONTAP 7-Modeを実行する環境でハイアベイラビリティ（HA）ペアの他方のノードにI/Oが送られないようにすることができます	いいえ
Disk.QFullSampleSize	すべての構成で 32 に設定されます。 この値を設定すると、 I/O エラーの防止に役立ちます。	いいえ
Disk.qFullThreshold	すべての構成で 8 に設定します。 この値を設定すると、 I/O エラーの防止に役立ちます。	いいえ
Emulex FC HBA タイムアウト	デフォルト値を使用します。	いいえ
QLogic FC HBA タイムアウト	デフォルト値を使用します。	いいえ
* iSCSI 設定 *		

パス選択ポリシー	すべての iSCSI パスで RR （ラウンドロビン）に設定されます。 この値を RR に設定すると、最適化されたすべてのアクティブなパスで負荷を分散できます。	いいえ
Disk.QFullSampleSize	すべての構成で 32 に設定されます。 この値を設定すると、I/Oエラーの防止に役立ちます	いいえ
Disk.qFullThreshold	すべての構成で 8 に設定します。 この値を設定すると、 I/O エラーの防止に役立ちます。	いいえ

VMware vSphere ESXi 7.0.1およびVMware vSphere ESXi 7.0.2を使用する場合、1-NFSの高度な設定オプションMaxQueueDepthが想定どおりに機能しないことがあります。参照してください "VMware KB 86331" を参照してください。

ONTAP ツールでは、 ONTAP FlexVol および LUN の作成時に特定のデフォルト設定も指定されます。

* ONTAP ツール*	デフォルト設定
Snapshot リザーブ（ -percent-snapshot-space ）	0
フラクショナルリザーブ（ -fractional-reserve ）	0
アクセス時間の更新（ -atime-update ）	いいえ
最小限の先読み（ -min-readahead ）	いいえ
スケジュールされたSnapshot	なし
ストレージ効率	有効
ボリュームギャランティ	なし（シンプロビジョニング）
ボリュームのオートサイズ	grow_shrink
LUN のスペースリザベーション	無効
LUN スペースの割り当て	有効

ハフオマンスノマルチハスセツテイ

現在使用可能なONTAPツールでは設定されていませんが、NetAppでは次の設定オプションを推奨しています。

- ハイパフォーマンスな環境で、または単一の LUN データストアでパフォーマンスをテストする場合は、ラウンドロビン（ VMW_PSP_RR ）パス選択ポリシー（ PSP ）の負荷分散設定をデフォルトの IOPS 設定 1000 から 1 に変更することを検討します。VMware の技術情報を参照 "2069356" 詳細については、
- vSphere 6.7 Update 1 では、 VMware がラウンドロビン PSP 用に新しいレイテンシの負荷分散メカニズムを導入しました。新しいオプションでは、 I/O に最適なパスを選択する際に、 I/O 帯域幅とパスレイテンシが考慮されますパス接続が異なる環境（あるパスのネットワークホップ数が別のパスよりも多い場合など）や、NetAppオールSANアレイシステムを使用している場合など、パス接続が同等でない環境で使用するとメリットがあります。を参照してください "パス選択プラグインとポリシー" を参照してください。

その他のドキュメント

vSphere 7を使用するFCPおよびiSCSIの詳細については、を参照してください。 "VMware vSphere 7.xとONTAPの併用"
vSphere 8を使用するFCPおよびiSCSIの詳細については、を参照してください。 "VMware vSphere 8.xとONTAPの併用"
vSphere 7を使用したNVMe-oFの詳細については、を参照してください。 "NVMe-oFの詳細については、「NVMe-oFホスト構成（ESXi 7.x with ONTAP）」を参照してください。"
vSphere 8を使用したNVMe-oFの詳細については、を参照してください。 "NVMe-oFの詳細については、「NVMe-oFホスト構成（ESXi 8.x with ONTAP）」を参照してください。"

ONTAPを備えた仮想ボリューム（VVOL）

概要

ONTAPは、20年以上にわたって業界をリードするVMware vSphere環境向けストレージ解決策であり、コストを削減しながら管理を簡易化する革新的な機能を継続的に追加しています。

本ドキュメントでは、VMware vSphere Virtual Volumes（VVOL）向けのONTAP 機能について説明します。最新の製品情報やユースケース、導入を合理化してエラーを削減するためのベストプラクティスなどを紹介します。

このドキュメントは、これまでに公開されていたテクニカルレポート_TR-4400：『VMware vSphere Virtual Volumes（vVol）with ONTAP _』を差し替えます。

ベストプラクティスは、ガイドや互換性リストなどの他のドキュメントを補うものです。ラボテストに基づいて開発されており、ネットアップのエンジニアやお客様は広範な現場経験を積んでいます。効果的またはサポートされている唯一の手法ではないかもしれませんが、一般的には、ほとんどのお客様のニーズを満たす最もシンプルなソリューションです。

本ドキュメントが更新され、vSphere 8.0 Update 1に搭載された新しいvVol機能がONTAP tools 9.12リリースでサポートされるようになりました。

Virtual Volumes（VVol）の概要

ネットアップは2012年にVMwareとの連携を開始し、vSphere APIs for Storage Awareness（VASA）for vSphere 5のサポートを開始しました。この初期のVASA Providerでは、プロファイルにストレージ機能を定義することができました。このプロファイルを使用すると、プロビジョニング時やポリシーへの準拠状況の確認時にデータストアをフィルタリングできます。時間の経過とともに、プロビジョニングの自動化を可能にする新しい機能が追加されたり、仮想ボリューム（VVol）が追加されたりして、個 々 のストレージオブジェクトが仮想マシンファイルと仮想ディスクに使用されたりします。これらのオブジェクトにはLUN、ファイルなどが含まれます。vSphere 8 - NVMe namespaces.NetAppは、2015年にvSphere 6でリリースされたVVOLのリファレンスパートナーとして、またvSphere 8でNVMe over Fabricsを使用したVVOLの設計パートナーとして、VMwareと緊密に連携しています。ネットアップでは、ONTAP の最新機能を活用できるように、VVOLの機能を継続的に強化しています。

注意が必要なコンポーネントは次のとおりです。

* VASA Provider *

VMware vSphereとストレージシステムの間の通信を処理するソフトウェアコンポーネントです。ONTAP の場合、VASA ProviderはONTAP Tools for VMware vSphere（ONTAP tools for VMware vSphere）と呼ばれるアプライアンスで実行されます。ONTAP toolsには、vCenterプラグイン、VMware Site Recovery Manager用のStorage Replication Adapter（SRA）、独自の自動化を構築するためのREST APIサーバも含まれています。ONTAP toolsを設定してvCenterに登録すると、ONTAP システムを直接操作する必要はほとんどなくなります。これは、必要なストレージのほぼすべてをvCenter UIから直接、またはREST APIによる自動化を通じて管理できるためです。
プロトコルエンドポイント（**PE**）
プロトコルエンドポイントは、ESXiホストとVVOLデータストアの間のI/Oのプロキシです。ONTAP VASA Providerは、VVOLデータストアのFlexVolごとに1つのプロトコルエンドポイントLUN（サイズ4MB）、またはデータストア内のFlexVolボリュームをホストしているストレージノードのNFSインターフェイス（LIF）ごとに1つのNFSマウントポイントを自動的に作成します。ESXiホストでは、これらのプロトコルエンドポイントは、個 々 のVVOL LUNや仮想ディスクファイルではなく直接マウントされます。プロトコルエンドポイントは、必要なインターフェイスグループやエクスポートポリシーとともにVASA Providerによって自動的に作成、マウント、アンマウント、および削除されるため、管理する必要はありません。
仮想プロトコルエンドポイント（**VPE**）
vSphere 8の新機能では、VVOLでNVMe over Fabrics（NVMe-oF）を使用する場合、プロトコルエンドポイントの概念はONTAP には関係ありません。代わりに、最初のVMの電源がオンになるとすぐに、各ANAグループのESXiホストによって仮想PEが自動的にインスタンス化されます。ONTAP では、データストアで使用されるFlexVol ボリュームごとにANAグループが自動的に作成されます。 VVOLにNVMe-oFを使用するもう1つの利点は、VASA Providerでバインド要求が不要であることです。代わりに、VVOLバインド機能はVPEに基づいてESXiホストが内部的に処理します。これにより、VVolのバインドストームがサービスに影響する可能性が低くなります。 詳細については、を参照してください "NVMeと仮想ボリューム" オン "VMware.com"
仮想ボリュームデータストア
仮想ボリュームデータストアは、VASA Providerで作成および管理されるVVolコンテナを表す論理データストアです。コンテナは、VASA Providerで管理されるストレージシステムからプロビジョニングされたストレージ容量のプールを表します。ONTAP toolsでは、1つのvVolデータストアに複数のFlexVol ボリューム（バッキングボリューム）を割り当てることができます。これらのvVolデータストアは、機能の異なるフラッシュシステムとハイブリッドシステムを組み合わせることで、ONTAP クラスタ内の複数のノードにまたがることができます。管理者は、プロビジョニングウィザードまたはREST APIを使用して新しいFlexVol ボリュームを作成できます。また、作成済みのFlexVol ボリュームがある場合は、元のストレージ用に選択できます。
仮想ボリューム（**VVol**）
VVOLは、VVOLデータストアに格納される実際の仮想マシンのファイルとディスクです。VVOL（単一）という用語は、単一の特定のファイル、LUN、またはネームスペースを指します。ONTAP は、データストアが使用するプロトコルに応じて、NVMeネームスペース、LUN、またはファイルを作成します。VVOLにはいくつかの異なるタイプがあり、最も一般的なものは、Config（メタデータファイル）、Data（仮想ディスクまたはVMDK）、Swap（VMの電源投入時に作成）です。VMware VM暗号化で保護されるvVolのタイプはOTHERになります。VMware VMの暗号化とONTAP ボリュームまたはアグリゲートの暗号化を混同しないでください。

ポリシーベースの管理

VMware vSphere APIs for Storage Awareness（VASA）を使用すると、VM管理者は、ストレージチームとやり取りすることなく、VMのプロビジョニングに必要なストレージ機能を簡単に使用できます。VASAがリリー

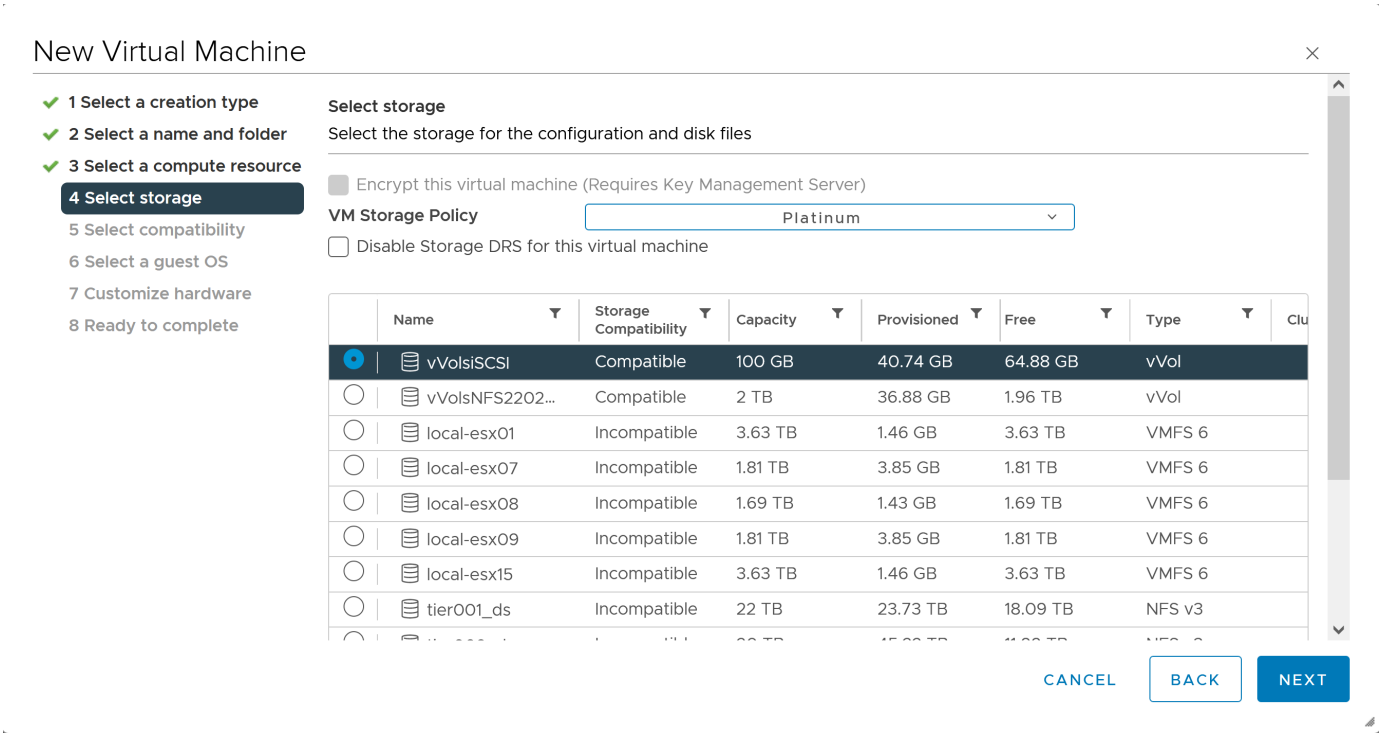
スされるまではVM管理者はVMストレージポリシーを定義できましたが、適切なデータストアを特定するためにはストレージ管理者と協力しなければなりませんでした。多くの場合、ドキュメントや命名規則を使用していました。VASAを使用すると、適切な権限を持つvCenter管理者は、vCenterユーザがVMのプロビジョニングに使用できる一連のストレージ機能を定義できます。VMストレージポリシーとデータストアストレージ機能プロファイルのマッピングにより、vCenterで互換性のあるデータストアのリストを表示して選択できるほか、ARIA（旧vRealize）AutomationやTanzu Kubernetes Gridなどの他のテクノロジを有効にして、割り当てられたポリシーからストレージを自動的に選択できます。このアプローチは、ストレージポリシーベースの管理と呼ばれます。ストレージ機能プロファイルとポリシーは従来のデータストアでも使用できますが、ここではVVOLデータストアに焦点を当てます。

次の2つの要素があります。

ストレージ機能プロファイル（**SCP**）
ストレージ機能プロファイル（SCP）は、ストレージテンプレートの形式です。これを使用すると、vCenterの管理者は、ONTAP でのそれらの機能の管理方法を実際に理解していなくても、必要なストレージ機能を定義できます。テンプレート形式のアプローチを採用することで、管理者は一貫した予測可能な方法でストレージサービスを簡単に提供できます。SCPで説明される機能には、パフォーマンス、プロトコル、Storage Efficiencyなどがあります。特定の機能はバージョンによって異なります。vCenter UIのONTAP Tools for VMware vSphereメニューを使用して作成します。REST APIを使用してSCPを作成することもできます。個々 の機能を選択して手動で作成することも、既存の（従来の）データストアから自動的に生成することもできます。
* VMストレージポリシー*
仮想マシンストレージポリシーは、vCenterの[Policies and Profiles]に作成されます。VVOLの場合は、NetApp VVOLストレージタイププロバイダから提供されるルールを使用してルールセットを作成します。ONTAP ツールを使用すると、個別のルールを強制的に指定するのではなく、SCPを選択するだけでシンプルなアプローチが可能になります。

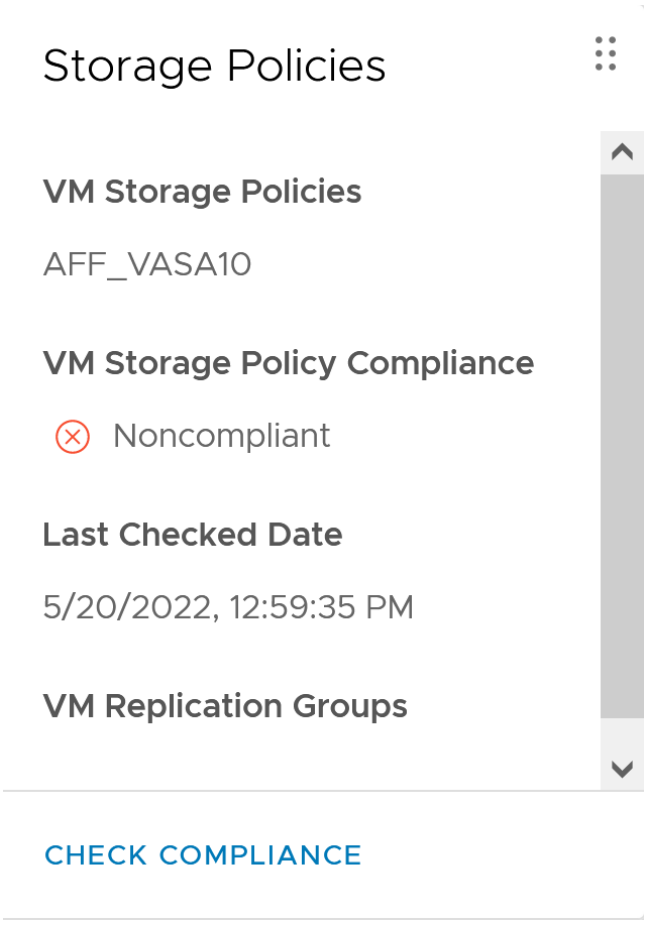
前述したように、ポリシーを使用すると、ボリュームのプロビジョニングタスクを合理化できます。適切なポリシーを選択するだけで、そのポリシーをサポートするvVolデータストアがVASA Providerに表示され、準拠している個 々 のFlexVol にvVolが配置されます（図1）。

ストレージポリシーを使用して**VM**を導入します



VMのプロビジョニングが完了すると、VASA Providerは準拠状況を継続的にチェックし、元のボリュームがポリシーに準拠しなくなったときにvCenterでアラームを生成してVM管理者に通知します（図2）。

VMストレージポリシーへの準拠



NetApp VVOLのサポート

ONTAPは、2012年の最初のリリースからVASA仕様をサポートしています。他のネットアップストレージシステムがVASAをサポートしている場合もありますが、本ドキュメントでは、現在サポートされているONTAP 9のリリースを中心に説明します。

ONTAP

NetAppは、AFF、ASA、FASシステムでのONTAP 9に加えて、ONTAP SelectでのVMwareワークロード、VMware Cloud on AWSでのAmazon FSx for NetApp、Azure VMware解決策でのNetApp、Google Cloud VMware EngineでのCloud Volumes Service、EquinixでのAzure NetApp Filesプライベートストレージをサポートしています。 ただし、特定の機能は、サービスプロバイダーおよび使用可能なネットワーク接続によって異なる場合があります。vSphereゲストから、これらの構成に格納されたデータやCloud Volumes ONTAPにアクセスすることもできます。

本書の発行時点では、ハイパースケーラ環境は従来のNFS v3データストアに限定されているため、VVOLは、オンプレミスのONTAP システム、または世界中のネットアップパートナーやサービスプロバイダがホストするオンプレミスシステムのすべての機能を提供するクラウド接続システムでのみ使用できます。

ONTAP の詳細については、を参照してください "ONTAP 製品ドキュメント"_

ONTAP およびVMware vSphereのベストプラクティスの詳細については、を参照してください "TR-4597"_

ONTAPでVVOLを使用するメリット

2015年にVMwareがVASA 2.0でVVOLをサポートするようになったとき、VMwareは「外付けストレージ（SAN / NAS）の新しい運用モデルを提供する統合管理フレームワーク」と表現しました。 この運用モデルには、ONTAP ストレージと組み合わせるメリットがいくつかあります。

ポリシーベースの管理

セクション1.2で説明したように、ポリシーベースの管理では、事前定義されたポリシーを使用してVMをプロビジョニングし、その後管理できます。これは、次のようなさまざまな方法でITの運用に役立ちます。

- 高速化。 ONTAP ツールにより、vCenter管理者がストレージプロビジョニング作業のためにストレージチームとチケットをオープンする必要がなくなります。ただし、vCenterとONTAP システムのONTAP tools RBACロールでは、必要に応じて特定の機能へのアクセスを制限することで、独立したチーム（ストレージチームなど）や同じチームによる独立したアクティビティを許可できます。
- *よりスマートなプロビジョニング。*ストレージシステムの機能をVASA APIを通じて公開できるため、VM管理者がストレージシステムの管理方法を理解しなくても、プロビジョニングワークフローで高度な機能を活用できます。
- プロビジョニングの高速化。 1つのデータストアでさまざまなストレージ機能をサポートし、VMポリシーに基づいてVMに応じて自動的に選択できます。
- *間違いを避けてください。*ストレージとVMのポリシーは事前に開発され、必要に応じて適用されます。VMをプロビジョニングするたびにストレージをカスタマイズする必要はありません。コンプライアンスアラームは、定義されたポリシーからストレージ機能が逸脱すると生成されます。前述したように、SCPは初期プロビジョニングを予測可能かつ反復可能にし、SCPに基づいてVMストレージポリシーを設定することで正確な配置を保証します。
- 容量管理の向上。 VASAおよびONTAP ツールを使用すると、必要に応じてストレージ容量を業界単位のアグリゲートレベルまで表示し、容量が不足し始めた場合に複数のレイヤからアラートを受け取ることができます。

最新の**SAN**で**VM**をきめ細かく管理

VMwareでは、ファイバチャネルとiSCSIを使用するSANストレージシステムが最初にESX向けにサポートされましたが、ストレージシステムから個 々 のVMファイルとディスクを管理する機能はありませんでした。代わりに、LUNがプロビジョニングされ、VMFSが個 々 のファイルを管理します。そのため、個 々 のVMストレージのパフォーマンス、クローニング、保護をストレージシステムで直接管理することは困難です。VVOLは、ONTAP の堅牢でパフォーマンスに優れたSAN機能により、NFSストレージを使用しているお客様がすでに利用しているストレージをきめ細かく制御します。

現在、vSphere 8とONTAP Tools for VMware vSphere 9.12以降では、従来のSCSIベースのプロトコルにVVOLで使用されていたきめ細かな制御機能が、NVMe over Fabricsを使用した最新のファイバチャネルSANで利用できるようになり、大規模環境でのパフォーマンスをさらに向上させることができます。vSphere 8.0 Update 1では、ハイパーバイザーストレージスタックでI/O変換を行うことなく、VVOLを使用して完全なエンドツーエンドのNVMe解決策 を導入できるようになりました。

優れたストレージオフロード機能

VAAIにはさまざまな処理がストレージにオフロードされますが、VASA Providerで対処できるギャップがいくつかあります。SAN VAAIでは、VMwareが管理するスナップショットをストレージシステムにオフロードできません。NFS VAAIはVM管理スナップショットをオフロードできますが、ストレージネイティブスナップショットを持つVMには制限事項があります。VVOLでは、個 々 のLUN、ネームスペース、または仮想マシンディスク用のファイルが使用されるため、ONTAP ではファイルやLUNのクローンを迅速かつ効率的に作成し、差分ファイルが不要になったVM単位のSnapshotを作成できます。NFS VAAIは、Storage vMotionのホット（電源をオンにした）移行用のクローン処理のオフロードもサポートしていません。従来のNFSデータストアでVAAIを使用する場合は、VMの電源をオフにして移行のオフロードを可能にする必要があります。ONTAPツールのVASA Providerを使用すると、ストレージ効率に優れたクローンをほぼ瞬時にホットデータとコールドデータの移行に使用できます。また、ほぼ瞬時にコピーを作成してVVOLのボリュームをまたがって移行することもできます。Storage Efficiencyにはこれらの大きなメリットがあるため、でVVOLワークロードを最大限に活用できる場合があります "容量削減保証" プログラム。同様に、VAAIを使用したボリューム間クローンで要件を満たせない場合は、VVOLでのコピー操作の向上により、ビジネス上の課題を解決できる可能性があります。

VVOLの一般的なユースケース

これらのメリットに加えて、VVOLストレージの一般的なユースケースを次に示します。

- 仮想マシンのオンデマンドプロビジョニング
  - プライベートクラウドまたはサービスプロバイダのIaaS：
  - ARIA（旧称vRealize）スイートやOpenStackなどによる自動化とオーケストレーションを活用できます
- ファーストクラスディスク（**FCD**）
  - VMware Tanzu Kubernetes Grid [TKG]の永続ボリューム。
  - 独立したVMDKライフサイクル管理を通じてAmazon EBSに似たサービスを提供
- 一時**VM**のオンデマンドプロビジョニング
  - テスト/開発ラボ
  - トレーニング環境

VVOLを最大限に活用すると（上記のユースケースなど）、具体的に次のような機能強化が実現します。

- クローンは、1つのボリューム内またはONTAP クラスタ内の複数のボリューム間ですばやく作成されます。これは、VAAIが有効な従来のクローンと比較して有利です。また、ストレージ効率にも優れています。ボリューム内のクローンには、ONTAPファイルクローンが使用されます。FlexCloneボリュームと同様に、ソースのVVOLファイル/ LUN /ネームスペースからの変更のみが格納されます。本番環境やその他のアプリケーションを目的とした長期的なVMを短時間で作成し、最小限のスペースでVMレベルの保護（VMware vSphere向けNetApp SnapCenter プラグイン、VMware管理スナップショットまたはVADPバックアップを使用）とパフォーマンス管理（ONTAP QoSを使用）を実現できます。
- VVOLは、vSphere CSIでTKGを使用する場合に理想的なストレージテクノロジであり、vCenter管理者が管理する個別のストレージクラスと容量を提供します。
- Amazon EBSに似たサービスは、FCDを介して提供できます。FCD VMDKは、その名前が示すように、vSphereのファーストクラスの市民であり、ライフサイクルが割り当てられているVMとは別に個別に管理できるためです。

ONTAP でVVOLを使用する

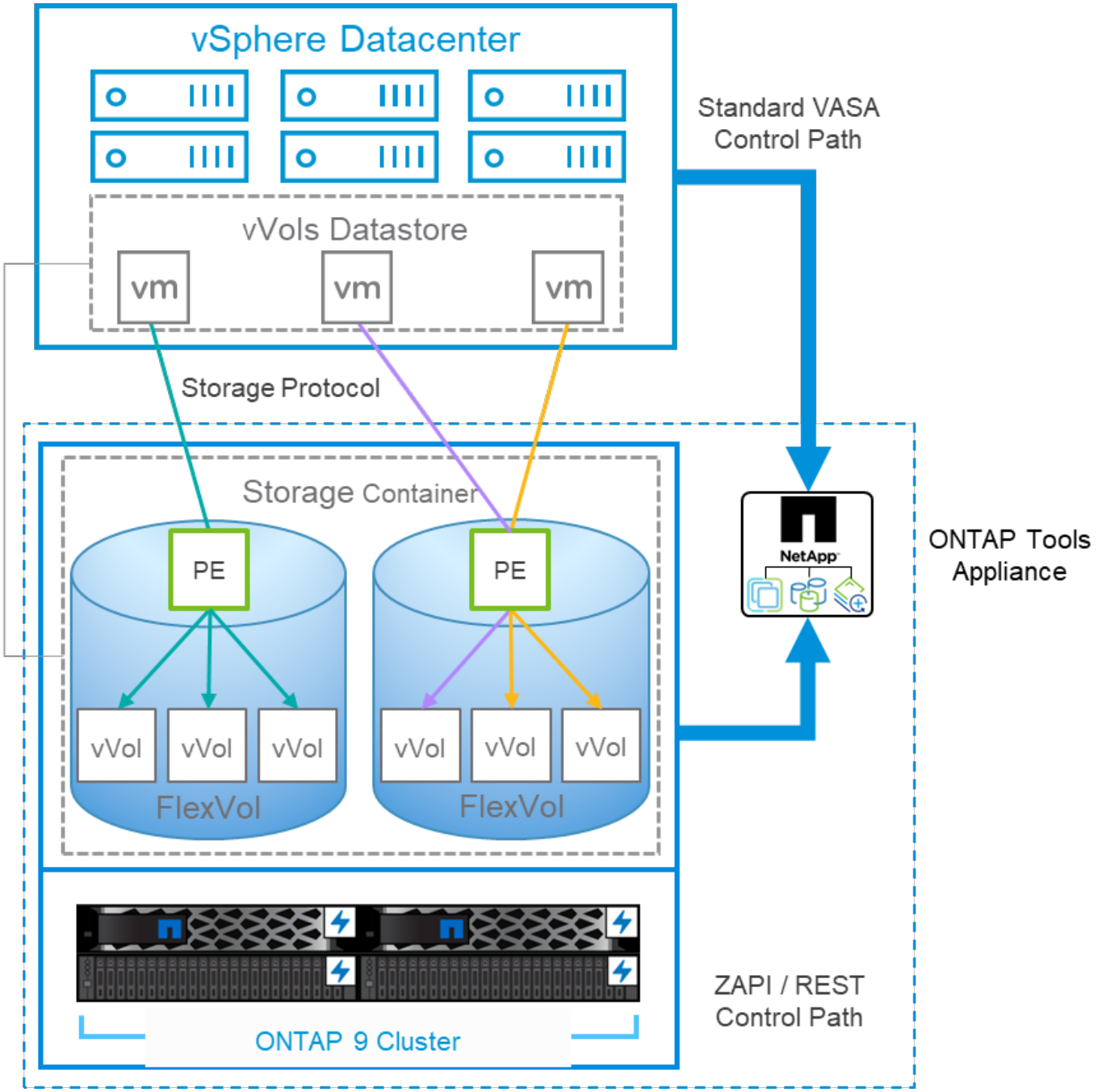
VVOLをONTAP で使用するための鍵は、ONTAP Tools for VMware vSphere仮想アプライアンスに含まれているVASA Providerソフトウェアです。

ONTAP ツールには、vCenter UI拡張機能、REST APIサーバ、Storage Replication Adapter for VMware Site Recovery Manager、Monitoring and Host構成ツール、VMware環境の管理に役立つ一連のレポートも含まれています。

製品およびドキュメント

ONTAPでVVOLを使用するために必要な追加製品は、ONTAP Oneに付属のONTAP FlexCloneライセンスとONTAP toolsアプライアンスだけです。最近リリースされたONTAP toolsは、ESXi上で動作する単一の統合アプライアンスとして提供され、これまで3種類のアプライアンスとサーバの機能を提供します。VVOLの場合、vSphereのONTAP 機能の一般的な管理ツールおよびユーザインターフェイスとして、ONTAP toolsのvCenter UI拡張機能またはREST APIを、特定のVVOL機能を提供するVASA Providerとともに使用することが重要です。SRAコンポーネントは従来のデータストアに含まれていますが、VMware Site Recovery ManagerはvVolにSRAを使用せず、代わりにvVolレプリケーションにVASAプロバイダを利用する新しいサービスをSRM 8.3以降に実装します。

iSCSIまたは**FCP**を使用する場合の**ONTAP tools VASA Provider**のアーキテクチャ



製品のインストール

新規インストールの場合は、仮想アプライアンスをvSphere環境に導入します。現在のリリースのONTAP toolsは自動的にvCenterに登録され、VASA Providerがデフォルトで有効になります。ESXiホストとvCenter Serverの情報に加えて、アプライアンスのIPアドレス設定の詳細も必要です。前述したように、VVOLに使用するすべてのONTAP クラスタには、ONTAP FlexCloneライセンスがあらかじめインストールされている必要があります。アプライアンスには可用性を確保するためのwatchdogが組み込まれています。ベストプラクティスとして、VMwareの高可用性機能とオプションのフォールトトレランス機能を使用して設定する必要があります。詳細については、セクション4.1を参照してください。ONTAP toolsアプライアンスまたはvCenter Serverアプライアンス（vCSA）をvVolストレージにインストールしたり移動したりしないでください。アプライアンスが再起動しない可能性があります。

ONTAP ツールのインプレースアップグレードは、NetApp Support Site （NSS）からダウンロードできるアップグレードISOファイルを使用してサポートされます。導入およびセットアップガイドの手順に従って、アプ

ライアンスをアップグレードします。

仮想アプライアンスのサイジングと構成の制限については、次のナレッジベースの記事を参照してください。
"『Sizing Guide for ONTAP tools for VMware vSphere』を参照してください"

製品ドキュメント

ONTAP ツールの導入に役立つ次のドキュメントを参照してください。

"完全なドキュメントリポジトリについては、次のリンクを参照してください。docs.netapp.com"

はじめに

- "リリースノート"
- "ONTAP Tools for VMware vSphereについて説明します"
- "ONTAP ツールクイックスタート"
- "ONTAP ツールを導入"
- "ONTAP ツールをアップグレードする"

ONTAP ツールを使用する

- "従来のデータストアをプロビジョニングする"
- "vVol データストアをプロビジョニングする"
- "ロールベースアクセス制御を設定する"
- "リモート診断を設定します"
- "ハイアベイラビリティを設定する"

データストアの保護と管理

- "従来のデータストアを保護" SRMを使用
- "VVOLベースの仮想マシンを保護" SRMを使用
- "従来のデータストアと仮想マシンを監視する"
- "VVol データストアと仮想マシンを監視する"

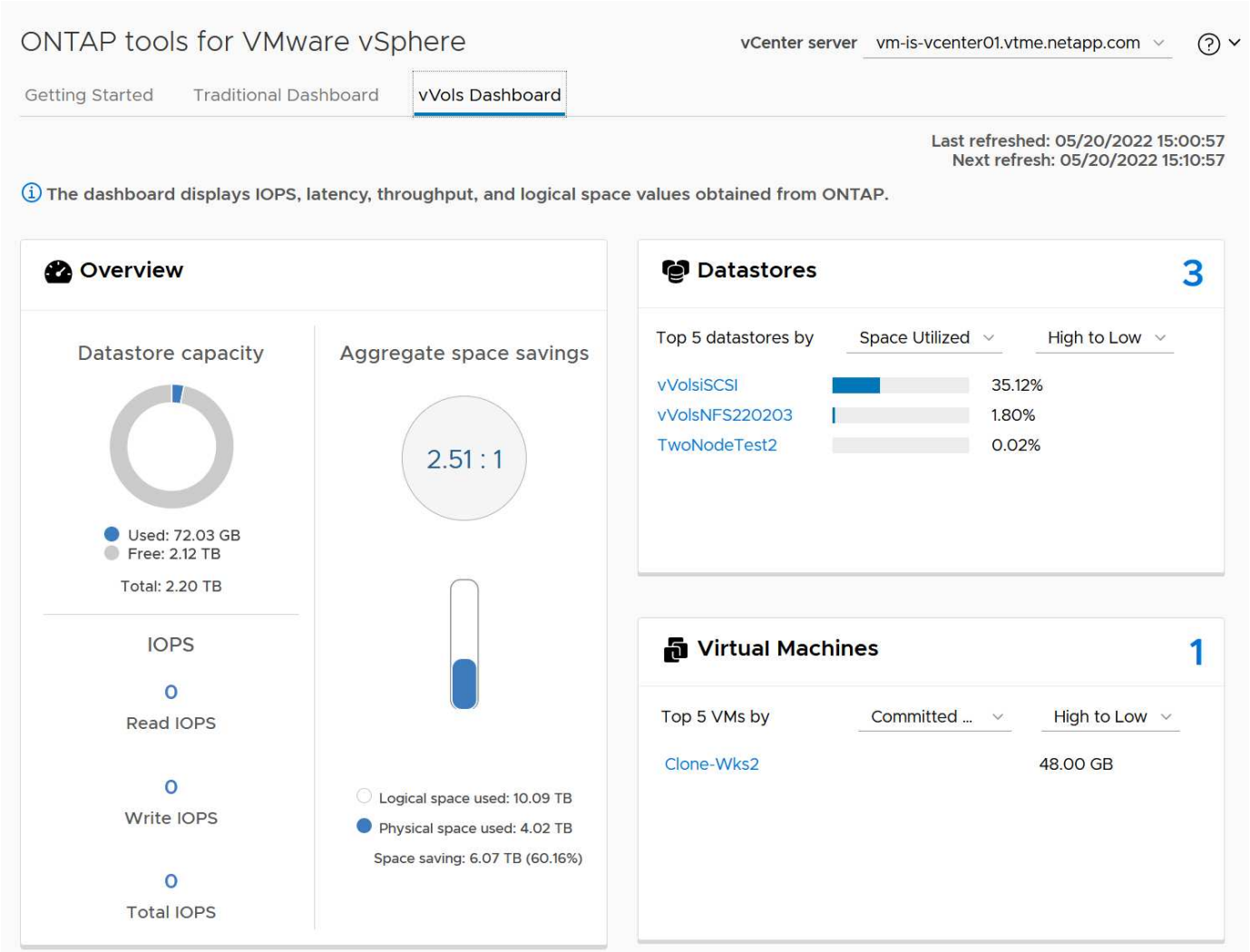
製品ドキュメント以外にも、役立つサポート技術情報アーティクルがあります。

- "『How to perform a VASA Provider Disaster Recovery - Resolution Guide』"

VASA Providerダッシュボード

VASA Providerには、個 々 のvVol VMのパフォーマンスと容量の情報が表示されたダッシュボードがあります。この情報は、VVOLファイルおよびLUNのONTAP から直接取得されます。上位5つのVMのレイテンシ、IOPS、スループット、アップタイム、上位5つのデータストアのレイテンシとIOPSなどが含まれます。ONTAP 9.7以降を使用している場合はデフォルトで有効になります。初期データが取得されてダッシュボードに表示されるまで、最大で30分かかることがあります。

ONTAP tools vVolダッシュボード



ベストプラクティス

vSphereでONTAP vVolを使用するのは簡単で、公開されているvSphereのメソッドに従います（使用しているバージョンのESXiに対応するVMwareのドキュメントの「vSphere Storage」の「Working with Virtual Volumes」を参照してください）。ここでは、ONTAP と併せて考慮すべき追加のプラクティスをいくつか紹介します。

制限

一般に、ONTAP でサポートされるVVOLの制限は、VMwareで定義されています（公開されているを参照） "構成の最大値"）。次の表は、ONTAP 固有のVVOLのサイズと数の制限をまとめたものです。必ずをチェックしてください "NetApp Hardware Universe の略" LUNとファイルの数とサイズの制限を更新

- ONTAP vVolの制限*

容量 / 機能	**SAN（SCSIまたはNVMe-oF）**	**NFS**
vVolの最大サイズ	62TiB *	62TiB *
FlexVol あたりの最大VVol数	一、〇 二四	20億です

容量 / 機能	**SAN（SCSIまたはNVMe-oF）**	**NFS**
ONTAP ノードあたりの最大VVol数	最大12,288 **	500億です
ONTAP ペアあたりの最大VVol数	最大24,576 **	500億です
ONTAP クラスタあたりの最大VVol数	最大98,304 **	特定のクラスタ制限はありません
最大QoSオブジェクト（共有ポリシーグループと個 々 のvVolサービスレベル）	ONTAP 9.3では12、000、ONTAP 9.4以降では40、000	

- サイズ制限はASA システム、またはONTAP 9.12.1P2以降を実行するAFF およびFAS システムによって異なります。
  - SAN vVol（NVMeネームスペースまたはLUN）の数はプラットフォームによって異なります。必ずをチェックしてください "NetApp Hardware Universe の略" LUNとファイルの数とサイズの制限を更新
- ONTAP ツールfor VMware vSphereのUI拡張機能またはREST APIを使用して、VVOLデータストア*およびプロトコルエンドポイントをプロビジョニングします。*

VVOLデータストアは一般的なvSphereインターフェイスを使用して作成することもできますが、ONTAPツールを使用すると、必要に応じてプロトコルエンドポイントが自動的に作成されます。また、ONTAPのベストプラクティスに従って、定義されたストレージ機能プロファイルに準拠したFlexVolボリュームが作成されます。ホスト/クラスタ/データセンターを右クリックし、ONTAP tools_and_Provision datastores_を選択します。ウィザードで目的のvVolオプションを選択するだけです。

- ONTAP Toolsアプライアンスまたはv Center Server Appliance（vCSA）は、管理対象のVVOLデータストアには絶対に保存しないでください。*

その結果、アプライアンスのリブートが必要になった場合、リブート中に自身のVVOLを再バインドできないため、アプライアンスのリブートが必要になることがあります。これらのデータは、別のONTAP ツールとvCenter環境で管理されるvVolデータストアに格納できます。

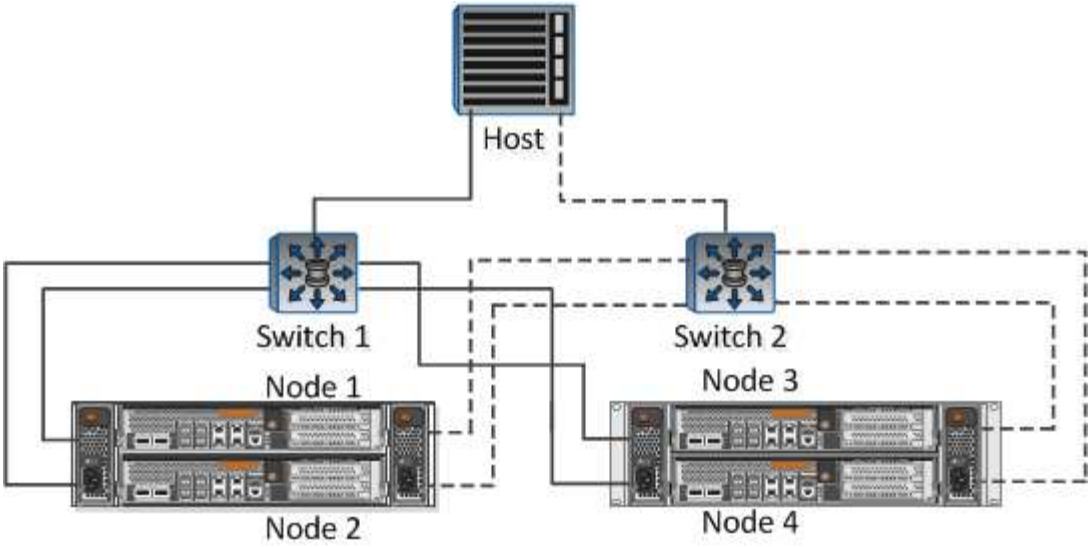
異なる**ONTAP** リリース間での**VVOL**処理は避けてください。

サポートされるストレージ機能（QoS、パーソナリティなど）はVASA Providerのリリースによって変更され、一部はONTAP リリースに依存します。ONTAP クラスタで異なるリリースを使用したり、リリースの異なるクラスタ間でVVolを移動したりすると、予期しない動作やコンプライアンスアラームが発生する可能性があります。

- VVOLにNVMe/FCまたはFCPを使用する前に、ファイバチャネルファブリックのゾーニングを設定してください。*

ONTAP tools VASAプロバイダは、管理対象のESXiホストで検出されたイニシエータに基づいて、FCPおよびiSCSI igroup、およびONTAP 内のNVMeサブシステムを管理します。ただし、ゾーニングを管理するためにファイバチャネルスイッチと統合することはできません。プロビジョニングを実行する前に、ベストプラクティスに従ってゾーニングを実行する必要があります。次に、4つのONTAPシステムに対する単一イニシエータゾーニングの例を示します。

単一イニシエータのゾーニング：



ベストプラクティスの詳細については、次のドキュメントを参照してください。

"_TR-4080『Best Practices for Modern SAN ONTAP 9』を参照してください"

"_TR-4684『Implementing and Configuring Modern SANs with NVMe-oF』を参照してください"

あなたの必要性に応じてあなたのバッキング**FlexVol**を計画しなさい。

VVOLデータストアに元のボリュームをいくつか追加して、ONTAP クラスタ全体にワークロードを分散したり、さまざまなポリシーオプションをサポートしたり、許可するLUNやファイルの数を増やしたりすることができます。ただし、最大限のストレージ効率が必要な場合は、すべてのバッキングボリュームを1つのアグリゲートに配置してください。また、クローニングのパフォーマンスを最大限に高める必要がある場合は、単一のFlexVol ボリュームを使用し、テンプレートまたはコンテンツライブラリを同じボリューム内に維持することを検討してください。VASA Providerは、移行、クローニング、Snapshotなど、多くのVVOLストレージ処理をONTAP にオフロードします。単一のFlexVol ボリューム内で実行すると、スペース効率に優れたファイルクローンが使用され、ほぼ瞬時に使用できます。この処理をFlexVol ボリューム間で実行すると、コピーをすぐに使用でき、インラインの重複排除と圧縮が使用されます。ただし、バックグラウンドの重複排除と圧縮を使用するボリュームでバックグラウンドジョブが実行されるまで、最大限のストレージ効率が回復されることはありません。ソースとデスティネーションによっては、一部の効率が低下する場合があります。

ストレージ機能プロファイル（**SCP**）はシンプルに。

必要のない機能は、anyに設定して指定しないでください。これにより、FlexVol ボリュームを選択または作成する際の問題を最小限に抑えることができます。たとえば、VASA Provider 7.1以前では、圧縮がデフォルトのSCP設定の[いいえ]のままになっていると、AFF システムであっても圧縮を無効にしようとします。

デフォルトの**SCP**をサンプルテンプレートとして使用して、独自の**SCP**を作成します。

付属のSCPはほとんどの汎用用途に適していますが、要件が異なる場合があります。

最大**IOPS**を使用して不明な**VM**やテスト**VM**を制御することを検討してください。

最大IOPSを使用すると、不明なワークロードのIOPSを特定のVVolに制限して、他の重要度の高いワークロードへの影響を回避できます。パフォーマンス管理の詳細については、表4を参照してください。

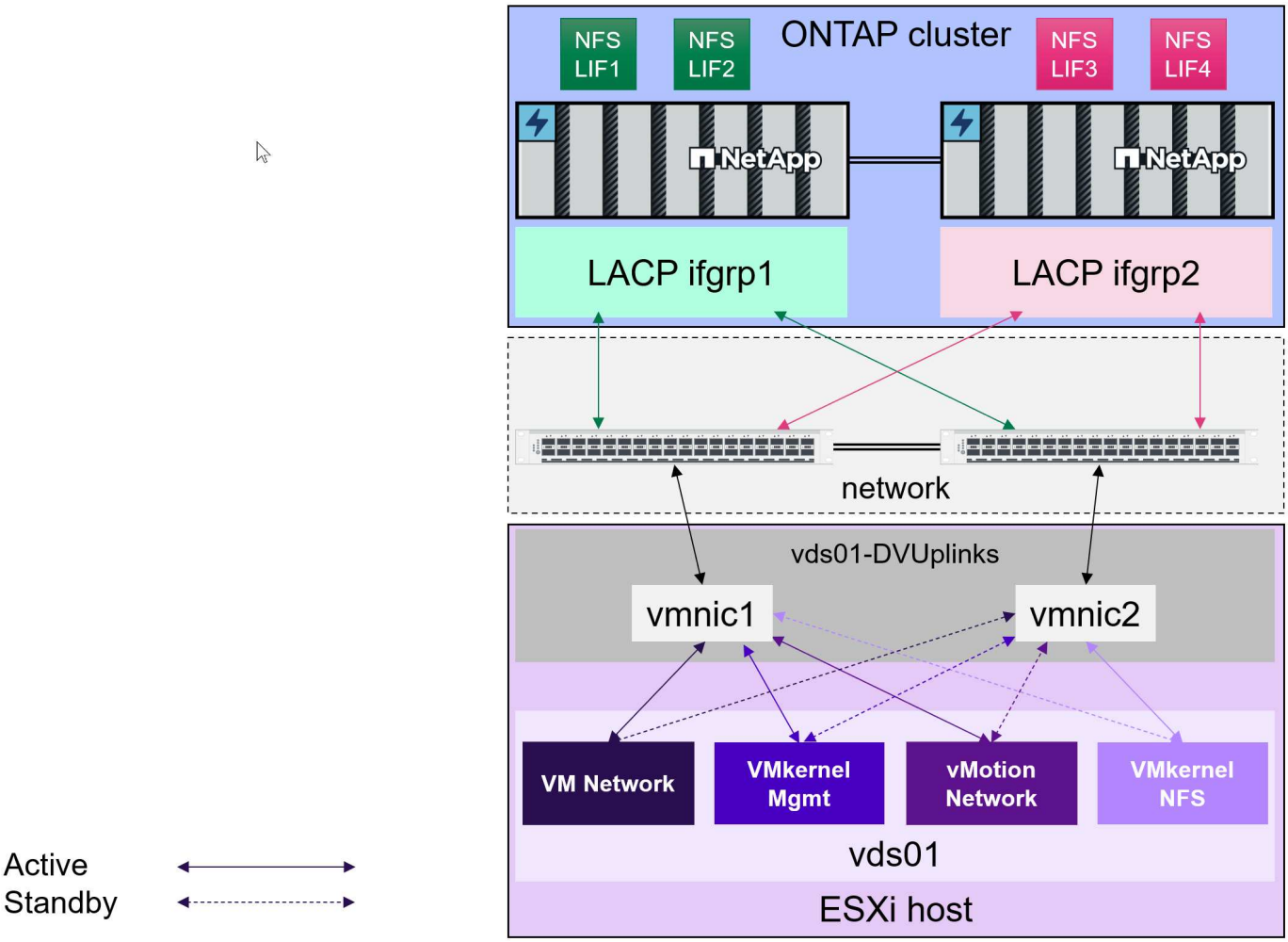
十分な数のデータ**LIF**があることを確認してください。
各HAペアのノードごとに少なくとも2つのLIFを作成します。ワークロードに応じて、さらに多くの処理が必

要になる場合があります。

すべてのプロトコルのベストプラクティスに従ってください。

選択したプロトコルに固有のNetAppおよびVMwareのその他のベストプラクティスガイドを参照してください。一般的に、上記以外の変更はありません。

- NFS v3経由でVVOLを使用したネットワーク構成の例*



vVolストレージの導入

VM用のVVOLストレージを作成するには、いくつかの手順を実行します。

従来のデータストアにONTAP を使用する既存のvSphere環境では、最初の2つの手順は必要ない場合があります。VMFSまたは従来のNFSベースのストレージの管理、自動化、レポート作成に、すでにONTAPツールを使用している場合があります。これらの手順については、次のセクションで詳しく説明します。

1. Storage Virtual Machine（SVM）とそのプロトコル設定を作成します。[NVMe/FC]、[NFSv3]、[NFSv4.1]、[iSCSI]、[FCP]、 またはそれらのオプションの組み合わせ。ONTAPのSystem Managerウィザードまたはクラスタシェルコマンドラインを使用できます。
   - スイッチ/ファブリック接続ごとにノードごとに少なくとも1つのLIFが必要です。FCP、iSCSI、またはNVMeベースのプロトコルを使用する場合は、ノードごとに2つ以上を作成することを推奨します。

- この時点でボリュームを作成することもできますが、_Provision Datastore_wizardで作成する方が簡単です。ただし、VMware Site Recovery ManagerでvVolレプリケーションを使用する場合は例外です。この方法を使用すると、既存のSnapMirror関係が設定された既存のFlexVol を使用した方が簡単です。QoSはSPBMとONTAP ツールで管理するため、VVOLに使用するボリュームでは有効にしないでください。

2. NetApp Support Site からダウンロードしたOVAを使用して、ONTAP Tools for VMware vSphereを導入します。
3. 環境に合わせてONTAP toolsを設定します。
   - ONTAP toolsの_Storage Systems_にONTAP クラスタを追加します
     - ONTAP toolsとSRAはクラスタレベルとSVMレベルの両方のクレデンシャルをサポートしますが、VASA Providerではストレージシステムのクラスタレベルのクレデンシャルのみがサポートされます。これは、VVOLに使用されるAPIの多くがクラスタレベルでしか使用できないためです。そのため、VVOLを使用する場合は、クラスタを対象としたクレデンシャルを使用してONTAPクラスタを追加する必要があります。
   - ONTAP データLIFがVMkernelアダプタとは異なるサブネットにある場合は、ONTAP toolsの設定メニューで、[Selected Subnets]リストにVMkernelアダプタのサブネットを追加する必要があります。デフォルトでは、ONTAP toolsはローカルサブネットへのアクセスのみを許可することでストレージトラフィックを保護します。
   - ONTAPツールには、事前定義されたポリシーがいくつか用意されています。これらのポリシーは、 ポリシーによるVMの管理 を参照してください。
4. vCenterの_Provision ONTAP tools_menuを使用して、_Provision datastore_wizardを起動します。
5. わかりやすい名前を指定し、目的のプロトコルを選択します。データストアの概要 も指定できます。
6. vVolデータストアでサポートするSCPを1つ以上選択します。これにより、プロファイルに一致しないONTAP システムがすべて除外されます。表示されたリストから、目的のクラスタとSVMを選択します。
7. ウィザードを使用して、指定したSCPごとに新しいFlexVol ボリュームを作成するか、適切なラジオボタンを選択して既存のボリュームを使用します。
8. vCenter UIの_PoliciesとProfiles_menuから、データストアで使用する各SCPのVMポリシーを作成します。
9. 「NetApp.clustered.Data.ONTAP.VP.vvol」ストレージルールセットを選択します。「NetApp.clustered.Data.ONTAP.VP.VASA10」ストレージルールセットは、vVol以外のデータストアでのSPBMサポート用です
10. ストレージ機能プロファイルは、VMストレージポリシーを作成するときに名前で指定します。この手順では、[replication]タブを使用してSnapMirrorポリシーの照合を設定し、[Tags]タブを使用してタグベースの照合を設定することもできます。選択できるようにするには、タグがすでに作成されている必要があります。
11. [Select storage]でVMストレージポリシーと互換性があるデータストアを選択して、VMを作成します。

従来のデータストアから**VVOL**への**VM**の移行

従来のデータストアからvVolデータストアへのVMの移行は、従来のデータストア間でVMを移動するだけです。VMを選択し、[Actions]リストから[Migrate]を選択し、移行タイプとして[*change storage only*]を選択します。移行コピー処理はvSphere 6.0以降ではSAN VMFSからVVOLへの移行でオフロードされますが、NAS VMDKからVVOLへの移行ではオフロードされません。

ポリシーによるVMの管理

ポリシーベースの管理でストレージプロビジョニングを自動化するには、次のことが必要です。

- ストレージ機能プロファイル（SCP）を使用して、ストレージ（ONTAP ノードとFlexVol ボリューム）の機能を定義します。
- 定義済みのSCPに対応するVMストレージポリシーを作成します。

VASA Provider 7.2以降では、機能とマッピングが簡易化され、以降のバージョンで継続的に改善されています。このセクションでは、この新しいアプローチに焦点を当てます。以前のリリースではサポートされていた機能の数が増え、個 々 にストレージポリシーにマッピングすることができましたが、このアプローチはサポートされなくなりました。

ストレージ機能プロファイルONTAP toolsリリース別の機能

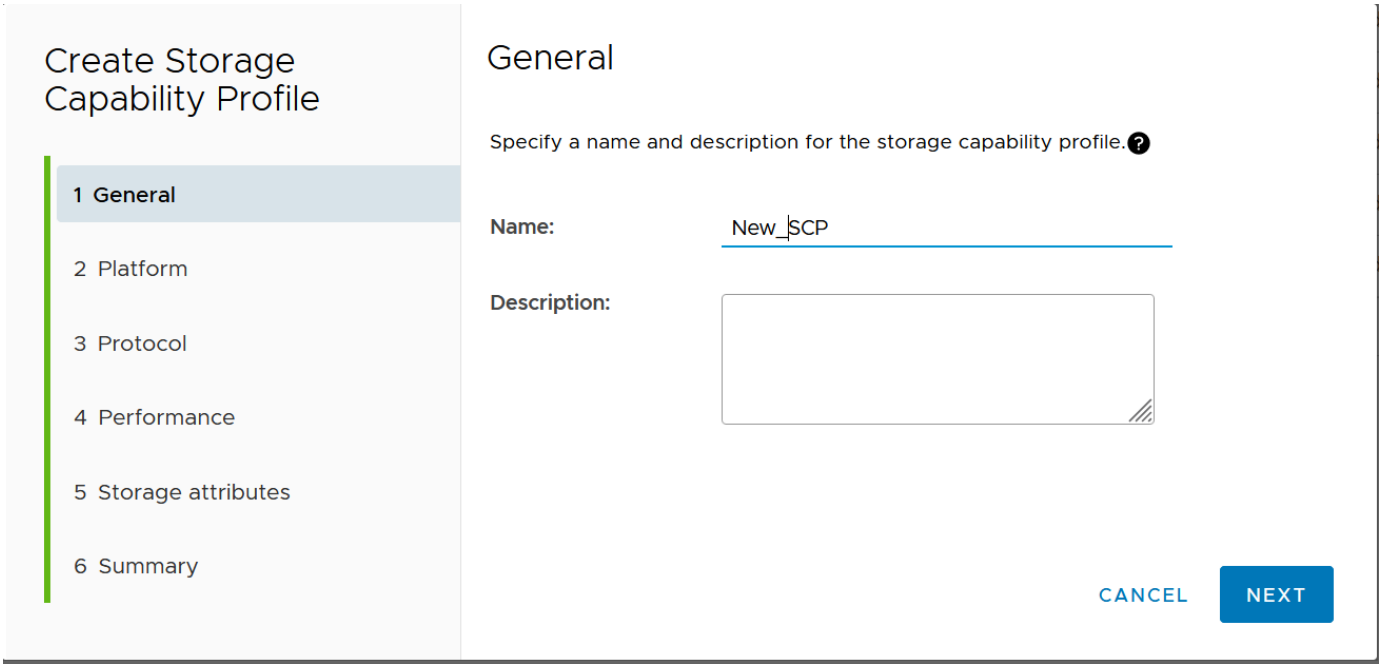
* SCP機能*	機能値	サポートされているリリース	* メモ *
* 圧縮 *	はい、いいえ、任意	すべて	7.2以降のAFF では必須です。
* 重複排除 *	はい、いいえ、任意	すべて	7.2以降のAFF では必須です。
* 暗号化 *	はい、いいえ、任意	7.2以降	暗号化されたFlexVol ボリュームを選択または作成します。ONTAP ライセンスが必要です。
* 最大 IOPS *	<number>	7.1以降ですが、違いがあります	7.2以降のQoSポリシーグループに表示されます。を参照してください ONTAP tools 9.10以降によるパフォーマンス管理 を参照してください。
パーソナリティ	略称はFAS	7.2以降	FAS には、ONTAP Select など、AFF以外のシステムも含まれます。AFF にはASA が含まれます。
プロトコル	NFS、NFS 4.1、iSCSI 、FCP、NVMe/FC、 任意	7.1以前、9.10以降	7.2-9.8は実質的に「任意」です。9.10以降では、NFS 4.1とNVMe/FC が元のリストに追加されました。
スペースリザベーション（シンプロビジョニング）	Thin（シン）、Thick（シック）、（任意）	すべて、違いを除いて	7.1以前ではシンプロビジョニングと呼ばれ、anyの値も使用できました。7.2ではスペースリザベーションと呼ばれていますすべてのリリースのデフォルトはシンです。

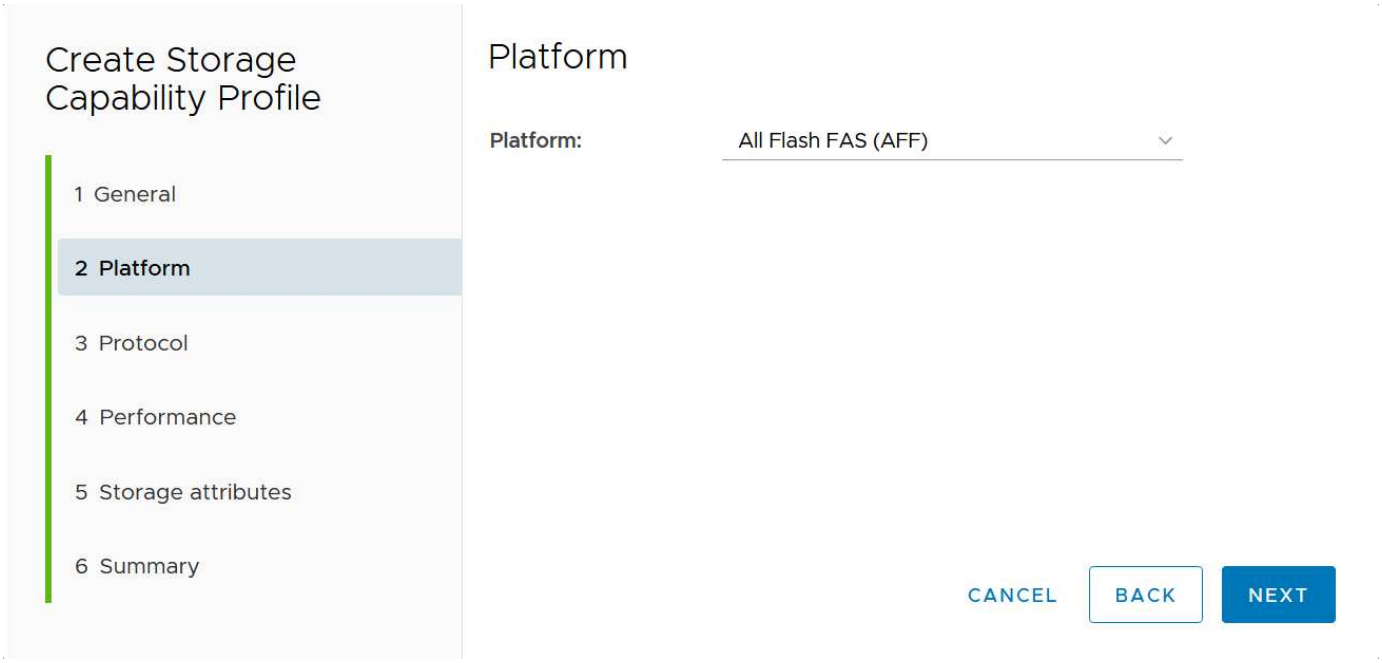
* SCP機能*	機能値	サポートされているリリース	* メモ *
* 階層化ポリシー *	[任意]、[なし]、[スナップショット]、[自動]	7.2以降	FabricPoolに使用- ONTAP 9.4以降を搭載したAFFまたはASAが必要です。NetApp StorageGRIDのようなオンプレミスのS3解決策 を使用しないかぎり、Snapshotのみが推奨されます。

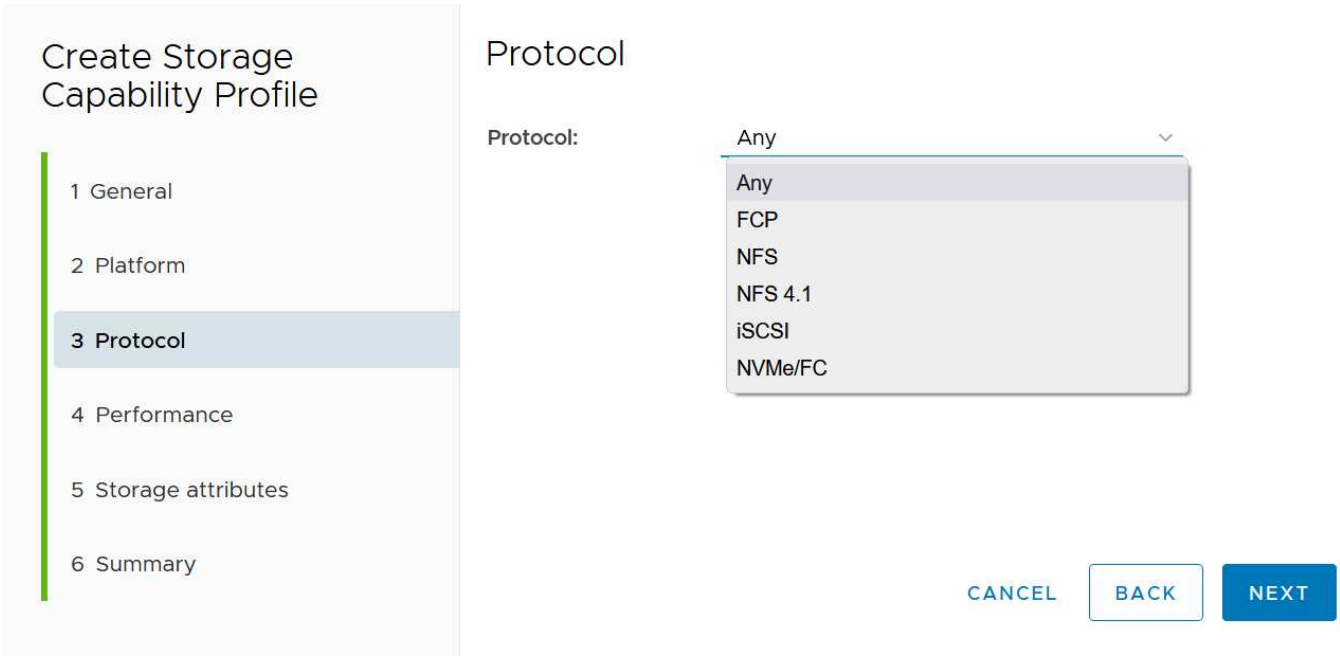
ストレージ機能プロファイルの作成

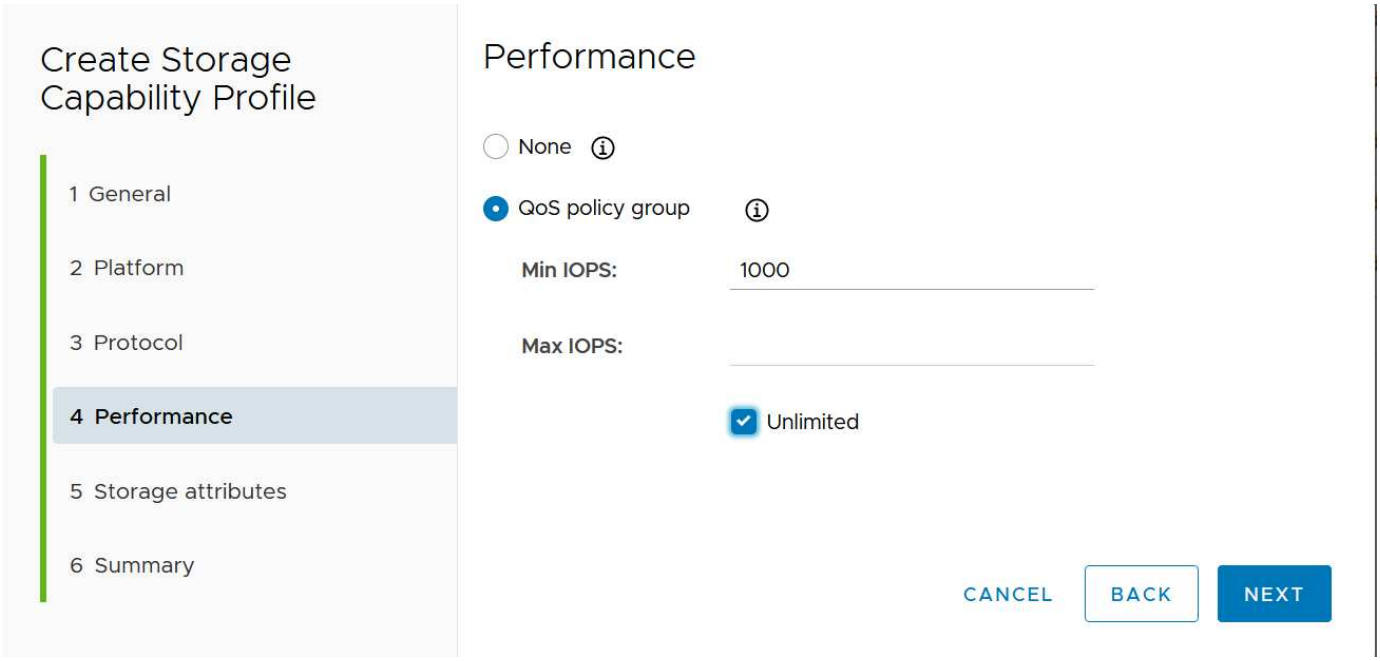
NetApp VASA Providerには、いくつかのSCPが事前定義されています。新しいSCPは、vCenter UIを使用して手動で作成することも、REST APIを使用した自動化を通じて作成することもできます。新しいプロファイルで機能を指定するか、既存のプロファイルをクローニングするか、既存の従来のデータストアからプロファイルを自動生成します。これは、ONTAP ツールのメニューを使用して行います。ストレージ機能プロファイル_を使用してプロファイルを作成またはクローニングし、ストレージマッピング_を使用してプロファイルを自動生成します。

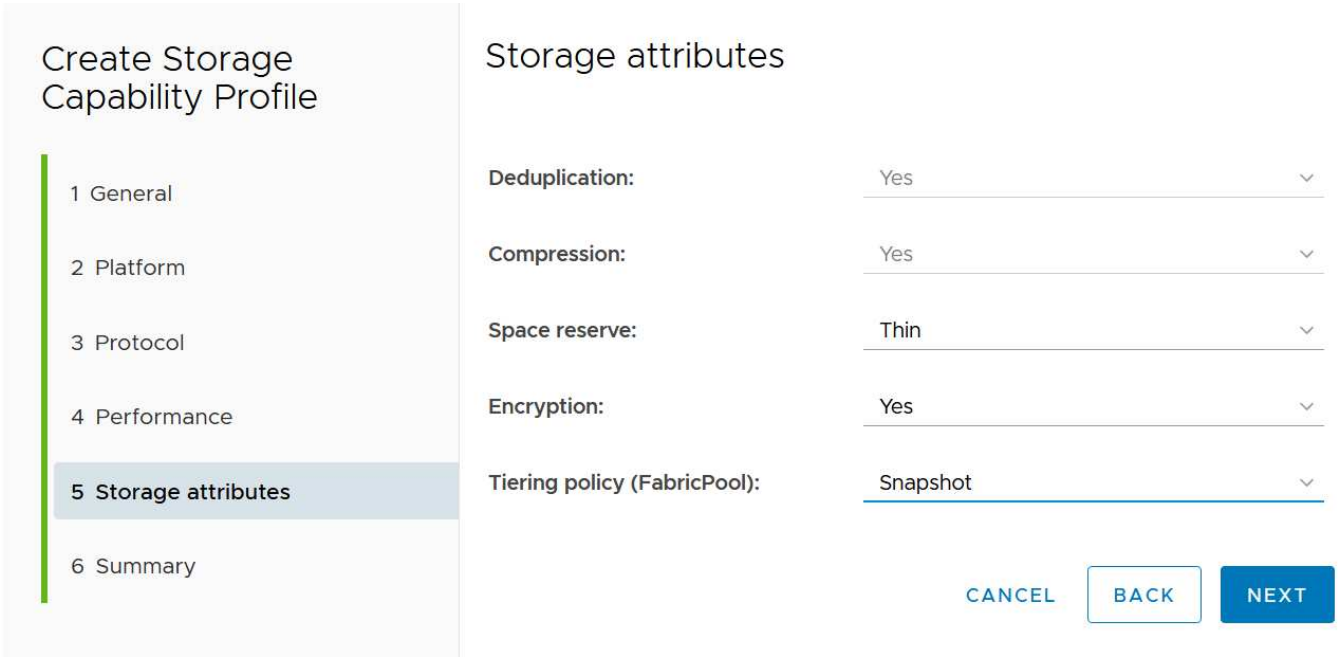
ONTAP tools 9.10以降のストレージ機能

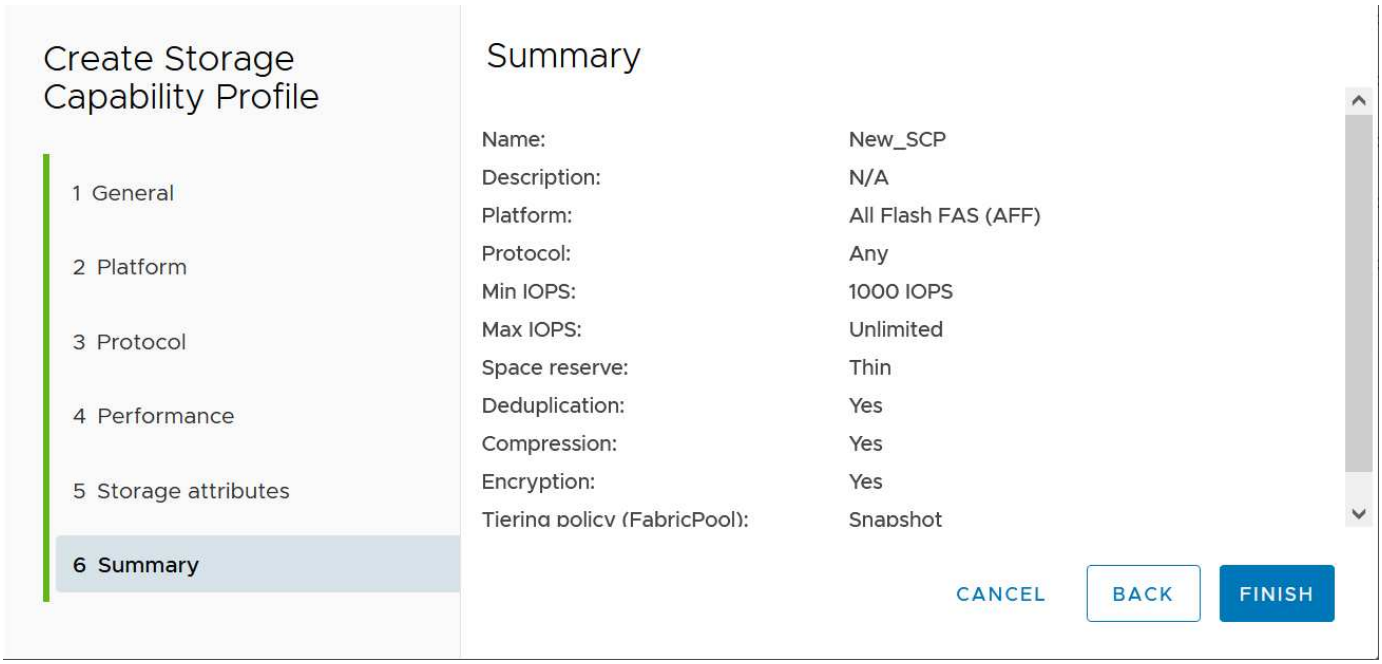



Create Storage Capability Profile
1 General
2 Platform
3 Protocol
4 Performance
5 Storage attributes
6 Summary
Platform
Platform:
All Flash FAS (AFF)
CANCEL
BACK
NEXT


Create Storage Capability Profile
1 General
2 Platform
3 Protocol
4 Performance
5 Storage attributes
6 Summary
Protocol
Protocol:
Any
Any
FCP
NFS
NFS 4.1
iSCSI
NVMe/FC
CANCEL
BACK
NEXT


Create Storage Capability Profile
1 General
2 Platform
3 Protocol
4 Performance
5 Storage attributes
6 Summary
Performance
None
QoS policy group
Min IOPS:
1000
Max IOPS:
Unlimited
CANCEL
BACK
NEXT


Create Storage Capability Profile
1 General
2 Platform
3 Protocol
4 Performance
5 Storage attributes
6 Summary
Storage attributes
Deduplication:
Yes
Compression:
Yes
Space reserve:
Thin
Encryption:
Yes
Tiering policy (FabricPool):
Snapshot
CANCEL
BACK
NEXT



- vVolデータストアを作成しています*
  必要なSCPを作成したら、そのSCPを使用してVVOLデータストア（および必要に応じてデータストア用のFlexVol ボリューム）を作成できます。ONTAP データストアを作成するホスト、クラスタ、またはデータセンターを右クリックし、_vVol tools>>_Provision Datastore_を選択します。データストアでサポートするSCPを1つ以上選択し、既存のFlexVol ボリュームから選択するか、データストア用に新しいFlexVol ボリュームをプロビジョニングします。最後に、データストアのデフォルトのSCPを指定します。このSCPは、ポリシーで指定されたSCPが設定されていないVMやスワップVVOL（ハイパフォーマンスなストレージは必要ありません）に使用されます。

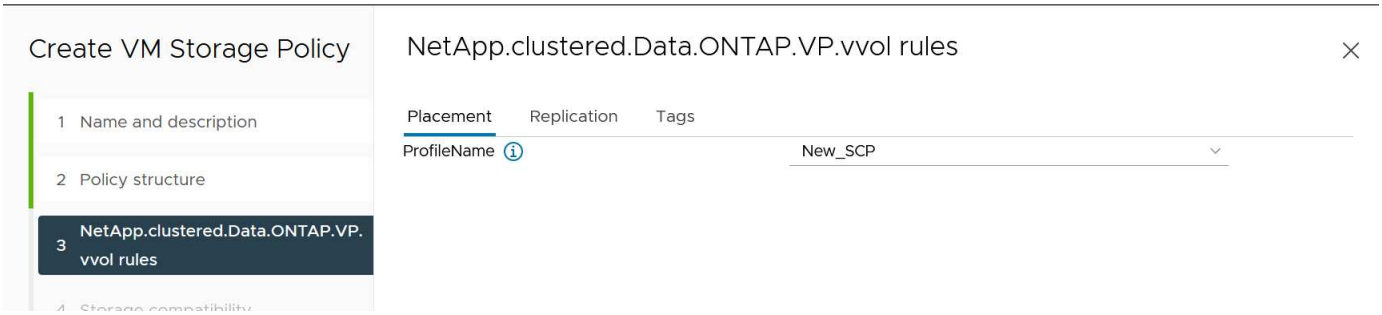
仮想マシンストレージポリシーを作成しています

仮想マシンストレージポリシーは、Storage I/O ControlやvSphere Encryptionなどのオプション機能を管理するためにvSphereで使用されます。また、VVOLでも使用され、特定のストレージ機能をVMに適用します。ポリシーを使用して特定のSCPをVMに適用するには、「NetApp.clustered.Data.ONTAP.VP.vVol」ストレージタイプと「ProfileName」ルールを使用します。ONTAP tools VASA Providerを使用した場合の例については、link：vmware-vvols-ontap.html #ベストプラクティス[NFS v3経由のVVOLを使用したネットワーク設定例]を参照してください。「NetApp.clustered.Data.ONTAP.VP.VASA10」ストレージのルールは、VVOLベース以外のデータストアで使用します。

以前のリリースも似ていますが、で説明しているように、 ストレージ機能プロファイルONTAP toolsリリース別の機能オプションは異なります。

作成したストレージポリシーは、に示すように、新しいVMのプロビジョニング時に使用できます。 "ストレージポリシーを使用してVMを導入します"。VASA Provider 7.2でパフォーマンス管理機能を使用する場合のガイドラインについては、を参照してください。 ONTAP tools 9.10以降によるパフォーマンス管理。

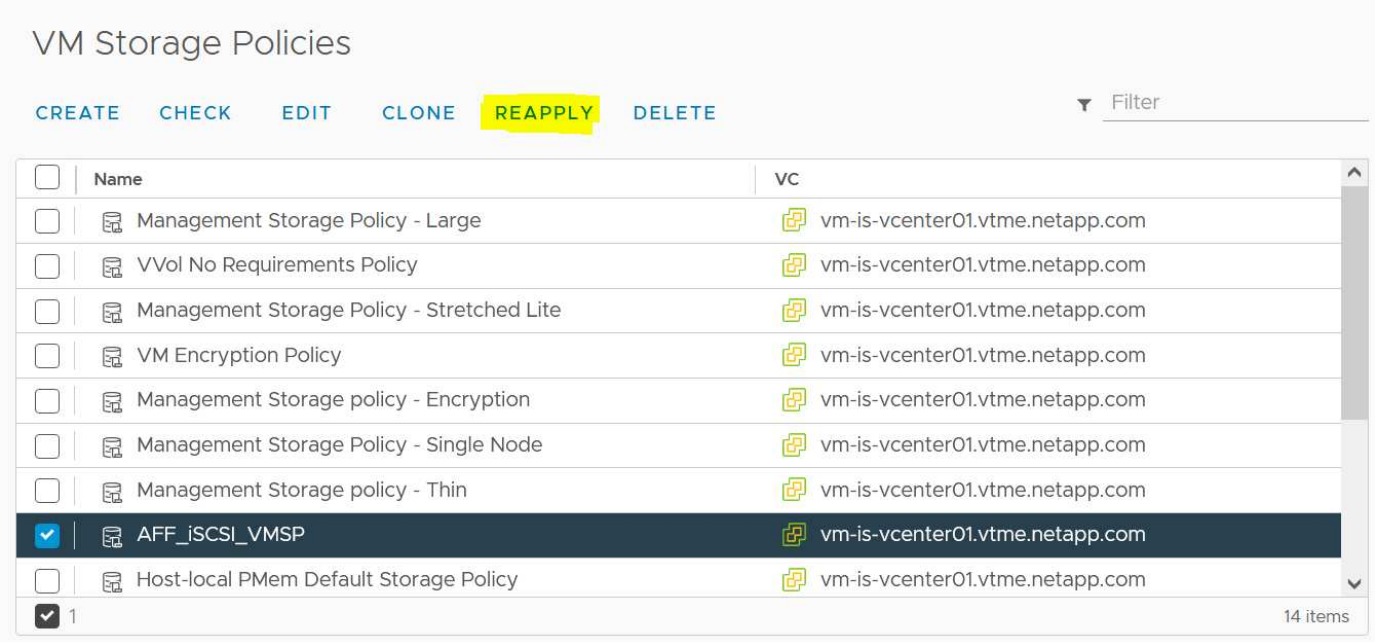
ONTAP tools VASA Provider 9.10を使用した**VM**ストレージポリシーの作成



ONTAP tools 9.10以降によるパフォーマンス管理

- ONTAP tools 9.10では、独自の分散配置アルゴリズムを使用して、vVolデータストア内の最適なFlexVolに新しいvVolが配置されます。指定したSCPと一致するFlexVol ボリュームに基づいて配置されます。これにより、データストアとバッキングストレージが、指定されたパフォーマンス要件を確実に満たすことができます。
- 最小IOPSや最大IOPSなどのパフォーマンス機能を変更するには、特定の構成に注意する必要があります。
  - *最小IOPSと最大IOPS *はSCPで指定し、VMポリシーで使用できます。
    - SCPでIOPSを変更しても、VMポリシーを編集してそれを使用するVMに再適用するまで、VVOLのQoSは変更されません（ ONTAP tools 9.10以降のストレージ機能）。または、必要なIOPSで新しいSCPを作成し、そのSCPを使用する（VMに再適用する）ようにポリシーを変更します。一般的には、サービス階層ごとに個別のSCPとVMストレージポリシーを定義し、VMのVMストレージポリシーを変更することを推奨します。
    - AFF とFAS のパーソナリティではIOPS設定が異なります。AFF では、MinとMaxの両方を使用できます。ただし、AFF以外のシステムで使用できるのは最大IOPSの設定のみです。
- 場合によっては、ポリシーの変更後（手動またはVASA ProviderとONTAP による自動）にVVOLの移行が必要になることがあります。
  - 一部の変更では移行は必要ありません（最大IOPSの変更など、前述のようにVMにすぐに適用できます）。
  - VVOLが格納されている現在のFlexVol でポリシーの変更をサポートできない場合（要求された暗号化ポリシーまたは階層化ポリシーがプラットフォームでサポートされていない場合など）は、vCenterでVMを手動で移行する必要があります。
- ONTAP toolsは、現在サポートされているバージョンのONTAP に対して、共有されていないQoSポリシーを個別に作成します。そのため、個 々 のVMDKにはそれぞれ独自のIOPSが割り当てられます。

VMストレージポリシーを再適用しています



VVOLを保護する

以降のセクションでは、VMware VVOLとONTAPストレージを使用する手順とベストプラクティスについて説明します。

VASA Providerの高可用性

NetApp VASA Providerは、vCenterプラグイン、REST APIサーバ（旧Virtual Storage Console[VSC]）、およびStorage Replication Adapterとともに仮想アプライアンスの一部として実行されます。VASA Providerを使用できない場合、VVOLを使用するVMは引き続き実行されます。ただし、新しいvVolデータストアを作成することはできず、vVolをvSphereで作成またはバインドすることもできません。vCenterはスワップVVolの作成を要求できないため、VVolを使用するVMの電源をオンにできません。また、vVolを新しいホストにバインドできないため、実行中のVMでvMotionを使用して別のホストに移行することはできません。

VASA Provider 7.1以降では、必要なときにサービスを利用できるようにするための新しい機能がサポートされています。VASA Providerと統合データベースサービスを監視する新しいwatchdogプロセスが含まれています。障害が検出されると、ログファイルが更新され、サービスが自動的に再起動されます。

vSphere管理者は、他のミッションクリティカルなVMをソフトウェア、ホストハードウェア、およびネットワークの障害から保護するのと同じ可用性機能を使用して、さらに保護を設定する必要があります。これらの機能を使用するために仮想アプライアンスで追加の設定を行う必要はありません。標準のvSphereアプローチを使用して設定するだけです。これらはネットアップによってテストされ、サポートされています。

vSphere High Availabilityは、障害発生時にホストクラスタ内の別のホストでVMを再起動するように簡単に構成できます。vSphere Fault Toleranceは、継続的にレプリケートされ、任意の時点でテイクオーバーできるセカンダリVMを作成することで、可用性を高めます。これらの機能の追加情報 はで使用できます "ONTAP tools for VMware vSphereのドキュメント（ONTAP toolsの高可用性の設定）"、およびVMware vSphereのドキュメント（「ESXiおよびvCenter ServerのvSphereの可用性」を参照）。

ONTAP tools VASA Providerは、VVOLの設定を管理対象のONTAP システムにリアルタイムで自動的にバックアップします。このシステムでは、VVOL情報がFlexVol ボリュームのメタデータに格納されます。何らかの理由でONTAP toolsアプライアンスが使用できなくなった場合でも、簡単かつ迅速に新しいアプライアンスを導入して設定をインポートできます。VASA Providerのリカバリ手順の詳細については、次の技術情報アーテ

ィクルを参照してください。

"『How to perform a VASA Provider Disaster Recovery - Resolution Guide』"

vVolレプリケーション

ONTAP をご利用のお客様の多くは、NetApp SnapMirrorを使用して従来のデータストアをセカンダリストレージシステムにレプリケートし、災害発生時にセカンダリシステムを使用して個 々 のVMやサイト全体をリカバリしています。ほとんどの場合、お客様はこの管理にソフトウェアツールを使用します。たとえば、VMware vSphere用NetApp SnapCenterプラグインなどのバックアップソフトウェア製品や、VMwareのSite Recovery Managerなどのディザスタリカバリ解決策（ONTAPツールのStorage Replication Adapterとともに使用）などです。

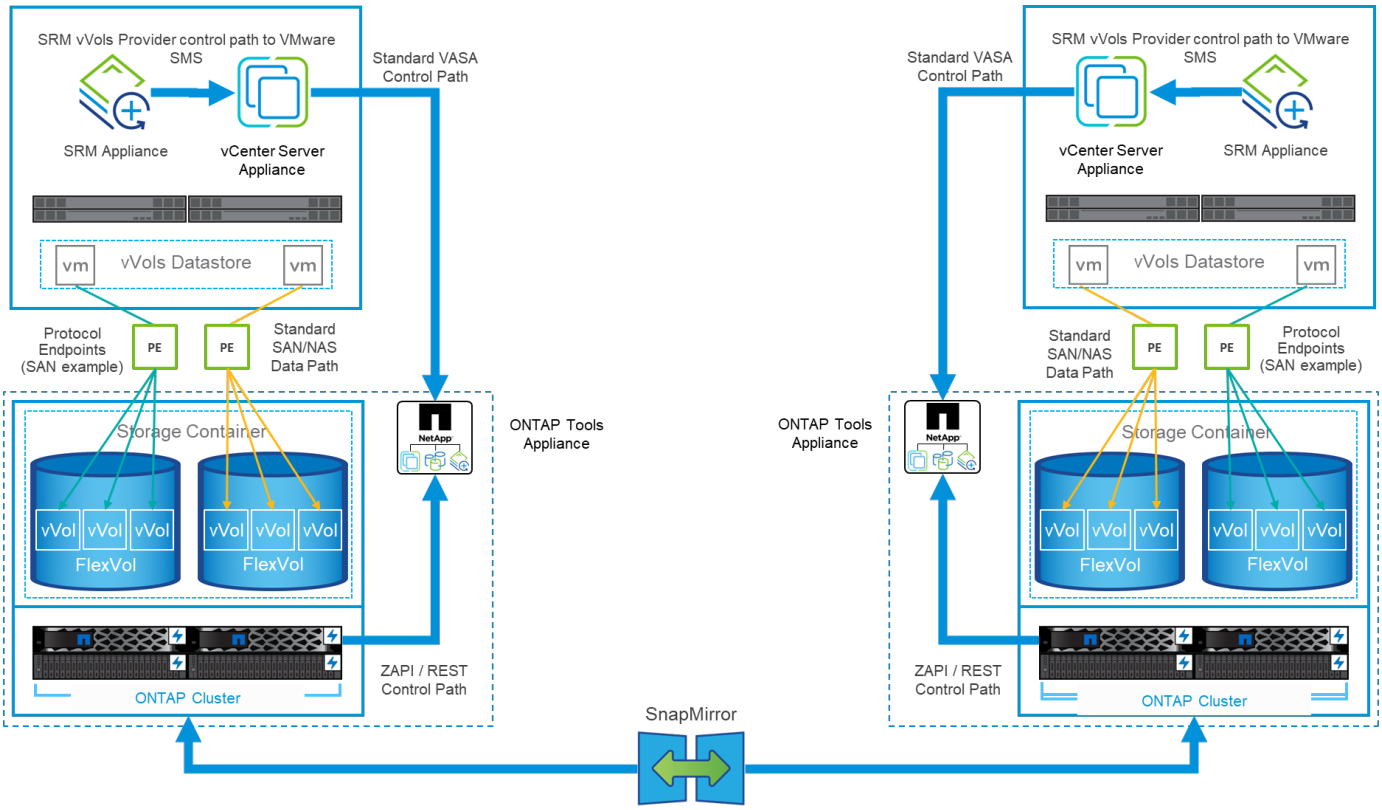
このソフトウェアツールの要件は、vVolレプリケーションの管理においてさらに重要になります。一部の機能はネイティブの機能で管理できます（たとえば、VMwareが管理するvVolのSnapshotは、高速で効率的なファイルクローンまたはLUNクローンを使用するONTAP にオフロードされます）が、一般的には、レプリケーションとリカバリを管理するためにオーケストレーションが必要です。VVOLに関するメタデータは、ONTAPとVASA Providerによって保護されますが、セカンダリサイトでメタデータを使用するには追加の処理が必要です。

ONTAP tools 9.7.1とVMware Site Recovery Manager（SRM）8.3リリースを併用すると、ディザスタリカバリと移行のワークフローオーケストレーションのサポートが追加され、NetApp SnapMirrorテクノロジのメリットを活用できるようになりました。

ONTAP tools 9.7.1を使用したSRMの初期リリースでは、FlexVolを事前に作成し、それらをVVOLデータストアのバッキングボリュームとして使用する前にSnapMirror保護を有効にする必要がありました。ONTAP tools 9.10以降では、このプロセスは不要になりました。既存のバッキングボリュームにSnapMirror保護を追加し、VMのストレージポリシーを更新して、SRMに統合されたディザスタリカバリと移行のオーケストレーション、自動化機能を備えたポリシーベースの管理を活用できるようになりました。

現在、ネットアップがサポートするvVol用のディザスタリカバリおよび移行自動化の解決策 はVMware SRMのみです。ONTAP ツールでは、vVolレプリケーションを有効にする前に、vCenterに登録されているSRM 8.3以降のサーバの有無が確認されます。 ONTAP ツールREST APIを活用して独自のサービスを作成することも可能です。

SRMを使用したvVolレプリケーション



MetroCluster のサポート

ONTAP toolsではMetroCluster のスイッチオーバーはトリガーされませんが、同じvSphere Metro Storage Cluster（vMSC）構成のVVol用NetApp MetroCluster システムではサポートされます。MetroCluster システムのスイッチオーバーは通常の方法で処理されます。

NetApp SnapMirrorビジネス継続性（SM-BC）はvMSC構成のベースとしても使用できますが、現時点ではVVOLではサポートされていません。

NetApp MetroCluster の詳細については、次のガイドを参照してください。

"*TR-4689 MetroCluster IP解決策 のアーキテクチャと設計*"

"*TR-4705 NetApp MetroCluster 解決策 のアーキテクチャと設計*"

"*VMware KB 2031038 NetApp MetroCluster によるVMware vSphereのサポート*"

vVolバックアップの概要

ゲスト内バックアップエージェントの使用、VMデータファイルのバックアッププロキシへの接続、VMware VADPなどの定義済みAPIの使用など、VMを保護する方法はいくつかあります。VVOLは同じメカニズムを使用して保護でき、多くのネットアップパートナーがVVOLを含むVMのバックアップをサポートしています。

前述したように、VMware vCenterで管理されるスナップショットは、スペース効率に優れた高速なONTAP ファイル/ LUNクローンにオフロードされます。これらは迅速な手動バックアップに使用できますが、vCenterでは最大32個のスナップショットに制限されています。vCenterを使用してスナップショットを作成し、必要に応じて元に戻すことができます。

SnapCenter Plugin for VMware vSphere（SCV）4.6以降では、ONTAP tools 9.10以降と組み合わせて使用す

ることで、vVolベースのVMのcrash-consistentバックアップおよびリカバリがサポートされるようになりました。SnapMirrorおよびSnapVault レプリケーションがサポートされたONTAP FlexVol ボリュームSnapshotを活用します。ボリュームあたり最大1023個のSnapshotがサポートされます。また、ミラーバックアップポリシーを使用したSnapMirrorを使用すると、保持期間の長いSnapshotをセカンダリボリュームに格納することもできます。

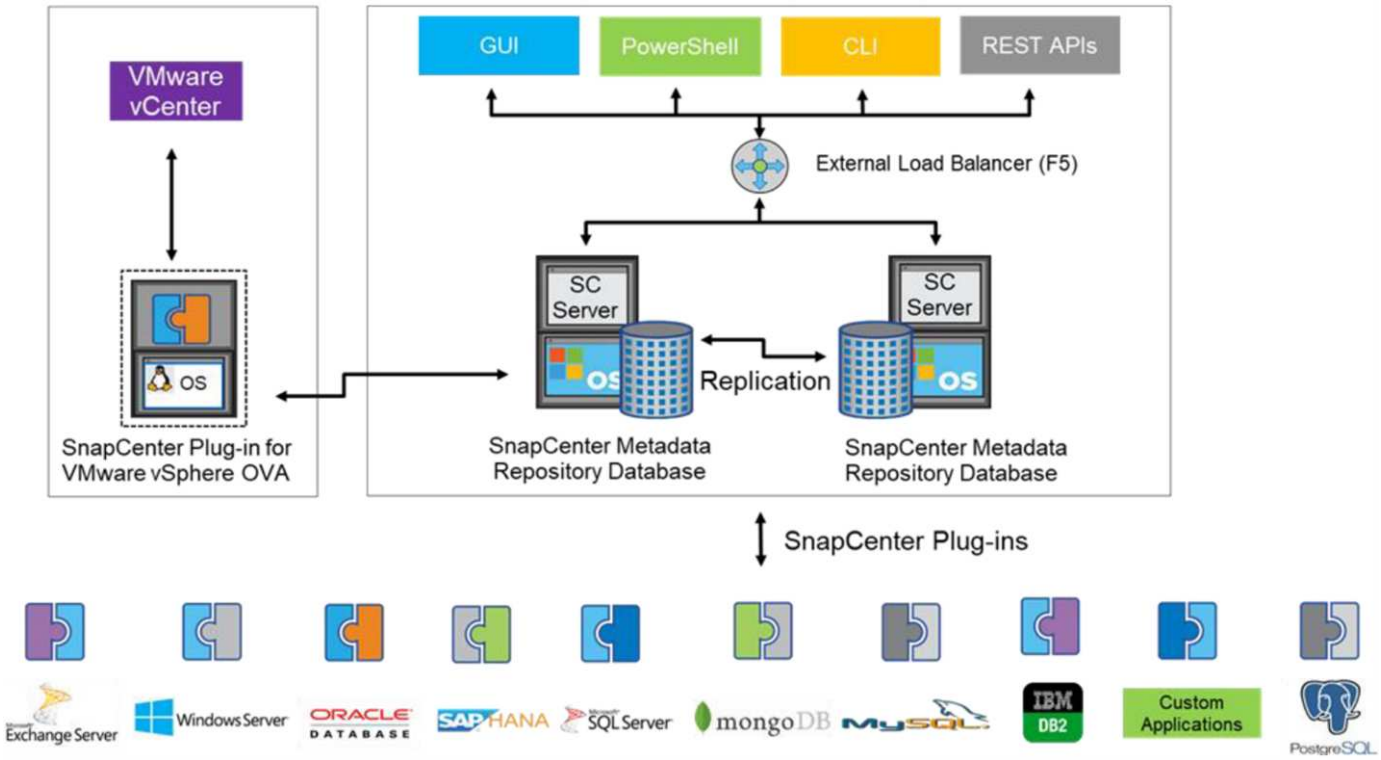
vSphere 8.0のサポートは、分離されたローカルプラグインアーキテクチャを使用するSCV 4.7で導入されました。vSphere 8.0U1のサポートがSCV 4.8に追加され、新しいリモートプラグインアーキテクチャに完全に移行しました。

VMware vSphere用のSnapCenter プラグインを使用したVVolバックアップ

NetApp SnapCenterでは、タグやフォルダに基づいてvVolのリソースグループを作成し、vVolベースのVMに対してONTAPのFlexVolベースのSnapshotを自動的に利用できるようになりました。これにより、環境内で動的にプロビジョニングされたVMを自動的に保護するバックアップ/リカバリサービスを定義できます。

SnapCenter Plugin for VMware vSphereは、vCenter拡張機能として登録されたスタンドアロンアプライアンスとして導入され、vCenter UIまたはREST APIを使用して管理され、バックアップ/リカバリサービスの自動化が可能です。

SnapCenter アーキテクチャ



本ドキュメントの執筆時点では、他のSnapCenterプラグインはまだVVOLをサポートしていないため、本ドキュメントではスタンドアロンの導入モデルについて説明します。

SnapCenter はONTAP FlexVol スナップショットを使用するため、vSphereへのオーバーヘッドは発生しません。また、vCenterで管理されているスナップショットを使用する従来のVMで発生する可能性のあるパフォーマンスの低下もありません。さらに、SCVの機能はREST APIを介して公開されるため、VMware ARIA Automation、Ansible、Terraformなどのツールや、標準のREST APIを使用できるその他のほぼすべての自動化ツールを使用して、自動化されたワークフローを簡単に作成できます。

SnapCenter REST API については、を参照してください "REST API の概要"

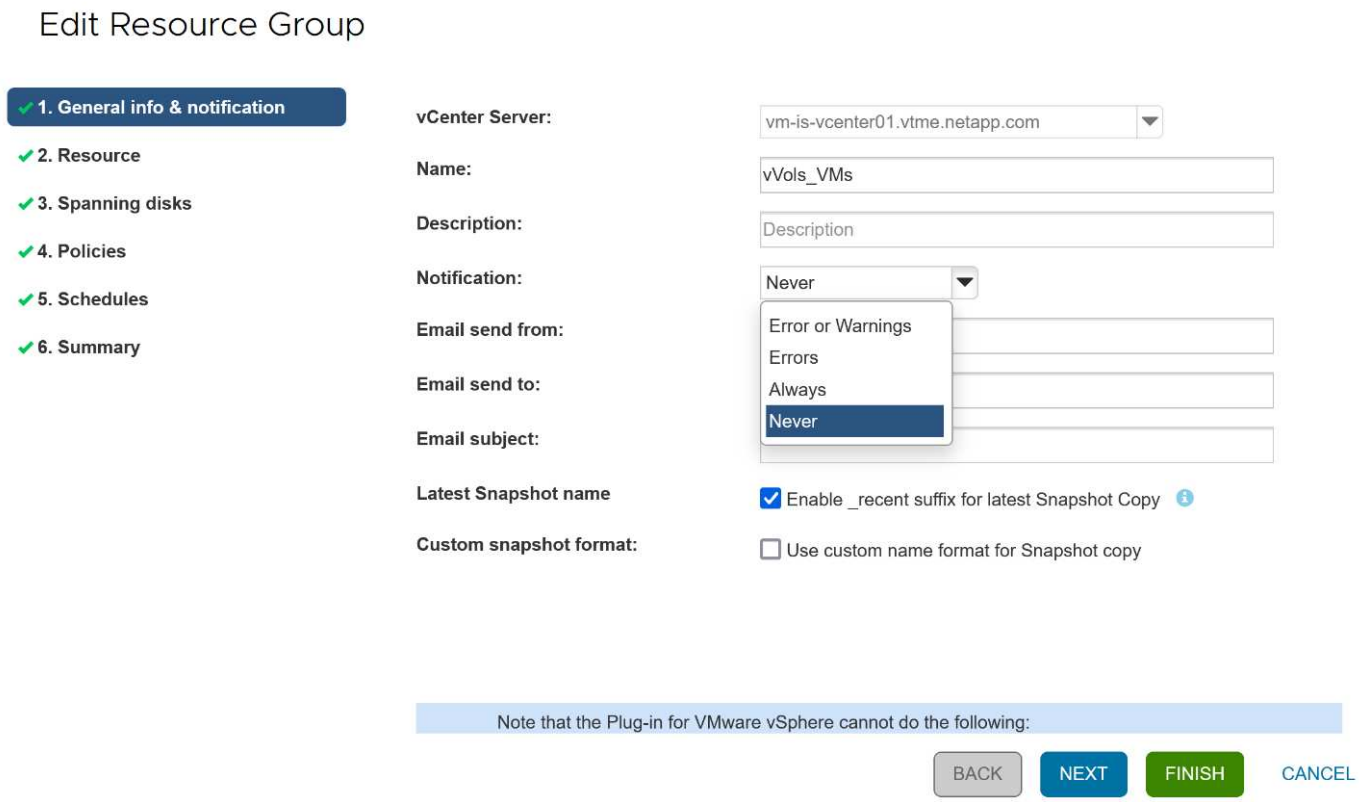
SnapCenter Plug-in for VMware vSphere REST API については、を参照してください "SnapCenter Plug-in for VMware vSphere REST API"

ベストプラクティス

SnapCenter 環境を最大限に活用するには、次のベストプラクティスを参考にしてください。

- SCVはvCenter Server RBACとONTAP RBACの両方をサポートしており、プラグインの登録時に自動的に作成される事前定義されたvCenterロールが用意されています。サポートされるRBACのタイプの詳細については、こちらを参照してください "こちらをご覧ください。"
  - vCenter UIを使用して、説明されている事前定義されたロールを使用して最小権限のアカウントアクセスを割り当てます "こちらをご覧ください"。
  - SnapCenter サーバでSCVを使用する場合は、_SnapCenterADMIN_ROLEを割り当てる必要があります。
  - ONTAP RBACは、SCVで使用するストレージシステムを追加および管理するために使用するユーザアカウントを指します。ONTAP RBACは、VVOLベースのバックアップには適用されません。ONTAP RBACとSCVの詳細については、こちらをご覧ください "こちらをご覧ください"。
- SnapMirrorを使用してバックアップデータセットを別のシステムにレプリケートし、ソースボリュームの完全なレプリカを作成します。前述したように、ソースボリュームのSnapshotの保持設定に関係なく、バックアップデータの長期保持にmirror-vaultポリシーを使用することもできます。どちらのメカニズムもVVOLでサポートされています。
- SCVではVVOL機能にONTAP Tools for VMware vSphereを使用する必要があるため、特定のバージョンの互換性については、必ずNetApp Interoperability Matrix Tool（IMT ）を参照してください
- VMware SRMでvVolレプリケーションを使用する場合は、ポリシーのRPOとバックアップスケジュールに注意してください
- 組織で定義された目標復旧時点（RPO）を満たす保持設定を使用してバックアップポリシーを設計
- バックアップの実行時にステータスが通知されるようにリソースグループに通知を設定します（下記の図10を参照）。

リソースグループの通知オプション



これらのドキュメントを使用して、**SCV**の使用を開始します

"SnapCenter Plug-in for VMware vSphere について説明します"

"SnapCenter Plug-in for VMware vSphere を導入"

トラブルシューティング

追加情報 には、いくつかのトラブルシューティングリソースが用意されています。

NetApp Support Site

NetApp Support Site には、ネットアップの仮想化製品に関するさまざまな技術情報アーティクルのほかに、の便利なランディングページも用意されています "VMware vSphere 用の ONTAP ツール" 製品：このポータルには、ネットアップコミュニティの記事、ダウンロード、テクニカルレポート、VMwareソリューションに関するディスカッションへのリンクが掲載されています。次のURLから入手できます。

"_ NetApp Support Site _"

その他の解決策 ドキュメントは、次のURLから入手できます。

"仮想化向けネットアップソリューション"

製品のトラブルシューティング

vCenterプラグイン、VASA Provider、Storage Replication Adapterなど、ONTAP ツールのさまざまなコンポーネントは、いずれもネットアップのドキュメントリポジトリにまとめられています。ただし、それぞれにKnowledge Baseのサブセクションがあり、特定のトラブルシューティング手順が記載されている場合があ

ります。これらは、VASA Providerで発生する可能性のある最も一般的な問題に対処します。

VASA ProviderのUIの問題

vCenter vSphere Web ClientでSerenityのコンポーネントに関する問題が発生し、VASA Provider for ONTAPのメニュー項目が表示されないことがあります。導入ガイドまたはこのナレッジベースのVASA Provider登録の問題の解決を参照してください "記事"。

vVolデータストアのプロビジョニングが失敗する

vVolデータストアの作成時にvCenterサービスがタイムアウトすることがあります。修正するには、vmware-spsサービスを再起動し、vCenterのメニュー（[Storage]>[New Datastore]）を使用してvVolデータストアを再マウントします。この問題については、『Administration Guide』のvCenter Server 6.5でvVolデータストアのプロビジョニングが失敗するという項を参照してください。

Unified Applianceをアップグレードすると、ISOのマウントに失敗します

vCenterのバグが原因で、Unified Applianceをあるリリースから次のリリースへアップグレードするために使用されるISOがマウントに失敗する可能性があります。ISOをvCenterのアプライアンスに接続できる場合は、このナレッジベースの手順に従ってください "記事" 解決するために。

VMware Site Recovery ManagerとONTAP

VMware Site Recovery ManagerとONTAP

ONTAPは、2002年に最新のデータセンターに導入されて以来、VMware vSphere環境向けストレージ解決策として業界をリードしてきました。また、コストを削減しながら管理を簡易化する革新的な機能を継続的に追加しています。

このドキュメントでは、業界をリードするVMwareのディザスタリカバリ（DR）ソフトウェアであるONTAP解決策for VMware Site Recovery Manager（SRM）について説明します。最新の製品情報とベストプラクティスを紹介し、導入の合理化、リスクの軽減、継続的な管理の簡素化を実現します。

> このドキュメントは、以前に公開されていたテクニカルレポート「TR-4900：VMware Site Recovery Manager」をONTAPに置き換えます。

ベストプラクティスは、ガイドや互換性ツールなどの他のドキュメントを補うものです。ラボテストに基づいて開発されており、ネットアップのエンジニアやお客様は広範な現場経験を積んでいます。推奨されるベストプラクティスがお客様の環境に適していない場合もありますが、一般に最もシンプルなソリューションであり、ほとんどのお客様のニーズに対応できます。

本ドキュメントでは、ONTAP Tools for VMware vSphere 9.12（NetApp Storage Replication Adapter[SRA]およびVASA Provider[VP]を含む）およびVMware Site Recovery Manager 8.7と組み合わせて使用した場合の、ONTAP 9の最近のリリースの機能を中心に説明します。

SRM で ONTAP を使用する理由

ONTAP ソフトウェアを基盤とするネットアップのデータ管理プラットフォームは、 SRM に最も広く採用されているストレージソリューションの一部です。理由はそれだけではありません。セキュアでハイパフォーマンスなユニファイドプロトコル（NASとSANを併用）データ管理プラットフォームで、業界を定義するストレージ効率、マルチテナンシー、サービス品質管理、スペース効率に優れたSnapshotによるデータ保

護、SnapMirrorによるレプリケーションを実現します。VMware ワークロードを保護するためにネイティブのハイブリッドマルチクラウド統合を活用し、多数の自動化ツールやオーケストレーションツールを簡単に利用できます。

SnapMirrorをアレイベースのレプリケーションに使用すると、実績のある成熟したONTAPのテクノロジを活用できます。SnapMirror を使用すると、 VM やデータストア全体ではなく、変更されたファイルシステムブロックのみをコピーして、データを安全かつ効率的に転送できます。重複排除、圧縮、コンパクションなどのスペース削減効果を活用できます。最新の ONTAP システムで、バージョンに依存しない SnapMirror が使用されるようになり、ソースとデスティネーションのクラスタを柔軟に選択できるようになりました。SnapMirror は、災害復旧のための最も強力なツールの 1 つとなりました。

従来の NFS 、 iSCSI 、ファイバチャネル接続データストア（現在は VVOL データストアをサポート）のいずれを使用している場合でも、 SRM は、ディザスタリカバリやデータセンター移行の計画とオーケストレーションに ONTAP の機能のメリットを活用する堅牢なファーストパーティ製品を提供します。

SRM での **ONTAP 9** の活用方法

SRM は、 ONTAP システムの高度なデータ管理テクノロジを活用して、 3 つの主要コンポーネントで構成される仮想アプライアンスである VMware vSphere 用 ONTAP ツールと統合します。

- vCenter プラグイン（旧 Virtual Storage Console （ VSC ））は、 SAN と NAS のどちらを使用している場合でも、ストレージ管理と効率化機能の簡易化、可用性の向上、ストレージコストと運用オーバーヘッドの削減を実現します。データストアのプロビジョニングのベストプラクティスを使用して、 NFS 環境およびブロックストレージ環境用の ESXi ホスト設定を最適化します。以上のメリットのために、 ONTAP ソフトウェアを実行するシステムで vSphere を使用する場合はこのプラグインを推奨します。
- VASA Provider for ONTAP は、 VMware vStorage APIs for Storage Awareness （ VASA ）フレームワークをサポートしています。VASA Provider では、 VM ストレージのプロビジョニングと監視に役立つように vCenter Server と ONTAP を接続します。VMware Virtual Volumes （ VVol ）のサポートと、ストレージ機能プロファイル（ VVol レプリケーション機能を含む）の管理、および個々の VM VVol のパフォーマンスの管理が可能になります。また、容量の監視やプロファイルへの準拠に関するアラームも生成されます。SRM と一緒に使用すると、 VASA Provider for ONTAP で VVOL ベースの仮想マシンをサポートできます。 SRM サーバに SRA アダプタをインストールする必要はありません。
- SRA は SRM と一緒に使用され、従来の VMFS データストアと NFS データストアの本番サイトとディザスタリカバリサイト間での VM データのレプリケーションを管理します。また、 DR レプリカの無停止テストにも使用できます。検出、リカバリ、再保護のタスクを自動化します。Windows SRM サーバおよび SRM アプライアンス用の SRA サーバアプライアンスと SRA アダプタの両方が含まれています。

SRM サーバに SRA アダプタをインストールして設定し、 VASA Provider で VVol 以外のデータストアを保護したり VVOL のレプリケーションを有効にしたりしたあとで、ディザスタリカバリ用に vSphere 環境を設定する作業を開始できます。

SRA と VASA Provider には、 SRM サーバ用のコマンド / 制御インターフェイスが用意されており、 VMware 仮想マシン（ VM ）を含む ONTAP FlexVol や、 SRA を保護する SnapMirror レプリケーションを管理できます。

SRM 8.3 以降では、 SRM サーバへの新しい SRM VVol Provider 制御パスが導入され、 SRA を使用せずに vCenter サーバおよびその経由で VASA Provider に通信できるようになりました。これにより、 SRM サーバは緊密に統合するための完全な API を提供するため、以前よりもはるかに ONTAP クラスタの制御を活用できました。

SRMでは、ネットアップ独自のFlexCloneテクノロジを使用して、システムを停止することなくDR計画をテストし、保護されたデータストアのクローンをDRサイトにほぼ瞬時に作成できます。SRM はサンドボックスを作成して安全にテストし、真の災害が発生した場合に組織とお客様を保護します。そのため、組織は災害時

にフェイルオーバーを実行できます。

実際に災害が発生した場合や、計画的な移行の場合でも、 SRM では、最終的な SnapMirror 更新（必要な場合）を使用して、データセットに最新の変更を送信できます。その後、ミラーを解除し、 DR ホストにデータストアをマウントします。この時点で、計画済みの戦略に基づいて、 VM の電源を任意の順序で自動的にオンにすることができます。

SRM と **ONTAP** などのユースケース：ハイブリッドクラウドと移行

SRM 環境に ONTAP の高度なデータ管理機能を統合することで、ローカルストレージオプションに比べて、拡張性とパフォーマンスが大幅に向上します。それだけではありませんが、ハイブリッドクラウドの柔軟性を備えています。ハイブリッドクラウドを使用すると、 FabricPool を使用して、未使用のデータブロックをハイパフォーマンスアレイから希望するハイパースケーラに階層化してコストを削減できます。これは、 NetApp StorageGRID などのオンプレミスの S3 ストアである可能性があります。また、 ONTAP Select （ CVO ）やを使用して、ソフトウェアで定義される Cloud Volumes ONTAP やクラウドベースの DR でエッジベースのシステムに SnapMirror を使用することもできます "Equinix 内の NetApp Private Storage" Amazon Web Services （ AWS ）、 Microsoft Azure 、 Google Cloud Platform （ GCP ）で、クラウド内に完全に統合されたストレージ、ネットワーク、コンピューティングサービスのスタックを構築できます。

その後、FlexCloneを使用すれば、ストレージの設置面積をほぼゼロに抑えながら、クラウドサービスプロバイダのデータセンター内でテストフェイルオーバーを実行できます。組織を保護することで、かつてないほどコストを削減できます。

SRM は、 SnapMirror を使用して、計画的な移行を実行することもできます。これにより、 VM を 1 つのデータセンターから別のデータセンターに効率的に転送したり、独自のデータセンターや、任意の数のネットアップパートナーサービスプロバイダを介して VM を転送したりできます。

導入のベストプラクティス

次のセクションでは、ONTAPとVMware SRMを使用した導入のベストプラクティスについて説明します。

SMT の **SVM** のレイアウトとセグメント化

ONTAP では、 Storage Virtual Machine （ SVM ）の概念を採用して、セキュアなマルチテナント環境で厳密にセグメント化します。ある SVM の SVM ユーザは、別の SVM のリソースにアクセスしたりリソースを管理したりすることはできませんこれにより、 ONTAP テクノロジを活用できます。ビジネスユニットごとに別々の SVM を作成して、同じクラスタ上で独自の SRM ワークフローを管理することで、全体的なストレージ効率を高めることができます。

SVM を対象としたアカウントと SVM 管理 LIF を使用して ONTAP を管理することを検討し、セキュリティ制御を強化するだけでなく、パフォーマンスも向上させます。SRA は、物理リソースを含むクラスタ全体のすべてのリソースを処理する必要がないため、 SVM を対象とした接続を使用する場合は本質的にパフォーマンスが向上します。その代わり、特定の SVM に抽象化された論理資産だけを認識する必要があります。

NAS プロトコルのみを使用する（ SAN アクセスなし）場合は、次のパラメータを設定することで、 NAS 向けに最適化された新しいモードを利用することもできます（ SRA と VASA は、アプライアンスで同じバックエンドサービスを使用するため）。

1. コントロールパネルにログインします。 `https://<IP address>:9083` [Web based CLI interface]をクリックします。
2. コマンドを実行します `vp updateconfig -key=enable.qtree.discovery -value=true`。

3. コマンドを実行します `vp updateconfig -key=enable.optimised.sra -value=true`。
4. コマンドを実行します `vp reloadconfig`。

VVOL に **ONTAP** ツールを導入する際の考慮事項について説明します

SRM で VVol を使用する場合は、クラスタを対象としたクレデンシャルとクラスタ管理 LIF を使用してストレージを管理する必要があります。これは、 VM ストレージポリシーに必要なポリシーを満たすためには、VASA Provider で基盤となる物理アーキテクチャを理解しておく必要があるためです。たとえば、オールフラッシュストレージを必要とするポリシーが設定されている場合、 VASA Provider では、どのシステムがオールフラッシュであるかを認識できる必要があります。

ONTAP Tools アプライアンスを管理している VVOL データストアに格納しないことを推奨します。その結果、アプライアンスがオフラインのためにアプライアンスのスワップ VVOL を作成できず、 VASA Provider の電源をオンにできなくなることがあります。

ONTAP 9 システムの管理に関するベストプラクティス

前述したように、クラスタまたは SVM を対象としたクレデンシャルと管理 LIF を使用して ONTAP クラスタを管理できます。パフォーマンスを最適化するには、VVOLを使用しないときは常にSVMを対象としたクレデンシャルの使用を検討してください。ただし、その場合は、いくつかの要件について確認しておく必要があります。また、機能の一部は失われます。

- デフォルトの vsadmin SVM アカウントには、 ONTAP ツールのタスクを実行するために必要なアクセスレベルがありません。そのため、新しい SVM アカウントを作成する必要があります。
- ONTAP 9.8以降を使用している場合はNetApp、ONTAP System Managerの[Users]メニューとONTAP toolsアプライアンスにあるjsonファイルを使用して、RBACの最小権限を持つユーザアカウントを作成することを推奨します。 `https://<IP address>:9083/vsc/config/`。管理者パスワードを使用して JSON ファイルをダウンロードしてください。これは SVM またはクラスタを対象としたアカウントに使用できます。

  ONTAP 9.6 以前を使用している場合は、で使用可能な RBAC User Creator （ RUC ）ツールを使用する必要があります "NetApp Support Siteの Toolchest"。

- vCenter UI プラグイン、 VASA Provider 、 SRA サーバはすべて完全に統合されたサービスであるため、vCenter UI で ONTAP ツール用のストレージを追加する場合と同じ方法で、 SRM で SRA アダプタにストレージを追加する必要があります。そうしないと、 SRA サーバが SRA アダプタ経由で SRM から送信された要求を認識しない可能性があります。
- SVM を対象としたクレデンシャルを使用している場合、 NFS パスのチェックは実行されませんこれは、物理的な場所が SVM から論理的に抽象化されているためです。ただしこれは原因の問題ではありません。最新の ONTAP システムで間接パスを使用してもパフォーマンスが著しく低下することはなくなりました。
- Storage Efficiency によるアグリゲートのスペース削減量が報告されないことがあります。
- サポートされている場合、負荷共有ミラーを更新することはできません。
- SVM を対象としたクレデンシャルで管理されている ONTAP システムでは、 EMS ロギングが実行されない場合があり

運用上のベストプラクティス

以降のセクションでは、VMware SRMとONTAPストレージの運用に関するベストプラク

ティスについて説明します。

データストアおよびプロトコル

- 可能であれば、必ず ONTAP ツールを使用してデータストアとボリュームをプロビジョニングしてください。ボリューム、ジャンクションパス、 LUN 、 igroup 、エクスポートポリシーが その他の設定は互換性のある方法で構成されます。
- SRM では、 ONTAP 9 で iSCSI 、ファイバチャネル、および NFS バージョン 3 をサポートしているのは、 SRA 経由のアレイベースのレプリケーションを使用している場合です。SRM は、従来のデータストアまたは VVOL データストアでの NFS バージョン 4.1 のアレイベースのレプリケーションをサポートしていません。
- 接続を確認するために、 DR サイトの新しいテスト用データストアをデスティネーション ONTAP クラスタからマウントしてアンマウントできることを必ず確認してください。データストアの接続に使用する各プロトコルをテストします。テスト用データストアは SRM の指示に従ってすべてのデータストアの自動化を実行するため、 ONTAP ツールを使用して作成することを推奨します。
- SAN プロトコルは各サイトで同機種にする必要があります。NFS と SAN を混在させることはできますが、 SAN プロトコルを 1 つのサイト内に混在させないでください。たとえば、サイト A では FCP を、サイト B では iSCSI を使用できますサイト A では、 FCP と iSCSI の両方を使用しないでくださいその理由は、 SRA がリカバリサイトに混在する igroup を作成しないため、 SRM が SRA に指定されたイニシエータリストをフィルタリングしないためです。
- 以前のガイドでは、データの局所性にLIFを作成することを推奨つまり、必ず、ボリュームを物理的に所有するノード上の LIF を使用してデータストアをマウントします。これは、 ONTAP 9 の最新バージョンでは必須ではなくなりました。可能なかぎり、クラスタを対象としたクレデンシャルを指定した場合でも、ONTAPツールではデータに対してローカルなLIF間で負荷を分散するように選択されますが、高可用性やパフォーマンスを確保するための必須要件ではありません。
- ONTAP 9では、オートサイズが緊急時に十分な容量を提供できない場合に、スペース不足が発生したときにSnapshotを自動的に削除してアップタイムを維持するように設定できます。この機能のデフォルト設定では、SnapMirrorで作成されたSnapshotは自動的に削除されません。SnapMirror Snapshotが削除されると、NetApp SRAは影響を受けたボリュームのレプリケーションを反転および再同期できません。ONTAPがSnapMirror Snapshotを削除しないようにするには、Snapshotの自動削除機能をtryに設定します。

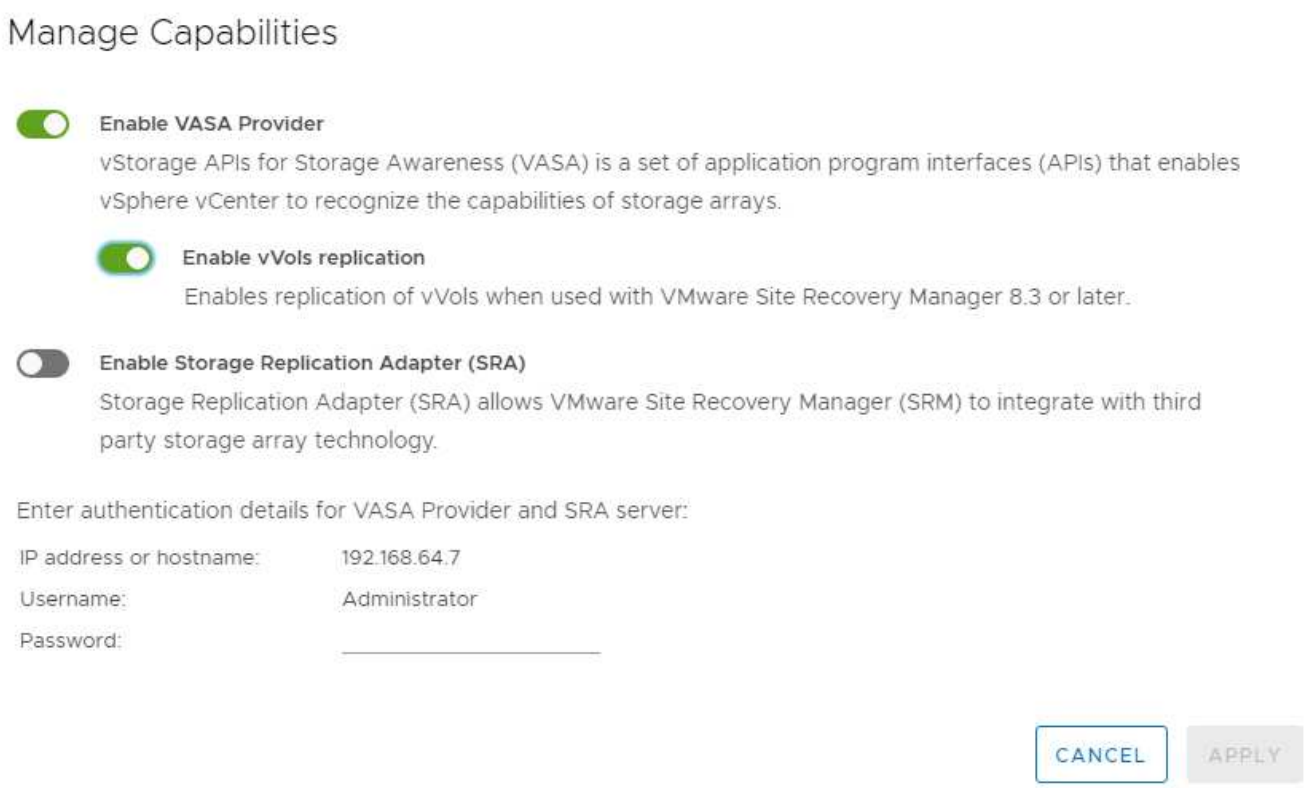
```
snap autodelete modify -volume -commitment try
```

- ボリュームのオートサイズの設定： `grow` SANデータストアヲフクムボリューム `grow_shrink` （NFSデータストアの場合）。の詳細を確認してください "ボリュームを自動的に拡張または縮小するための設定"。
- SRMは、データストアの数が少なく、保護グループがリカバリプランで最小化されている場合に最適なパフォーマンスを発揮します。したがって、RTOが重要なSRMで保護された環境では、VM密度の最適化を検討する必要があります。
- Distributed Resource Scheduler（DRS）を使用して、保護対象のESXiクラスタとリカバリESXiクラスタの負荷を分散します。フェイルバックを計画している場合、再保護を実行すると、以前に保護されていたクラスタが新しいリカバリクラスタになります。DRSは、両方向への配置のバランスをとるのに役立ちます。
- SRMでIPカスタマイズを使用するとRTOが増加する可能性があるため、可能な場合は使用しないでください。

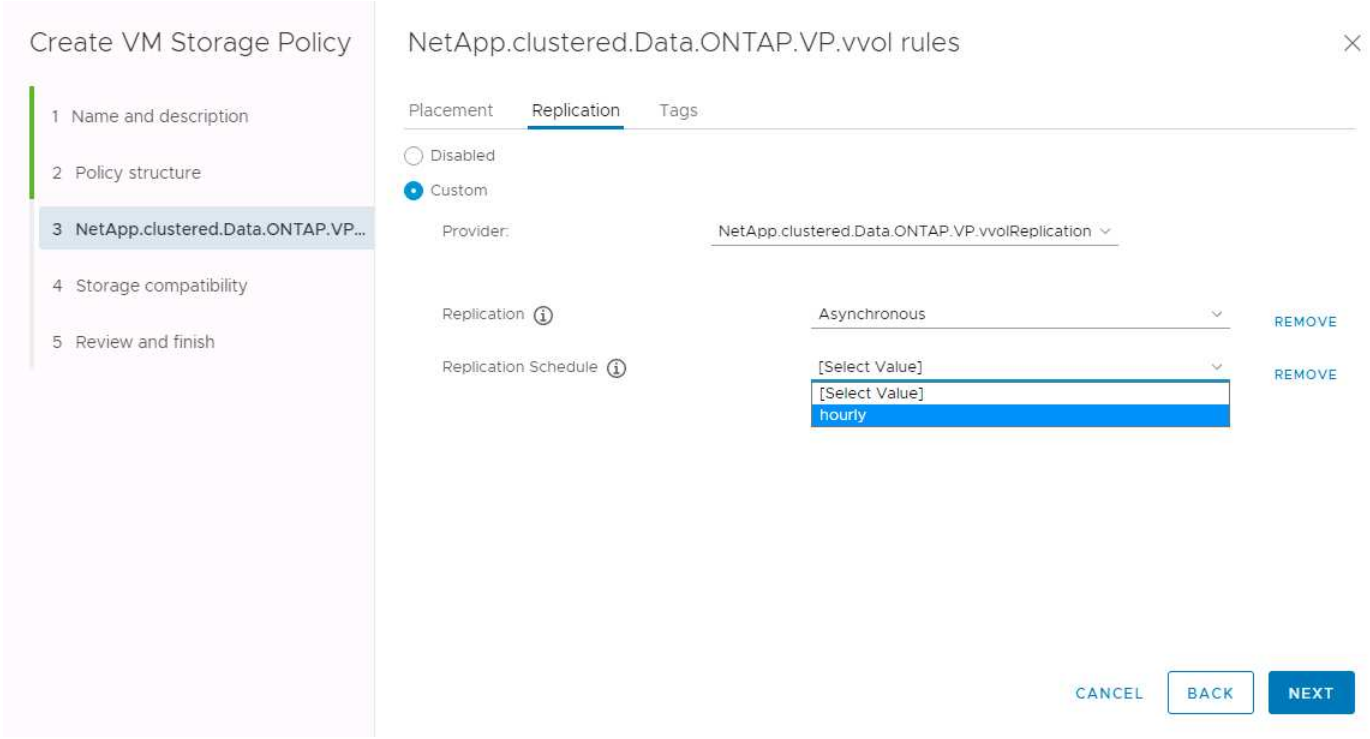
Storage Policy Based Management（SPBM；ストレージポリシーベースの管理）とVVOL

SRM 8.3以降では、vVolデータストアを使用したVMの保護がサポートされます。SnapMirror スケジュールは、次のスクリーンショットに示すように、 ONTAP のツール設定メニューで VVOL のレプリケーションが有効になっている場合、 VASA Provider によって VM ストレージポリシーに公開されます。

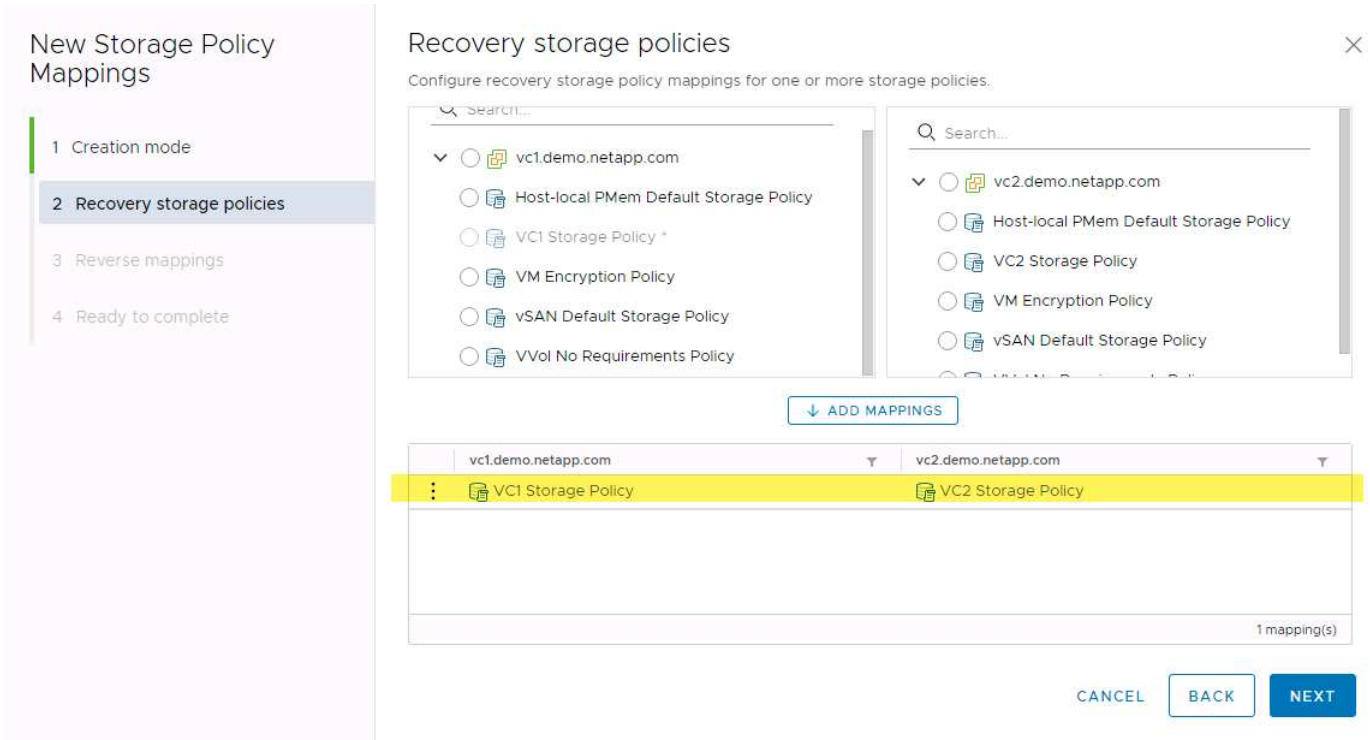
次の例は、vVolレプリケーションを有効にする方法を示しています。



次のスクリーンショットは、 VM ストレージポリシーの作成ウィザードに表示される SnapMirror スケジュールの例を示しています。



ONTAP VASA プロバイダでは、異なるストレージへのフェイルオーバーがサポートされます。たとえば、システムは、エッジの場所にある ONTAP Select からコアデータセンターの AFF システムにフェイルオーバーできます。ストレージの類似性に関係なく、レプリケーションが有効な VM ストレージポリシーのストレージポリシーマッピングとリバースマッピングを常に設定して、リカバリサイトで提供されるサービスが期待される要件を満たしていることを確認する必要があります。次のスクリーンショットは、ポリシーマッピングの例を示しています。



VVOL データストア用にレプリケートされたボリュームを作成します

以前の VVOL データストアとは異なり、レプリケートされた VVOL データストアはレプリケーションを有効にして最初から作成する必要があります。また、 SnapMirror 関係を持つ ONTAP システムで事前に作成されたボリュームを使用する必要があります。そのためには、クラスタピアリングや SVM ピアリングなどの設定を事前に行う必要があります。これらのアクティビティはONTAP管理者が実行する必要があります。これにより、複数のサイトでONTAPシステムを管理する担当者とvSphereの運用を主に担当する担当者が厳密に分離されます。

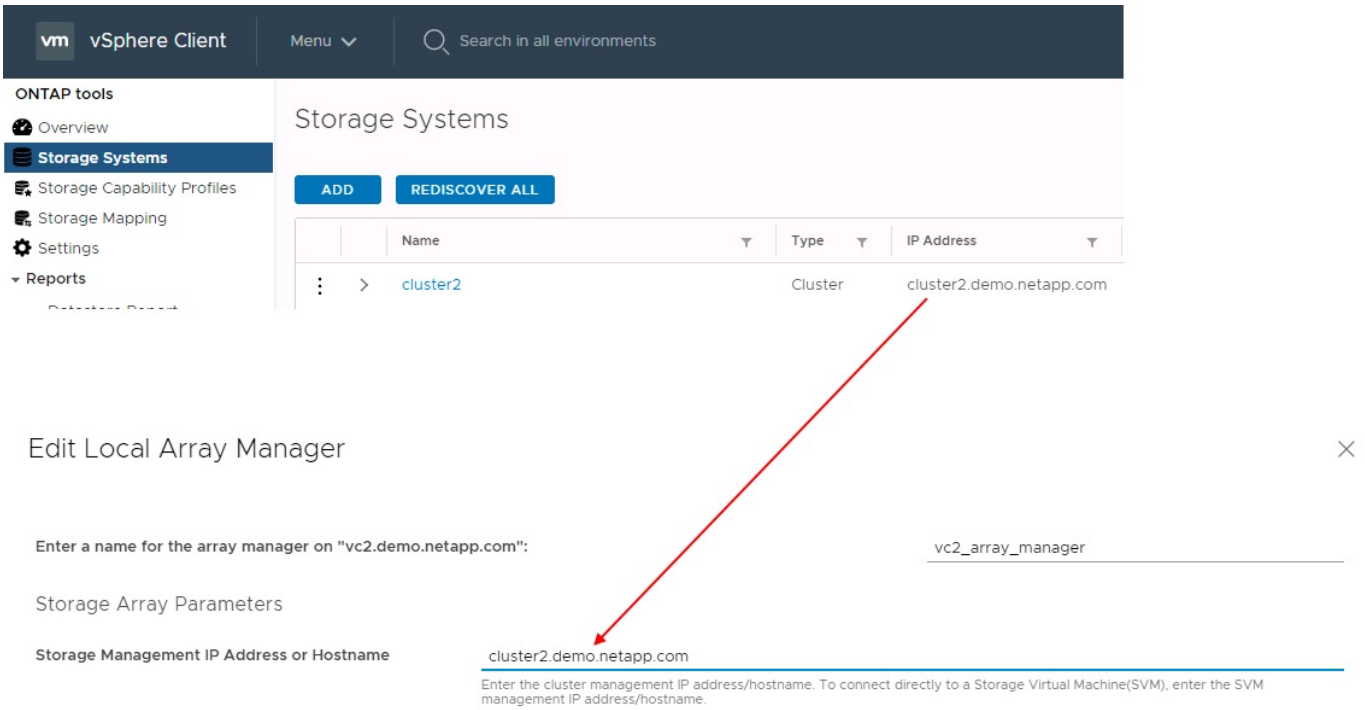
これは、 vSphere 管理者の代わりに新たな要件となります。ボリュームは ONTAP ツールの範囲外に作成されるため、定期的な再検出スケジュール期間が設定されるまで ONTAP 管理者が行った変更を認識することはありません。そのため、 VVOL で使用するボリュームまたは SnapMirror 関係を作成したときは常に再検出を実行することを推奨します。次のスクリーンショットに示すように、ホストまたはクラスタを右クリックし、ONTAP tools]>[Update Host and Storage Data]を選択します。

VVOL と SRM については、 1 つ注意が必要です。保護された VM と保護されていない VM を同じ VVOL データストアに混在させないでください。これは、 SRM を使用して DR サイトにフェイルオーバーする場合、保護グループに属する VM のみが DR でオンラインになるためです。そのため、再保護（ SnapMirror を DR から本番環境に戻して再保護）する際に、フェイルオーバーされなかった VM が上書きされて、貴重なデータが含まれる可能性があります。

アレイペアについて

アレイペアごとにアレイマネージャが作成されます。SRM ツールと ONTAP ツールでは、クラスタクレデンシャルを使用している場合でも、各アレイペアリングを SVM の範囲で実行します。これにより、管理対象に割り当てられている SVM を基に、各テナント間で DR ワークフローを分割できます。特定のクラスタに対して複数のアレイマネージャを作成し、非対称にすることができます。異なる ONTAP 9 クラスタ間でファンアウトまたはファンインを実行できます。たとえば、クラスタ 1 の SVM A と SVM B をクラスタ 2 の SVM C に、クラスタ 3 の SVM D に、またはその逆にレプリケートできます。

SRM でアレイペアを設定する場合は、 ONTAP ツールに追加するのと同じ方法でアレイペアを SRM に追加する必要があります。つまり、アレイペアは同じユーザ名、パスワード、および管理 LIF を使用する必要があります。これは、 SRA がアレイと正しく通信するための要件です。次のスクリーンショットは、 ONTAP ツールでのクラスタの表示方法と、アレイマネージャへのクラスタの追加方法を示しています。



複製グループについて

レプリケーショングループには、同時にリカバリされる仮想マシンの論理集合が含まれます。レプリケーショングループは、 ONTAP ツール VASA Provider で自動的に作成されます。ONTAP の SnapMirror レプリケーションはボリュームレベルで実行されるため、ボリューム内のすべての VM が同じレプリケーショングループに属します。

レプリケーショングループについて考慮する必要がある要素と、 FlexVol ボリュームに VM を分散する方法にはいくつかの要素があります。類似するVMを同じボリュームにグループ化すると、アグリゲートレベルの重複排除機能がない古いONTAPシステムでストレージ効率を高めることができますが、グループ化するとボリュームのサイズが大きくなり、ボリュームのI/Oの同時実行数が少なくなります。最新のONTAPシステムでは、同じアグリゲート内のFlexVolボリュームにVMを分散することで、パフォーマンスとストレージ効率の最適なバランスを実現できます。その結果、アグリゲートレベルの重複排除が活用され、複数のボリューム間でI/Oの並列化が促進されます。保護グループ（以下で説明）には複数のレプリケーショングループを含めることができるため、ボリューム内の VM を 1 つにまとめてリカバリできます。このレイアウトの欠点は、Volume SnapMirrorではアグリゲートの重複排除が考慮されないため、ブロックがネットワーク経由で複数回送信される可能性があることです。

レプリケーショングループの最後の考慮事項の 1 つは、各グループがその性質によって論理整合グループになることです（ SRM 整合グループと混同しないようにしてください）。これは、ボリューム内のすべての VM が同じ Snapshot を使用して同時に転送されるためです。したがって、相互に整合性が必要な VM がある場合は、同じ FlexVol に格納することを検討してください。

保護グループについて

保護グループでは、 VM とデータストアをグループ単位で定義し、グループをまとめて保護サイトからリカバリします。保護対象サイトとは、通常の安定状態での運用中、保護グループで構成された VM が存在する場所です。SRM には保護グループの複数のアレイマネージャが表示される場合がありますが、保護グループは複数のアレイマネージャにまたがることはできません。このため、異なる SVM 上の複数のデータストアに VM ファイルをまたがって配置することはできません。

リカバリ・プランについて

リカバリプランでは、同じプロセスでリカバリする保護グループを定義します。同じリカバリプランに複数の保護グループを設定できます。また、リカバリプランの実行オプションを増やすには、 1 つの保護グループを複数のリカバリプランに含めることもできます。

リカバリプランを使用すると、 SRM 管理者は、 VM を優先グループ 1 （最大）から 5 （最小）に割り当てて、リカバリワークフローを定義できます。デフォルトは 3 （中）です。優先度グループ内で、 VM に依存関係を設定できます。

たとえば、データベースにMicrosoft SQL Serverを使用するティア1のビジネスクリティカルなアプリケーションがあるとします。したがって、優先度グループ 1 に VM を配置することにします。優先度グループ 1 では、サービスの提供順序の計画を開始します。Microsoft Windows ドメイン・コントローラを起動してから Microsoft SQL Server を起動してください。アプリケーション・サーバの前にオンラインになっている必要があります。依存関係は特定の優先度グループ内でのみ適用されるため、これらすべてのVMを優先度グループに追加してから依存関係を設定します。

アプリケーションチームと連携してフェイルオーバーシナリオに必要な処理の順序を把握し、それに応じてリカバリ計画を作成することを強く推奨します。

テストフェイルオーバー

ベストプラクティスとして、保護対象の VM ストレージの構成を変更する場合は、必ずテストフェイルオーバーを実行してください。これにより、災害が発生した場合に、Site Recovery Managerが予想されるRTOターゲット内でサービスをリストアできると信頼できます。

特に VM ストレージの再設定後にゲストアプリケーションの機能を確認することを推奨します。

テストリカバリ処理を実行すると、 VM 用の ESXi ホストにプライベートテスト用のバブルネットワークが作成されます。ただし、このネットワークは物理ネットワークアダプタに自動的には接続されないため、 ESXi ホスト間の接続は提供されません。DR テスト時に異なる ESXi ホストで実行されている VM 間の通信を可能にするために、 DR サイトの ESXi ホスト間に物理プライベートネットワークを作成します。テスト用ネットワークがプライベートであることを確認するために、テスト用のバブルネットワークを物理的に分離するか、 VLAN や VLAN タギングを使用して分離します。このネットワークは本番用ネットワークから分離する必要があります。 VM がリカバリされると、実際の本番用システムと競合する可能性のある IP アドレスを持つ本番用ネットワークに配置することはできなくなります。SRM でリカバリプランを作成する際、テスト中に VM を接続するためのプライベートネットワークとして、作成したテストネットワークを選択できます。

テストが検証されて不要になったら、クリーンアップ処理を実行します。クリーンアップを実行すると、保護されている VM が初期状態に戻り、リカバリプランが Ready 状態にリセットされます。

フェイルオーバーに関する考慮事項

サイトのフェイルオーバーに関しては、このガイドに記載されている処理の順序に加えて、その他にもいくつかの考慮事項があります。

競合する問題の 1 つに、サイト間のネットワークの違いがあります。環境によっては、プライマリサイトと DR サイトで同じネットワーク IP アドレスを使用できる場合があります。この機能は、拡張仮想 LAN （ VLAN ）または拡張ネットワークセットアップと呼ばれます。それ以外の環境では、プライマリサイトと DR サイトで別々のネットワーク IP アドレス（異なる VLAN など）を使用する必要があります。

VMware では、この問題を解決する方法をいくつか提供しています。1 つは、 VMware NSX -T Data Center のようなネットワーク仮想化テクノロジーです。ネットワークスタック全体を運用環境からレイヤ 2 ～ 7 に

抽象化し、より移植性の高いソリューションを実現します。の詳細を確認してください "SRMでのNSX-Tオプション"。

SRM では、リカバリ時に VM のネットワーク設定を変更することもできます。この再設定には、IPアドレス、ゲートウェイアドレス、DNSサーバ設定などの設定が含まれます。リカバリ時に個 々 のVMに適用されるさまざまなネットワーク設定は、リカバリプランのVMのプロパティ設定で指定できます。

VMware の dr-ip-customizer というツールを使用すると、リカバリプランで複数の VM のプロパティを個別に編集しなくても、 SRM で VM ごとに異なるネットワーク設定を適用できます。このユーティリティの使用方法については、を参照してください。 "VMwareのドキュメント"。

再保護

リカバリ後、リカバリサイトが新しい本番用サイトになります。リカバリ処理によって SnapMirror レプリケーションが解除されたため、新しい本番用サイトは今後の災害から保護されません。新しい本番用サイトは、リカバリ後すぐに別のサイトで保護することを推奨します。元の本番サイトが運用されている場合、 VMware 管理者は、元の本番サイトを新しいリカバリサイトとして使用して新しい本番サイトを保護できるため、保護の方向を実質的に変えることができます。再保護は、致命的でない障害でのみ使用できます。そのため、元の vCenter Server 、 ESXi サーバ、 SRM サーバ、および対応するデータベースを最終的にリカバリ可能な状態にする必要があります。使用できない場合は、新しい保護グループと新しいリカバリプランを作成する必要があります。

フェイルバック

フェイルバック処理は、基本的に以前とは異なる方向のフェイルオーバーです。ベストプラクティスとして、フェイルバックを実行する前に、元のサイトが許容可能なレベルの機能に戻っていること、つまり元のサイトにフェイルオーバーしていることを確認することを推奨します。元のサイトが侵害されたままの場合は、障害が十分に修正されるまでフェイルバックを遅らせる必要があります。

フェイルバックのもう 1 つのベストプラクティスとして、再保護の完了後、および最終フェイルバックの実行前に、常にテストフェイルオーバーを実行することがあります。これにより、元のサイトに配置されたシステムで処理が完了できるかどうかを確認できます。

元のサイトを再保護する

フェイルバック後、再保護を再度実行する前に、すべての関係者にサービスが正常に戻ったことを確認する必要があります。

フェイルバック後の再保護を実行すると、基本的に環境は最初の状態に戻り、 SnapMirror レプリケーションが本番用サイトからリカバリサイトに再度実行されます。

レプリケーショントポロジ

ONTAP 9 では、クラスタの物理コンポーネントはクラスタ管理者には見えますが、クラスタを使用しているアプリケーションやホストからは直接見えません。物理コンポーネントは共有リソースのプールを提供し、このリソースプールから論理クラスタリソースが構築されます。アプリケーションとホストは、ボリュームと LIF を含む SVM 経由でのみデータにアクセスします。

VMware vCenter Site Recovery Manager では、各 NetApp SVM がアレイとして扱われます。SRM は、特定のアレイ間（または SVM から SVM ）のレプリケーションレイアウトをサポートしています。

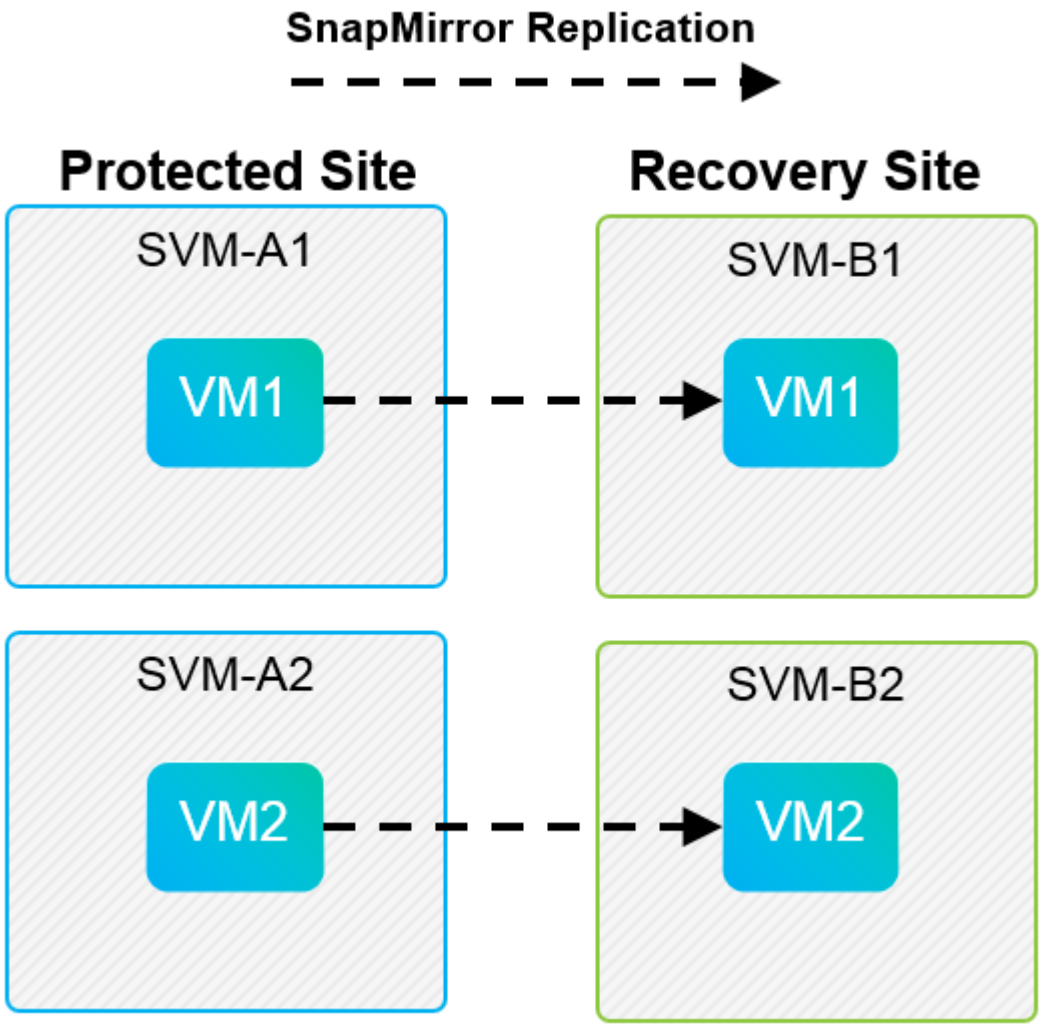
1 つの VM が、次の理由から、複数の SRM アレイ上で仮想マシンディスク（ VMDK ）または RDM を所有することはできません。

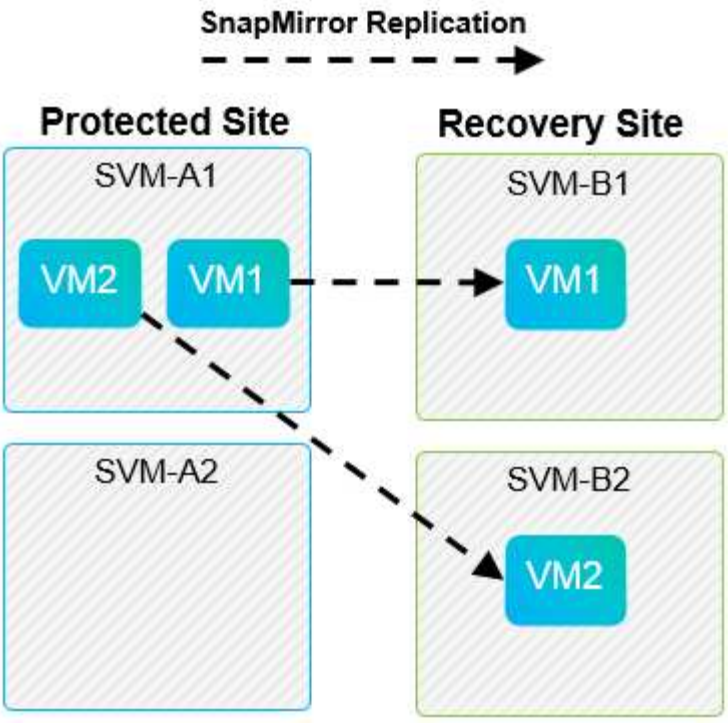
- SRM は SVM のみを認識し、個々の物理コントローラは認識しません。
- SVM は、クラスタ内の複数のノードにまたがる LUN とボリュームを制御できます。

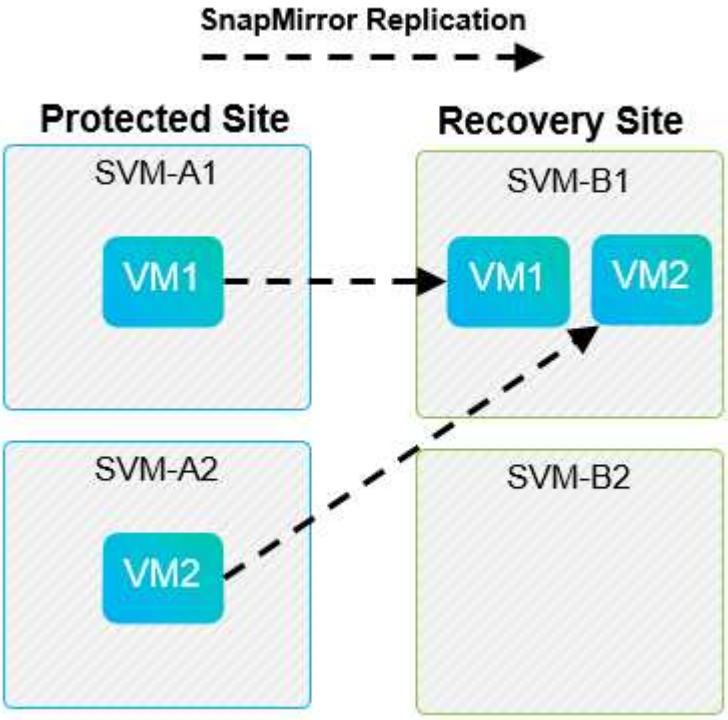
ベストプラクティス
サポートされるかどうかを判断するには、このルールに注意してください。 SRM と NetApp SRA を使用して VM を保護するには、 VM のすべての部分が 1 つの SVM 上にのみ存在する必要があります。このルールは、保護対象サイトとリカバリサイトの両方に適用されます。

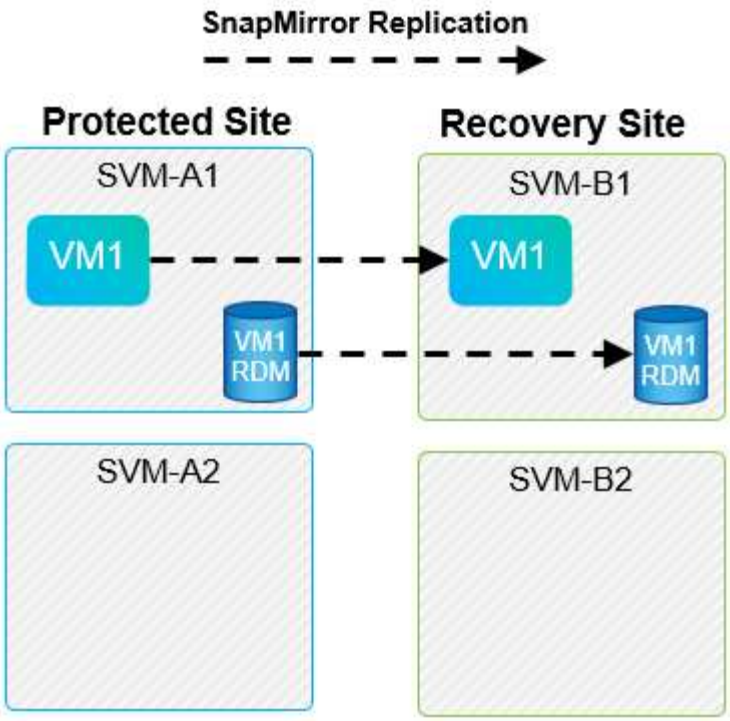
サポートされる **SnapMirror** レイアウト

次の図は、 SRM と SRA でサポートされる SnapMirror 関係のレイアウトシナリオを示しています。レプリケートされたボリューム内の各 VM は、各サイトの 1 つの SRM アレイ（ SVM ）上のデータのみを所有します。



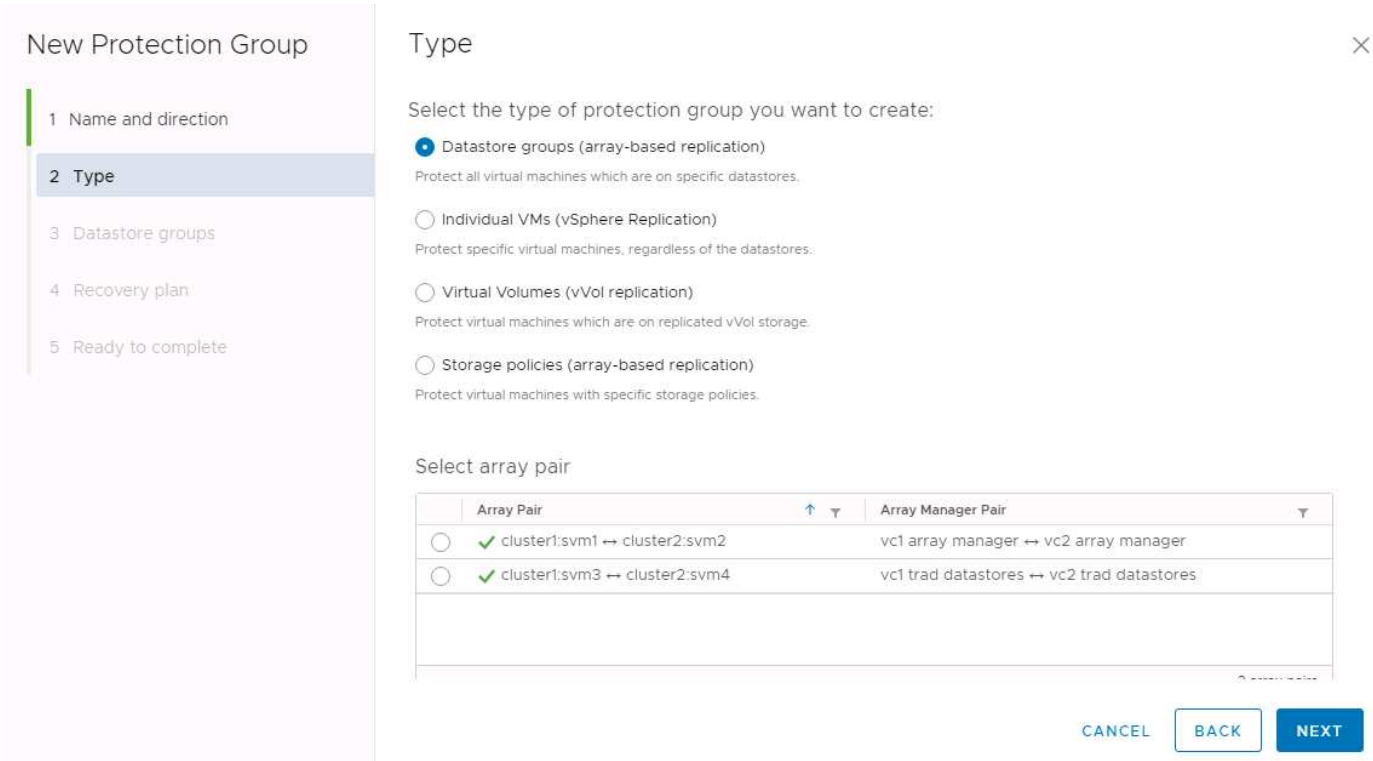

SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
SVM-B1
VM2
VM1
VM1
SVM-A2
SVM-B2
VM2


SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
SVM-B1
VM1
VM1
VM2
SVM-A2
SVM-B2
VM2



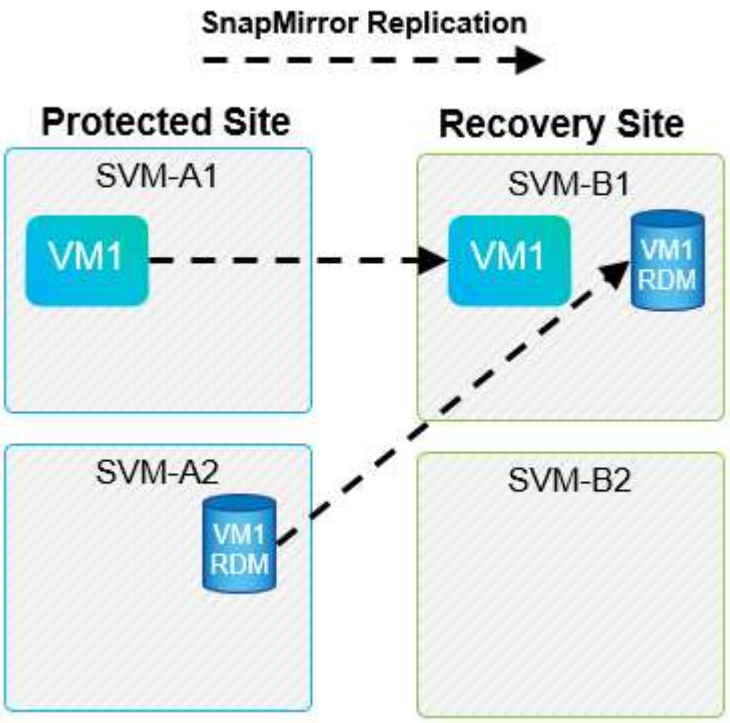
サポートされている **Array Manager** レイアウト

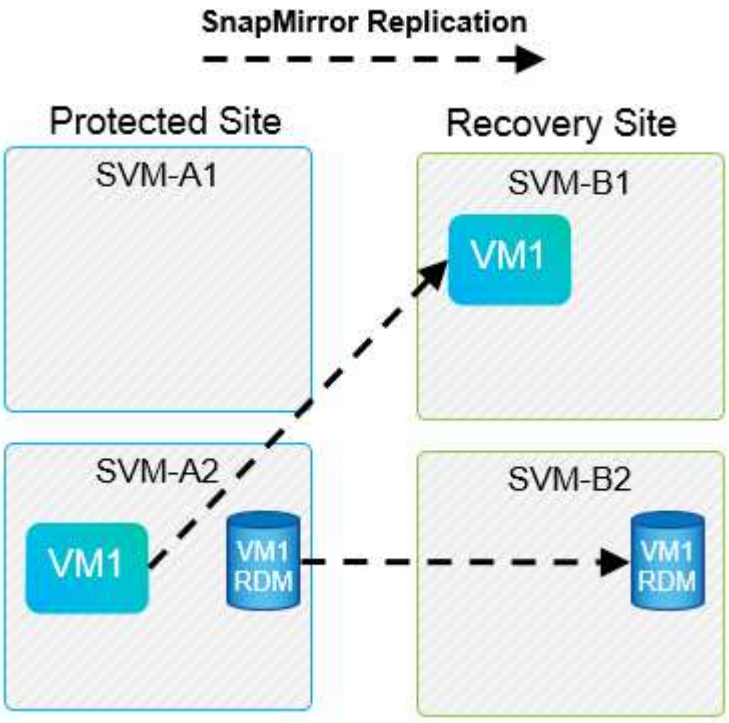
次のスクリーンショットに示すように、 SRM でアレイベースレプリケーション（ ABR ）を使用すると、保護グループは単一のアレイペアに分離されます。このシナリオでは、 `SVM1` および `SVM2` ピア関係を設定する `SVM3` および `SVM4` リカバリサイトで。ただし、保護グループを作成するときに選択できるアレイペアは 2 つのうちの 1 つだけです。



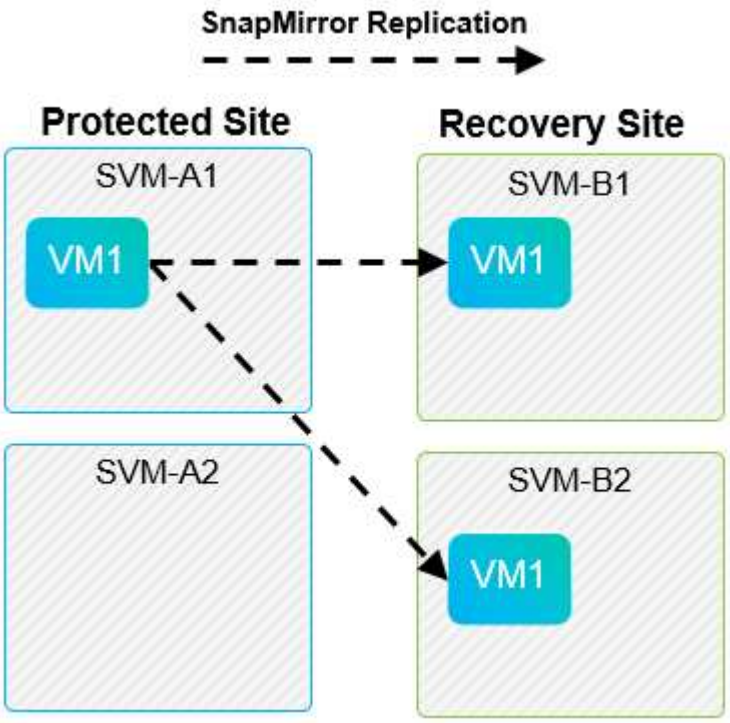
サポートされないレイアウトです

サポート対象外の構成では、個々の VM が所有する複数の SVM にデータ（ VMDK または RDM ）があります。次の図に示す例では、 `VM1` SRMで保護を設定できません。理由： `VM1` 2つのSVM上のデータがあります。



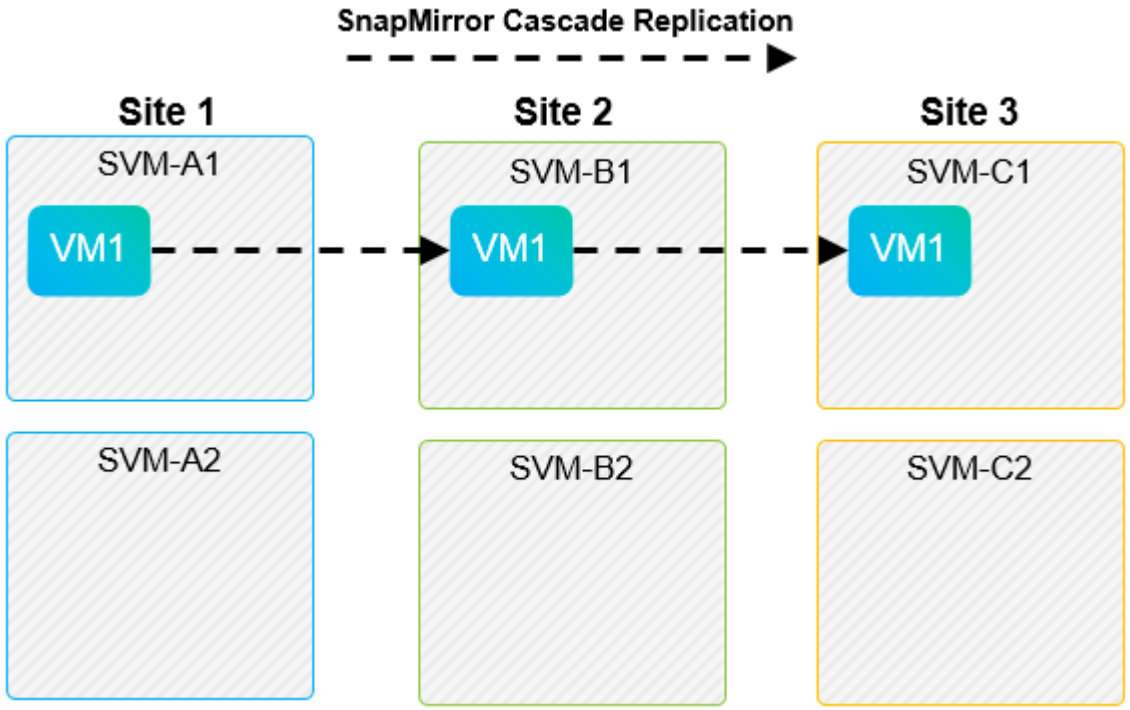


1 つのネットアップボリュームを 1 つのソース SVM から同じ SVM または異なる SVM の複数のデスティネーションにレプリケートするレプリケーション関係は、 SnapMirror ファンアウトと呼ばれます。SRM ではファンアウトはサポートされていません。次の図の例では、 `VM1` SnapMirrorを使用して2つの異なる場所にレプリケートされるため、SRMで保護を設定できません。



SnapMirror カスケード

SnapMirror でソースボリュームをデスティネーションボリュームにレプリケートし、そのデスティネーションボリュームを SnapMirror で別のデスティネーションボリュームにレプリケートする SnapMirror 関係のカスケードを、 SRM ではサポートしていません。次の図に示すシナリオでは、 SRM を使用してサイト間のフェイルオーバーを実行することはできません。



SnapMirror と SnapVault

NetApp SnapVault ソフトウェアを使用すると、ネットアップストレージシステム間でエンタープライズデータをディスクベースでバックアップできます。SnapVault と SnapMirror は同じ環境内に共存できますが、SRM でサポートされているのは、 SnapMirror 関係のフェイルオーバーだけです。

NetApp SRAは、 `mirror-vault` ポリシータイプ。

SnapVault は ONTAP 8.2 で一から再構築されました。以前の Data ONTAP 7-Mode で使用されていたユーザは共通点に注意する必要がありましたが、このバージョンの SnapVault では主に拡張機能が追加されています。大きな進歩の 1 つは、 SnapVault 転送時にプライマリデータの Storage Efficiency を維持できることです。

アーキテクチャの重要な変更点は、 7-Mode SnapVault の場合と同様に、 ONTAP 9 の SnapVault でも qtree レベルではなくボリュームレベルでレプリケートされる点です。つまり、 SnapVault 関係のソースはボリュームでなければならず、そのボリュームは SnapVault セカンダリシステム上の独自のボリュームにレプリケートされる必要があります。

SnapVaultを使用する環境では、プライマリストレージシステム上に特別な名前のスナップショットが作成されます。実装されている構成に応じて、SnapVaultスケジュールまたはNetApp Active IQ Unified Managerなどのアプリケーションを使用して、名前付きSnapshotをプライマリシステムに作成できます。プライマリシステムで作成された名前付きSnapshotがSnapMirrorデスティネーションにレプリケートされ、そこからSnapVaultデスティネーションに保存されます。

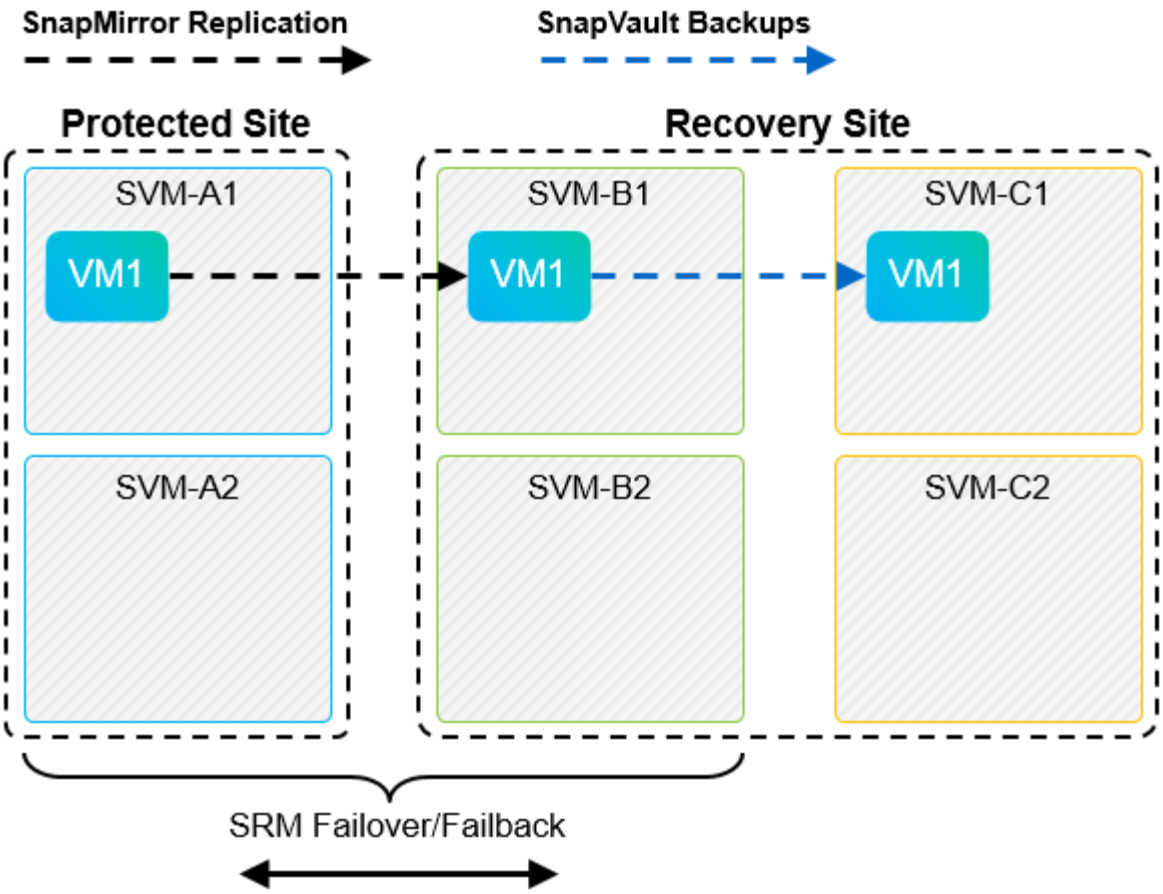
ソースボリュームは、ボリュームが DR サイトの SnapMirror デスティネーションにレプリケートされるカスケード構成で作成でき、そこから SnapVault デスティネーションに保存されます。ファンアウト関係では、一方のデスティネーションが SnapMirror デスティネーション、もう一方が SnapVault デスティネーションであるソースボリュームも作成できます。ただし、 SRM フェイルオーバーまたはレプリケーションの反転時に、SRA は、 SnapMirror デスティネーションボリュームを SnapVault のソースとして使用するように SnapVault 関係を自動では再設定しません。

SnapMirror および SnapVault for ONTAP 9 の最新情報については、を参照してください "TR-4015 ：『 SnapMirror Configuration Best Practice Guide for ONTAP 9 』"

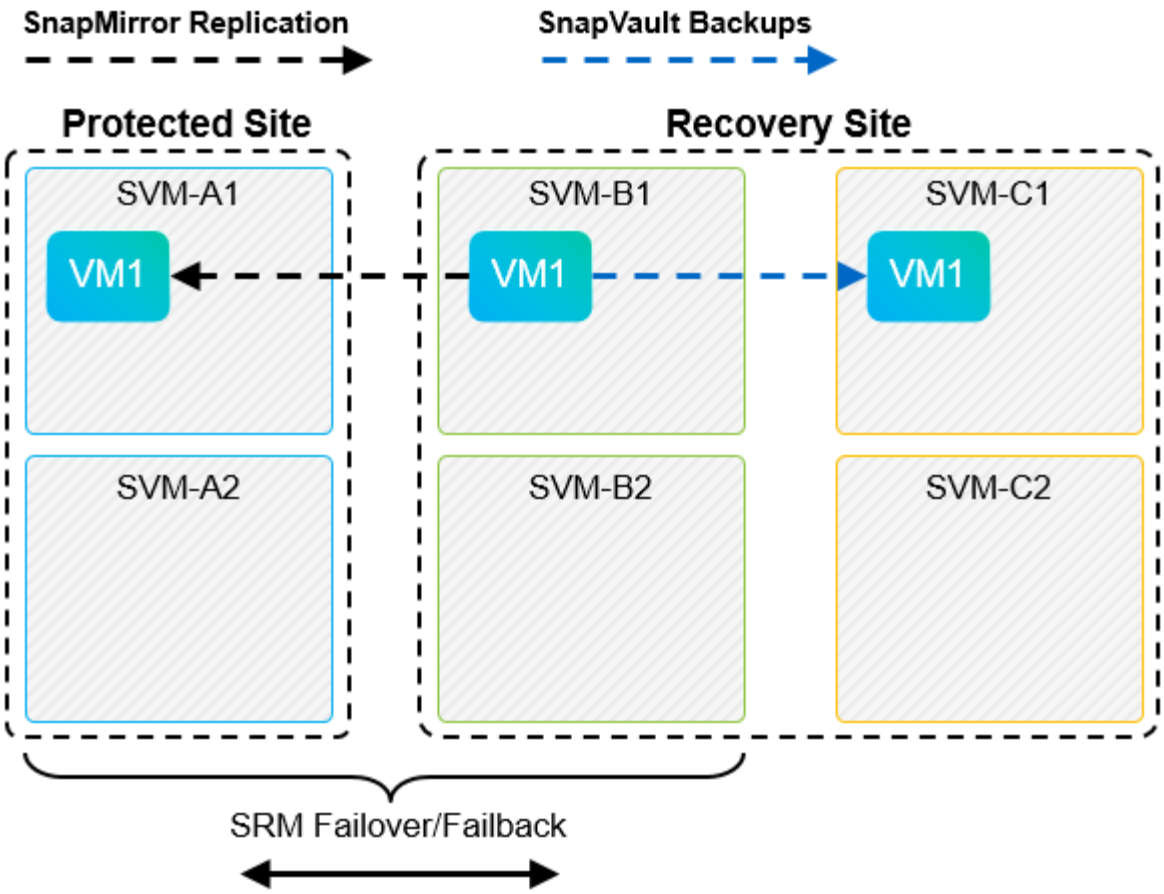
ベストプラクティス
SnapVault と SRM を同じ環境で使用する場合、通常は DR サイトの SnapMirror デスティネーションから SnapVault バックアップを実行する、 SnapMirror から SnapVault へのカスケード構成を使用することを推奨します。災害が発生すると、この構成によってプライマリサイトにアクセスできなくなります。リカバリサイトに SnapVault デスティネーションを配置すると、フェイルオーバー後に SnapVault バックアップを再設定して、リカバリサイトで SnapVault バックアップを継続できるようになります。

VMware 環境では、各データストアに Universal Unique Identifier （ UUID ）が割り当てられ、各 VM には一意の Managed Object ID （ MOID ）が割り当てられます。SRM は、フェイルオーバーやフェイルバックの実行時にこれらの ID を維持しません。SRM はフェイルオーバーでデータストア UUID と VM MOID を維持しないため、これらの ID に依存するアプリケーションは SRM フェイルオーバーのあとに再設定する必要があります。たとえば、 SnapVault レプリケーションを vSphere 環境と調整する NetApp Active IQ Unified Manager などがあります。

次の図に、 SnapMirror から SnapVault へのカスケード構成を示します。SnapVault デスティネーションがプライマリサイトの停止の影響を受けない DR サイトまたは第 3 のサイトにある場合、フェイルオーバー後にバックアップを続行できるように環境を再設定できます。



次の図は、 SRM を使用して SnapMirror レプリケーションをプライマリサイトに反転したあとの構成を示しています。SnapMirror ソースから SnapVault バックアップが実行されるように環境が再設定されている。このセットアップは、 SnapMirror SnapVault のファンアウト構成です。



SRM でフェイルバックを実行し、 SnapMirror 関係が再度反転されると、本番環境のデータはプライマリサイトに戻ります。SnapMirror と SnapVault のバックアップにより、 DR サイトへのフェイルオーバー前と同じ方法でこのデータを保護できるようになりました。

Site Recovery Manager 環境での **qtree** の使用

qtree は、 NAS のファイルシステムクォータを適用可能な特殊なディレクトリです。ONTAP 9 では qtree を作成でき、 SnapMirror でレプリケートされたボリュームに配置できます。ただし、 SnapMirror では、個々の qtree のレプリケーションまたは qtree レベルのレプリケーションは実行できません。すべての SnapMirror レプリケーションは、ボリュームレベルで実行されます。このため、 SRM で qtree を使用することは推奨されません。

FC と **iSCSI** の混在環境

サポート対象の SAN プロトコル（ FC 、 FCoE 、 iSCSI ）の場合、 ONTAP 9 は LUN サービスを提供します。 LUN サービスの提供とは、 LUN を作成して、接続されているホストにマッピングする機能です。クラスタは複数のコントローラで構成されるため、個々の LUN へのマルチパス I/O で管理される論理パスが複数あります。ホスト上で Asymmetric Logical Unit Access （ ALUA ；非対称論理ユニットアクセス）が使用されるため、 LUN への最適なパスが選択され、データ転送用にアクティブになります。LUN への最適パスが変わった場合（格納先ボリュームが移動された場合など）、 ONTAP 9 は自動的にこの変更を認識し、システムを停止することなく調整します。最適パスが利用できなくなった場合、 ONTAP は無停止で他の利用可能なパスに切り替えることができます。

VMware SRM と NetApp SRA の環境では、一方のサイトで FC プロトコルを使用し、もう一方のサイトで iSCSI プロトコルを使用できます。ただし、 FC 接続のデータストアと iSCSI 接続のデータストアを同じ ESXi ホストで混在させたり、同じクラスタ内の別のホストで使用したりすることはできません。この構成は SRM ではサポートされていません。 SRM フェイルオーバーまたはテストフェイルオーバーの実行中、 SRM

は要求に応じて ESXi ホストのすべての FC イニシエータと iSCSI イニシエータを含めます。

ベストプラクティス
SRM と SRA では、保護サイトとリカバリサイト間での FC プロトコルと iSCSI プロトコルの混在をサポートしています。ただし、各サイトで FC または iSCSI のどちらかのプロトコルを 1 つだけ使用し、同じサイトで両方のプロトコルを使用することはできません。1 つのサイトに FC プロトコルと iSCSI プロトコル両方を設定する必要がある場合、一部のホストで iSCSI を使用し、他のホストで FC を使用することを推奨します。また、 VM がどちらか一方のホストグループまたは他方のホストグループにフェイルオーバーするように設定されるように、 SRM リソースマッピングを設定することも推奨します。

VVol レプリケーションを使用する場合の SRM のトラブルシューティング

SRM で VVOL レプリケーションを使用する場合、 SRA と従来のデータストアで使用するワークフローは大きく異なります。たとえば、アレイマネージャの概念はありません。そのため、 `discoverarrays` および `discoverdevices` コマンドは表示されません。

トラブルシューティングを行う場合は、以下に示す新しいワークフローについて理解しておくと役立ちます。

1. queryReplicationPeer ： 2 つのフォールトドメイン間のレプリケーション契約を検出します。
2. queryFaultDomain ：障害ドメインの階層を検出します。
3. queryReplicationGroup ：ソースドメインまたはターゲットドメインに存在するレプリケーショングループを検出します。
4. syncReplicationGroup ：ソースとターゲット間でデータを同期します。
5. queryPointInTimeReplica ：ターゲット上のポイントインタイムレプリカを検出します。
6. testFailoverReplicationGroupStart ：テストフェイルオーバーを開始します。
7. testFailoverReplicationGroupStop ：テストフェイルオーバーを終了します。
8. promoteReplicationGroup ：テスト中のグループを本番環境に昇格します。
9. prepareFailoverReplicationGroup ：災害復旧の準備をします。
10. FailoverReplicationGroup ：ディザスタリカバリを実行します。
11. revertReplicateGroup ：逆方向のレプリケーションを開始します。
12. queryMatchingContainer: 指定されたポリシーを使用したプロビジョニング要求を満たす可能性のあるコンテナを（ホストまたはレプリケーショングループとともに）検索します。
13. queryResourceMetadata ： VASA Provider からすべてのリソースのメタデータを検出し、リソース利用率を回答として queryMatchingContainer 関数に返すことができます。

VVOL レプリケーションの設定時に表示される最も一般的なエラーは、 SnapMirror 関係を検出できないエラーです。これは、ボリュームおよび SnapMirror 関係が ONTAP ツールを対象としたものではないためです。そのため、 SnapMirror 関係が常に完全に初期化されていることを確認し、レプリケートされた VVOL データストアを作成する前に両方のサイトの ONTAP ツールで再検出を実行することを推奨します。

追加情報

このドキュメントに記載されている情報の詳細については、以下のドキュメントや Web

サイトを参照してください。

- TR-4597 ：『 VMware vSphere for ONTAP 』
"https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vsphere-overview.html"
- TR-4400 ：『 VMware vSphere Virtual Volumes with ONTAP 』
"https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vvols-overview.html"
- TR-4015 ：『 SnapMirror Configuration Best Practice Guide for ONTAP 9 』
https://www.netapp.com/media/17229-tr4015.pdf?v=127202175503P
- RBAC User Creator for ONTAP の略
"https://mysupport.netapp.com/site/tools/tool-eula/rbac"
- VMware vSphere リソース用の ONTAP ツール
"https://mysupport.netapp.com/site/products/all/details/otv/docsandkb-tab"
- VMware Site Recovery Manager のドキュメント
"https://docs.vmware.com/en/Site-Recovery-Manager/index.html"

を参照してください "Interoperability Matrix Tool （ IMT ）" NetApp Support Siteで、本ドキュメントに記載されている製品や機能のバージョンがお客様の環境でサポートされるかどうかを確認してください。NetApp IMT には、ネットアップがサポートする構成を構築するために使用できる製品コンポーネントやバージョンが定義されています。サポートの可否は、お客様の実際のインストール環境が公表されている仕様に従っているかどうかによって異なります。

ONTAPを使用したvSphere Metroストレージクラスタ

ONTAPを使用したvSphere Metroストレージクラスタ

VMwareの業界をリードするvSphereハイパーバイザーは、vSphere Metro Storage Cluster（vMSC）と呼ばれるストレッチクラスタとして導入できます。

vMSCソリューションは、NetApp®MetroCluster™とSnapMirrorアクティブ同期（旧称SnapMirrorビジネス継続性（SMBC））の両方でサポートされており、1つ以上の障害ドメインで全体的な停止が発生した場合に高度なビジネス継続性を提供します。さまざまな障害モードへの耐障害性は、どの設定オプションを選択するかによって異なります。

vSphere環境向けの継続的可用性ソリューション

ONTAPのアーキテクチャは、柔軟性と拡張性に優れたストレージプラットフォームであり、データストアにSAN（FCP、iSCSI、NVMe-oF）サービスとNAS（NFS v3およびv4.1）サービスを提供します。NetApp AFF、ASA、FASの各ストレージシステムは、ONTAPオペレーティングシステムを使用して、ゲストストレージアクセス用にS3、SMB / CIFSなどの追加プロトコルを提供します。

NetApp MetroClusterは、ネットアップのHA（コントローラフェイルオーバーまたはCFO）機能を使用してコントローラ障害から保護します。また、ローカルSyncMirrorテクノロジ、災害時のクラスタフェイルオーバー（オンデマンドのコントローラフェイルオーバーまたはCFOD）、ハードウェアの冗長性、地理的な分離によって高レベルの可用性を実現します。SyncMirrorは、アクティブにデータを提供しているローカルプレックス（ローカルシェルフ上）と、通常はデータを提供していないリモートプレックス（リモートシェルフ上）の2つのプレックスにデータを書き込むことで、MetroCluster構成の2つの部分にわたってデータを同期的にミラーリングします。ハードウェアの冗長性は、コントローラ、ストレージ、ケーブル、スイッチ（ファブリックMetroClusterで使用）、アダプタなど、MetroClusterのすべてのコンポーネントで確保されています。

NetApp SnapMirrorアクティブ同期は、FCPおよびiSCSI SANプロトコルを使用してデータストアをきめ細かく保護するため、優先度の高いワークロードのみを選択的に保護できます。アクティブ/スタンバイ解決策であるNetApp MetroClusterとは異なり、ローカルサイトとリモートサイトの両方にアクティブ/アクティブアクセスを提供します。現時点では、アクティブ同期は非対称解決策であり、一方が他方よりも優先されるため、パフォーマンスが向上します。これにはAsymmetric Logical Unit Access（ALUA；非対称論理ユニットアクセス）機能が使用され、どのコントローラを優先するかがESXiホストに自動的に通知されます。ただし、NetAppでは、アクティブな同期によって完全対称アクセスがまもなく有効になることが発表されています。

2つのサイトにVMware HA / DRSクラスタを作成するために、ESXiホストをvCenter Server Appliance（vCSA）で使用および管理します。vSphere管理ネットワーク、vMotion®ネットワーク、および仮想マシンネットワークは、2つのサイト間の冗長ネットワークを介して接続されます。HA / DRSクラスタを管理するvCenter Serverは両方のサイトのESXiホストに接続でき、vCenter HAを使用して設定する必要があります。

を参照してください "vSphere Clientでクラスタを作成および構成する方法" をクリックしてvCenter HAを設定します。

また、 "VMware vSphere Metro Storage Cluster Recommended Practices"。

vSphere Metro Storage Clusterとは

vSphere Metro Storage Cluster（vMSC）は、仮想マシン（VM）とコンテナを障害から保護する認定済みの構成です。これは、ストレッチストレージの概念とESXiホストのクラスタを使用して実現されます。ESXiホストは、ラック、建物、キャンパス、さらには都市など、さまざまな障害ドメインに分散されます。NetApp MetroClusterとSnapMirrorのアクティブな同期ストレージテクノロジは、それぞれホストクラスタに対してRPO=0またはNearRPO=0の保護を提供するために使用されます。vMSCの構成は、物理的または論理的な「サイト」全体に障害が発生した場合でも、データを常に利用できるように設計されています。vMSC構成に含まれるストレージデバイスは、vMSC認定プロセスを完了したあとに認定されている必要があります。サポートされているすべてのストレージデバイスは、 "VMwareストレージ互換性ガイド"。

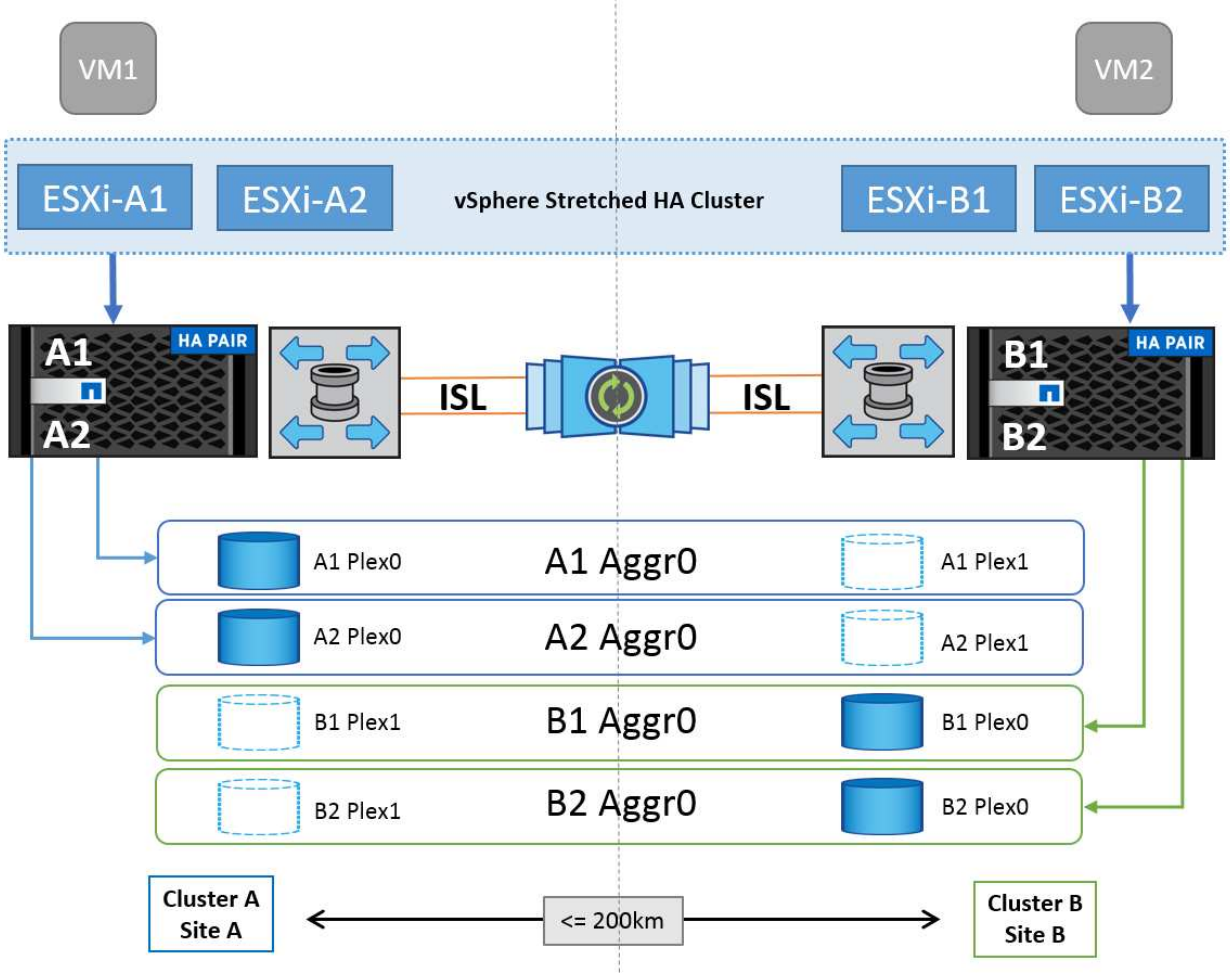
vSphere Metro Storage Clusterの設計ガイドラインの詳細については、次のドキュメントを参照してください。

- "NetApp MetroClusterによるVMware vSphereのサポート"
- "NetApp SnapMirrorビジネス継続性によるVMware vSphereのサポート" （SnapMirrorアクティブ同期）

レイテンシの考慮事項に応じて、NetApp MetroClusterを導入してvSphereで使用できます。

- ストレッチMetroCluster
- ファブリックMetroCluster

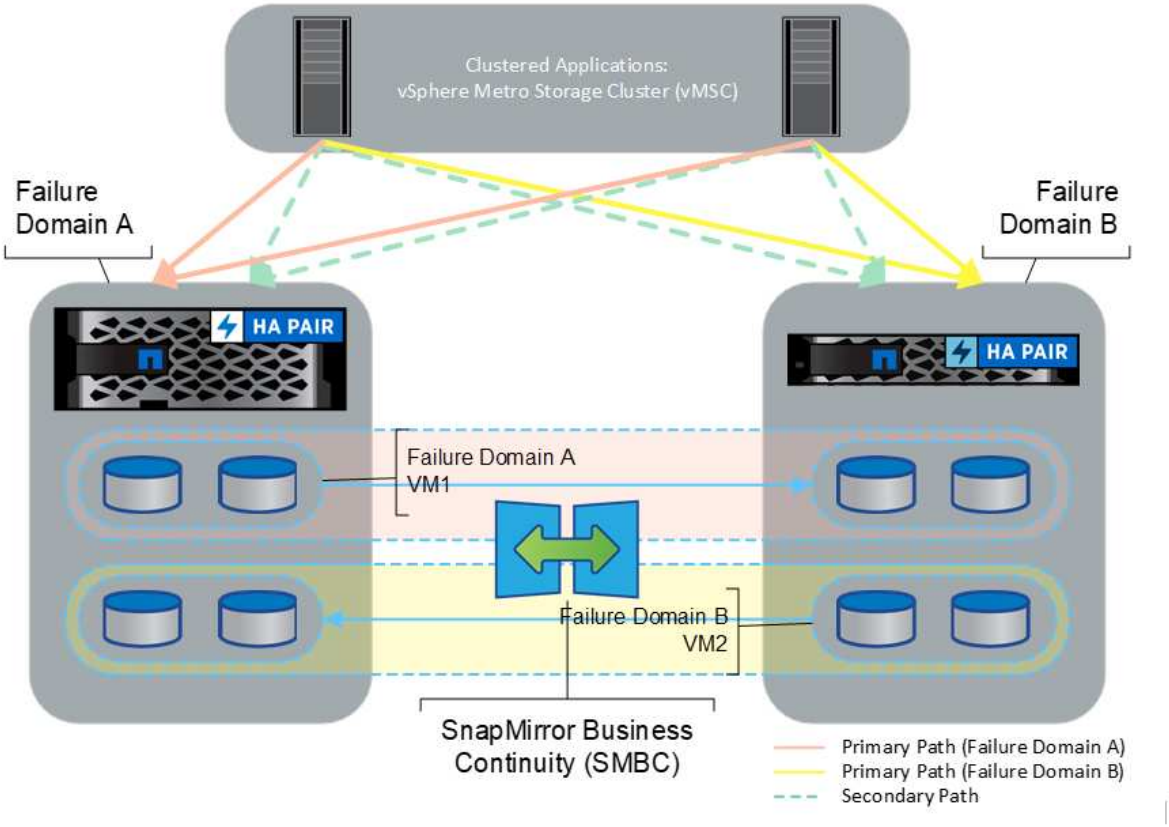
次の図は、ストレッチMetroClusterのトポロジ図の概要を示しています。



を参照してください "MetroCluster のドキュメント" を参照してくださいMetroCluster。

SnapMirror Active Syncは、2つの方法で導入することもできます。

- 非対称
- 対称（ONTAP 9.14.1でのプライベートプレビュー）



を参照してください "ネットアップのドキュメント" を参照し、SnapMirror Active Syncの設計と導入に関する情報を確認してください。

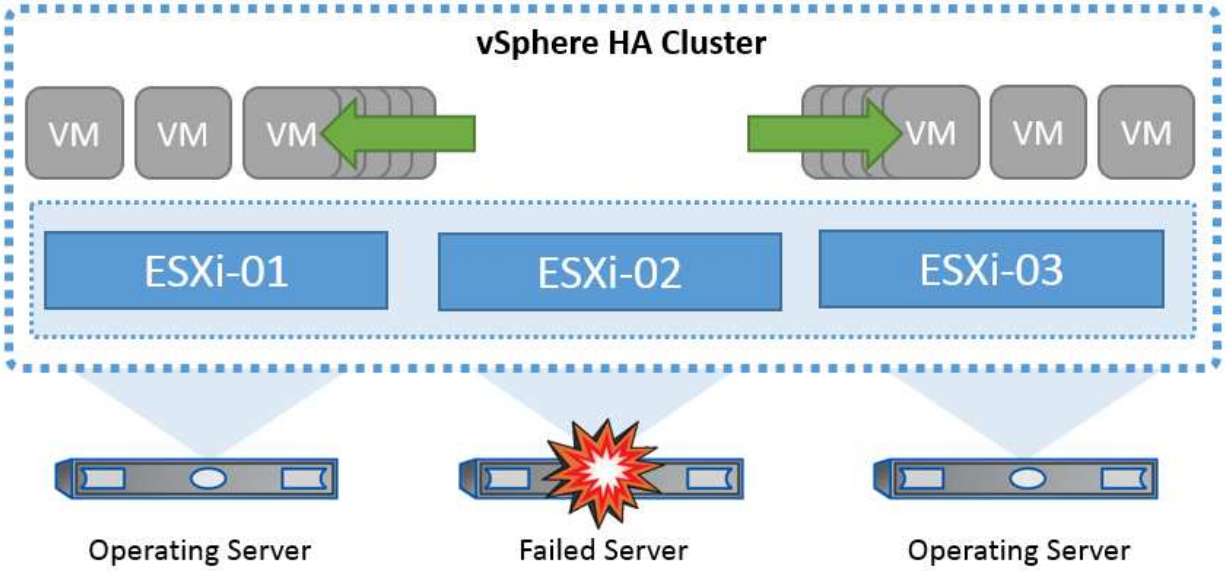
VMware vSphere解決策の概要

vCenter Server Appliance（vCSA）は、管理者がESXiクラスタを効果的に運用できるようにする、強力な一元管理システムであり、vSphere用の単一コンソールです。VMプロビジョニング、vMotion処理、High Availability（HA；高可用性）、Distributed Resource Scheduler（DRS；分散リソーススケジューラ）、Tanzu Kubernetes Gridなどの主要な機能を簡易化します。VMwareクラウド環境に欠かせないコンポーネントであり、サービスの可用性を考慮して設計する必要があります。

vSphereの高可用性

VMwareのクラスタテクノロジは、ESXiサーバを仮想マシンの共有リソースプールにグループ化し、vSphere High Availability（HA；高可用性）を提供します。vSphere HAは、仮想マシンで実行されるアプリケーションに対して、使いやすく高可用性を提供します。クラスタでHA機能を有効にすると、いずれかのESXiホストが応答しなくなったり分離されたりした場合に、各ESXiサーバが他のホストとの通信を維持します。 HAクラスタは、そのESXiホストで実行されていた仮想マシンのリカバリを、クラスタ内の残りのホスト間でネゴシエートできます。ゲストオペレーティングシステムに障害が発生すると、vSphere HAは影響を受ける仮想マシンを同じ物理サーバ上で再起動します。vSphere HAを使用すると、計画的停止の削減、計画外停止の防止、システム停止からの迅速なリカバリが可能になります。

vSphere HAクラスタ：障害が発生したサーバからVMをリカバリします。



VMware vSphereはNetApp MetroClusterまたはSnapMirrorのアクティブ同期を認識しないため、vSphereクラスタ内のすべてのESXiホストが、ホストおよびVMグループのアフィニティ構成に応じてHAクラスタ処理の対象となるホストとして認識されることを理解しておくことが重要です。

ホスト障害の検出

HAクラスタが作成されるとすぐに、クラスタ内のすべてのホストが選択対象になり、いずれかのホストがマスターになります。各スレーブはマスターに対してネットワークハートビートを実行し、マスターはすべてのスレーブホストに対してネットワークハートビートを実行します。vSphere HAクラスタのマスターホストは、スレーブホストの障害を検出する役割を果たします。

検出された障害のタイプによっては、ホストで実行されている仮想マシンのフェイルオーバーが必要になる場合があります。

vSphere HAクラスタでは、次の3種類のホスト障害が検出されます。

- 障害-ホストが機能を停止しました。
- 分離-ホストがネットワークから分離されます。
- パーティション-ホストとマスターホストとのネットワーク接続が失われます。

マスターホストは、クラスタ内のスレーブホストを監視します。この通信は、1秒ごとにネットワークハートビートを交換して行われます。マスターホストは、スレーブホストからのハートビートの受信を停止すると、ホストの稼働状況を確認してから、ホストに障害が発生したことを宣言します。マスターホストが実行する活性チェックでは、スレーブホストがいずれかのデータストアとハートビートを交換しているかどうかを確認します。また、マスターホストは、管理IPアドレスに送信されたICMP pingにホストが応答するかどうかをチェックして、単にマスターノードから隔離されているか、ネットワークから完全に隔離されているかを検出します。これは、デフォルトゲートウェイに対してpingを実行することによって行われます。隔離アドレスを手動で指定することで、隔離検証の信頼性を高めることができます。

ベストプラクティス

NetAppでは、隔離アドレスを少なくとも2つ追加し、各アドレスをサイトローカルにすることを推奨しています。これにより、隔離検証の信頼性が向上します。

ホスト隔離時の対応

[Isolation Response]はvSphere HAの設定で、vSphere HAクラスタ内のホストが管理ネットワーク接続を失い、実行は継続した場合に仮想マシンでトリガーされる処理を決定します。この設定には、[Disabled]、[Shut Down and Restart VMs]、[Power Off and Restart VMs]の3つのオプションがあります。

[Shut Down]は、[Power Off]よりも優れています。[Power Off]では、最新の変更がディスクにフラッシュされたり、トランザクションがコミットされたりしません。仮想マシンが300秒以内にシャットダウンされない場合は、電源がオフになります。待機時間を変更するには、詳細オプションdas.isolationshutdowntimeoutを使用します。

HAは隔離時の対応を開始する前に、vSphere HAマスターエージェントがVM構成ファイルが格納されたデータストアを所有しているかどうかを確認します。そうでない場合、VMを再起動するマスターがないため、ホストは隔離時の対応をトリガーしません。ホストはデータストアの状態を定期的にチェックして、マスターロールを持つvSphere HAエージェントがデータストアを要求しているかどうかを判断します。

ベストプラクティス

NetAppでは、[Host Isolation Response]を[Disabled]に設定することを推奨しています。

ホストがvSphere HAマスターホストから分離またはパーティショニングされ、ハートビートデータストアまたはpingを介してマスターと通信できなくなると、スプリットブレイン状態が発生することがあります。マスターは、隔離されたホストの停止を宣言し、クラスタ内の他のホスト上のVMを再起動します。仮想マシンのインスタンスが2つ実行され、そのうちの1つだけが仮想ディスクの読み取りまたは書き込みを実行できるため、スプリットブレイン状態が発生します。VM Component Protection（VMCP）を設定することで、スプリットブレイン状態を回避できるようになりました。

VMコンポーネント保護（**VMCP**）

vSphere 6で強化されたHA関連機能の1つにVMCPがあります。VMCPは、ブロック（FC、iSCSI、FCoE）とファイルストレージ（NFS）のAll Paths Down（APD）状態とPermanent Device Loss（PDL）状態からの保護を強化します。

Permanent Device Loss（PDL）

PDLとは、ストレージデバイスに永続的に障害が発生した場合、または管理上削除されて元に戻ることがない場合に発生する状態です。NetAppストレージアレイは、デバイスが永続的に失われたことを宣言するSCSIセンスコードをESXiに発行します。vSphere HAの[Failure Conditions and VM Response]セクションで、PDL状態が検出されたあとの応答を設定できます。

ベストプラクティス

NetAppでは、[Response for Datastore with PDL]を[* Power off and restart VMs]に設定することを推奨しています。この状態が検出されると、vSphere HAクラスタ内の正常なホストでVMが即座に再起動されます。

すべてのパスがダウン（**APD**）

APDは、ストレージデバイスがホストからアクセスできなくなり、アレイへのパスが使用できなくなった場合に発生する状態です。ESXiは、これをデバイスの一時的な問題とみなし、再び使用可能になることを想定しています。

APD状態が検出されると、タイマーが開始されます。140秒後、APD状態が正式に宣言され、デバイスはAPDタイムアウトとしてマークされます。140秒が経過すると、[Delay for VM Failover APD]で指定された分数が

カウントされます。指定した時間が経過すると、影響を受ける仮想マシンが再起動されます。必要に応じて異なる方法（[Disabled]、問題Events]、[Power Off and Restart VMs]）で応答するようにVMCPを設定できます。

ベストプラクティス

NetAppでは、[Response for Datastore with APD]を「* Power off and restart VMs（conservative）*」に設定することを推奨しています。

保守的とは、HAがVMを再起動できる可能性を示します。[Conservative]に設定すると、APDの影響を受けるVMは、別のホストで再起動できることがわかっている場合にのみ再起動されます。アグレッシブの場合、HAは他のホストの状態を認識していなくてもVMの再起動を試行します。その結果、VMが配置されているデータストアにアクセスできるホストがないと、VMが再起動されない可能性があります。

タイムアウトになる前にAPDステータスが解決され、ストレージへのアクセスが回復した場合は、明示的に設定していないかぎり、仮想マシンが不要に再起動されることはありません。環境がAPD状態から回復した場合でも応答が必要な場合は、[Response for APD Recovery After APD Timeout]を[Reset VMs]に設定する必要があります。

ベストプラクティス

NetAppでは、[Response for APD Recovery After APD Timeout]を[Disabled]に設定することを推奨します。

NetApp MetroCluster向け**VMware DRS**の実装

VMware DRSは、クラスタ内のホストリソースを集約する機能で、主に仮想インフラストラクチャ内のクラスタ内での負荷分散に使用されます。VMware DRSは、クラスタ内でロードバランシングを実行するために、主にCPUリソースとメモリリソースを計算します。vSphereはストレッチクラスタリングを認識しないため、両方のサイトのすべてのホストをロードバランシングの対象とします。サイト間トラフィックを回避するために、NetAppでは、VMの論理的な分離を管理するDRSアフィニティルールを設定することを推奨しています。これにより、サイト全体に障害が発生しないかぎり、HAとDRSでローカルホストのみが使用されるようになります。

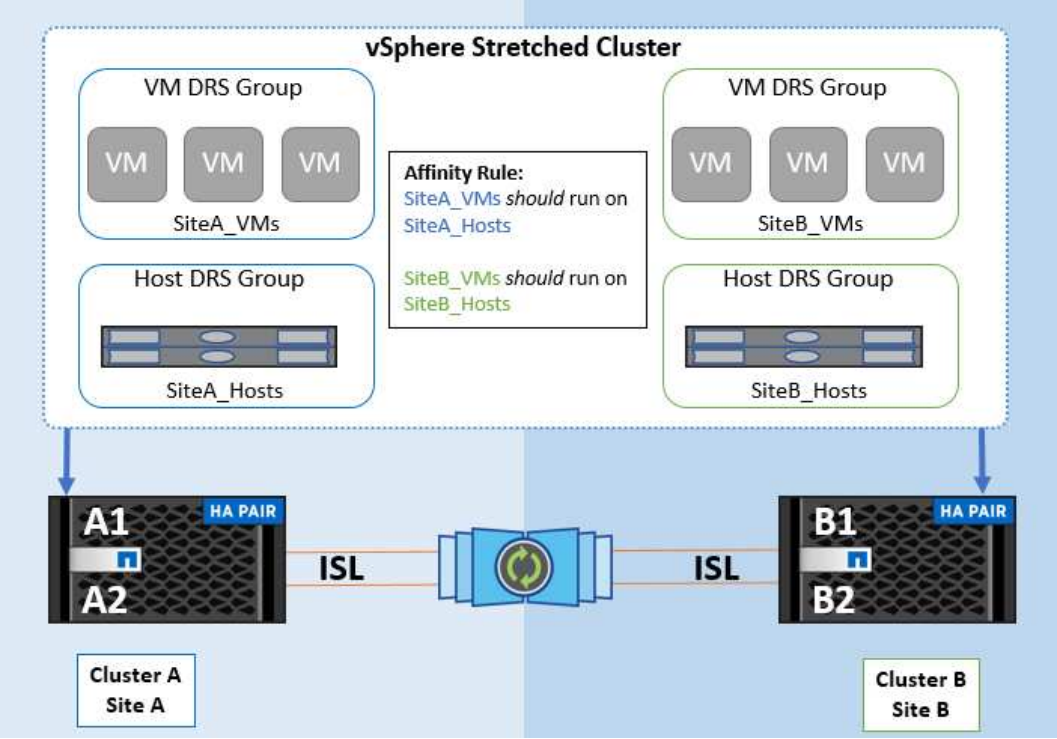
クラスタ用のDRSアフィニティルールを作成する場合は、仮想マシンのフェイルオーバー時にvSphereがそのルールを適用する方法を指定できます。

vSphere HAのフェイルオーバー動作を指定できるルールには、次の2種類があります。

- VMの非アフィニティルールでは、フェイルオーバー処理中に指定した仮想マシンが分離されたままになります。
- VMホストアフィニティルールは、フェイルオーバー処理中に、指定した仮想マシンを特定のホストまたは定義されたホストグループのメンバーに配置します。

VMware DRSのVMホストアフィニティルールを使用すると、サイトAとサイトBを論理的に分離して、特定のデータストアのプライマリ読み取り/書き込みコントローラとして設定されたアレイと同じサイトのホストでVMを実行できます。また、VMホストアフィニティルールを使用すると、仮想マシンはストレージに対してローカルなままになり、サイト間でネットワーク障害が発生した場合に仮想マシンの接続が確保されます。

次に、VMホストグループとアフィニティルールの例を示します。



ベストプラクティス

NetAppでは、障害が発生した場合にvSphere HAによって違反されるため、「must」ルールではなく「should」ルールを実装することを推奨しています。「must」ルールを使用すると、サービスが停止する可能性があります。

サービスの可用性は常にパフォーマンスより優先されるべきです。データセンター全体で障害が発生した場合、「must」ルールではVMホストアフィニティグループからホストを選択する必要があり、データセンターが使用できなくなっても仮想マシンは再起動されません。

NetApp MetroClusterでのVMware Storage DRSの実装

VMware Storage DRS機能を使用すると、データストアを1つのユニットに集約し、Storage I/O Controlのしきい値を超えた場合に仮想マシンディスクのバランスを調整できます。

Storage I/O Controlは、Storage DRS対応のDRSクラスタではデフォルトで有効になっています。Storage I/O Controlを使用すると、I/Oの輻輳時に仮想マシンに割り当てるストレージI/Oの量を管理者が制御できるため、重要度の高い仮想マシンを優先してI/Oリソースを割り当てることができます。

Storage DRSは、Storage vMotionを使用して、データストアクラスタ内の別のデータストアに仮想マシンを移行します。NetApp MetroCluster環境では、仮想マシンの移行をそのサイトのデータストア内で制御する必要があります。たとえば、サイトAのホストで実行されている仮想マシンAを移行する場合は、サイトAのSVMのデータストア内で移行するのが理想的です。そうしないと、仮想ディスクの読み取り/書き込みはサイト間リンクを介してサイトBから行われるため、仮想マシンは引き続き動作しますが、パフォーマンスは低下します。

ベストプラクティス

NetAppでは、ストレージサイトのアフィニティに従ってデータストアクラスタを作成することを推奨しています。つまり、サイトAに対するサイトアフィニティが設定されたデータストアクラスタと、サイトBに対するサイトアフィニティが設定されたデータストアを混在させないでください。

Storage vMotionを使用して仮想マシンを新規にプロビジョニングまたは移行するたびに、NetAppそれらの仮想マシンに固有のすべてのVMware DRSルールを手動で更新することを推奨します。これにより、ホストとデータストアの両方について、サイトレベルで仮想マシンのアフィニティが確保され、ネットワークとストレージのオーバーヘッドが削減されます。

vMSC設計および実装ガイドライン

本ドキュメントでは、ONTAPストレージシステムを使用するvMSCの設計と実装のガイドラインについて説明します。

NetAppストレージ構成

NetApp MetroCluster（MCC構成）のセットアップ手順については、次のWebサイトを参照してください。"MetroCluster のドキュメント"。SnapMirrorアクティブ同期の手順については、 "SnapMirror のビジネス継続性機能の概要"。

一度MetroClusterを設定すると、従来のONTAP環境を管理するようなものになります。Storage Virtual Machine（SVM）は、コマンドラインインターフェイス（CLI）、System Manager、Ansibleなどのさまざまなツールを使用してセットアップできます。SVMを設定したら、通常の運用に使用する論理インターフェイス（LIF）、ボリューム、論理ユニット番号（LUN）をクラスタに作成します。これらのオブジェクトは、クラスタピアリングネットワークを使用してもう一方のクラスタに自動的にレプリケートされます。

MetroClusterを使用していない場合は、SnapMirrorアクティブ同期を使用して、異なる障害ドメインにある複数のONTAPクラスタ間で、データストア単位でのきめ細かな保護とアクティブ/アクティブアクセスを実現できます。SnapMirrorアクティブ同期では、整合グループを使用して1つ以上のデータストア間で書き込み順序の整合性が確保されます。また、アプリケーションとデータストアの要件に応じて、複数の整合グループを作成することもできます。整合グループは、複数のデータストア間でのデータ同期が必要なアプリケーションに特に役立ちます。SnapMirror Active Syncでは、rawデバイスマッピング（RDM）とゲスト内iSCSIイニシエータを使用するゲスト接続ストレージもサポートされます。整合グループの詳細については、を参照してください。 "整合グループの概要"。

SnapMirrorアクティブ同期を使用するvMSC構成の管理は、MetroClusterとは多少異なります。まず、これはSANのみの構成であり、SnapMirrorのアクティブな同期でNFSデータストアを保護することはできません。次に、両方の障害ドメインのレプリケートされたデータストアにアクセスできるように、両方のLUNのコピーをESXiホストにマッピングする必要があります。

VMware vSphere HA の場合

vSphere HAクラスタの作成

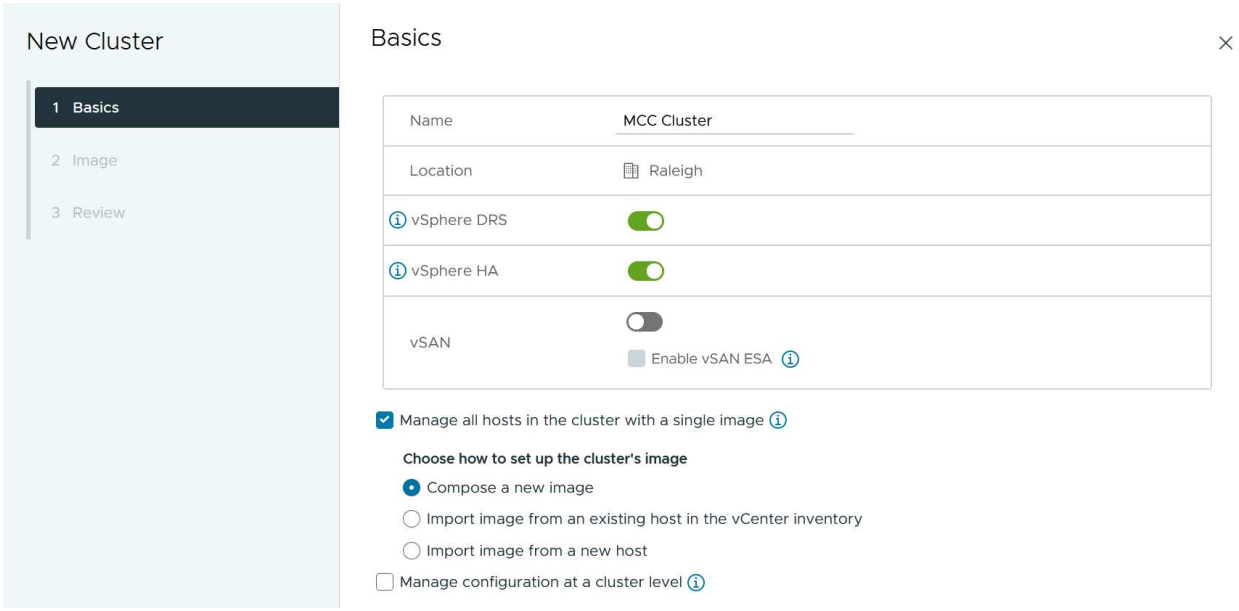
vSphere HAクラスタの作成は複数の手順で構成されます。詳細については、を参照してください。"docs.vmware.comのvSphere Clientでクラスタを作成および構成する方法"。つまり、最初に空のクラスタを作成してから、vCenterを使用してホストを追加し、クラスタのvSphere HAなどの設定を指定する必要があります。

*注：*このドキュメントには、このドキュメントより優先されるものはありません。 "VMware vSphere Metro Storage Cluster Recommended Practices"

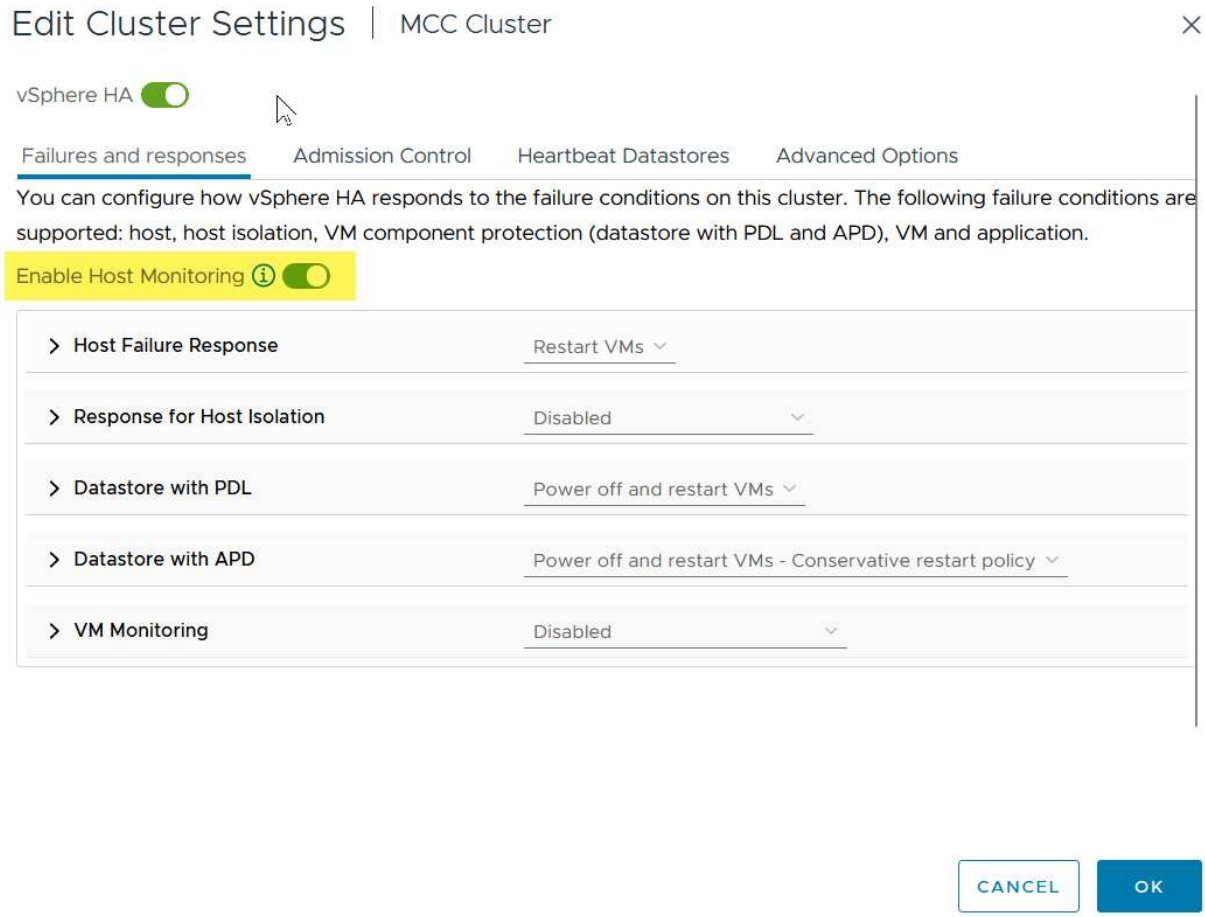
HAクラスタを設定するには、次の手順を実行します。

1. vCenter UIに接続します。
2. [Hosts and Clusters]で、HAクラスタを作成するデータセンターを選択します。

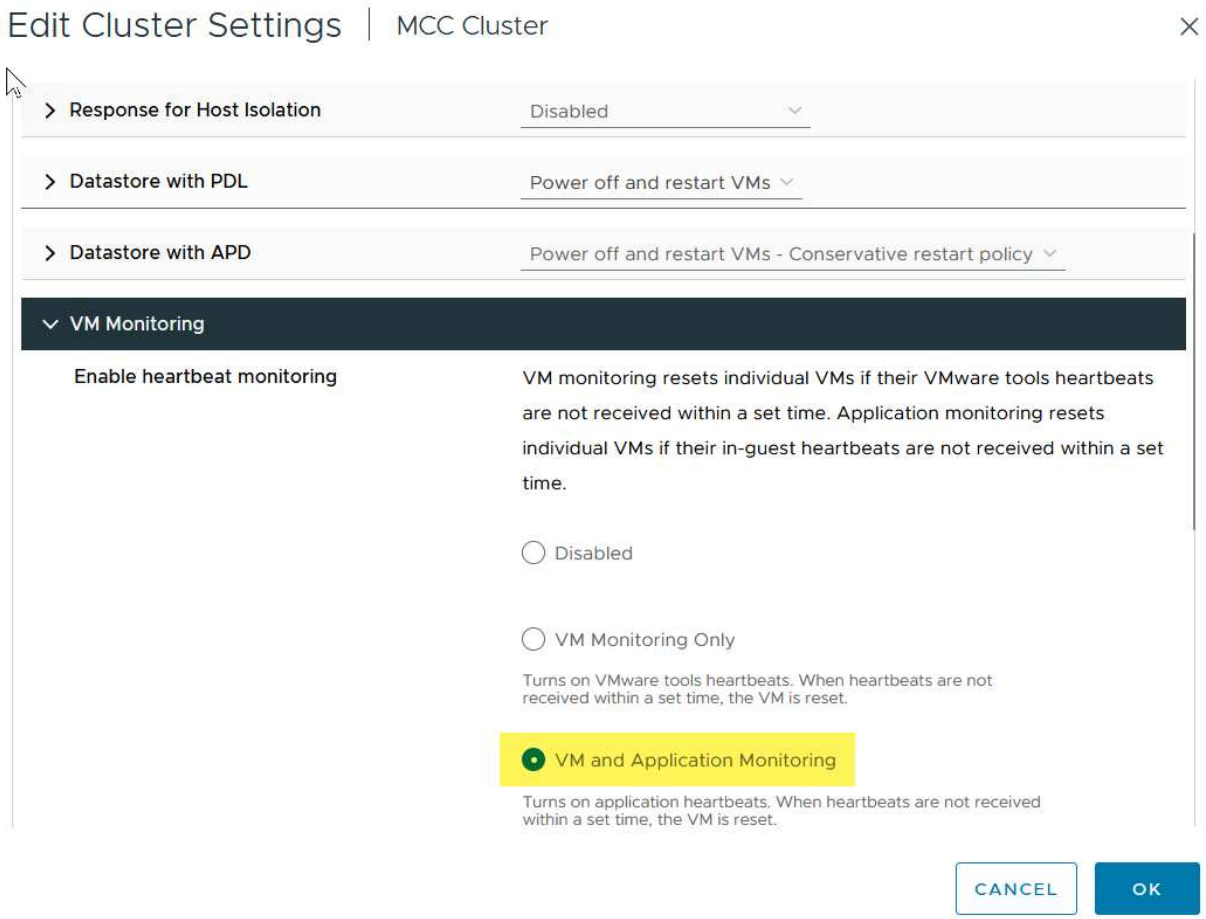
3. データセンターオブジェクトを右クリックし、[New Cluster]を選択します。[Basics]で、vSphere DRSとvSphere HAが有効になっていることを確認します。ウィザードの手順を実行します。



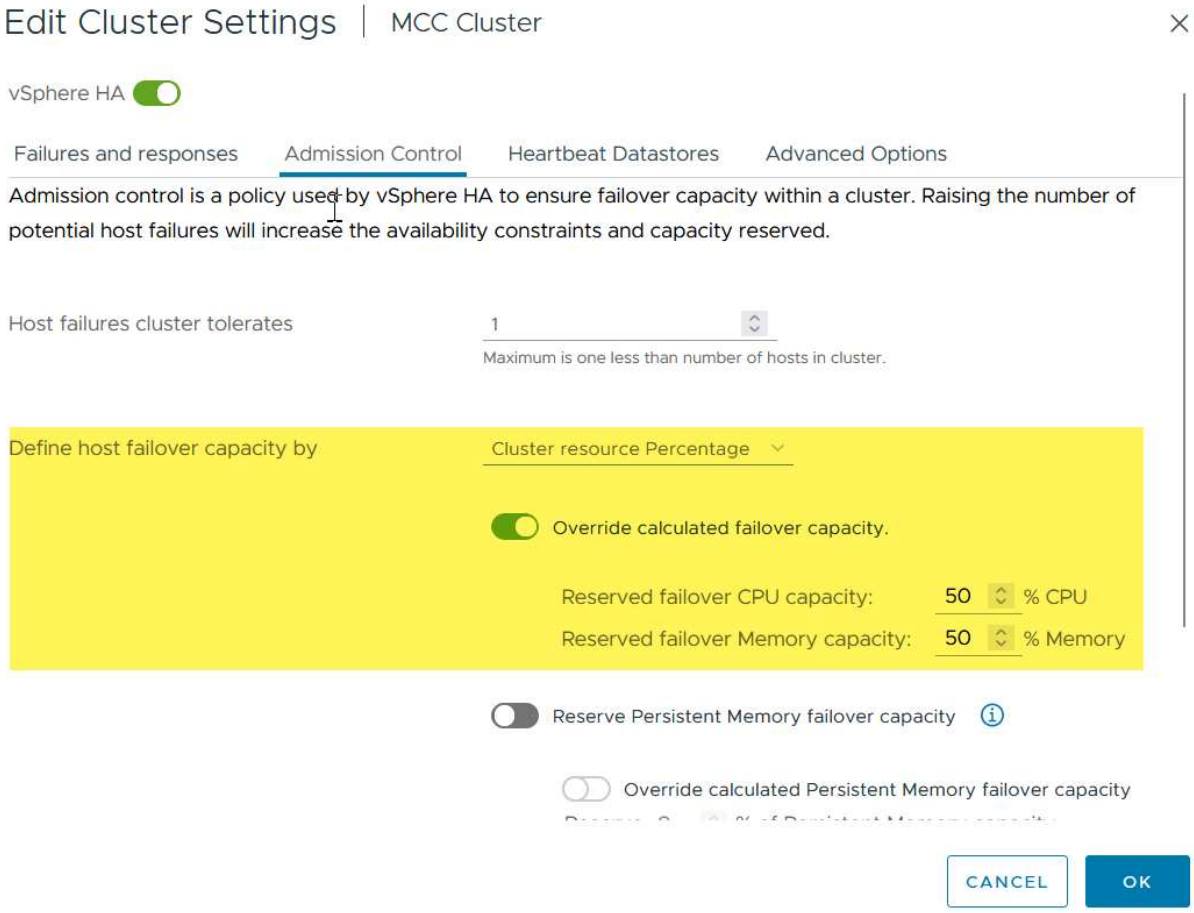
1. クラスタを選択し、[Configure]タブに移動します。[vSphere HA]を選択し、[edit]をクリック
2. [Host Monitoring]で、[Enable Host Monitoring]オプションを選択します。



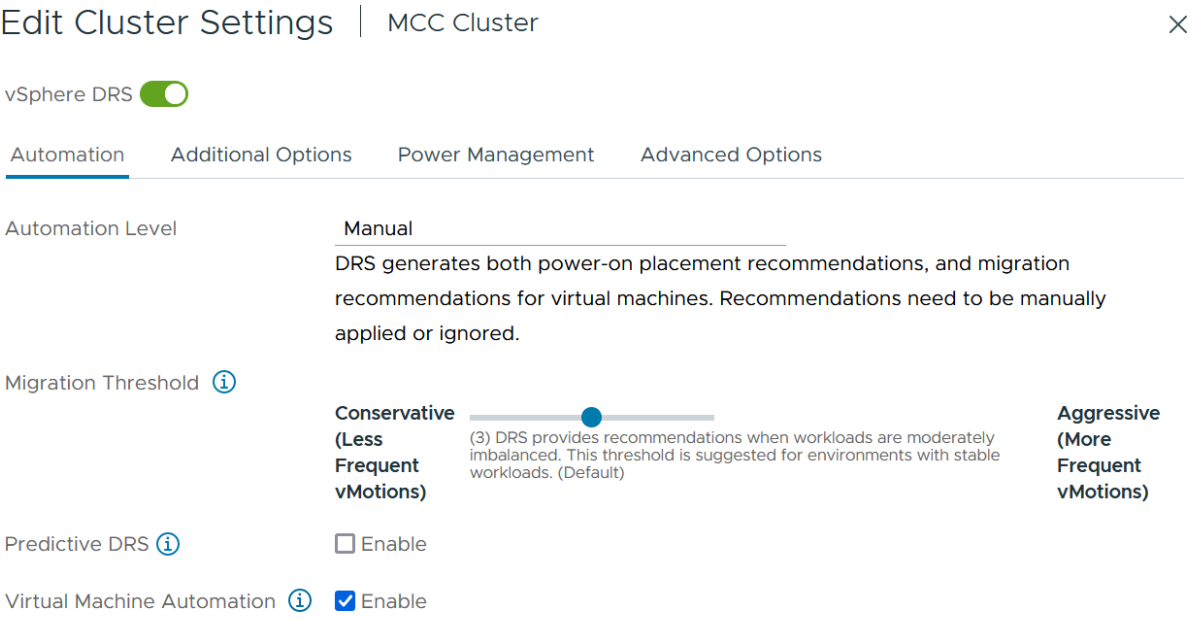
1. [Failures and Responses]タブの[VM Monitoring]で、[VM Monitoring Only]オプションまたは[VM and Application Monitoring]オプションを選択します。



1. [Admission Control]で、[HA Admission Control]オプションを[cluster resource reserve]に設定し、50%のCPU/MEMを使用します。



1. [OK]をクリックします。
2. [DRS]を選択し、[編集]をクリックします。
3. アプリケーションで必要な場合を除き、自動化レベルを手動に設定します。



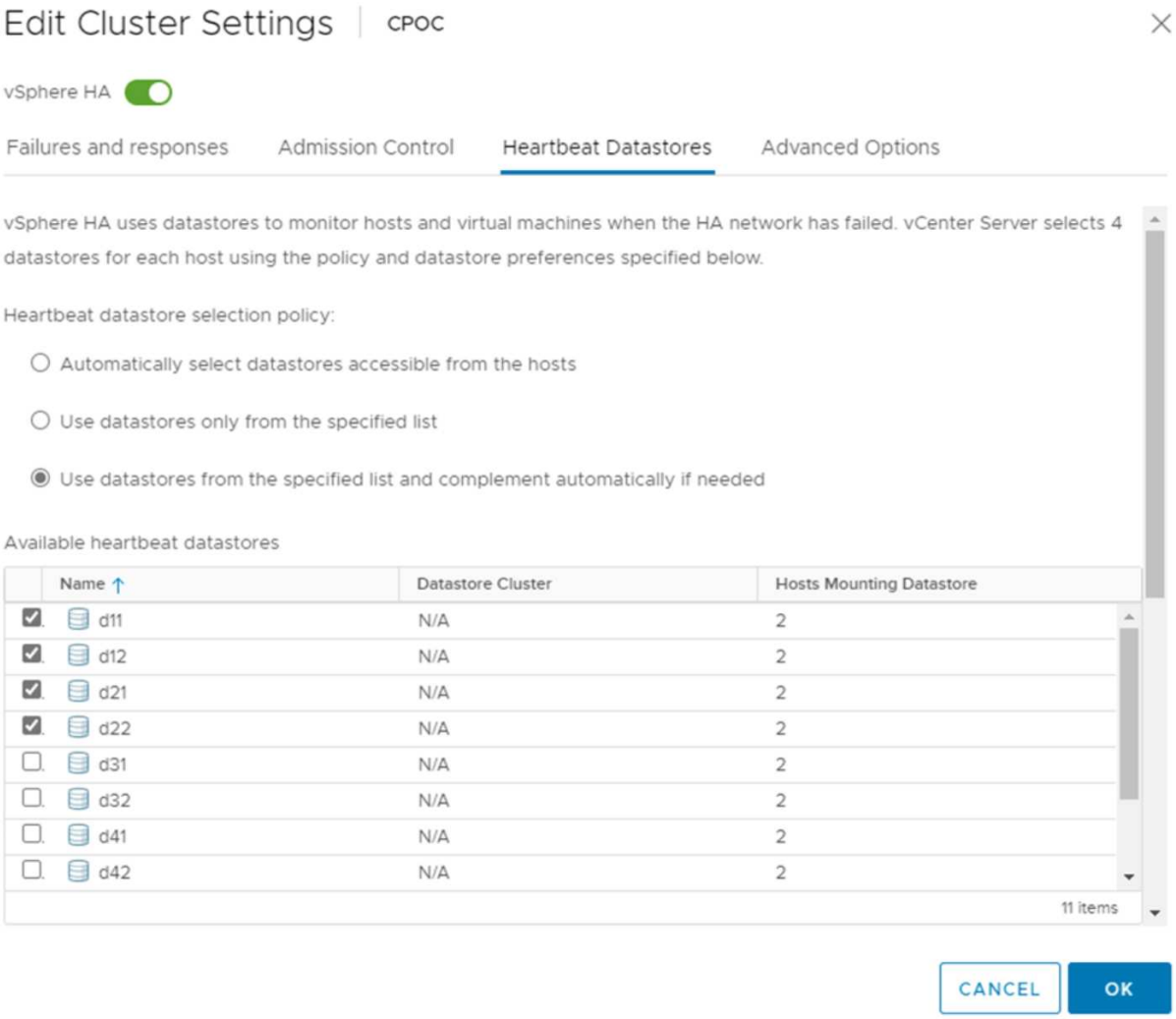
1. VMコンポーネント保護を有効にします。を参照してください。 "docs.vmware.com"。
2. MCCを使用するvMSCでは、次のvSphere HAの追加設定が推奨されます。

失敗	応答
ホスト障害です	VMの再起動
ホストの分離	無効
Permanent Device Loss（PDL；永続的デバイス損失）のあるデータストア	VMの電源をオフにして再起動する
すべてのパスがダウンしているデータストア（APD）	VMの電源をオフにして再起動する
ゲストが鼓動しない	VMのリセット
VM再起動ポリシー	VMの重要度に応じて決定
ホスト隔離時の応答	VMのシャットダウンと再起動
PDLを使用したデータストアの応答	VMの電源をオフにして再起動する
APDを使用するデータストアの応答	VMの電源をオフにして再起動する（控えめ）
APDのVMフェイルオーバーの遅延	3分
APDタイムアウトによるAPDリカバリの応答	無効
VM監視の感度	プリセット高

ハートビート用のデータストアの設定

vSphere HAでは、管理ネットワークに障害が発生した場合、データストアを使用してホストと仮想マシンを監視します。vCenterでのハートビートデータストアの選択方法を設定できます。ハートビート用のデータストアを設定するには、次の手順を実行します。

1. [Datastore Heartbeating]セクションで、[Use Datastores from the Specified List and Complement Automatically if Needed]を選択します。
2. vCenterで使用するデータストアを両方のサイトから選択し、[OK]を押します。

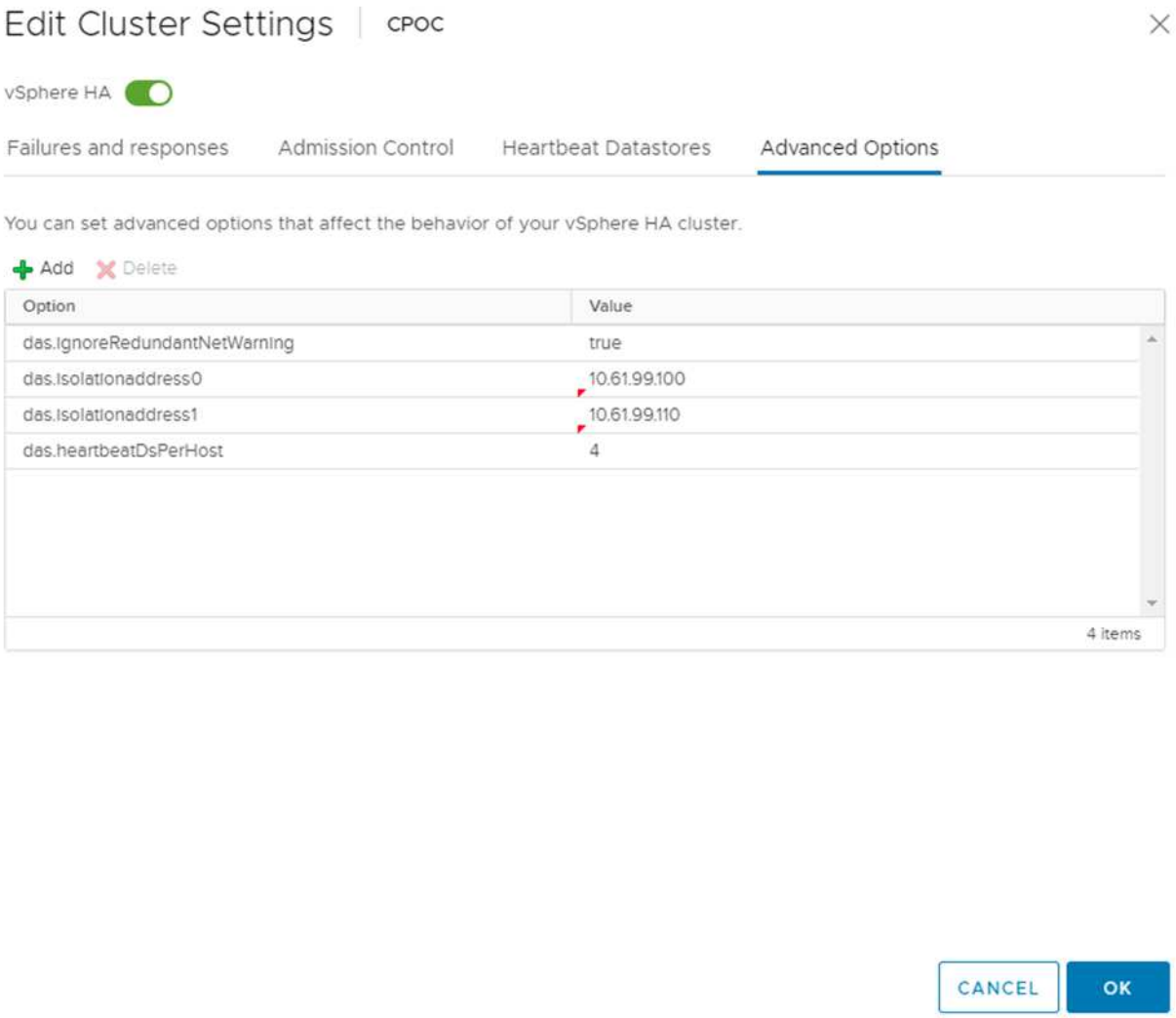


詳細オプションの設定

ホスト障害の検出

HAクラスタ内のホストがネットワークまたはクラスタ内の他のホストに接続できなくなると、分離イベントが発生します。デフォルトでは、vSphere HAは管理ネットワークのデフォルトゲートウェイをデフォルトの分離アドレスとして使用します。ただし、ホストがpingを実行するための追加の隔離アドレスを指定して、隔離応答をトリガーするかどうかを判断することができます。pingを実行できる隔離IPをサイトごとに1つずつ追加します。ゲートウェイIPは使用しないでください。使用するvSphere HAの詳細設定はdas.isolationaddressです。この目的には、ONTAPまたはメディエーターのIPアドレスを使用できます。

を参照してください "core.vmware.com" 詳細については、_を参照してください。



das.heartbeatDsPerHostという詳細設定を追加すると、ハートビートデータストアの数を増やすことができます。4つのハートビートデータストア（HB DSS）（サイトごとに2つ）を使用します。[Select from List but complent]オプションを使用します。これは、1つのサイトで障害が発生してもHB DSSが2つ必要になるためです。ただし、MCCやSnapMirrorのアクティブな同期で保護する必要はありません。

を参照してください "core.vmware.com" 詳細については、_を参照してください。

NetApp MetroCluster向けVMware DRSアフィニティ

このセクションでは、MetroCluster環境内のサイト/クラスタごとに、VMとホストのDRSグループを作成します。次に、VMホストアフィニティをローカルストレージリソースとアライメントするようにVM\Hostルールを設定します。たとえば、サイトAのVMがVMグループsitea_vmsに属し、サイトAのホストがホストグループsitea_hostsに属しているとします。次に、VM\Hostルールで、sitea_vmsをsitea_hostsのホストで実行するように記述します。

ベストプラクティス

- NetAppでは、「Must Run on Hosts in Group」という仕様ではなく、「Should Run on Hosts in Group」という仕様を使用することを強く推奨しています。サイトAのホストで障害が発生した場合、vSphere HAを使用してサイトAのVMをサイトBのホストで再起動する必要がありますが、後者の仕様では、HAがサイトBのVMを再起動することは難しいルールであるため許可されていません。前者の仕様はソフトルールで

あり、HAが発生した場合は違反となるため、パフォーマンスではなく可用性が確保されます。

*注：*仮想マシンがVMとホストのアフィニティルールに違反したときにトリガーされるイベントベースのアラームを作成できます。vSphere Clientで、仮想マシンの新しいアラームを追加し、イベントトリガーとして[VM is violating VM-Host Affinity Rule]を選択します。アラームの作成と編集の詳細については、を参照してください。 "vSphereの監視とパフォーマンス" ドキュメント

DRSホストグループの作成

サイトAとサイトBに固有のDRSホストグループを作成するには、次の手順を実行します。

1. vSphere Web Clientで、インベントリ内のクラスタを右クリックし、[Settings]を選択します。
2. [VM\Host Groups]をクリックします。
3. 追加をクリックします。
4. グループの名前を入力します（例：sitea_hosts）。
5. [Type]メニューから[Host Group]を選択します。
6. [Add]をクリックし、サイトAから目的のホストを選択して[OK]をクリックします。
7. 同じ手順を繰り返して、サイトBのホストグループをもう1つ追加します。
8. [OK] をクリックします。

DRS VMグループの作成

サイトAとサイトBに固有のDRS VMグループを作成するには、次の手順を実行します。

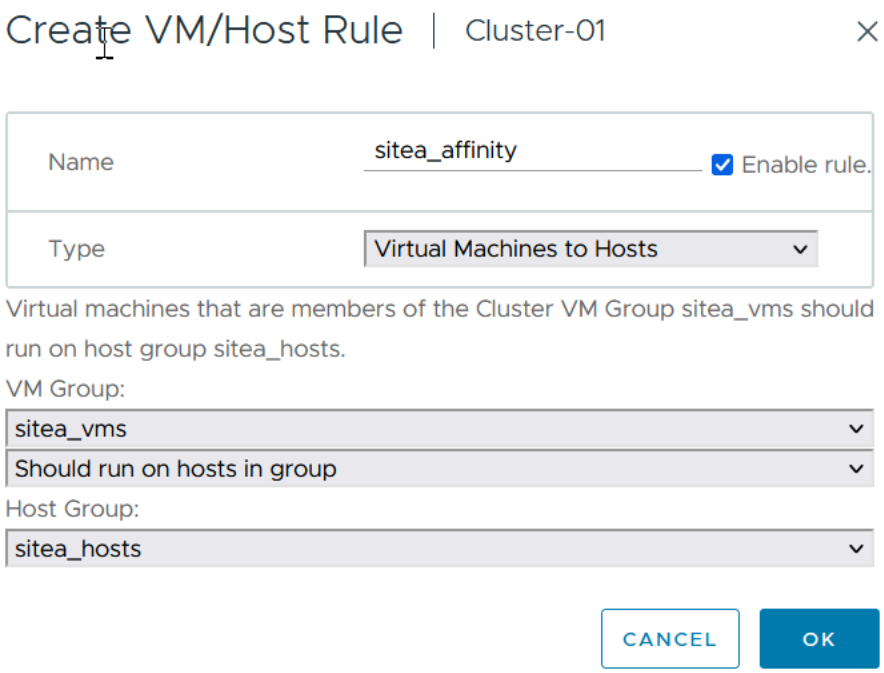
1. vSphere Web Clientで、インベントリ内のクラスタを右クリックし、[Settings]を選択します。
2. [VM\Host Groups]をクリックします。
3. 追加をクリックします。
4. グループの名前を入力します（例：sitea_vms）。
5. [Type]メニューから[VM Group]を選択します。
6. [Add]をクリックし、サイトAから目的のVMを選択して[OK]をクリックします。
7. 同じ手順を繰り返して、サイトBのホストグループをもう1つ追加します。
8. [OK] をクリックします。

VMホストルールの作成

サイトAとサイトBに固有のDRSアフィニティルールを作成するには、次の手順を実行します。

1. vSphere Web Clientで、インベントリ内のクラスタを右クリックし、[Settings]を選択します。
2. [VM\Host Rules]をクリックします。
3. 追加をクリックします。
4. ルールの名前を入力します（例：sitea_affinity）。
5. Enable Ruleオプションがオンになっていることを確認します。
6. [Type]メニューから[Virtual Machines to Hosts]を選択します。

7. VMグループを選択します（例：sitea_vms）。
8. ホストグループを選択します（例：sitea_hosts）。
9. 同じ手順を繰り返して、サイトBのVM\Hostルールをもう1つ追加します。
10. [OK] をクリックします。



NetApp MetroCluster向けVMware vSphere Storage DRS

データストアクラスタの作成

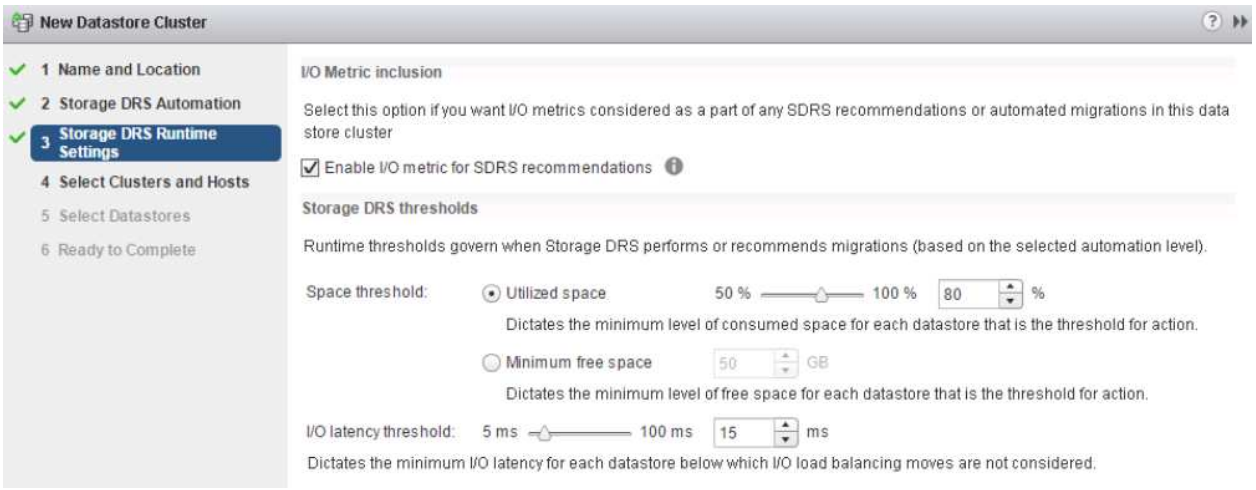
各サイトのデータストアクラスタを設定するには、次の手順を実行します。

1. vSphere Web Clientを使用して、[Storage]の下にあるHAクラスタが配置されているデータセンターに移動します。
2. データセンターオブジェクトを右クリックし、[Storage]>[New Datastore Cluster]を選択します。
3. [Turn on Storage DRS]オプションを選択し、[Next]をクリックします。
4. すべてのオプションを[No Automation（Manual Mode）]に設定し、[Next]をクリックします。

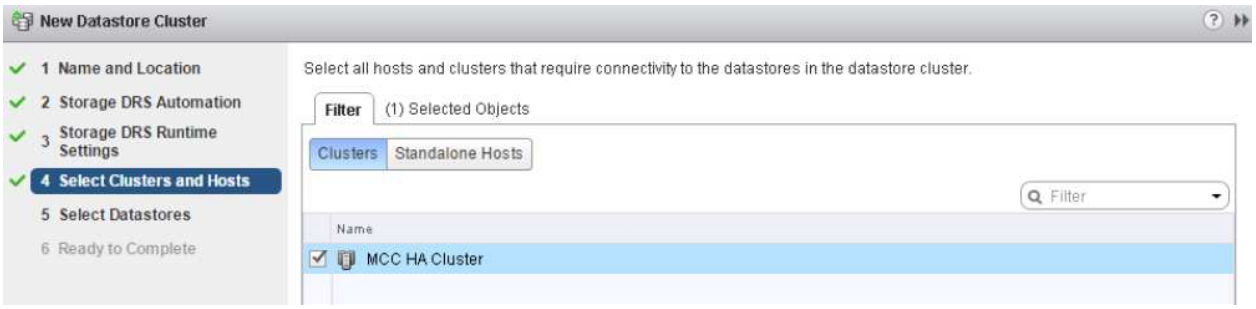
ベストプラクティス

- NetAppでは、移行が必要になるタイミングを管理者が判断して制御できるように、Storage DRSを手動モードで設定することを推奨しています。

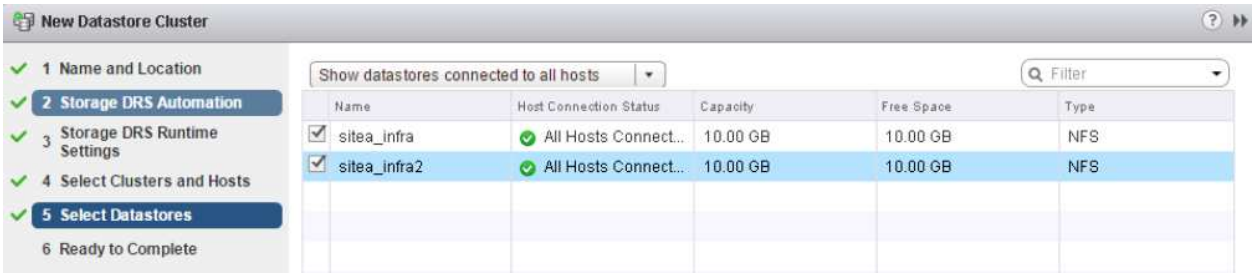
1. [Enable I/O Metric for SDRS Recommendations]チェックボックスがオンになっていることを確認します。メトリック設定はデフォルト値のままにできます。



1. HAクラスタを選択し、[Next]をクリックします。



1. サイトAに属するデータストアを選択し、[Next]をクリックします。



1. オプションを確認し、[完了]をクリックします。
2. 同じ手順を繰り返してサイトBのデータストアクラスタを作成し、サイトBのデータストアのみが選択されていることを確認します。

vCenter Serverの可用性

vCenter Server Appliance（VCSA）はvCenter HAで保護する必要があります。vCenter HAでは、アクティブ/パッシブHAペアに2つのVCSAを導入できます。障害ドメインごとに1つ。vCenter HAの詳細については、"docs.vmware.com"。

計画的イベントと計画外イベントの耐障害性

NetApp MetroClusterとSnapMirrorのアクティブ同期は、NetAppハードウェアとONTAP®ソフトウェアの高可用性とノンストップオペレーションを強化する強力なツールです。

これらのツールは、ストレージ環境全体をサイト全体で保護し、データの可用性を確保します。スタンドアロンサーバ、高可用性サーバクラスタ、Dockerコンテナ、仮想サーバのいずれを使用している場合でも、NetAppテクノロジは、停電、冷却装置の障害、ネットワーク接続の障害、ストレージアレイのシャットダウン、または運用上のエラーが原因で全体が停止した場合でも、ストレージの可用性をシームレスに維持します。

MetroClusterとSnapMirrorのアクティブな同期では、計画的または計画外のイベントが発生した場合に、次の3つの基本的な方法でデータを継続できます。

- 冗長コンポーネントによる単一コンポーネント障害からの保護
- ローカルのHAテイクオーバー：1台のコントローラに影響するイベントに対応
- 完全なサイト保護–ストレージおよびクライアントのアクセスをソースクラスタからデスティネーションクラスタに移動することで、サービスを迅速に再開します。

つまり、1つのコンポーネントで障害が発生してもシームレスに運用が継続され、障害が発生したコンポーネントを交換すると自動的に冗長運用に戻ります。

シングルノードクラスタ（通常はONTAP Selectなどのソフトウェア定義バージョン）を除くすべてのONTAPクラスタには、テイクオーバーとギブバックと呼ばれるHA機能が組み込まれています。クラスタ内の各コントローラが別のコントローラとペアリングされ、HAペアが形成されます。これらのペアにより、各ノードはストレージにローカルで接続されます。

テイクオーバーは、データサービスを維持するために一方のノードがもう一方のノードのストレージをテイクオーバーする自動プロセスです。ギブバックは、通常動作に戻る逆のプロセスです。テイクオーバーは、ハードウェアのメンテナンス時やONTAPのアップグレード時などに計画的に行うことも、ノードのパニックやハードウェア障害による計画外で行うこともできます。

テイクオーバー時に、MetroCluster構成のネットワーク接続型ストレージ論理インターフェイス（NAS LIF）が自動的にフェイルオーバーされます。ただし、ストレージエリアネットワークLIF（SAN LIF）はフェイルオーバーせず、引き続き論理ユニット番号（LUN）への直接パスを使用します。

HAのテイクオーバーとギブバックの詳細については、 "HAペアの管理の概要"。この機能は、MetroClusterまたはSnapMirrorのアクティブな同期に固有ではないことに注意してください。

MetroClusterによるサイトのスイッチオーバーは、一方のサイトがオフラインになった場合、またはサイト全体のメンテナンスのために計画的に実行された場合に実行されます。オフラインになったクラスタのストレージリソース（ディスクおよびアグリゲート）の所有権がもう一方のサイトに引き継がれ、障害が発生したサイトのSVMがディザスタサイトでオンラインになって再起動されます。その際、クライアントとホストのアクセス用にIDは保持されます。

SnapMirrorのアクティブな同期では、両方のコピーが同時にアクティブに使用されるため、既存のホストは引き続き動作します。サイトのフェイルオーバーを正しく実行するには、NetAppメディエーターが必要です。

MCCを使用するvMSCの障害シナリオ

以降のセクションでは、vMSCおよびNetApp MetroClusterシステムで発生したさまざまな障害シナリオで想定される結果について説明します。

単一のストレージパス障害

このシナリオでは、HBAポート、ネットワークポート、フロントエンドデータスイッチポート、FCケーブル、イーサネットケーブルなどのコンポーネントで障害が発生すると、ストレージデバイスへの特定のパスがESXiホストによって停止とマークされます。HBA /ネットワーク/スイッチポートで耐障害性を提供してストレージデバイスに複数のパスが設定されている場合は、ESXiがパススイッチオーバーを実行するのが理想的です。この間、ストレージデバイスへの複数のパスを提供することでストレージの可用性が確保されるため、仮想マシンは影響を受けずに実行され続けます。

*注：*このシナリオではMetroClusterの動作に変更はなく、すべてのデータストアがそれぞれのサイトで引き続き実行されます。

ベストプラクティス

NFS / iSCSIボリュームを使用している環境ではNetApp、NFS vmkernelポート用に少なくとも2つのネットワークアップリンクを標準vSwitchに設定し、NFS vmkernelインターフェイスが分散vSwitchにマッピングされているポートグループに設定することを推奨します。NICチーミングは、アクティブ/アクティブまたはアクティブ/スタンバイのいずれかで設定できます。

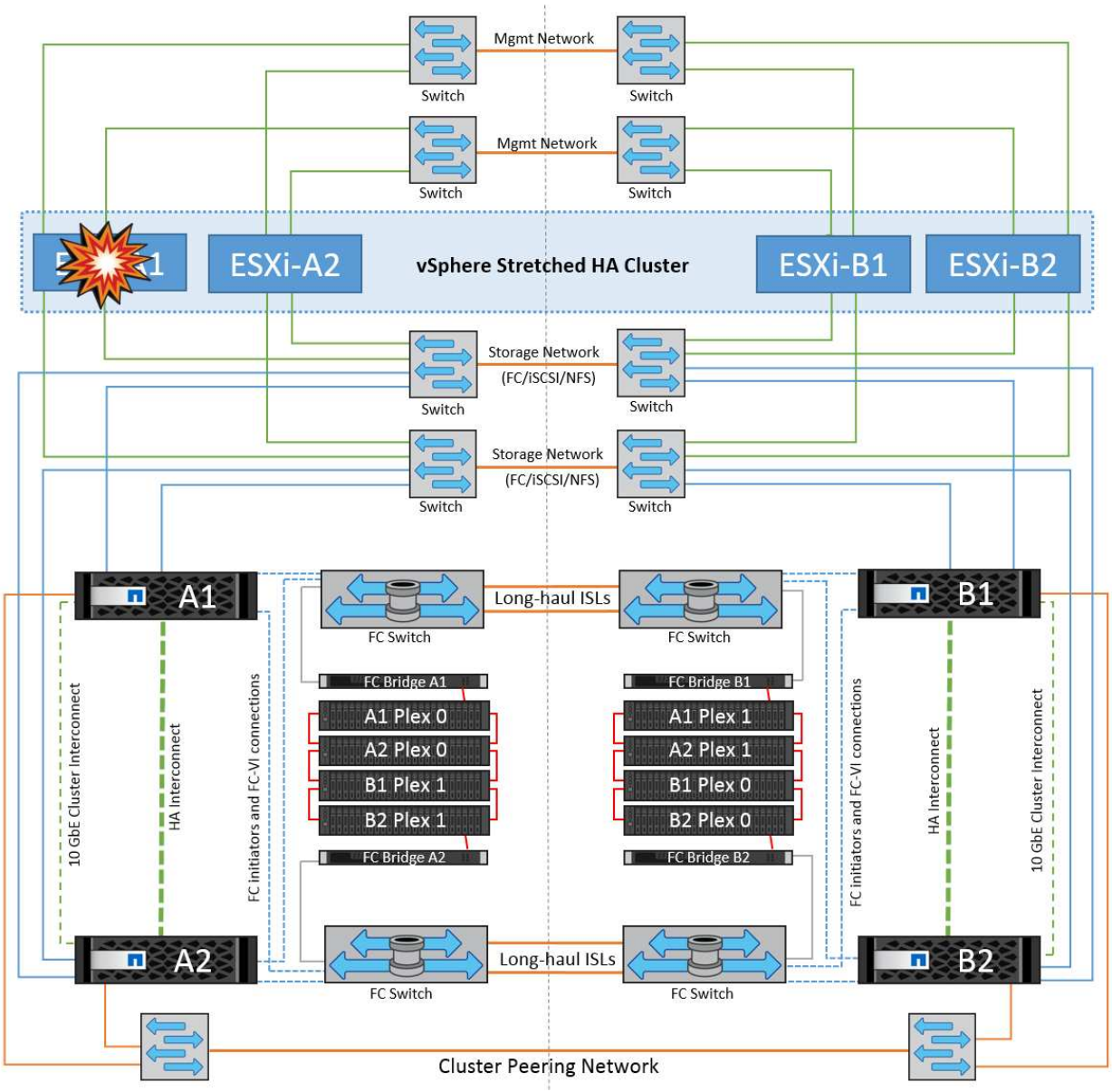
また、iSCSI LUNの場合は、vmkernelインターフェイスをiSCSIネットワークアダプタにバインドしてマルチパスを設定する必要があります。詳細については、vSphereストレージのドキュメントを参照してください。

ベストプラクティス

ファイバチャネルLUNを使用する環境でNetAppは、HBAを少なくとも2つ搭載し、HBA /ポートレベルでの耐障害性を保証することを推奨します。NetAppでは、ゾーニングを設定するためのベストプラクティスとして、単一のイニシエータから単一のターゲットへのゾーニングも推奨しています。

新規および既存のすべてのNetAppストレージデバイスにポリシーが設定されるため、Virtual Storage Console（VSC）を使用してマルチパスポリシーを設定する必要があります。

単一のESXiホスト障害



このシナリオでは、ESXiホストで障害が発生すると、VMware HAクラスタのマスターノードがネットワークハートビートを受信しなくなるため、ホスト障害を検出します。ホストが本当に停止しているのか、ネットワークパーティションだけなのかを判別するために、マスターノードはデータストアハートビートを監視し、ハートビートがない場合は、障害が発生したホストの管理IPアドレスに対してpingを実行して最終チェックを実行します。これらのチェックがすべて無効の場合、マスターノードはこのホストを障害が発生したホストであると宣言し、この障害が発生したホストで実行されていたすべての仮想マシンが、クラスタ内の残りのホストでリブートされます。

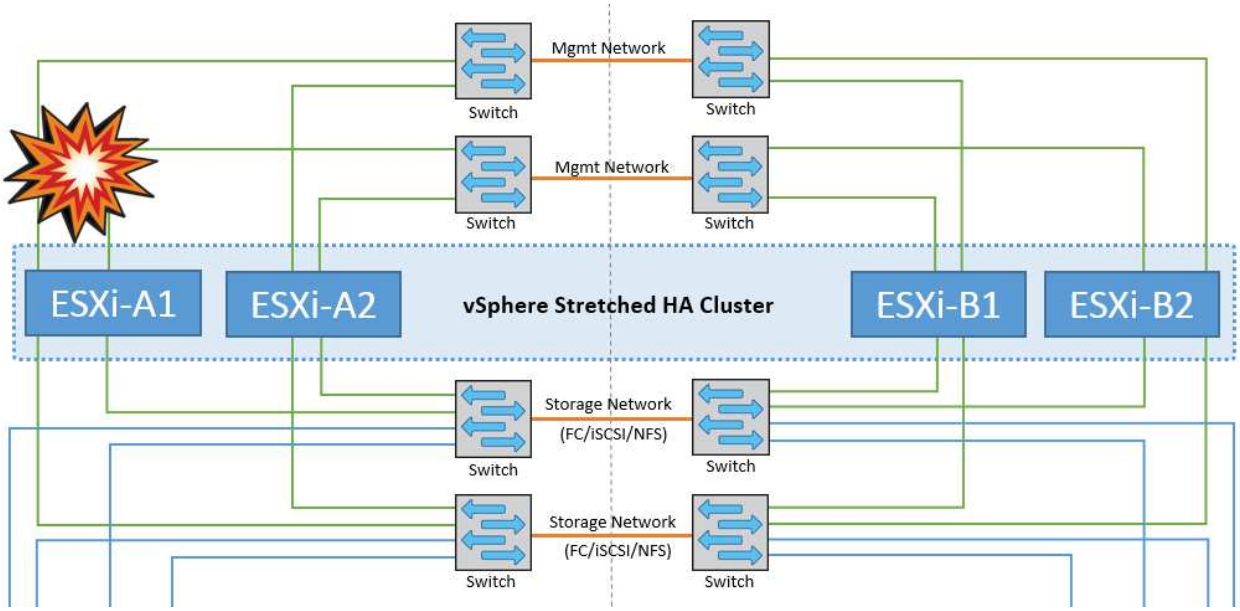
DRSのVMとホストのアフィニティルールが設定されている場合（VMグループsitea_vmsのVMはホストグループsitea_hostsのホストを実行する必要があります）、HAマスターは最初にサイトAで使用可能なリソースを確認します。サイトAに使用可能なホストがない場合、マスターはサイトBのホストでVMの再起動を試みます。

ローカルサイトのリソースに制約がある場合は、もう一方のサイトのESXiホストで仮想マシンが起動される可能性があります。ただし、DRSのVMとホストのアフィニティルールに違反した場合は、仮想マシンをローカルサイトの稼働しているESXiホストに移行することで修正されます。DRSが手動に設定されている場合、NetAppはDRSを起動し、推奨事項を適用して仮想マシンの配置を修正することを推奨します。

このシナリオではMetroClusterの動作に変更はなく、すべてのデータストアがそれぞれのサイトで引き続き実

行されます。

ESXiホストの分離

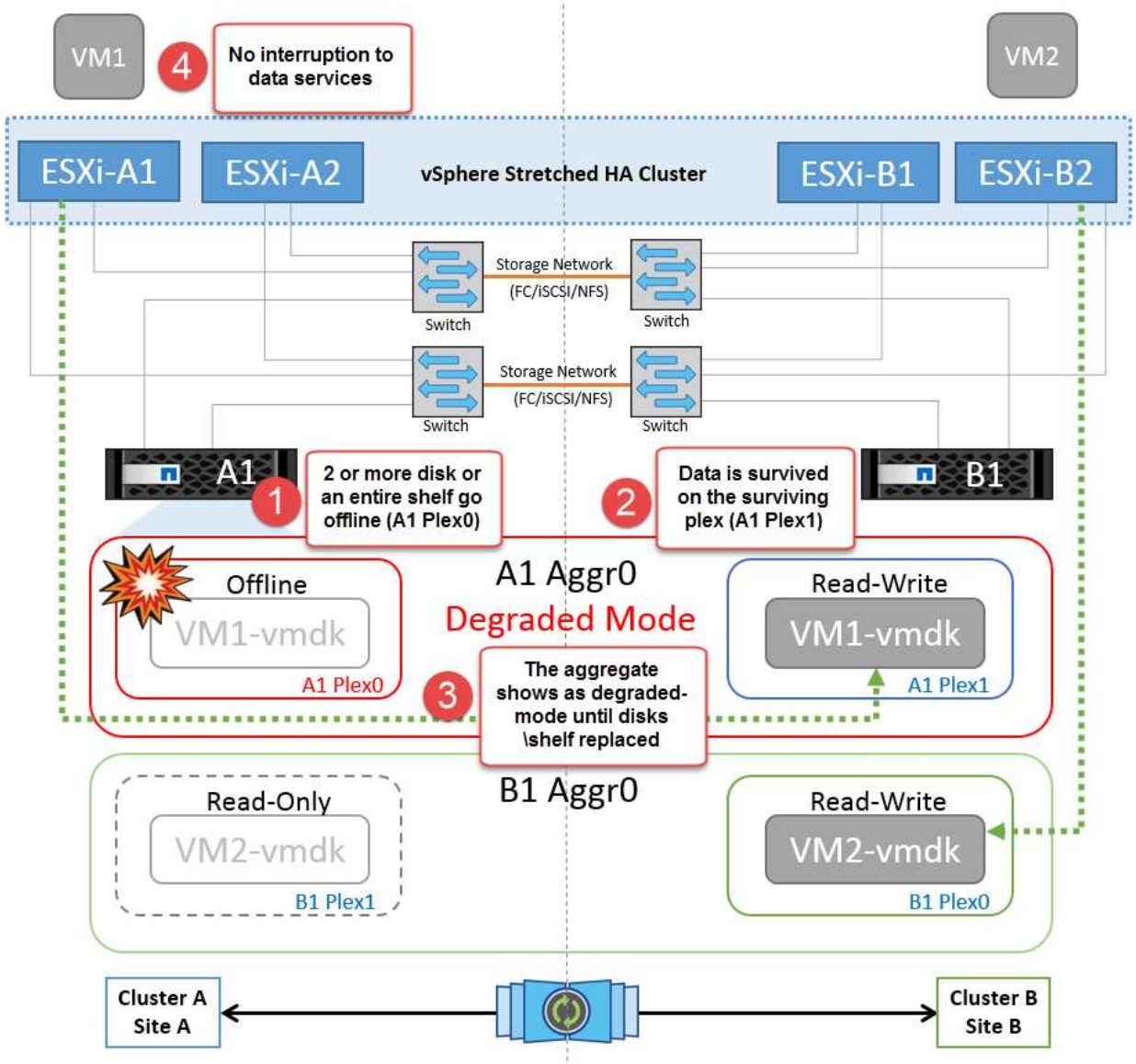


このシナリオでは、ESXiホストの管理ネットワークが停止すると、HAクラスタ内のマスターノードがハートビートを受信しなくなり、このホストがネットワークから分離された状態になります。障害が発生したか、隔離されているだけかを判別するために、マスターノードはデータストアハートビートの監視を開始します。ホストが存在する場合、ホストはマスターノードによって分離されていると宣言されます。構成されている隔離時の対応に応じて、ホストは仮想マシンの電源をオフにするか、シャットダウンするか、仮想マシンの電源をオンにしたままにするかを選択できます。分離応答のデフォルトの間隔は30秒です。

このシナリオではMetroClusterの動作に変更はなく、すべてのデータストアがそれぞれのサイトで引き続き実行されます。

ディスクシェルフの障害

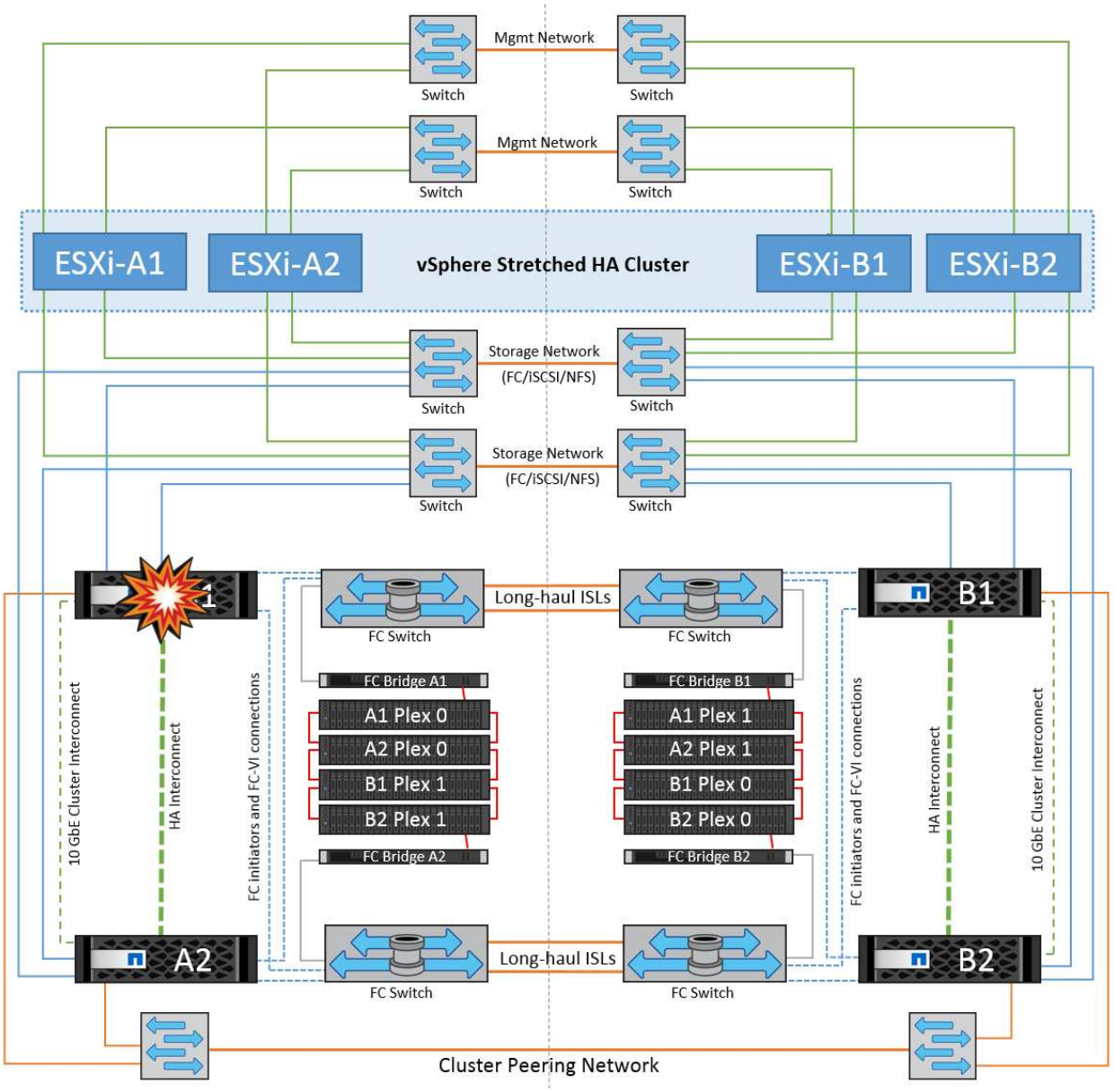
このシナリオでは、3本以上のディスクまたはシェルフ全体で障害が発生しています。データは、データサービスを中断することなく、稼働しているプレックスから提供されます。ディスク障害は、ローカルまたはリモートのプレックスに影響する可能性があります。アクティブなプレックスが1つしかないため、アグリゲートはデグレードモードになります。障害が発生したディスクを交換すると、影響を受けたアグリゲートが自動的に再同期されてデータが再構築されます。再同期後、アグリゲートは自動的に通常のミラーモードに戻ります。単一のRAIDグループ内の3本以上のディスクで障害が発生した場合は、プレックスを最初から再構築する必要があります。



*注：*この間、仮想マシンのI/O処理への影響はありませんが、データはISLリンクを介してリモートのディスクシェルフからアクセスされるため、パフォーマンスが低下します。

単一のストレージコントローラ障害

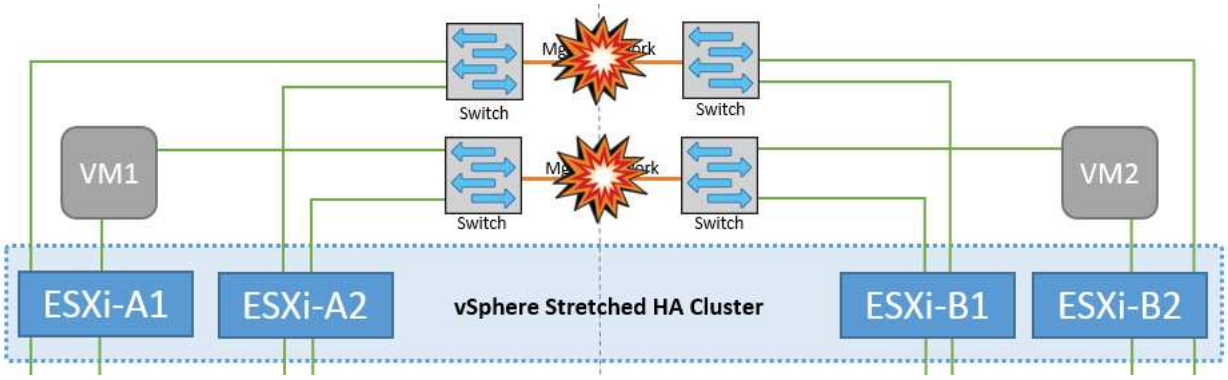
このシナリオでは、一方のサイトの2台のストレージコントローラのどちらかで障害が発生します。各サイトにHAペアがあるため、一方のノードで障害が発生すると、もう一方のノードへのフェイルオーバーが透過的かつ自動的にトリガーされます。たとえば、ノードA1に障害が発生した場合、そのストレージとワークロードは自動的にノードA2に転送されます。すべてのプレックスが引き続き使用可能なため、仮想マシンに影響はありません。2つ目のサイトのノード（B1とB2）は影響を受けません。また、クラスタ内のマスターノードは引き続きネットワークハートビートを受信するため、vSphere HAによる処理は行われません。



フェイルオーバーがローリングディザスタ（ノードA1からA2にフェイルオーバー）の一部である場合に、その後A2またはサイトA全体で障害が発生すると、災害後にサイトBでスイッチオーバーが発生する可能性があります。

スイッチ間リンクの障害

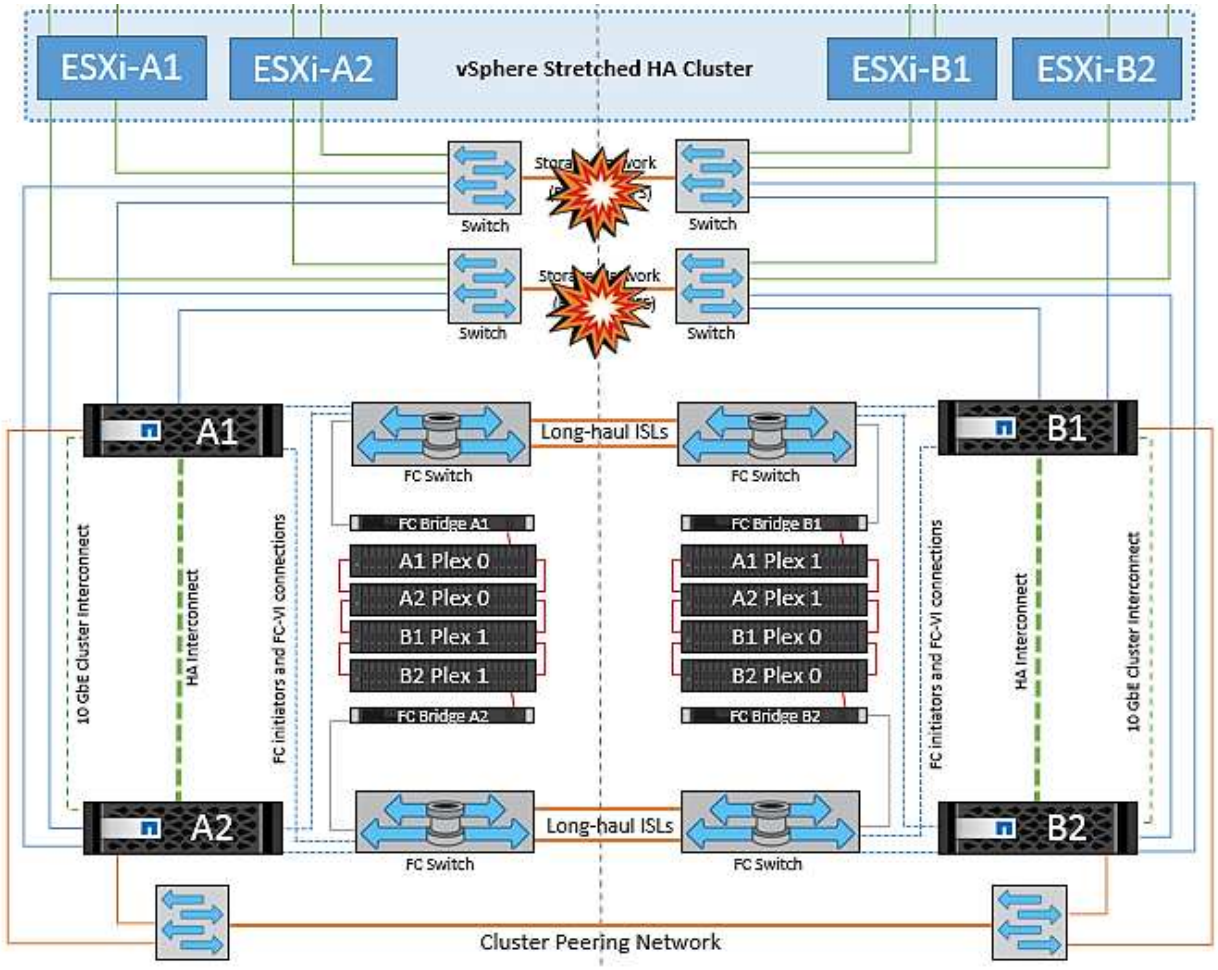
管理ネットワークでのスイッチ間リンク障害



このシナリオでは、フロントエンドホスト管理ネットワークのISLリンクで障害が発生し、サイトAのESXiホストがサイトBのESXiホストと通信できなくなります。これにより、特定のサイトのESXiホストからHAクラスタ内のマスターノードにネットワークハートビートを送信できなくなるため、ネットワークが分割されます。そのため、パーティションのために2つのネットワークセグメントがあり、各セグメントにマスターノードがあり、特定のサイト内でVMがホスト障害から保護されます。

*注：*この間、仮想マシンは実行されたままであり、このシナリオではMetroClusterの動作に変更はありません。すべてのデータストアがそれぞれのサイトで引き続き実行されます。

ストレージネットワークのスイッチ間リンク障害

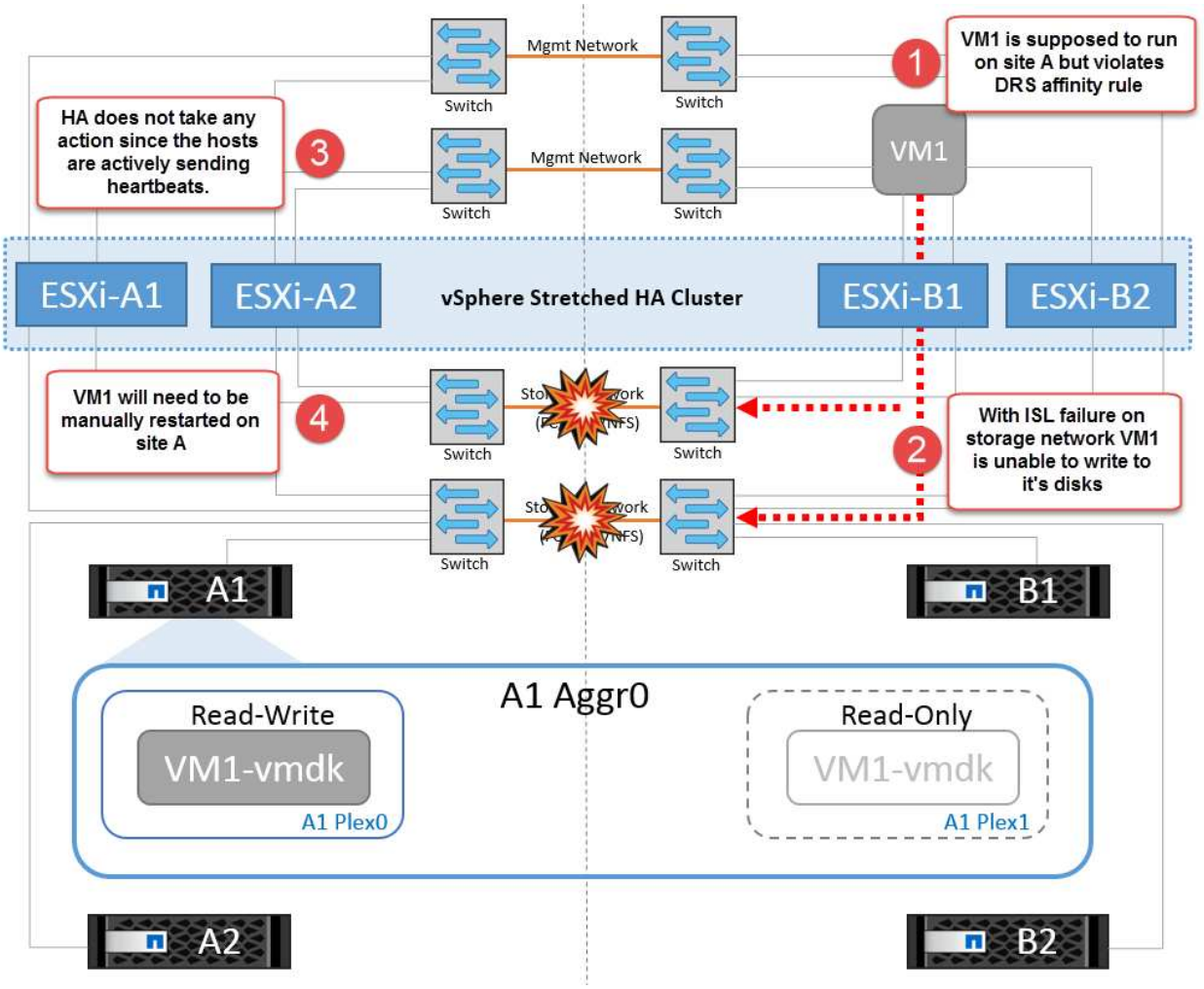


このシナリオでは、バックエンドストレージネットワークのISLリンクで障害が発生すると、サイトAのホストはサイトBのクラスタBのストレージボリュームまたはLUNにアクセスできなくなります。その逆も同様です。VMware DRSルールは、ホストとストレージサイトのアフィニティによって、サイト内で影響を与えるこ

となく仮想マシンを実行できるように定義されています。

この間、仮想マシンはそれぞれのサイトで実行されたままになり、このシナリオではMetroClusterの動作に変更はありません。すべてのデータストアがそれぞれのサイトで引き続き実行されます。

何らかの理由でアフィニティルールに違反した場合（ローカルクラスタAのノードにディスクが配置されているサイトAから実行されていたVM1がサイトBのホストで実行されている場合など）、仮想マシンのディスクにISLリンクを介してリモートからアクセスされます。ISLリンクで障害が発生すると、ストレージボリュームへのパスが停止し、その仮想マシンが停止するため、サイトBで実行されているVM1はディスクに書き込むことができなくなります。この場合、ホストからハートビートがアクティブに送信されるため、VMware HAによる処理は行われません。これらの仮想マシンは、それぞれのサイトで手動で電源をオフにしてオンにする必要があります。次の図は、VMがDRSアフィニティルールに違反していることを示しています。

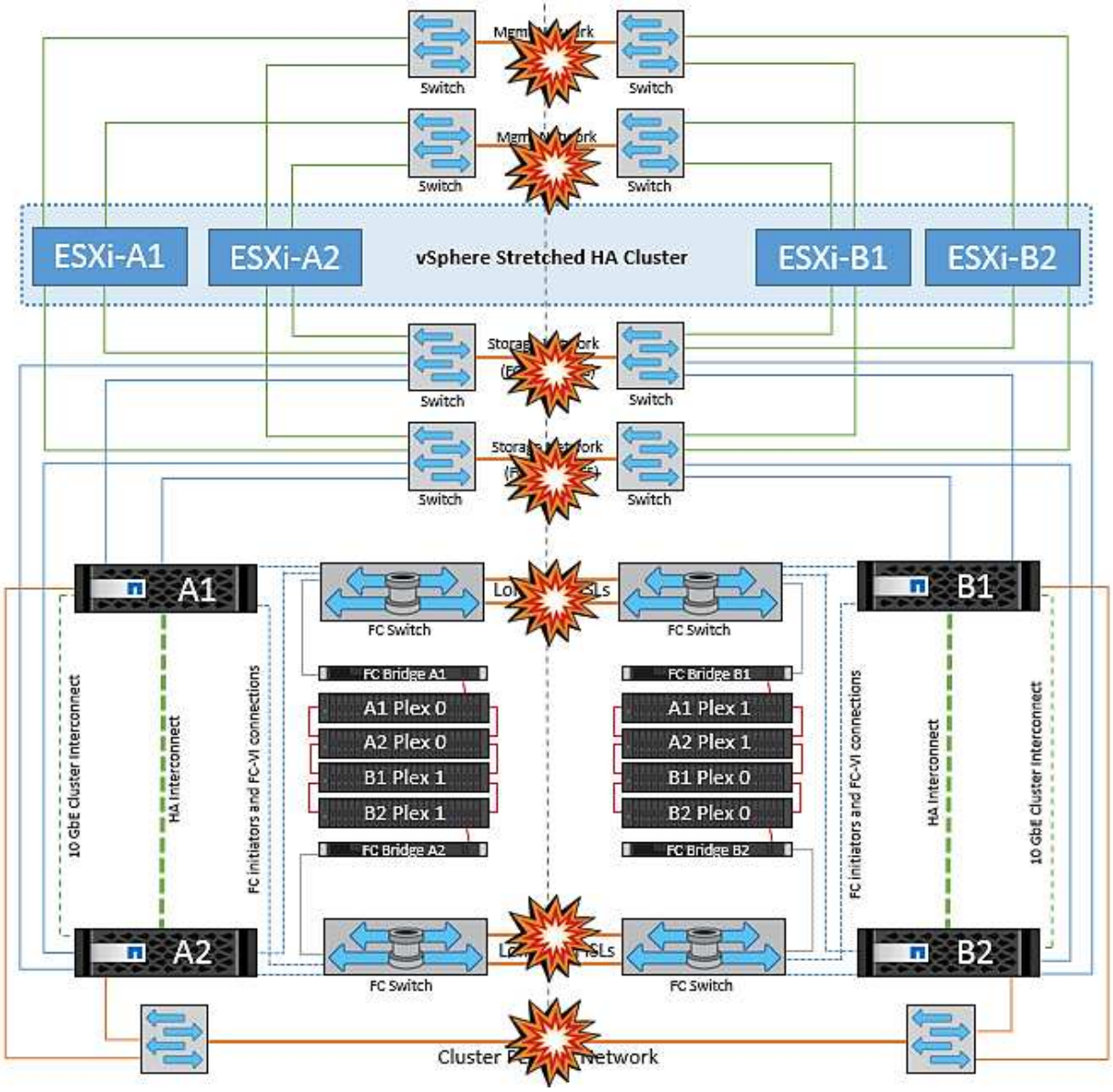


すべてのスイッチ間障害またはデータセンターの完全なパーティション

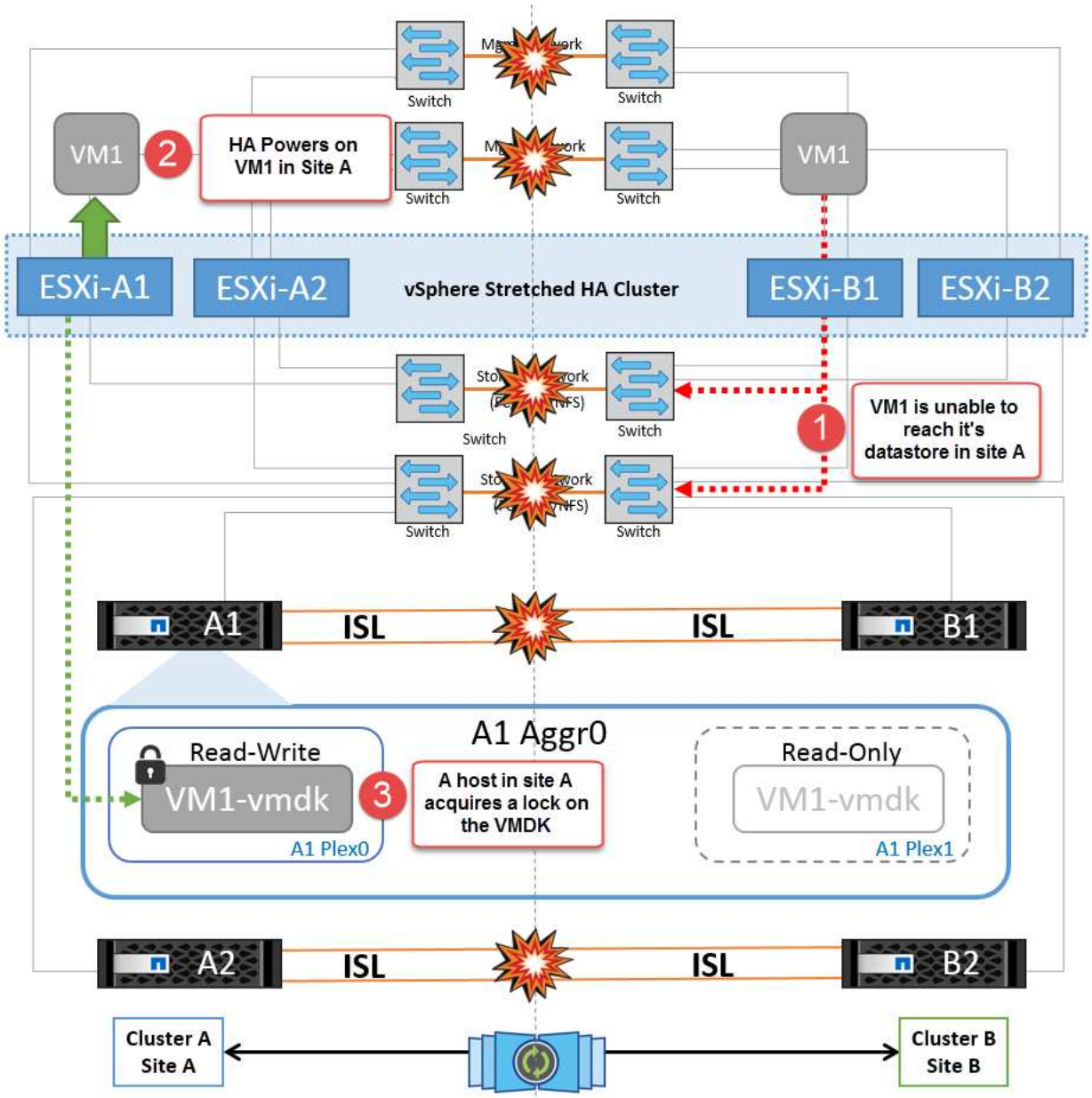
このシナリオでは、サイト間のすべてのISLリンクが停止し、両方のサイトが相互に分離されます。管理ネットワークやストレージネットワークでのISL障害などのシナリオで説明したように、ISL全体で障害が発生しても仮想マシンは影響を受けません。

ESXiホストがサイト間でパーティショニングされると、vSphere HAエージェントがデータストアハートビートをチェックし、各サイトでローカルのESXiホストがデータストアハートビートを対応する読み書き可能なボリューム/ LUNに更新できるようになります。サイトAのホストは、ネットワーク/データストアハートビートがないため、サイトBの他のESXiホストで障害が発生したと見なします。サイトAのvSphere HAはサイトBの仮想マシンの再起動を試行しますが、ストレージISLの障害が原因でサイトBのデータストアにアクセスで

きなくなるため、再起動は失敗します。同様の状況がサイトBでも繰り返されます。



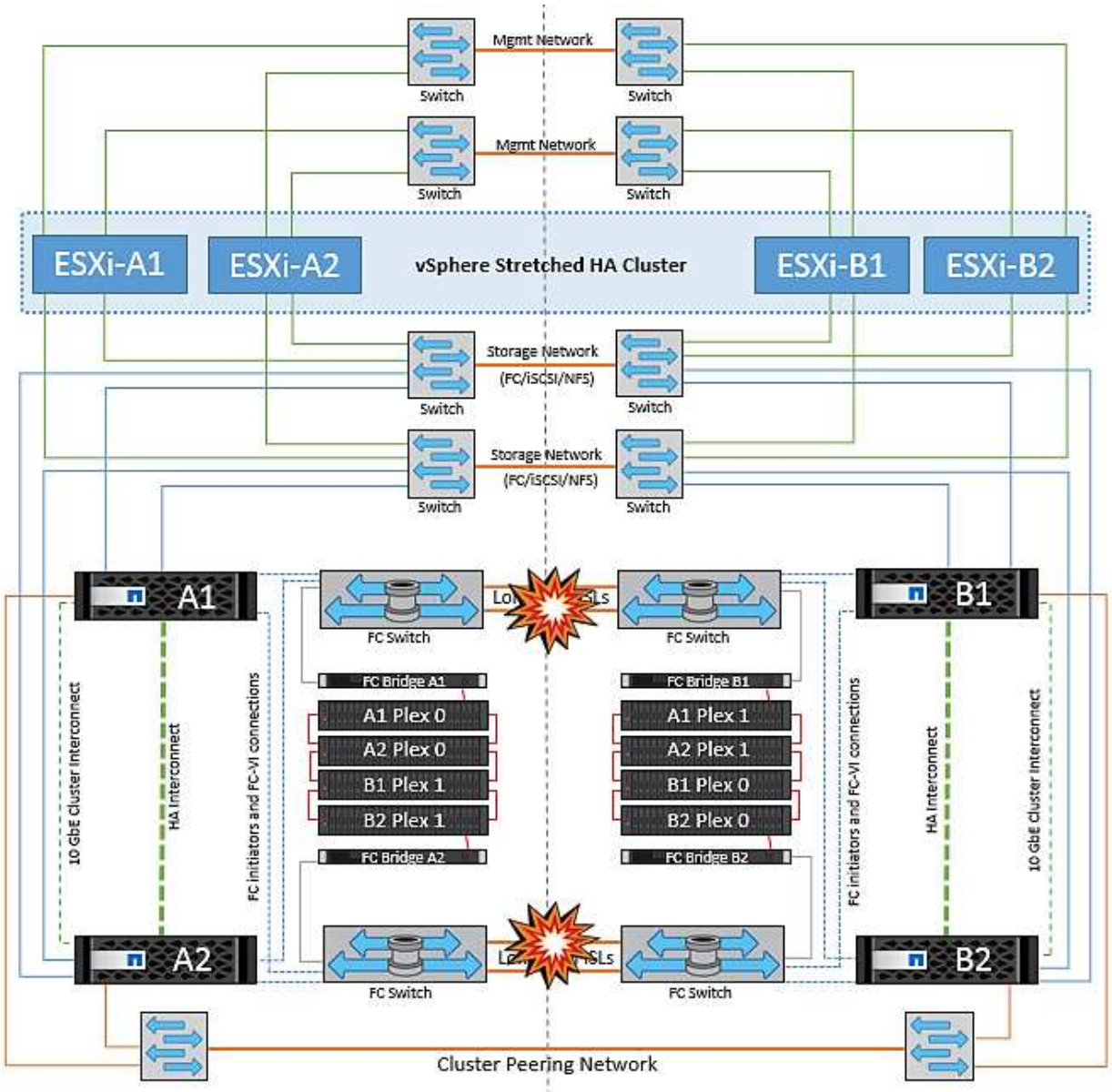
NetAppでは、DRSルールに違反した仮想マシンがないかどうかを確認することを推奨しています。リモートサイトから実行されている仮想マシンはデータストアにアクセスできないため停止し、vSphere HAはその仮想マシンをローカルサイトで再起動します。ISLリンクがオンラインに戻ると、同じMACアドレスで仮想マシンのインスタンスが2つ実行されることはないため、リモートサイトで実行されていた仮想マシンが強制終了されます。



NetApp MetroClusterの両方のファブリックのスイッチ間リンク障害

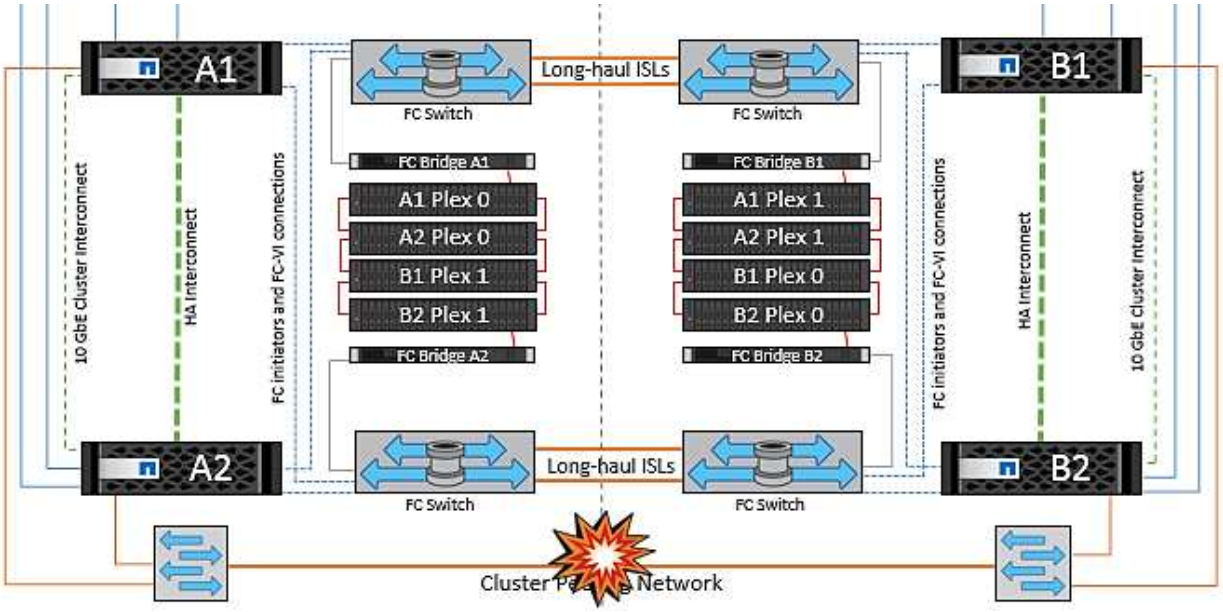
1つ以上のISLで障害が発生した場合、トラフィックは残りのリンクを経由して続行されます。両方のファブリックのすべてのISLで障害が発生し、ストレージとNVRAMのレプリケーション用のサイト間のリンクがなくなった場合、各コントローラはローカルデータの提供を継続します。少なくとも1つのISLをリストアすると、すべてのプレックスの再同期が自動的に実行されます。

すべてのISLが停止したあとに発生した書き込みは、もう一方のサイトにミラーリングされません。そのため、構成がこの状態のときに災害時にスイッチオーバーを実行すると、同期されていなかったデータが失われます。この場合、スイッチオーバー後のリカバリを手動で行う必要があります。ISLが長期間使用できなくなる可能性がある場合は、災害時のスイッチオーバーが必要な場合にデータ損失のリスクを回避するために、すべてのデータサービスをシャットダウンすることができます。この処理を実行するかどうかは、少なくとも1つのISLが使用可能になる前にスイッチオーバーが必要な災害が発生する可能性と比較して判断する必要があります。また、ISLで連鎖的に障害が発生した場合は、すべてのリンクで障害が発生する前に、いずれかのサイトへの計画的スイッチオーバーをトリガーすることもできます。



ピアクラスタのリンク障害

ピアクラスタのリンクで障害が発生した場合、ファブリックのISLはアクティブなままであるため、データサービス（読み取りと書き込み）は両方のサイトで両方のプレックスに対して継続されます。クラスタ設定の変更（新しいSVMの追加、既存のSVMでのボリュームやLUNのプロビジョニングなど）は、もう一方のサイトに伝播できません。これらはローカルのCRSメタデータボリュームに保持され、ピアクラスタリンクのリストア時にもう一方のクラスタに自動的に伝播されます。ピアクラスタのリンクがリストアされる前に強制スイッチオーバーが必要な場合は、スイッチオーバープロセスの一環として、サバイバーサイトにあるメタデータボリュームのリモートレプリケートコピーから、未処理のクラスタ構成変更が自動的に再生されます。



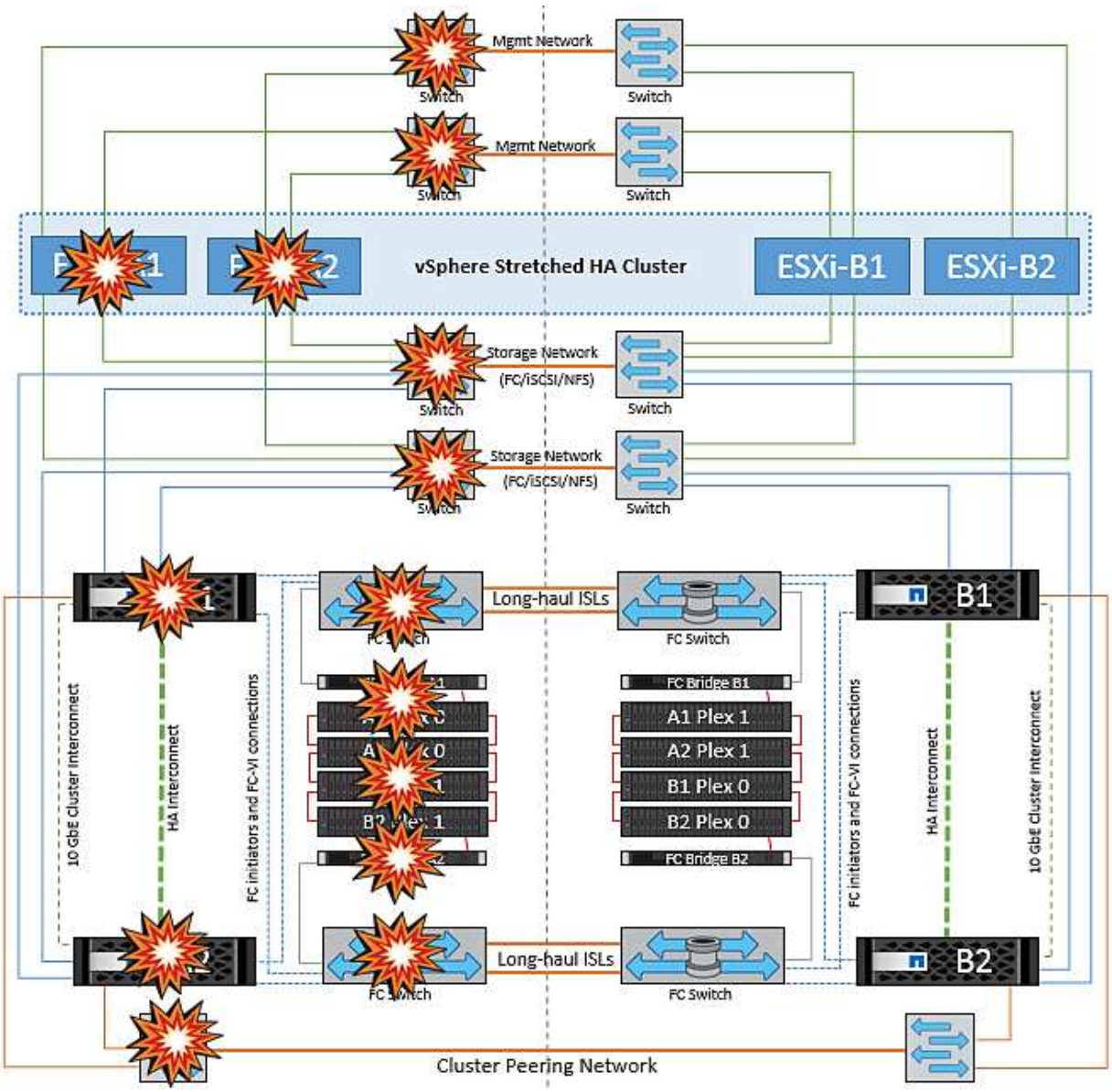
サイト全体の障害

サイトA全体で障害が発生した場合、サイトAのESXiホストが停止しているため、サイトBのESXiホストはサイトAのESXiホストからネットワークハートビートを受信しません。サイトBのHAマスターは、データストアハートビートが存在しないことを確認し、サイトAのホストで障害が発生したことを宣言して、サイトAの仮想マシンをサイトBで再起動しようとします。この間に、ストレージ管理者はスイッチオーバーを実行して障害が発生したノードのサービスをサバイバーサイトで再開し、サイトAのすべてのストレージサービスをサイトBでリストアします。サイトAのボリュームまたはLUNがサイトBで使用可能になると、HAマスターエージェントはサイトAの仮想マシンをサイトBで再起動しようとします。

vSphere HAマスターエージェントがVMの再起動（VMの登録と電源投入を含む）に失敗した場合、遅延後に再起動が再試行されます。再起動の間隔は、最大30分まで設定できます。vSphere HAは、再起動を最大試行回数（デフォルトでは6回）試行します。

注： HAマスターは、Placement Managerが適切なストレージを検出するまで再起動の試行を開始しません。したがって、サイト全体で障害が発生した場合は、スイッチオーバーの実行後に再起動が試行されます。

サイトAがスイッチオーバーされた場合は、サバイバーサイトBのいずれかのノードで障害が発生しても、サバイバーノードにフェイルオーバーすることでシームレスに対応できます。この場合、4つのノードの作業は1つのノードだけで実行されます。この場合のリカバリでは、ローカルノードへのギブバックを実行します。その後、サイトAがリストアされるとスイッチバック処理が実行され、構成の安定した運用が再開されます。



製品のセキュリティ

VMware vSphere 用の ONTAP ツール

ONTAP Tools for VMware vSphereを使用したソフトウェアエンジニアリングでは、次のセキュアな開発アクティビティを採用しています。

- * 脅威モデリング。 * 脅威モデリングの目的は、ソフトウェア開発ライフサイクルの早い段階で、機能、コンポーネント、または製品のセキュリティ上の欠陥を発見することです。脅威モデルとは、アプリケーションのセキュリティに影響するすべての情報を構造化したものです。本質的に、これはセキュリティの観点から見たアプリケーションとその環境です。
- * Dynamic Application Security Testing （ DAST ）。 * このテクノロジーは、実行中のアプリケーションで脆弱な状態を検出するように設計されています。DAST は、 Web 対応アプリケーションの公開 HTTP および HTML インターフェイスをテストします。
- * サードパーティーのコード通貨。 * オープンソース・ソフトウェア（ OSS ）を使用したソフトウェア開発の一環として、製品に組み込まれた OSS に関連するセキュリティ上の脆弱性に対処する必要がありま

す。これは継続的な取り組みです。新しい OSS バージョンには、いつでも新たに検出された脆弱性が報告される可能性があります。

- * 脆弱性スキャン。 * 脆弱性スキャンは、お客様にリリースされる前にネットアップ製品の一般的なセキュリティの脆弱性と既知のセキュリティの脆弱性を検出するためのものです。
- * ペネトレーションテスト。 * ペネトレーションテストは、システム、 Web アプリケーション、またはネットワークを評価して、攻撃者によって悪用される可能性のあるセキュリティの脆弱性を検出するプロセスです。ネットアップでのペネトレーションテスト（ペンテスト）は、承認された信頼できる第三者企業のグループが実施します。テスト範囲には、高度な攻撃方法やツールを使用した悪意のある侵入者やハッカーと同様のアプリケーションまたはソフトウェアに対する攻撃の開始が含まれます。

製品のセキュリティ機能

ONTAP Tools for VMware vSphereの各リリースには、次のセキュリティ機能が含まれています。

- * ログインバナー。 * SSH はデフォルトでは無効になっており、 VM コンソールから有効になっている場合は 1 回限りのログインしか許可されません。ユーザがログインプロンプトでユーザ名を入力すると、次のログインバナーが表示されます。
- 警告： * このシステムへの不正アクセスは禁止されており、法律で訴追されます。このシステムにアクセスすることで、不正な使用が疑われる場合に、ユーザーのアクションが監視される可能性があることに同意したものとみなされます。

  ユーザがSSHチャネルを介したログインを完了すると、次のテキストが表示されます。

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- * ロールベースアクセス制御 (RBAC) 。 * ONTAP ツールには、次の 2 種類の RBAC 制御が関連付けられています。
  - vCenter Server 標準の権限
  - vCenter プラグインに固有の権限。詳細については、を参照してください "リンクをクリックしてください"。
- * 暗号化された通信チャネル。 * すべての外部通信は、バージョン 1.2 の TLS を使用して HTTPS 経由で行われます。
- * 最小限のポート露出。 * 必要なポートのみがファイアウォールで開かれています。

  次の表に、オープンポートの詳細を示します。

TCP v4 / V6 ポート番号	方向（ **Direction** ）	機能
8143	インバウンド	REST API 用の HTTPS 接続
8043	インバウンド	HTTPS 接続

TCP v4 / V6 ポート番号	方向（ Direction ）	機能
9060	インバウンド	HTTPS 接続 SOAP over https 接続に使用されます クライアントがONTAP tools APIサーバに接続できるようにするには、このポートを開く必要があります。
22	インバウンド	SSH （デフォルトでは無効）
9080	インバウンド	HTTPS 接続 - VP および SRA - ループバックからの内部接続のみ
9083年だ	インバウンド	HTTPS 接続 - VP および SRA SOAP over https 接続に使用されます
一一六二	インバウンド	VP SNMP トラップパケット
1527年	内部のみ	Derby データベースポート。このコンピュータとそれ自体の間のみ、外部接続は許可されません — 内部接続のみ
443	双方向	ONTAP クラスタへの接続に使用します

- * 認証局（ CA ）署名証明書のサポート。 * VMware vSphere 用の ONTAP ツールは CA 署名証明書をサポートしています。を参照してください "こちらの技術情報アーティクル" を参照してください。
- * 監査ログ。 * サポートバンドルはダウンロード可能で、非常に詳細です。ONTAP ツールは、すべてのユーザログインおよびログアウトアクティビティを個別のログファイルに記録します。VASA API 呼び出しは、専用の VASA 監査ログ（ローカルの cxf.log ）に記録されます。
- * パスワードポリシー。 * 次のパスワードポリシーが適用されます。
  - パスワードはどのログファイルにも記録されません。
  - パスワードはプレーンテキストで伝達されません。
  - パスワードは、インストールプロセスで設定します。
  - パスワード履歴は設定可能なパラメータです。
  - パスワードの最小有効期間は 24 時間に設定されます。
  - パスワードフィールドの自動入力は無効です。
  - ONTAP ツールは、保存されているすべてのクレデンシャル情報を SHA256 ハッシュで暗号化し

SnapCenterプラグインVMware vSphere

NetApp SnapCenter Plug-in for VMware vSphereのソフトウェアエンジニアリングでは、次のような安全な開発作業を行います。

- * 脅威モデリング。 * 脅威モデリングの目的は、ソフトウェア開発ライフサイクルの早い段階で、機能、コンポーネント、または製品のセキュリティ上の欠陥を発見することです。脅威モデルとは、アプリケー

ションのセキュリティに影響するすべての情報を構造化したものです。本質的に、これはセキュリティの観点から見たアプリケーションとその環境です。

- *動的アプリケーションセキュリティテスト(DAST)。*実行中のアプリケーションの脆弱な状態を検出するように設計されたテクノロジ。DAST は、 Web 対応アプリケーションの公開 HTTP および HTML インターフェイスをテストします。
- *サードパーティのコード通貨。*ソフトウェアの開発およびオープンソースソフトウェア（OSS）の使用の一環として、製品に組み込まれているOSSに関連するセキュリティの脆弱性に対処することが重要です。これは、OSSコンポーネントのバージョンに、いつでも新たに検出された脆弱性が報告される可能性があるため、継続的な取り組みです。
- * 脆弱性スキャン。 * 脆弱性スキャンは、お客様にリリースされる前にネットアップ製品の一般的なセキュリティの脆弱性と既知のセキュリティの脆弱性を検出するためのものです。
- *ペネトレーションテスト。*ペネトレーションテストは、システム、Webアプリケーション、またはネットワークを評価して、攻撃者によって悪用される可能性のあるセキュリティの脆弱性を検出するプロセスです。ネットアップでのペネトレーションテスト（ペンテスト）は、承認された信頼できる第三者企業のグループが実施します。このテスト範囲には、高度な攻撃方法やツールを使用した悪意のある侵入者やハッカーなどのアプリケーションやソフトウェアに対する攻撃の開始が含まれます。
- *製品セキュリティインシデント対応アクティビティ。*セキュリティの脆弱性は社内外で発見され、タイムリーに対処しなければ、ネットアップの評判に深刻なリスクをもたらす可能性があります。このプロセスを容易にするために、Product Security Incident Response Team（PSIRT）は脆弱性を報告して追跡します。

製品のセキュリティ機能

NetApp SnapCenter Plug-in for VMware vSphereの各リリースには、次のセキュリティ機能が含まれています。

- 制限付きシェルアクセス。 SSHはデフォルトで無効になっており、1回限りのログインはVMコンソールから有効にした場合にのみ許可されます。
- *ログインバナーのアクセス警告*ログインプロンプトにユーザ名を入力すると、次のログインバナーが表示されます。
- 警告： * このシステムへの不正アクセスは禁止されており、法律で訴追されます。このシステムにアクセスすることで、不正な使用が疑われる場合に、ユーザーのアクションが監視される可能性があることに同意したものとみなされます。

  ユーザがSSHチャネルを介したログインを完了すると、次の出力が表示されます。

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- * ロールベースアクセス制御 (RBAC) 。 * ONTAP ツールには、次の 2 種類の RBAC 制御が関連付けられています。
  - vCenter Server標準の権限。

- VMware vCenterプラグインに固有の権限。詳細については、を参照してください "ロールベースアクセス制御（ RBAC ）"。

- *暗号化された通信チャネル。*すべての外部通信は、TLSを使用してHTTPS経由で行われます。
- * 最小限のポート露出。 * 必要なポートのみがファイアウォールで開かれています。

次の表に、オープンポートの詳細を示します。

TCP v4 / V6ポート番号	機能
8144	REST API 用の HTTPS 接続
8080 です	OVA GUIでのHTTPS接続
22	SSH（デフォルトでは無効）
3306	mysql（内部接続のみ。外部接続はデフォルトで無効）
443	nginx（データ保護サービス）

- 認証局（**CA**）署名証明書のサポート。 SnapCenter Plug-in for VMware vSphereは、CA署名証明書の機能をサポートしています。を参照してください "SnapCenter Plug-in for VMware vSphere（SCV）にSSL証明書を作成/インポートする方法"。
- *パスワードポリシー。*次のパスワードポリシーが有効です。
  - パスワードはどのログファイルにも記録されません。
  - パスワードはプレーンテキストで伝達されません。
  - パスワードは、インストールプロセスで設定します。
  - クレデンシャル情報はすべてSHA256ハッシュを使用して保存されます。
- *基本オペレーティングシステムイメージ。*この製品は、アクセス制限とシェルアクセスが無効になったOVA用のDebianベースOSに同梱されています。これにより、攻撃のフットプリントが削減されます。すべてのSnapCenter リリースベースのオペレーティングシステムには、最大限のセキュリティを適用できる最新のセキュリティパッチが適用されています。

ネットアップでは、SnapCenter Plug-in for VMware vSphereアプライアンスに関連するソフトウェア機能およびセキュリティパッチを開発し、その後、バンドルソフトウェアプラットフォームとしてお客様にリリースします。ネットアップでは、これらのアプライアンスにはLinuxのサブシステムに固有の依存関係と独自のソフトウェアが含まれているため、サブオペレーティングシステムを変更しないことを推奨します。これは、ネットアップアプライアンスに影響を及ぼす可能性が高いためです。これは、ネットアップがアプライアンスをサポートできるかどうかに影響します。アプライアンスはセキュリティ関連の問題にパッチを適用するためにリリースされているため、最新のコードバージョンをテストして導入することを推奨します。

『Security Hardening Guide for ONTAP tools for VMware vSphere』

『Security Hardening Guide for ONTAP tools for VMware vSphere』

『ONTAP tools for VMware vSphereセキュリティ強化ガイド』には、最もセキュアな設定を行うための包括的な手順が記載されています。

これらのガイドは、アプライアンス自体のアプリケーションとゲストOSの両方に適用されます。

ONTAP Tools for VMware vSphereインストールパッケージの整合性の検証

ONTAP toolsインストールパッケージの整合性を検証するには、2つの方法があります。

1. チェックサムの確認
2. シグネチャの検証

チェックサムは、OTVインストールパッケージのダウンロードページで提供されています。ダウンロードしたパッケージのチェックサムを、ダウンロードページに表示されているチェックサムと照合して確認する必要があります。

ONTAP tools OVAの署名の確認

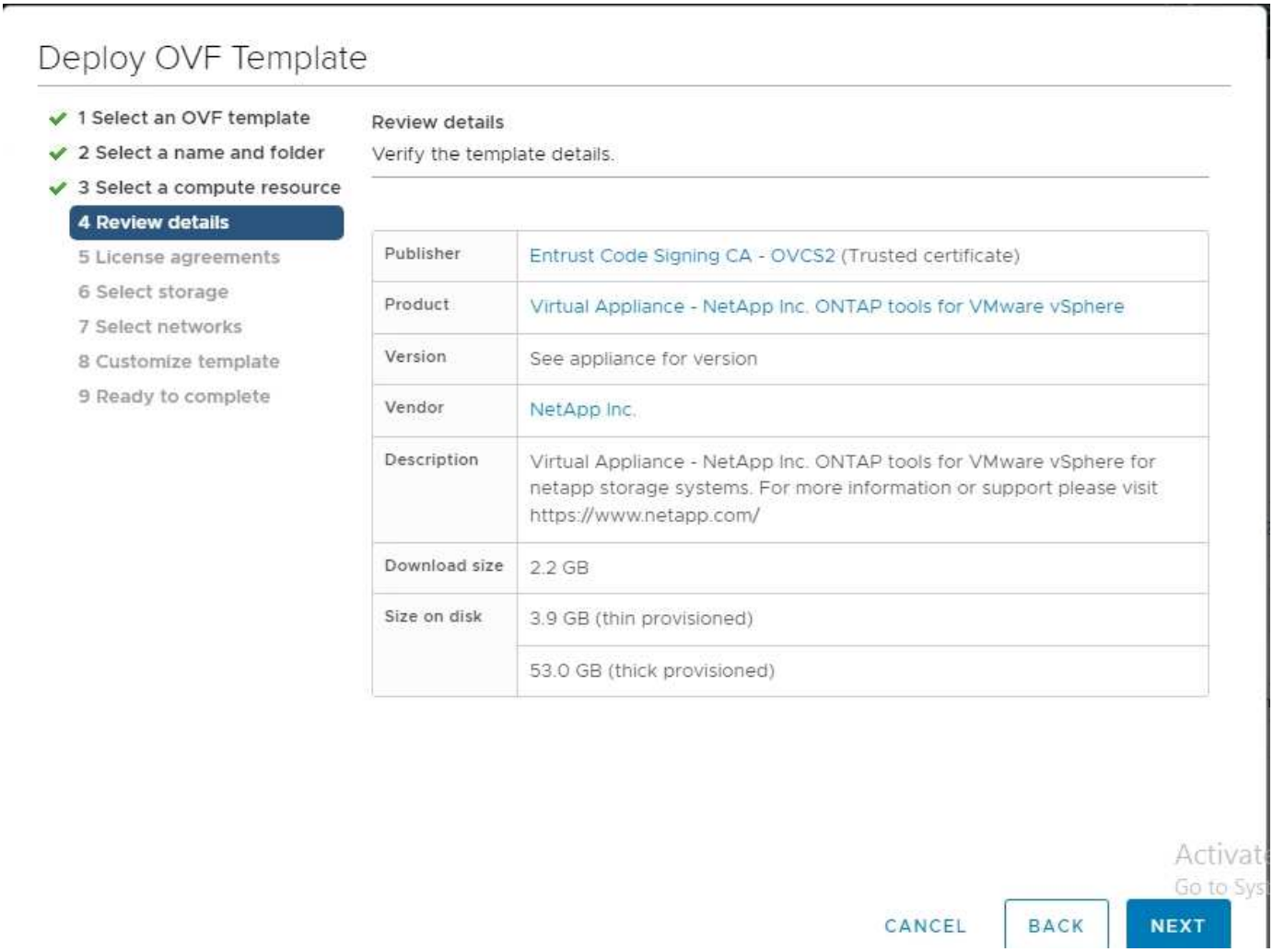
vAppインストールパッケージはtarball形式で提供されます。このtarballには、仮想アプライアンスの中間証明書とルート証明書、READMEファイル、OVAパッケージが含まれています。READMEファイルには、vApp OVAパッケージの整合性を検証する方法が記載されています。

また、提供されたルート証明書と中間証明書をvCenterバージョン7.0U3E以降にアップロードする必要があります。 vCenterのバージョン7.0.1から7.0.U3Eの場合、証明書を検証する機能はVMwareではサポートされていません。vCenterバージョン6.xの証明書はアップロードする必要はありません。

信頼されたルート証明書の**vCenter**へのアップロード

1. VMware vSphere ClientでvCenter Serverにログインします。
2. administrator@vsphere.localまたはvCenter Single Sign-On Administratorsグループの別のメンバーのユーザ名とパスワードを指定します。インストール時に別のドメインを指定した場合は、administrator@mydomainとしてログインします。
3. 証明書管理ユーザーインターフェイスに移動します。a.[ホーム]メニューから[管理]を選択します。B[証明書]で、[証明書管理]をクリックします。
4. プロンプトが表示されたら、vCenter Serverのクレデンシャルを入力します。
5. [信頼されたルート証明書]で、[追加]をクリックします。
6. [browse]をクリックし、証明書の.pemファイル（otv_ova_inter_root_cert_chain.pem）の場所を選択します。
7. 追加をクリックします。証明書がストアに追加されます。

を参照してください "証明書ストアへの信頼されたルート証明書の追加" を参照してください。（OVAファイルを使用して）vAppを導入する際、vAppパッケージのデジタル署名は[Review details]ページで確認できます。ダウンロードしたvAppパッケージが正規のものである場合は、[発行者]列に[信頼された証明書]と表示されます（次のスクリーンショットを参照）。



ONTAP tools ISOおよび**SRA tar.gz**の署名の確認

NetAppは、製品ダウンロードページでコード署名証明書をお客様と共有し、OTV-ISOおよびsra.tgzの製品zipファイルも提供しています。

コード署名証明書から、ユーザーは次のように公開鍵を抽出できます。

```
 #> openssl x509 -in <code-sign-cert, pem file> -pubkey -noout > <public-key name>
```

公開鍵を使用して、以下のように`ISO`および`tgz`製品`zip`の署名を検証する必要があります。

```
 #> openssl dgst -sha256 -verify <public-key> -signature <signature-file> <binary-name>
```

例

```
#> openssl x509 -in OTV_ISO_CERT.pem -pubkey -noout > OTV_ISO.pub
#> openssl dgst -sha256 -verify OTV_ISO.pub -signature netapp-ontap-tools-
for-vmware-vsphere-9.12-upgrade-iso.sig netapp-ontap-tools-for-vmware-
vsphere-9.12-upgrade.iso
Verified OK  => response
```

ポートとプロトコル

ここでは、ONTAP tools for VMware vSphereサーバと、管理対象のストレージシステム、サーバ、その他のコンポーネントなどのエンティティ間の通信に必要なポートとプロトコルを示します。

OTVに必要なインバウンドおよびアウトバウンドポート

次の表に、ONTAP toolsが適切に機能するために必要なインバウンドポートとアウトバウンドポートを示します。表に記載されているポートだけがリモートマシンからの接続用に開いていることを確認し、他のすべてのポートはリモートマシンからの接続用にブロックする必要があります。これにより、システムのセキュリティと安全性が確保されます。

次の表に、オープンポートの詳細を示します。

* TCP v4/V6ポート番号*	* 方向 *	機能
8143	インバウンド	REST API 用の HTTPS 接続
8043	インバウンド	HTTPS 接続
9060	インバウンド	HTTPS接続+ SOAP over HTTPS接続に使用+ クライアントがONTAP tools APIサーバに接続できるようにするには、このポートを開く必要があります。
22	インバウンド	SSH （デフォルトでは無効）
9080	インバウンド	HTTPS 接続 - VP および SRA - ループバックからの内部接続のみ
9083年だ	インバウンド	HTTPS接続- VPおよびSRA+ SOAP over HTTPS接続に使用
一一六二	インバウンド	VP SNMP トラップパケット
8443	インバウンド	リモートプラグイン
1527年	内部のみ	Derbyデータベースポート、このコンピュータとそれ自体の間のみ、外部接続は許可されません-内部接続のみ
8150	内部のみ	ログ整合性サービスはポートで実行されます
443	双方向	ONTAP クラスタへの接続に使用します

Derbyデータベースへのリモートアクセスの制御

管理者は、次のコマンドを使用してDerbyデータベースにアクセスできます。ONTAP toolsのローカルVMとリ

モートサーバからアクセスするには、次の手順を実行します。

```
java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
connect 'jdbc:derby://<OTV-
IP>:1527//opt/netapp/vpserver/vvoldb;user=<user>;password=<password>';
```

例：

```
root@UnifiedVSC:~# java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
ij version 10.15
ij> connect 'jdbc:derby://localhost:1527//opt/netapp/vpserver/vvoldb;user=app;password=[redacted]';
ij> show tables;
TABLE_SCHEM         |TABLE_NAME                     |REMARKS
------------------------------------------------------------------------
SYS                 |SYSALIASES                     |
SYS                 |SYSCHECKS                      |
SYS                 |SYSCOLPERMS                    |
SYS                 |SYSCOLUMNS                     |
SYS                 |SYSCONGLOMERATES               |
SYS                 |SYSCONSTRAINTS                 |
SYS                 |SYSDEPENDS                     |
SYS                 |SYSFILES                       |
SYS                 |SYSFOREIGNKEYS                 |
SYS                 |SYSKEYS                        |
SYS                 |SYSPERMS                       |
```

ONTAP Tools for VMware vSphereアクセスポイント（ユーザ）

ONTAP Tools for VMware vSphereをインストールすると、次の3種類のユーザが作成され、使用されます。

1. システムユーザ：rootユーザアカウント
2. アプリケーションユーザ：管理者ユーザ、maintユーザ、およびdbユーザアカウント
3. サポートユーザ：diagユーザアカウント

1.システムユーザ

システム(root)ユーザは、基盤となるオペレーティングシステム(Debian)にインストールされたONTAPツールによって作成されます。

- ONTAP toolsのインストールにより、デフォルトのシステムユーザ"root"がDebian上に作成されます。デフォルトでは無効になっており、「メンテナンス」コンソールから個別に有効にすることができます。

2.アプリケーションユーザ

ONTAP toolsでは、アプリケーションユーザの名前はローカルユーザです。これらは、ONTAP toolsアプリケーションで作成されたユーザです。次の表に、アプリケーションユーザのタイプを示します。

* ユーザー *	* 概要 *
管理者ユーザ	ONTAP toolsのインストール時に作成され、ONTAP toolsの導入時にユーザがクレデンシャルを指定します。ユーザは「maint」コンソールで「password」を変更できます。パスワードの有効期限は90日で、ユーザは同じパスワードを変更する必要があります。

* ユーザー *	* 概要 *
メンテナンスユーザ	ONTAP toolsのインストール時に作成され、ONTAP toolsの導入時にユーザがクレデンシャルを指定します。ユーザは「maint」コンソールで「password」を変更できます。これはメンテナンスユーザで、メンテナンスコンソールの処理を実行するために作成されます。
データベースユーザ	ONTAP toolsのインストール時に作成され、ONTAP toolsの導入時にユーザがクレデンシャルを指定します。ユーザは「maint」コンソールで「password」を変更できます。パスワードの有効期限は90日で、ユーザは同じパスワードを変更する必要があります。

3.サポートユーザ（**diag**ユーザ）

ONTAP toolsのインストール中に、サポートユーザが作成されます。このユーザを使用して、サーバで問題や停止が発生した場合にONTAPツールにアクセスしたり、ログを収集したりできます。デフォルトでは、このユーザは無効になっていますが、「メンテナンス」コンソールからアドホックで有効にすることができます。このユーザーは一定期間後に自動的に無効になることに注意することが重要です。

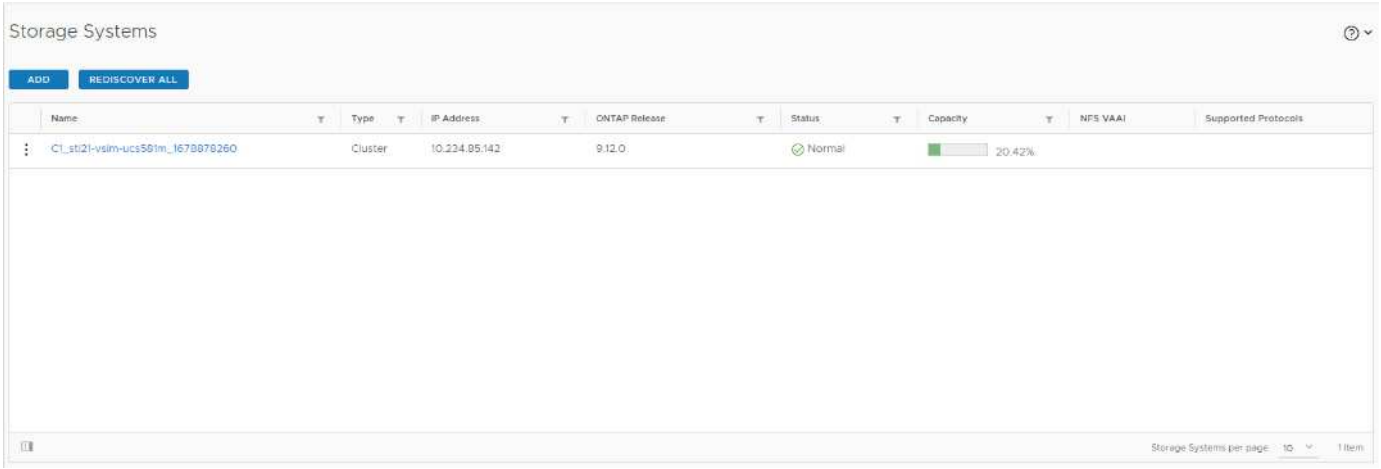
相互**TLS**（証明書ベースの認証）

ONTAPバージョン9.7以降では、相互TLS通信がサポートされます。ONTAP Tools for VMwareおよびvSphere 9.12以降では、新しく追加したクラスタとの通信に相互TLSが使用されます（ONTAPのバージョンによって異なります）。

ONTAP

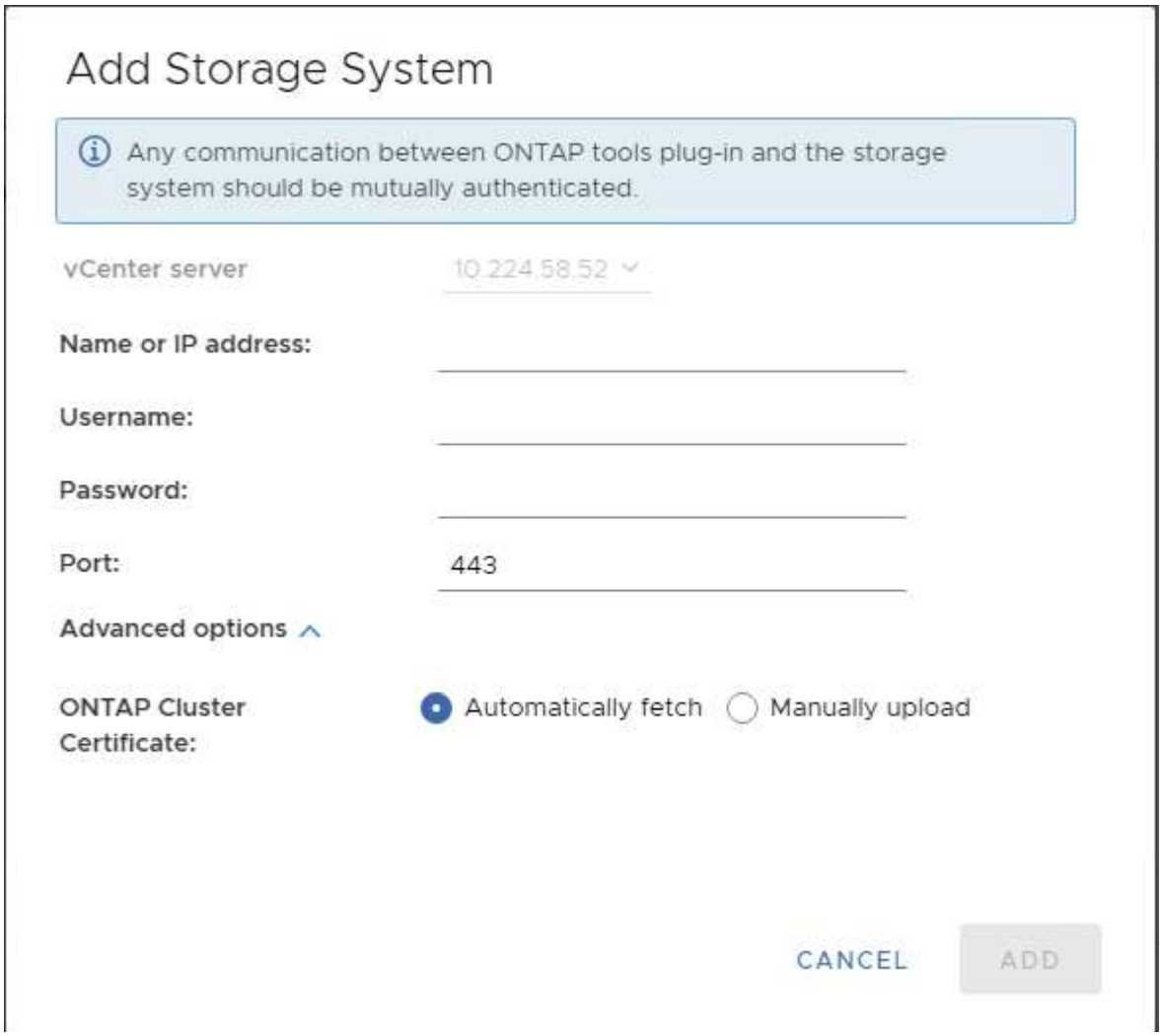
以前に追加されたすべてのストレージシステム：アップグレード中に、追加されたすべてのストレージシステムが自動信頼され、証明書ベースの認証メカニズムが設定されます。

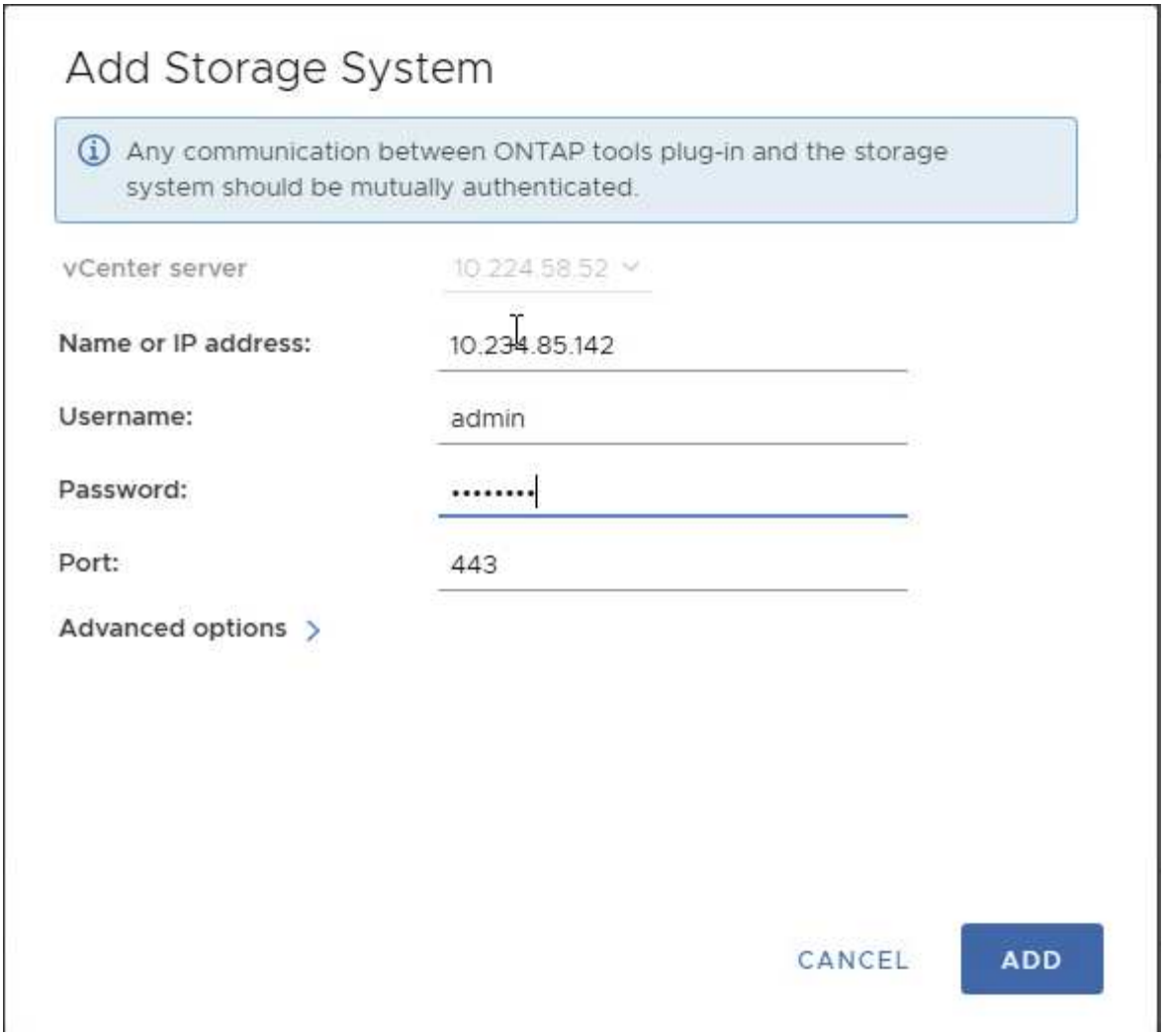
下のスクリーンショットのように、[クラスタセットアップ]ページには、各クラスタに対して設定されたMutual TLS（証明書ベースの認証）のステータスが表示されます。

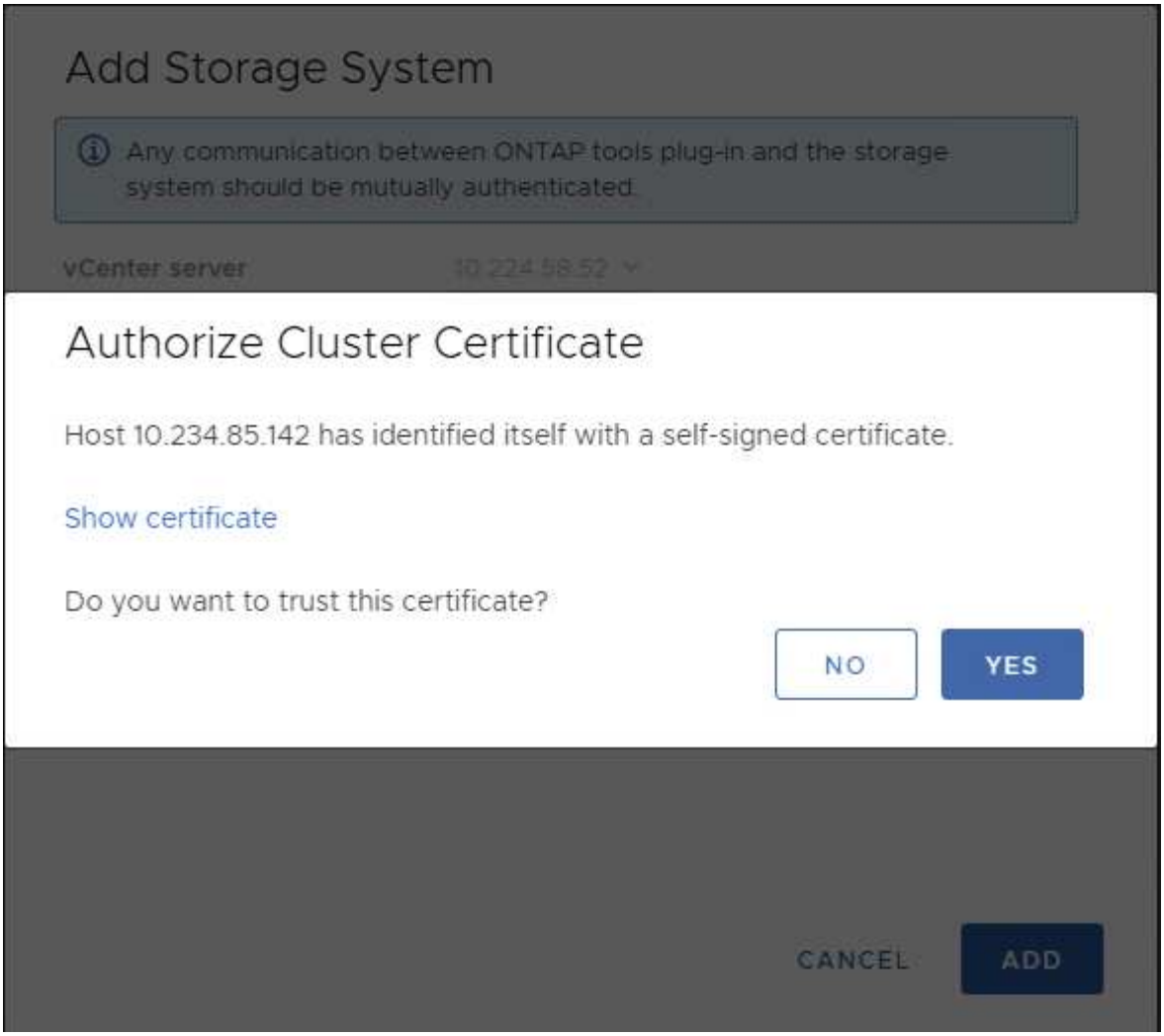


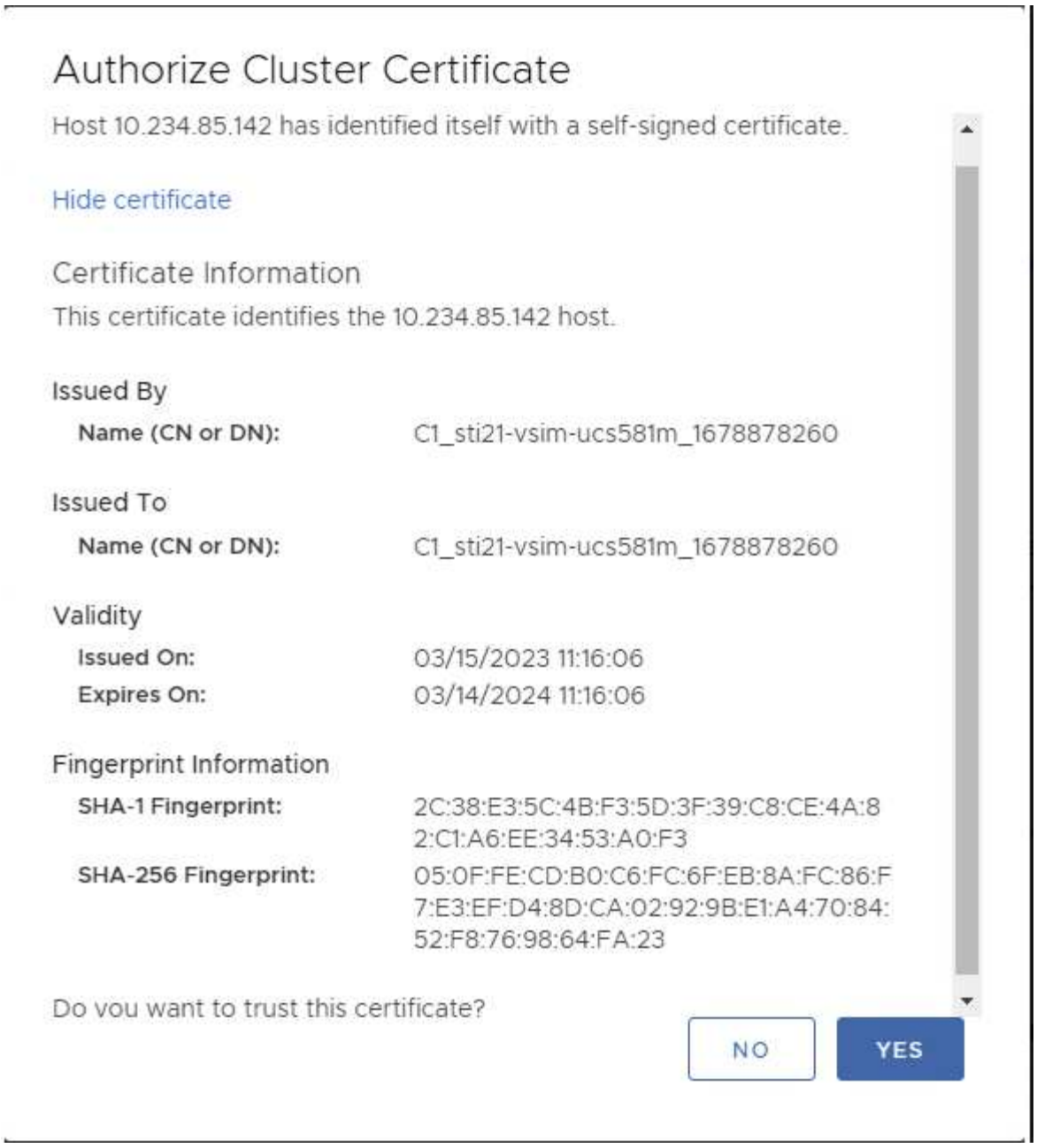
クラスタの追加

クラスタ追加のワークフロー中に、追加するクラスタがMTLSをサポートしている場合、MTLSはデフォルトで設定されます。ユーザはこの設定を行う必要はありません。次のスクリーンショットは、クラスタの追加時にユーザに表示される画面を示しています。


Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
10.224.58.52
Name or IP address:
Username:
Password:
Port:
443
Advanced options
ONTAP Cluster Certificate:
Automatically fetch
Manually upload
CANCEL
ADD


Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
10.224.58.52
Name or IP address:
10.234.85.142
Username:
admin
Password:
Port:
443
Advanced options
CANCEL
ADD

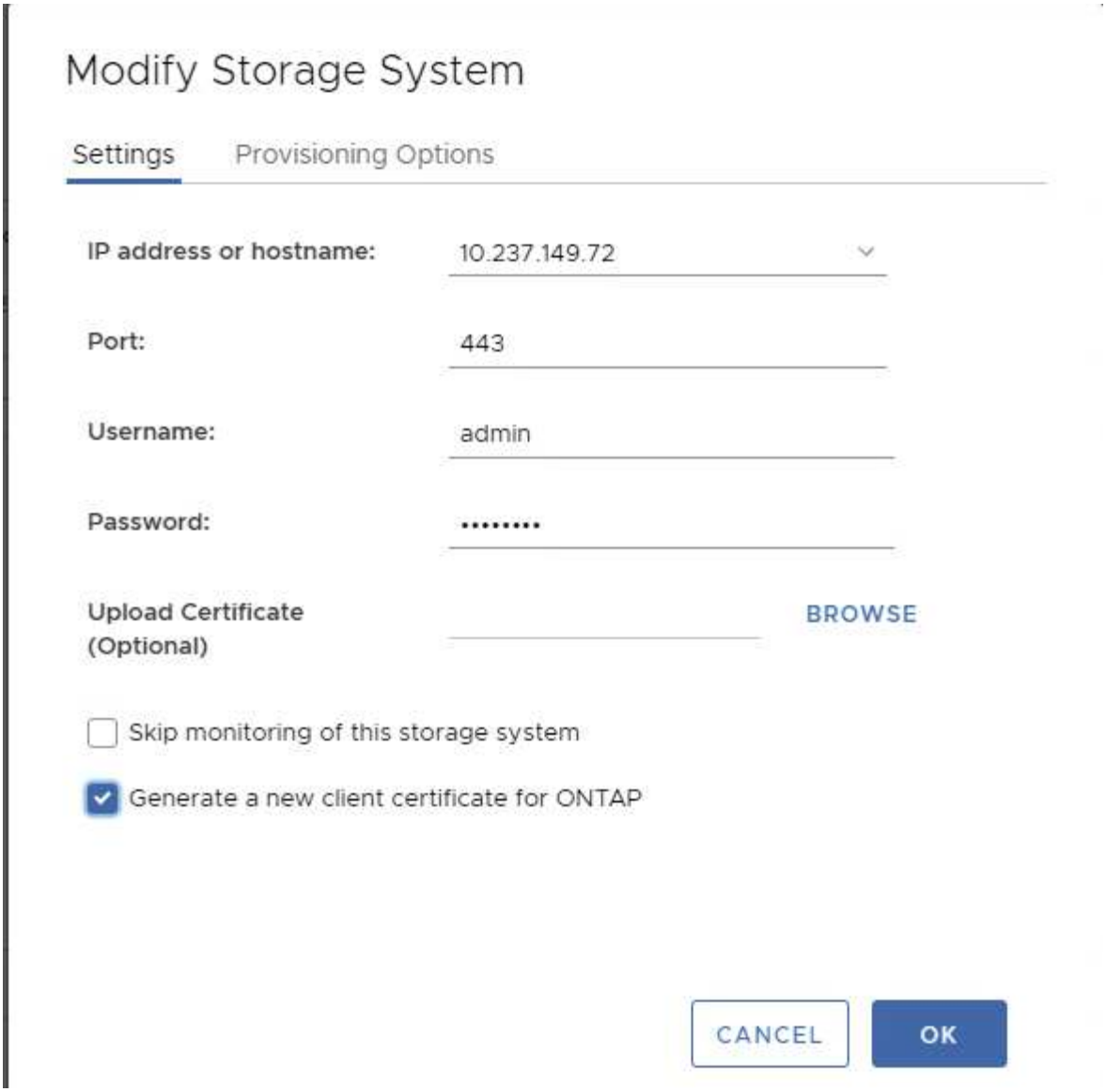

Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
Authorize Cluster Certificate
Host 10.234.85.142 has identified itself with a self-signed certificate.
Show certificate
Do you want to trust this certificate?
NO
YES
CANCEL
ADD

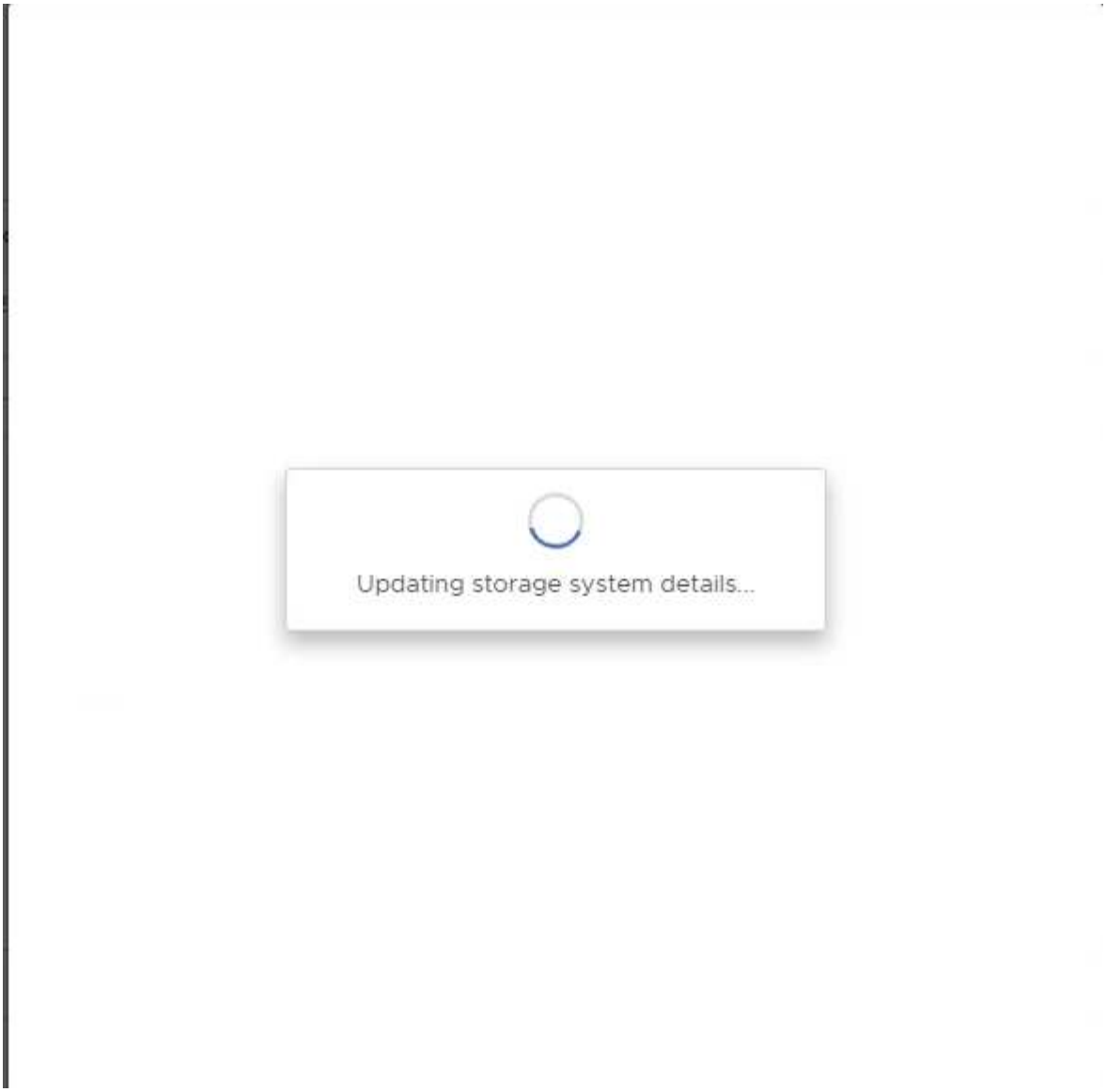


クラスタの編集

クラスタの編集処理には、次の2つのシナリオがあります。

- ONTAP証明書の有効期限が切れた場合、ユーザは新しい証明書を取得してアップロードする必要があります。
- OTV証明書の有効期限が切れた場合は、チェックボックスをオンにして証明書を再生成できます。
  - ONTAPの新しいクライアント証明書を生成します。_


Modify Storage System
Settings
Provisioning Options
IP address or hostname:
10.237.149.72
Port:
443
Username:
admin
Password:
Upload Certificate (Optional)
BROWSE
Skip monitoring of this storage system
Generate a new client certificate for ONTAP
CANCEL
OK



ONTAP toolsのHTTPS証明書

デフォルトでは、ONTAP toolsは、Web UIへのHTTPSアクセスを保護するために、インストール時に自動的に作成される自己署名証明書を使用します。ONTAP toolsには次の機能があります。

1. HTTPS証明書の再生成

ONTAP toolsのインストール時に、HTTPS CA証明書がインストールされ、証明書がキーストアに格納されます。ユーザは、maintコンソールを使用してHTTPS証明書を再生成することができます。

上記のオプションは、'アプリケーション設定'→'証明書の再生成'に移動することで*_maint_console*でアクセスできます。

ログインバナー

ユーザがログインプロンプトにユーザ名を入力すると、次のログインバナーが表示され

ます。SSHはデフォルトで無効になっており、VMコンソールから有効にすると1回限りのログインしか許可されないことに注意してください。

```
WARNING: Unauthorized access to this system is forbidden and will be
prosecuted by law. By accessing this system, you agree that your actions
may be monitored if unauthorized usage is suspected.
```

ユーザが SSH チャネルを介したログインを完了すると、次のテキストが表示されます。

```
Linux UnifiedVSC 5.10.0-21-amd64 #1 SMP Debian 5.10.162-1 (2023-01-21)
x86_64
```

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
```

```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

非アクティブ時のタイムアウト

不正アクセスを防止するために、非アクティブタイムアウトが設定されます。このタイムアウトは、許可されたリソースを使用している間、一定期間非アクティブなユーザを自動的にログアウトします。これにより、許可されたユーザーのみがリソースにアクセスできるようになり、セキュリティの維持に役立ちます。

- デフォルトでは、vSphere Clientセッションはアイドル状態が120分続くと閉じます。そのため、ユーザは再度ログインしてクライアントの使用を再開する必要があります。タイムアウト値を変更するには、webclient.propertiesファイルを編集します。 vSphere Clientのタイムアウトを設定できます。 "vSphere Clientのタイムアウト値の設定"
- ONTAP toolsのWeb-CLIセッションのログアウト時間は30分です。

ユーザあたりの最大同時要求数（ネットワークセキュリティ保護::DOS攻撃）

デフォルトでは、ユーザあたりの最大同時要求数は48です。ONTAP toolsのrootユーザは、環境の要件に応じてこの値を変更できます。この値は、**DoS**攻撃に対するメカニズムを提供するため、非常に大きな値に設定しないでください。

ユーザーは、最大同時セッション数やサポートされているその他のパラメーターを*_/opt/netapp/vscserver/etc/dosfilterParams.json _*ファイルで変更できます。

フィルタを設定するには、次のパラメータを使用します。

- ***delayMs***：レート制限を超えたすべての要求が考慮されるまでの遅延（ミリ秒単位）。要求を拒否するには-1を指定します。
- ***throttlems***:セマフォの非同期待機時間
- ***maxRequestms***：この要求の実行を許可する期間。
- ***ipWhitelist***：レート制限されないIPアドレスのカンマ区切りリスト。（vCenter、ESXi、SRAのIP）
- ***maxRequestsPerSec***：1秒あたりの接続からの最大要求数。

dosfilterParamsファイルのデフォルト値:

```
{"delayMs":"-1",
"throttleMs":"1800000",
"maxRequestMs":"300000",
"ipWhitelist":"10.224.58.52",
"maxRequestsPerSec":"48"}
```

ネットワークタイムプロトコル（**NTP**）の設定

ネットワーク時間設定の不一致が原因で、セキュリティの問題が発生する場合があります。このような問題を防ぐには、ネットワーク内のすべてのデバイスに正確な時間設定があることを確認することが重要です。

仮想アプライアンス

NTPサーバは、仮想アプライアンスのメンテナンスコンソールから設定できます。 ユーザは、*System Configuration⇒_* Add new NTP Server_optionでNTPサーバの詳細を追加できます。

デフォルトでは、NTPのサービスはntpdです。これはレガシーサービスであり、場合によっては仮想マシンでは適切に機能しません。

*** Debian ***

Debianでは、ユーザは/etc/ntp.confファイルにアクセスしてNTPサーバの詳細を確認できます。

パスワードポリシー

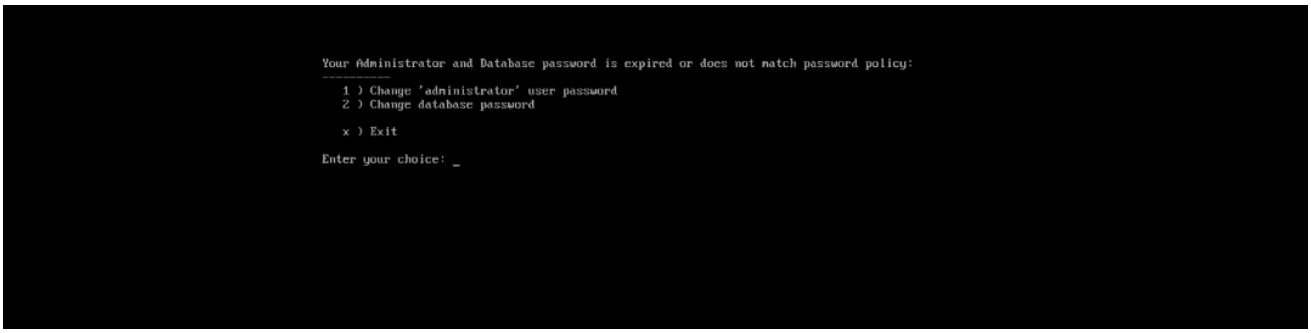
ONTAPツールを初めて導入するユーザ、またはバージョン9.12以降にアップグレードするユーザは、管理者ユーザとデータベースユーザの両方に対して、強力なパスワードポリシーに従う必要があります。導入プロセス中に、新しいユーザにパスワードの入力を求めるプロンプトが表示されます。バージョン9.12以降にアップグレードするBrownfieldユーザの場合は、メンテナンスコンソールで強力なパスワードポリシーに従うオプションを使用できます。

- ユーザがmaintコンソールにログインすると、パスワードが複雑なルールセットに照らしてチェックされ、従わなかった場合、ユーザは同じパスワードをリセットするように求められます。
- パスワードのデフォルトの有効期間は90日です。75日が経過すると、ユーザはパスワードを変更するため

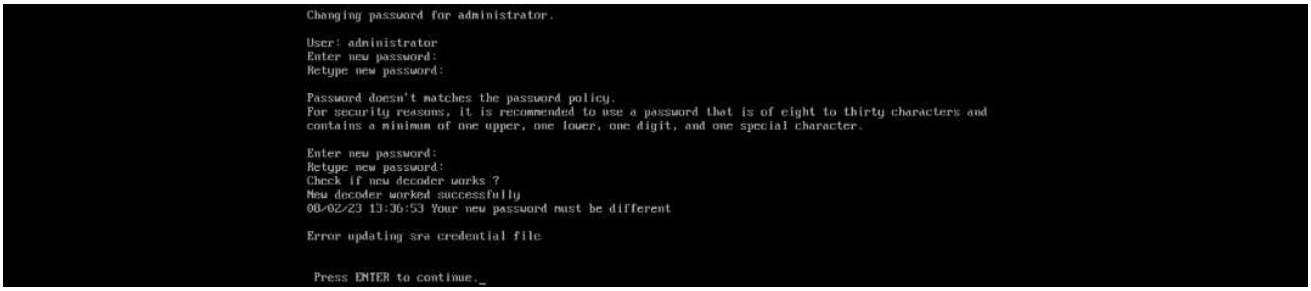
の通知を受け取り始めます。

- サイクルごとに新しいパスワードを設定する必要があります。システムは最後のパスワードを新しいパスワードとして受け取りません。
- ユーザがmaintコンソールにログインするたびに、メインメニューをロードする前に、次のスクリーンショットのようなパスワードポリシーがチェックされます。

- パスワードポリシーまたはONTAP tools 9.11以前からのアップグレードセットアップに従っていないことが検出された場合。パスワードをリセットするための次の画面が表示されます。



- ユーザが弱いパスワードを設定しようとするか、最後のパスワードをもう一度入力すると、次のエラーが表示されます。



法的通知

著作権に関する声明、商標、特許などにアクセスできます。

著作権

"https://www.netapp.com/company/legal/copyright/"

商標

NetApp 、 NetApp のロゴ、および NetApp の商標ページに記載されているマークは、 NetApp, Inc. の商標です。その他の会社名および製品名は、それぞれの所有者の商標である場合があります。

"https://www.netapp.com/company/legal/trademarks/"

特許

ネットアップが所有する特許の最新リストは、次のサイトで入手できます。

https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf

プライバシーポリシー

"https://www.netapp.com/company/legal/privacy-policy/"

オープンソース

通知ファイルには、ネットアップソフトウェアで使用されるサードパーティの著作権およびライセンスに関する情報が記載されています。

ONTAP

"ONTAP 9.13.1に関する注意事項"
"ONTAP 9.12.1に関する注意事項"
"ONTAP 9.12.0の注意事項"
"ONTAP 9.11.1の通知です"
"ONTAP 9.10.1 での通知"
"ONTAP 9.10.0に関する注意事項"
"ONTAP 9.9.1 に関する注意事項"
"ONTAP 9.8 に関する注意事項"
"ONTAP 9.7 の場合の注意事項"
"ONTAP 9.6に関する注意事項"
"ONTAP 9.5 では次の点に注意"
"ONTAP 9.4 の注意事項"
"ONTAP 9.3 での注意"
"ONTAP 9.2に関する注意事項"
"ONTAP 9.1に関する注意事項"

MCC IP向けONTAPメディエーター

"9.9.1 ONTAP Mediator for MCC IPに関する通知"
"9.8 ONTAP Mediator for MCC IPに関する通知"
"9.7 ONTAP Mediator for MCC IPに関する通知"