高可用性アーキテクチャ
ONTAP Select

NetApp
May 07, 2026

This PDF was generated from https://docs.netapp.com/ja-jp/ontap-select-9181/concept_ha_config.html on May 07, 2026. Always check docs.netapp.com for the latest.

目次

高可用性アーキテクチャ

ONTAP Select ハイアベイラビリティ構成

高可用性オプションを確認して、環境に最適なHA構成を選択してください。

顧客はアプリケーションのワークロードをエンタープライズクラスストレージアプライアンスからコモディティ ハードウェア上で実行されるソフトウェアベースのソリューションに移行し始めていますが、回復力と耐障害性に関する期待とニーズは変わっていません。復旧目標時点（RPO）をゼロとするHAソリューションは、インフラストラクチャスタック内のいずれかのコンポーネントの障害によるデータ損失から顧客を保護します。

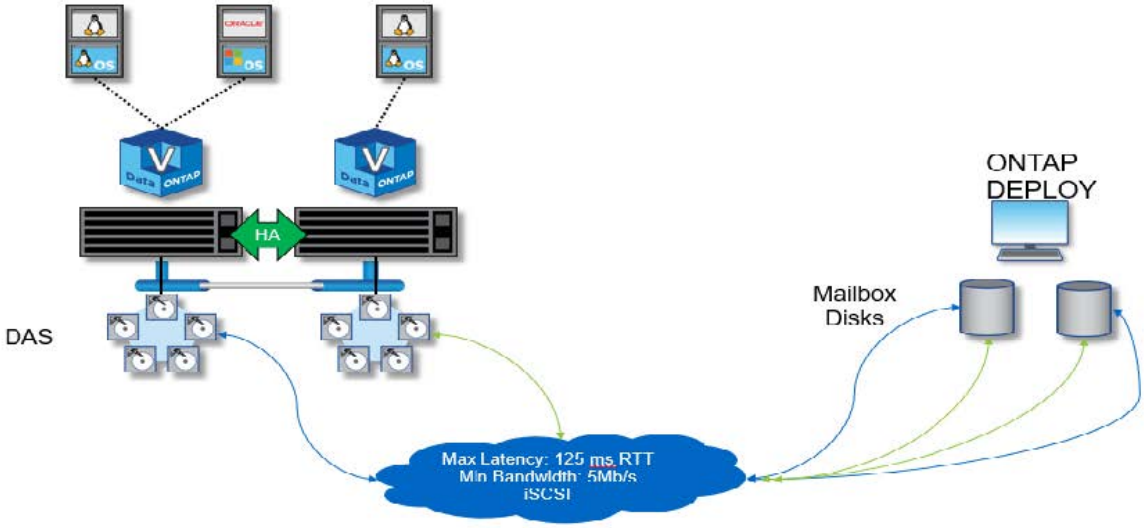
SDS市場の大部分は、シェアードナッシング ストレージの概念に基づいて構築されており、ソフトウェア レプリケーションにより、異なるストレージ サイロにユーザー データの複数のコピーを保存することでデータの復元力が提供されます。ONTAP Selectは、ONTAPが提供する同期レプリケーション機能（RAID SyncMirror）を使用してクラスター内にユーザー データの追加コピーを保存することで、この前提に基づいて構築されています。これはHAペアのコンテキスト内で発生します。すべてのHAペアは、ユーザー データのコピーを2つ保存します。1つはローカル ノードによって提供されるストレージ上に、もう1つはHAパートナーによって提供されるストレージ上に保存されます。ONTAP Selectクラスター内では、HAおよび同期レプリケーションは結び付けられており、2つの機能を切り離したり、独立して使用したりすることはできません。その結果、同期レプリケーション機能はマルチノード オファリングでのみ利用可能です。

> ONTAP Selectクラスタでは、同期レプリケーション機能はHA実装の機能であり、非同期SnapMirrorまたはSnapVaultレプリケーションエンジンの代替ではありません。同期レプリケーションは、HAとは独立して使用することはできません。

ONTAP Select HA導入モデルには、マルチノードクラスタ（4ノード、6ノード、8ノード、10ノード、または12ノード）と2ノードクラスタの2つがあります。2ノードONTAP Selectクラスタの顕著な特徴は、スプリットブレインシナリオを解決するために外部メディエーターサービスを使用することです。ONTAP Deploy VMは、構成するすべての2ノードHAペアのデフォルトのメディエーターとして機能します。

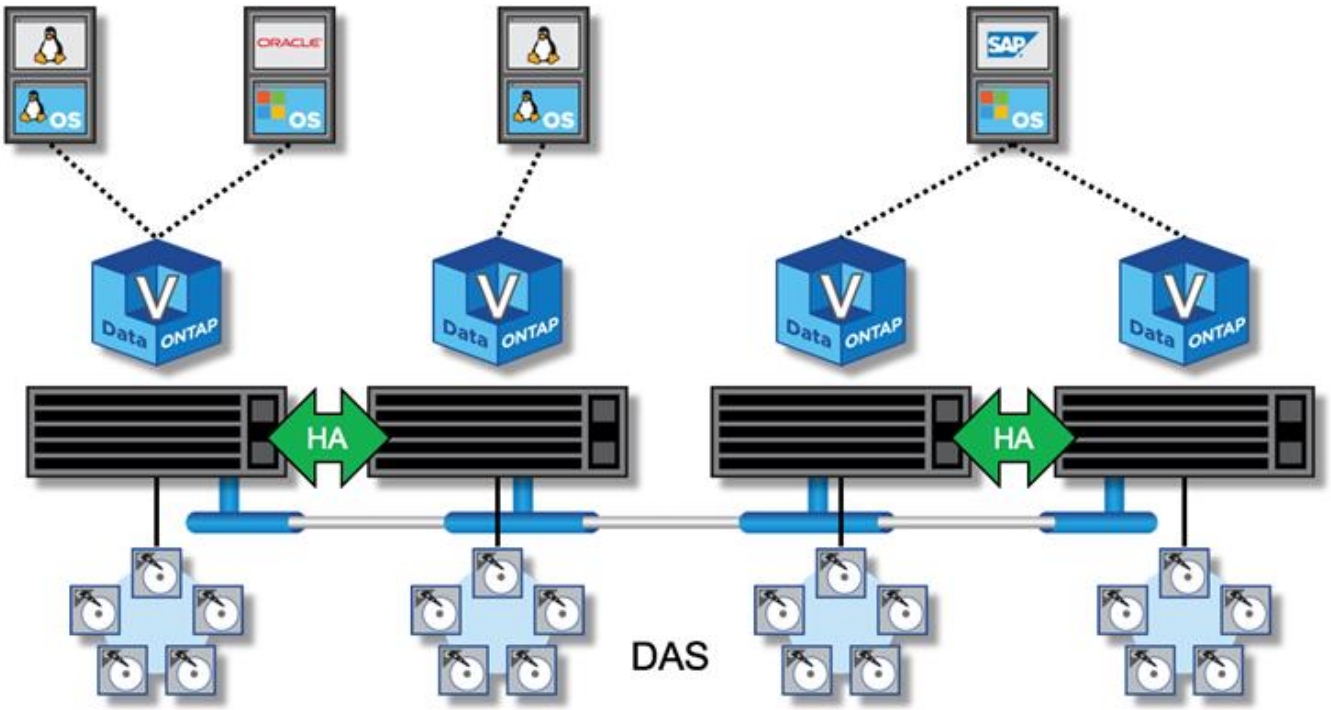
2つのアーキテクチャを次の図に示します。

リモートメディエーターとローカル接続ストレージを使用する**2**ノード**ONTAP Select**クラスタ



2ノードONTAP Selectクラスタは、1つのHAペアと1つのメディエーターで構成されます。HAペア内では、各クラスタノード上のデータアグリゲートが同期的にミラーリングされ、フェイルオーバーが発生してもデータ損失は発生しません。

*ローカル接続ストレージを使用した4ノードONTAP Selectクラスタ
*



- 4ノードONTAP Selectクラスタは2組のHAペアから構成されます。6ノード、8ノード、10ノード、および12ノードのクラスタは、それぞれ3、4、5、および6つのHAペアで構成されます。各HAペア内では、各クラスタノード上のデータアグリゲートが同期的にミラーリングされるため、フェイルオーバーが発生した場合でもデータの損失は発生しません。
- DASストレージを使用する場合、物理サーバ上に存在できるONTAP Selectインスタンスは1つだけで

す。ONTAP Selectは、システムのローカルRAIDコントローラへの非共有アクセスを必要とし、ストレージへの物理的な接続がなければ不可能な、ローカルに接続されたディスクを管理するように設計されています。

2ノードHAとマルチノードHAの比較

FAS アレイとは異なり、HA ペア内の ONTAP Select ノードは IP ネットワーク経由でのみ通信します。つまり、IP ネットワークは単一障害点（SPOF）であり、ネットワーク パーティションやスプリット ブレイン シナリオからの保護が設計の重要な側面になります。マルチノード クラスタは、3 つ以上の存続ノードによってクラスタ クォーラムを確立できるため、単一ノードの障害に耐えることができます。2ノードクラスタは、ONTAP Deploy VM によってホストされるメディエータ サービスに依存して同じ結果を実現します。

ONTAP SelectノードとONTAP Deploy mediatorサービス間のハートビート ネットワーク トラフィックは最小限で回復力があるため、ONTAP Deploy VMをONTAP Select 2ノードクラスタとは別のデータセンターでホストできます。

> ONTAP Deploy VMは、2ノードクラスタのメディエーターとして機能する場合、そのクラスタの不可欠な部分となります。メディエーターサービスが利用できない場合、2ノードクラスタはデータの提供を継続しますが、ONTAP Selectクラスタのストレージフェイルオーバー機能は無効になります。したがって、ONTAP Deployメディエーターサービスは、HAペア内の各ONTAP Selectノードと常に通信を維持する必要があります。クラスタクォーラムが適切に機能するためには、最低5Mbpsの帯域幅と最大125msの往復時間（RTT）レイテンシが必要です。

メディエーターとして機能するONTAP Deploy VMが一時的または永続的に使用できない場合、セカンダリONTAP Deploy VMを使用して2ノードクラスタのクォーラムをリストアできます。この構成では、新しいONTAP Deploy VMはONTAP Selectノードを管理できませんが、クラスタクォーラムアルゴリズムには正常に参加します。ONTAP SelectノードとONTAP Deploy VM間の通信は、IPv4上のiSCSIプロトコルを使用して行われます。ONTAP Selectノード管理IPアドレスがイニシエータで、ONTAP Deploy VMのIPアドレスがターゲットです。したがって、2ノードクラスタを作成する際に、ノード管理IPアドレスにIPv6アドレスを使用することはできません。ONTAP Deployでホストされるメールボックスディスクは、2ノードクラスタの作成時に自動的に作成され、適切なONTAP Selectノード管理IPアドレスにマスクされます。設定全体がセットアップ時に自動的に実行されるため、それ以上の管理操作は必要ありません。クラスタを作成するONTAP Deployインスタンスが、そのクラスタのデフォルトのメディエーターになります。

元のメディエーターの場所を変更する必要がある場合は、管理操作が必要です。元のONTAP Deploy VMが失われた場合でも、クラスタクォーラムをリカバリすることは可能です。ただし、NetAppでは、2ノードクラスタがインスタンス化されるたびにONTAP Deployデータベースをバックアップすることを推奨しています。

2ノードHA対2ノードストレッチHA（MetroCluster SDS）

2ノードのアクティブ / アクティブHAクラスタをより長い距離に拡張し、各ノードを異なるデータセンターに配置することが可能です。2ノードクラスタと2ノードストレッチクラスタ（MetroCluster SDSとも呼ばれる）の唯一の違いは、ノード間のネットワーク接続距離です。

2ノードクラスタは、両方のノードが300m以内の距離で同じデータセンター内に配置されているクラスタとして定義されます。一般的に、両方のノードは同じネットワーク スイッチまたはスイッチ間リンク（ISL）ネットワーク スイッチのセットへのアップリンクを持っています。

2ノードMetroCluster SDSは、物理的に300m以上離れたノード（異なる部屋、異なる建物、異なるデータセンター）で構成されるクラスタとして定義されます。さらに、各ノードのアップリンク接続は、それぞれ別のネットワークスイッチに接続されます。MetroCluster SDSは専用ハードウェアを必要としません。ただし、環境はレイテンシに関する要件（RTTが最大5ms、ジッターが最大5ms、合計10ms）を満たす必要があります。

MetroCluster SDSはプレミアム機能であり、プレミアムライセンスまたはプレミアムXLライセンスが必要です。プレミアムライセンスは、小規模および中規模の仮想マシンの作成に加え、HDDおよびSSDメディアの作成もサポートしています。Premium XLライセンスは、NVMeドライブの作成もサポートしています。

> MetroCluster SDSは、ローカル接続ストレージ（DAS）と共有ストレージ（vNAS）の両方に対応しています。vNAS構成では、ONTAP Select VMと共有ストレージ間のネットワークにより、通常、固有のレイテンシが高くなります。MetroCluster SDS構成では、共有ストレージのレイテンシを含め、ノード間の遅延を最大10msに抑える必要があります。つまり、Select VM間のレイテンシだけを測定するだけでは不十分です。なぜなら、これらの構成では共有ストレージのレイテンシも無視できないからです。

ONTAP Select HA RSMとミラーアグリゲート

RAID SyncMirror（RSM）、ミラーアグリゲート、および書き込みパスを使用してデータ損失を防止します。

同期レプリケーション

ONTAP HAモデルは、HAパートナーという概念に基づいて構築されています。ONTAP Selectは、ONTAPに搭載されているRAID SyncMirror（RSM）機能を使用してクラスタノード間でデータブロックを複製し、HAペア全体にユーザデータのコピーを2つ分散させることで、このアーキテクチャを非共有汎用サーバ環境に拡張します。

2ノードクラスタメディエータを使用すると、2つのデータセンターにまたがることができます。詳細については、セクション"2ノード伸長HA（MetroCluster SDS）のベストプラクティス"を参照してください。

ミラーアグリゲート

ONTAP Selectクラスタは2~12個のノードで構成されます。各HAペアには、IPネットワークを介してノード間で同期的にミラーリングされたユーザデータのコピーが2つ含まれています。このミラーリングはユーザに対して透過的であり、データ アグリゲートの作成プロセス中に自動的に設定されるデータ アグリゲートのプロパティです。

ONTAP Selectクラスタ内のすべてのアグリゲートは、ノードのフェイルオーバーが発生した場合のデータ可用性を確保し、ハードウェア障害が発生した場合のSPOFを回避するために、ミラーリングする必要があります。ONTAP Selectクラスタ内のアグリゲートは、HAペアの各ノードから提供される仮想ディスクから構築され、以下のディスクを使用します：

- ローカルディスクセット（現在のONTAP Selectノードによって提供）
- ミラーリングされたディスクセット（現在のノードのHAパートナーによって提供される）

> ミラーリングされたアグリゲートを構築するために使用されるローカルディスクとミラーディスクは、同じサイズでなければなりません。これらのアグリゲートは、plex 0およびplex 1と呼ばれます（それぞれローカルミラーペアとリモートミラーペアを示します）。実際のplex番号は、インストール環境によって異なる場合があります。

このアプローチは、標準的なONTAPクラスタの動作方法とは根本的に異なります。これは、ONTAP Selectクラスタ内のすべてのルートディスクとデータディスクに適用されます。アグリゲートには、データのローカルコピーとミラーコピーの両方が含まれます。したがって、N個の仮想ディスクを含むアグリゲートは、N/2個のディスク分の固有のストレージを提供します。これは、データの2番目のコピーがそれぞれ固有のディスク

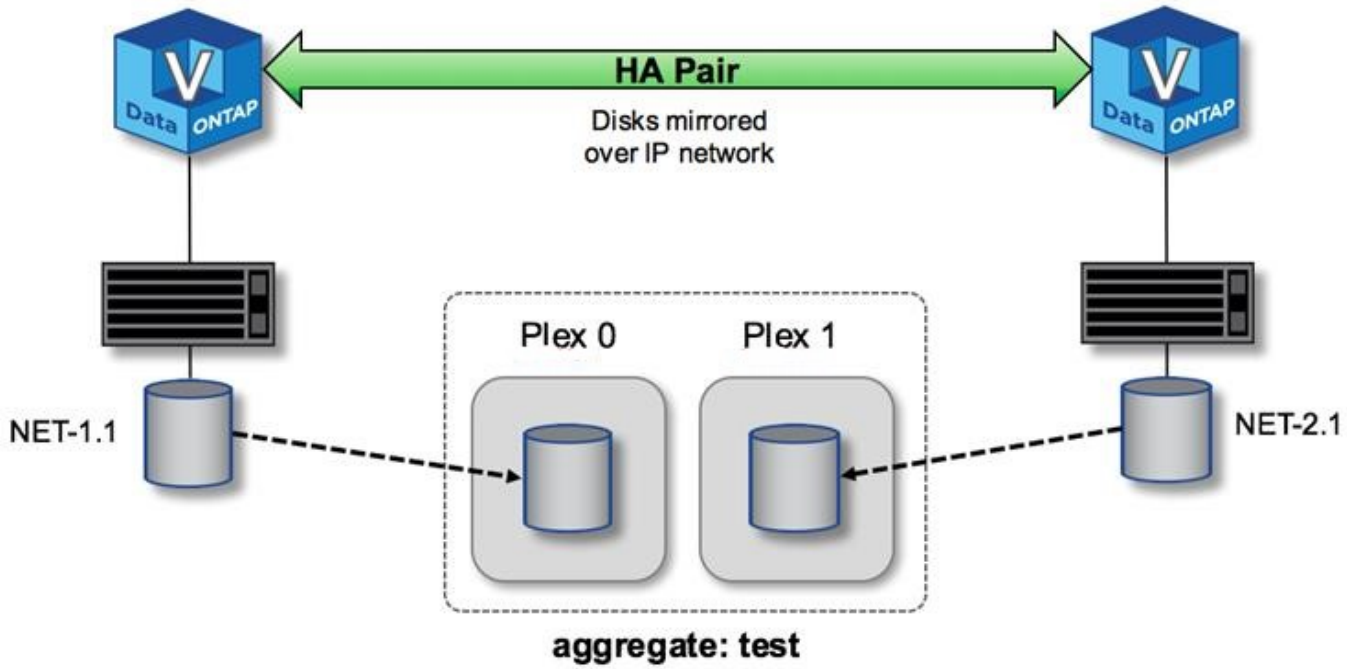
上に存在するためです。

次の図は、4ノードONTAP Selectクラスタ内のHAペアを示しています。このクラスタ内には、両方のHAパートナーのストレージを使用する単一のアグリゲート（test）が存在します。このデータ アグリゲートは、2つの仮想ディスクセットで構成されています：ONTAP Select所有クラスタノード（Plex 0）から提供されるローカルセットと、フェイルオーバーパートナー（Plex 1）から提供されるリモートセットです。

Plex 0は、すべてのローカルディスクを格納するバケットです。Plex 1は、ミラーディスク、つまりユーザーデータの2つ目の複製コピーを保存する役割を担うディスクを格納するバケットです。アグリゲートを所有するノードはPlex 0にディスクを提供し、そのノードのHAパートナーはPlex 1にディスクを提供します。

次の図は、2つのディスクを持つミラーアグリゲートを示しています。このアグリゲートの内容は2つのクラスタノード間でミラーリングされており、ローカルディスクNET-1.1はPlex 0バケットに配置され、リモートディスクNET-2.1はPlex 1バケットに配置されています。この例では、アグリゲートtestは左側のクラスタノードによって所有されており、ローカルディスクNET-1.1とHAパートナーミラーディスクNET-2.1を使用しています。

ONTAP Selectミラーアグリゲート



ONTAP Selectクラスターがデプロイされると、システム上に存在するすべての仮想ディスクは自動的に適切なプレックスに割り当てられるため、ディスク割り当てに関してユーザーによる追加の手順は必要ありません。これにより、ディスクが誤って別のプレックスに割り当てられることを防ぎ、最適なミラーディスク構成を実現します。

書き込みパス

クラスタノード間でのデータブロックの同期ミラーリングと、システム障害時のデータ損失なしという要件は、ONTAP Selectクラスタ内で受信した書き込みが伝播する際の経路に大きな影響を与えます。このプロセスは2つの段階で構成されます：

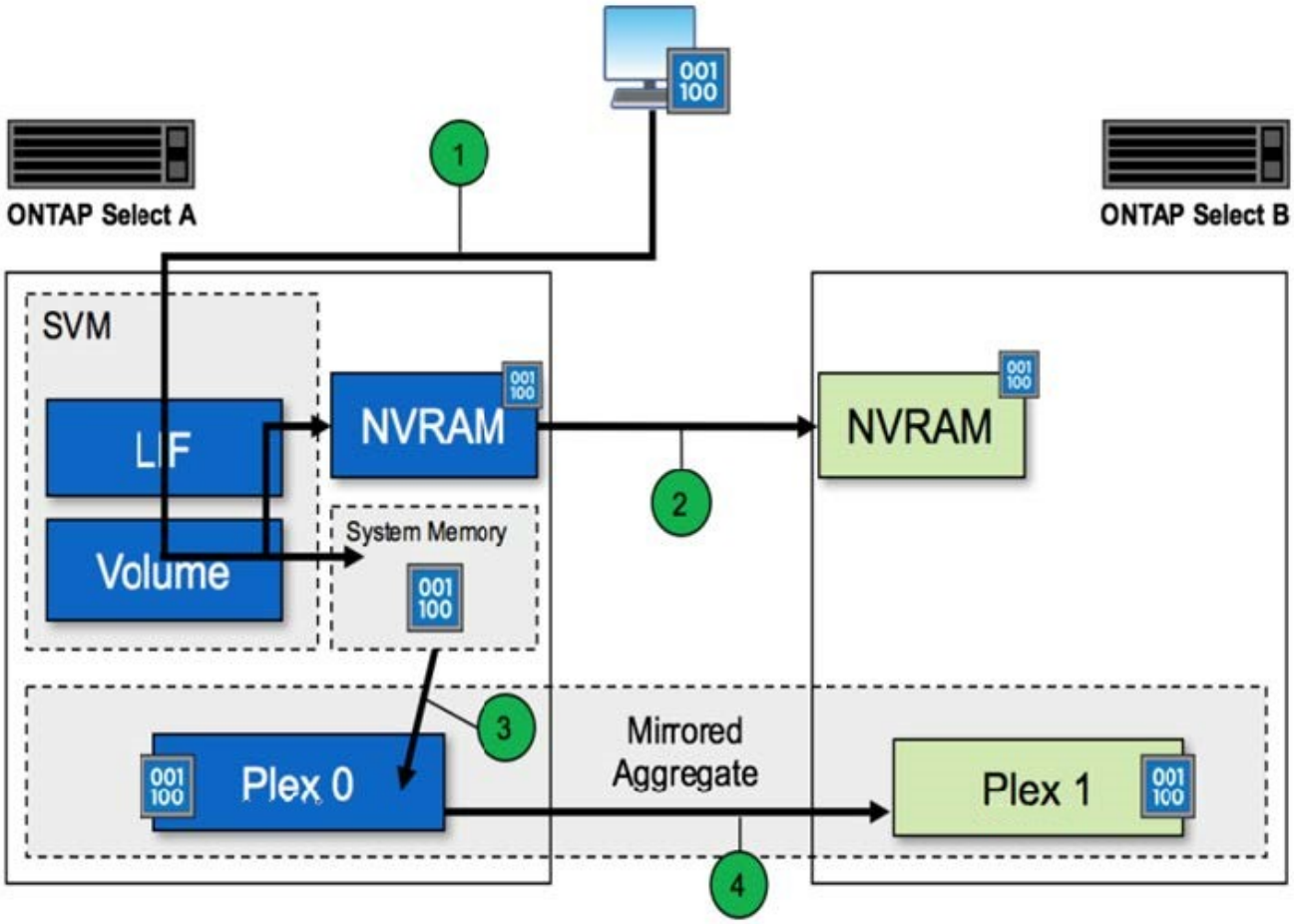
- 謝辞
- デステージング

ターゲット ボリュームへの書き込みはデータ LIF を介して行われ、クライアントに確認応答が返される前に、ONTAP Select ノードのシステム ディスクに存在する仮想化された NVRAM パーティションにコミットされます。HA 構成では、追加の手順が発生します。これらの NVRAM 書き込みは、確認応答される前に、ターゲット ボリュームの所有者の HA パートナーに即座にミラーリングされます。このプロセスにより、元のノードでハードウェア障害が発生した場合でも、HA パートナー ノード上のファイル システムの整合性が保証されます。

書き込みがNVRAMにコミットされた後、ONTAPは定期的にこのパーティションの内容を適切な仮想ディスクに移動します。このプロセスはデステージングと呼ばれます。この処理は、対象ボリュームを所有するクラスタノード上で一度だけ実行され、HAパートナー上では実行されません。

次の図は、ONTAP Selectノードへの受信書き込み要求の書き込みパスを示しています。

*ONTAP Select書き込みパスワークフロー
*



受信書き込み確認応答には、以下の手順が含まれます：

- 書き込みは、ONTAP SelectノードAが所有する論理インターフェースを介してシステムに入ります。
- 書き込みはノードAのNVRAMにコミットされ、HAパートナーであるノードBにミラーリングされます。
- I/O要求が両方のHAノードに存在すると、要求はクライアントに確認応答されます。

ONTAP SelectのNVRAMからデータ アグリゲートへのデステージング（ONTAP CP）には、次の手順が含まれます。

- 書き込みは仮想NVRAMから仮想データ アグリゲートにデステージされます。
- ミラーエンジンは、ブロックを両方のプレックスに同期的に複製します。

ONTAP Select HAはデータ保護を強化します

高可用性（HA）ディスクハートビート、HAメールボックス、HAハートビート、HAフェイルオーバー、およびギブバックは、データ保護を強化するために機能します。

ディスクハートビート

ONTAP Select HAアーキテクチャは、従来のFASアレイで使用されている多くのコードパスを活用していますが、いくつかの例外があります。これらの例外の1つは、ディスクベースのハートビートの実装です。これは、ネットワークの分離によってスプリットブレイン動作が発生するのを防ぐためにクラスタノードが使用する、ネットワークベースではない通信方法です。スプリットブレインシナリオは、通常ネットワーク障害によって引き起こされるクラスタのパーティショニングの結果であり、各側が相手側がダウンしていると判断し、クラスタリソースの引き継ぎを試みます。

エンタープライズクラスのHA実装は、このようなシナリオに適切に対応できなければならない。ONTAPは、カスタマイズされたディスクベースのハートビート方式によってこれを実現します。これはHAメールボックスの役割であり、クラスタノードがハートビートメッセージを送信するために使用する物理ストレージ上の場所です。これは、クラスターが接続性を判断し、フェイルオーバー発生時のクォーラムを定義するのに役立ちます。

共有ストレージHAアーキテクチャを使用するFASアレイでは、ONTAPは以下の方法でスプリットブレインの問題を解決します：

- SCSIの永続的予約
- 永続的なHAメタデータ
- HAインターコネクト経由で送信されるHA状態

ただし、ONTAP Selectクラスタのシェアードナッシングアーキテクチャでは、ノードは自身のローカルストレージのみを参照でき、HAパートナーのストレージは参照できません。したがって、ネットワーク分割によってHAペアの両側が分離される場合、クラスタクォーラムとフェイルオーバー動作を決定するための前述の方法は利用できなくなります。

既存の脳分離検出・回避方法は使用できないものの、共有なし環境の制約に適合する仲介方法が依然として必要である。ONTAP Selectは既存のメールボックスインフラストラクチャをさらに拡張し、ネットワーク分断が発生した場合の仲介手段として機能できるようにする。共有ストレージが利用できないため、仲介はNAS経由でメールボックスディスクにアクセスすることによって行われます。これらのディスクは、2ノードクラスタのメディエーターを含め、iSCSIプロトコルを使用してクラスター全体に分散されています。したがって、クラスタノードはこれらのディスクへのアクセスに基づいて、インテリジェントなフェイルオーバーの判断を下すことができる。ノードがHAパートナー以外のノードのメールボックスディスクにアクセスできる場合、そのノードは正常に稼働している可能性が高い。

> メールボックスアーキテクチャと、クラスタクォーラムとスプリットブレインの問題を解決するディスクベースのハートビート方式は、マルチノードバリアントのONTAP Selectで4つの独立したノードまたは2ノードクラスタ用のメディエーターのいずれかが必要な理由です。

HAメールボックスポスティング

HAメールボックスアーキテクチャは、メッセージポストモデルを使用します。クラスタノードは、一定間隔で、ノードが稼働中であることを示すメッセージを、メディエーターを含むクラスタ内のすべての他のメールボックスディスクに投稿します。正常なクラスタ内では、どの時点においても、クラスタノード上の単一のメールボックスディスクに、他のすべてのクラスタノードから送信されたメッセージが保存されています。

各Selectクラスタノードには、共有メールボックスへのアクセス専用の仮想ディスクが接続されています。このディスクは、ノード障害やネットワーク分断が発生した場合にクラスタの仲介手段として機能することを主な機能としているため、メディエーターメールボックスディスクと呼ばれています。このメールボックスディスクには、各クラスタノード用のパーティションが含まれており、他のSelectクラスタノードによってiSCSIネットワーク経由でマウントされます。これらのノードは定期的に、メールボックスディスクの適切なパーティションにヘルスステータスを送信します。クラスタ全体に分散配置されたネットワークアクセス可能なメールボックスディスクを使用することで、到達可能性マトリックスを通じてノードの状態を推測できます。例えば、クラスタノードAとBはクラスタノードDのメールボックスには投稿できますが、ノードCのメールボックスには投稿できません。さらに、クラスタノードDはノードCのメールボックスに投稿できないため、ノードCはダウンしているかネットワークから隔離されている可能性が高く、引き継ぐ必要があります。

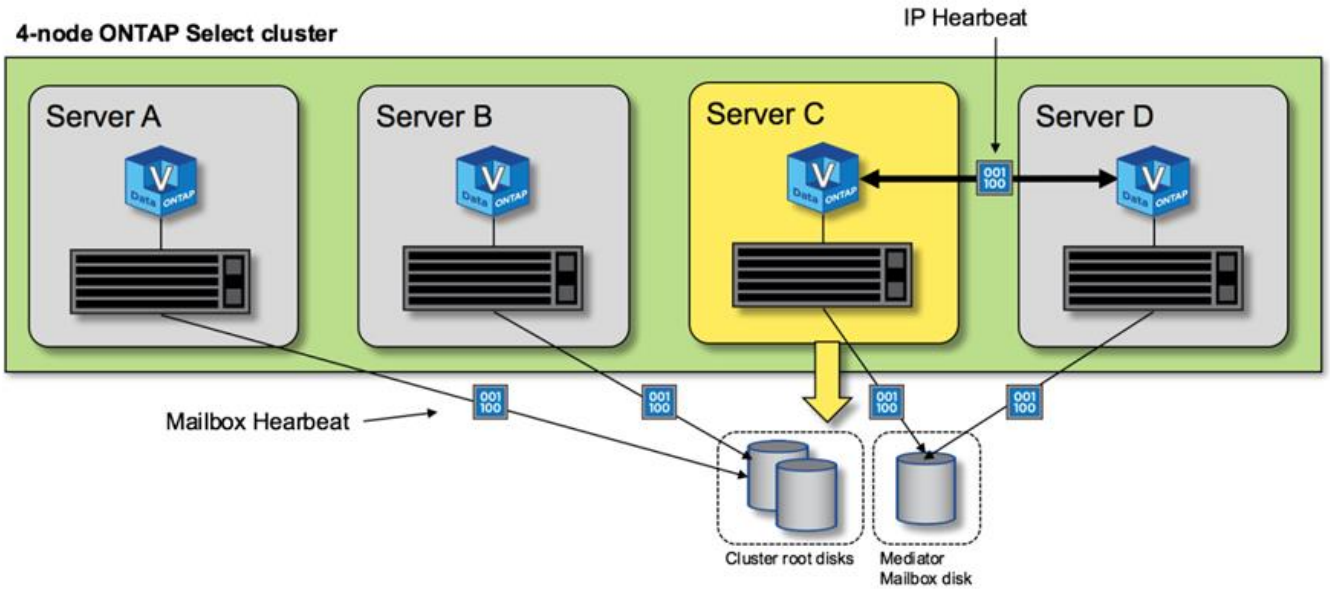
HAハートビート

NetApp FASプラットフォームと同様に、ONTAP SelectはHAインターコネクトを介してHAハートビートメッセージを定期的に送信します。ONTAP Selectクラスタ内では、これはHAパートナー間に存在するTCP/IPネットワーク接続を介して実行されます。さらに、ディスクベースのハートビートメッセージは、メディエーターメールボックスディスクを含む、すべてのHAメールボックスディスクに送信されます。これらのメッセージは数秒ごとに送信され、定期的に読み戻されます。これらの送受信頻度により、ONTAP SelectクラスタはFASプラットフォームで使用可能なものと同じ時間枠である約15秒以内にHA障害イベントを検出できます。ハートビートメッセージが読み取れなくなると、フェイルオーバーイベントがトリガーされます。

次の図は、単一のONTAP Selectクラスタノード（ノードC）の視点から、HAインターコネクトおよびメディエーターディスクを介してハートビートメッセージを送受信するプロセスを示しています。

ネットワークハートビートは、HAインターコネクトを介してHAパートナーであるノードDに送信され、ディスクハートビートはクラスタノードA、B、C、およびDのすべてのメールボックスディスクを使用します。

4ノードクラスタにおける**HA**ハートビート：定常状態



HAのフェイルオーバーとギブバック

フェイルオーバー操作中、生存ノードは、HAパートナーのデータのローカルコピーを使用して、ピアノードのデータ提供責任を引き継ぎます。クライアントI/Oは中断なく継続できますが、データへの変更は、ギブバックが行われる前にレプリケートされる必要があります。ONTAP Selectは強制的なギブバックをサポートしていません。強制的なギブバックを行うと、生存ノードに保存されている変更が失われるためです。

同期バック処理は、リブートしたノードがクラスタに再参加すると自動的にトリガーされます。同期バックに必要な時間は、いくつかの要因によって異なります。これらの要因には、レプリケートする必要がある変更の数、ノード間のネットワークレイテンシ、および各ノードのディスクサブシステムの速度が含まれます。同期バックに必要な時間が、自動ギブバックウィンドウの10分を超える可能性があります。この場合、同期バック後に手動でのギブバックが必要です。同期バックの進捗状況は、次のコマンドを使用して監視できます：

```
storage aggregate status -r -aggregate <aggregate name>
```

著作権に関する情報

Copyright © 2026 NetApp, Inc. All Rights Reserved. Printed in the U.S.このドキュメントは著作権によって保護されています。著作権所有者の書面による事前承諾がある場合を除き、画像媒体、電子媒体、および写真複写、記録媒体、テープ媒体、電子検索システムへの組み込みを含む機械媒体など、いかなる形式および方法による複製も禁止します。

ネットアップの著作物から派生したソフトウェアは、次に示す使用許諾条項および免責条項の対象となります。

このソフトウェアは、ネットアップによって「現状のまま」提供されています。ネットアップは明示的な保証、または商品性および特定目的に対する適合性の暗示的保証を含み、かつこれに限定されないいかなる暗示的な保証も行いません。ネットアップは、代替品または代替サービスの調達、使用不能、データ損失、利益損失、業務中断を含み、かつこれに限定されない、このソフトウェアの使用により生じたすべての直接的損害、間接的損害、偶発的損害、特別損害、懲罰的損害、必然的損害の発生に対して、損失の発生の可能性が通知されていたとしても、その発生理由、根拠とする責任論、契約の有無、厳格責任、不法行為（過失またはそうでない場合を含む）にかかわらず、一切の責任を負いません。

ネットアップは、ここに記載されているすべての製品に対する変更を随時、予告なく行う権利を保有します。ネットアップによる明示的な書面による合意がある場合を除き、ここに記載されている製品の使用により生じる責任および義務に対して、ネットアップは責任を負いません。この製品の使用または購入は、ネットアップの特許権、商標権、または他の知的所有権に基づくライセンスの供与とはみなされません。

このマニュアルに記載されている製品は、1つ以上の米国特許、その他の国の特許、および出願中の特許によって保護されている場合があります。

権利の制限について：政府による使用、複製、開示は、DFARS 252.227-7013（2014年2月）およびFAR 5252.227-19（2007年12月）のRights in Technical Data -Noncommercial Items（技術データ - 非商用品目に関する諸権利）条項の(b)(3)項、に規定された制限が適用されます。

本書に含まれるデータは商用製品および / または商用サービス（FAR 2.101の定義に基づく）に関係し、データの所有権はNetApp, Inc.にあります。本契約に基づき提供されるすべてのネットアップの技術データおよびコンピュータ ソフトウェアは、商用目的であり、私費のみで開発されたものです。米国政府は本データに対し、非独占的かつ移転およびサブライセンス不可で、全世界を対象とする取り消し不能の制限付き使用権を有し、本データの提供の根拠となった米国政府契約に関連し、当該契約の裏付けとする場合にのみ本データを使用できます。前述の場合を除き、NetApp, Inc.の書面による許可を事前に得ることなく、本データを使用、開示、転載、改変するほか、上演または展示することはできません。国防総省にかかる米国政府のデータ使用権については、DFARS 252.227-7015(b)項（2014年2月）で定められた権利のみが認められます。

商標に関する情報

NetApp、NetAppのロゴ、http://www.netapp.com/TMに記載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。