



NetAppとDremioの次世代ハイブリッド氷山レイクハウスソリューション

NetApp artificial intelligence solutions

NetApp
February 12, 2026

目次

NetAppとDremioの次世代ハイブリッド氷山レイクハウスソリューション	1
NetAppとDremioの次世代ハイブリッドIceberg Lakehouseソリューション	1
ソリューションの概要	1
NetAppの概要	1
Dremioの概要	2
Dremio とNetApp のハイブリッド Iceberg Lakehouse ソリューションは顧客にどのような価値をもたらしますか?	3
テクノロジー要件	3
展開手順	4
ソリューション検証の概要	15
顧客ユースケース	22
NetApp ActiveIQ のユースケース	22
自動車部品販売の顧客ユースケース。	23
まとめ	24
詳細情報の入手方法	24

NetAppとDremioの次世代ハイブリッド氷山レイクハウスのソリューション

NetAppとDremioの次世代ハイブリッドIceberg Lakehouseソリューション

このドキュメントでは、ONTAP S3、NAS、StorageGRIDなどのNetAppストレージコントローラからのさまざまなデータソースを使用したDremioの展開の詳細について説明します。展開中に、TPC-DSベンチマークツールを使用して、さまざまなソースに対して99個のSQLクエリを実行しました。このドキュメントでは、NetApp内の顧客ユースケースや、自動車部品販売の顧客に関連するユースケースについても説明します。

ソリューションの概要

Hybrid Iceberg Lakehouseソリューションは、データレイクの顧客が直面する顧客の課題に対処するための独自の利点を提供します。Dremio Unified LakehouseプラットフォームとNetApp ONTAP、StorageGRID、NetApp Cloudソリューションを活用することで、企業はビジネス運営に大きな価値を付加できます。このソリューションは、NetAppソースを含む複数のデータソースへのアクセスを提供するだけでなく、全体的な分析パフォーマンスを向上させ、企業がビジネスの成長につながるビジネスインサイトを獲得できるよう支援します。

NetAppの概要

- ONTAPやStorageGRIDなどのNetAppの製品は、ストレージとコンピューティングの分離を可能にし、特定の要件に基づいて最適なリソース利用を実現します。この柔軟性により、お客様はNetAppストレージソリューションを使用してストレージを独自に拡張できるようになります。
- NetAppのストレージコントローラを活用することで、顧客はNFSおよびS3プロトコルを使用してベクターデータベースにデータを効率的に提供できます。これらのプロトコルは、顧客データの保存を容易にし、ベクターデータベースインデックスを管理するため、ファイルやオブジェクトの方法を通じてアクセスされるデータの複数のコピーが不要になります。
- NetApp ONTAPは、AWS、Azure、Google Cloudなどの主要なクラウドサービスプロバイダー全体でNASおよびオブジェクトストレージのネイティブサポートを提供します。この幅広い互換性により、シームレスな統合が保証され、顧客データのモビリティ、グローバルなアクセス性、災害復旧、動的なスケラビリティ、高パフォーマンスが実現します。

StorageGRID

業界をリードする当社のオブジェクトストレージstorageGRIDは、自動データ配置、柔軟な導入オプション、階層化された消去コーディングによる比類のない耐久性を実現する強力なポリシーエンジンを提供します。単一の名前空間で数十億のオブジェクトとペタバイトのデータをサポートするスケーラブルなアーキテクチャを備えています。このソリューションはハイブリッドクラウド統合を可能にし、主要なクラウドプラットフォームへのデータ階層化を可能にします。2019年のIDC Marketscape Worldwide Object-Based Vendor Assessmentでリーダーとして認められました。

さらに、storageGRID は、ソフトウェア定義のオブジェクトストレージ、地理的冗長性、およびマルチサイト機能により、大規模な非構造化データの管理に優れています。ポリシーベースの情報ライフサイクル管理が組み込まれており、ミラーリングや検索などのクラウド統合機能を提供します。Common Criteria、NF203 Digital Safe Component、ISO/IEC 25051、KPMG、Cohasset Compliance Assessment など、さまざまな認証を取得しています。

要約すると、NetApp storageGRID は、大規模な非構造化データを効率的に管理するための強力な機能、スケーラビリティ、ハイブリッドクラウド統合、コンプライアンス認定を提供します。

NetApp ONTAP

NetApp ONTAPは、幅広いエンタープライズ機能を提供する堅牢なストレージソリューションです。これには、アプリケーションの一貫性と改ざん防止を備えた即時バックアップを提供するスナップショットが含まれています。SnapRestore はオンデマンドでのバックアップのほぼ瞬時の復元を可能にし、SnapMirror は統合されたりリモートバックアップおよび災害復旧機能を提供します。このソリューションには、自律ランサムウェア保護 (ARP) も組み込まれており、複数の管理者による検証、FIPS 認定による保存データの暗号化、転送中のデータの暗号化、多要素認証 (MFA)、ロールベースのアクセス制御 (RBAC) などの機能によってデータのセキュリティを確保します。包括的なログ記録、監査、オンボードおよび外部キー管理、安全な消去、複数のテナントの安全な管理により、データのセキュリティとコンプライアンスがさらに強化されます。

NetApp ONTAP には、低い総所有コストで、高いレベルの整合性、パフォーマンス、保持力を備えた規制に準拠したデータ保持を提供するSnapLockも搭載されています。NetApp ONTAP 9 と完全に統合されており、悪意のある行為、不正な管理者、ランサムウェアに対する保護を提供します。

このソリューションには、転送中および保存中のデータの暗号化のための NSE/NVE 暗号化、多要素管理者アクセス、および複数管理者検証が含まれます。Active IQ はAI を活用した予測分析と修正アクションを提供し、QoS はサービス品質のワークロード制御を保証します。管理と自動化の統合は、SysMgr/GUI/CLI/API を通じて直感的に行えます。FabricPool は自動データ階層化を可能にし、インラインデータ圧縮、重複排除、および圧縮を通じて効率性を実現します。NetApp は、顧客に費用をかけずにワークロード効率の目標を達成することを保証します。

NetApp ONTAP は、NVMe/FC、FC、NVMe/TCP、iSCSI、NFS、SMB、S3 などのさまざまなプロトコルをサポートし、統合ストレージソリューションを実現します。全体として、NetApp ONTAP は、多様なストレージニーズを満たすための広範なエンタープライズ機能、強力なセキュリティ、コンプライアンス、効率性、汎用性を提供します。

Dremioの概要

Dremio は、セルフサービス分析と AI のための統合レイクハウスプラットフォームです。Dremio 統合分析プラットフォームは、従来のデータウェアハウスソリューションの数分の 1 のコストで、レイクハウスの柔軟性、拡張性、パフォーマンスを実現し、ユーザーをデータに近づけます。Dremio は、「シフトレフト」分析を可能にして、複雑でコストのかかるデータ統合と ETL を排除し、データ移動のないシームレスなエンタープライズ規模の分析を実現します。Dremio には以下の機能もあります:

- ユニバーサルなセマンティックレイヤーと緊密に統合された高性能な SQL クエリエンジンによって実現される使いやすいセルフサービス分析により、クラウドとオンプレミスの両方ですべてのデータの接続、管理、分析が容易になります。
- Dremio の Apache Iceberg ネイティブのレイクハウス管理機能は、データ検出を簡素化し、データ最適化を自動化して、Git にヒントを得たデータバージョン管理による高性能分析を実現します。
- Dremio はオープンソースとオープンスタンダードに基づいて構築されており、企業がロックインを回避し、イノベーションの態勢を維持できるようにします。エンタープライズ企業は、あらゆるワークロードで最高の価格性能比を備えた、最も使いやすいレイクハウスプラットフォームとして Dremio を信頼して

います。

Dremio と NetApp のハイブリッド Iceberg Lakehouse ソリューションは顧客にどのような価値をもたらしますか？

- **データ管理とアクセス性の向上:** Dremio は、組織がデータ レイクから直接高速にデータをクエリできるようにするデータ レイクハウス プラットフォームでよく知られています。一方、NetApp は、クラウド データ サービスとデータ ストレージ ソリューションの大手プロバイダーです。この共同オファーにより、お客様は企業のデータを効率的かつ効果的に保存、管理、アクセス、分析するための包括的なソリューションを利用できるようになります。
- **パフォーマンスの最適化:** NetApp のデータ ストレージに関する専門知識と Dremio のデータ処理およびデータ最適化の機能を組み合わせたパートナーシップにより、データ操作のパフォーマンスを向上させ、レイテンシを削減し、ビジネス インサイトを迅速に得るソリューションが提供されます。Dremio は、NetApp 独自の社内 IT 分析インフラストラクチャにもパフォーマンス上のメリットをもたらしました。
- **スケーラビリティ:** Dremio と NetApp はどちらも、拡張性を考慮して設計されたソリューションを提供しています。この共同ソリューションは、顧客に高度にスケーラブルなデータ ストレージ、データ管理、および分析環境を提供します。ハイブリッド Iceberg Lakehouse 環境では、Dremio SQL クエリ エンジンと NetApp StorageGRID を組み合わせることで、比類のないスケーラビリティ、同時実行性、クエリ パフォーマンスを実現し、あらゆるビジネスの分析ニーズに対応できます。
- **データ セキュリティとガバナンス:** 両社とも、データ セキュリティとガバナンスに重点を置いています。これらを組み合わせることで、強力なセキュリティとデータ ガバナンス機能が提供され、データが保護され、データ ガバナンスの要件が満たされることが保証されます。ロールベースのきめ細かなアクセス制御、包括的な監査、エンドツーエンドのデータシステム、統合 ID 管理、広範なコンプライアンスとセキュリティ フレームワークを備えた SSO などの機能により、企業の分析データ環境が安全に管理されます。
- **コスト効率:** Dremio のデータ レイク エンジンと NetApp のストレージ ソリューションと統合することで、顧客はデータ管理とデータ移動に関連するコストを削減できます。組織は、従来のデータ レイク環境から、NetApp と Dremio で構成されたより最新のレイクハウス ソリューションに移行することもできます。このハイブリッド Iceberg Lakehouse ソリューションは、高速クエリ パフォーマンスと市場をリードするクエリ同時実行性を提供し、TCO を削減し、ビジネス洞察を得るまでの時間を短縮します。

テクノロジー要件

このドキュメントで実行された検証には、以下に概説するハードウェアおよびソフトウェア構成が使用されました。これらの構成は、環境を設定するためのガイドラインとして役立ちます。ただし、具体的なコンポーネントは個々の顧客の要件に応じて異なる場合があります。ことに注意してください。

ハードウェア要件

ハードウェア	詳細
NetApp AFFストレージアレイ HA ペア	<ul style="list-style-type: none">• A800• ONTAP 9.14.1• 48 x 3.49TB SSD-NVM• 2 つの S3 バケット: Dremio メタデータと顧客データ。

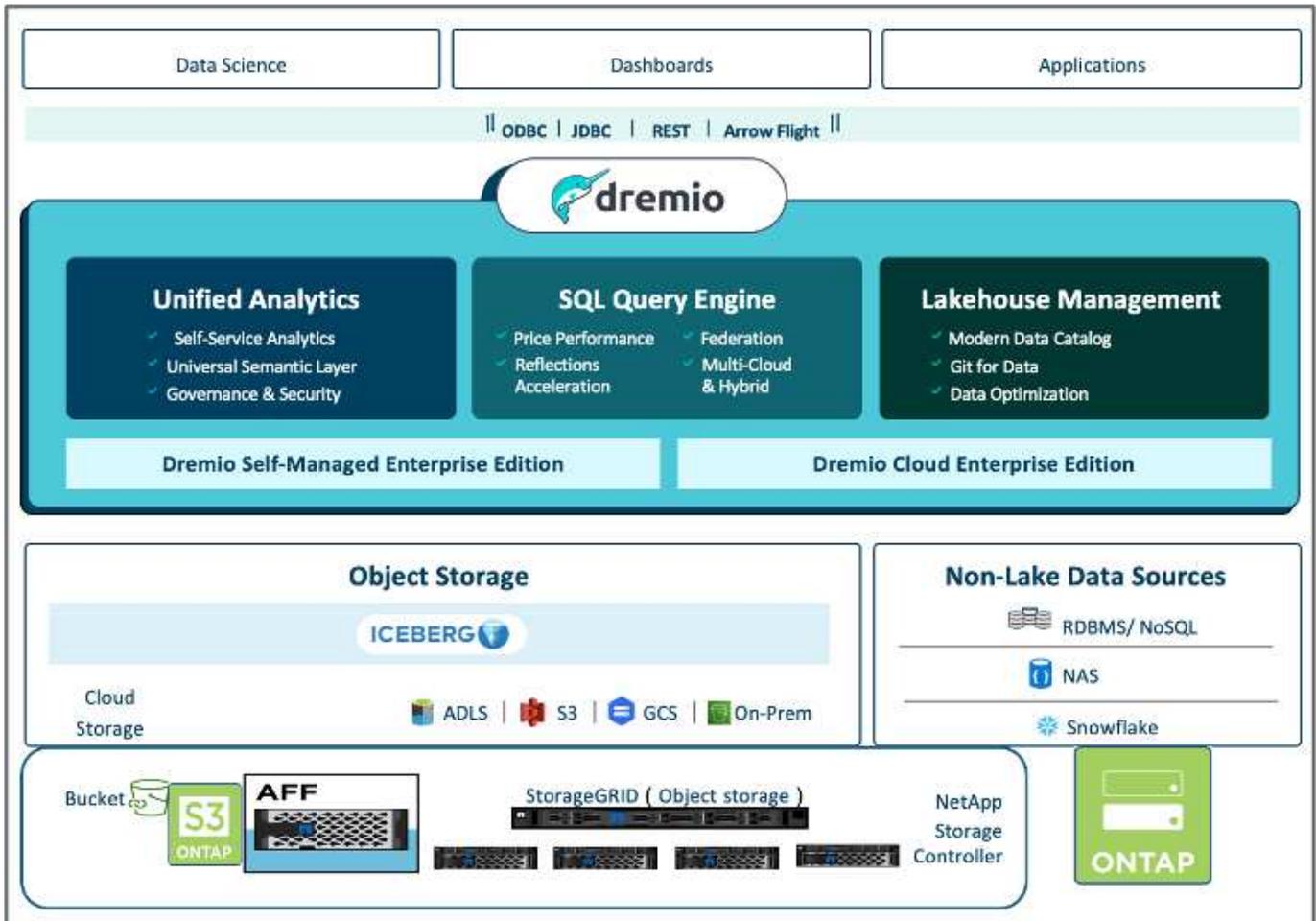
ハードウェア	詳細
富士通 PRIMERGY RX2540 M4 x 4	<ul style="list-style-type: none"> • CPU×64 • インテル Xeon Gold 6142 CPU @ 2.60GHz • 256 GM 物理メモリ • 1 x 100GbE ネットワークポート
ネットワーク	<ul style="list-style-type: none"> • 100ギガビットイーサネット
StorageGRID	* 1 x SG100、3xSGF6024 * 3 x 24 x 7.68TB * 2 つの S3 バケット:Dremio メタデータと顧客データ。

ソフトウェア要件

ソフトウェア	詳細
ドレミオ	<ul style="list-style-type: none"> • バージョン - 25.0.3-202405170357270647-d2042e1b • エンタープライズエディション
オンプレミス	<ul style="list-style-type: none"> • 5ノードのDremioクラスター • マスターコーディネーター1名とエグゼキューター4名

展開手順

このリファレンスアーキテクチャの検証では、1つのコーディネーターと4つのエグゼキューターで構成されるDremio構成を使用しました。



NetAppのセットアップ

- ストレージシステムの初期化
- ストレージ仮想マシン (SVM) の作成
- 論理ネットワークインターフェースの割り当て
- NFS、S3 の設定とライセンス

NFS (ネットワーク ファイル システム) の場合は、次の手順に従ってください。1. NFSv4 または NFSv3 用の Flex Group ボリュームを作成します。この検証のセットアップでは、48 台の SSD を使用しました。1 台の SSD はコントローラのルート ボリューム専用で、47 台の SSD は NFSv4 用に分散されています。Flex Group ボリュームの NFS エクスポート ポリシーに、Dremio サーバー ネットワークに対する読み取り/書き込み権限があることを確認します。

1. すべての Dremio サーバーでフォルダーを作成し、各 Dremio サーバーの論理インターフェイス (LIF) を介してこのフォルダーに Flex Group ボリュームをマウントします。

S3 (Simple Storage Service) の場合は以下の手順に従ってください。

1. 「vserver object-store-server create」コマンドを使用して、HTTP が有効で管理ステータスが「up」に設定されたオブジェクトストアサーバーをセットアップします。HTTPS を有効にしてカスタムリスナーポートを設定するオプションがあります。
2. 「vserver object-store-server user create -user <username>」コマンドを使用して、オブジェクトストア

サーバー ユーザーを作成します。

3. アクセス キーとシークレット キーを取得するには、次のコマンドを実行します: "set diag; vserver object-store-server user show -user <username>". ただし、今後はこれらのキーはユーザー作成プロセス中に提供されるか、REST API 呼び出しを使用して取得できるようになります。
4. 手順 2 で作成したユーザーを使用して object-store-server グループを確立し、アクセスを許可します。この例では、「FullAccess」を指定しました。
5. タイプを「S3」に設定して 2 つの S3 バケットを作成します。1 つは Dremio 構成用、もう 1 つは顧客データ用です。

飼育員のセットアップ

Dremio が提供する Zookeeper 構成を使用できます。この検証では、別の Zookeeper を使用しました。この Web リンクに記載されている手順に従いました。 <https://medium.com/@ahmetfurkandemir/distributed-hadoop-cluster-1-spark-with-all-dependencies-03c8ec616166>

Dremio のセットアップ

この Web リンクに従って、tar ボール経由で Dremio をインストールしました。

1. Dremio グループを作成します。

```
sudo groupadd -r dremio
```

2. dremio ユーザーを作成します。

```
sudo useradd -r -g dremio -d /var/lib/dremio -s /sbin/nologin dremio
```

3. Dremio ディレクトリを作成します。

```
sudo mkdir /opt/dremio
sudo mkdir /var/run/dremio && sudo chown dremio:dremio /var/run/dremio
sudo mkdir /var/log/dremio && sudo chown dremio:dremio /var/log/dremio
sudo mkdir /var/lib/dremio && sudo chown dremio:dremio /var/lib/dremio
```

4. tar ファイルをダウンロードしてください <https://download.dremio.com/community-server/>

5. Dremio を /opt/dremio ディレクトリに解凍します。

```
sudo tar xvf dremio-enterprise-25.0.3-202405170357270647-d2042e1b.tar.gz
-C /opt/dremio --strip-components=1
```

6. 構成フォルダーへのシンボリック リンクを作成します。

```
sudo ln -s /opt/dremio/conf /etc/dremio
```

7. サービス構成を設定します (SystemD セットアップ)。

- a. dremio デーモンのユニット ファイルを /opt/dremio/share/dremio.service から /etc/systemd/system/dremio.service にコピーします。
- b. システムを再起動する

```
sudo systemctl daemon-reload
```

- c. 起動時に dremio が起動するようにします。

```
sudo systemctl enable dremio
```

8. コーディネーター上で Dremio を構成します。詳細については、Dremio 構成を参照してください。

- a. Dremio.conf

```
root@hadoopmaster:/usr/src/tpcds# cat /opt/dremio/conf/dremio.conf

paths: {
  # the local path for dremio to store data.
  local: "${DREMIO_HOME}"/dremiocache"

  # the distributed path Dremio data including job results,
  downloads, uploads, etc
  #dist: "hdfs://hadoopmaster:9000/dremiocache"
  dist: "dremioS3:///dremioconf"
}

services: {
  coordinator.enabled: true,
  coordinator.master.enabled: true,
  executor.enabled: false,
  flight.use_session_service: false
}

zookeeper: "10.63.150.130:2181,10.63.150.153:2181,10.63.150.151:2181"
services.coordinator.master.embedded-zookeeper.enabled: false
root@hadoopmaster:/usr/src/tpcds#
```

- b. コアサイト.xml

```

root@hadoopmaster:/usr/src/tpcds# cat /opt/dremio/conf/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

      http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
  implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>fs.dremioS3.impl</name>
    <value>com.dremio.plugins.s3.store.S3FileSystem</value>
  </property>
  <property>
    <name>fs.s3a.access.key</name>
    <value>24G4C1316APP2BIPDE5S</value>
  </property>
  <property>
    <name>fs.s3a.endpoint</name>
    <value>10.63.150.69:80</value>
  </property>
  <property>
    <name>fs.s3a.secret.key</name>
    <value>Zd28p43rgZaU44PX_ftT279z9nt4jBSro97j87Bx</value>
  </property>
  <property>
    <name>fs.s3a.aws.credentials.provider</name>
    <description>The credential provider type.</description>
    <value>org.apache.hadoop.fs.s3a.SimpleAWSCredentialsProvider</value>
  </property>
  <property>
    <name>fs.s3a.path.style.access</name>
    <value>>false</value>

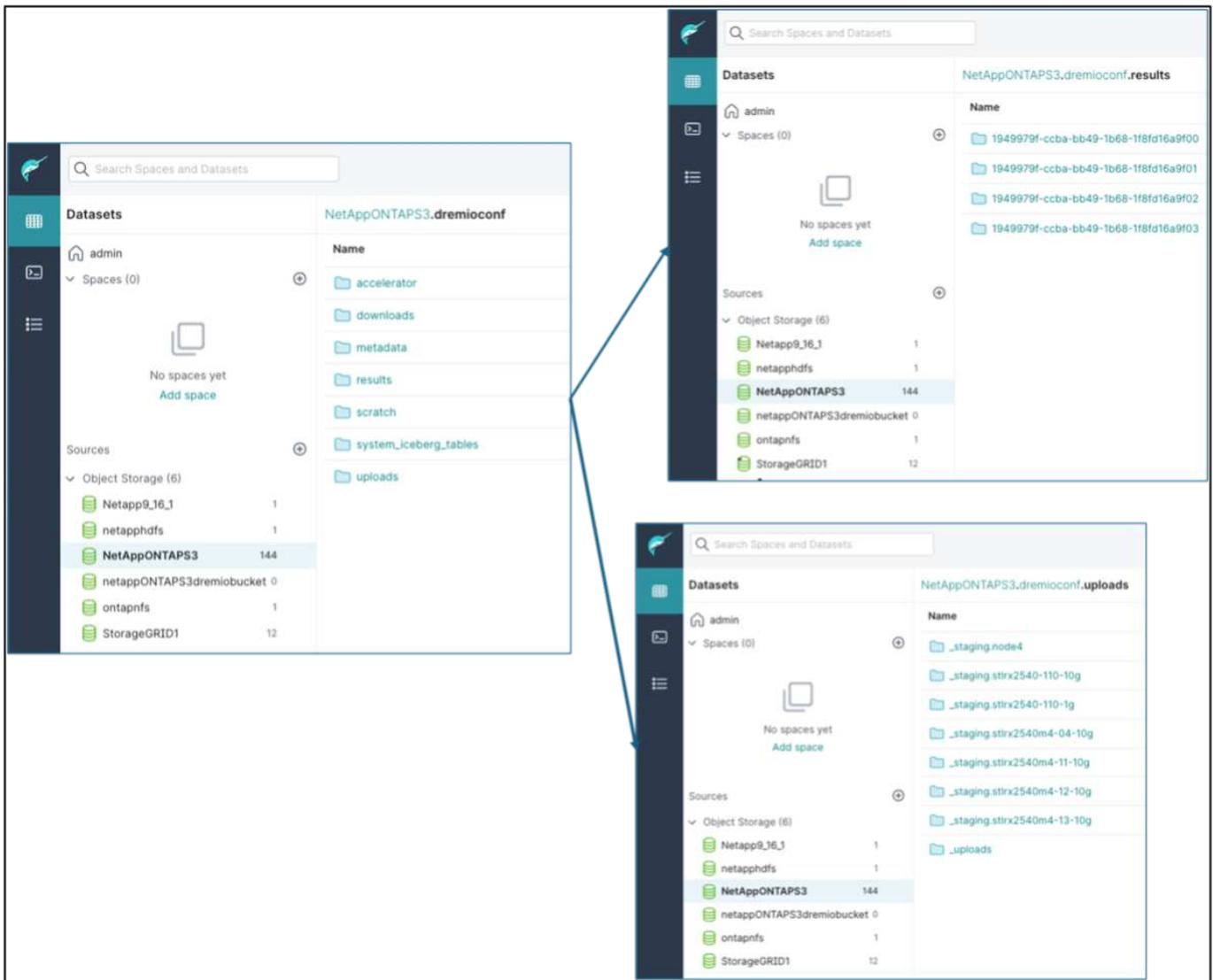
```

```

    </property>
  <property>
    <name>hadoop.proxyuser.dremio.hosts</name>
    <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.dremio.groups</name>
    <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.dremio.users</name>
    <value>*</value>
  </property>
  <property>
    <name>dremio.s3.compat</name>
    <description>Value has to be set to true.</description>
    <value>>true</value>
  </property>
  <property>
    <name>fs.s3a.connection.ssl.enabled</name>
    <description>Value can either be true or false, set to true
to use SSL with a secure Minio server.</description>
    <value>>false</value>
  </property>
</configuration>
root@hadoopmaster:/usr/src/tpcds#

```

9. Dremio 構成はNetAppオブジェクトストレージに保存されます。私たちの検証では、「dremioconf」バケットは ontap S3 バケットに存在します。下の図は、「dremioconf」S3 バケットの「scratch」および「uploads」フォルダの詳細を示しています。



1. 実行者で Dremio を構成します。私たちのセットアップには 3 つのエグゼキュータがあります。
 - a. dremio.conf

```

paths: {
  # the local path for dremio to store data.
  local: "${DREMIO_HOME}"/dremiocache"

  # the distributed path Dremio data including job results,
  downloads, uploads, etc
  #dist: "hdfs://hadoopmaster:9000/dremiocache"
  dist: "dremioS3:///dremioconf"
}

services: {
  coordinator.enabled: false,
  coordinator.master.enabled: false,
  executor.enabled: true,
  flight.use_session_service: true
}

zookeeper: "10.63.150.130:2181,10.63.150.153:2181,10.63.150.151:2181"
services.coordinator.master.embedded-zookeeper.enabled: false

```

b. Core-site.xml – コーディネーターの構成と同じです。



NetApp は、Datalake および Lakehouse 環境の主要なオブジェクト ストレージ ソリューションとしてStorageGRID を推奨しています。さらに、ファイル/オブジェクトの二重性のためにNetApp ONTAPが採用されています。このドキュメントでは、お客様のご要望に応じてONTAP S3 のテストを実施し、データ ソースとして正常に機能していることが確認されています。

複数のソースの設定

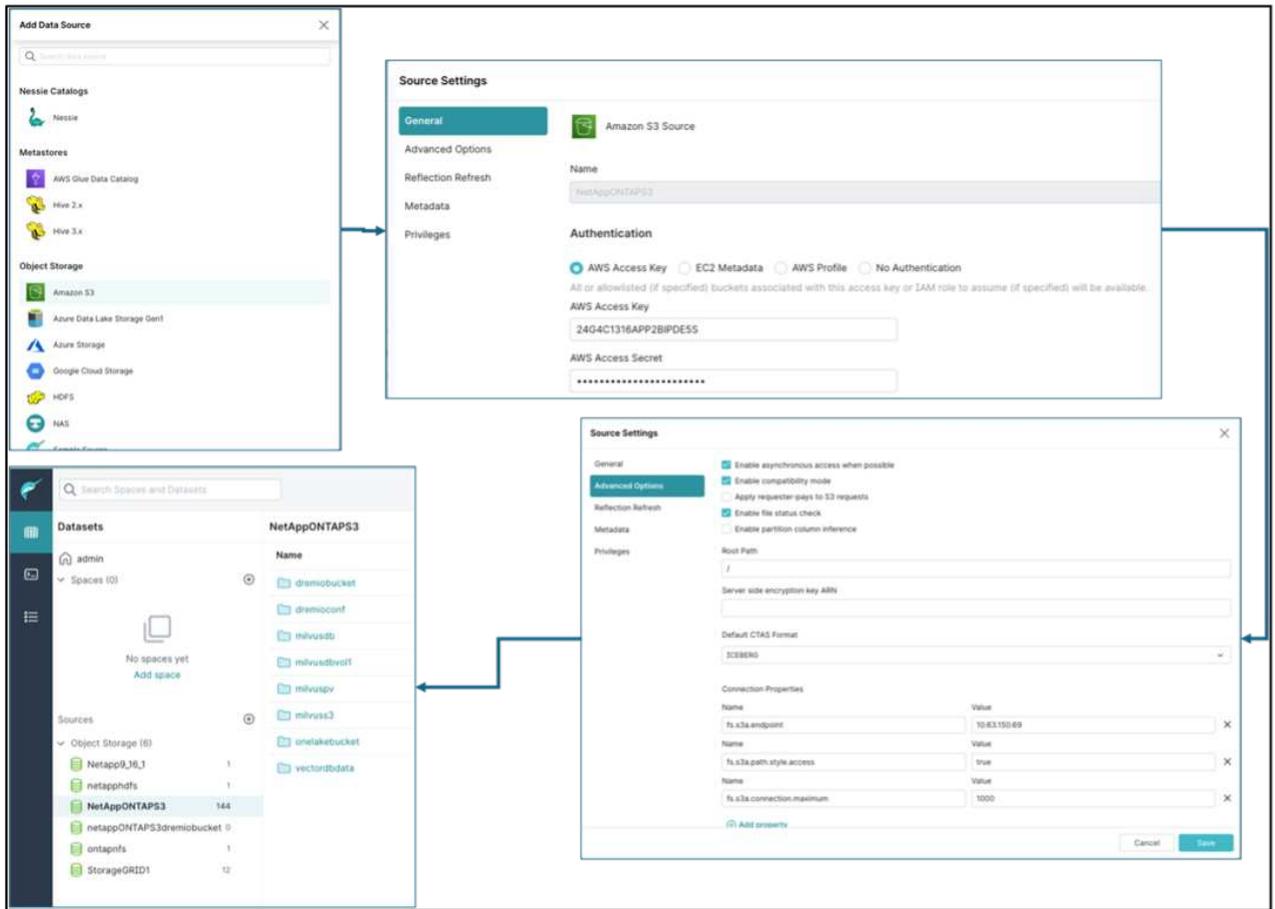
1. Dremio でONTAP S3 と storageGRID を s3 ソースとして設定します。
 - a. Dremio ダッシュボード → データセット → ソース → ソースの追加。
 - b. 一般セクションでAWSアクセスキーとシークレットキーを更新してください
 - c. 詳細オプションで互換モードを有効にし、以下の詳細で接続プロパティを更新します。 ontap S3 または storageGRID のいずれかのNetAppストレージ コントローラからのエンドポイント IP/名前。

```

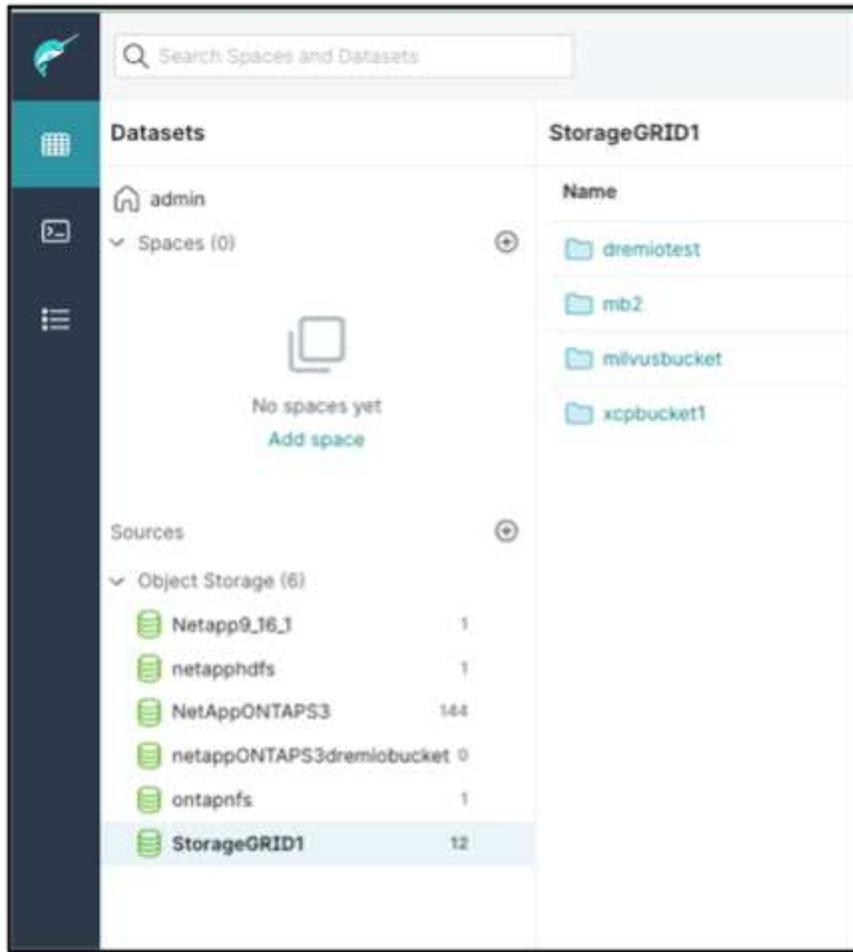
fs.s3a.endpoint = 10.63.150.69
fs.s3a.path.style.access = true
fs.s3a.connection.maximum=1000

```

- d. 可能な場合はローカル キャッシュを有効にする、可能な場合に使用する利用可能なキャッシュの合計の最大パーセント = 100
- e. 次に、 NetAppオブジェクト ストレージからバケットのリストを表示します。



f. storageGRID バケットの詳細のサンプルビュー



2. Dremio で NAS (具体的には NFS) をソースとして設定します。

a. Dremio ダッシュボード → データセット → ソース → ソースの追加。

b. 一般セクションで、名前と NFS マウント パスを入力します。NFS マウント パスが Dremio クラスタ内のすべてのノード上の同じフォルダーにマウントされていることを確認してください。

Add Data Source

Search data source

Nessie Catalogs

- Nessie

Metastores

- AWS Glue Data Catalog
- Hive 2.x
- Hive 3.x

Object Storage

- Amazon S3
- Azure Data Lake Storage Gen1
- Azure Storage
- Google Cloud Storage
- HDFS
- NAS**

New NAS Source

General

Advanced Options

Reflection Refresh

Metadata

Privileges

NAS Source

Name:

Connection

Mount Path ⁱ:

Search Spaces and Datasets

Datasets

admin

Spaces (0)

No spaces yet
[Add space](#)

Sources

Object Storage (6)	
Netapp9_16_1	1
netapphdfs	1
NetAppONTAPS3	144
netappONTAPS3dremiobucket	0
ontapnfs	1
StorageGRID1	12

ontapnfs

Name
csvfile_from_dataset
results

+

```

root@hadoopmaster:~# for i in hadoopmaster hadoopnode1 hadoopnode2
hadoopnode3 hadoopnode4; do ssh $i "date;hostname;du -hs
/opt/dremio/data/spill/ ; df -h //dremionfsdata "; done
Fri Sep 13 04:13:19 PM UTC 2024
hadoopmaster
du: cannot access '/opt/dremio/data/spill/': No such file or directory
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.69:/dremionfsdata 2.1T    921M   2.0T   1% /dremionfsdata
Fri Sep 13 04:13:19 PM UTC 2024
hadoopnode1
12K /opt/dremio/data/spill/
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.69:/dremionfsdata 2.1T    921M   2.0T   1% /dremionfsdata
Fri Sep 13 04:13:19 PM UTC 2024
hadoopnode2
12K /opt/dremio/data/spill/
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.69:/dremionfsdata 2.1T    921M   2.0T   1% /dremionfsdata
Fri Sep 13 16:13:20 UTC 2024
hadoopnode3
16K /opt/dremio/data/spill/
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.69:/dremionfsdata 2.1T    921M   2.0T   1% /dremionfsdata
Fri Sep 13 04:13:21 PM UTC 2024
node4
12K /opt/dremio/data/spill/
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.69:/dremionfsdata 2.1T    921M   2.0T   1% /dremionfsdata
root@hadoopmaster:~#

```

ソリューション検証の概要

このセクションでは、複数のソースから SQL テスト クエリを実行して機能を検証し、NetAppストレージへのスピルオーバーをテストおよび検証しました。

オブジェクトストレージのSQLクエリ

1. dremio.envでサーバーあたりのメモリを250GBに設定する

```
root@hadoopmaster:~# for i in hadoopmaster hadoopnode1 hadoopnode2
hadoopnode3 hadoopnode4; do ssh $i "hostname; grep -i
DREMIO_MAX_MEMORY_SIZE_MB /opt/dremio/conf/dremio-env; cat /proc/meminfo
| grep -i memtotal"; done
hadoopmaster
#DREMIO_MAX_MEMORY_SIZE_MB=120000
DREMIO_MAX_MEMORY_SIZE_MB=250000
MemTotal:          263515760 kB
hadoopnode1
#DREMIO_MAX_MEMORY_SIZE_MB=120000
DREMIO_MAX_MEMORY_SIZE_MB=250000
MemTotal:          263515860 kB
hadoopnode2
#DREMIO_MAX_MEMORY_SIZE_MB=120000
DREMIO_MAX_MEMORY_SIZE_MB=250000
MemTotal:          263515864 kB
hadoopnode3
#DREMIO_MAX_MEMORY_SIZE_MB=120000
DREMIO_MAX_MEMORY_SIZE_MB=250000
MemTotal:          264004556 kB
node4
#DREMIO_MAX_MEMORY_SIZE_MB=120000
DREMIO_MAX_MEMORY_SIZE_MB=250000
MemTotal:          263515484 kB
root@hadoopmaster:~#
```

2. dremio.conf ファイルとストレージの詳細で、スピルオーバーの場所 (`${DREMIO_HOME}/dremiocache`) を確認します。

```

paths: {
  # the local path for dremio to store data.
  local: "${DREMIO_HOME}"/dremiocache"

  # the distributed path Dremio data including job results, downloads,
  uploads, etc
  #dist: "hdfs://hadoopmaster:9000/dremiocache"
  dist: "dremioS3:///dremioconf"
}

services: {
  coordinator.enabled: true,
  coordinator.master.enabled: true,
  executor.enabled: false,
  flight.use_session_service: false
}

zookeeper: "10.63.150.130:2181,10.63.150.153:2181,10.63.150.151:2181"
services.coordinator.master.embedded-zookeeper.enabled: false

```

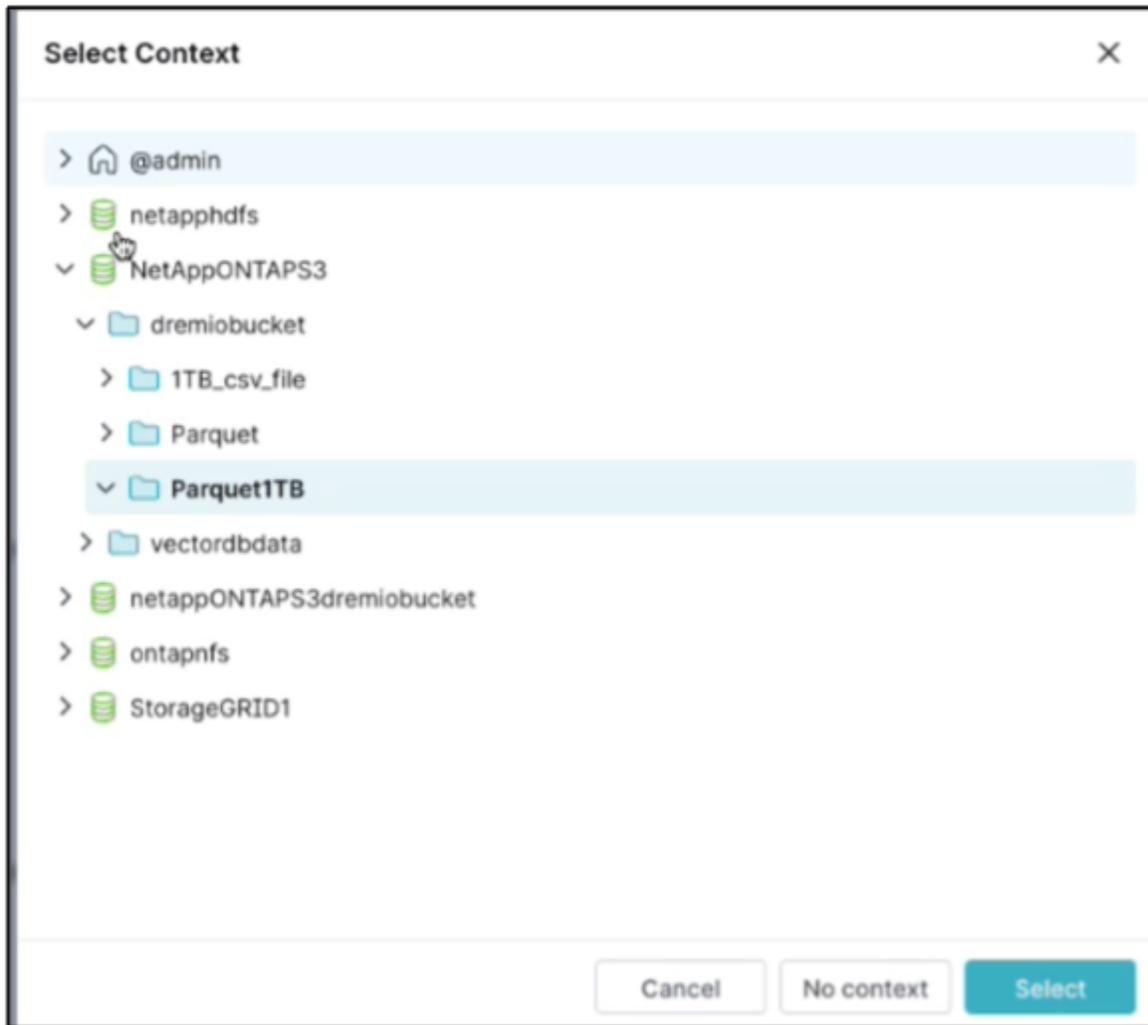
3. Dremioのスピルオーバー場所をNetApp NFSストレージに向ける

```

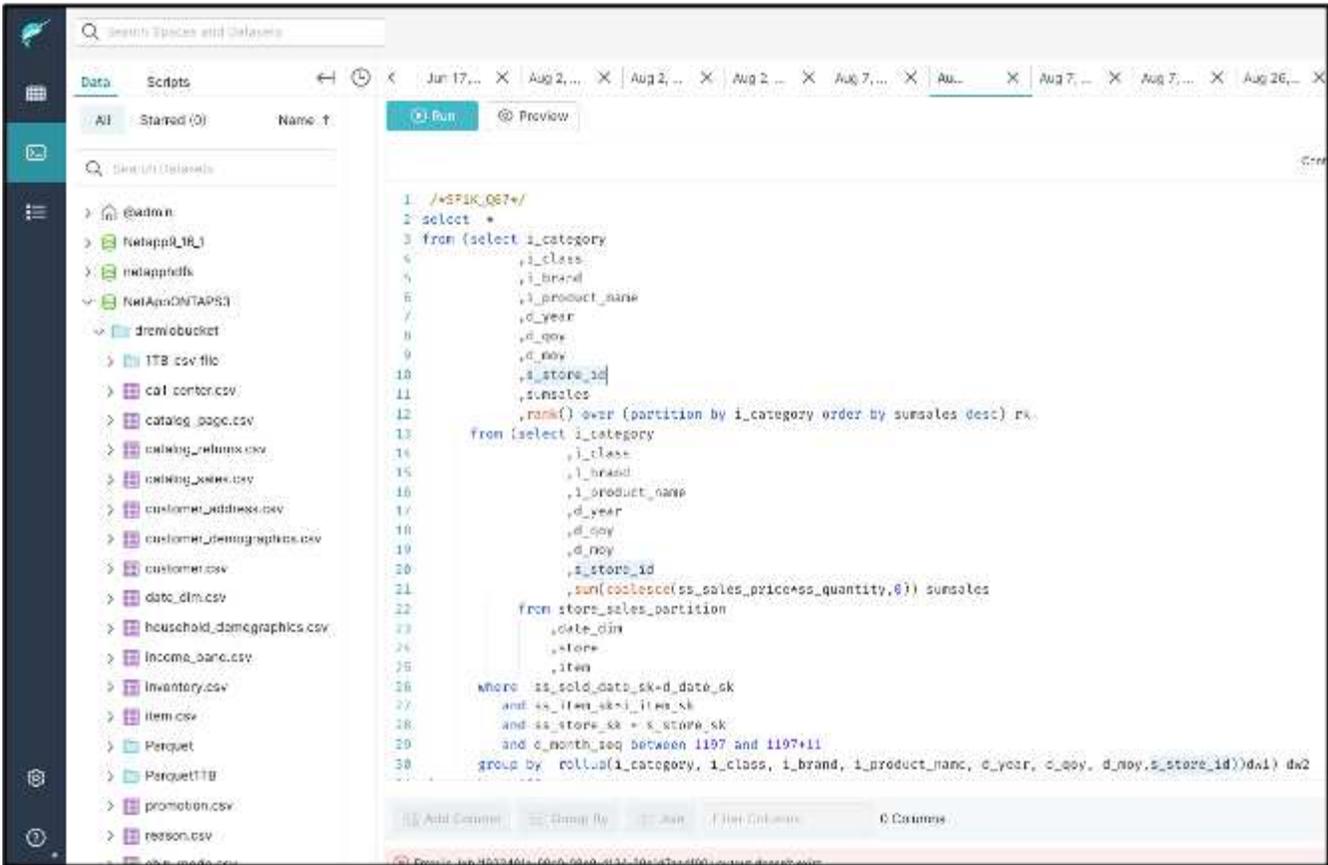
root@hadoopnode1:~# ls -ltrh /dremiocache
total 4.0K
drwx----- 3 nobody nogroup 4.0K Sep 13 16:00 spilling_stlrx2540m4-12-
10g_45678
root@hadoopnode1:~# ls -ltrh /opt/dremio/dremiocache/
total 8.0K
drwxr-xr-x 3 dremio dremio 4.0K Aug 22 18:19 spill_old
drwxr-xr-x 4 dremio dremio 4.0K Aug 22 18:19 cm
lrwxrwxrwx 1 root root 12 Aug 22 19:03 spill -> /dremiocache
root@hadoopnode1:~# ls -ltrh /dremiocache
total 4.0K
drwx----- 3 nobody nogroup 4.0K Sep 13 16:00 spilling_stlrx2540m4-12-
10g_45678
root@hadoopnode1:~# df -h /dremiocache
Filesystem                Size      Used Avail Use% Mounted on
10.63.150.159:/dremiocache_hadoopnode1  2.1T    209M   2.0T   1% /dremiocache
root@hadoopnode1:~#

```

4. コンテキストを選択します。私たちのテストでは、ONTAP S3にあるTPCDSによって生成されたparquetファイルに対してテストを実行しました。Dremioダッシュボード → SQLランナー → コンテキスト → NetAppONTAPS3 → Parquet1TB



1. DremioダッシュボードからTPC-DSクエリ67を実行する



1. すべての実行プログラムでジョブが実行されていることを確認します。Dremioダッシュボード → ジョブ → <jobid> → 生のプロファイル → EXTERNAL_SORTを選択 → ホスト名

Raw Profile

04-xx-04 - FILTER

04-xx-05 - WINDOW

04-xx-06 - EXTERNAL_SORT

Thread	Setup Time	Process Time	Wait Time	Max Batches	Max Records	Peak Memory	Hostname	Record Processing Rate	Operator State	Last Schedule Time
04-00-06	0.000s	0.000s	0.000s	0	0	129KB	strx2540-110-10g	0	CAN_CONSUME	16:35:54
04-01-06	0.000s	0.000s	0.000s	0	0	129KB	strx2540m4-04-10g	0	CAN_CONSUME	16:35:54
04-02-06	0.000s	0.000s	0.000s	0	0	129KB	strx2540m4-12-10g	0	CAN_CONSUME	16:35:54
04-03-06	0.017s	0.000s	0.000s	0	0	129KB	strx2540m4-13-10g	0	CAN_CONSUME	16:35:54
04-04-06	0.000s	0.000s	0.000s	0	0	129KB	strx2540-110-10g	0	CAN_CONSUME	16:35:54
04-05-06	0.000s	0.000s	0.000s	0	0	129KB	strx2540m4-04-10g	0	CAN_CONSUME	16:35:54
04-06-06	0.027s	0.000s	0.000s	0	0	129KB	strx2540m4-12-10g	0	CAN_CONSUME	16:35:54
04-07-06	0.003s	0.000s	0.000s	0	0	129KB	strx2540m4-13-10g	0	CAN_CONSUME	16:35:54

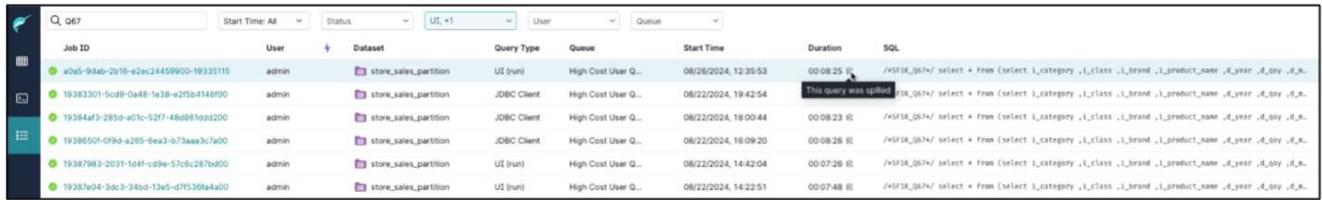
1. SQL クエリが実行されている場合、NetAppストレージ コントローラ内のデータ キャッシュの分割フォルダーを確認できます。

```

root@hadoopnode1:~# ls -ltrh /dremiocache
total 4.0K
drwx----- 3 nobody nogroup 4.0K Sep 13 16:00 spilling_stlrx2540m4-12-10g_45678
root@hadoopnode1:~# ls -ltrh /dremiocache/spilling_stlrx2540m4-12-10g_45678/
total 4.0K
drwxr-xr-x 2 root daemon 4.0K Sep 13 16:23 1726243167416

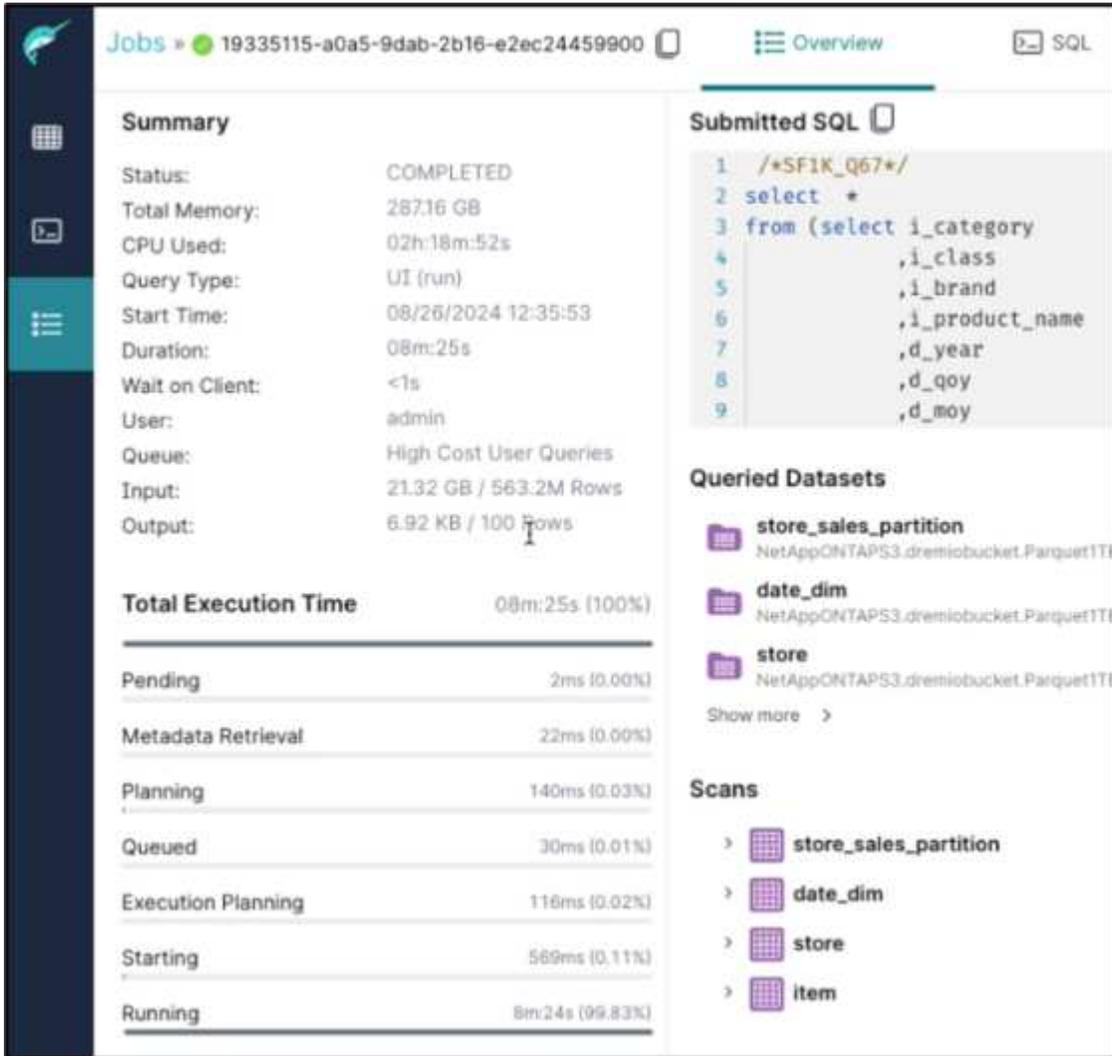
```

2. SQLクエリはスピルオーバーで完了しました



Job ID	User	Dataset	Query Type	Queue	Start Time	Duration	SQL
19383301-5cd9-0a48-1e38-e275b414900	admin	store_sales_partition	JDBC Client	High Cost User Q...	08/22/2024, 19:42:54	00:08:25	/*SF1K_Q67*/ select * from (select i_category ,i_class ,i_brand ,i_product_name ,d_year ,d_qoy ,d_moy
193844f3-285d-a07c-5277-48d8816d9200	admin	store_sales_partition	JDBC Client	High Cost User Q...	08/22/2024, 18:00:44	00:08:23	/*SF1K_Q67*/ select * from (select i_category ,i_class ,i_brand ,i_product_name ,d_year ,d_qoy ,d_moy
1938650f-0f9d-a265-6ea3-673aaa3c7a00	admin	store_sales_partition	JDBC Client	High Cost User Q...	08/22/2024, 16:09:20	00:08:26	/*SF1K_Q67*/ select * from (select i_category ,i_class ,i_brand ,i_product_name ,d_year ,d_qoy ,d_moy
19387983-2031-164f-cd9e-57c6c287bd00	admin	store_sales_partition	UI (run)	High Cost User Q...	08/22/2024, 14:42:04	00:07:26	/*SF1K_Q67*/ select * from (select i_category ,i_class ,i_brand ,i_product_name ,d_year ,d_qoy ,d_moy
19387a04-5ac3-346d-13e5-d7f53694a00	admin	store_sales_partition	UI (run)	High Cost User Q...	08/22/2024, 14:22:51	00:07:48	/*SF1K_Q67*/ select * from (select i_category ,i_class ,i_brand ,i_product_name ,d_year ,d_qoy ,d_moy

3. ジョブ完了の概



Jobs * 19335115-a0a5-9dab-2b16-e2ec24459900 Overview SQL

Summary

Status: COMPLETED
Total Memory: 287.16 GB
CPU Used: 02h:18m:52s
Query Type: UI (run)
Start Time: 08/26/2024 12:35:53
Duration: 08m:25s
Wait on Client: <1s
User: admin
Queue: High Cost User Queries
Input: 21.32 GB / 563.2M Rows
Output: 6.92 KB / 100 Rows

Total Execution Time

Total Execution Time	08m:25s (100%)
Pending	2ms (0.00%)
Metadata Retrieval	22ms (0.00%)
Planning	140ms (0.03%)
Queued	30ms (0.01%)
Execution Planning	116ms (0.02%)
Starting	569ms (0.11%)
Running	6m:24s (99.83%)

Submitted SQL

```
1 /*SF1K_Q67*/  
2 select *  
3 from (select i_category  
4         ,i_class  
5         ,i_brand  
6         ,i_product_name  
7         ,d_year  
8         ,d_qoy  
9         ,d_moy
```

Queried Datasets

- store_sales_partition
NetAppONTAPS3.dremiobucket.ParquetTTB
- date_dim
NetAppONTAPS3.dremiobucket.ParquetTTB
- store
NetAppONTAPS3.dremiobucket.ParquetTTB

Show more >

Scans

- store_sales_partition
- date_dim
- store
- item

要。

4. こぼれたデータのサイズを確認す

EXTERNAL_SORT 04-06



Runtime	1.68m (100%)
Startup	49.09ms (0.05%)
Processing	39.62s (39.36%)
IO Wait	1.02m (60.6%)

Overview/Main

Batches Processed:	104333
Records Processed:	387.6M
Peak Memory:	199 MB
Bytes Sent:	44 GB
Number of Threads:	180

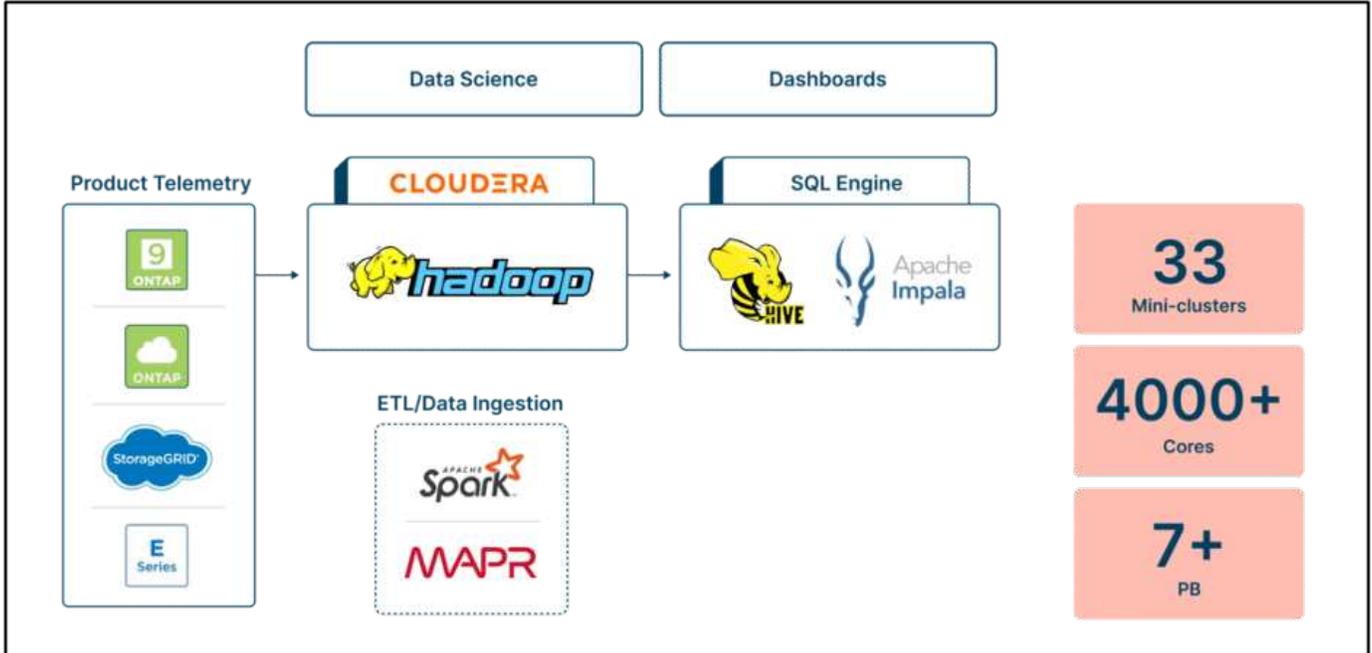
Operator Statistics

Merge Time Nanos:	0ns
Spill Count:	360
Spill Time Nanos:	37.68m
Total Spilled Data Size:	20,339,702,765
Batches Spilled:	97,854

NAS およびStorageGRIDオブジェクト ストレージにも同じ手順が適用されます。

顧客ユースケース

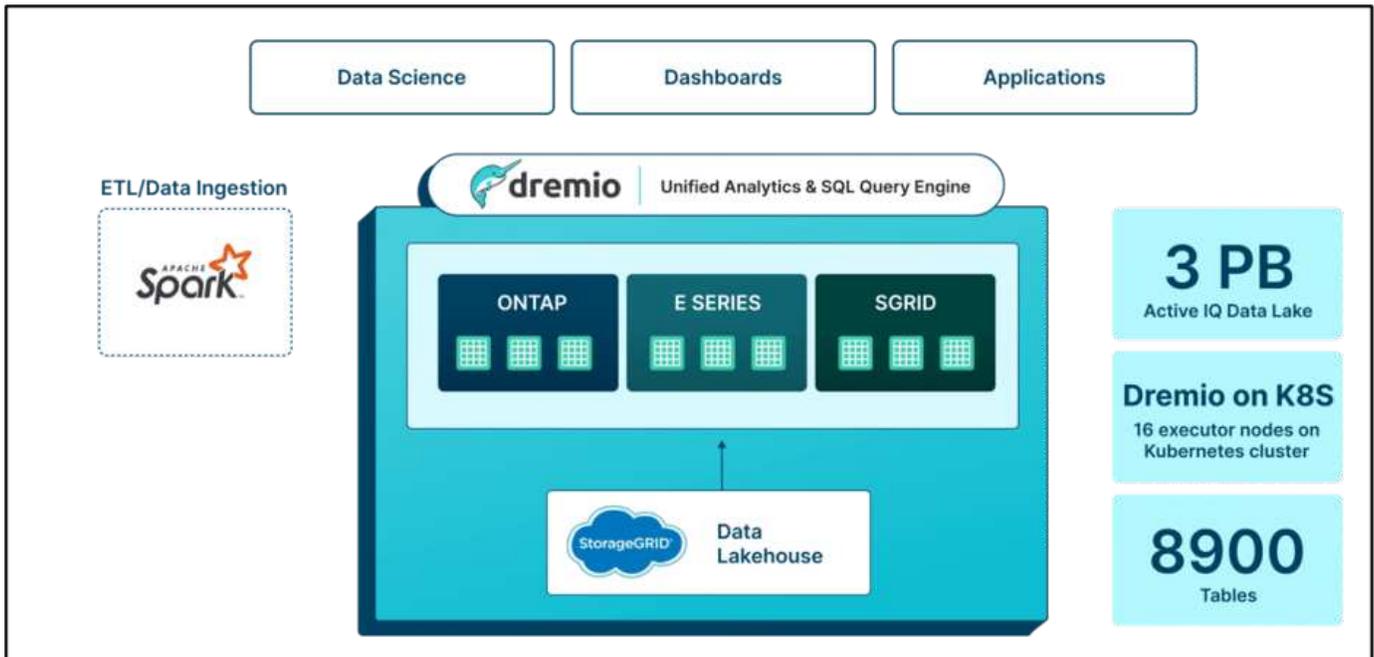
NetApp ActiveIQ のユースケース



課題: 当初は多数のユースケースをサポートするために設計された NetApp 独自の社内Active IQソリューションは、社内ユーザーと顧客の両方に包括的なサービスを提供するソリューションへと進化しました。しかし、基盤となる Hadoop/MapR ベースのバックエンド インフラストラクチャでは、データの急速な増加と効率的なデータ アクセスの必要性により、コストとパフォーマンスに関する課題が生じていました。ストレージを拡張すると、不要なコンピューティング リソースが追加され、コストが増加します。

さらに、Hadoop クラスターの管理には時間がかかり、専門知識が必要でした。データのパフォーマンスと管理の問題により状況はさらに複雑になり、クエリには平均 45 分かかり、誤った構成によりリソースが不足するようになりました。これらの課題に対処するため、NetApp は既存のレガシー Hadoop 環境に代わる手段を模索し、Dremio 上に構築された新しい最新ソリューションによってコストが削減され、ストレージとコンピューティングが分離され、パフォーマンスが向上し、データ管理が簡素化され、きめ細かな制御が可能になり、災害復旧機能も提供されると判断しました。

解決
:



Dremioにより、NetAppは段階的なアプローチでHadoopベースのデータインフラストラクチャを最新化し、統合分析のロードマップを提供できるようになりました。データ処理に大幅な変更を必要とする他のベンダーとは異なり、Dremioは既存のパイプラインとシームレスに統合し、移行時の時間と費用を節約しました。NetAppは、完全にコンテナ化された環境に移行することで、管理オーバーヘッドを削減し、セキュリティを強化し、耐障害性を強化しました。Dremioは、Apache IcebergやArrowなどのオープンエコシステムを採用することで、将来性、透明性、拡張性を確保しました。

Dremioは、Hadoop/Hiveインフラストラクチャの代替として、セマンティックレイヤーを通じて二次的なユースケース向けの機能を提供しました。既存のSparkベースのETLおよびデータ取り込みメカニズムはそのままに、Dremioは、重複のないデータの検出と探索を容易にする統合アクセスレイヤーを提供しました。このアプローチにより、データ複製係数が大幅に削減され、ストレージとコンピューティングが分離されました。

メリット: Dremioを使用することで、NetAppはデータ環境におけるコンピューティング消費とディスク容量要件を最小限に抑え、大幅なコスト削減を実現しました。新しいActive IQデータレイクは、3ペタバイトのデータを保持する8,900個のテーブルで構成されています。以前のインフラストラクチャでは、7ペタバイト以上でした。Dremioへの移行には、33個のミニクラスターと4,000個のコアからKubernetesクラスター上の16個のエグゼキューターノードへの移行も含まれていました。コンピューティングリソースが大幅に減少したにもかかわらず、NetAppはパフォーマンスの顕著な向上を実現しました。Dremioを介してデータに直接アクセスすることで、クエリの実行時間が45分から2分に短縮され、予測メンテナンスと最適化のための洞察を得るまでの時間が95%高速化されました。この移行により、コンピューティングコストが60%以上削減され、クエリが20倍以上高速化され、総所有コスト(TCO)が30%以上節約されました。

自動車部品販売の顧客ユースケース。

課題: この世界的な自動車部品販売会社では、経営陣と企業財務計画および分析グループが販売レポートの統合ビューを取得できず、個々の事業ラインの販売指標レポートを読み取って統合するしかありませんでした。その結果、顧客は少なくとも1日前のデータに基づいて意思決定を行うようになりました。新しい分析情報を取得するには、通常4週間以上かかります。データパイプラインのトラブルシューティングにはさらに多くの時間が必要となり、すでに長いタイムラインにさらに3日以上が追加されます。レポート開発プロセスとレポートのパフォーマンスが遅いため、アナリストコミュニティは、新しいビジネスインサイトを見つけたり、新しいビジネス行動を推進したりするのではなく、データの処理または読み込みが完了するまで継続的に待機する必要がありました。これらの問題のある環境は、さまざまな事業部門の多数の異なるデータベースで構成されており、多数のデータサイロが発生していました。遅くて断片化された環境では、アナリストが単一の真実のソースではなく独自の真実を導き出す方法が多すぎるため、データガバナンスが複雑になりま

した。このアプローチでは、データ プラットフォームと人件費で 190 万ドル以上かかりました。レガシープラットフォームを維持し、データ要求に応えるには、年間 7 人のフィールド テクニカル エンジニア (FTE) が必要でした。データ要求の増加に伴い、データ インテリジェンス チームは、将来のニーズに合わせてレガシー環境を拡張することができませんでした。

ソリューション: NetApp Object Store で大規模な Iceberg テーブルをコスト効率よく保存および管理します。Dremio のセマンティック レイヤーを使用してデータ ドメインを構築し、ビジネス ユーザーがデータ製品を簡単に作成、検索、共有できるようにします。

顧客へのメリット: • 既存のデータアーキテクチャを改善および最適化し、洞察を得るまでの時間を4週間から数時間に短縮 • トラブルシューティング時間を3日から数時間に短縮 • データプラットフォームと管理コストを38万ドル以上削減 • 年間2 FTEのデータインテリジェンス作業を削減

まとめ

結論として、この技術レポートでは、ONTAP S3、NAS、StorageGRIDなどのNetAppストレージコントローラーのさまざまなデータソースと組み合わせた、Dremioを使用したq Hybrid Iceberg Lakehouseの包括的な展開の詳細について説明しました。展開プロセスは正常に実行され、TPC-DS ベンチマーク ツールを使用して、さまざまなデータソースにわたって 99 個の SQL クエリが実行されました。このレポートでは、NetApp内の顧客の使用事例も調査し、多様なビジネス要件を満たすDremioの汎用性と有効性を実証しました。さらに、自動車部品販売の顧客を対象とする特定のユースケースが検討され、データ分析と洞察のためにDremioを活用する実際的なアプリケーションと利点が強調されました。

全体として、このドキュメントは、NetAppストレージコントローラーを使用したDremioの導入と使用方法を理解するための貴重なリソースとして機能し、さまざまな業界でデータ主導の意思決定と最適化を推進する機能と可能性を示しています。

詳細情報の入手方法

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、次のドキュメントや Web サイトを参照してください。

- Zookeeperのインストール

<https://medium.com/@ahmetfurkandemir/distributed-hadoop-cluster-1-spark-with-all-dependencies-03c8ec616166>

- ドレミオ

<https://docs.dremio.com/current/get-started/cluster-deployments/deployment-models/standalone/standalone-tarball/>

- storageGRIDを使用したDremioの設定

<https://docs.netapp.com/us-en/storagegrid-enable/tools-apps-guides/configure-dremio-storagegrid.html#configure-dremio-data-source>

- NetAppのユースケース

<https://www.dremio.com/customers/netapp/>

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权を有し、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权については、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。