

AWS Black Belt Online Seminar

AWS DataSync

佐藤 真也

Solutions Architect

2024/09



AWS Black Belt Online Seminar とは

- 「サービス別」「ソリューション別」「業種別」などのテーマに分け、アマゾン ウェブ サービス ジャパン合同会社が提供するオンラインセミナーシリーズです
- AWS の技術担当者が、AWS の各サービスやソリューションについてテーマごとに動画を公開します
- 以下の URL より、過去のセミナー含めた資料などをダウンロードすることができます
 - <https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-service-cut/>
 - <https://www.youtube.com/playlist?list=PLzWGOASvSx6FIwIC2X1nObr1KcMCBBlqY>



ご感想は X (Twitter) へ！ハッシュタグは以下をご利用ください
#awsblackbelt

内容についての注意点

- 本資料では資料作成時点のサービス内容および価格についてご説明しています。AWS のサービスは常にアップデートを続けているため、最新の情報は AWS 公式ウェブサイト (<https://aws.amazon.com/>) にてご確認ください
- 資料作成には十分注意しておりますが、資料内の価格と AWS 公式ウェブサイト記載の価格に相違があった場合、AWS 公式ウェブサイトの価格を優先とさせていただきます
- 価格は税抜表記となっています。日本居住者のお客様には別途消費税をご請求させていただきます
- 技術的な内容に関しましては、有料の [AWS サポート窓口](#) へお問い合わせください
- 料金面でのお問い合わせに関しましては、[カスタマーサポート窓口](#) へお問い合わせください (マネジメントコンソールへのログインが必要です)

自己紹介

佐藤 真也

アマゾン ウェブ サービス ジャパン合同会社
技術統括本部 フィナンシャルサービス インダストリ 技術本部
保険ソリューション部



好きな AWS サービス

- Amazon Simple Storage Service (S3)
- Amazon FSx シリーズ



前提知識と本セミナーの対象者

前提知識

- AWS のグローバルインフラストラクチャや基本サービスの知識
- AWS のストレージサービスの概要

対象者

- AWS へのデータ移行方法を学びたい方
- AWS DataSync の機能を深く知りたい方

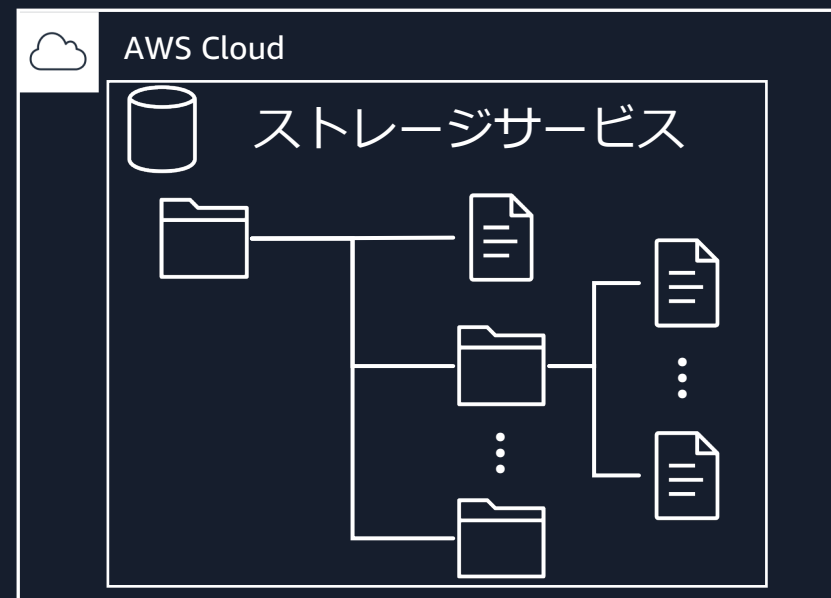
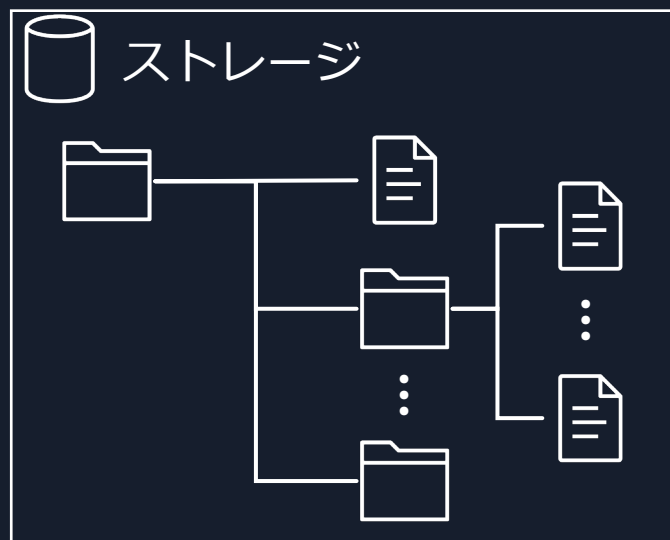
アジェンダ

1. AWS DataSync とは
2. AWS DataSync Agent
3. AWS DataSync タスク
4. AWS DataSync を用いたデータ転送の流れ
5. 他のデータ転送ソリューションとの比較
6. 補足と注意点

AWS DataSync とは

大量のデータを転送するときの課題

送信元ストレージから、LAN や WAN を経由して AWS クラウドへ辿り着くまでに多くの課題に直面する



送信元ストレージのボトルネック

LAN のボトルネック

WAN のボトルネック

送信先ストレージのボトルネック

送信対象のファイル整理

伝送路の暗号化

データとメタデータの整合性検証

スケジューリング・エラーハンドリング・パフォーマンスの向上

+ スクリプトの開発とデプロイ

AWS DataSync とは

データの転送に関するマネージドサービスで、AWS ストレージサービスとオンプレミス、他クラウドサービス間でデータの移動が「簡単に」「高速に」「安全に」「低コストで」実現できる



簡単なデータ転送



高速なデータ転送



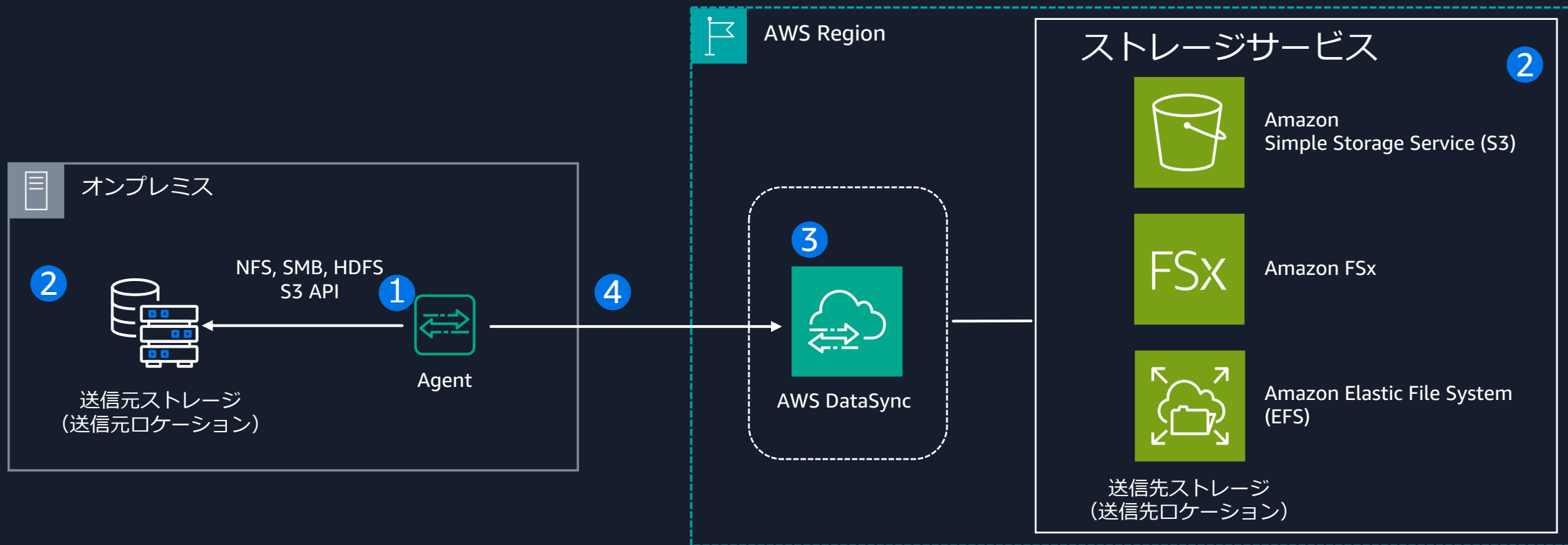
セキュリティ
高い信頼性



低コスト

AWS DataSync 全体像: データ転送の流れ

1. オンプレミスに **Agent** を配置する
2. 送信元と送信先ストレージを、それぞれ**送信元ロケーション**と**送信先ロケーション**として設定する
3. 転送スケジュールなどを定義した**タスク**を設定する
4. その後**タスクを実行**し、Agent は送信元ロケーションからデータを読み取り、AWS DataSync へ送信する
AWS DataSync は送信先ロケーションへデータを保存する



簡単なデータ転送



Agent をデプロイしタスクを作成し実行するだけで、データを転送できる



Vmware



Hyper-V



KVM

Agent

容易にセットアップできる仮想マシンプラットフォームで、データ転送を実行する



タスクで

- 送信元ロケーションと送信先ロケーション
- 送信対象ファイルの整理
- データの整合検証方法、スケジュール、帯域制限などを設定する

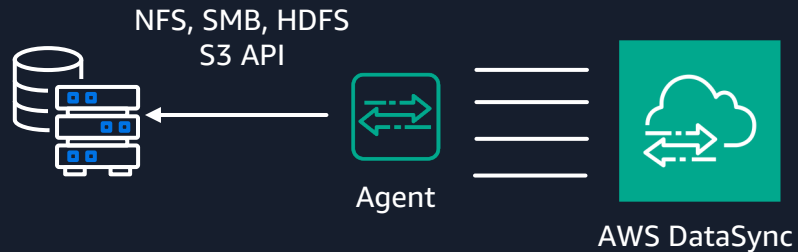
タスク実行

特定のスケジュールで特定のフォルダ・ファイルだけを、帯域制限しながら自動的にデータを転送し、データの整合性も検証する

高速なデータ転送

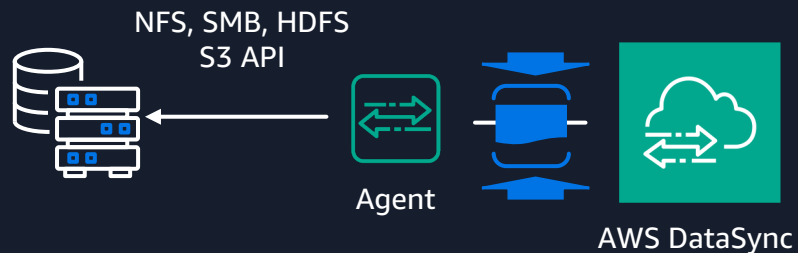


Agent は AWS DataSync ヘデータを転送する時、マルチセッション・圧縮・ロードバランス機能を自動的に用いるため、高速なデータ転送が実現できる



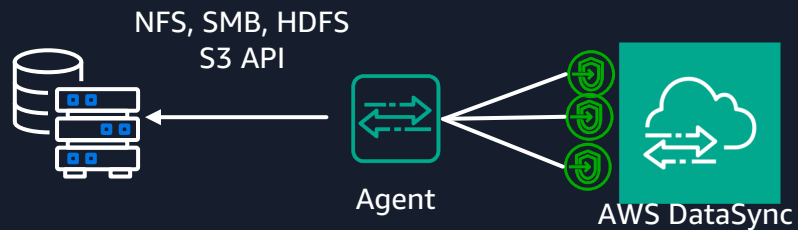
マルチセッション

Agent と AWS DataSync の間では、同時に複数の TCP セッションを生成し、並列にデータを転送する



圧縮

Agent と AWS DataSync の間では、データを圧縮し転送を効率化できる



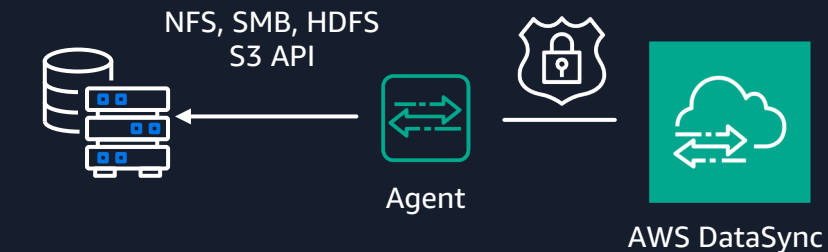
ロードバランス

AWS DataSync 側に複数のエンドポイントを自動生成し、Agent からのデータ転送を負荷分散できる

セキュリティと高い信頼性



AWS DataSync はデータの暗号化とエラーハンドリングに対応している



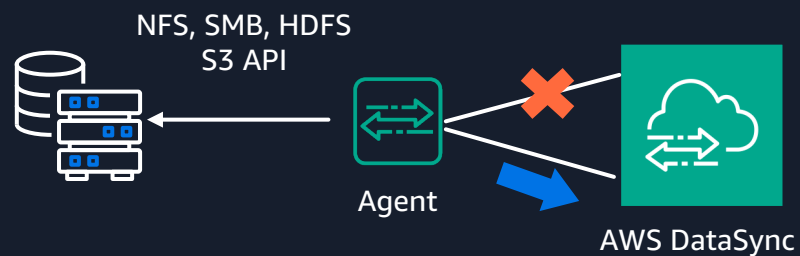
伝送路の暗号化

Agent と AWS DataSync の間では、TLS で暗号化できる



保管時の暗号化

送信先の Amazon S3 や Amazon EFS、Amazon FSx シリーズではデータ保管時に暗号化することができる



エラーハンドリング

ネットワークの切断などで転送に失敗した場合、エラーを記録しつつ、次のスケジュールタスクで再送することができる

低コスト



AWS DataSync は 1 GB あたり 0.0125 USD のデータの転送料金で、データ転送に便利な機能を独自開発・デプロイすることなく利用でき、開発・運用コストを削減できる

データ転送に便利な機能の例

お客様が担当する作業

AWS が提供する機能

送信対象のファイル整理

スケジューリング

データの整合性検証

帯域制限

マルチセッション

圧縮

転送データの暗号化

スケジューリング

エラーハンドリング

スクリプトの開発

スクリプトでの運用

要件に応じて
開発が必要

送信対象のファイル整理

スケジューリング

データの整合性検証

帯域制限

マルチセッション

圧縮

転送データの暗号化

スケジューリング

エラーハンドリング

Agent をデプロイ

AWS DataSync

AWS DataSync と同じアカウント上の AWS サービスを送信先に指定する

送信元ロケーション	送信先ロケーション
<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※	<ul style="list-style-type: none">お客様管理の NFS/SMB/HDFS ストレージまたは、S3 API 互換ストレージ他クラウドストレージサービスAWS Snowball Edge S3 互換ストレージ
<ul style="list-style-type: none">お客様管理の NFS/SMB/HDFS ストレージまたは、S3 API 互換ストレージ他クラウドストレージサービスAWS Snowball Edge S3 互換ストレージ	<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※
<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※	<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※
Amazon S3	Amazon S3 on Outposts
Amazon S3 on Outposts	Amazon S3

※ Amazon FSx for NetApp ONTAP、Windows File Server、Open ZFS、Lustre



AWS DataSync と異なるアカウント上の AWS サービスを送信先に指定する

送信元ロケーション	送信先ロケーション
<ul style="list-style-type: none">NFS ロケーションとして設定した Amazon EFS ※1SMB ロケーションとして設定した Amazon FSx for Windows File Server ※1	<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※2
Amazon S3	<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※2
<ul style="list-style-type: none">Amazon S3Amazon EFSAmazon FSx シリーズ※2	Amazon S3
お客様管理の NFS/SMB/HDFS ストレージまたは、S3 API 互換ストレージ	Amazon S3

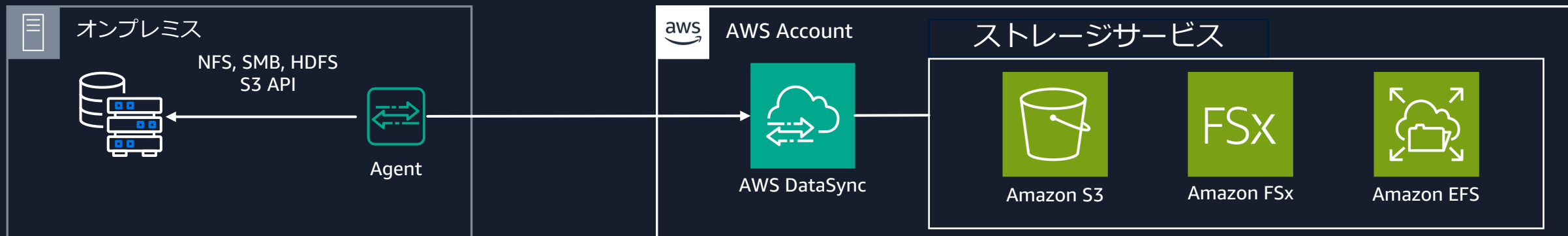
※1 Amazon EFS を送信元ロケーションとして選択するのではなく、NFS を送信元ロケーションとして一旦設定する。その上で、Amazon EFS のファイルシステムのドメイン名を NFS サーバのドメイン名として設定する。Amazon FSx for Windows File Server の場合も同様である。

※2 Amazon FSx for NetApp ONTAP、Windows File Server、Open ZFS、Lustre

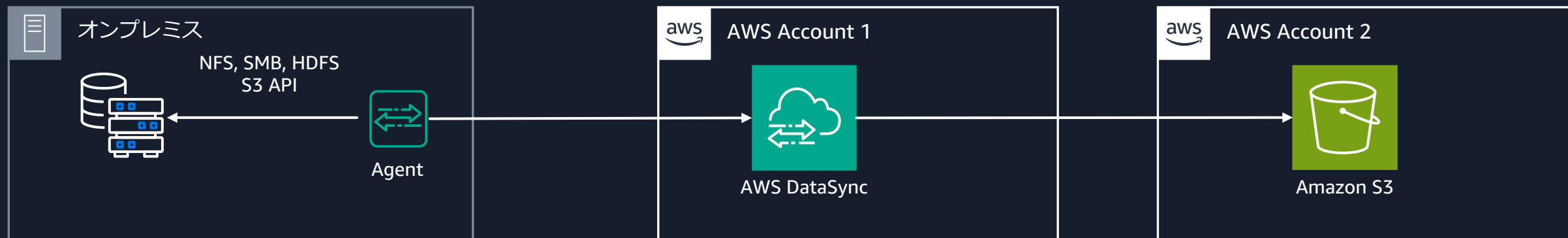


データ転送アーキテクチャの例

AWS DataSync と同じアカウント上の AWS サービスを送信先ロケーションに指定する場合



AWS DataSync と異なるアカウント上の AWS サービスを送信先ロケーションに指定する場合

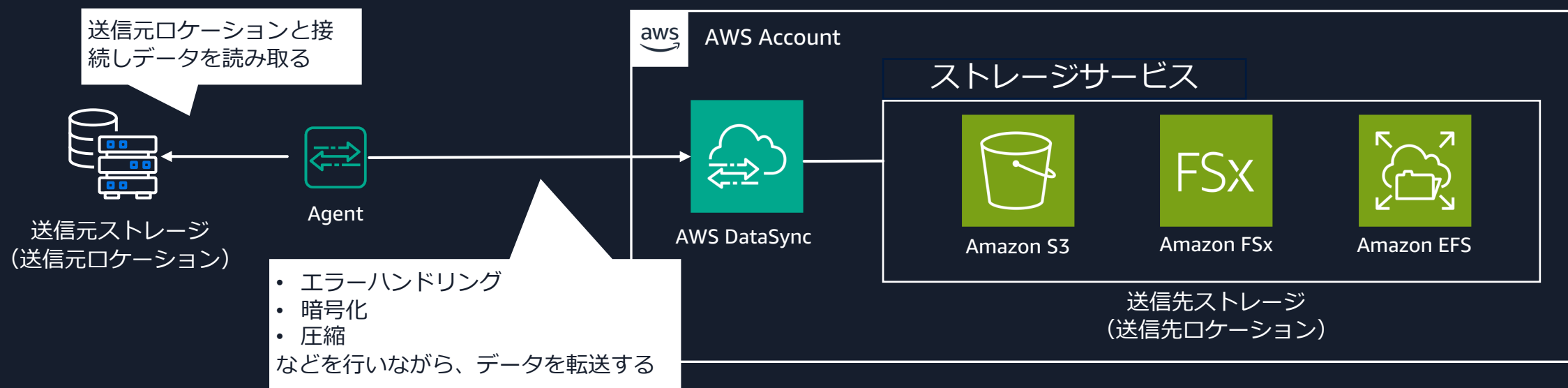


AWS DataSync Agent



AWS DataSync Agent とは

仮想マシンアプライアンスで、送信元ストレージからデータを読み取り、AWS DataSync へデータを転送する



AWS DataSync Agent の要件

Agent はハイパーバイザまたは Amazon EC2、Amazon Outposts 上で実行できる



Vmware



Hyper-V



KVM

対応ハイパーバイザ:

- VMware ESXi (version 6.5, 6.7, 7.0, or 8.0)
- Linux Kernel-based Virtual Machine (KVM)
- Microsoft Hyper-V (version 2012 R2, 2016, or 2019)

ディスク容量と RAM の要件:

- VM イメージとシステムデータのインストールに利用する 80 GB のディスク容量
- 1 回のタスクで取り扱うファイル数が 2,000 万ファイル以下の場合 32 GB RAM
2,000 万ファイルを超える場合 64 GB RAM

詳細: <https://docs.aws.amazon.com/datasync/latest/userguide/agent-requirements.html>



Amazon EC2



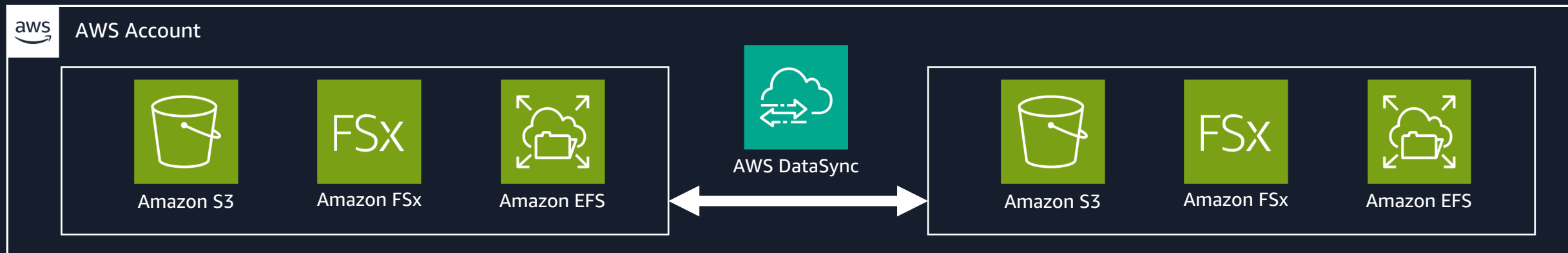
Amazon EC2:

- AWS DataSync Agent を含む AMI を使用する
- 1 回のタスクで取り扱うファイル数が 2,000 万ファイル以下の場合 m5.2xlarge
2,000 万ファイルを超える場合 m5.4xlarge

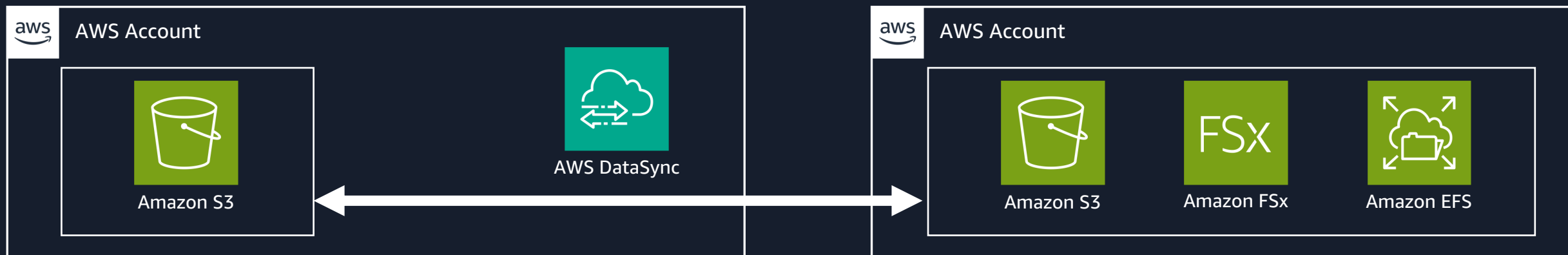
AWS DataSync Agent が不要なケース

次のパターンはクロスリージョンを含み、Agent が不要である。記載がないパターンは Agent が必須となる

- 同じアカウントにおける AWS ストレージサービス間でのデータ転送



- Amazon S3 と他の AWS ストレージサービス間で、クロスアカウントデータ転送

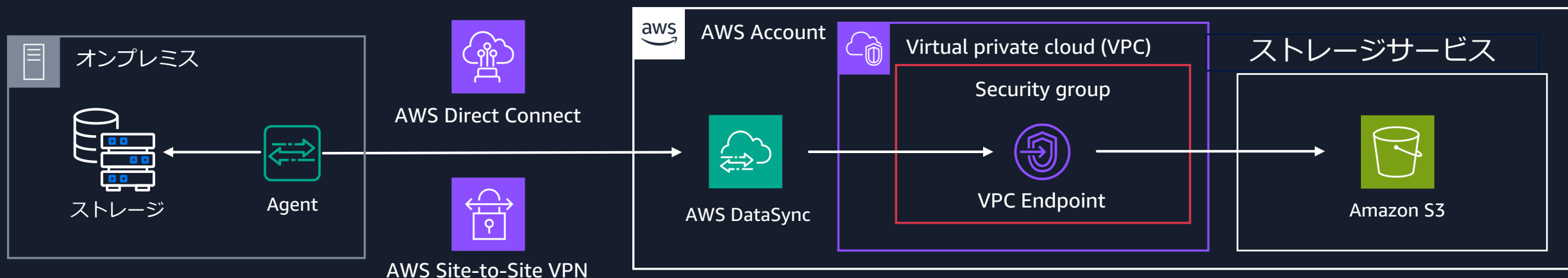


AWS DataSync のエンドポイント

AWS DataSync と Agent が通信するエンドポイントを要件に合わせて選択できる

検証方法	説明
パブリックエンドポイント	インターネット経由で、AWS DataSync と通信する
FIPS サービスエンドポイント	インターネット経由で、AWS DataSync と通信する。FIPS に準拠したプロセスでデータを転送する
VPC エンドポイント	Amazon VPC 経由で AWS DataSync と通信する

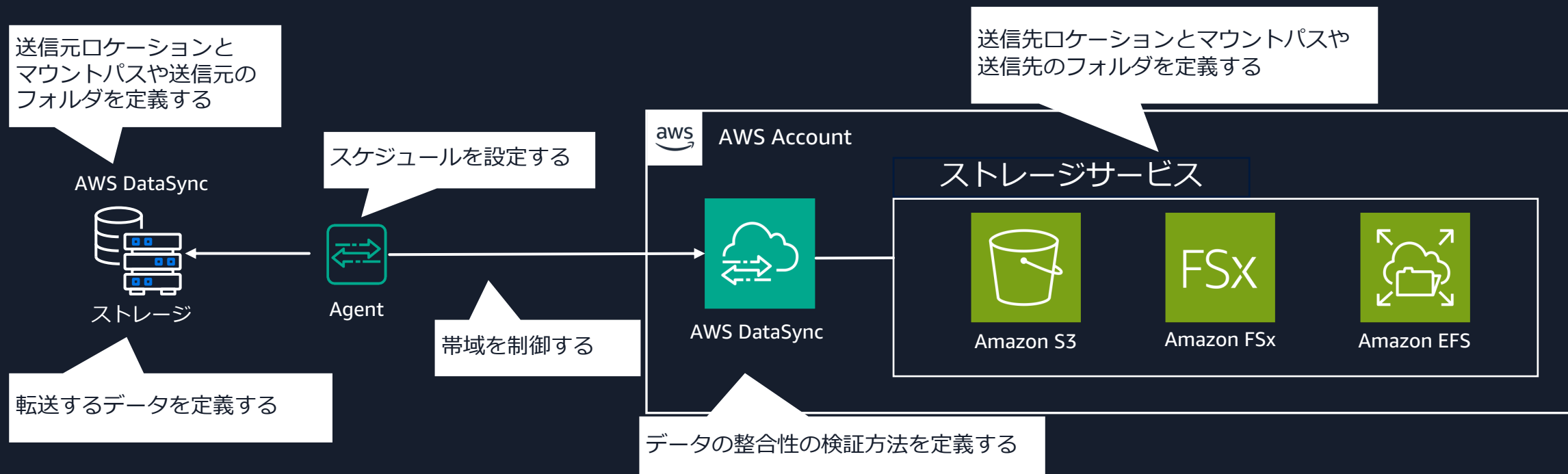
VPC エンドポイントを利用した際のデータ転送アーキテクチャ



AWS DataSync タスク

AWS DataSync タスクとは

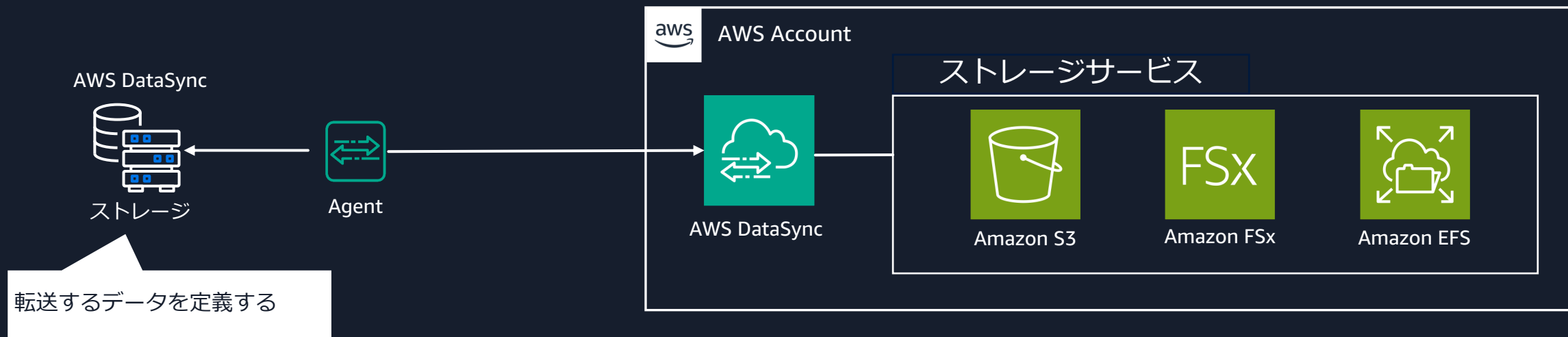
AWS DataSync が「どこから」「どこへ」「何を」「どのように」転送するかを設定するタスクを実行することで、AWS DataSync はデータの転送をスケジュールに従って開始する



転送するデータを定義する

送信元ロケーションから送信先ロケーションへデータを送信する場合に、次の設定ができる

- マニフェストによる送信対象データのリスト化
- フィルタを用いて特定の名前のファイルやフォルダのみ送信または送信から除外
- データの同期方法



マニフェスト

マニフェストはファイルをリスト化し、特定のファイルだけを送信できる

- 指定した送信元ロケーションの起点となるフォルダを基準に、ファイル名をリスト化する
送信対象のファイルが複数ある場合には改行する

NFS ストレージの場合

```
photos/picture1.png  
photos/picture2.png  
photos/picture3.png
```

Amazon S3 の場合、Version ID を含むことができる

```
photos/picture1.png,111111  
photos/picture2.png,121212  
photos/picture3.png
```

- ファイル名にカンマや改行、引用符を含む場合は、「"」を用いる

ファイル名が
「filename,with,commas.txt」の場合

```
"filename,with,commas.txt"
```

ファイル名に改行が入る場合

```
"Thank  
You."
```

ファイル名が
「filename"with"commas.txt」の場合

```
filename""with""quotes.txt
```

- マニフェストを csv または txt ファイルとして、AWS DataSync と同じアカウント/リージョンの Amazon S3 バケットへと保存する

マニフェストの注意点および TIPS

- ディレクトリやフォルダ単位で指定することはできない。ファイル名を含めたフルパスを指定する
- 4,096 文字または 1,024 バイトを超えるファイル名やオブジェクトパスは定義できない
- 重複したファイル名やオブジェクトパスを定義できない
- フィルタと併用することはできない
- マニフェストを用いる場合、送信先ロケーションのファイルは削除しない。そのため、データ転送時のオプションとして「Keep deleted files」を選択できず、デフォルトで保持される
- マニフェスト使用時には、AWS DataSync のタスク設定時に「s3:GetObject」と「s3:GetObjectVersion」の権限を与える必要がある

その他の注意事項は[こちら](#)を参照する

フィルタ

特定のパス、名前、拡張子を記述することで、そのパターンに一致するデータを転送することができる
また、転送から除外するパターンも記述できる

- フィルタで利用できる特殊文字

特殊文字	意味
* (ワイルドカード)	前方または後方一致を表す。例、 /movies* は /movies も /movies_folder も含む
(パイプ)	OR を表す。例、 *.tmp *.temp は tmp または temp の拡張子を持つファイルを含む
¥ (バックスラッシュ)	*, , ¥ に対するエスケープ文字。例、 ¥*.temp と ¥¥.temp はそれぞれ *.temp と ¥.temp というファイル名を表す。

- 送信元ロケーションの起点となるフォルダを基準にパス、名前、拡張子を記述する
この際、起点となるフォルダ直下のフォルダは、 / から記述する
- 複数の記述を行う場合には、 | で区切る必要がある。ただし、AWS DataSync コンソールから設定する場合には不要である

フィルタの例

Includes [Info](#)
Specify the files, objects, and folders that you want to transfer.

Add pattern

/prod ×

/prod/ ×

/prod* ×

/Prod ×

Excludes – オプション [Info](#)
Specify the files, objects, and folders that you don't want to transfer.

Add pattern

/prod/temp ×

/prod/temp* ×

/prod/dev/temp1.text ×

*.temp ×

- 起点となるフォルダ直下の `/prod` フォルダ、`/prod` と前方一致するフォルダまたはファイル、`/Prod` フォルダが転送対象となる
 - `/prod` と `/prod/` は同じ意味で解釈される
 - 大文字と小文字は区別される
- 起点となるフォルダ直下の `/prod/temp` フォルダ、`/prod/temp` と前方一致するフォルダまたはファイル、`/prod/dev/temp1.text`、`.temp` ファイルは転送対象から除外される

フィルタの注意点および TIPS

- ワイルドカードを用いた後方一致は、除外パターンを記述するときのみ利用できる
*.text のような後方一致を転送対象として記述することはできない
- 送信元または送信先ロケーションに Amazon S3 を選択した場合、/ というオブジェクトキーをファイルシステムにおけるフォルダとして扱う
- デフォルトで以下のフォルダの転送を AWS DataSync は除外する
 - .snapshot
 - /.aws-datasync および /.awssync
 - /.zfs

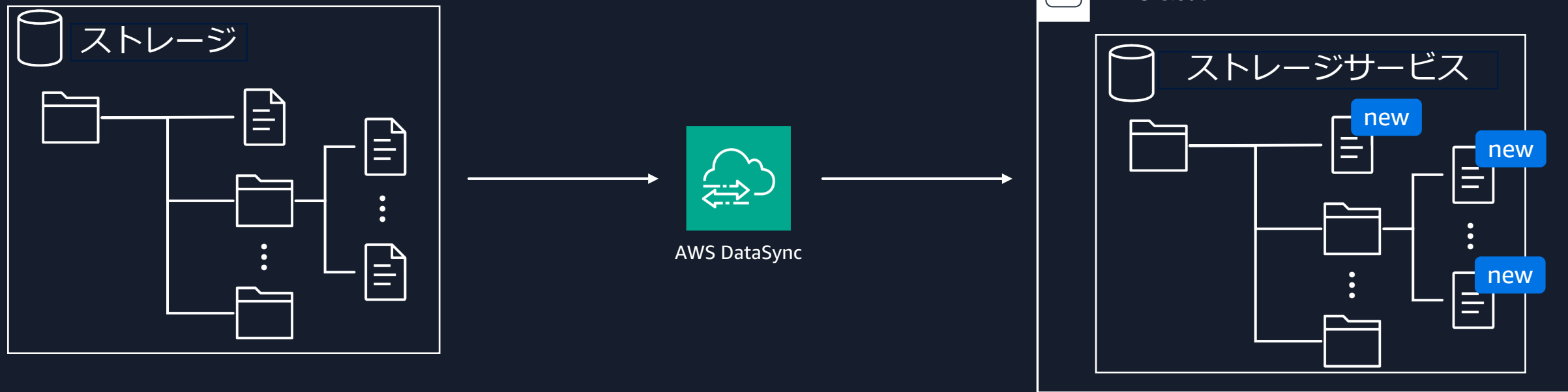
その他の注意事項は[こちら](#)を参照する

AWS DataSync のデータ同期方法 (1)

AWS DataSync ではタスクを設定する際に、送信元と送信先ロケーションで差分が存在するときの挙動として「すべてのデータを送信する」か「変更したデータだけ送信する」か定義できる

すべてのデータを送信する場合

すべてのデータを送信元ロケーションから送信先ロケーションへと送信し、送信先ロケーション上のデータを上書きする



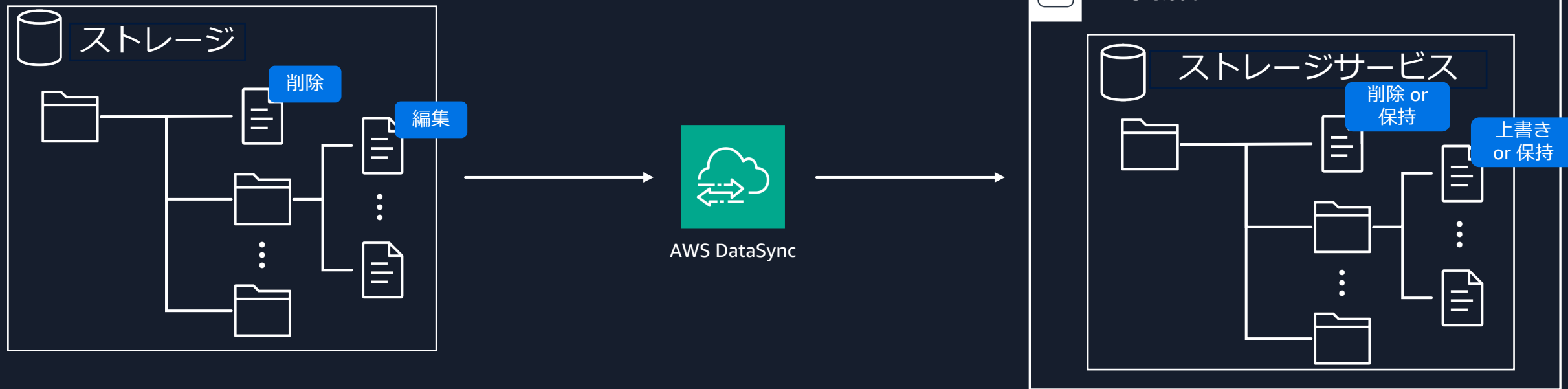
AWS DataSync のデータ同期方法 (2)

AWS DataSync ではタスクを設定する際に、送信元と送信先ロケーションで差分が存在するとき「すべてのデータを送信する」か「変更したデータだけ送信する」か定義できる

変更したデータだけ送信する場合

AWS DataSync はストレージ間のデータとメタデータをスキャンした上で、差分を判断する

- 送信元ストレージでデータを編集した場合、送信先ロケーション上で上書きするか否か選択できる
- 送信元ストレージでデータを削除した場合、送信先ロケーションで削除しない設定ができる (Keep deleted files)



AWS DataSync におけるデータの整合性の検証

3つの検証方法があり、(1) Verify only the data transferred を推奨する

検証方法	説明
(1) Verify only the data transferred	<ul style="list-style-type: none">データの転送中、チェックサムに基づいて整合性を検証する送信処理完了後、送信したデータとそのメタデータのチェックサムをストレージ間で比較し検証する
(2) Verify all data in the destination	<ul style="list-style-type: none">データの転送中、チェックサムに基づいて整合性を検証する送信処理完了後、すべてのデータとそのメタデータのチェックサムをストレージ間で比較し、完全に同期していることを確認する
(3) Check integrity during transfer	データの転送中のみ、チェックサムに基づいて整合性を検証する

(2) Verify all data in the destination の注意点/TIPS

- ストレージ間で完全に同期していることを確認できるものの、ストレージヘルクエストが発生する
例えば Amazon S3 を用いた場合、リクエストにかかる料金に注意する
- マニフェストを利用する場合には、記載したファイルのみ整合性を検証する
- Amazon S3 Glacier Flexible Retrieval や Glacier Deep Archive をロケーションに用いる場合には使えない

AWS DataSync におけるメタデータの扱い (1)

コピーされるメタデータは送信元と送信先ロケーションの組み合わせに依存する
ここでは一部の例のみ取り扱い、すべての組み合わせは[こちら](#)を参照する

送信元	送信先	コピーされるメタデータ※1
<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	<ul style="list-style-type: none">Amazon S3	<ul style="list-style-type: none">ファイルとディレクトリの更新日時ファイルとディレクトリのアクセス日時※2User ID と group IDPOSIX permission

送信元	送信先	コピーされるメタデータ
<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	<ul style="list-style-type: none">ファイルとディレクトリの更新日時ファイルとディレクトリのアクセス日時※2User ID と group IDPOSIX permission

AWS DataSync におけるメタデータの扱い (2)

コピーされるメタデータは送信元と送信先ロケーションの組み合わせに依存する
ここでは一部の例のみ取り扱い、すべての組み合わせは[こちら](#)を参照する

送信元	送信先	コピーされるメタデータ
<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	<ul style="list-style-type: none">ファイルのアクセス/更新/作成日時ファイル所有者の SIDRead-only、system などの属性※NTFS DACL/SACL

送信元	送信先	コピーされるメタデータ
<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	<ul style="list-style-type: none">Amazon S3	<ul style="list-style-type: none">デフォルトの POSIX メタデータが作成され、Windows のメタデータは破棄される

タスクを作成する (1)

送信元ロケーションの作成

送信元のロケーションのオプション

送信元のロケーションを作成または選択する

データ転送元のロケーションを選択する

☒ 新しいロケーションを作成する ☐ 既存のロケーションを選択する

設定

ロケーションタイプ

Amazon EFS ファイルシステム ▼

リージョン

US West (Oregon) us-west-2 ▼

ファイルシステム

▼ 

そのほか、マウントパス、サブネット、セキュリティグループを設定する

送信先ロケーションの作成

送信先ロケーションのオプション

送信先のロケーションを作成または選択する

データ転送先のロケーションを選択する

☒ 新しいロケーションを作成する ☐ 既存のロケーションを選択する

設定

ロケーションタイプ

Simple Storage Service (Amazon S3) ▼

リージョン

US West (Oregon) us-west-2 ▼

S3 バケット

us-west-2 ▼ 

us-west-2 の Simple Storage Service (Amazon S3) バケットである必要があります。

そのほか、Amazon S3 のストレージクラスやフォルダ、AWS IAM ロールを設定する

タスクを作成する (2)

送信対象のデータを定義する

特定のデータを転送対象に設定する場合には、フィルタまたはマニフェストを使用する

Task name – オプション

Source data options [Info](#)

Contents to scan
Specify the data in your source location that you want to transfer to your destination.

Excludes – オプション [Info](#)
Specify the files, objects, and folders that you don't want to transfer.

Add pattern

OR

Task name – オプション

Source data options [Info](#)

Contents to scan
Specify the data in your source location that you want to transfer to your destination.

☐ Using filters
Scan files, objects, and folders matching include patterns.
Exclude those that match any exclude patterns.

☒ Using a manifest
Scan files or objects declared in a manifest.

マニフェストファイル [Info](#)
Simple Storage Service (Amazon S3) のマニフェストファイルを選択します。マニフェストファイルの各ファイルまたはオブジェクトパスは、次の条件を満たす必要があります。(bulletedList) たとえば、/my-folder/my-text.txt がタスクの送信元ロケーションのすぐ下にある場合は、/my-folder/my-text.txt をマニフェストファイルの個別の行として指定します。

オブジェクトのバージョン

S3 URI

バージョンを選...

表示

S3 を参照

Format: s3://bucket/prefix/object

マニフェストのアクセスロール [Info](#)
S3 からマニフェストファイルを読み取るために DataSync が使用するロールを選択します。

自動生成する

Role must allow:

タスクを作成する (3)

Transfer options

Transfer mode
Determines whether DataSync transfers only the data and metadata that differ between the source and the destination.

Transfer only data that has changed ▼

Verification
Verify the data in your destination after transfer. DataSync always performs data integrity checks while transferring and writing data.

Verify only the data transferred ▼

Bandwidth limit
Allocate the maximum network bandwidth to be utilized by this task.

☐ Use available

50 MIB/s ▼

☒ Keep deleted files
Keep files in destination even when deleted from source

☒ Overwrite files
Overwrite files in destination with files from source

▼ 追加の設定

☒ Copy ownership
Maintain user and group ID

☒ Copy permissions
Maintain existing permissions

☒ Copy timestamps
Maintain access time and modification time

☒ Queuing
Allow this task to be started and queued until a previous execution completes.

データの同期方法を定義する

データの整合性の検証方法を定義する

帯域を制御する

変更したデータだけ送信する際の同期方法を定義する

Amazon EFS 上のファイルのメタデータをコピーする

キューの有無を設定する
以前のタスクが完了するまで
キューに保存できる

タスクを作成する（4）

スケジュールの設定では、毎時などの頻度や Cron 式に基づいた設定ができる

Schedule

Specify when you want this task to run.

頻度

カスタム ▼

☒ Cron 式
Use for task schedules that run on a specific time and day, such as 8 AM on the first Wednesday of every month.

☐ Rate expression
Use for task schedules that run on a regular interval, such as every 12 hours.

Cron 式

複雑なスケジュールには cron 式を使用します。 [詳細と構文の例を参照してください](#)

0/60 ***?*

☒ Enable schedule

AWS DataSync では 1 時間未満の間隔でデータの転送をスケジュールすることはできないので注意する

※ 1 時間未満の間隔で、AWS DataSync を用いたデータの転送が必須な場合、以下の方法で実現できる

- タスクを複数作成する
- 定期的なタスクの実行をスクリプトなどで実装する

タスクを作成する (5)

タスク実行のレポートやログを保存できる

Task report - オプション

Get a detailed report for your DataSync transfer task, including specific file names and how much data you moved. Amazon S3 pricing applies. [詳細はこちら](#)

Report type

None ▼

Logging

ログレベル

CloudWatch Logs に記録される詳細レベルを定義します。CloudWatch Logs の料金は、DataSync によって生成されるログに適用されます。

転送エラーなどの基本情報をログに記録する ▼

CloudWatch Log グループ

エラーやその他の障害などのタスク実行アクティビティをログに記録します。

▼

自動生成する

☒ CloudWatch リソースポリシーを作成する

ロググループにログをアップロードするには、DataSync に十分なアクセス許可を付与するリソースポリシーが必要です。 [詳細はこちら](#)

設定することで、指定した Amazon S3 バケットへと次のいずれかを配信する

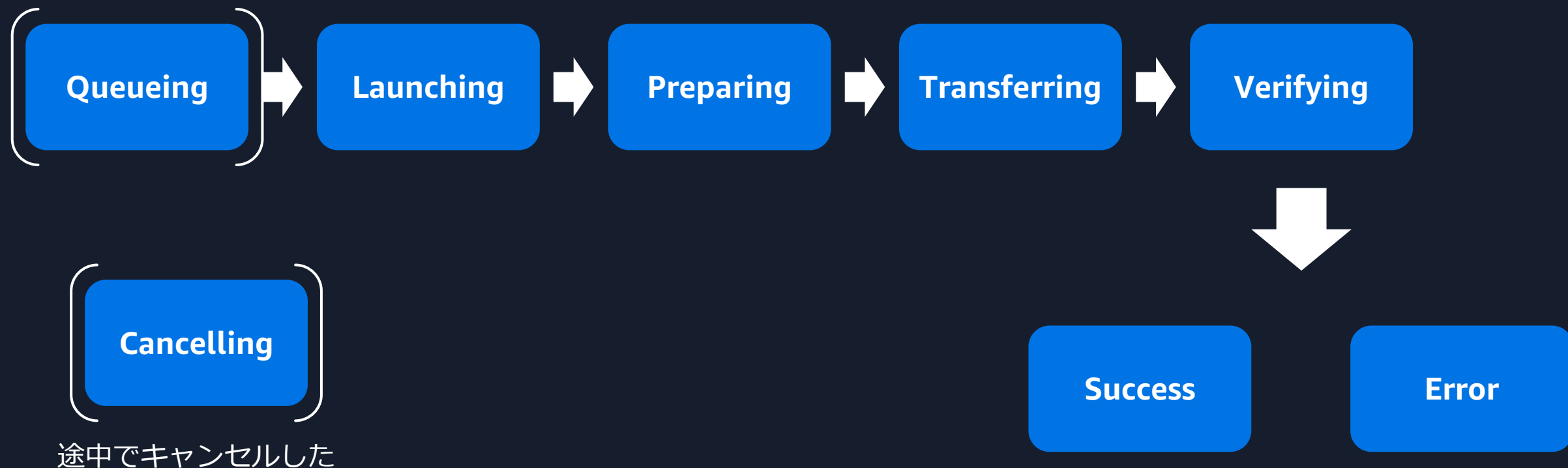
- ファイル数などの概要レポート
- ファイルのリストなどの詳細レポート

Amazon CloudWatch Logs へ配信する
ログレベルを設定できる

- エラーなどの基本情報
- 転送した全てのファイル

タスクを実行した際のステータス

タスクを実行した際、Agent が別のタスクを実行中の場合にはタスクのステータスが **Queueing** となる
Launching ではタスクを開始しており、その後 **Preparing** の段階で転送対象のデータを決定する
AWS DataSync は **Transferring** ではデータを転送し、データ転送完了後に **Verifying** で整合性を検証する



AWS DataSync を用いたデータ転送の流れ

(例) NFS ストレージから Amazon S3 へのデータ移行

想定ケース

Amazon EC2 上に Agent をデプロイし、NFS サーバから Amazon S3 へと AWS DataSync を通じてデータを転送する。なお、今回は NFS サーバとして Amazon EFS を利用する※



NFS サーバの起点となるディレクトリ直下に、次のようなメタデータを持つファイルとディレクトリを作成する

```
drwxr-xr-x. 2 0 0 6144 Aug 23 07:30 dir1
-rw-r--r--. 1 0 0 1024 Aug 23 07:30 test.dat
```

POSIX permission

User ID と group ID

ファイルの更新日時

※ Amazon EFS は Agent を用いずにロケーションの設定が可能であり、ロケーション作成時に「Amazon EFS」が候補として表示される。しかし、ここでは任意の NFS ストレージを模擬しているため、「NFS」としてロケーションを登録する。

Agent のインストール/アクティベート (1)

Amazon EC2 上へと Agent をインストールする

(1) 以下のようなコマンドを発行し、AWS DataSync Agent が含まれる AMI ID を取得する

```
aws ssm get-parameter --name /aws/service/datasync/ami --region リージョン名
```

(2) 取得した AMI を用いて、m5.2xlarge タイプのインスタンスを起動する
この際、HTTP 80 のポートおよび NFS サーバとの通信に必要なポートを開放する
今回はパブリックエンドポイント経由で AWS DataSync と通信するため、Public IP アドレスを付与する

注意: 2,000 万ファイルを超える場合には m5.4xlarge を用意する

Agent のインストール/アクティベート (2)

AWS DataSync のコンソールにおいて、エージェントを作成する

エージェントを作成する

エージェントをデプロイする

セルフマネージドストレージと AWS ストレージサービスの間でデータを転送するには、エージェント VM をデプロイして開始していることを確認してください。エージェントは、オンプレミスのハイパーバイザーまたは Amazon EC2 のいずれかで実行できます。

ハイパーバイザー

Amazon EC2

ユーザーガイドに記載されている AMI を使用して EC2 インスタンスを作成します。 [詳細はこちら](#)

サービスエンドポイント

エンドポイントタイプ

このエージェントが接続するサービスエンドポイントを選択します。 [詳細はこちら](#)

US West (Oregon) の公開サービスエンドポイント

アクティベーションキー

アクティベーションは、デプロイしたエージェントを AWS アカウントに安全に関連付けます。 [詳細はこちら](#)

エージェントのアクティベーションキーを取得する方法を選択します。 [詳細はこちら](#)

- ☒ エージェントからアクティベーションキーを自動的に取得する
ブラウザは HTTP 経由でエージェントに接続し、一意のアクティベーションキーを取得します。

- ☐ エージェントのアクティベーションキーを手動で入力する
アクティベーションキーが既にあり、手動で入力する必要がある場合は、このオプションを選択します。

エージェントのアドレス

DataSync エージェントがデプロイされたら、そのドメイン名または IP アドレスを入力します。キーを取得する をクリックすると、ブラウザはこのアドレスに接続して、一意のアクティベーションキーを取得します。

http://

このアドレスにブラウザからアクセスできることを確認してください。

キャンセル

キーを取得する

Agent のインストール/アクティベート (3)

アクティベーションキーが自動的に取得でき、エージェントの作成に成功する


サービスエンドポイント

エンドポイントタイプ
このエージェントが接続するサービスエンドポイントを選択します。 [詳細はこちら](#)

US West (Oregon) の公開サービスエンドポイント

アクティベーションキー

アクティベーションは、デプロイしたエージェントを AWS アカウントに安全に関連付けます。 [詳細はこちら](#)



エージェントからアクティベーションキーを正常に取得しました。

アクティベーションキー


エージェントのアドレス

[DataSync](#) > [エージェント](#) > bb-agent

bb-agent

編集 削除

詳細

サービスエンドポイント	エージェントのステータス	Creation time
公開 US West (Oregon)	オンライン	2024年8月23日 16:49:37 JST
Agent ARN		
 [Redacted]		

タスクの設定

送信元ロケーションを Amazon EFS ではなく、今回は NFS とする
メタデータが送信先ロケーションである Amazon S3 へとコピーされるように設定する

ロケーションタイプ

ネットワークファイルシステム (NFS) ▼

リージョン

US West (Oregon) us-west-2

Transfers to or from NFS must stay in the current Region.

エージェント

選択したエージェントは NFS サーバーにアクセスできる必要があります。

1 つ以上のエージェントを選択する ▼

bb-agent (agent-02dd3b9fc7c43002b) 公開 - US West (Oregon) ✕

NFS サーバー

ドメイン名または IP アドレス。

マウントパス

NFS サーバーによってエクスポートされたマウントパス、またはエクスポートされたパスのサブディレクトリ。

/

▶ 追加の設定

Transfer options

Transfer mode

Determines whether DataSync transfers only the data and metadata that differ between the source and the destination.

Transfer only data that has changed ▼

Verification

Verify the data in your destination after transfer. DataSync always performs data integrity checks while transferring and writing data.

Verify only the data transferred ▼

Bandwidth limit

Allocate the maximum network bandwidth to be utilized by this task.

☒ Use available

☒ Keep deleted files

Keep files in destination even when deleted from source

☒ Overwrite files

Overwrite files in destination with files from source

▼ 追加の設定

☒ Copy ownership

Maintain user and group ID

☒ Copy permissions

Maintain existing permissions

☒ Copy timestamps

Maintain access time and modification time

☒ Queueing

Allow this task to be started and queued until a previous execution completes.

タスクの開始と完了

タスクを実行する。必要に応じて、履歴からログやステータスを確認する

bb-task

開始 ▼

停止

編集

削除

概要

タスクのステータス

最新の実行

Task ARN

🟢 利用可能

-

📄

bb-task

開始 ▼

停止

編集

削除

概要

タスクのステータス

最新の実行

作成日

🟢 利用可能

2024年8月23日 17:05:14 JST

2024年8月23日 17:04:49 JST

Task ARN

📄

ロケーション

オプション

スケジュール

タグ

Monitoring

履歴

履歴 (1)

編集

停止

フィルター条件 ▼

履歴には、各タスクの実行が表示され、実行結果に関する情報が含まれます。

🔍

タスク実行を検索する

< 1 >

実行 ID	開始時刻 ▼	期間 ▼	ステータス ▼	MiB/秒 ▼	ファイル/秒 ▼
○ exec-0406aa30dab5bdf87	2024年8月23日 17:05:14 JST	4 秒	🟢 成功	0	1.89

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

48

タスクの完了

Amazon S3 上に NFS サーバ上のファイルとディレクトリがコピーされていることが確認できる
また、メタデータも意図した通りに保存される

<input type="checkbox"/>	名前 ▲	タイプ ▼	最終更新日時 ▼	サイズ ▼	ストレージクラス ▼
<input type="checkbox"/>	 .aws-datasync-metadata	aws-datasync-metadata	2024/08/23 05:08:16 PM JST	0 B	スタンダード
<input type="checkbox"/>	 dir1/	フォルダ	-	-	-
<input type="checkbox"/>	 test.dat	dat	2024/08/23 05:08:16 PM JST	1.0 KB	スタンダード

メタデータ (7)

メタデータは、名前-値 (キーと値) のペアとして提供されるオプションの情報です。 [詳細](#)

タイプ	キー	値
システム定義	Content-Type	application/octet-stream
ユーザー定義	x-amz-meta-user-agent	aws-datasync/3.8.2831.0-b4a75425
ユーザー定義	x-amz-meta-file-atime	1724398224584000000ns
ユーザー定義	x-amz-meta-file-owner	0
ユーザー定義	x-amz-meta-file-permissions	100644
ユーザー定義	x-amz-meta-file-group	0
ユーザー定義	x-amz-meta-file-mtime	1724398224584000000ns

- ファイルのアクセス日時
- User ID
- POSIX permission
- group ID
- ファイルの更新日時

他のデータ転送ソリューションとの比較

オンプレミスから AWS へのデータ移行方法

オンライン転送とオフライン転送に別れる

オンライン転送では、AWS マネージドサービス・ツール・スクリプトでの実装という選択肢がある

オンライン転送



スクリプト
robocopy
aws s3
copy



AWS DataSync



Amazon FSx for
NetApp ONTAP
SnapMirror

データ移行を
早く開始したい

安定したネットワーク
帯域が確保できる

Amazon FSx シリーズや
Amazon EFS へ移行したい

移行時のボトルネックは
オンプレミスと AWS 間の
ネットワーク帯域である

移行中にデータを
オフラインにできる

データ移行のスケジュールに
余裕がある

オフライン転送



AWS Snowball Edge
Storage Optimized
210 TB

AWS Snowball Edge を利用する際は
メタデータの取り扱いに注意する

オンライン転送の選択肢

AWS DataSync はエラーハンドリング、データの整合性検証、スケジューリングなどの便利な機能をマネージドサービスとして提供するため、利用することで**楽にデータを AWS へ送信できることがポイント**である

ただし、オンライン転送の手段は、移行元と移行先ストレージ、データ転送の要件に応じて選択する

例 1: NFS サーバから Amazon EFS へと楽にデータを同期したい

AWS DataSync ではスケジュールを設定でき、データの同期方法や整合性確認も機能として提供する

例 2: Amazon FSx for NetApp ONTAP へとデータを送信したい

転送元ストレージが ONTAP である場合、スナップショット単位でブロック転送が可能な SnapMirror の活用が適するケースが多い。AWS DataSync や robocopy はファイル単位での送信となる

例 3: Amazon S3 へとデータを数分ごとの間隔で定期的に送信したい

aws s3 コマンドを用いたスクリプトベースのデータ転送基盤を構築する

※ AWS DataSync のタスクをスケジュール以外の仕組みで実行すれば、実現できる

オンライン転送の選択肢

オンライン転送の手段は、移行元と移行先ストレージ、データ転送の要件に応じて選択する

例えば、移行元と移行先ストレージが同じでも、要件が変われば当然適したツールは変わる

例 4: NFS サーバから Amazon S3 へとデータを転送したい。ただし、メタデータを保存したい

AWS DataSync を用いることで、NFS サーバ上のファイルやディレクトリのメタデータが Amazon S3 上でも保存される

一方で、aws s3 コマンドを用いてメタデータを保存することは困難である

参考

送信元	送信先	コピーされるメタデータ
<ul style="list-style-type: none">NFS	<ul style="list-style-type: none">Amazon S3	<ul style="list-style-type: none">ファイルとディレクトリの更新日時/アクセス日時User ID と group ID、POSIX permission

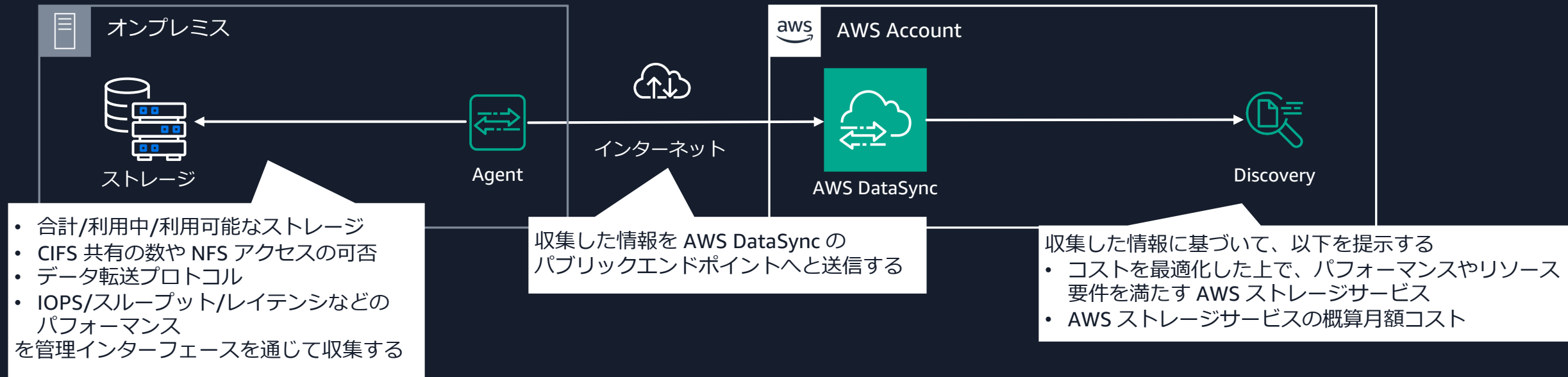
補足と注意点

補足と注意点一覧

1. ストレージの検出
2. Agent と送信元ストレージの組み合わせパターン
3. 複数の Agent を用いることのメリット
4. Agent はどこに配置すべきか
5. 開いたファイルとロックされたファイルの扱い
6. リンクとディレクトリの扱い
7. SMB サーバからデータを転送する際の注意

ストレージの検出（オプション）

AWS DataSync によるストレージの検出を利用し、移行に関する推奨情報を取得できる Agent はオンプレミスストレージの情報を収集し、AWS DataSync へと送信する

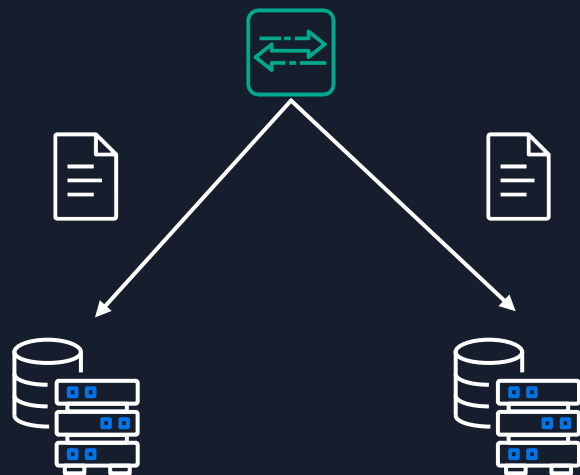


- 本機能を利用する際、AWS DataSync への通信にはパブリックエンドポイントが利用される
- 正確な推奨情報の出力には、14 日間以上の情報収集を推奨する
- Amazon FSx for NetApp ONTAP の Single AZ 構成、Amazon EFS の One Zone ストレージクラス、Amazon FSx for Windows File Server の Single AZ 構成はストレージの検出では考慮しない
- Amazon VPC 内に Agent を配置し、Amazon VPC 内部のリソースに対するストレージの検出もできる

Agent と送信元ストレージの組み合わせパターン

1 vs N

単一の Agent に対して、
複数のストレージを設定する



複数のタスクを設定する
タスクは必要に応じ、
キューに入れられる

1 vs 1

単一の Agent に対して、
単一のストレージを設定する



N vs 1

複数の Agent に対して
単一のストレージを設定する



大規模データの移行が
高速に実現できる

複数の Agent を用いることのメリット

Agent ごとに異なるタスクを設定し並列でデータを転送することで、大規模データの送信を高速化できる

タスクごとに送信元ストレージのマウントパスを分けたり、マニフェストやフィルタを利用することで、データセットのタスクごとの区分を明確化できる

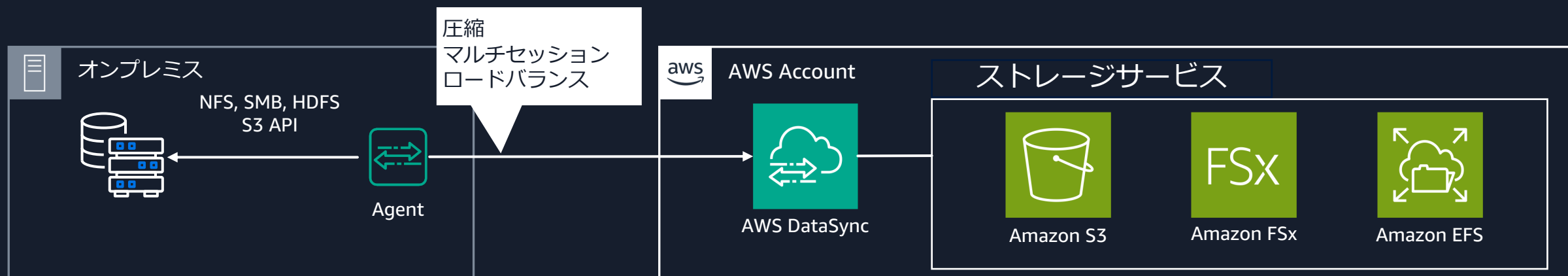


注意点

- 送信元ストレージやネットワークへの負荷が発生する
- Agent はそれぞれ別のタスクを設定するため、可用性を高くすることとは関係がない
- ストレージの検出を行う場合には、1つのストレージに対して Agent は1つのみ設定できる

Agent はどこに配置すべきか

オンプレミスからのデータ移行の場合、送信元ロケーションと近い場所へ配置する
データ送信時に AWS DataSync が実装する圧縮・マルチセッション・ロードバランスの効果範囲が広がる



仮に Amazon EC2 上に Agent を配置すると、オンプレミスの送信元ストレージと Agent は NFS や SMB などの通信となり、AWS DataSync が提供する圧縮・マルチセッション・ロードバランスの恩恵を受けられない

開いたファイルとロックされたファイルの扱い

開いているファイルとロックされたファイルを AWS DataSync の転送対象とするとき注意する

開いたファイル

通常、AWS DataSync は開いたファイルを送信できる
ただし、転送中にデータが送信元ストレージで編集された場合、データの整合性検証時に不一致と判断する
最新のバージョンを送信するためには、タスクを再度実行する

ロックされたファイル

AWS DataSync はファイルのロックまたはロック解除はできない
AWS DataSync がロックされたファイルを開けない場合、そのファイルの転送をスキップし、エラーを記録する

ハードリンクの扱い

ハードリンク、シンボリックリンク、ディレクトリの扱いは送信元と送信先の組み合わせに依存する
一例のみ紹介し、すべてのパターンは[こちら](#)を参照する

ハードリンクの扱いの例

送信元	送信先	ハードリンクの扱い
<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	ハードリンクは保存される
<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	ハードリンクはサポートしない。データ転送は完了するものの、ログにエラーが記録される
<ul style="list-style-type: none">任意のストレージ	<ul style="list-style-type: none">Amazon S3	ハードリンクとその参照先は別のオブジェクトとして保存される。 Amazon S3 上でハードリンクを変更しなければ※、AWS DataSync を用いて NFS、Amazon EFS、Amazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS へ復元した時、ハードリンクも復元できる

※ 例えば、Amazon S3 から参照先ファイルをダウンロードし、再度アップロードした場合、復元時にハードリンクは復元できない。参照先とは異なる inode 番号を示す



ハードリンクを NFS から Amazon S3 へコピー

Amazon EFS 上の source ディレクトリ上に hello.text とそのハードリンクを作成する



```
5930988602823351302 -rwxrwxrwx. 2 root root 6 Aug 24 10:30 hello.lnk  
5930988602823351302 -rwxrwxrwx. 2 root root 6 Aug 24 10:30 hello.text
```

同じ inode 番号

AWS DataSync でこれらのファイルを Amazon S3 へ送信すると、ハードリンクと参照先が作成される

<input type="checkbox"/>	 hello.lnk	lnk	2024/08/24 07:34:20 PM JST	6.0 B	スタンダード
<input type="checkbox"/>	 hello.text	text	2024/08/24 07:34:20 PM JST	6.0 B	スタンダード

Amazon EFS 上の hello.text を変更し再度タスクを実行すると、共に更新される

<input type="checkbox"/>	 hello.lnk	lnk	2024/08/24 07:39:22 PM JST	6.0 B	スタンダード
<input type="checkbox"/>	 hello.text	text	2024/08/24 07:39:22 PM JST	6.0 B	スタンダード

Amazon EFS 上の target ディレクトリ上へ、AWS DataSync で復元するとハードリンクは保持されている

```
9934453109177352605 -rwxrwxrwx. 2 root root 6 Aug 24 10:35 hello.lnk  
9934453109177352605 -rwxrwxrwx. 2 root root 6 Aug 24 10:35 hello.text
```

同じ inode 番号

aws s3 コマンドで復元すると、ハードリンクは保持されないので注意する

シンボリックリンクとディレクトリの扱い

ハードリンク、シンボリックリンク、ディレクトリの扱いは送信元と送信先の組み合わせに依存する
一例のみ紹介し、すべてのパターンは[こちら](#)を参照する

シンボリックリンクの扱いの例

送信元	送信先	シンボリックリンクの扱い
<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	<ul style="list-style-type: none">NFSAmazon EFSAmazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS	シンボリックリンクは保存される
<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	<ul style="list-style-type: none">SMBAmazon FSx for NetApp ONTAP (SMB)/ Windows File Server	シンボリックリンクはサポートしない 転送は完了するが、ログにエラーが記録される
<ul style="list-style-type: none">任意のストレージ	<ul style="list-style-type: none">Amazon S3	ターゲットパスは保存される NFS、Amazon EFS、Amazon FSx for NetApp ONTAP (NFS)/ Lustre/OpenZFS へ復元した時、リンクも復元できる

ディレクトリの扱いの例

送信元	送信先	ディレクトリの扱い
<ul style="list-style-type: none">任意のストレージ	<ul style="list-style-type: none">Amazon S3	<p>/ がパスの末尾となる空のオブジェクトとして、登録される test という空のディレクトリを送信した場合の挙動の例:</p> <div><div><input type="checkbox"/>  test/</div><div>フォルダ</div></div>

SMB サーバからデータを転送する際の注意


SMB サーバをマウントし、ファイル・フォルダ・メタデータを読み取れる権限が必要である
特に Microsoft Active Directory を利用している場合、次の権限が必要である

ユーザー権限	説明
Restore files and directories (SE_RESTORE_NAME)	オブジェクト所有者、権限、ファイルメタデータ、NTFS DACLs を AWS DataSync がコピーする際に必要である 通常は、Domain Admins または Backup Operators グループに所属している
Manage auditing and security log (SE_SECURITY_NAME)	AWS DataSync が NTFS SACLs をコピーする際に利用する 通常は Domain Admins グループに所属している

DFS 名前空間はサポートしない

AWS DataSync の SMB サーバへのアクセスには、Kerberos 認証は使えない。代わりに NTLM 認証を用いる

ファイル共有名が日本語の場合、ロケーションの作成ができない（2024 年 8 月 23 日現在）



ソースロケーションの作成に失敗しました

リクエスト ID

9ce4b016-80ea-45e2-bc2a-464721620d58

アクション

datasync:CreateLocationFsxWindows

ステータスコード

400

API 応答

1 validation error detected: Value '/日本語共有' at 'subdirectory' failed to satisfy constraint: Member must satisfy regular expression pattern: ^[a-zA-Z0-9_-\.\(\)\\$\p{Zs}]+\$

まとめ

- AWS DataSync を用いることで、AWS ストレージサービスとオンプレミス、他クラウドサービス間でデータの移動を実現できる
- データの整合性の検証・差分同期・スケジューリング・帯域制御・エラーハンドリング・圧縮などのデータ転送時に便利な機能を、AWS DataSync はマネージドサービスとして提供する
- データを送信する際には、送信元と送信先ストレージの組み合わせや要件に応じて適切な手段を選択する

Thank you!