Amazon Simple Storage Service (Amazon S3)

データ保護編

佐藤 真也

Amazon Web Service Japan G.K.
Solutions Architect
2023/04

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
- 動画を一時停止・スキップすることで、興味がある分野・項目だけの聴講も可能、スキマ時間の学習にもお役立ていただけます
- 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/

内容についての注意点

- 本資料では 2023 年 4 月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。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

自己紹介

名前：佐藤 真也 (Sato Shinya)

所属：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 金融ソリューション本部
保険ソリューション部

好きなAWSサービス：

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

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フルマネージドサービスの概念
- AWS IAM、Amazon VPC などの基盤となるサービスの知識
- Amazon S3 入門編あるいは同等の知識※

対象者

- Amazon S3 でどのようにデータを保護するか気になる方

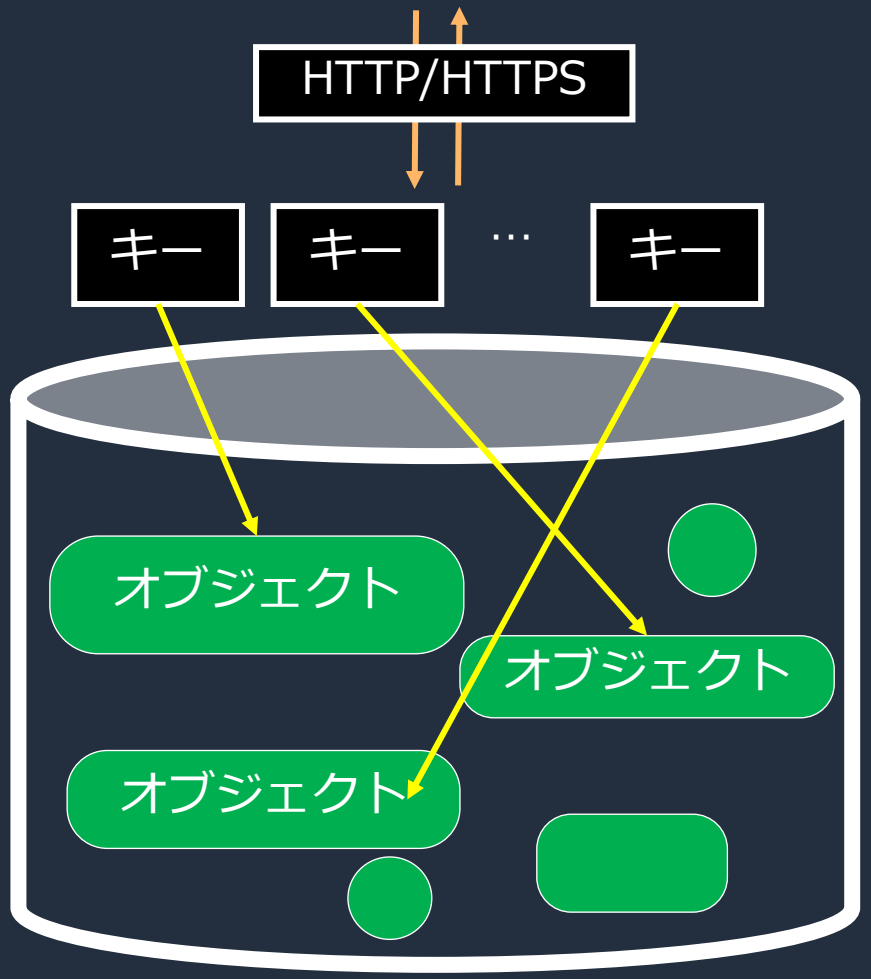
※参考リンク:
Amazon S3 入門編: https://www.youtube.com/watch?v=wQ8ZDvoMSno
Amazon S3 セキュリティ編: https://www.youtube.com/watch?v=VutHE2vSvFo&t=1s
Amazon S3 ユースケース編: https://www.youtube.com/watch?v=uuK-VaQLrzg

アジェンダ

1. Amazon S3 の概要
2. オブジェクトのバージョニングとロック機能
3. AWS Backup の利用
4. レプリケーションによるデータ保護
5. データの整合性の検証
6. まとめ

Amazon S3 の概要

オブジェクトストレージとは



特徴

- HTTP/HTTPS でアクセス
- 一意のキーに対するオブジェクト（データ）が存在
- 階層構造を取るファイルストレージとは異なり、フラットな構造

メリット

- スケールが容易で、大容量のデータ保存が可能
- オブジェクト単位でのアクセス制御
- 高い可用性と耐障害性
- 独自にカスタマイズできるメタデータを追加可能

AWS のストレージサービス

OBJECT

Amazon S3

BLOCK

Amazon EBS

FILE

Amazon EFS

Amazon FSx for NetApp ONTAP

Amazon FSx for Windows File Server

Amazon FSx for Lustre

Amazon FSx for OpenZFS

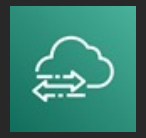
Amazon File Cache

BACKUP

AWS Backup

DATA TRANSFER AND MIGRATION

AWS Storage Gateway



AWS DataSync

AWS Transfer Family

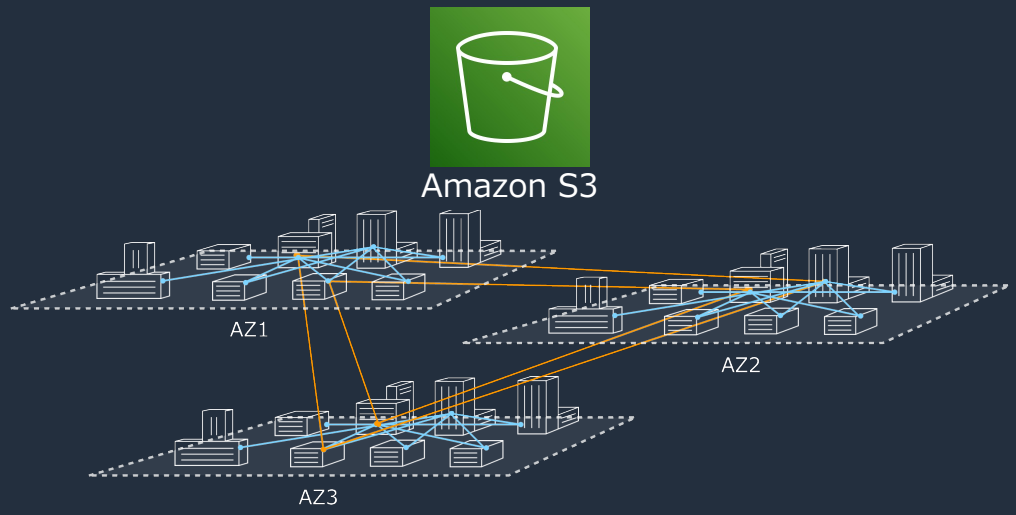
AWS Snowball

AWS Snowcone

Amazon S3 とは

高いパフォーマンスと可用性、そして低コストが特徴なオブジェクトストレージ
2006 年に登場してから、現在に至るまでのイノベーションが積み重なった歴史あるサービス

- 耐久性
  - 99.999999999%（イレブンナイン）
  - 最低 3 つのアベイラビリティゾーン（AZ）で冗長化
- スケーラビリティ
  - 無制限のデータ保存
  - ただし、1 オブジェクトは最大 5 TB
- 低コスト
- セキュリティ
  - アクセス制御とログ監査
- データの保護
  - 誤削除から守る機能
- アクセシビリティ
  - HTTP/HTTPS でアップロード/ダウンロード/変更/削除といった操作が可能
- 様々な AWS サービスとの連携



Amazon S3 の特徴などは FAQ にて詳解: https://aws.amazon.com/jp/s3/faqs/

Amazon S3 の 2018 年までの主要アップデート

2006 年
- ・バケットポリシー
- ・バージョニング
- ・マルチパートアップロード

をサポート

S3 登場

2010 年

S3 Glacier
をサポート

2011 年
- ・東京リージョン

で利用可能に
- ・SSE-S3
- ・ライフサイクル機能

をサポート

2012 年

2014 年
- ・SSE-C
- ・SSE-KMS

をサポート

2015 年
- ・クロスリージョン
  レプリケーション
- ・VPC エンドポイント
- ・S3 標準 IA

をサポート

2016 年
- ・S3 Transfer Acceleration
- ・オブジェクトのタグ付け
- ・S3 インベントリ
- ・ストレージクラス分析

をサポート

2017 年

デフォルト暗号化
をサポート

2018 年
- ・S3 Select
- ・S3 Intelligent-Tiering
- ・オブジェクトロック
- ・ブロックパブリックアクセス

をサポート

Amazon S3 の 2019 年以降の主要アップデート

2019 年
- バッチオペレーション
- S3 Glacier Deep Archive
- アクセスアナライザー

をサポート

2020 年上期

バッチオペレーションによる
- オブジェクトロック
- タグ付

をサポート

2020 年下期
- Object Ownership
- S3 Storage Lens
- 強力な整合性
- Outposts 上での S3

をサポート

2021 年上期
- Private Link
- S3 Object Lambda

をサポート

2021 年下期
- マルチリージョンアクセスポイント
- S3 Intelligent-Tiering における Archive Instant Access
- S3 Glacier Instant Retrieval

をサポート

2022 年上期
- AWS Backup for S3
- 新たなチェックサムアルゴリズム

をサポート

2022 年下期
- マルチリージョンアクセスポイントのアクティブ/パッシブ構成

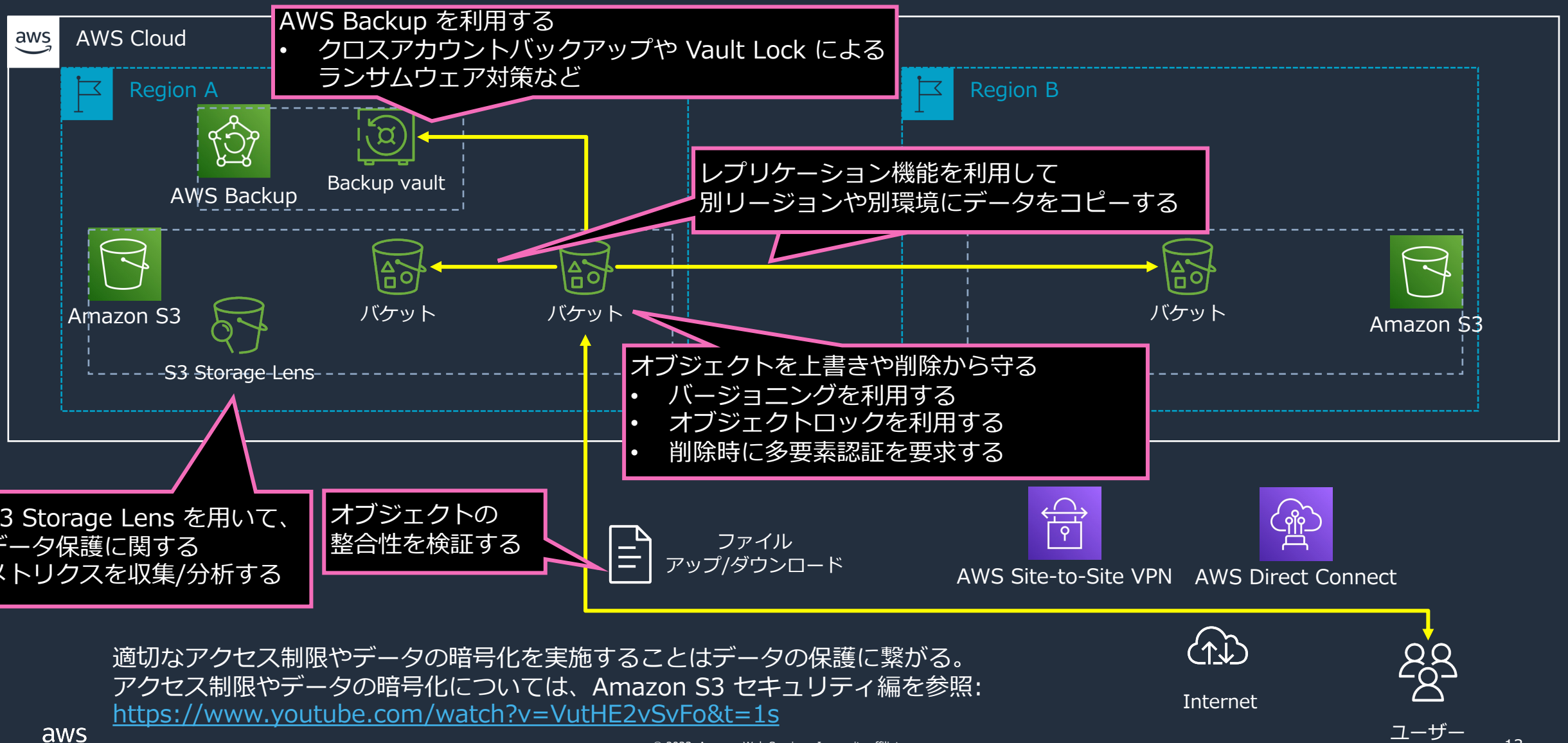
をサポート
- S3 Glacier からの取り出し時のスループットが最大 10 倍

と高速化

2023 年
- デフォルト暗号化が自動的に有効
- マルチリージョンアクセスポイントのクロスアカウントアクセス

をサポート

S3 におけるデータ保護のポイント



適切なアクセス制限やデータの暗号化を実施することはデータの保護に繋がる。
アクセス制限やデータの暗号化については、Amazon S3 セキュリティ編を参照:
https://www.youtube.com/watch?v=VutHE2vSvFo&t=1s

オブジェクトのバージョニングとロック機能

バージョニングとは

同じバケット内部でオブジェクトの複数のバージョンを保持する方法。バケットレベルで設定し、設定以降に作成/上書きされたオブジェクトはバージョン ID が付与される※

メリット

- オブジェクトを削除/上書きした場合にも、復元ができる
  - 削除処理を行なった場合、オブジェクトが削除されるのではなく、代わりに削除マーカーが挿入される。削除マーカーが最新のオブジェクトのバージョンとなる
  - 上書きされると、上書きしたものが最新のオブジェクトのバージョン

注意

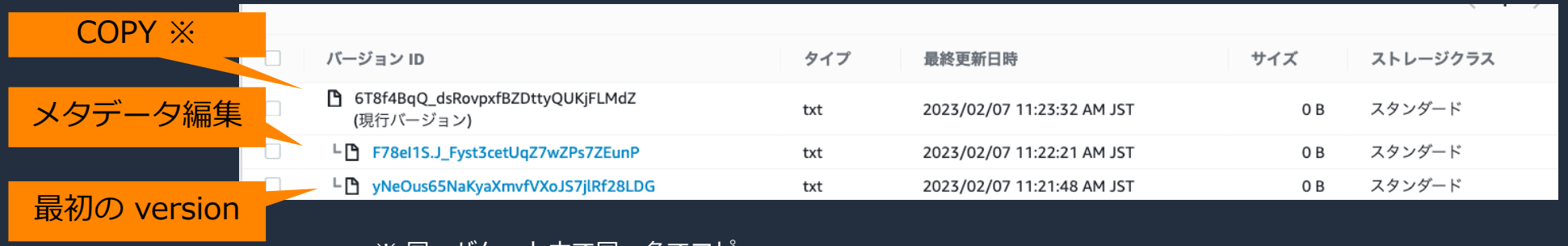
- オブジェクトのバージョン分が課金の対象となる
  - 3 つのバージョンを保持していた場合、3 つのオブジェクトに対して課金される。
  - バージョン 1/2/3 がそれぞれ 10/20/30 KB の場合、合計 60 KB 課金される。
    これは各バージョンは以前のバージョンとの差分ではなく、完全なオブジェクトであるため

※バージョニング設定前に存在したオブジェクトのバージョン ID は null

バージョニングの仕組み

バージョニングを有効にすると、一意のバージョン ID を自動的に生成する

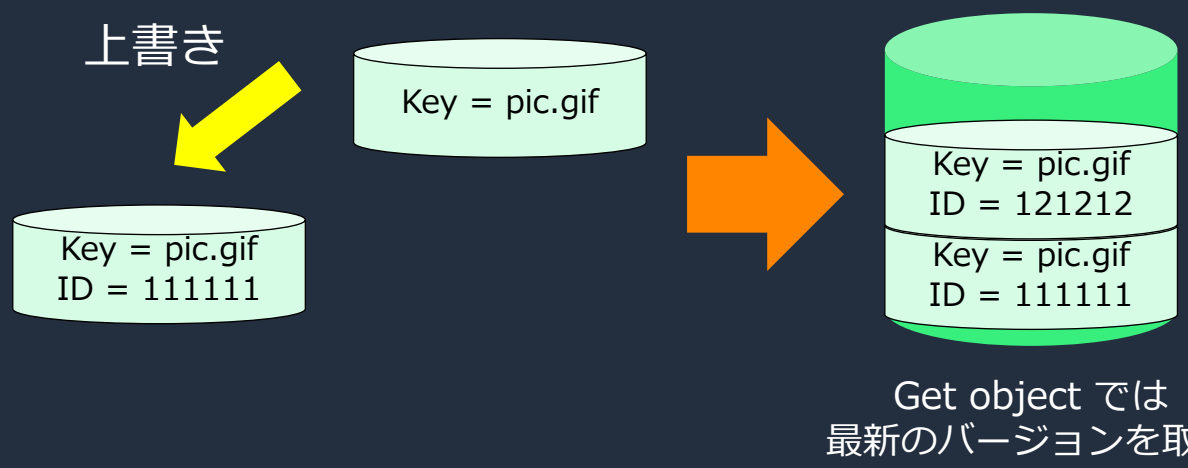
バージョニングが有効でない場合、バージョン ID は存在するが値は null となる
バージョン ID の生成は S3 のみができ、編集はできない
COPY コマンドやメタデータの編集でも新しいバージョンが作成される



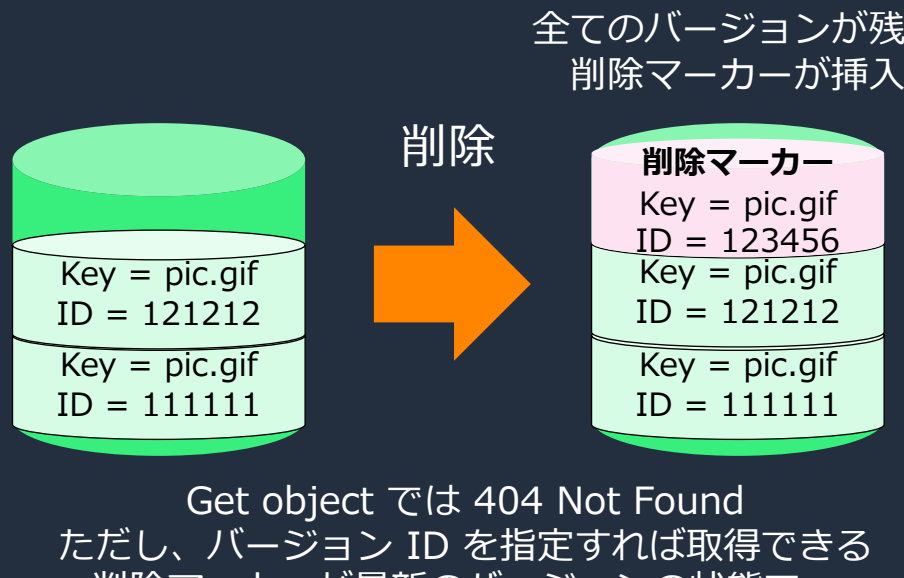
※ 同一バケット内で同一名でコピー

```
[redacted] % aws s3 cp s3://shinya-sato-bb/dummy.txt s3://shinya-sato-bb/
copy: s3://shinya-sato-bb/dummy.txt to s3://shinya-sato-bb/dummy.txt
```

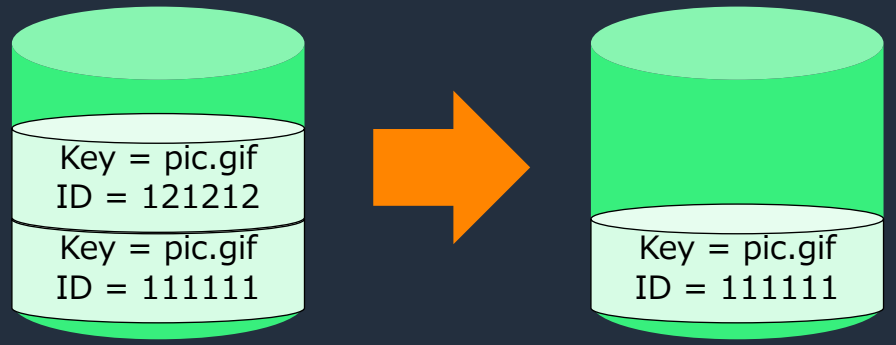
バージョニングのワークフロー



Get object では
最新のバージョンを取得



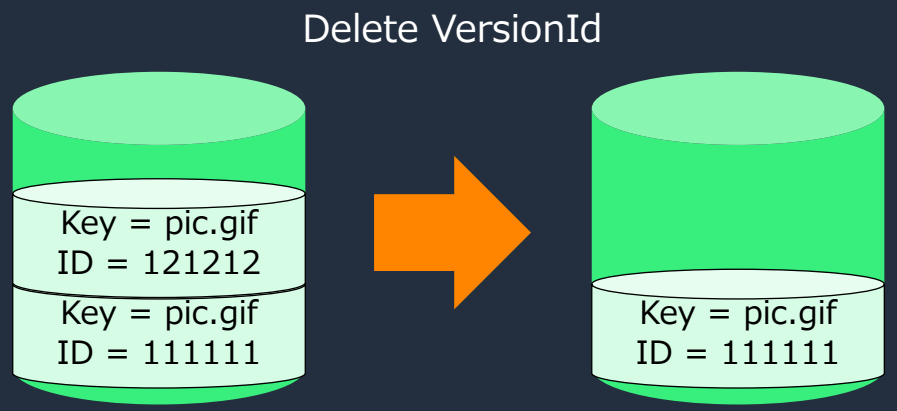
Get object では 404 Not Found
ただし、バージョン ID を指定すれば取得できる
削除マーカーが最新のバージョンの状態で、
削除処理を行うと、別の削除マーカーがさらに追加される



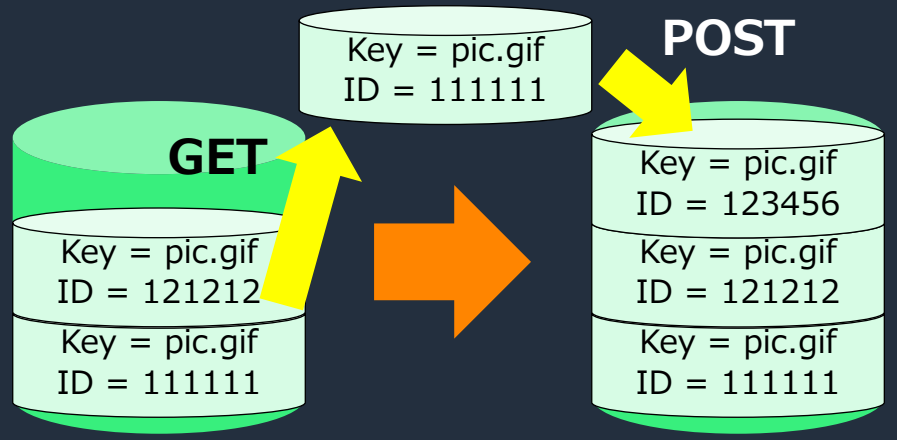
- オブジェクトの完全な削除には、削除するバージョンを指定する
  - null の場合、バージョン ID として null を渡す必要がある
- 削除マーカーの挿入がなく、永久に削除できる
- S3 バケットの所有者のみが特定のバージョンを永久に削除できる

以前のバージョンの復元方法

- オブジェクトの新しいバージョンを完全に削除する
  - 以前のバージョンが最新のバージョンになる

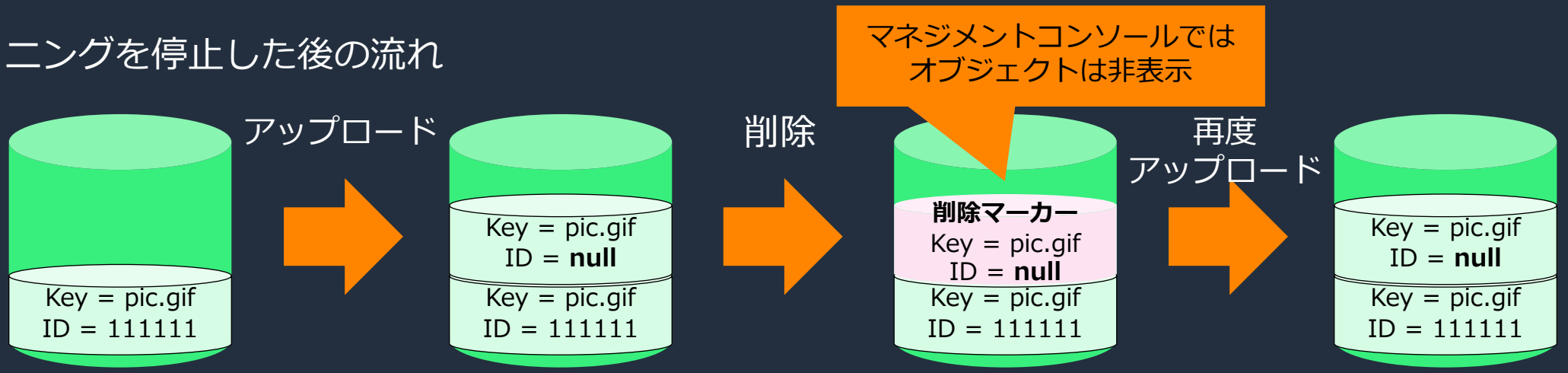


- 以前のバージョンを取得し、同じバケットにアップロードする
  - コピーされたオブジェクトが最新となり、全てのオブジェクトバージョンが保持される



バージョニングを停止した時の挙動

バージョニングを停止した後の流れ



バージョン ID は null として管理される。バージョニングを停止した後にアップロード/削除/再度アップロードしたオブジェクトは同一の null ID として管理される

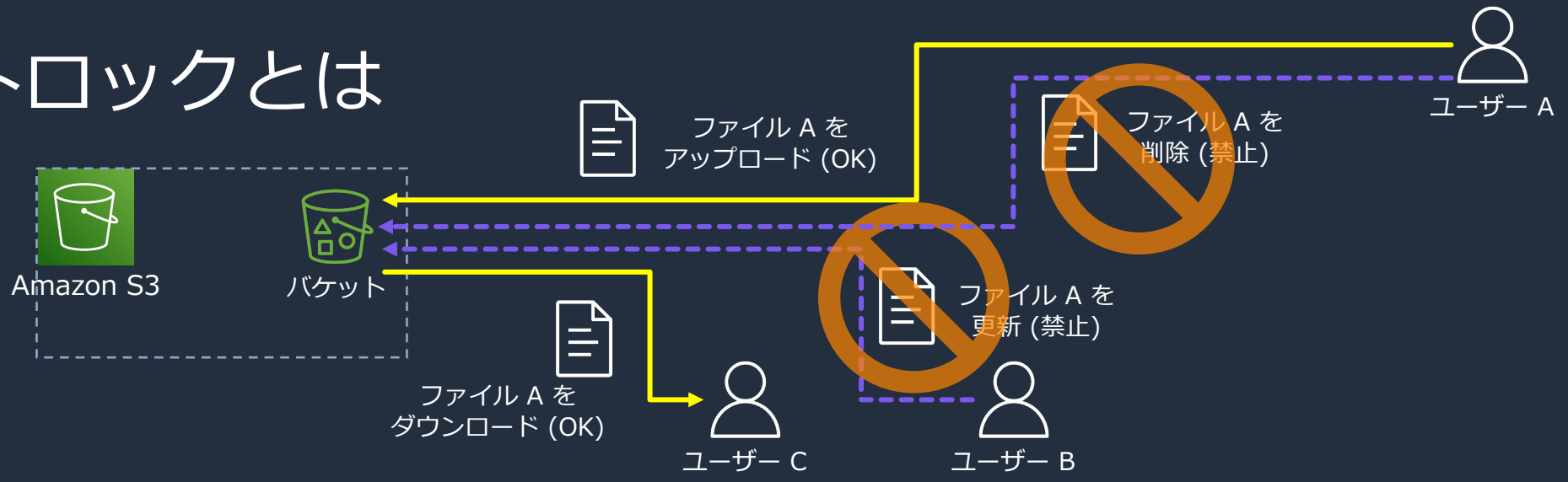
アップロード

	バージョン ID	タイプ
	null (現行バージョン)	txt
	6T8f4BqQ_dsRovpxfBZDttyQUKjFLMdZ	txt
	F78el1S.J_Fyst3cetUqZ7wZPs7ZEunP	txt
	yNeOus65NaKyaXmvfVXoJS7jlRf28LDG	txt

削除し、再度アップロード

	バージョン ID	タイプ
	null (現行バージョン)	txt
	6T8f4BqQ_dsRovpxfBZDttyQUKjFLMdZ	txt
	F78el1S.J_Fyst3cetUqZ7wZPs7ZEunP	txt
	yNeOus65NaKyaXmvfVXoJS7jlRf28LDG	txt

オブジェクトロックとは



オブジェクトロックはバージョニングされたバケットのみに適用できる Write Once Read Many（WORM）機能で、削除/上書きを一定期間/無期限に防止できる※

※変更/削除処理を反映するよう見えるが、実際にはバージョニングにより保存されている

2 つのリテンションモードがある

- ガバナンスモード
  - s3:BypassGovernanceRetention の権限を持ち、かつ、x-amz-bypass-governance-retention:true のヘッダーを含む場合を除き、上書き/削除/設定の変更ができない
  - 特定権限があれば、削除/設定変更ができるため、コンプライアンスモードのテストとしても利用できる
- コンプライアンスモード
  - いかなるユーザーも上書き/削除/設定の変更を行うことができない

保持期間とリーガルホールド

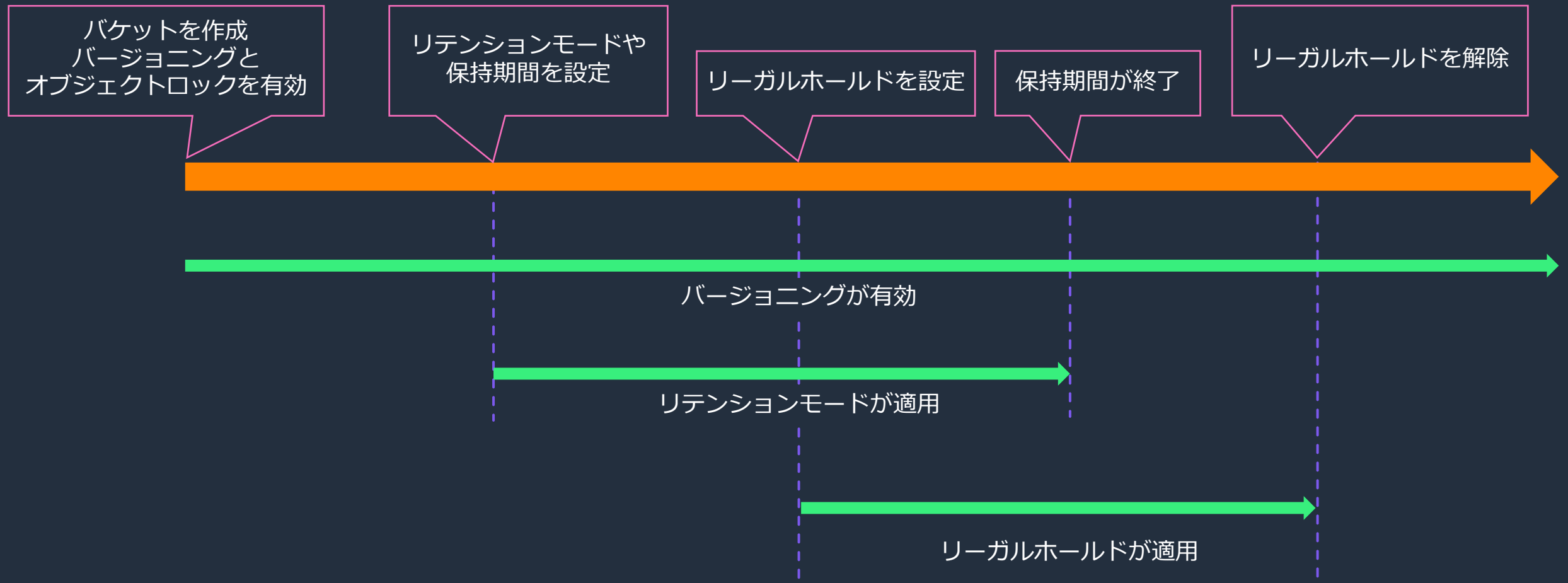
保持期間
- 個々のオブジェクトバージョンに対して適用される
- この期間、ガバナンス/コンプライアンスモードが適用され、オブジェクトを上書き/削除できない
- 保持期間の終了後、リーガルホールドを適用しない限り、オブジェクトを上書き/削除できる

リーガルホールド
- オブジェクトの上書きと/削除を防止する
- 保持期間とは独立に設定され有効期間はない。s3:PutObjectLegalHold 許可を持つ任意のユーザーが自由に適用/解除できる
- オブジェクトロックが有効になっているバケットに対して設定ができるが、リテンションモードや保持期間とは独立に設定される
- 保持期間中はリテンションモードが適用され、保持期間後はリーガルホールドが適用される

保持期間とリーガルホールドの関係

	リーガルホールドを適用する	リーガルホールドを適用しない
保持期間終了前	• ガバナンス/コンプライアンスモードが適用される • s3:PutObjectLegalHold 許可を持つ任意のユーザーは自由にリーガルホールドを設定できる	
保持期間終了後	s3:PutObjectLegalHold 許可を持つ任意のユーザーが自由に解除するまで WORM 機能が引き続き適用される	自由にオブジェクトを操作できる

有効になるデータ保護機能



オブジェクトロックでの挙動: 削除処理

オブジェクトロックを有効にした状態でテスト

オブジェクトをアップロード

	バージョン ID	タイプ	最終更新日時
	NTDMXt21RE3vxRd96UX9FQsVWat4PaNj (現行バージョン)	txt	2023/02/07 11:38:49 AM JST

オブジェクトを削除すると、マネジメントコンソールからはオブジェクトが非表示になり、削除処理が反映されているように見える。しかし、再度アップロードして確認すると、削除処理時には削除マーカーが追加され、実際には以前のバージョンは保持されていることが確認できる

	バージョン ID	タイプ	最終更新日時
	5qMRCgFW47NPsadB1SLTllyMSr5W34m8 (現行バージョン)	txt	2023/02/07 11:41:05 AM JST
	OkYaye1PNpXHmo6YgKAcVBPzrVI5RXZP	削除マーカー	2023/02/07 11:40:51 AM JST
	NTDMXt21RE3vxRd96UX9FQsVWat4PaNj	txt	2023/02/07 11:38:49 AM JST

バージョンを指定して削除（完全削除）はできない

⊗ 削除に失敗しました (1 オブジェクト, 0 B)

名前でオブジェクトを検索

1

名前	フォルダ	バージョン ID	タイプ	最終更新日時	サイズ	エラー
dummy.txt	-	5qMRCgFW47NPsadB1SLTllyMSr5W34m8	txt	2023/02/07 11:41:05 AM JST	0 B	⊗ アクセスが拒否されました

オブジェクトロックでの挙動: リーガルホールドの変更

リーガルホールドの変更は反映される

リーガルホールドの設定は
リテンションモードや保持期間とは独立
に適用することや設定変更できる

オブジェクトロックの注意点

- オブジェクトロックはバケット作成時に設定し、作成後の有効化/無効化はできない
- バケット作成後に、保持期間の有効化/無効化/変更やリテンションモードの変更はできる
  これらの設定は個々のオブジェクトに適用される
  - 保持期間やリテンションモードを変更した場合には、既存のオブジェクトは従来の設定に従う一方で、新規のオブジェクトには変更後の設定が反映される
- オブジェクトロックの有効にはバケットのバージョニングが必須
  - バージョニングはオブジェクトロックが有効になっているバケットでは無効化できない
- AWS KMS キーでサーバーサイド暗号化を行う場合には、キーの保護も検討する
  - キーを破棄することで、オブジェクトは保持され続けるものの復号できなくなる
    （自動キーローテーションのように暗号化マテリアルを保存する必要がある）

オブジェクトロックのユースケース

- オブジェクト単位で一定期間、上書き/削除を防止する
  - 監査やコンプライアンス目的
  - ランサムウェア対策

- リーガルホールドを適用することで、保持期間の終了後も解除するまで無制限で「監査やコンプライアンス目的」/「ランサムウェア対策」のため、オブジェクトを上書き/削除から保護する

AWS Backup の利用

AWS Backup とは

AWS Backup

AWS の各サービスのバックアップの実行とバックアップデータの一元的な管理を提供

集中型の管理

「バックアッププラン」、「バックアップルール」、「Backup Vault」を定義

バックアップの自動化

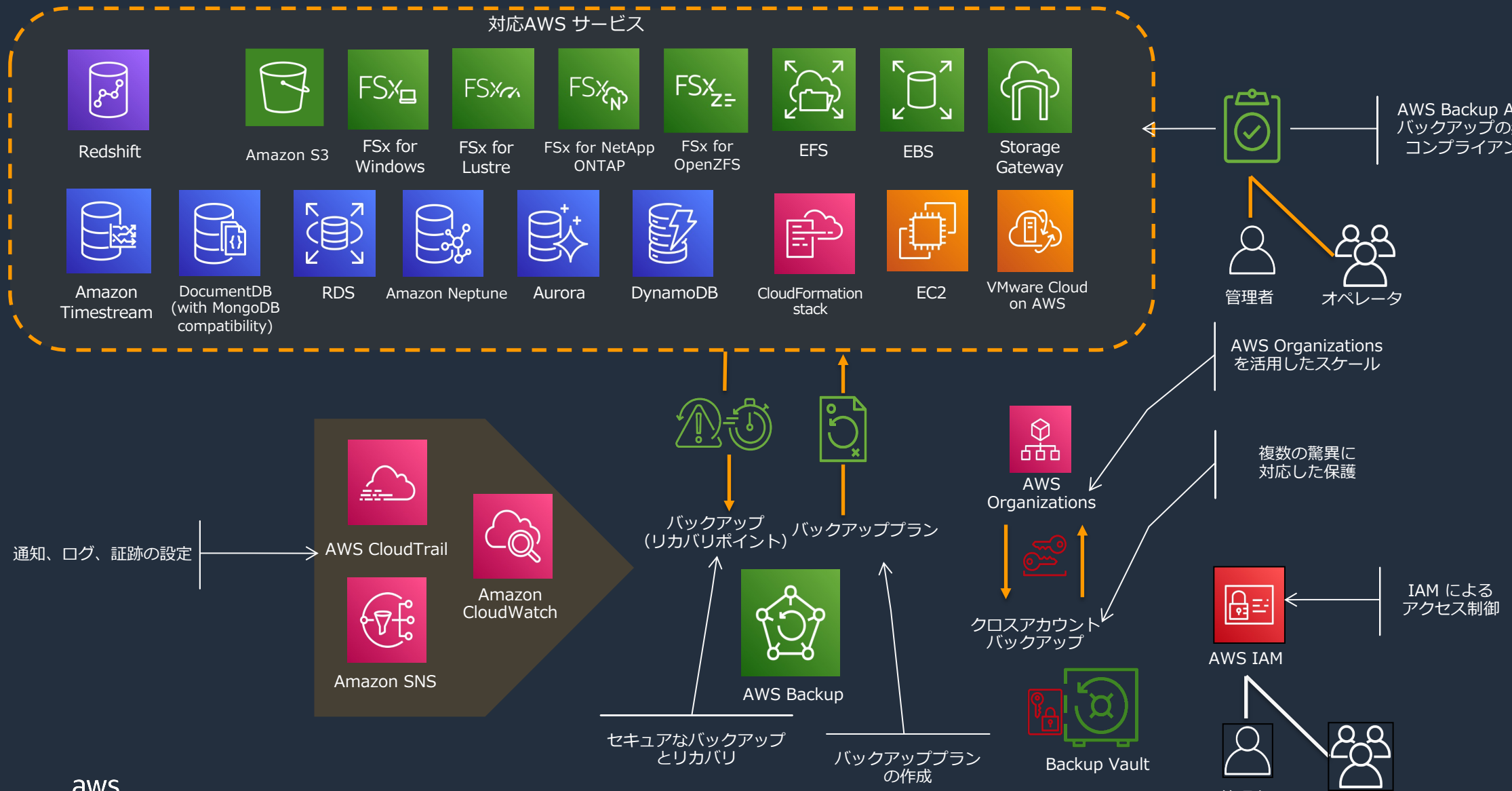
- 「バックアップスケジュール」を定義
- CloudTrail や SNS と連携

コンプライアンス

- IAM でアクセス権限を管理
- 複数のコンプライアンス標準に準拠 (PCI-DSS 含む)

- クラウドネイティブなバックアップと AWS Storage Gateway を統合したハイブリッドなバックアップを提供
- ポリシーベースおよびタグベースのバックアップ
- 自動化されたバックアップスケジューリング
- バックアップの暗号化
- クロスアカウント、クロスリージョンのバックアップ
- 自動バックアップリテンション管理

AWS Backup 全体像

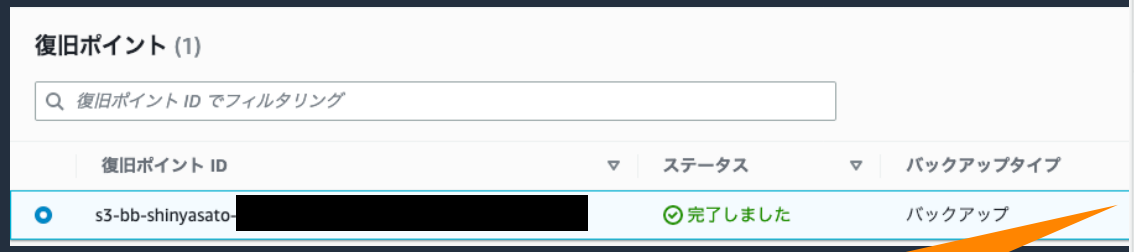


AWS Backup for S3: バケットの保護

AWS Backup を利用して、S3 バケットをバケット単位で Vault へとバックアップできる

AWS Backup for S3: バケットのリストア

AWS Backup を利用して、S3 バケットをバケット単位/オブジェクト単位でリストアできる



バケット全体またはオブジェクト/フォルダ単位でのリストアができる

同じリージョンでバージョニングが有効になっている既存のバケットへのリストアができる

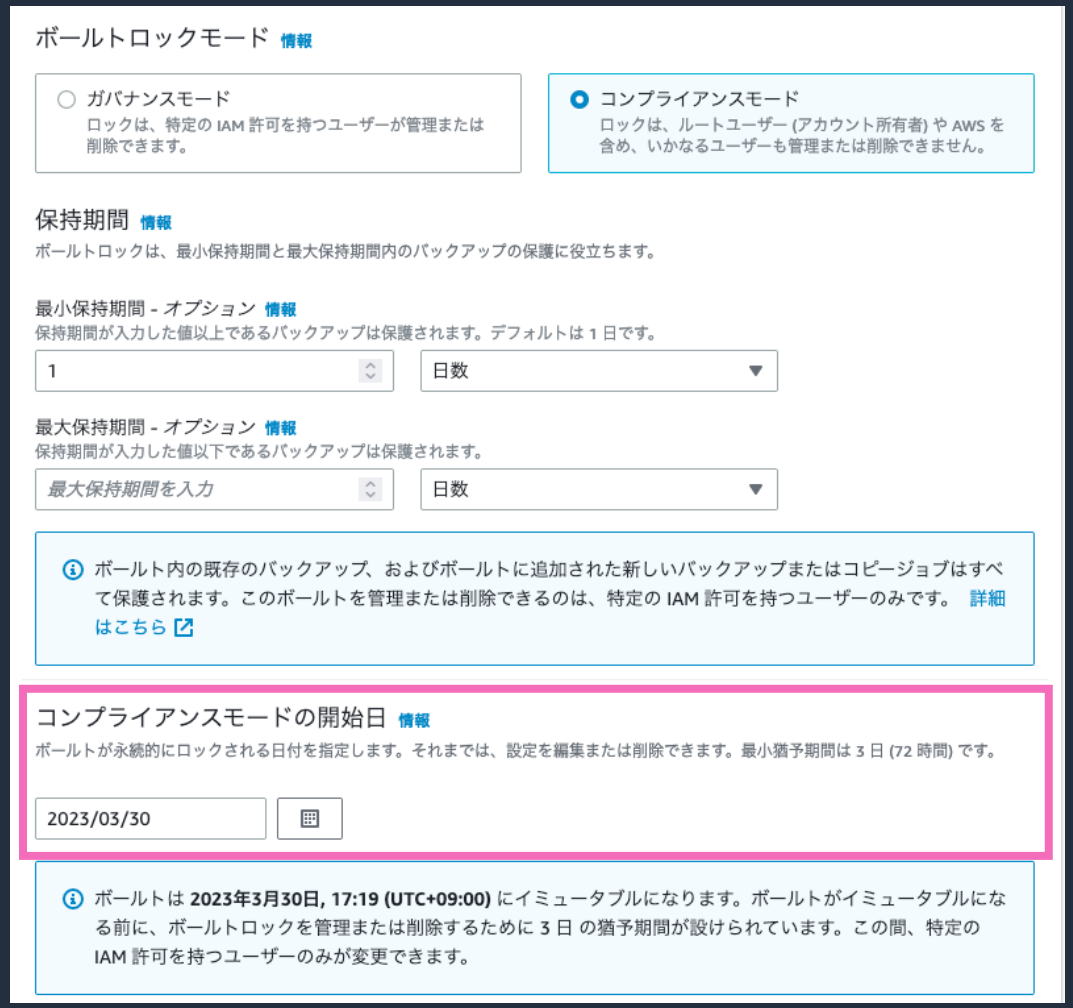
AWS Backup for S3: 特徴と注意点

- バックアップの対象となるバケットはバージョニングを有効化する必要がある
- オブジェクト単位での増分（インクリメンタル）バックアップ
  - 1 GB のオブジェクトのうち、1 KB 分だけ変更された場合にも、1 GB の新しいバックアップが作成される
- 定期的なバックアップとポイントインタイムリカバリ用の継続的なバックアップが選択できる
- クロスリージョン/アカウントバックアップが利用できる
  - ポイントインタイムリカバリと併用はできない

その他参考情報: https://docs.aws.amazon.com/ja_jp/aws-backup/latest/devguide/s3-backups.html

AWS Backup Vault Lock

- S3 のオブジェクトロック同様に、ガバナンスモードとコンプライアンスモードが設定できる
  - コンプライアンスモードでは適用開始までの猶予期間を設定できる
- S3 同様にリーガルホールドを設定できる
- Vault 単位で設定することができ、復旧ポイントが削除されることを防止する
- 最小保持期間と最大保持期間
  - 範囲内の保持期間となる復旧ポイントを保護する
  - 例、最小保持期間が 2 日で、最大保持期間が 2 週間の場合、保持期間が 2 日から 2 週間の復旧ポイントのみ保護される



AWS Backup for S3: ユースケース

- S3 バケット/フォルダ単位でバックアップ/リストアできる
  - 復旧地点を選択することで、素早くリストアできる
- Vault 単位で、WORM 機能を利用できる
  - ランサムウェアなどの脅威への対策
  - コンプライアンス要件を満たす

Vault 単位で、WORM 機能が利用できる　　オブジェクト 単位で、WORM 機能が利用できる

- 単一のバックアップポリシーを使用して、S3 を含めたサービスのバックアップを一元的に管理できる
- AWS Backup Audit Manager により、バックアップ頻度や保持期間などの要件を監査できる

レプリケーションによる データ保護

オブジェクトのレプリケーション 1

- S3 バケット間でオブジェクトを非同期にコピーできる
  - レプリケーションを設定した後に追加、変更、削除されたオブジェクトがレプリケーションされる
  - 非同期コピーであるため、即座にレプリケートされるわけではない
  - オブジェクトの作成時刻やバージョン ID などの全てのメタデータを保持しながらレプリケートする
- ソースバケットとターゲットバケットは、アカウント/リージョンは同じ/異なる場合でも利用できる
  - つまり、クロスアカウント/クロスリージョンレプリケーションができる

必要な準備
- ソースとターゲットバケットは共にバージョニングを有効化する
- レプリケーションルールには適切な権限を付与する
  - SSE-KMS の場合には、レプリケーションルールにアタッチするロールには、ソース/ターゲットで使用するキーへのアクセスを許可する。また、キーポリシーを確認し、レプリケーションルールにアタッチするロールへのアクセス許可を確認する
- クロスアカウントレプリケーションの場合、ターゲットバケットのバケットポリシーで、ソースバケットの所有者にアクセス許可を与える

オブジェクトのレプリケーション 2

レプリケーションのオプションを追加で設定することや、ターゲットバケットは別のストレージクラスを選択できる

99.99% のオブジェクトを 15 分以内のレプリケーションすることもオプションで設定できる※

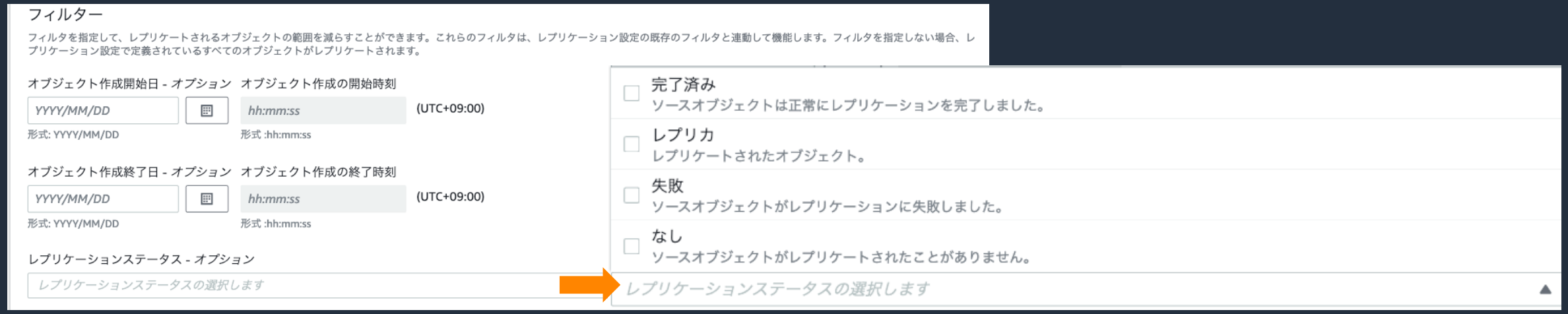
※レプリケーションデータの転送速度が 1 Gbps を超えた期間には適用されない。

別のストレージクラスを選択できる

	ストレージクラス	用に設計	アベイラビリティーゾーン
◉	スタンダード	ミリ秒単位のアクセスが可能で、アクセス頻度の高いデータ (1か月に 1 回以上)	≥ 3
○	Intelligent-Tiering	アクセスパターンが変化したり不明であるデータ	≥ 3
○	標準 - IA	ミリ秒単位のアクセスが可能で、アクセス頻度の低いデータ (1か月に 1 回)	≥ 3
○	1 ゾーン -IA	1 つのアベイラビリティーゾーンに保存され、ミリ秒単位のアクセスが可能な再利用可能でアクセス頻度の低いデータ (1 か月に 1 回)	1
○	Glacier Instant Retrieval	ミリ秒単位で瞬時に取得可能で、アクセスが四半期に一度の存続期間が長いアーカイブデータ	≥ 3
○	Glacier Flexible Retrieval (旧 Glacier)	取得時間が数分から数時間で、アクセスが 1 年に一度の存続期間が長いアーカイブデータ	≥ 3
○	Glacier Deep Archive	取得時間が数時間で、アクセスが 1 年に 1 回未満の存続期間が長いアーカイブデータ	≥ 3
○	低冗長化	非クリティカルでアクセス頻度の高い、ミリ秒単位のアクセスが可能なデータ (S3 標準はコスト効率が高いため、推奨されません)	≥ 3

バッチレプリケーションとは

- レプリケーションの設定を行われる前に存在するオブジェクト、以前にレプリケートされたオブジェクト、レプリケーションに失敗したオブジェクトをレプリケートできる
  - 新しいレプリケーションルールを設定して実行する
  - 既存のレプリケーションルールを用いて実行する
- フィルターを用いて、オブジェクトの作成日などを指定しレプリケートする対象を制限できる
- 注意点
  - ソースバケットにオブジェクトと削除マーカーの複数のバージョンがあるとき、削除マーカーを先にレプリケートする可能性がある。ライフサイクルポリシーによっては、削除マーカーが期限切れとしてマークされた時点で、レプリケート前にターゲットバケットからオブジェクトが削除される
  - ターゲットバケットからバージョン ID を指定して削除されたオブジェクトは再度レプリケーションできない



レプリケーションの考慮事項 1

- ターゲットバケットへレプリケートされないオブジェクト
  - 別のソースバケットからレプリケートされたオブジェクト
  - S3 Glacier Flexible Retrieval/S3 Glacier Deep Archive に保存されているオブジェクト

多段レプリケーションはできない

複数のバケットへレプリケートする

レプリケーションの考慮事項 2

- デフォルト設定では削除マーカーはレプリケートされない。特にソース/ターゲットバケットのライフサイクル設定が異なる場合には注意する。たとえば、ソースバケットでは削除マーカーが挿入されている一方で、ターゲットバケットでは、削除マーカーが挿入されずバケット間に差異が生じる可能性がある。

- Object Ownership を有効化する。
  - 有効にしていない場合、デフォルトではレプリケート元のオブジェクトの所有者もレプリカの所有者になる

- レプリケーションが完了しオブジェクトが利用可能になるまでの時間は、サイズにより異なる
  - ライフサイクルルールは、ターゲットバケットで利用可能になった時間ではなく、作成時間が適用される

レプリケーションのユースケース

- クロスリージョンレプリケーション
  - 遠く離れた地域にデータをレプリケートすることで、コンプライアンス要件を満たす
  - Disaster Recovery
  - エンドユーザーが地理的に分散している時、レイテンシを小さくすることができる

- クロスアカウントレプリケーション
  - ログを 1 つのバケットに集約することができる
  - 本番/分析/テスト環境などといった環境をまたがり、データだけをレプリケートすることができる

データの整合性の検証

チェックサムの利用

チェックサムは、デジタルフィンガープリントの一種（コンテンツの一意性を確認するための値）
チェックサムを利用することで、アップロード/ダウンロードするデータの整合性を検証できる

押さえておくべき用語
- MD5
- ETag
- 追加のチェックサム

MD5 を利用したデータの整合性の検証 1

1. MD5 のダイジェスト（ハッシュ）値を計算する
2. Content-MD5 ヘッダーとして、アップロード時に MD5 ダイジェスト値を引き渡す
3. アップロードされたオブジェクトの整合性を S3 が確認する

ファイルの MD5 ダイジェスト値を計算
バイナリ形式で出力したものを、base16 でエンコード

```
[ec2-user@ip-10-0-12-31 ~]$ echo "Hello World!" >> dummy.txt
[ec2-user@ip-10-0-12-31 ~]$ openssl md5 -binary dummy.txt | base64
jd2L5LF5pSmvpfL/rkuYWA==
[ec2-user@ip-10-0-12-31 ~]$ aws s3api put-object --bucket s3-bb-shinyasato --key dummy.txt --body dummy.txt --content-md5 jd2L
5LF5pSmvpfL/rkuYWA==
{
    "ETag": "\"8ddd8be4b179a529afa5f2ffae4b9858\"",
    "ServerSideEncryption": "AES256"
}
```

アップロードを確認
ETag は 16 進数形式の MD5 ダイジェスト値

```
[ec2-user@ip-10-0-12-31 ~]$ openssl md5 -hex dummy.txt
MD5(dummy.txt)= 8ddd8be4b179a529afa5f2ffae4b9858
```

※ ETag はオブジェクトの特定のバージョンの識別子を示すレスポンスヘッダーで、オブジェクト自体の変更を反映し、メタデータの変更時には反映されない

MD5 を利用したデータの整合性の検証 2

MD5 ダイジェスト値が異なる場合にはエラーが発生する
今回は、引き渡す MD5 ダイジェスト値を意図的に変更し、擬似的なファイルの改竄を発生させた

```
[ec2-user@ip-10-0-12-31 ~]$ echo "Hello World!" >> dummy.txt
[ec2-user@ip-10-0-12-31 ~]$ openssl md5 -binary dummy.txt | base64
jd2L5LF5pSmvpfL/rkuYWA==
[ec2-user@ip-10-0-12-31 ~]$ aws s3api put-object --bucket s3-bb-shinyasato --key dummy.txt --body dummy.txt --content-md5 jd2L
5LF5pSmvpfL/rkuYWA==
{
    "ETag": "\"8ddd8be4b179a529afa5f2ffae4b9858\"",
    "ServerSideEncryption": "AES256"
}
```

```
[ec2-user@ip-10-0-12-31 ~]$ aws s3api put-object --bucket s3-bb-shinyasato --key dummy.txt --body dummy.txt --content-md5 jd2L
5LF5pSmvpfL/rkuYWA==1

An error occurred (InvalidDigest) when calling the PutObject operation: The Content-MD5 you specified was invalid.
```

MD5 を利用したデータの整合性の検証 3

アップロードすると、自動的に ETag が MD5 の値になる

```
[ec2-user@ip-10-0-12-31 ~]$ aws s3api head-object --bucket s3-bb-shinyasato --key dummy.txt
{
    "AcceptRanges": "bytes",
    "LastModified": "2023-03-24T05:11:20+00:00",
    "ContentLength": 13,
    "ETag": "\"8ddd8be4b179a529afa5f2ffae4b9858\"",
    "ContentType": "text/plain",
    "ServerSideEncryption": "AES256",
    "Metadata": {}
}
```

- ETag は PUT Object/POST Object/Copy や マネジメントコンソール上でオブジェクトが作成されたとき、MD5 ダイジェスト値となる。この場合、保存または計算された MD5 ダイジェスト値と ETag を比較することで、整合性を検証することができる
- ただし、マルチパートアップロードを用いた場合や、SSE-C/SSE-KMS を設定時には、ETag は MD5 ダイジェストではない。この場合には、“--metadata md5=” などのコマンドを用いて、MD5 ダイジェストを別途保存することでデータの整合性の検証ができる

マルチパートアップロードを利用した場合の例

マルチアップロードの対象となるサイズのファイルを作成

```
[ec2-user@ip-10-0-12-31 ~]$ dd if=/dev/zero of=./dummy.txt bs=1M count=512
512+0 records in
512+0 records out
536870912 bytes (537 MB, 512 MiB) copied, 5.33334 s, 101 MB/s
[ec2-user@ip-10-0-12-31 ~]$ openssl md5 -hex dummy.txt
MD5(dummy.txt)= aa559b4e3523a6c931f08f4df52d58f2
[ec2-user@ip-10-0-12-31 ~]$ aws s3 cp ./dummy.txt s3://s3-bb-shinyasato/ --metadata md5=aa559b4e3523a6c931f08f4df52d58f2
upload: ./dummy.txt to s3://s3-bb-shinyasato/dummy.txt
[ec2-user@ip-10-0-12-31 ~]$ aws s3api head-object --bucket s3-bb-shinyasato --key dummy.txt
{
    "AcceptRanges": "bytes",
    "LastModified": "2023-03-24T05:28:04+00:00",
    "ContentLength": 536870912,
    "ETag": "\"6d1954a8c7d6f09434c1ba4745a86869-64\"",
    "ContentType": "text/plain",
    "ServerSideEncryption": "AES256",
    "Metadata": {
        "md5": "aa559b4e3523a6c931f08f4df52d58f2"
    }
}
```

MD5 ダイジェスト値を計算

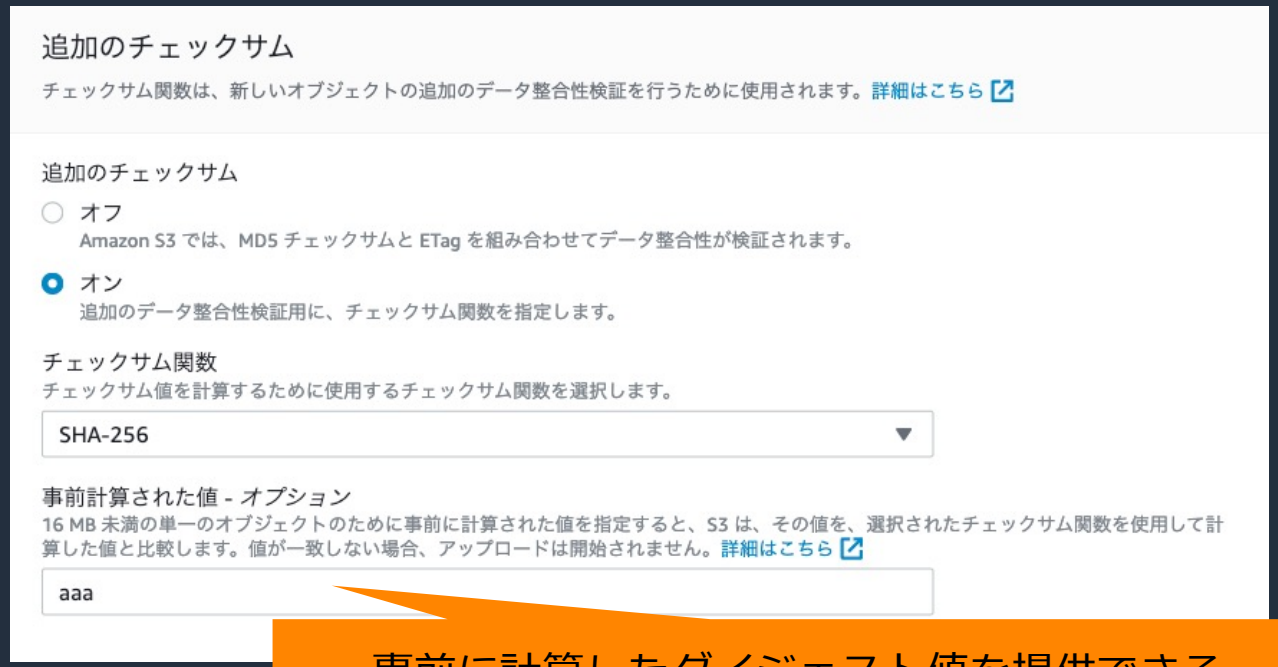
MD5 ダイジェスト値を
メタデータを引き渡す

ETag は MD5 ダイジェスト値とは異なる

メタデータとして MD5 ダイジェスト値を保存

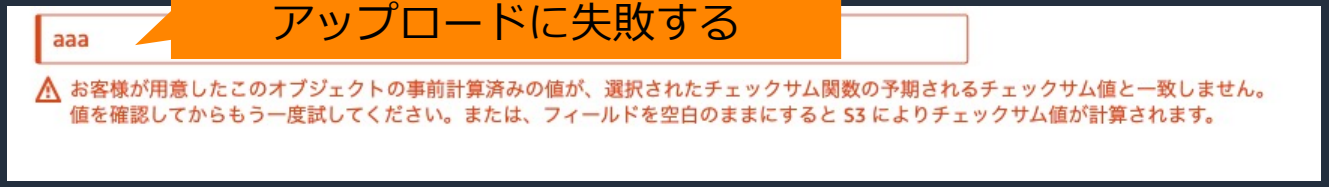
MD5 以外のチェックサムを利用した整合性の検証

- CRC32/CRC32C/SHA-1/SHA-256 をサポート



事前に計算したダイジェスト値を提供できる

ダイジェスト値が異なると
アップロードに失敗する

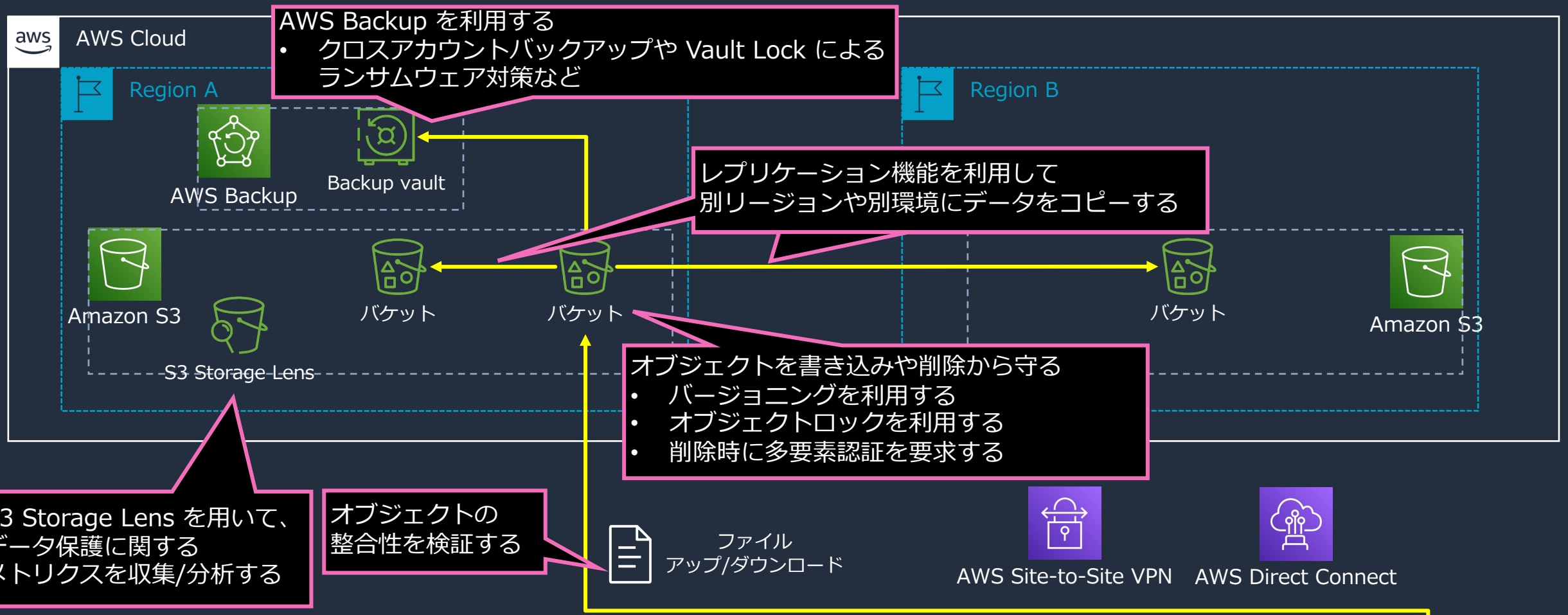


アップロードに成功するとチェックサムが保存される
注意: マルチパートアップロードを利用した場合、オブジェクトの
ダイジェスト値とは異なる値が保存される※

※参考:
https://docs.aws.amazon.com/ja_jp/AmazonS3/latest/userguide/checking-object-integrity.html

まとめ

S3 におけるデータ保護のポイント（再掲）



適切なアクセス制限やデータの暗号化を実施することはデータの保護に繋がる。
アクセス制限やデータの暗号化については、Amazon S3 セキュリティ編を参照:
https://www.youtube.com/watch?v=VutHE2vSvFo&t=1s

まとめ

- Amazon S3 は高い耐久性/低コスト/セキュアなオブジェクトストレージ
- データのバージョニングとオブジェクトロックを使い、データの誤削除や上書きを防ぐ。
  特に WORM 機能を提供するオブジェクトロックはランサムウェアに対して有効となる
- AWS Backup を利用することで、バケット/フォルダ単位でバックアップ/リストア/保護できる
- レプリケーション機能を利用して、コンプライアンス要件を満たすことや DR に利用できる
- チェックサム機能を利用することで、データの改竄を検知することができる
- アクセス制御を適切に設定することで、データの保護はさらに強力になる

本資料に関するお問い合わせ・ご感想

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

https://aws.amazon.com/jp/premiumsupport/

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）

https://console.aws.amazon.com/support/home#/case/create?issueType=customer-service

具体的な案件に対する構成相談は、後述する個別相談会をご活用ください

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

その他コンテンツのご紹介

ウェビナーなど、AWSのイベントスケジュールをご参照いただけます

https://aws.amazon.com/jp/events/

ハンズオンコンテンツ

https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-hands-on/

AWS 個別相談会

AWSのソリューションアーキテクトと直接会話いただけます

https://pages.awscloud.com/JAPAN-event-SP-Weekly-Sales-Consulting-Seminar-2021-reg-event.html

Thank you!