

AWS Builders Online Series

T2-4

デモで解説！今すぐ使える AWSへのデータ連携パターン

吉田 成利

アマゾン ウェブ サービス ジャパン合同会社
テクノロジーパートナーシップ本部
Data & Analytics シニアパートナーソリューションアーキテクト



自己紹介

吉田 成利 | Michitoshi Yoshida

シニアパートナーソリューションアーキテクト

データベースやデータ活用を専門に、
パートナー様のトレーニングや技術支援を担当

好きな AWS のサービス：

Amazon Relational Database Service

Amazon Aurora, Amazon S3



ぜひハッシュタグ **#AWSBuilders** をつけて、
本イベント／セッションで学んだことを呟いてください！
チャットまたは Twitter でのご質問もお待ちしてます！

本セッションについて

対象：

- AWS でデータ分析を始めようとされている方
- データ分析基盤の構築にあたり、データをどう集めるかを検討されている方

本セッションで学べること：

- データの種類、分析に必要なデータの鮮度に応じた、データ連携のパターン
- データ連携を簡単に、すばやく実現できる AWS サービス



アジェンダ

- ・データ連携におけるポイントと
代表的な AWS へのデータ連携パターン
- ・データベースからのデータ連携
- ・SaaS アプリケーションからのデータ連携
- ・今後に向けて



データ連携におけるポイントと 代表的な AWS へのデータ連携パターン



データ分析を取り巻くトレンド



データ容量の
爆発的増加



新たなデータソース
からの取り込み



データ種類の
多様化



多くのユーザー
からの利用

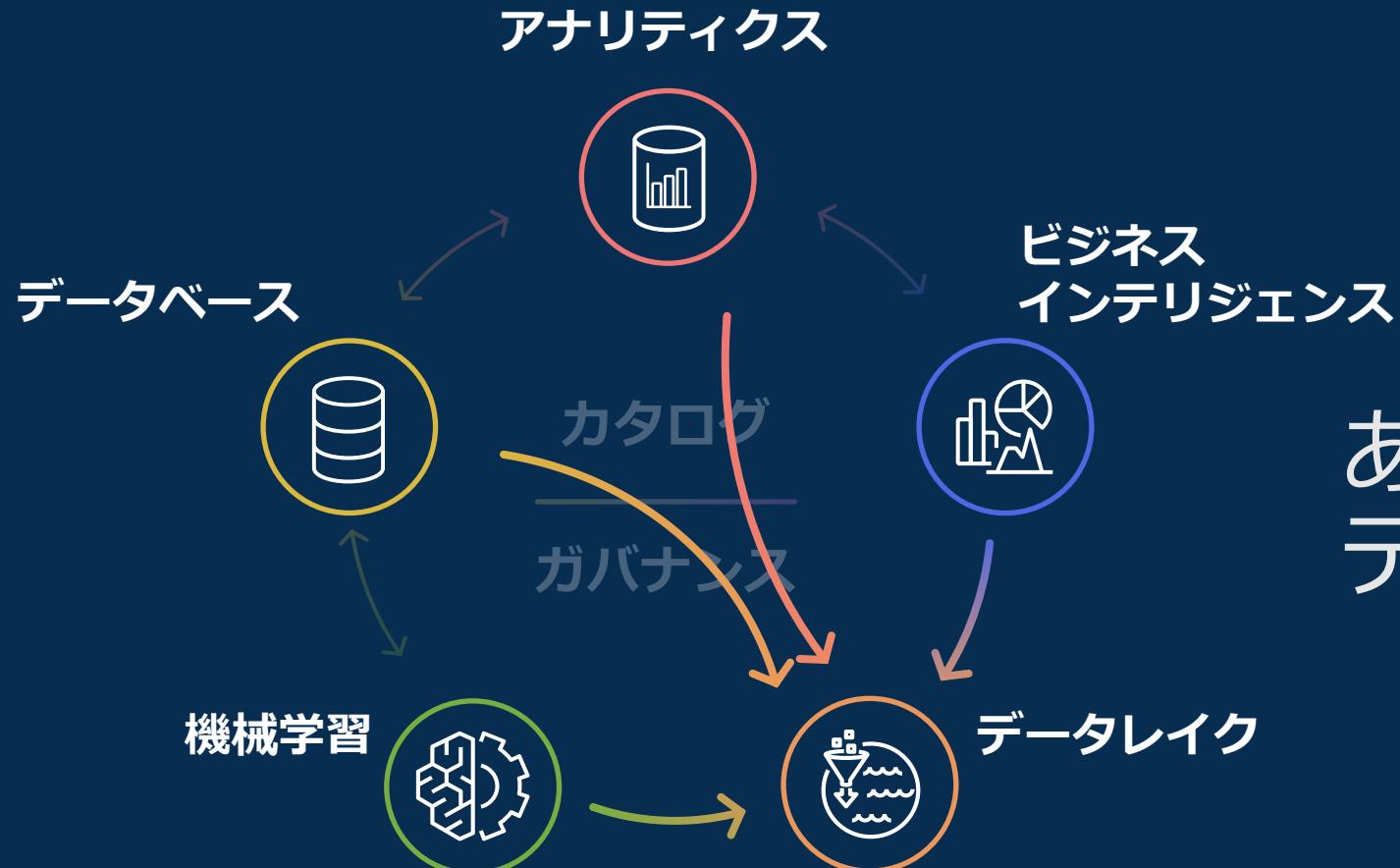


多くのアプリケーシ
ョンからの分析

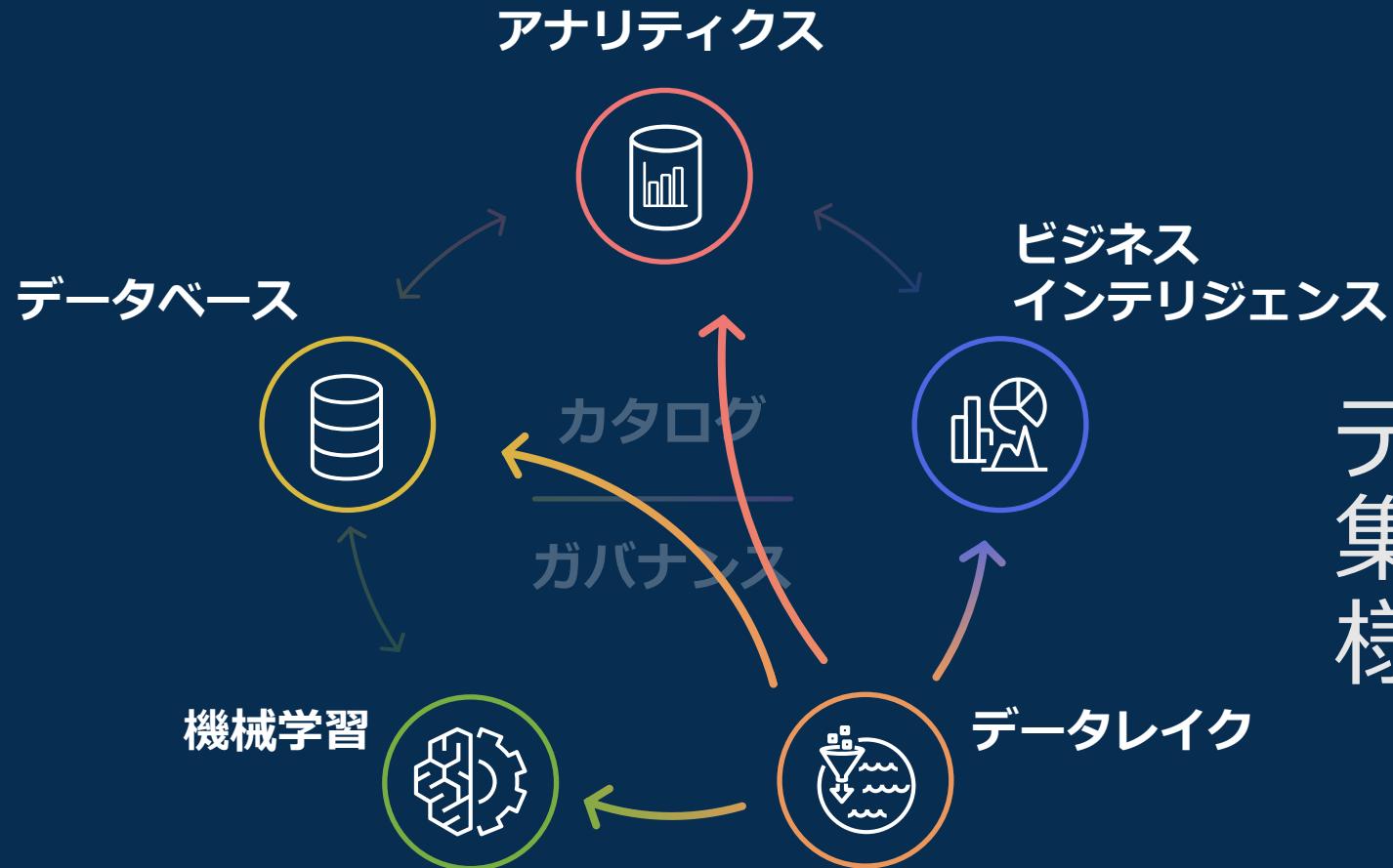


Amazon S3 データレイク





あらゆるデータが
データレイクへ



データレイク上に
集めたデータを
様々なサービスで活用

主なデータソース

- ・ **ファイル**: CSV/XML/JSON、画像ファイル、社内ドキュメント等
- ・ **ストリーミング**: IoT センサーデータ、ログ、ビデオストリーム 等
- ・ **3rd パーティーデータ**: 天候データ、株価データ等
- ・ **データベース**: リレーショナルデータベース (RDBMS) 、 NoSQL
- ・ **SaaS アプリケーション**: CRM, ERP アプリケーション等



主なデータソース

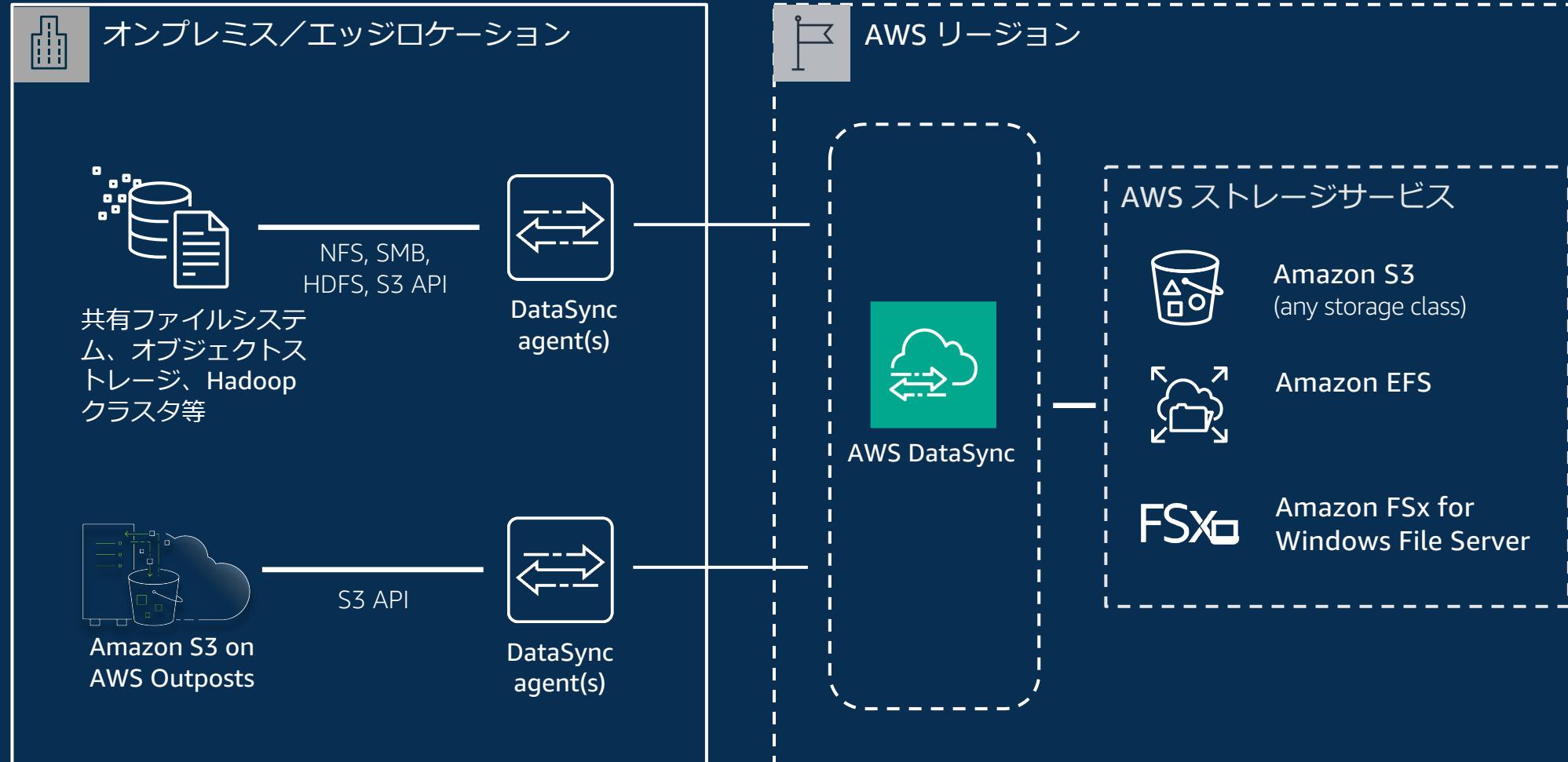
- **ファイル**: CSV/XML/JSON、画像ファイル、社内ドキュメント等
- **ストリーミング**: IoT センサーデータ、ログ、ビデオストリーム 等
- **3rd パーティーデータ**: 天候データ、株価データ等
- **データベース**: リレーショナルデータベース (RDBMS) 、NoSQL
- **SaaS アプリケーション**: CRM, ERP アプリケーション等



ファイルサーバからのデータ連携

AWS DataSync

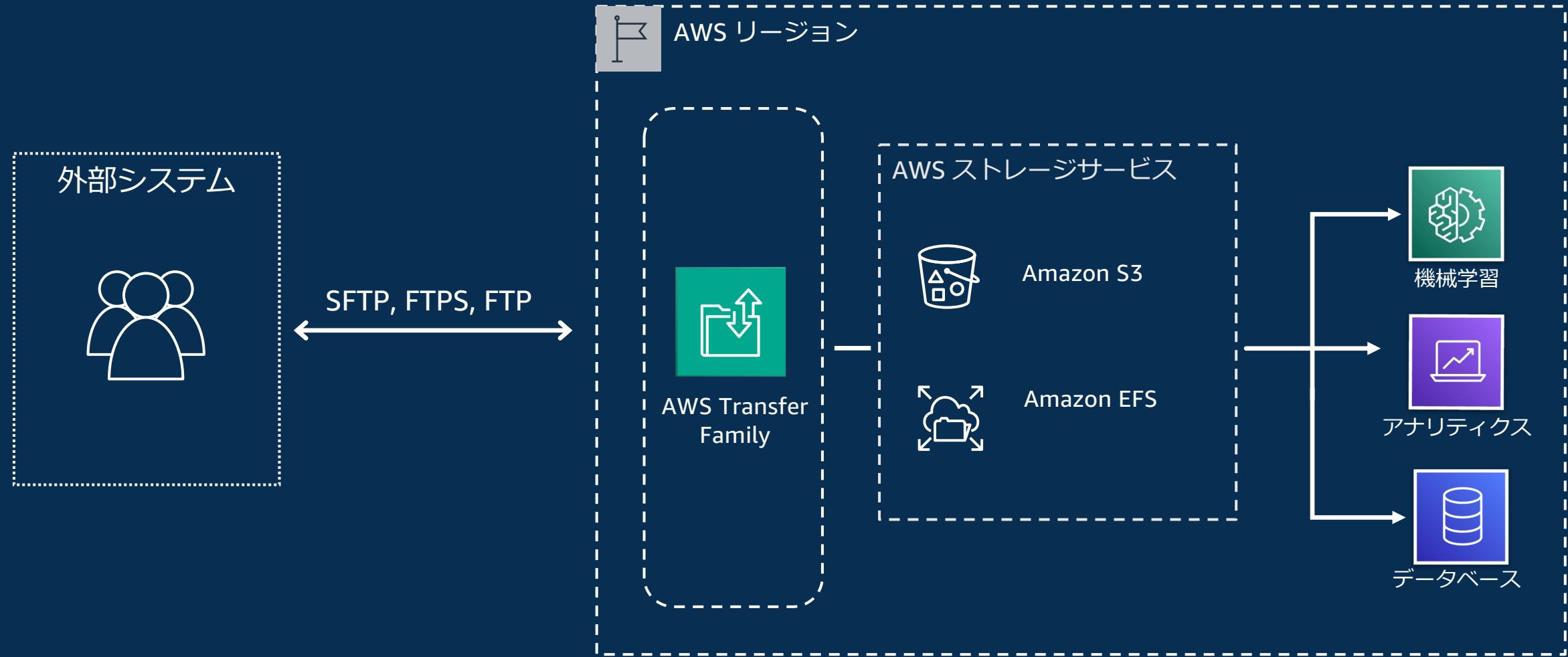
ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



外部システムからのファイル連携

AWS Transfer Family

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



ストリーミングデータの連携

Amazon Kinesis Data Streams / Amazon MSK

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



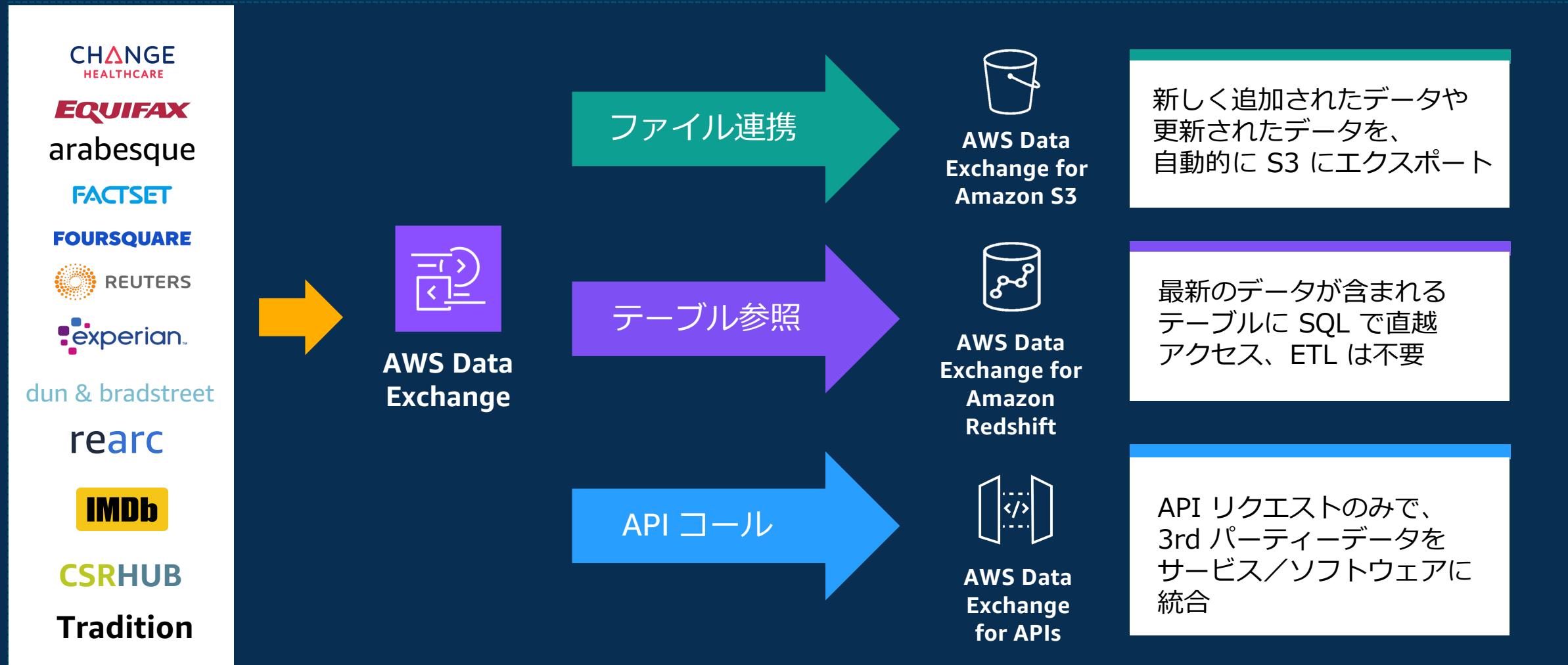
- ・ストリーミングやメッセージキュー Apache Kafka のマネージドサービス



3rd パーティーデータの連携

AWS Data Exchange

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



主なデータソース

- ・ ファイル: CSV/XML/JSON、画像ファイル、社内ドキュメント等
- ・ ストリーミング: IoT センサーデータ、ログ、ビデオストリーム 等
- ・ 3rd パーティーデータ: 天候データ、株価データ等
- ・ データベース: リレーショナルデータベース (RDBMS) 、 NoSQL
- ・ SaaS アプリケーション: CRM, ERP アプリケーション等



データベースからのデータ連携



データ連携の2大パターン

必要なデータの鮮度によって、データ連携の方法を選択

バッチ

日次～数時間ごとに実行

ストリーミング

新規のデータをほぼ
リアルタイムに連携

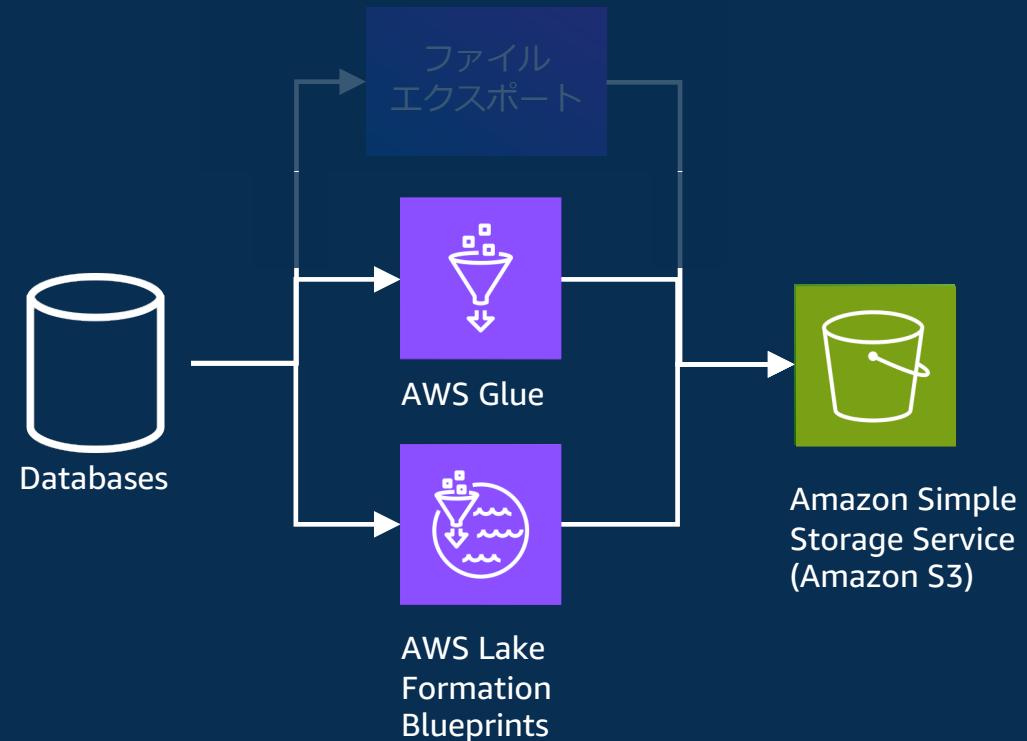


バッチでのデータ連携に使えるサービス

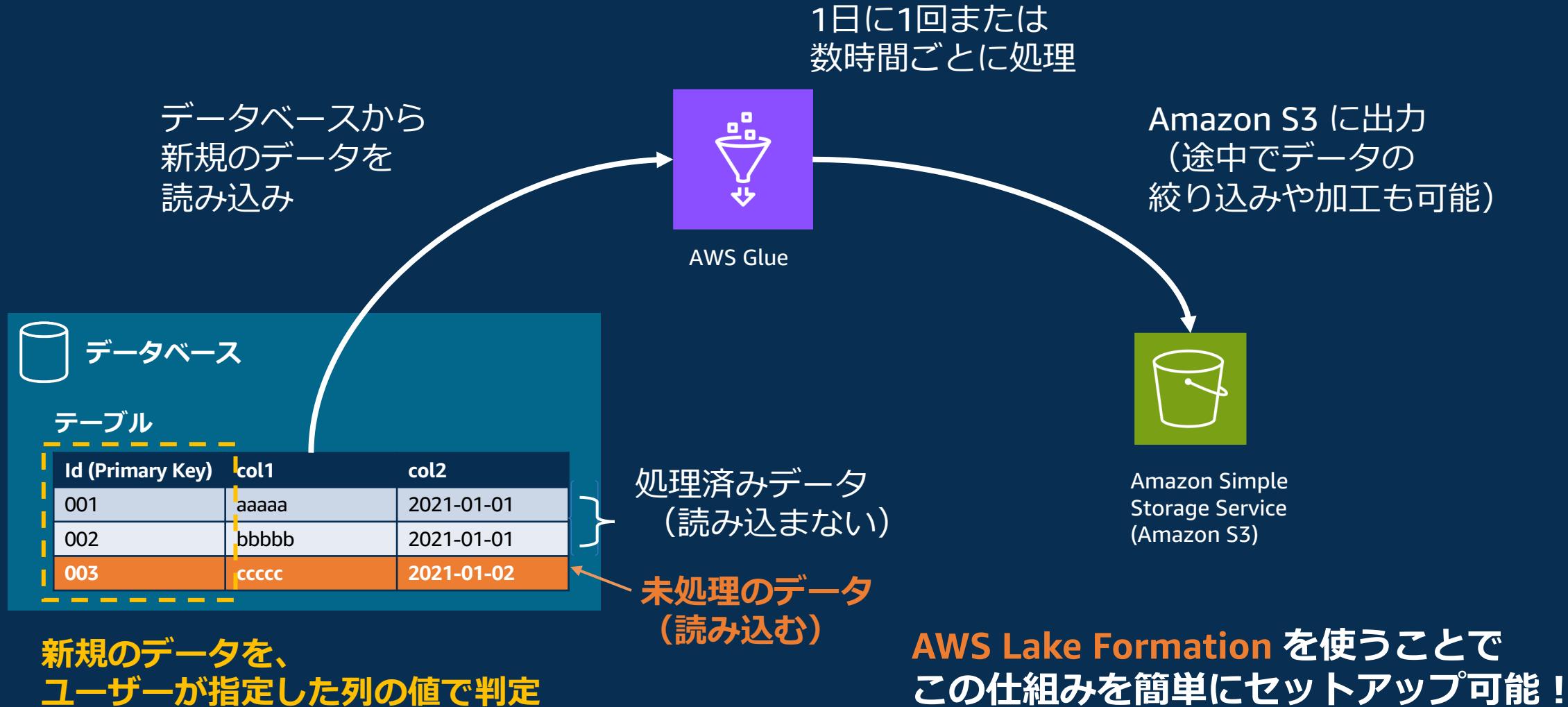
AWS Glue / AWS Lake Formation

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション

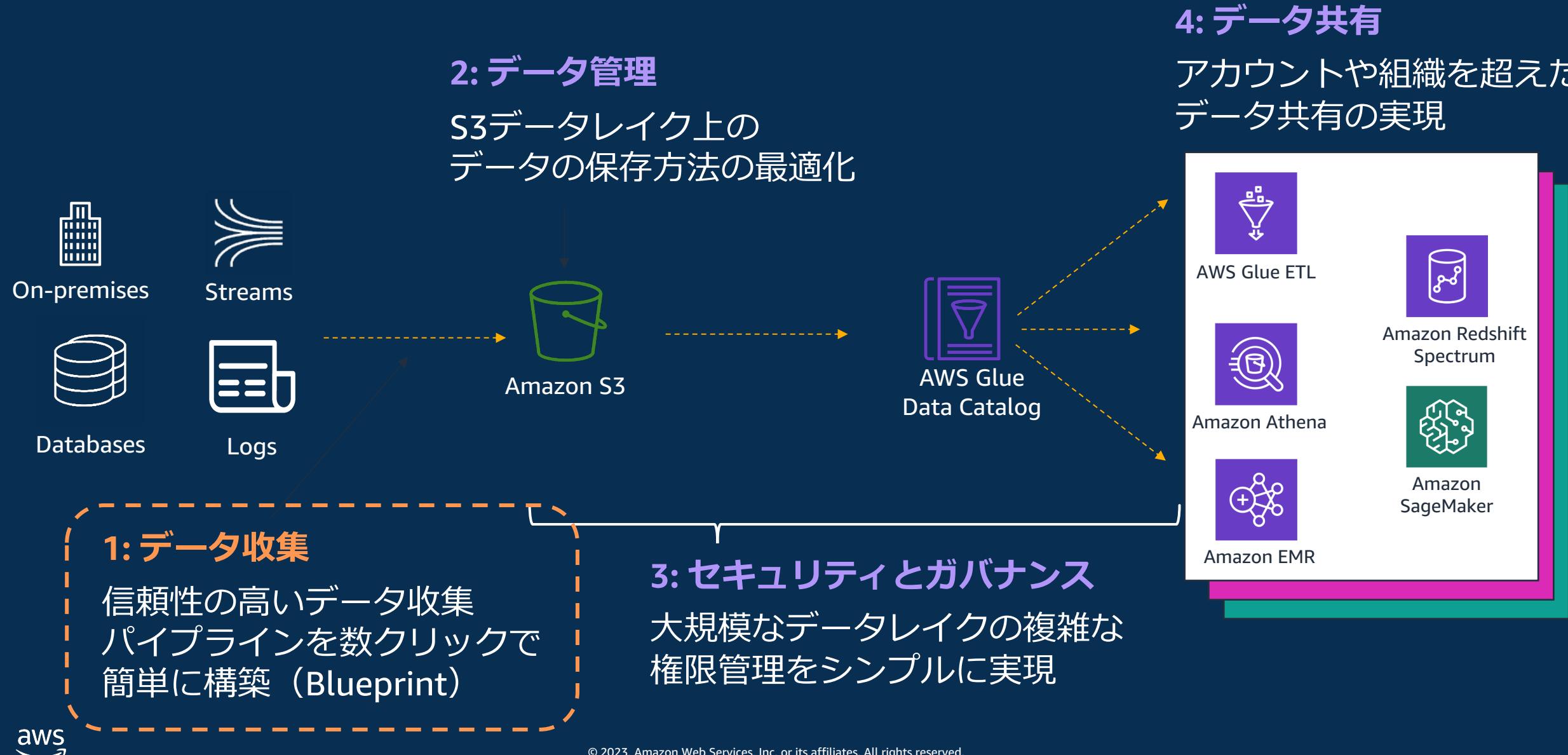
バッチ
(日次～数時間ごと)



AWS Glue を使ったバッチ連携 (1日に1回～数時間ごと)



AWS Lake Formation を使ったデータ収集の構築自動化



Blueprint : データ収集の仕組みを数クリックで構築

データレイクへの取込みを簡単に実現する「データ管理のテンプレート」
汎用的なユースケースで必要な AWS Glue のワークフロー、クローラー、ジョブを生成

- **Database snapshot**

- データベースから
- 表の**全データ**を取り込み

- **Incremental database**

- データベースから
- 新データのみ、**差分**で取り込み

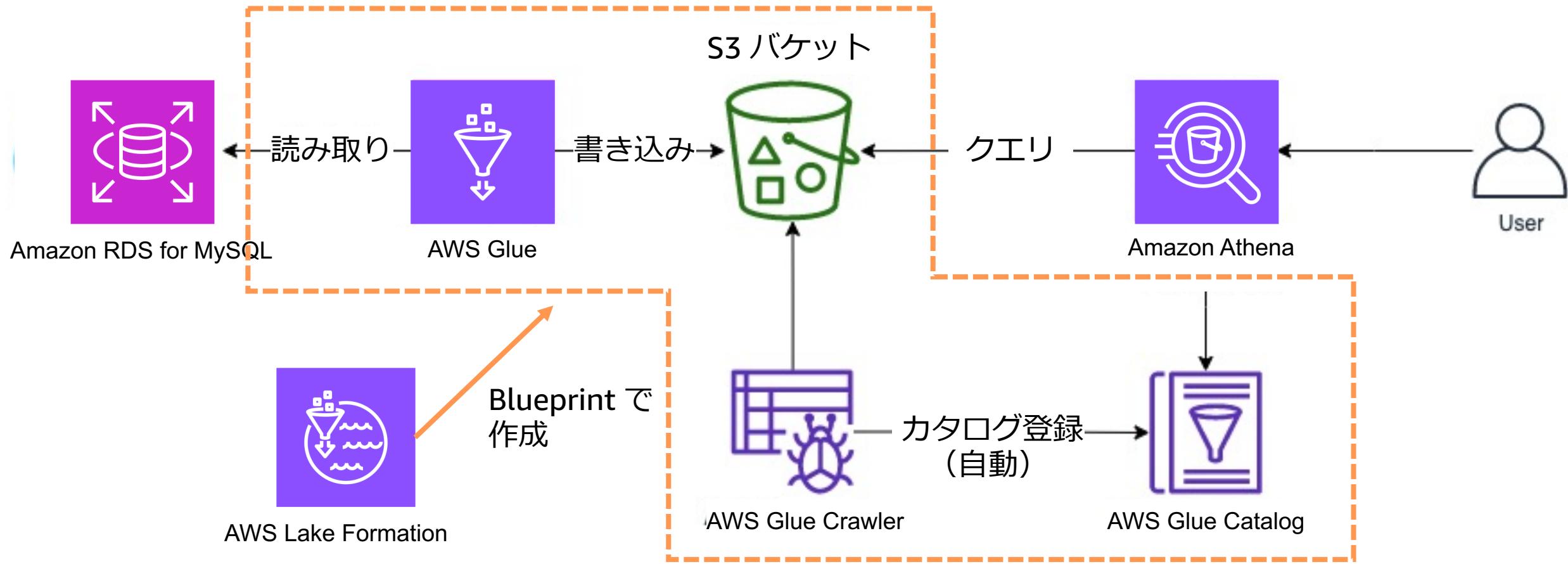
Blueprint type

Configure a blueprint to create a workflow.

- Database snapshot
Bulk load data to your data lake from MySQL, PostgreSQL, Oracle, and Microsoft SQL Server databases.
- Incremental database
Load new data to your data lake from MySQL, PostgreSQL, Oracle, and SQL Server databases.
- AWS CloudTrail
Bulk load data from AWS CloudTrail sources.
- Classic Load Balancer logs
Load data from Classic Load Balancer logs.
- Application Load Balancer logs
Load data from Application Load Balancer logs.



デモ：Blueprint を使ったデータ収集機能の構築



aws | サービス | 検索 [オプション+S] | バージニア北部 ▾

Cloud9 Step Functions Batch RDS S3 Systems Manager CloudWatch

コンソールのホーム 情報

デフォルトレイアウトにリセット + ウィジェットを追加

最近アクセスしたサービス 情報

- AWS Lake Formation
- CloudFormation
- AWS Glue
- Amazon AppFlow
- RDS
- S3
- Step Functions
- Athena
- CloudWatch
- VPC
- Amazon Redshift

[すべてのサービスを表示](#)

AWS へようこそ

- AWS の開始方法** AWS を最大限に活用するために基礎を学び、有益な情報を見つけましょう。
- トレーニングと認定** AWS のエキスパートから学び、スキルと知識を深めましょう。
- AWS の最新情報?** 新しい AWS のサービス、機能、およびリージョンについてご覧ください。

AWS Health 情報

ヘルスデータなし

これは、AWS Health へのアクセス権限がないことが原因の可能性があります。アカウント管理者に連絡してください。

[AWS Health に移動](#)

コストと使用状況 情報

コストと使用状況なし

これは、AWS Cost Manager を設定していないか、許可が付与されていないことが原因である可能性がある

[AWS コスト管理に移動](#)

デモのアーキテクチャ

```

graph LR
    A[Amazon RDS for MySQL] --> B[AWS Glue]
    B --> C[S3 バケット]
    B --> D[Amazon Athena]
    B --> E[AWS Glue Catalog]
    F[AWS Lake Formation] -- Blueprint で作成 --> G[AWS Glue Catalog]
    
```

Amazon RDS for MySQL → AWS Glue → S3 バケット
 AWS Glue → Amazon Athena
 AWS Glue → AWS Glue Catalog
 AWS Glue Catalog → User
 AWS Glue Catalog → AWS Lake Formation
 AWS Lake Formation → Blueprint で作成 → AWS Glue Catalog
 AWS Glue Catalog → カタログ登録(自動) → AWS Glue Catalog

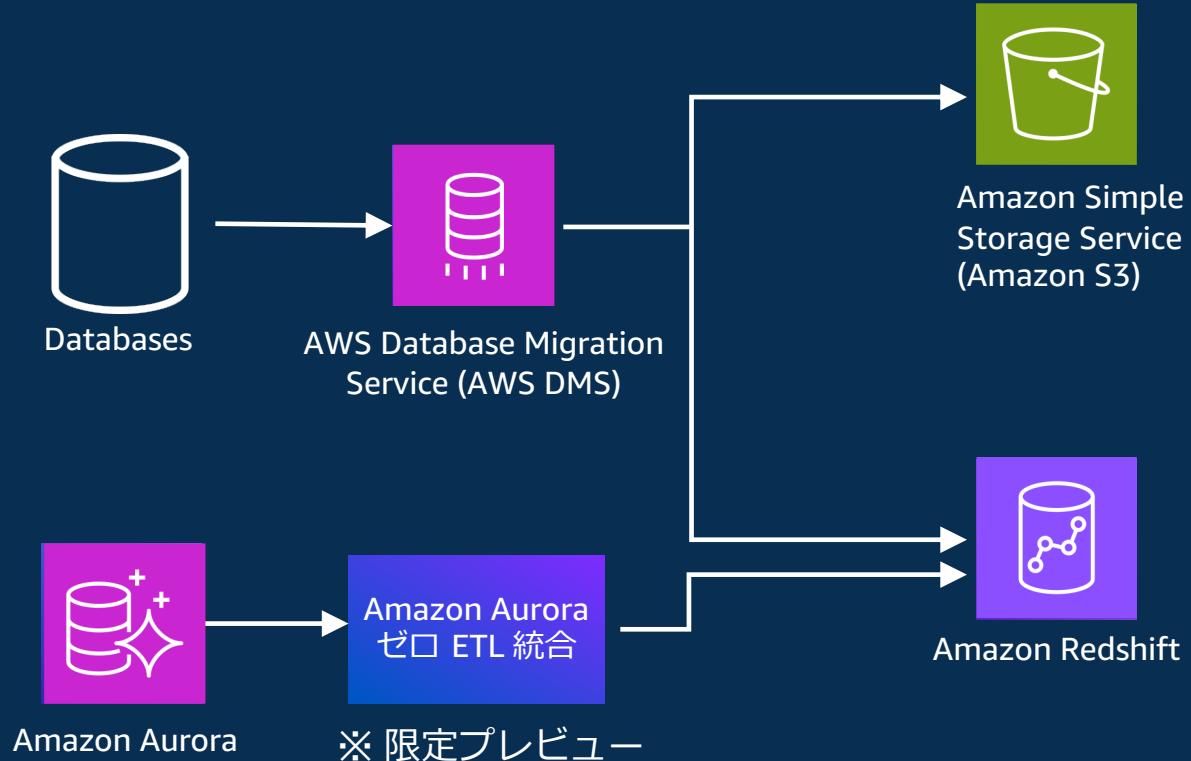
CloudShell フィードバック 言語 © 2023, Amazon Web Services, Inc. またはその関連会社

ニアリアルタイム連携に使えるサービス

ビジネス上のクリティカルなイベントを、分析基盤にすぐに連携

ファイル
ストリーミング
3rdパーティデータ
データベース
SaaS アプリケーション

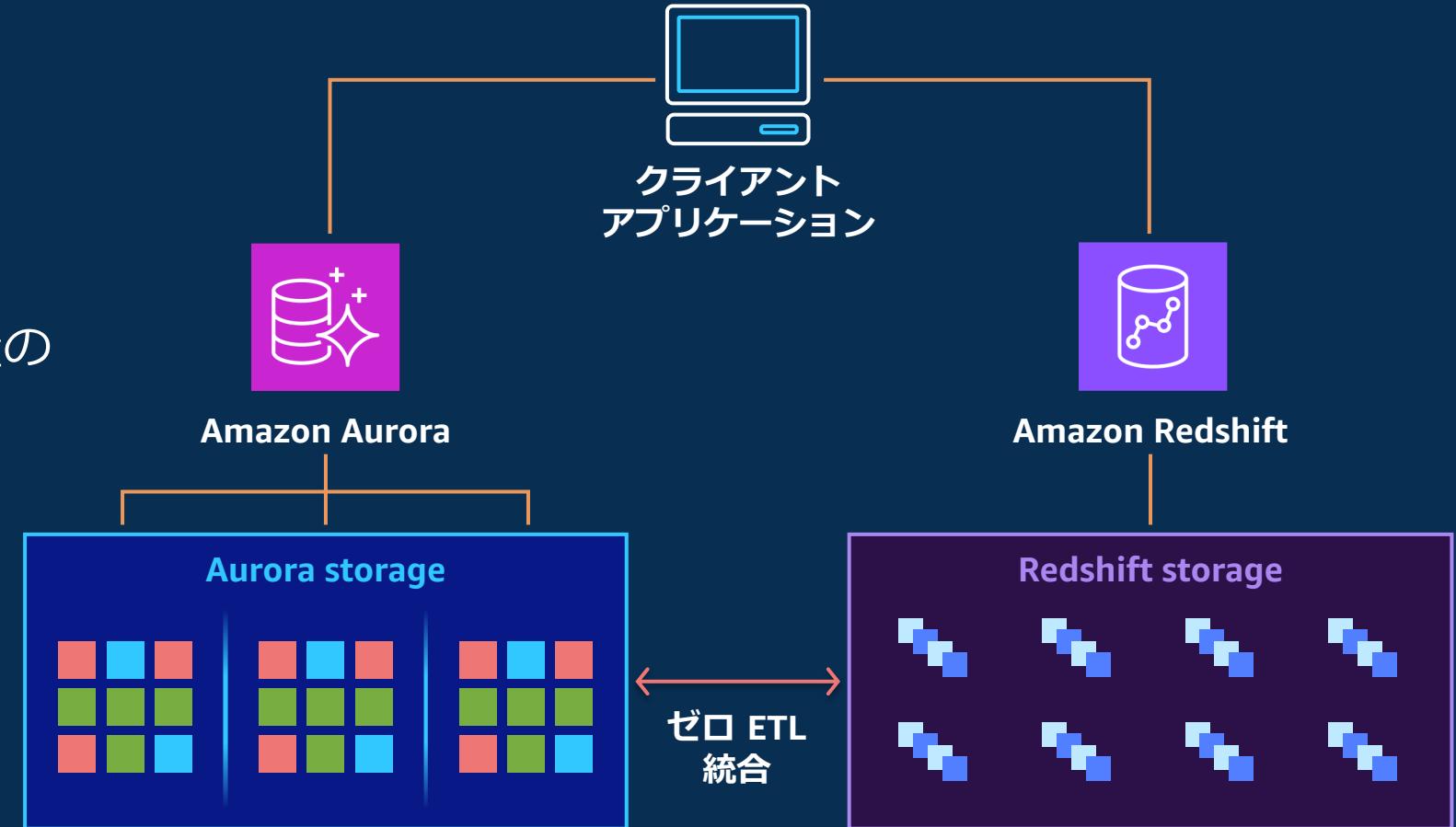
ストリーミング (ニアリアルタイム)



Amazon Aurora と Amazon Redshift のゼロ ETL 統合 (限定プレビュー)

ニアリアルタイムでのデータ連携をシンプルに実現

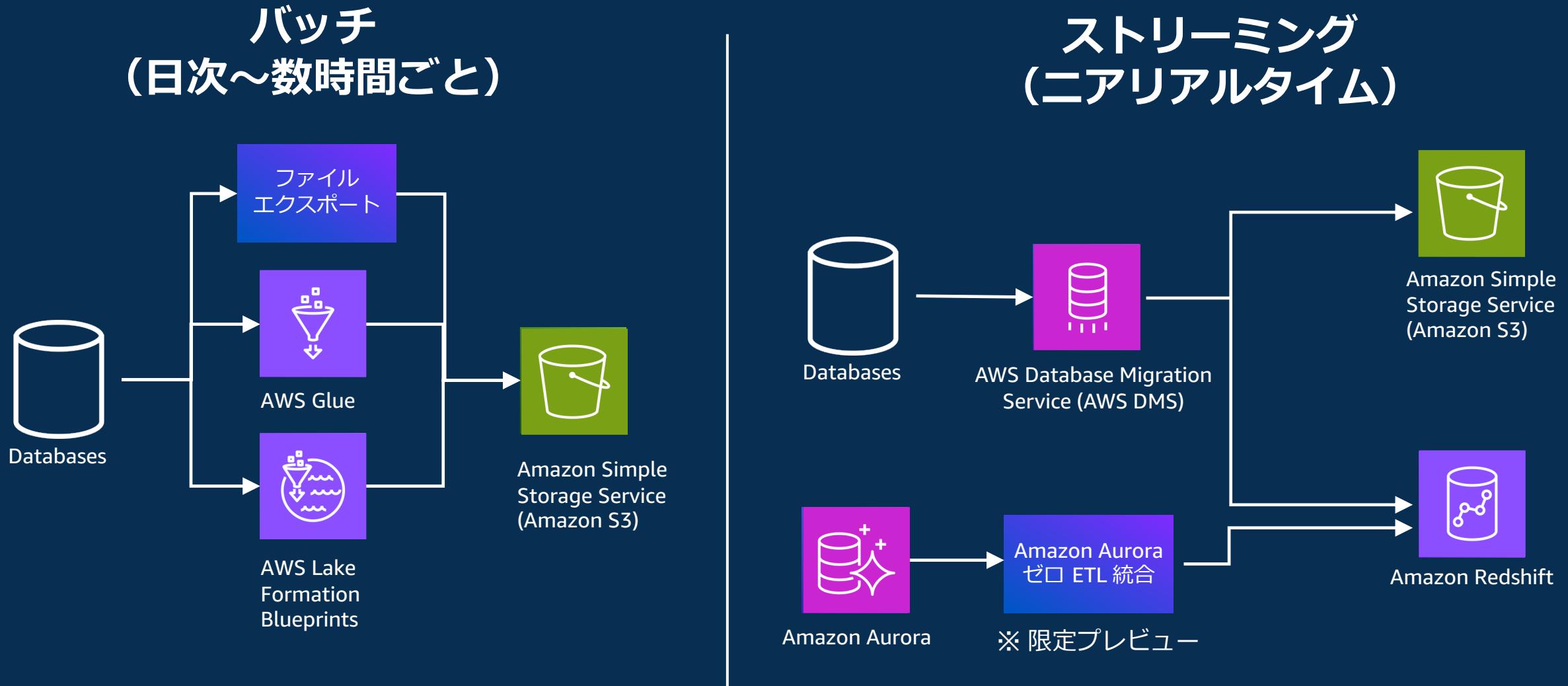
- 複雑化しやすいデータベースとデータウェアハウスの間の ETL パイプラインを排除
- Amazon Aurora で発生した大量のトランザクションデータを、ニアリアルタイムでの分析や機械学習に利用可能
- 複数の Aurora クラスタのデータを Redshift に集約して分析することも可能



データベースからの連携パターン

AWS Glue / AWS Lake Formation / AWS Database Migration Service (DMS)

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



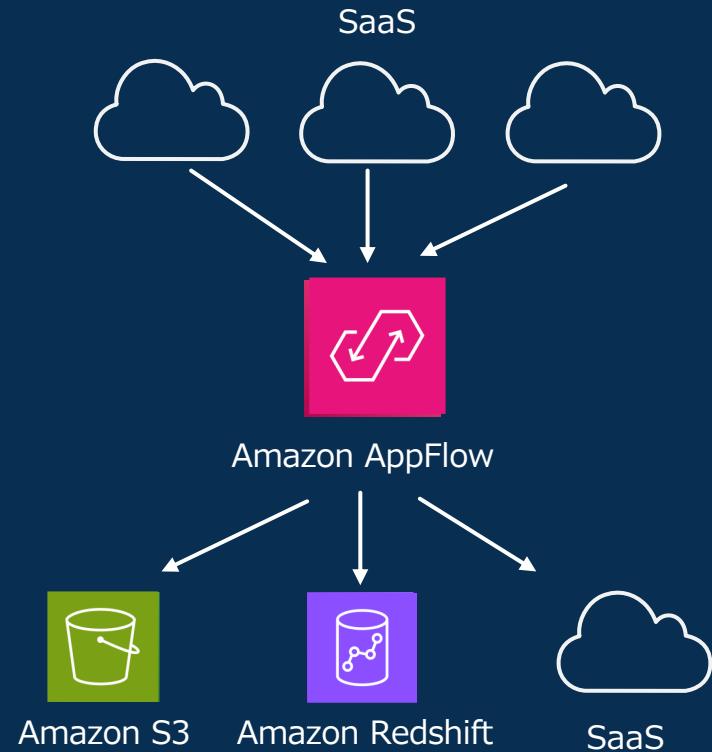
SaaS アプリケーションからの データ連携



SaaS データ連携 : Amazon AppFlow

わずか数回のクリックで SaaS アプリケーションと AWS 間の双方向のデータフローを可能にするアプリケーション統合サービス

ファイル
ストリーミング
3rd パーティーデータ
データベース
SaaS アプリケーション



ノーコードでデータ連携機能を構築



初期コスト／サーバー管理が不要



データフローを数分で作成・変更



大量のレコードを高速・安全に転送可能

55 以上の SaaS や AWS サービスと連携可能

サービス	送信元?	送信先?
Amazon Connect	Yes	No
Amazon EventBridge	No	Yes (Limited)
Amazon Honeycode	No	Yes
Lookout for Metrics	No	Yes
RDS for PostgreSQL	No	Yes
Amazon Redshift	No	Yes
Amazon S3	Yes	Yes
Amplitude	Yes	No
CircleCI	Yes	No
Datadog	Yes	No
DocuSign Monitor	Yes	No
Domo	Yes	No
Dynatrace	Yes	No
Facebook Ads	Yes	No
Facebook Page Insights	Yes	No
Freshdesk	Yes	No
GitHub	Yes	No
GitLab	Yes	No
Google Ads	Yes	No

サービス	送信元?	送信先?
Google Analytics	Yes	No
Google Analytics 4	Yes	No
Google Search Console	Yes	No
HubSpot	Yes	No
Infor Nexus	Yes	No
Instagram Ads	Yes	No
Jira Cloud	Yes	No
LinkedIn Ads	Yes	No
LinkedIn Pages	Yes	No
Mailchimp	Yes	No
Marketo	Yes	No
Mixpanel	Yes	No
MS SharePoint Online	Yes	No
MS Teams	Yes	No
Okta	Yes	No
Productboard	Yes	No
QuickBooks Online	Yes	No
Recharge	Yes	No
SendGrid	Yes	No
Salesforce	Yes	Yes

サービス	送信元?	送信先?
Salesforce Marketing Cloud	Yes	No
Salesforce Pardot	Yes	No
SAP OData	Yes	Yes (Limited)
SendGrid	Yes	No
ServiceNow	Yes	No
Singular	Yes	No
Slack	Yes	No
Snowflake	No	Yes
Stripe	Yes	No
Trend Micro	Yes	No
Typeform	Yes	No
Upsolver	No	Yes
Veeva	Yes	No
Zendesk	Yes	Yes
Zendesk Chat	Yes	No
Zendesk Cell	Yes	No
Zendesk Sunshine	Yes	No
Zoom	Yes	No



AWS PrivateLink を使った SaaS との閉域通信

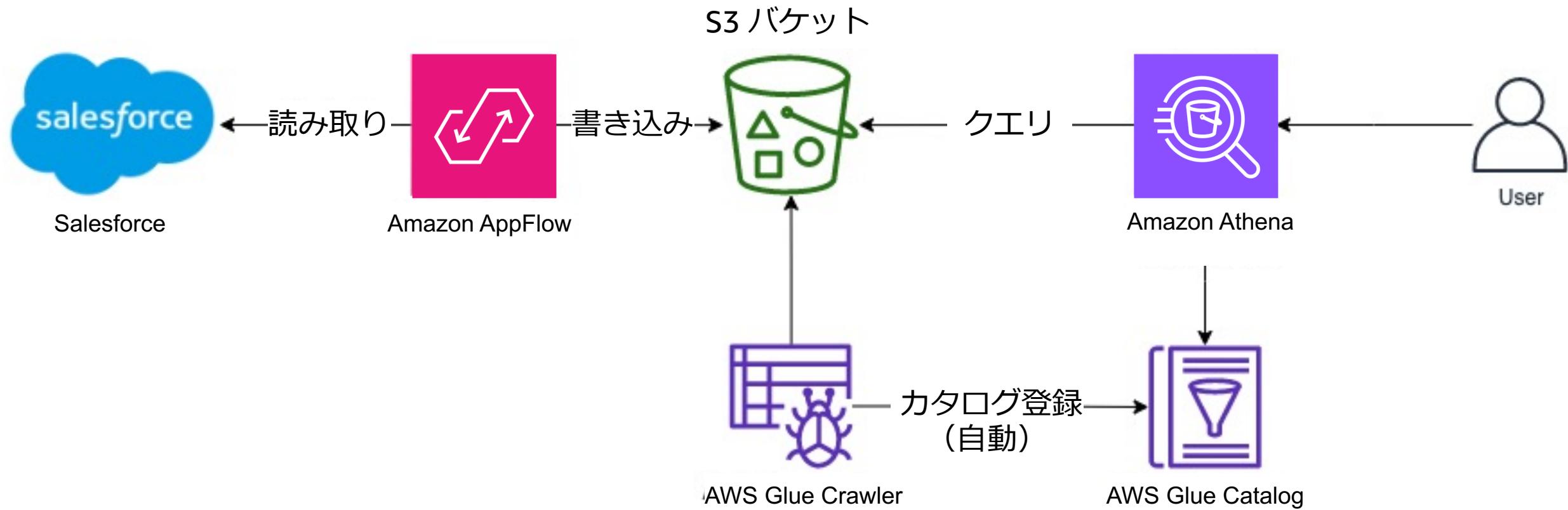
特定の「SaaS とリージョンの組合せ」において、インターネット (Public) を介さずに AppFlow サービスが管理する PrivateLink 経路でデータ転送の閉域通信を利用可能

PrivateLink 対応 SaaS	送信元?	送信先?
Salesforce		Yes Yes
SAP		Yes Yes (Limited)
Singular		Yes No
Snowflake		No Yes
Trend Micro		Yes No

※ 当該 SaaS アカウントにて、個別の前提条件を満たし、
PrivateLink 接続に対応する構成・有効化されていることが前提
例：Salesforce の場合、Private Connect アクセス可能かつ
Salesforce アカウント側に「Manager Metadata」と「Manage External
Connections」権限が必要



デモ：Salesforce から Amazon S3 へのデータ連携



aws サービス 検索 [オプション+S] 東京

Cloud9 Step Functions Batch EC2 RDS S3 Systems Manager CloudWatch

コンソールのホーム 情報

デフォルトルエイアウトにリセット +

最近アクセスしたサービス 情報

- S3
- AWS Glue
- Athena
- Amazon AppFlow
- CloudFormation
- EC2
- Amazon Redshift
- VPC
- CloudWatch
- AWS Lake Formation
- Step Functions
- RDS

すべてのサービスを表示

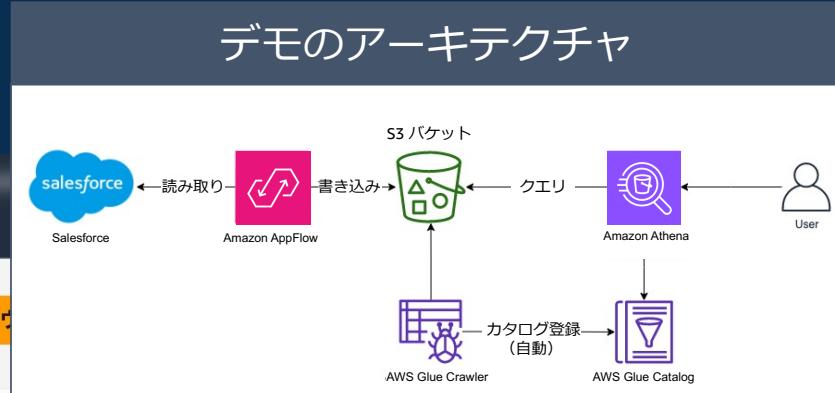
AWS Health 情報

未解決の問題 0 過去 7 日間

スケジュールされている変更

コストと使用状況 情報

CloudShell フィードバック 言語 © 2023, Amazon Web Services, Inc. またはその関連会社。 プライバシー 用語 Cookie の設定



まとめ



まとめ

- 分析基盤に、どんなデータを、どうやって集めるかを考えるうえで、**目的や得たいビジネス効果、利用するユーザのスキルから逆算**するアプローチが有効
- AWS では、**必要なデータの形式や鮮度に応じた、豊富なデータ連携のサービスや機能**を提供
- AWS Lake Formation や Amazon AppFlow といった、**データ収集の仕組みを簡単に実現するサービス**を利用することで、データから価値を得るまでのリードタイムを短縮可能



本セッションでご紹介した AWS サービス

- ファイル: [AWS DataSync](#) / [AWS Transfer Family](#)
- ストリーミング: [Amazon Kinesis](#) /
[Amazon Managed Streaming for Apache Kafka](#)
- 3rd パーティデータ: [AWS Data Exchange](#)
- データベース: [AWS Glue](#) / [AWS Lake Formation](#) / [AWS DMS](#)
- SaaS アプリケーション: [Amazon AppFlow](#)



今後に向けて

[AWS Lake Formation Workshop](#)で、
Blueprint を使ったデータベースからの
データ連携や、データレイクの
権限管理を体験！

AWS Lake Formation Workshop > Lake Formation Basics > Hydrate Data Lake > Blueprints

Blueprints

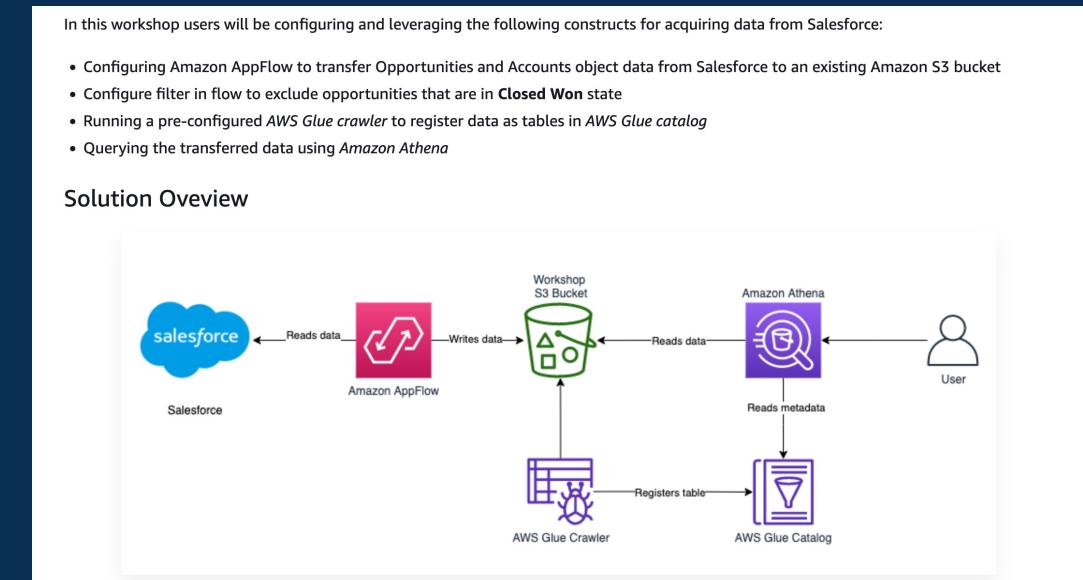
Blueprints enable data ingestion from common sources using automated workflows. At high level, Lake Formation provides two type of blueprints:

- **Database blueprints:** This blueprints help ingest data from MySQL, PostgreSQL, Oracle, and SQL server databases to your data lake. You can ingest either as bulk load snapshot, or incrementally load new data over time.
- **Log file blueprints:** Ingest data from popular log file formats from AWS CloudTrail, Elastic Load Balancer, and Application Load Balancer logs.

In this exercise, we will use database snapshot as blueprint and will ingest entire TPC database to your data lake.

• Click on the **Blueprints** option from the left navigation panel and then click on **Use blueprint** button.

[Amazon AppFlow Workshop](#) で、
Salesforce をはじめとした
SaaS アプリケーションから
AWS へのデータ連携を体験！



Thank you!



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS TRAINING & CERTIFICATION

600+ ある AWS Skill Builder の無料デジタルコースで学ぼう

30 以上の AWS ソリューションの中から、自分にもっとも関係のあるクラウドスキルとサービスにフォーカスし、自習用のデジタル学習プランとランプアップガイドで学ぶことができます。

自分に合ったスキルアップ方法で学ぼう

[EXPLORE.SKILLBUILDER.AWS »](https://explore.skillbuilder.aws)



あなたのクラウドスキルを AWS 認定で証明しよう

業界で認められた資格を取得して、スキルアップの一歩を踏み出しましょう。AWS Certified Cloud Practitioner の取得方法と、準備に役立つ AWS のリソースをご覧ください。

[受験準備のためのリソースにアクセスしよう »](#)



AWS Builders Online Series に ご参加いただきありがとうございます

楽しんでいただけましたか? ぜひアンケートにご協力ください。
本日のイベントに関するご意見/ご感想や今後のイベントについての
ご希望や改善のご提案などがございましたら、ぜひお聞かせください。



aws-apj-marketing@amazon.com



twitter.com/awscloud_jp



facebook.com/600986860012140



<https://www.youtube.com/user/AmazonWebServicesJP>



<https://www.linkedin.com/showcase/aws-careers/>



twitch.tv/aws

