



Amazon Redshift Overview

AWS Black Belt Online Seminar

池田 敬之

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/>
 - <https://www.youtube.com/playlist?list=PLzWGOASvSx6FIwIC2X1nObr1KcMCBBBlqY>

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おります。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：池田 敬之

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 西日本ソリューション部

ソリューションアーキテクト

経歴：日系SIerにてデータベースの導入・運用

->外資系データベースベンダーにてプリセールス

->関西の企業様担当ソリューションアーキテクト @AWS

好きなAWSサービス： Amazon Redshift



本セミナーの対象者

データベースの導入や運用に携わるデータベース管理者

AWS 環境におけるデータウェアハウスに関心がある方

Amazon Redshift の機能を網羅的に学びたい方

アジェンダ

1. データ分析における背景と課題
2. Amazon Redshift とは
3. Amazon Redshift の様々な機能
4. 料金体系
5. まとめ

データ分析における 背景と課題

大規模データ分析における課題



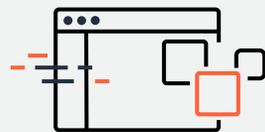
多種多様なソースとデータタイプ



様々な分析ニーズへの対応



データ量とベロシティ



パフォーマンスが出ない



システム管理が困難



拡張するのが複雑



増加する予測不可能なコスト

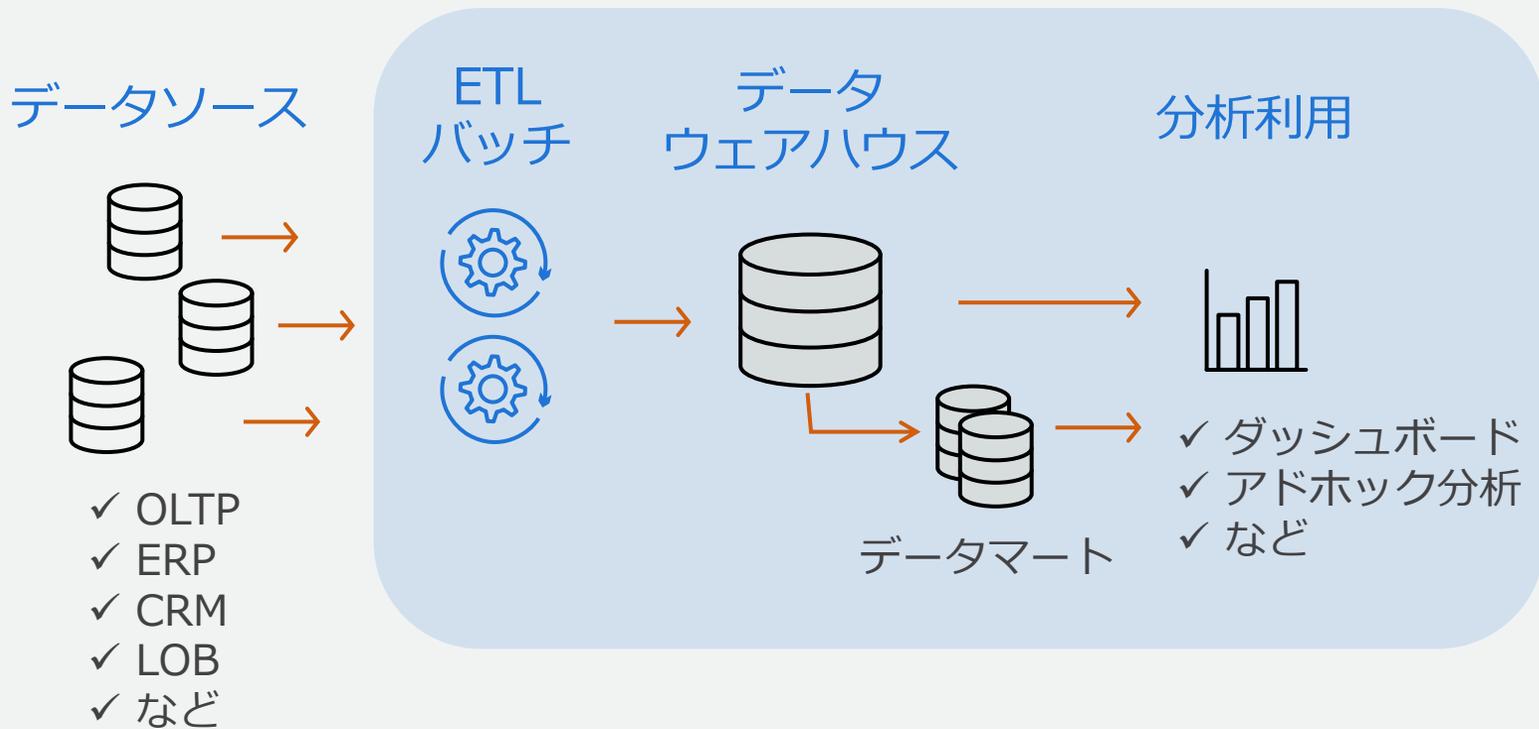


柔軟性のないツール



セキュリティ コンプライアンス

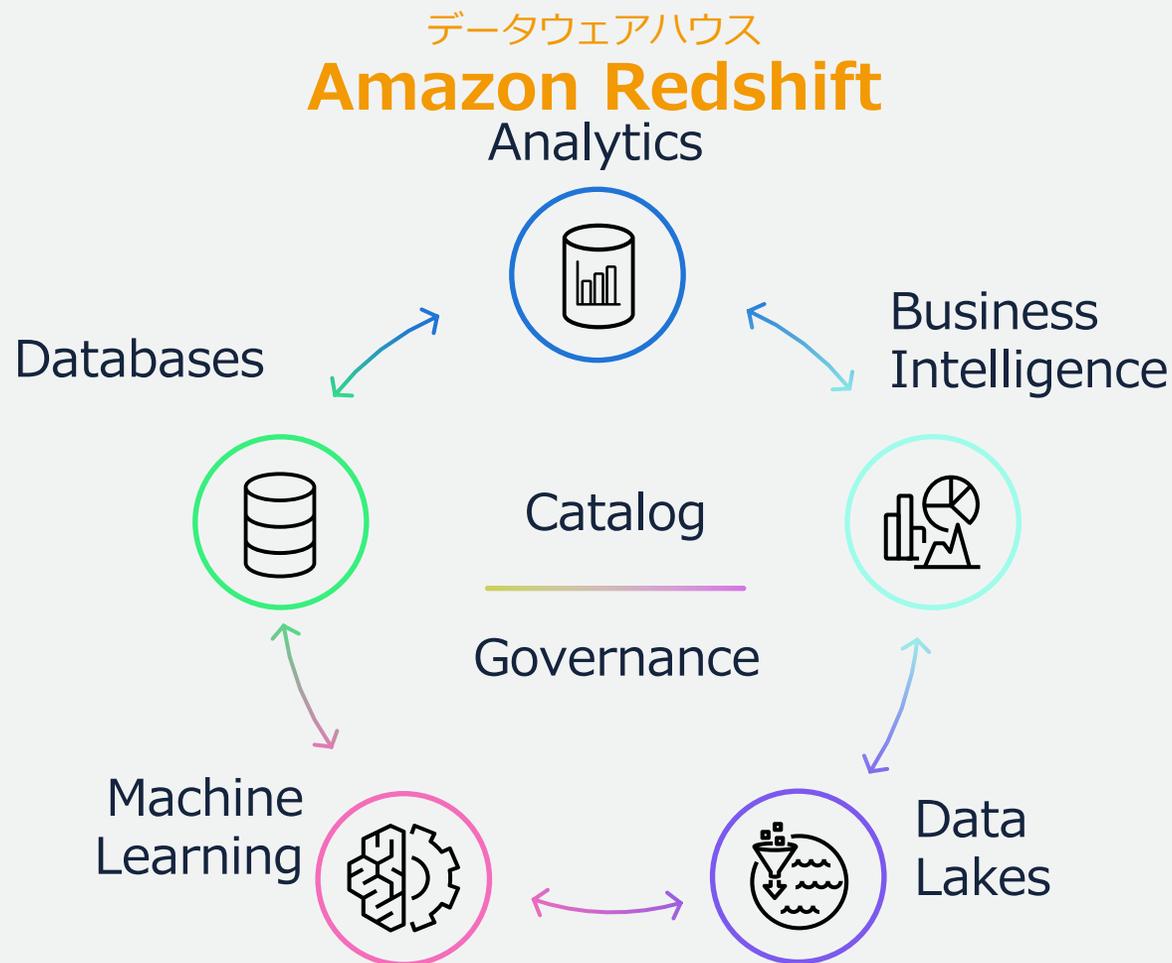
従来型 データ分析アーキテクチャ



- ソースシステムからETLツールでバッチデータ収集・変換
- データはデータウェアハウスに蓄積
- 必要に応じデータマート生成
- SQLクライアントによるアドホック分析やBI ツール等で可視化

Modern data architecture on AWS

AWS 上で実現する為の5つの柱



- ① サイロを無くし、単一のビューでデータにアクセス
- ② スケーラブルでハイパフォーマンスな基盤を低価格で提供
- ③ 機械学習でエンドユーザーに対して新たな体験を提供
- ④ セキュリティとガバナンスを満たすアクセス手法
- ⑤ 全てのユーザーが自由にデータを活用できる環境

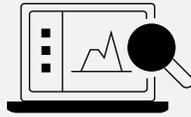
Amazon Redshift とは

Amazon Redshift

数万人の顧客を抱え、最も広く使用されているクラウドデータウェアハウス

Easy analytics for everyone

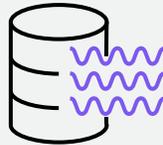
誰でも簡単に
分析できる



インフラストラクチャを気にすることなく、数秒でデータからインサイトを得ることに集中

Analyze all your data

全てのデータを
分析する



データサイロを打ち破り、あらゆるデータに対してデータの移動なしにリアルタイムな分析や予測分析が可能

Best price performance
at any scale

コストパフォーマンスの
高さ



他のクラウドデータウェアハウスに比べて最大 3 倍のコストパフォーマンスを実現し、動的にスケーリングして複雑で重要なワークロードのクエリ速度を向上

多様なユースケースに 適応可能な データウェアハウスの 選択肢



Amazon
Redshift
Provisioned

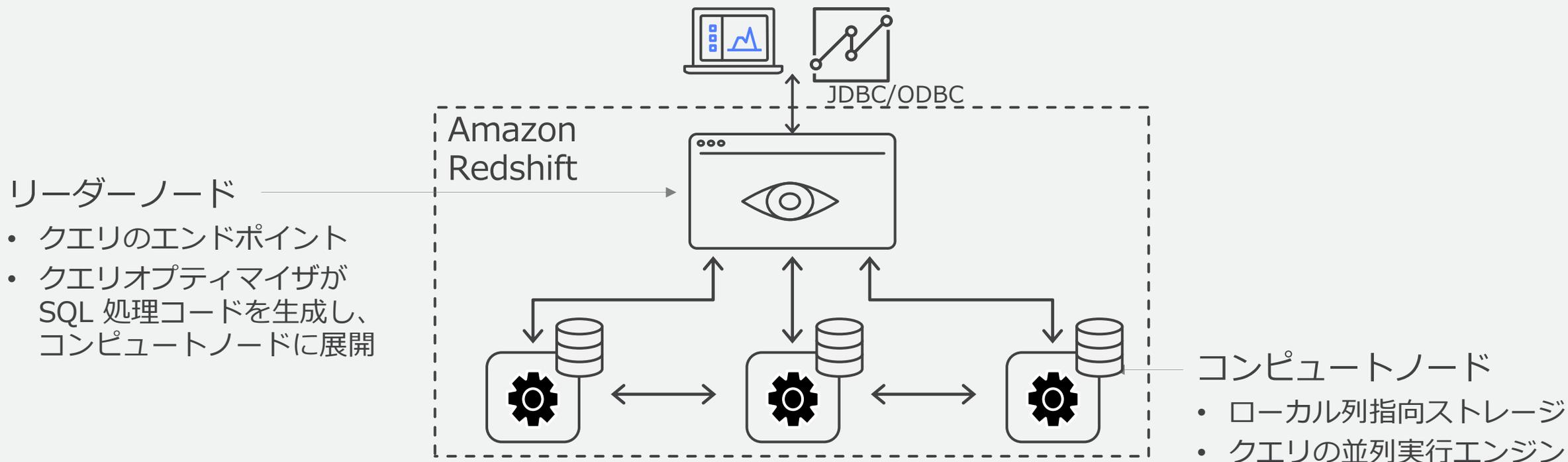
粒度の細かい制御とカスタム可能な
プロビジョニング



Amazon
Redshift
Serverless

インフラストラクチャ管理が不要で
簡単かつクイックにスタート可能

Amazon Redshift Provisioned のアーキテクチャ (DC2)



シェアードナッシング + MPP (Massively Parallel Processing) アーキテクチャ
データを複数のコンピューターノードに分散させ、各ノードで並列処理することで分析クエリに対する処理スループットを向上させるアプローチ

Amazon Redshift Provisioned のアーキテクチャ (RA3)

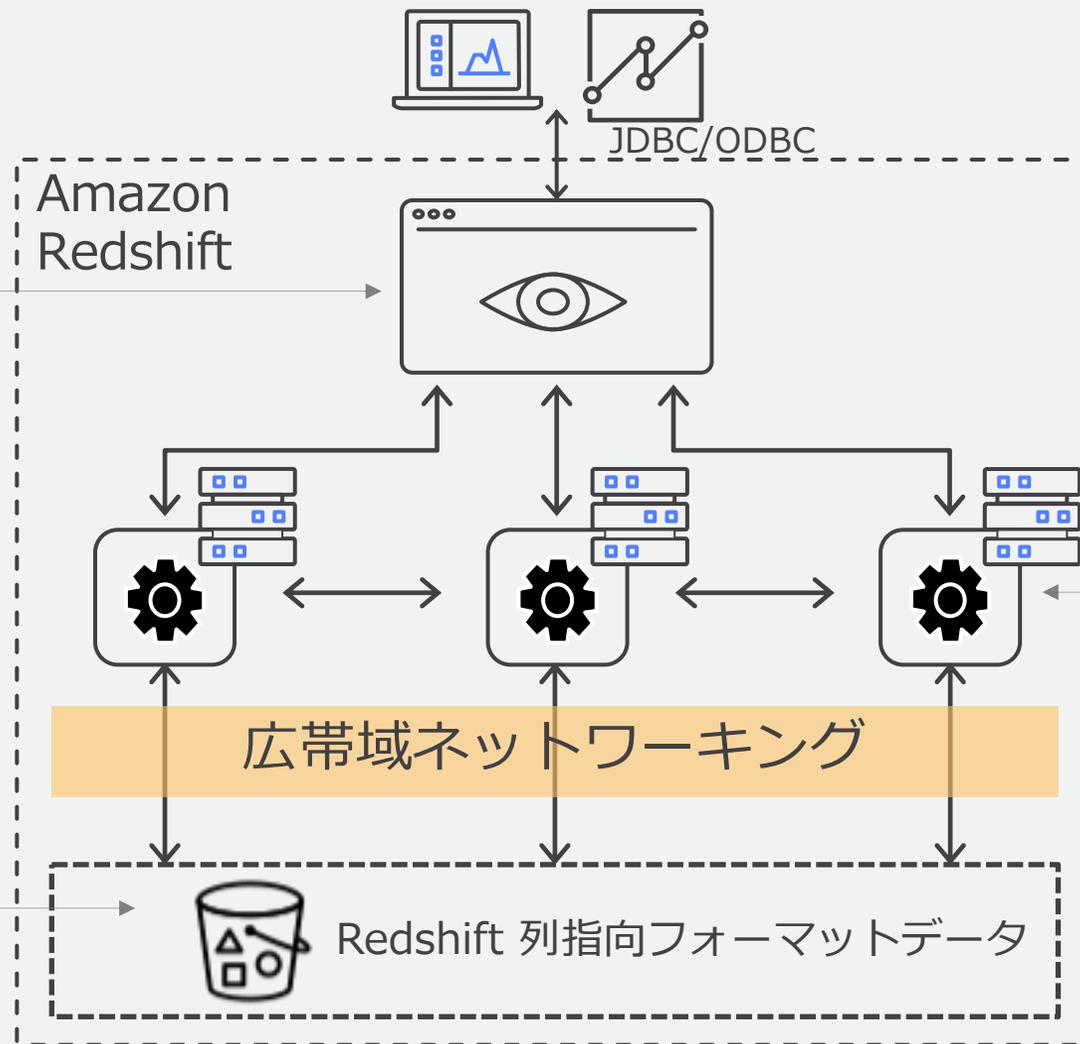
コンピューターノードとストレージの分離

リーダーノード

- クエリのエンドポイント
- クエリオプティマイザが SQL 処理コードを生成し、コンピューターノードに展開

マネージドストレージ

- Redshift 管理 S3 バケット
- データの永続ストレージ

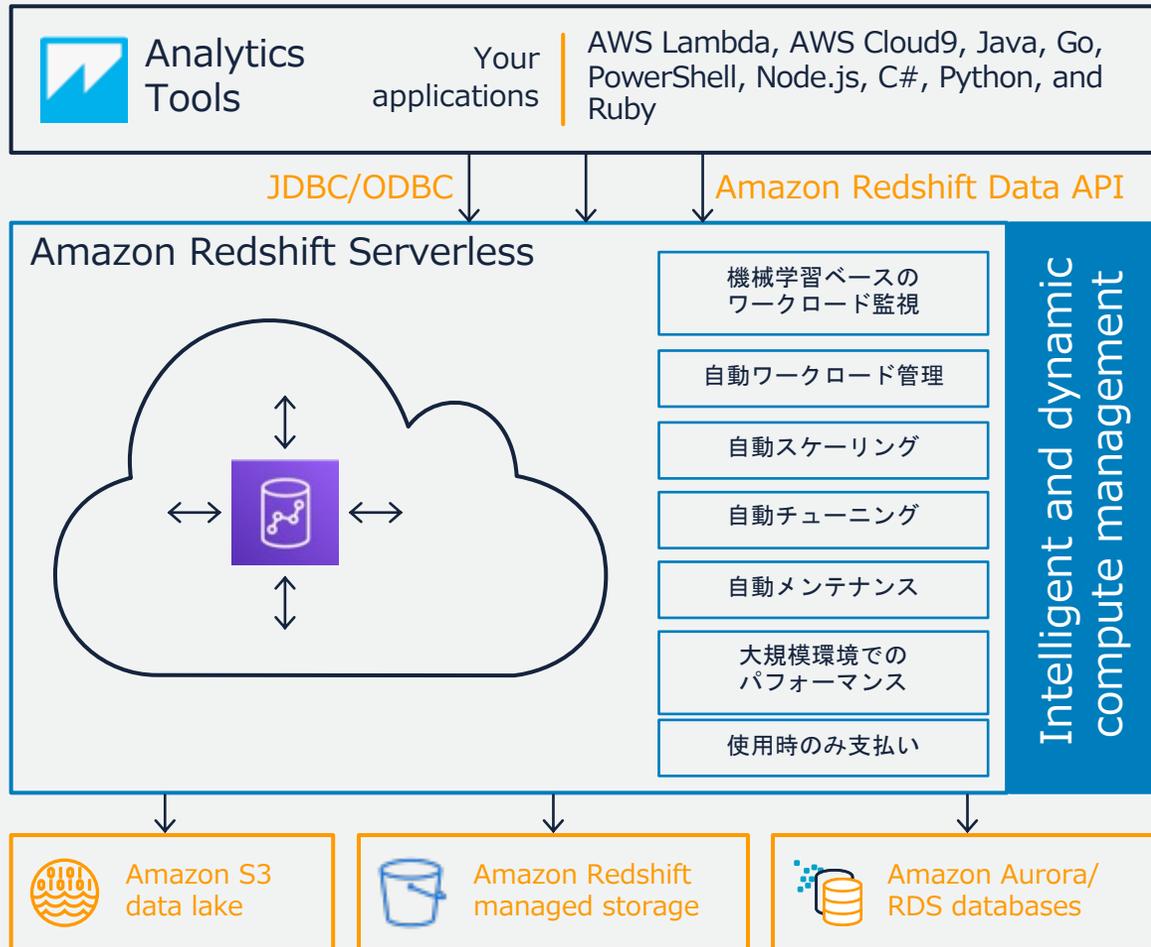


コンピューターノード

- 高速ローカル SSD キャッシュ + 大容量 RAM 搭載 + 広帯域ネットワーク
- クエリの並列実行エンジン

Amazon Redshift Serverless

クラスタの管理が不要となる新しい選択肢



ノードタイプ、ノード数、ワークロード管理、スケーリングなどを制御する必要はなく、サービスが自動実行。データを投入してすぐに分析処理を開始できる

Amazon Redshift Provisionedと同じ SQL が実行可能

処理能力はワークロードに応じて自動的にスケールアウト・ダウンし、非アクティブのときは自動停止

利用量に応じた従量課金で利用できる

東京、バージニア、オハイオ、オレゴン、フランクフルト、ほか15リージョンで一般提供 (大阪は未提供)

Amazon Redshift のユースケース

主に大容量データを高速に集計・分析する必要があるワークロードに活用



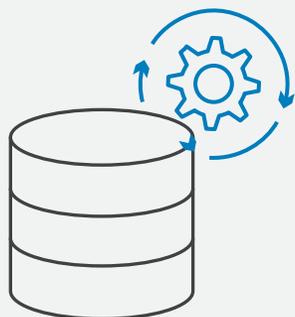
経営ダッシュボード



定型レポートニング



アドホック分析



ETL/バッチ

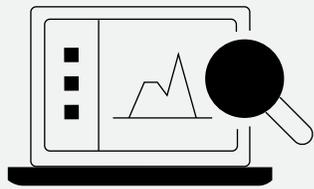


予測分析
(機械学習)



DaaS
(Data as a Service)

Amazon Redshift の 様々な機能



Easy analytics for everyone

誰でも簡単に
分析できる

Amazon Redshift Provisionedの開始ステップ

- 1画面の入力作業ですぐに起動
- ノード数やタイプは後から変更可能

クラスターを作成 情報

クラスターのノード構成

クラスター設定

クラスター識別子

これは、クラスターを識別する一意のキーです。

redshift-cluster-1

識別子は 1~63 文字にする必要があります。有効な文字は a~z (小文字のみ) および - (ハイフン) です。

このクラスターを何に使用する予定ですか?

本番稼働

高速で一貫したパフォーマンスを最高の料金で提供するよう設定します。

無料トライアル

Amazon Redshift について学習できるように設定します。期限が Amazon Redshift クラスターを作成したことがない場合、この設定は一定期間無料になります。

クラスターのサイズを選択する

を選択します。

選択のヘルプ

ノードの種類 情報

CPU、RAM、ストレージ容量、およびドライブタイプの要件を満たすノードの種類を選択します。

ra3.xlarge

ノードの数

必要なノードの数を入力します。

2

範囲 (2~32)

設定の概要 情報

ra3.xlarge | 2 個のノード

概算費用の確認

\$5,523.84/月

オンデマンドコンピューティング料金の見積もり

リザーブドノードを購入することで、コストの 60% 超を節約できます。詳細はこちら [🔗](#)

256 TB

最大圧縮ストレージ

RA3 は Redshift マネージド型ストレージにデータを格納します。各 RA3 ノードは、最大 64 TB の圧縮データ容量をマネージドストレージに取得し、最適なクエリパフォーマンスを確保します。

\$0.026/GB/月

推定ストレージ料金

RA3 クラスターの実行時に、マネージドストレージに保存したデータ量に対してのみ料金をお支払いいただけます。

追加設定 デフォルトを使用

これらの設定はオプションであり、クラスターの使用を開始するためにデフォルト設定が定義されています。これらの設定を今すぐ変更するには、[Use defaults] をオフにしてください。

ネットワーク

デフォルト VPC (vpc) と デフォルトサブネットを使用。

セキュリティ

default (sg-0433273d00f8be087) クラスターセキュリティグループの使用。

バックアップ

自動スナップショットは、約 8 時間ごと、またはデータ変更の 1 ノードあたり 5 GB ことのいずれか早いほうで作成されます。

設定

データベース暗号化なしの default.redshift-1.0 パラメータグループを使用。

メンテナンス

現在のメンテナンストラックを使用。

クラスターの詳細設定

キャンセル

クラスターを作成

Amazon Redshift Serverlessの開始ステップ

- デフォルトセッティングでサーバーレスエンドポイントを有効化してすぐに分析可能
- 名前空間（ストレージ）とワークグループ（コンピューター）の設定を行う
- ノード数やタイプなどは指定はない

名前空間 情報

名前空間は、データベースオブジェクトとユーザーのコレクションです。データプロパティには、データベース名とパスワード、許可、および暗号化とセキュリティが含まれます。

名前空間の名前
これは、名前空間を定義する一意の名前です。

名前は 3~64 文字である必要があります。有効な文字は、a~z (小文字のみ)、0~9 (数字)、-(ハイフン) です。

▼ データベース名とパスワード

データベース名
Amazon Redshift サーバーレス環境の最初のデータベースの名前。

名前は 1~64 文字の英数字 (小文字のみ) にする必要があります。予約語にすることはできません。

管理者ユーザーの認証情報
デフォルトの管理者ユーザー認証情報として提供される IAM 認証情報。新しい管理者のユーザー名とパスワードを追加するには、管理者ユーザーの認証情報をカスタマイズします。

管理者ユーザー認証情報をカスタマイズ
デフォルトの IAM 認証情報を使用するには、このオプションをオフにします。

ワークグループ 情報

ワークグループは、エンドポイントの作成元となるコンピューティングリソースのコレクションです。コンピューティングプロパティには、ネットワークとセキュリティの設定が含まれます。

ワークグループの名前
これは、ワークグループを定義する一意の名前です。

名前は 3~64 文字である必要があります。有効な文字は、a~z (小文字のみ)、0~9 (数字)、-(ハイフン) です。

▼ ネットワークとセキュリティ

Virtual Private Cloud (VPC)
この VPC は、このデータベースの仮想ネットワーク環境を定義します。

VPC セキュリティグループ
この VPC セキュリティグループは、VPC で使用できるサブネットと IP 範囲を定義します。

サブネット
指定されたデータベースに関連付けられている、選択した VPC のサブネット。

拡張された VPC のルーティング
このオプションをオンにすると、サーバーレスデータベースとデータリポジトリ間のネットワークトラフィックがインターネットではなく VPC 経由でルーティングされます。

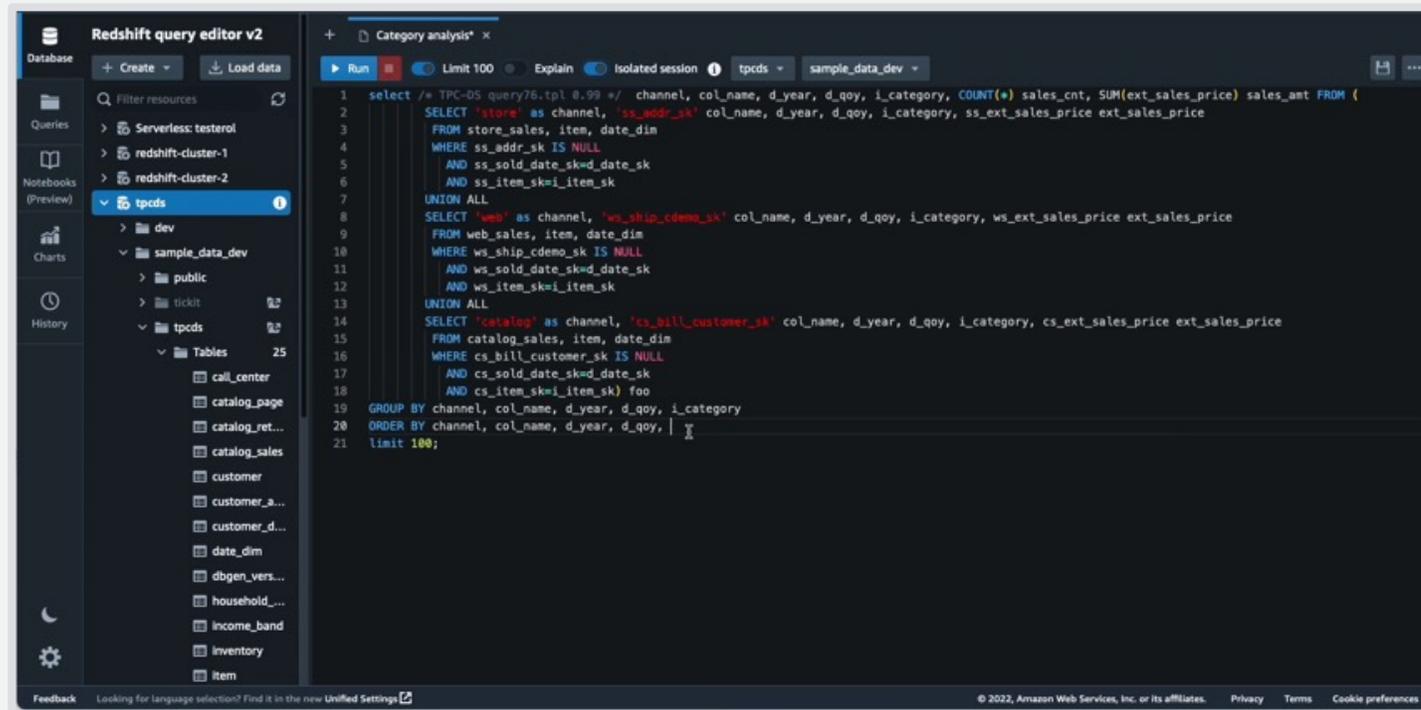
拡張された VPC ルーティングをオンにする

キャンセル

設定を保存

Amazon Redshift query editor v2

- ✓ 事前セットアップ不要ですぐに使える
Web ベースの新しいクエリエディタ
- ✓ Amazon Redshift の管理コンソールへのアクセス権限を必要とすることなく、SSO で簡単にアクセス
- ✓ GUI 操作でテーブルの作成やS3 もしくはローカルファイルのデータを簡単にロード可能
- ✓ SQL 文を実行したり、クエリ結果のグラフ化を行うことも可能
- ✓ Markdown形式と複数SQL入力に対応したSQL Notebooks もサポート



自動化されたパフォーマンスチューニング

機械学習ベースの自動最適化により人手によるチューニングを最小限に

物理データ配置やその最適化を自動化

データやワークロードの増加に応じて
ピーク性能を最適化

機械学習を活用し、変化するワークロード
に適応

パフォーマンスチューニングの自動化



Automatic
vacuum
delete



Automatic
distribution keys



Automatic
sort keys



Auto
workload
manager



Automatic
table sort



Automatic column
compression
encoding



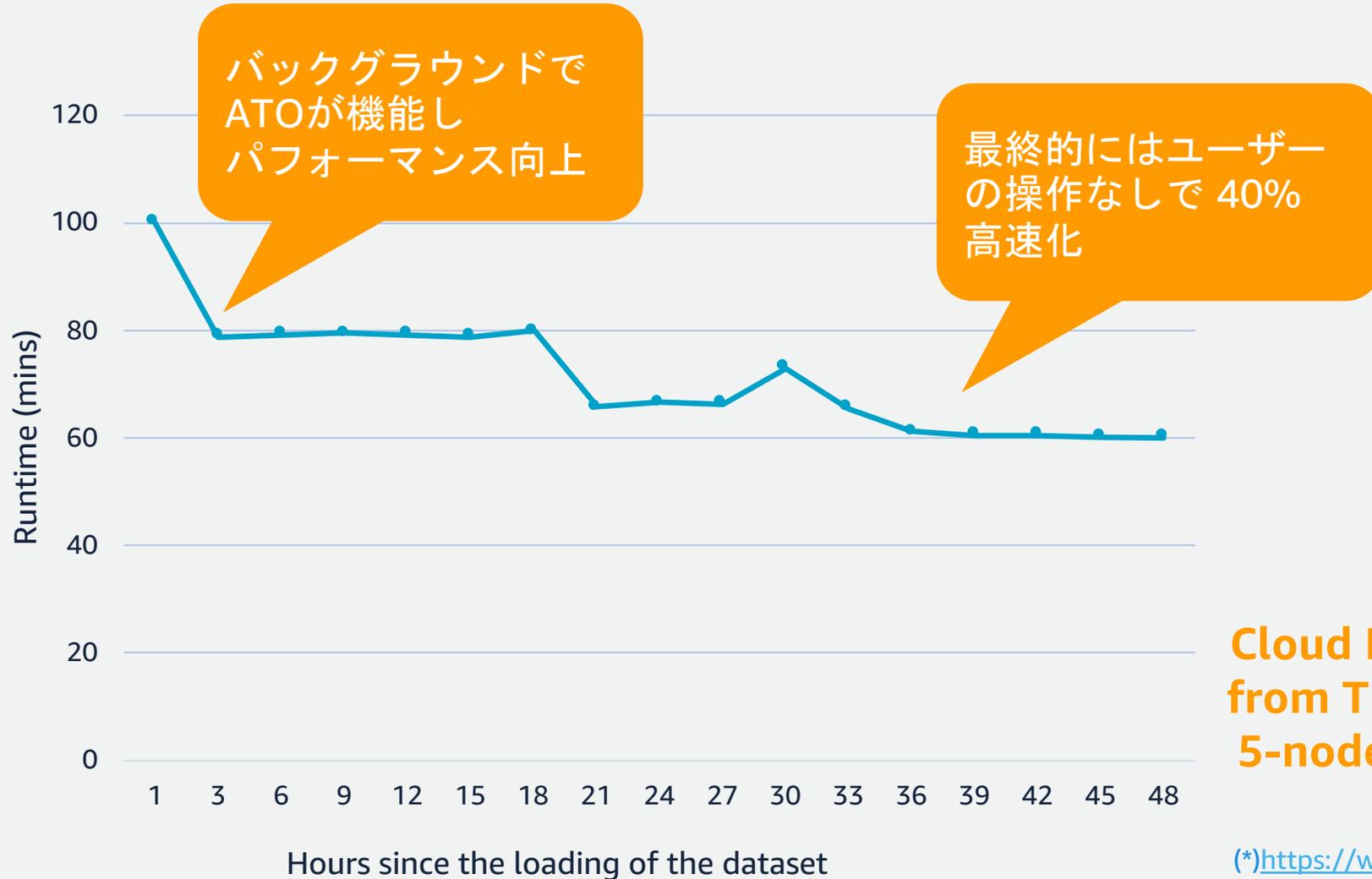
Auto Analyze



Auto refresh & rewrite
Materialized Views

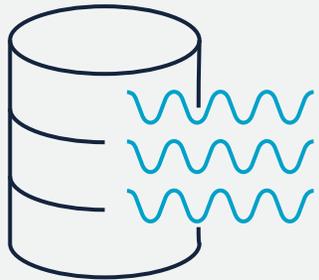
自動テーブル最適化

(Automatic Table Optimization: ATO)



Cloud DW derived from TPC-H* 30TB 5-node RA3.16XL

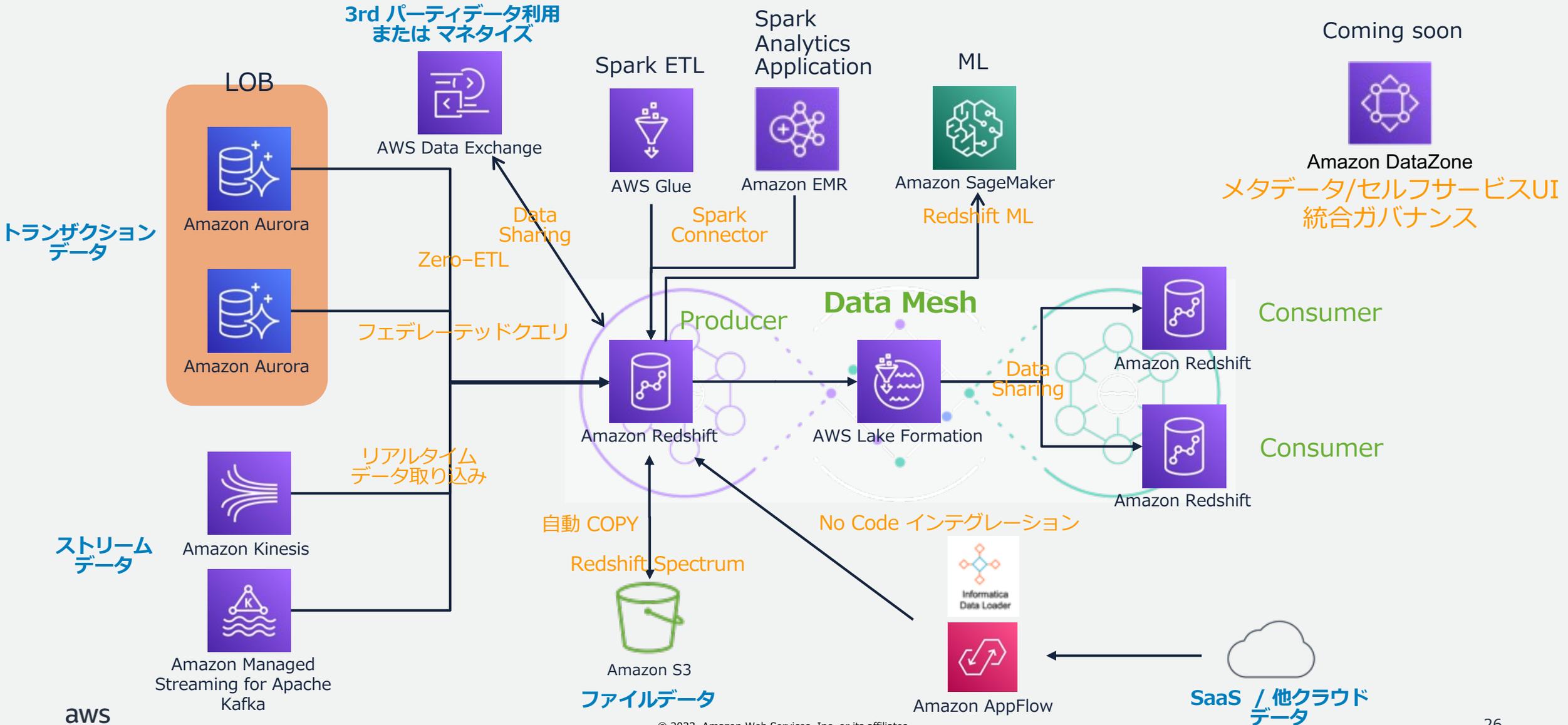
(*)<https://www.tpc.org/tpch/>



Analyze all your data

全てのデータを
分析する

Redshift を中心とした次世代データ基盤 構成例



Amazon Redshift Spectrum

データレイクへのクエリ

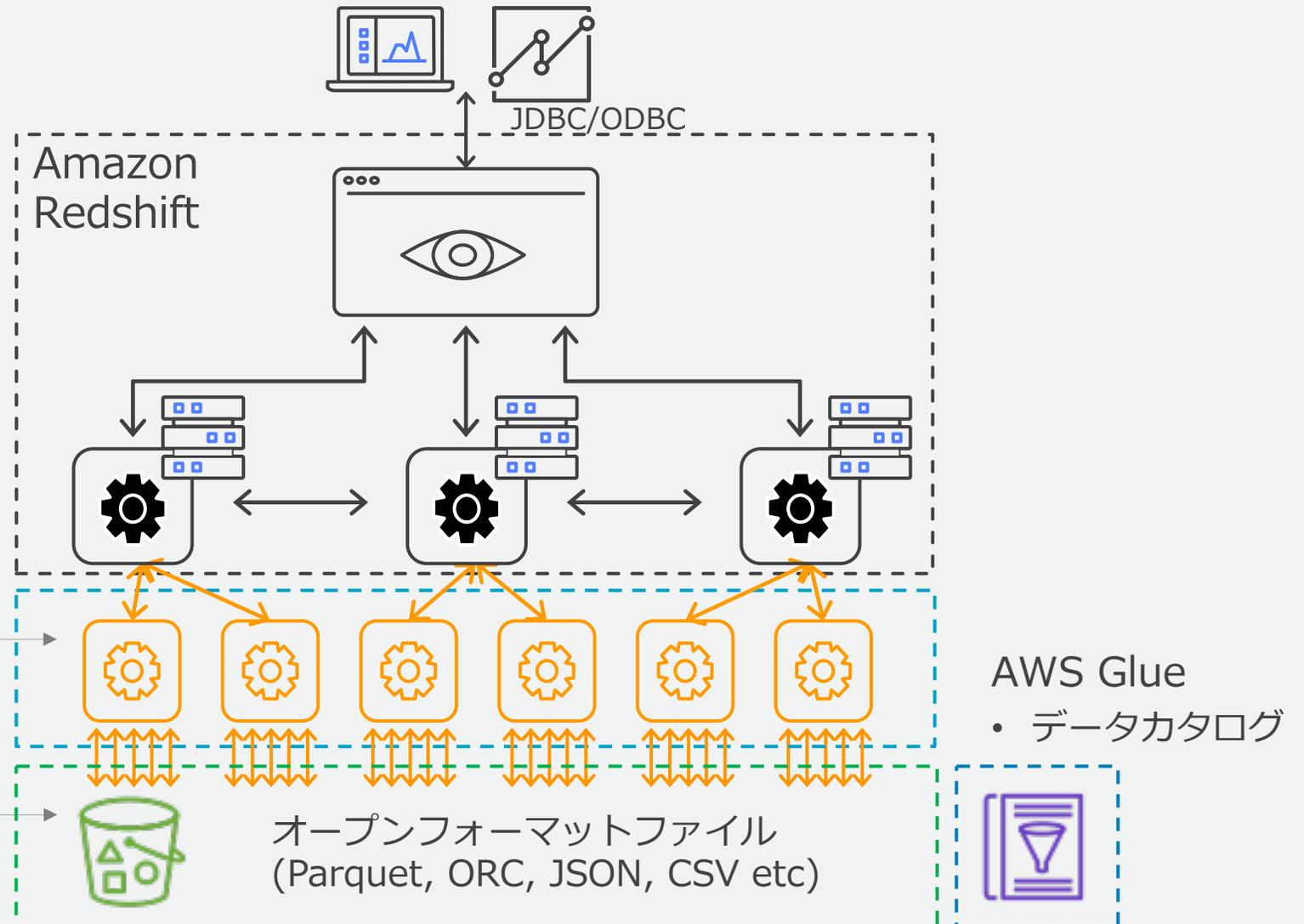
- ✓ データロードせずに S3 上のオープンフォーマットデータへ直接クエリ可能
- ✓ エクサバイト級データへも、高速なデータ取り出しと一貫したパフォーマンスを実現

Amazon Redshift Spectrum

- S3 上のファイルに対する並列クエリ実行エンジン

データレイク

- ユーザー管理 S3 バケット



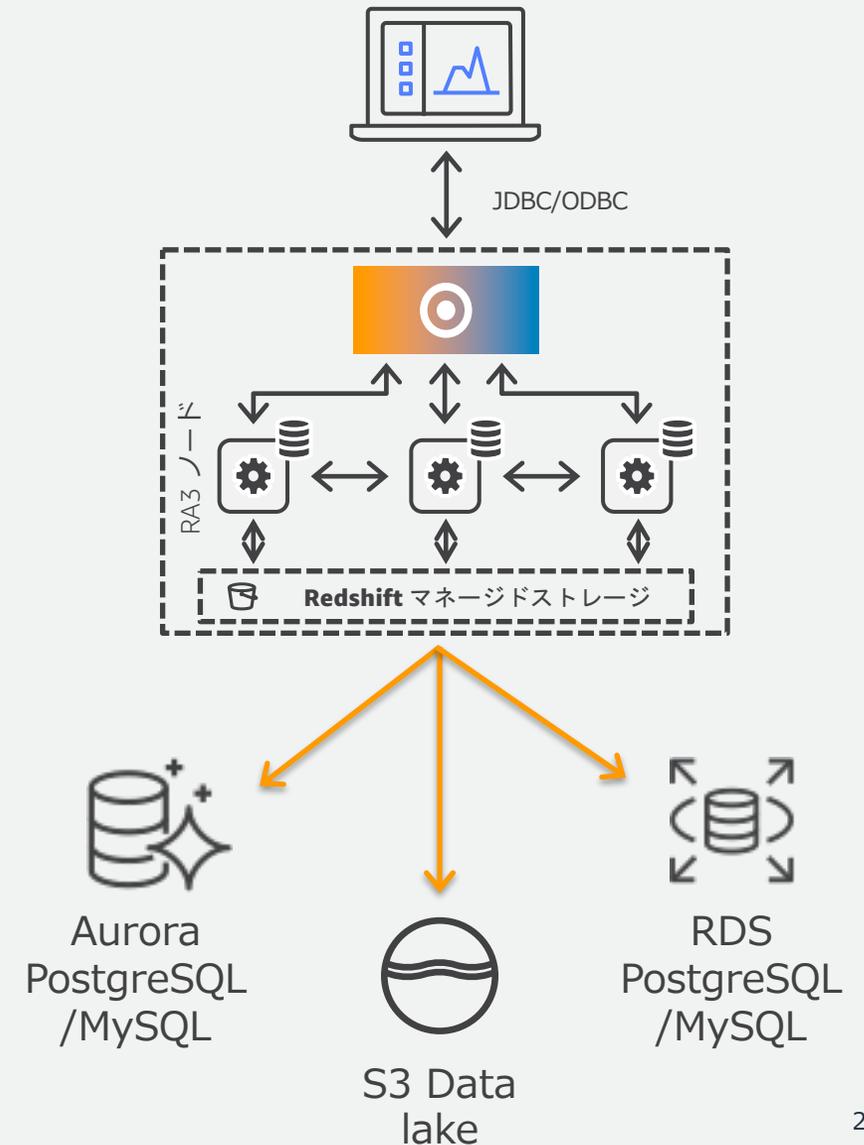
フェデレーテッドクエリ

RDS/Aurora の PostgreSQL/MySQL に対して
直接クエリ可能

データ移動なしにライブデータを分析

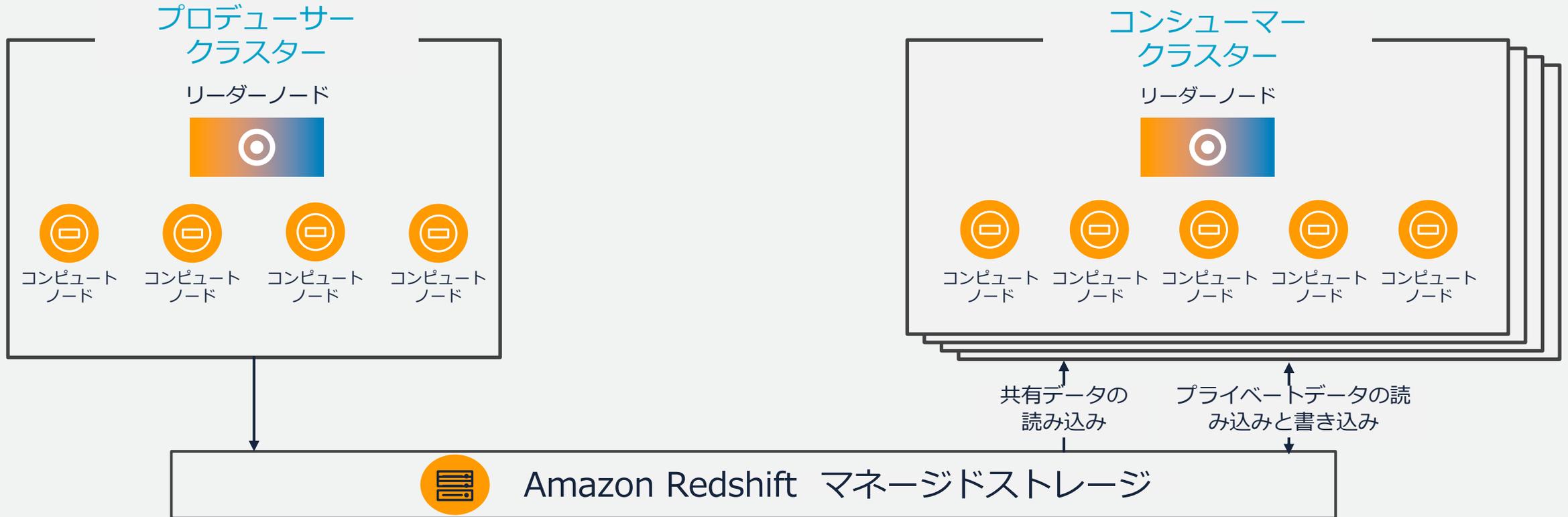
データウェアハウス、データレイク、
オペレーショナルデータベースのデータを
統合して分析

高い性能でセキュアにデータアクセス



Amazon Redshift Data Sharing

Redshift クラスター間でセキュアに簡単にデータを共有することが可能

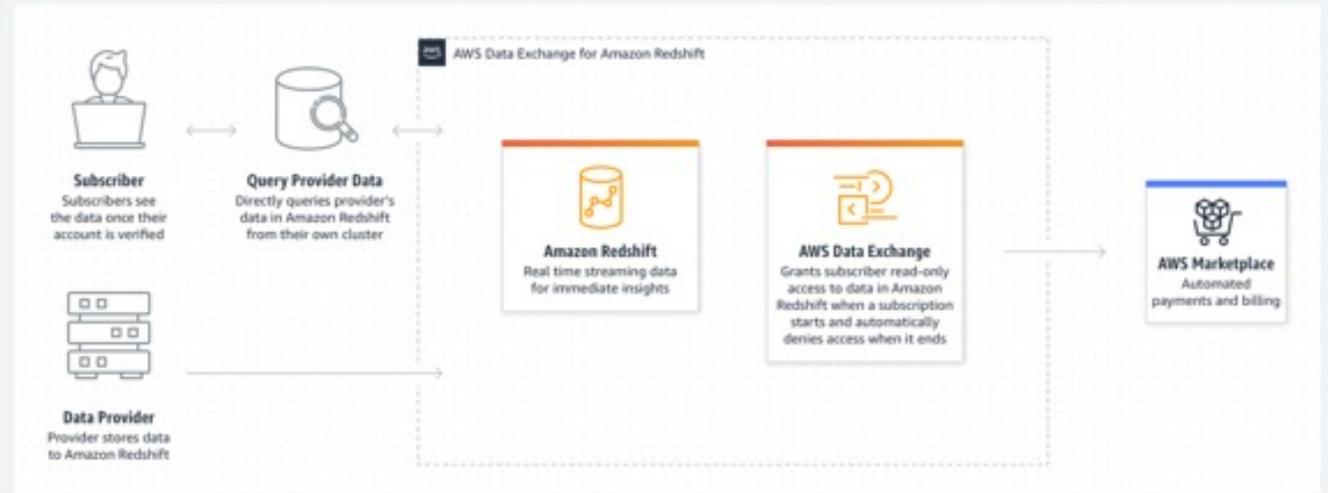


- 共有データにアクセスするプロデューサー/コンシューマーそれぞれのワークロードを分離
- プロデューサー/コンシューマーごとに個別に料金をお支払い
- プロデューサーを一時停止してもコンシューマーから共有データにアクセス可能
- Amazon Redshift Serverless, Provisioned (RA3)でサポート
- クロスアカウント、クロスリージョンでのデータ共有も可能。

AWS Data Exchange for Amazon Redshift

3rdパーティデータをETLなしで容易にインテグレーション

- ✓ お客様データをサブスクリプション型で提供/購読可能とする
AWS Data Exchange と統合
- ✓ AWS Data Exchange で検索/購読したデータを、ETLなしで自社データと容易に組み合わせて利用可能
- ✓ AWS Data Exchangeで連携されるデータは、購読期間のみアクセス許可
- ✓ 購読に関わるエンタイトルメント・課金・支払いは、AWS Marketplaceを通じて自動化



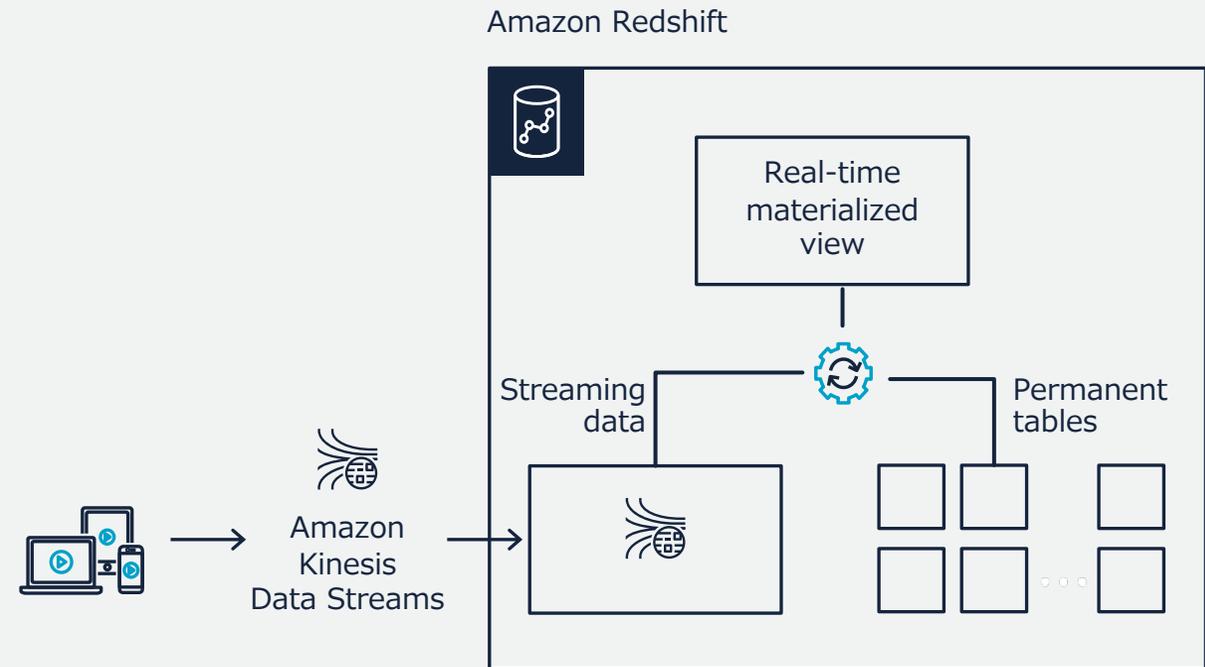
<https://aws.amazon.com/redshift/features/aws-data-exchange-for-amazon-redshift/>

※Amazon Redshift Serverless, Provisioned (RA3)でサポート

Amazon Redshift Streaming Ingestion

ストリーミングエンジンとのネイティブな統合で、素早く簡単にデータ挿入

- ✓ Kinesis Data Streams (KDS) および Amazon Managed Streaming for Apache Kafka (MSK) からのストリーミング取り込みをサポート
- ✓ Amazon S3 でステージングする必要なく、数百MB/秒の取り込みを秒単位のレイテンシーで実現
- ✓ KDSまたはMSKからデータを取り込み、マテリアライズド・ビューを直接作成するパイプラインを、SQLで作成可能



Amazon S3 からの自動 COPY

S3 に新規追加されたオブジェクトを ETL パイプラインなしに Redshift へ自動 COPY



COPY JOB 設定例

```
COPY public.target_table
FROM 's3://mybucket-bucket/staging-folder'
IAM_ROLE 'arn:aws:iam::123456789012:role/MyLoadRoleName'
JOB CREATE my_copy_job_name
AUTO ON;
```

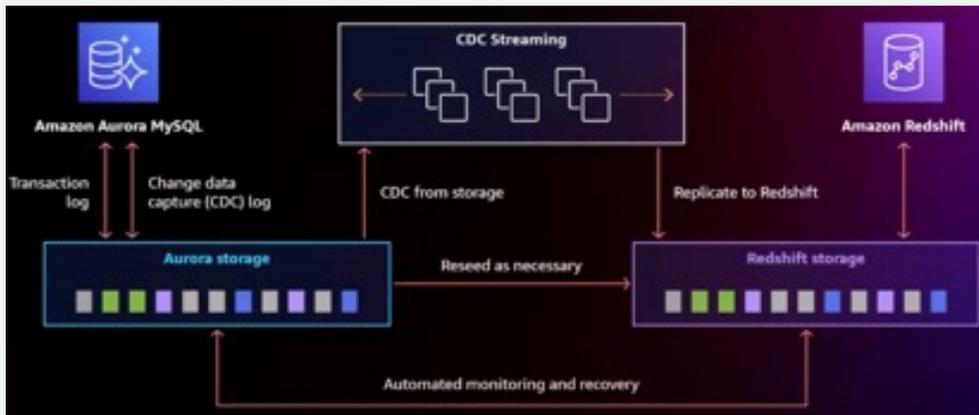
- SQL で COPY JOB を設定すると、S3 への新規ファイル格納をトリガーに指定されたテーブルへ自動 COPY
- COPY JOB の実行状況はシステムテーブルで確認可能
- 外部の ETL サービス不要でイベントドリブンのデータ連携を容易に実現できる
- 対応リージョン：
 - US East (N. Virginia), US East (Ohio), US West (Oregon), Asia Pacific (Tokyo), Europe (Ireland) and Europe (Stockholm).
- 2023.5.1までプレビュー実施中

Amazon Aurora との Zero-ETL 統合

Aurora に蓄積されたデータを ETL パイプラインなしにニアリアルタイムで Redshift で分析



- ETL パイプライン不要で複数の Aurora (MySQL) データベースからニアリアルタイム連携
- Amazon Aurora に書き込まれてから数秒以内に Redshift に複製
- Redshift ML、マテリアライズドビュー、データ共有、フェデレーテッドアクセスなどを活用して、Aurora のトランザクションデータを分析
- ストレージレイヤー統合により伴う DB インスタンスへのパフォーマンス影響なくデータ統合を実現
- バージニア北部リージョンにて限定プレビューを実施中（要申し込み）



Data Sharing の AWS Lake Formation 統合

データメッシュにおける一元化されたアクセス制御



- Redshift data sharing のアクセス制御を Lake Formation で一元管理可能に
- 行レベルと列レベルのきめ細かい権限管理により、セキュリティとガバナンスを強化
- 手動のスク립ト作成や複雑なクエリは不要
- ポリシーを一度定義すれば、複数のコンシューマークラスタに対して一貫して適用可能
- 対応リージョン：
 - Asia Pacific (Tokyo), US East (Ohio), US East (N. Virginia), US West (Oregon), Europe (Ireland), Europe (Stockholm)

Table permissions

Table permissions
Choose specific access permissions to grant.

<input checked="" type="checkbox"/> Select	<input type="checkbox"/> Insert	<input type="checkbox"/> Delete	<input type="checkbox"/> Super
<input type="checkbox"/> Describe	<input type="checkbox"/> Alter	<input type="checkbox"/> Drop	This permission is the union of all the individual permissions to the left, and supersedes them.

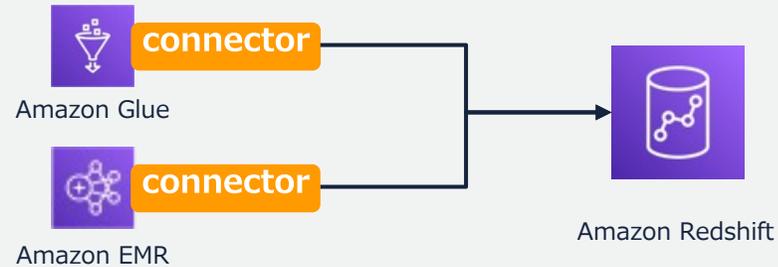
Grantable permissions
Choose the permission that may be granted to others.

<input type="checkbox"/> Select	<input type="checkbox"/> Insert	<input type="checkbox"/> Delete	<input type="checkbox"/> Super
<input type="checkbox"/> Describe	<input type="checkbox"/> Alter	<input type="checkbox"/> Drop	This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

<https://aws.amazon.com/jp/blogs/big-data/centrally-manage-access-and-permissions-for-amazon-redshift-data-sharing-with-aws-lake-formation/>

Apache Spark インテグレーション

Spark アプリケーションから Redshift のデータへ高速アクセス



PySparkによる読み書き例

```
import boto3
from pyspark.sql import SQLContext

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshifhost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .load()

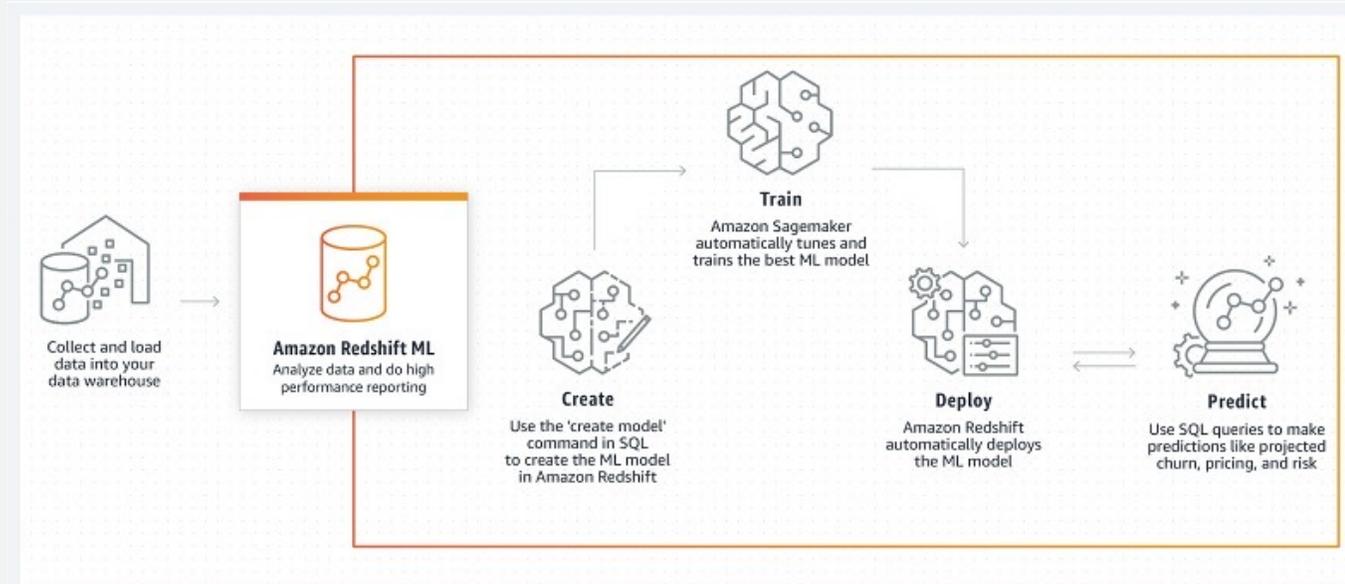
df.write \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName_copy") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .mode("error") \
    .save()
```

- パフォーマンスとセキュリティが強化され、従来のコネクタと比較しパフォーマンスが最大で10倍高速化するオープンソースプロジェクトのコネクタ
 - <https://github.com/spark-redshift-community/spark-redshift>
- 任意の Notebook 環境または Spark Job から、data frame / Spark SQL を利用して Redshift のデータを利用した Spark アプリケーションを構築可能
 - ※ Spark アプリケーションは Java , Scala , Pythonなど任意の言語で
- Amazon EMR および AWS Glue はこのコネクタを標準搭載
 - ※ 対応バージョンは EMR 6.9, EMR Serverless, Glue 4.0
- IAM ベース認証によるセキュリティ向上

Amazon Redshift ML

SQL 経由で Amazon SageMaker と連携し、機械学習モデルの作成・トレーニングが可能

- ✓ ユースケース：製品のリコメンデーション、不正防止、顧客離反の削減など
- ✓ SQL で機械学習モデルの作成、トレーニング、デプロイ
- ✓ 機械学習アルゴリズムは自動選択 または XGBoost/K-MEANS/MLP/Linear Learnerを指定可能
※カスタムモデルインポート (BYOM) もサポート
- ✓ モデルの前処理、作成、トレーニング、デプロイを自動で実行
- ✓ 推論モデルを Amazon Redshift 上にデプロイし SQL ステートメントの一部としてユーザー定義関数を呼び出すように推論を実行可能

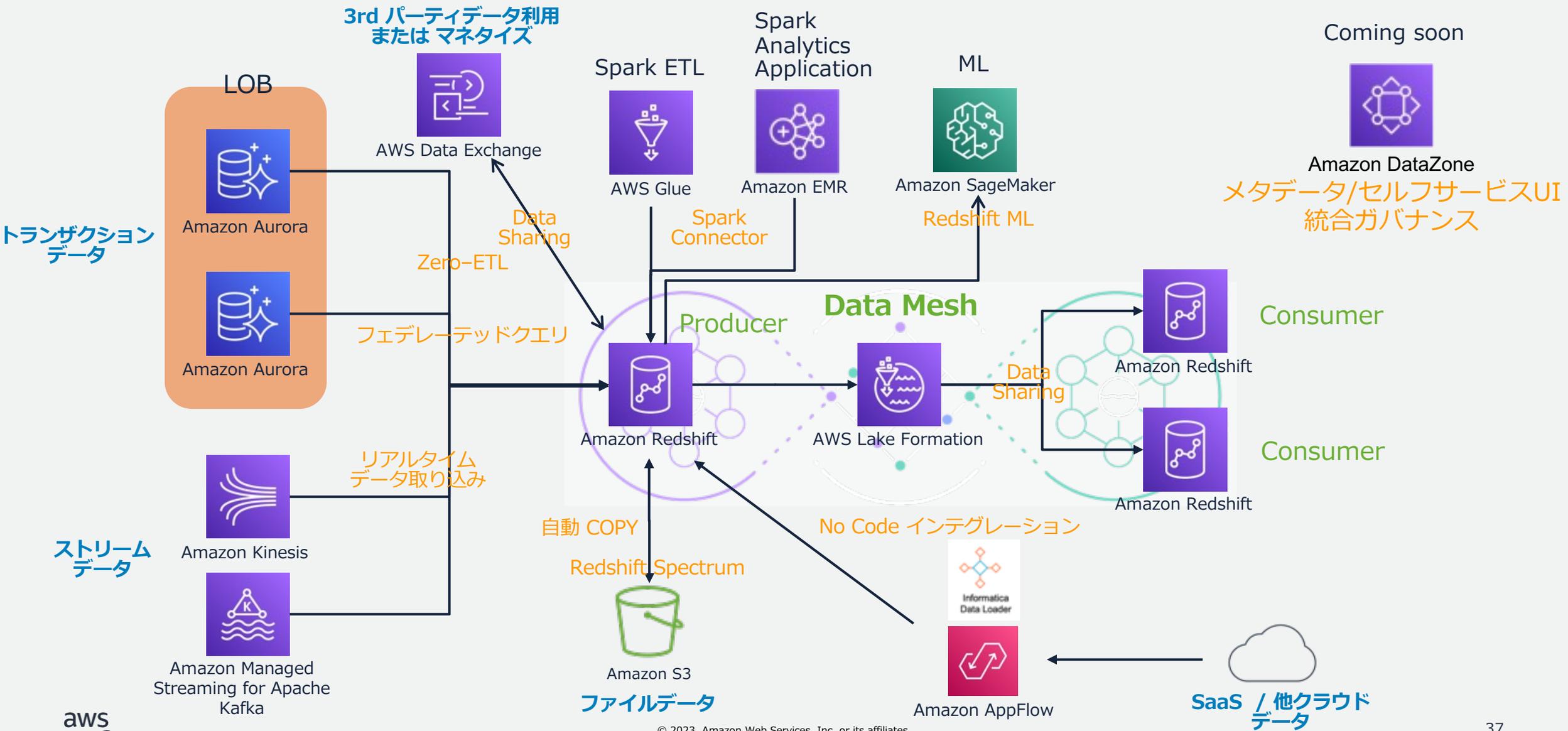


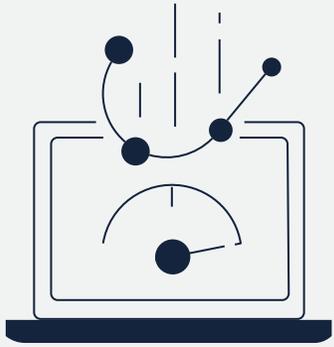
```
CREATE MODEL customer_churn
FROM (SELECT c.age, c.zip,
c.monthly_spend, c.monthly_cases,
c.active AS label
FROM customer_info_table c)
TARGET label
FUNCTION predict_customer_churn
```

<https://aws.amazon.com/jp/about-aws/whats-new/2021/05/aws-announces-general-availability-of-amazon-redshift-ml/>



[再掲]Redshift を中心とした次世代データ基盤 構成例





Best price performance at any scale

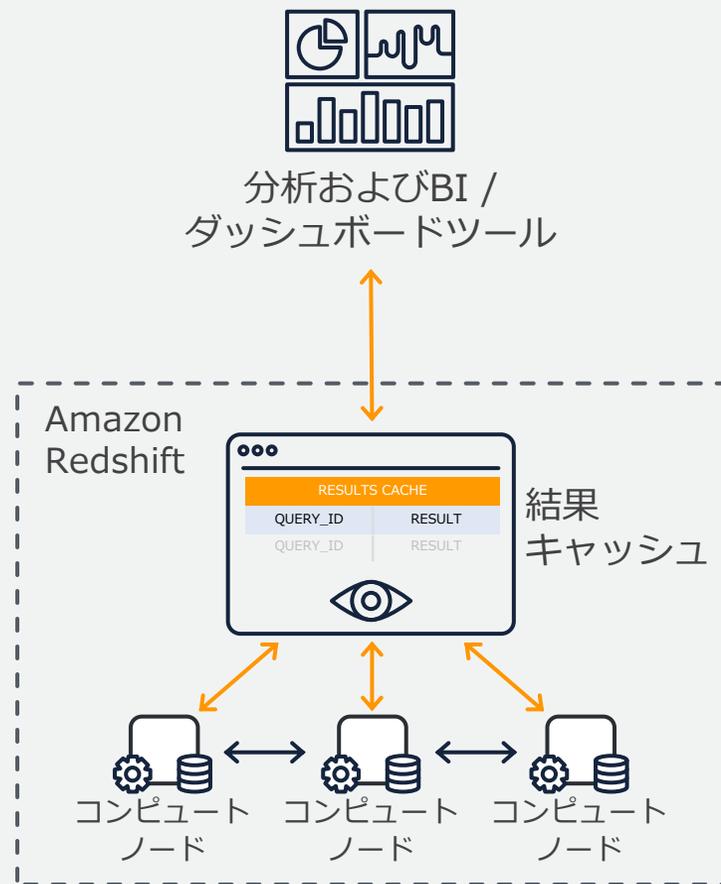
コストパフォーマンス の高さ

高いコストパフォーマンスを実現する Redshift の特長

1. 分析ワークロードに特化したアーキテクチャデザイン
 - 列指向ストレージとクエリオプティマイザ
 - 複数のコンピューターノードによる SQL の並列分散実行
 - 結果キャッシュにより、繰り返し実行されるクエリを 1 秒未満で実行
2. クエリスループットを最大化するため、常にワークロードを最適化
 - ベースライン性能が自動的にかつ継続的に向上していく
 - テーブルデザインやメンテナンス処理、ワークロード管理の自動最適化
 - 実行時間の短いクエリのスループットを高速化
3. クエリ特性に合わせたチューニングも可能
 - 分散、圧縮、ソート、マテリアライズドビュー
 - リソース優先度制御など、豊富なワークロードマネージメント機能

結果セットのキャッシュ(リザルトキャッシュ)

繰り返し実行されるクエリを 1 秒未満で実行



結果セットのキャッシュ機能

- 1 クエリはリーダーノードにて受付
- 2 リーダーノード内のキャッシュにクエリ結果が含まれている場合、コンピュートノード上での処理を伴わずに返される
- 3 クエリ結果がキャッシュに存在しない場合、コンピュートノード上でクエリが実行されて、その結果がキャッシングされる

結果セットのキャッシュ機能によって Amazon Redshift クラスターの処理に余裕が生じて、クエリ全般でパフォーマンスが向上する

マテリアライズドビューによるクエリ高速化

頻繁に実行するクエリパターンを高速化

- 結合、フィルタ、集計、射影
- ベーステーブルとは異なる分散キーやソートキーを指定できる
- 外部表の参照も可能

ETL/BI パイプラインの簡素化

- データは自動でリフレッシュされる

クエリリライト

- アプリケーションからはテーブルに対してクエリするだけで必要に応じて実行計画をRewrite し、マテリアライズドビューを参照

Auto MV : 自動マテリアライズドビュー

- Redshiftが継続的にワークロードをモニターし、機械学習によりパフォーマンス面で最適なマテリアライズドビューを自動的に生成・管理

マテリアライズドビュー

loc_sales	
loc	total_sales
SF	12.00
NY	10.00

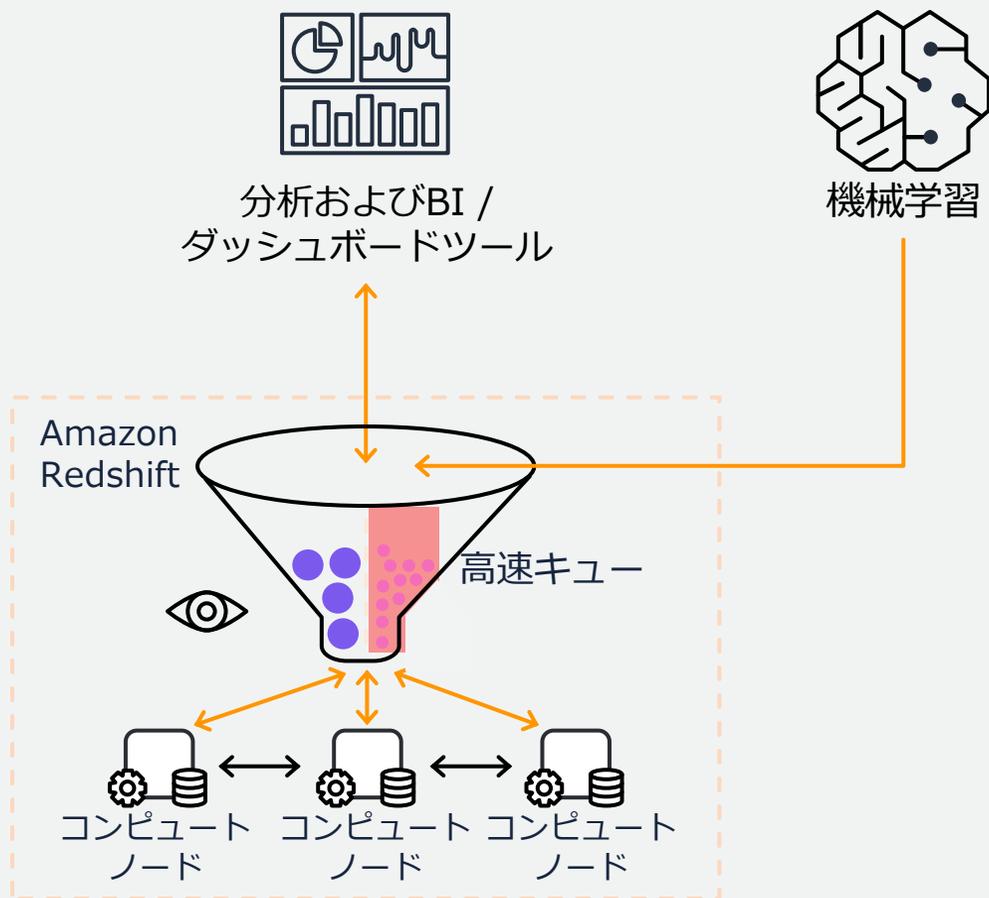
“地域ごとの
総売上は?”

store_info		
store	owner	loc
s1	Joe	SF
s2	Ann	NY
s3	Lisa	SF

sales			
item	store	cust	price
i1	s1	c1	12.0
i2	s2	c1	3.0
i3	s2	c2	7.0

ショートクエリアクセラレーション (SQA)

実行時間の短いクエリは、通常のキューとは別の領域で処理



SQA (Short Query Acceleration) の機能

- 1 機械学習によってクエリの実行時間を予測する
- 2 ショートクエリと判断されたクエリは専用の高速キューにルーティングされる
- 3 リソースはショートクエリのために動的に確保される

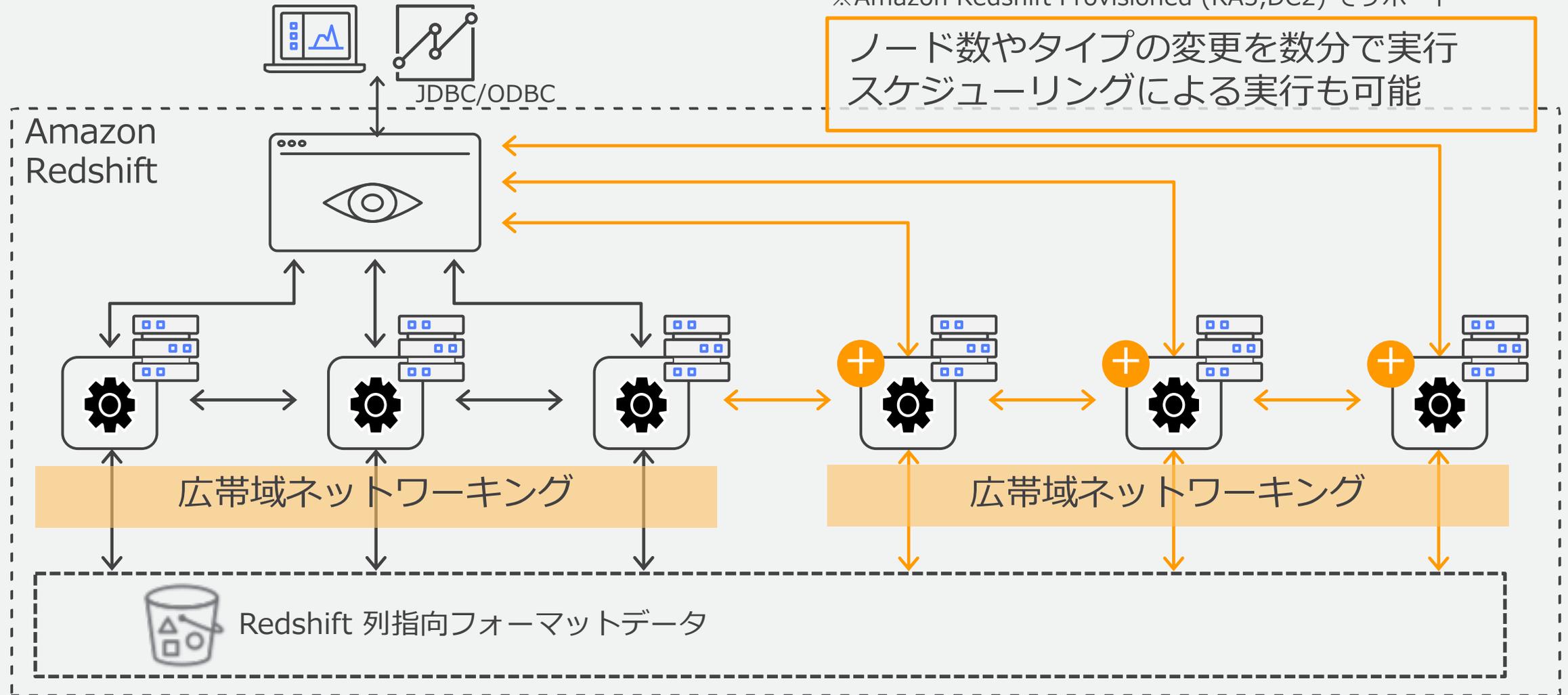
ショートクエリの実行がヘビークエリによって邪魔をされないための仕組み

Elastic Resize (伸縮自在なサイズ変更)

必要な性能に合わせてノード構成を変更

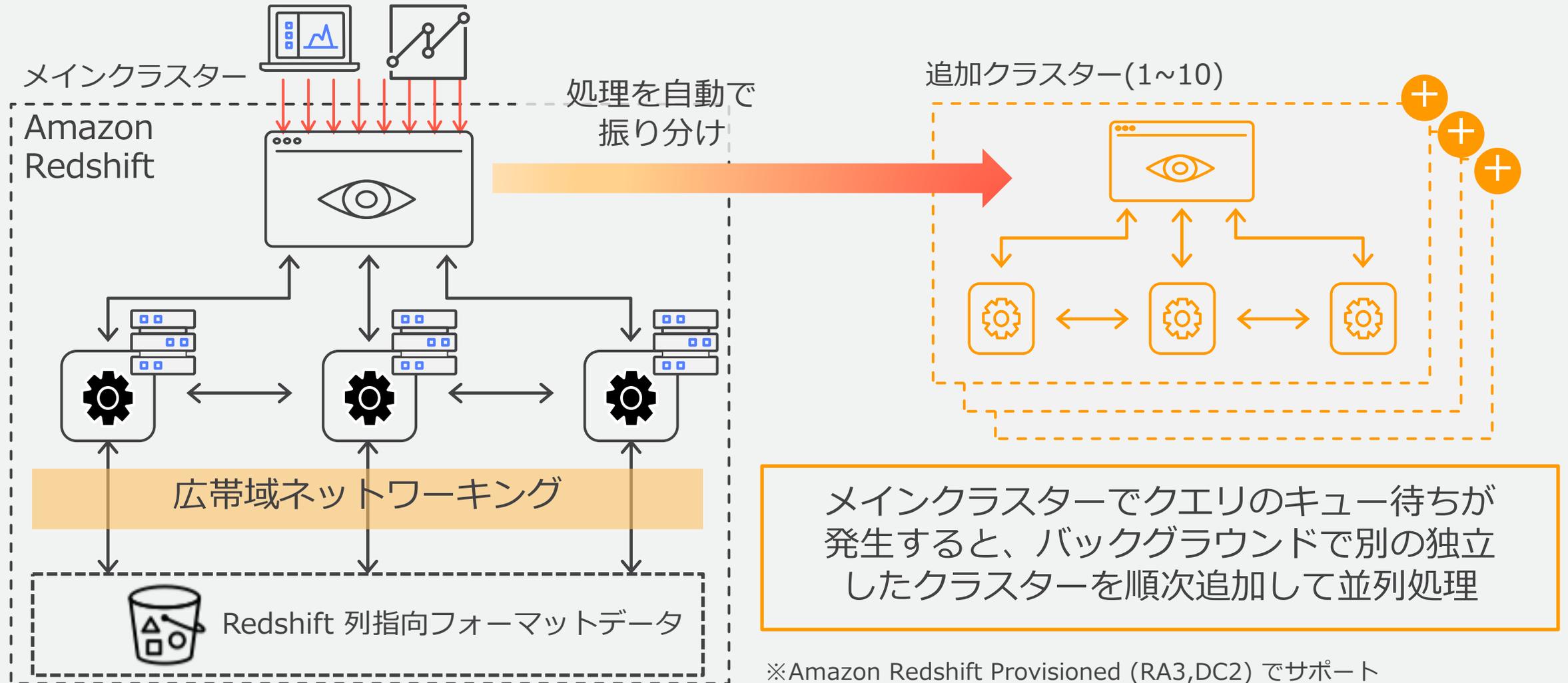
※Amazon Redshift Provisioned (RA3,DC2) でサポート

ノード数やタイプの変更を数分で実行
スケジューリングによる実行も可能



同時実行スケーリング

ピーク時にコンピュートを自動拡張

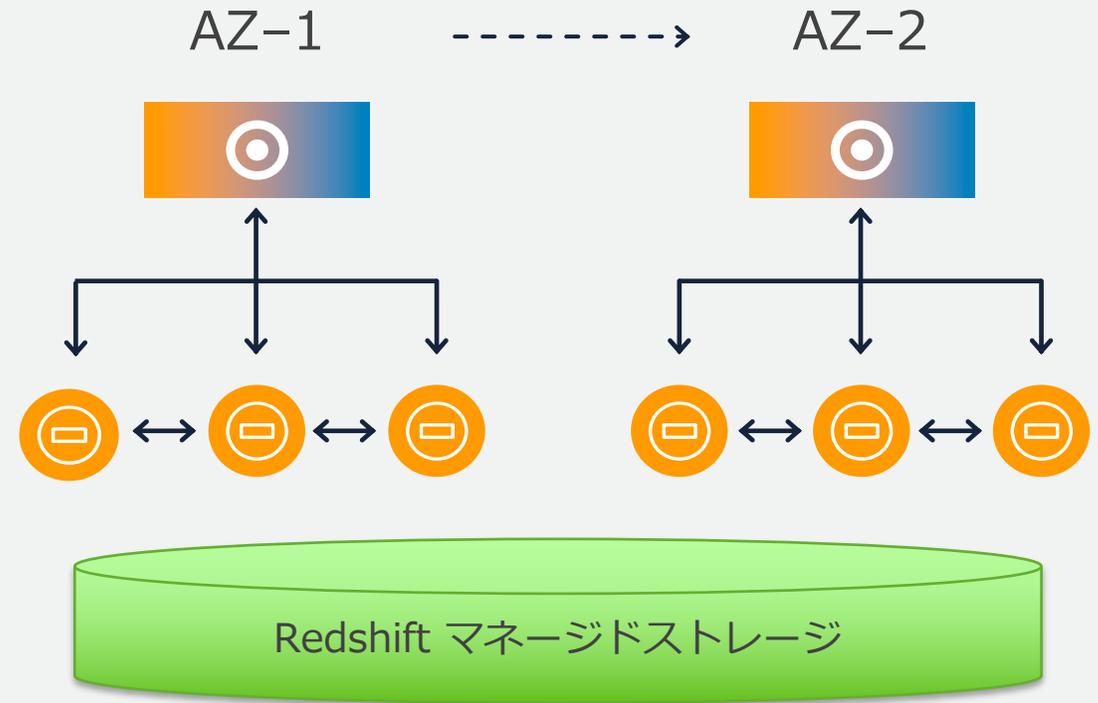


可用性とセキュリティ

クラスター再配置

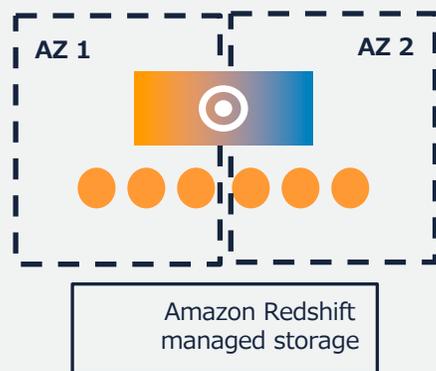
クラスターを別の AZ にフェールオーバー可能

- ✓ データロスなしにリカバリー (RPO = 0)
- ✓ スナップショットからのリストア不要
- ✓ オンデマンドフェールオーバー
- ✓ クラスターは別の AZ にオンデマンドで作成されるため、スタンバイレプリカクラスター不要
- ✓ RA3 インスタンスでのサポート



Multi-AZ 配置

ビジネスクリティカルアプリケーション向けの高可用性と耐障害性



- 1つのエンドポイントを持つ単一のデータウェアハウスとしてアクセスでき、ワークロード処理を複数の AZ に自動的に分散
- AZ障害の場合
 - ✓ 手動操作不要で自動フェイルオーバー
 - ✓ データ損失なし
 - ✓ 単一のエンドポイントのため、障害時も接続先切り替え不要
- サポートはRA3クラスターのみ
- 対応リージョン
 - US East (N. Virginia), US East (Ohio), US West (Oregon), **Asia Pacific (Tokyo)**, Europe (Ireland) and Europe (Stockholm)

追加コストなしに高いセキュリティ機能を利用可能



エンドツーエンドのデータ暗号化



IAM 連携、シングルサインオン、多要素認証



ネットワーク分離のための Amazon VPC



柔軟なアクセスコントロール (ユーザー/グループにテーブルや列単位での権限制御)



監査ロギングと AWS Cloudtrail 統合



Email や Amazon SNS によるイベント通知



多くの認定および準拠

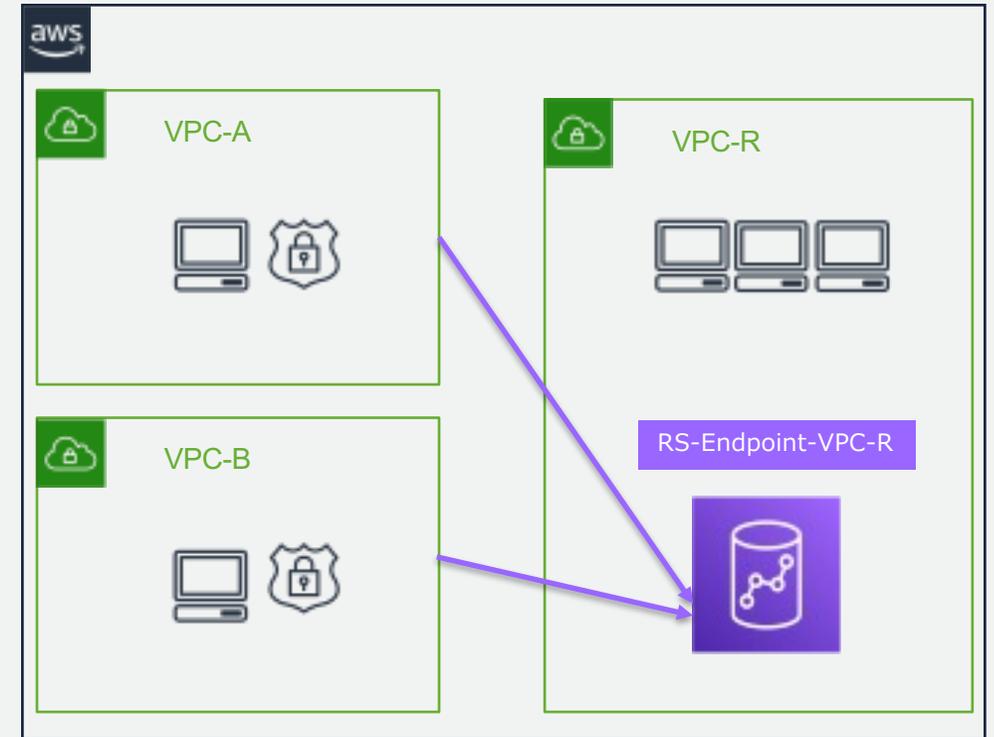
SOC 1/2/3, PCI-DSS, FedRAMP, HIPAA



Amazon Redshift マネージド VPC エンドポイント

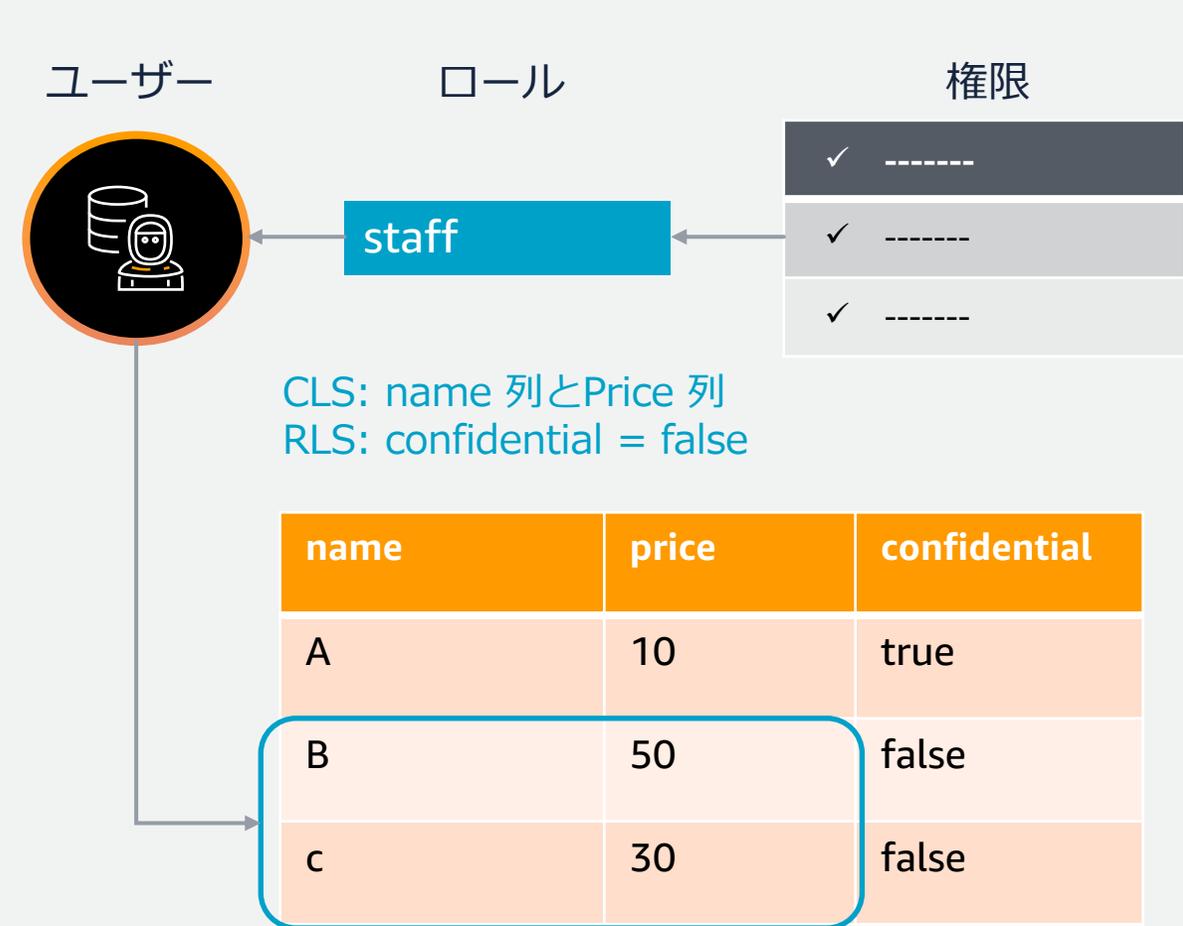
別のVPCから Amazon Redshift へプライベートにアクセス

- ✓ 複数のVPCから Amazon Redshift へプライベート接続でセキュアにアクセス
※異なるアカウント間の接続も可能
- ✓ VPC to VPC のアクセスが、インターネットを経由せずに実現
- ✓ どのVPCからのアクセスを許可するかを設定可能
- ✓ Amazon Redshift Provisioned の場合は、RA3インスタンスタイプで、かつクラスター再配置が有効になっていること



データベースにおけるアクセスコントロール

- テーブル、データベース、スキーマ、関数、プロシージャ、言語に対してアクセスコントロール設定可能
- 列レベルでのアクセスコントロール (CLS)
 - 特定の列のみにアクセスを制限する
- 行レベルでのアクセスコントロール (RLS)
 - 定義されたポリシーに基づいて、テーブル内のデータの特定のレコードのみにアクセス
- ロール単位のアクセス制御 (RBAC)
 - 権限のコレクションであるロールにユーザをアサインし、データベース 権限の管理を簡素化



参考:

https://docs.aws.amazon.com/ja_jp/redshift/latest/dg/r_GRANT.html

動的データマスキング

センシティブなデータを権限に応じて動的にマスキング表示

実テーブル

ID	Geo-location	Name	Phone number
123	WA	Ana	123-456-3568
124	NY	Alice	123-457-8901
125	WA	Bruce	123-458-1234


Admin Role



マスキングポリシー

例) Geo-location = NY 以外のPhone numberは***でマスクする

ポリシーに従ってマスクされた表示内容

ID	Geo-location	Name	Phone number
123	WA	Ana	****-****-****
124	NY	Alice	123-457-8901
125	WA	Bruce	****-****-****

条件判定用
カラム

マスク対象
カラム


NY Analyst Role

- 基になるデータを変更またはコピーせずに、RBACにより権限に応じて表示方法だけを動的に切り替え可能
- どのカラムをどのようにマスクするか？をマスキングポリシーで定義
 - ポリシーは複数設定でき、PRIORITYも定義可能
設定状況はシステムビューで確認可能
- 具体的なマスク処理方法はUDFで自由に拡張可能
- 特定のカラムの値の条件に応じて、マスク対象カラムのマスキングルールを柔軟に制御できる
 - 例) カラムAの値がXXの時のみ、カラムBをマスク
- 対応リージョン：
 - US East (Ohio), US East (N. Virginia), US West (Oregon), **Asia Pacific (Tokyo)**, Europe (Ireland), Europe (Stockholm)

その他の便利な機能

Amazon Redshift RSQL

Amazon Redshift へアクセス可能なコマンドラインクライアント

- Amazon Redshift 固有の機能の追加セットを備えた PostgreSQL psql コマンドラインツールの機能をサポート
- ソートキーや Redshift Spectrum テーブルなど、Amazon Redshift 固有のオブジェクトの属性を表示可能
- IF (¥ELSEIF, ¥ELSE, ¥ENDIF)、¥GOTO、¥LABEL などの拡張されたコントロールフローコマンドも利用可能で、既存のオンプレミス DWH で利用されてきたバッチスクリプトの置き換えも容易に

```
$ rsql -h test.xxxxxxx.ap-northeast-1.redshift-serverless.amazonaws.com -U admin -d dev -p 5439
Password for user admin:
(test) admin@dev=# \d customer
```

Column	Type	Collation	Nullable	Default Value	Encoding	DistKey	SortKey
c_custkey	bigint		NO		none	t	1
c_name	character varying(25)	case_sensitive	NO		lzo	f	0
c_address	character varying(40)	case_sensitive	NO		lzo	f	0
c_nationkey	integer		NO		az64	f	0
c_phone	character(15)	case_sensitive	NO		lzo	f	0
c_acctbal	numeric(12,2)		NO		az64	f	0
c_mktsegment	character(10)	case_sensitive	NO		lzo	f	0
c_comment	character varying(117)	case_sensitive	NO		lzo	f	0

```
Diststyle: KEY
Backup: YES
Created: 2022-09-27 14:24:22.597181
Unique Constraints:
"customer_pkey" PRIMARY KEY (c_custkey)

(test) admin@dev=# \! cat test.sql
select count(*) as cnt from lineitem \gset
select :cnt as cnt;
\if :cnt > 100
  \goto LABELB
\endif

\label LABELA
\remark 'this is label LABELA'
\label LABELB
\remark 'this is label LABELB'

(test) admin@dev=# \i test.sql
cnt
-----
18000048306
(1 row)

\label LABELA ignored
\label LABELB processed
this is label LABELB
```

https://docs.aws.amazon.com/ja_jp/redshift/latest/mgmt/rsql-query-tool.html

ストアドプロシージャ

- 一連の SQL クエリと論理オペレーションを実行可能
- ストアドプロシージャは SELECT クエリに加えてデータ定義言語 (DDL) とデータ操作言語 (DML) を組み込むことができる
- PL/pgSQL フォーマットのストアドプロシージャに対応
- プロシージャはデータベースに保存され、適切な実行権限を持つすべてのユーザーが使用
- AWS Schema Conversion Tool(SCT) による自動変換にも対応

```
CREATE OR REPLACE PROCEDURE test_sp1(f1 int, f2 varchar)
AS $$
BEGIN
    RAISE INFO 'f1 = %, f2 = %', f1, f2;
END;
$$ LANGUAGE plpgsql;

call test_sp1(5, 'abc');
INFO: f1 = 5, f2 = abc
CALL
```

Amazon Redshift Data API

Python, Go, Java, Javascript といった言語から呼び出せるデータアクセスのための API を提供

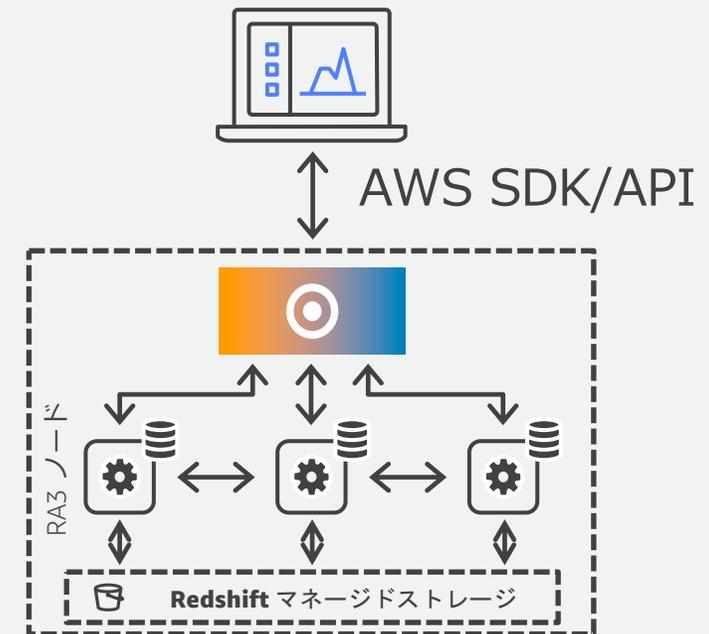
AWS Lambda や AWS AppSync からのクエリ、ロード、アンロードを容易に

ドライバ設定やコネクション管理不要

IAM や Secrets Manager を利用して接続

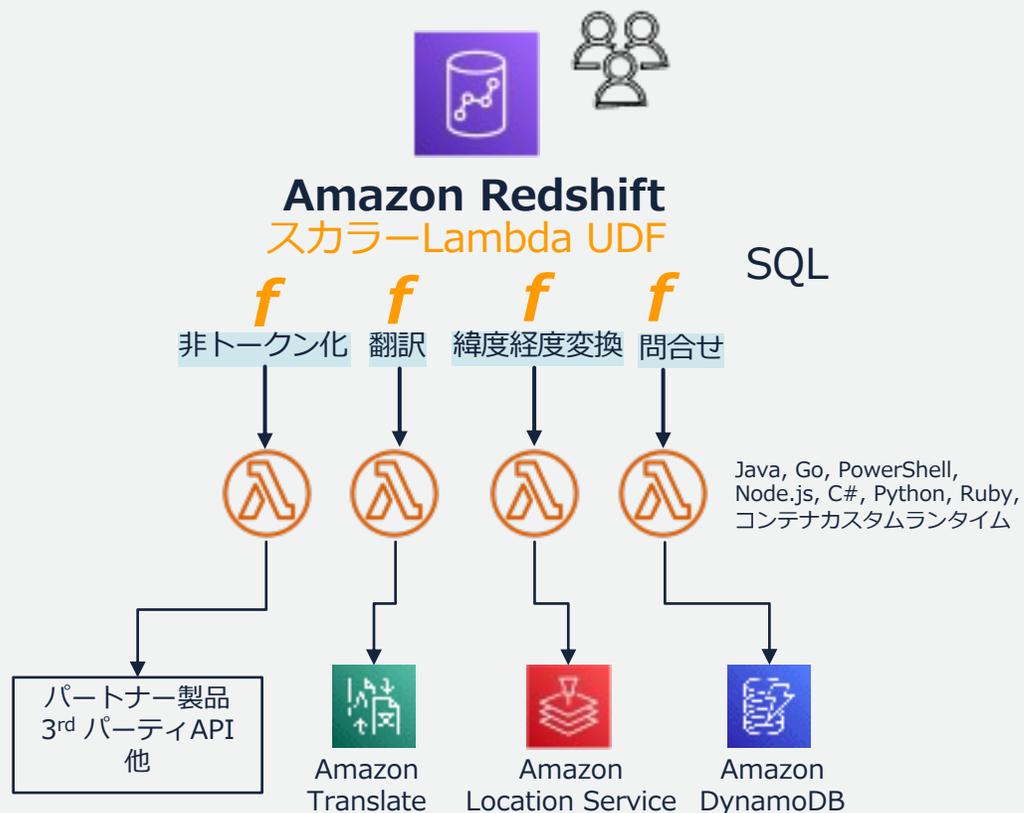
batch-execute-statement を使用して複数のステートメントを実行することも可能

```
aws redshift-data execute-statement
--ClusterIdentifier [CLUSTER ID]
--Database [DATABASE]
--Sql [QUERY]
--SecretArn [CREDENTIALS_ARN]
```



スカラー Lambda UDFs

Redshift から様々な外部サービスの機能をSQLで簡単利用



- AWS Lambda で定義されたカスタム関数を SQL クエリの一部として使用可能
- Redshift の標準機能で実現できないことも SQL で呼び出す関数として拡張可能
- カスタム関数例
 - パートナー製品と連携したデータ暗号化/トークン化
 - AWS AI サービス API と連携した翻訳・自然言語処理
 - Amazon Location Service API と連携したジオコーディング
 - DynamoDB や他のデータソースへの問い合わせ
 - など

Example uses of user-defined functions (UDFs)

<https://docs.aws.amazon.com/redshift/latest/dg/udf-example-uses.html>

SUPER データ型：半構造化データサポート

JSON データを簡単・効率的かつ強力的に処理

高速な行指向データの取り込み

PartiQL クエリ言語を活用した使いやすいSQL
拡張機能により、スキーマレスのネストされた
データへのアクセスを実現

SUPER 型のデータをマテリアライズドビュー化
することによる高速な列指向分析

customers テーブル

id INTEGER	name SUPER	phones SUPER
1	{"given": "Jane", "family": "Doe"}	[{"type": "work", "num": "9255550100"}, {"type": "cell", "num": "6505550101"}]
2	{"given": "Richard", "family": "Roe"}	[{"type": "work", "num": "5105550102"}]

```
SELECT name.given AS firstname, ph.num  
FROM customers c, c.phones ph  
WHERE ph.type = 'cell';
```

```
firstname | num  
-----+-----  
"Jane"   | 6505550101
```

空間データ(spatial data)

大容量の空間データとビジネスデータを
シームレスに統合し、高速に分析

データ型として**GEOMETRY型**と**GEOGRAPHY型**
をサポート

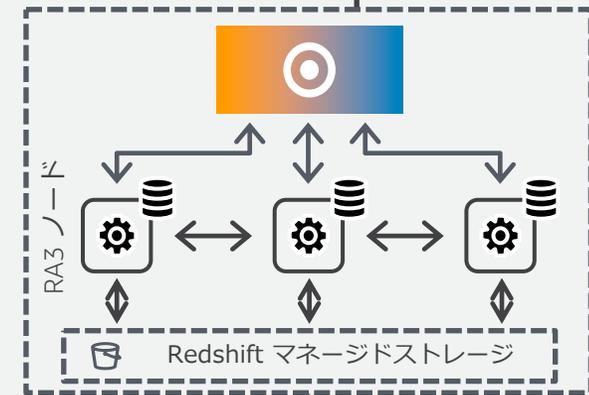
空間データのインポート、エクスポート、
アクセスおよび処理を実行可能

空間関数

https://docs.aws.amazon.com/ja_jp/redshift/latest/dg/geospatial-functions.html



Insert ↓ ↑ Select



Copy



料金体系

<https://aws.amazon.com/jp/redshift/pricing/>

Amazon Redshift の主な課金要素

課金要素	Amazon Redshift Provisioned	Amazon Redshift Serverless
コンピューティング	<p>インスタンス起動時間 (クラスタ利用)</p> <p>Redshift Spectrum 利用料 (データスキャン量に対する従量制)</p> <p>同時実行スケーリング利用料 (有効化および実行時。無料枠あり)</p>	<p>RPU 実行時間 (ワークロード実行時のコンピューティング利用)</p> <p>(データレイククエリやスケーリング 実行時も RPU 実行時間でカウント)</p>
ストレージ	<p>RMS ストレージ使用量 (RA3 インスタンスタイプ のみ)</p>	<p>RMS ストレージ使用量</p>

※ RPU : Redshift Processing Unit

※ その他として、Redshift ML、バックアップストレージ、データ転送、スカラーLambda UDF利用など利用に応じて課金



Amazon Redshift Provisioned

Redshift インスタンス起動時間 + ストレージ使用量(RA3のみ)

- コンピュートノード数 x 1 時間あたりの価格 (リーダーノードは課金対象外)
- リザーブドインスタンス (1 年、3 年) の利用で単価を大幅に下げることが可能

※東京リージョンの料金を記載。コンピュート料金は、1 時間未満の時間は 1 秒単位で請求されます。

	vCPU	メモリ	ストレージ	I/O	デフォルト スライス数	ノード数	コンピュート 料金 (*)	ストレージ 料金 (*)
RA3 with Redshift Managed Storage(RMS)								
ra3.xlplus	4	32 GiB	32TB RMS ※1	650 MB/s	2	1~16 ※2	\$1.278 /h	\$0.0261 /GB-Month
ra3.4xlarge	12	96 GiB	128TB RMS	2.0 GB/s	4	2~32 ※2	\$3.836 /h	
ra3.16xlarge	48	384 GiB	128TB RMS	8.0 GB/s	16	2~128	\$15.347 /h	
DC2 - Dense Compute								
dc2.large	2	15 GiB	0.16TB SSD	0.6 GB/s	2	1~32	\$0.314 /h	
dc2.8xlarge	32	244 GiB	2.56TB SSD	7.5 GB/s	16	2~128	\$6.095 /h	

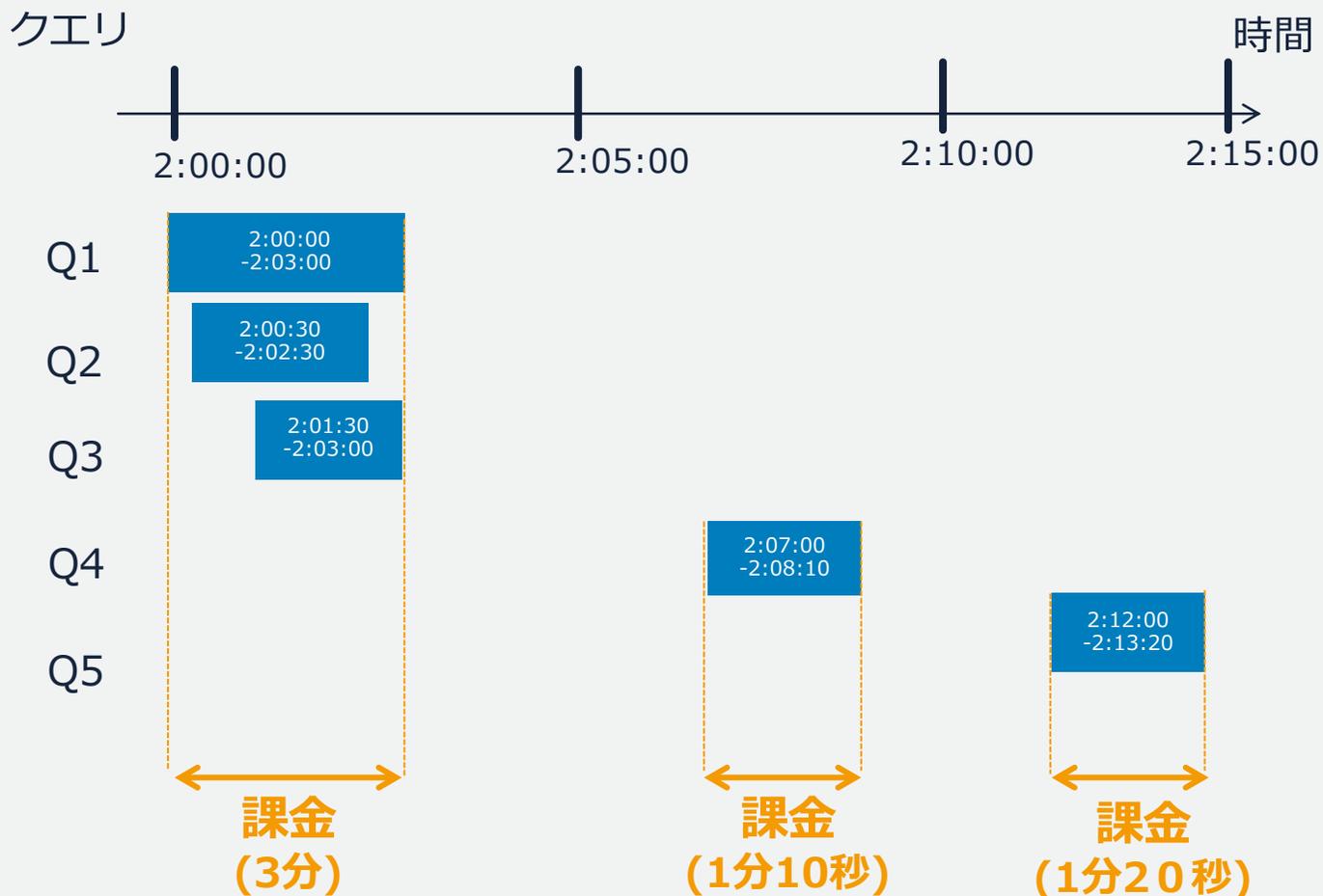
※1 ra3.xlplusのRMSストレージ上限は、1ノード 構成の場合のみ最大 4TBまで、2ノード 以上のクラスタの場合は1ノード あたり最大32TBとなります

※2 作成時の最大ノード数は表記の通りですが、Elastic Resizeを使うことによって、xlplusでは最大32ノード、4xlargeでは最大64ノードまで増加させることができます。

Amazon Redshift Serverless

- 演算処理: 使用したRPU(Redshift Processing Unit) に応じて秒単位課金 (東京リージョンの場合\$0.494/RPU/時間。最低料金時間は60秒)
- 1 RPU=16GiB メモリ
- オプションで消費するRPUの基本容量を、8 RPU から 512 RPU まで、8 RPU 単位 (8、16、24...512) で設定可能
- データロードやクエリ実行などクラスタ使用中の時間に対しての課金となり、起動時間やアイドル中の時間に対する料金は不要
- 同時実行スケーリング、データレイククエリの料金は Serverless の料金に含まれており、別途料金は不要
- ストレージ: Amazon Redshiftマネージドストレージとスナップショットストレージの料金が発生

Amazon Redshift Serverless の課金イメージ



- Redshift Processing Unit (RPU) 消費に応じた秒単位課金※ + Redshift マネージドストレージ (RMS) 消費の重量課金
- 同時実行スケールリングやデータレイクへのクエリも包含 (別途追加料金なし)
- アイドル時間は課金対象外

課金対象期間	クエリ実行時間
@2:03	3 分 (for Q1, Q2, Q3)
@2:09	1 分 10 秒 (for Q4)
@2:14	1 分 20 秒 (for Q5)
課金対象合計	5 分 30 秒

※60 秒の最低料金あり。データウェアハウスの起動時間には料金はかかりません。

まとめ

Amazon Redshift とは、

何万人ものお客様が使用しているフルマネージド型のペタバイト規模のクラウドデータウェアハウス

あらゆる規模のあらゆるデータを

簡単、迅速、安全、かつコスト効率の高い方法で分析可能



Easy analytics for everyone

誰でも簡単に
分析できる



Analyze all your data

全てのデータを
分析する



Best price performance
at any scale

コストパフォーマンスの
高さ

参考資料

Amazon Redshift ドキュメント

https://docs.aws.amazon.com/ja_jp/redshift/latest/mgmt/welcome.html

Amazon Redshift の最新情報

<https://aws.amazon.com/jp/redshift/whats-new/>

[AWS Black Belt Online Seminar] Amazon Redshift 運用管理

[資料](#) [動画](#)

Amazon Web Services ブログ (Category: Amazon Redshift)

<https://aws.amazon.com/jp/blogs/news/category/database/amazon-redshift/>

Amazon Redshift Immersion Labs

<https://catalog.us-east-1.prod.workshops.aws/workshops/9f29cdba-66c0-445e-8cbb-28a092cb5ba7/en-US/>

Amazon Redshift Utils

<https://github.com/awslabs/amazon-redshift-utils>

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!