



その ETL パイプラインもういらなにかも？ zero-ETL 総まとめ！

2023/12/21

AWS Solutions Architect
Shuhei Fukami

自己紹介

Name :
深見修平(Shuheï Fukami)

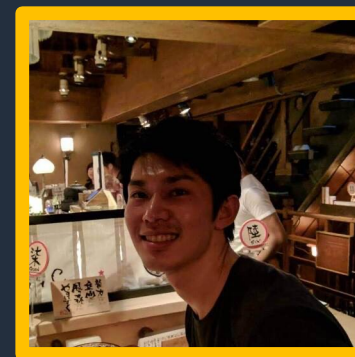
Title :
Solutions Architect

趣味 :
サッカー(多大なブランクあり)
自転車(本格ではない)
特技 :
ピザ焼き
(窯があるのに焼く人間がない時ぜひお呼びください)

好きなAWSサービス :
Amazon OpenSearch Service, Amazon Redshift Serverless



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



注意事項

本日紹介する機能の中には2023年12月現在で public preview や limited preview のものが含まれています。

- Amazon DynamoDB zero-ETL integration with Amazon Redshift (limited preview)
→ previre の申し込みは[こちら](#)
- Amazon RDS for MySQL zero-ETL integration with Amazon Redshift(Public Preview)
- Amazon Aurora PostgreSQL zero-ETL integration with Amazon Redshift(Public Preview)
- Amazon OpenSearch Service zero-ETL integration with Amazon S3 (Public Preview)

アジェンダ

- re:Invent 2023 で発表された zero-ETL
- Amazon Redshift zero-ETL integration
- OpenSearch Service zero-ETL integration
- Demo
- まとめ

re:Invent 2023 で発表された zero-ETL

Amazon Redshift

- Amazon Aurora PostgreSQL zero-ETL integration with Amazon Redshift (Public Preview)
- Amazon RDS for MySQL zero-ETL integration with Amazon Redshift (Public Preview)
- Amazon DynamoDB zero-ETL integration with Amazon Redshift (Limited Preview)

Amazon OpenSearch Service

- Amazon DynamoDB zero-ETL integration with Amazon OpenSearch (GA)
- Amazon OpenSearch Service zero-ETL integration with Amazon S3 (Public Preview)



Amazon Redshift zero-ETL integrations



従来の ETL ソリューションにまつわる課題



ETLジョブの**構築・運用**が必要

スキーマ変更などの操作の際に
複雑なデータの再構築

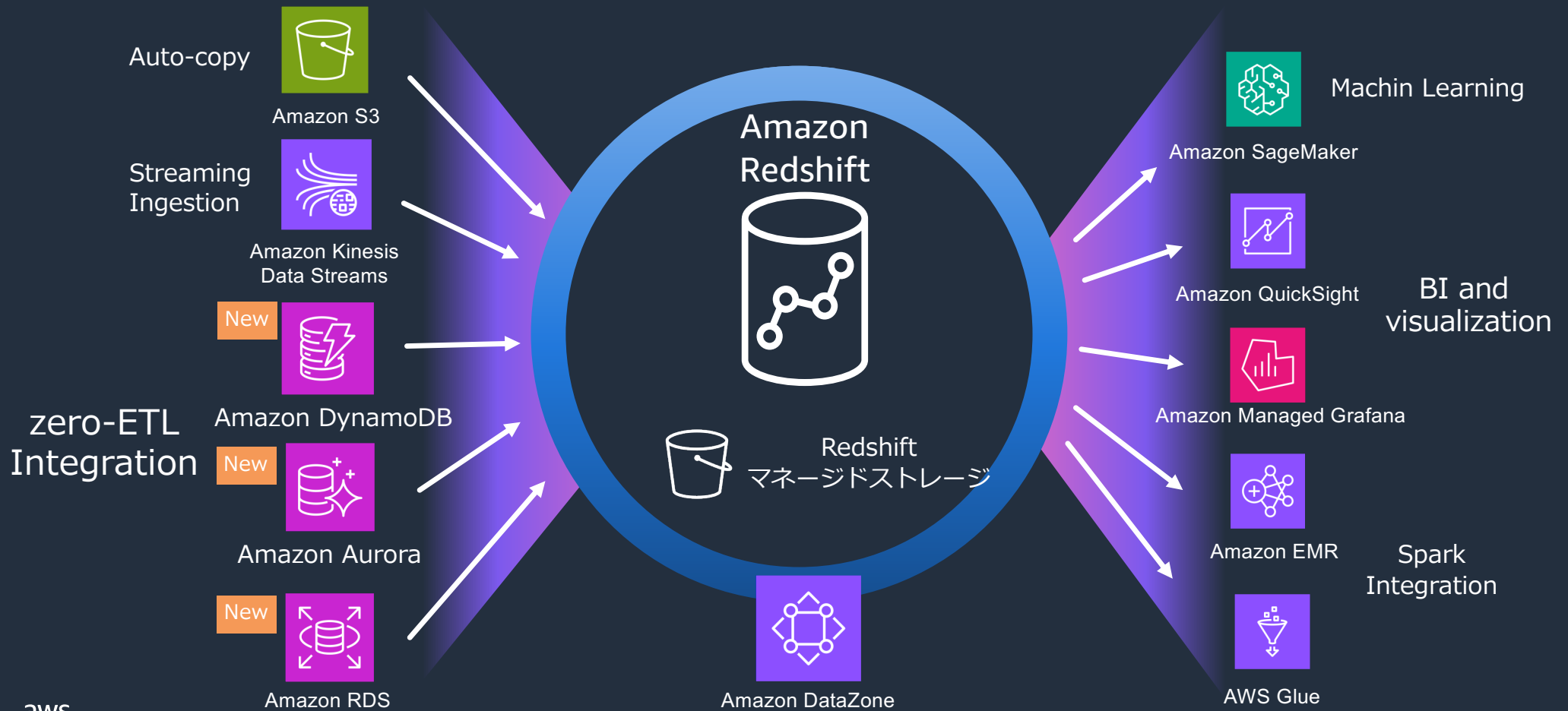
十分なインサイトを得るには
**完全で一貫性のある
最新のデータを扱うことが重要**

Amazon Redshift の zero-ETL のビジョン



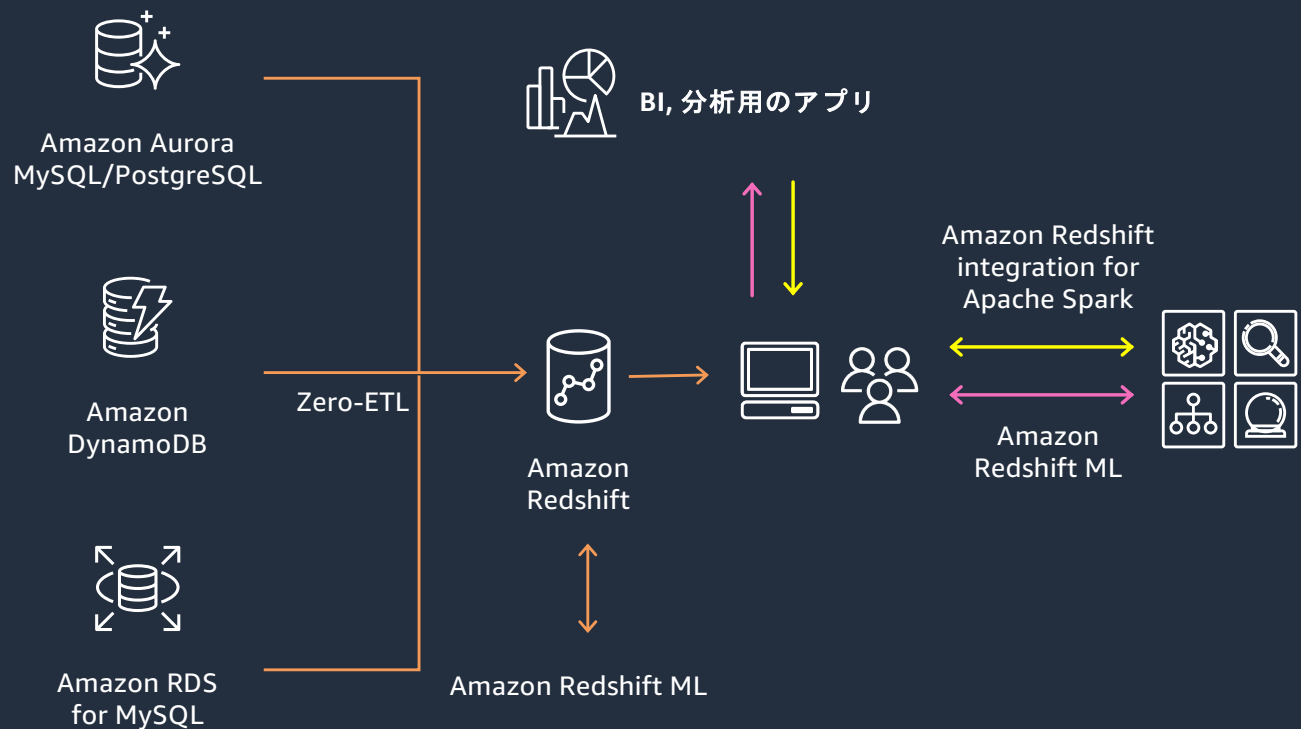
パイプラインの管理無しに、ペタバイト級のトランザクションデータの
ニアリアルタイム分析を可能にする簡単でセキュアな方法

Redshift と様々なサービスとのシームレスなデータ統合が可能に



Amazon Redshift zero-ETL integrations

3つの新規のデータベースをサポート



多数のデータベースを一つのデータウェアハウスで統合

Amazon Aurora MySQL (GA),

Amazon Aurora PostgreSQL
(public preview)

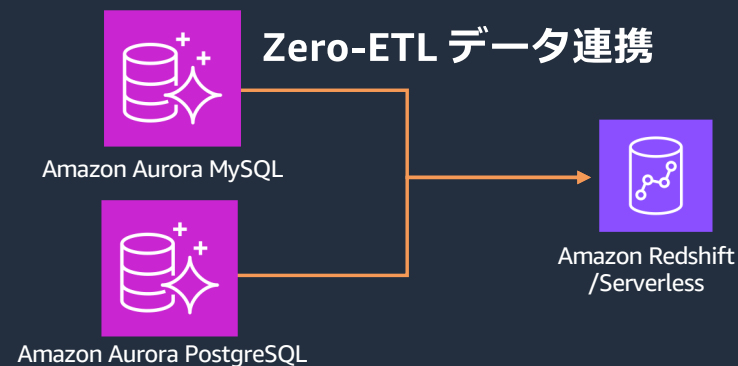
Amazon DynamoDB
(limited preview)

Amazon RDS for MySQL
(public preview)

Amazon Aurora との zero-ETL 統合の概要

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

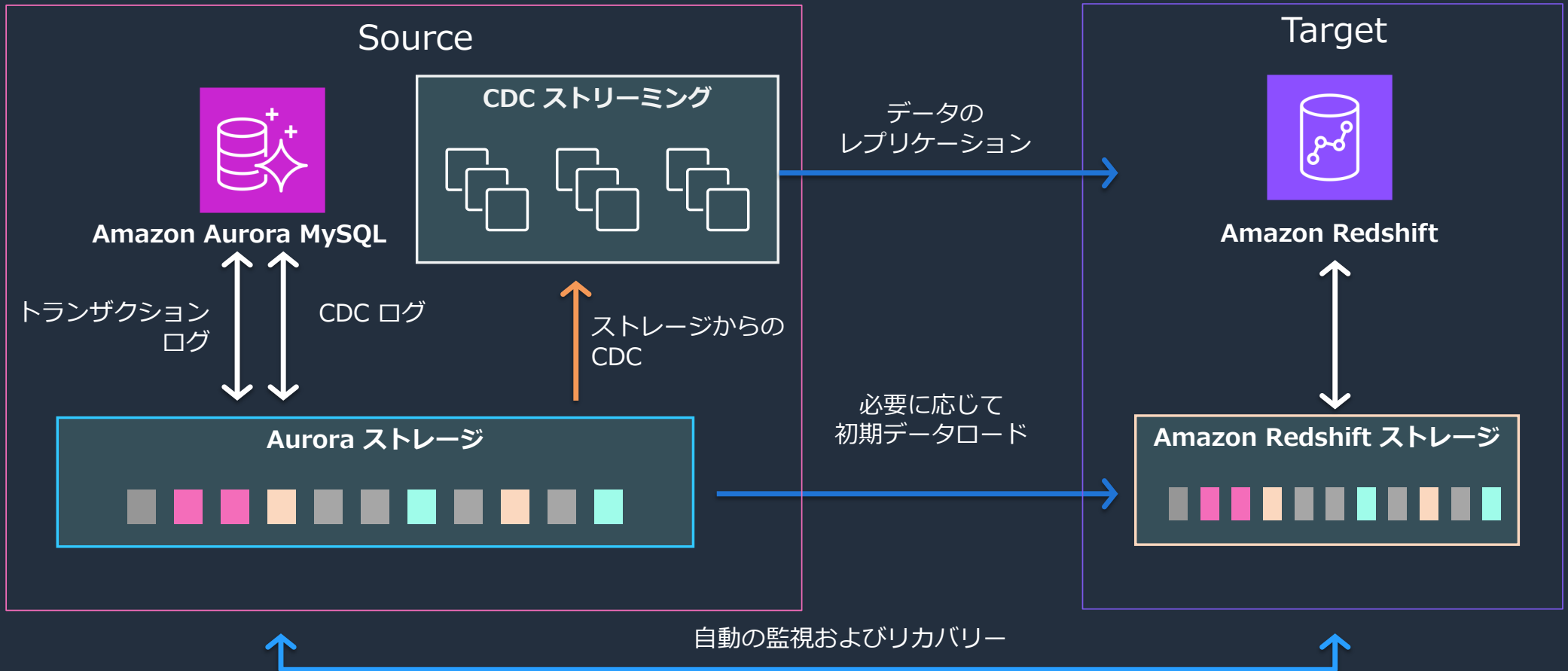
- ✓ ETL パイプライン不要で
複数の Aurora データベースから連携
- ✓ Amazon Aurora に書き込まれてから数秒以内に
Amazon Redshift にデータを複製
- ✓ Amazon Redshift 組み込み機能と合わせて活用し
Aurora のトランザクションデータを
パフォーマンス効率よく分析
- ✓ zero-ETL 自体の料金は無料



※binlog 有効化とenhanced binlog の利用が必要

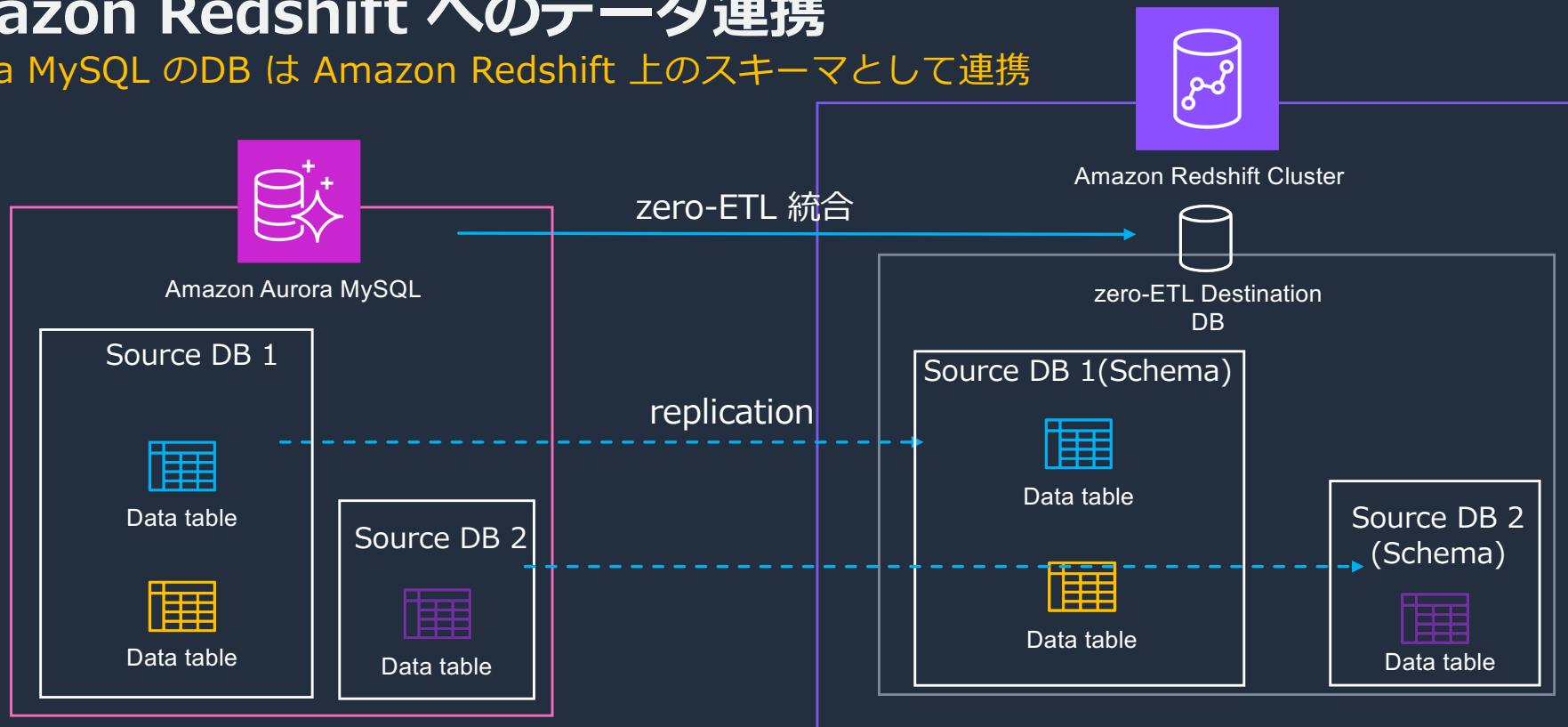
ストレージレイヤーでの統合

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



zero-ETL による Aurora MySQL から Amazon Redshift へのデータ連携

Aurora MySQL のDB は Amazon Redshift 上のスキーマとして連携

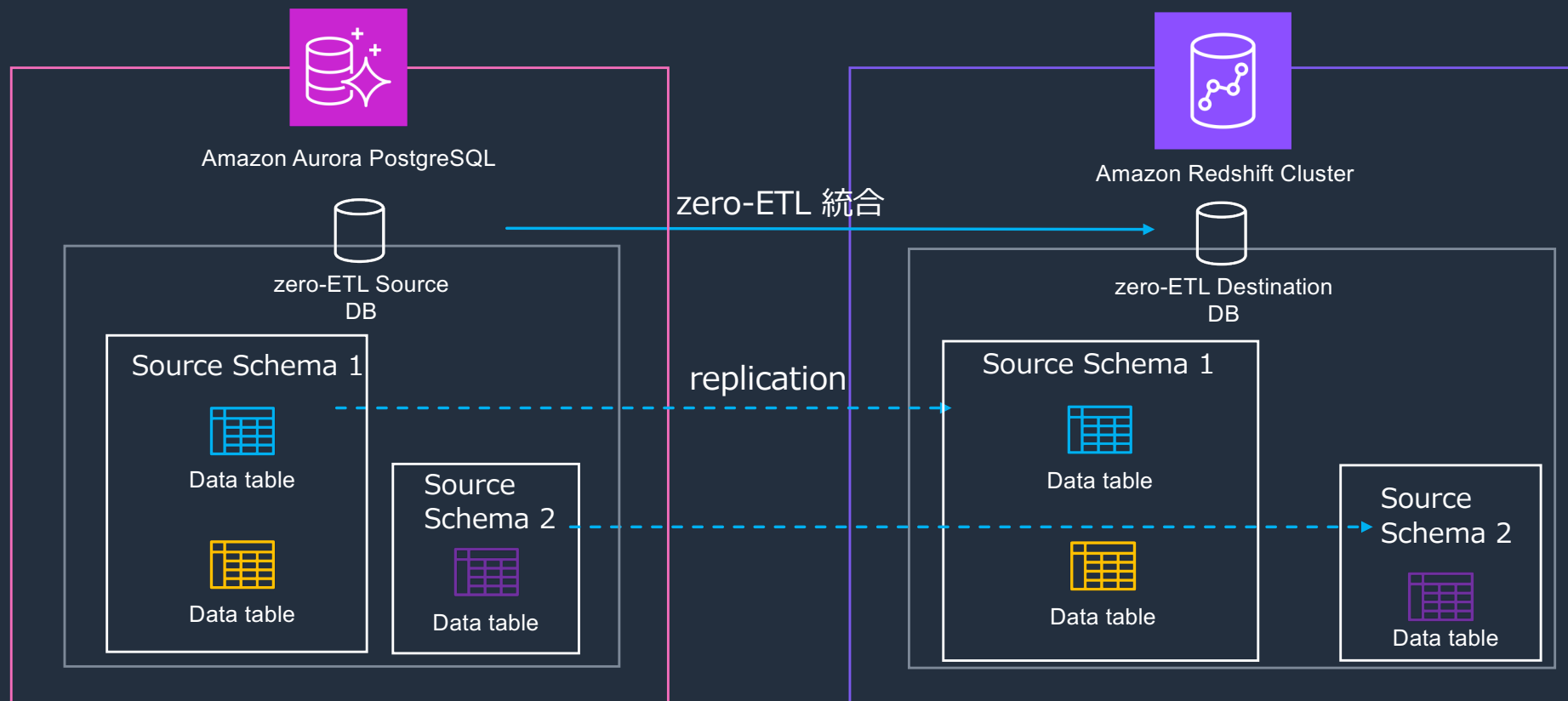


zero-ETL で連携したテーブルは **Read Only**

ただし、クロスデータベースクエリにより別のDBからマテリアライズビューの作成などの操作が可能



zero-ETL による Aurora PostgreSQL から Amazon Redshift へのデータ連携

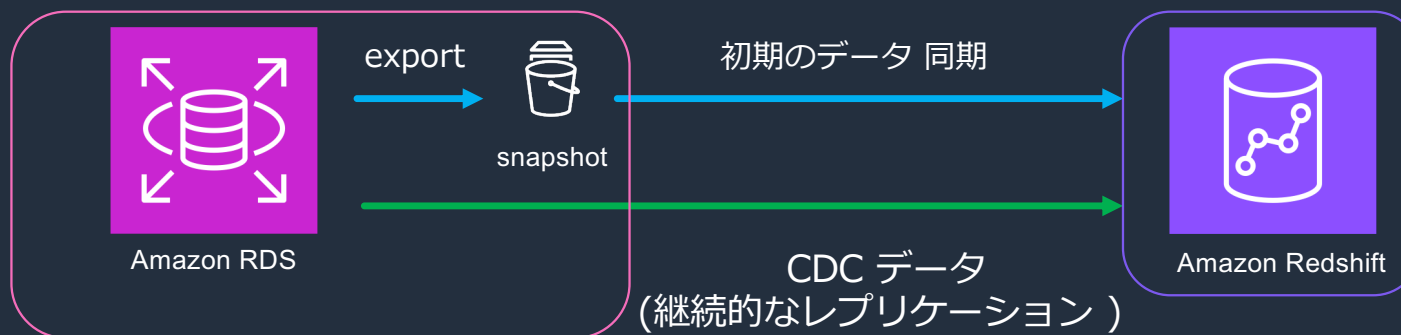


zero-ETL で連携したテーブルは **Read Only**

ただし、クロスデータベースクエリにより別のDBからマテリアライズビューの作成などの操作が可能 ¹⁶



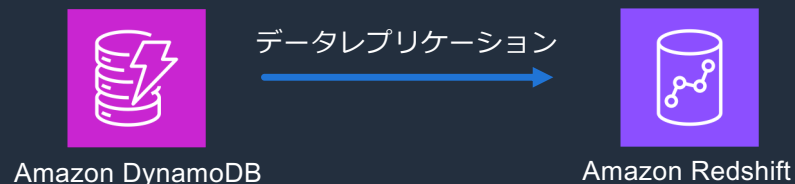
Amazon RDS for MySQL zero-ETL integration with Amazon Redshift (Public Preview)



- ✓ snapshot をエクスポートし、それをもとに初期のデータの同期を行う
再同期の際にはスナップショットをエクスポートをしておいて再実行
- ✓ CDC データを用いた継続的なレプリケーション
- ✓ 東京リージョンを含む6つのリージョンで Preview
- ✓ RDS for MySQL zero-ETL の料金は
スナップショット費用 + CDC データの転送量

https://aws.amazon.com/rds/mysql/pricing/#Zero-ETL_integration_costs_.28Preview.29

Amazon DynamoDB zero-ETL integration with Amazon Redshift



- ✓ DynamoDB のデータをシームレスに Amazon Redshift に連携
- ✓ DynamoDB のワークロードに影響を与えず実現
- ✓ ETL pipeline の構築に複雑なコンポーネントの組み合わせが不必要に
- ✓ Limited preview in the US East (Ohio) region

Preview の申請はこちらから → → → → → →

<https://pages.awscloud.com/DynamoDB-zeroETL-AmazonRedshift-Preview.html>

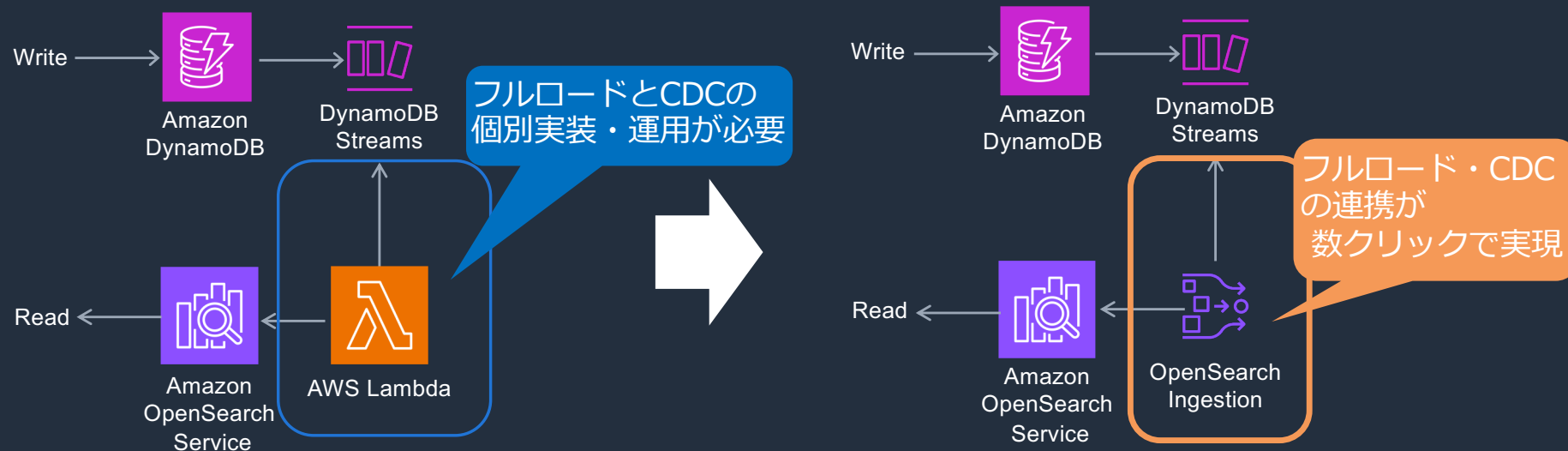


Amazon OpenSearch Service zero-ETL integrations



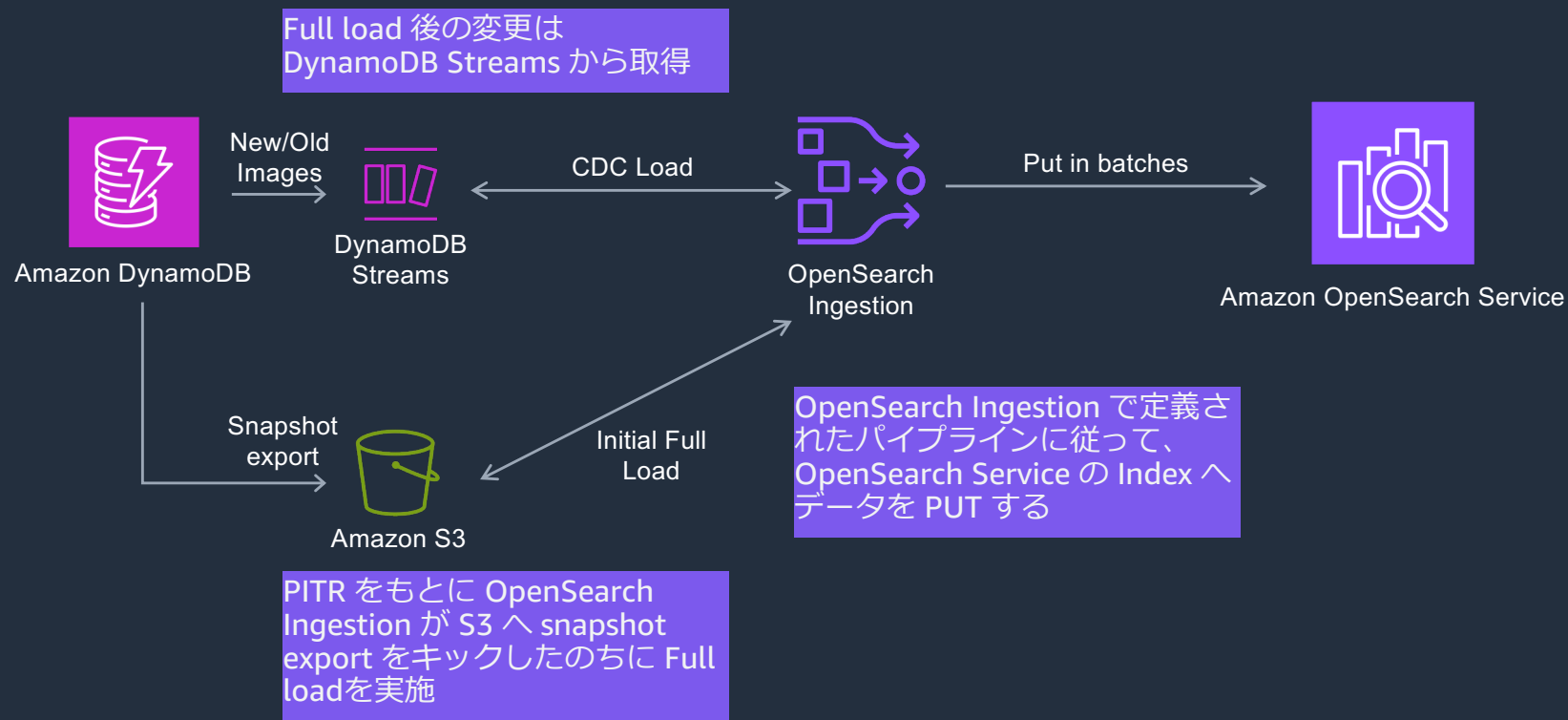
Amazon DynamoDB zero-ETL integration with Amazon OpenSearch

一般利用開始



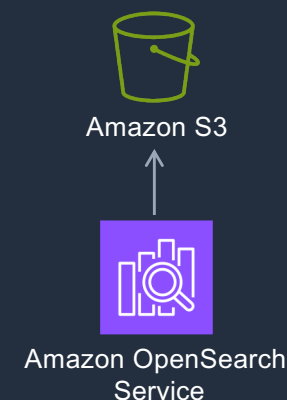
- 数回のクリックで、DynamoDB から OpenSearch Service にデータを同期できる zero-ETL サービス
 - ETL コードやインフラの管理が不要になり、OpenSearch との連携が簡単に
- 料金は DDB Streams、PITRリカバリ、OpenSearch Ingestion の OCU 料金
- Amazon OpenSearch Ingestion が利用可能な全リージョンで一般提供開始

Amazon DynamoDB zero-ETL integration の構成イメージ



Amazon OpenSearch Service zero-ETL integration with Amazon S3

- S3 にあるデータに対して Amazon OpenSearch Service で直接クエリすることが可能に
- S3 にあるログデータ (VPC Flow ログ, ELB ログなど) に対してインシデントなどがあつたときに分析と可視化が簡単に実施可能になった
- 料金は OpenSearch Service Direct Query の OCU (分単位課金)
- 東京含め 6 リージョンで利用可能



Demo



まとめ

- Amazon Redshift では3つ、OpenSearch Service では2つの新たな zero-ETL が登場
- 用途の違うデータストア間でシームレスにデータを連携し、ワークロードに適したサービスを選択することが容易に実現
- 既存のETLを置き換えることで運用を軽減
- 複数のソースに散らばるデータを1つに集めることが容易になり、横断での分析を行う基盤の実現が容易に