



# Amazon Aurora と Amazon Redshift の Zero-ETL 統合のご紹介

Shota Sato

Analytics Specialist Solution Architect

# 自己紹介

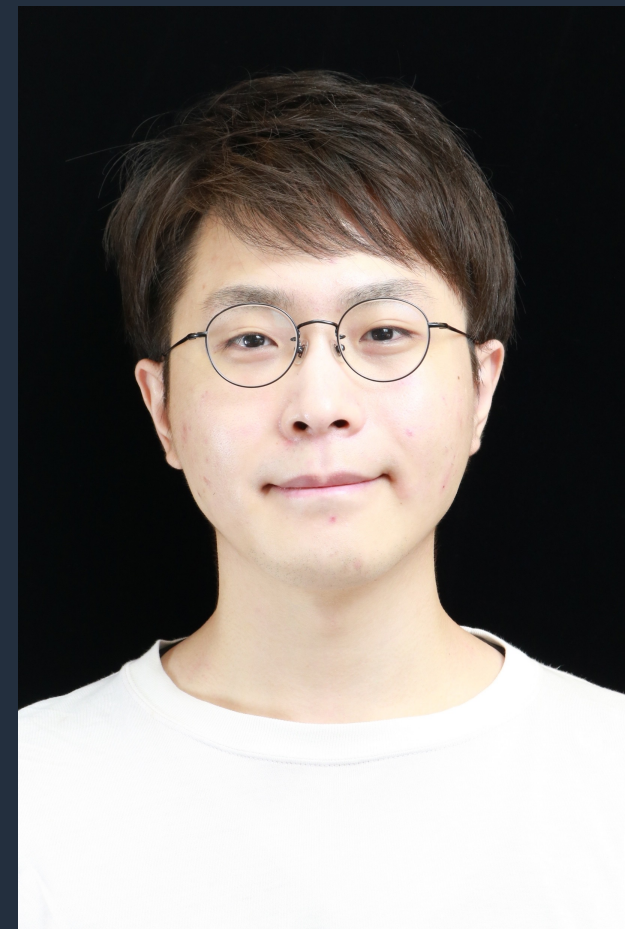
## 佐藤 祥多

アマゾン ウェブ サービス ジャパン 合同会社  
データ事業本部ポートフォリオ  
スペシャリストソリューション部  
アナリティクススペシャリストソリューションアーキテクト

ゲーム業界やインターネットメディア業界のお客様に  
対してアナリティクス領域全般をサポート

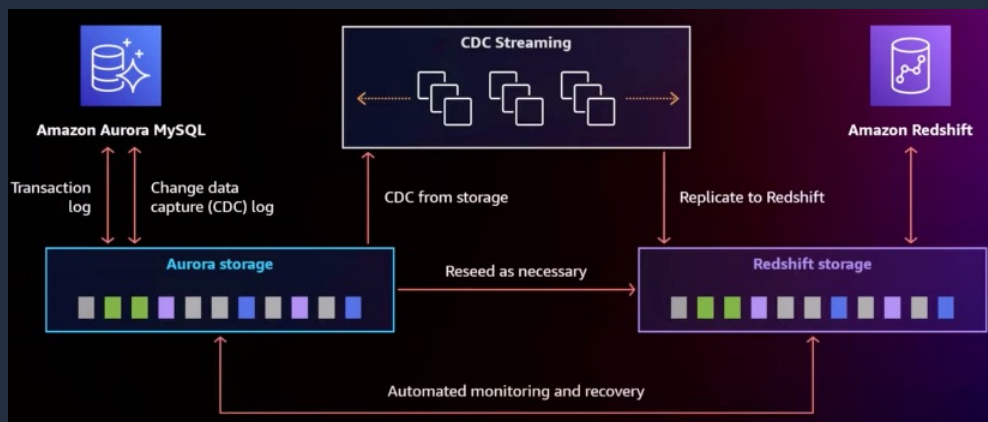
好きな AWS サービス

Amazon Simple Storage Service (Amazon S3),  
AWS Lake Formation, Amazon DataZone



# 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

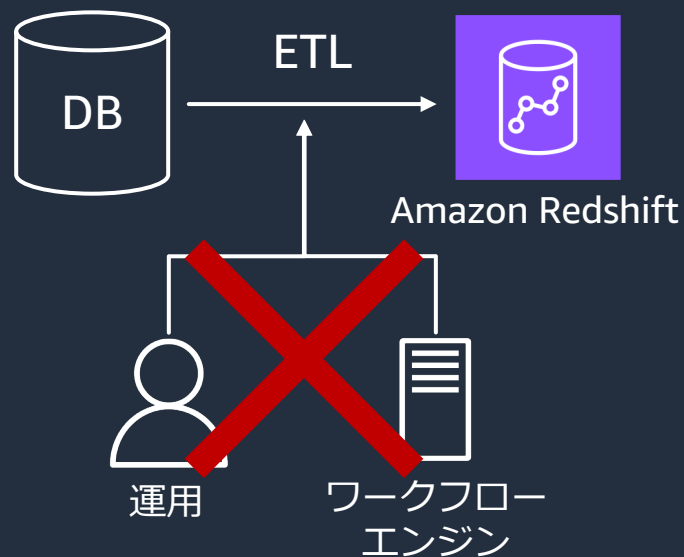


- ETL パイプライン不要で 複数の Aurora データベースからニアリアルタイム連携
- Amazon Aurora に書き込まれてから**数秒以内に Redshift に複製**
- Redshift 組み込み機能である ML、マネリアライズドビュー、データ共有、フェデレーテッドアクセスなどを活用して、Aurora のトランザクションデータを分析
- ストレージレイヤー統合により伴う**DBインスタンスへのパフォーマンス影響なくデータ統合を実現**
- **他クラウドへの DTO コスト**が問題になっているケースや **Aurora 上でのレポーティング**がアプリに影響がでているケースに有用

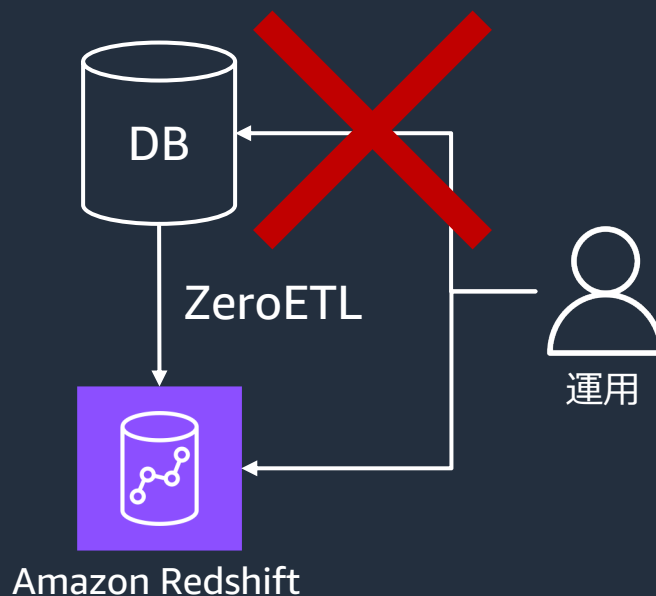
# どういったケースに有効か

運用軽減、負荷軽減、リアルタイムの可視化に有効

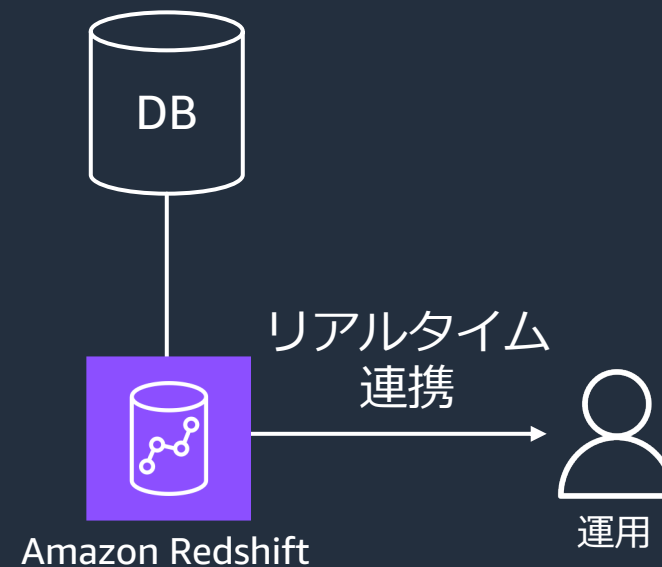
## ETL の運用軽減



## DB への負荷軽減

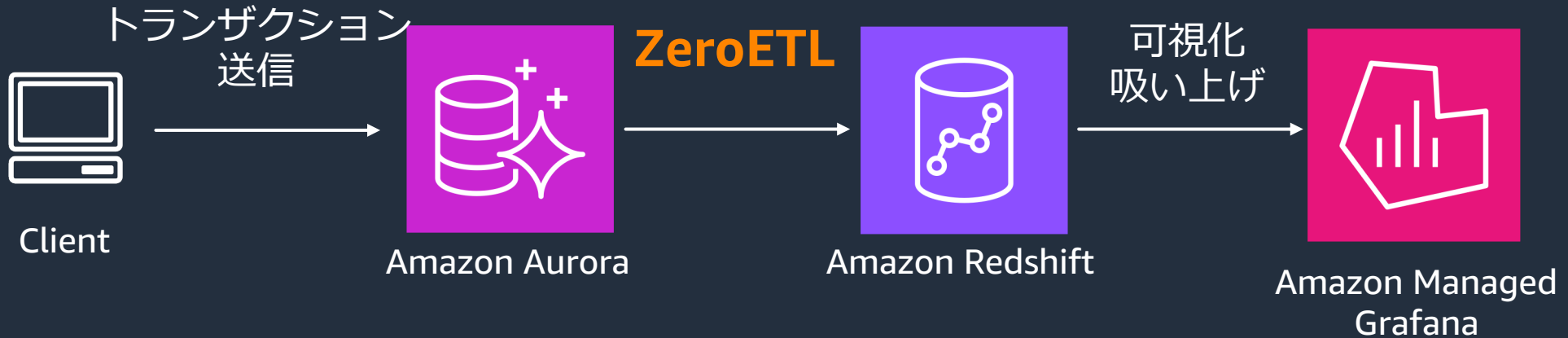


## リアルタイムでの可視化



# デモシナリオ

ゲーム内の仮想通貨の動きをリアルタイムで可視化



# デモシナリオ

仮説：イベントやガチャを導入することで、仮想通貨の購入・消費の変化をリアルタイムで確認して効果を測定する

仮想通貨の購入：小  
仮想通貨の消費：小



イベント・ガチャ  
導入



仮想通貨の購入：大？  
仮想通貨の消費：大？

購入や消費の変化は予想どおりか？

確認ポイント：

- どのくらい手軽にデータ連携が可能か
- どのくらいのスピードで連携されるか

# 制約

## Public Preview のため、制約は多い

- To share data from an Amazon Redshift destination database (producer) to an Amazon Redshift consumer data warehouse, your consumer data warehouse must be on the PREVIEW\_2023 track. For more information, see [Data sharing concepts for Amazon Redshift](#).
- You can't perform pause, resume, or classic resize operations on Amazon Redshift provisioned clusters with integrations configured.
- You can't enable enhanced VPC support on data warehouses with integrations configured.
- You can't create an integration if the target Amazon Redshift provisioned cluster is paused. In addition, you can't pause a provisioned Amazon Redshift cluster if it has an existing integration.
- If you remove a source DB cluster as an authorized integration source for a Amazon Redshift data warehouse, all associated integrations are deleted.
- You can't delete a Amazon Redshift data warehouse that has existing zero-ETL integrations. You must delete all associated integrations first.
- You can't delete a destination database when an integration is associated with the database.
- The destination database is read-only. Materialized views are only supported when used in cross-database queries. For more information, see [Querying data across databases](#).
- You can't create tables, views, or materialized views in the destination database. You can't query tables in the target data warehouse that are in the Resync state.

### Unsupported data types

If a table in your source Aurora DB cluster includes an unsupported data type, the table goes out of sync and isn't consumable by the Amazon Redshift target. You must manually fix the table to resume streaming between the source and target.

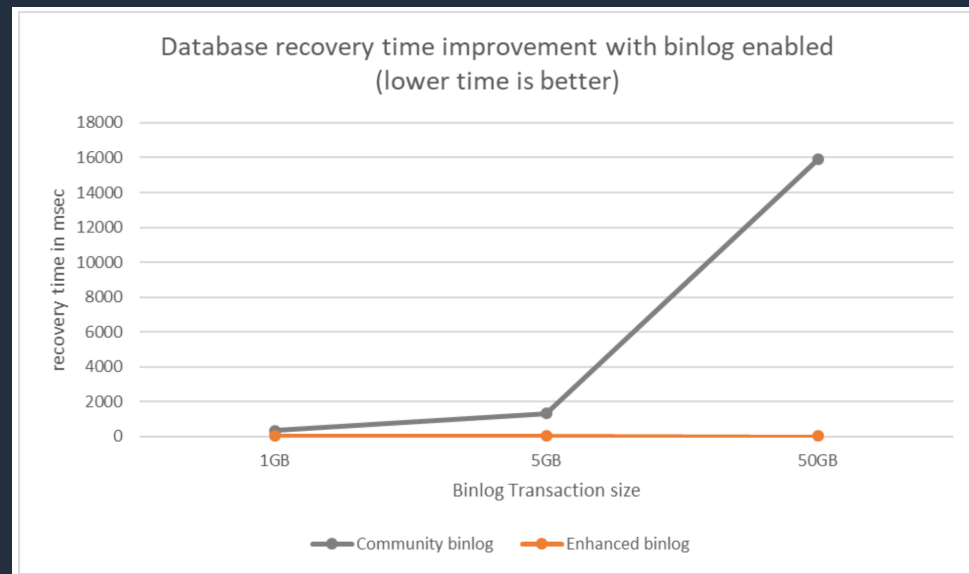
The following data types aren't supported for Aurora zero-ETL integrations with Amazon Redshift:

- BINARY, VARBINARY
- BIT
- JSON
- GEOMETRY, POINT, LINestring, POLYGON, MULTIPOINT, MULTILINESTRING, MULTIPOLYGON, GEOMETRYCOLLECTION
- TINYBLOB, BLOB, MEDIUMBLOB, LONGBLOB

<https://docs.aws.amazon.com/AmazonRDS/latest/AuroraUserGuide/zero-etl.html>

# よくある質問

- Aurora への負荷や影響は？
  - 拡張 Binlog を利用するため、Aurora MySQL クラスタの負荷影響は最小限に抑えられる
    - Aurora MySQLでは Binlog を用いないレプリケーションを行っているため、明確な用途がない場合は Binlog を無効にすることを推薦
  - Aurora MySQL クラスタから Redshift への初期ロード・CDC によるデータベースインスタンスへの負荷はかからない
  - Crash recovery の時間も拡張 Binlog により大幅に改善







**Thank you!**