



データ活用のベストプラクティスと それを支える **AWS** アナリティクスサービス

志村 誠, アナリティクスソリューションアーキテクト
アマゾン ウェブ サービス ジャパン合同会社



志村 誠

アマゾン ウェブ サービス ジャパン合同会社
アナリティクスソリューション部
部長, アナリティクスソリューションアーキテクト

幅広い業種・業界のお客様に対して, アナリティクス
サービスの活用を技術的側面から支援

なぜデータ分析をするのでしょうか？

ビジネス課題に対して データにもとづく意思決定をするため

カンや思いこみに頼らない，データに裏付けされた事実による判断
仮説に基づく新しい施策の効果を，データを用いて検証
次に何をすべきかを判断するために，まず現状を把握

データに基づく意思決定に必要なもの

意思決定（仮説検証）に役立つ指標

定量的に計測可能なメトリクス

指標を出すための元データ

最終的にメトリクスに変換可能な大量の生データ

元データを指標に落とすための仕組み

膨大かつ未整形の生データを分析可能な形に変換し、適切なツールを選択



指標と元データの例



既存のマーケティングでは顧客に有効にアプローチできない

- 指標 新商品売上やプロモーションの反応率
- データ EC サイトや既存店舗の売上データ, 各種ログ



顧客のサービス継続率が低下してきている

- 指標 顧客満足度やサービス継続率
- データ ユーザマスタ, CRM マスタ, サービス利用ログ



取引の監査に膨大な工数がかかり適切に行えていない

- 指標 監査に要する時間や検知数
- データ 過去の取引情報

ひととおり分析できるようになりました
これで万事 **OK**？

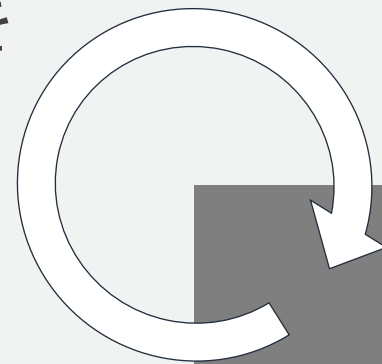
仮説検証の結果，新しい次の仮説が生まれる

顧客属性や購買実績だけでなく，オンラインでの行動パターンの影響を見たい
キャンペーンメールのおすすめ商品を改善したら，サービス解約率はどうなる？
より詳細なトランザクション情報があれば，不正取引の検知率を上げられる？

最初から明確な仮説を持つのが困難な場合も

- まずは構築した基盤上でデータを集計・可視化して、どのような施策を実施するべきか、というあたりをつける。簡単な仮説検証のプロセスをまわす

得られたデータをもとに、
徐々に施策と評価指標を
明確にしていく



機械学習による
予測や判断の自動化

応用的な分析手法
ストリーム分析・可視化

さまざまな集計軸で分析
BI ツールや SQL

データの蓄積とダッシュボードでの可視化

データ分析のステージや目的に応じて
使うツールやデータ，変換方法が変わる

よくあるデータウェアハウス + **BI** ダッシュボード

今みえている要件に対しては、これで満たせる。パフォーマンス検証も済んだ。でもその先は？ 優先度の高いバッチ処理も追加したくなったら？ 他の部署も同様の分析をしたくなったら？ 新しいデータを追加する必要がでてきたら？ そもそも新しい仮説を検証するのが SQL だと難しいときどうする？



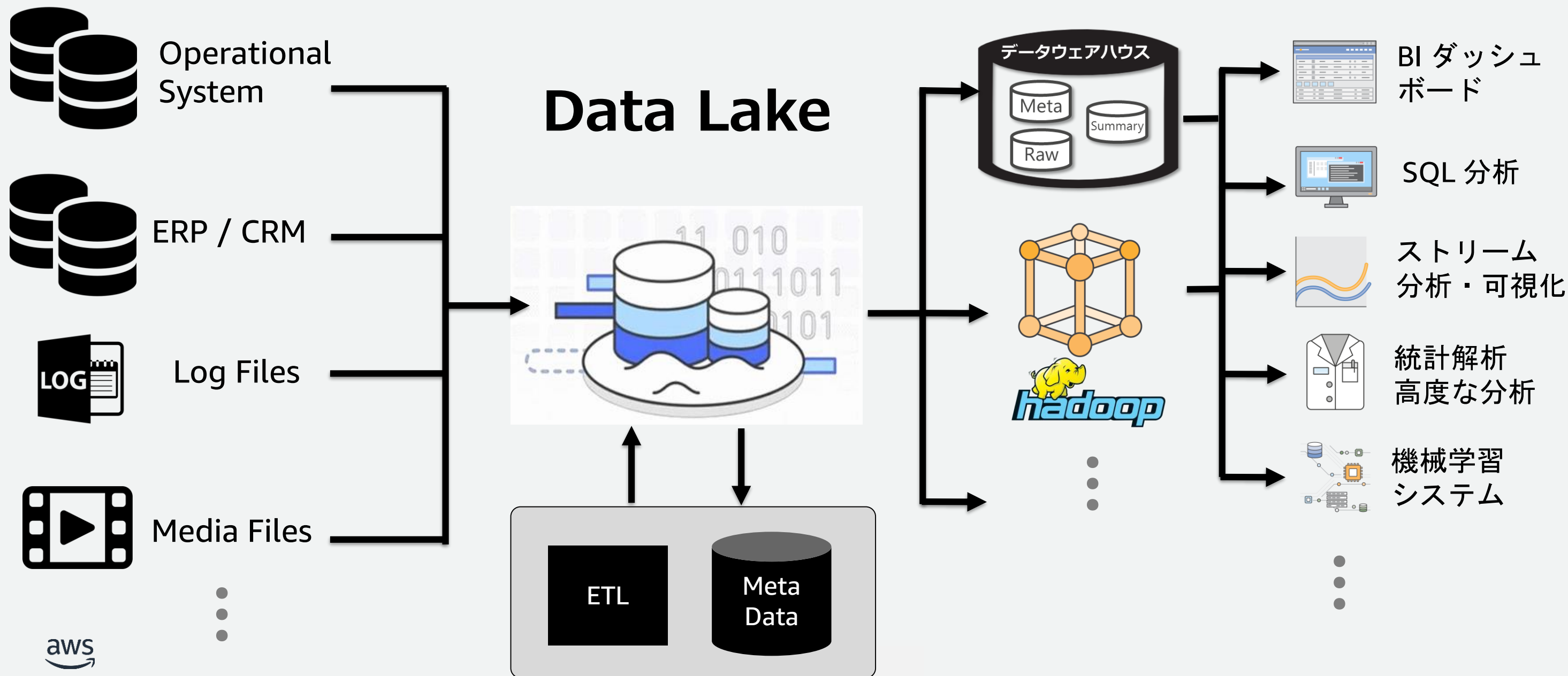
Right tool for the job

ツールに合わせて業務を決めてしまうのではなく、検証したい仮説、達成したいビジネスゴールに合わせて適切な指標・ツール・データを選択

- 主要なKPI の可視化 → BI ツール
- 探索的な仮説の発見 → SQL
- 不正取引の監視 → リアルタイム検知 & ダッシュボード
- 広告効果の精緻な検証 → ノートブックによる統計解析
- 商品のレコメンド → 機械学習を組み込んだシステム

利用したいツールに合わせて
適切な形にデータを加工整形しなければならない
そのため、大量の生データを保持する必要がある

データ分析の可能性を広げるデータレイク



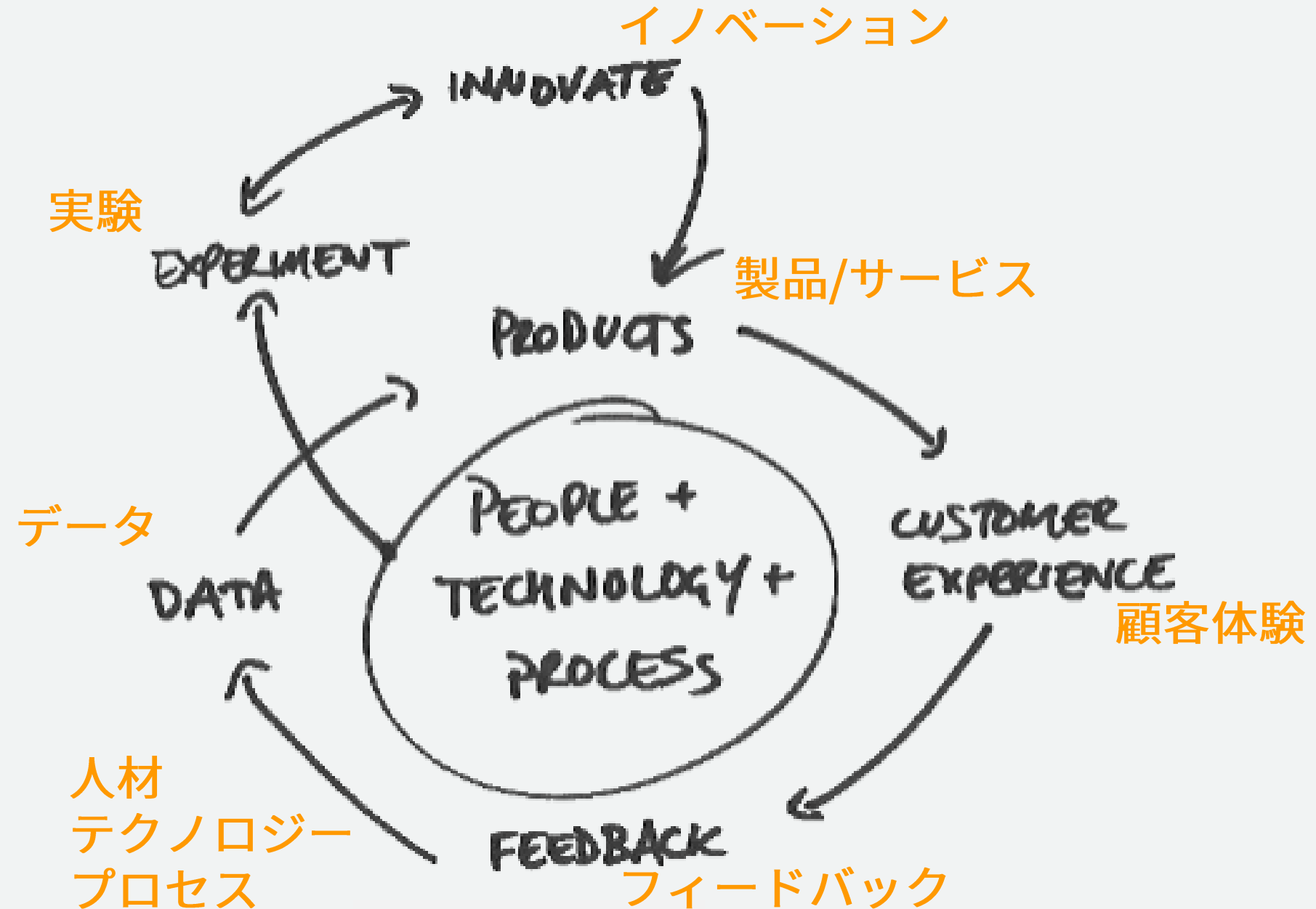
データ分析の可能性を広げるデータレイク



Amazon における データ活用の考え方とその実践

データの活用サイクルを回す

Data Flywheel



Amazon の多様なビジネスで求められるデータウェアハウス

多様で膨大なデータ分析のために、かつては世界最大規模の Oracle クラスターを利用していた

項目	値
コアデータサイズ	5PB+
総ストレージサイズ	35PB+
アクティブデータセット数	38,000+
ユーザー数	80,000+
クエリ数	200,000/day
ジョブ数	900,000/day

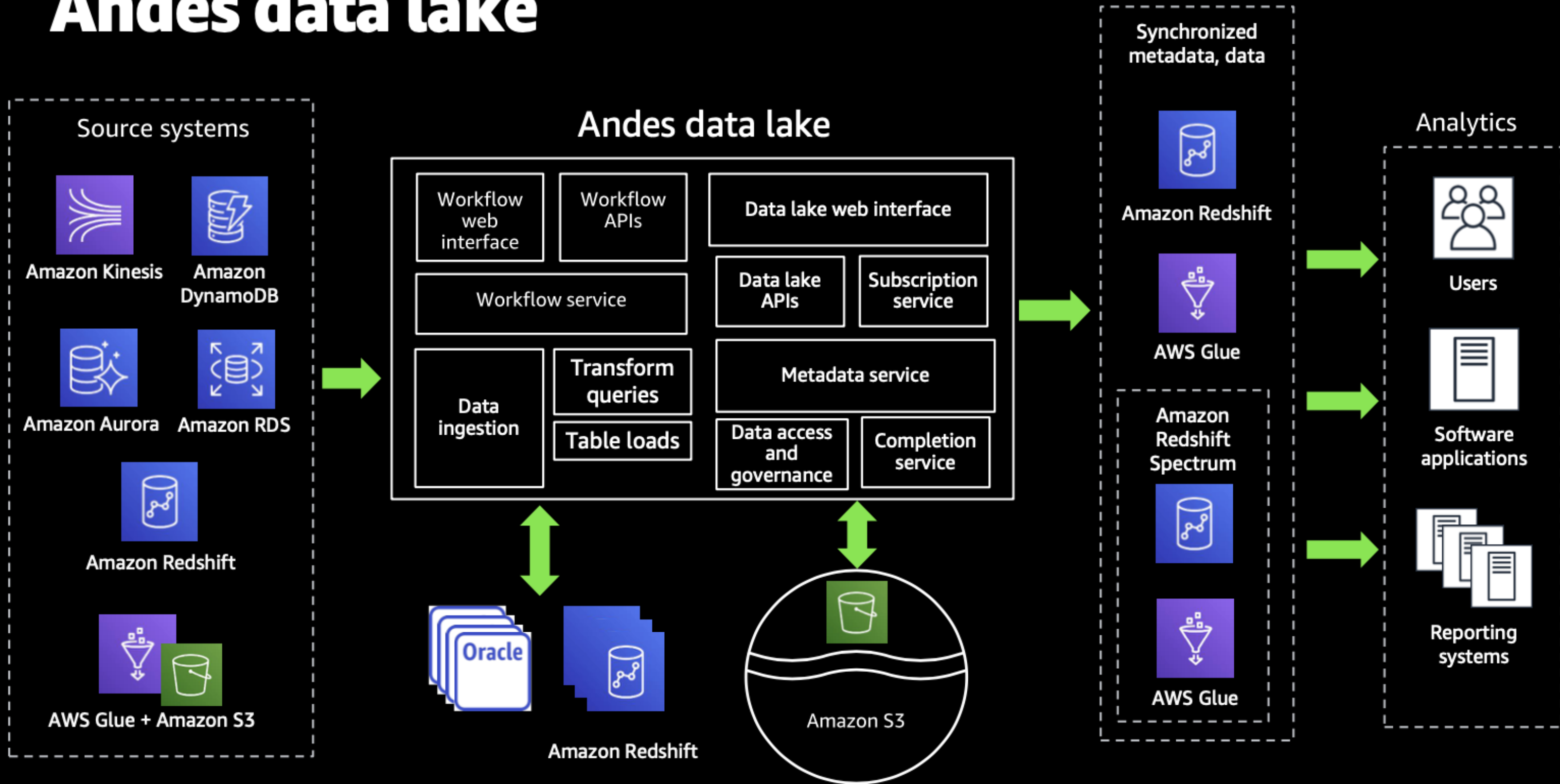


プロジェクト「Andes」

ゴールは 3 つ

1. Amazon のビジネスに合わせて拡張できるエコシステムを持つこと
2. 進化を可能にするオープンアーキテクチャ
必要な仕事に対してエンジニアが適したツールを選択できる
3. ビジネス需要を満たすサービスを構築および集約できるようにする
データレイクを構成するすべてのサービスには、ビルディングブロックとして提供される AWS サービスを組み合わせ、進化させる

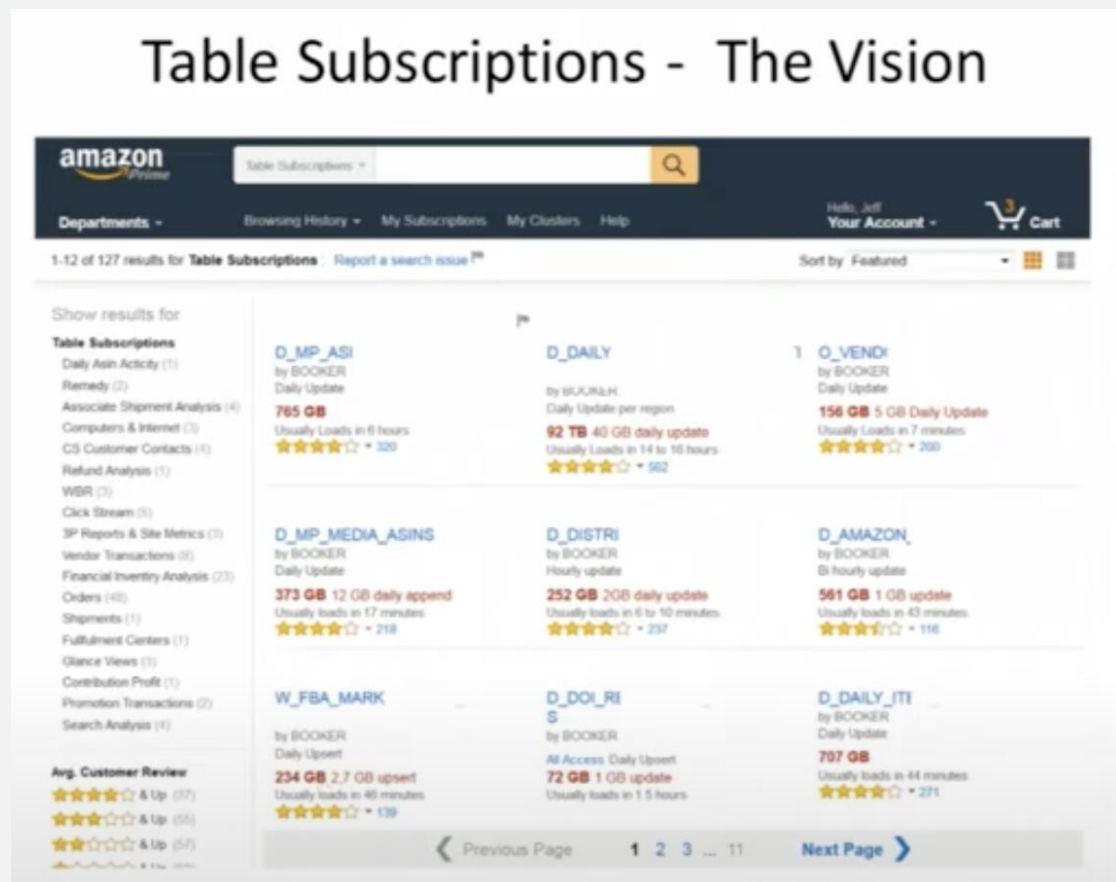
Andes data lake



Andes Data Lake 利用イメージ

要件

- テーブル名, フィールド名あるいは, ビジネス用語を入力して検索できる
- データセットが表示される
- データは部門, 業務などで絞り込みができる
- テーブル詳細では, スキーマなどを確認でき、必要なデータが含まれているかどうかユーザーが判断できる
- Subscribe すると, データの配置場所を指定できる
- 一度 Subscribe したデータに変更が発生すると, 通知を受けとれる

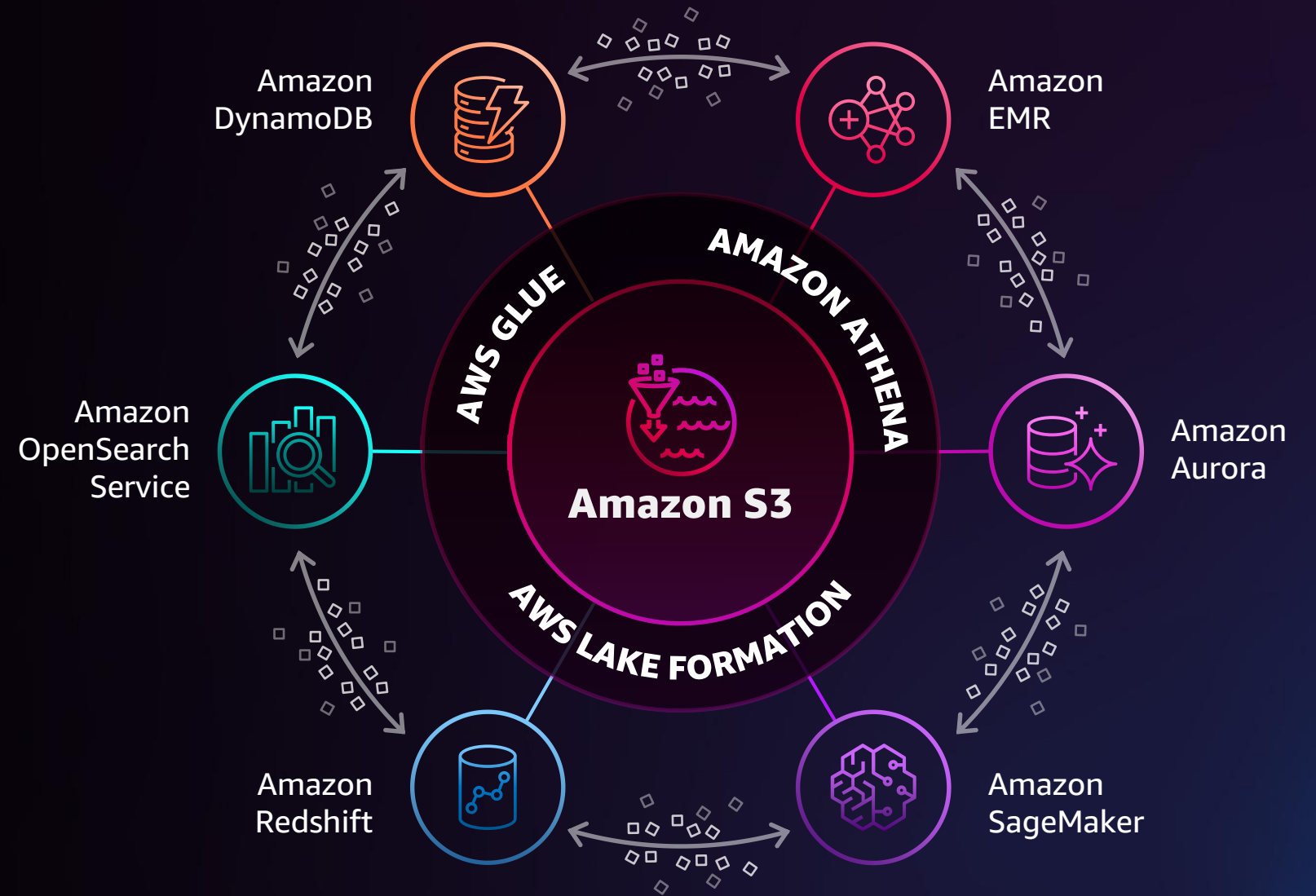


* Working Backwards を行ったときのイメージ



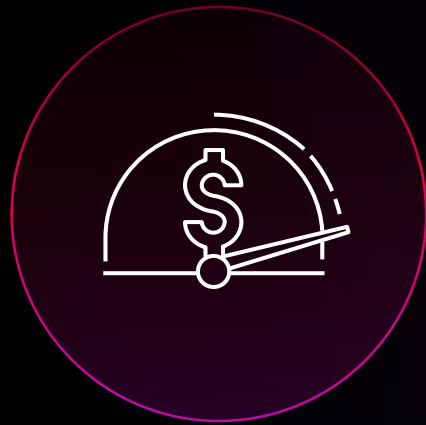
AWS アナリティクスサービスの 全体像とはじめの一步のご紹介

モダンデータ ストラテジー on AWS





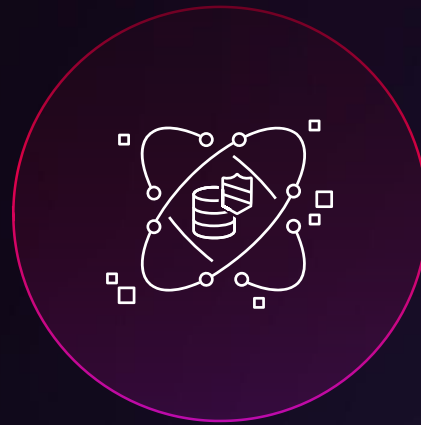
スケーラブルな
データレイク



パフォーマンスと
コストの観点から
サービスを使い分ける



サーバーレスと
使いやすさ



統合された
データアクセス
セキュリティ
ガバナンス



機械学習との統合



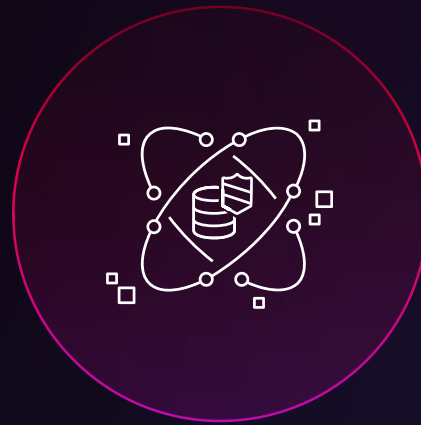
スケーラブルな
データレイク



パフォーマンスと
コストの観点から
サービスを使い分ける



サーバーレスと
使いやすさ

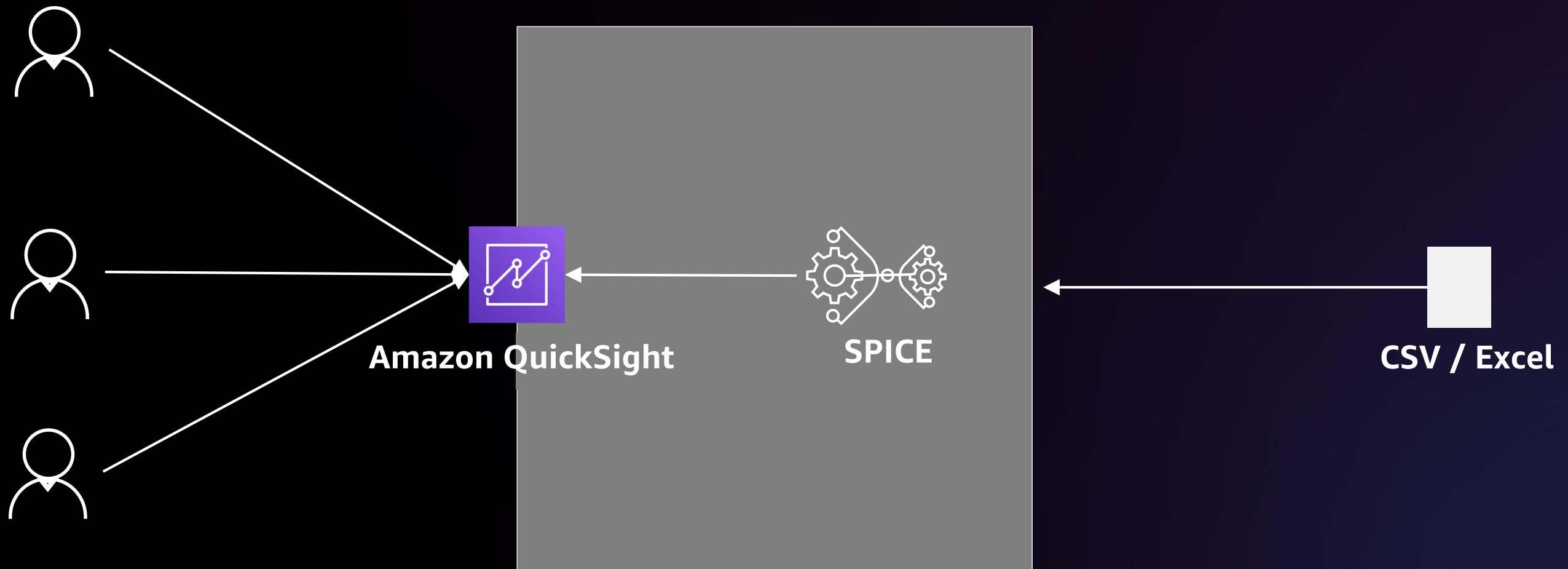


統合された
データアクセス
セキュリティ
ガバナンス



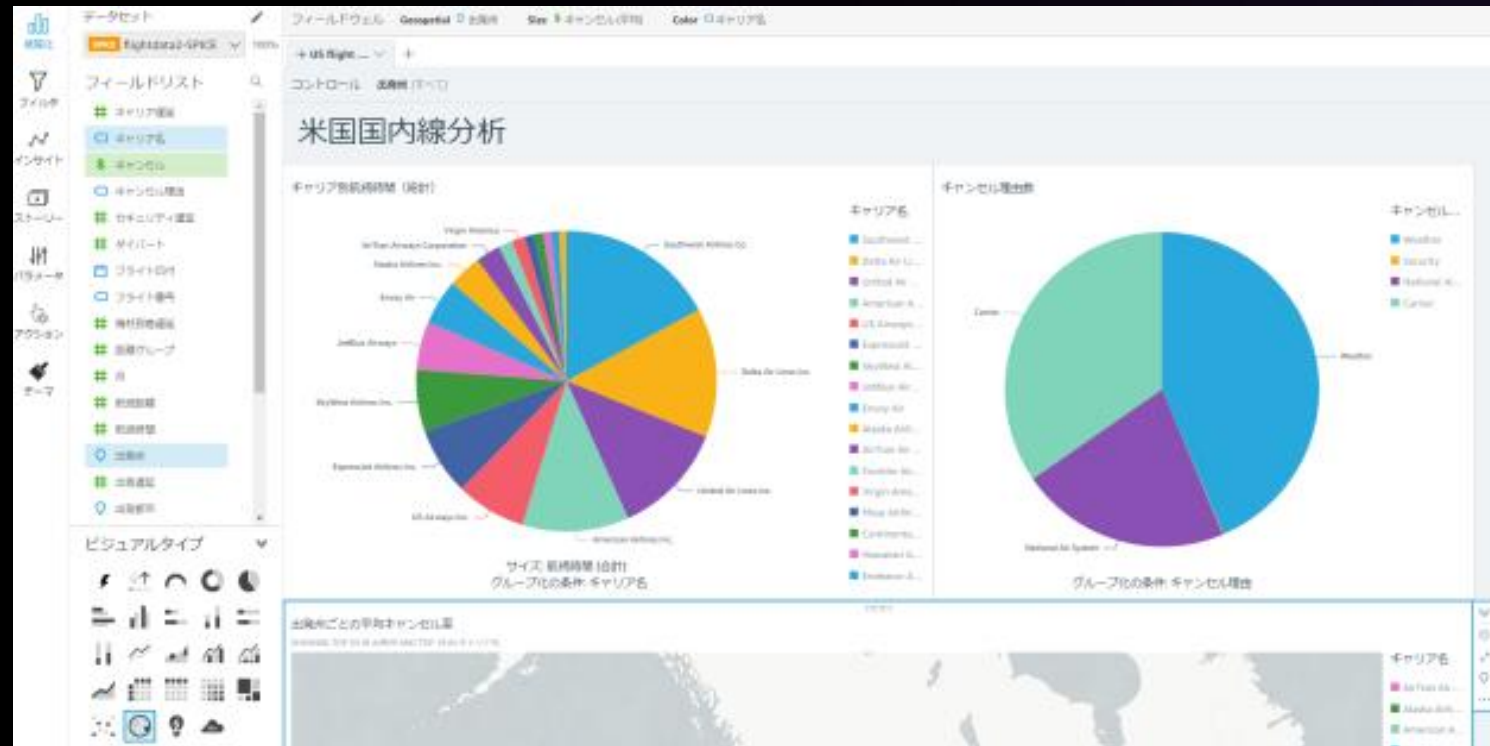
機械学習との統合

サーバーレス BI によるお手軽な可視化

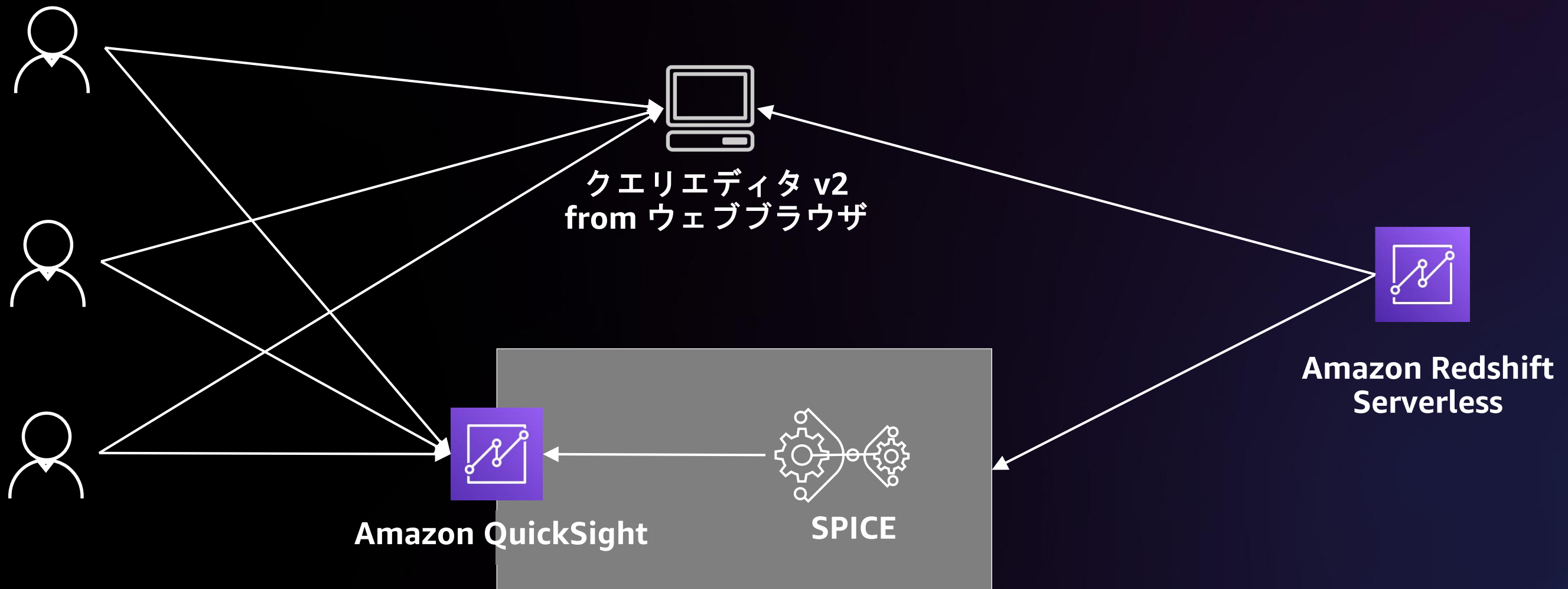


Amazon QuickSight

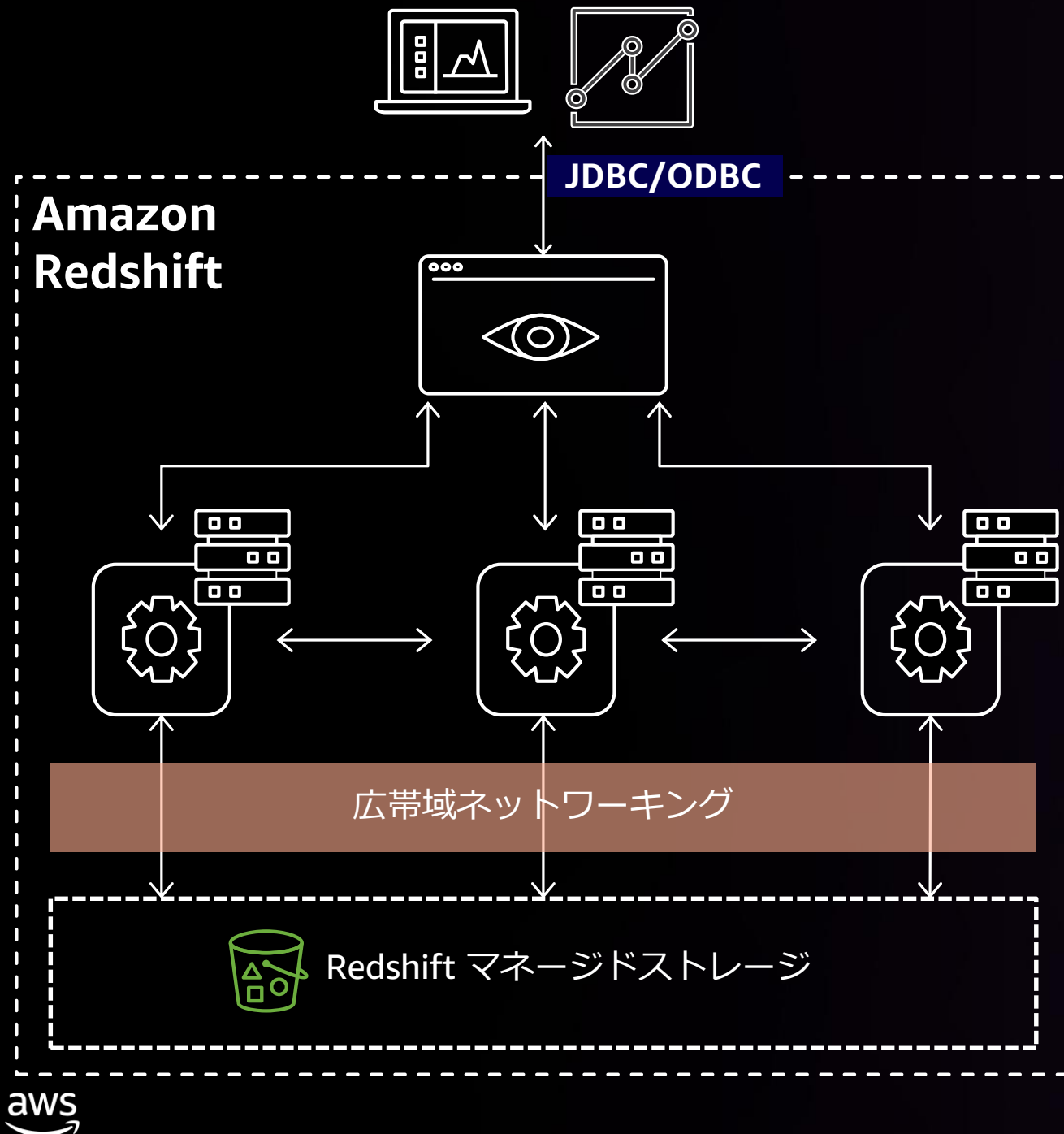
- すぐに使い始められる, サーバーレスの BI サービス
- データをインメモリ DB の SPICE に取り込んで, すばやく可視化
- 作成したダッシュボードをウェブ上で他の人に共有
- さまざまなデータに接続可能



サーバーレスデータウェアハウスによる クエリ分析ソリューション



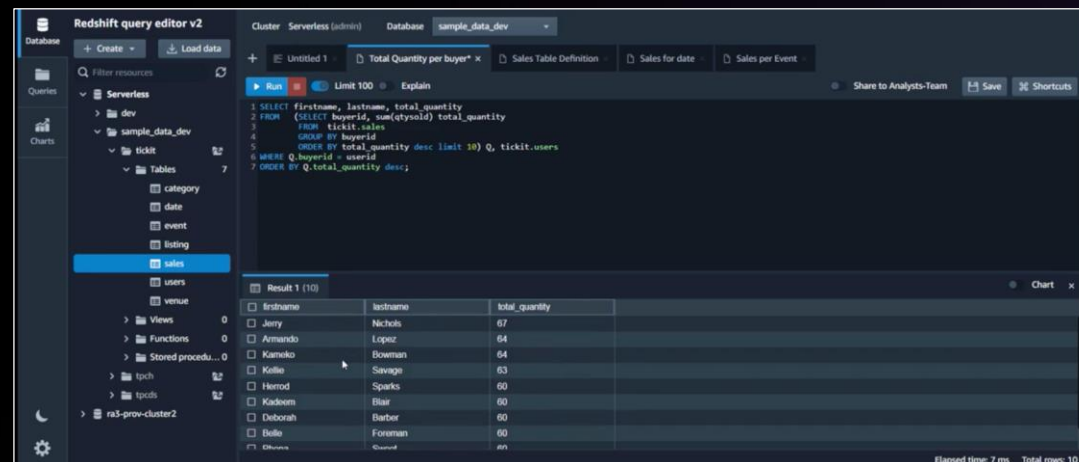
Amazon Redshift



- 構造化データだけでなく、JSON などの半構造化データも高速に分析できる、クラウドのデータウェアハウス
- 数分でクラスタが立ち上がり、チューニングやメンテナンスの必要なく簡単に使える。数千ものクエリを同時にさばける
- データレイク/他 AWS サービスとの高い親和性を持つ

Amazon Redshift **Serverless** (Public Preview)


- クラスターのプロビジョニングや管理の必要なく，サーバーレスの形でクエリが実行できる．ワークロードに合わせて，裏側でコンピューティングリソースがスケール．使わないときは自動停止
- 既存の Redshift の主要機能はすべてそのまま使えるため，純粋にインフラ部分を気にせず利用可能
- クエリ利用に応じた従量課金で，使わない間は課金なし．変動の大きなワークロードに最適



Get started with Amazon Redshift Serverless (Preview)

To start using Amazon Redshift Serverless (Preview), set up your Serverless endpoint and create a database.

Configuration

 Your data is encrypted by default with an AWS owned key. To choose a different key, choose **Customize settings**.

☒ Use default settings

Default settings have been defined to help you get started. You can change them at any time.

☐ Customize settings

Customize your settings for your specific needs.

Database name and password

Database name
dev

Admin user credentials
IAM credentials provided

Network and security

Virtual private cloud (VPC)
[vpc-e46a1289](#)

AWS KMS encryption
AWS owned KMS key

VPC security group
[sg-46629f29](#)

Audit logging
Off

Subnet

subnet-e76a128a, subnet-e16a128c, subnet-e66a128b

Permissions

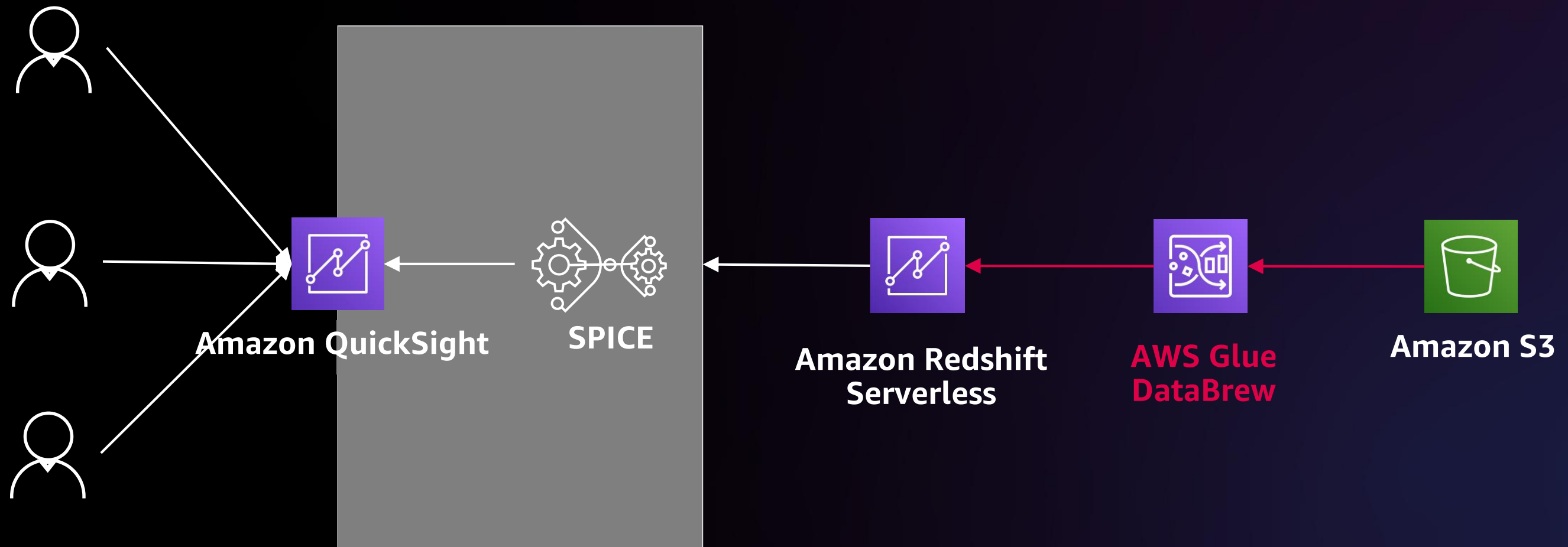
Default IAM role
Not assigned

IAM roles
Not assigned

Cancel

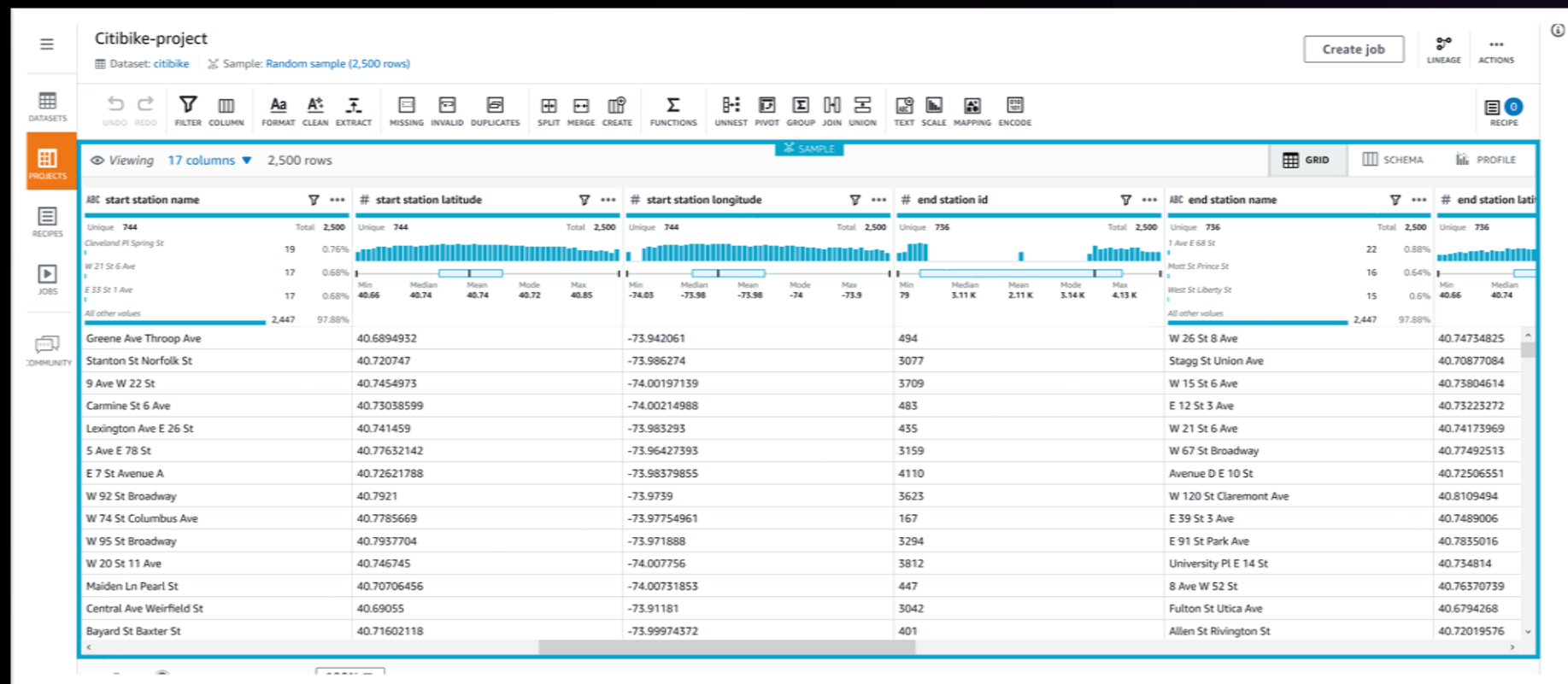
Create serverless endpoint

ビジュアルデータ加工ツールによる 前処理ソリューション



AWS Glue Databrew

- GUI でインタラクティブに前処理を記述できるデータ加工ツール
- データのプロファイリングや分布，カラム同士の相関等も確認可能
- サンプルングデータで記述した処理を，大規模データにジョブとして適用





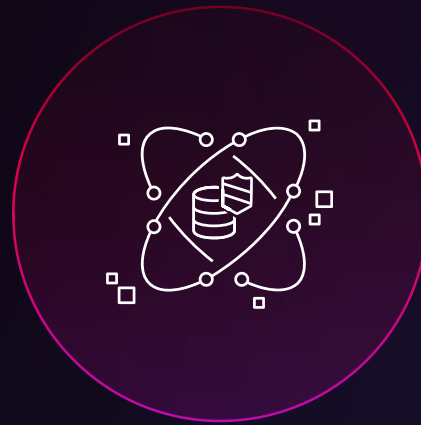
スケーラブルな
データレイク



パフォーマンスと
コストの観点から
サービスを使い分ける



サーバーレスと
使いやすさ



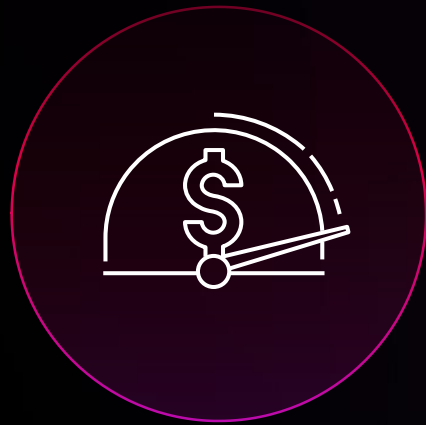
統合された
データアクセス
セキュリティ
ガバナンス



機械学習との統合



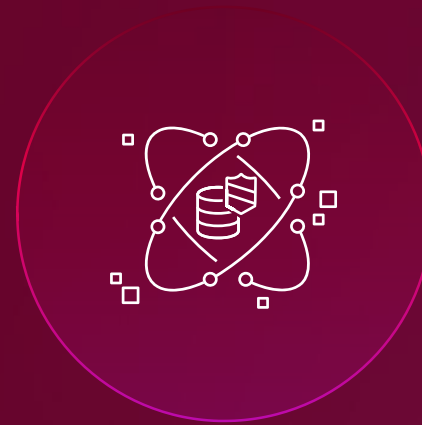
スケーラブルな
データレイク



パフォーマンスと
コストの観点から
サービスを使い分ける



サーバーレスと
使いやすさ

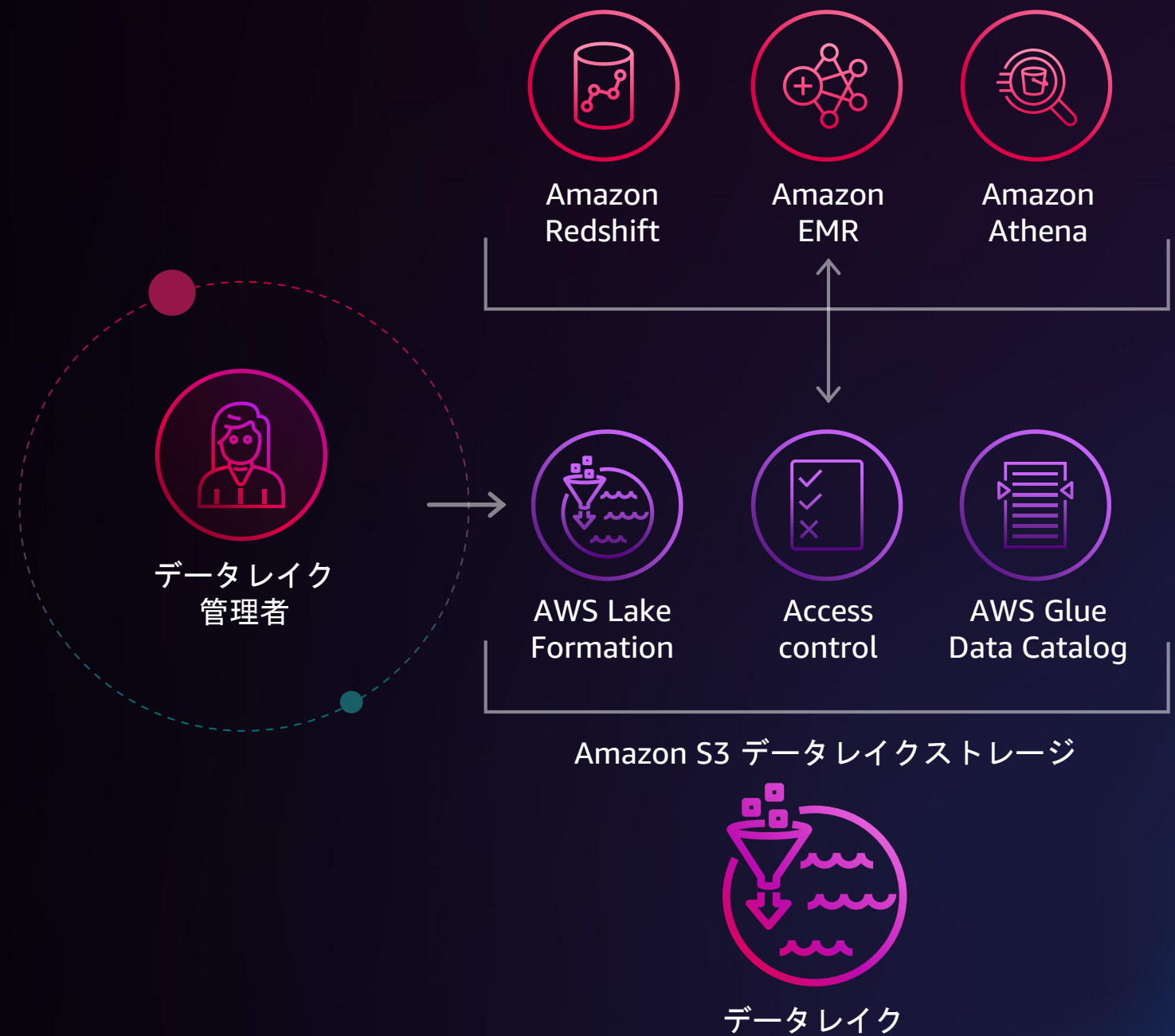


統合された
データアクセス
セキュリティ
ガバナンス



機械学習との統合

Lake Formation による シンプルなセキュリティ管理



AWS Glue: データ統合に関する豊富な機能

サーバーレスデータ統合サービス

スケーラブルな
データ取り込みの仕組み



ビルトインのデータ変換機能

Spark / Python の実行エンジン



モニタリング

統合・一元化された
データガバナンス



AWS Glue データカタログ



AWS Glue クローラー



Lake Formation

データソースへの接続と
データの取り込み



AWS Glue コネクター



AWS Glue コネクター
マーケットプレイス



多様なインターフェース

ユーザーの生産性と
Data Ops



個々のペルソナに応じたツール

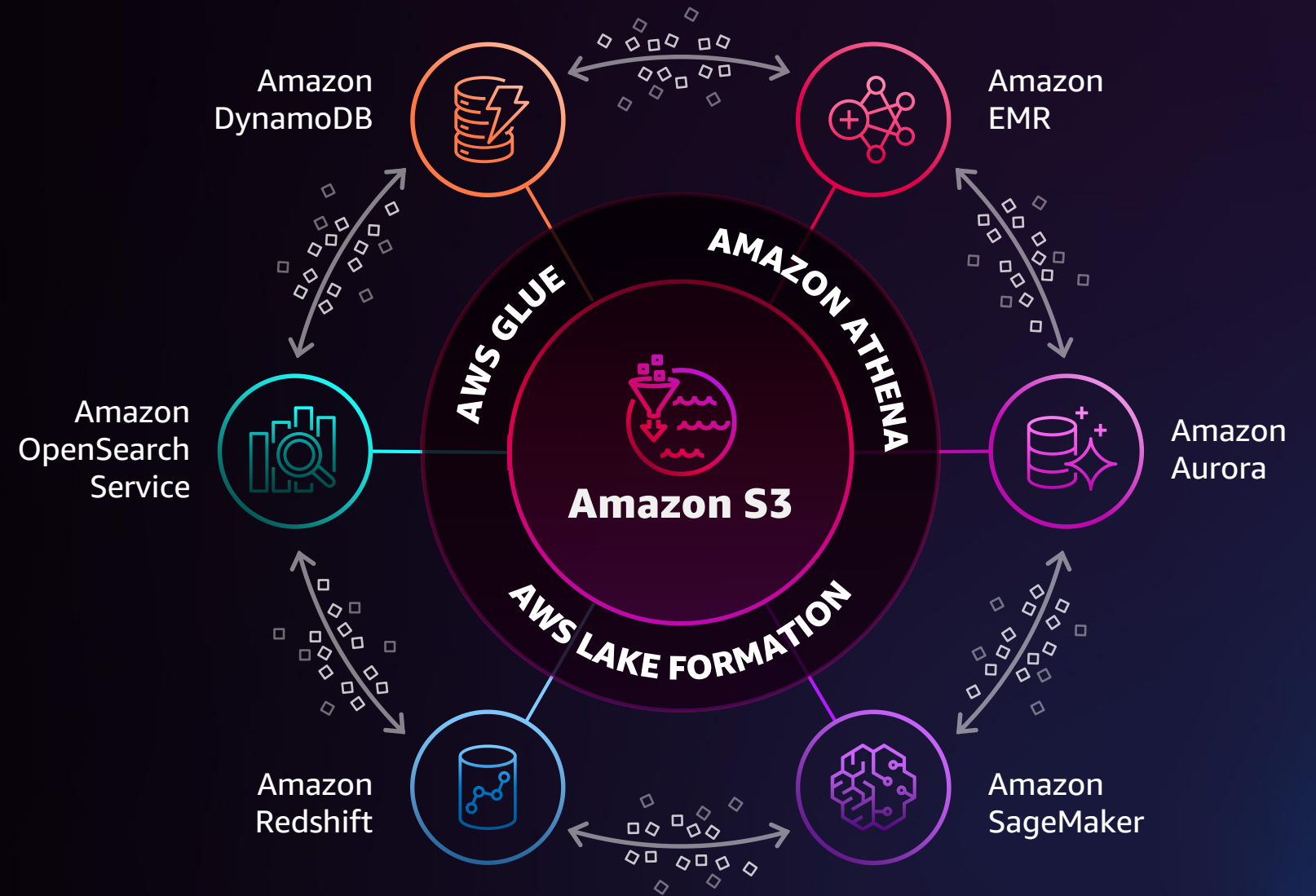


生産性を高めるためのツール群



Data Ops のためのツール群

モダンデータ ストラテジー on AWS





Thank you!

アマゾン ウェブ サービス ジャパン合同会社