



# [AWS Black Belt Online Seminar]

## 小売現場のデータを素早くビジネス に活用するAWSデータ基盤

リテールシリーズ

Solutions Architect 松本 和久  
2021/06

AWS 公式 Webinar  
<https://amzn.to/JPWebinar>



過去資料  
<https://amzn.to/JPArchive>



# AWS Black Belt Online Seminar とは



「サービス別」「ソリューション別」「業種別」のそれぞれのテーマに分け、アマゾン ウェブ サービス ジャパン株式会社が主催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の各サービスについてテーマごとに動画を公開します
- お好きな時間、お好きな場所でご受講いただけるオンデマンド形式です
- 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

# 内容についての注意点

- 本資料では2021年6月収録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(<http://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。
- AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at <http://aws.amazon.com/agreement/>. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

# 自己紹介

## 松本 和久

アマゾン ウェブ サービス ジャパン  
ソリューションアーキテクト  
エンタープライズ(流通・小売)のお客様を担当

### 好きなAWSサービス

- Amazon Connect



# アジェンダ

- 小売業界でデータ活用が求められる理由
- データ活用における課題
- データを活用できる基盤とは
- データ基盤の4要素に見る技術的な手立て
- まとめ

# 小売業界でデータ活用が 求められる理由

# 小売業界のチャレンジ

## 人口減少、少子高齢化

### 市場規模の縮小

新規顧客獲得のための差別化、  
既存顧客維持のためのエンゲージメント

### 人手不足

人力・直感ベースのプロセスの可視化、  
効率化と接客の自動化

## 不確実性

### ビジネスモデルの多様化

空間中心の既存モデルからタッチポイント  
中心のモデル、ハイブリッドモデルへ変化への  
対応

### テクノロジーの進化

デジタル化が目的ではなく、ユースケースを  
実現させるためのテクノロジー活用への  
ビジョンとリソース

### 消費者ニーズの多様化

しつこくない、コネクタされたパーソナライ  
ゼーションへの高い期待が今は当たり前

## コロナ禍

### 安全問題

お客様や従業員を守りながら、ビジネスの  
持続可能性(CSR)が課題に

### 供給問題

可視化、効率化、Real Time対応の  
必要性が鮮明に

# AWSにおける小売業界向けの6分野

## ストア



- 店舗：運用効率化、摩擦のない顧客経験
- Eコマース：俊敏性、弾力性

## カスタマー エンゲージメント



- カスタマーインサイト
- マーケティングROI
- パーソナライゼーション
- カスタマーセンター

## データ



- カスタマーインサイト
- レコメンデーション
- 運用効率化
- 予測
- 可視化、最適化

## サプライチェーン



- ウェアハウスマネジメント
- ロジスティクス
- 運送マネジメント
- 在庫管理

## マーチャンダイジング



- スペースプランニング
- カテゴリーマネジメント
- アロケーション
- 最適化

## エンタープライズ コア アプリケーション



- マイグレーションによるコスト削減
- モダナイゼーション（マイクロサービス、サーバーレス、目的別データベース）



# 小売業のビジネスを牽引するドライバー

データからのインサイトなしでは成り立たない時代へ



ニーズに合う商品を  
適切なチャネルで、  
適切な価格で提供



顧客生涯価値  
(CLV)の  
向上



マージン向上  
のための  
コスト管理



イノベーションと  
結果の加速

# データはあるが、インサイトが足りない



**61%**

の小売業者は、  
現在起こっていることに  
反応するため  
に十分な速さで  
情報入手ができない



**88%**

の小売業者は、  
アクションにつなぐ  
インサイトを  
生み出すのに  
苦労している



**43%のみ**

の小売業者が  
持つデータが  
アクションに  
つなげられる  
データである

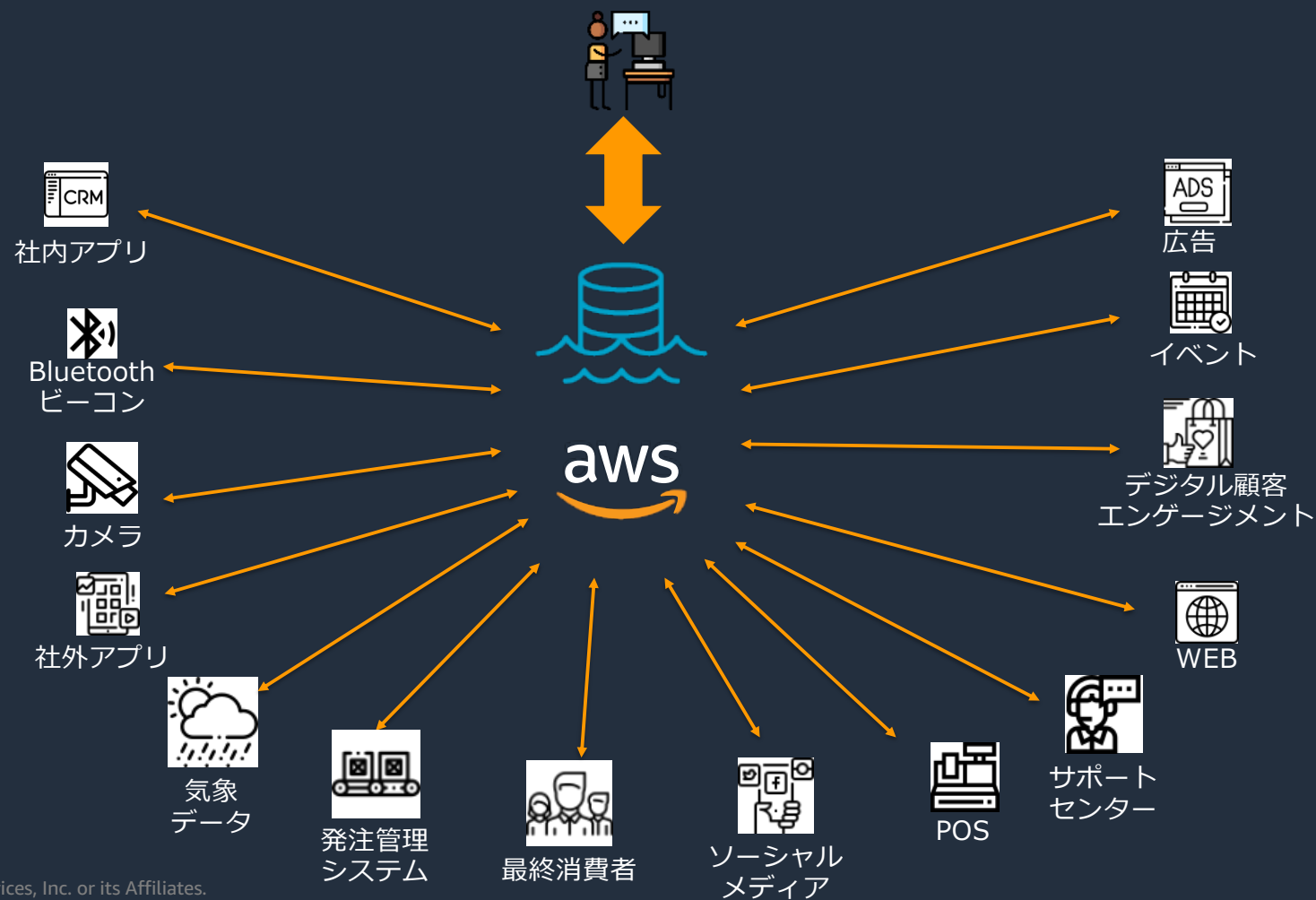


**16%のみ**

の小売業者が  
データを活用して  
競争優位性を  
特定して  
発展させている

# データ活用における課題と対策

# カスタマージャーニー視点で活用するビッグデータ



# 課題：ビッグデータの価値を生み出すことへの壁

データの品質が悪い



88%

の小売業者は、  
アクションにつなげるのに  
問題を感じている

単一のビューになってない



51%

の小売業者は、  
消費者の統一された姿を  
持っていない

データへのアクセスが困難



51%

の小売業者は、  
システム間でデータを  
共有できない

# 課題：ビッグデータの価値を生み出すことへの壁

データの品質が悪い



88%

の小売業者は、  
アクションにつなげるのに  
問題を感じている

単一のビューになってない



51%

の小売業者は、  
消費者の統一された姿を  
持っていない

データへのアクセスが困難



51%

の小売業者は、  
システム間でデータを  
共有できない

適切なプラットフォームがないと、データからインサイトは得られないまま

# 対策：データレイクで大きなビジネスインパクトをドライブ

## 消費者インサイトの増加



**66%**

の小売業者は  
インサイトによる  
消費者理解が飛躍的に  
増えたと感じている

## 売り上げの増加



**64%**

の小売業者は  
売り上げの増加を実感

## マルチチャネル



**54%**

複数のチャネルからの  
売り上げの増加

適切なプラットフォームを構築すると、インサイトを得られるようになる

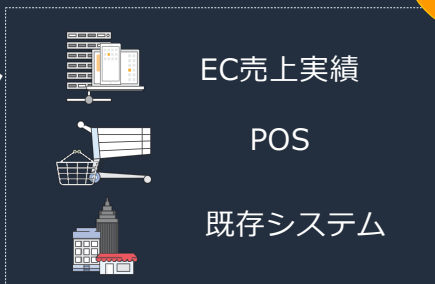
# データを活用できる基盤とは



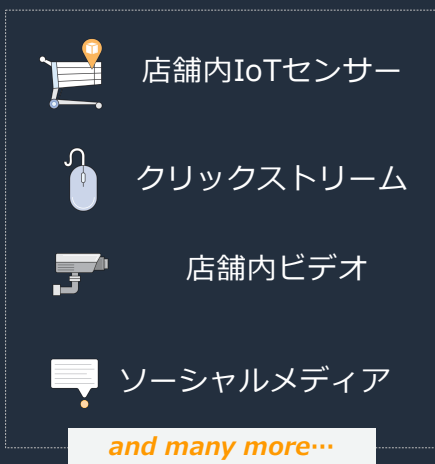


# 小売データ活用の将来像

## トラディショナルなデータ

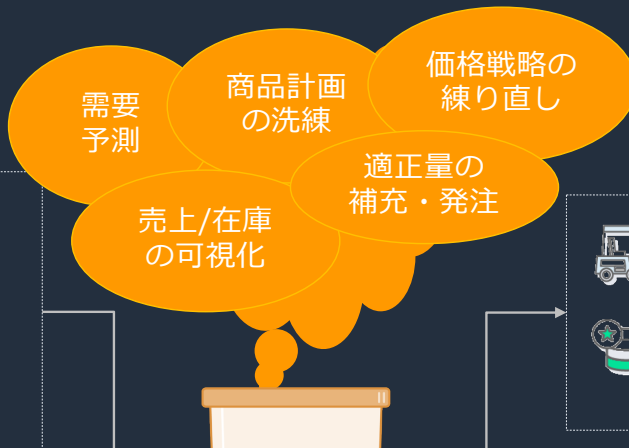


## 新しいタイプのデータ



## データソースを統合

データレイク



## トラディショナルなデータ分析



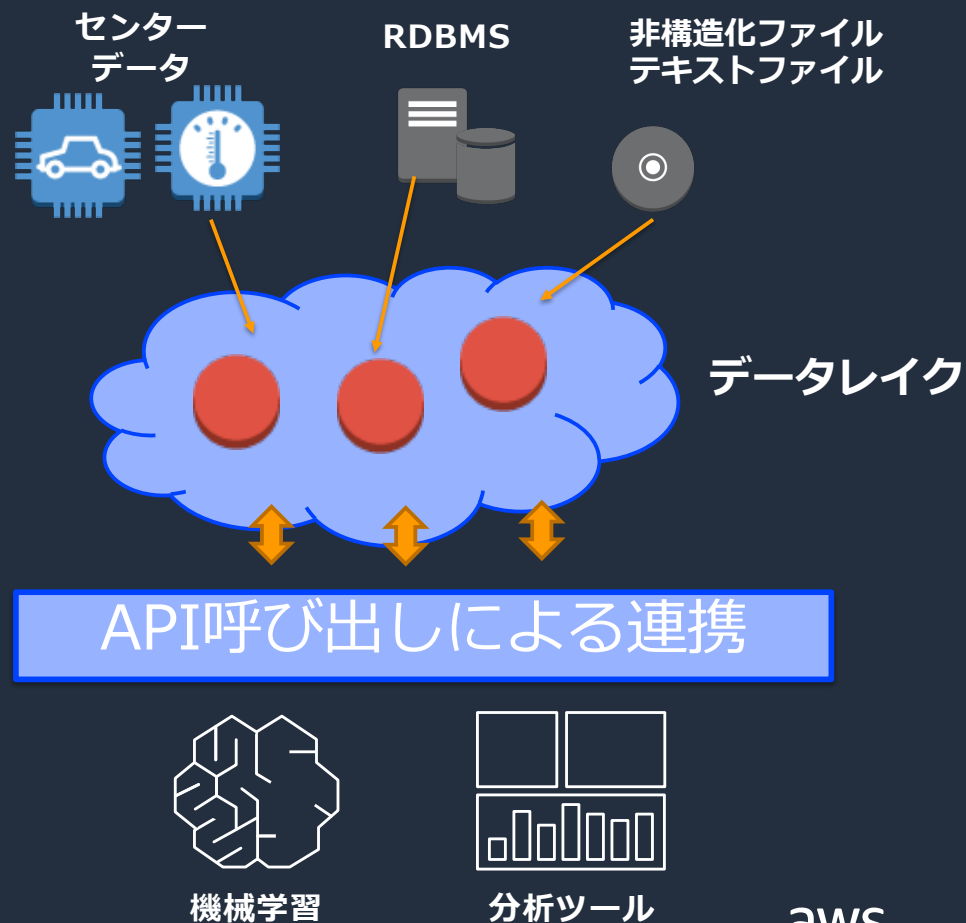
## アドバンスドなオンデマンドのデータ分析



# データレイクとは

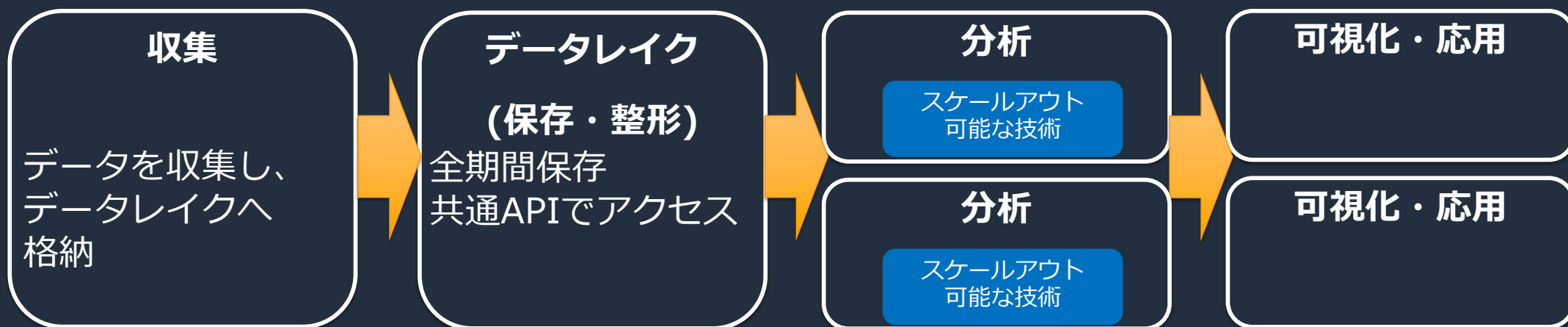
- 多様なデータを一元的に保存
- データを失わない
- サイズ制限からの開放
- 決められた方法（API）ですぐにアクセスできる

→システム全体のハブ

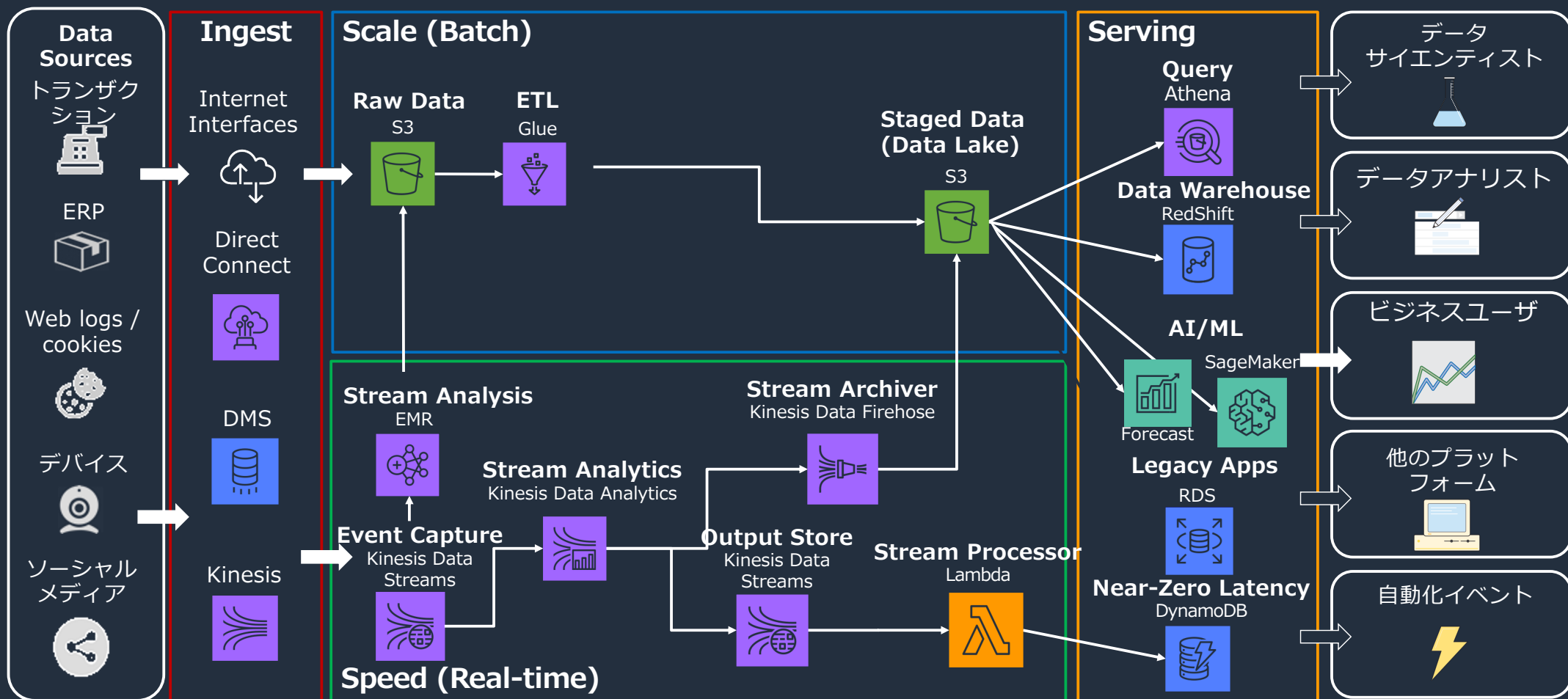


# データレイクを中心としたデータ分析基盤

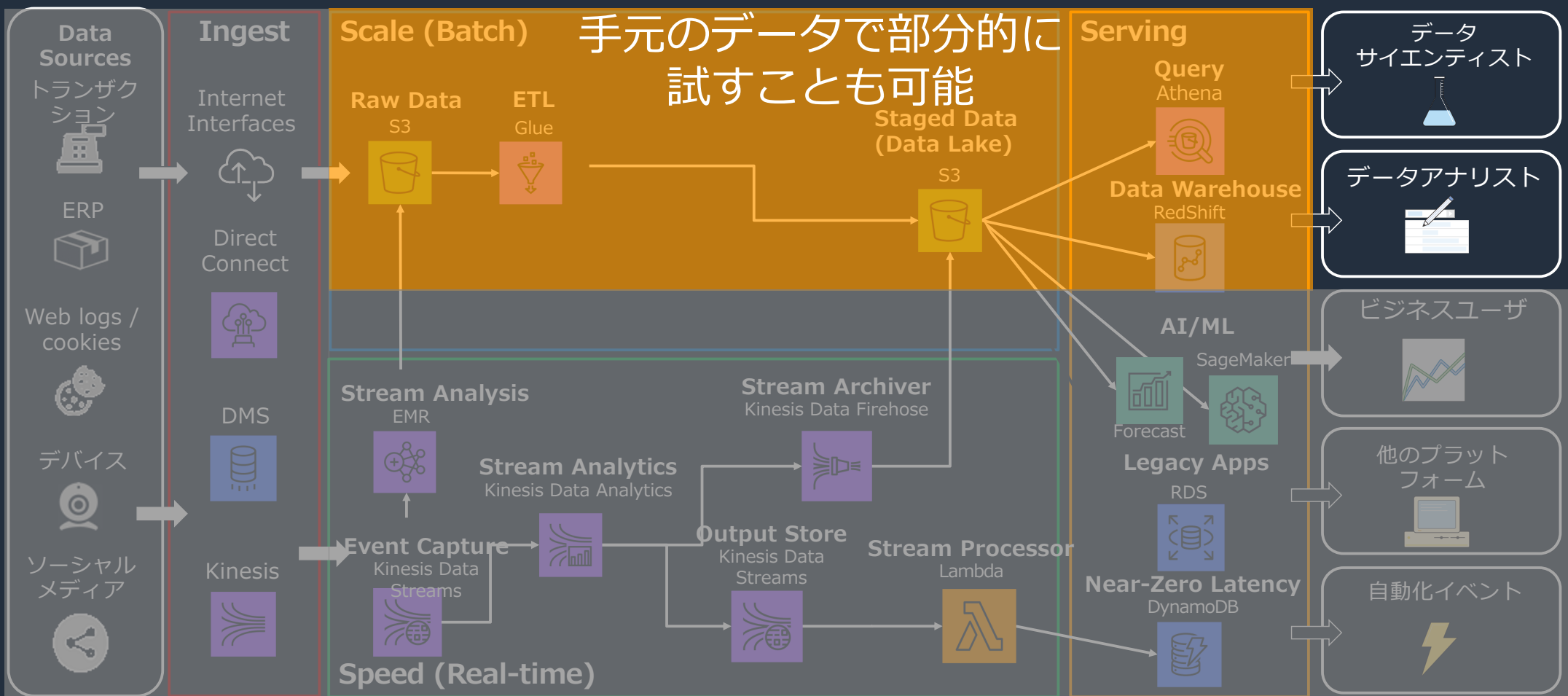
- 保存したデータのカタログ作成やETLを行う
- 必要なところに、必要な技術を適用する



# AWSが考えるモダンデータアーキテクチャ



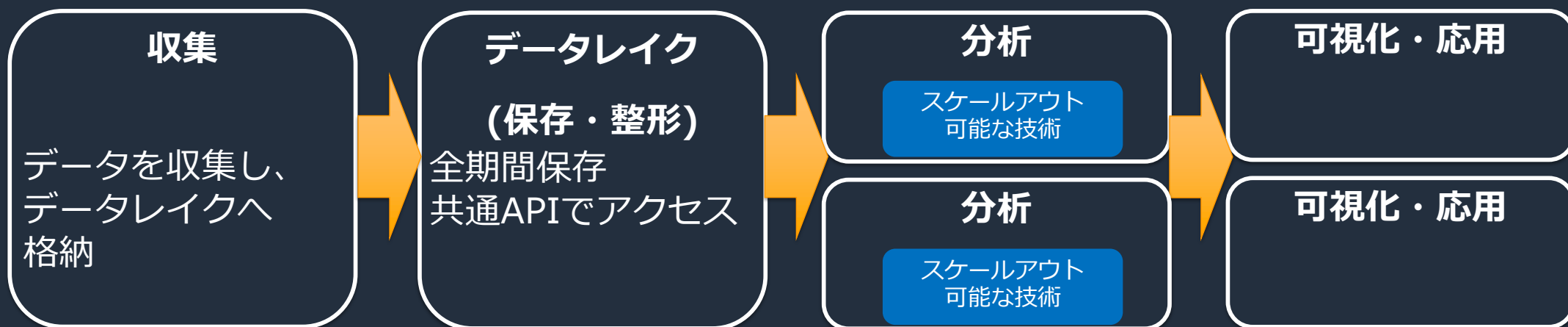
# AWSが考えるモダンデータアーキテクチャ



# データ基盤の4要素に見る 技術的な手立て

# データレイクを中心としたデータ分析基盤（再掲）

- 保存したデータのカタログ作成やETLを行う
- 必要なところに、必要な技術を適用する



# データを収集する



# 売上データをリアルタイムに収集する

## 課題

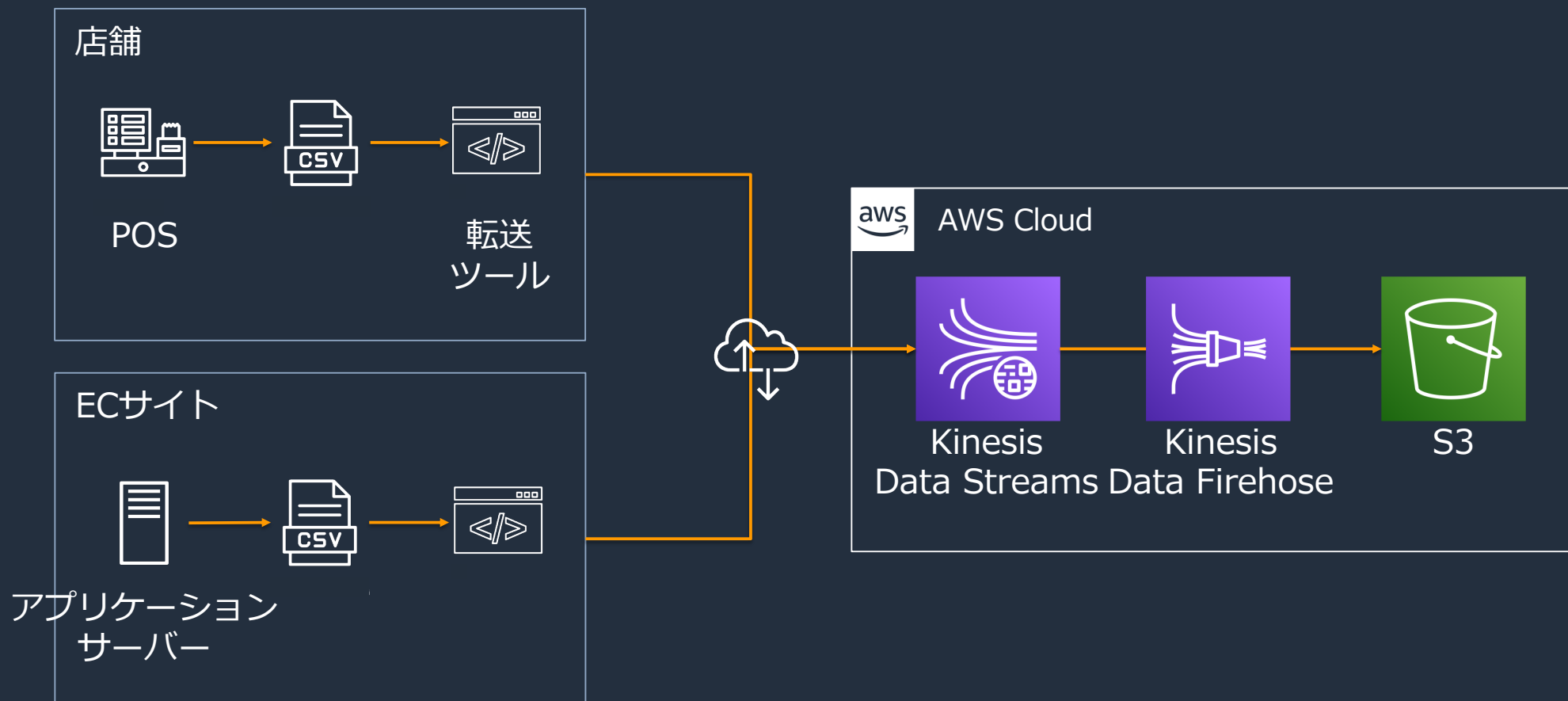
- 毎日のバッチ処理で集計しているため、在庫変動に時間がかかり、機会損失が生じている
- 売上データをリアルタイムに収集して在庫と連動する仕組みが必要

## 収集の要件

- 売上/在庫をリアルタイムに可視化するために、
- すべてのPOSデータを、
- リアルタイム or ニアリアルタイムに、
- POS/ストコンから、
- 既存ネットワークを経由してクラウドへ送る

# 売上データをストリームとして送る

収集



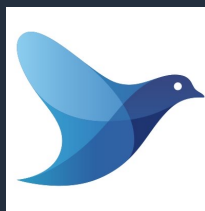
# 転送ツールの選択

収集



Kinesis Agent

- AWSサービスとの統合が容易
- Agent自体のメトリクスを管理可能



Fluentd

- ログ管理OSSのデファクト
- 様々なプラグインが用意されている



SDK

- 好みの言語で開発可能
- カスタムロジックを実装可能

# Kinesis Agent for Microsoft Windows 設定例

収集

```
{
  "Sources": [
    {
      "Id": "PosLog",
      "SourceType": "DirectorySource",
      "Directory": "C:¥¥MyService¥¥logs",
      "FileNameFilter": "*.csv",
      "RecordParser": "Delimited",
      "Delimiter": ",",
      "TimeZoneKind": "local",
      "Headers": "Recodrd-Date,Record-Time,StoreId,PosId,UUID,ItemId,Num",
      "TimestampField": "{Date} {Time}",
      "TimestampFormat": "yyyy/MM/dd HH:mm:ss",
      "InitialPosition": "Bookmark"
    }
  ],
}
```

Source宣言でログデータを収集する場所と対象を指定する

[次ページへ](#)

# Kinesis Agent for Microsoft Windows 設定例

収集

続き

```
"sinks": [  
  {  
    "Id": "MyKinesisFirehoseSink",  
    "SinkType": "KinesisFirehose",  
    "Region": "ap-northeast-1",  
    "StreamName": "PosLogFirehoseStream",  
    "Format": "json"  
  }  
],  
"pipes": [  
  {  
    "Id": "ApplicationLogTotestKinesisFirehoseSink",  
    "SourceRef": "PosLog",  
    "SinkRef": "MyKinesisFirehoseSink"  
  }  
]  
}
```

Sink宣言でログデータを送信する場所と形式を指定する

Pipe宣言でSourceをSinkに接続する

[https://docs.aws.amazon.com/ja\\_jp/kinesis-agent-windows/latest/userguide/configuring-kinesis-agent-windows.html](https://docs.aws.amazon.com/ja_jp/kinesis-agent-windows/latest/userguide/configuring-kinesis-agent-windows.html)

© 2021, Amazon Web Services, Inc. or its Affiliates.



# Kinesis Agent for Microsoft Windows での出力例

収集

Sink宣言でFormatをJSONLにした結果

```
{"Record-Date":"2021/05/02","Record-  
Time":"11:00:00","StoreId":"00001","PosId":"001","UUID":"aaaa1234-bb01-cc23-  
abcd01234567","ItemId":"1","Num":"1"}  
{"Record-Date":"2021/05/02","Record-  
Time":"11:01:00","StoreId":"00001","PosId":"001","UUID":"aaaa1234-bb01-cc23-  
abcd2345678","ItemId":"2","Num":"2"}  
{"Record-Date":"2021/05/02","Record-  
Time":"11:02:00","StoreId":"00001","PosId":"001","UUID":"aaaa1234-bb01-cc23-  
abcd3456789","ItemId":"3","Num":"3"}
```

# 店舗とAWSのネットワーク接続の選択

収集

VPNで直接接続

店舗にVPN装置を配置し、AWS Transit Gatewayを介してVPCに接続する

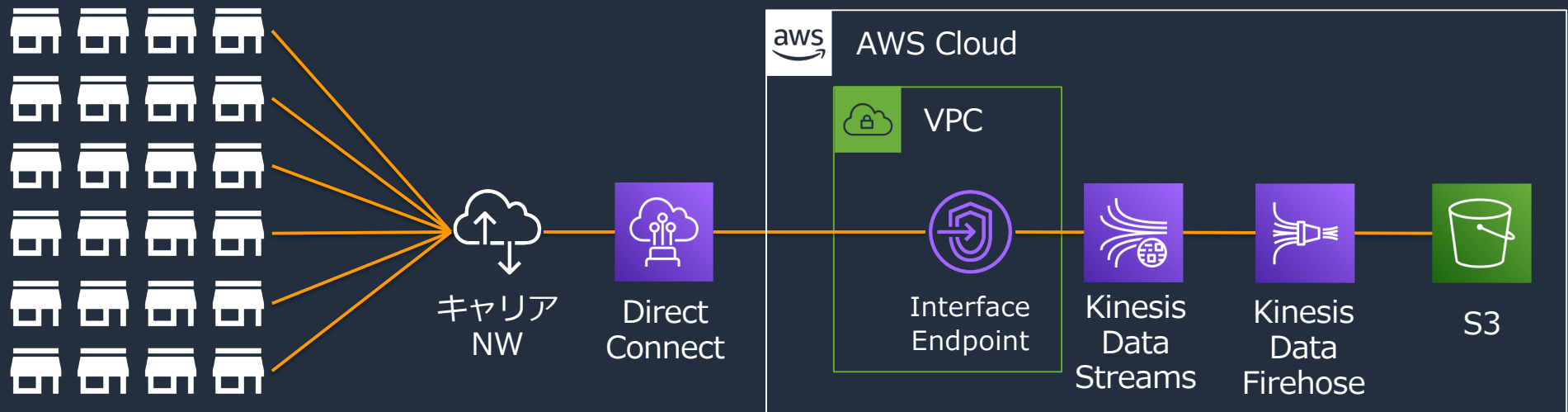
キャリアNWを介して接続

店舗が接続する既存のキャリアNWからAWS Direct Connectを経由して、VPCに接続する

# 店舗とのネットワーク接続

## キャリアNWを介して接続

- キャリアNWからDirect Connectに接続し、経路を束ねる
- 既存NWを用いるため、店舗側に設備投資がかからない

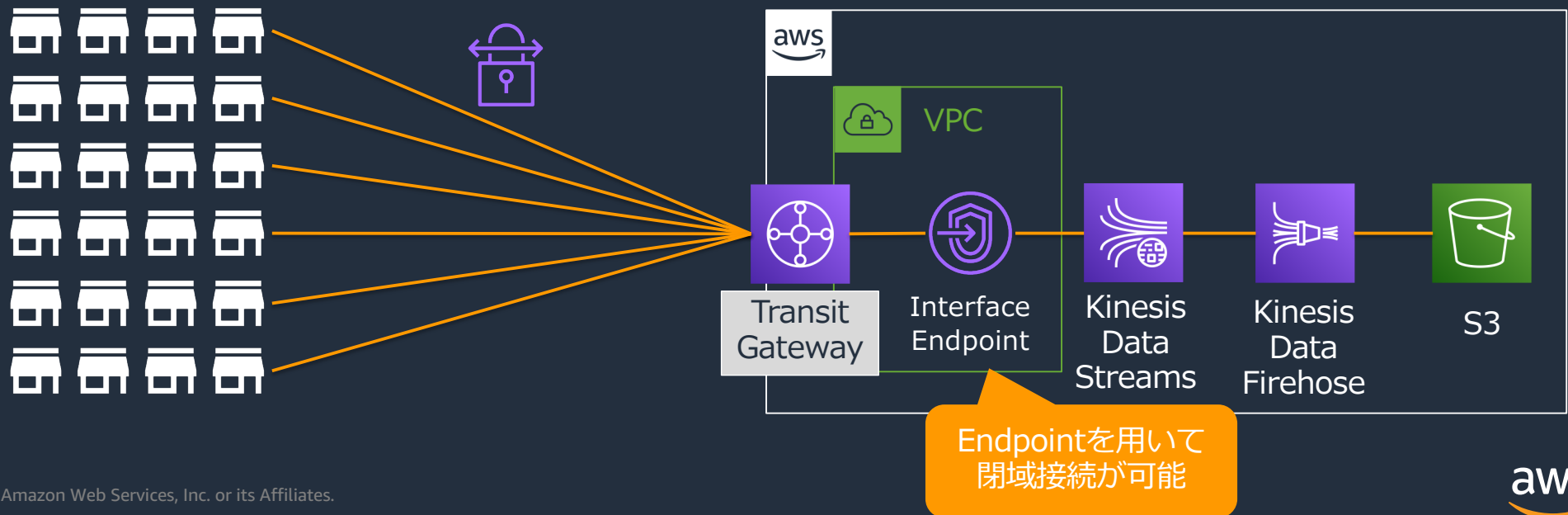




# 店舗とのネットワーク接続

## VPNで直接接続

- Transit Gatewayには最大5,000VPN拠点(VPC, Direct Connectを拠点数に含む)まで接続可能
- NWのキャパシティを考慮する要素が少ない(店舗ごとのVPN接続のみ)



Q) POS/ストコンの接続先がhostsファイルで固定指定されている場合はAWSへの接続の可用性設計をどうする？

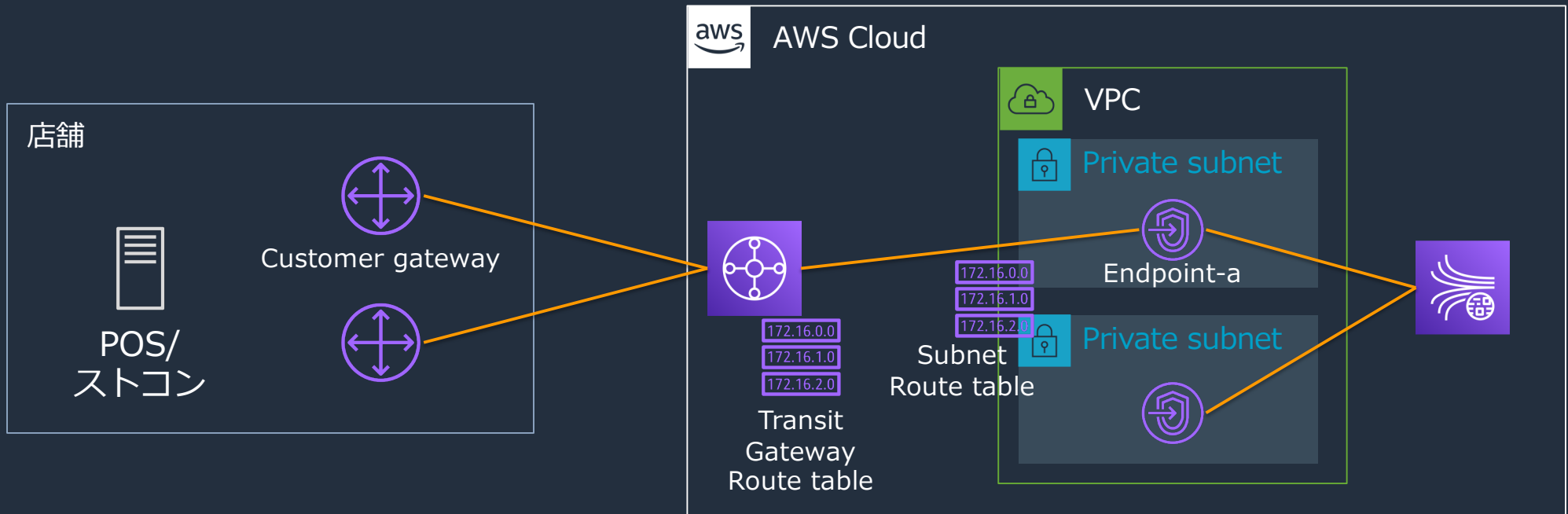
# 店舗とのネットワーク接続

## Answer

- 本当にPOS/ストコンから名前解決ができないかを確認する
- どうしてもできない場合は、Transit Gateway上に固定の仮想IPを作成し、これに接続する

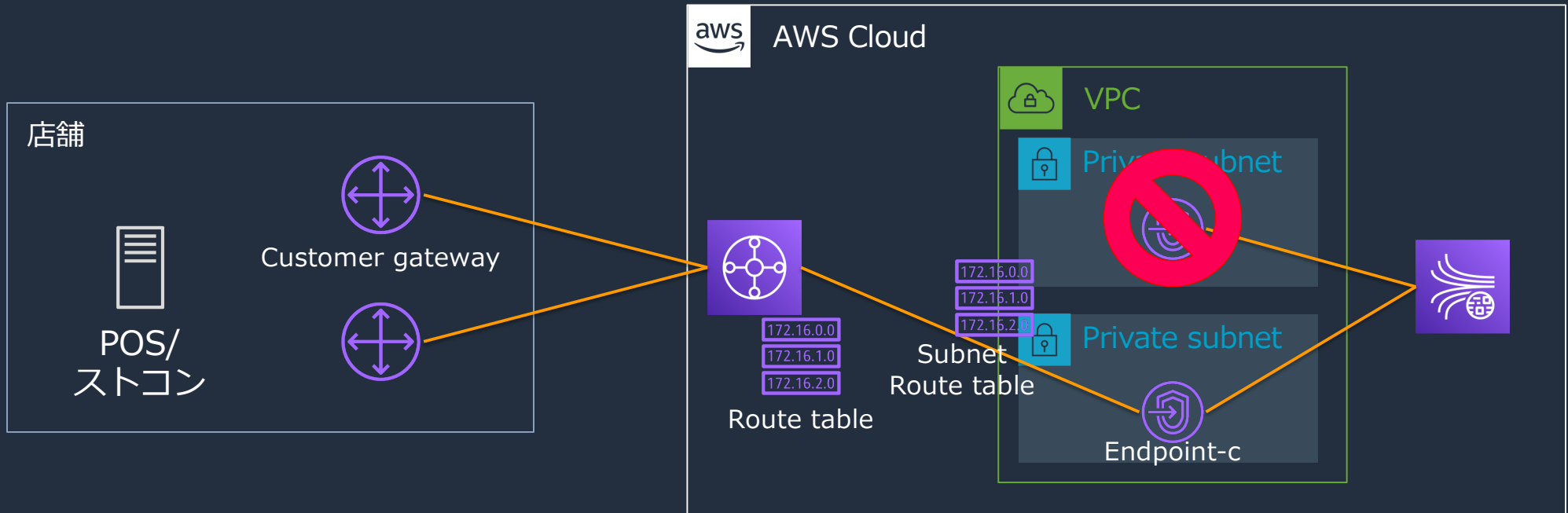
# 店舗とのネットワーク接続

- 店舗からはTransit Gateway上の固定の仮想IP(VPC外のアドレス)に接続し、Transit GatewayのRoute tableでVPCに静的ルーティングする
- VPCサブネットのRoute tableで仮想IPから片側のEndpointに静的ルーティングする



# 店舗とのネットワーク接続

- 障害時はサブネットのRoute tableを書き換える
- 障害を自動検知する仕組みをAWS Lambdaを用いて実装する



# データを保存・整形する

# 売上データを保存する

## 課題

- 売上データを保管するPOS/ストコンの可用性が低い
- 長期間にわたるデータ保存もしたいが容量に限りがある

## 保存の要件

- 売上/在庫をリアルタイムに可視化するために、
- すべてのPOSデータを、
- 消失すること無く、
- 数年にわたって保存する

# データレイクに最適なAmazon S3

任意の量のデータを高い耐久性で保存できるオブジェクトストレージ

- 店舗からの生データを任意の形式で保存
- 容量の上限無し
- 高い耐久性: 99.999999999%
- 低コスト: \$0.025/GB/月\*
- 多様な権限管理と、暗号化によるセキュリティ
- 様々なプログラム言語やサービスとAPIで連携

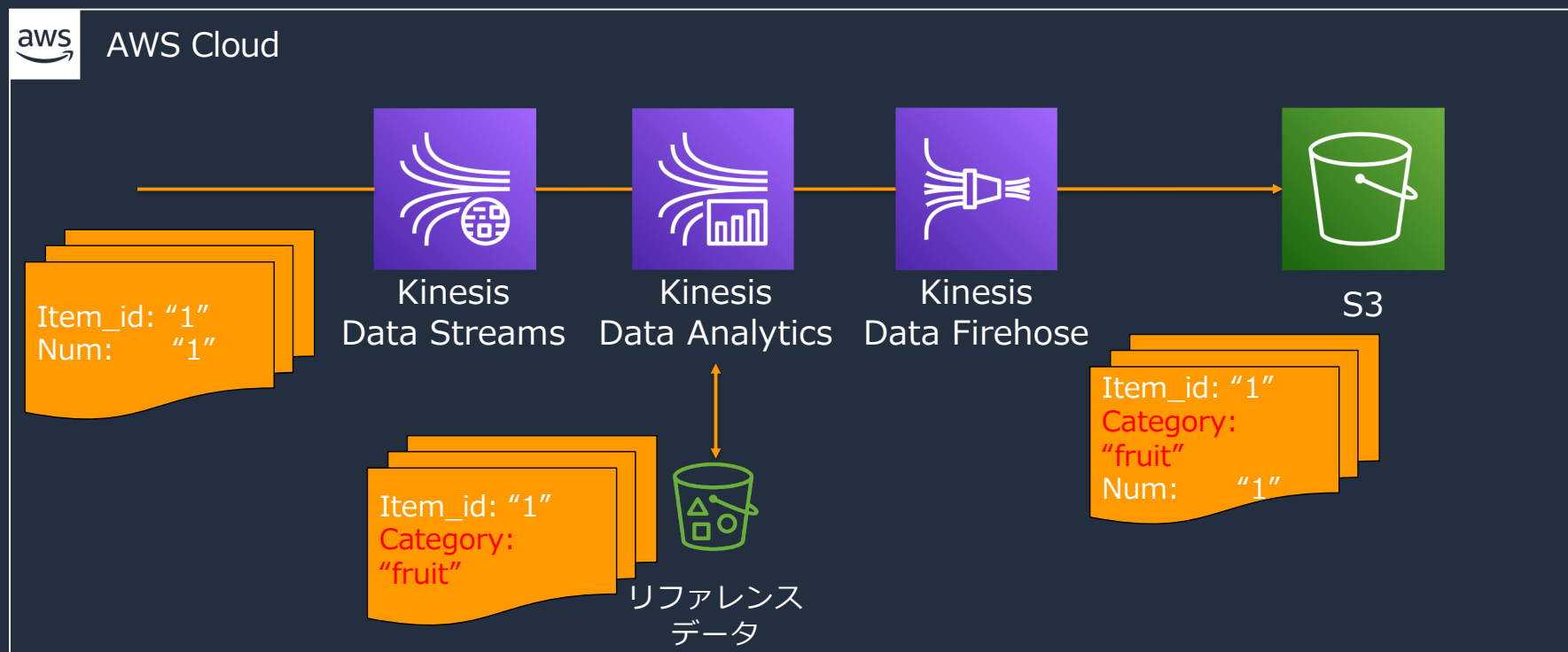


\* 東京リージョン、標準クラスの利用



# S3に保存する前に一次処理を行うことも可能

Amazon Kinesis Data Analyticsによる前処理



[https://docs.aws.amazon.com/ja\\_jp/kinesisanalytics/latest/dev/how-it-works-input.html#source-reference](https://docs.aws.amazon.com/ja_jp/kinesisanalytics/latest/dev/how-it-works-input.html#source-reference)

# S3に保存する前に一次処理を行うことも可能

保存・整形

ストリームデータ

```
{"ItemId":"1","Num":"2"}
```

リファレンスデータ

```
{"ItemId":"1","Category":"fruit"}
```

Kinesis Data AnalyticsアプリケーションのSQL

```
CREATE OR REPLACE STREAM "DESTINATION_SQL_STREAM" (  
  "ItemId" VARCHAR(64),  
  "Num" INT  
  "Category" VARCHAR(64)  
);  
  
CREATE OR REPLACE PUMP "STREAM_PUMP" AS INSERT INTO "DESTINATION_SQL_STREAM"  
  SELECT STREAM  
    s."ItemId"  
    ,s."Num"  
    ,r."Category"  
  FROM "SOURCE_SQL_STREAM_001" s  
  LEFT OUTER JOIN  
    "REFERENCE_DATA" r ON s."ItemId" = r."ItemId";
```

# 売上データを整形する

保存・整形

## 課題

- サイズの小さい、多数のJSONファイルのままだと、大量データの分析に時間がかかる

## 処理の要件

- 分析クエリを短時間でコスト効率よく実行できること

# 売上データを整形する

保存・整形

## 課題

- サイズの小さい、多数のJSONファイルのままだと、大量データの分析に時間がかかる

## 処理の要件

- 分析クエリを短時間でコスト効率よく実行できること

### [解決策]

- JSON形式から列指向のデータ形式に変換する
- 小さなファイルをひとまとめにする
- 日付でパーティションを作成する

# 列指向フォーマットとは

カラム（列単位）でデータをまとめて保存するデータフォーマット

## メリット1) OLAP 系の分析クエリを効率的に実行できる

- 多くの分析クエリは、一度のクエリで一部のカラムしか使用しない

行指向

1	2	3	4	5	6

列指向

1	2	3	4	5	6

## メリット2) I/O の効率があがる

- 圧縮と同時に使うことで I/O 効率がさらに向上
- カラムごとに分けられてデータが並んでいる
- 同じカラムは、似たような中身のデータが続くため、圧縮効率がよくなる

行指向

1	2	3	4	5	6

列指向

1	2	3	4	5	6

# コンパクションとは

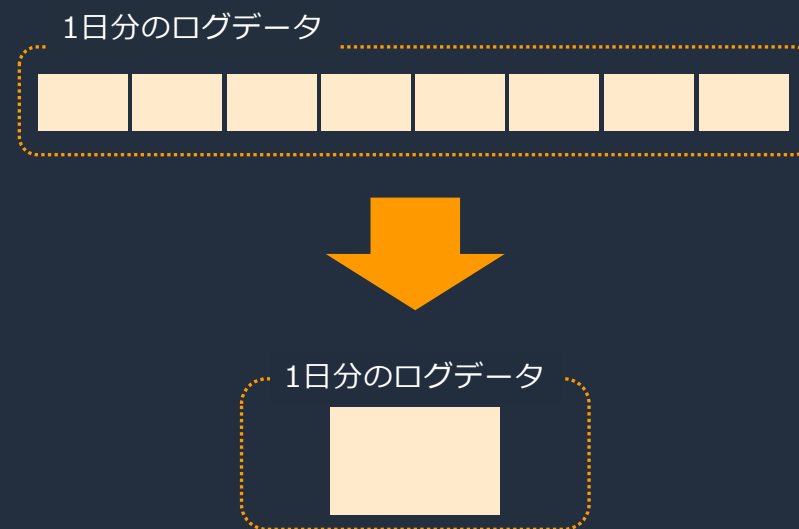
サイズの小さい複数ファイルを分析のしやすい単位に集約

## ファイル数が多い場合の問題点

- Hadoop系の分散エンジンでは、読み込む必要のあるブロックごとにタスクが生成される。その為、データが少ないブロックへの処理ではオーバーヘッドが発生しパフォーマンスが低下する。

## コンパクション後のメリット

- データ分析前にファイルをひとまとめにしておくことで、オーバーヘッドが減少し分析時のパフォーマンスが向上する。

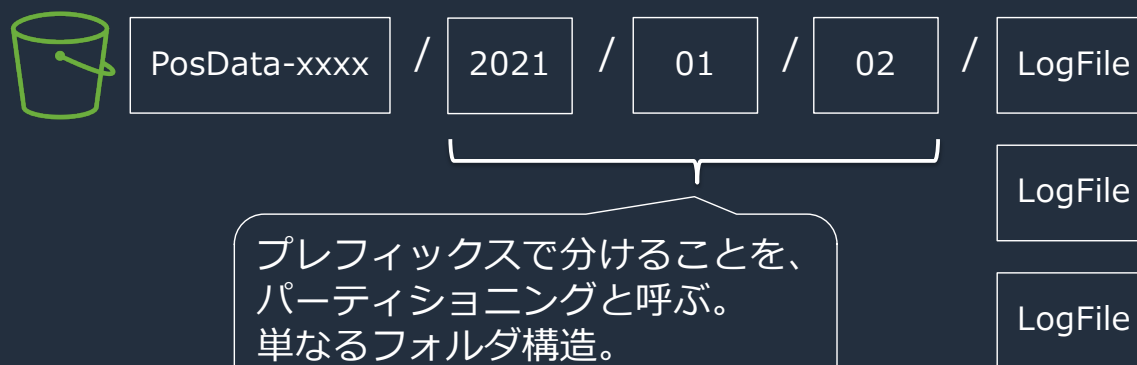
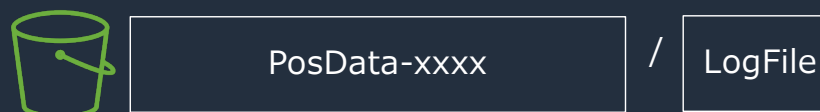


# パーティショニングとは

分析によく使う単位でファイルを分析して格納

## パーティショニングをしない場合の問題点

- 特定のシステムのログを検索したい場合、全てのLogFileに対してスキャンが掛かる。その為、読み込むデータ量が多くなり、パフォーマンスに影響がある。



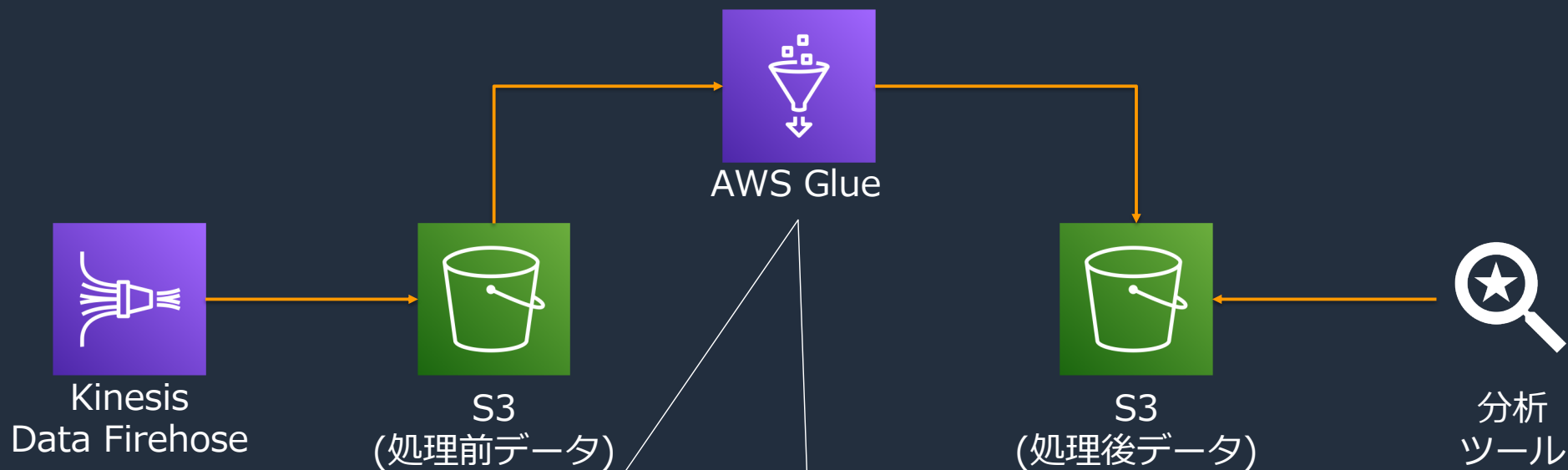
プレフィックスで分けることを、パーティショニングと呼ぶ。単なるフォルダ構造。

## パーティショニング後のメリット

- 検索条件にシステム名を含めることで、特定のファイルにのみアクセスする。不要データへのアクセスが減る為、パフォーマンスが向上する。

# AWS Glueでデータを整形する

保存・整形



解決策：

1. フォーマットを列指向フォーマットに変換する
2. 大量ファイルをコンパクションする
3. 適切な単位でパーティショニングする



# AWS Glueでデータを整形する

保存・整形

処理から試す  
なら既存シス  
テムのDBから  
抽出したデー  
タを用いる



解決策：

1. フォーマットを列指向フォーマットに変換する
2. 大量ファイルをコンパクションする
3. 適切な単位でパーティショニングする

# AWS Glueでデータを整形する

保存・整形

Glueでのデータ整形の詳細は、下記の AWS Black Belt Online Seminarの資料をご参照ください

[AWS Black Belt Online Seminar]  
AWS Glue ~Glue Studioを使ったデータ  
変換のベストプラクティス~

動画:

<https://youtu.be/xRszN4Tb4uM>

資料:

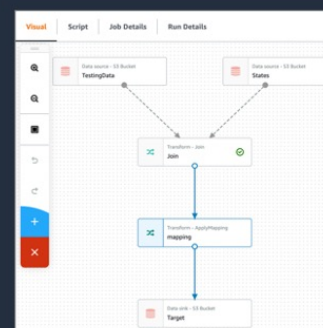
[https://d1.awsstatic.com/webinars/jp/pdf/services/20210330\\_AWSBlackBelt2021\\_AWSGlueStudio.pdf](https://d1.awsstatic.com/webinars/jp/pdf/services/20210330_AWSBlackBelt2021_AWSGlueStudio.pdf)

## AWS Glue Studioを使用したETL JOBの作成

- ・ マネージメントコンソールより、AWS Glue を選択

### AWS Glue Studio

ETL ジョブの作成、実行、監視を容易にする視覚的なインターフェース



#### ビジュアルオーサリング

コードを書くことなくETLジョブを作成可能

#### サーバーレス

ビッグデータ処理能力を利用可能

#### シンプルビュー

単一画面でジョブを管理可能

#### カスタマイズ可能

任意のコード(Python、Scala、Java)で ETL を高度化可能

<https://aws.amazon.com/jp/blogs/big-data/making-etl-easier-with-aws-glue-studio/>

© 2021, Amazon Web Services, Inc. or its Affiliates.

# 処理したデータを管理する

## 課題

- 処理したデータ種類が多くなると、何がどこにあり、誰がアクセスして良いかの管理が煩雑になる

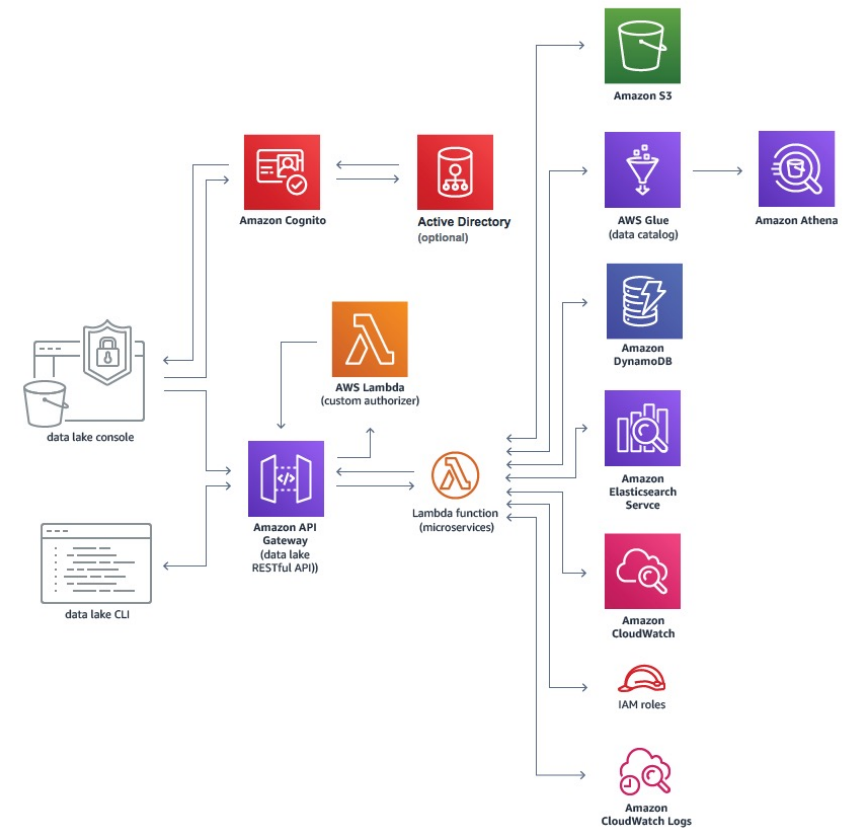
## 要件

- データカタログを用いて、登録したデータを参照できる
- 登録したデータに対して、適切なアクセス権を管理できる

# Datalake on AWS solution

保存・整形

- データレイク(S3)に登録するデータを簡単にタグ付け、検索、共有、変換、分析、制御するために必要なAWSサービスを構成するソリューション
- ユーザーがデータセットの検索や閲覧ができるコンソールがデプロイされる



<https://aws.amazon.com/jp/solutions/implementations/data-lake-solution/>

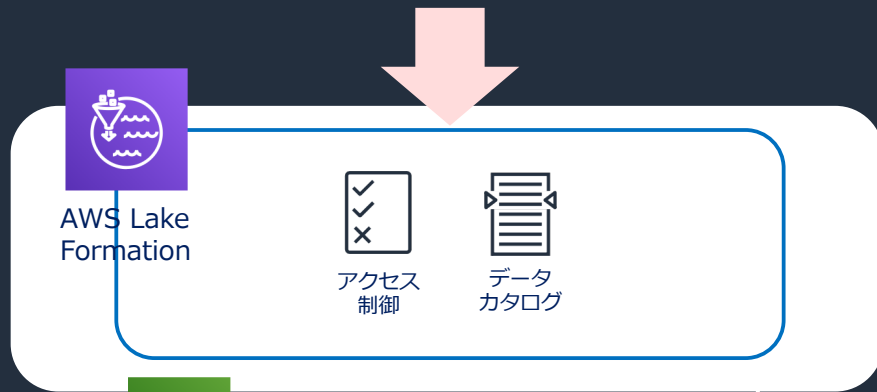
# AWS Lake Formationによる権限管理

保存・整形



管理者

1. Lake Formation 上でユーザーのアクセス制御を設定

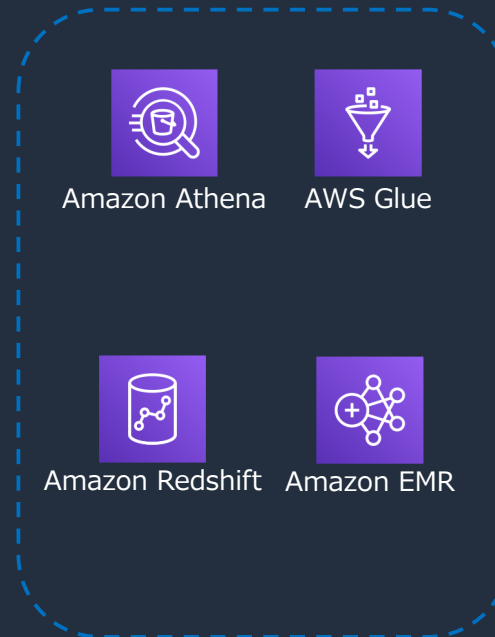


Amazon S3



00001011110110100011110010110010  
1000110000 データレイク 1001010111  
101110010 ストレージ 100101011  
1011111011010001111001011  
10000101001100001001  
101101000111

3. Lake Formationがメタデータとアクセス権限を統合管理し、データへのアクセスを認可する



2. ユーザーは、任意のサービスからデータへアクセス

# AWS Lake Formationによる権限管理

保存・整形

## 列レベルのアクセス制御の指定

Column - optional  
Choose filter type

Include columns

Q |

None

Include columns

Exclude columns

Table permissions  
Choose the specific access permissions to grant.

Select all  Alter  Insert  Drop

Delete  Select



ユーザー 1

一部の列のみ  
アクセス可能

- テーブルおよび列レベルの権限付与 (行レベルはプレビュー中)
- タグベースの権限管理も可能

Column name	Data type
marketplace	string
customer_id	bigint
review_id	string
product_id	string
product_parent	bigint
product_title	string
star_rating	string
helpful_votes	bigint
total_votes	bigint
vine	string
verified_purchase	string
review_headline	string
review_body	string
review_date	string
product_category	string



ユーザー 2

全ての列に  
アクセス可能

# AWS Lake Formationによる権限管理

保存・整形

Lake Formationでの権限管理の詳細は、下記の AWS Black Belt Online Seminarの資料をご参照ください

[AWS Black Belt Online Seminar]  
AWS Lake Formation

動画:

[https://youtu.be/riS\\_STGzZHI](https://youtu.be/riS_STGzZHI)

資料:

[https://d1.awsstatic.com/webinars/jp/pdf/services/20191001\\_BlackBelt\\_LakeFormation\\_A.pdf](https://d1.awsstatic.com/webinars/jp/pdf/services/20191001_BlackBelt_LakeFormation_A.pdf)

## アクセス時の連携イメージ

エンドサービスは  
S3 から直接関連  
データを検索

2 テーブルへの  
アクセスリクエスト

## データアクセス許可

テーブルの基になるデータを読み書きするために、プリンシパルにテーブルに対するデータアクセス許可をシンプルな Grant/Revoke で与えます。  
(例 SELECT, INSERT)

Grant 選択

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Confidential and Trademark

# データを分析する



# 分析の例

分析

AWSのAnalytics、DBサービスを活用して様々なユースケースに対応可能

カテゴリ	ユースケース	利用サービス/機能
Analytics	大規模な売上データで複数の分析業務を実施する	Amazon Redshift Data Sharing
DB	売上データから商品の相関関係を分析する	Amazon Neptune

# 同じ売上データで複数の分析業務を実施する

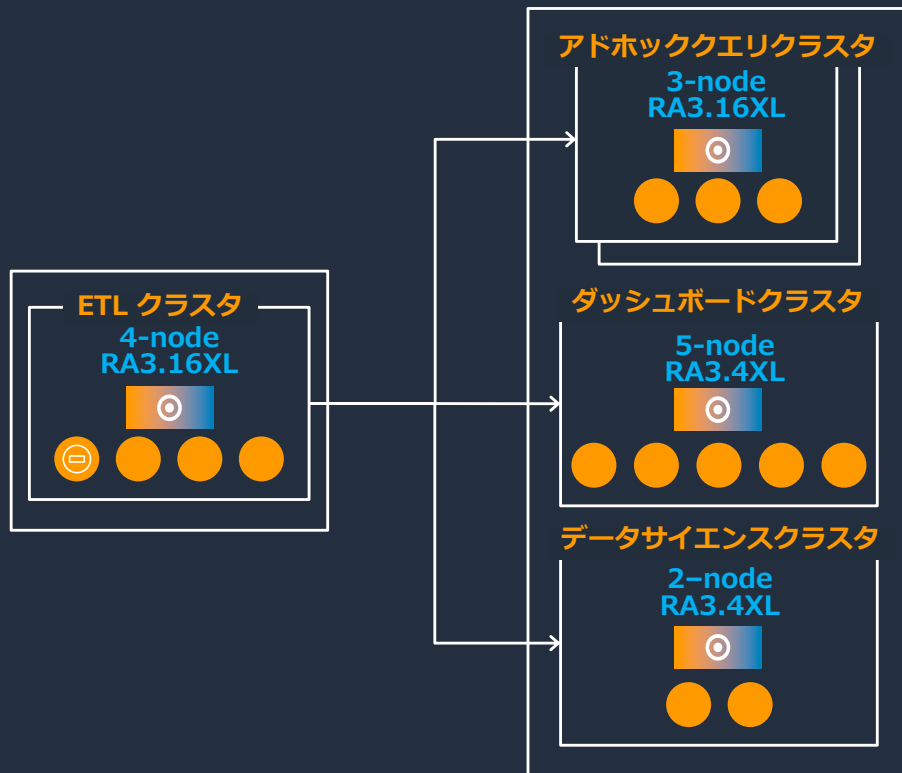
## Amazon Redshift Data Sharing



- プロデューサークラスターが書き込んだ共有データを、コンシューマークラスターがread onlyで共有することが可能
- データの移動がなく、アクセス権の管理と共有状況の監査により、セキュアなデータ共有を実現

# Data sharing のユースケース例

様々なビジネスクリティカルなワークロードを、簡単でコストパフォーマンスよくサポート



新しい分析ワークロードを素早く利用可能に

パフォーマンス要件に応じて個々のクラスタのサイズ・性能を最適化

データセットを共有しながら、ワークロードを分離

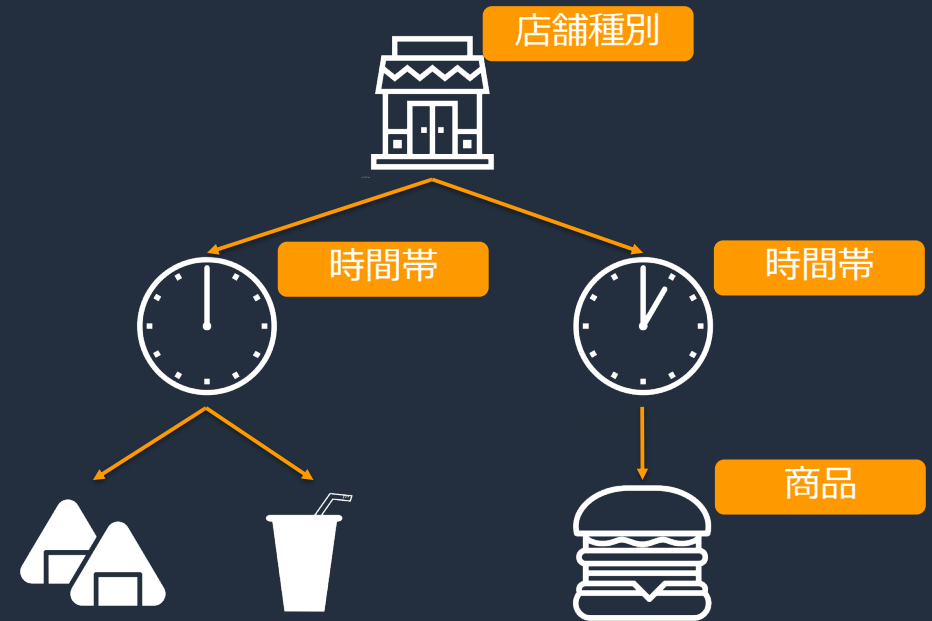
個々のワークロードごとに費用請求

クラスタは必要に応じて停止/再開

# 売上データから商品の相関関係を分析する

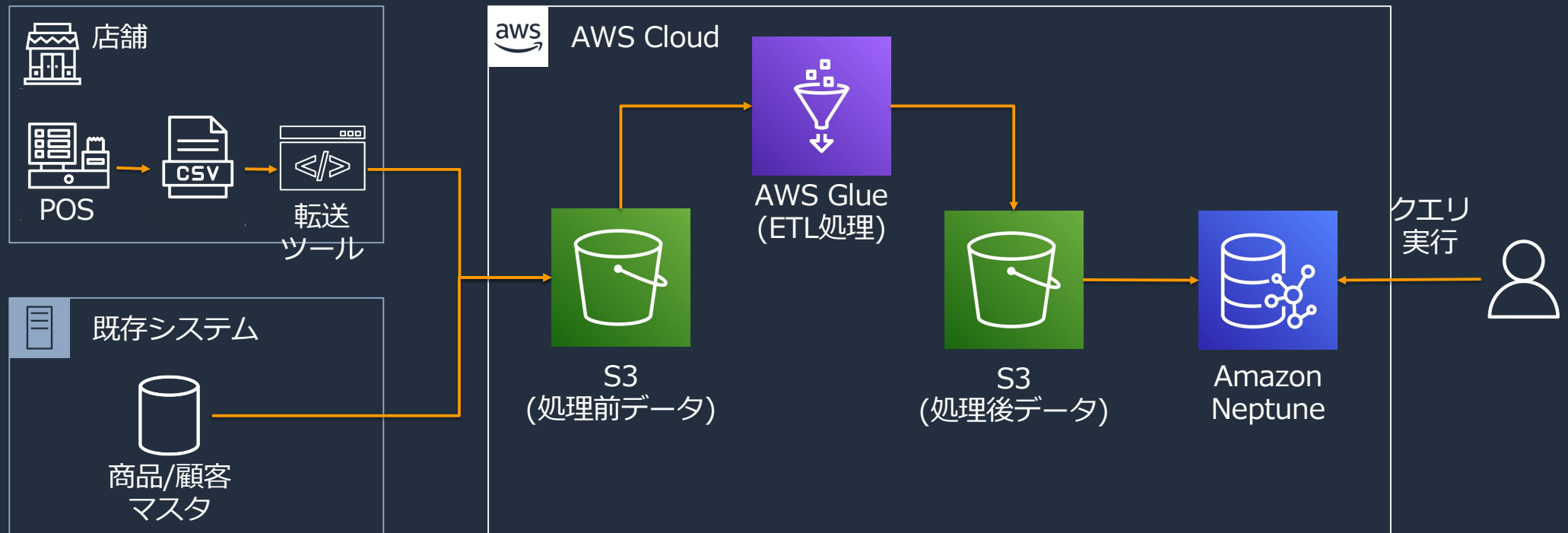
Amazon Neptuneによる商品グラフの作成

- 商業地にある店舗Aの、
- 昼の時間帯に、
- おにぎりAとセットで購入される商品は何か？



# 売上データから商品の相関関係を分析する

## Amazon Neptuneによる商品グラフの作成



# Amazon Neptune Advanced Design Pattern

グラフデータのモデリングプロセスの詳細は、下記のセミナー資料をご参照ください

[Purpose Build Database Day]  
Amazon Neptune  
Advanced Design Pattern

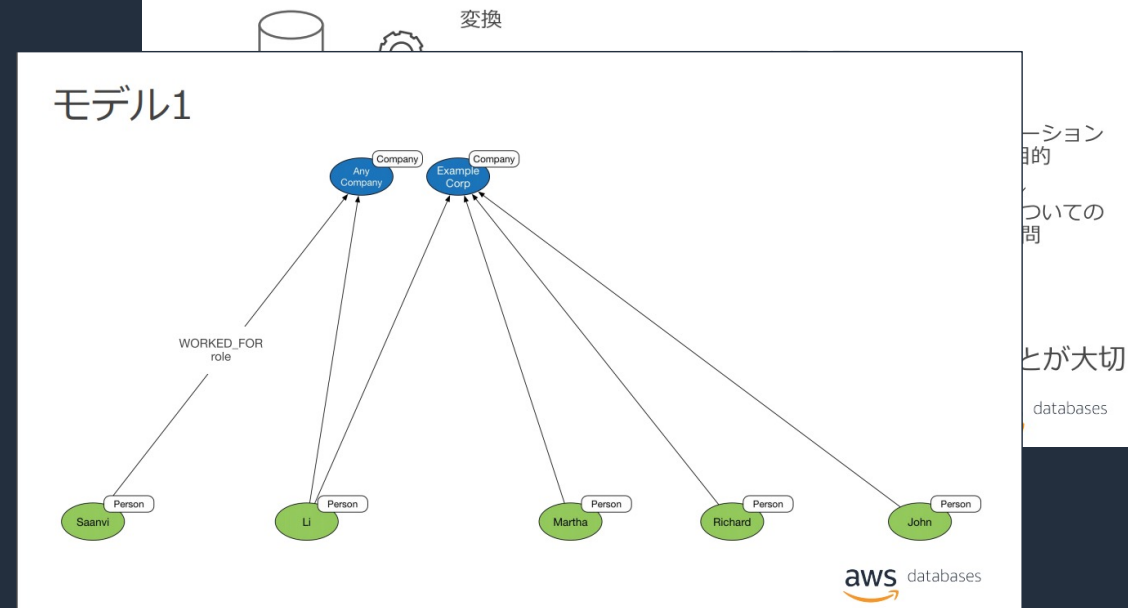
動画:

<https://aws.amazon.com/jp/about-aws/events/2020/purpose-built-databases-week-archive/>

資料:

[https://pages.awscloud.com/rs/112-TZM-766/images/Session%205%20-%20Neptune Advanced Design Pattern\\_rev.pdf](https://pages.awscloud.com/rs/112-TZM-766/images/Session%205%20-%20Neptune%20Advanced%20Design%20Pattern_rev.pdf)

Working backwards - クエリを想定したデザイン



# Amazon Neptune

Amazon Neptuneの詳細は、下記のAWS Black Belt Online Seminarの資料をご参照ください

[AWS Black Belt Online Seminar]  
Amazon Neptune

動画:

<https://youtu.be/XDTr6U0fcWE>

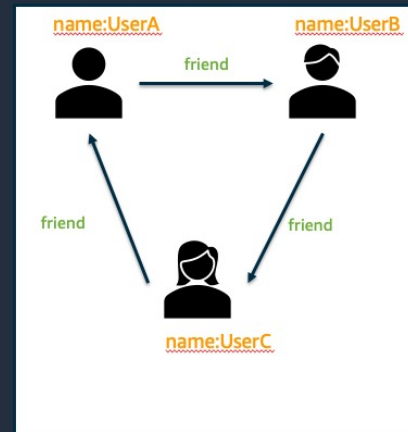
資料:

[https://d1.awsstatic.com/webinars/jp/pdf/services/20200714\\_AWS\\_BlackBelt\\_AmazonNeptune.pdf](https://d1.awsstatic.com/webinars/jp/pdf/services/20200714_AWS_BlackBelt_AmazonNeptune.pdf)

データの投入

パルクでデータを投入する方法は2種類

Gremlin 101



"UserA"の友達であるユーザを探したい

```
g.V().has('name', 'UserA').out('friend')
```

⇒ ['UserB']

© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

aws

aws

# データを可視化・応用する



# 可視化・応用の例

可視化・応用

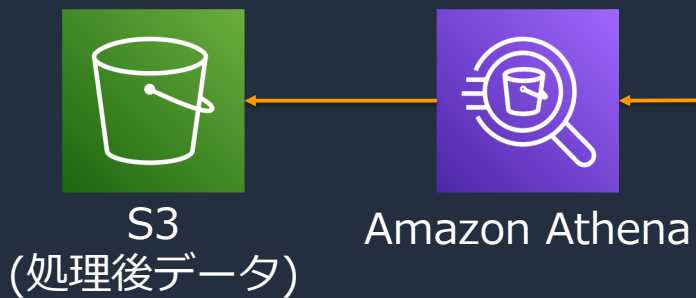
AWSのAnalytics、AI/MLサービスを活用して様々なユースケースに対応可能

カテゴリ	ユースケース	利用サービス/機能
Analytics	売上や在庫データを可視化する	<ul style="list-style-type: none"><li>Amazon Athena</li><li>Amazon Redshift</li><li>Amazon QuickSight</li></ul>
AI/ML	SQLでテーブルデータから売上や在庫の時系列予測をする	Amazon Redshift ML
	CSVデータから売上や在庫の時系列予測をする	Amazon Forecast
	売上高のようなビジネスメトリクスを監視して、突然の変化を把握する	Amazon Lookout for Metrics

# 売上や在庫データを可視化する

Athenaで集計・結合・加工してQuickSightに取り込み

ダッシュボード/レポートを共有



A screenshot of the Amazon QuickSight dashboard. At the top center is the QuickSight logo (a purple square with a white line graph icon) and the text 'Amazon QuickSight'. Below this is a screenshot of the 'S3 Analytics - Storage Class Analysis' dashboard. The dashboard includes a 'Fields list' on the left with filters for ConfigId, Date, Filter, ObjectAge, ObjectAgeForSIATrans..., RecommendedObject..., and StorageClass. It also shows 'Visual types' like bar, line, and pie charts. The main area contains four charts: 'Data Retrieved by Age Group (STANDARD)', 'Storage Consumed by Age Group (STANDARD)', 'Storage Consumed by Age Group (STANDARD\_IA)', and 'Data Retrieved by Age Group (STANDARD\_IA)'. Each chart shows data over time (Date (DAY)) for different ObjectAge groups (000-029, 030-044, 045-059, 060-074, 075-089, 090-179, 180-364, 365+).



# 売上や在庫データを可視化する

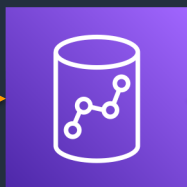
Amazon QuickSightからAmazon Redshiftへのクエリ実行

ダッシュボード/レポートを共有

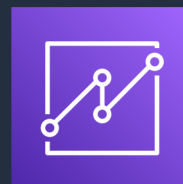
大規模なデータを  
ロード



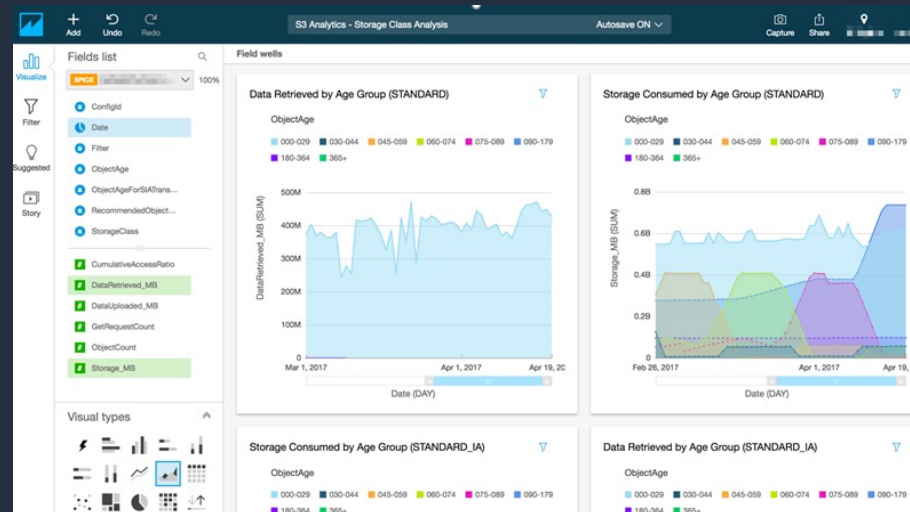
S3  
(処理後データ)



Amazon Redshift



Amazon QuickSight



# Amazon QuickSightセルフハンズオンキット

## 販売管理ダッシュボード編

予算売上・粗利KPI表示や損益管理、明細表の作り方など、小売業務ですぐに使えるハンズオンキット



<https://aws.amazon.com/jp/blogs/news/amazon-quicksight-handson-202006/>

# How to analyze retail data

Amazon QuickSightのBIダッシュボードで小売データを分析する

どのような分析手順でインサイトを抽出するかを解説

ブログ:

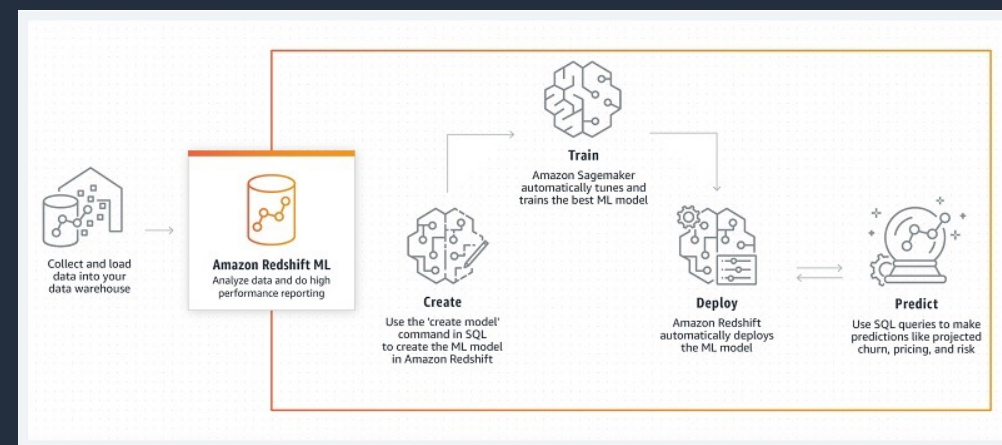
<https://aws.amazon.com/jp/blogs/news/quicksight-dashboard-analysis-retail/>



# SQLでテーブルデータから売上や在庫の時系列予測をする

## Amazon Redshift ML

- ✓ 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 を指定可能
- ✓ 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呼び出すように推論を実行可能



```
CREATE MODEL demo_ml.customer_churn
FROM (SELECT c.age, c.zip, c.monthly_spend,
c.monthly_cases, c.active FROM
customer_info_table c)
TARGET c.active;
```

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

# Amazon Redshift ML のしくみ

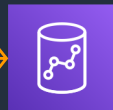
可視化・応用

## トレーニング



```
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
```

Autopilot が実行され  
モデルを返却



Amazon Redshift



Amazon SageMaker

シンプルなSQLコマンドでモデルを作成, トレーニング, デプロイ

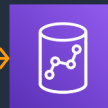
SageMaker Autopilot を使用したモデルの自動選択, 前処理, トレーニング (アルゴリズム指定も可)

トレーニングされたモデルは, Amazon Redshift 上にコンパイルされ, 以後ユーザーは SQL を使用して推論を行うことが可能

## 推論



```
SELECT n.id, n.firstName, n.lastName,
predict_customer_churn(n.age,c.zip,..)
AS activity_prediction
FROM new_customers n
WHERE n.marital_status = 'single'
```



Amazon Redshift

構築済みモデルを使用して推論を実行  
(処理は Amazon Redshift 内で実行)

# Amazon Redshift ML

可視化・応用

Amazon Redshift MLの詳細は、下記のセミナー資料をご参照ください

「SQLでやってみよう  
データ分析と機械学習」

動画:

<https://youtu.be/RZk0n6CTJGI>

資料:

[https://pages.awscloud.com/rs/112-TZM-766/images/20210325-03\\_Data-Analytics-SQL-ML.pdf](https://pages.awscloud.com/rs/112-TZM-766/images/20210325-03_Data-Analytics-SQL-ML.pdf)

## ML モデルを使用して推論を実行

推論は UDF として利用可能

SELECT customer\_id,

## ML モデルの作成とトレーニング

トレーニングデータをテーブル名または SELECT クエリとして指定

TARGET として実際に予測したい列を指定

FUNCTION にはあとで推論を行う際に使用されるための予測関数名を指定

```
CREATE MODEL customer_churn
```

```
FROM (SELECT c.age AS feat_1, c.zip AS feat_2,  
c.monthly_spend AS feat_3, c.monthly_cases AS  
feat_4, c.active AS label  
FROM customer_info_table c)
```

```
TARGET label
```

```
FUNCTION predict_customer_churn
```

© 2021, Amazon Web Services, Inc. or its Affiliates.

aws

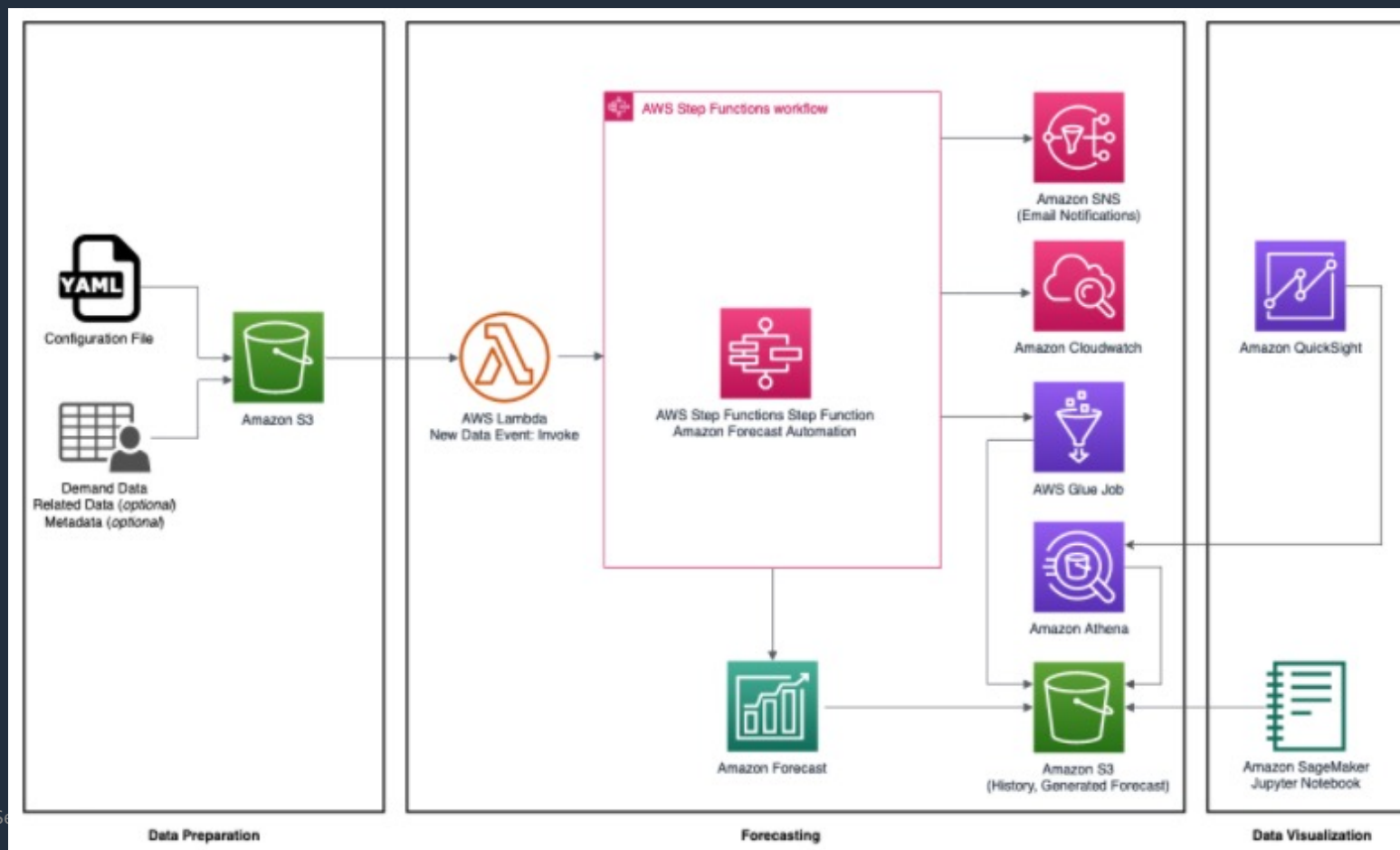
aws

aws



# CSVデータから売上や在庫の時系列予測をする

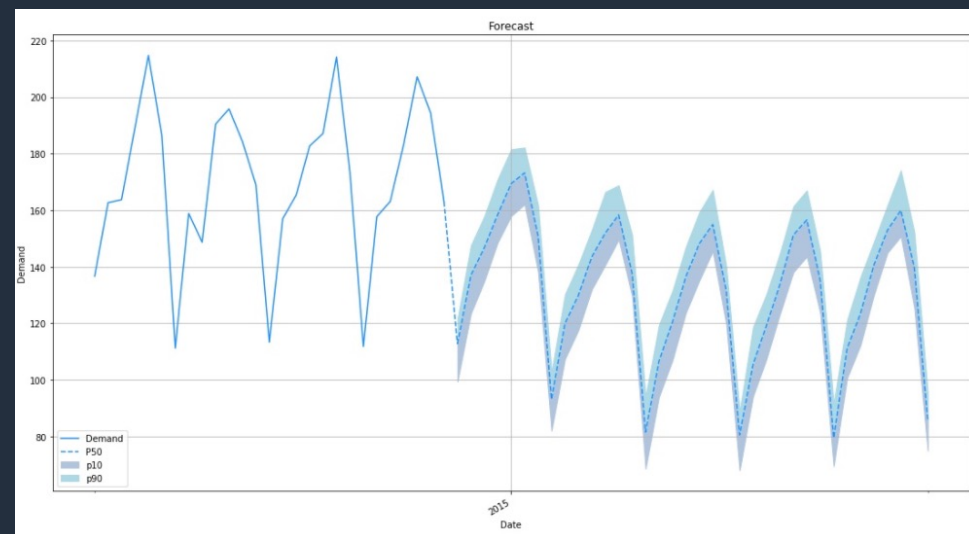
Amazon Forecastを中心としたソリューションをAWS CloudFormationで実装可能



# CSVデータから売上や在庫の時系列予測をする

Amazon Forecastを中心としたアーキテクチャを数分で実装可能

- 予測結果はQuickSightのダッシュボードで確認可能
- 予測はディメンション(店舗の場所など)またはアイテムレベルのメタデータ(ブランド、サイズ、色など)間で比較可能
- 複数の分位での予測結果を表示



<https://aws.amazon.com/jp/solutions/implementations/improving-forecast-accuracy-with-machine-learning/#>

# CSVデータから売上や在庫の時系列予測をする

予測の準備 ~構成ファイルをS3にアップロード~

```
retail_experiment_2:
  DatasetGroup:
    Domain: RETAIL

  Datasets:
    - Domain: RETAIL
      DatasetType: TARGET_TIME_SERIES
      DataFrequency: D
      TimestampFormat: yyyy-MM-dd
      Schema:
        Attributes:
          - AttributeName: item_id
            AttributeType: string
          - AttributeName: timestamp
            AttributeType: timestamp
          - AttributeName: demand
            AttributeType: float
```

Amazon Forecastでサポートする ドメイン  
(ユースケース)を指定

時系列予測の機械学習に用いるトレーニングデータの構成(タイプ、時系列の間隔、時刻フォーマット、スキーマ)を指定

# CSVデータから売上や在庫の時系列予測をする

予測の準備 ~構成ファイルをS3にアップロード~

```
Predictor:
  AlgorithmArn:
arn:aws:forecast:::algorithm/Deep_AR_Plus
  MaxAge: 604800 # one week
  PerformHPO: True
  ForecastHorizon: 30
  FeaturizationConfig:
    ForecastFrequency: D
  Featurizations:
    - AttributeName: price
      FeaturizationPipeline:
        - FeaturizationMethodName: filling
          FeaturizationMethodParameters:
            futurefill: median
            middlefill: median
            backfill: median
```

- Amazon Forecastが時系列予測を行うために使用する予測モデルの構成を指定する
- 構成パラメータはAmazon Forecastの [CreatePredictor API](#)で指定できるもののすべてをサポート

# CSVデータから売上や在庫の時系列予測をする

予測の準備 ~構成ファイルをS3にアップロード~

Forecast:

ForecastTypes:

- "0.10"
- "0.50"
- "0.90"

Amazon Forecastが時系列予測を行う際の構成(分位)を指定する

## 分位の指定の目的

- 予測の上限と下限を設けることで実際の値がそれらの間にあると予測できる
- 在庫不足が超過よりリスクが高くなる場合、中央値(P50, 上記の"0.50")より高い分位(例: P65)の方が有効のように判断する
- 分位は最大5つまで指定可能

# CSVデータから売上や在庫の時系列予測をする

予測の作成 ~時系列データをS3にアップロード~

- 準備で指定したデータセットのフォーマットで作成したCSVファイルをS3にアップロードする
- アップロードをトリガーとして予測のためのワークフローが開始される

Solution document

<https://docs.aws.amazon.com/solutions/latest/improving-forecast-accuracy-with-machine-learning/welcome.html>

# CSVデータから売上や在庫の時系列予測をする

Amazon Forecastの詳細は、下記の AWS Black Belt Online Seminarの資料をご参照ください

[AWS Black Belt Online Seminar]  
Amazon Forecast

動画:  
<https://youtu.be/-Bq5f7jFKoE>  
資料:

[https://d1.awsstatic.com/webinars/jp/pdf/services/20200128\\_AWS\\_BlackBelt\\_AmazonForecast.pdf](https://d1.awsstatic.com/webinars/jp/pdf/services/20200128_AWS_BlackBelt_AmazonForecast.pdf)

## データの準備 TARGET\_TIME\_SERIES


例 (RETAIL ドメインの場合)

### Amazon Forecast のワークフロー概要

Amazon Forecast

#### ② データセットグループの作成

##### データセットグループ

データの準備	予測子の作成	予測の作成
③ スキーマの定義  ④ データセットの作成 ⑤ 各種データのインポート ・ TARGET_TIME_SERIES ・ ITEM_METADATA ・ RELATED_TIME_SERIES	⑥ 予測子の作成 ・ 予測期間の設定 ・ アルゴリズムの選択 ・ パラメータの設定  いわゆるトレーニングが実行され、モデルが作成される  ⑦ 予測子の評価 各指標を参考に評価を実施	⑧ 予測の作成 学習済みモデルを利用し、予測（推論エンドポイント）を作成する  予測はAPI経由での取得、または、マネジメントコンソール上で確認することができる

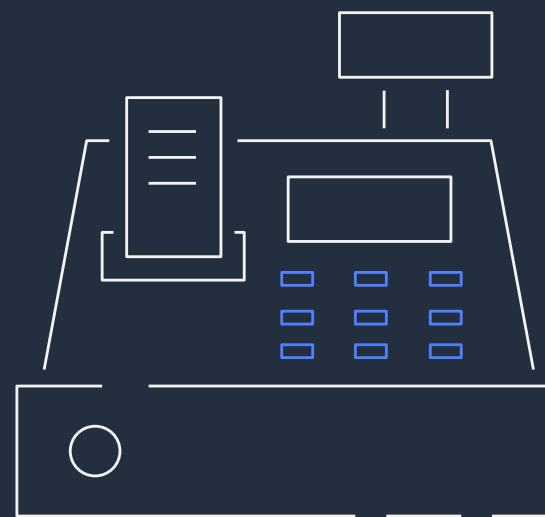
© 2020, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

aws

# ビジネスメトリクスを監視して、突然の変化を把握する

Amazon Lookout for Metricsでの自動検知

在庫や商品価格、販促、コンバージョン率などの角度から監視して、売上やマージンについてのインサイトを得る

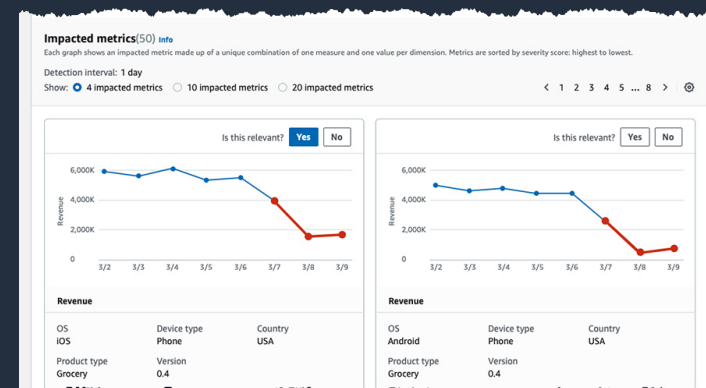
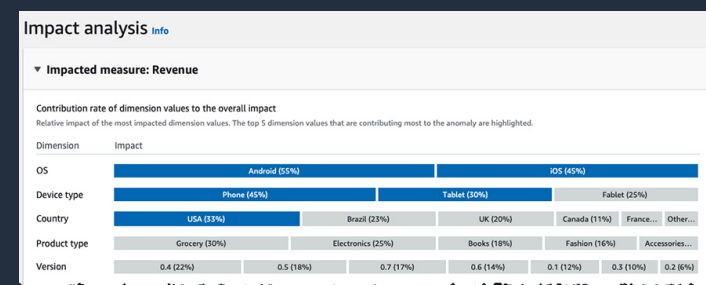




# Amazon Lookout for Metrics

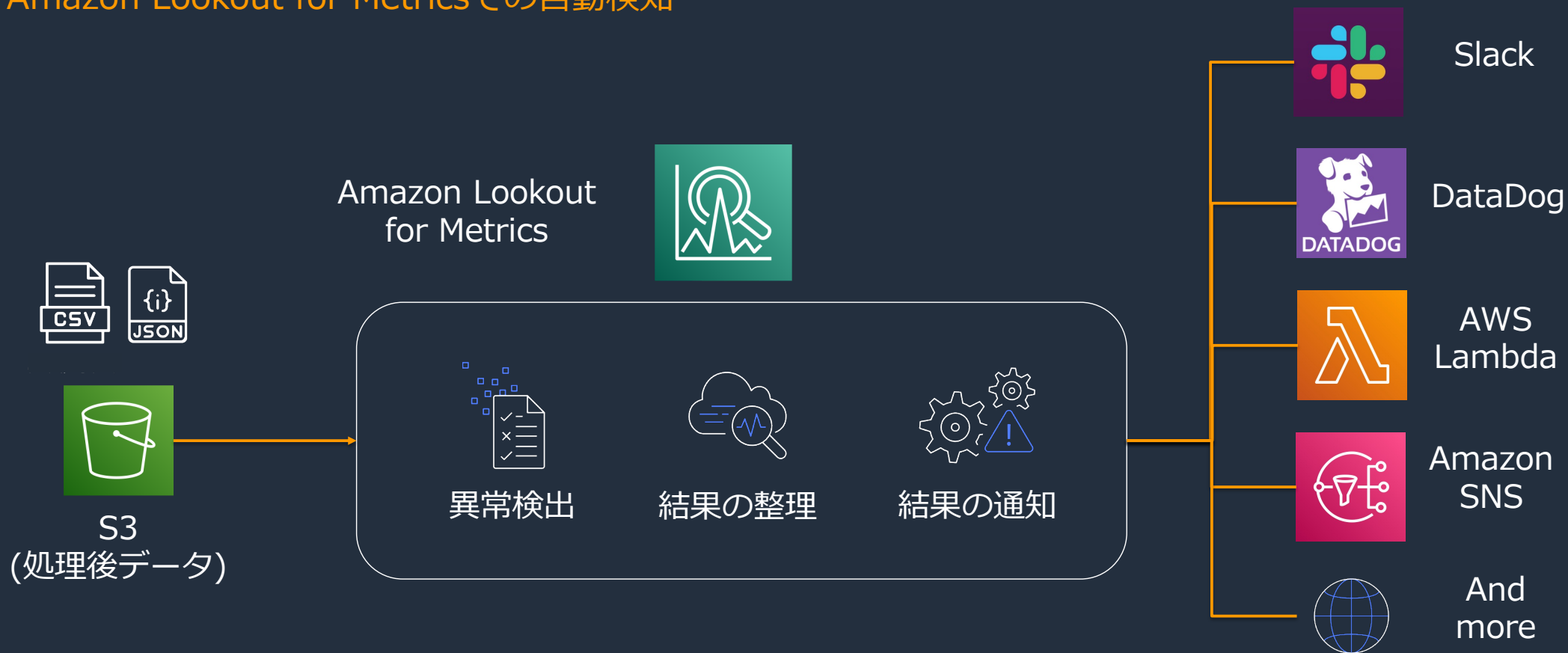
Amazon が利用しているものと同じ機械学習の技術を利用し、ビジネス上のメトリクスにおける異常値を検知する

- データに機械学習を適用し高精度で異常を検知する  
利用者側に機械学習の知識は必要ない
- 季節性のあるメトリクスにも対応し、  
ビジネスの健全性チェックや問題の検出、  
改善点の抽出を高精度に実現する
- S3 や RDS、Redshift、CloudWatch をはじめ  
SaaS からのデータも利用可能  
自動的に最も正確な予測モデルが構成され、  
チェックを開始できる
- 東京を含む9のリージョンで使用可能



# ビジネスメトリクスを監視して、突然の変化を把握する

Amazon Lookout for Metricsでの自動検知



# ビジネスメトリクスを監視して、突然の変化を把握する

## Amazon Lookout for Metricsの開始方法



### Step 1 Detectorの作成

時系列データ内の外れ値を検出するMLモデル(Detector)を作成する。検出間隔などを指定する



### Step 2 Metricsの定義

プライマリのMeasures、セカンダリのDimensionsのMetricsをDataset(CSV or JSON)に含まれるデータから指定する



### Step 3 Detectorの起動

データ分析を開始するためにDetectorを起動する

Get started

<https://github.com/aws-samples/amazon-lookout-for-metrics-samples>

# ビジネスメトリクスを監視して、突然の変化を把握する

## 2つの検出モード

### Backtesting

既存の時系列データを用いて Lookout for Metrics で検出する異常値の評価ができる。

Continuous で運用する前の評価として利用できる

```
s3://my-lookoutmetrics-dataset-123456789012/  
  historical/data-202009.jsonl  
  historical/data-202010.jsonl  
  historical/data-202011.jsonl
```

### Continuous

継続的に時系列データを Lookout for Metrics に渡して、リアルタイムに異常値を検出できる

```
s3://my-lookoutmetrics-dataset-123456789012/  
  continuous/20201225/1520/data.jsonl  
  continuous/20201225/1525/data.jsonl  
  continuous/20201225/1530/data.jsonl
```

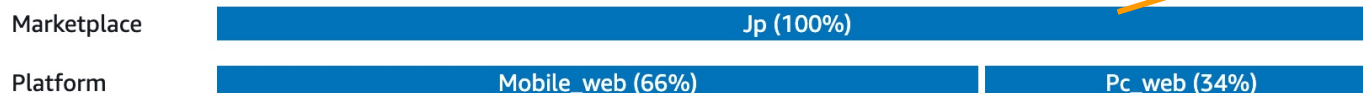
# ビジネスメトリクスを監視して、突然の変化を把握する

Amazon Lookout for Metrics で異常の確認

Contribution rate of dimension values to the overall impact

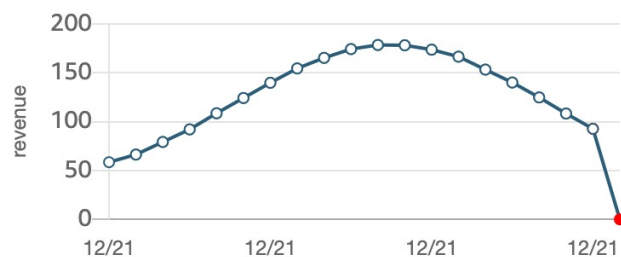
Relative impact of the most impacted dimension values

Dimension Value (Sum of dimension values equals 100%)



異常を自動的にグループ化し、ランク付ける

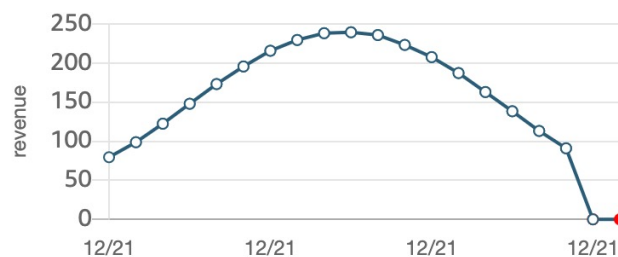
Is this anomaly relevant?  Yes  No



revenue

marketplace platform  
jp mobile\_web

Is this anomaly relevant?  Yes  No



revenue

marketplace platform  
jp pc\_web

本当に異常なのかをフィードバックできる

影響度が大きい  
メトリクスのグラフ

# まとめ

# 本セッションのまとめ

- 小売現場で発生するデータからインサイトを得てビジネスに活用することがますます求められている
- データを活用するために、データレイクを中心とした分析基盤が必要である
- 収集、保存、分析、可視化・応用が一連の流れだが、ユースケースに応じて必要なところから始められる
- ユースケースに合わせたAWSサービス、ソリューションをクイックに利用できる

## 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>

88

具体的な案件に対する構成相談は、後述する個別技術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



# AWS の日本語資料の場所「AWS 資料」で検索



日本担当チームへお問い合わせ サポート 日本語 ▾ アカウント ▾

コンソールにサインイン

製品 ソリューション 料金 ドキュメント 学習 パートナー AWS Marketplace その他 🔍

## AWS クラウドサービス活用資料集トップ

アマゾン ウェブ サービス (AWS) は安全なクラウドサービスプラットフォームで、ビジネスのスケールと成長をサポートする処理能力、データベースストレージ、およびその他多種多様な機能を提供します。お客様は必要なサービスを選択し、必要な分だけご利用いただけます。それらを活用するために役立つ日本語資料、動画コンテンツを多数ご提供しております。(本サイトは主に、AWS Webinar で使用した資料およびオンデマンドセミナー情報を掲載しています。)

[AWS Webinar お申込 »](#)

[AWS 初心者向け »](#)

[業種・ソリューション別資料 »](#)

[サービス別資料 »](#)

<https://amzn.to/JPArchive>



# AWS のハンズオン資料の場所「AWS ハンズオン」で検索



お問い合わせ サポート 日本語 アカウント

[今すぐ無料サインアップ](#)

製品 ソリューション 料金 ドキュメント 学ぶ パートナーネットワーク AWS Marketplace イベント さらに詳しく見る

## AWS ハンズオン資料

AWS をステップバイステップでお試しいただくのに役立つ動画および資料を掲載しています。

その他の資料は以下をご覧ください。

[初心者向けの資料](#)

[サービス別の資料](#)

[AWS オンラインセミナースケジュール](#)

[AWS クラウドサービス活用資料集トップ](#)

### AWS 初心者向けハンズオン

AWS 初心者向けに「AWS Hands-on for Beginners」と題し、初めて AWS を利用する方や、初めて対象のサービスに触る方向けに、操作手順の解説動画を見ながら自分のペースで進められるハンズオンをテーマごとにご用意しています。

<https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-hands-on/>



# AWS Well-Architected 個別技術相談会

毎週“W-A個別技術相談会”を実施中

- AWSのソリューションアーキテクト(SA)に  
対策などを相談することも可能

- **申込みはイベント告知サイトから**  
(<https://aws.amazon.com/jp/about-aws/events/>)

**AWS イベント** で[検索]

Big Data





ご視聴ありがとうございました