



# 運用いらずで手軽に分析！

サーバレスで始める  
AWSのAnalyticsサービス

平間 大輔

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

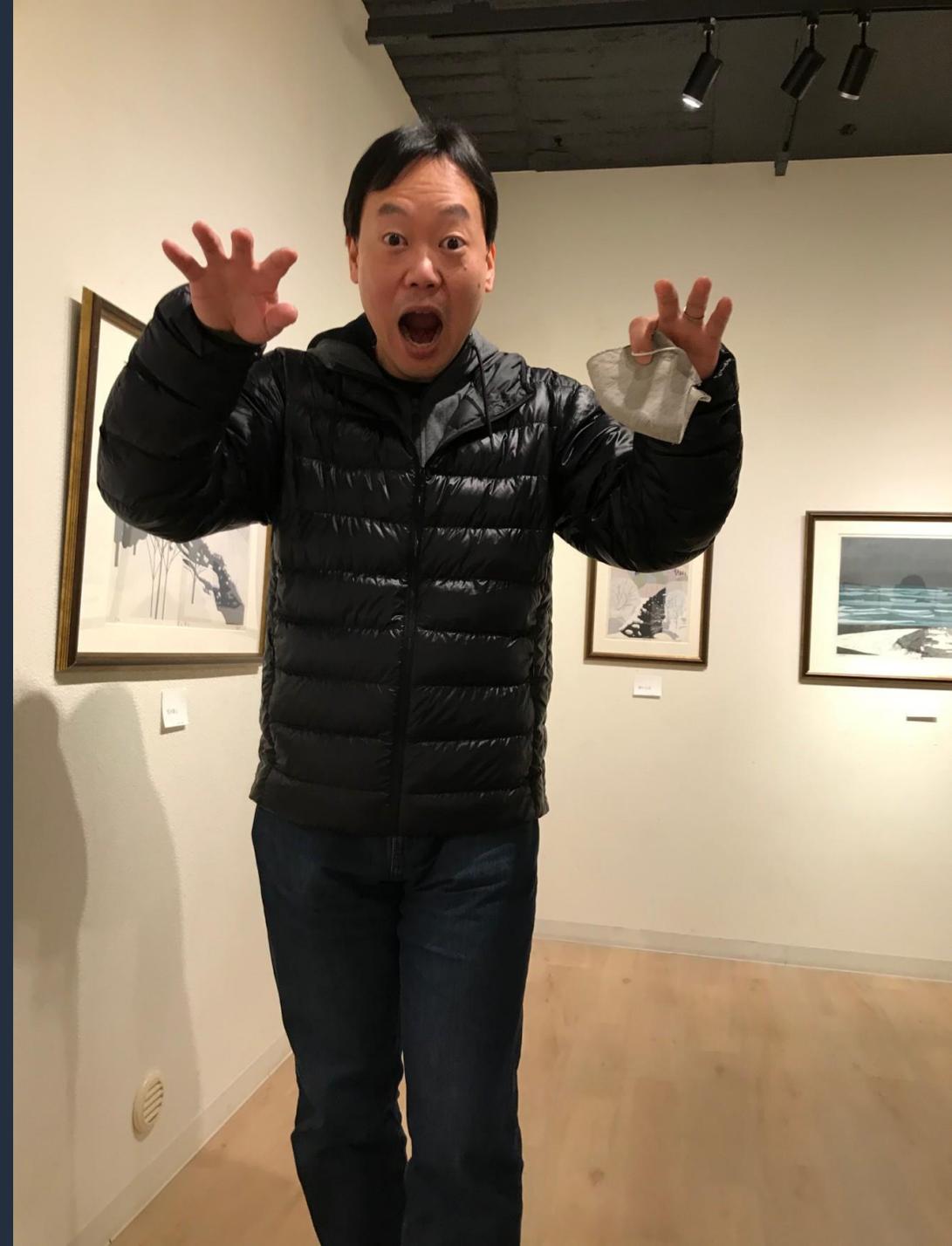
# 自己紹介

## 平間 大輔

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  
アナリティクス分野の技術支援を担当

好きな AWS サービス

- Amazon Redshift
- Amazon QuickSight



# データ活用の重要性と課題

# データ活用における共通課題



飛躍的な  
成長



新たな  
情報源



多様性の  
増加



更に多くの  
ユーザが利用



複数の  
アプリケーションを  
利用した分析

# 最新のデータ戦略を考える



# AWSでの モダンな データ戦略



# AWS Analytics の柱



スケーラブルな  
データレイク



パフォーマンス  
とコストを重視  
した専用設計



サーバレスで  
使いやすい



統合された  
データアクセス、  
セキュリティ、  
ガバナンス



組み込みの  
機械学習

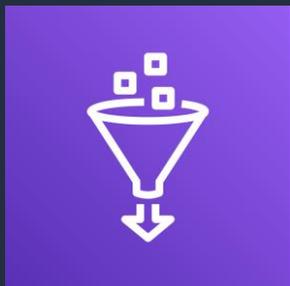
# AWSはクラウドにおけるデータ分析のための サーバーレスオプションが最も多い



# サーバレスでデータ加工



# AWS Glue



データ収集、カタログ化、データ加工を一気通貫して行うことができる  
統合的なデータインテグレーションサービス



スケーラブル



低コスト



インフラメンテ不要



複雑な処理に対応



バッチとストリームを同じコードで対応



他のシステムと容易に連携

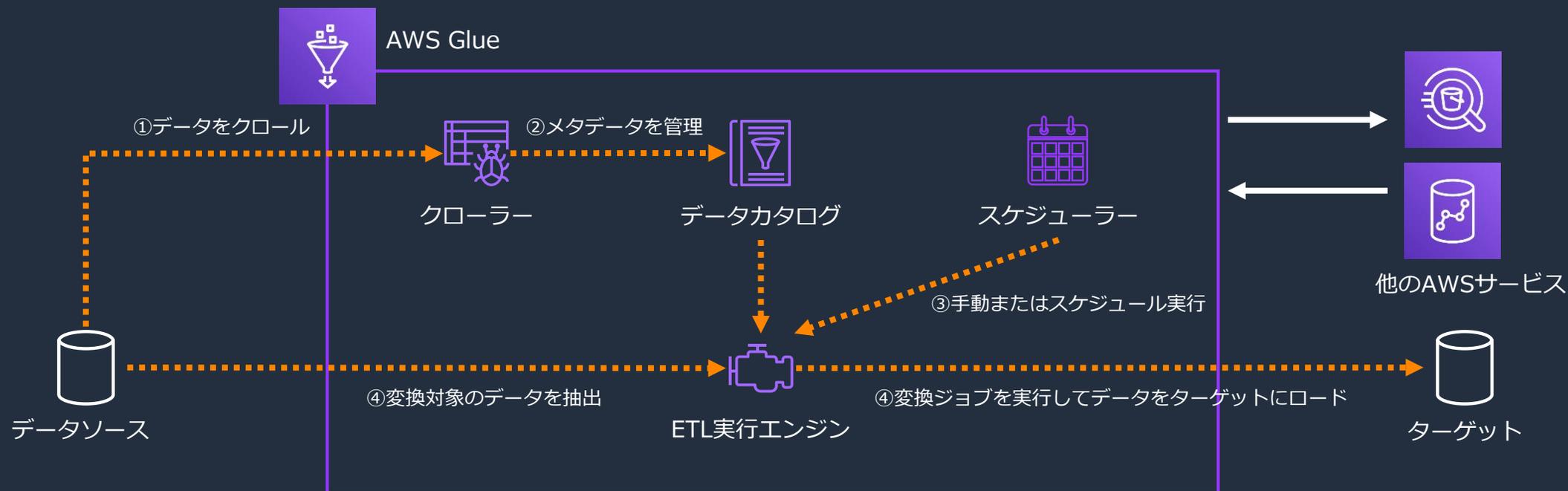


ロックインなし



Notebookでの開発

# AWS Glueの全体像



## 概要

- ① クローラーにてデータソースのメタデータをクロールして、データカタログに登録・更新
- ② データカタログにてメタデータを管理
- ③ スケジューラーにてジョブの実行タイミングを定義
- ④ データソースからデータを抽出し、ETL実行エンジンにてジョブをサーバーレスで実行  
(ジョブはPython(PySpark)、Scala、またはPython Shellを選択)

# AWS Glue

利用者のスキルセットやユースケースにあわせて適切な開発ツールを選択可能

## AWS Glue Studio



ETLディベロッパー

## AWS Glue DataBrew



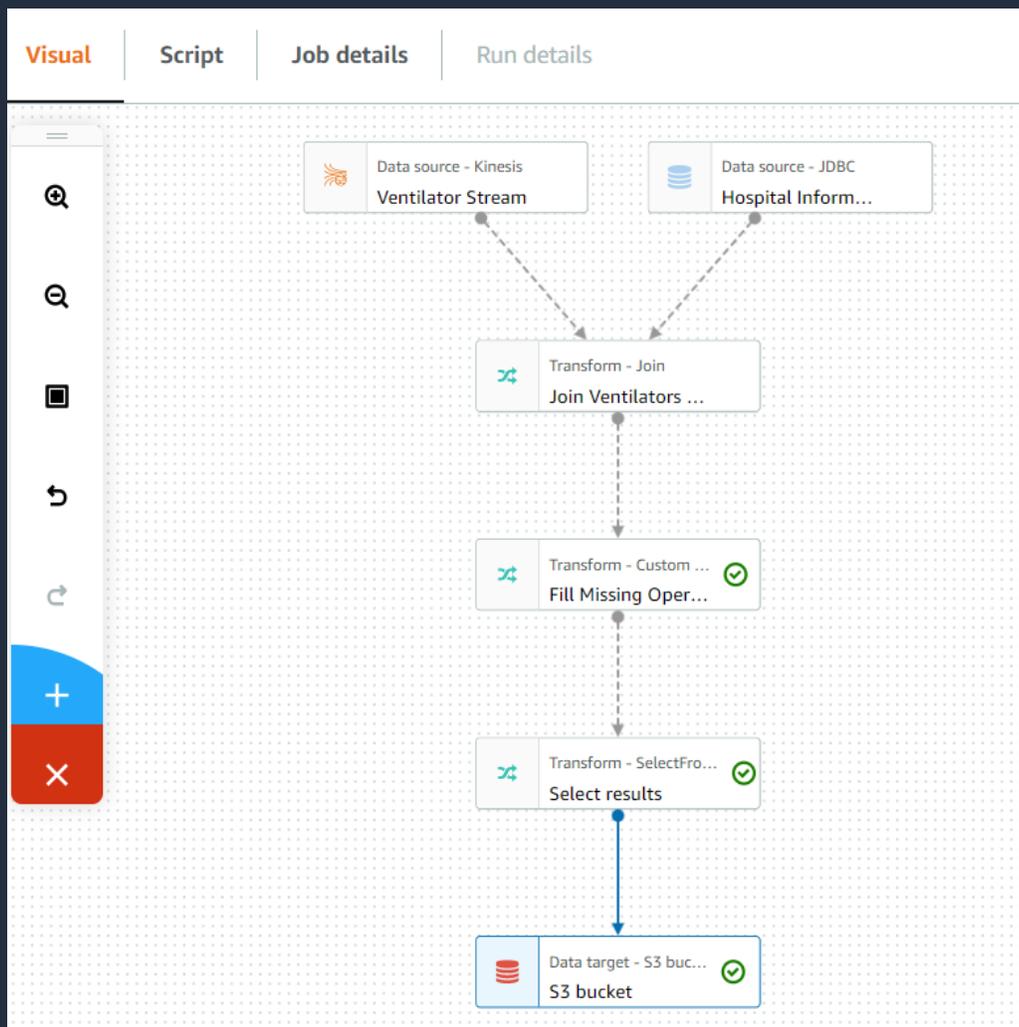
ビジネスアナリスト  
データサイエンティスト

# AWS Glue Studio

## Glueジョブを構築するGUI



## ETLディベロッパー



コードを記述せずにETLジョブを視覚的にオーサリング

コンソールから数千のジョブを監視

分散処理実施のための学習コストを削減

再利用可能なコードを使った高度な変換

# AWS Glue DataBrew

分析やマシンラーニングのデータ準備ツール



ビジネスアナリスト  
データサイエンティスト

The screenshot displays the AWS Glue DataBrew interface. The top section shows a dataset named 'dataset-met-objects' with 500 rows. Below this, a 'Data profile overview' is shown for a dataset named 'fifa-players-data'. The overview includes a summary of 15,465 total rows and 23 total columns. It also shows data types (8 Integer, 15 String), missing cells (8%), and duplicate rows (0%). A 'Correlations' heatmap is visible, showing the relationship between variables like 'softfa\_id', 'age', 'height\_cm', 'weight\_kg', 'overall', 'potential', 'team\_jersey\_number', and 'contract\_valid\_until'.

リッチなビジュアルインターフェイス  
でデータを整形・正規化

250以上の組み込み変換機能を選択し  
タスクを自動化

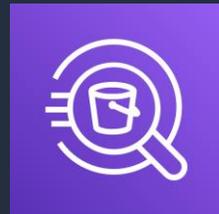
データパターンと異常値を把握するための  
データプロファイル機能

大規模なデータセットを操作可能

# サーバレスで分析クエリ

# Amazon Athena

## サーバーレスでインタラクティブなクエリ実行サービス



データベース: qshandson

テーブルとビューのフィルタリング...

▼ テーブル (1) テーブルの作成

- ▼ flightdata2 (パーティション化)
- flight\_date (date)
- flight\_number (string)
- origin\_city (string)
- origin\_state (string)
- destination\_city (string)
- destination\_state (string)
- departure\_delay (decimal(22,6))
- arrival\_delay (decimal(22,6))
- cancelled (decimal(22,6))
- diverted (decimal(22,6))
- air\_time (decimal(22,6))
- distance (decimal(22,6))
- distance\_group (decimal(22,6))
- carrier\_delay (decimal(22,6))
- weather\_delay (decimal(22,6))
- security\_delay (decimal(22,6))
- late\_aircraft\_delay (decimal(22,6))
- origin\_airport\_name (string)
- cancellation\_reason (string)
- destination\_airport\_name (string)
- carrier\_name (string)
- year (int) (パーティション化)
- month (int) (パーティション化)

新しいクエリ 1 | 新しいクエリ 2 | **新しいクエリ 3** | 新しいクエリ

```
1 SELECT year, count(*) from flightdata2 GROUP BY year;
```

クエリの実行 | 名前を付けて保存 | 作成 (実行時間: 4.32 秒, ス...

クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します

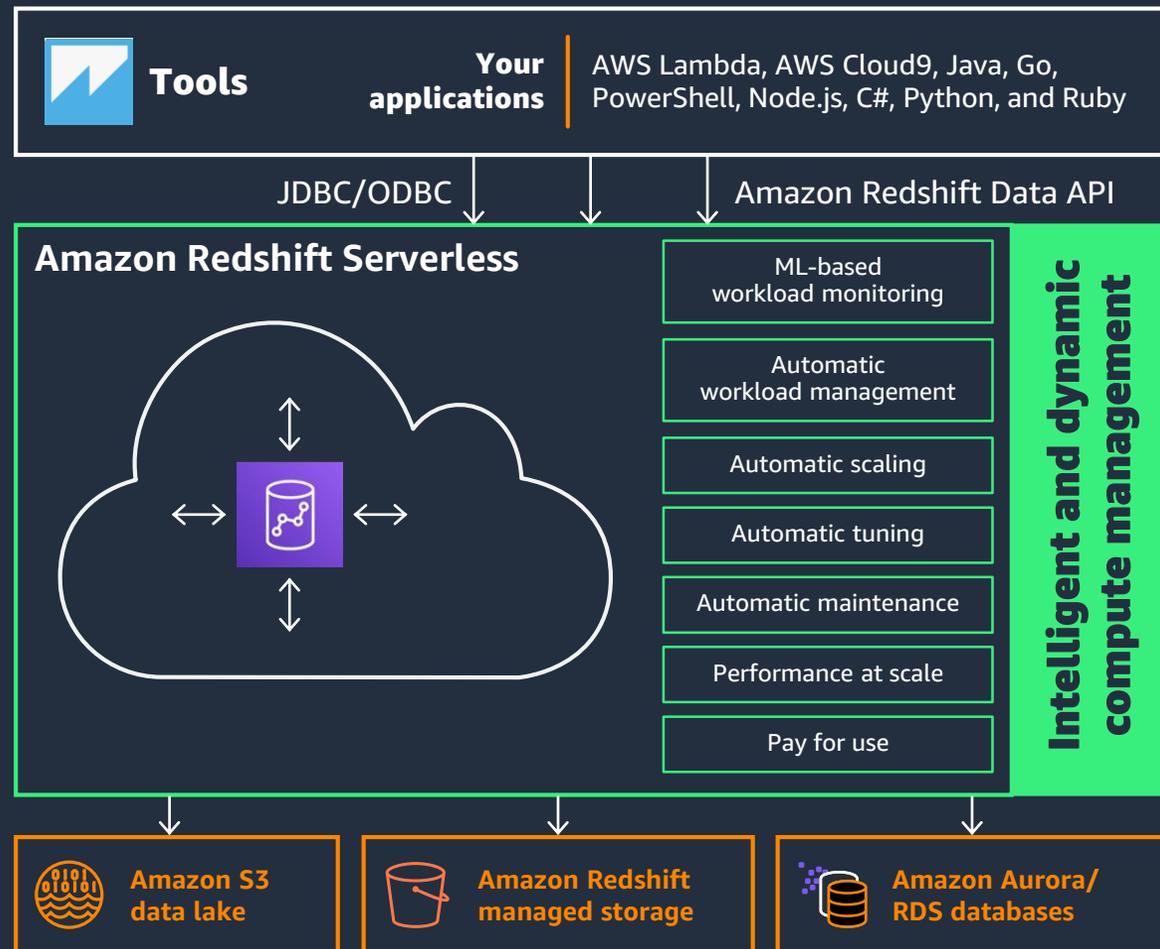
結果

	year
1	2011
2	2010
3	2014
4	2012
5	2013

- S3上に保存したファイルにSQLを実行可能に
- Prestoベース、ANSI SQL
- 実行されたクエリに対してのみ課金
- 課金はデータ圧縮後のスキャン量に対し実施。クエリごとのコストを30%~90%削減

# Amazon Redshift Serverless

## クラスタの管理をせず高速DWHがすぐ利用可能に



- ノードタイプ、ノード数、ワークロード管理、スケールアップなどを実行する必要はなく、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
- 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

### Amazon Redshift SQL のすべての機能が利用可能

- Security and user management
- Data lake queries
- Complex joins
- Federated query
- Semi-structured data
- Durability and transactional guarantees
- Data sharing
- JDBC/ODBC and Data API
- Machine learning functions
- And more

# Amazon Redshift Serverless

Provisioned(クラスター)型よりサーバーレスがフィットするユースケース

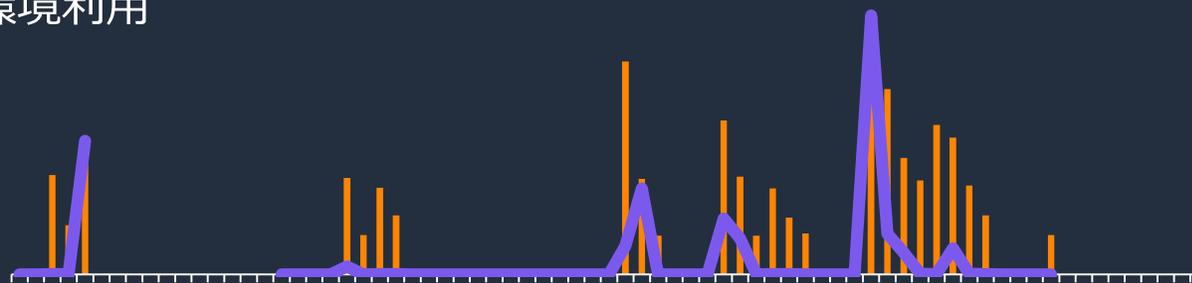
- インフラを意識せず すぐ簡単に分析開始

例) アドホックビジネス分析、dev/test環境利用

- 未知かつ予測が難しい

ワークロード

例) アドホック分析



- 定期実行ワークロード

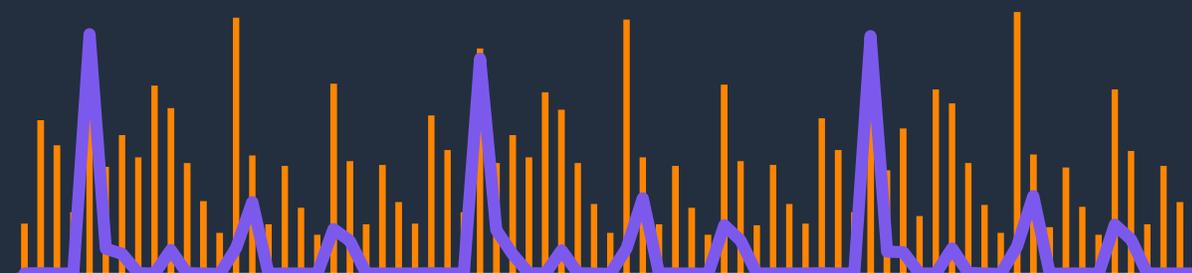
例) 定期ETLバッチジョブ



- 定常的にスパイクを伴う

ワークロード

例) 定常実行ジョブとデータサイエンティストの高度なアドホッククエリの混在



© 2022, Amazon Web Services, Inc. or its affiliates.

query\_count scanned\_data

# 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
- 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

New!

- Markdown形式と複数SQL入力に対応したSQL Notebooks もサポート

※2022/6現在プレビュー

The screenshot displays the Amazon Redshift query editor v2 interface. The main window shows a SQL query: `1 select * from date limit 10`. Below the query, the results are displayed in a table with columns: dateid, caldate, day, week, and month. The table contains 10 rows of data for dates from 2008-01-10 to 2008-01-23. Below the table, there is a section for adding a Markdown cell. In the bottom right corner, there is a pie chart visualization showing the distribution of sales across various categories: Books (13%), Home (12%), Women (12%), Electronics (11%), Music (11%), Shoes (8%), Sports (8%), Jewelry (5%), Men (5%), and Children (5%).

dateid	caldate	day	week	month
1831	2008-01-10	FR	2	JAN
1836	2008-01-11	SA	3	JAN
1837	2008-01-11	SA	3	JAN
1840	2008-01-14	TU	3	JAN
1843	2008-01-17	FR	3	JAN
1845	2008-01-19	SU	4	JAN
1846	2008-01-20	MO	4	JAN
1847	2008-01-21	TU	4	JAN
1849	2008-01-23	TH	4	JAN

# 参考) 各種クエリサービスの使い分け



Amazon  
Athena



Amazon  
Redshift  
Serverless



Amazon  
Redshift  
Provisioned

用途

簡易な処理

簡易～複雑な処理

(例: 複数の結合やサブクエリ、  
レコード単位での UPDATE/DELETE)

簡易～複雑な処理

(例: 複数の結合やサブクエリ、  
レコード単位での UPDATE/DELETE)

規模

低頻度または予測不能  
ワークロード  
(小～中規模)

低頻度または予測不能  
ワークロード  
(中～大規模)  
RPU 32～512

予測可能  
ワークロード  
(中～超大規模)  
最大 ra3.16xlarge 128 ノード

基盤管理

データレイク構築  
自動スケール

自動

プロビジョニング  
詳細なチューニング可

費用

クエリスキャン量  
(別途ストレージサービス使用料)

ワークロード実行時間  
RMS ストレージ使用量

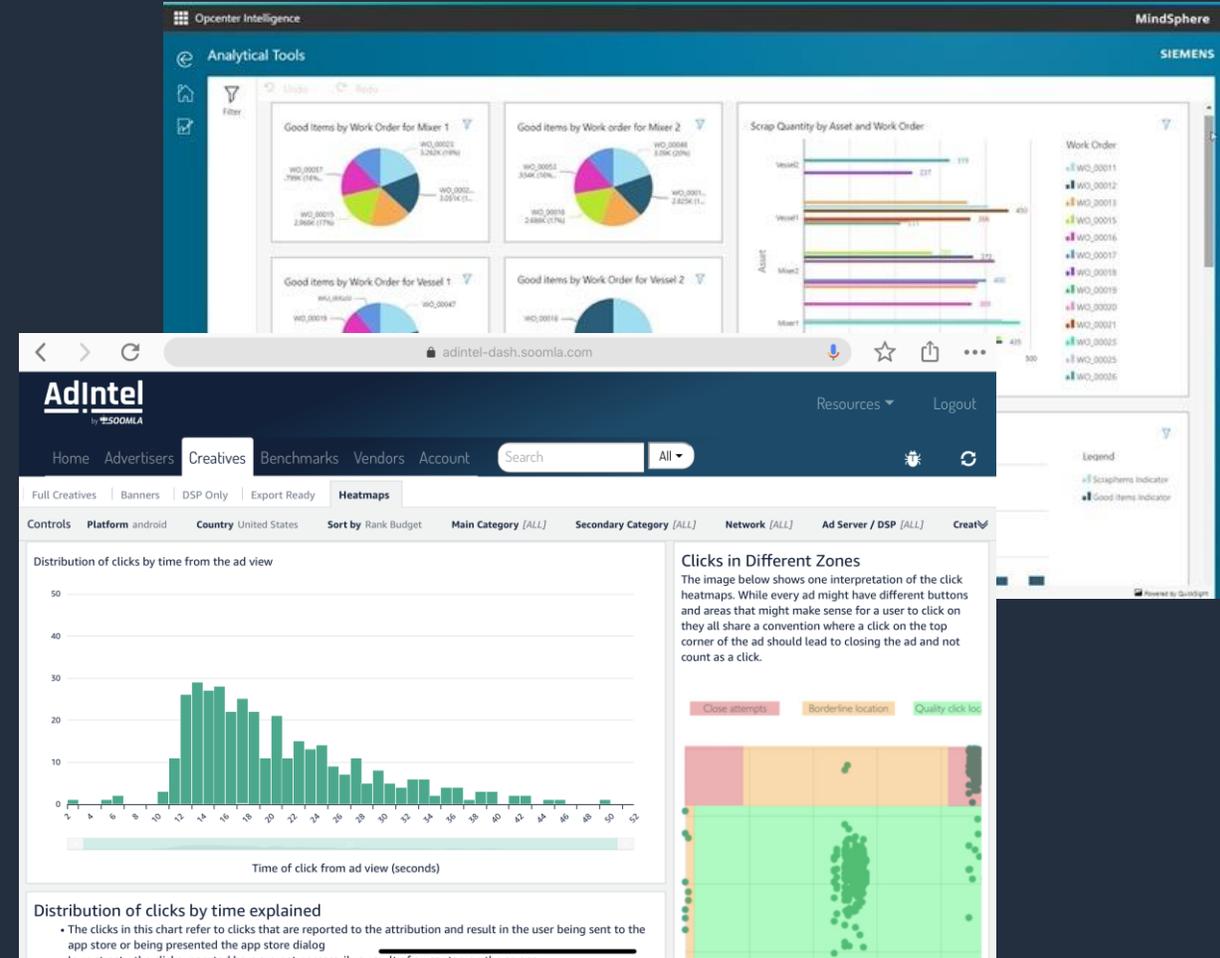
クラスタ稼働時間  
(リザーブドインスタンス適応可能)  
RMS ストレージ使用量  
クエリスキャン量  
(Spectrum 利用時)

# サーバレスで可視化

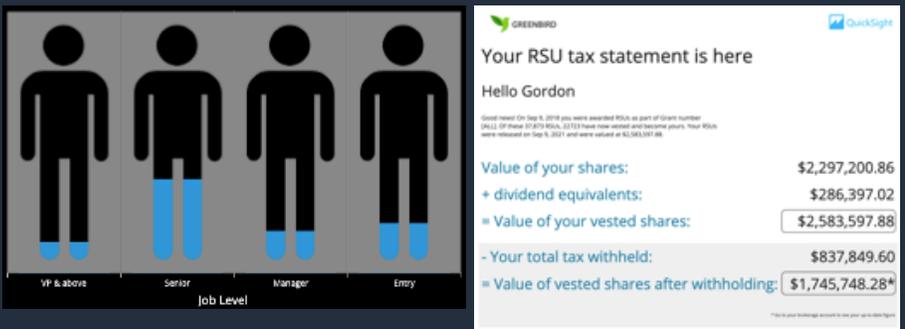
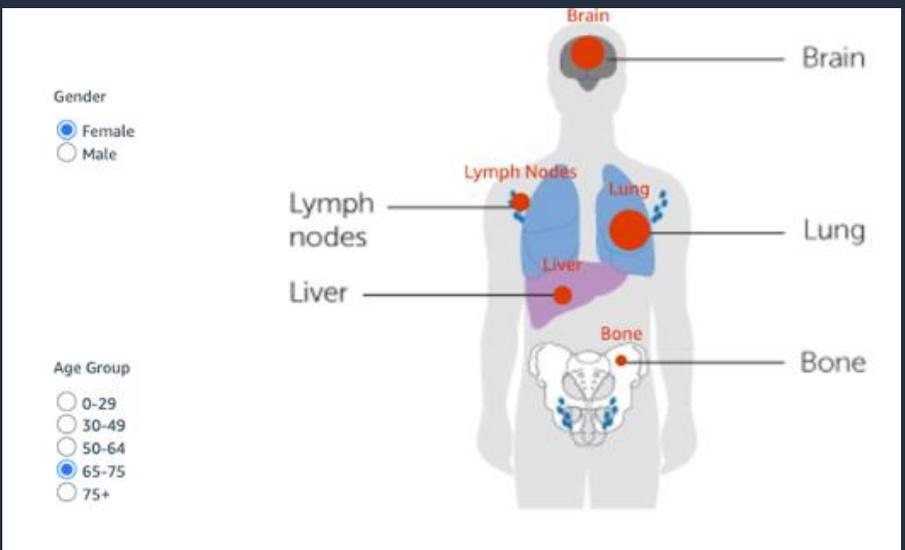


# Amazon QuickSight

- クラウドに最適化されたBI
- サーバーレス = サーバー管理不要
- 月単位のサブスクリプション
- 閲覧者は従量課金 = 使わなければゼロ円
- 柔軟なアクセス権限管理



# 直感的で インタラクティブな 分析サービス

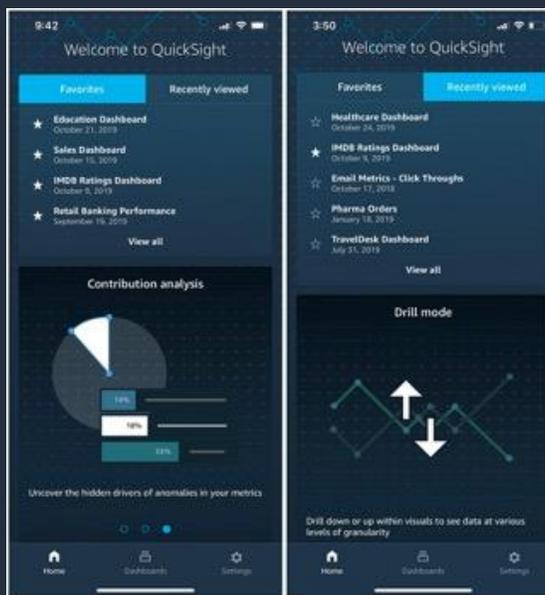


# いつでも、どこでも、閲覧できるBI

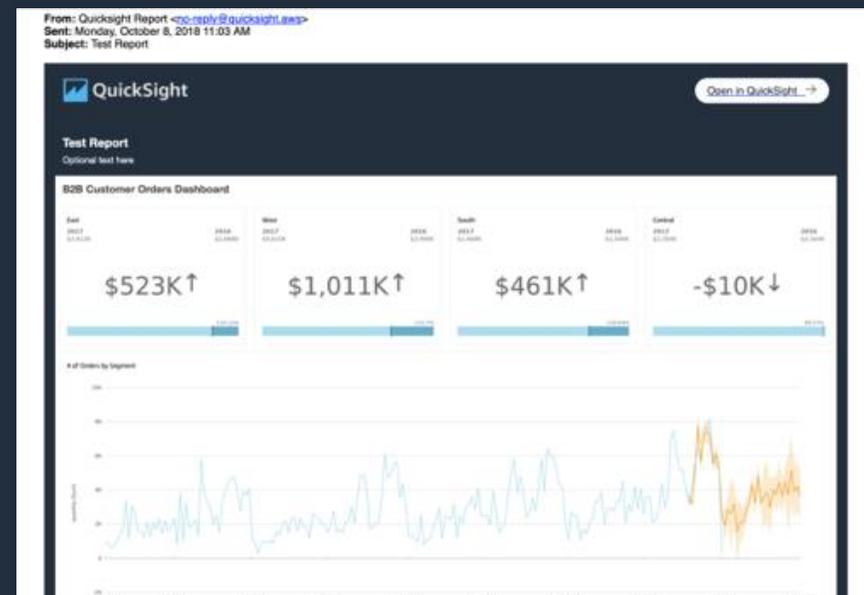
PC



モバイル

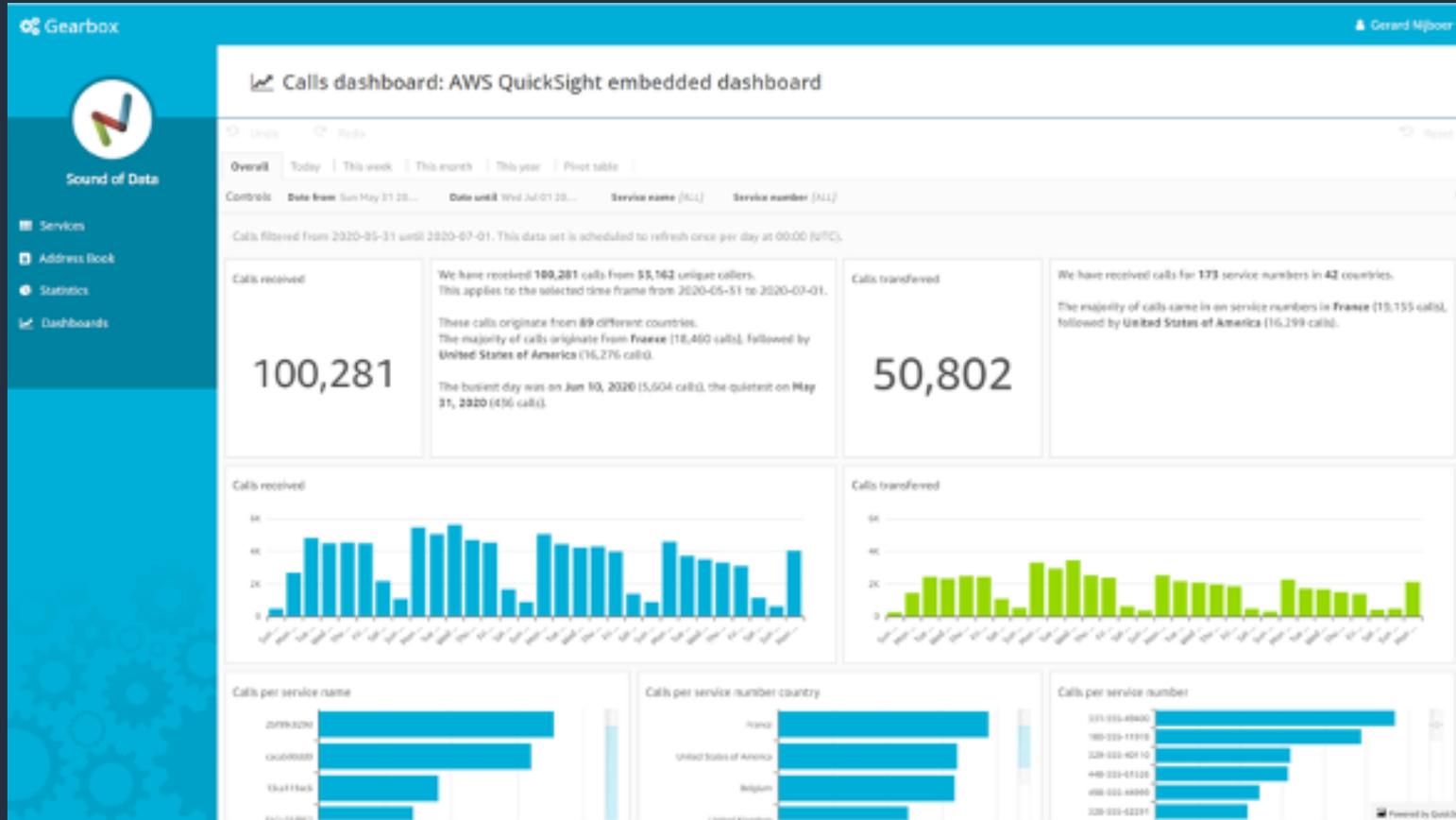


Emailレポート



# 組み込みアナリティクス

- ・ インタラクティブな分析機能を社内ポータルサイトやSaaSに組み込みことも可能



sound of data様 事例

<https://aws.amazon.com/jp/quicksight/embedded-analytics/>

# 機械学習機能のML Insightで分析を深める



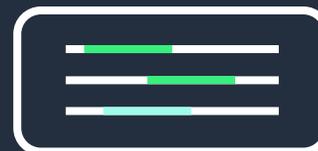
## 異常検知

異常な動きを検知し、膨大なデータの中からその要因を探り出す



## 予測

ポイント&クリックの簡単操作で予測を実現



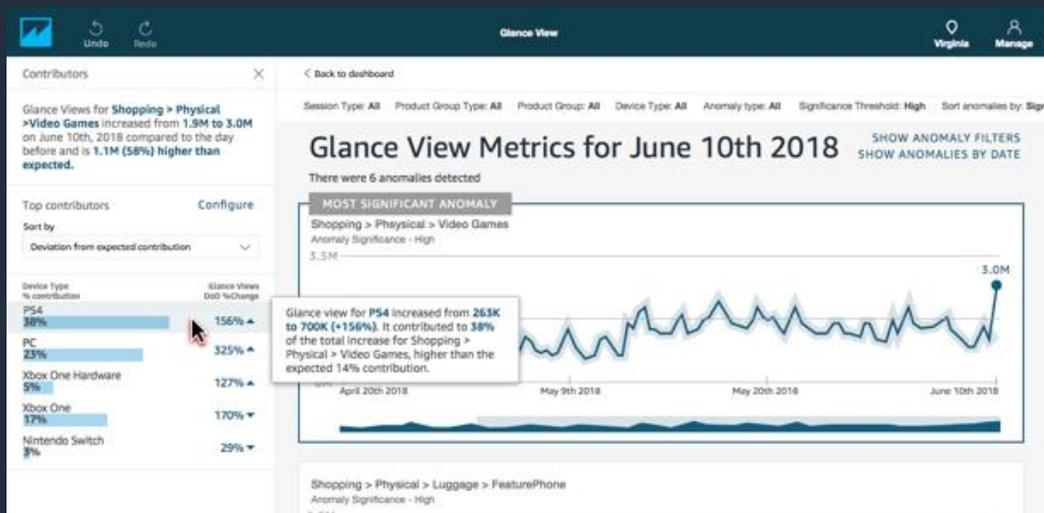
## 自動ナラティブ

分かりやすい文章でデータの状況を表現



## ML予測連携

SageMakerで作成したモデルと連携し可視化



# まとめ

# サーバレスのサービスで手軽に分析を始めよう

- AWSのAnalytics サービスにはサーバレスオプションが豊富
- データのカタログ化からデータ加工処理の開発・実行までをAWS Glueがカバー
- データへのクエリは、利用状況に応じてAmazon AthenaとAmazon Redshift Serverlessを使い分け
- Amazon QuickSightで手軽にリッチな可視化を実現



**Thank you!**