

AWS でデータ分析ならここから！ はじめての Amazon Redshift Serverless

野上 恭平

アマゾン ウェブ サービス ジャパン合同会社
AWS 技術統括本部 インターネットメディアソリューション部
ソリューションアーキテクト

野上恭平

所属とロール

アマゾン ウェブ サービス ジャパン合同会社
技術統括本部
インターネットメディアソリューション部
ソリューションアーキテクト

普段の業務

Web 系のお客様を中心に
AWS に関する様々な技術支援を実施

好きな AWS サービス

Amazon Redshift, Amazon QuickSight



Agenda

- データ分析のオーバービュー
- よくある課題と Redshift による解決
 - どのように Redshift にアクセスするのか？
 - どのようにしてデータを取り込むのか？
 - JSON 形式のデータはどう取り扱えるのか？
- Next Step

データ分析のオーバービュー

データ活用のユースケース

オペレーション改善

需要と財務の予測
プロセスの自動化と最適化
サプライチェーンの最適化
不正行為の検知
異常検知

マーケティング最適化

Customer 360
ターゲティングの改善
パフォーマンスの理解向上
チャーン分析

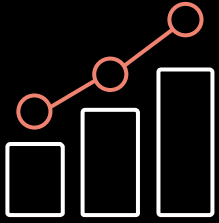
顧客体験

パーソナライゼーション
コンタクトセンター・インテリジェンス
カスタマー・インサイトに基づく製品開発

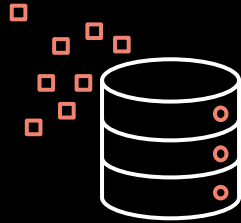
アプリケーション開発

データベースのアプリケーション
開発業務

データ活用の要求が増えるにともなって データ基盤に求められる要件が広がっている



増加する
データへの対応



新しいデータ
ソースへの対応



多様なデータの
取り込み



多様なスキル
セットのユーザ

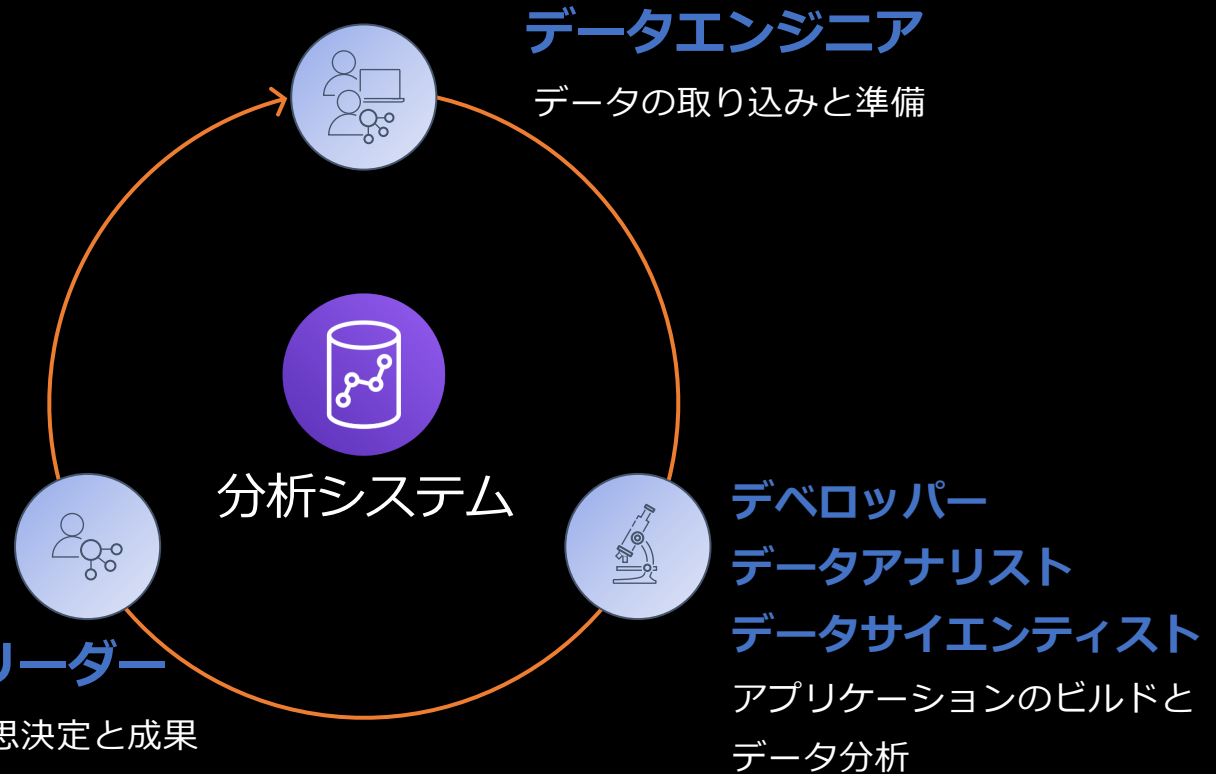


さまざまな
アプリケーション
との接続

分析システムに求められる要素

拡張が容易で、
さまざまなデータに対応し、
多様なユーザーが簡単に利用できる
分析システムにより
データをより効果的に活用する

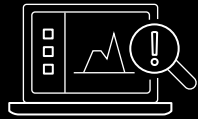
ビジネスリーダー
ビジネス上の意思決定と成果



Amazon Redshift

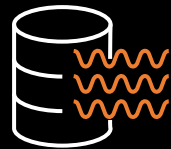
数万人の顧客を抱え、最も広く使用されているクラウドデータウェアハウス

誰でも簡単に
分析できる



インフラストラクチャを気にすることなく、数秒でデータからインサイトを得ることに集中

全てのデータを
分析する



データベース、データレイク、データウェアハウスにまたがる複雑でスケーリングされたデータに対して、リアルタイムの予測分析を実行するインサイトを取得

コストパフォーマンス
の高さ



他のクラウドデータウェアハウスに比べて、最大 3 倍のコストパフォーマンスを実現し、動的にスケーリングして複雑で重要なワークロードのクエリ速度を向上

Amazon Redshift Serverless

Public Preview

シンプルなユーザー体験



クラスターを管理することなく、
データ分析の実行やスケーリングが可能に

Redshiftと同じ機能を持ち、 高いパフォーマンス



Redshift の豊富な SQL の機能を利用でき、
業界をリードするの価格性能比を実現

インテリジェントに スケールする実行環境



一貫した高速なパフォーマンスを提供するために、
容量を自動的に調整する

使った分だけの従量課金



コンピューター料金はSQLの実行時間に応じて
秒単位でのお支払い、アイドル時間の料金はかからない

Redshift Serverless は簡単に始められる

The image shows a composite screenshot of the Amazon Redshift Serverless interface. On the left, the console page 'Get started with Amazon Redshift Serverless (Preview)' is visible, featuring a 'Try Amazon Redshift Serverless (Preview)' button. On the right, a callout box states 'Get to powerful insights fast' and 'The Amazon Redshift serverless experience makes it easy for customers to run and scale analytics without having to provision and manage their data warehouse. Simply load and query data.' In the foreground, the 'Redshift query editor v2' is open, displaying a SQL query for a TPC-DS subset. A green notification banner at the top of the editor reads 'tpcds sample data is loaded successfully.' The query editor interface includes a sidebar with a file tree, a central SQL editor with a 'Run' button, and a bottom status bar.

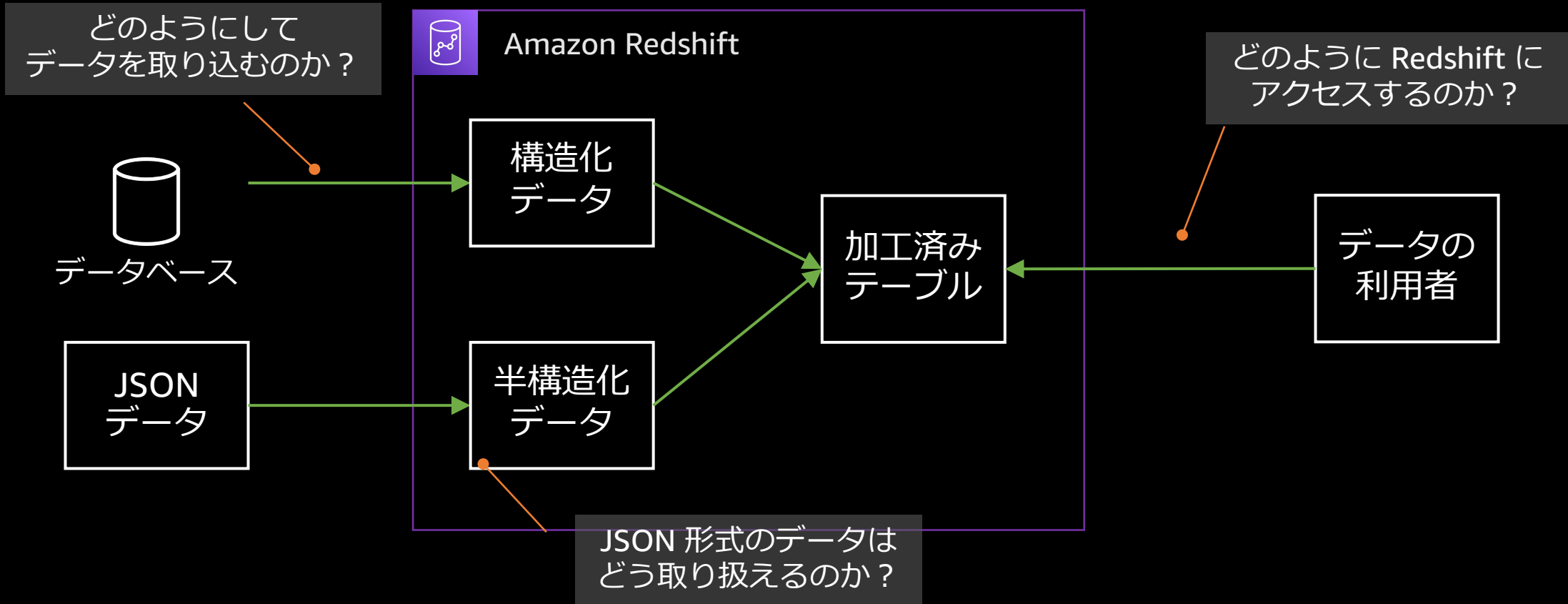
1 Redshift Serverless を有効化

2 デフォルト設定を確認

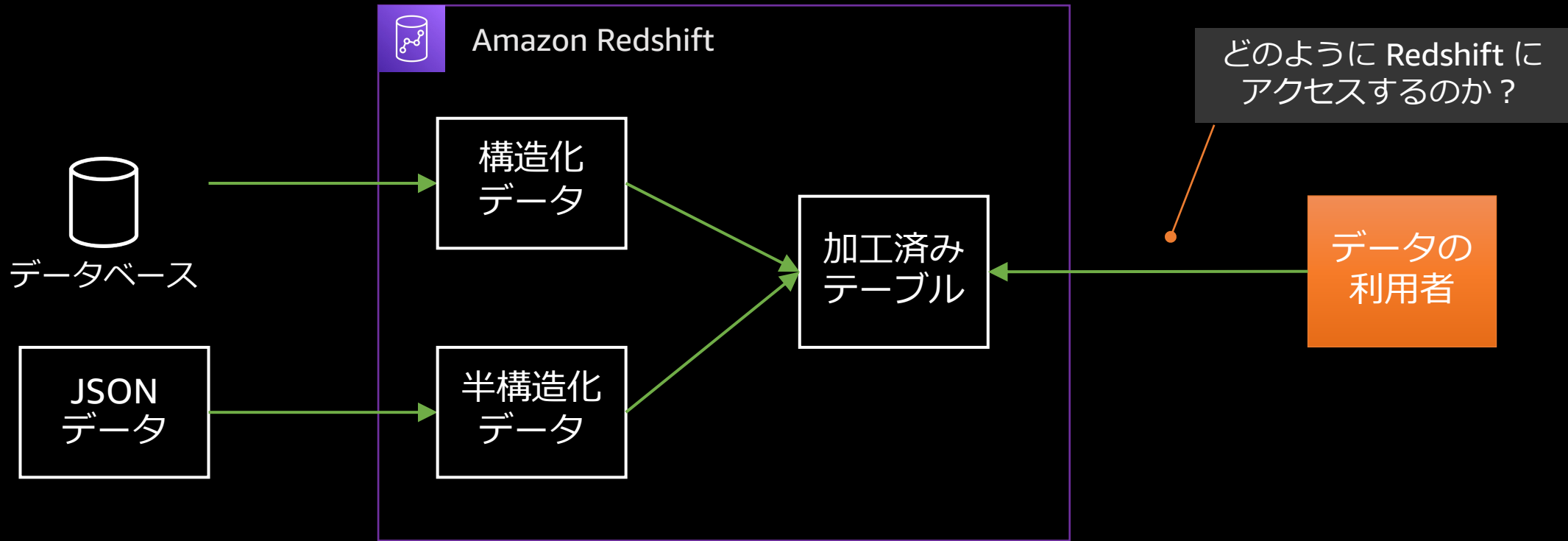
3 ブラウザでクライアントにアクセス

よくある課題と Redshift による解決

Amazon Redshift Serverless でのデータ分析



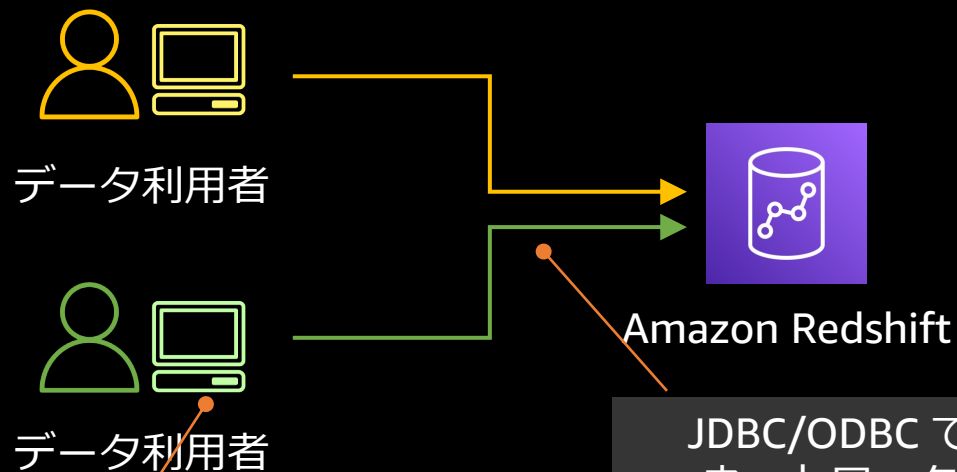
Amazon Redshift Serverless でのデータ分析



どのように Redshift にアクセスするのか？

よくある 課題

- Redshift にアクセスして SQL を実行するには、クラスター、クライアント間でネットワークの疎通をとる必要がある。
- データの利用者は、ローカル PC 上でクライアントをセットアップする必要がある。



各 PC 上でクライアントの
セットアップが必要

JDBC/ODBC で接続するのに
ネットワークの疎通が必要

Amazon Redshift Query Editor v2

データ探索と SQL による分析のためのブラウザベースの分析ツール

The screenshot displays the Amazon Redshift Query Editor v2 interface. On the left, a navigation pane shows the database structure, including tables like 'daily_historical_sales' and 'catalog_sales'. The main area contains a SQL query editor with a query that selects various aggregate functions (avg, sum) from the 'catalog_sales' table, grouped by category, zip, and state. Below the query editor, the 'Result 1 (100)' section shows a pie chart visualization. The pie chart is titled 'Click to enter Plot title' and displays data for various categories: Books (13%), Home (12%), Women (12%), Electronics (11%), Music (11%), Shoes (11%), Sports (9%), Jewelry (8%), Men (8%), and Children (5%). A legend on the right side of the pie chart lists these categories with corresponding colors. A '+ Trace' button is visible in the top right of the visualization panel, and a red arrow points to it. The interface also includes a 'Run' button, 'Limit 100' setting, and 'Explain' option.

セットアップなしで利用できる
ブラウザベースの SQL クライアント

ツールの管理が不要

SQL の編集と実行、データの可視化、
共同作業をサポート

Query Editor v2 が解決する課題をBreakDown

データの投入にも SQL が必要

表形式の結果表示はできても、
グラフの描画ができない

定型クエリの定義ができない

他の分析者との共同作業が難しい

Query Editor v2 が解決する課題をBreakDown

データの投入にも SQL が必要

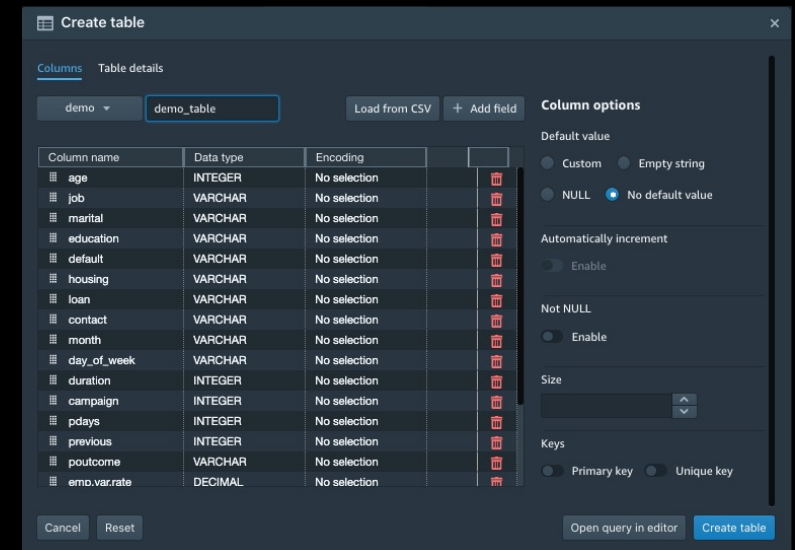
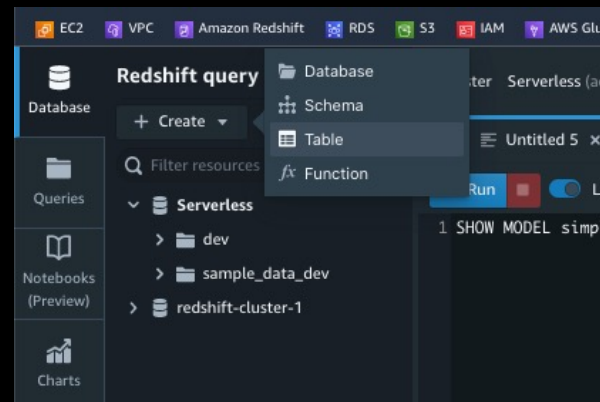
表形式の結果表示はできても、
グラフの描画ができない

定型クエリの定義ができない

他の分析者との共同作業が難しい

画面操作でのデータインポートエクスポート

- ローカルの CSV ファイルを直接アップロードして **テーブル定義を自動的に識別**し、テーブル作成
- Amazon S3にアップロードしたファイルをGUIからロード可能
- クエリ結果を CSV, JSON 形式でエクスポート



Query Editor v2 が解決する課題をBreakDown

データの投入にも SQL が必要

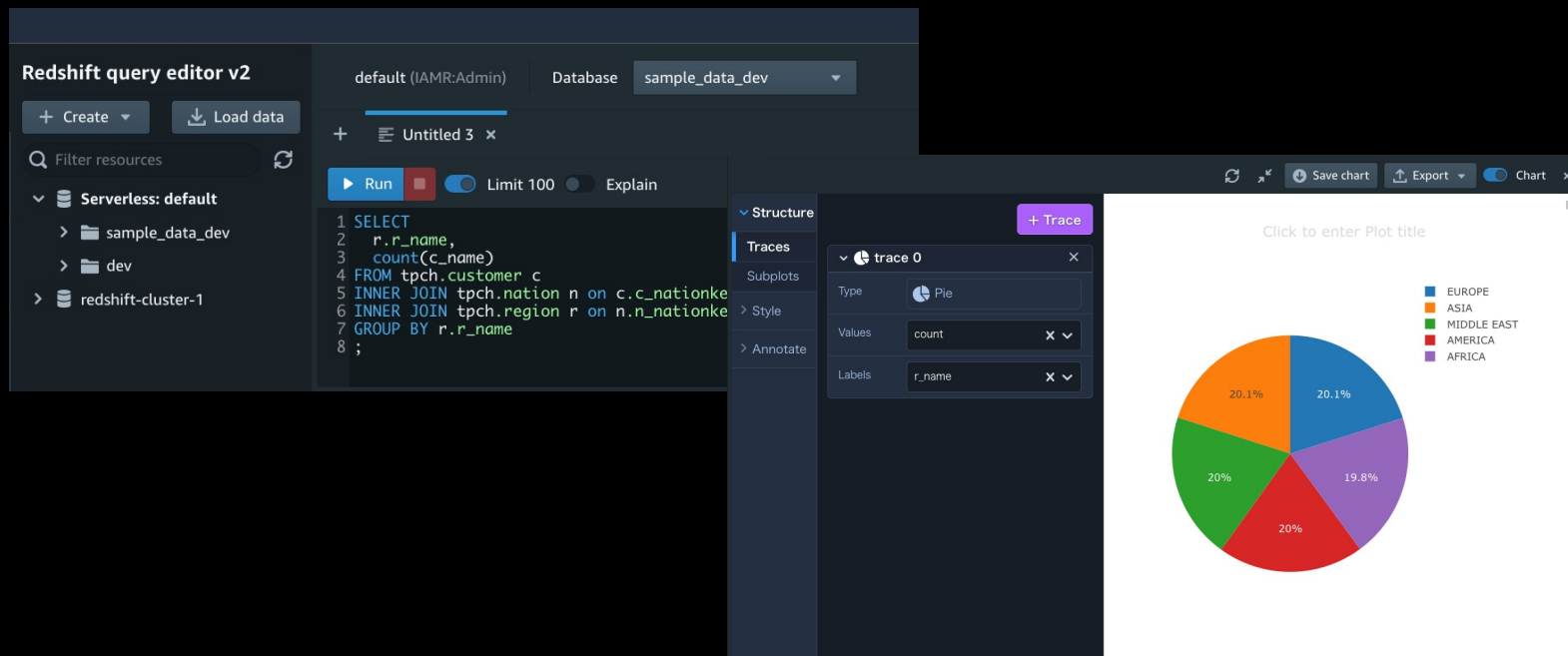
表形式の結果表示はできても、
グラフの描画ができない

定型クエリの定義ができない

他の分析者との共同作業が難しい

組み込みの可視化機能

- クエリ結果を、表形式だけではなくグラフでの表示が可能
- 散布図、折れ線グラフ、円グラフなど11個の可視化をサポート



Query Editor v2 が解決する課題をBreakDown

データの投入にも SQL が必要

表形式の結果表示はできても、
グラフの描画ができない

定型クエリの定義ができない

他の分析者との共同作業が難しい

パラメータ化されたクエリ

- よく使う定型的なクエリを保存可能
- クエリの一部をパラメータ化して実行時に条件を指定できる

The screenshot displays the AWS Query Editor v2 interface. At the top, the cluster is identified as 'redshift-cluster-1' and the database as 'dev'. A query editor window shows a SQL query with a parameterized value: `select public.sales.sellerid, sum(pricepaid) from sales where salesid > ${salesid} group by sellerid;`. Below the query, there are buttons for 'Run', 'Limit 100', 'Explain', 'Save', 'Share', and 'Shortcuts'. A 'Run query form' dialog box is open, showing a table with 'Parameters' and 'Value' columns. The 'salesid' parameter is set to '100'. The dialog has 'Cancel', 'Reset form', and 'Run now' buttons. In the background, a table of results is visible with columns 'sellerid' and 'sum'. The footer of the interface includes 'Feedback', 'English (US)', and copyright information: '© 2008 - 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use Cookie preferences'.

Parameters	Value
salesid	100

sellerid	sum
37614	4957
13005	4785
18692	5368
6805	673
7700	7700

Query Editor v2 が解決する課題をBreakDown

データの投入にも SQL が必要

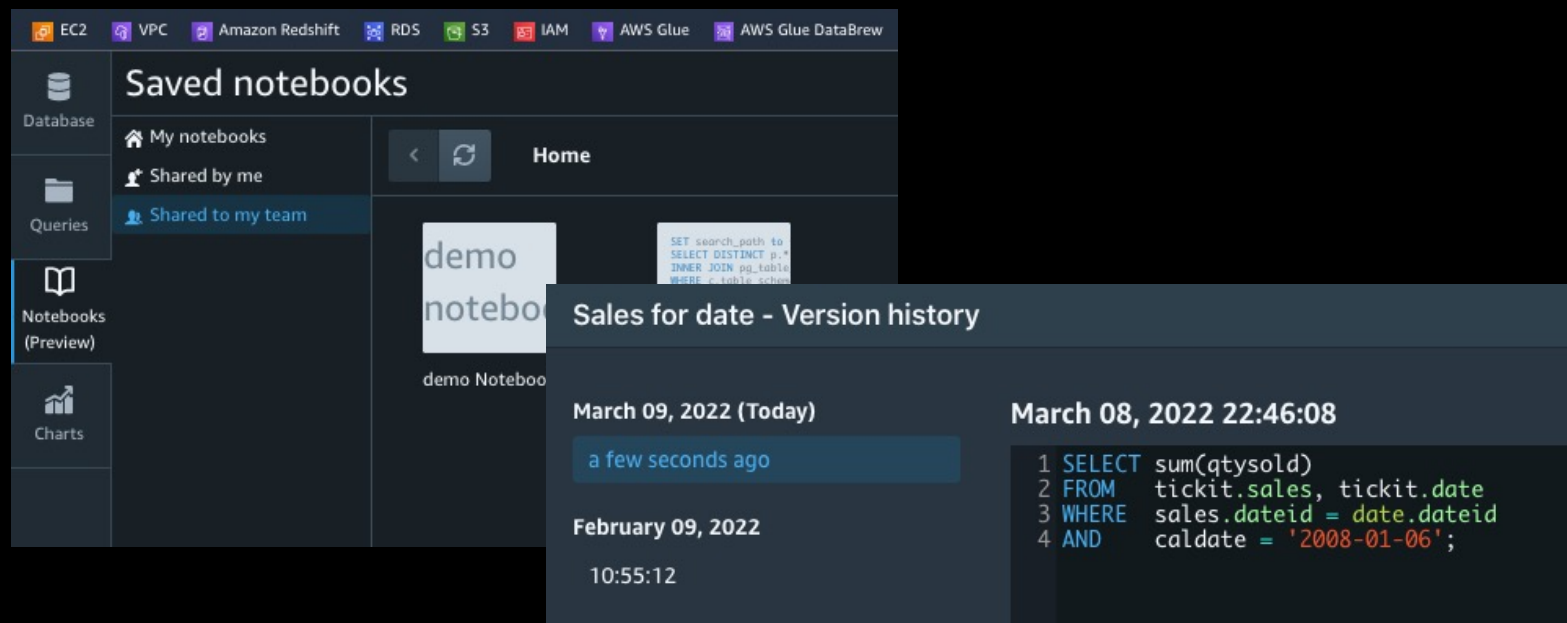
表形式の結果表示はできても、
グラフの描画ができない

定型クエリの定義ができない

他の分析者との共同作業が難しい

チーム共有

- 作成したクエリやグラフ、Notebook はチーム内で共有できる (NotebookはPublic Preview中)
- クエリ資産は変更履歴を自動管理できる



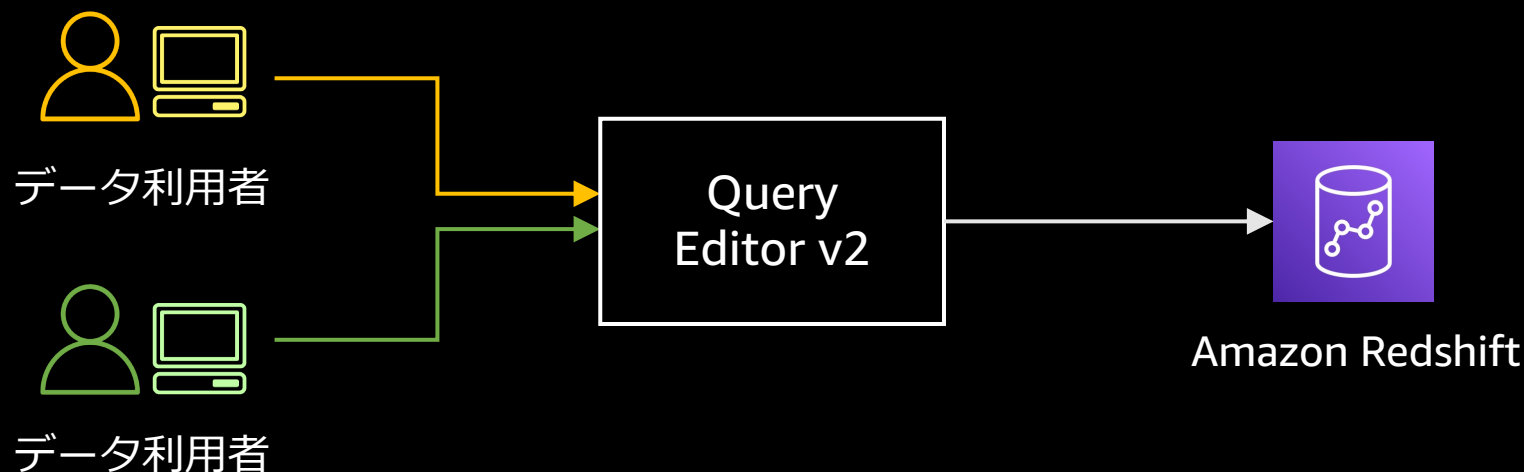
The screenshot displays the AWS Query Editor v2 interface. The top navigation bar includes services like EC2, VPC, Amazon Redshift, RDS, S3, IAM, AWS Glue, and AWS Glue DataBrew. The left sidebar shows navigation options: Database, Queries, Notebooks (Preview), and Charts. The main content area is titled 'Saved notebooks' and shows a 'demo notebook' selected. A 'Version history' overlay is open, showing a list of versions for the notebook 'Sales for date'. The most recent version is from 'March 09, 2022 (Today)' at 'a few seconds ago'. A previous version from 'March 08, 2022 22:46:08' is also visible, with its SQL query displayed:

```
1 SELECT sum(qtysold)
2 FROM tickit.sales, tickit.date
3 WHERE sales.dateid = date.dateid
4 AND caldate = '2008-01-06';
```

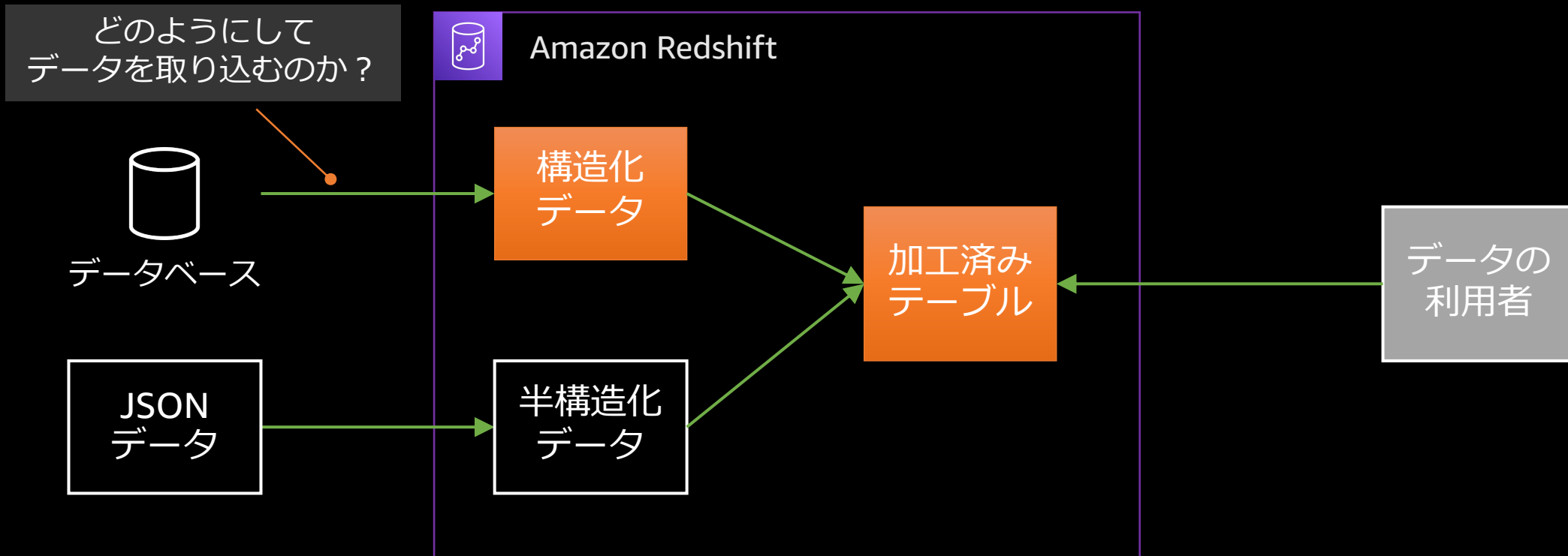
Query Editor v2 でのアクセス

課題の 解決策

- ネットワークで直接接続することなく、Redshift に接続して分析を実行できる。
- 利用者は、ブラウザのみでアクセスして、分析に関わるさまざまなタスクを実施できる。



Amazon Redshift Serverless でのデータ分析



どのようにしてデータを取り込むのか？

よくある 課題

- データベースから Amazon Simple Storage Service(S3) へデータを出力した後、S3 から Redshift に取り込みするのが一般的だが、シンプルに Amazon RDS のデータを取り込む場合でもバッチ処理が必要になる。



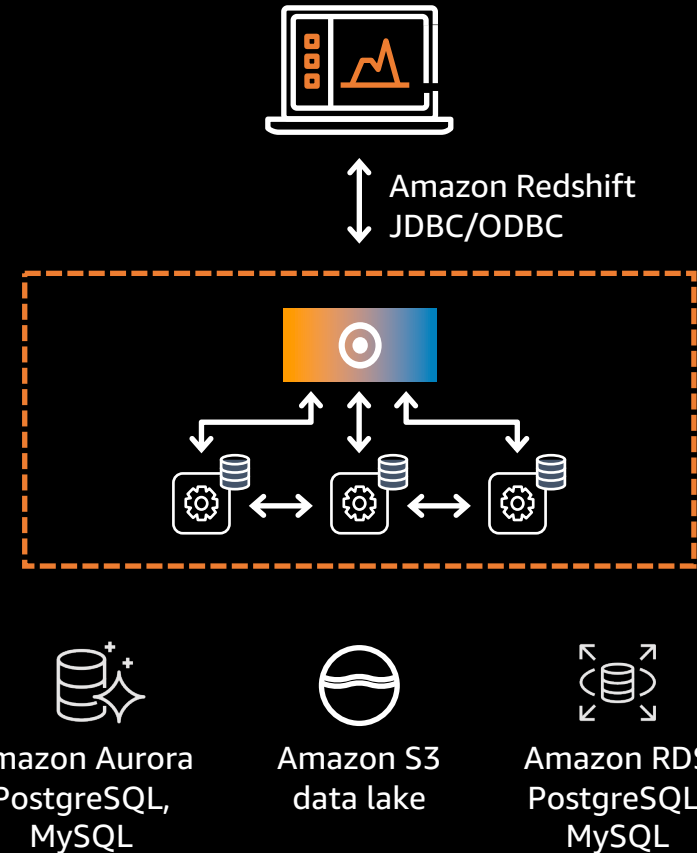
Amazon Redshift Federated Query

さまざまなデータベースをまたがったデータ統合

事前にデータの移動や取り込みをせず、
最新のデータでの分析

Amazon Redshift から Amazon RDS や
Amazon Aurora 上のデータへアクセス

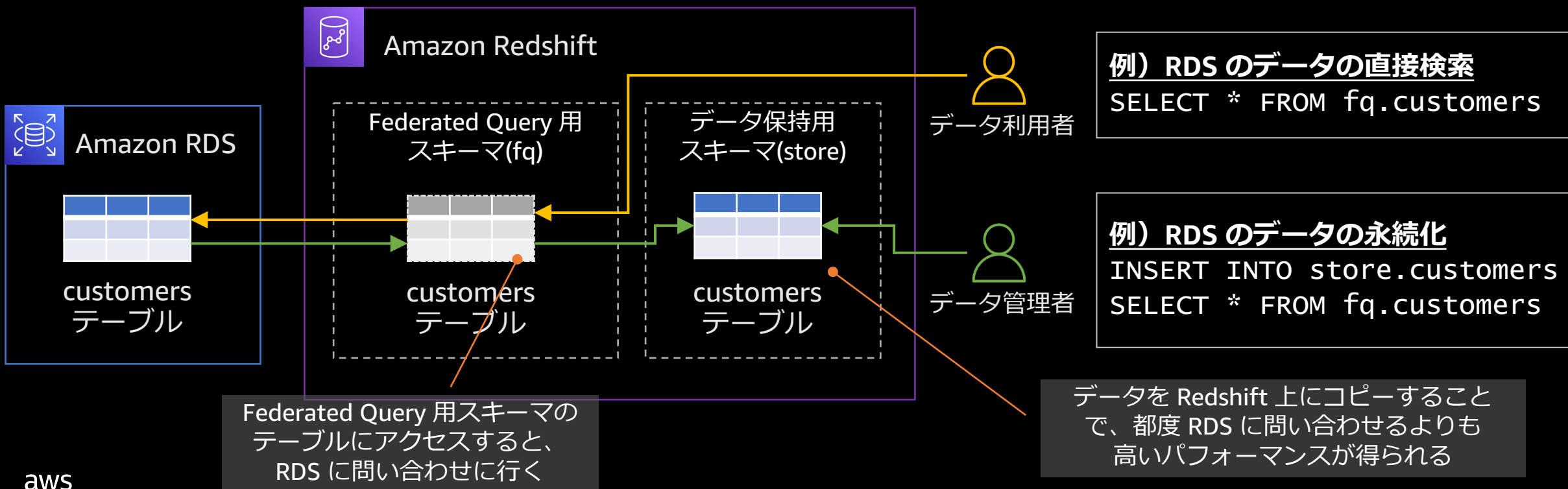
複数のデータベースにまたがるクエリや
データの結合を容易に実現



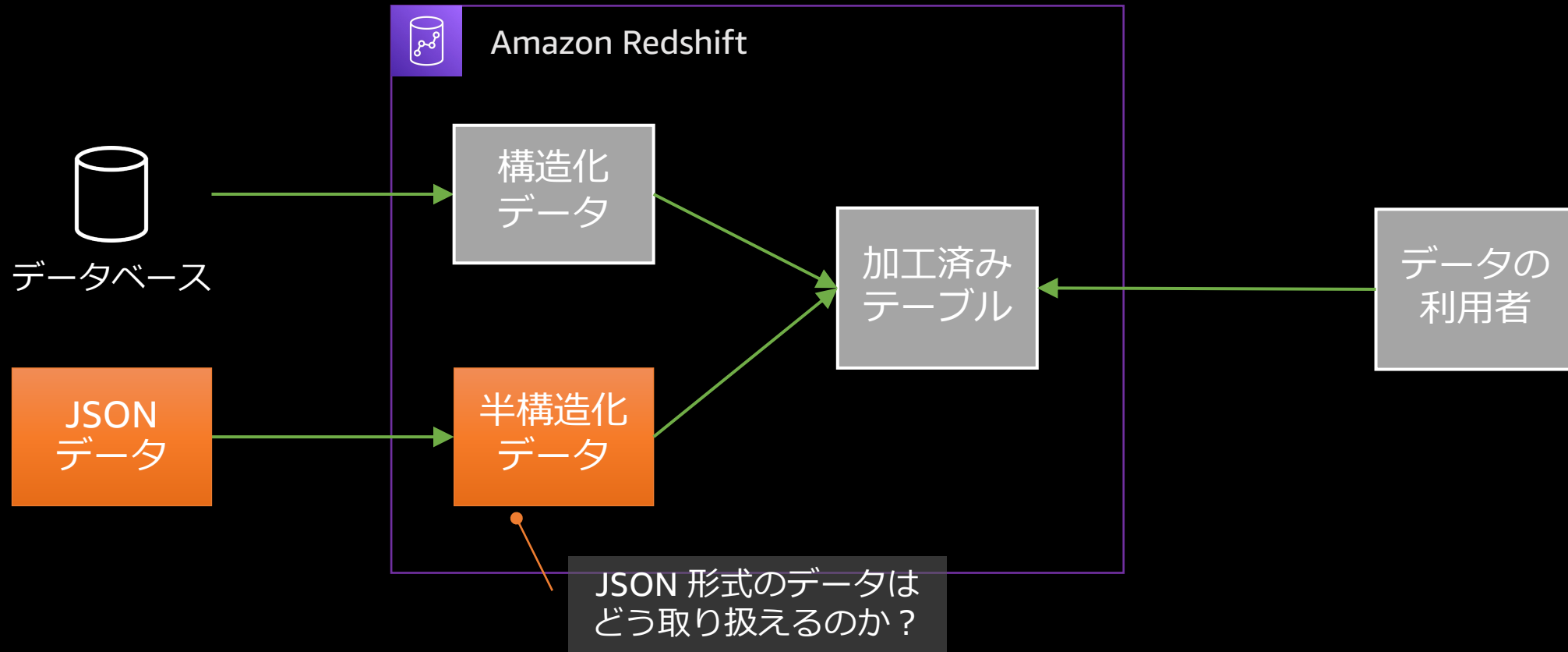
Federated Query によるデータ連携

課題の 解決策

- Federated Query 用のスキーマを使って、透過的に外部テーブルにアクセス。
- クエリが都度 RDS へ送信されないように、Redshift 上のテーブルとしてデータを保持できる。



Amazon Redshift Serverless でのデータ分析



JSON 形式のデータはどう取り扱えるのか？

よくある 課題

- Web のアクセスログなど、JSON 形式で保持している。
- JSON データをリレーショナルデータベースで使う場合は変換が必要。

```
[{
  "id": 1,
  "name": {"first": "Jane", "last": "Doe"},
  "phones": [
    {"type": "work", "num": "9255550100"},
    {"type": "cell", "num": "6505550101"}
  ]
}, {
  "id": 2,
  "name": {"first": "Jack", "last": "Roe"},
  "phones": [
    {"type": "work", "num": "5105550102"}
  ]
}]
```

表へどのように
マッピングするか？

SUPER 型による JSON データのサポート

ネイティブで半構造化データを取り扱える

SUPER 型として列を定義すれば、
JSON データに対して簡単にクエリを実行

データの取り込み時は
JSON ファイルをそのまま取り込み

ネストした JSON に対しても
スキーマを定義せずに SQL でアクセス

id	name	phones
INTEGER	SUPER	SUPER
1	{"first": "Jane", "last": "Doe"}	[{"type": "work", "num": "9255550100"}, {"type": "cell", "num": "6505550101"}]
2	{"first": "Jack", "last": "Doe"}	[{"type": "work", "num": "5105550102"}]

クエリ例)

```
SELECT
  name.first AS firstname, ph.num
FROM
  customers,
  customers.phones ph
WHERE ph.type = 'cell';
```

フィールドに
直接アクセス

配列の要素を展開

結果)

firstname	num
Jane	6505550101

SUPER 型での JSON データの活用

課題の 解決策

- JSON の全体、または、一部を SUPER 型として定義。
- JSON 文字列を展開する関数を利用せず、直接フィールドにアクセス。

```
[{
  "id": 1,
  "name": {"first": "Jane", "last": "Doe"},
  "phones": [
    {"type": "work", "num": "9255550100"},
    {"type": "cell", "num": "6505550101"}
  ]
}, {
  "id": 2,
  "name": {"first": "Jack", "last": "Doe"},
  "phones": [
    {"type": "work", "num": "5105550102"}
  ]
}]
```

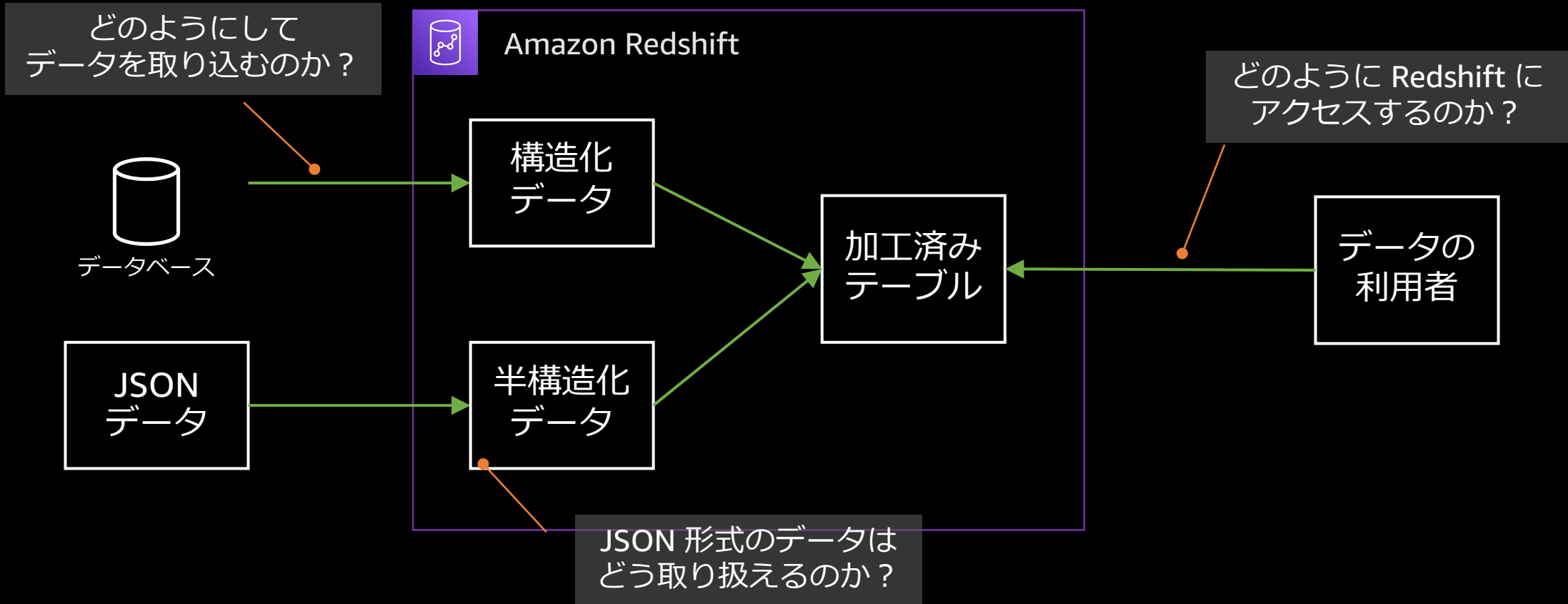
全体を
SUPERで定義

data(SUPER)
{ "id": 1, "name": { "first": "Jane", "last": "Doe" }, "phones": [{ "type": "work", "num": "9255550100" }, { "type": "cell", "num": "6505550101" }] }
{ "id": 2, "name": { "first": "Jack", "last": "Doe" }, "phones": [{ "type": "work", "num": "5105550102" }] }

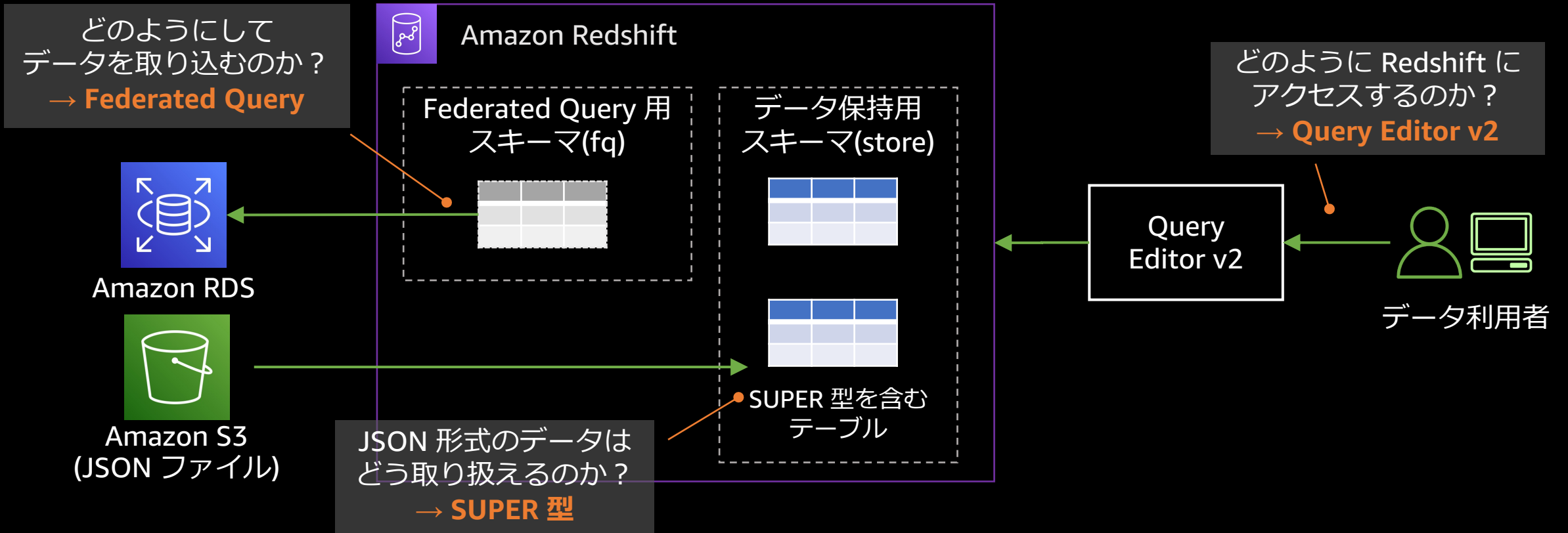
ドット繋ぎで必要なフィールドにアクセス可能
例) SELECT **data.name.first** FROM table;

first
Jane
Jack

Amazon Redshift Serverless でのデータ分析



Amazon Redshift Serverless でのデータ分析



Demo



Demoのシナリオ

- Redshift Serverlessを起動し、 Query Editor v2にアクセス。
- サンプルテーブルにクエリを実行し、可視化。
- SUPER型のテーブルを作成して、S3からJSONデータを取り込む。
- SUPER型のテーブルをクエリして、JSONデータにアクセスする。

Demo

The screenshot shows the Amazon Redshift website interface in Japanese. At the top, there is a navigation bar with the AWS logo, a 'サービス' (Services) menu, a search bar containing 'サービス、特徴、ブログ およびドキュメントなどを検索 [オプション+S]', and a location dropdown set to '東京'. Below the navigation, the main content area is titled 'アナリティクス' (Analytics) and features a large heading 'Amazon Redshift' followed by the sub-headline '迅速、簡単、セキュアな分析を大規模に行うことにより、価値の創出を加速します。' (Accelerate value creation by performing large-scale, fast, simple, and secure analysis). A descriptive paragraph follows, explaining that Amazon Redshift simplifies data warehouse management and analysis. To the right, a callout box titled '強力なインサイトをすばやく得る' (Get powerful insights quickly) highlights the serverless experience and includes a button 'Amazon Redshift サーバーレスを試す' (Try Amazon Redshift serverless). Below this, another callout box titled 'クラスターのプロビジョニングと管理' (Cluster provisioning and management) describes the ease of creating clusters and includes a 'クラスターを作成' (Create cluster) button. At the bottom left, a '仕組み' (How it works) section features a video thumbnail titled 'Introduction to Data Warehousing on AWS with Ama...' with a 'Copy link' button. The footer contains a 'フィードバック' (Feedback) link, a language selection prompt, copyright information for 2022 Amazon Web Services, Inc., and links for 'プライバシー' (Privacy), '用語' (Glossary), and 'Cookie の設定' (Cookie settings).

aws サービス

サービス、特徴、ブログ およびドキュメントなどを検索 [オプション+S]

東京

アナリティクス

Amazon Redshift

迅速、簡単、セキュアな分析を大規模に行うことにより、価値の創出を加速します。

Amazon Redshift により、データウェアハウスを管理する必要がなくなり、分析の実行とスケールアップが容易になります。運用データベース、データレイク、データウェアハウス、数千ものサードパーティーのデータセットにわたってすべてのデータの予測分析を行い、インサイトがリアルタイムで得られます。

強力なインサイトをすばやく得る

Amazon Redshift のサーバーレスエクスペリエンスにより、お客様はデータウェアハウスをプロビジョニングして管理することなく、簡単に分析を実行およびスケールできます。データのロードとクエリを行うだけです。

Amazon Redshift サーバーレスを試す

仕組み

aws Introduction to Data Warehousing on AWS with Ama... Copy link

クラスターのプロビジョニングと管理

数回のクリックで、Amazon Redshift でプロビジョニングされた最初のクラスターを数分で作成できます。

クラスターを作成

フィードバック 言語の選択をお探ですか? 新しい Unified Settings で見つけてください。

© 2022, Amazon Web Services, Inc. またはその関連会社。 プライバシー 用語 Cookie の設定

より発展的なデータ活用をするには？

過去の Builders Online Series で Redshift Serverless および連携するサービスについて解説。

- クラウド BI ツールである Amazon QuickSight での高度な可視化
- データ準備ツールである AWS Glue DataBrew でのデータ準備



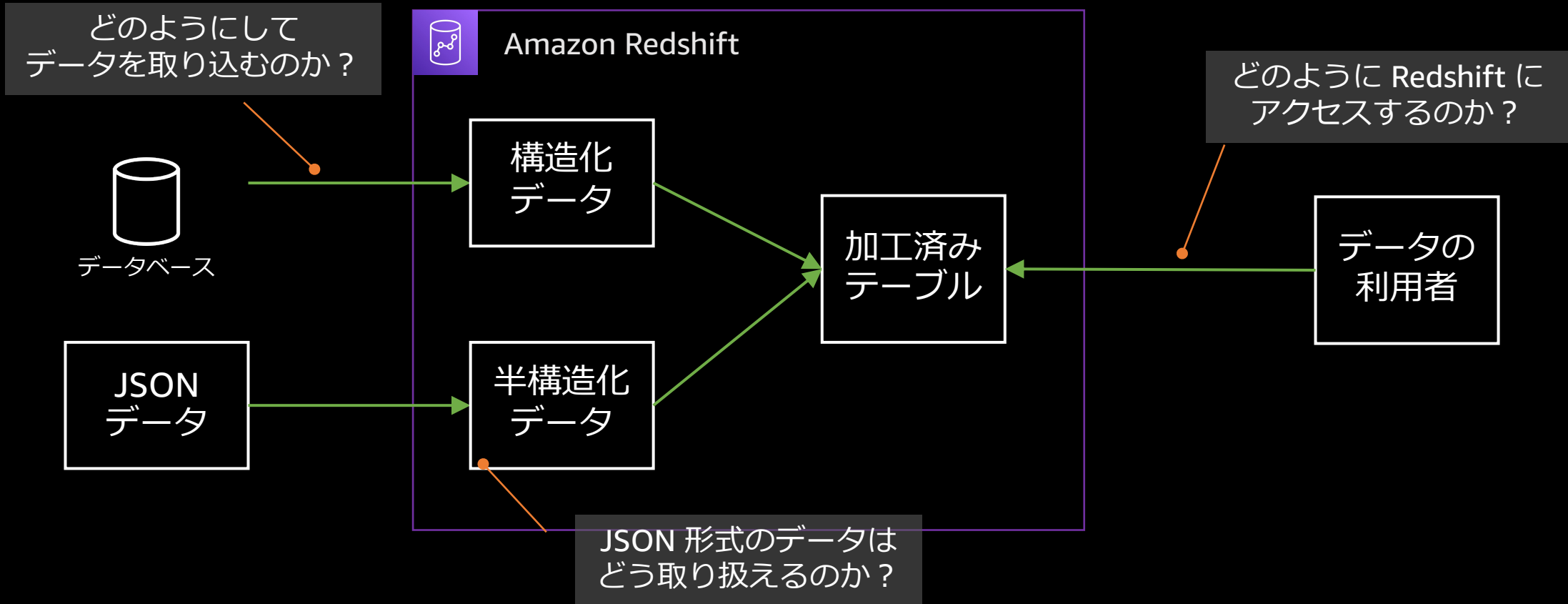
<https://aws.amazon.com/jp/events/aws-builders-online-series-archive/2022-Q1/>

https://pages.awscloud.com/rs/112-TZM-766/images/BOS41_AWS-Builders-Online-Series_2022-Q1_Presentation-Deck.pdf



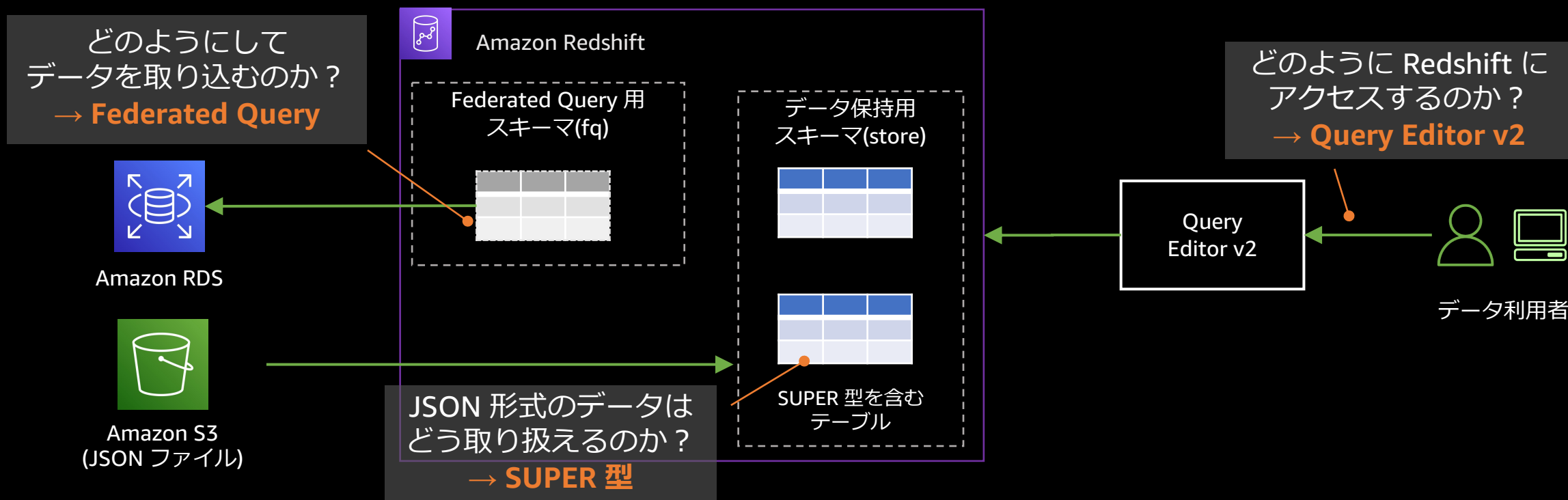
まとめ

Amazon Redshift Serverless でのデータ分析

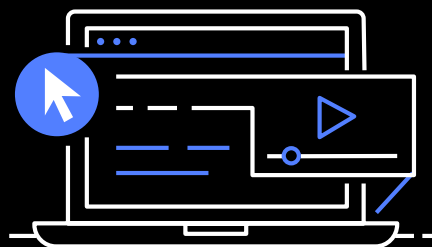


まとめ

Amazon Redshift Serverless はすぐにデータ分析を始められるサービス。
簡単に分析を始めるための機能を3つご紹介しました。



AWS デジタルトレーニング



実力、自信、信頼性を
高め、業界で認められ
た資格で差をつけよう

デジタル学習

- [スキルビルダー](#) – AWS のエキスパートが開発した数百のデジタルトレーニングを自分のスケジュールで学習できます
- [Cloud Quest](#) - AWS Cloud Quest は、実践的なクラウド経験を積み、AWSクラウドのスキルを身につけることができる、初めてで唯一のロールプレイングゲームです

認定試験準備ためのリソース

- [Cloud Practitioner](#) - AWS Certified Cloud Practitioner 取得に役立つリソースをご紹介します
- [Developer – Associate](#) – AWS Certified Developer – Associate 取得に役立つリソースをご紹介します

AWS Builders Online Series に ご参加いただきありがとうございます

楽しんでいただけましたか? ぜひアンケートにご協力ください。
本日のイベントに関するご意見/ご感想や今後のイベントについてのご希望や改善のご提案などがございましたら、ぜひお聞かせください。



aws-apj-marketing@amazon.com



twitter.com/awscloud_jp



facebook.com/600986860012140



<https://www.youtube.com/user/AmazonWebServicesJP>



<https://www.linkedin.com/showcase/aws-careers/>



twitch.tv/aws

Thank you!