

データ分析の最初の一歩 ～ Amazon Redshift Serverless & Amazon QuickSight によるサーバーレス分析 ～

志村 誠

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
アナリティクスソリューションアーキテクト



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

データ分析全体のオーバービュー

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

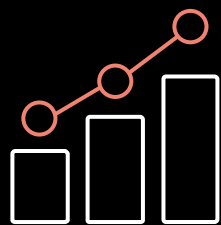
顧客体験

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

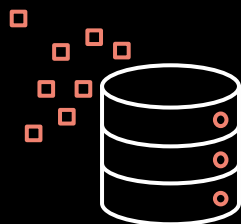
アプリケーション開発

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



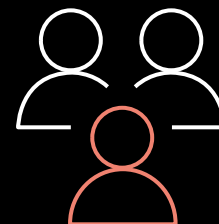
増加する
データへの対応



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

Amazon S3 データレイク



NEW

Serverless and on-demand analytics

PREVIEW AVAILABLE TODAY



AMAZON REDSHIFT
SERVERLESS



AMAZON EMR
SERVERLESS



AMAZON MSK
SERVERLESS



AMAZON KINESIS
ON-DEMAND

Adam Selipsky Keynote より

https://www.youtube.com/watch?v=WGA2P_oH5Xc

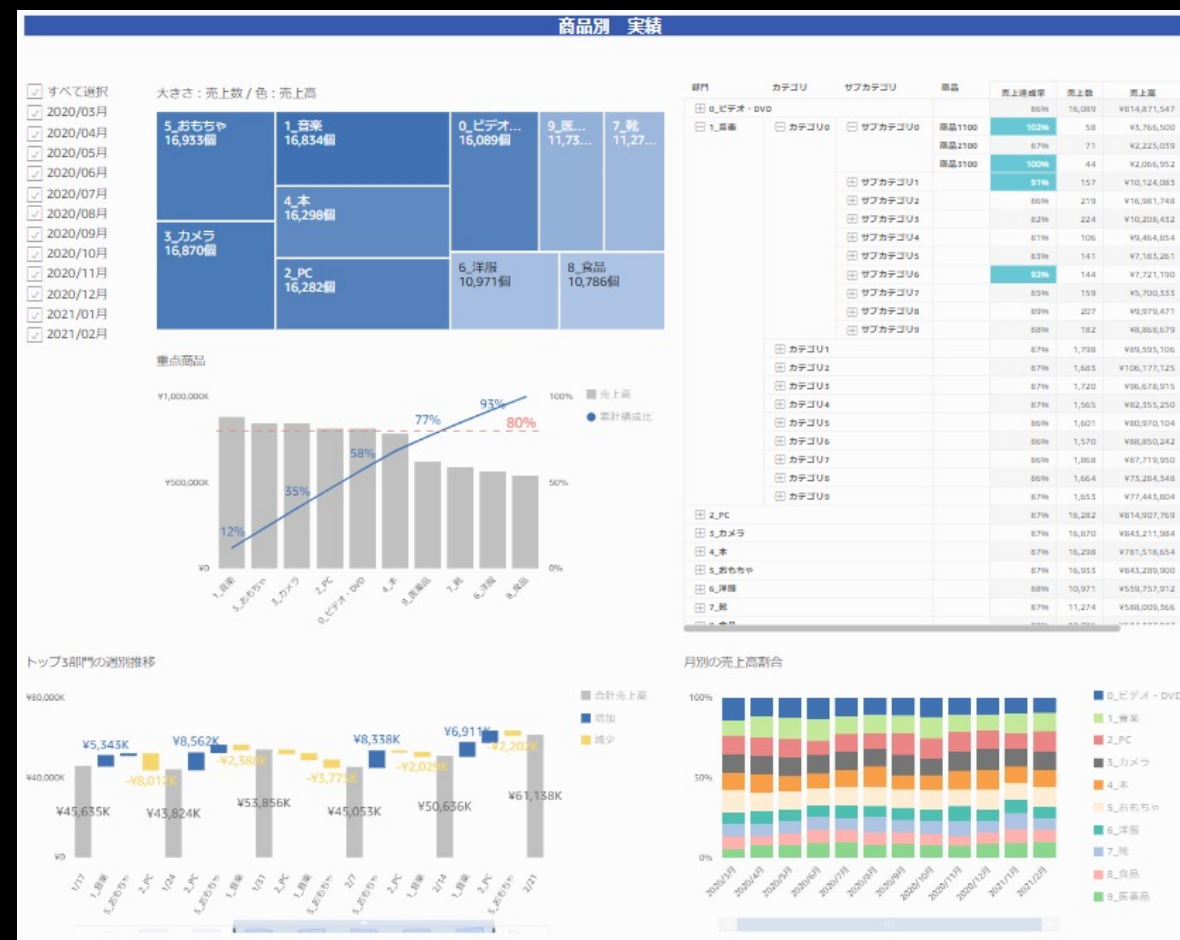
最初の一歩



手元のデータを可視化してインサイトを得る

- 経験や直感だけでなく、ファクトに基づいて判断できると、ビジネスの精度がよりあがる
- 例えば小売業なら、売り上げデータを店舗や商品カテゴリ軸で集計して、傾向を掴む。時系列推移からトレンドを予想するなど...

グラフにしたり表にまとめたり、気になるところを簡単に深掘りできると非常に便利！



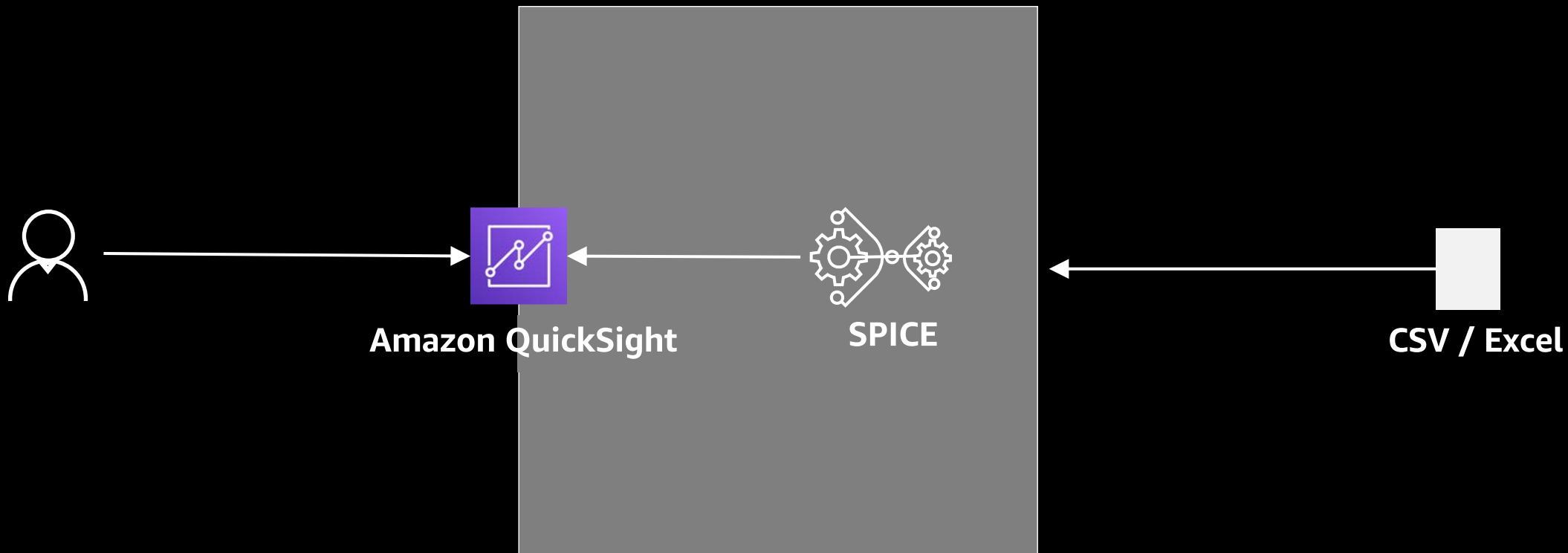
データの可視化にまつわる課題

- 自分の PC のローカル環境だと、少量のデータしか扱えない
- ローカルで作成したグラフを他の人と共有するのが大変
- グラフ作成のために、毎週同じような作業を繰り返し手作業で行っている
- 可視化ツールのライセンスを持っている人が限られており、共有するのが難しい

などなど...

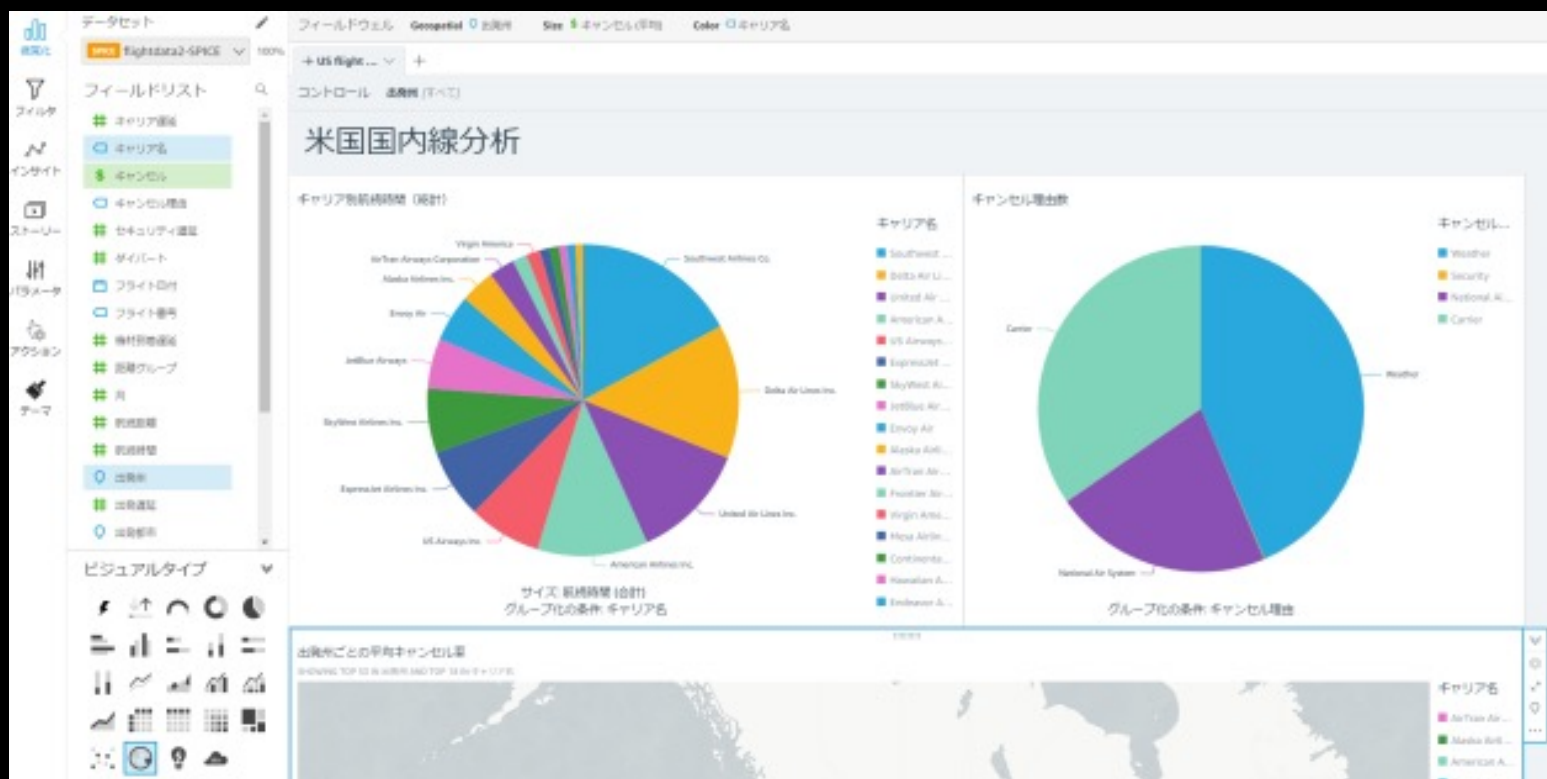


サーバーレス BI による可視化ソリューション



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



Amazon QuickSight クイックスタート

AWS Management Console

AWS services

▼ Recently visited services

Billing	Athena	AWS License Manager	Control Tower
Amazon Redshift	Amazon OpenSearch Service (successor to Amazon Elasticsearch Service)	Amazon SageMaker	CloudFormation
AWS Glue DataBrew	S3	IAM	VPC
Step Functions		AWS Glue	AWS Organizations

▶ All services

Build a solution

Get started with simple wizards and automated workflows.

Launch a virtual machine With EC2 2-3 minutes 	Build a web app With Elastic Beanstalk 6 minutes 	Build using virtual servers With Lightsail 1-2 minutes 	Register a domain With Route 53 3 minutes
Connect an IoT device With AWS IoT 5 minutes 	Start migrating to AWS With AWS MGN 1-2 minutes 	Start a development project With CodeStar 5 minutes 	Deploy a serverless microservice With Lambda, API Gateway 2 minutes

▶ See more

Getting Started with AWS

Learn the fundamentals and start building on AWS now. [Get Started](#)

Stay connected to your AWS resources on-the-go

AWS Console Mobile App now supports four additional regions. Download the AWS Console Mobile App to your iOS or Android mobile device. [Learn more](#)

Explore AWS

Free AWS Training
Advance your career with AWS Cloud Practitioner Essentials—a free, six-hour, foundational course. [Learn more](#)

AWS Certification Resources
Explore the resources available to help you prepare for your AWS Certification. [Learn more](#)

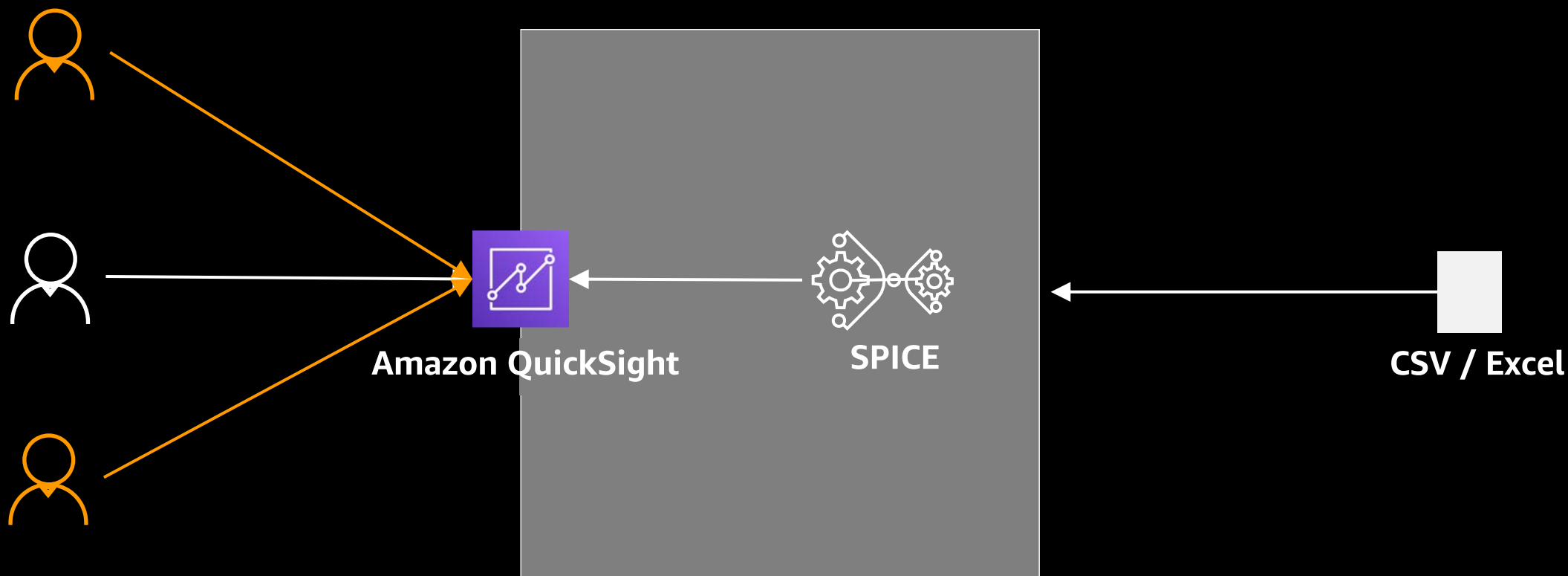
AWS Training
Free digital courses to help you develop your skills. [Learn more](#)

AWS Certification
Propel your career forward with AWS Certification. [Learn more](#)

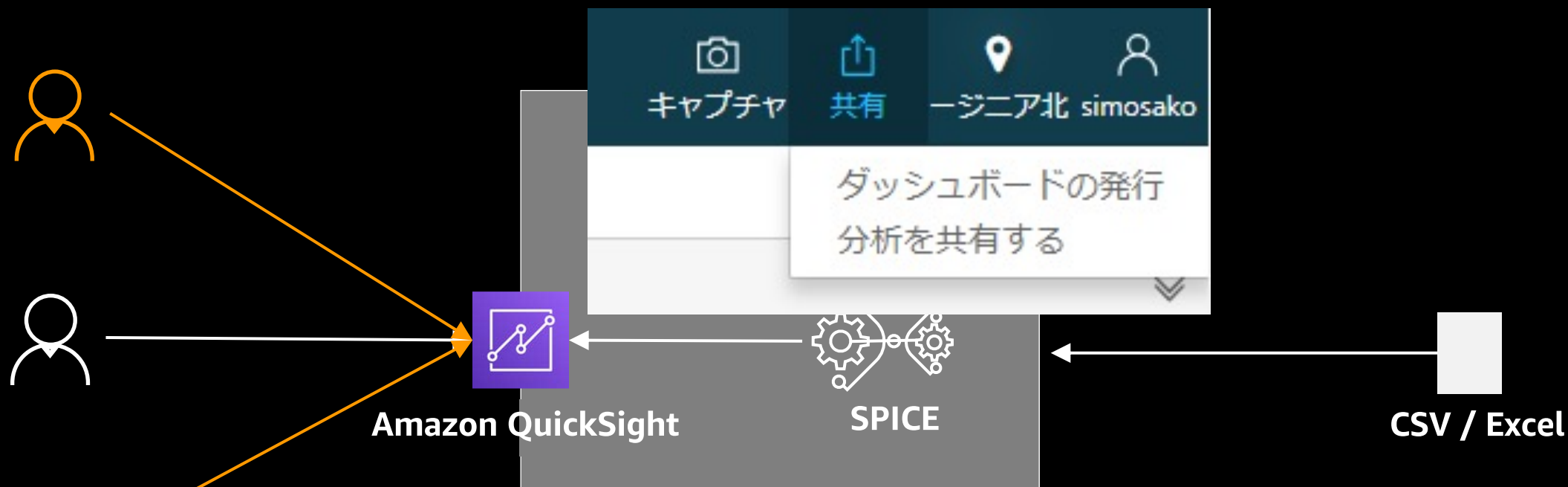
Have feedback?

[Submit feedback](#) to tell us about your experience with the AWS Management Console.

作成したグラフを他の人に簡単に共有

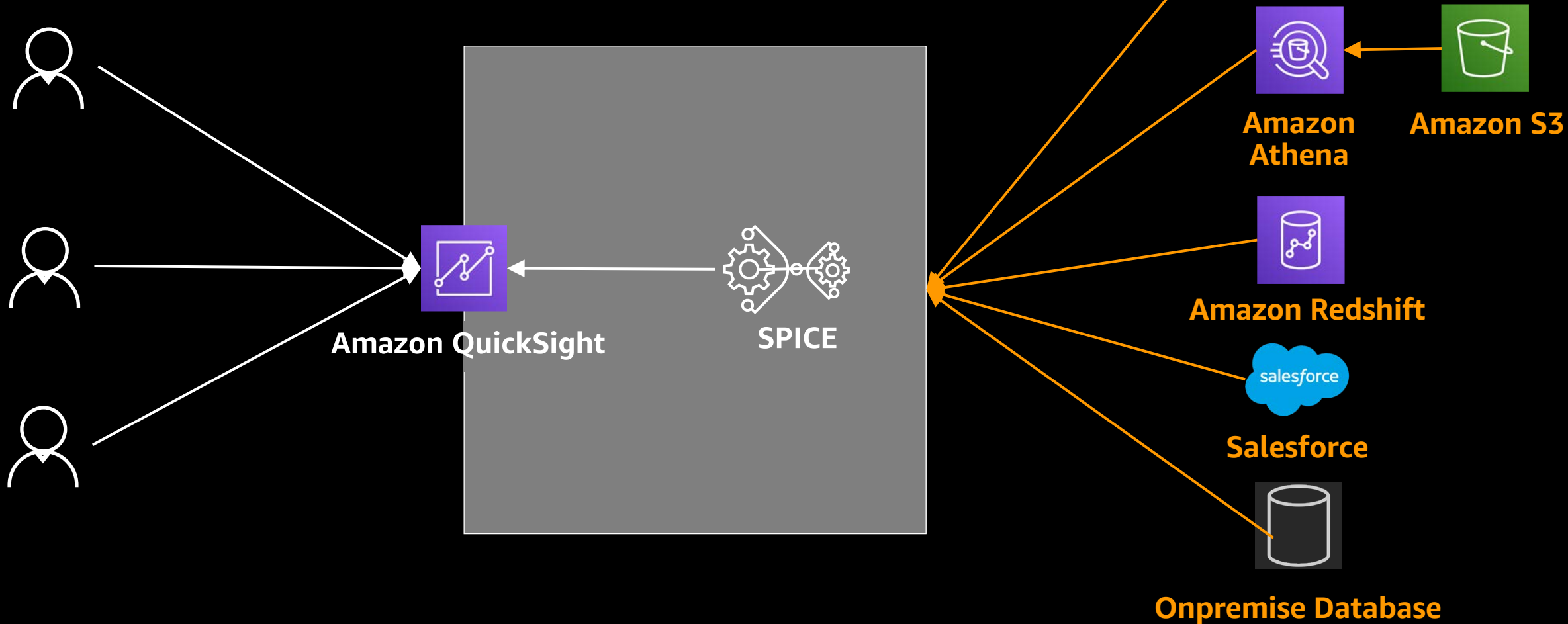


作成したグラフを他の人に簡単に共有



- グループやユーザーを指定して、ほかの人に共有できる
- ダッシュボードの内容を HTML メールとして定期配信も可能
- 従量課金の Reader ライセンス（\$0.3/Session, 月最大 \$5.0*）により、低コストで多くのユーザーにライセンスを展開

さまざまなデータソースに接続可能



さまざまなデータソースに接続可能

<h3>オンプレミス</h3> <p>オンプレミスにセキュアに接続。S3やEXCELファイルのアップロードに対応</p>  <ul style="list-style-type: none"> • Excel • CSV • Teradata • MySQL • SQL Server • PostgreSQL • Oracle 	<h3>AWSクラウド</h3> <p>ConnectS3上のデータやVPC内のデータベースにセキュアに接続</p>   <ul style="list-style-type: none"> • Redshift • RDS • S3 • Athena • Aurora • Teradata • MySQL • Oracle • SQL Server • Timestream • Amazon Elasticsearch Service • Presto • Spark • PostgreSQL • MariaDB • Snowflake • IoT Analytics 	<h3>SaaS</h3> <p>3rdパーティーのSaaSに接続</p>      <ul style="list-style-type: none"> • Salesforce • Square • Adobe Analytics • Jira • ServiceNow • Twitter • Github
--	---	--

AWS 上のさまざまなデータベースだけでなく、オンプレミスのデータベースや、SaaS サービスをデータソースとして利用可能



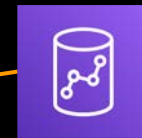
Amazon RDS



Amazon Athena



Amazon S3



Amazon Redshift



Salesforce



Onpremise Database

次の一歩

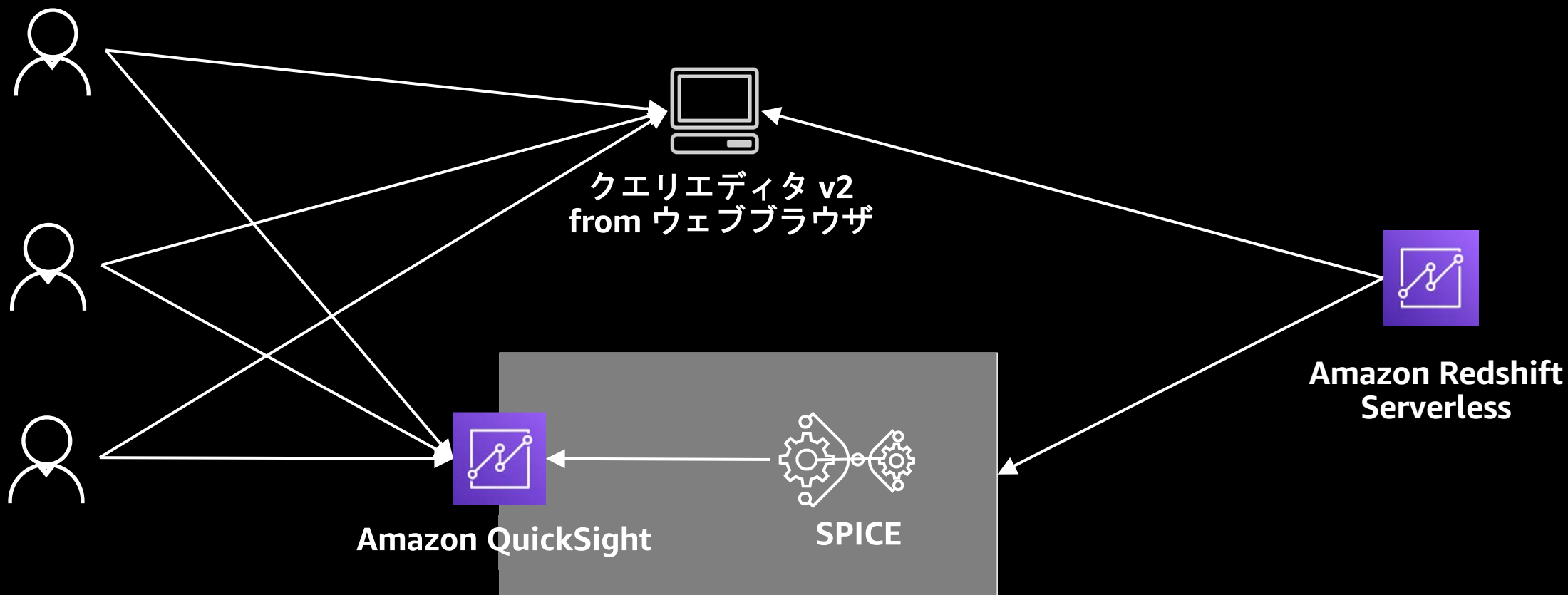


もっといろいろやりたくなったら？

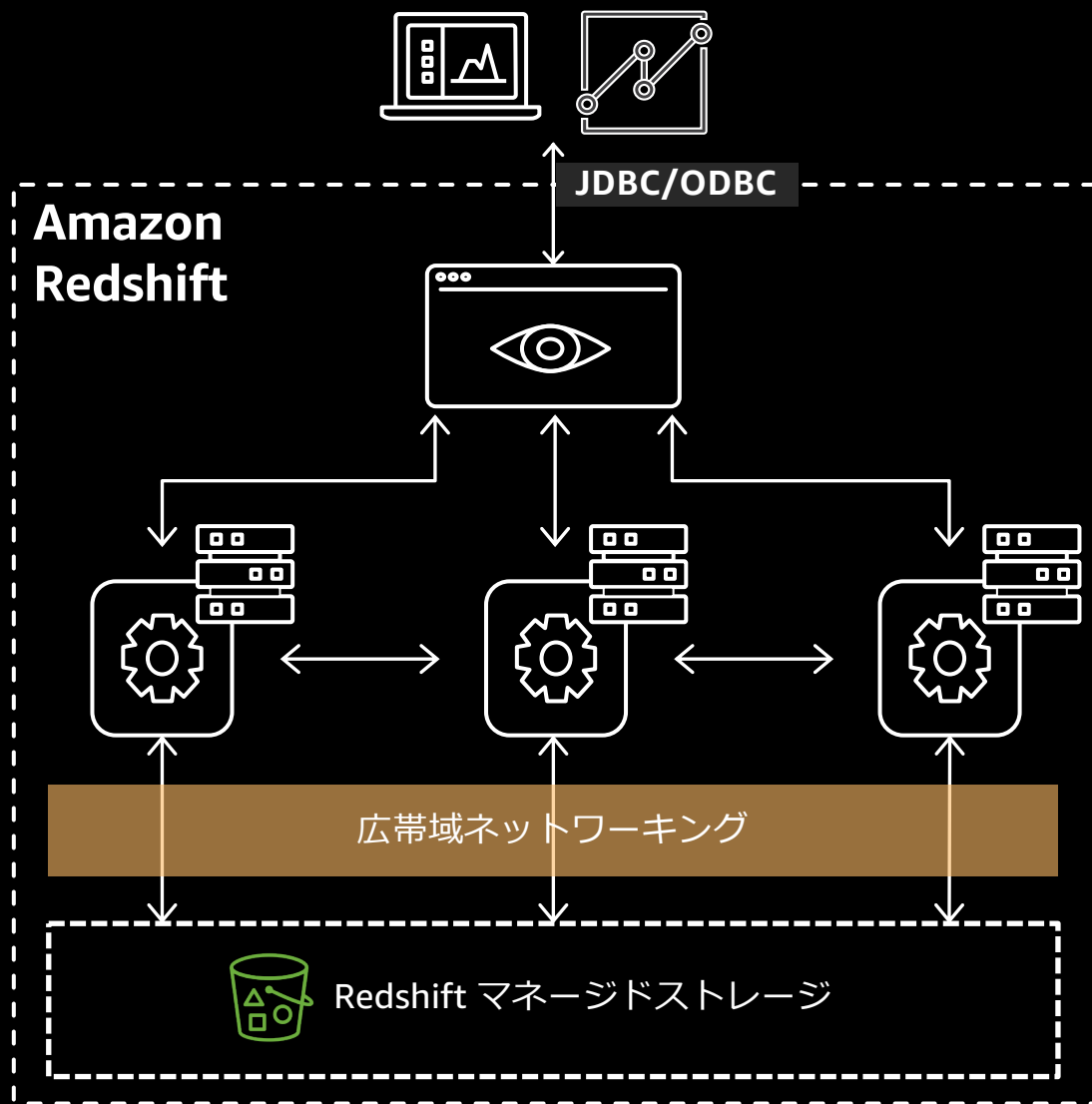
たとえば...

- 扱う対象のデータが大きすぎて、SPICE に入りきらない
- 可視化する前に、生データに対して加工集計を行っておきたい
- 可視化だけでなく、SQL を使って分析したい

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
- 数分でクラスターが立ち上がり、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**

ブラウザから Redshift Serverless にクエリ

The screenshot displays the Amazon Redshift Serverless console dashboard. The left sidebar contains navigation options: 'Serverless dashboard', 'Query editor', 'Serverless configuration', 'Monitoring' (with sub-items for query and database monitoring, resource monitoring, and datashares), 'Provisioned clusters dashboard', and 'Documentation'. The main content area is titled 'Serverless dashboard' and includes a 'Query data' button. It features a 'Resource summary' section showing 2 databases and 0 snapshots. Below this is a 'Query summary' section for the last hour, displaying '6 queries'. Two bar charts are shown: one for 'All completed and failed queries' and another for 'All running and queued queries'. The first chart shows 6 completed queries and 0 failed queries across six time intervals. The second chart shows 0 running and 0 queued queries across four time intervals. A 'View details' button is present for the query summary.

Resource summary
Viewing data from Asia Pacific (Tokyo) region.

Resource	Count
Databases	2
Snapshots	0

Query summary
Viewing data for the last 1 hour.

6 queries

Filter displayed data: All completed and failed queries

Queries

Time (UTC)	Completed	Failed
Dec 9 3:44 AM	7	0
Dec 9 3:54 AM	6	0
Dec 9 4:04 AM	6	0
Dec 9 4:14 AM	7	0
Dec 9 4:24 AM	7	0
Dec 9 4:34 AM	6	0

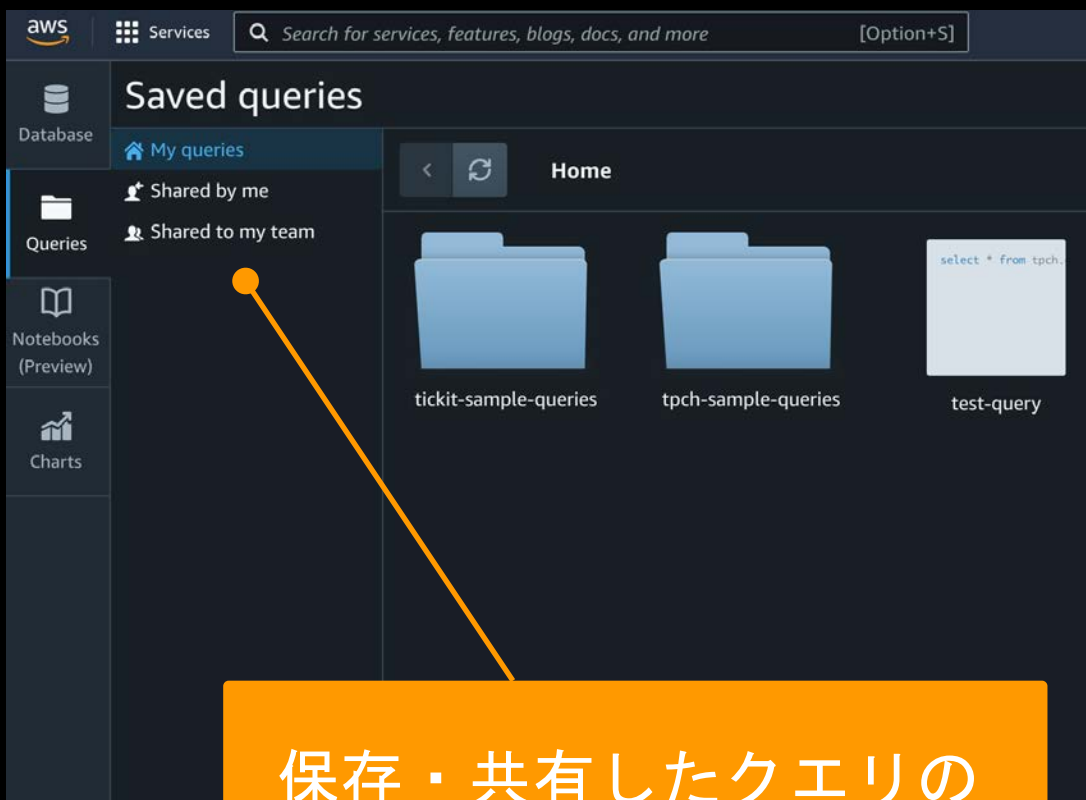
Filter displayed data: All running and queued queries

Queries

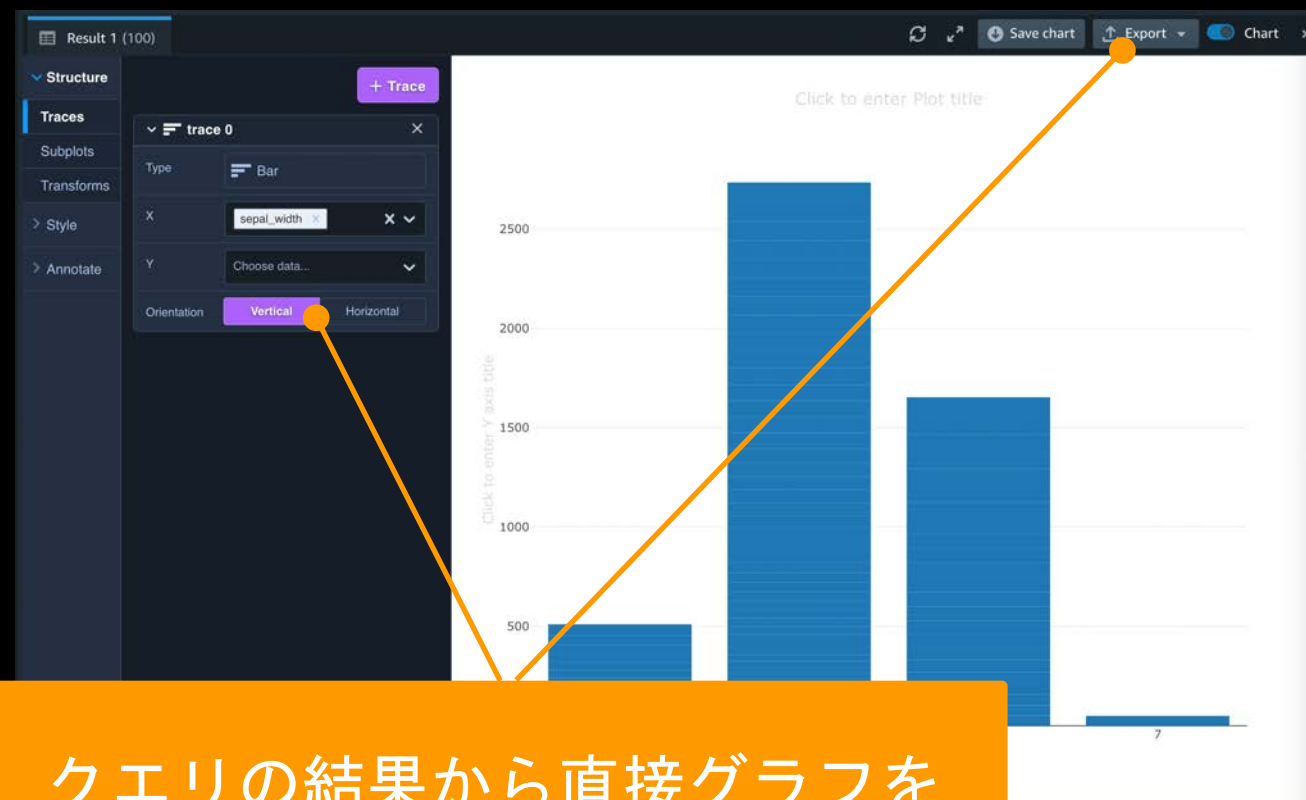
Time (UTC)	Running	Queued
Dec 9 3:45 AM	0	0
Dec 9 4:00 AM	0	0
Dec 9 4:15 AM	0	0
Dec 9 4:30 AM	0	0

RPU capacity used

結果の可視化やクエリの共有



保存・共有したクエリの一覧を確認



クエリの結果から直接グラフを作成. 画像として出力も可能

Redshift の外にあるデータにも直接クエリ

The screenshot displays the AWS Redshift console interface. On the left, the 'Create schema' dialog is open, showing the 'External' schema type selected. The 'PostgreSQL' source is chosen in the Glue Data Catalog, with 'dev' as the source database name and 'public' as the source schema name. The main console area shows the 'Redshift query editor v2' for the 'sales_db' database. The 'store_sales_ods' table is selected in the resource tree. The query editor contains a SQL query that joins data from 'ext_postgres_ods.inventory_ods' with local Redshift data. The query results are displayed in a table below the editor.

```
6
7 -- QUERY POSTGRES ODS DATA
8 SELECT *
9 FROM ext_postgres_ods.inventory_ods;
10
11 -- COMBINE POSTGRES ODS DATA with LOCAL, WH DATA in REDSHIFT SERVERLESS
12 SELECT d_year as year,
13        ('Q' || d_qoy) as quarter,
14        i_category as item_category,
15        count(*) as orders
16 FROM
17 ext_postgres_ods.store_sales_ods s, date_dim d, item i
18 WHERE s.ss_sold_date_sk = d.d_date_sk
19       and s.ss_item_sk = i.i_item_sk
20       and s.ss_transaction_dt > current_date -1
21 GROUP BY 1,2,3
22 ORDER BY 4 desc;
```

year	quarter	item_category	orders
2021	Q3	Sports	26
2021	Q3	Home	25
2021	Q3	Men	25
2021	Q3	Women	22
2021	Q3	Electronics	21
2021	Q3	Children	20
2021	Q3	Jewelry	20
2021	Q3	Music	19

PostgreSQL, MySQL, S3 を外部スキーマとして登録

スキーマをエディタ上で参照. クエリも実行できる

QuickSight から Redshift Serverless にクエリ

The screenshot displays the Amazon QuickSight interface in Japanese. On the left is a navigation sidebar with options: 'お気に入り' (Favorites), '最新' (Latest), 'マイフォルダ' (My Folders), '共有フォルダ' (Shared Folders), 'ダッシュボード' (Dashboards), '分析' (Analysis), and 'データセット' (Data Sets). The '分析' (Analysis) option is selected. The main area is titled '分析' (Analysis) and contains three analysis cards:

- lineitem analysis**: Updated 1 day ago. The chart area shows a value of 1,500,000.
- 販売データ analysis**: Updated 2 years ago. The chart area shows a line graph.
- sales analysis**: Updated 3 years ago. The chart area shows a bar chart.

At the top right of the main area, there is a dropdown menu for '最終更新日 (新しい順)' (Last updated (newest first)) and a button for '新しい分析' (New Analysis).

Redshift Serverless のなにがいいか？

- サーバーレスのため、クラスターの管理や運用、チューニングといった作業から完全に解放される。そのため**誰でも簡単に使い始められる**。独立したクエリエディタ経由でブラウザからアクセスし、すぐにクエリを投げられる。GUIベースでのテーブル作成、データロードもサポート
- **利用量が事前に予測できない、変動が大きいワークロードで高いコストパフォーマンスを発揮**。重たいクエリが大量に実行される場合、自動でスケールして対応してくれる一方、利用しない間は課金もされない
- **既存の Redshift の優れた機能がそのまま使える**。JSON データを扱える SUPER 型、Materialized View、Spectrum、Federated Query、強固なセキュリティ、データシェアリングなど

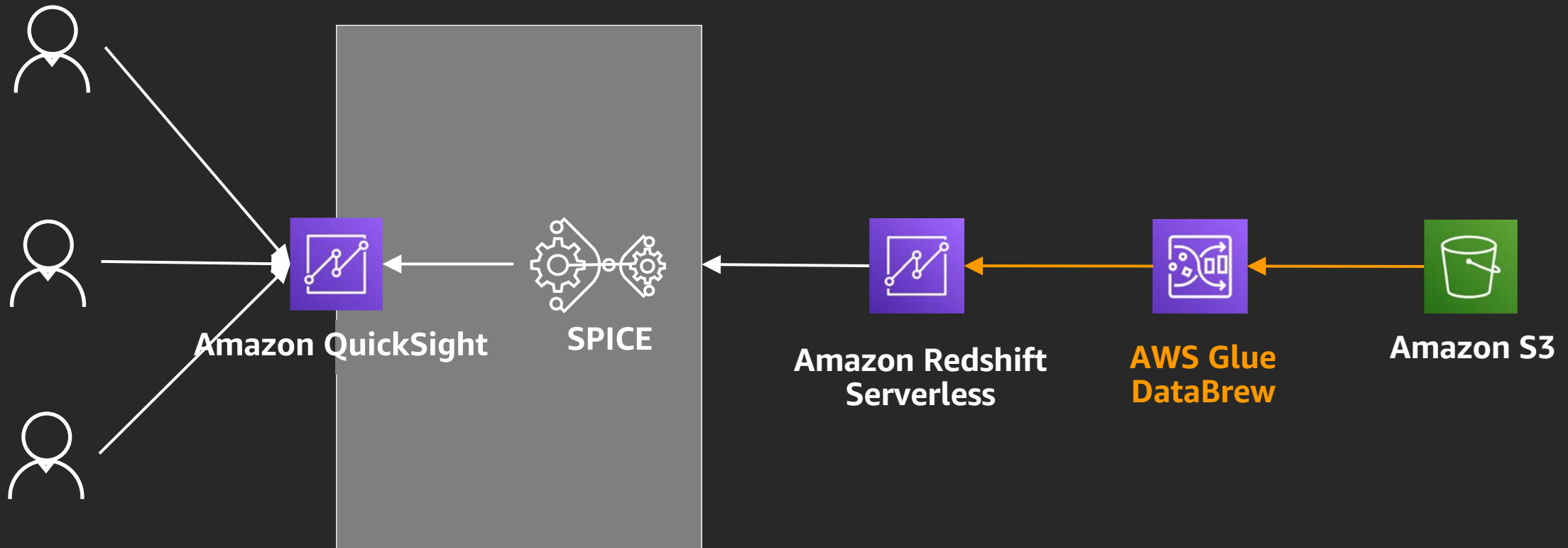
おまけ: さらにその次の一歩

そのままクエリできない汚いデータがある場合は？

たとえば...

- カラムの中身に大文字と小文字が混在してたり，スペルミスの値が含まれているため，中身を確認して正しい値にしたい
- 特定カラムの分布におかしな偏りがないか確認してから取り込みたい
- SQL でやるのはしんどい複雑な変換処理を手軽にやりたい

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



まとめ

まとめ

- AWS の分析サービスを活用することで、**誰でも簡単に**データ分析をスタートすることができる
- 可視化、クエリの実行、そして前処理を、ビジュアルベースでインタラクティブに行える。サーバーレスなので、インフラのことを考える必要がない
- 興味を持たれた方は、ぜひ実際に自分の手でさわって試してみてください！

実際に試してみるためのリソース

Amazon QuickSight

- ベーシック / 販売管理 / 埋め込みハンズオン: <https://aws.amazon.com/jp/blogs/news/amazon-quicksight-handson-202006/>
- Amazon DynamoDB データの可視化ハンズオン: <https://aws.amazon.com/jp/blogs/news/athena-federated-query-dynamodb-quicksight/>
- 小売データのダッシュボード & インサイト抽出: <https://aws.amazon.com/jp/blogs/news/quicksight-dashboard-analysis-retail/>

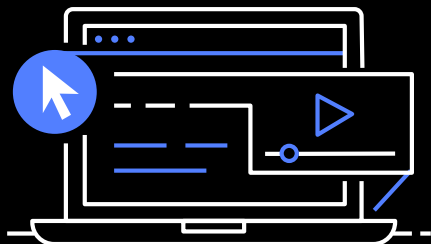
Amazon Redshift Serverless

- 紹介ブログ: <https://aws.amazon.com/blogs/aws/introducing-amazon-redshift-serverless-run-analytics-at-any-scale-without-having-to-manage-infrastructure/>
- デモ動画: <https://www.youtube.com/watch?v=1jSD1UKciYo>

AWS Glue DataBrew

- ハンズオン: <https://aws-dataengineering-day.workshop.aws/1300.html>
- デモ動画: <https://www.youtube.com/watch?v=G8o5ekfbBO4>
- 紹介資料: https://pages.awscloud.com/rs/112-TZM-766/images/20210325-02_Data-Analytics-DataBrew.pdf

AWS デジタルトレーニング



学習方法を柔軟に選ぶことができます。

550以上の無料デジタルトレーニングでクラウドのスキルを身につけることができ、クラスルームトレーニングでさらに深く学ぶこともできます。

おすすめのコース

- **[AWS Cloud Practitioner Essentials \(Japanese\)](#)**

このコースは、特定の技術領域ではなく、アマゾン ウェブ サービス (AWS) クラウドを全体的に理解したい方を対象としています。受講者は、AWS クラウドの概念、AWS のサービス、セキュリティ、アーキテクチャ、料金、サポートについて学習し、AWS クラウドについての知識を深めます。このコースは、AWS Certified Cloud Practitioner 試験の準備にも役立ちます。

- **[Getting Started with AWS Security, Identity, and Compliance \(Japanese\)](#)**

このコースでは、AWS のセキュリティテクノロジー、ユースケース、メリット、サービスの概要について説明します。AWS のセキュリティ、アイデンティティ、コンプライアンスのカテゴリにおけるさまざまなサービスを紹介します。このコースを修了すると、クラウド内のセキュリティの重要性を理解し、データの保護に使用する AWS のサービスを選択できるようになります。

- **[Moving to Managed Databases on AWS \(Japanese\)](#)**

リフトアンドシフト戦略は、クラウドを利用しようとする組織が一般的に採用する方法です。このコースでは、AWS ヒーローである Alex DeBrie が、アマゾン ウェブ サービス (AWS) においてセルフマネージドデータベースからフルマネージドデータベースソリューションに移行することのメリットについて説明します。また、AWS クラウドに移行する際に考慮すべき重要な要素についても学習します。最後に、この Moving to Managed Databases on AWS の 8 つのコースのカリキュラムの一環として、一般的なデータベースの移行に関するユースケースのコ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!