



aws INNOVATE

DATA EDITION

23 August, 2022

Data virtualization and unifying disparate data sources with Amazon Athena

Indrajit Ghosalkar

Senior Solutions Architect

Amazon Web Services



Agenda

- Overview of data virtualization
- Data virtualization architecture
- Amazon Athena Federated Query
- **Demo** - To build a unified view of your disparate data sources
- Key takeaways

Overview of data virtualization



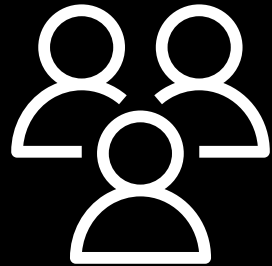
You

(Enterprise Data Architect / Consultant)



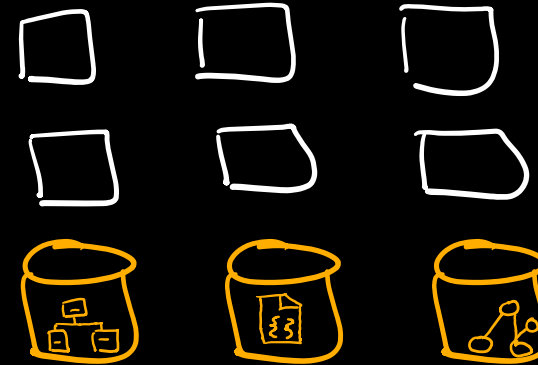
**OCTANK
INTERNATIONAL**

Overview of data virtualization



IT Teams @
Octank International

Microservices



Fit-for-purpose Databases



Data integration challenges



You

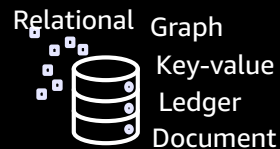
(Enterprise Data Architect / Consultant)

" Hi, we have a lot of different data sources, databases, complex and slow ETL pipelines, multiple data integration tools.

All this is very resource, time and cost intensive.. What to do ? "



CTO / Data Teams@
Octank International



Modern data
landscape diversity



Time to value



Operations



Data
governance



Data
life cycling

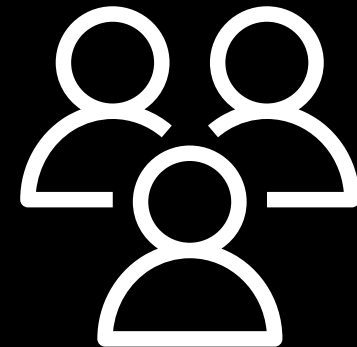
Overview of data virtualization



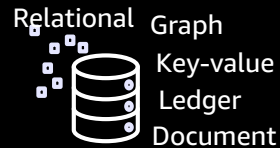
You

(Enterprise Data Architect / Consultant)

"Data virtualization"



CTO / Data Teams@
Octank International



Modern data
landscape diversity



Time to value



Operations

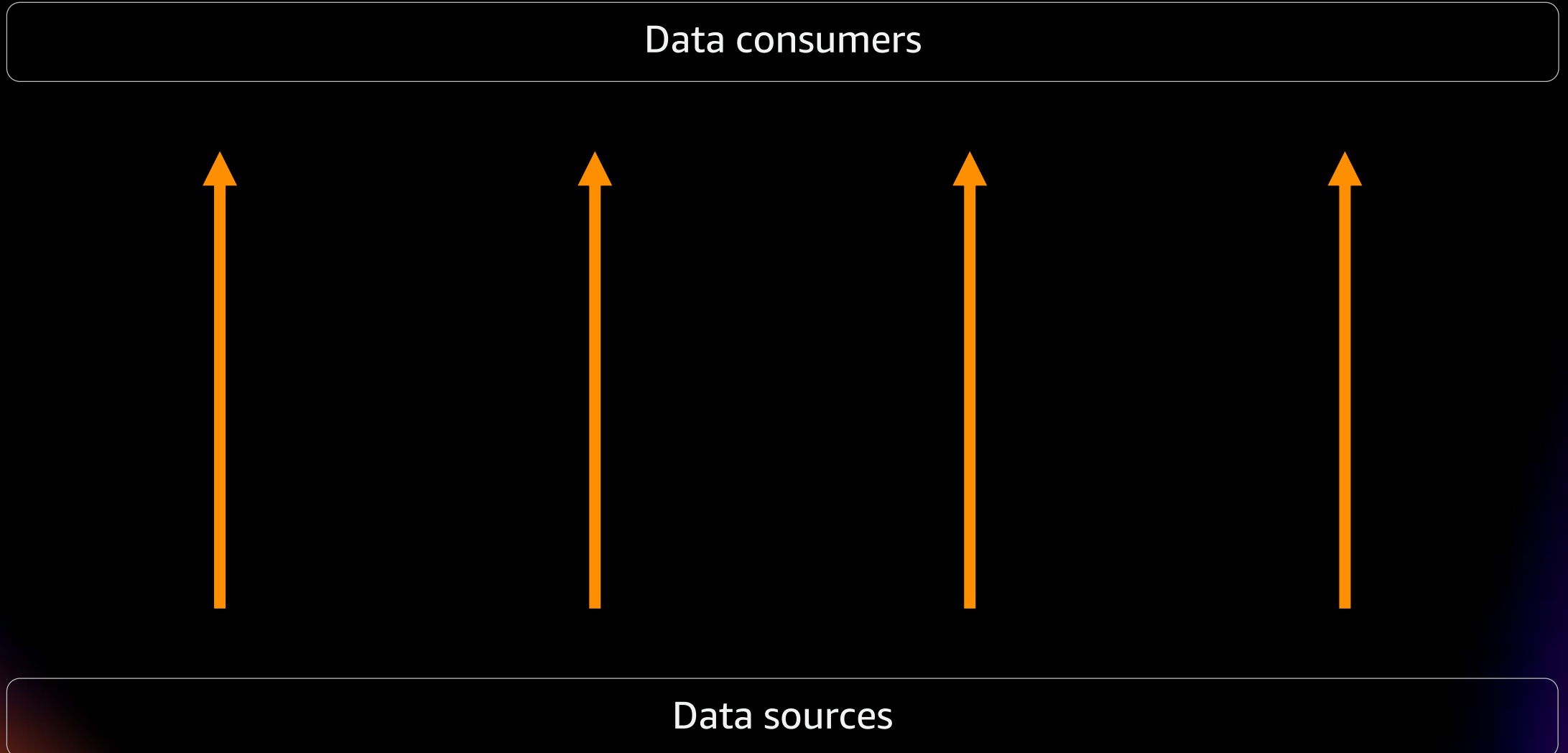


Data
governance



Data
life cycling

Data virtualization architecture



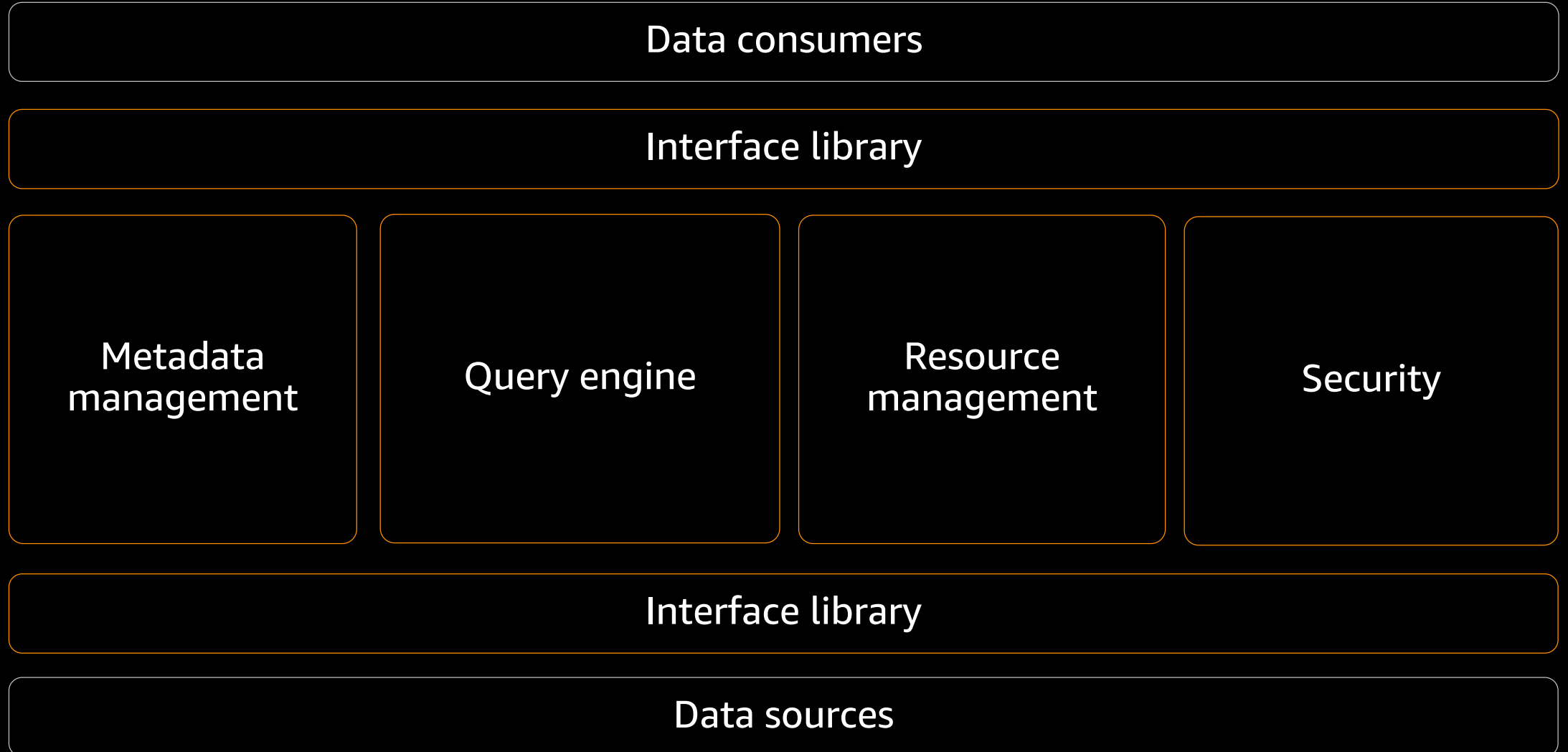
Data virtualization architecture

Data consumers

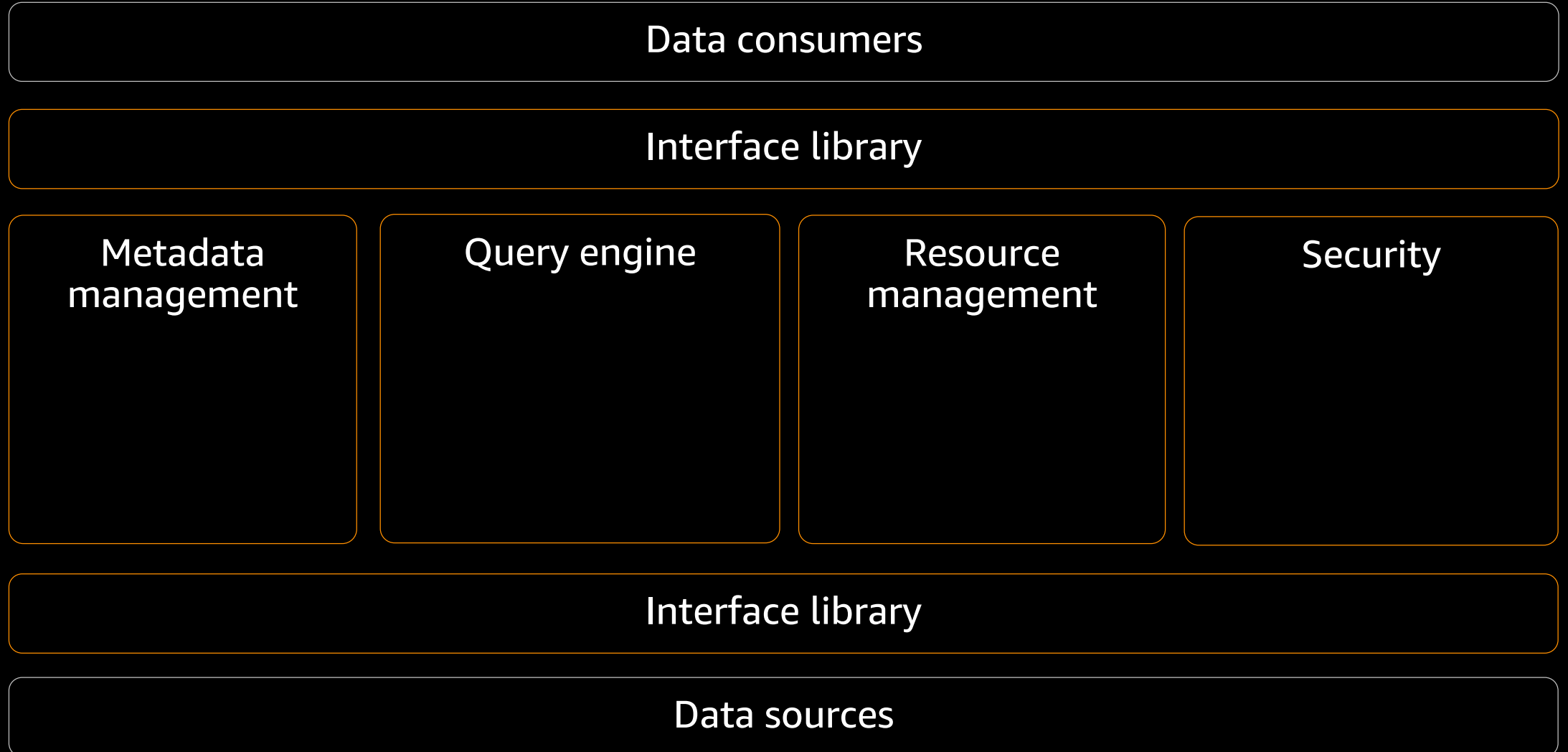
Data virtualisation

Data sources

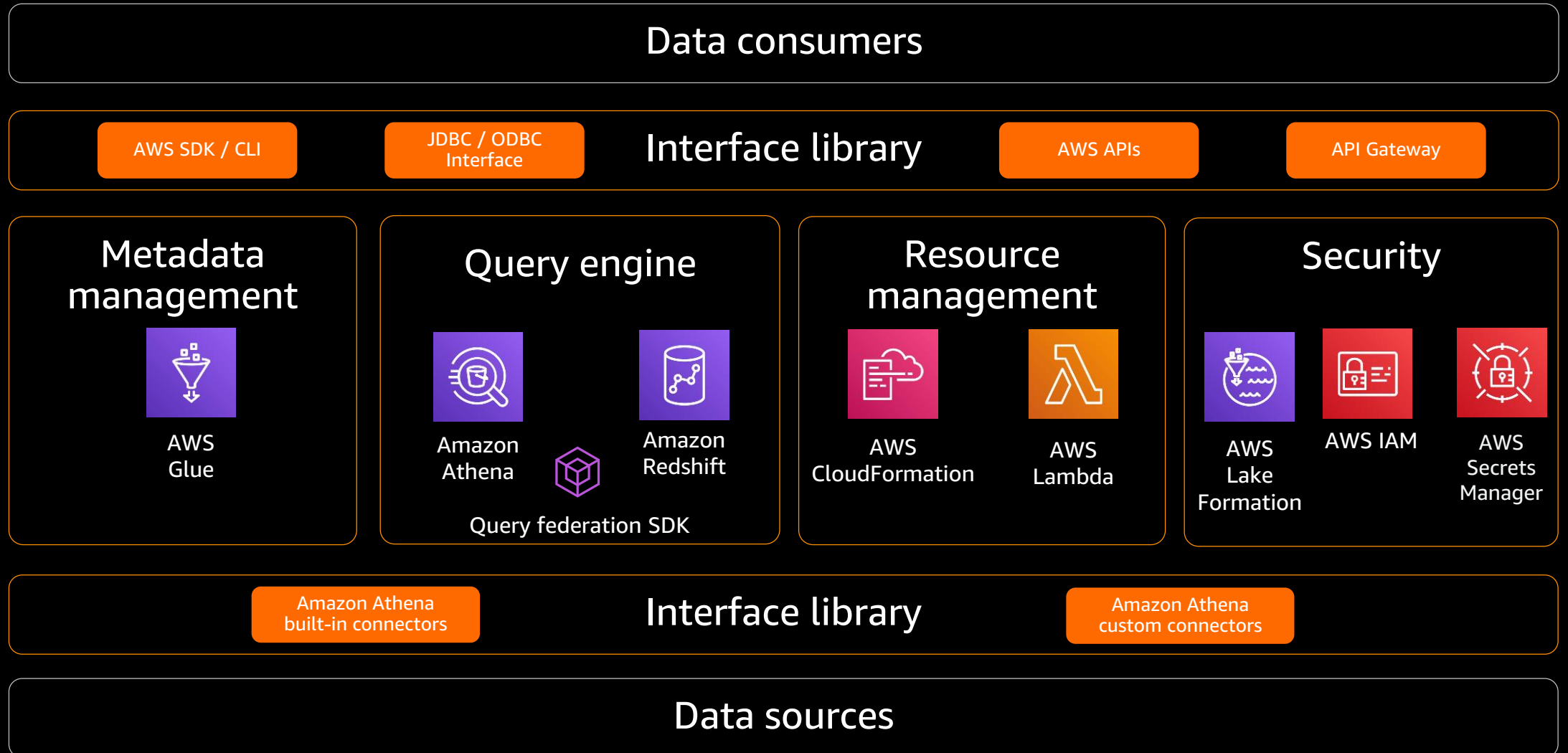
Data virtualization architecture



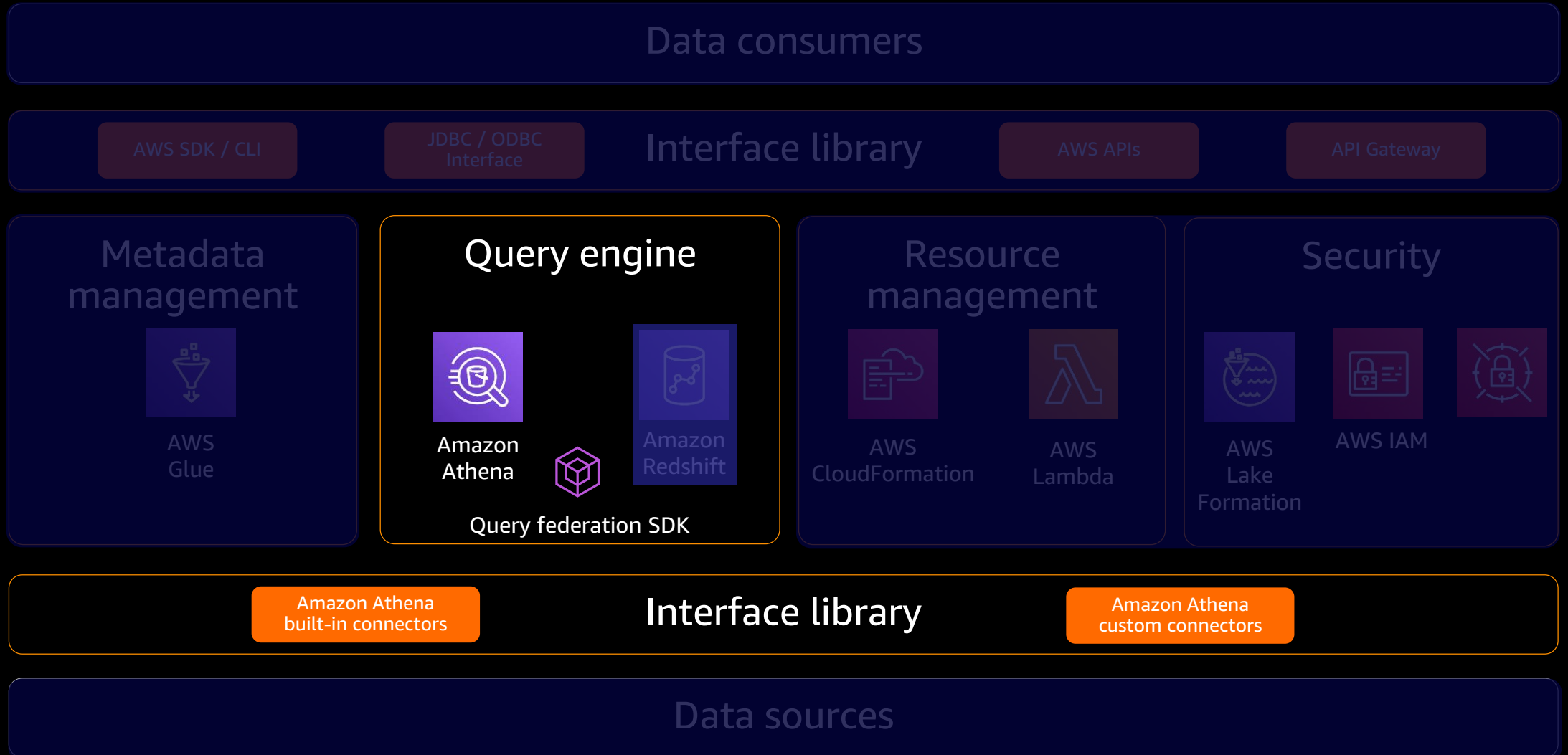
Data virtualization architecture



Data virtualization architecture

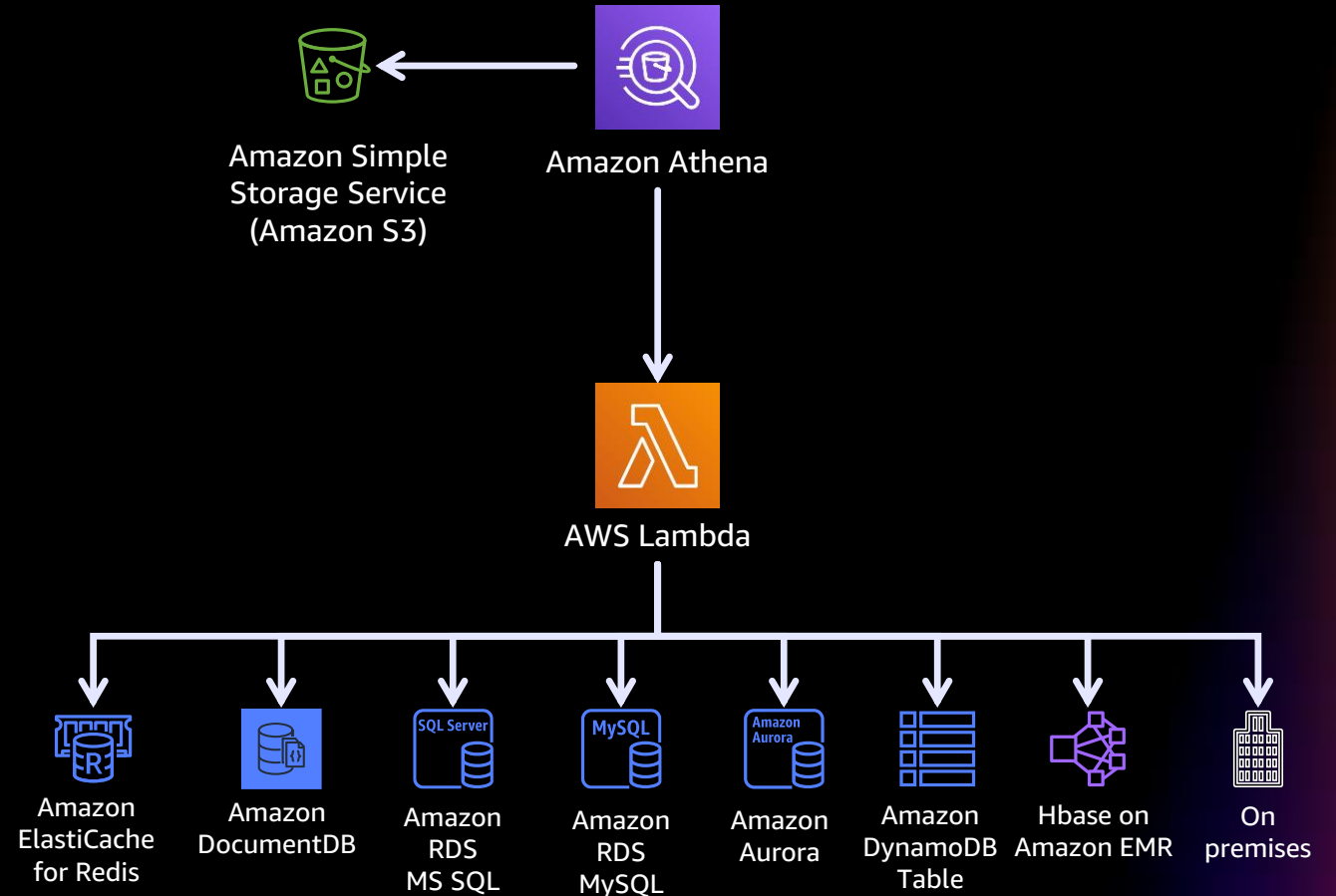


Data virtualization architecture



What is a federated query?

- Serverless query execution
- Run query across relational, non-relational, object, or custom data sources
- Run query across on-premises or cloud data sources



Federated query is simple to use

1

Deploy data
source connector

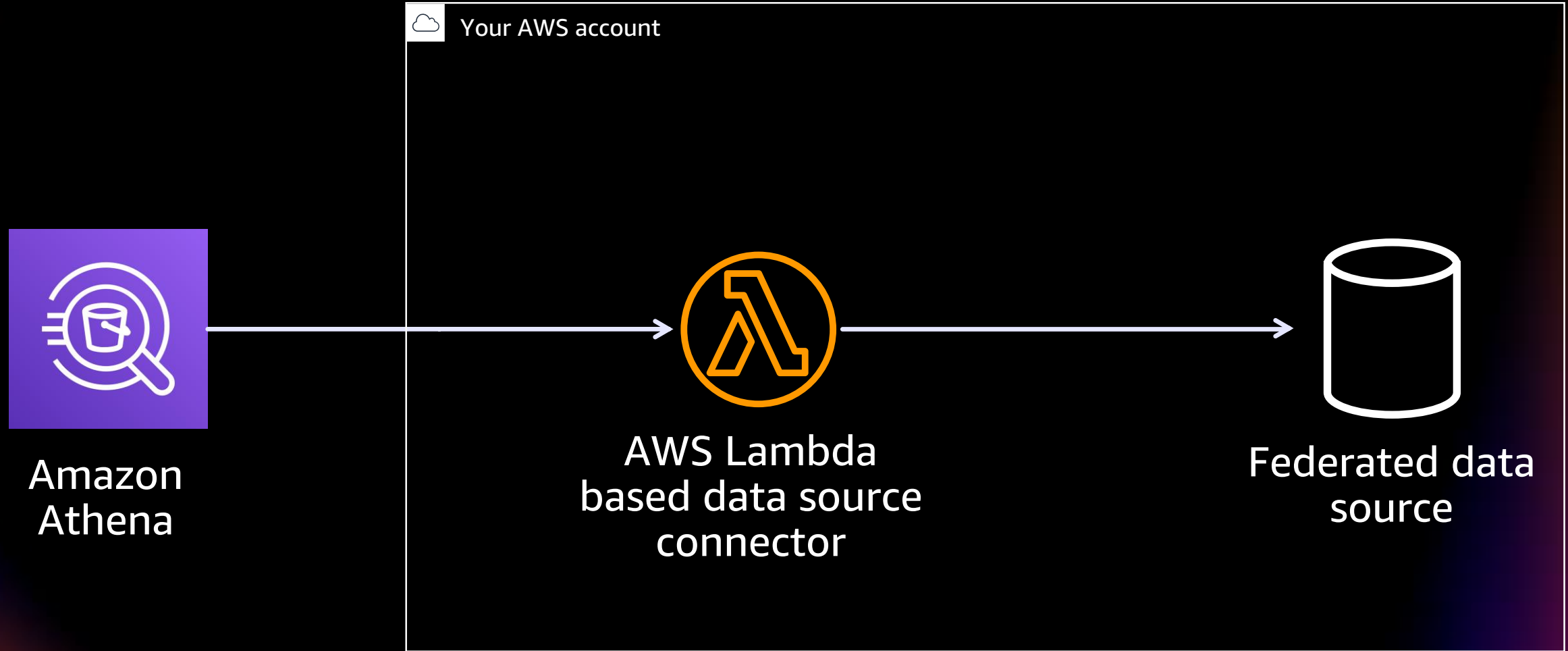
2

Register data source
connector
Specify a catalogue
name

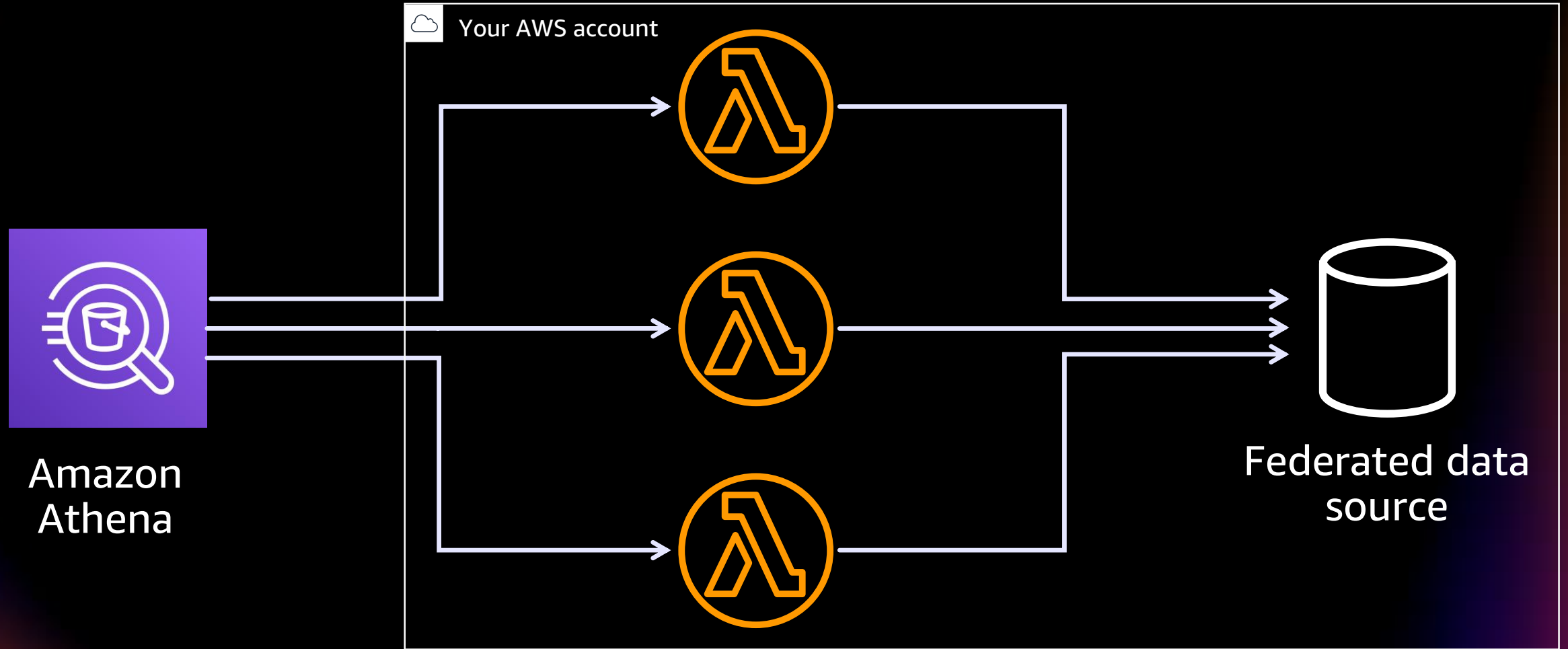
3

Write SQL
Query

Anatomy of a federated query



Anatomy of a federated query

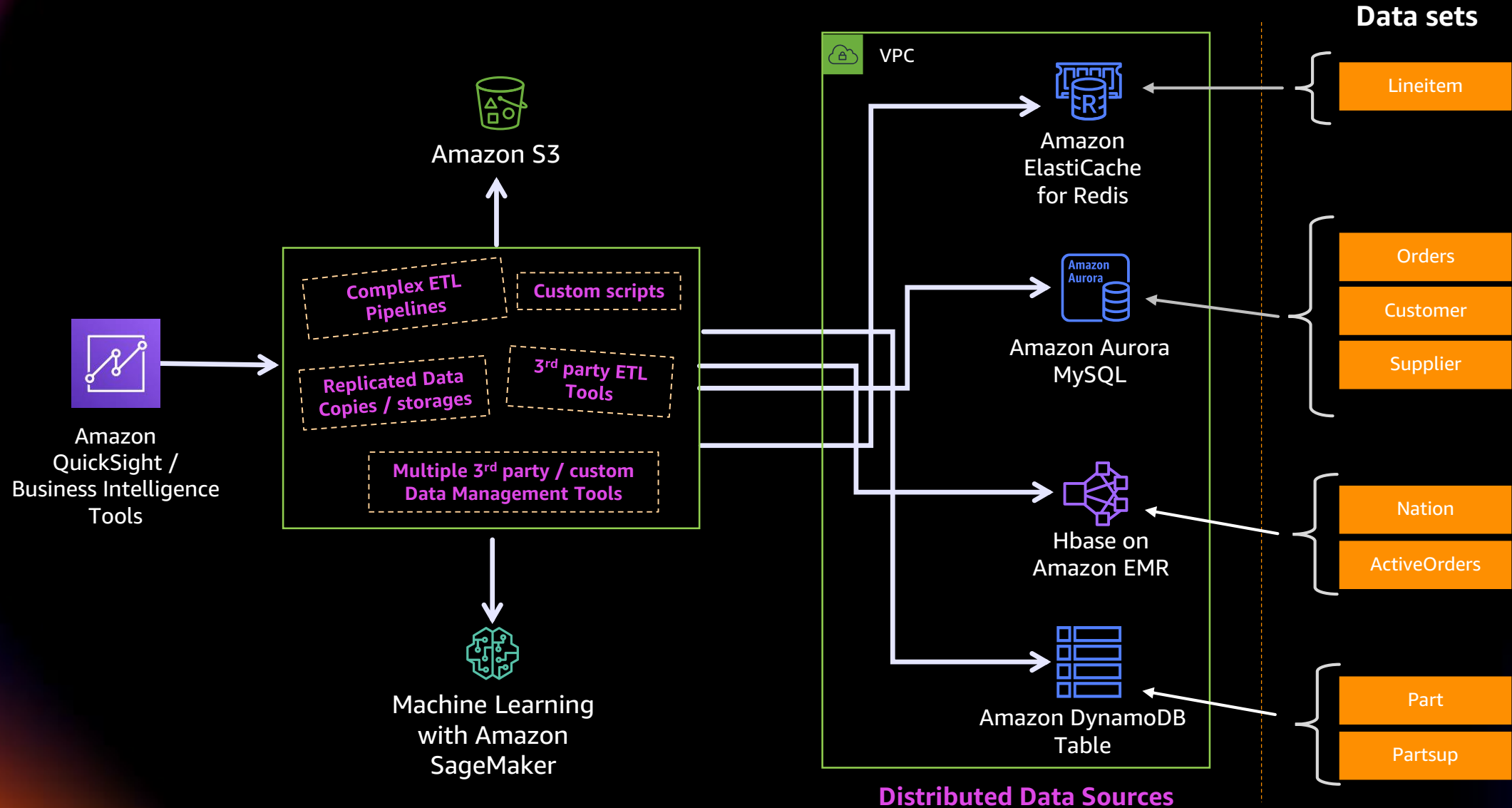


Let's get into the demo

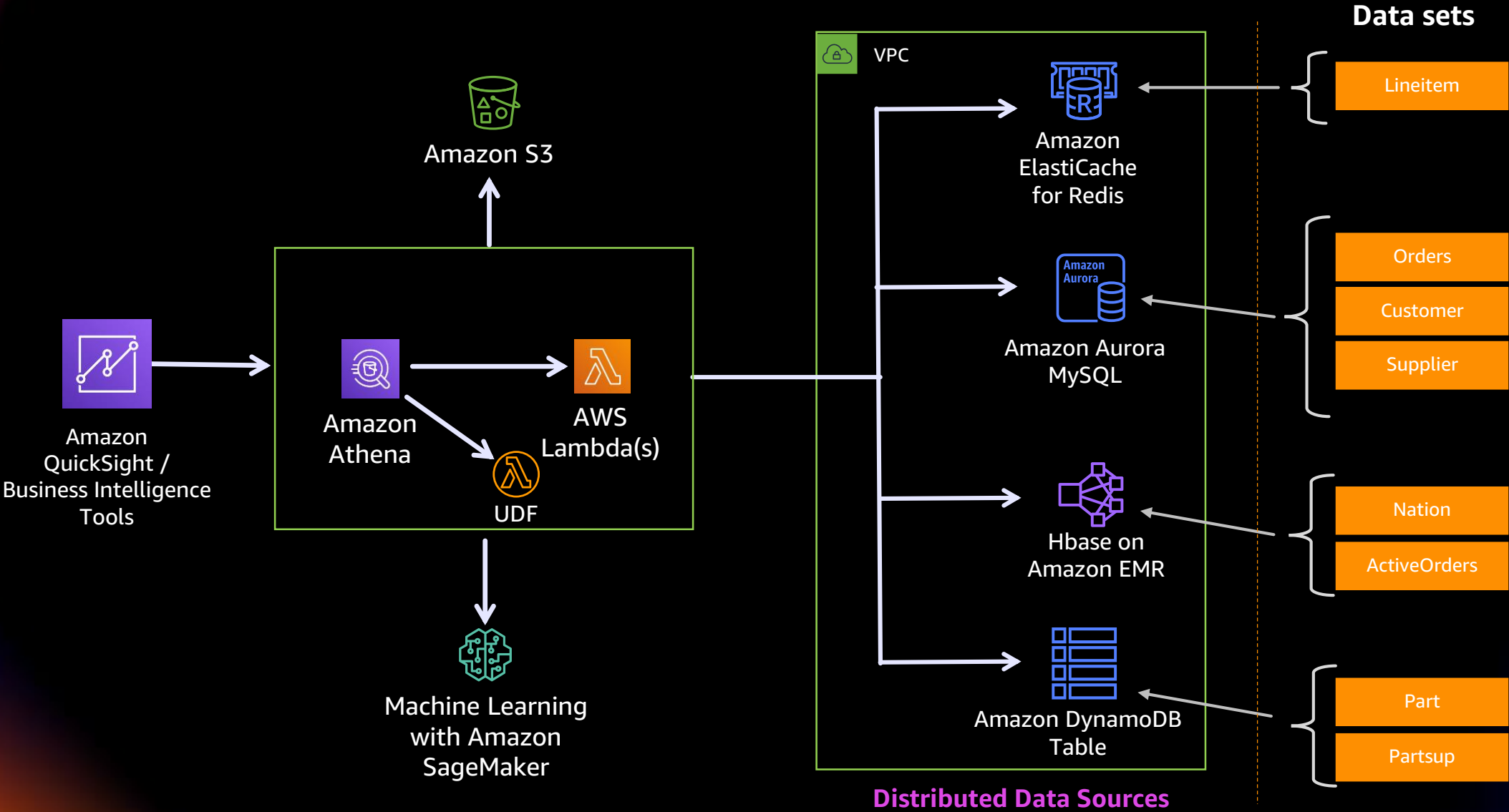
What you need to know for demo

- Basics of **Amazon Athena** and **AWS Lambda**
- Some knowledge of **SQL** and writing **SQL queries**
- High-level knowledge of **Data visualization** or business Intelligence dashboards.

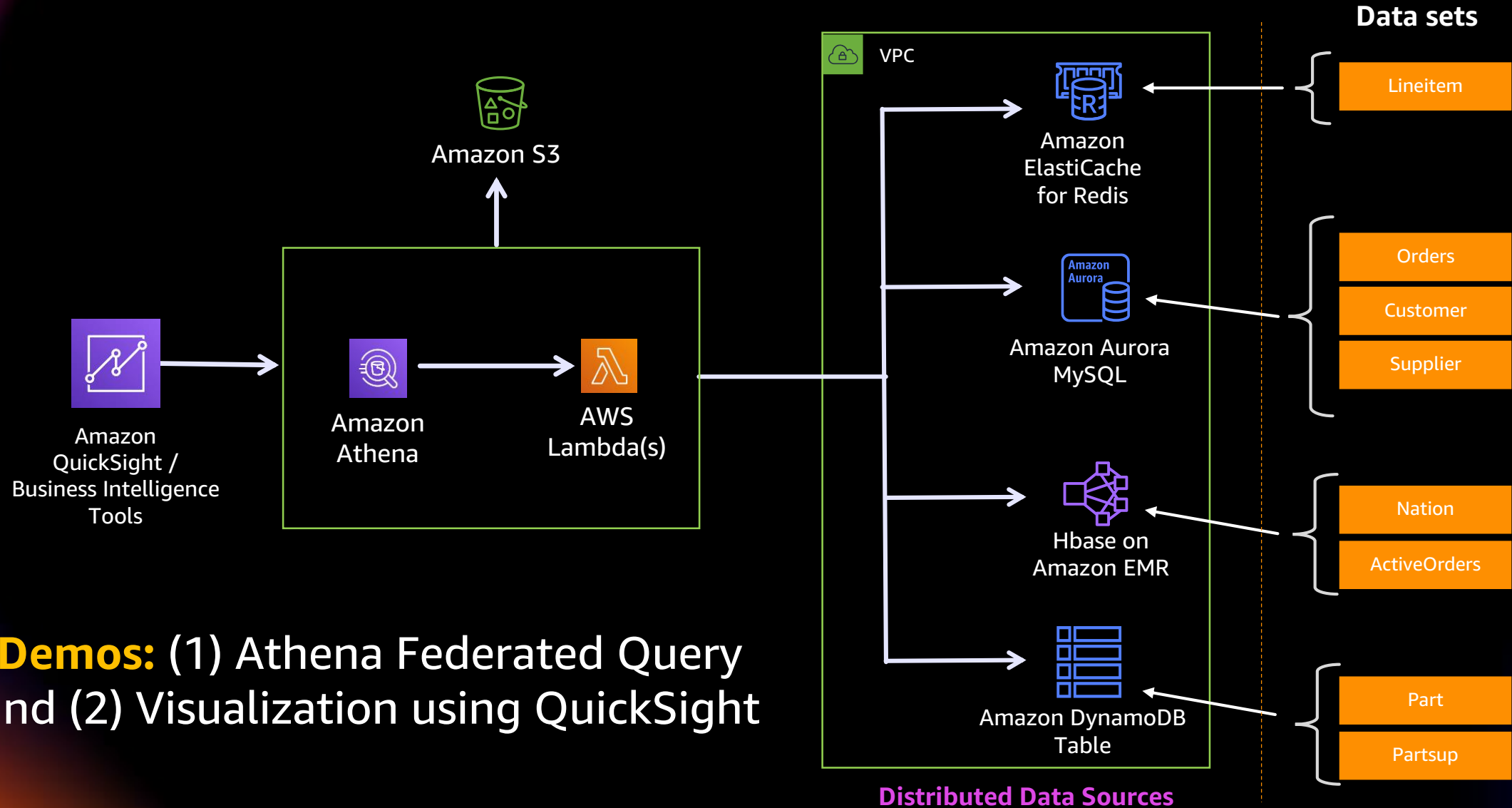
Octank International – **Current** data architecture



Octank International – **Future** data architecture

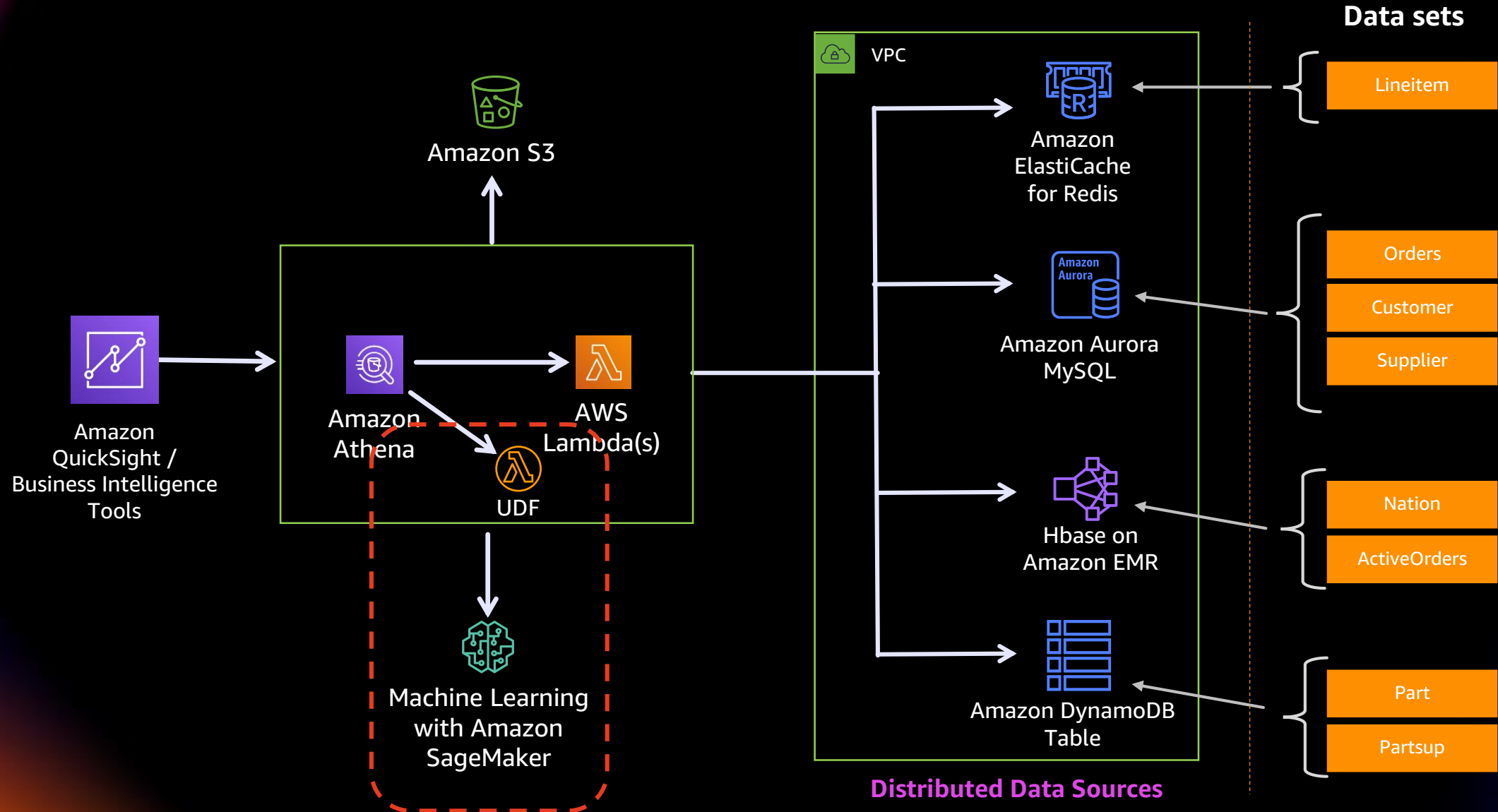


Octank International – Demo data architecture



Demos: (1) Athena Federated Query and (2) Visualization using QuickSight

Amazon Athena features – UDFs and ML Integration



Demo 1 – Amazon Athena Federated Queries

Data source connectors available today

AVAILABLE VIA AWS SERVERLESS APPLICATION REPOSITORY AND GITHUB



JDBC compliant
data source



Amazon
Neptune



Amazon
DocumentDB



Amazon
Timestream



Amazon
DynamoDB



AWS CMDB
(resource
inventory)



Redis on Amazon
ElastiCache



Amazon
CloudWatch



Amazon
CloudWatch
Metrics



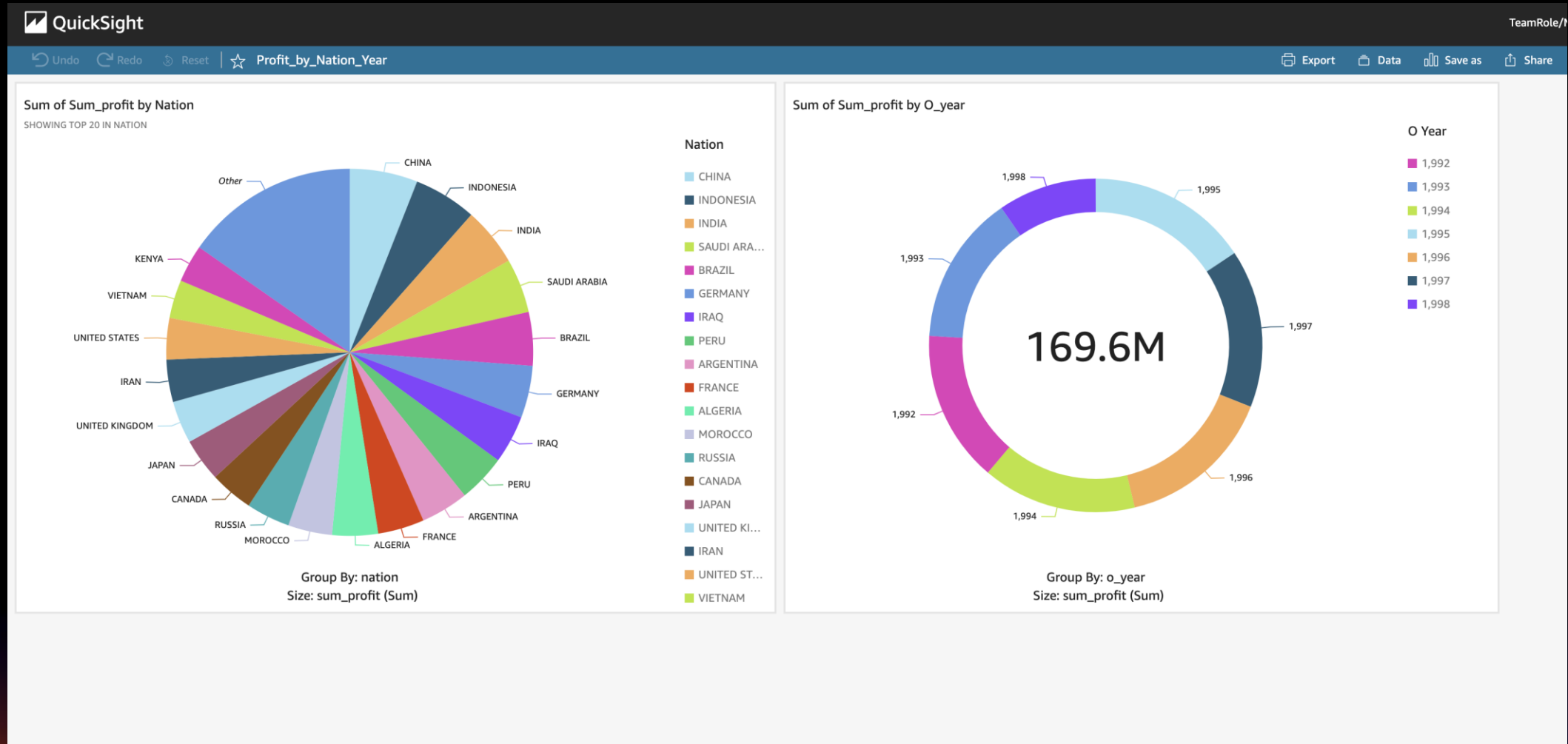
Amazon OpenSearch
Service



Amazon EMR

Demo 2 – Unified view of disparate data sources

Demo - Unified view of disparate data sources



Amazon Athena Features

User Defined Functions (UDFs) and Integrated Machine Learning (ML)

Expanding SQL syntax with Amazon Athena User Defined Functions (UDF)



Write once



Deploy once



Invoke as many times
as needed in a query

Integrate Machine Learning with Amazon Athena



Federated Athena query to select data from any data source



Transform data using UDFs in Amazon Athena



Train and deploy model on Amazon SageMaker



Write UDF to pre or post process data



Run inference on data from any data source

Recap – What have we learned?

- The need for data virtualization
- What are Amazon Athena Federated Queries
- How to get started on integrating Amazon Athena with your data sources and build unified view on Amazon QuickSight
- What are Amazon Athena User Defined Functions (UDFs) and support for Machine Learning (ML) integration in Amazon Athena

Learn



[Cloud-native data
virtualisation
whitepaper](#)

Build



[Amazon Athena
Query Federation
on GitHub](#)

Deploy



AWS Serverless
Application Repository

[AWS Serverless
Application
Repository](#)

Visit the AWS Data resource hub

A modern data strategy can help you manage, act on, and react to your data so you can make better decisions, respond faster, and uncover new opportunities. Dive deeper with these resources today.

- Harness data to reinvent your organization
- In unpredictable times, a data strategy is key
- Make data a strategic asset
- Rewiring your culture to be data-driven
- Put your data to work with a modern analytics approach
- ... and more!



<https://tinyurl.com/data-hub-aws>

[Visit resource hub](#)

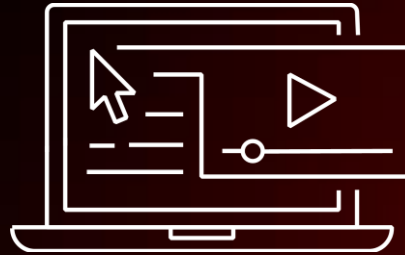
AWS Training and Certification for Data and Analytics



AWS Data & Analytics FREE Training Resources

Discover how to harness data, one of the world's most valuable resources, and innovate at scale.

<https://bit.ly/3Ntlhy7>



AWS Data Analytics Learning Plan

This learning plan expose you to the fastest way to get answers from all your data to all your users. It can also help prepare you for the AWS Certified Data Analytics - Specialty certification exam.

<https://bit.ly/3wBVjD1>



AWS Certified Data Analytics - Specialty

Earning AWS Certified Data Analytics – Specialty validates expertise in using AWS data lakes and analytics services.

<https://go.aws/3lwF0RR>

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey**.
Let us know what you thought of today's event and how we can improve the event experience for you in the future.



aws-apj-marketing@amazon.com



twitter.com/AWSCloud



facebook.com/AmazonWebServices



youtube.com/user/AmazonWebServices



slideshare.net/AmazonWebServices



twitch.tv/aws

Thank you!