
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT327

Modernize your analytics architecture with Amazon Athena

Daniel Chen

Senior Product Manager, Athena
Amazon Web Services

Saurabh Bhutyani

Principal Analytics Specialist Solutions Architect
Amazon Web Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What to expect

01

UNDERSTAND

Amazon Athena and data source connectors

02

LEARN

Modern analytics architecture pattern

03

APPLY

How to implement with Athena?

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Common challenges we hear from customers

"It's difficult and expensive to create data product with administrative overhead"

"Sharing data is difficult"

"I wish to focus on innovating with data, not on maintaining and administering a data lake"

"My data sits across siloed data sources, want to query them from one place"

"Current data architecture is complex and monolithic"

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

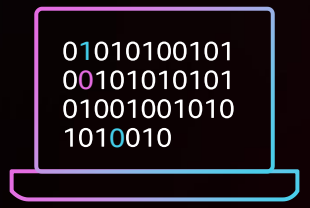
Companies expect more value from their data



Growing exponentially



From new sources



Increasingly diverse formats



Used by many internal roles

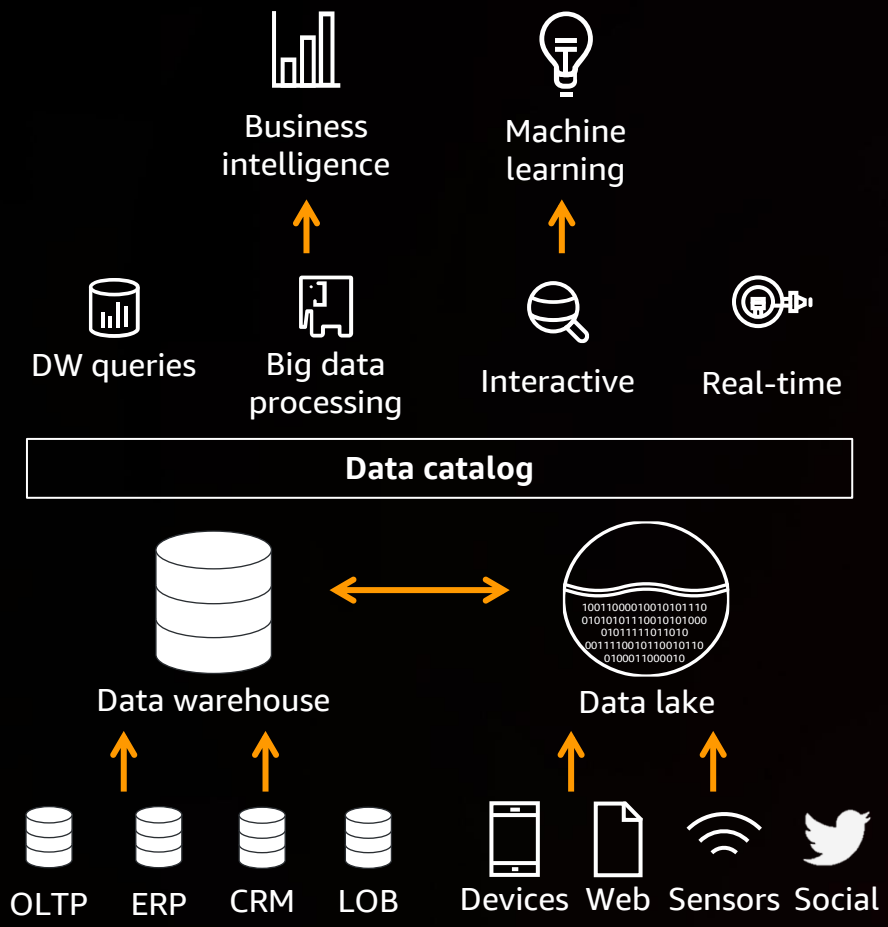


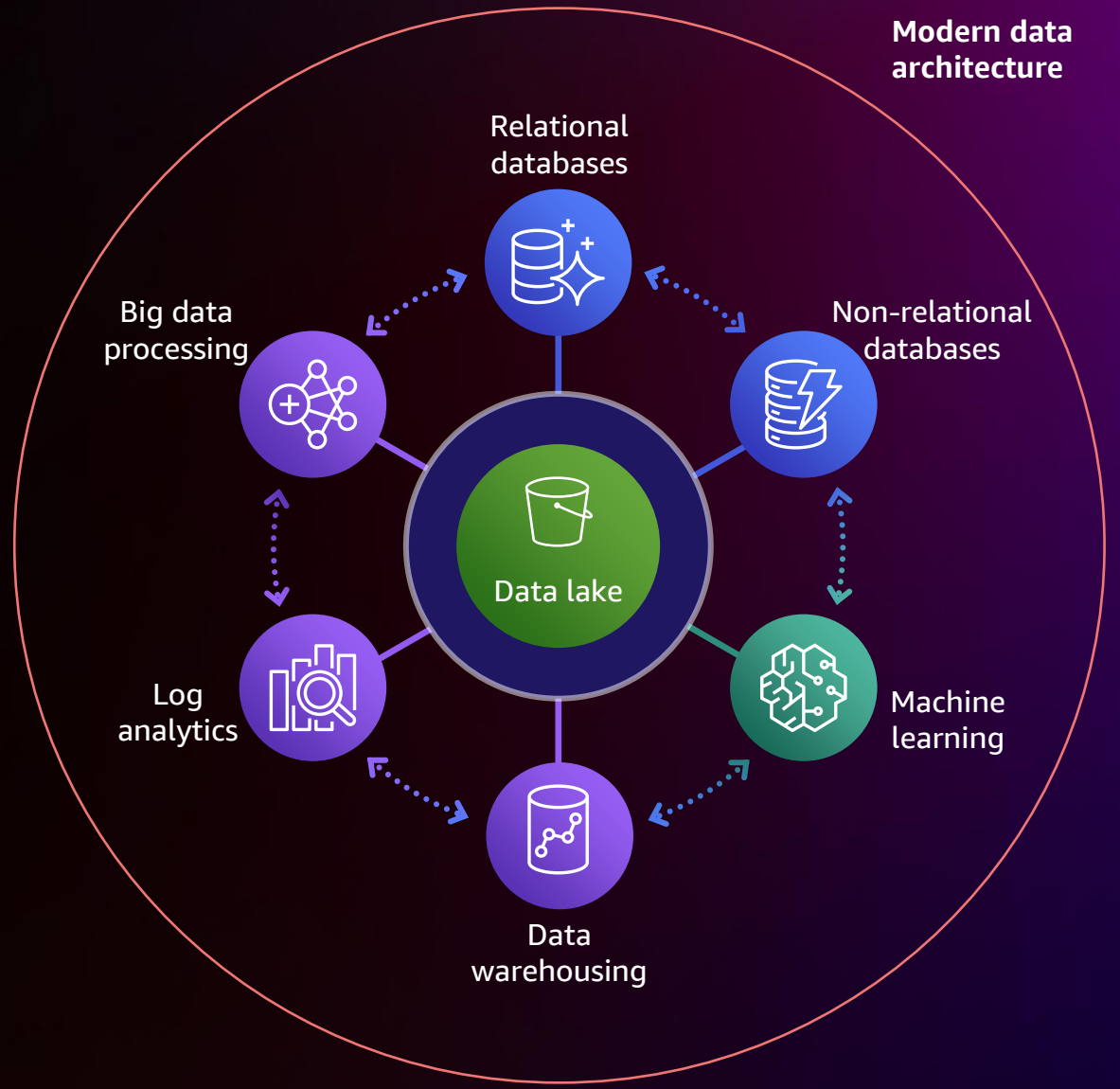
Analytics and machine learning

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data lakes and modern data architecture





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Athena

Interactive analytics for all your data



Easy to get started



Interactive analytics

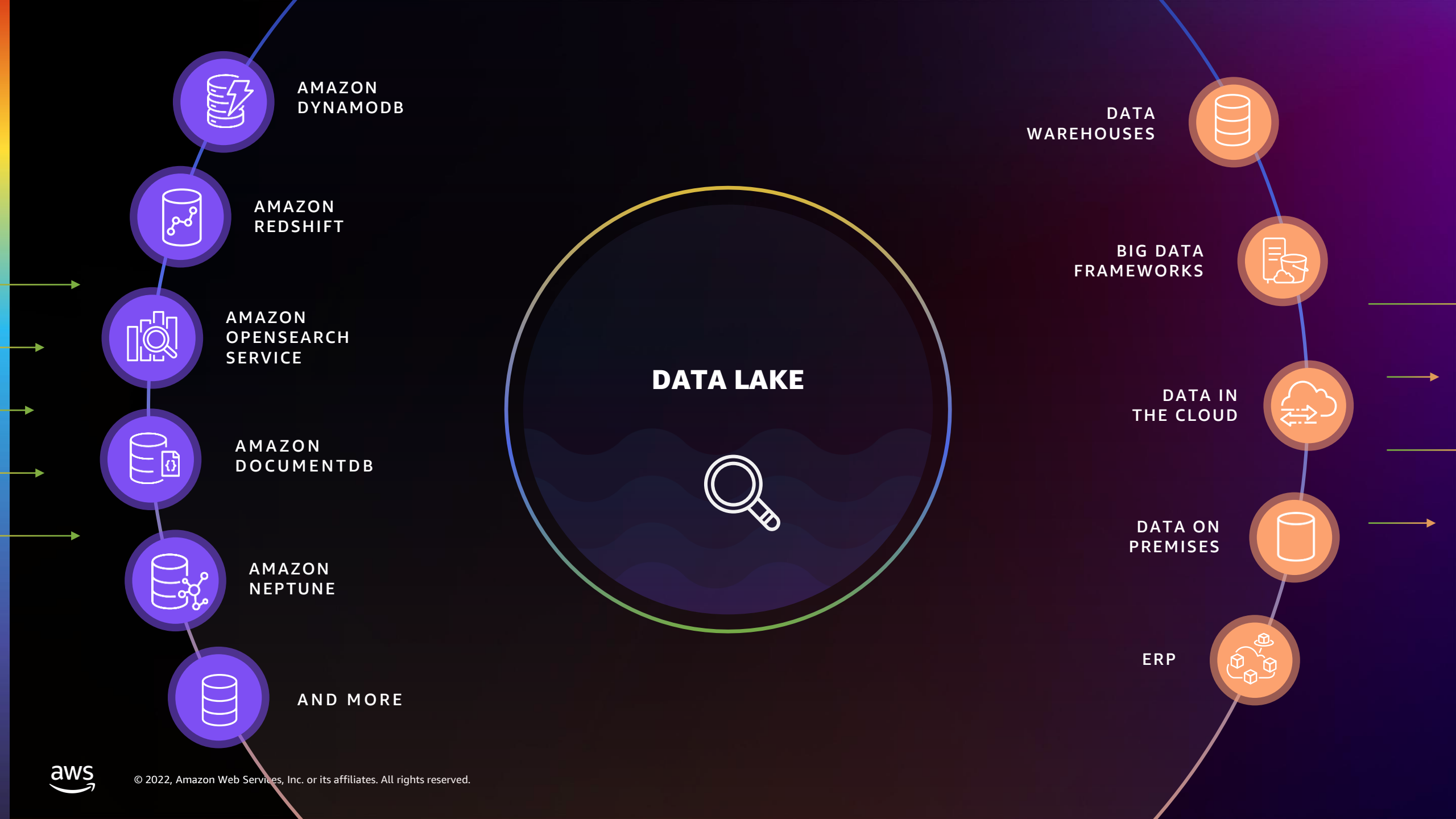


Open and flexible



Cost-effective

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Federated Query

DATA SOURCE CONNECTORS FOR IN-PLACE ANALYTICS



Integrate external databases with your data lake

No data movement

Flexible and easy way to ingest data for analysis

Process data in place

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Supported data sources

AWS

- Amazon Redshift
- Amazon DynamoDB
- Amazon RDS
- Amazon DocumentDB
- Amazon OpenSearch Service
- Amazon Timestream
- Amazon ElastiCache
- Amazon Neptune
- Amazon CloudWatch
- Amazon CloudWatch Metrics

THIRD PARTY

NEW

- Azure Data Lake Storage Gen2
- Azure Synapse
- Microsoft SQL Server
- Cloudera
- Google BigQuery
- Teradata
- SAP HANA
- Hortonworks
- Snowflake
- Oracle
- IBM Db2
- Redis
- HBase
- MySQL
- PostgreSQL
- Vertica
- TPC-DS
- Custom

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data source connectors use cases



Interactive query

Fast, scalable interactive queries on petabyte-scale data lakes



Business intelligence

Visualizations and scheduled reporting on your data lake



Data workflows

Lightweight data workflows to prepare data for downstream use cases



Analytics apps

Build applications that use Athena to surface insights



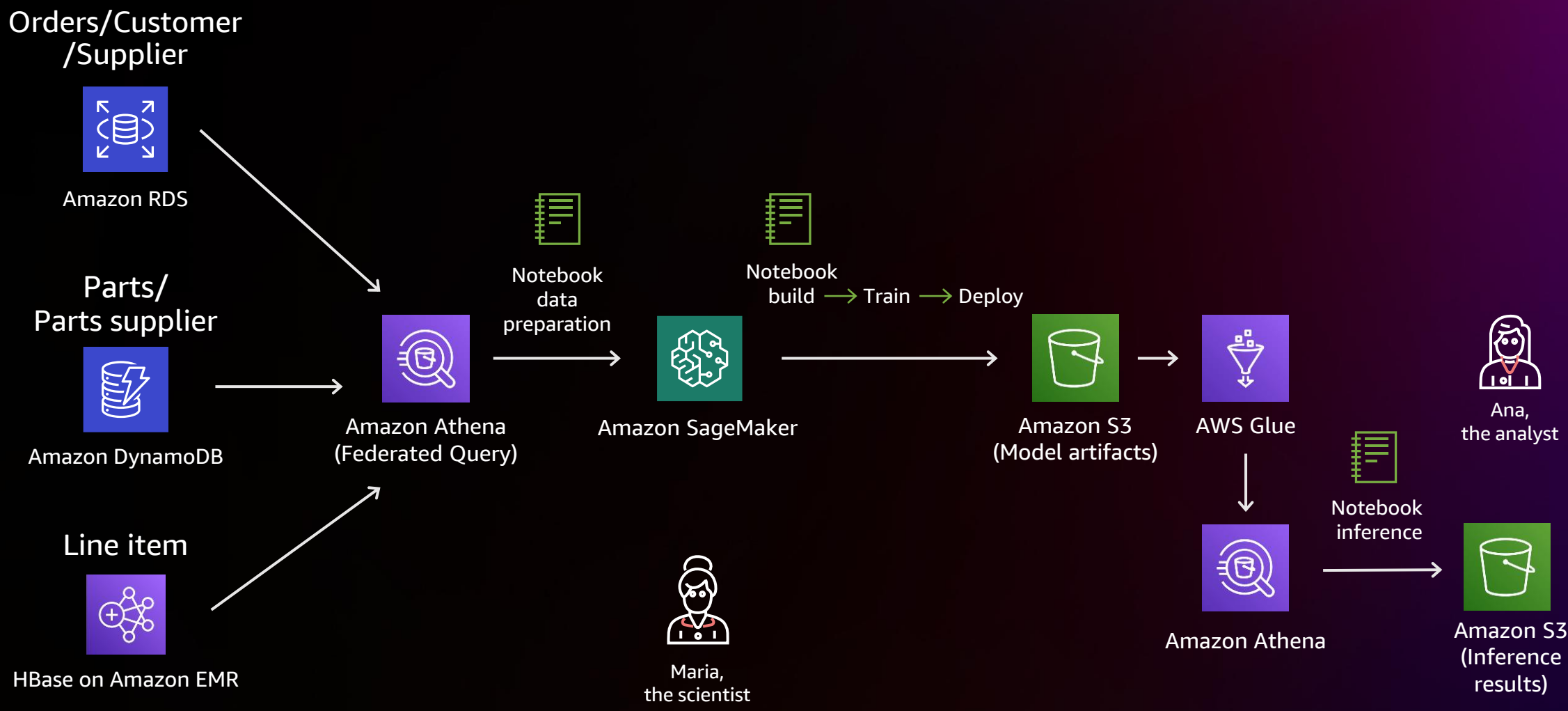
Machine learning

Combine raw data from multiple sources to train more diverse models

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Operationalize ML models with Athena



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data architecture pattern: Data virtualization

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data virtualization

"Data virtualization enables distributed databases and multiple heterogeneous data stores to be **accessed and viewed as a single database.**"

– Data Management Association International

Simplify data management

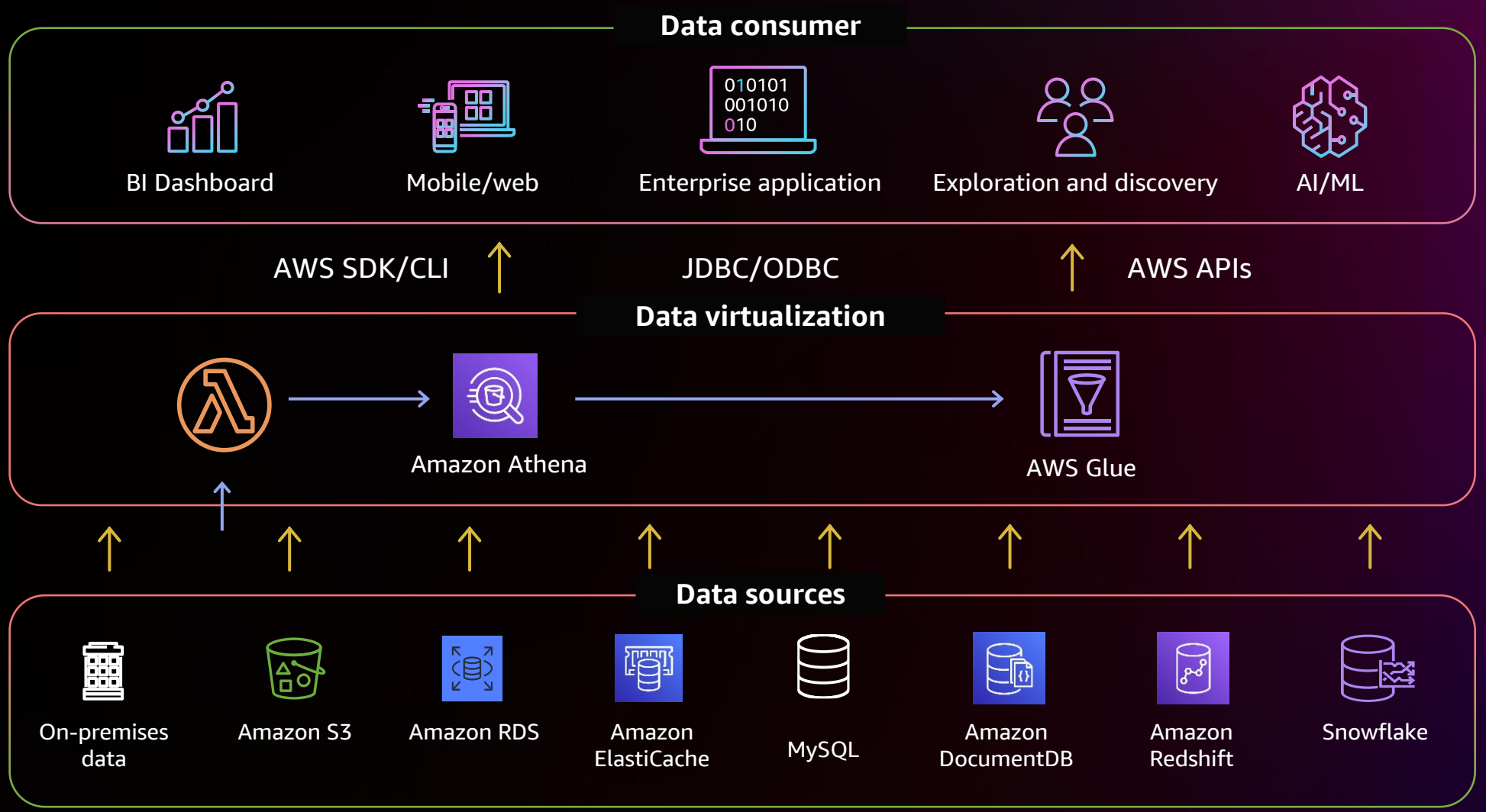
Improve collaboration

Reduce costs

Centralized access

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data virtualization with Athena



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data architecture pattern: Data mesh

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data mesh: four core principles



Domain ownership



Data as a product



Self-serve data platform

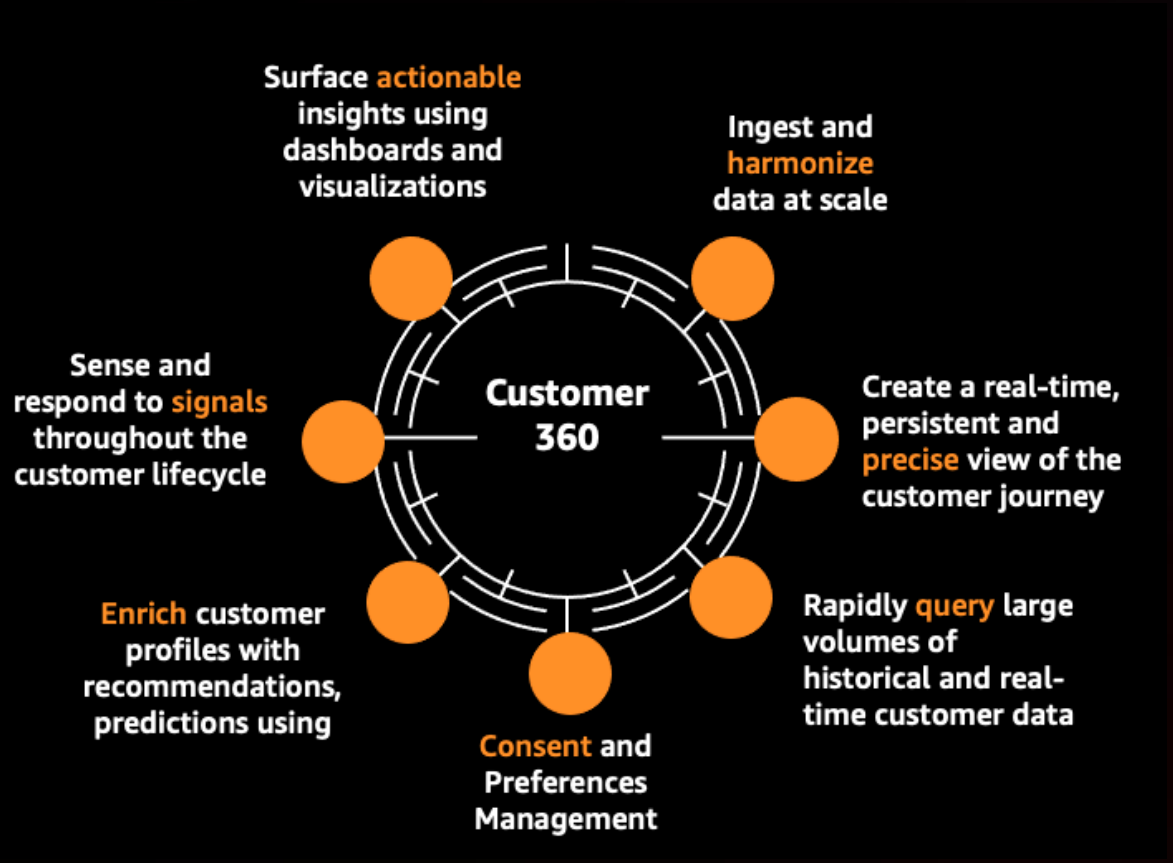


Federated computational governance

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

https://www.thoughtworks.com/radar/techniques/data-mesh

What is a data product?



Read-only analytical data organized by domains

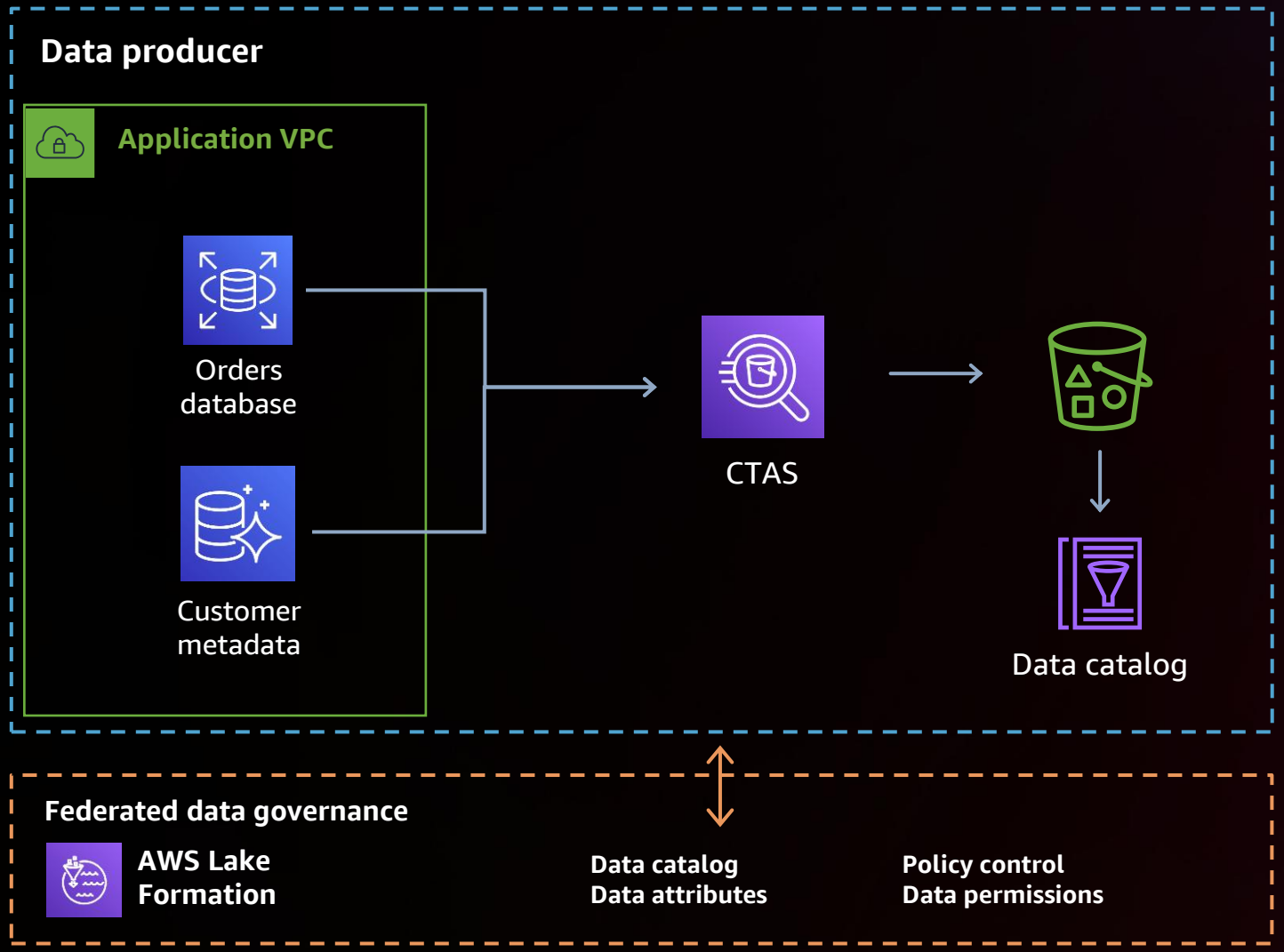
Broad, cohesive collections of related data aligned to business use cases and goals

Each data product is autonomous and data team manages their data products independently

Data products provide a data contract to ease data sharing

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Create data product with Athena



Data domain producers ingest data into their respective S3 buckets through a set of pipelines that they own and operate

Producers are data owners responsible for the **full lifecycle of the data** under their control

Producers **catalog** datasets to make them discoverable by consumers

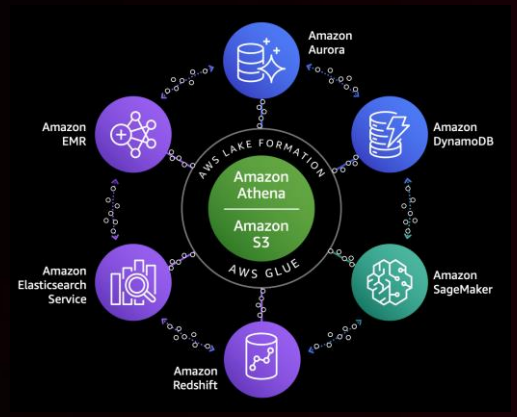
Producers control access to data through **Federated data governance** model to enforce **fine grained access control**

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The end result

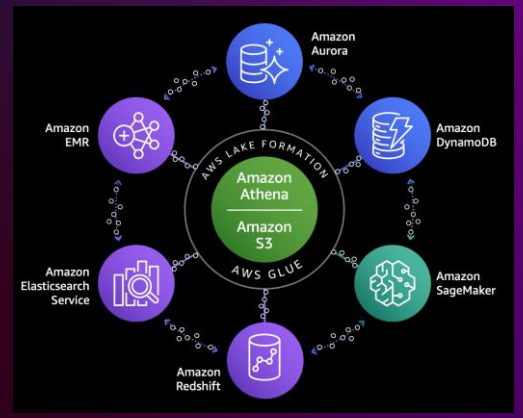
With a data mesh architecture, we aim to:

- **Accelerate** the **independent** delivery of business domain data lakes
- Increase data security and compliance within domains
- Enable **self-service** data product creation, discovery, and subscription
- Allow consumers to **transparently** access data products
- **Monitor, audit, and govern** access over time



Producer domain

Access data



Consumer domain

Discover data product
Request subscription

Create data product
Approve subscription



Data mesh



Monitor
Audit
Govern

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data mesh common patterns

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

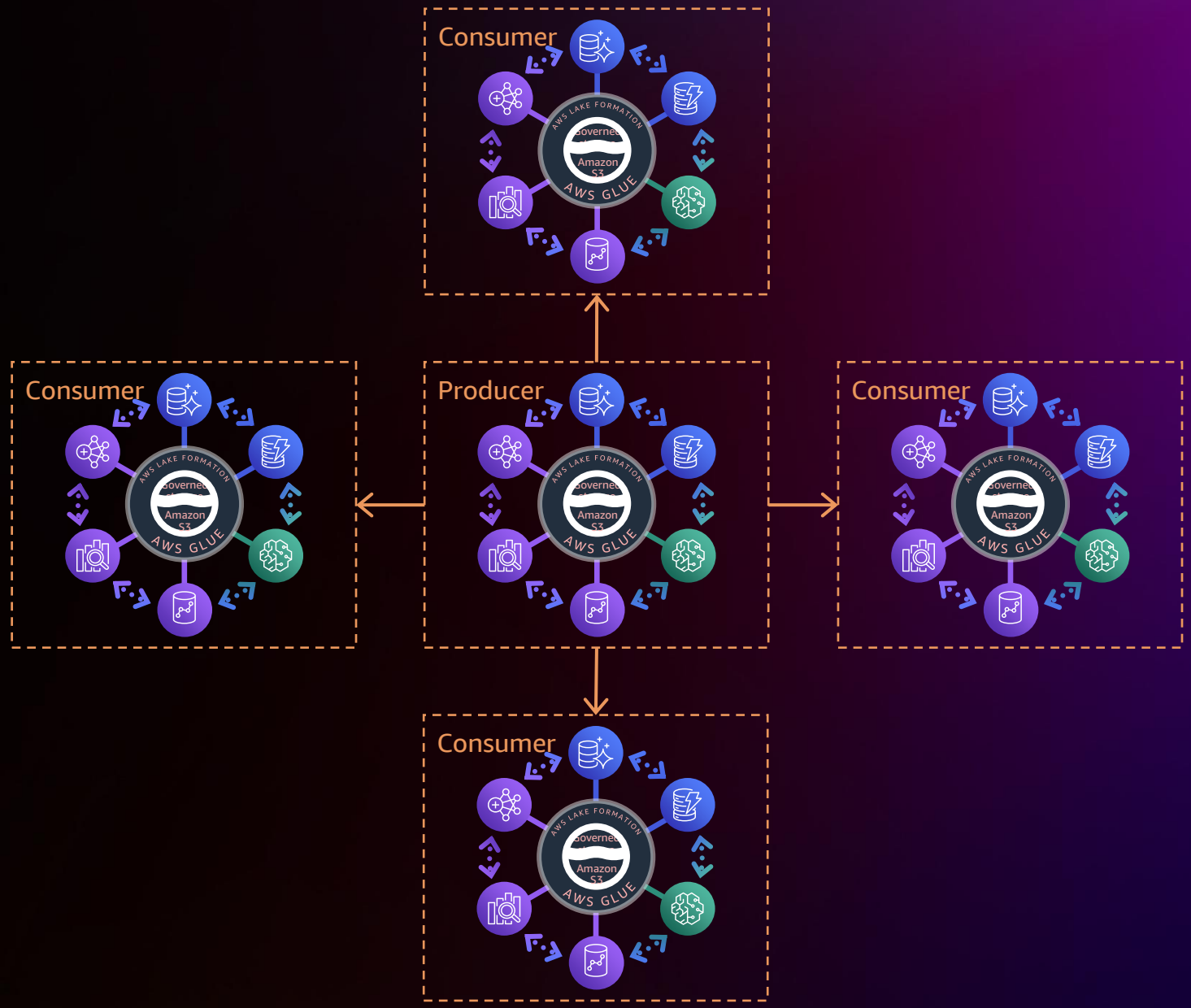
Hub and spoke

Producer shares directly with consumers

Decentralized governance

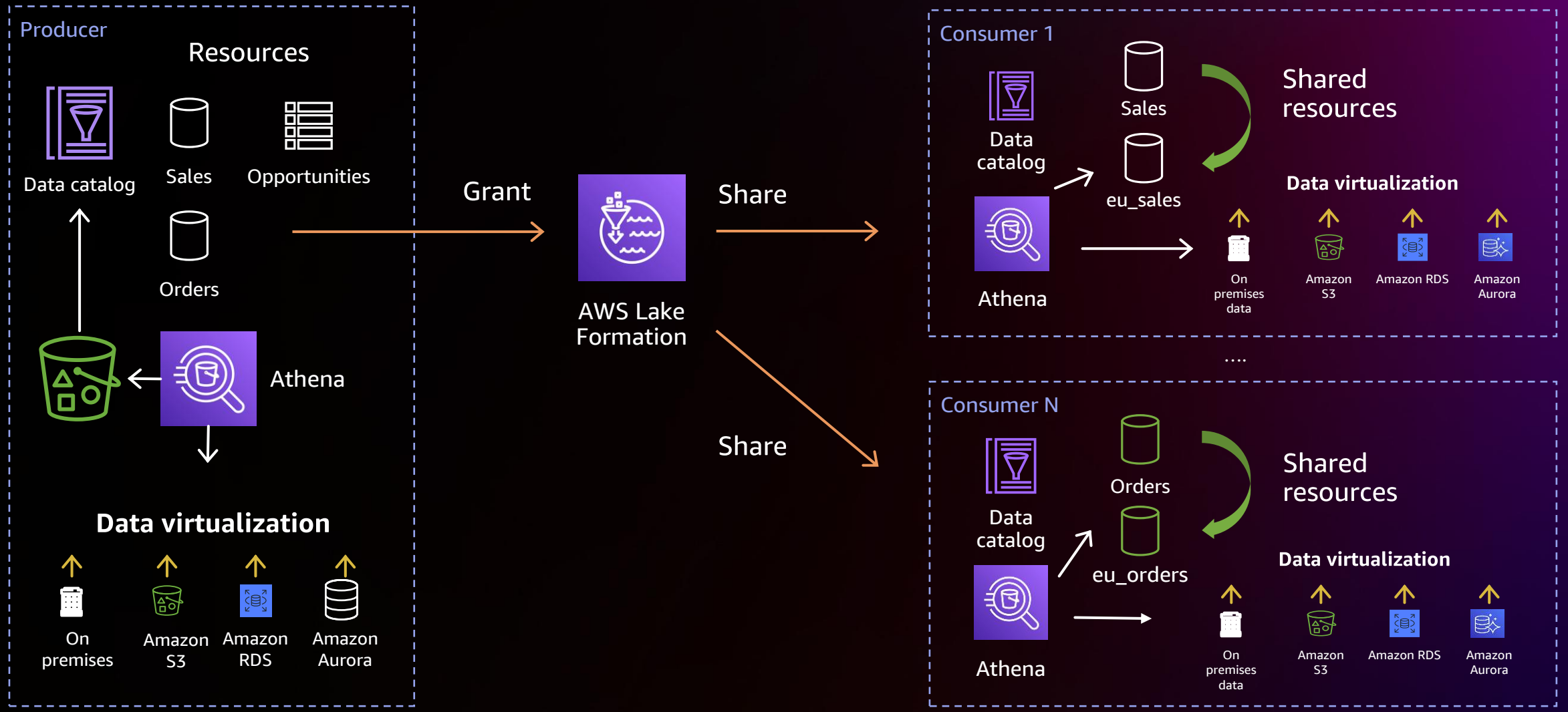
No central data catalog

Simple to set up and get started



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data mesh with Athena: Hub and spoke



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

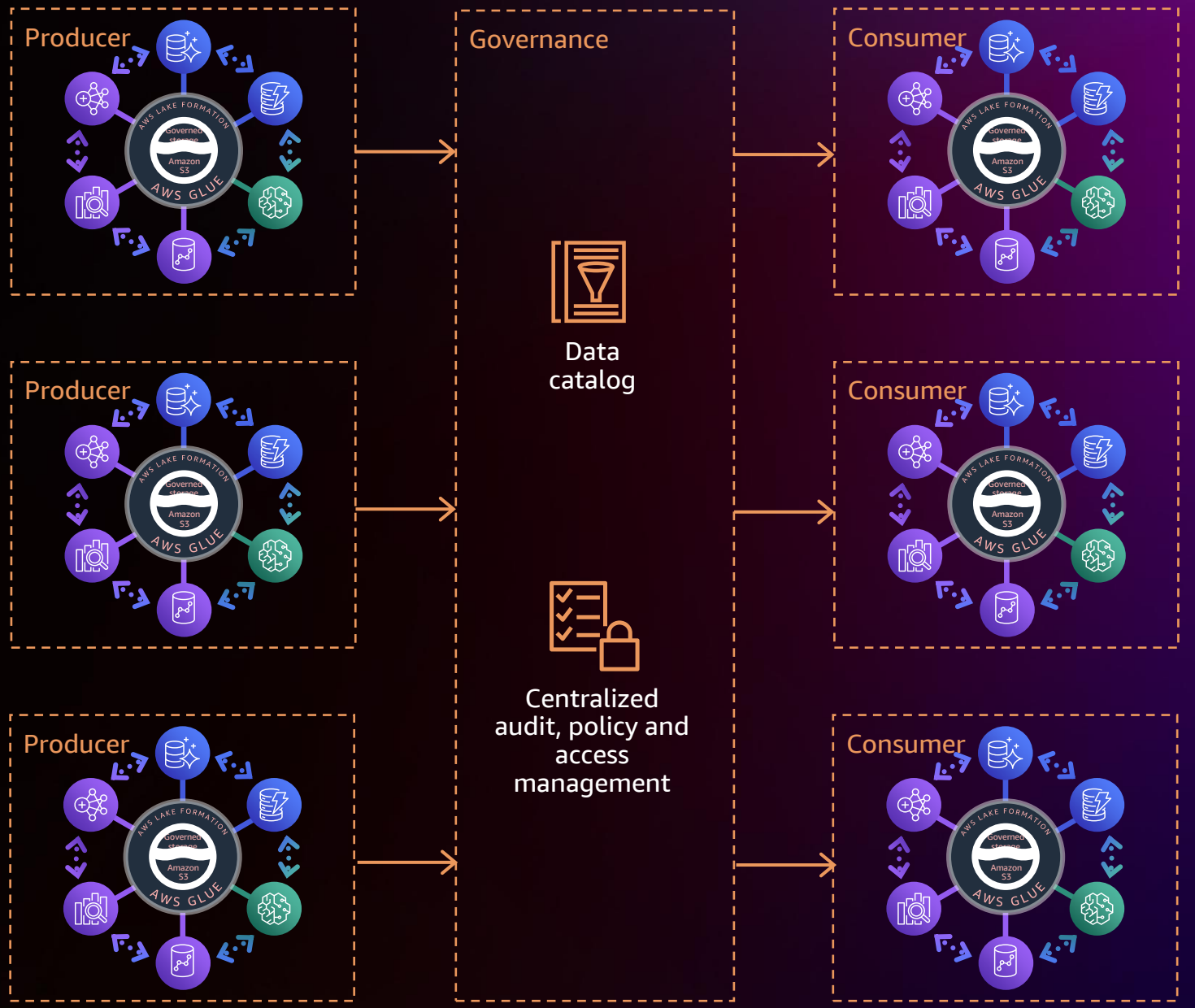
Centralized governance

Producer shares via central catalog

Centralized governance and audit

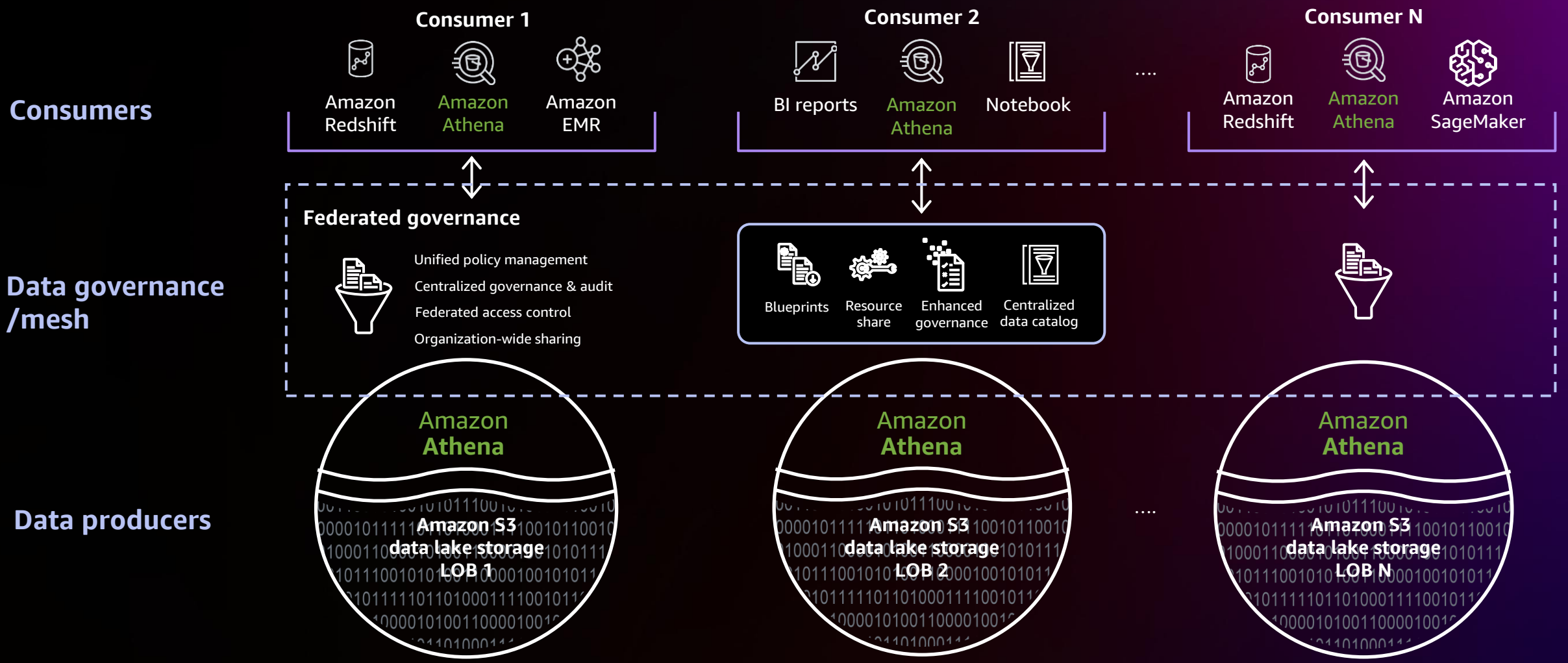
Federated access control

Unified policy management



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data mesh with Athena: centralized governance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Daniel Chen
danchenx@amazon.com

Saurabh Bhutyani
bhutyani@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.