
AWS
re:Invent
NOV. 27 – DEC. 1, 2023 | LAS VEGAS, NV

ANT331

Build an end-to-end data strategy for analytics and generative AI

Chanu Damarla
(he/him)
Principal Product Manager, New analytics initiatives
AWS

Ramkumar Nottath
(he/him)
Principal Solutions Architect, Analytics
AWS

Kiran Ramineni
(he/him)
VP
Single Family, Cloud, Data, ML/AI, and Infrastructure Architecture
Fannie Mae

Agenda

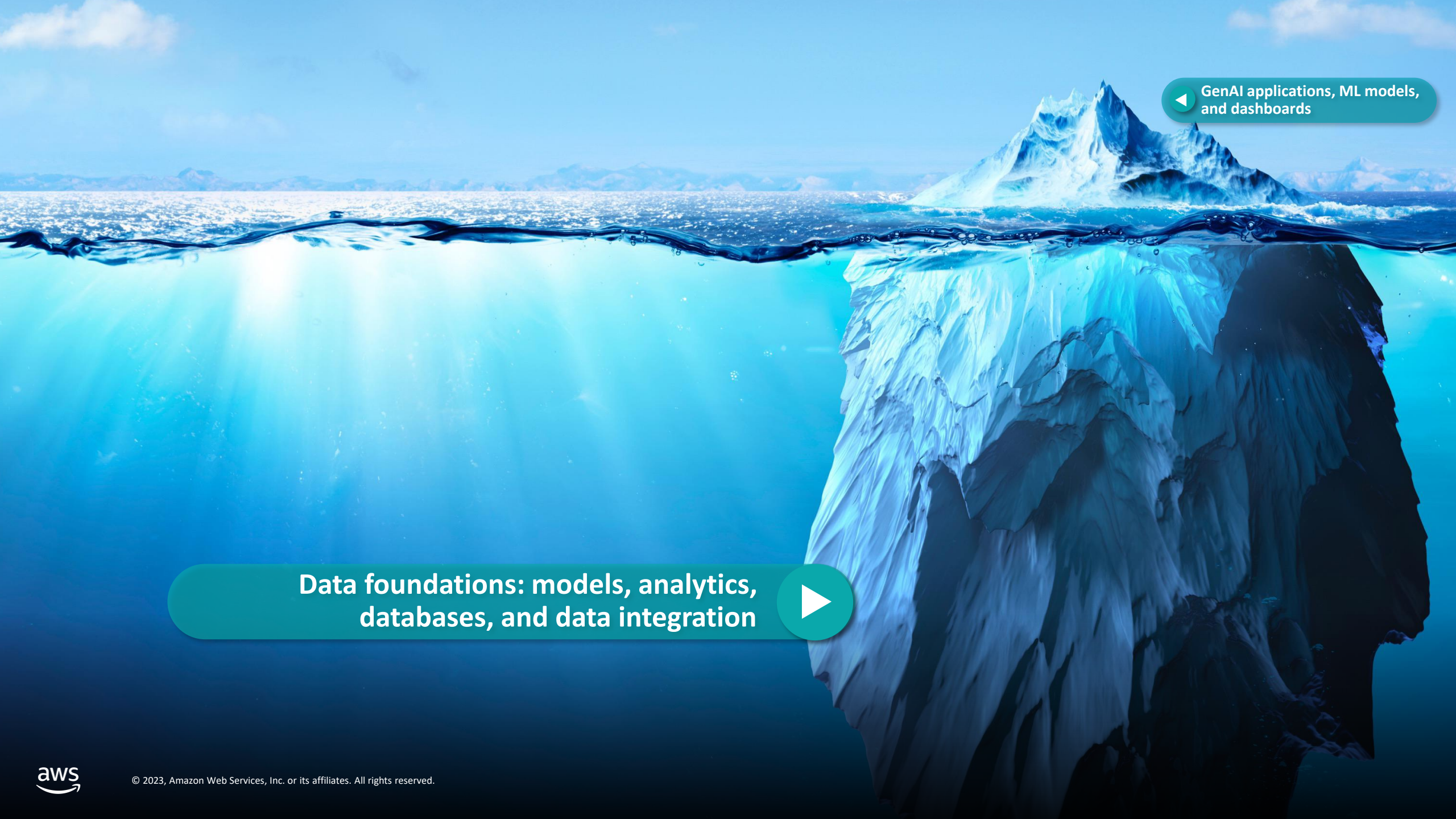
Why do you need an end-to-end data strategy?

How does AWS help you build your end-to-end data strategy?

Demo of building an end-to-end system with AWS

How Fannie Mae build their end-to-end data strategy using AWS

Conclusion



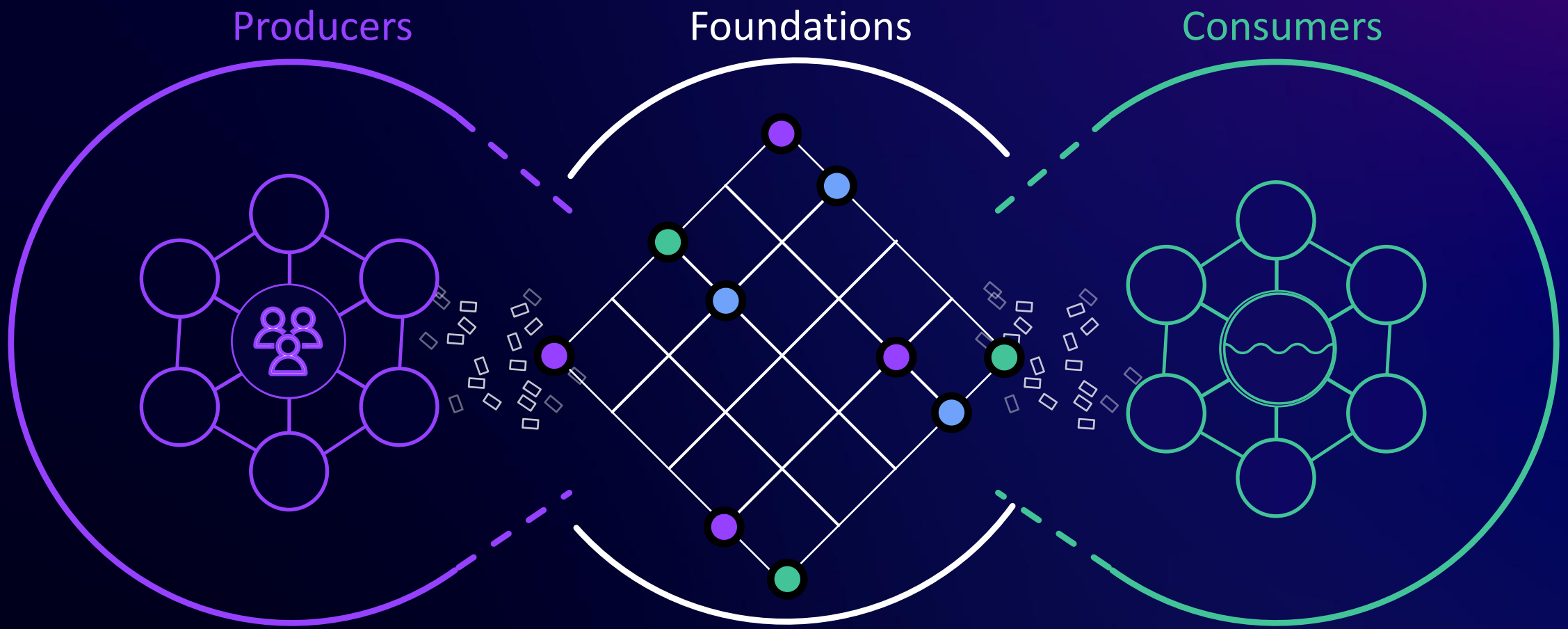
Deriving insights from data can be challenging

Data silos

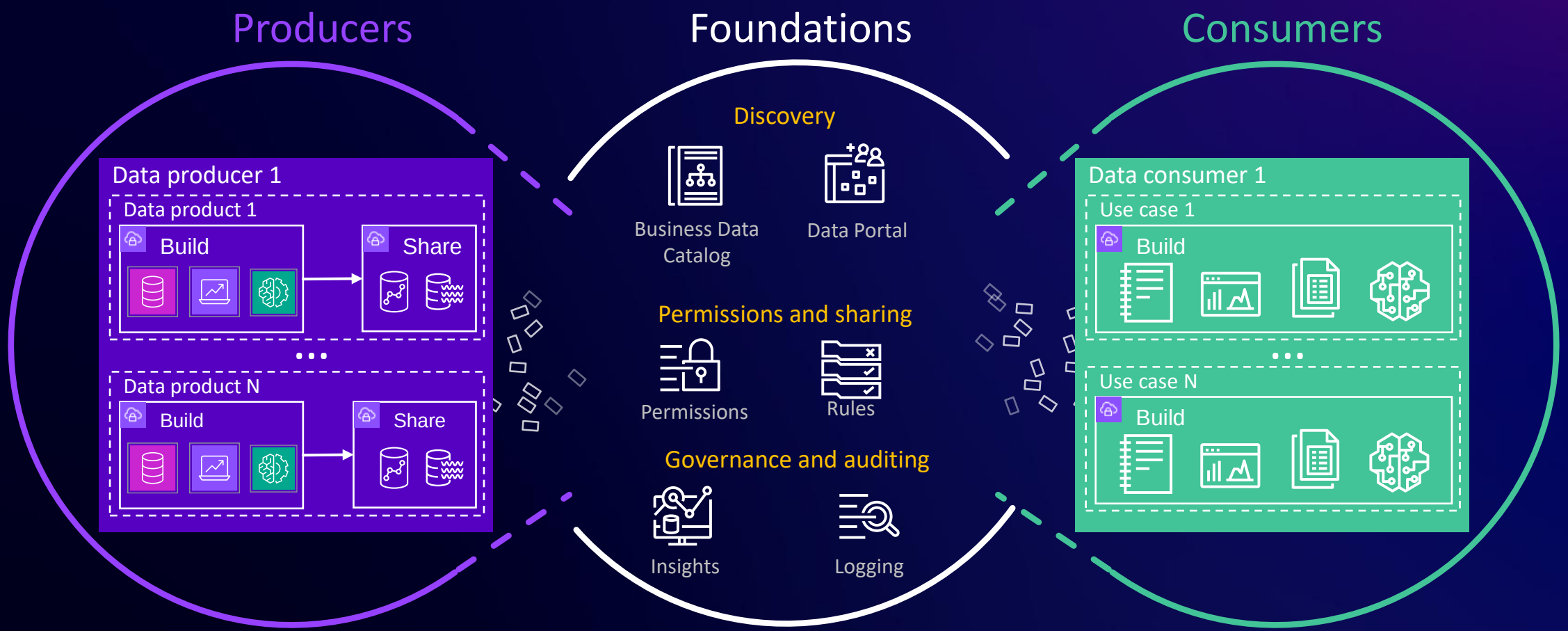
People silos

Business silos

End-to-end data strategy



End-to-end data architecture



Multi-account approach

Infra & SLAs

Standardize

Governance and compliance

Multi-account approach

Make data accessible

ANT336 | Governed data sharing with Amazon DataZone and AWS Lake Formation

Agenda

Why do you need an end-to-end data strategy?

How does AWS help you build your end-to-end data strategy?

Demo of building an end-to-end system with AWS

How Fannie Mae build their end-to-end data strategy using AWS

Conclusion

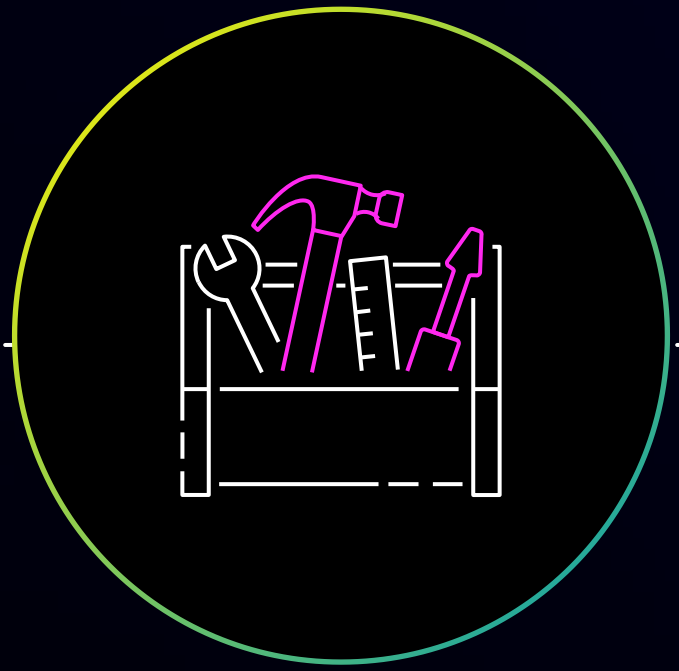
Purpose-built for

performance and cost

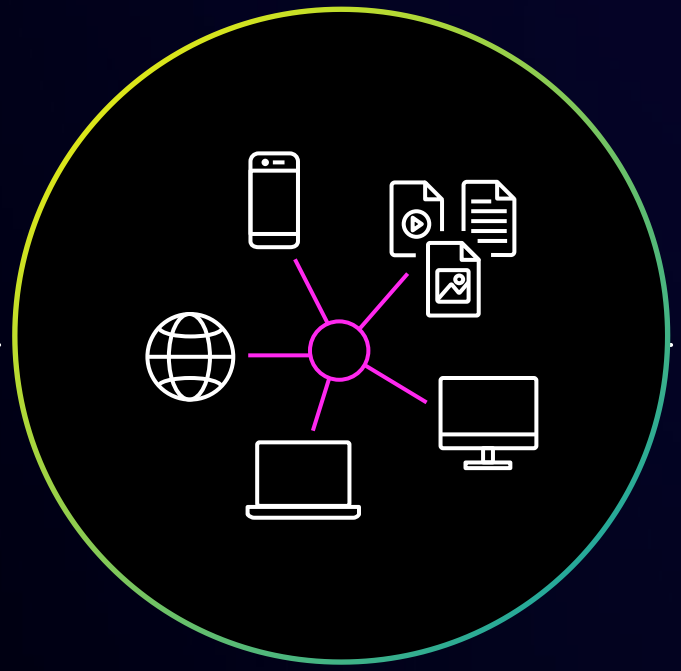
One size does not fit all

Generated @ https://partyrock.aws

Comprehensive set of tools to



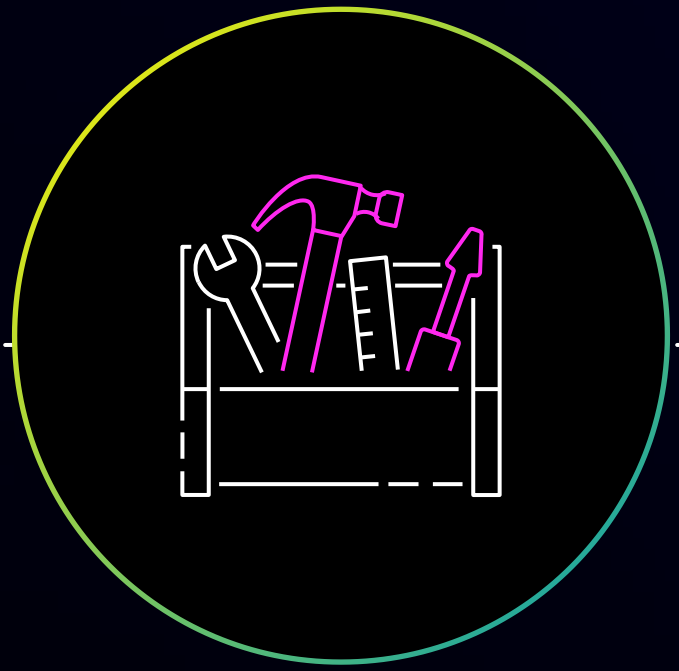
Utilize



Integrate

Govern

Comprehensive set of tools to

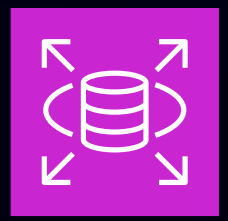


Utilize

Integrate

Govern

Complete set of relational and purpose-built databases



Amazon RDS

Amazon Aurora

KEY-VALUE

Amazon DynamoDB

DOCUMENT

Amazon DocumentDB

CACHING

Amazon ElastiCache

GRAPH

Amazon Neptune

TIME-SERIES

Amazon Timestream

LEDGER

Amazon QLDB

WIDE COLUMN

Amazon Keyspaces

MEMORY

Amazon MemoryDB for Redis

DAT212 | Future-proof you applications with AWS databases.

Amazon Aurora

COMMERCIAL-GRADE CLOUD NATIVE DATABASE DELIVERED AS A MANAGED SERVICE

Drop-in compatibility with MySQL and PostgreSQL

Simplicity and cost-effectiveness of open-source databases

Throughput and availability of commercial databases

Simple pay-as-you-go pricing

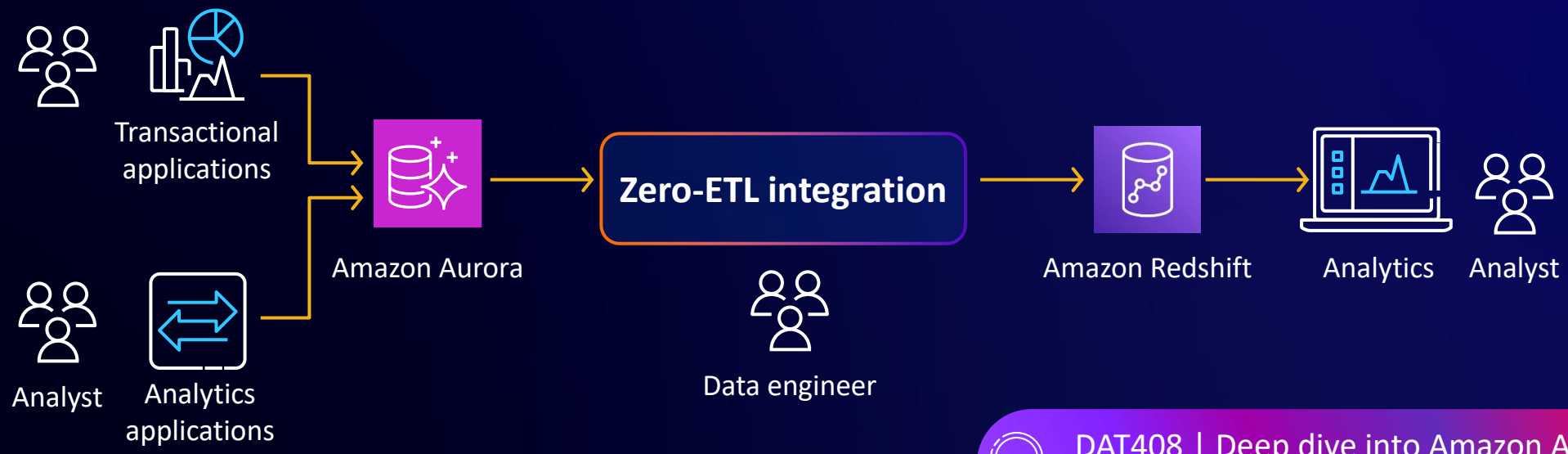
DAT408 | Deep dive into Amazon Aurora and its innovations

Amazon Aurora

COMMERCIAL-GRADE CLOUD NATIVE DATABASE DELIVERED AS A MANAGED SERVICE

Recent launch

Amazon Aurora MySQL zero-ETL integration with Amazon Redshift enables near real-time analytics on petabyte scale transactional data



DAT408 | Deep dive into Amazon Aurora and its innovations

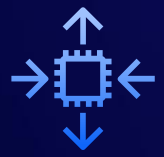
Comprehensive set of analytics services

INTERACTIVE QUERY	BIG DATA PROCESSING	REAL-TIME ANALYTICS	DATA WAREHOUSING	DATA INTEGRATION	BUSINESS INTELLIGENCE	OPERATIONAL ANALYTICS
						
Amazon Athena	Amazon EMR	Amazon Kinesis Amazon MSK Amazon Managed Service for Apache Flink	Amazon Redshift	AWS Glue	Amazon QuickSight	Amazon OpenSearch Service
SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS	SERVERLESS

Amazon Redshift

THE BEST PRICE-PERFORMANCE FOR CLOUD DATA WAREHOUSING

Analyze all your data



Easy, secure, and reliable

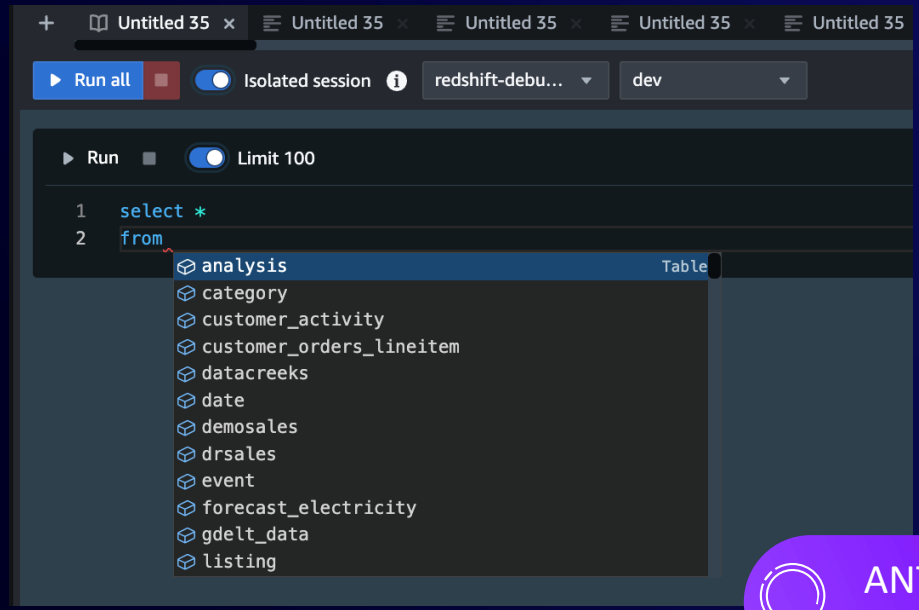
Price-performance at any scale

Amazon Redshift

THE BEST PRICE-PERFORMANCE FOR CLOUD DATA WAREHOUSING

Recent launch

Build queries more efficiently and accurately with syntax and error checking in Amazon Redshift Query Editor v2



ANT325 | Amazon Redshift: A decade of innovation in cloud data warehousing

Amazon EMR

EASILY RUN BIG DATA PROCESSING FRAMEWORKS LIKE SPARK, HIVE, PRESTO, OR FLINK

Fully managed and customizable

Latest version support

Multiple flexible and versatile deployment models

Best price performance

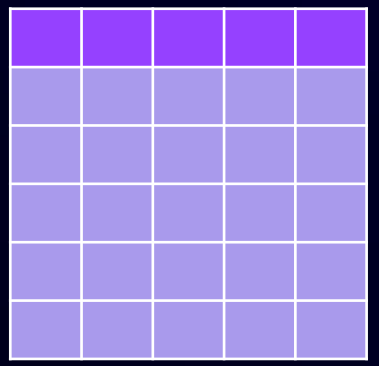
ANT204 | What's new in Amazon EMR and Amazon Athena?

Amazon EMR

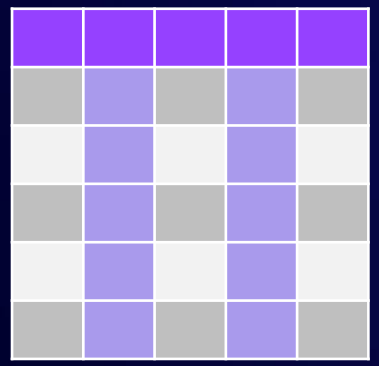
EASILY RUN BIG DATA PROCESSING FRAMEWORKS LIKE SPARK, HIVE, PRESTO, OR FLINK

Recent launch

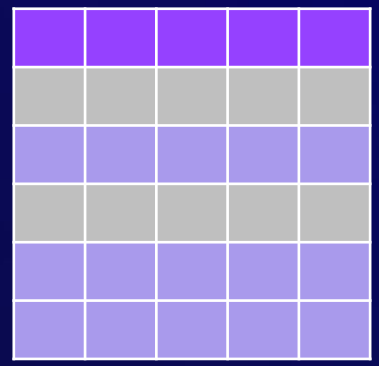
Fine-grained access control with AWS Lake Formation permissions for Spark jobs on Amazon EMR



Table



Column



Row

Amazon QuickSight

CLOUD-NATIVE BI SOLUTION
FOR ILLUMINATING
ORGANIZATIONAL INSIGHTS

Auto-scaling and serverless

Internal and/or external users

Deeply integrated with AWS services

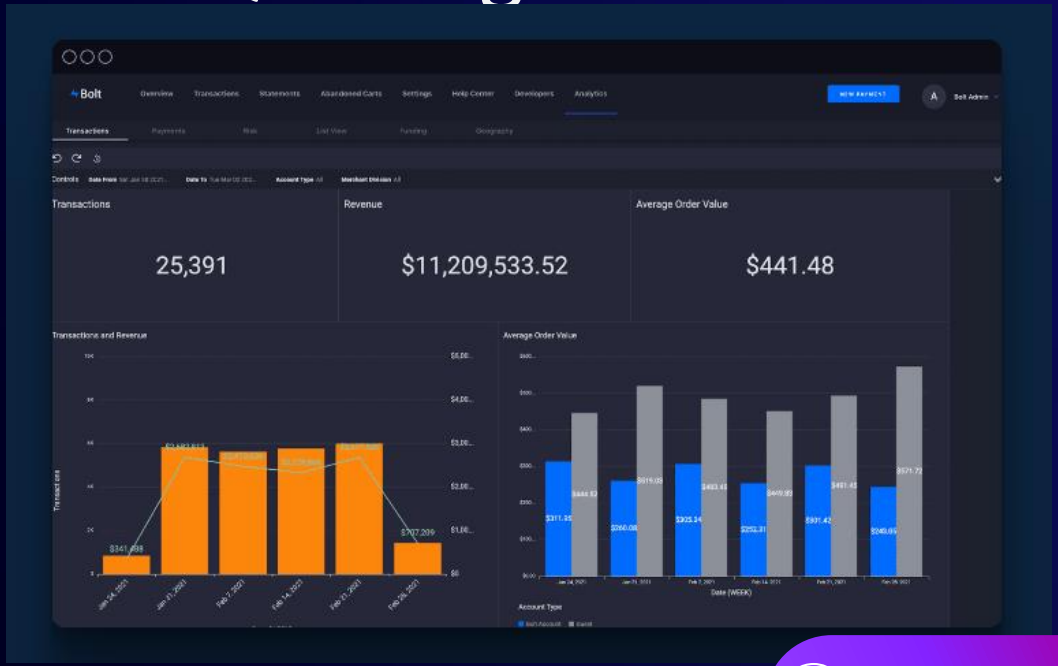
Augmented insights on-demand

Amazon QuickSight

CLOUD-NATIVE BI SOLUTION FOR ILLUMINATING ORGANIZATIONAL INSIGHTS

Recent launch

Democratize data access with embedded Amazon QuickSight dashboards



Amazon OpenSearch Service

REAL-TIME SEARCH, MONITORING, AND ANALYSIS OF OPERATIONAL DATA

Managed: Increase operational excellence and use a popular open source solution

Secure: Audit and secure your data with a data center and network architecture and built-in certifications

Observability: Systematically detect potential threats and react to a system's state through an open source solution for machine learning, alerting, and visualization

Cost conscious: Optimize time and resources for strategic work

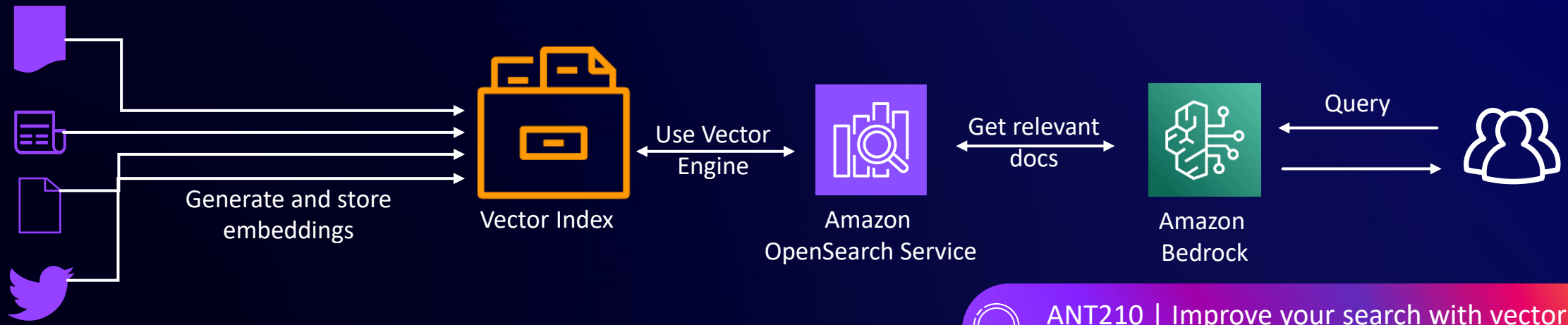
ANT301 | What's new in Amazon OpenSearch Service

Amazon OpenSearch Service

REAL-TIME SEARCH, MONITORING, AND ANALYSIS OF OPERATIONAL DATA

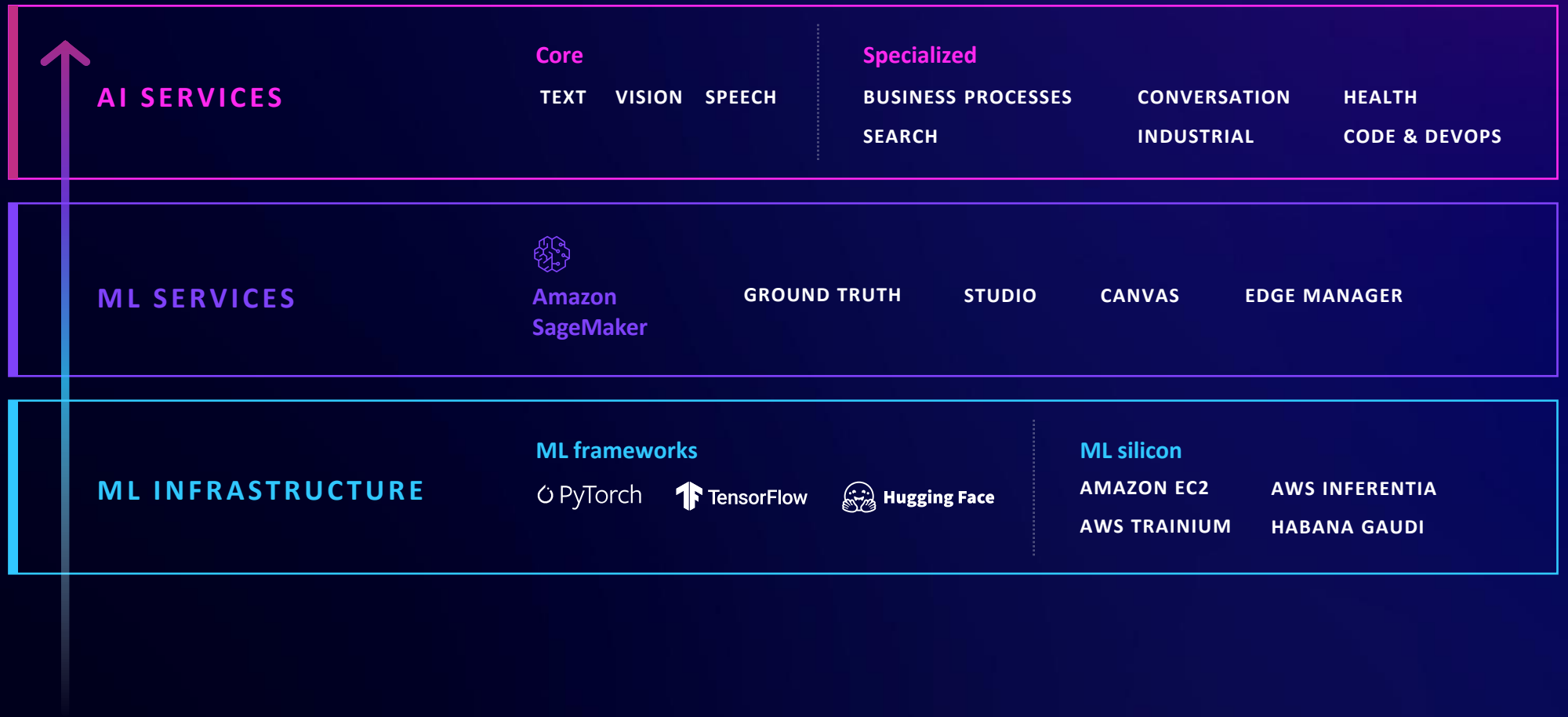
Recent launch

Vector Engine for Amazon OpenSearch Service Serverless: Simple, scalable, high-performance vector storage and search capability without having to manage infrastructure with



ANT210 | Improve your search with vector capabilities in OpenSearch Service

Broad set of machine learning capabilities



Amazon Bedrock

THE EASIEST WAY TO BUILD AND SCALE GENERATIVE AI APPLICATIONS WITH FOUNDATION MODELS

Accelerate development of generative AI applications using FMs through an API, without managing infrastructure

Choose FMs from Amazon, AI21 Labs, Anthropic, Cohere, Meta, and Stability AI to find the right FM for your use case

Privately customize FMs using your organization's data

AIM218 | Build your first generative AI application with Amazon Bedrock.

aws

Amazon Bedrock

Choice of foundation models

AI21labs

JURASSIC-2

Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch

ANTHROP\C

CLAUDE 2

LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems

co:here

COMMAND

Text generation model for business applications like summarization, copywriting, dialog, extraction, and question answering

Meta AI

LLAMA 2

Pre-trained and fine-tuned LLMs for natural language tasks like question answering and reading comprehension

stability.ai

SDXL 1.0

Generation of unique, realistic, high-quality images, art, logos, and designs

amazon

AMAZON TITAN

Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings and search

aws

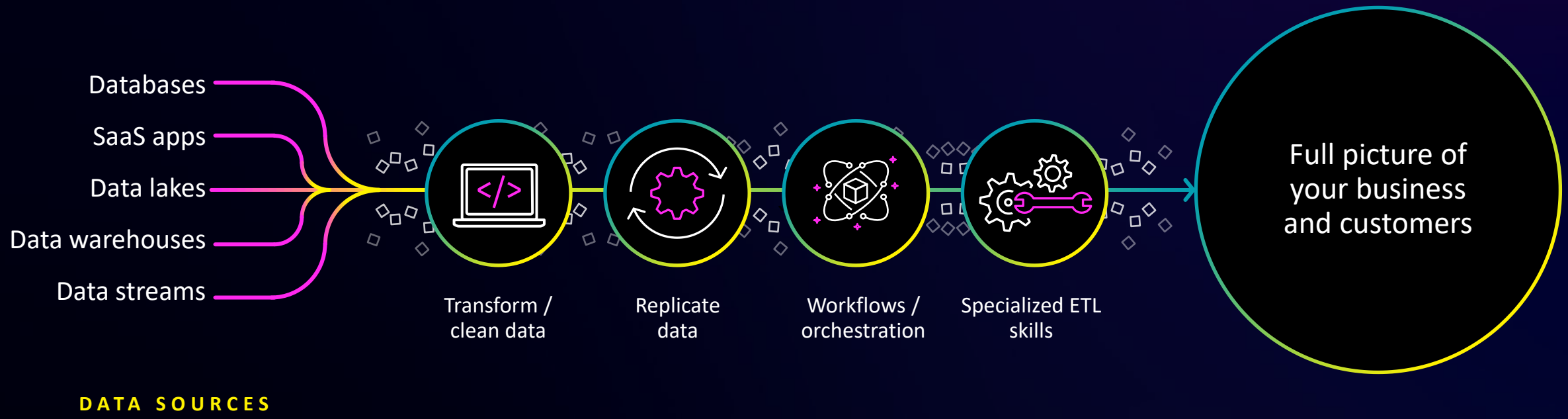
Comprehensive set of tools to

Utilize

Integrate

Govern

Data integration gives you a full picture



Use-case specific data integration services

ETL & Data Integration

AWS Glue

Migration

AWS Database Migration Service

Marketplace

AWS Data Exchange

Orchestration

Amazon Managed Apache Airflow

Orchestration

AWS Step Functions

SaaS Apps

Amazon AppFlow

Eventing

Amazon EventBridge

Data Transfer

AWS Transfer Family

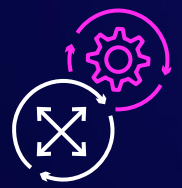
Data Sync

AWS DataSync

AWS Glue

SIMPLE, SCALABLE, AND SERVERLESS DATA INTEGRATION

Integrate data faster



Automate at scale

No servers to manage

Built on Spark, Python, and Ray

AWS Glue

SIMPLE, SCALABLE, AND SERVERLESS DATA INTEGRATION

Recent launch

ETL AI coding assistant in AWS Glue Notebooks, powered by Amazon CodeWhisperer. Empowers everyone to build DI pipelines using natural language

```
# Write Spark DataFrame into Redshift
def write_spark_df_to_redshift(spark_df, redshift_table_name, redshift_conn):
    spark_df.write.format("jdbc").options(
        url=redshift_conn.url,
        driver=redshift_conn.driver,
        dbtable=redshift_table_name,
        user=redshift_conn.user,
        password=redshift_conn.password
    ).mode("append").save()
```


DOP202 | Realizing the developer productivity benefits of Amazon CodeWhisperer.

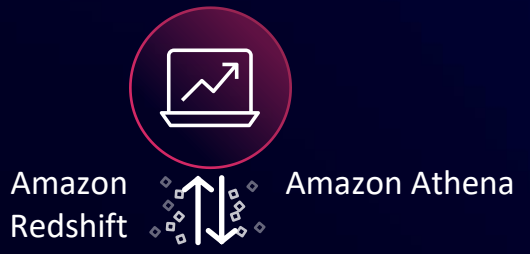
INVESTING IN A

Zero ETL

FUTURE

Investing in a zero-ETL future by

Accessing data in place via federated queries



Moving analytics and ML closer to the data

Amazon Aurora ML

Amazon Athena ML

Amazon Redshift ML

Amazon QuickSight ML

Building point-to-point integrations

ANT332 | Customize Amazon Athena to integrate with new data sources.

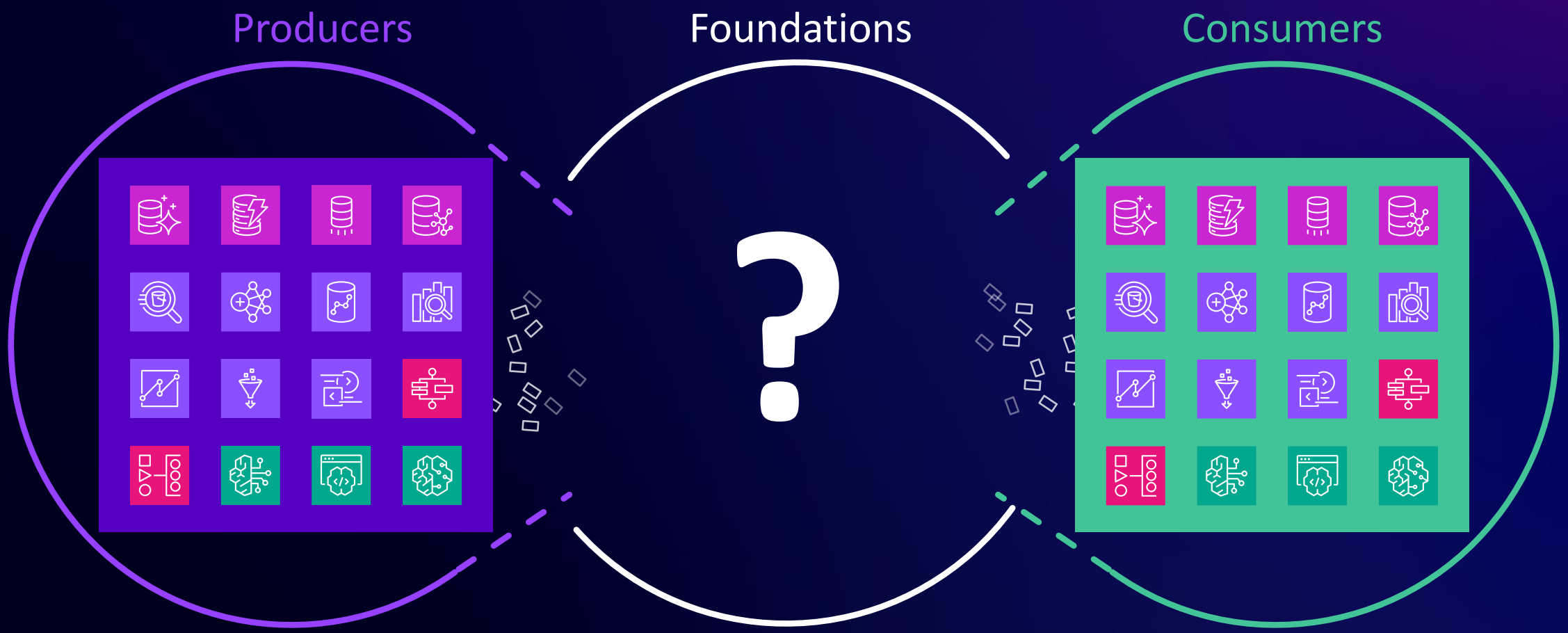
Comprehensive set of tools to

Utilize

Integrate

Govern

What about the foundations?



Built-in governance tools

Amazon SageMaker **ML governance**

Governance and auditability for end-to-end ML development

AWS **Lake Formation**

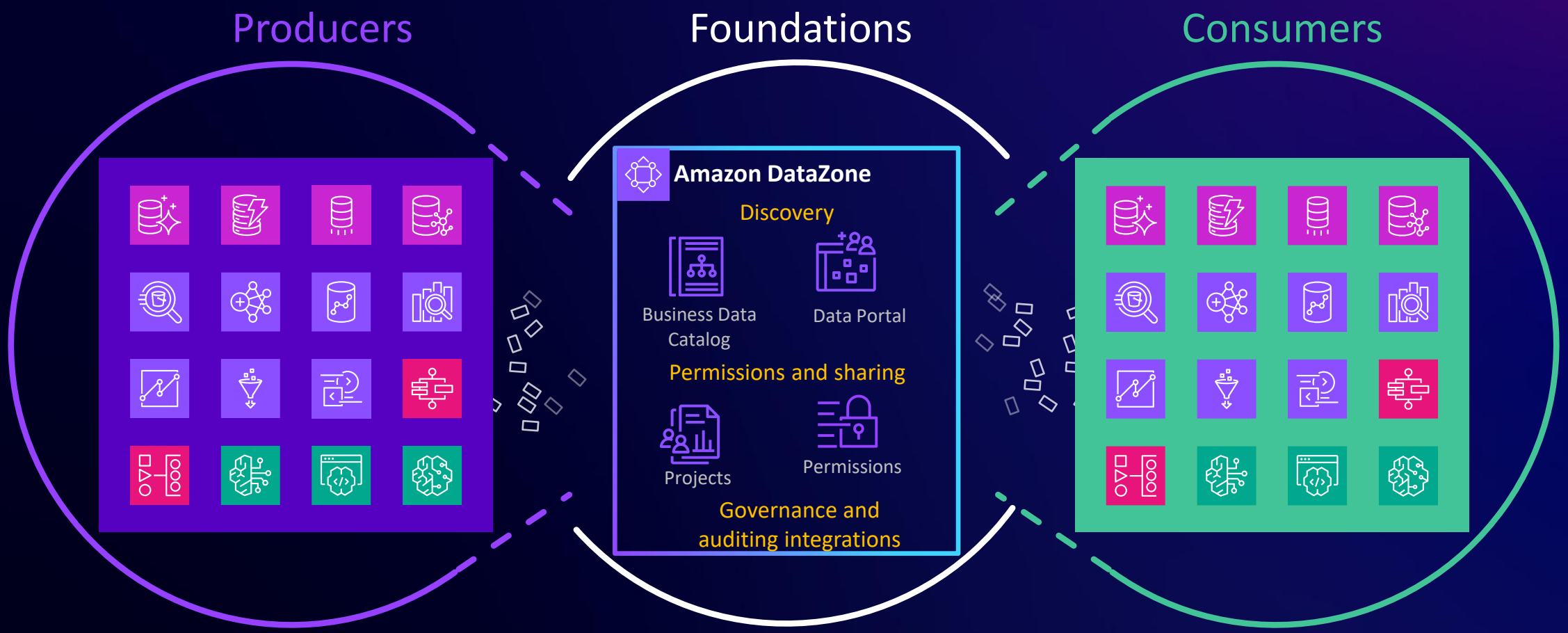
Govern data within your data lake on Amazon S3

Amazon **DataZone**

Catalog, discover, share, and govern data across the organization

ANT303 | What's new in AWS Lake Formation?

What is in Amazon DataZone?



Agenda

Why do you need an end-to-end data strategy?

How does AWS help you build your end-to-end data strategy?

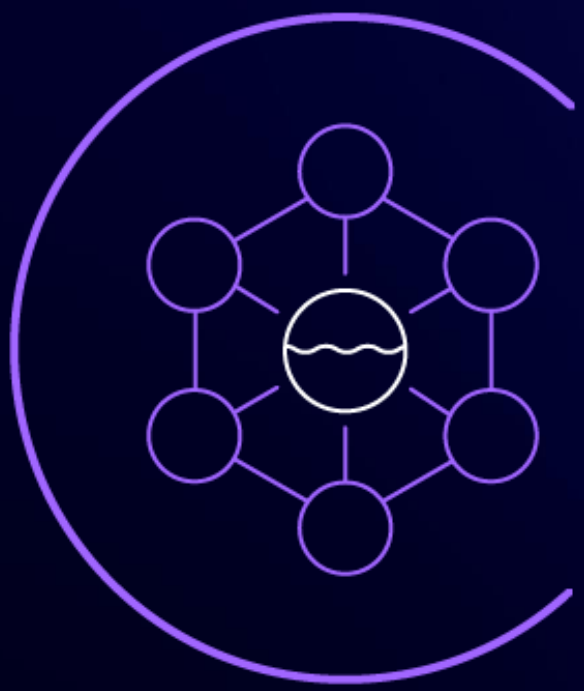
Demo of building an end-to-end system with AWS

How Fannie Mae build their end-to-end data strategy using AWS

Conclusion

End-to-end demo…

Producers



Build a curated data set using a ZETL integration between Aurora MySQL and Amazon Redshift

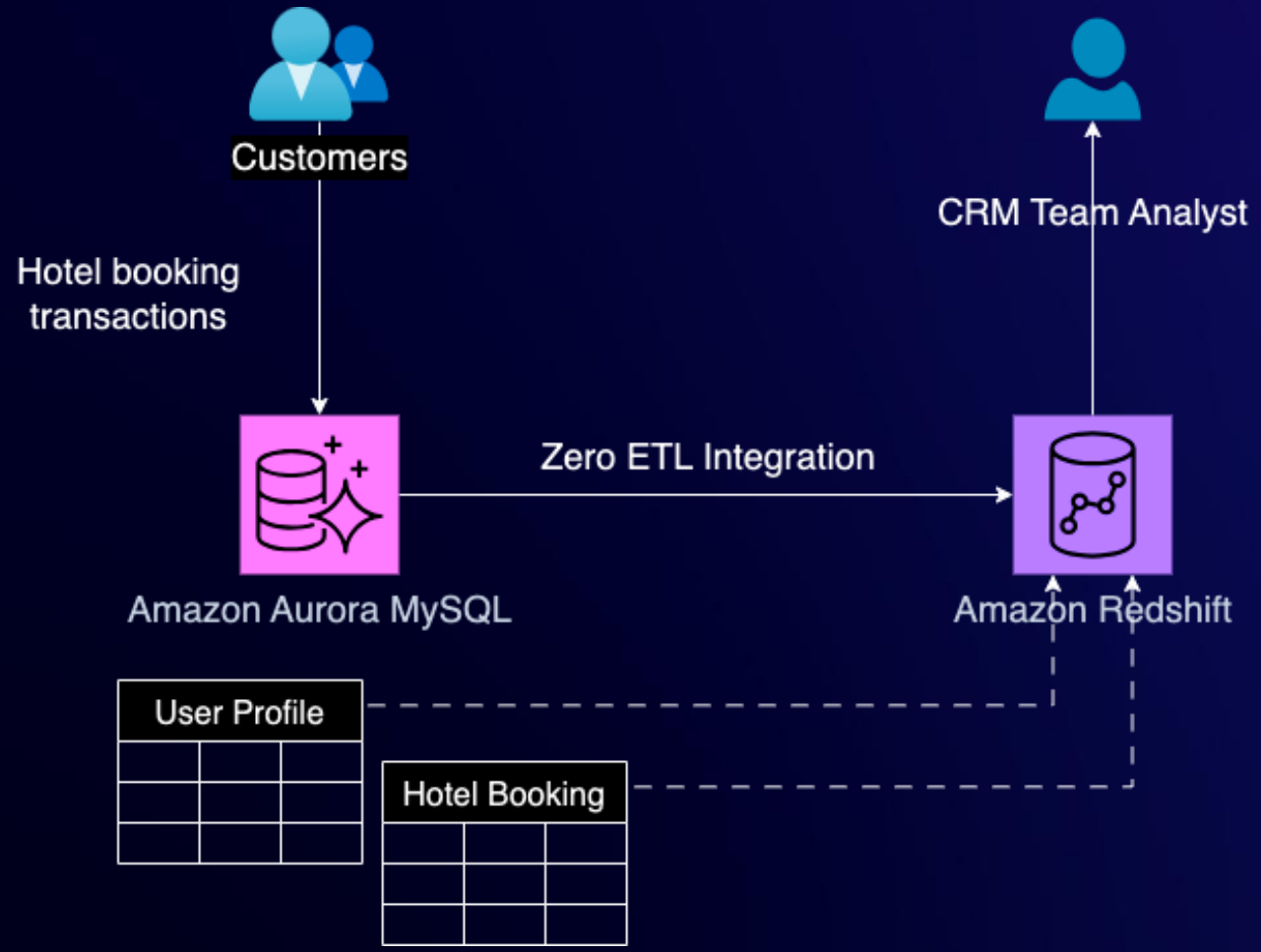
Technology

Share the curated data set across your organization

Consumers

Build a generative AI application using the curated data set

Demo: Create a curated data set…



Reference Blog

Console Home Info

Reset to default layout | + Add widgets

Recently visited Info

- Amazon DataZone
- Amazon Redshift
- Amazon Bedrock
- RDS
- S3
- Cloud9
- CloudShell
- IAM
- Lambda
- EC2
- CloudFormation
- Secrets Manager

View all services

Welcome to AWS

Getting started with AWS
Learn the fundamentals and find valuable information to get the most out of AWS.

Training and certification
Learn from AWS experts and advance your skills and knowledge.

What's new with AWS?
Discover new AWS services, features, and Regions.

AWS Health Info

Open issues
0 Past 7 days

Scheduled changes
0 Upcoming and past 7 days

Other notifications
0 Past 7 days

Go to AWS Health

Cost and usage Info

No cost and usage data

You haven't configured AWS Cost Explorer or you do not have permission.

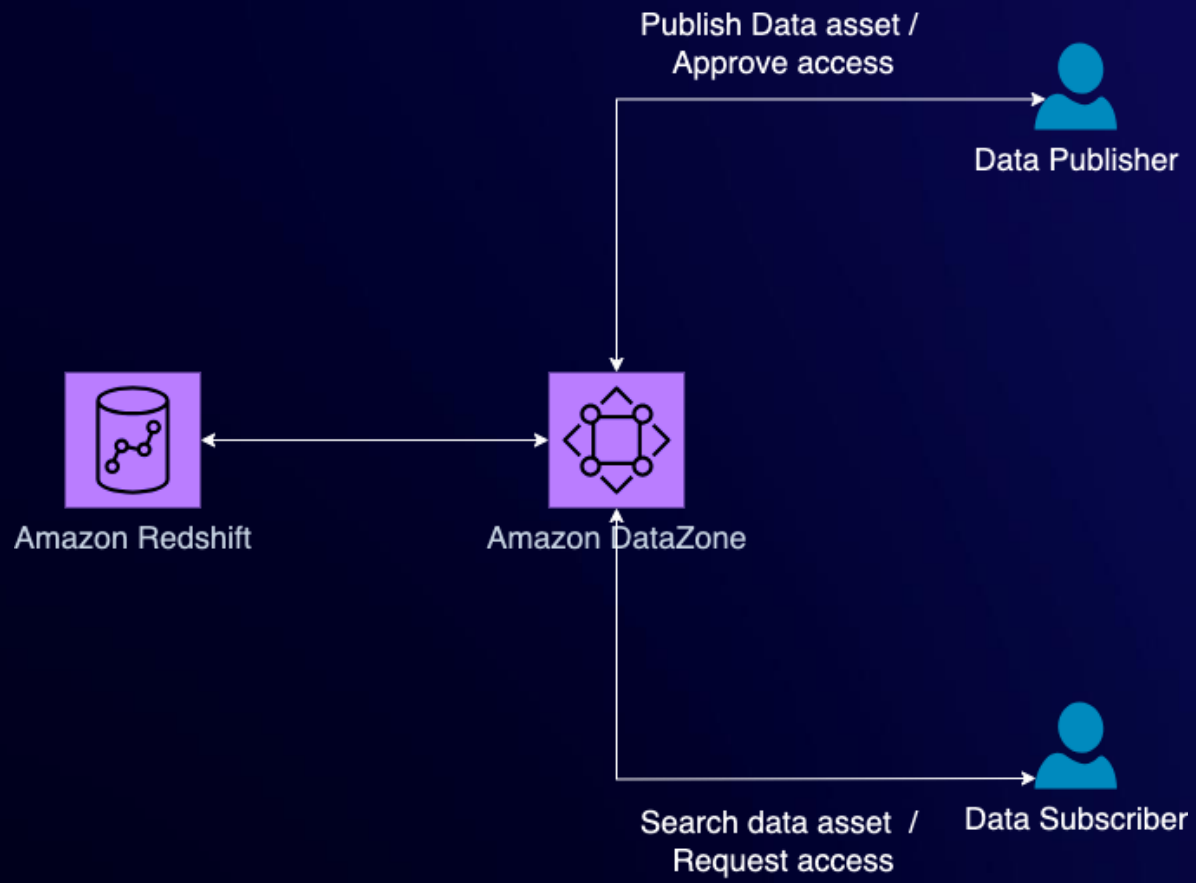
Turn on Cost Explorer

Build a solution Info

Start building with simple wizards and automated workflows.

- Launch a virtual machine — With EC2 (2 mins)
- Register a domain — With Route 53 (3 mins)
- Start a development project — With CodeStar (5 mins)
- Build a web app — With AWS App Runner (5 mins)
- Deploy a serverless microservice — With Lambda (2 mins)
- Start migrating to AWS — With AWS MGN (2 mins)
- Host a static web app — With AWS Amplify Console (2 mins)
- Build SQL Server on AWS — With high availability (HA and FCI) (2 mins)
- Deploy SAP on AWS — With NetWeaver and HANA (with HA) (1…

Demo: Sharing data across the organization



Reference Blog

CUSTOMER_PUBL...

Search Assets

Producer

Customer_Publisher

OVERVIEW | DATA | ENVIRONMENTS | MEMBERS

Projects / Customer_Publisher

Customer_Publisher

ACTIONS

Created by: Admin • Created on: Nov 20, 2023, 05:31:05 PM • Project ID: 66dwdeevlkigpz

No description

Project overview

PUBLISHED ASSETS	SUBSCRIBED ASSETS	DATA SOURCES	ENVIRONMENTS	MEMBERS
0	0	0	0	2

Working with Projects

Find data

Find data in the catalog and get access to it through a project.

VIEW SUBSCRIBED DATA

VIEW PENDING REQUESTS

Publish data

Publish data from your data source into the catalog through a project.

CREATE DATA SOURCE

VIEW INCOMING REQUESTS

Manage data catalog

Build your business metadata using glossary terms and metadata forms.

VIEW METADATA FORMS

VIEW GLOSSARIES

Set up environments

Create project environments to use and publish data.

CREATE ENVIRONMENT

CREATE PROFILE

ANALYTICS TOOLS

No tools

Create an environment to work with your data.

CREATE ENVIRONMENT

Send feedback

Generic | SA | Blogs | Demo | Hiring | SKO

Consumer

CUSTOMER_PUBL...

Search Assets

Customer_Publisher

OVERVIEW | DATA | ENVIRONMENTS | MEMBERS

← Back

User Profile

Technical name: user_profile • Asset Type: amazon.datazone.RedshiftTableAssetType

No description

Published listing for data asset **User Profile** is being created

ACTIONS

BUSINESS METADATA | SCHEMA | HISTORY

README

HIDE

Add a readme section

A readme will help users understand this page better

CREATE README

GLOSSARY TERMS What's this?

No terms found

Annotate this asset with terms

ADD TERMS

METADATA FORMS What's this?

Amazon Redshift table

Account ID	537175266406
Database name	trip_data_zero_etl
Storage type	Serverless
Region	us-east-1

ASSET DETAILS

PUBLISHED REVISION
3

LATEST INVENTORY REVISION
3

OWNING PROJECT
Customer_Publisher

SUBSCRIPTION APPROVAL
Required EDIT

LAST MODIFIED BY
SYSTEM

UPDATED AT
Nov 20, 2023, 08:51:57 PM

CREATED BY
SYSTEM

CREATED AT
Nov 20, 2023, 08:51:25 PM

Send feedback

Back

User Profile

Technical name: **user_profile** • Asset Type: **RedshiftTableAssetType**

No description

ACTIONS

Your subscription request has been submitted. You will be able to access the data once the request is approved

BUSINESS METADATA | SCHEMA | MY SUBSCRIPTION

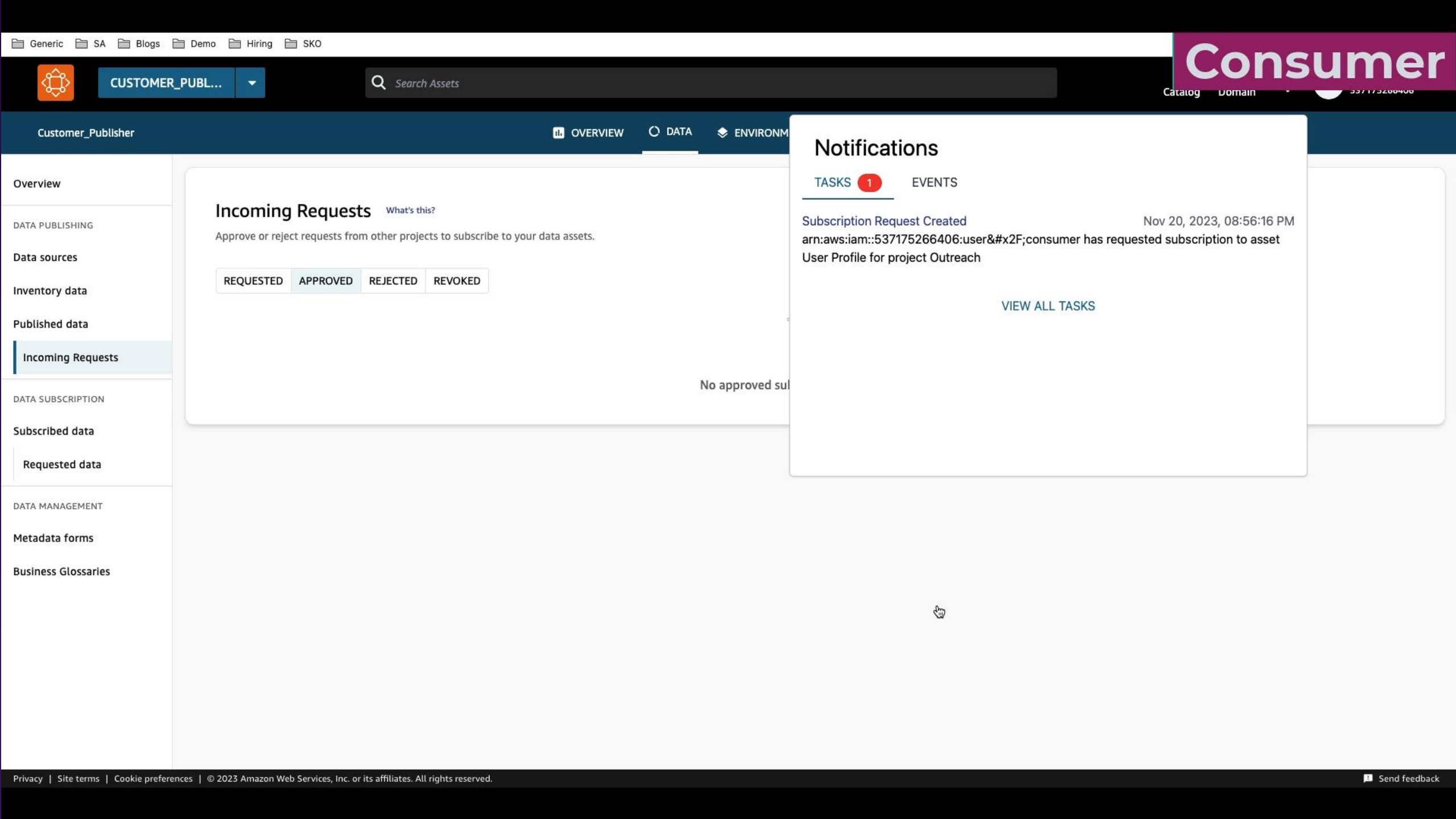
REQUEST DETAILS

SUBSCRIBER
Outreach

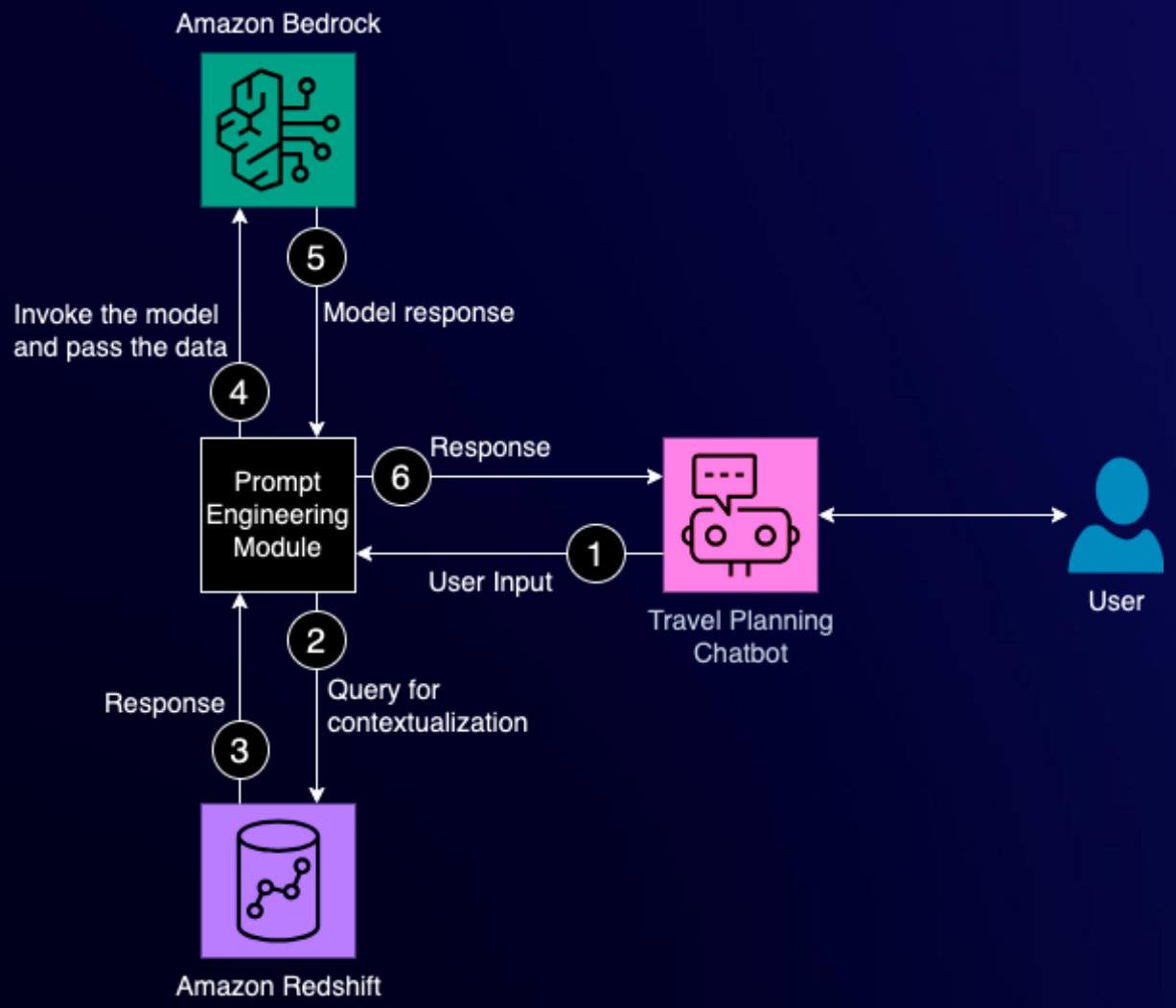
REQUESTOR
consumer

REQUEST DATE
11/20/2023, 8:56:14 PM

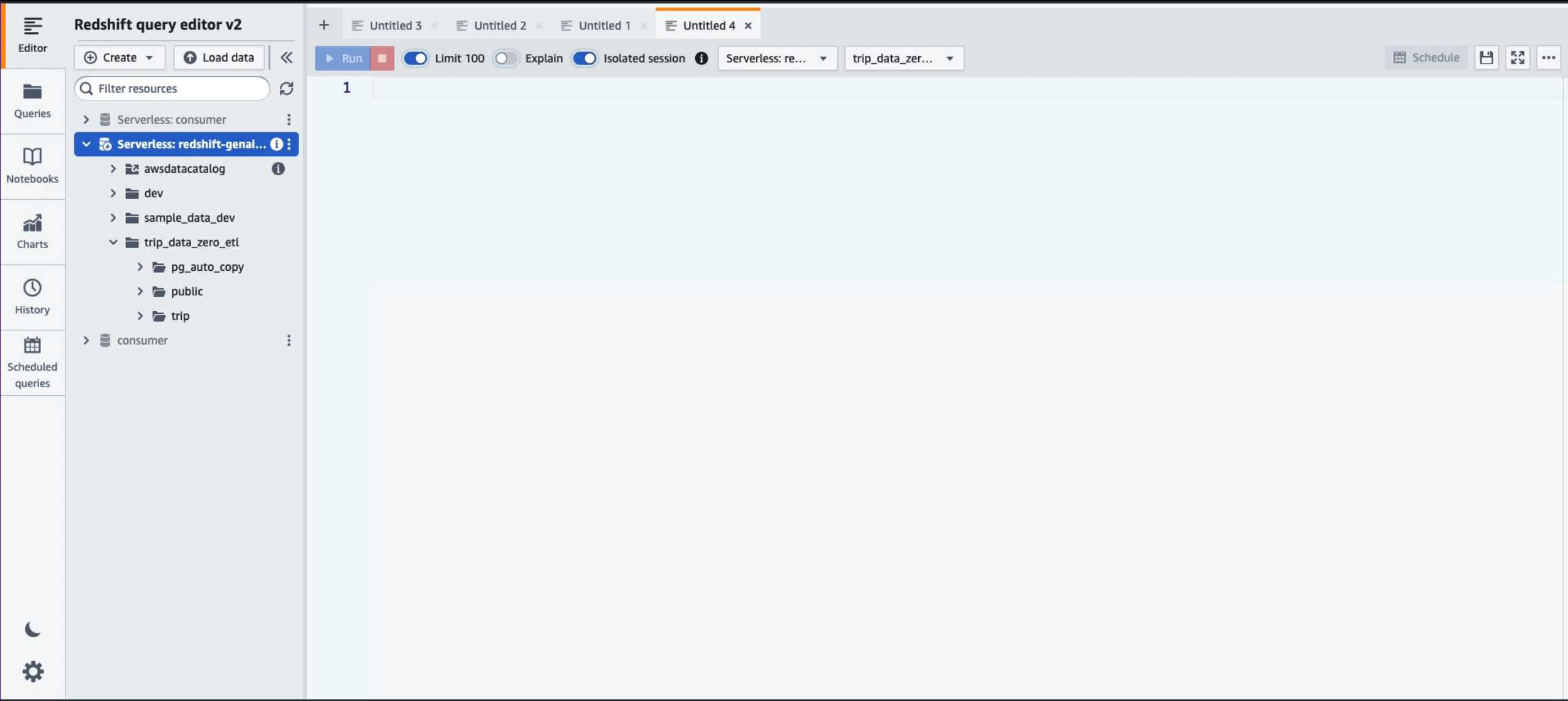
REASON FOR ACCESS
For analysis


Generic
SA
Blogs
Demo
Hiring
SKO
Consumer
CUSTOMER_PUBL...
Search Assets
Catalog
Domain
Customer_Publisher
OVERVIEW
DATA
ENVIRONM
Overview
DATA PUBLISHING
Data sources
Inventory data
Published data
Incoming Requests
DATA SUBSCRIPTION
Subscribed data
Requested data
DATA MANAGEMENT
Metadata forms
Business Glossaries
Incoming Requests
What's this?
Approve or reject requests from other projects to subscribe to your data assets.
REQUESTED
APPROVED
REJECTED
REVOKED
No approved sul
Notifications
TASKS 1
EVENTS
Subscription Request Created
Nov 20, 2023, 08:56:16 PM
arn:aws:iam::537175266406:user/consumer has requested subscription to asset User Profile for project Outreach
VIEW ALL TASKS
Privacy | Site terms | Cookie preferences | © 2023 Amazon Web Services, Inc. or its affiliates. All rights reserved.
Send feedback

Demo: Generative AI application

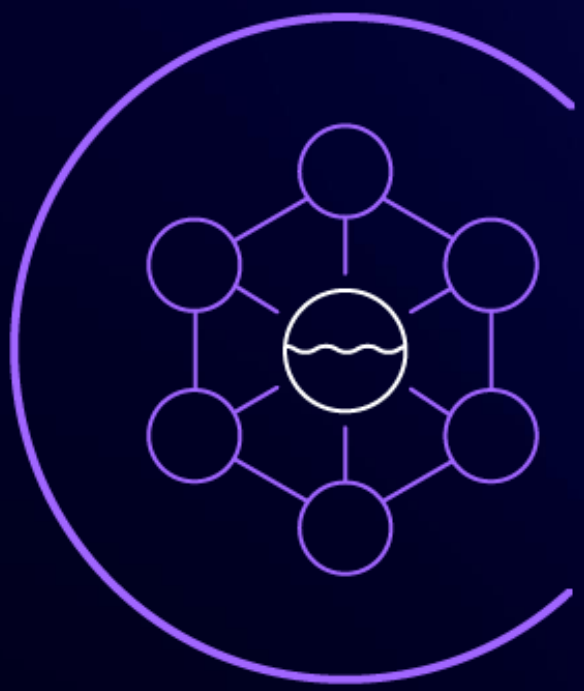


Reference Blog


Editor
Queries
Notebooks
Charts
History
Scheduled queries
Redshift query editor v2
Create
Load data
Filter resources
Serverless: consumer
Serverless: redshift-genai...
awsdatacatalog
dev
sample_data_dev
trip_data_zero_etl
pg_auto_copy
public
trip
consumer
Untitled 3
Untitled 2
Untitled 1
Untitled 4
Run
Limit 100
Explain
Isolated session
Serverless: re...
trip_data_zer...
Schedule
1
CloudShell
Feedback
© 2023, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

End-to-end demo…

Producers



Build a curated data set using a ZETL integration between Aurora MySQL and Amazon Redshift

Technology

Share the curated data set across your organization

Consumers

Build a generative AI application using the curated data set

Agenda

Why do you need an end-to-end data strategy?

How does AWS help you build your end-to-end data strategy?

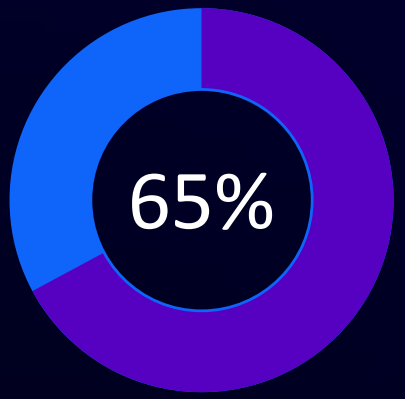
Demo of building an end-to-end system with AWS

How Fannie Mae build their end-to-end data strategy using AWS

Conclusion

Fannie Mae, America's valued housing partner

Our mission: To facilitate equitable and sustainable access to homeownership and quality, affordable rental housing across America



- **Homeownership in United States**
- **~$1.4T liquidity in 2020**

~1 in 4

Homes financed by Fannie Mae

The data: Everything you can think of that is submitted with a mortgage, including large amounts of personally identifiable information (PII)

Fannie Mae is an integral part of the US housing market

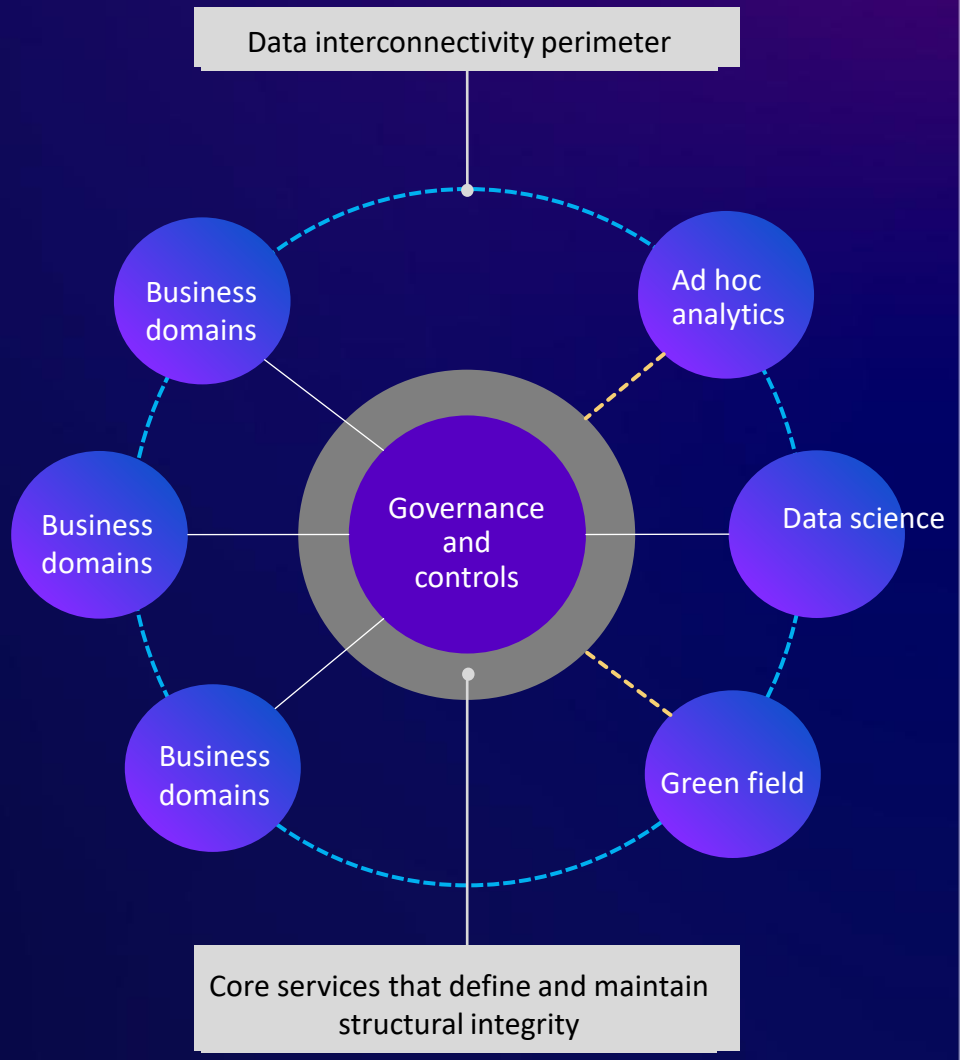
Implementation at Fannie Mae

Fannie Mae has incorporated the data mesh principles in its data strategy **well before the term "data mesh" became a buzzword in the industry**

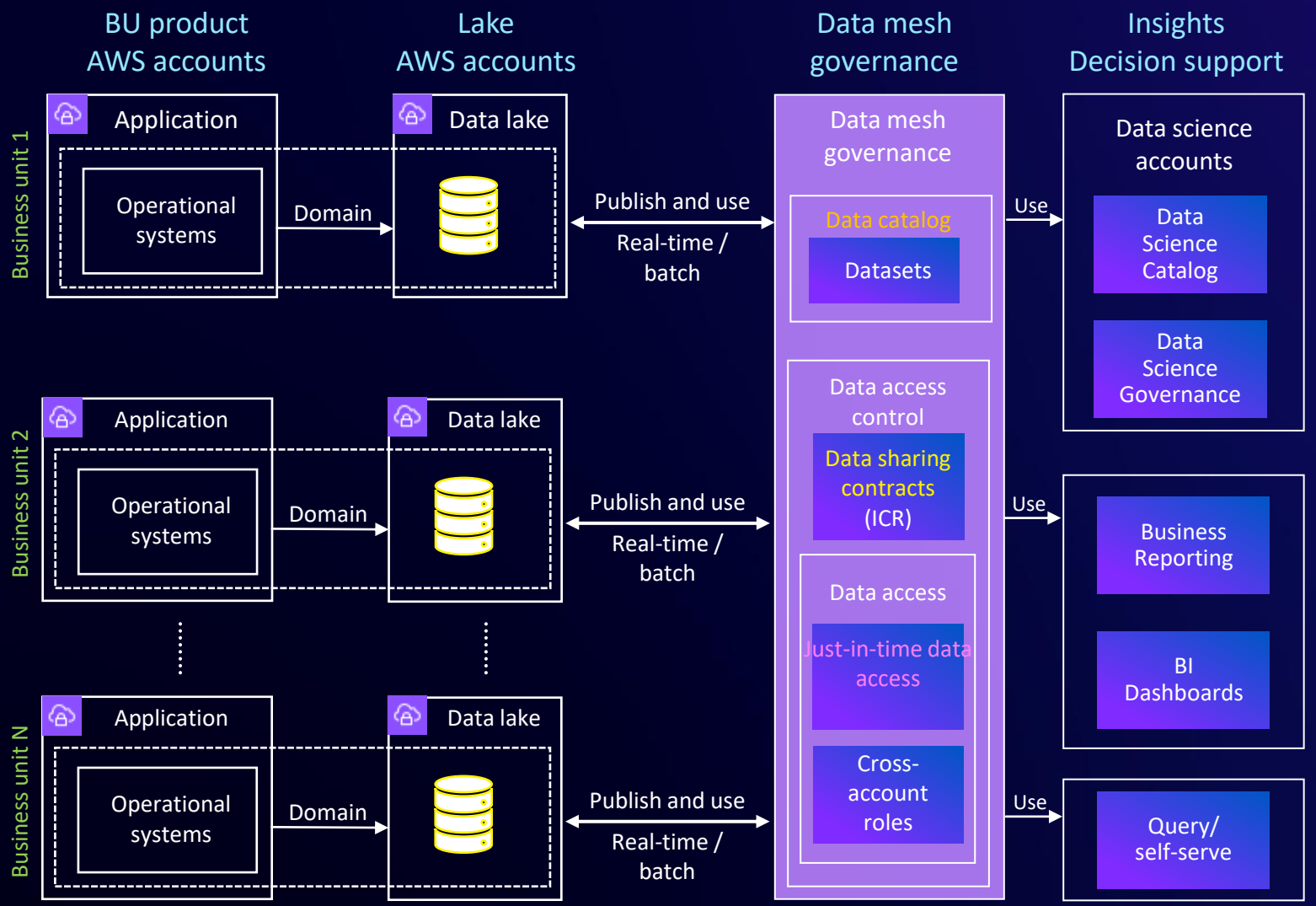
Enterprise data lake (EDL) at Fannie Mae is a multi-account/ hub-and-spoke-based distributed data management architecture – promote data-as-a-product concept

While the hub enables a **centralized oversight and governance controls**, the spokes allow for a **BU-specific data hydration**, management policies

Centralized oversight helps Fannie Mae stay in adherence to **strict compliance and interoperability standards** while ensuring the data is discoverable

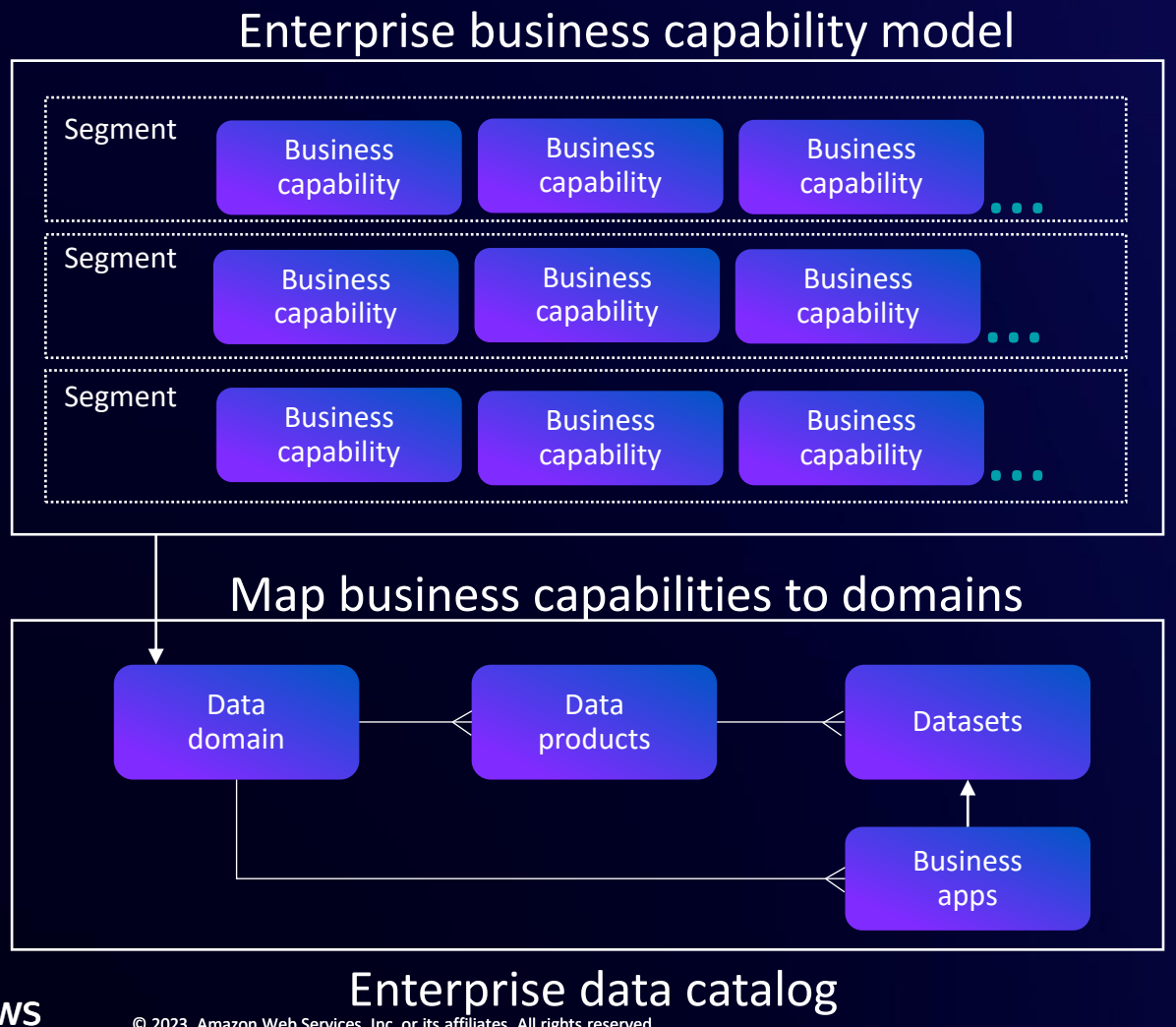


Data mesh architecture



- Multi-account AWS structure to segregate business units
- Business units manage their data, code, and infrastructure
- Self-service data infrastructure for data producers and consumers
- Enterprise wide data catalog for data discovery and access
- Data sharing contracts enforce governance and provide lineage
- One-click data access using just-in-time data access system

Recommendations for data modeling and domain owners



Features

Promote data-as-a-product thinking

Establish ownership and accountability by business domains

Use business capabilities to establish bounded context

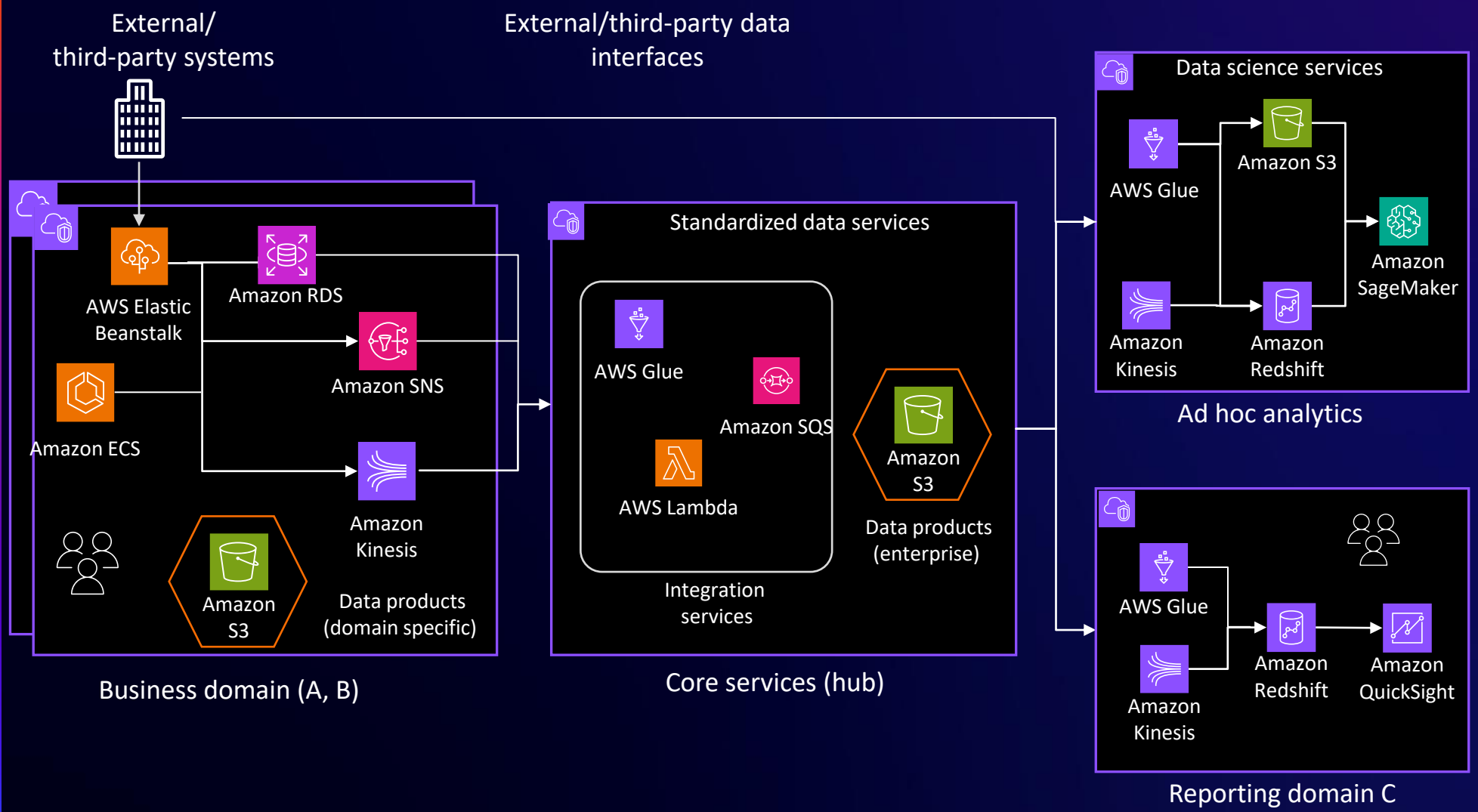
Key responsibilities of domain owner

Enable data products availability and empower ownership of data and its metadata

Serve as a go-to person for consumers' data engagement needs

Catalog datasets and enrich them with business context

Use case: Relationship between the domains



Features

Decouple management of data and technology by embracing data mesh concepts

Integrate data products across domains for seamless interoperability

Aligned to ELDM standards – discoverable through common vocabulary, regardless of the physical model

Associate ownership, access management, and metadata at the product level

Benefits

Fannie Mae

Agility and scalability

Empowers decentralized data operations, enhancing speed to market, scalability, and business domain agility

Faster access to critical data

Enables easy access and time to market through a decentralized and self-service infrastructure

Visibility for cross-functional utilization

Optimizes data management by decentralizing data ownership and distribution among cross-functional domains

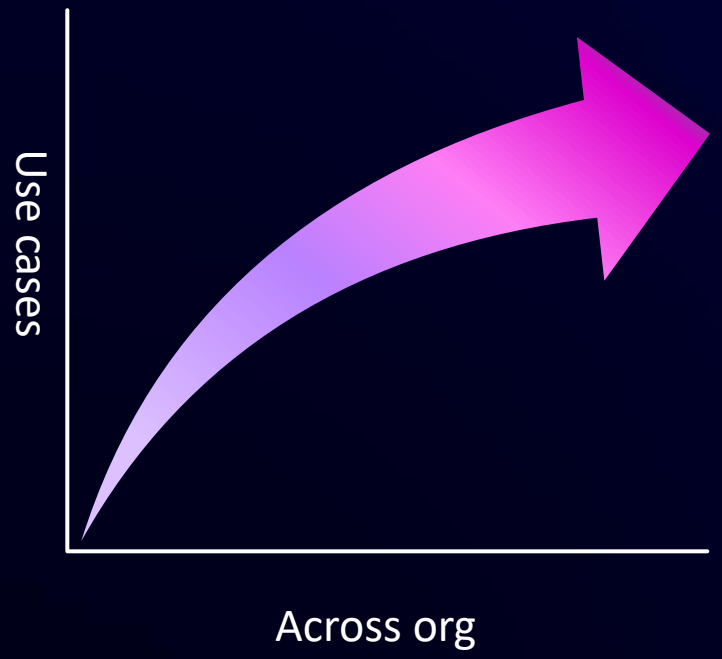
Independence

Prevents being locked into a single data platform or product

aws

What's ahead?

Drive adoption across org and use-cases



Evaluate new capabilities to simplify architecture

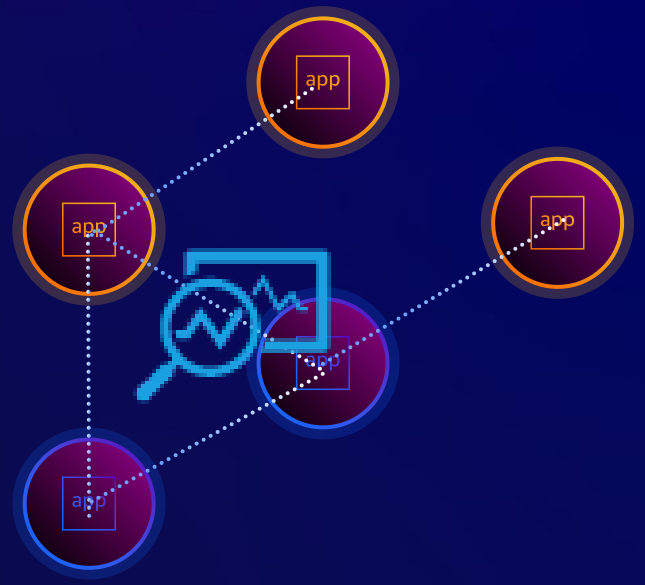
Amazon DataZone

AWS Glue

Amazon Redshift ML

Amazon Bedrock

Improve governance with zero-trust capabilities



Agenda

Why do you need an end-to-end data strategy?

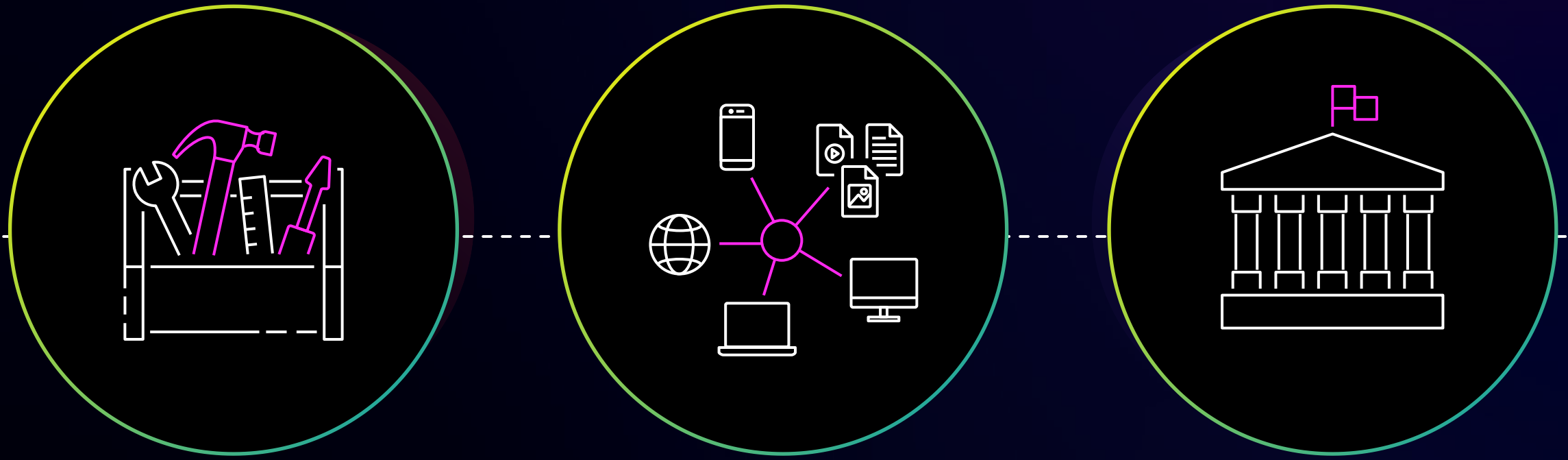
How does AWS help you build your end-to-end data strategy?

Demo of building an end-to-end system with AWS

How Fannie Mae build their end-to-end data strategy using AWS

Conclusion

Break down data silos and build solid foundations



Data silos

Utilize

People silos

Integrate

Business silos

Govern

You’ve just completed an AWS Analytics Superheroes session

Scan QR code to learn more about the AWS Analytics Superheroes

Thank you!

Please complete the session survey in the mobile app