AI Conclave

Online

AIOT303

Building data foundations for generative AI software companies

Madhavi Watve
Senior Solutions Architect
AWS India

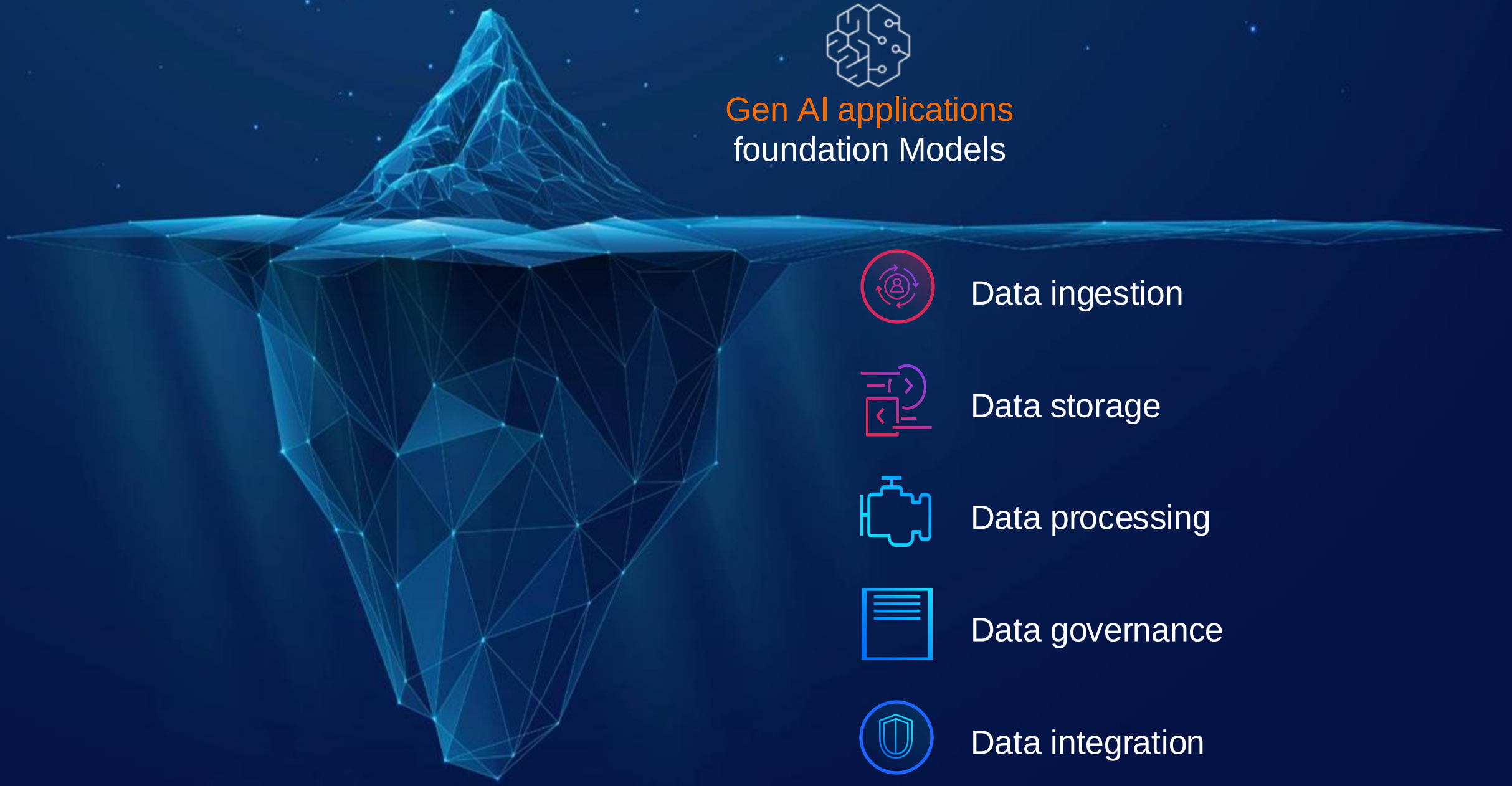
Agenda

- What is data foundation for generative AI?
- Why AWS?
- Generative AI applications patterns

Generative AI

"Organizations are not just talking about generative AI – they're investing time, money and resources to move it forward and drive business outcomes,"

"Generative AI is now on CEOs' and boards' agendas as they seek to take advantage of the transformative potential of this technology."


Gen AI applications
foundation Models
Data ingestion
Data storage
Data processing
Data governance
Data integration

AWS services to build data pipeline for generative AI

Data ingestion

AWS Database Migration Service (AWS DMS)

Amazon Kinesis

Amazon Managed Streaming for Apache Kafka (Amazon MSK)

Amazon AppFlow

AWS Storage Gateway

Zero-ETL integrations

NEW

Amazon SageMaker Lakehouse

Amazon Redshift

Zero-ETL: fully managed integrations by AWS

AWS data sources

- Amazon DynamoDB
- Aurora PostgreSQL
- Aurora MySQL
- RDS for MySQL

Applications

- Salesforce
- SAP
- ServiceNow
- Salesforce Pardot
- Zendesk
- Instagram Ads
- Facebook Ads
- Zoho CRM

Data storage

Data lake and LakeHouse

Amazon Simple Storage Service (Amazon S3)

Amazon S3 Tables

Amazon Redshift

Amazon SageMaker Lakehouse

Databases – Relational and NoSQL

Amazon Relational Database Service (Amazon RDS)

Amazon MemoryDB for Redis

Amazon DynamoDB

Amazon OpenSearch Service

Amazon Aurora

Amazon DocumentDB (with MongoDB compatibility)

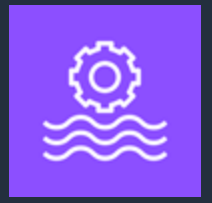
Amazon Neptune

Data processing and preparation

Data processing

AWS Glue

Amazon EMR



Amazon Managed Service for Apache Flink

Data Catalog

Crawler

AWS Glue Data Catalog

Data preparation Using AIML

Amazon SageMaker Ground Truth

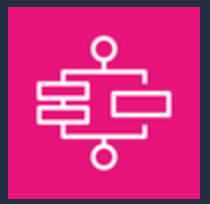
Amazon Textract

Amazon Transcribe

Amazon Translate

Data integration

Data workflows



AWS Step Functions

Amazon EventBridge

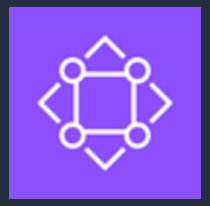
AWS Lambda

Generative AI data workflows

Amazon Bedrock Agents

Amazon Bedrock Knowledge Bases

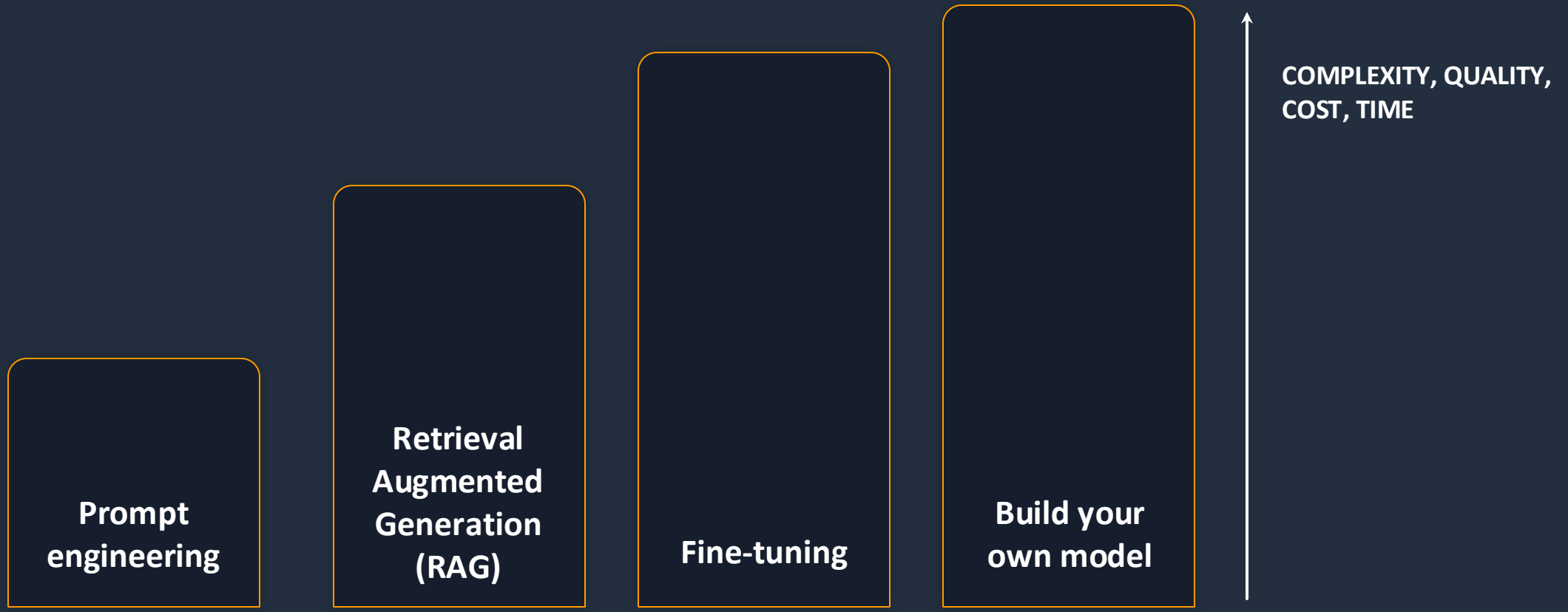
Data governance

Data governance in generative AI applications	 Amazon DataZone	 Amazon Bedrock Guadrails	 AWS Lake Formation
Overall data governance	 AWS CloudTrail	 AWS Control Tower	 AWS Identity and Access Management (IAM)

Generative AI application patterns

Using your data in generative AI applications

COMMON APPROACHES



Pattern 1: Prompt engineering

Review: “Earnings per share have beaten analyst expectations"

What is the sentiment?

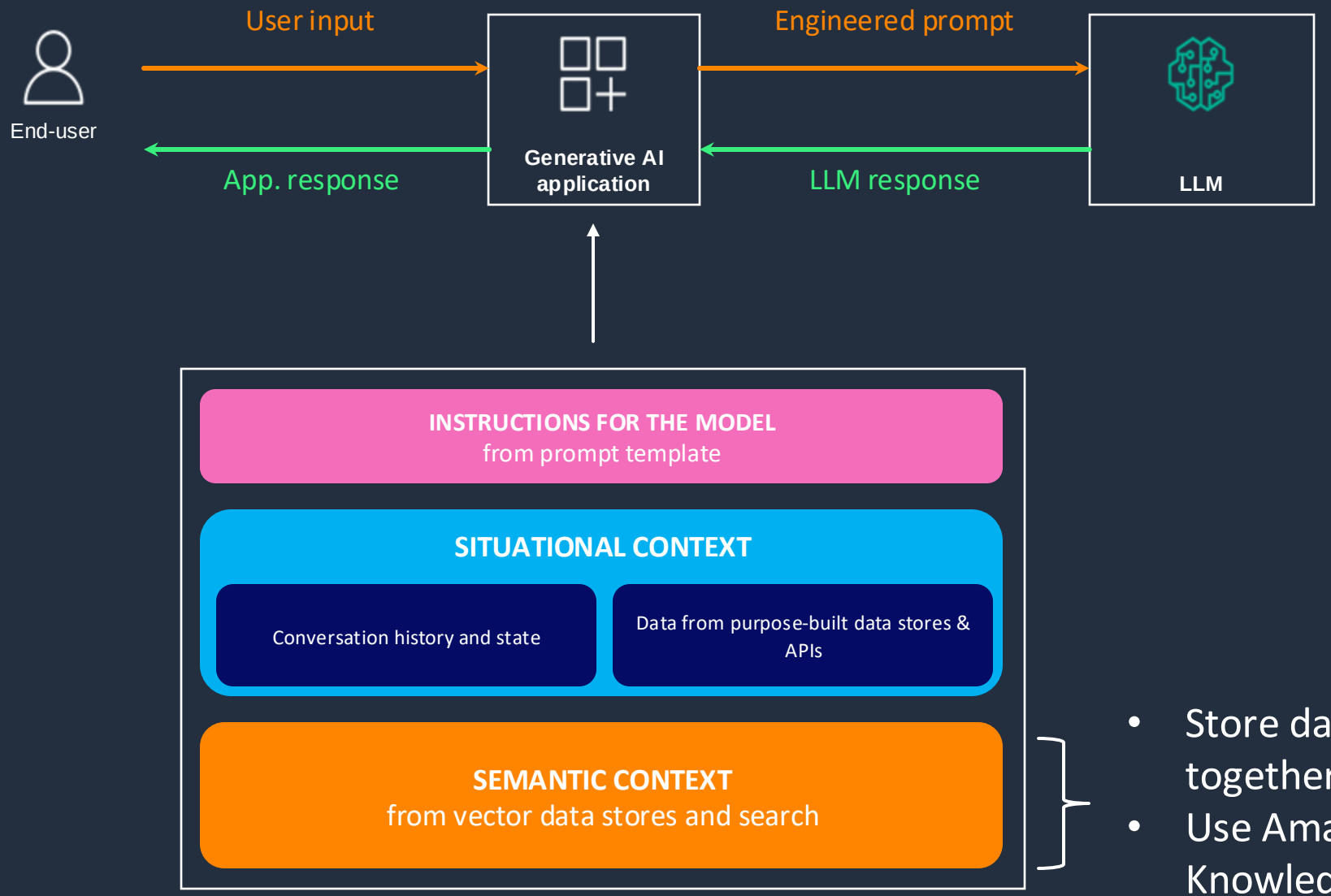
Input

Output

The text explains that earnings have been expectations, that is generally a good signal in financial reporting, therefore the review is positive.

Amazon Bedrock supports **intelligent prompt routing & prompt caching**

Pattern 2: Retrieval Augmented Generation (RAG)



- Store data and vectors together
- Use Amazon Bedrock Knowledge Bases

Storing data and vectors together

Use familiar tools that meet your requirements

Avoid additional licensing and management

Provide a faster experience to end users

Reduce the need for data sync and movement

Vector enabled data stores

Amazon OpenSearch Service

Amazon Aurora And RDS For PostgreSQL

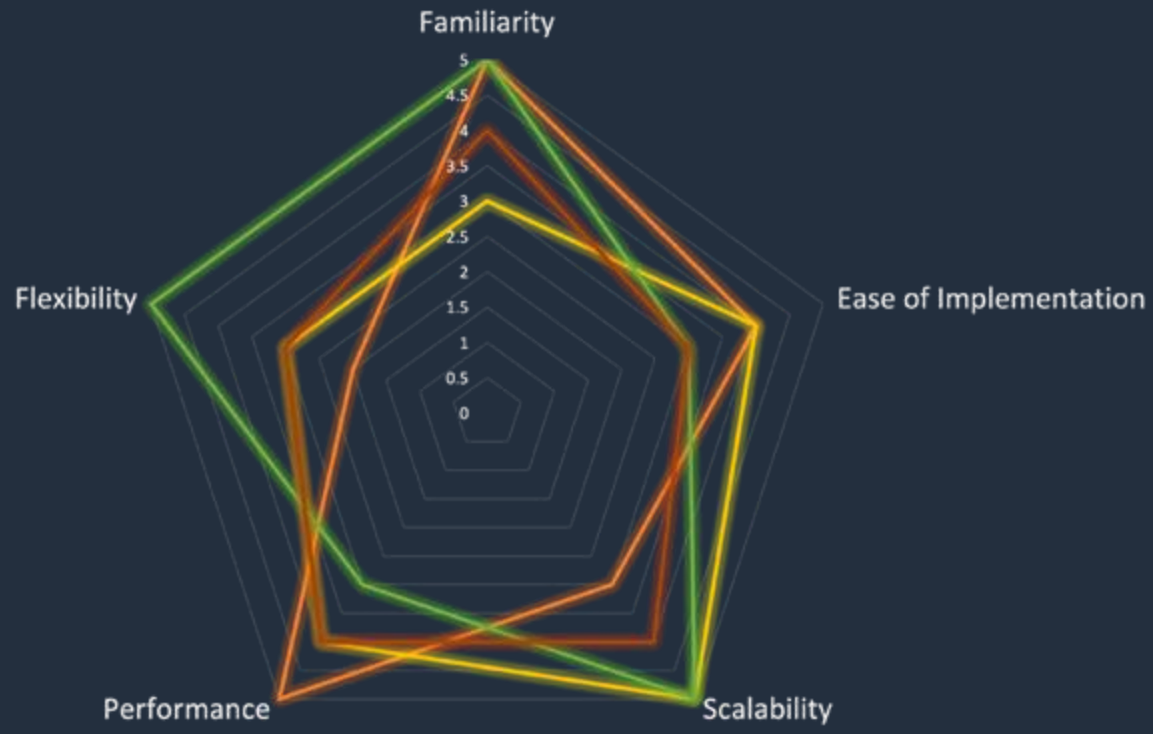
Amazon MemoryDB for Redis

Amazon DocumentDB (with MongoDB compatibility)

Amazon Neptune

Choosing the right vector data store

- Familiarity
- Ease of implementation
  - Abstractions
  - Integrations
- Scalability
  - Vector dimensions supported
  - Number of embeddings
- Performance
  - Queries per second (QPS)
  - Recall rate
- Flexibility



Amazon Bedrock Knowledge Bases

- Natively supported Knowledge bases:


Amazon OpenSearch Service


Amazon Aurora And RDS For PostgreSQL


Amazon MemoryDB for Redis


Amazon DocumentDB (with MongoDB compatibility)


Pinecone


Redis Enterprise Cloud


Amazon Neptune

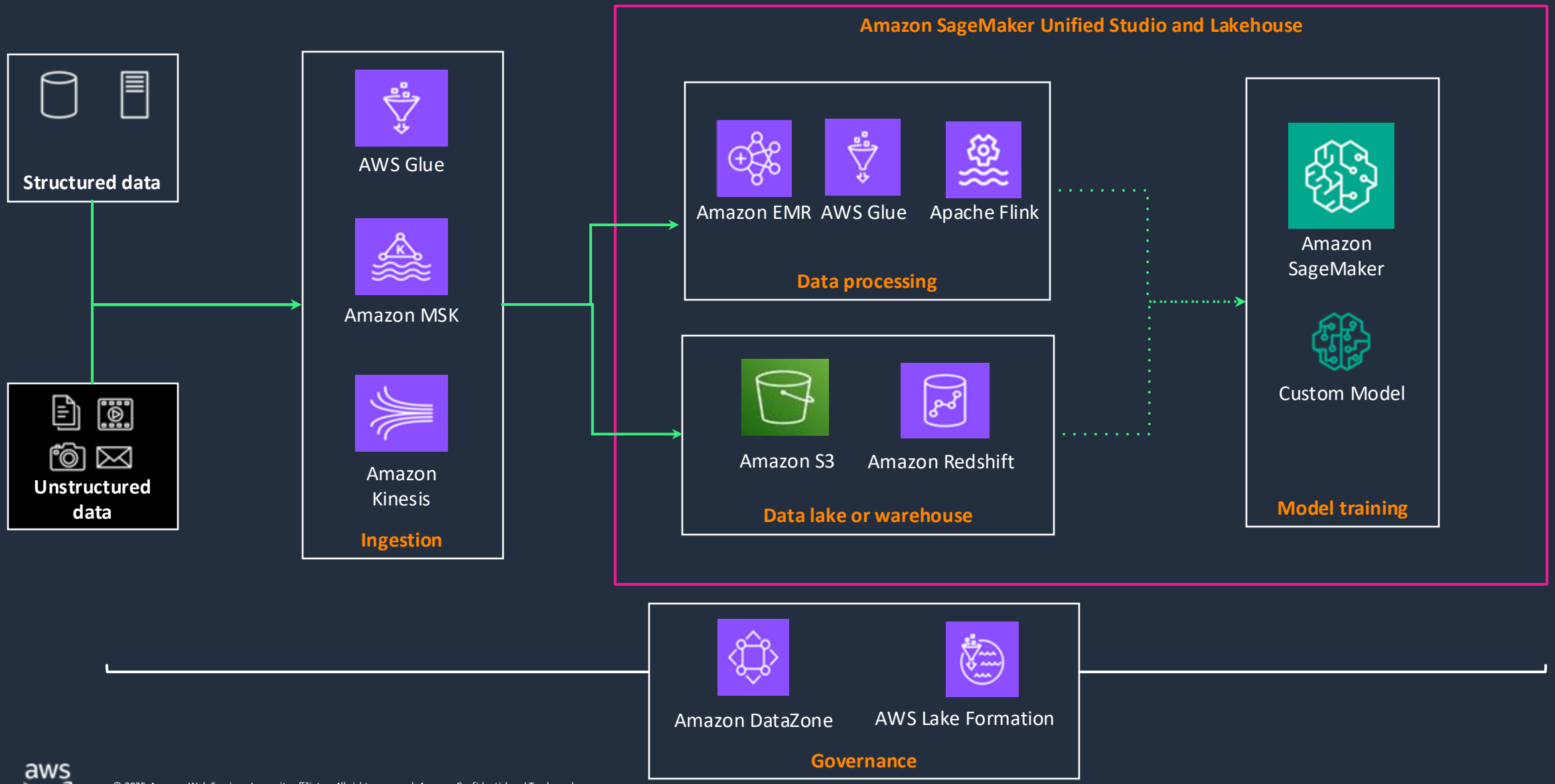

Amazon Kendra GenAI Index

- Streamlines the entire RAG workflow
- In-built support for session context management
- Source citations

Amazon Bedrock Knowledge Bases – New features

1) **Amazon Bedrock Data Automation (preview)**
   - Extract, transform and generate structured data from multi-modal content
   - unified, API-driven experience
2) **GraphRAG (preview)**
   - Generate knowledge graphs linking relationships across data sources
   - Enhance transparency of source information for better fact verification
3) **Integration with Amazon Kendra GenAI Index**
   - Build digital assistants and intelligent search experiences more efficiently and effectively
4) **Structured data retrieval**
   - Incorporate contextual information from a variety of structured data sources

Pattern 3: Model training/Fine-tuning



Amazon SageMaker AI

➢ Amazon SageMaker Unified studio (Preview)

- Build with all your data and tools for analytics and AI in a single environment

➢ Amazon SageMaker Lakehouse

- Unifies all your data across data lakes and data warehouses
- Access and query your data with all Apache Iceberg–compatible tools and engines
- zero-ETL : near real time ingestion from operational databases and applications
- Access and query data in-place with federated query capabilities

➢ Amazon SageMaker Catalog

- Securely discover, govern, and collaborate among data and apps

Summary

Data is the differentiator

Generic generative AI

Generative AI that knows your business and your customer

Putting it all together

Generative AI applications

Amazon Bedrock Knowledge Bases

Data processing and conversion across modalities

- Amazon Textract
- Amazon SageMaker Ground Truth
- Amazon Rekognition
- Amazon Translate
- Amazon Transcribe

Analytics

- Amazon Redshift
- Amazon EMR Spark, Hadoop
- AWS Glue Scala, Python
- Amazon Athena
- Amazon Managed Service Apache Flink

Databases with Vector capability

- Amazon Aurora PostgreSQL
- Amazon RDS PostgreSQL
- Amazon Kendra GenAI Index
- Amazon OpenSearch Service
- Amazon Neptune Graph
- Amazon DocumentDB with MongoDB
- Amazon MemoryDB for Redis

Data lake warehouse

- Amazon S3/Glacier
- Amazon Redshift
- Amazon SageMaker Lakehouse

Data governance

- Lake Formation
- DataZone

Data Movement

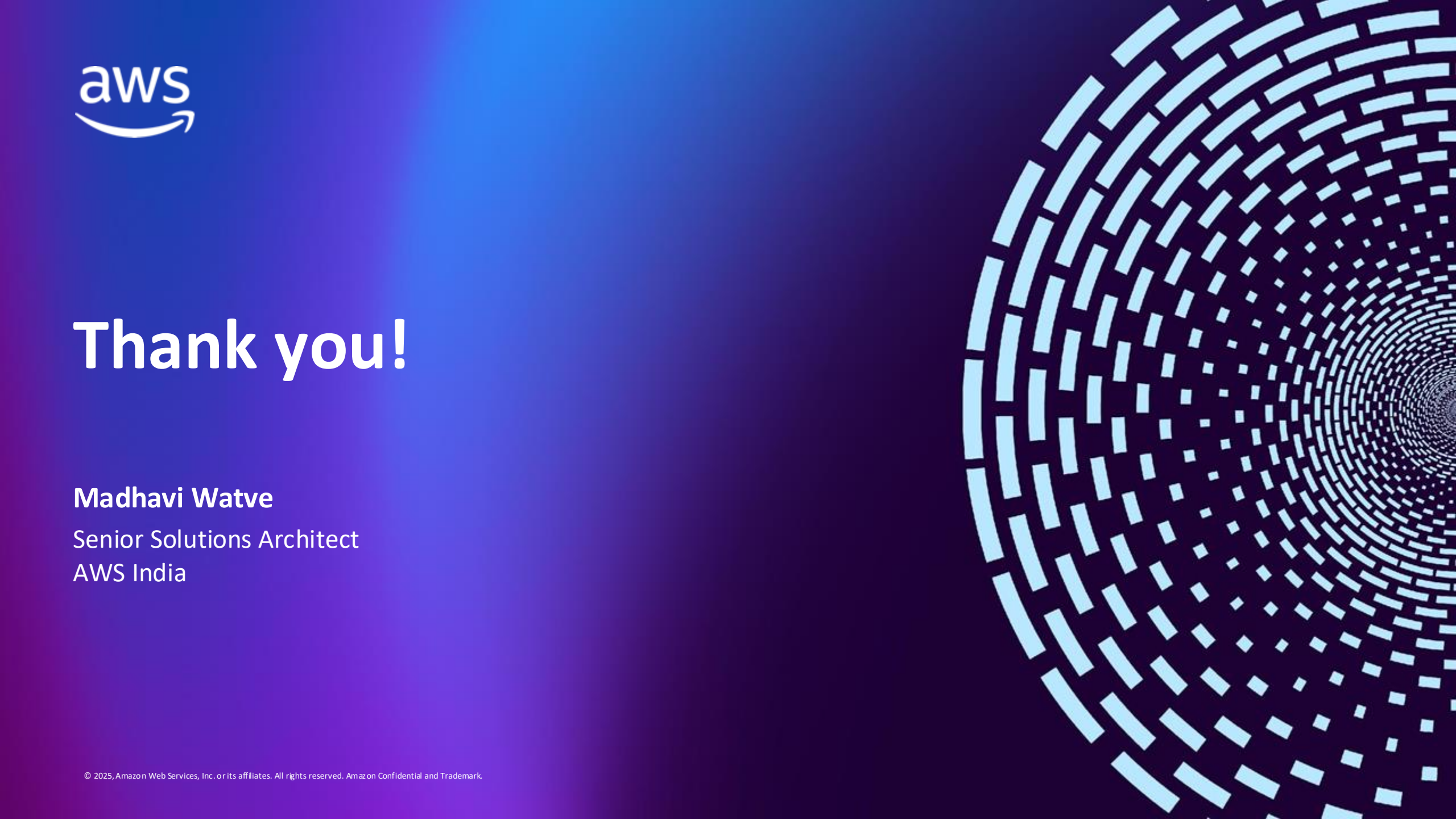
Database Migration Service | Zero-ETL Integrations | Kinesis Data Firehose | Kinesis Data Streams | Managed Streaming for Kafka

Amazon SageMaker Unified Studio

Amazon SageMaker Lake house

Amazon SageMaker Feature store

Amazon SageMaker Ground Truth



Thank you!

Madhavi Watve
Senior Solutions Architect
AWS India