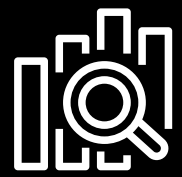
# Analytics Deep Dive: Observability, ETL, Big Data and Governance with AWS

Imtiaz (Taz) Sayed (He/Him)

Head of WW Data Analytics Community
AWS



# Amazon OpenSearch Service

## Observability and Operational Analytics

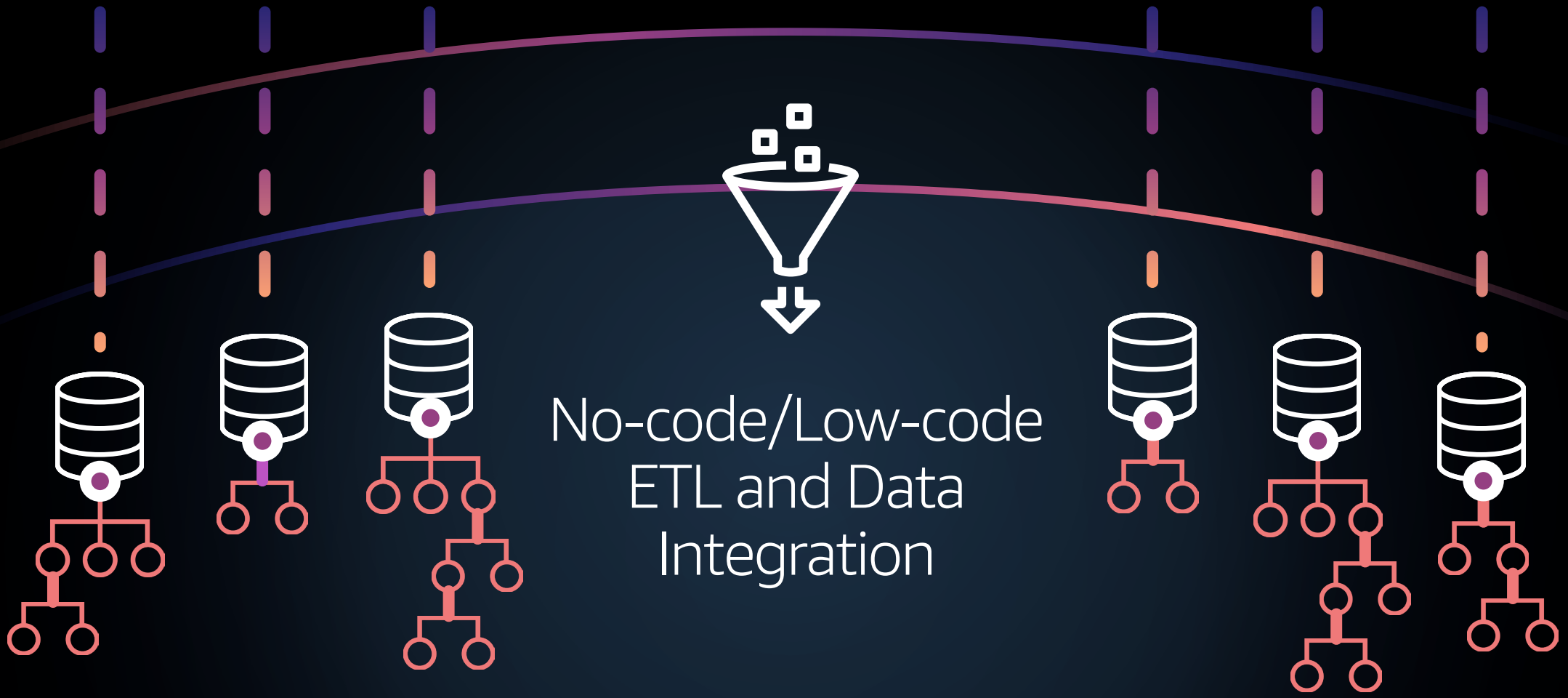
Fully managed

Log and search analytics

Detect potential threats

Cost effective

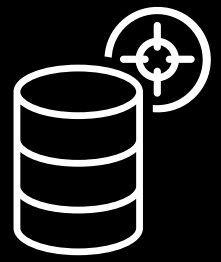
# AWS **Glue**



No-code/Low-code ETL and Data Integration

# AWS **Lake Formation**

## DATA GOVERNANCE



Simplified S3 policies

Fine grained permissions

Cross-account shares

# Observability and operational analytics with AWS Workshop

Raghavarao Sodabathina(He/Him)

Enterprise Solutions Architect
AWS

# Amazon OpenSearch Service

- A managed service that makes it easy to **deploy, operate,** and **scale OpenSearch** and legacy Elasticsearch clusters in the AWS Cloud
- Built-in **observability tooling** such as **trace analytics, event analytics,** and **log analytics** that help you gain insights from **logs, metrics** and **traces**
- Comes with **anomaly detection** features that detect abnormal events in your data and integrate with **built-in alerting**
- **Integrated with** Amazon Kinesis Data Firehose, Amazon MSK, Amazon CloudWatch Logs, and other tooling

# Workshop Setup

1. Access the lab guide: https://catalog.us-east-1.prod.workshops.aws/workshops/f0213896-4dd9-494a-89c5-f7886b45ed4a/en-US/

2. Navigate to https://dashboard.eventengine.run/ and use the event hash: XXXX for AWS led events

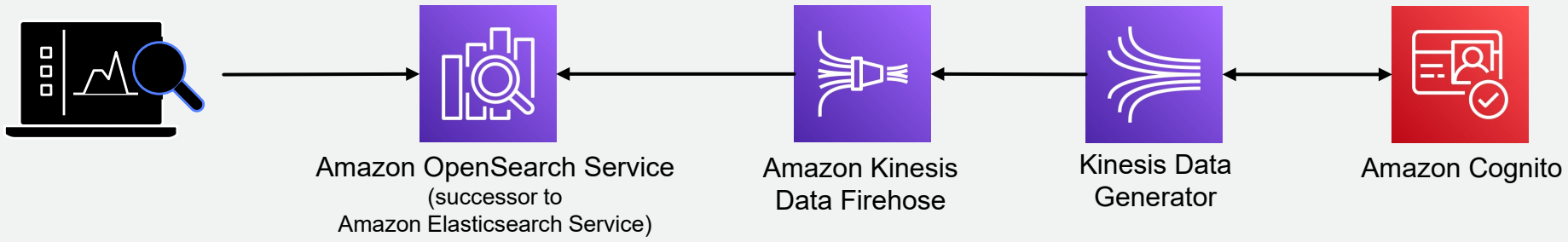
3. You can also complete this workshop on your own AWS account by following above lab guide.

# Workshop Architecture

Kinesis Data Generator (KDG)

OpenSearch Dashboards for visualization

Test streaming anomaly detection



# KDG template

Simulates Apache web logs

Time-based data

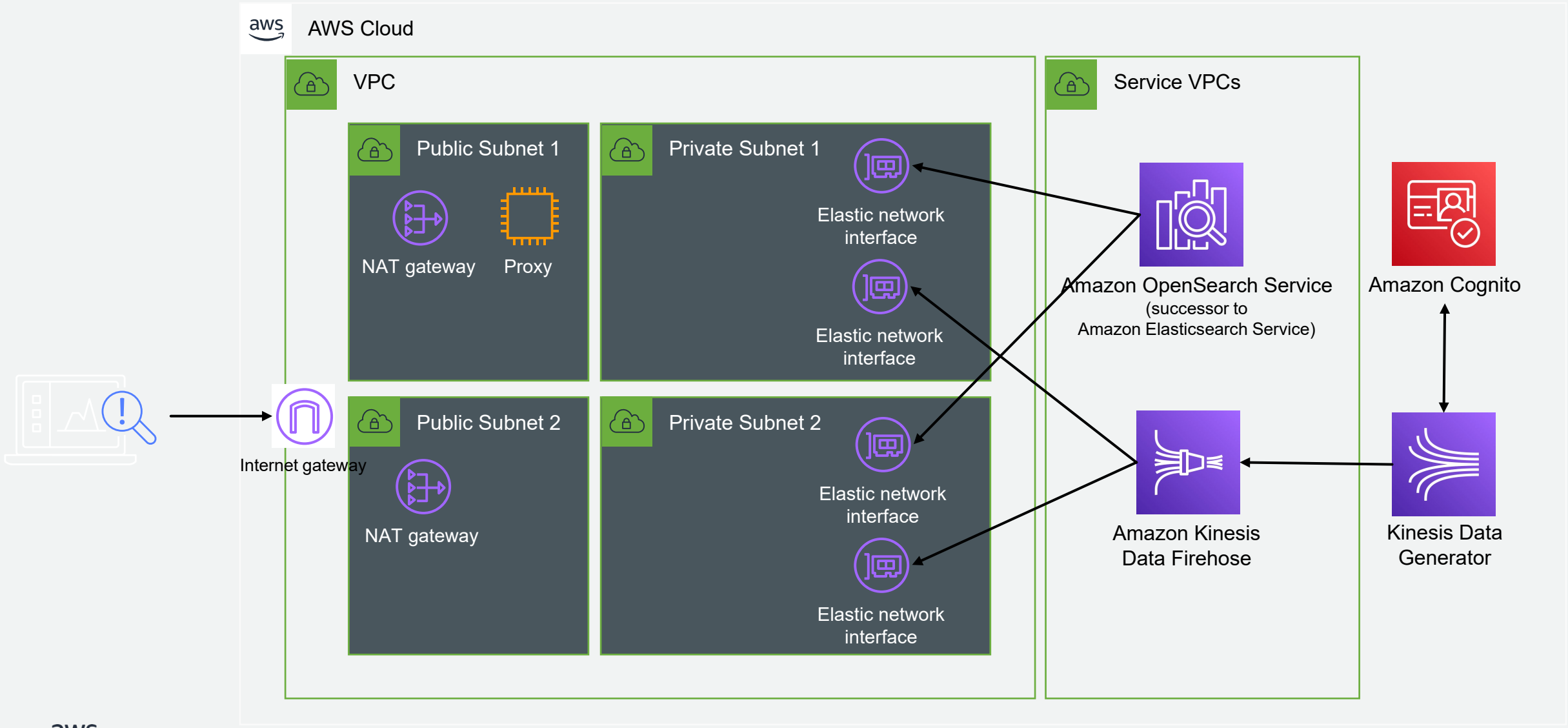
Randomized REST methods

Randomized status codes

Simulated geo information

```
{
    "host": "{{internet.domainName}}",
    "ident": "-",
    "auth": "-",
    "@timestamp": "{{date.utc("YYYY-MM-DDTHH:mm:ss")}}",
    "bytes": {{random.number(1000000)}},
    "verb": "{{random.weightedArrayElement(
        {
            "weights": [0.8,0.1,0.1],
            "data": ["GET","PUT","POST"]
        }
    )}}",
    "url": "{{internet.url}}",
    "http": "HTTP/1.1",
    "status": {{random.weightedArrayElement(
        {
            "weights": [0.9,0.05,0.05],
            "data": ["200","404","503"]
        }
    )}},
    "agent": "{{internet.userAgent}}",
    "clientip": "{{internet.ip}}",
    "geo": {
        "srcdest": "{{address.countryCode}}:{{address.countryCode}}",
        "src": "{{address.countryCode}}",
        "dest": "{{address.countryCode}}",
        "coordinates": {
            "lat": "{{address.latitude}}",
            "lon": "{{address.longitude}}"
        }
    }
}
```

# Lab Architecture



# Resources

**Stay up to date!**
www.aws.amazon.com/opensearch-service

**Contribute to OpenSearch project**
opensearch.org

**Amazon OpenSearch Service Immersion Days**
Deep dive through online training and hands-on labs.
searchservices-ww-gtm@amazon.com


**Roadmap**
OpenSearch

**Documentation**
Developer Guide

**Partners**
Get help from AWS Partner Network

**Stay connected**

Twitter: @OpenSearchProj
https://twitter.com/OpenSearchProj

LinkedIn:
https://www.linkedin.com/company/opensearch-project

Facebook:
https://www.facebook.com/OpenSearchProject

Twitch:
https://www.twitch.tv/OpenSearchProject

Meetup:
https://www.meetup.com/OpenSearch/

# Lake Formation Demo

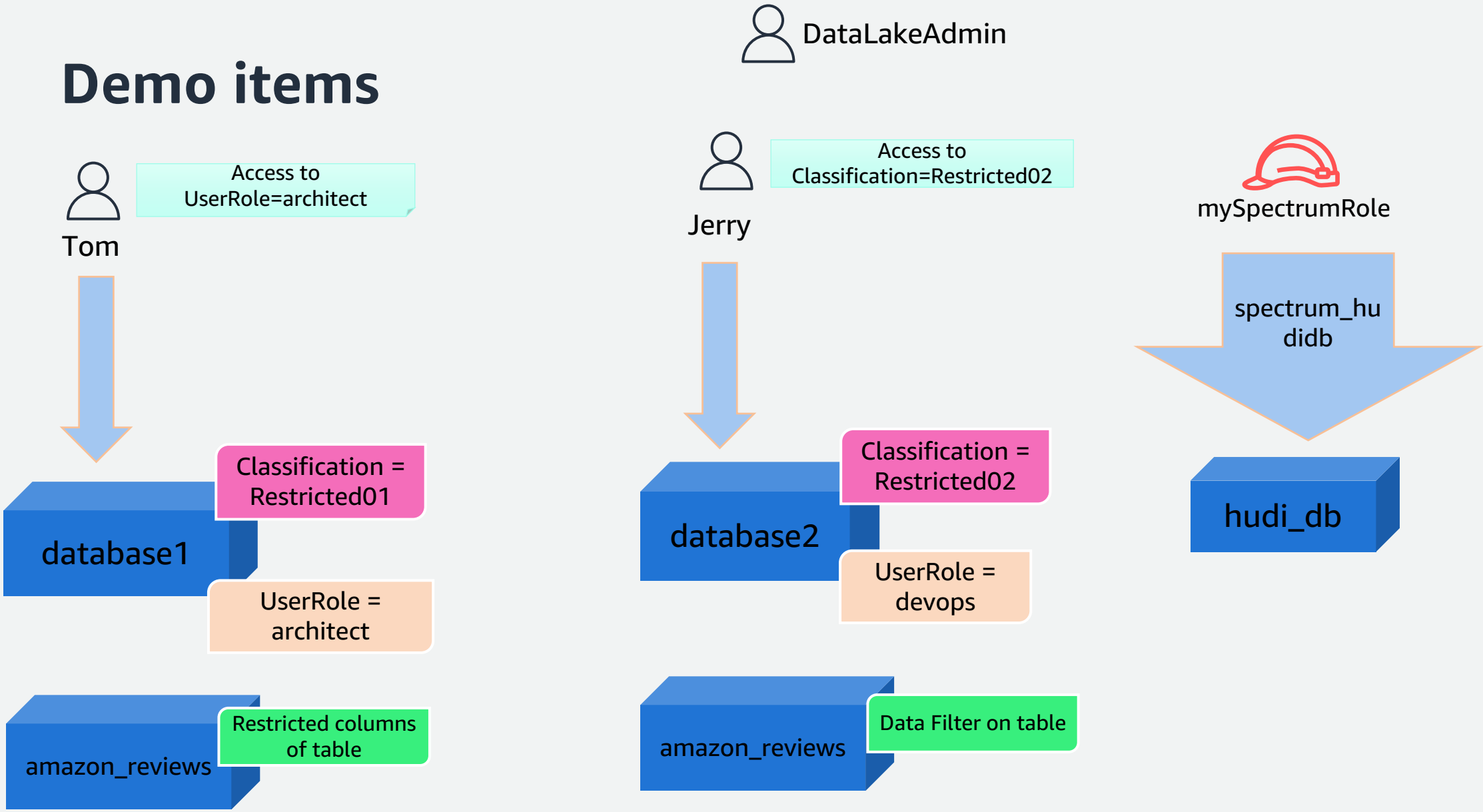
Aarthi Srinivasan

Senior Big Data Architect
AWS Lake Formation

# Demo agenda

- Table permissions using named resource method.
  - Restricted column permissions
  - Data filters
- LF-Tags based permissions
- Show Permissions using Amazon Athena
- Show Permissions using Redshift Spectrum

# Demo items



# AWS Glue - Data Integration for every user

Ramkumar Nottath (He/Him)

Sr. Solutions Architect – Analytics
AWS

# AWS Glue
# Serverless Data Integration for complex workloads

## Serverless
No infrastructure to maintain. Allocate needed compute power and run jobs

## Data Integration for every user
Development environments catered to different skillsets - visual ETL development for Data Engineers, notebook styled development for Data Scientists, and no code development for Data Analysts

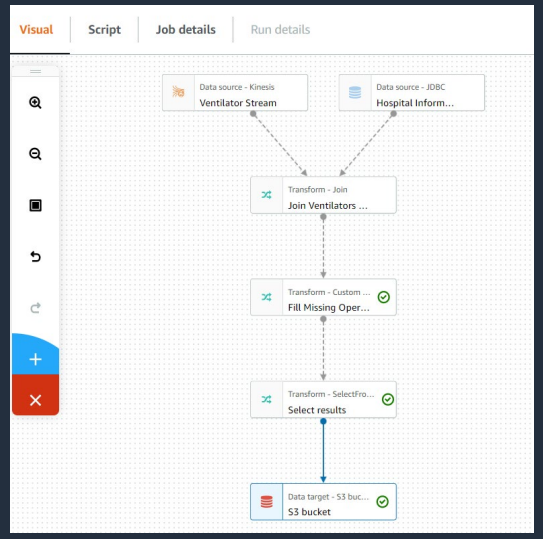
## Cost-effective
All-in-one pricing model is 55% cheaper than other cloud data integration options

## Handles complex workloads
Connect to various data sources, process petabytes of data in real-time, includes batch and event driven modes

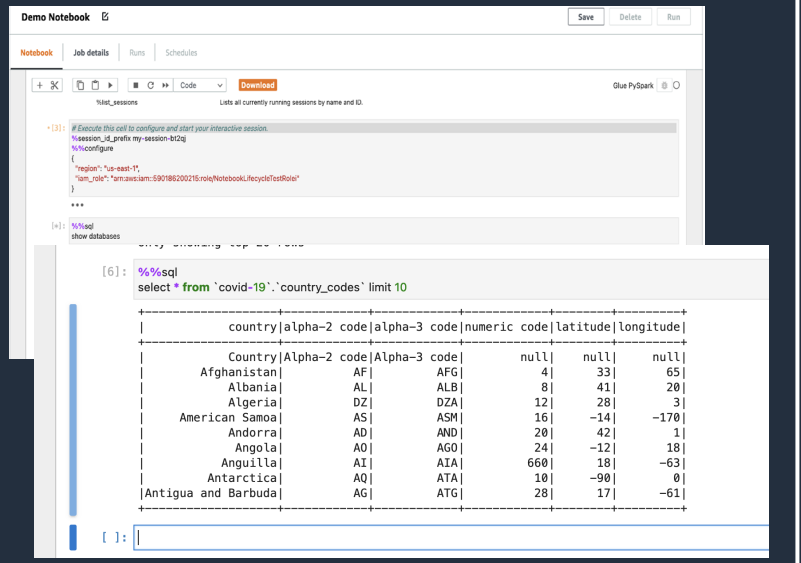
## No lock-in
Develop data integration pipelines in open source SparkSQL, PySpark, and Scala

# Built for multiple personas

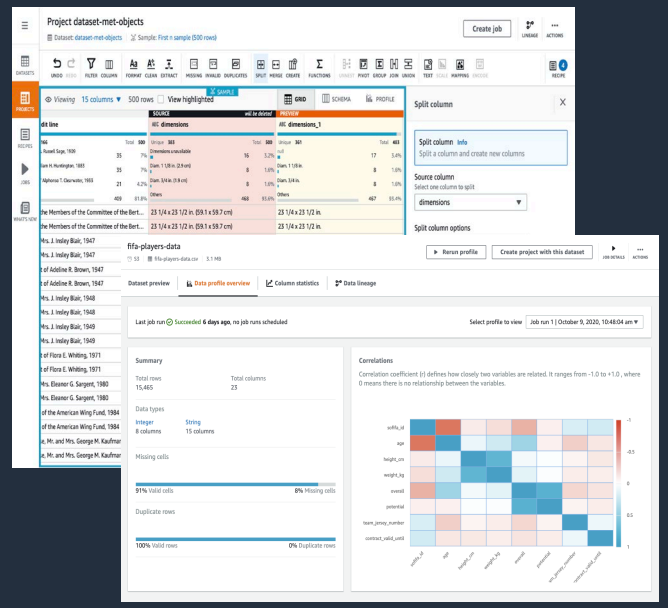
## Customers migrate to promote self service




Glue Studio for Data Engineers who prefer visual low-code development experience




Glue Notebooks & interactive sessions for Data Engineers who prefer code based experience




Glue Databrew for Analysts who prefer a no-code experience

# Thank you!