



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

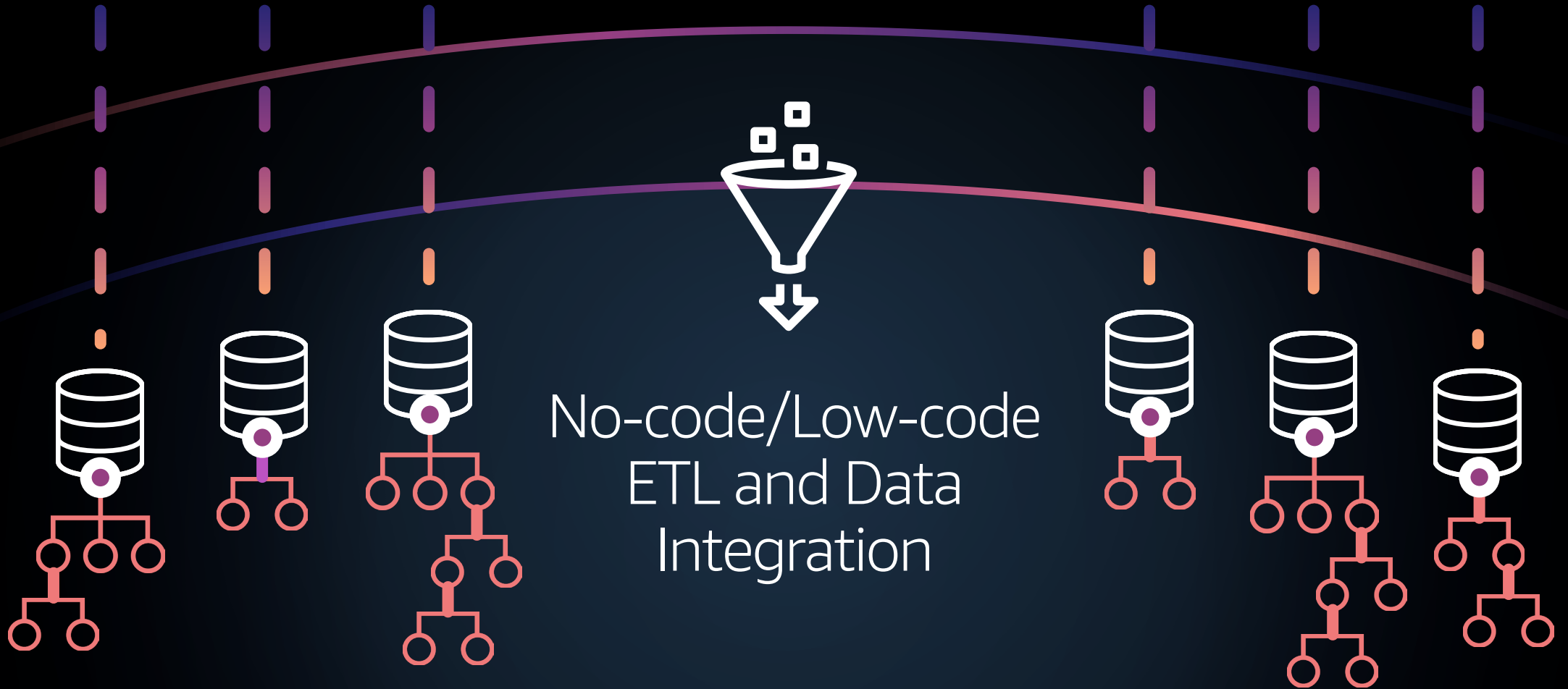
Fully managed

Log and search analytics

Detect potential threats

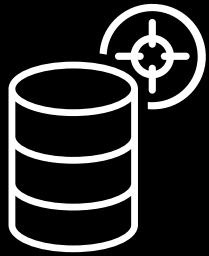
Cost effective

AWS Glue



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

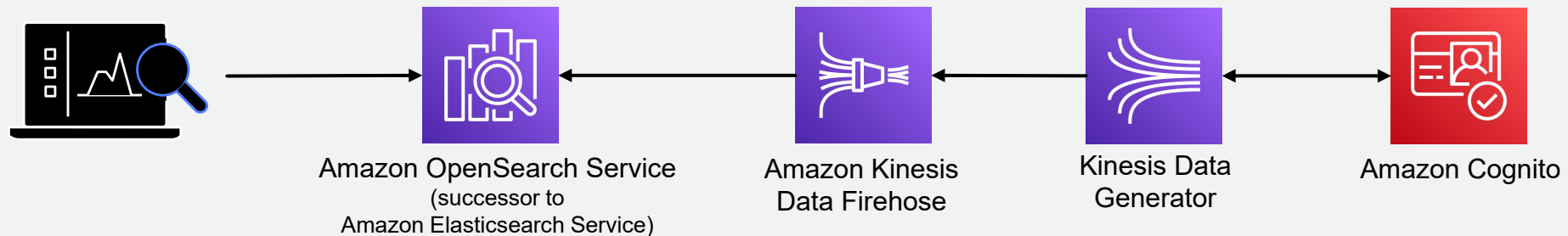
1. Access the lab guide: <https://catalog.us-east-1.prod.workshops.aws/workshops/f0213896-4dd9-494a-89c5-f7886b45ed4a/en-US/>
2. Navigate to <https://dashboard.eventengine.run/> and use the event hash: XXXX for AWS led events
3. You can also complete this workshop on your own AWS account by following 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

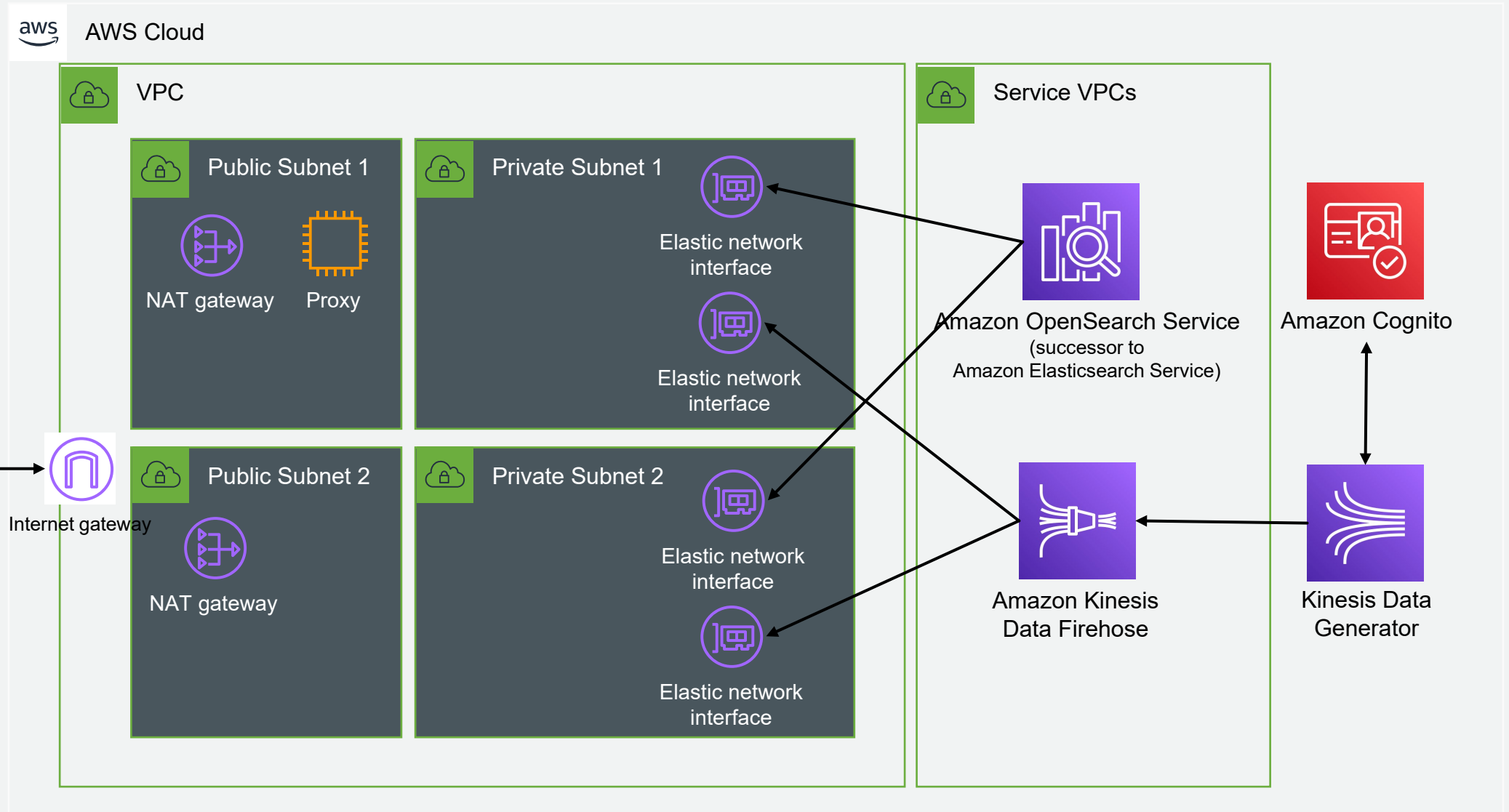
Randomized status codes

Simulated geo information

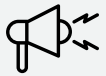


```
"host": "{{internet.domainName}}",
"ident": "-",
"auth": "-",
"@timestamp": "{{date.utc("YYYY-MM-DDTHH:mm:ss")}}",
"bytes": {{random.number(1000000)}}},
"verb": "{{random.weightedArrayElement(
  {
    "weights": [0.8,0.1,0.1],
    "data": ["GET","PUT","POST"]
  }
)}}",
"url": "{{internet.url}}",
"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

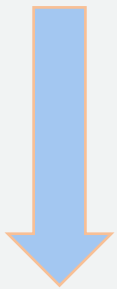
- Table permissions using named resource method.
 - Restricted column permissions
 - Data filters
- LF-Tags based permissions
- Show Permissions using Amazon Athena
- Show Permissions using Redshift Spectrum

 DataLakeAdmin

Demo items

 Access to
UserRole=architect

Tom




Classification =
Restricted01

database1

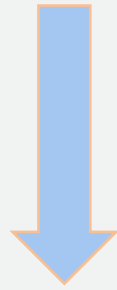
UserRole =
architect

Restricted columns
of table

amazon_reviews

 Access to
Classification=Restricted02

Jerry



Classification =
Restricted02

database2

UserRole =
devops

Data Filter on table

amazon_reviews


mySpectrumRole

spectrum_hu
didb

hudi_db





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



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

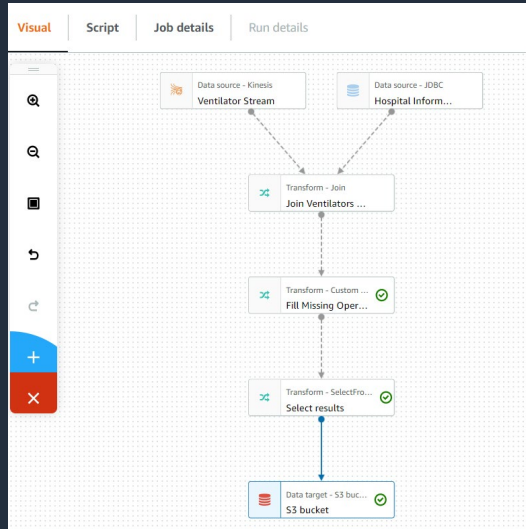


No lock-in

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

The screenshot displays the Glue Notebooks interface. A code cell contains a SQL query: `select * from 'aws-logs-1-uk-east-1':"iam_role"-"arn:aws:iam:590186200115:role/Notebook/RecycleTestRole";`. Below the code, a terminal window shows the execution of a SQL query: `select * from 'aws-logs-1-uk-east-1':"iam_role"-"arn:aws:iam:590186200115:role/Notebook/RecycleTestRole";`. The results are displayed in a table format:

country	alpha-2 code	alpha-3 code	numeric code	latitude	longitude
Afghanistan	AF	AFG	4	33	65
Albania	AL	ALB	9	41	20
Algeria	DZ	DZA	12	28	3
American Samoa	AS	ASM	16	-14	-170
Andorra	AD	AND	20	42	1
Angola	AO	AGO	24	-12	18
Anguilla	AI	AIA	660	18	-63
Antarctica	AQ	ATA	10	-90	0
Antigua and Barbuda	AG	ATG	28	17	-61

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

The screenshot shows the Glue DataBrew console. It displays a job configuration for 'Project dataset-met-objects'. The job is a 'Split column' job with 15 columns and 500 rows. The configuration includes a 'Split column info' section with 'Split column' and 'Source column' options. Below the configuration, a 'Correlations' section shows a heatmap of correlation coefficients between variables. The heatmap is a 15x15 grid with a color scale from -1.0 (blue) to 1.0 (red). The diagonal elements are all 1.0. The off-diagonal elements are mostly 0.0, indicating no correlation between variables. The console also shows a 'Summary' section with 'Total rows: 15,465' and 'Total columns: 23'.

Glue DataBrew for Analysts who prefer a no-code experience





Thank you!