**DAP IMPLEMENTATION (Prasanna Kumar Venkatesan - 2307558)**

**Step 05 : Data Enriching**

**Figure 1 :** **Creating Database in Glue.** A computer screen shot of a computer

Description automatically generated

Created Databases for the dataset in datactalog, and assigning the data to it.

**Figure 2: Creation of Crawler**

A screenshot of a computer

Description automatically generatedThe crawler has the dataset of the S3 bucket which is high quality data when you run you can see the dataset combines and gives you the output you look for

**Figure 3: SQL Query through Athena**

A screenshot of a computer

Description automatically generated

In this we open Athena and run the dataset to set the query and give a query to combaine the dataset with commands to it.

**Step 06: Data Protection:**

**Figure 04: Creating KMS for Bucket**

A screenshot of a computer

Description automatically generated

In this we are creating KMS for each bucket in S3 so we created KMS for all the 3 buckets in S3 (RAW,TRF,CUR) so added the screenshot for the single implementation. So we creating a unique key for the S3 bucket and start working on it the changed the key from default to the key we created for the bucket

**Figure 05: Encryption of the default key:**

A screenshot of a computer

Description automatically generated

**Figure 06: Encryption of default version for S3 Bucket:**

A screenshot of a computer

Description automatically generated

Once the encryption of the key is done the version of the S3 Bucket is also changed and updated it.

**Figure 07: Replication rule for all the bucket:**

A screenshot of a computer

Description automatically generated

Once the KMS is implemented and executed we need to duplicate the dataset for backup so we need to create a replication rule for all the buckets

**Step 07: Data Governance**

**Figure 08: Data Governance Pipline**

A screenshot of a computer

Description automatically generated

In this we check the Data quality & Control Implemented on the dataset. So we have the clean dataset we want to make sure we have the quality dataset so by implementing some quality question to the data pipeline we are assessing the quality of the dataset and storing it to a file.

Figure 09: Data Governance Pipeline Job Run Results.

A screenshot of a computer

Description automatically generated

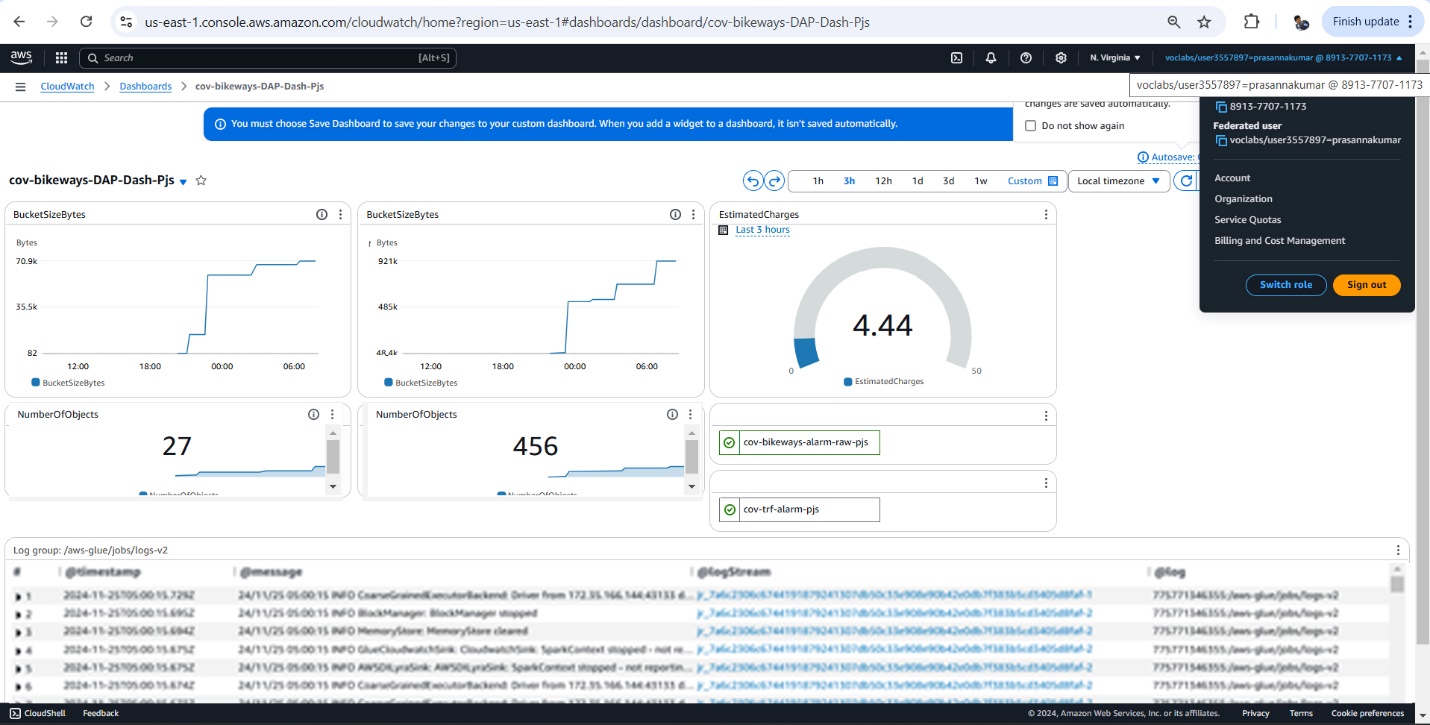
**Figure 10: Results Stored in S3 Transform Bucket**

A computer screen with text

Description automatically generated

**Step 08: Data Observability**

**Figure 11: Dashboard showing DAP for the City of Vancouver Bikeways dataset.**



In this dashboard it showing the DAP of City of Vancouver Bikeways dataset with the number and standards and there usages of storage in S3 and also showing the output and showing the alarm of the dataset.

**Figure 12: Cloud Trail showing the tracking user Activity**

A screenshot of a computer

Description automatically generated

After Creating the Cloud Watch, we created the Cloud Trail to show the tracking of the user activity in the AWS by showing the end users what we read and write in the storage by giving them the access to the storage.