Orders\_Data\_Tracking\_Pipeline

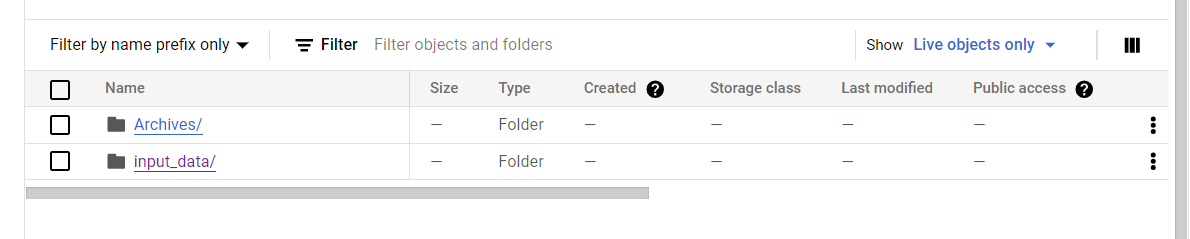
Tech Stacks:

1. Databricks
2. GCP bucket
3. GCP

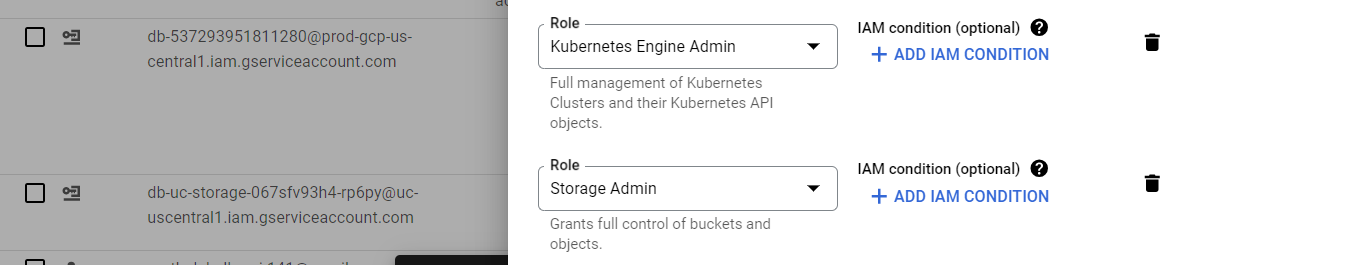
Steps Involved: -

1.Bucket Creation

Create bucket named “Order\_Tracking\_Bucket ” consisting of folders ‘input\_data’ and ‘archives’. ‘input\_data’ will be loaded with the daily data and once transformation is applied on the daily received data the data will be moved to ‘Archives’.

2. Accessing GCP storage:-

i. To access storage bucket we need to give admin access such as ‘Access Secrets of GCP project’ and GCP connector jar files.

Give full admin access to GCP buckets by setting up ‘Storage Admin’ role to the databrics service account. 

iii. Visit ‘Service accounts’ tab on right side of console and from there from managed key option download the key in the json foramat.

iv. Download GCS connector jar file to connect with to cloud storage.

v.gcs connector is also required to connect with gcs.

3. Developed two scripts which ‘stage\_logistic\_tracking\_job’ where daily data is stored and after procesing csv file uploaded in ‘input\_data’ folder, table is created in delta location upon which ‘staging\_order\_tracking’ table is created. The processed file is then moved to archives folder.

4.Another script ‘target\_logistic\_tracking\_upsert’ reads the data from the ‘staging\_order\_tracking’ table and performs upsert operation where the daily status is updated target\_order\_tracking which contains all records with updated status.