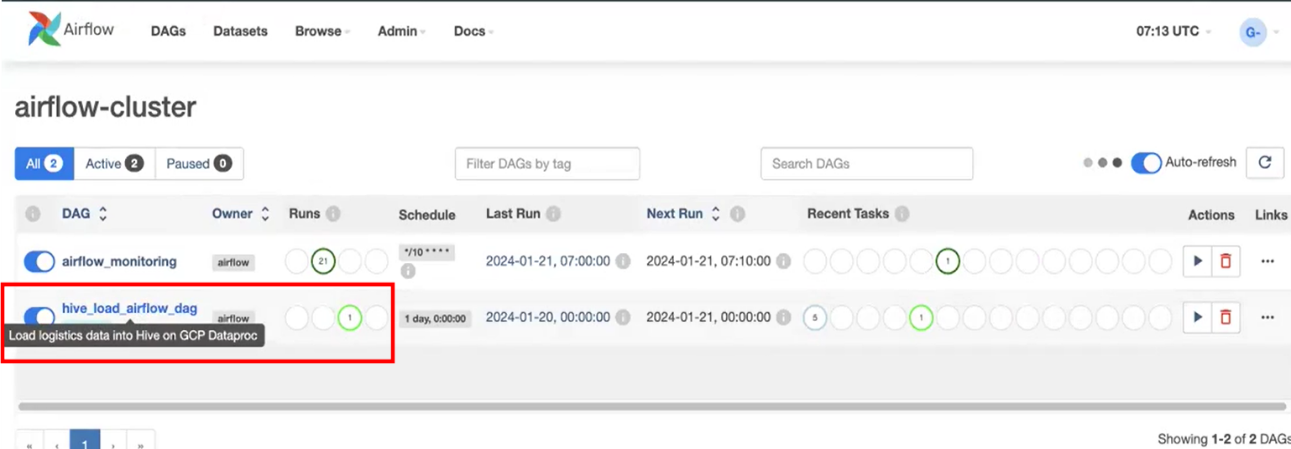
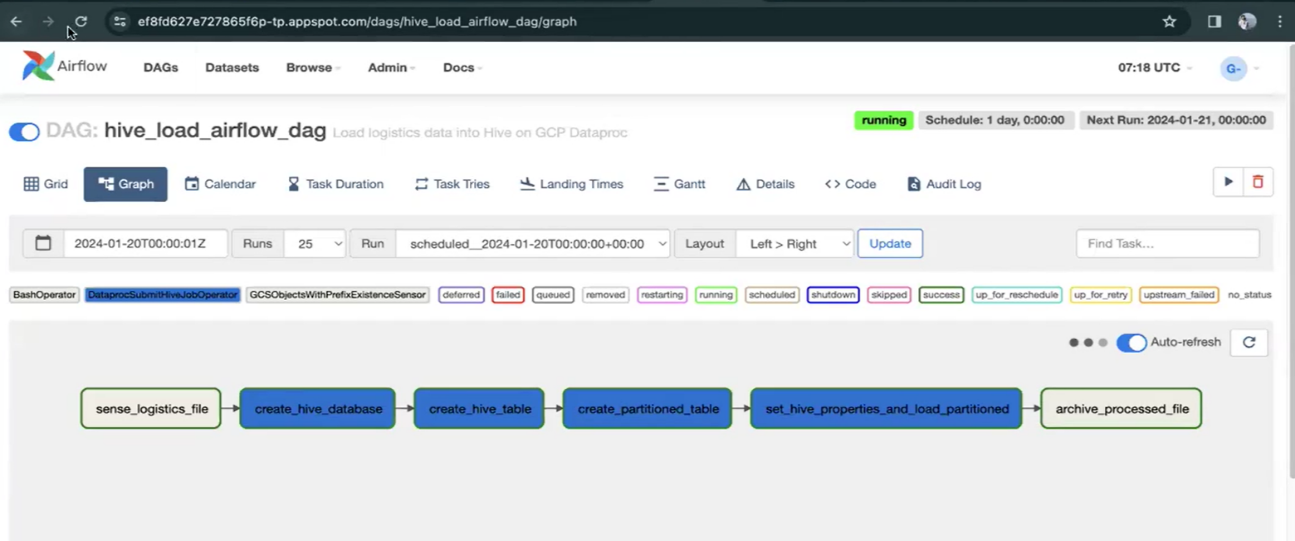
Logistics Data Warehouse Management

1. we had a gcs bucket named 'logistic-raw' where we were receiving daily csv file with file name having date in it.



1. next set up an airflow job used to listen the file received in bucket. We perfomed certain stages in airflow job which includes create database, create external stage table, create partition table, loading data into partition table and then running archieve.

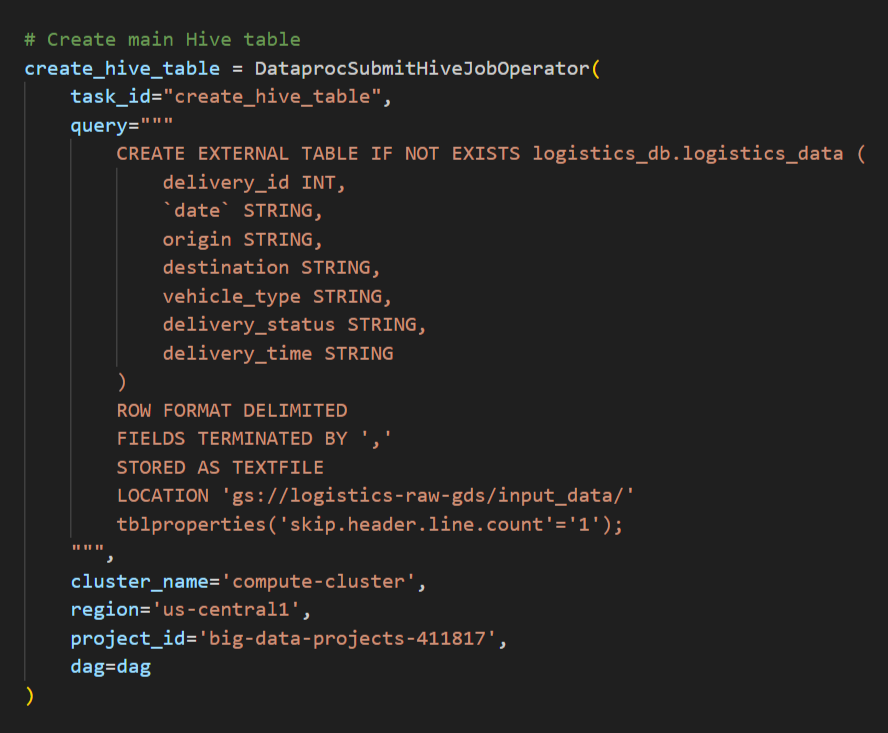


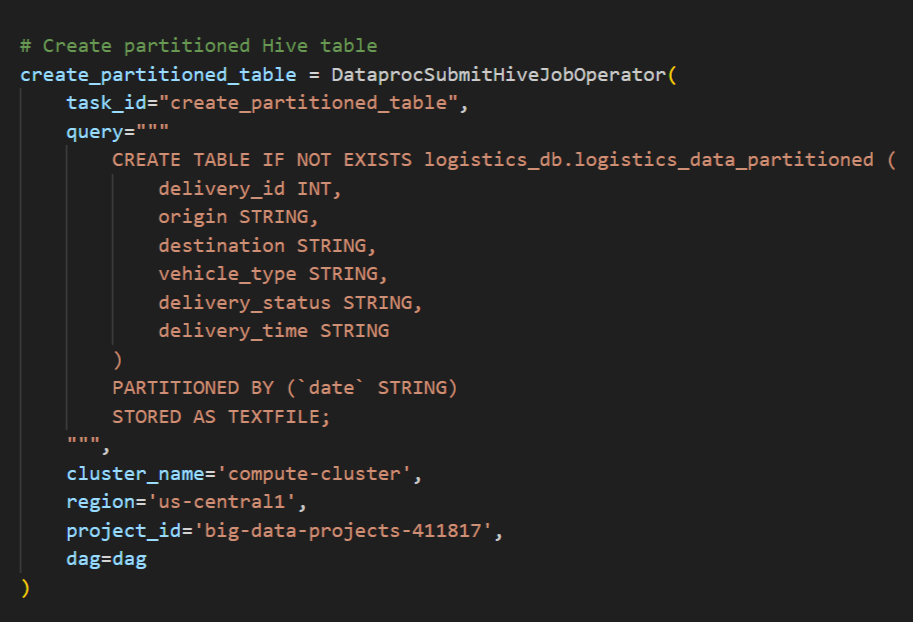


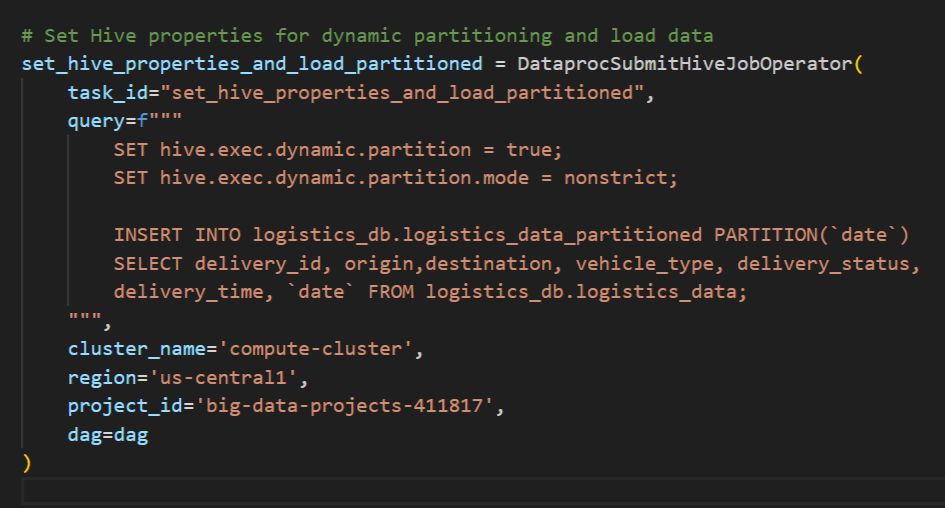
1. In airflow dag job, we were using gcs cloud storage sensor here to sense the file landing in the gcs storage bucket and DataprocSubmitHiveJobOperator to run, submit hive queries on the dataproc cluster.



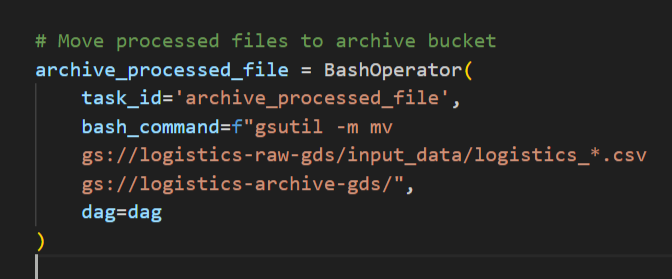
1. We created external stage table on the 'logistic-raw' bucket path. We ensured that only delta data or current data should be present in this path. External table was being used in order to ingest the data in partitioned form into the hive partition table.







1. We have another bucket 'logistics-archieve' where we used to move daily file from 'logistic-raw' after file is processed. In this way we archieved processed file using the BashOperator.



1. Here, the query results for hive partitioned table

