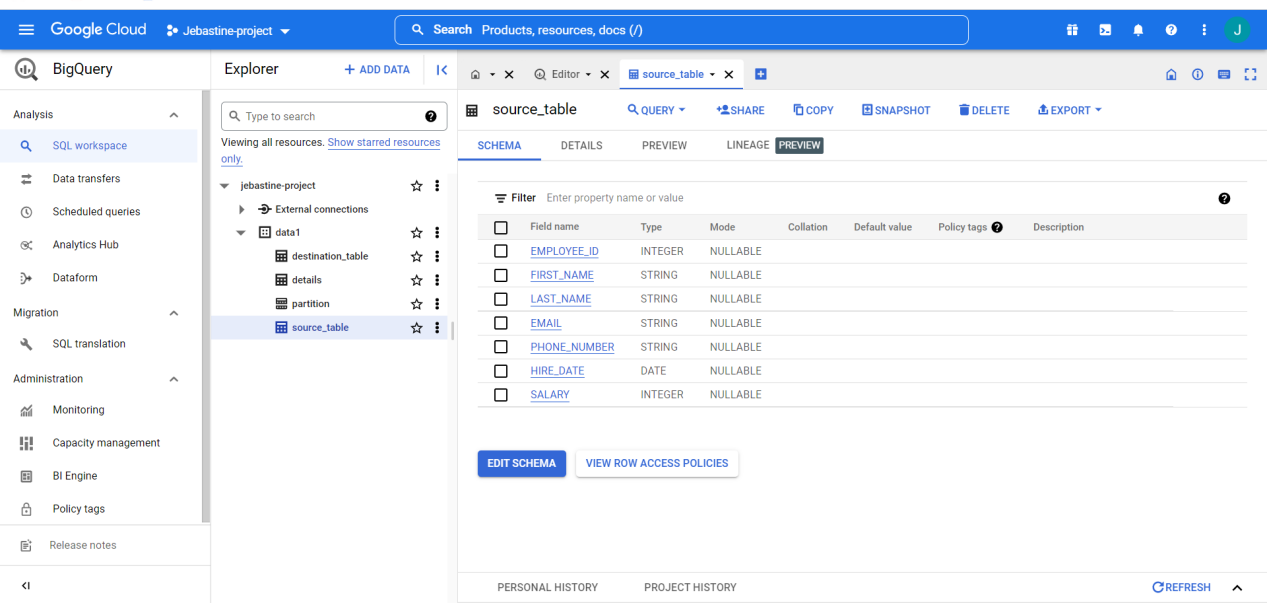
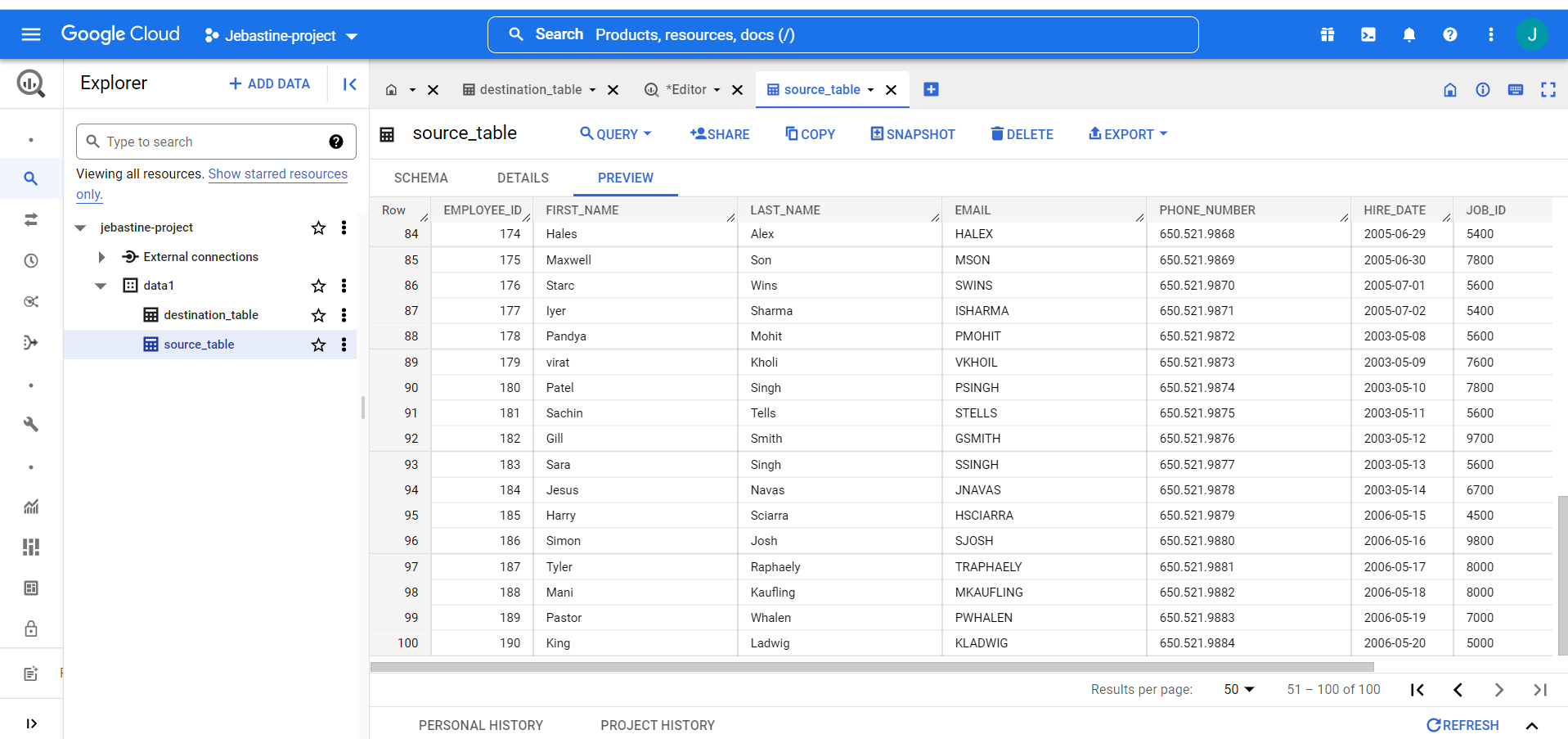
1.Creating the table:

Source table schema:

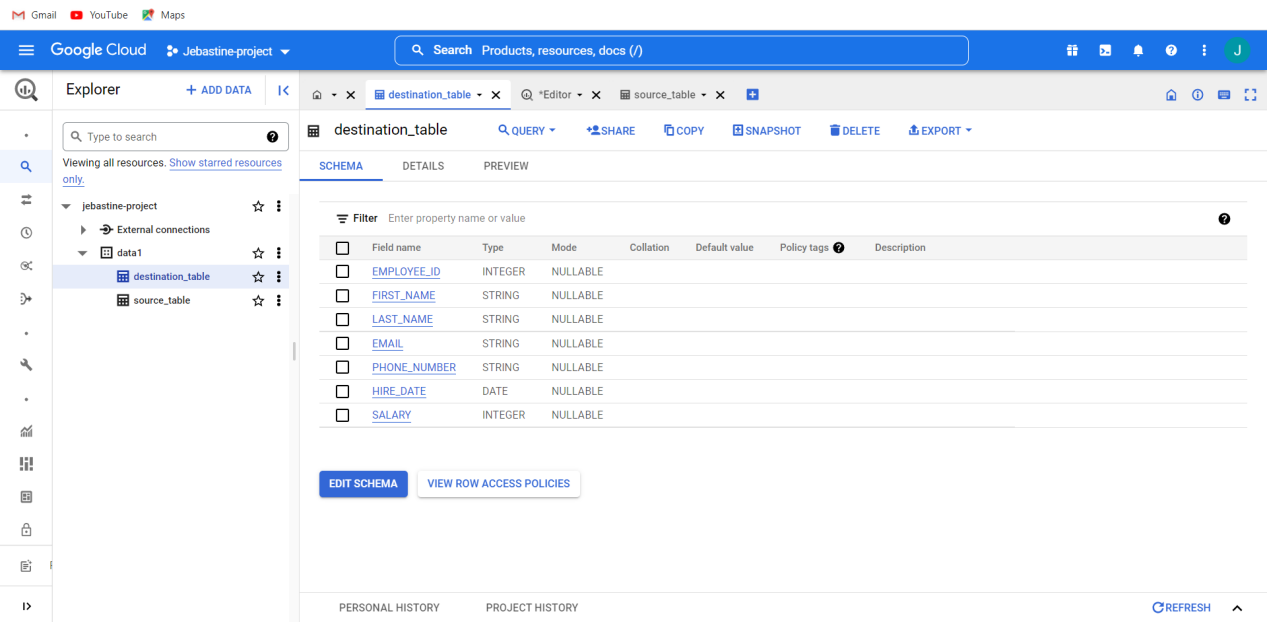


Source table preview

100 records:

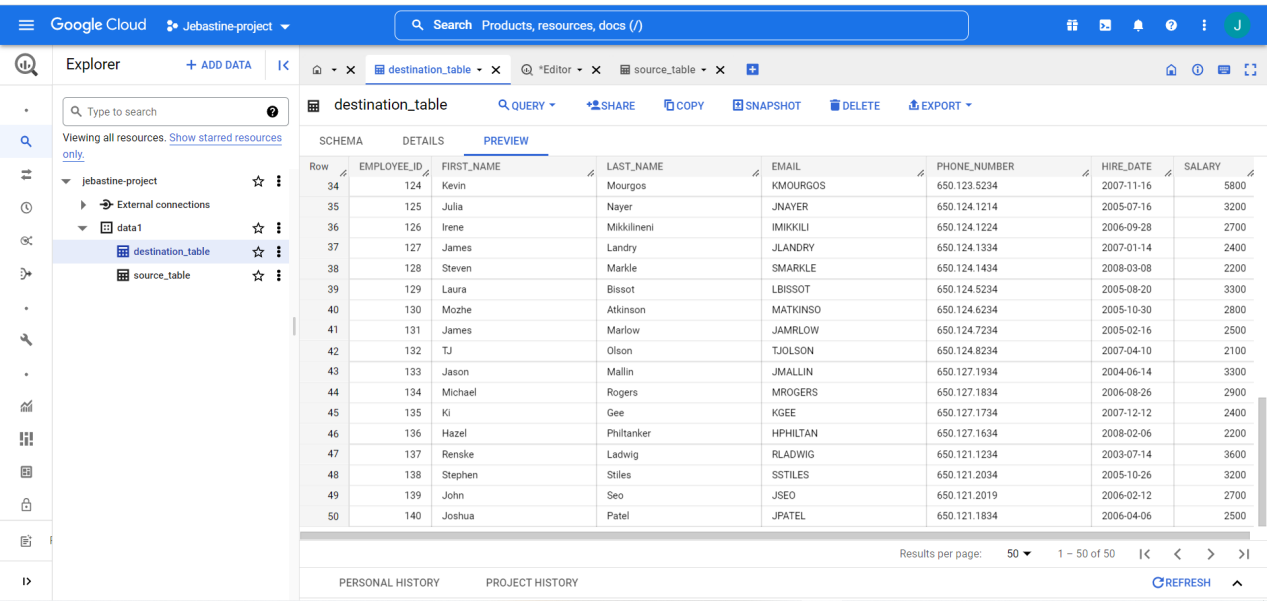


Destination table schema:



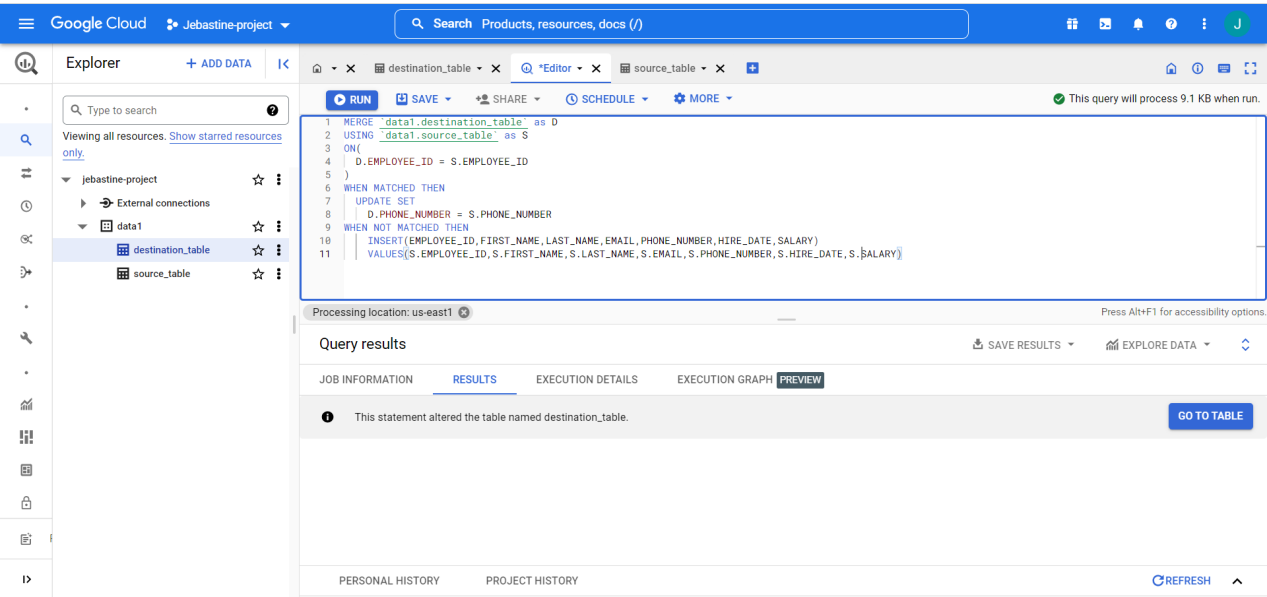
Destination table preview

Where there are only 50 records:

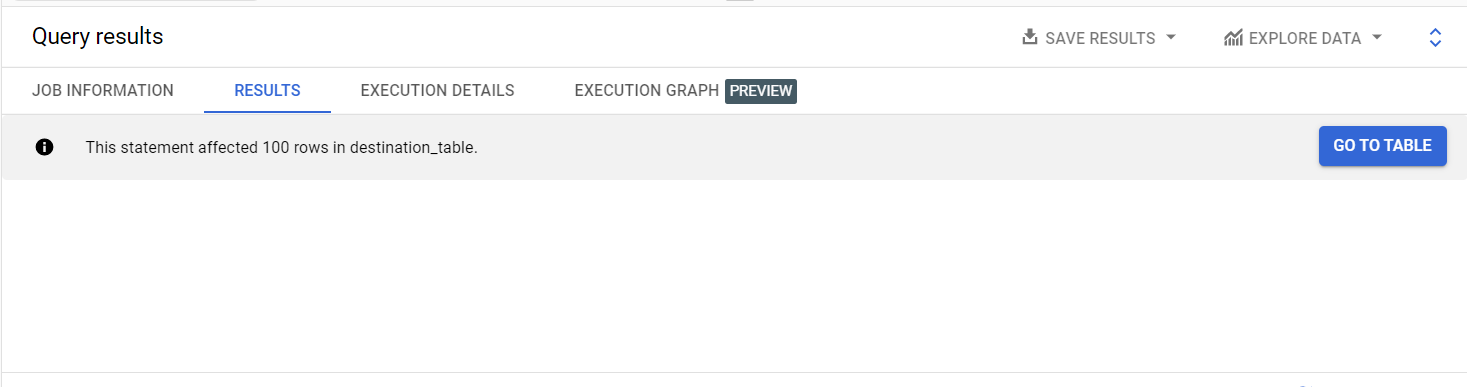


Create merge statement to load the source table data into destination table using upsert scenario:

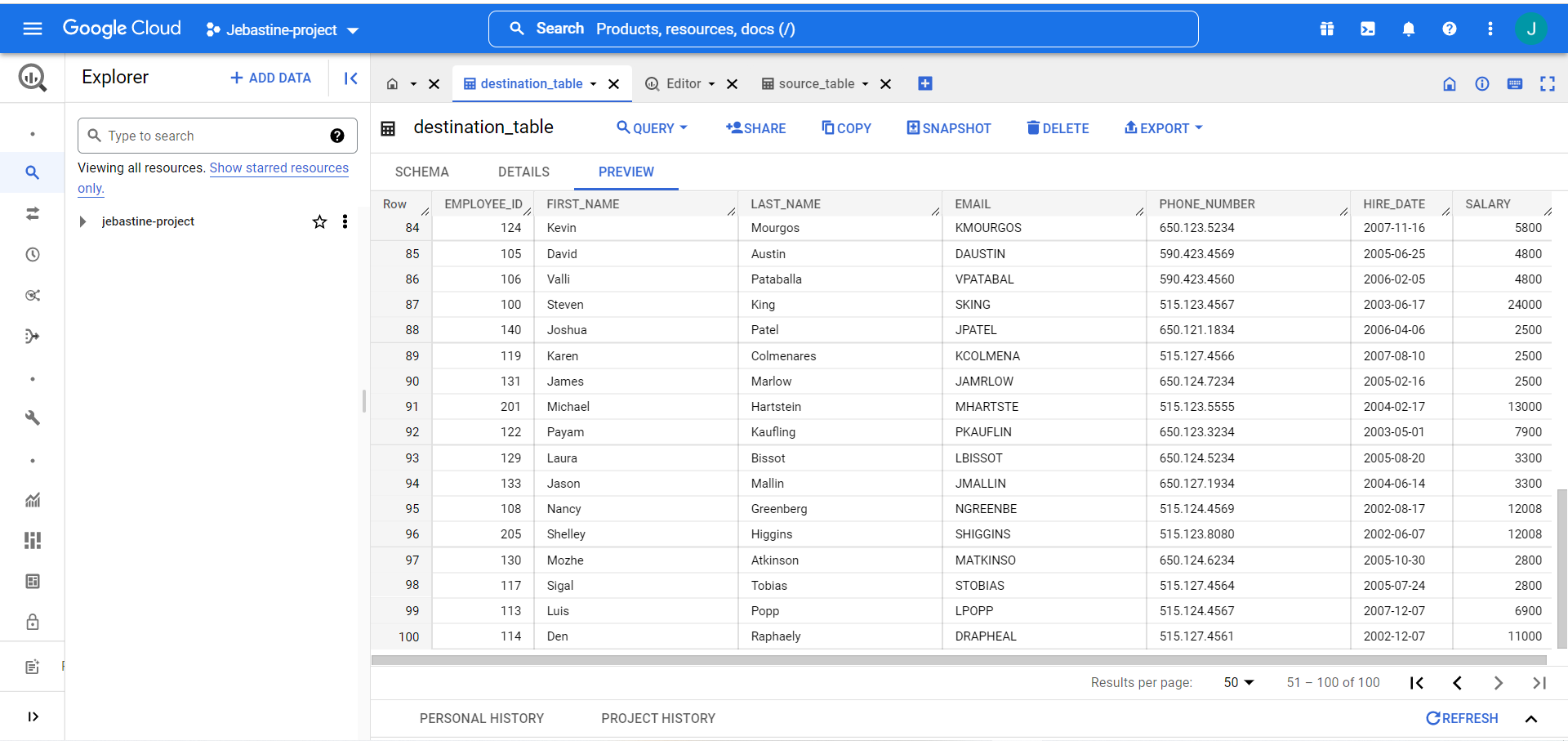
Query:



Destination table is affected:



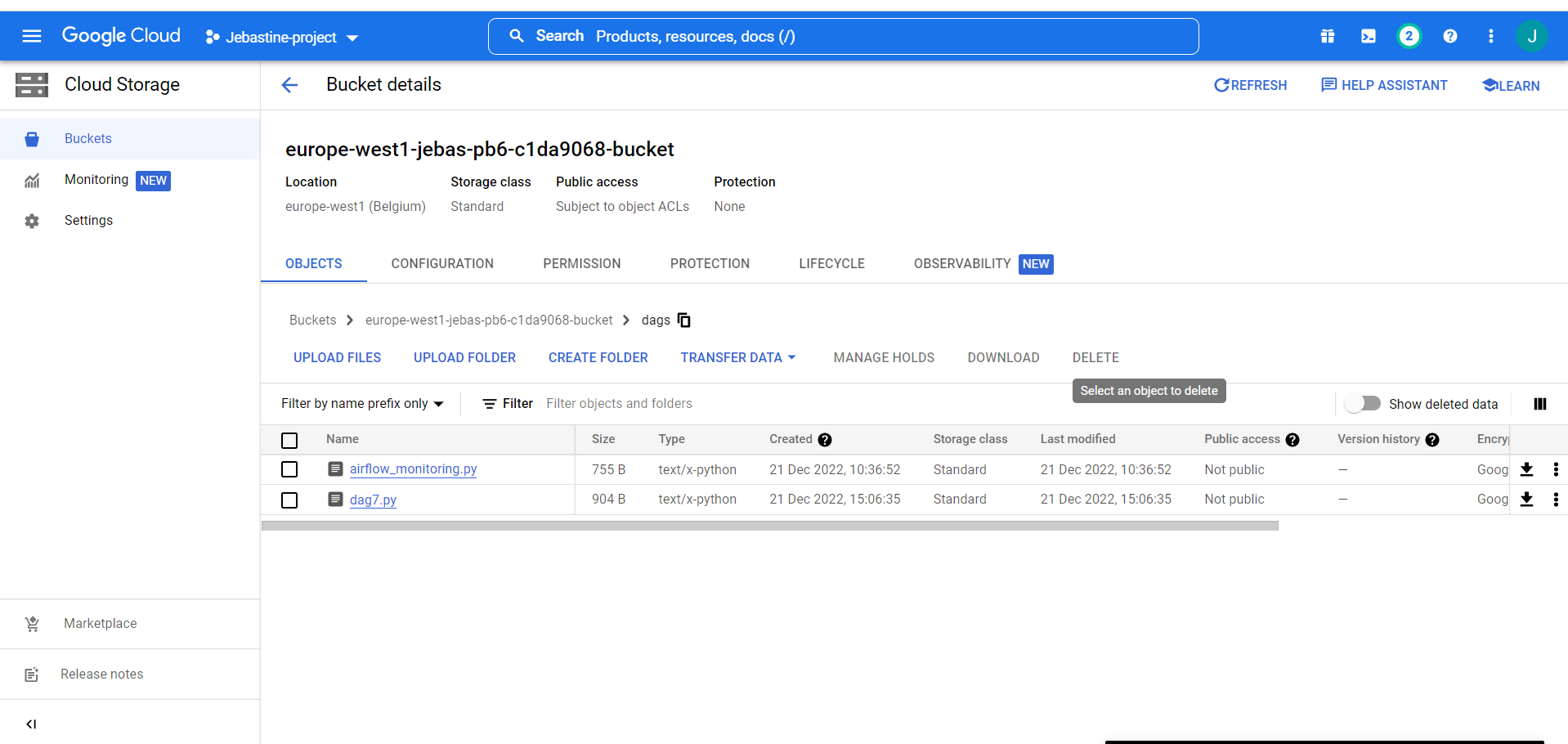
Merge and upsert done:



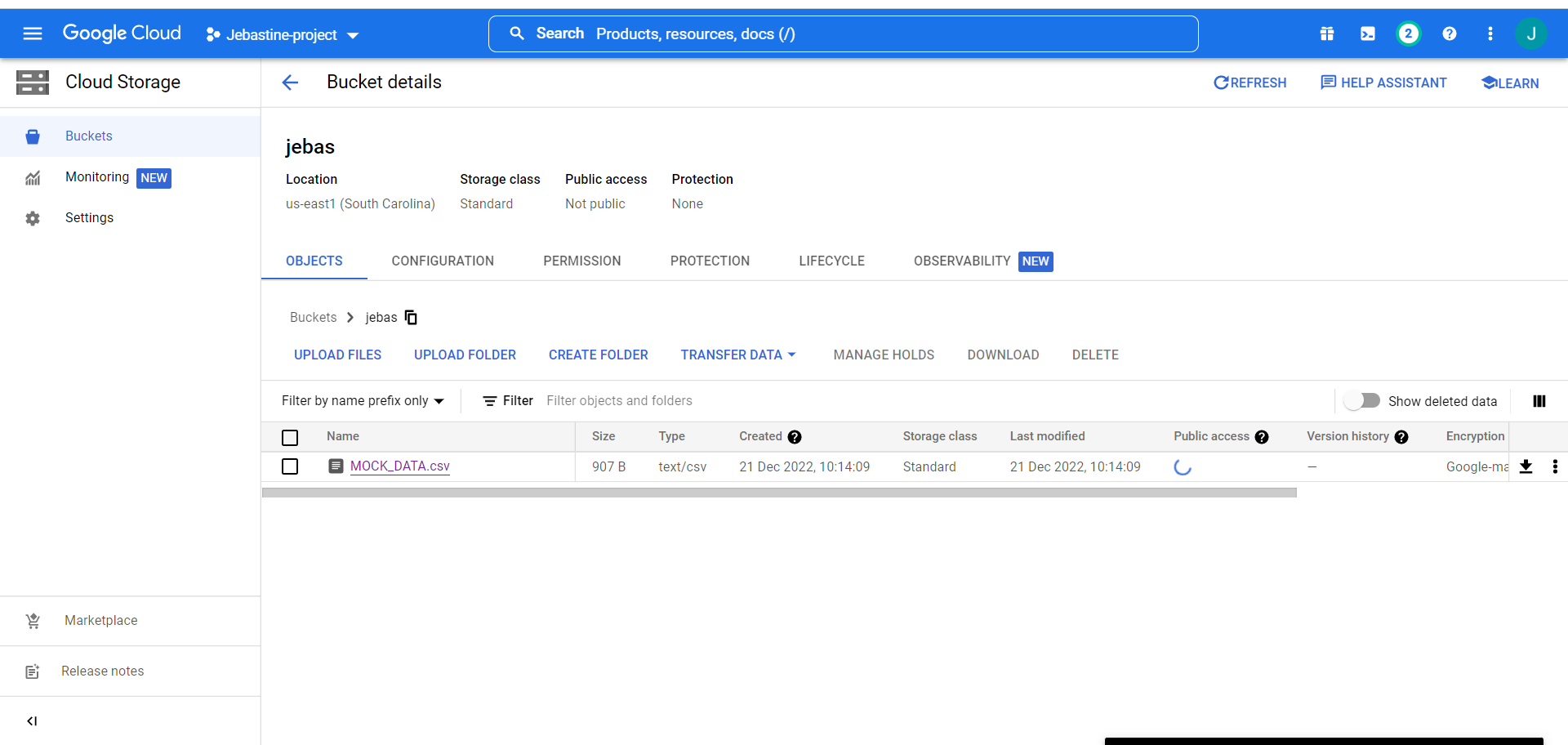
2.Create an airflow dag which will load the data from GCS bucket to Bigquery table.

Using composer creating airflow in gcp:

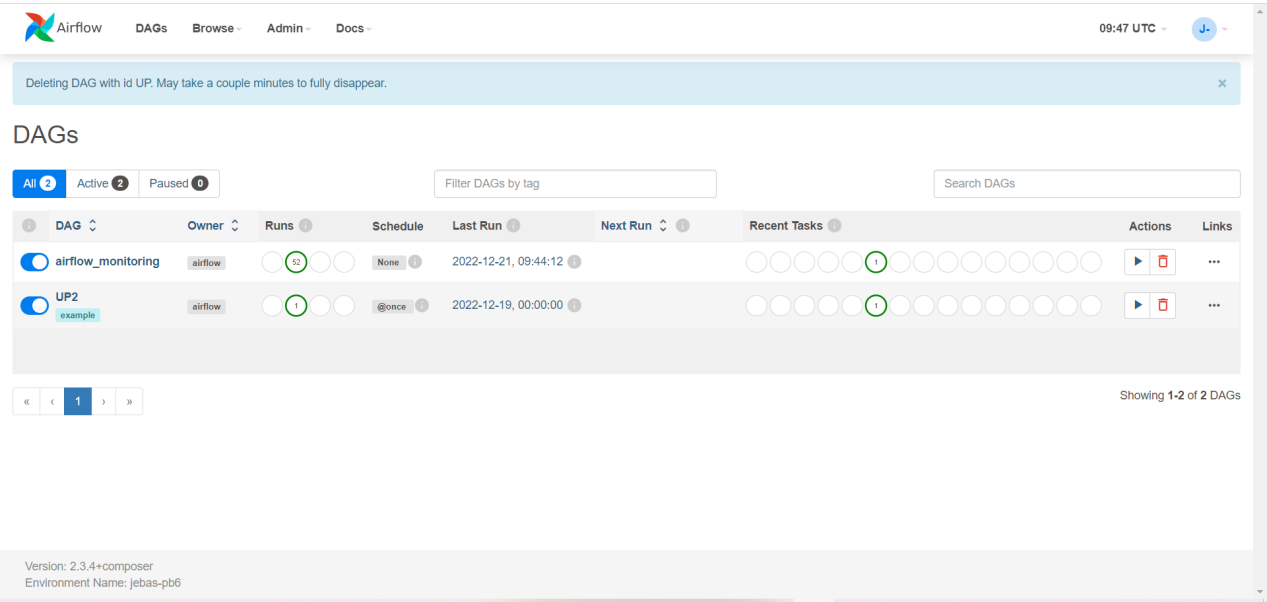
Dag folder:



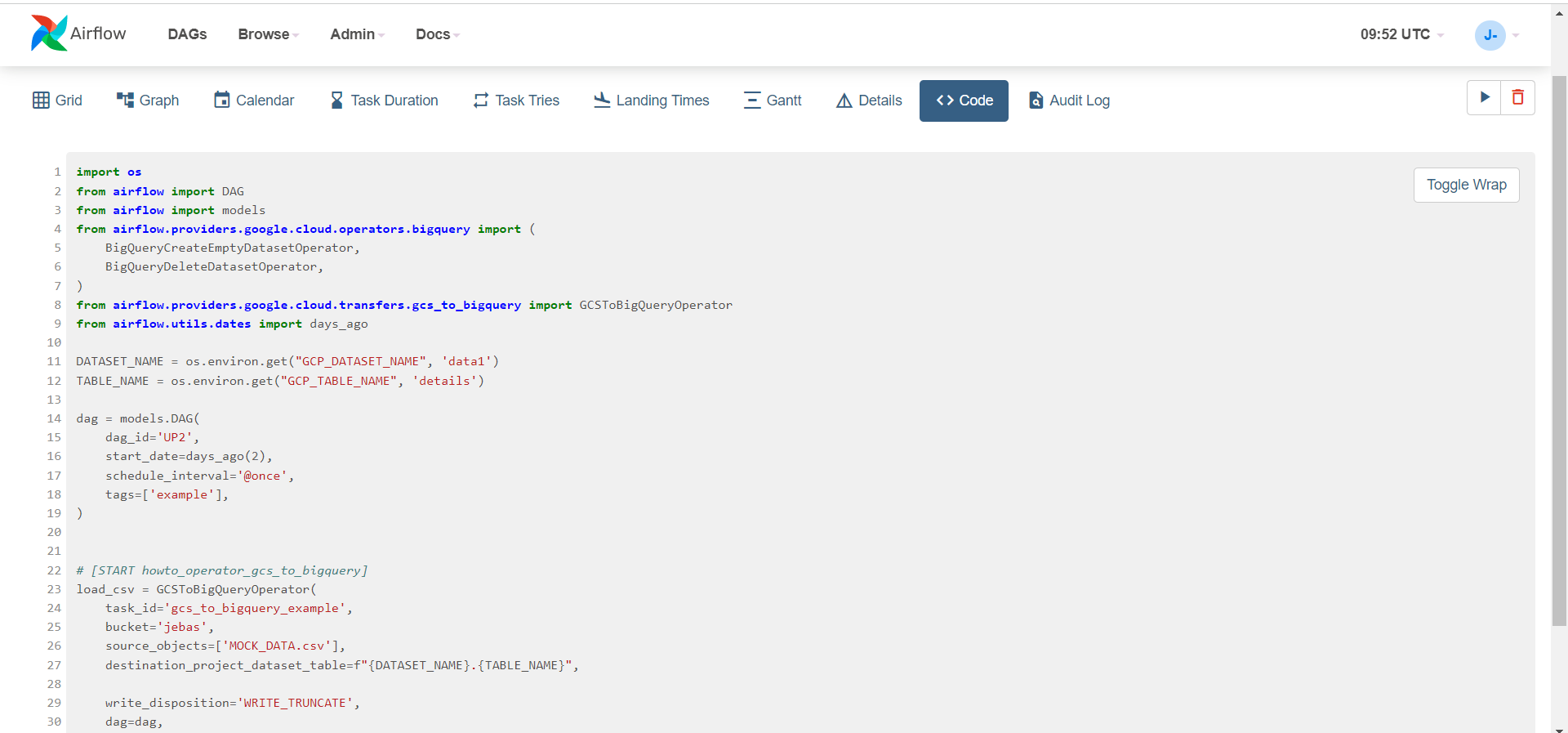
Csv folder stored in the bucket:



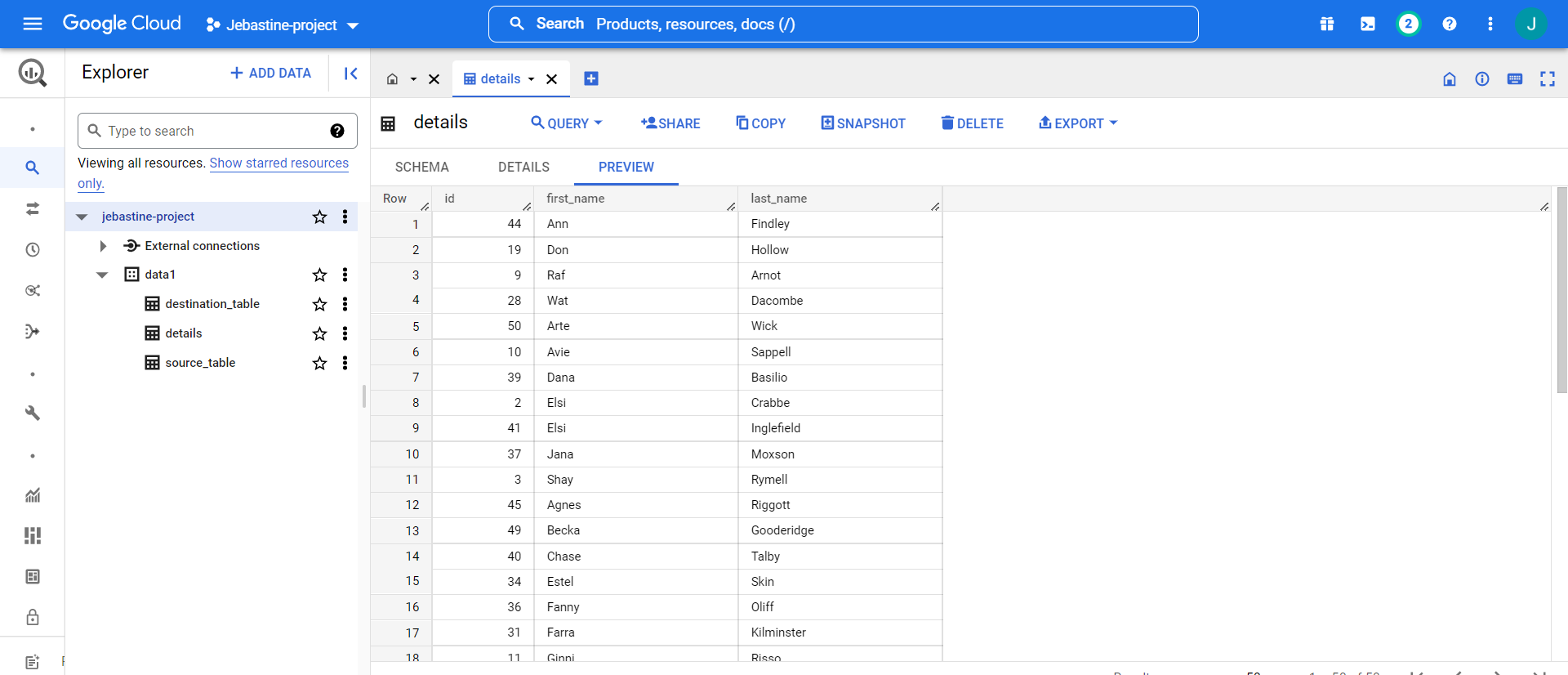
Dag is success:



Code:

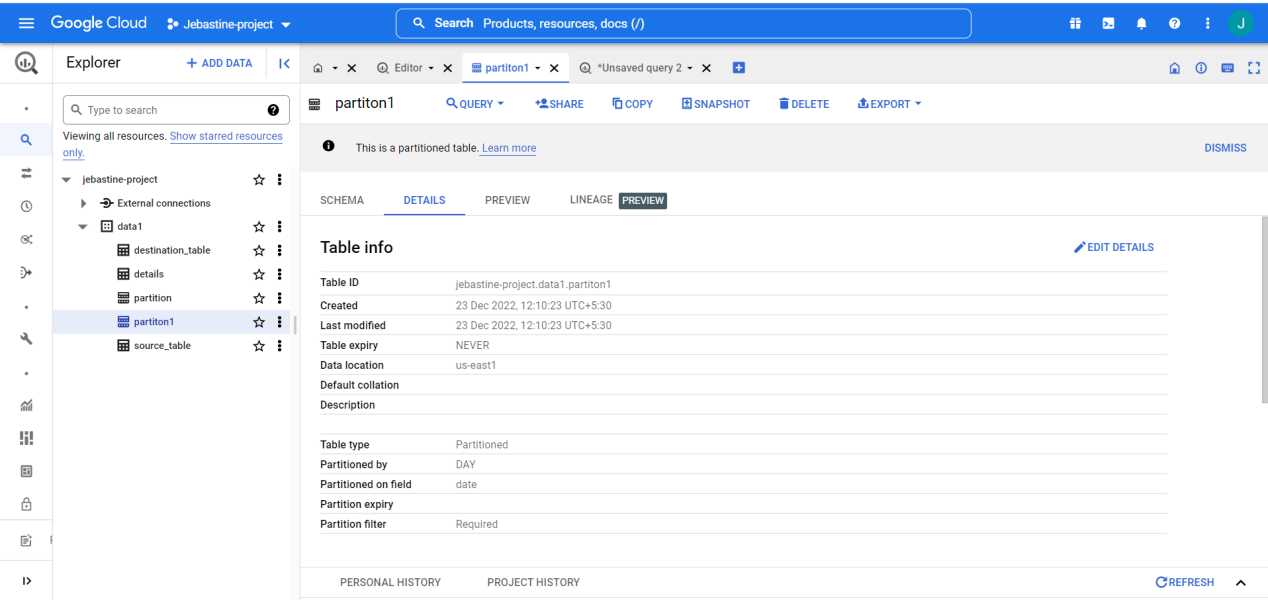


Successfully loaded:



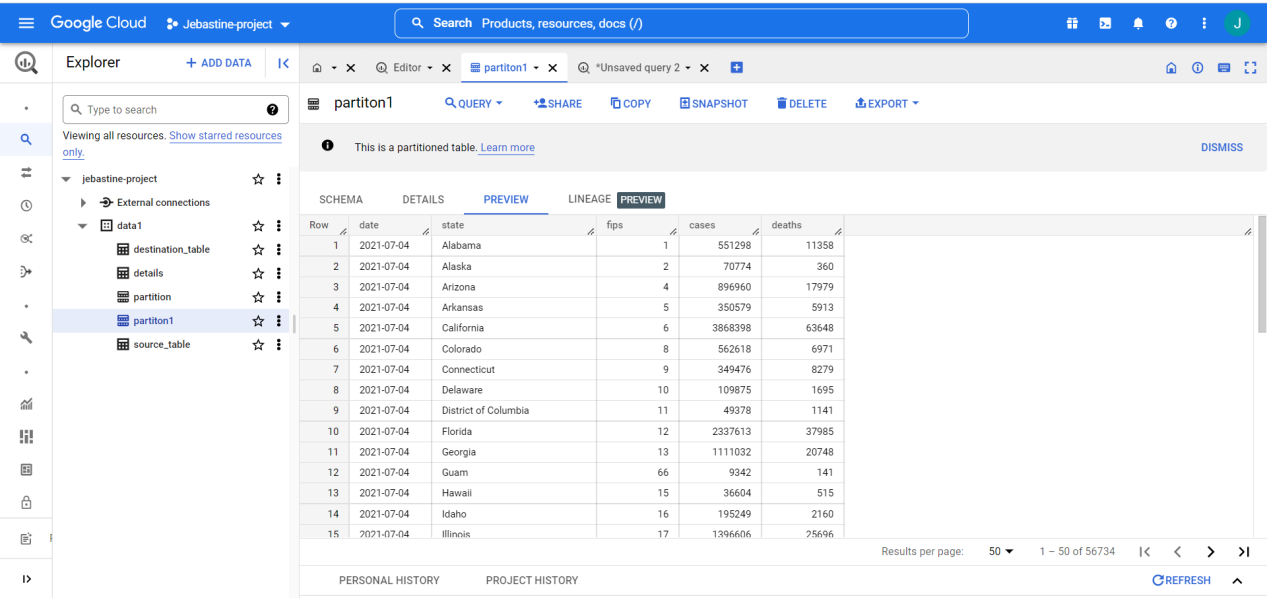
3.Upload high volume of data in big query and try to optimise it's read performance by using appropriate partitioning and clustering.

Data for partitioning by day:

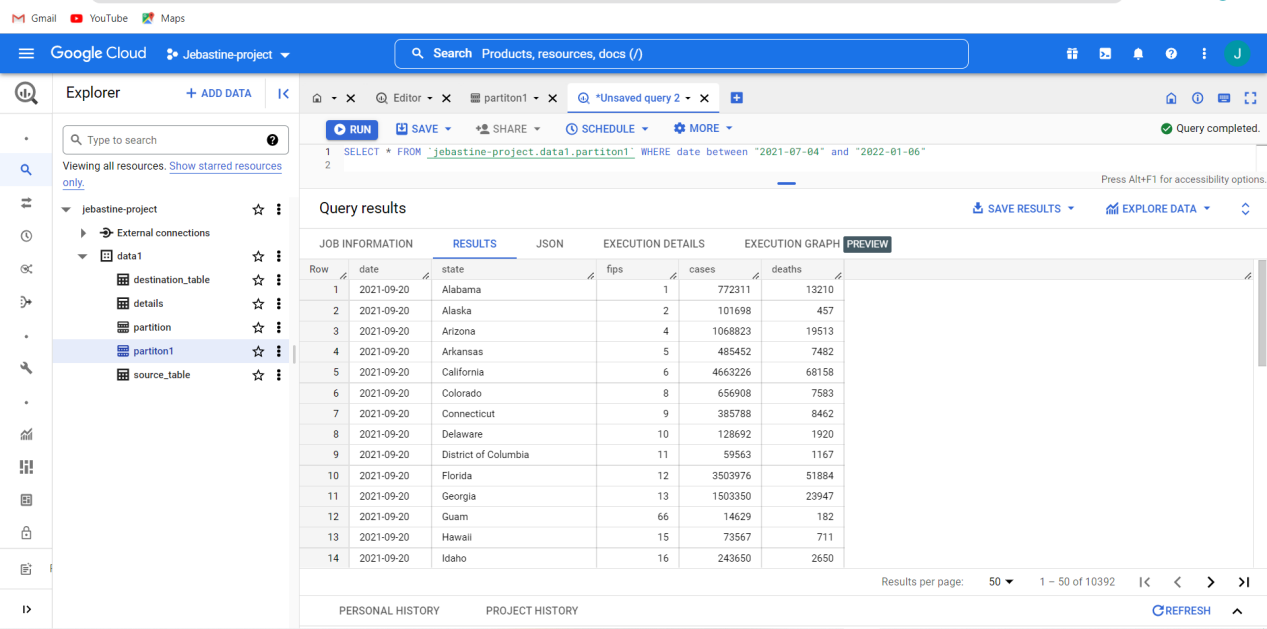


Volume of data:

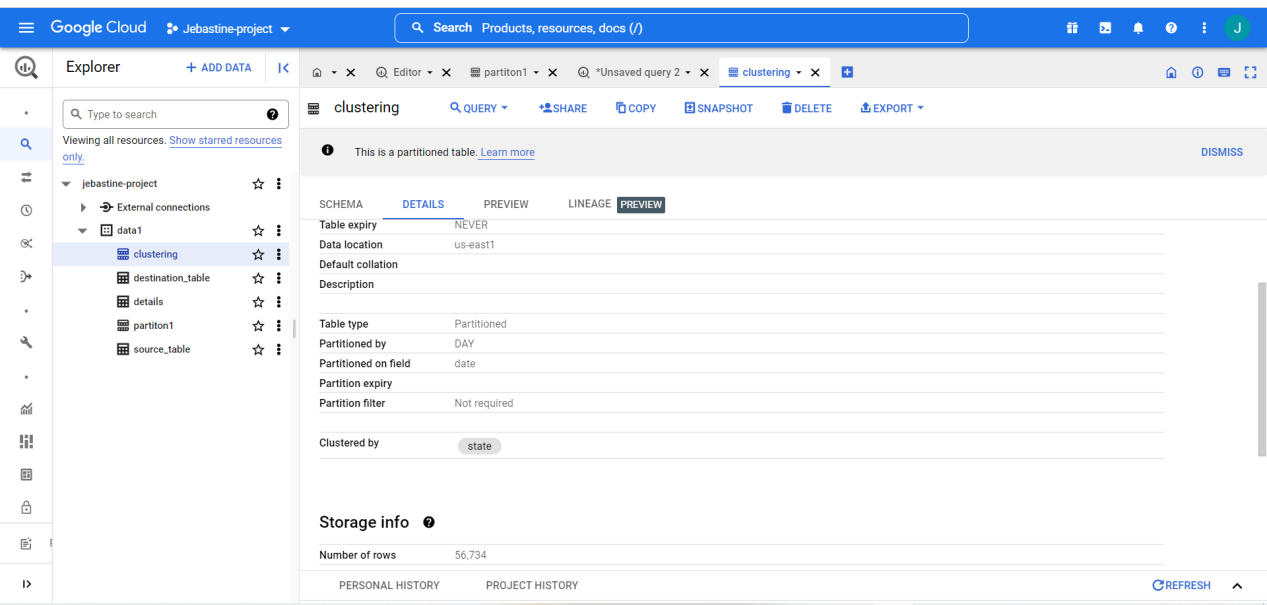
56,734 rows:



Partitioning based on date:



Clustering based on state:



Query :

Successful:

