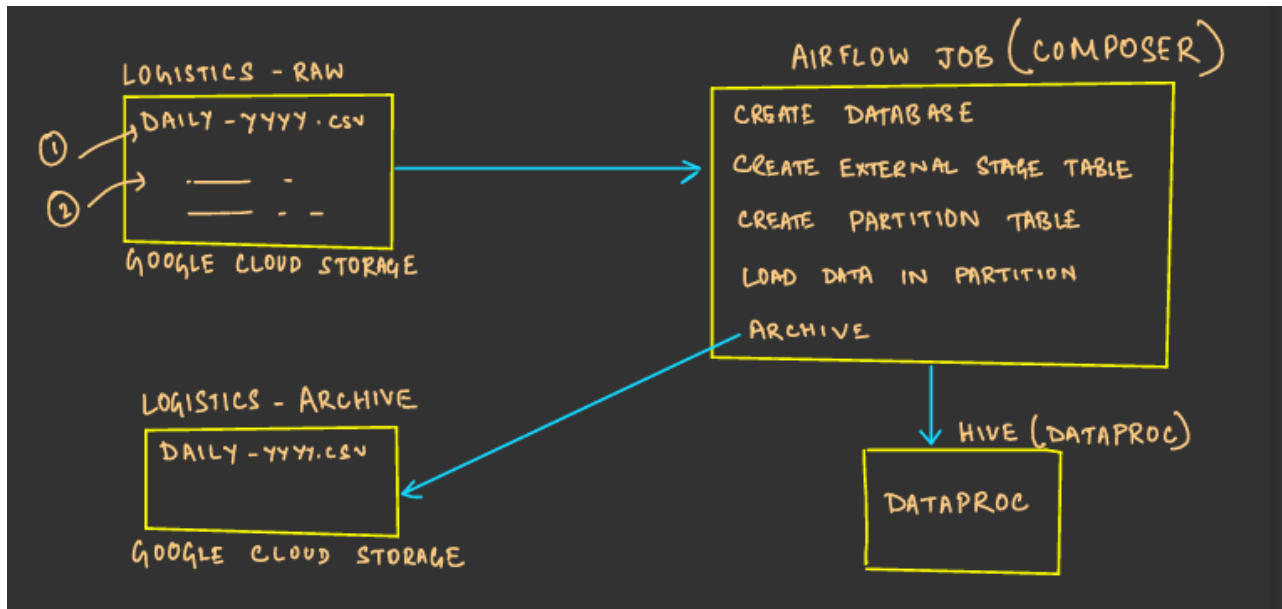


➤ Logistics Data Warehouse management

Tech stack : [Google Cloud Storage](#), [Composer](#), [Dataproc](#), [Hive](#)

Project design:



Project execution steps:

- Airflow job (composer) has been scheduled. New files arrives in Google Cloud Storage input bucket, it will sense the logistics file through sensors written in airflow code.
- In Dataproc cluster, Hive Databases, tables will get created , data will get loaded in partitioned table and the file will be moved to archive folder after processing. So, in GCS input bucket always the new file will be present before processing.
- Daily files will be processed in an automated manner following the archival process.

Airflow job execution:

