


Project 3

① Daily airline data ingest

Data

↳ dim table of airport codes → redshift

↳ daily flight data

- ↳ aircraft ca
- ↳ origin airport
- ↳ destination code
- ↳ dept-delay
- ↳ arr-delay

53

fact table in redshift

- ↳ aircraft
- ↳ dept-airport-name
- ↳ arr-airport-name
- ↳ dept-city
- ↳ arr-city
- ↳ dept-state
- ↳ arr-state
- ↳ dept-delay
- ↳ arr-delay

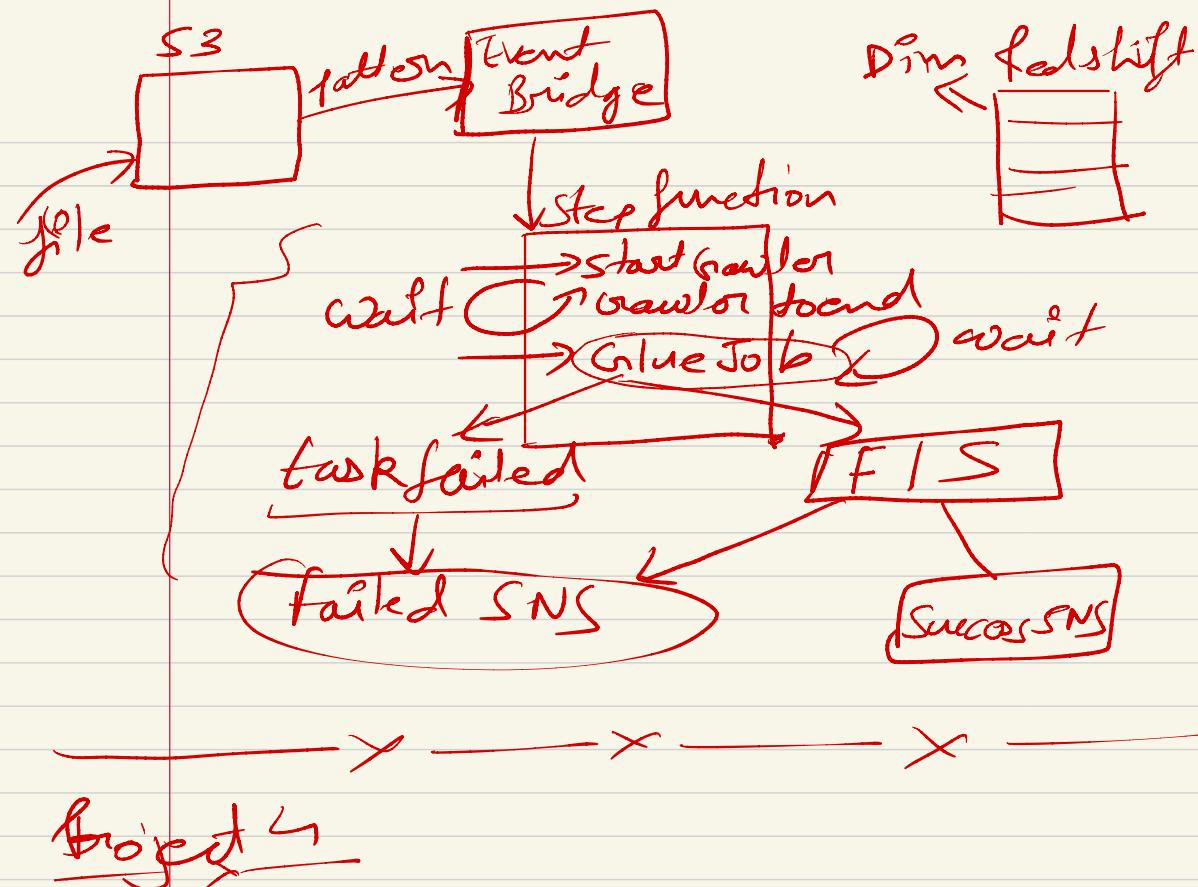
What we need to achieve

- ① daily incremental load in redshift fact table
- ② As soon as file land in S3, we need to start process

Tech Stack

- ① S3
- ② S3 CloudTrail Notification
- ③ Event Bridge pattern rule
- ④ Glue crawler
- ⑤ Glue ETL
- ⑥ SNS for notification
- ⑦ Redshift warehouse
- ⑧ Step function for orchestrate

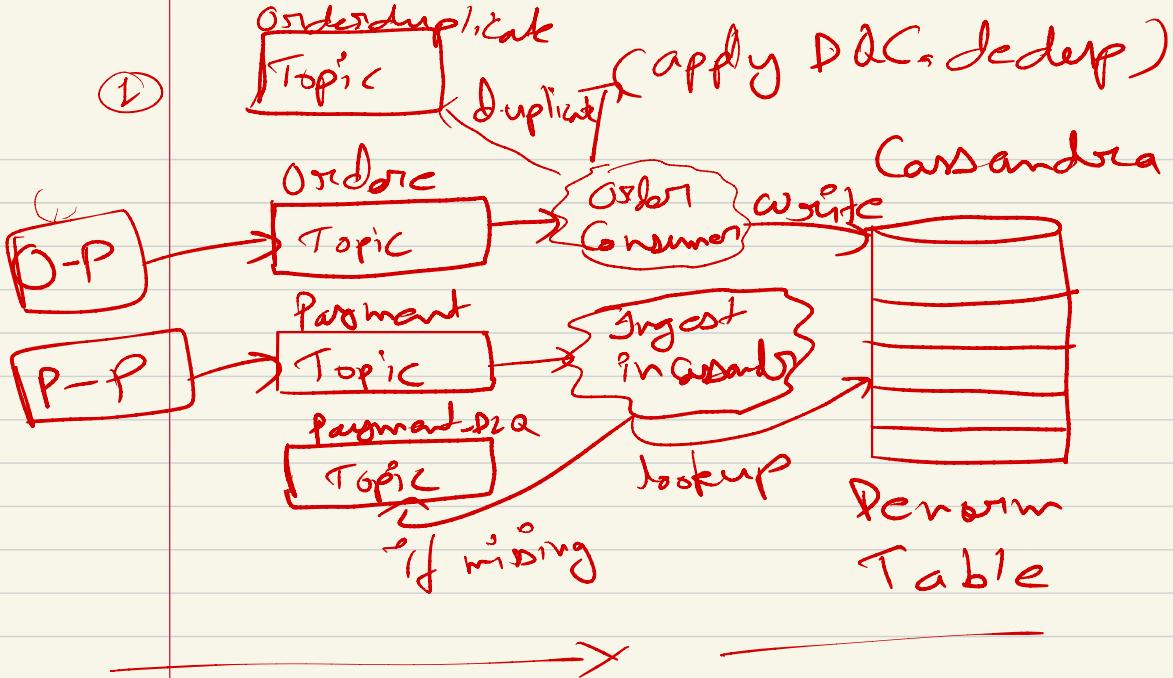
Architecture



Ecom store data processing

Tech Stack

- ① GCP Pub-Sub
- ② Python Producer/Consumer
- ③ DLQ
- ④ Cassandra with docker



Projects

logistics data warehouse
management

Tech Stack

- ① GCP Bucket
- ② Airflow
- ③ Hive Operators in Airflow
- ④ Dataproc cluster
- ⑤ Hive

Airflow Job

