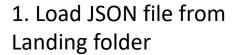
## **Data Engineer Case Study**

### **One Time Activities**

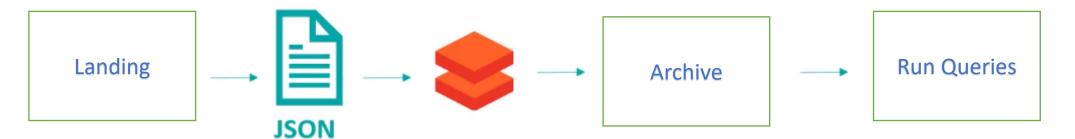
- Create "Landing" folder.
  - A separate service will download new files in this folder.
- Create "Archive" folder
  - Files will be moved from Landing to Archive folder after they get processed.
- Create "processing" and "presentation" databases.
- Create Delta tables for processing db.

# **Data Engineer Case Study**

### **Architecture Diagram**



3. Move saved files
From Landing to Archive
Folder



Clean and Save JSON file in multiple tables (Processing Layer)

4. Run Queries and Aggregation on tables saved in step #2.Also save complex aggregation Results in tables (Presentation Layer)

## **Data Engineer Case Study**

### **Orchestration**

- We can create Airflow Dag to orchestrate Data Pipeline.
  - We can create dependency of tasks in Airflow using set\_downstream and set\_upstream methods.
  - We can create fixed daily schedule for Airflow Dag to trigger it daily at given time.
  - We can also have another Dag created which can keep checking landing folder for new files and when new files will come it will trigger our pipeline Dag.