

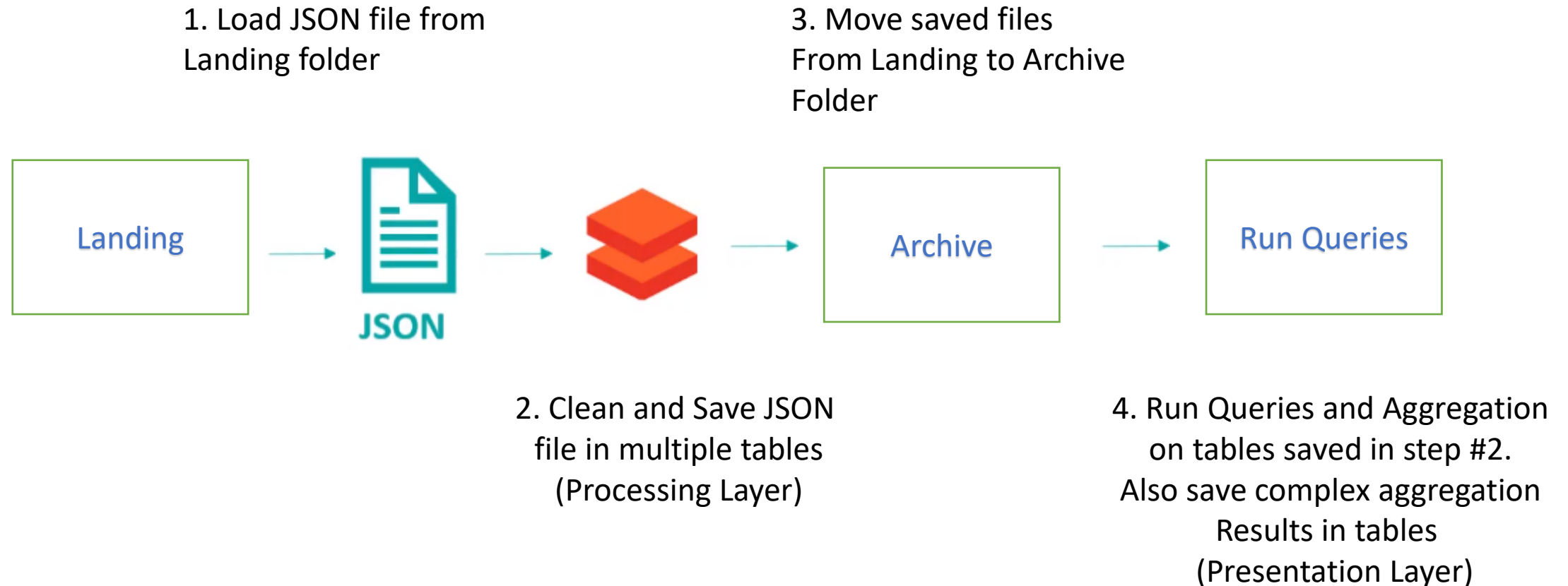
Data Engineer Case Study

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



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.