Syed Hamza Jamil (29454)

Saad Ullah (29416)

Architecture of Data Warehouse and Its Logic

Data Lake: AWS S3

Data Warehouse: Redshift

ETL Tool: Airflow

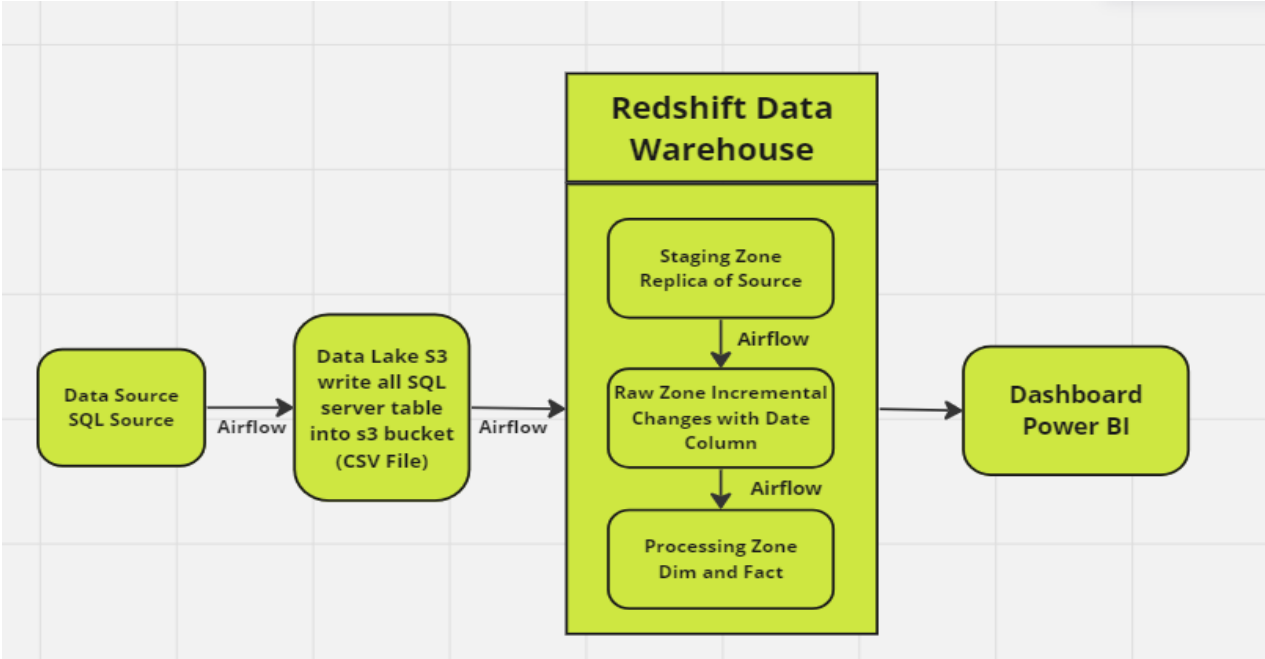
Dimension: SCD Type 1

Fact: Fact_inventory_transaction and Fact_orderdetails

Raw zone: full dump plus incrementally

Logic: In first time we will load the full dump data into warehouse then we will load or add incrementally data logic (insert and update).

Data Warehouse Architecture Diagram



In this data warehouse project there are four steps are bellows

Writing Source SQL Table in S3 Data Lake using Airflow

The purpose of this step to move data from source to s3 using airflow, once csv file is created we are able to move data fast into staging zone on redshift

S3 to Redshift Staging zone using Airflow

In this step we will fetch the data from s3 data lake and load these data into redshift staging zone using airflow.

Staging zone to raw zone using Airflow

In this step we will fetch the incrementally data from staging zone to raw zone, its means we will load incrementally data into raw zone with date column using airflow.

Raw zone to processing zone using Airflow

In this step we will fetch the data from raw zone to processing zone and convert these tables data into dim and fact tables using airflow for dim creation we are use SCD type 1.

Power BI Dashboard using Redshift Data Source

In last step, I connected redshift data source in Power BI Dashboard for creating dashboard