Global BnB Data Platform

Source

insideairbnb.com

Object Type: Gzipped CSV Files

Interface: Webscraping using Python

Raw Data Storage (Landing Zone)

Google Cloud Storage

Object Type: Gzipped CSV Files

Interface: Snowpipe (event-driven)

Snowflake Data Warehouse

Bronze Schema

Raw Data

Object Type: Tables

Load Type: Incremental Load

Insertion Type: Update and Insert

Transformations: Each row of data is inserted as such from the external stage

dbt

Silver Schema

Cleaned Data

Object Type: Tables

Load Type: Incremental Load

Insertion Type: Update and Insert

Transformations: Data is cleaned

dbt

Gold Schema

Business Ready Data

Object Type: Tables and Views

Transformations: Joins and Aggregations using the Silver Tables. Split into Dimension and Fact Tables and Views. (Star Schema)

dbt and all ETL Scripts are orchestrated using Airflow