Case Study: Real Estate | Data Engineering

Scenario:

As a Data Engineer employed by a real estate firm, your primary responsibility is to construct a data pipeline that efficiently extracts information from **various tables**, including leads and sales data. This pipeline will operate on a **daily basis**, leveraging the power of Airflow for seamless automation. Your role extends to **configuring** Airflow and **crafting** the necessary database tables to facilitate smooth **data ingestion**.

Data Tables:

Leads

Sales

Task:

- **Configure** Airflow locally.
- **Design** tables (**Star Schema**) on the Data Warehouse side and **create** an Entity-Relationship Diagram (**ERD**).
- **Develop** a data pipeline to **extract** Leads/Sales data.
- The **data pipeline** should also **Transform** the extracted data as necessary and **ingest** it into the database.

  Note: Input data and output data(tables) can be treated as files, simulating tables.

  Bonus: Establish a **local database** such as: SQL Lite, load sales and leads data into tables under a schema named ”**data_source**” , and configure Airflow to read from this database. Create output tables in the database under a schema named "**dwh**" and use Airflow to write data into the database.

Delivered Files:

- **DAG** as python file
- **ERD** Diagram