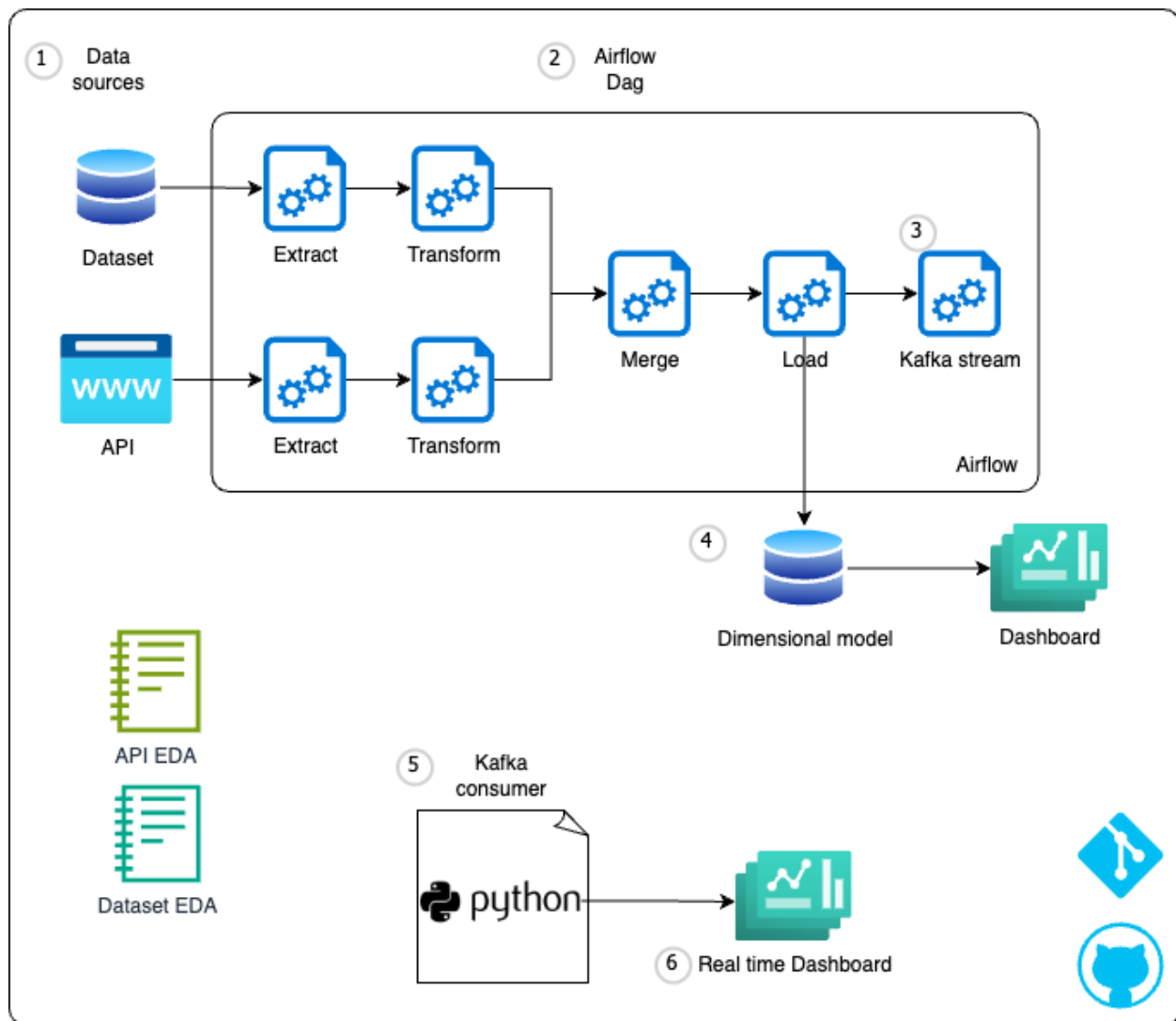# ETL class project - Final delivery

Introduction

The final delivery of the project course has six steps (as in the figure) described as follows:

1. Data sources: two different data sources selected in previous deliveries, the first one from a dataset, the second one from an API

2. Airflow DAG: A DAG with different tasks to extract, transform and load, the load task has to store the dimensional model into a database

3. Streaming: A streaming Kafka task to stream a metric from the Fact table

4. Dashboard: Get the data from the dimensional model created in the Airflow ETL DAG and create a dashboard in your preferred tool (Power BI , Looker Studio)

5. Kafka Consumer: In a python APP, instance the Kafka consumer to get the metric from the step 3

6. Real time dashboard: Connect the visualization tool you selected to the python app with the Kafka consumer and create a real time dashboard to visualize the streaming of the data

> NOTE: Have in mind that the whole project has to be in GitHub and you need to include the respective EDA notebooks created in previous deliveries

What is Expected

Is expected to have the complete pipeline working from the step 1 to the 6, evidences of the data in the database, dashboards and EDA notebooks , GitHub repository

Technologies

We expect you to use in this challenge:

- Python
- Jupiter Notebook
- Database (you choose)
- AirFlow

- Kafka

- CSV files and API consumption

- Visualization tool

- Git/Github