Coal Mining ELT Pipeline

Karunia Perjuangan Mustadl'afin

I. Background Problem

A coal mining company aims to optimize its mining operations using their production data. They also want to integrate external factors to further optimize analysis and build predictive analytic models in the form of forecasting. To solve that problem, the provided solution has to implement a unified data warehouse that collects data from various sources and transforms them with ETL/ELT pipeline so that the information can be presented as BI dashboards to help stakeholders devise strategic decisions for the coal mining company.

II. Overview

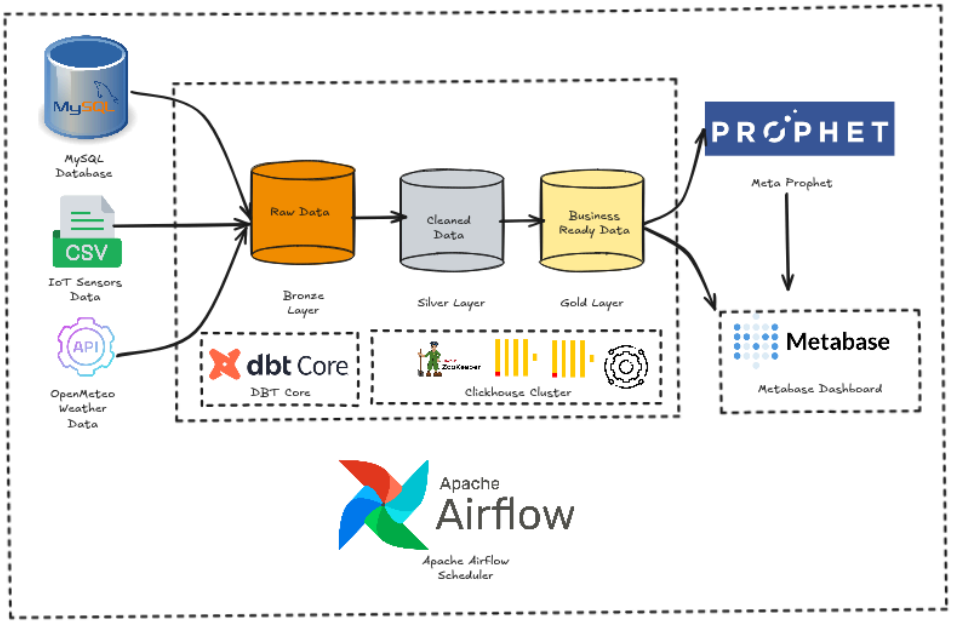
In this solution, we used the ELT approach which positioned transformation in the data warehouse instead of data source. This can help reduce the cost of the transformation as columnar based data warehouses are more efficient than row based RDBMS services in joins and aggregations.

This ELT pipeline mainly encompassed some data components:

a. **Data Sources**: The various forms of processed data sources: the operational data came from MySQL, the equipment monitoring data came from IoT sensors which served as a CSV file, and the external weather API which was extracted from OpenMeteo API service.
b. **Clickhouse**: An open-source state of the art OLAP database for processing large amounts of data.
c. **DBT Core**: A library used to perform data transformation automatically.
d. **Apache Airflow**: A library that can be used to schedule and run ELT processes sequentially.
e. **Prophet**: A library that can be used to predict future data based on existing data trends.
f. **Metabase**: A library that can be used to display business intelligence dashboards with minimal efforts.

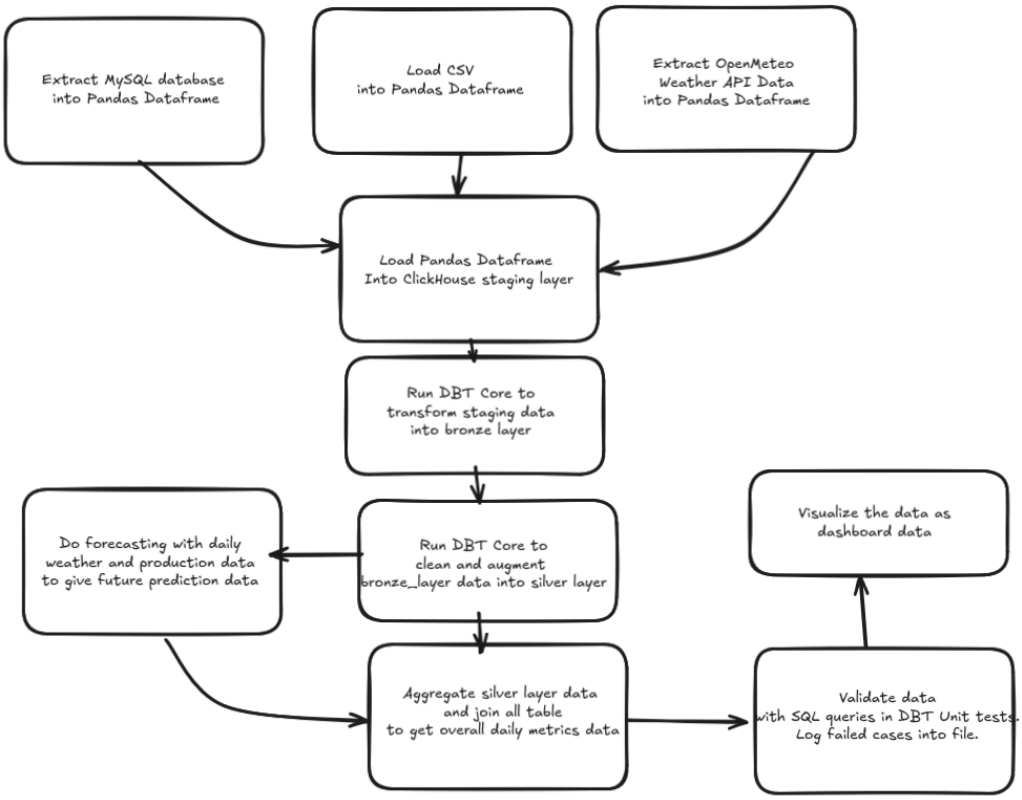
III. Architecture Design

The architecture diagram that represents the system interactions between components is shown below:



IV. ELT Process

The process for getting the data is shown below:



V. Data Validation

For validating data, DBT Core can be used to build unit tests in forms of SQL queries. The unit tests can raise exceptions during the transformations process if the constraints are not satisfied in the input and output of the transformations. The unit tests checks the condition below:

1. Whether equipment_utilization in the final daily metrics table is between 0 and 1
2. Whether every mining production day has their respective weather data matches
3. Whether every total production data every day is a non-negative number.

VI. Repository

The repository is link is https://github.com/karuniaperjuangan/coal-mining-synapsis

Main point of interests:

a. Main DAG file: Link
b. Python Extraction codes: Link
c. DBT transformations models and queries: Link
d. DBT Unit Tests: Link
e. Docker Compose file: Link