

Project Documentation: Lakehouse Data Pipeline

Overview

This project implements a structured data pipeline following the Medallion Architecture (Staging → Bronze → Silver → Gold) using PySpark and Lakehouse in Fabric. The pipeline handles raw HR_data, transforming it through different stages and making it analytics-ready for tools like Power BI.

Initial Setup

- Created four lakehouses representing each layer: Staging, Bronze, Silver, and Gold.
- Created a notebook that stores all lakehouse paths in variables to ensure dynamic access. This makes it easier to update paths without modifying each code block.

Layers and Transformation Steps

1. Staging to Bronze

Created a Notebook [Staging to bronze.ipynb](#) to raw data from staging lakehouse to bronze lake house in parquet format

2. Bronze to Silver

Reading the file loaded in bronze and performing cleaning and duplicating

And typecasting on the file and further converted it into delta table in

Silver lakehouse using [Bronze to silver.ipynb](#) file

3. Silver to Gold

Read the delta table from silver lakehouse and performed various transformations to create dimensions from the delta table by adding surrogacy key and required transformations and sending it to gold lakehouse using

[Silver_to_gold_dim.ipynb](#)

4. Fact Table Creation

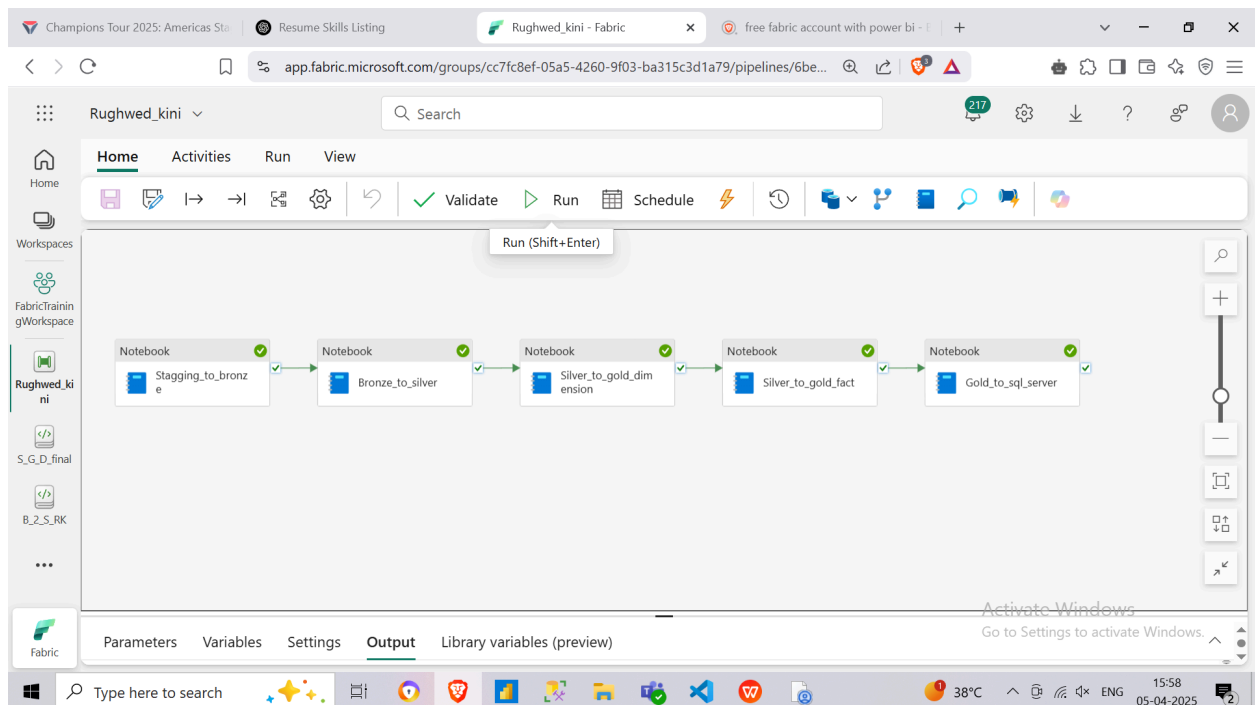
Reading the delta table as well as dimension table created I created fact_table in gold by join and transformations by using , [silver_to_gold_fact.ipynb](#)

5. Data Export to SQL Database

Now in order to load the data from gold lakehouse I create a azure sql server
And configured the connection using jdbc by using [Gold_to_ssms.ipynb](#)

6. Creating pipeline to automate the transformations

Now to automate all this process from ingestion to sql server i created a pipeline with help of this notebooks.



7. Creating Power bi dashboard

After successfully running the pipeline my date is readily available at my sql server from which i can load it into power bi desktop create dashboards from it for data driven decision taking.

