**Microsoft Fabric Project Documentation- Mohammeduzair Sayyed**

**Project Title: Enterprise Data Warehouse Implementation using Microsoft Fabric (Medallion Architecture)**

**Objective:** Design and implement a robust end-to-end data pipeline using Microsoft Fabric based on the Medallion Architecture, including Bronze, Silver, and Gold layers, with integrated reporting through Power BI.

**1. Project Architecture (Medallion Layers)**

**Bronze Lakehouse:**

* Raw data ingestion from ADLS Gen 2(Product.csv, Region.csv, Sales.csv).
* Used pipelines to copy each file into Lakehouse Bronze.
* Files had different delimiters. Couldn’t configure all types in one pipeline.
* 3 pipelines used. Invoke legacy used to combine them and use at once.
* Logging implemented to track file name, ingestion time, row count.

**Silver Lakehouse:**

* Data transformed and converted into Delta Tables using Notebook.
* File-specific transformations applied.
* Created Delta tables:
  + product\_delta
  + region\_delta
  + sales\_delta

**Gold Warehouse (Goal initially):**

* Intended for final tables, aggregations, views, and fact generation.
* SQL Scripts created:
  + Table creation scripts for product, region, sales.

**Issue:**

* Warehouse copy operations failed due to type mismatch (e.g., Parquet UTF8 vs FLOAT).
* COPY INTO failed, leading to alternate approach using Notebooks.
* Data Warehouse generation failed.
* SQL endpoint generation failed in lakehouse, couldn’t connect to PowerBI

**Gold Warehouse (Success):**

* **Used a pre-existing warehouse with SQL analytics endpoint generated.**
* **Used ‘Script’ feature to create empty tables pre-emptively into the warehouse with the same schema as the files uploaded.**
* **Used casting and converted all the attributes to VARCHAR to prevent any type mismatch issues.**
* **Used ‘copy data’ feature to copy delta tables from silver lakehouse into gold warehouse.**

**2. Key Learnings and Challenges**

**Interesting Findings:**

* Warehouse scripting in Fabric requires each CREATE VIEW or PROCEDURE to be the first statement in a batch.
* CREATE TABLE and IF NOT EXISTS do not work together in Fabric.
* Additional columns feature is very tricky. Whitespace is unacceptable and has a very vague error pointer. Eg. If column name is ‘Ingestion Time’ it should be ‘IngestionTime’.
* COPY INTO in Warehouse has strict data type compatibility issues with Parquet files.(UTF-8000) issues.
* Connecting Power BI via Warehouse or Lakehouse (Direct Lake) is seamless if SQL endpoints are generated. (Updated version below.)
* Fabric Notebooks can’t directly connect to Warehouse — only to Lakehouses.
* Fabric doesn’t allow delta tables directly into warehouses and doesn’t convert them intuitively into tables.

**Difficulties Faced:**

* COPY INTO Warehouse failing due to FLOAT column being read as UTF8 BYTE\_ARRAY.
* Warehouse table creation syntax needed to be precise (no IF NOT EXISTS).
* SQL views and procedures had strict syntactic placement issues.
* Mismatch between expected formats in Warehouse and actual Delta storage in Silver layer.
* SQL cannot be performed if analytics isn’t generated.
* Unable to get delta tables from Silver Lakehouse to Gold Lakehouse using notebooks.
* Couldn’t convert delta tables back to files.
* PowerBI cannot be connected due to above error
* No warehouse can be created in Fabric.(mostly)
* No lakehouse created (by me) had any sort of SQL analytics endpoint generated.
* Connecting to PowerBI is not reliable.
* Used warehouse url to connect to PowerBI,unable.
* Used import data and selected Azure SQL, still not working.
* Data pane doesn’t show any data,unable to perform live visualizations.
* Tried multiple ways to connect PowerBI but not possible. Connection succeeds but data pane is empty.
* Triggers are tricky in Fabric. Connection with ADLS Gen 2 should dictate the trigger to be Blob event but details are unaccepted.
* Subscription name and blob/lake name are not accepted. Whereas the same are accepted in copy data activity under source.

**3. Final Approach**

**Data Flow:**

1. **Ingest from ADLS Gen 2to Bronze** using Fabric Pipelines.
2. **Transform in Silver** using Notebooks (Delta format).
3. **~~Aggregate and Join in Notebook (PySpark + Spark SQL)~~** ~~using data from Silver.(not done)~~
4. **~~Write back aggregated data to Gold Lakehouse as Delta Tables~~**~~.~~
5. Created a new pipeline with scripts feature and created empty tables in warehouse.

6. Used Copy data to ingest delta tables into the empty tables from silver lakehouse to gold warehouse.

**~~Final Notebook Tasks:~~**

* ~~Join sales, product, region tables.~~
* ~~Perform aggregations:~~
* ~~Save results to Gold Lakehouse Delta Tables.~~

**4. Power BI Integration**

* Tried to connect to Gold Lakehouse via Direct Lake mode.
* Build reports using aggregated Delta tables.

**5. Files and Assets Created**

| **Asset Type** | **Name** | **Layer** |
| --- | --- | --- |
| Pipeline | ingest\_product\_to\_bronze | Bronze |
| Pipeline | ingest\_region\_to\_bronze | Bronze |
| Pipeline | ingest\_sales\_to\_bronze | Bronze |
| Delta Table | product\_delta | Silver |
| Delta Table | region\_delta | Silver |
| Delta Table | sales\_delta | Silver |
| SQL Table | product | Gold |
| SQL Table | region | Gold |
| SQL Table | sales | Gold |
| ~~Notebook~~ | ~~silver\_to\_gold\_aggregator~~ | ~~Silver input / Gold output~~ |

**Status:**

* Bronze layer completed.
* Silver layer completed.
* ~~Aggregations finalized.~~
* ~~Gold tables created via Notebook.~~
* ~~Final copy to Gold Lakehouse successful.~~
* Gold warehouse ready with views,data and aggregations.
* Power BI connection ready.

**Next Steps:**

* Final visualizations in Power BI.(Unable to connect)
* Unable to add a trigger to the pipeline to fully automate the pipeline.