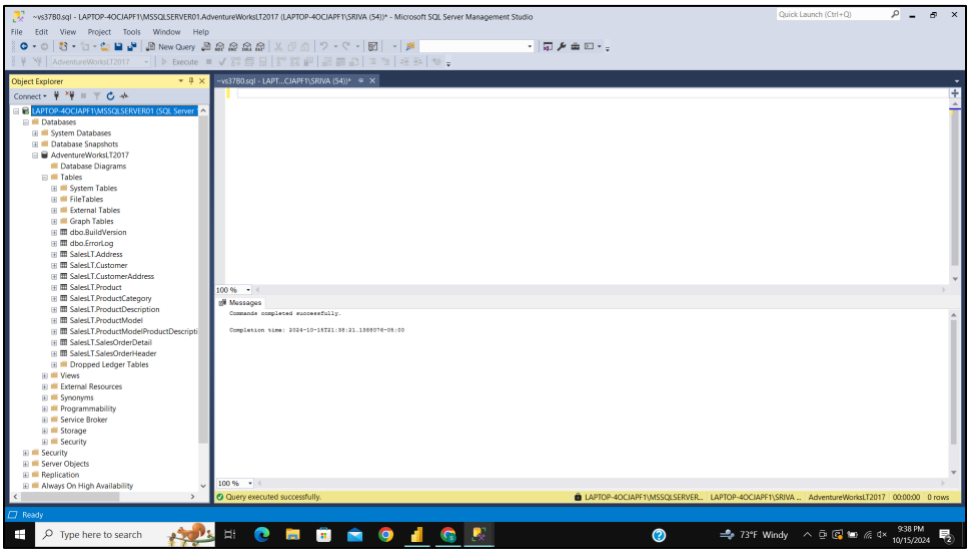
ETL project on Azure Cloud Platform

- SRI VARDHAN GOUD PADALA

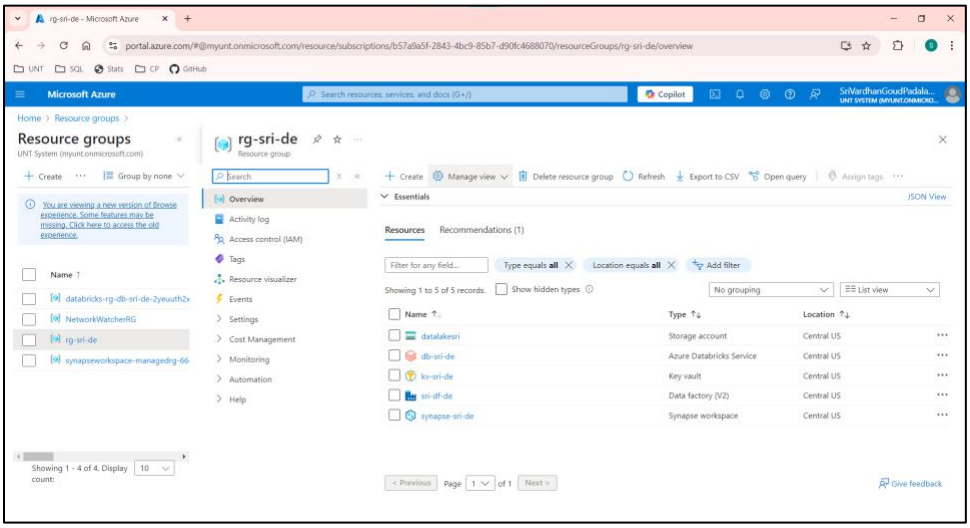
Setting up on-prem database

- Imported AdventureWorksLT2017 database to SQL server
- Created a user and login creds and assigned permission
- Stored the login creds in azure key vault in secrets section



Setting up Azure resources

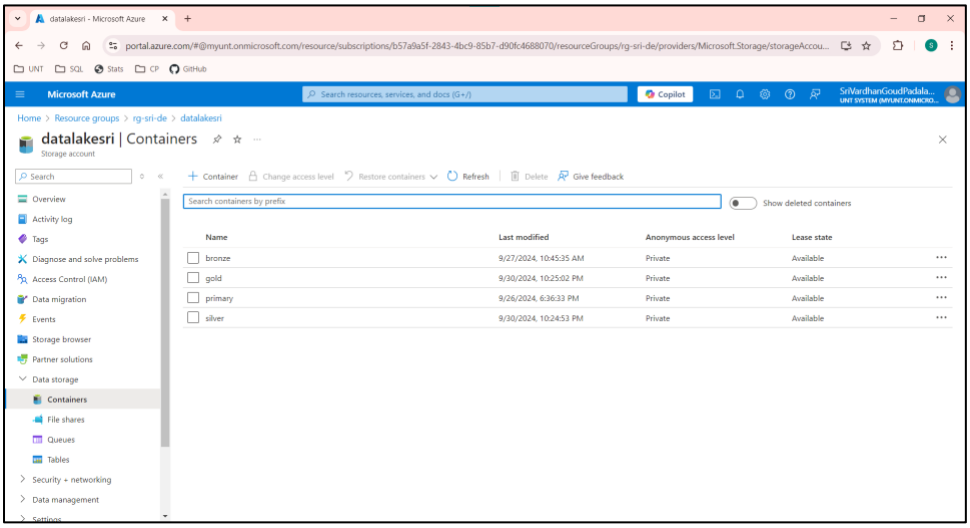
- Azure data lake gen 2 (provided by synapse analytics)
- Azure synapse analytics
- Azure data bricks
- Azure data factory
- Azure key vault

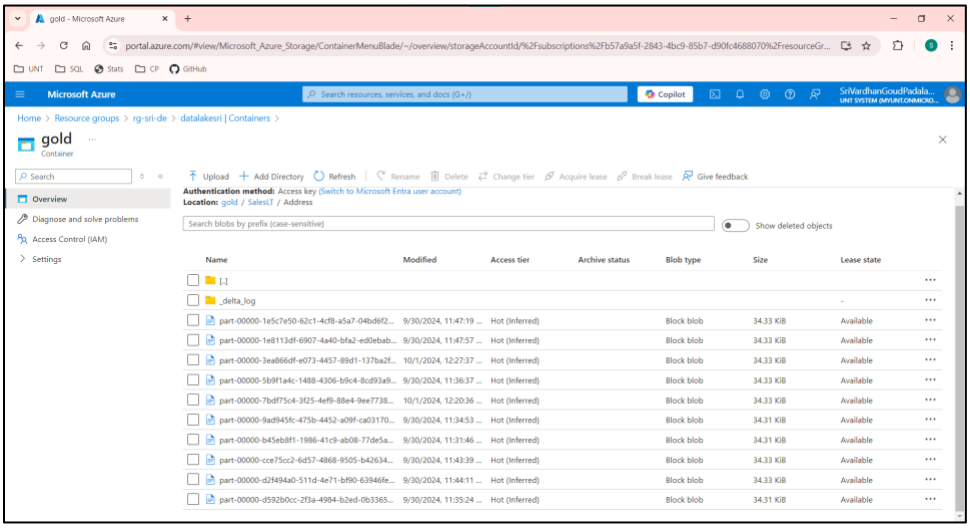


1.Azure Data Lake Gen 2

Created 3 different folders following a lake house architecture

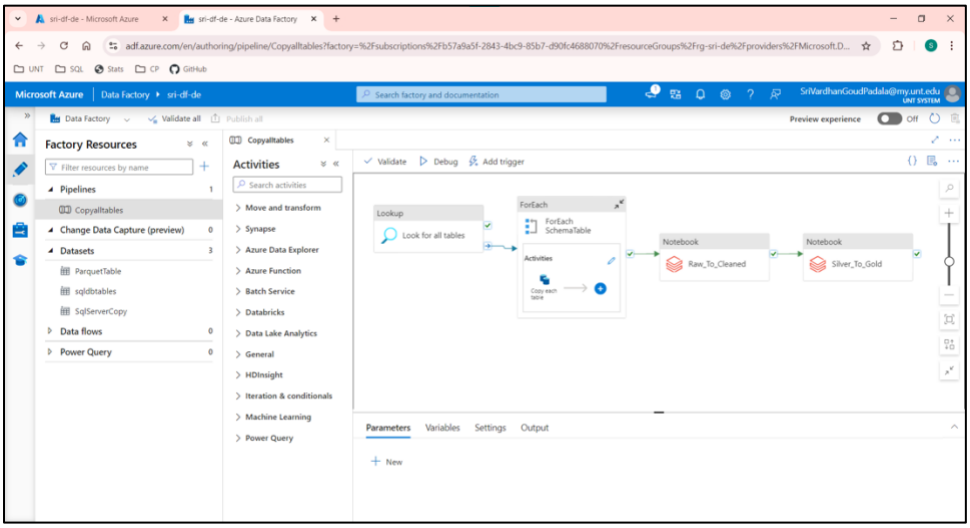
- Bronze: Exact Copy of Data ingested from on-prem database i.e. raw format of data
- Silver: Transformed data
- Gold: Aggregated Data

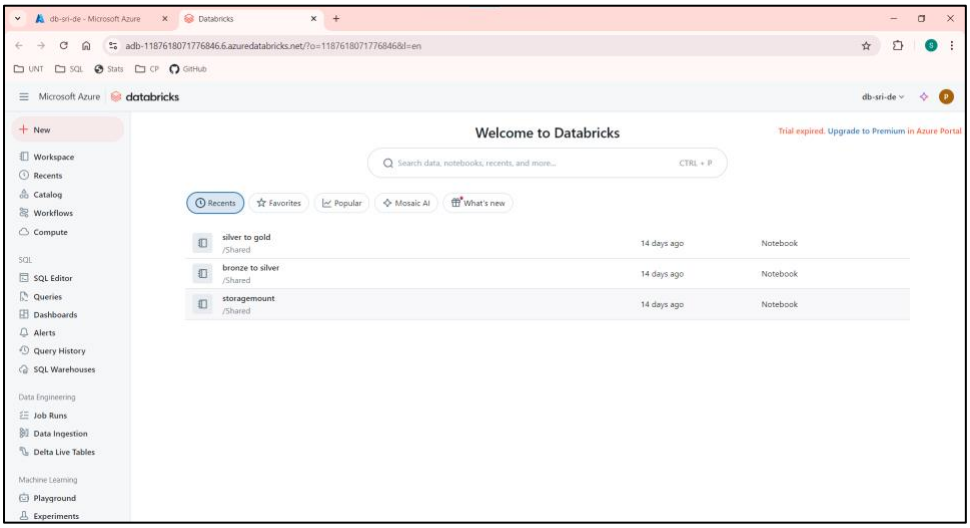




2. Azure Data Factory

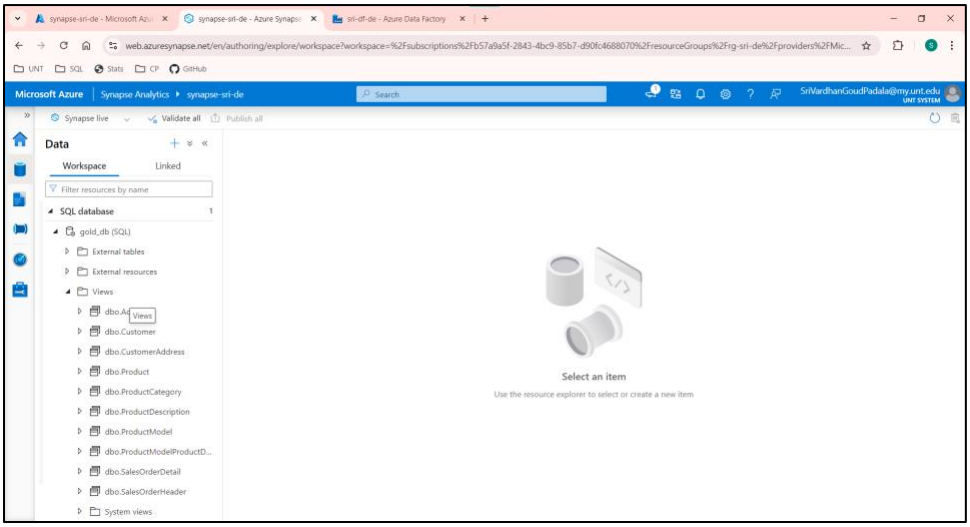
- Extracted Data from on-prem using copy data tool in ADF.
- Read the format of files and transferred in the same format to Azure Data Lake Gen2
- Utilized Databricks for transformations and cleaning.
- Stored the transformed data to Data lake

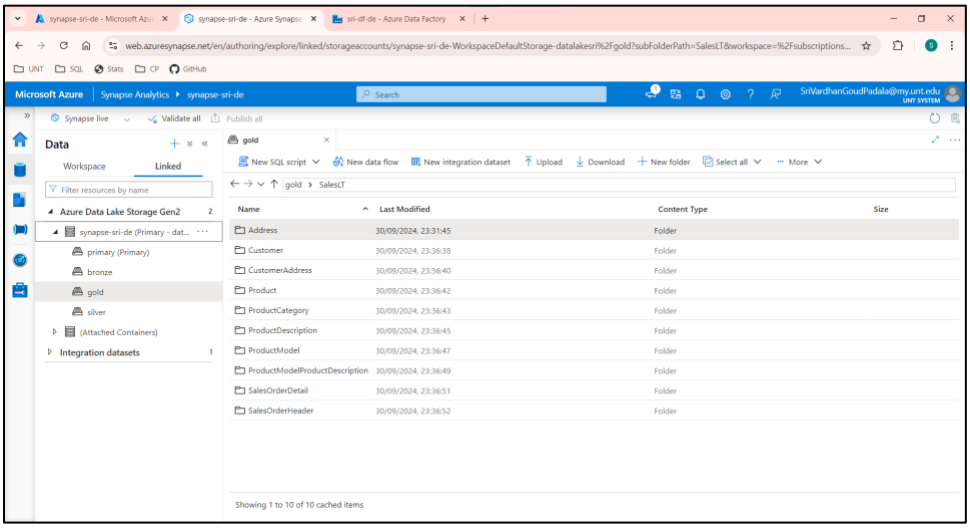




3.AZURE SYNAPSE ANALYTICS

- For further Analysis I transferred data to a synapse analytics Workspace.





NEXT STEPS

1. Analyze data in Synapse using serverless SQL pool.
2. Create visualization in power BI using data from synapse analytics workspace.