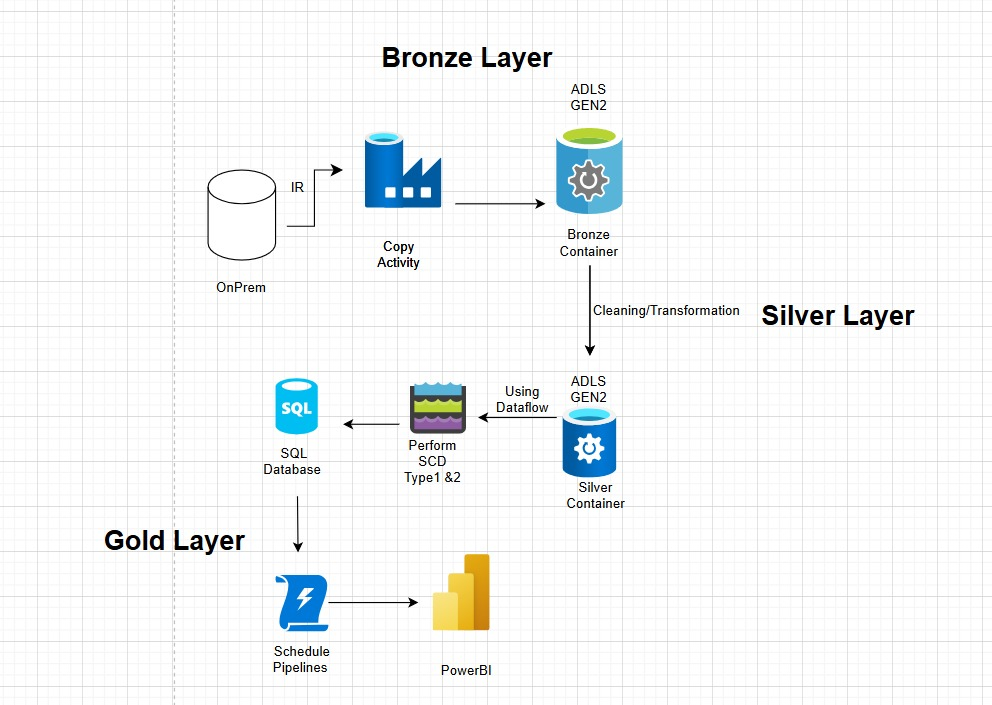
**Boot Camp Project – 1 Data Pipeline for Customer Account Analysis**

Master Pipeline

**Step 1: Data Ingestion (Backend Storage to Raw/Bronze Container)**

* **ADF Copy Activity:**
  + Configured to transfer data from the backend team's storage account to the raw (bronze) container in the data lake.
  + **Source Files:**
    - accounts.csv
    - customers.csv
    - loan\_payments.csv
    - loans.csv
    - transactions.csv
  + **Sink:** Raw (bronze) container in your ADLS account.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

Fig 1 : Integration Run time and Linked Service

A screenshot of a computer

AI-generated content may be incorrect.

Fig 2: Bronze Layer Pipeline

A screenshot of a computer

AI-generated content may be incorrect.

Fig 1 – Meta Data

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.Fig 3: Contents inside the Copy Activity

**Results**:

A screenshot of a computer

AI-generated content may be incorrect.

Files Stores in the CSV Format.

**Step 2: Data Cleaning with ADF Dataflows**

* **Read Data:**
  + Dataflows read from all five sources in the raw (bronze) container.
* **Remove Null Values:**
  + Used aggregate transformation, window, or assert activities to identify and remove null Values.
* **Data Transformation:**
  + Applied schema enforcement (avoiding inferSchema), data type conversions, and handled missing values.
  + Output stored in CSV Files for optimized querying.

A screenshot of a computer program

AI-generated content may be incorrect.

Fig 4: Silver Layer

A screenshot of a computer

AI-generated content may be incorrect.

Source of the File

A screenshot of a computer

AI-generated content may be incorrect.

Transformations to remove null values

A screenshot of a computer

AI-generated content may be incorrect.

Alter Row

**Step 3: Dataflows Using SCD Type Techniques (SCD 1 and SCD 2)**

* **Data Movement:**
  + Implemented Slowly Changing Dimension (SCD) Type 1 and Type 2 logic within Dataflows.
  + Data moved through the following layers:
    - Local → Bronze
    - Bronze → Silver
    - Silver → Gold
* **Upsert to SQL DB:**
  + Final cleaned and transformed data upserted into Azure SQL Database tables.
* **Scheduling:**
  + Pipelines scheduled for automated, repeatable data processing.

A screenshot of a computer

AI-generated content may be incorrect.

Fig 5: Gold Layer

A screenshot of a computer

AI-generated content may be incorrect.

Pipeline for Accounts, Customers under SCD Type1

A diagram of a company

AI-generated content may be incorrect.

Pipeline for Transactions, Loan payments, Loans

A screenshot of a computer

AI-generated content may be incorrect.

Transformations

A screenshot of a computer

AI-generated content may be incorrect.

Added Hash Key

A screenshot of a computer

AI-generated content may be incorrect.

Look Up Activity

A screenshot of a computer

AI-generated content may be incorrect.

Conditional Split

A screenshot of a computer

AI-generated content may be incorrect.

Audit the Columns

A screenshot of a computer

AI-generated content may be incorrect.

Alter Row

A screenshot of a computer

AI-generated content may be incorrect.

Sink

A screenshot of a computer

AI-generated content may be incorrect.

Union

A screenshot of a computer

AI-generated content may be incorrect.

Auditing the Columns

A screenshot of a computer

AI-generated content may be incorrect.

Sink for Insert

**Results**

**A screenshot of a phone

AI-generated content may be incorrect.**

Customers table output

A screenshot of a computer

AI-generated content may be incorrect.

Transaction Table data

**Step 4: Data Visualization with Power BI**

* **Data Source:**
  + Connected Power BI to Azure SQL Database tables.
* **Report Development:**
  + Built interactive visuals and dashboards based on processed data.
* **Publishing:**
  + Published the Power BI report to Fabric Workspace for sharing and collaboration.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

We need to add the Server Name

A screenshot of a computer

AI-generated content may be incorrect.

Select the Required Tables from the server

A screenshot of a computer screen

AI-generated content may be incorrect.A screenshot of a computer screen

AI-generated content may be incorrect.Fig 6: Accounts Dashboard

Fig 7: Customers Dashboard

A screenshot of a graph

AI-generated content may be incorrect.A screenshot of a graph

AI-generated content may be incorrect.Fig 8: Loan Payments Dashboard

Fig 9: Loans Dashboard

**Triggers:**

* **Scheduled the Trigger at 9 AM every day.**

A screenshot of a computer

AI-generated content may be incorrect.

Fig 10: Scheduled Trigger

**Master Pipeline Structure**

* **Master Pipeline:**
  + Built using ADF's Execute Pipeline activity.
  + Sequentially calls child pipelines for each layer (Bronze, Silver, Gold) to ensure orderly data processing.

**A screenshot of a computer

AI-generated content may be incorrect.**

Fig 11: Master Pipeline

Git Hub – luckyaniketh08@gmail.com