**Databricks PySpark Script for Handling Slowly Changing Dimensions (SCD Type 1)**

**Overview**

This script is designed to process and update a Delta table in Databricks using Azure Blob Storage. It implements **SCD Type 1**, ensuring that new records are inserted while existing records are updated based on a hash key mechanism.

Upload files to storage account(I have used Blob storage ) A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

* **Mount Azure Blob StorageA screenshot of a computer code

  Description automatically generatedMounts Azure Blob Storage** to Databricks.
* **Security Concern:** Use **Azure Key Vault** instead of hardcoding credentials.

**2. List Mounted Storage Locations & Files**

* A screenshot of a computer

  Description automatically generatedLists all **mounted storage locations**.
* Lists **files inside the mounted storage**.

**3. Read a CSV File into a Data FrameA screenshot of a computer

Description automatically generated**

* Reads a CSV file from **Azure Blob Storage**.
* **option("header", "true")** ensures column names are taken from the first row.
* **option("inferSchema", "true")** enables automatic schema detection.
* **4. Create a Delta Table (if not exists)** **A screenshot of a computer

  Description automatically generated**Creates a **Delta table** for customer data.
* Stores data in **Delta format**, enabling ACID transactions & time travel. A screenshot of a computer

  Description automatically generated

**5. Define Source and Target PathsA screenshot of a computer

Description automatically generated**

* Uses a **Databricks Widget** to accept filedate as input.
* Constructs **source and target paths dynamically**.

## **6. Read Source Data and Generate Hash KeyA screenshot of a computer Description automatically generatedA screenshot of a computer Description automatically generated7. Load Delta Table and Perform Anti-Join**

* **A screenshot of a computer

  Description automatically generatedA screenshot of a computer

  Description automatically generated**Loads **existing Delta table** as a DataFrame.
* **Anti-Join** filters out records that already exist **unchanged**.
* Keeps **only new or updated records**.
* **8. Merge Changes into the Delta TableA screenshot of a computer code

  Description automatically generatedMerges new and updated records into Delta table**.
* **Updates existing records** where ID matches.
* **Inserts new records** where no match is found.

## **9. Display the Updated Delta Table**

A screenshot of a computer

Description automatically generatedDay2

A screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screen shot of a computer

Description automatically generated