**Data handling with SQL Database, ADLS gen2 and Databricks**

**Objective:**

* Build a data pipeline to extract data from an SQL database, clean it by removing duplicates, and store it in Azure Data Lake Storage (ADLS) Gen2 using Databricks.
* Create SCD TYPE 1 logic using the Delta Tables in Databricks

**Task**

1. **Extract Data from SQL Database**
   * Establish a connection to the SQL database using JDBC.
   * Retrieve and verify the data.
2. A screenshot of a computer

   Description automatically generatedNow go to key vault , object 🡪 secretes🡪create a secrete for sql url , password and usernameA screenshot of a computer

   Description automatically generatedgo to key vault setting 🡪 properties, copy vault URL and Resource ID A screenshot of a computer

   Description automatically generatedNow launch Databricks , Create secrete scope for key vault using URL and resource ID from A screenshot of a computer

   Description automatically generatedNow check if the scope is created successfully or notA screenshot of a computer

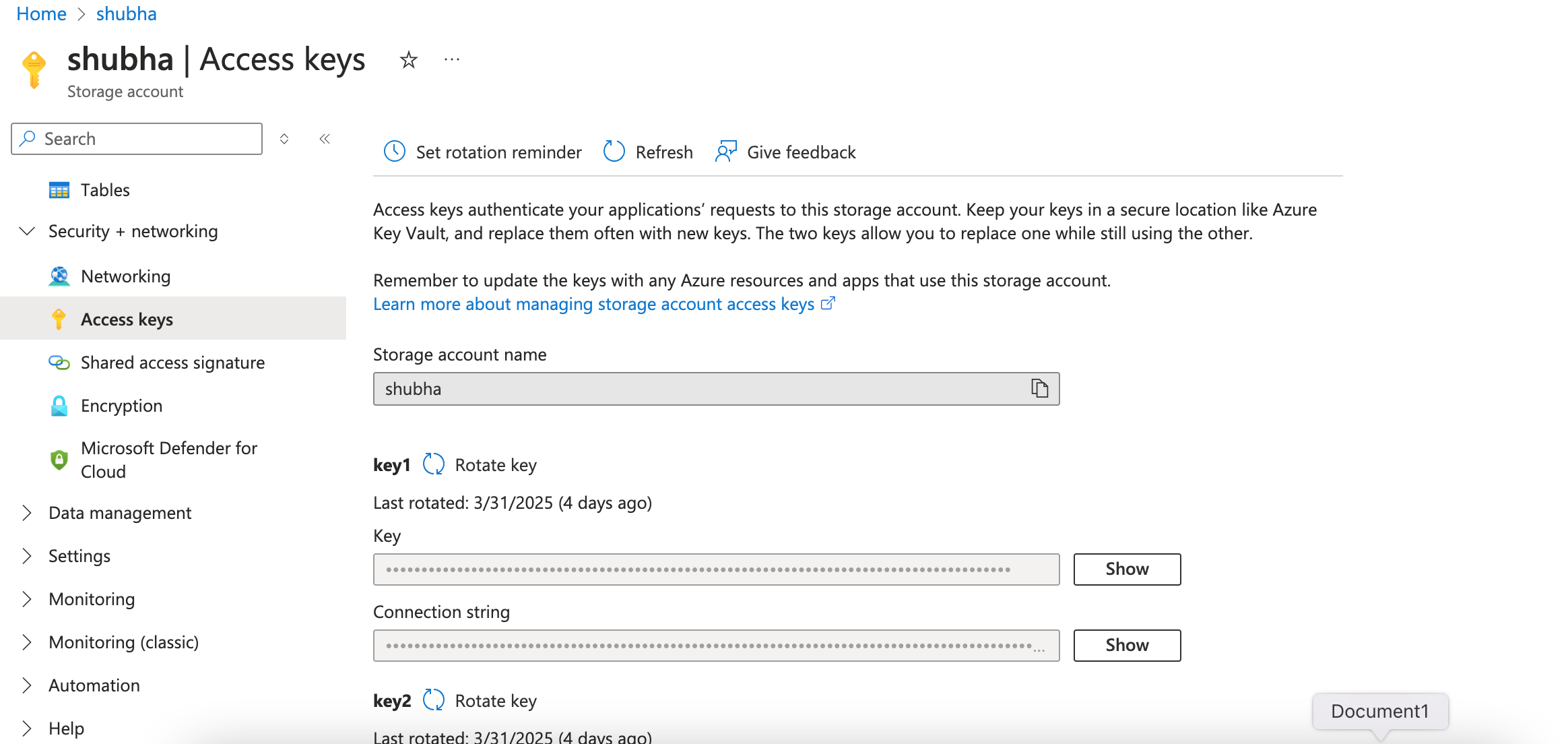
   Description automatically generatedNow go to SQL Database and create credit table and insert few records A screenshot of a computer

   Description automatically generatedNow create a SQL connection in DatabricksA screenshot of a computer program

   Description automatically generatedget data from sql using dataframeA screenshot of a computer

   Description automatically generated**Set Up ADLS Gen2 Mount Point**
   * Create an ADLS Gen2 container.
   * Configure a Databricks mount point for ADLS Gen2 access.

To create a mount to adls gen2 storage account, create a secrete key with access key

check the connectionA screenshot of a computer

Description automatically generated

create a folder in adls gen 2 A screenshot of a computer

Description automatically generatedNow mount adls gen2 storage account A close-up of a computer

Description automatically generated

A white background with a white line

Description automatically generated with medium confidence Now write a data frame which we created in adls gen2

A screen shot of a computer code

Description automatically generatedRead the data which we added in gen2 A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedRemove null A screenshot of a computer

Description automatically generatedRemove duplicate values A screenshot of a computer

Description automatically generatedNow writing this new data frame working into different folder A screen shot of a computer code

Description automatically generatedRead and display the cleaned data A screenshot of a computer program

Description automatically generatedNow create a SCD type1 logic using cleaned data

Now create a delta table A white screen with red text

Description automatically generated A screenshot of a computer

Description automatically generatedNow read the clean data and assign it to the dataframe A screenshot of a computer

Description automatically generatedcreate a hashkey for each column A screenshot of a computer

Description automatically generated A screenshot of a computer program

Description automatically generated A screenshot of a computer program

Description automatically generated A screen shot of a computer program

Description automatically generated A screenshot of a computer

Description automatically generated