**Project Description:**

A blue rectangle with white text

Description automatically generated

In this mini project, I utilized Azure Data Factory to develop a pipeline that transfers incremental data from an Azure SQL Database table to Azure Data Lake storage. The process involves several key steps:

1. Setting up the Data Lake Storage account: Prepare a storage solution to maintain the watermark value, which is crucial for tracking data updates.
2. Creating Azure SQL Database
3. Creating a Data Factory.
4. Creating Linked Services.
5. Configuring Datasets.
6. Building a Pipeline.
7. Executing the Pipeline: Understand how to initiate the pipeline operation.
8. Monitoring the Process: Learn how to track the pipeline's performance and output

Run below queries on SQL database:

CREATE TABLE product\_data

(

ProductID int,

ProductName varchar(255),

LastModifyTime datetime

);

INSERT INTO product\_data

(ProductID, ProductName, LastModifyTime)

VALUES

(1, 'UltraHD 4K Television','2021-07-15 10:30:00 AM'),

(2, 'Bluetooth Stereo Headphones','2021-07-16 11:45:00 AM'),

(3, 'Smartphone 128GB','2021-07-17 01:10:00 PM'),

(4, 'Wireless Charging Pad','2021-07-18 02:25:00 PM'),

(5, 'Noise Cancelling Earbuds','2021-07-19 03:40:00 PM'),

(6, 'Portable External SSD 1TB','2021-07-20 08:15:00 AM'),

(7, 'Smartwatch Fitness Tracker','2021-07-21 09:50:00 AM'),

(8, 'Gaming Laptop 16GB RAM','2021-07-22 04:05:00 PM'),

(9, 'Professional DSLR Camera','2021-07-23 05:20:00 PM'),

(10, 'Robot Vacuum Cleaner','2021-07-24 07:35:00 PM');

CREATE TABLE watermarktable (TableName varchar(255), WatermarkValue datetime);

INSERT INTO watermarktable VALUES ('product\_data','1/1/2010 12:00:00 AM');

CREATE PROCEDURE usp\_update\_watermark @LastModifiedtime datetime, @TableName varchar(50)

AS

BEGIN

UPDATE watermarktable

SET [WatermarkValue] = @LastModifiedtime

WHERE [TableName] = @TableName

END

*# Note: The stored procedure usp\_update\_watermark is designed to update a watermark value in a database table, typically used for tracking the most recent modification times of various tables*

*#parameters: @LastModifiedtime datetime: This parameter represents the timestamp that was used as the new watermark. It should be the most recent modification time that you want to record for a specific table. @TableName varchar(50): This parameter specifies the name of the table for which the watermark is being updated. It identifies the target table in the watermark tracking system*

*# Purpose and Use*

*#Tracking Changes: This procedure is typically used in systems where tracking the last modification time of various tables is important. For instance, in data synchronization scenarios, ETL processes, or data warehousing, where you need to keep track of when each table was last updated.*

*#Watermark Table: The watermarktable is assumed to be a system or utility table designed to store watermark information for one or more tables. It should have at*

*Note: use to create a unique file name  @CONCAT('Incremental-', pipeline().RunId, .txt')*

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

select \* from [dbo].[product\_data] where LastModifytime > '@{activity('LookupOldWaterMarkActivity').output.firstRow.WatermarkValue}' and LastModifytime <= '@{activity('LookupNewWaterMarkActivity').output.firstRow.NewWatermarkvalue}'

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

@{activity('LookupNewWaterMarkActivity').output.firstRow.NewWatermarkvalue}  
@{activity('LookupOldWaterMarkActivity').output.firstRow.TableName}

Run the pipeline

then insert data

INSERT INTO [dbo].[product\_data] VALUES (11, 'newdata','9/6/2023 2:23:00 AM')

INSERT INTO [dbo].[product\_data] VALUES (12, 'newdata','9/7/2023 9:01:00 AM')

Check the result: Below are screenshots confirming the pipeline ran successfully.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated