Case study1:

**Case Study 1: Retail Sales Data Processing (using Azure blob storage and Azure Data Factory)**

**Client:** XYZ Retail Inc.

**Objective:** To streamline the process of ingesting, transforming, and storing retail sales data for reporting and analysis.

**Background:** XYZ Retail Inc. collects daily sales data from multiple store locations in CSV format. This data needs to be processed, aggregated, and transformed before being loaded into a new storage container for analysis.

**Requirements:**

1. **Data Storage:**
   * **Source Container:** source-data in Azure Blob Storage where raw CSV files will be uploaded.
   * **Destination Container:** processed-data in Azure Blob Storage where transformed CSV files will be stored.
2. **Data Processing:**
   * **Data Transformation:** Perform basic transformations such as data cleaning (e.g., removing null values), aggregating sales data by date, and filtering irrelevant columns.
   * **Data Movement:** Copy the processed data from the source container to the destination container.
3. **Automation:**
   * **Pipeline Execution:** Automate the ingestion, transformation, and storage process using Azure Data Factory.
   * **Scheduling:** Run the pipeline daily to process new data files.

**High-Level Steps:**

**1. Set Up Azure Blob Storage**

1. **Create Storage Accounts:**
   * **Storage Account Name:** xyzretailstorage
   * **Resource Group:** RetailDataGroup
   * **Region:** Choose an appropriate region (e.g., East US)
2. **Create Containers:**
   * **Source Container:** source-data
   * **Destination Container:** processed-data
3. **Upload Sample Data:**
   * Upload sample CSV files with sales data to the source-data container for initial testing.

**2. Configure Azure Data Factory**

1. **Create Data Factory:**
   * **Data Factory Name:** RetailDataFactory
   * **Resource Group:** RetailDataGroup
   * **Region:** Same as storage account (e.g., East US)
2. **Create Linked Services:**
   * **Azure Blob Storage Linked Service:**
     + Configure connection to xyzretailstorage.
     + Use the storage account connection string.
3. **Create Datasets:**
   * **Source Dataset:**
     + Data source: source-data container
     + Format: CSV
     + Schema: Configure based on the CSV file
   * **Destination Dataset:**
     + Data source: processed-data container
     + Format: CSV
     + Schema: Define or infer schema based on transformation.

**3. Build and Configure Pipeline**

1. **Create Pipeline:**
   * **Pipeline Name:** SalesDataPipeline
   * **Activities:**
     + **Copy Data Activity:**
       - **Source:** Link to source-data dataset.
       - **Sink:** Link to processed-data dataset.
       - **Transformation:** Optionally, configure transformations like column mapping and data filtering.
2. **Optional Data Flow:**
   * Create a Data Flow if more complex transformations are needed.
   * Add steps for data cleaning (e.g., removing null values) and aggregation (e.g., summing sales by date).

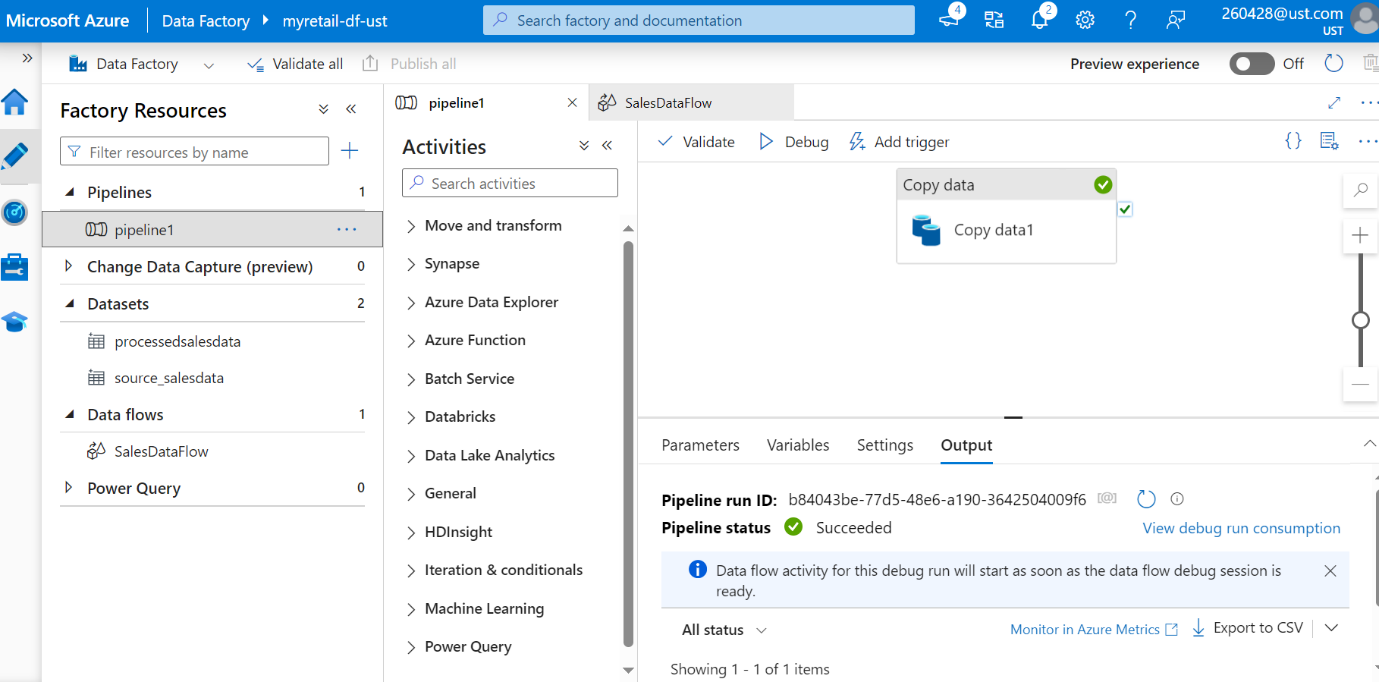
**4. Automate and Monitor**

1. **Create Trigger:**
   * **Trigger Type:** Schedule
   * **Frequency:** Daily
   * **Start Time:** Set the start time according to your business requirements (e.g., after daily data uploads)
2. **Monitor Pipeline:**
   * Check the **"Monitor"** section in Azure Data Factory to track the execution and ensure successful data processing.

Azure data factory:

A screenshot of a computer

Description automatically generated



o/p:

A screenshot of a computer

Description automatically generated