Architecture diagram for pipeline

The overview of the process is

1. Linked Service & Blob Account (Source)

- Linked Service: Acts as a connection configuration in Azure Data Factory (ADF) to external sources like Azure Blob Storage.
- Blob Account: Contains the source files (`sales_data.csv`, `product_master.csv`) in a specific container or folder.
- ADF uses this linked service to read the files during ingestion.

2. Scheduled Trigger

- Purpose: Automates the pipeline execution.
- Set to run daily, hourly, or on a specific schedule.

3. Ingest CSV Files from Blob Storage

- Uses Copy Activity to move or load the raw CSV files from Blob Storage to:
  - Staging Layer in Azure Data Lake
  - Or directly into Azure SQL DB for small datasets
- This step ensures data is available for transformation and analysis.

4. Data Cleaning & Transformation (ADF Data Flow / Azure Databricks)

This is the heart of the processing logic.

- ADF Data Flow (No-code) or Databricks Notebook (Code-based, scalable)

5. Store Results in Data Lake / Azure SQL DB

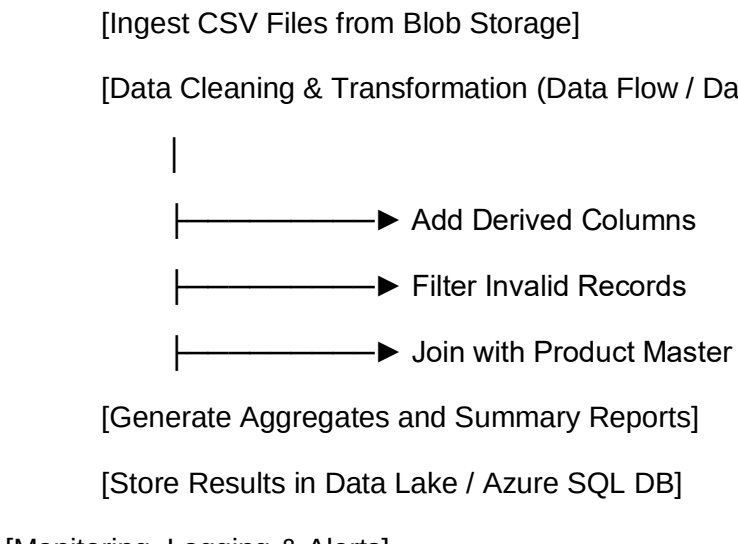
- Store processed output in:
  - A curated layer in Azure Data Lake Storage (ADLS) (for downstream analytics)
  - Or in Azure SQL Database (for dashboarding with Power BI)
- Files can be saved as CSV, or table records.

6. Monitoring, Logging & Alerts

- Built-in ADF Monitoring dashboard shows run status, duration, errors.

This is simple **Architecture diagram**

[Scheduled Trigger]


[Ingest CSV Files from Blob Storage]
[Data Cleaning & Transformation (Data Flow / Databricks)]
Add Derived Columns
Filter Invalid Records
Join with Product Master
[Generate Aggregates and Summary Reports]
[Store Results in Data Lake / Azure SQL DB]
[Monitoring, Logging & Alerts]