**Azure Data Factory ELT Demo Lab**

**Scenario**

The business receives **raw sales CSV files** in **Azure Blob Storage**.

* You must **Extract** and **Load** these raw files into an **Azure SQL Database staging table** using ADF.
* Then you will **Transform** the data directly in SQL (ELT) by running stored procedures for aggregation.

**Step 1: Prepare Resources**

1. **Azure Blob Storage**
   * Create a container named raw.
   * Upload a file named sales.csv:

SaleID,Product,Quantity,Price,Date

1,Laptop,2,500,2025-09-01

2,Mouse,10,20,2025-09-01

3,Laptop,1,500,2025-09-02

4,Keyboard,5,50,2025-09-02

1. **Azure SQL Database**
   * Create database SalesELT.
   * Create two tables:
2. -- Staging table (raw load)
3. CREATE TABLE SalesRaw (
4. SaleID INT,
5. Product VARCHAR(50),
6. Quantity INT,
7. Price DECIMAL(10,2),
8. SaleDate DATE
9. );
10. -- Final summary table
11. CREATE TABLE ProductSalesSummary (
12. ProductName VARCHAR(50),
13. TotalQuantity INT,
14. TotalSales DECIMAL(18,2)
15. );
16. **Stored Procedure for Transformation**
17. CREATE PROCEDURE sp\_AggregateSales
18. AS
19. BEGIN
20. INSERT INTO ProductSalesSummary (ProductName, TotalQuantity, TotalSales)
21. SELECT
22. Product,
23. SUM(Quantity),
24. SUM(Quantity \* Price)
25. FROM SalesRaw
26. GROUP BY Product;
27. END;

**Step 2: Create Linked Services in ADF**

1. **Blob Storage Linked Service** (account key auth).
2. **Azure SQL DB Linked Service** (SQL auth).
3. Test both connections.

**Step 3: Create Datasets**

1. **Blob Dataset** → CSV file in raw container.
2. **SQL Dataset** → SalesRaw table.

**Step 4: Create Pipeline (Extract & Load)**

1. Create a new pipeline named PL\_ELT\_SalesLoad.
2. Add a **Copy Data Activity**:
   * **Source:** Blob dataset (CSV).
   * **Sink:** SQL dataset (SalesRaw).
   * Enable **Table Auto Create = False** (we already created staging table).
3. Publish the pipeline.

**Step 5: Add Transformation Step (SQL Stored Proc)**

1. After the Copy Activity, add a **Stored Procedure Activity**.
2. Configure:
   * Linked Service: Azure SQL DB.
   * Stored Procedure: sp\_AggregateSales.
3. Connect Copy → Stored Proc.

**Step 6: Run & Monitor**

1. Manually **trigger pipeline run**.
2. Go to **Monitor Hub** → confirm Copy + Stored Proc execution.
3. Run SQL query to validate results:
4. SELECT \* FROM ProductSalesSummary;

Expected Output:

| **ProductName** | **TotalQuantity** | **TotalSales** |
| --- | --- | --- |
| Laptop | 3 | 1500.00 |
| Mouse | 10 | 200.00 |
| Keyboard | 5 | 250.00 |

**Step 7: Add Trigger**

1. Create a **Schedule Trigger** to run pipeline daily at midnight.
2. Associate with PL\_ELT\_SalesLoad.

**Step 8: Error Handling & Retry**

1. On the Copy Activity → set **Retry = 3, Interval = 60 sec**.
2. Break the connection string → rerun to see retries.
3. Fix connection and rerun successfully.

**Step 9: Cost Estimation**

1. Open **Monitor Hub → Activity Runs → Check Duration & IR usage**.
2. Use **Azure Pricing Calculator** to estimate cost:
   * 1 Copy Activity + 1 Stored Procedure Activity per day.
   * 30 runs per month.