Hands-On Lab 2: Aggregate Transformation in Azure Data Factory

Lab Created by : Dr. Sandeep Kumar Sharma

What This Lab Will Do

This lab focuses ONLY on **Aggregation** operations using Mapping Data Flows.
You will load a simple Orders dataset and calculate total quantity and total revenue (quantity × price) grouped by product.

This is a very clean, simple, single-purpose lab.

Step 1 — Prepare Storage and Upload File

Upload the following file into folder `lab2/` in your Storage Account.

orders_extended.csv

```
order_id,product,quantity,unit_price
1,Keyboard,2,500
2,Mouse,1,300
3,Keyboard,1,500
4,Monitor,2,7000
5,Mouse,3,300
```

Step 2 — Create Dataset in ADF

1. Go to **Author → Datasets → New**.
2. Create dataset **DelimitedText**.
3. Name: `ds_orders_lab2`
4. File: `lab2/orders_extended.csv`
5. First row header = True.

Step 3 — Create Mapping Data Flow

1. Go to **Data Flows → New mapping data flow**.
2. Name it: **df_aggregate_lab2**.

Add Source

- Add a source named `src_orders`.
- Select dataset: `ds_orders_lab2`.

Add Derived Column (Compute revenue)

We first compute order_value.

1. Add **Derived Column** transformation.
2. Add new column:
3. Name: `order_value`
4. Expression: `toInteger(quantity) * toInteger(unit_price)`

Add Aggregate Transformation

1. Add **Aggregate** after Derived Column.
2. Group By:
3. `product`
4. Aggregations:
5. `total_quantity` = `sum(toInteger(quantity))`
6. `total_revenue` = `sum(order_value)`
7. `order_count` = `count()`

The aggregation will output one row per product.

Step 4 — Add Sink

1. Add **Sink**.
2. Create dataset pointing to `lab2/output/aggregate/`.
3. Name: `ds_output_lab2`.
4. Keep mapping as default.

Step 5 — Debug & Run

1. Enable **Data Flow Debug**.
2. Check **Aggregate** preview.
3. Publish and Trigger the pipeline.

Expected Output

```
product,total_quantity,total_revenue,order_count
Keyboard,3,1500,2
Mouse,4,1200,2
Monitor,2,14000,1
```

Lab Completed

This lab covered ONLY the **Aggregate transformation** — grouping and computing sums/counts.

Move to Lab 3 for Conditional Splits & Other Functions.