Hands-On Lab 4: Window Transformations in Azure Data Factory

Trainer: Dr. Sandeep Kumar Sharma

What This Lab Will Do

This lab will teach you ONLY one concept — **Window Transformations** in Azure Data Factory Mapping Data Flows.

You will learn how to: - Apply **row ranking** (ROW_NUMBER) - Calculate **running totals** - Create **moving averages**

All using a very small, simple dataset.

Step 1 — Prepare Storage and Upload File

Upload this file into `lab4/` folder in your Storage Account.

sales.csv

```
day,sales_amount
1,100
2,200
3,150
4,300
5,250
```

Step 2 — Create Dataset in ADF

1. Go to **Author → Datasets → New**.
2. Choose **DelimitedText**.
3. Name: `ds_sales_lab4`
4. File path: `lab4/sales.csv`
5. First row header = True.

Step 3 — Create Mapping Data Flow

Go to **Data Flows → New Mapping Data Flow**. Name: **df_window_lab4**

Add Source

- Name: `src_sales`
- Dataset: `ds_sales_lab4`

Part A — Row Number Using Window Function

Goal: Assign row numbers based on the `day` sequence.

Steps

1. Add **Window** transformation.
2. Name it `win_rownum`.
3. Window settings:
4. Window type: **Row rank**
5. Ordered by: `day` ascending
6. Output column will be: `rowNumber`.

Part B — Running Total

Goal: Compute cumulative sales.

Steps

1. Add another **Window** transformation.
2. Name: `win_running_total`.
3. Window function type: **Aggregate**.
4. Group By: none.
5. Ordered By: `day` ascending.
6. Aggregation:
7. `running_total = sum(sales_amount)` with **Window framing**:
    - *From:* Unbounded Preceding
    - *To:* Current Row

Part C — Moving Average (Last 3 Days)

Goal: Compute average sales for a 3-day sliding window.

Steps

1. Add **Window** transformation.
2. Name: `win_moving_avg`.
3. Window type: **Aggregate**.
4. Ordered by: `day` ascending.
5. Aggregation:
6. `moving_avg = avg(sales_amount)`.
7. Window frame:
8. *From:* 2 rows before
9. *To:* Current Row

Step 4 — Add Sink

1. Add **Sink** transformation.
2. Create dataset: `ds_output_lab4` → write to folder `lab4/output/window/`.
3. Keep mapping default.

Step 5 — Debug & Run

1. Enable **Data Flow Debug**.
2. Preview Window function results.
3. Publish.
4. Trigger pipeline.

Expected Output

```
day,sales_amount,rowNumber,running_total,moving_avg
1,100,1,100,100
2,200,2,300,150
3,150,3,450,150
4,300,4,750,216.67
5,250,5,1000,233.33
```

Lab Completed

This lab covered ONLY Window functions: ✔️Row Number
✔️Running Total
✔️Moving Average

Next lab will cover Pivot/Unpivot transformations.