Hands-On Lab 3: Conditional Split, Lookups, and Basic Functions in Azure Data Factory

Lab Created by : Dr. Sandeep Kumar Sharma

What This Lab Will Do

This lab focuses ONLY on three simple transformations:

1. **Conditional Split** – route rows into different outputs based on rules.
2. **Lookup Transformation** – enrich rows with small reference data.
3. **Basic Functions** – use expressions like `toInteger`, `concat`, `iif`, etc.

Step 1 — Prepare Storage and Upload Files

Upload the following files into the folder `lab3/`.

1. orders.csv

```
order_id,customer_id,order_value
1,101,2000
2,102,500
3,103,12000
4,101,800
5,104,25000
```

2. customer_tiers.csv

```
customer_id,customer_tier
101,Gold
102,Silver
103,Bronze
```

Note: Customer 104 has no tier — useful to show Lookup with "no match".

Step 2 — Create Datasets in ADF

Create these datasets:

- `ds_orders_lab3` → lab3/orders.csv
- `ds_customer_tiers_lab3` → lab3/customer_tiers.csv
- `ds_output_high_lab3` → lab3/output/high_value/
- `ds_output_low_lab3` → lab3/output/low_value/
- `ds_output_lookup_lab3` → lab3/output/lookup_enriched/

Format = Delimited Text First Row Header = True

Step 3 — Create Mapping Data Flow

Go to **Data Flows → New Mapping Data Flow**. Name it: **df_other_transforms_lab3**

Add Source

- Name: `src_orders`
- Dataset: `ds_orders_lab3`

Part A — Conditional Split

Goal

Separate high-value orders (> 10,000) from low-value orders.

Steps

1. Add **Conditional Split** transformation.
2. Define rules:
3. **HighValue**: `toInteger(order_value) >= 10000`
4. **LowValue**: `true()` (catch-all)

Add Sinks

- Connect HighValue output → Sink to folder `output/high_value/`
- Connect LowValue output → Sink to folder `output/low_value/`

Part B — Lookup Transformation

Goal

Enrich orders with `customer_tier` using lookup file.

Steps

1. From `src_orders`, add a **Lookup** transformation.
2. Lookup dataset: `ds_customer_tiers_lab3`.
3. Lookup condition:
4. `src_orders.customer_id == ds_customer_tiers_lab3.customer_id`
5. Lookup behavior: **Allow no match**.

Add Sink

- Sink dataset: `ds_output_lookup_lab3`

Orders with no tier (like customer_id 104) will have NULL tier.

Part C — Use Basic Functions

Examples of functions you can test in a **Derived Column** transformation:

1. `status_flag`

```
iif(order_value >= 10000, 'High', 'Normal')
```

2. `order_value_double`

```
toInteger(order_value) * 2
```

3. `customer_label`

```
concat('Cust-', customer_id)
```

Add a **Derived Column** transformation if you want to apply these.

Step 4 — Debug & Run

1. Enable **Debug Mode**.

2. Test Conditional Split outputs.
3. Preview Lookup results.
4. Publish.
5. Trigger pipeline.

Expected Outputs

High Value Orders

Orders ≥ 10,000:

```
order_id,customer_id,order_value
3,103,12000
5,104,25000
```

Low Value Orders

```
order_id,customer_id,order_value
1,101,2000
2,102,500
4,101,800
```

Lookup Enriched Output

```
order_id,customer_id,order_value,customer_tier
1,101,2000,Gold
2,102,500,Silver
3,103,12000,Bronze
4,101,800,Gold
5,104,25000,
```

Lab Completed

This lab covered three basic transformations clearly and simply: ✔️Conditional Split
✔️Lookup
✔️Basic Functions (Derived Columns)