Hands-On Lab 6: Flatten & Parse JSON in Azure Data Factory

Lab Created by : Dr. Sandeep Kumar Sharma

Concept Overview (Before HOL)

Modern systems often store data in **nested JSON format**. ADF Mapping Data Flows provide: - **Parse JSON** → Converts JSON strings into structured columns - **Flatten** → Expands arrays into individual rows

You use these when working with: - API responses - Logs - Semi-structured documents - IoT device data

This lab teaches how to extract meaningful tabular data from nested JSON.

What This Lab Will Do

You will: 1. Parse a JSON column into structured fields 2. Flatten an array inside the JSON 3. Output clean rows

---

Step 1 — Upload File to Storage

Upload this file into `lab6/` .

orders_nested.json

```
{
  "orders": [
    { "order_id": 1, "customer": "A", "items": [ {"product": "Keyboard", "qty":
2}, {"product": "Mouse", "qty": 1} ] },
    { "order_id": 2, "customer": "B", "items": [ {"product": "Monitor", "qty":
1} ] }
  ]
}
```

Step 2 — Create Dataset

- Dataset: `ds_orders_json_lab6`
- File: `lab6/orders_nested.json`

- Format: **JSON**

Step 3 — Create Mapping Data Flow

Name: **df_flatten_json_lab6**

Add Source

- Source: `src_json`
- Dataset: `ds_orders_json_lab6`

The source will load the root object.

Part A — Parse JSON

1. Add **Derived Column** transformation
2. Add new column:
3. Name: `parsed_orders`
4. Expression: `fromJson(byName('orders'), 'array<struct<order_id:int, customer:string, items:array<struct<product:string, qty:int>>>')`

Now `parsed_orders` is a structured array.

Part B — Flatten Array of Orders

1. Add **Flatten** transformation
2. Unroll by: `parsed_orders`

Output will produce rows:

```
order_id, customer, items
```

Part C — Flatten Items Inside Each Order

1. Add second **Flatten** transformation
2. Unroll by: `parsed_orders.items`

Final output will look like:

```
order_id,customer,product,qty
1,A,Keyboard,2
1,A,Mouse,1
2,B,Monitor,1
```

Step 4 — Add Sink

Sink dataset: `lab6/output/flatten/` Format: CSV

Step 5 — Debug & Run

- Start debug session
- Inspect Flatten output
- Publish
- Trigger

Lab Completed

This lab covered: ✔️Parsing JSON strings ✔️Flattening nested arrays ✔️Extracting structured tabular data from JSON