Hands-On Lab 8: Handling Nulls & Error Rows in Azure Data Factory

Lab Created by : Dr. Sandeep Kumar Sharma

Concept Overview (Before HOL)

In real ETL pipelines, data is often **dirty**, **incomplete**, or **invalid**. ADF Mapping Data Flows provides tools to: - Handle **NULL values** safely - Replace missing data with defaults - Divert invalid rows using **Error Handling** - Perform safe type conversions (`toIntegerSafe`, `toStringSafe`)

This lab demonstrates how data engineers clean and rescue bad rows.

What This Lab Will Do

You will learn: 1. How to replace NULLs using `iif(isNull())` 2. How to use **Derived Column** to handle missing values 3. How to use the **Error Handling** tab in each transformation 4. How to create a clean output + bad records output

---

Step 1 — Upload File

Upload this file to `lab8/`.

customers_dirty.csv

```
customer_id,name,age
101,Sandeep,34
102,,29
103,Varun,abc
104,Meena,
```

Problems: - Missing names - Non-numeric age (“abc”) - Null age

---

Step 2 — Create Dataset

Dataset name: `ds_customers_lab8` File: `lab8/customers_dirty.csv` Format: CSV

Step 3 — Create Mapping Data Flow

Name: **df_nulls_errors_lab8**

Add Source

- Name: `src_customers`
- Dataset: `ds_customers_lab8`

Part A — Replace NULLs

Add **Derived Column** transformation and define:

Replace missing name

```
clean_name = iif(isNull(name) or name == '', 'UNKNOWN', name)
```

Safe conversion for age

```
clean_age = toIntegerSafe(age)
```

Replace null ages

```
clean_age = iif(isNull(clean_age), 0, clean_age)
```

Part B — Error Row Handling

In the **Derived Column** → *Error Handling* panel: - Choose **Redirect rows** - Redirect invalid transformation rows to a new stream: `error_stream`

Example errors caught: - Invalid number conversion for age - Missing required columns

---

Part C — Add Sinks

1. **Clean records sink**
2. Path: `lab8/output/clean/`

3. **Error records sink**

4. Path: `lab8/output/errors/`

Step 4 — Debug & Run

1. Enable Debug
2. Verify clean and error outputs
3. Publish & Trigger

Expected Output

Clean Output

```
customer_id,clean_name,clean_age
101,Sandeep,34
102,UNKNOWN,29
104,Meena,0
```

Error Output

```
customer_id,name,age,errorMessage
103,Varun,abc,"Failed to convert age"
```

Lab Completed

This lab covered: ✔️NULL handling
✔️Safe conversions
✔️Redirecting error rows
✔️Producing clean + bad data outputs