



Document: Metadata-Driven ADF Pipeline – Ingest Excel Sales Files



GOAL

Create **one reusable Azure Data Factory (ADF) pipeline** that dynamically ingests **Sales Excel files** from **multiple source systems** (e.g., SAP, Senior) into **raw SQL tables** using metadata stored in a config table.



DATABASE TABLES

1. raw.IngestionConfig

Column	Purpose
SourceSystem	e.g., 'SAP', 'Senior'
FolderPath	ADLS path to the folder
FilePattern	e.g., <code>SAP_Sales_*.xlsx</code>
RawTable	Target raw table name
IsActive	If set to 1, this config is active

Used to drive folder, pattern, and table dynamically.

2. raw.Sap_Sales and raw.Senior_Sales

```
CREATE TABLE raw.Sap_Sales (  
  OrderID INT,  
  Customer NVARCHAR(100),  
  Amount DECIMAL(18,2),  
  OrderDate DATE,  
  IngestedOn DATETIME DEFAULT GETDATE()  
);
```

```
CREATE TABLE raw.Senior_Sales (  
  OrderID INT,  
  Client NVARCHAR(100),  
  Total DECIMAL(18,2),  
  Date DATE,
```

```
    IngestedOn DATETIME DEFAULT GETDATE()  
);
```

3. `raw.FileIngestionLog`

```
CREATE TABLE raw.FileIngestionLog (  
    LogId INT PRIMARY KEY IDENTITY(1,1),  
    SourceSystem NVARCHAR(50),  
    FileName NVARCHAR(255),  
    Status NVARCHAR(50),  
    ProcessedOn DATETIME DEFAULT GETDATE()  
);
```

PIPELINE: `pl_ingest_sales_data`

Pipeline Parameter

Name	Type	Example
<code>SourceSystem</code>	String	'SAP'

STEP 1: Lookup1 - Fetch Metadata

Query:

```
SELECT * FROM raw.IngestionConfig  
WHERE SourceSystem = '@pipeline().parameters.SourceSystem' AND IsActive = 1
```

Output:

```
{  
  "FolderPath": "Sales/SAP",  
  "FilePattern": "SAP_Sales_*.xlsx",  
  "RawTable": "raw.Sap_Sales"  
}
```

STEP 2: Get Metadata1 - List All Files in Folder

Dataset path: `@activity('Lookup1').output.firstRow.FolderPath` **Output:**

```
{
  "childItems": [
    {"name": "SAP_Sales_20240620.xlsx"},
    {"name": "SAP_Sales_20240621.xlsx"}
  ]
}
```

STEP 3: ForEach - Loop over Files

Inside loop:

► Get Metadata2 - Fetch Last Modified

Inputs: `folderPath` from Lookup1, `fileName` = `@item().name` **Output:**

```
{
  "lastModified": "2024-06-21T08:15:00Z"
}
```

Used to determine the latest file.

STEP 4: Set Variable - IsValidFileName

```
@if(
  startsWith(variables('LatestFileName'),
    activity('Lookup1').output.firstRow.FilePatternPrefix),
  'true',
  'false'
)
```

STEP 5: Lookup2 - Check If File Already Processed

```
SELECT COUNT(1) AS FileExists
FROM raw.FileIngestionLog
```

```
WHERE FileName = '@variables('LatestFileName')'
AND SourceSystem = '@pipeline().parameters.SourceSystem'
```

Output:

```
{
  "FileExists": 0
}
```

 STEP 6: If Condition - Proceed to Process

```
@and(
  equals(variables('IsValidFileName'), 'true'),
  equals(activity('Lookup2').output.firstRow.FileExists, 0)
)
```

 STEP 7 (TRUE): CopyToRaw - Load Data

Source Dataset: Excel file\ **Sink Dataset:** SQL Table\ **Table Name:**

```
@activity('Lookup1').output.firstRow.RawTable
```

Uses Auto Mapping.

 STEP 8 (TRUE): Stored Proc - Log Success

```
EXEC raw.sp_log_file_ingestion
  @SourceSystem = '<SourceSystem>',
  @FileName = '<LatestFileName>',
  @Status = 'Success';
```

 ELSE: Log as Skipped

```
EXEC raw.sp_log_file_ingestion
  @SourceSystem = '<SourceSystem>',
```

```
@FileName = '<LatestFileName>',  
@Status = 'Skipped';
```

TRIGGERS

Name	Time	Parameter
Trigger_SAP_8AM	8:00 AM	SAP
Trigger_Senior_12PM	12:00 PM	Senior

Each passes SourceSystem to the same pipeline.

SMART FEATURES

Feature	Benefit
Metadata-driven design	Add new source systems easily without new pipelines
One pipeline	Centralized, reusable, low maintenance
Dynamic ADLS access	Folder and file detection handled automatically
Auto Mapping in Copy	No custom column mapping required
Logging mechanism	Auditable, idempotent ingestion process
Parameterized triggers	Separate schedules per source without duplication

SCALING TO NEW SOURCES

To add a new source system like "Oracle":

1. Add entry in `raw.IngestionConfig`
2. Create `raw.Oracle_Sales` table
3. Create trigger `Trigger_Oracle_6PM` (pass `SourceSystem='Oracle'`)
4. Done — pipeline will ingest automatically.