

# ***Dynamic pipeline design***

## **Project Assignment - 03**

Aurabrato Ghosh

Azure Data Engineering

### **Table of Contents**

Objective .....	2
Step-by-Step Implementation .....	2
Step 2: Metadata Control Table .....	3
Step 3: Integration Runtime Setup .....	4
Step 4: Linked Services .....	4
Step 5: Pipeline Design in Synapse .....	5
Step 6: Testing & Validation .....	6
Errors & Resolutions .....	7
Conclusion .....	7
References .....	7

## Objective

Design a dynamic pipeline to:

- Extract data from 5 on-premises SQL Server tables.
- Load the extracted data into Azure Data Lake Storage Gen2.
- Copy the data back to a designated C: drive folder on the local machine.
- Use a metadata control table to drive logic.
- Configure Logic App for failure notifications.
- Automate VM creation and deletion for data movement tasks.

## Step-by-Step Implementation

### Step 1: On-Prem SQL Server Setup

- Created a new SQL Server VM using a trial subscription.
- Installed SSMS, created database Project3, and added 5 tables:
  - Students
  - Courses
  - Enrollments
  - Payments
  - Tickets
- Inserted 3–4 sample rows in each table.

```
-- Create Students table
CREATE TABLE dbo.Students (
    StudentID INT PRIMARY KEY,
    FullName NVARCHAR(100),
    Email NVARCHAR(100)
);
INSERT INTO dbo.Students VALUES
(1, 'Alice Johnson', 'alice@example.com'),
(2, 'Bob Smith', 'bob@example.com'),
(3, 'Charlie Davis', 'charlie@example.com');
```

**Image 1: SQL query to create Students table and insert data**

```
-- Create Courses table
CREATE TABLE dbo.Courses (
    CourseID INT PRIMARY KEY,
    CourseName NVARCHAR(100),
    Instructor NVARCHAR(100)
);
INSERT INTO dbo.Courses VALUES
(101, 'Data Engineering 101', 'Dr. Stone'),
(102, 'Cloud Computing Basics', 'Prof. Lin'),
(103, 'Azure Fundamentals', 'Dr. Rao');
```

**Image 2: SQL query to create Courses table and insert data**

```
-- Create Enrollments table
CREATE TABLE dbo.Enrollments (
    EnrollmentID INT PRIMARY KEY,
    StudentID INT,
    CourseID INT,
    EnrollmentDate DATE
);
INSERT INTO dbo.Enrollments VALUES
(1001, 1, 101, '2025-05-01'),
(1002, 2, 102, '2025-05-02'),
(1003, 3, 103, '2025-05-03');
```

**Image 3: SQL query to create Enrollments table and insert data**

```
-- Create Payments table
CREATE TABLE dbo.Payments (
    PaymentID INT PRIMARY KEY,
    StudentID INT,
    Amount DECIMAL(10, 2),
    PaymentDate DATE
);
INSERT INTO dbo.Payments VALUES
(501, 1, 150.00, '2025-05-10'),
(502, 2, 200.00, '2025-05-12'),
(503, 3, 175.00, '2025-05-13');
```

**Image 4: SQL query to create Payments table and insert data**

```
-- Create Tickets table
CREATE TABLE dbo.Tickets (
    TicketID INT PRIMARY KEY,
    StudentID INT,
    Issue NVARCHAR(255),
    Status NVARCHAR(50)
);
INSERT INTO dbo.Tickets VALUES
(301, 1, 'Login issue with portal', 'Resolved'),
(302, 2, 'Unable to access course material', 'Open'),
(303, 3, 'Billing query', 'In Progress');
```

**Image 5: SQL query to create Tickets table and insert data**

## Step 2: Metadata Control Table

- Created dbo.MetadataControl with columns: TableName, SchemaName, FileName, FolderPath.
- Inserted rows for the 5 tables with folder and file paths.

```
CREATE TABLE dbo.MetadataControl (
    TableName NVARCHAR(100),
    SchemaName NVARCHAR(50),
    FileName NVARCHAR(100),
    FolderPath NVARCHAR(200)
);

INSERT INTO dbo.MetadataControl VALUES
('Students', 'dbo', 'students.csv', 'project3/students/'),
('Courses', 'dbo', 'courses.csv', 'project3/courses/'),
('Enrollments', 'dbo', 'enrollments.csv', 'project3/enrollments/'),
('Payments', 'dbo', 'payments.csv', 'project3/payments/'),
('Tickets', 'dbo', 'tickets.csv', 'project3/tickets/');

SELECT * FROM dbo.MetadataControl
```

**Image 6: Metadata table Creation SQL query in VM SSIS**

### Step 3: Integration Runtime Setup

- Created a Self-Hosted Integration Runtime named SQL-OnPrem.
- Installed and registered the runtime inside the VM using Key1.
- Ensured service was connected.

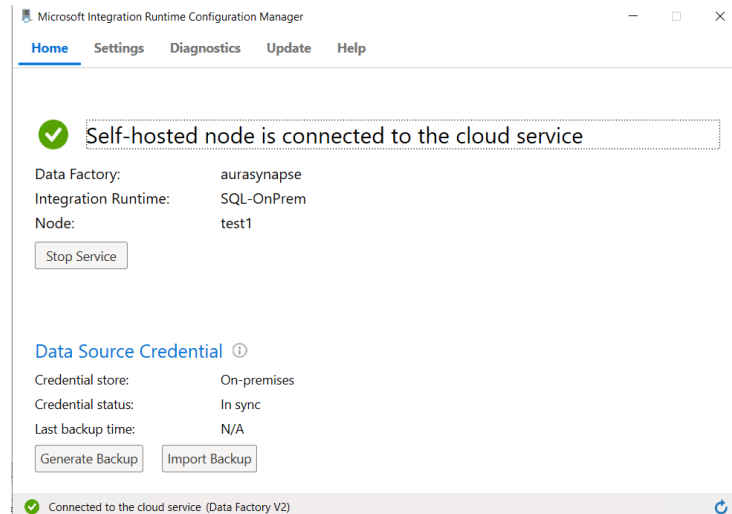


Image 7: SQL SH integration on VM

### Step 4: Linked Services

- Linked Service 1: On-prem SQL (uses Self-hosted IR)
- Linked Service 2: ADLS Gen2 (AutoResolve IR)

Name \*

LS\_SQL\_Project3\_OnPrem

Description

Connect via integration runtime \* ⓘ

✓ SQL-OnPrem

Version

☒ 2.0 (Recommended) ☐ 1.0

Server name \*

localhost

Database name \*

Project3

Authentication type

SQL authentication

User name \*

aughosh

Image 8: On-premises SQL Server linked service

Name \*  
LS\_ADLSGen2

Description

Connect via integration runtime \* ⓘ  
✓ AutoResolveIntegrationRuntime

Authentication type  
Account key

Account selection method ⓘ  
☐ From Azure subscription
 ☒ Enter manually

URL \*  
https://aura1adls.dfs.core.windows.net/

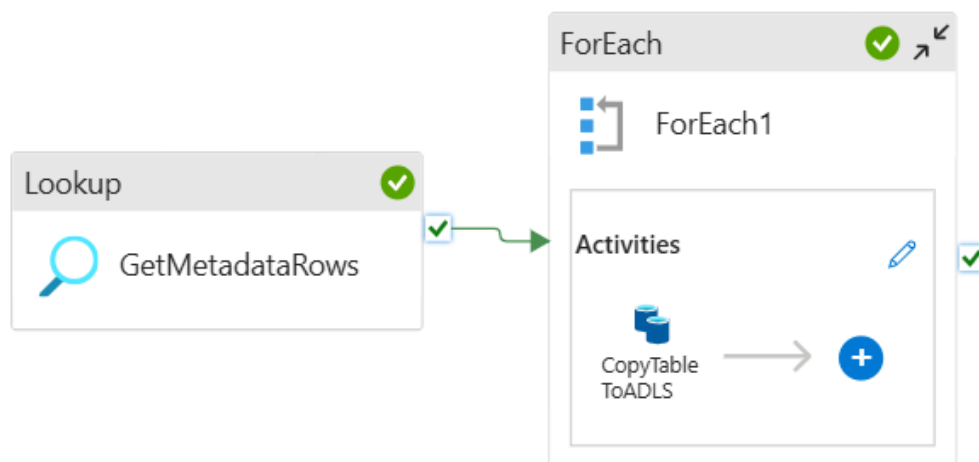
Storage account key Azure Key Vault

Storage account key \*  
.....

**Image 9: Azure ADLS gen2 linked service**

## Step 5: Pipeline Design in Synapse

- Created a pipeline named PL\_Project3\_Load.
- Components used:
  - Lookup Activity to read metadata rows.
  - ForEach Activity to loop through each table.
  - Copy Activity to copy from on-prem SQL to ADLS.



**Image 10: Pipeline design in Azure Synapse Analytics**

Step 6: Testing & Validation

- Executed pipeline and verified:
  - Files landed in the right folders in ADLS
  - File names followed metadata configuration
- Confirmed rows matched from SQL tables.

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓
CopyTableToADLS	✔ Succeeded	Copy data	6/16/2025, 4:36:20 AM	10s	SQL-OnPrem
CopyTableToADLS	✔ Succeeded	Copy data	6/16/2025, 4:36:06 AM	13s	SQL-OnPrem
CopyTableToADLS	✔ Succeeded	Copy data	6/16/2025, 4:35:51 AM	14s	SQL-OnPrem
CopyTableToADLS	✔ Succeeded	Copy data	6/16/2025, 4:35:39 AM	10s	SQL-OnPrem
CopyTableToADLS	✔ Succeeded	Copy data	6/16/2025, 4:35:18 AM	19s	SQL-OnPrem
ForEach1	✔ Succeeded	ForEach	6/16/2025, 4:35:17 AM	1m 15s	
GetMetadataRows	✔ Succeeded	Lookup	6/16/2025, 4:35:03 AM	13s	SQL-OnPrem

Image 11: Successful pipeline execution run

sink1 > exported-tables

Authentication method: Access key (Switch to Microsoft Entra user account)

Showing all 4 items

<input type="checkbox"/>	Name	Last modified
<input type="checkbox"/>	📁 [..]	
<input type="checkbox"/>	📄 customer.csv	6/11/2025, 2:56:08 AM
<input type="checkbox"/>	📄 orders.csv	6/11/2025, 2:56:34 AM
<input type="checkbox"/>	📄 products.csv	6/11/2025, 2:56:58 AM
<input type="checkbox"/>	📄 suppliers.csv	6/11/2025, 2:57:22 AM

Image 12: Successful copy of tables from On-premises SQL Server to Azure ADLS Gen2 storage on cloud

## Errors & Resolutions

- OutOfMemoryException
  - Cause: Copying large datasets with default config
  - Resolution: Reduced dataset volume; restarted host
- No process is on the other end of the pipe
  - Cause: SQL Server not reachable
  - Resolution: Restarted Integration Runtime and verified connection
- Runtime shows Unavailable
  - Cause: Old IR registration active
  - Resolution: Deleted and re-registered IR with correct name

## Conclusion

This project successfully demonstrates a dynamic metadata-driven pipeline design, leveraging self-hosted integration runtime to bridge on-premises and cloud data infrastructure.

## References

- Azure Synapse Pipelines Documentation  
<https://learn.microsoft.com/en-us/azure/synapse-analytics/pipelines/overview>
- Self-hosted Integration Runtime in Azure Data Factory  
<https://learn.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime>
- How to Register Integration Runtime on a Virtual Machine  
<https://learn.microsoft.com/en-us/azure/data-factory/configure-self-hosted-integration-runtime-windows>
- Create and Manage Linked Services in Synapse  
<https://learn.microsoft.com/en-us/azure/synapse-analytics/data-integration/concepts-linked-services>
- Azure Logic Apps for Notifications  
<https://learn.microsoft.com/en-us/azure/logic-apps/logic-apps-overview>