Data Ingestion and Loading in Azure Synapse Pipeline

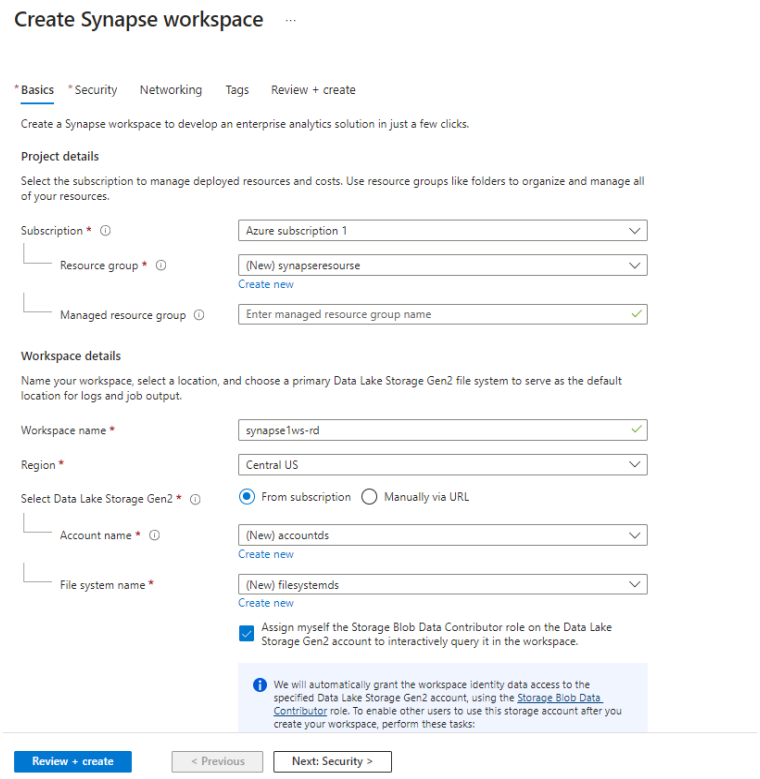
Design and implement a Synapse Pipeline that ingests data from an Azure SQL Database into Azure Data Lake Storage (ADLS), followed by loading the data into a Synapse Dedicated SQL Pool for analysis and processing.

Step 1: Open Azure Portal

Step 2: Go to search and type “Azure Synapse Analytics” and select “create”

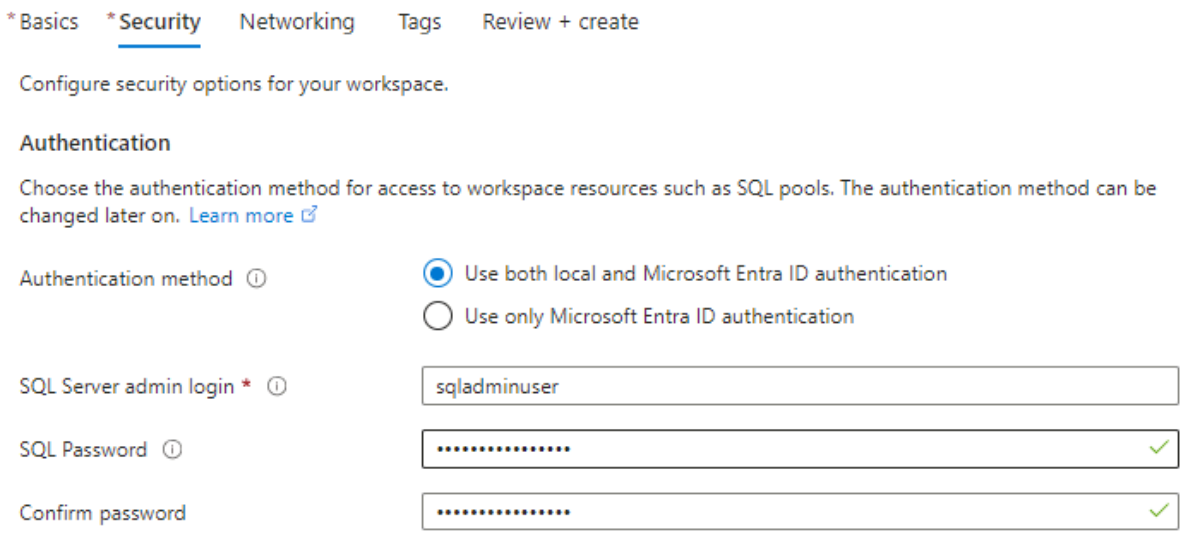
Step 3: **Basics**

- Resource group - Create New - “synapseresourse”
- Workspace name: “synapse1ws-rd”
- Region: “Central US”
- Select Data Lake Storage Gen2: “From Subscription”
- Account name - create new - “accountds”
- File system name - create new - “filesystemds”
- Next Security.

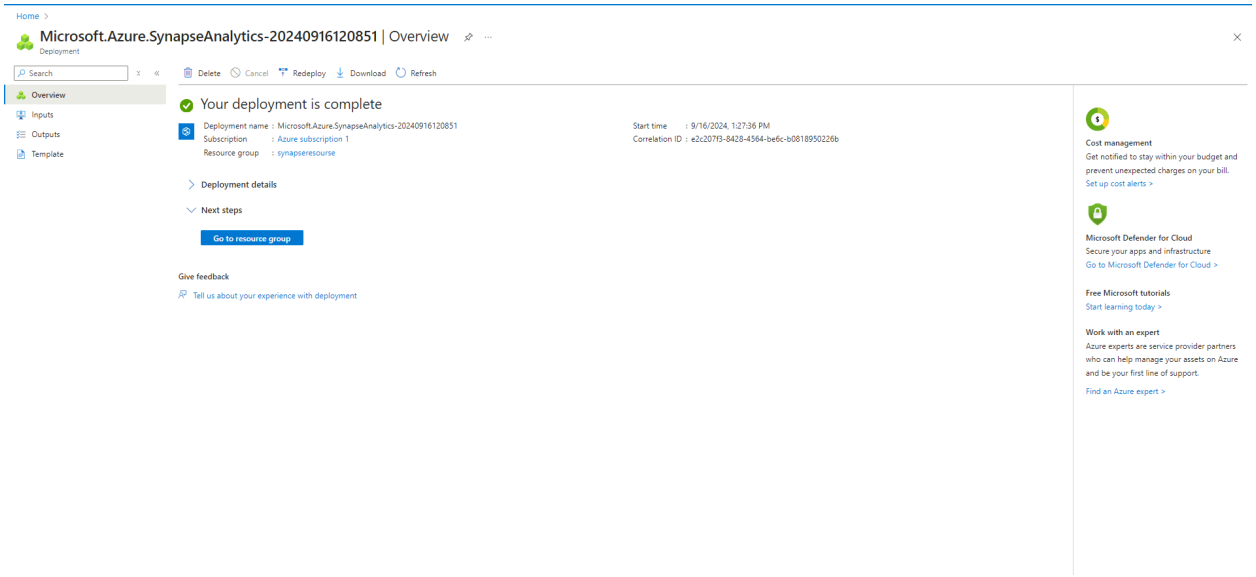


Step 4: **Security**

- Authentication method: “Use both local and Microsoft Entra ID authentication”
- SQL Server admin login: “sqladminuser”
- Password: ”****”
- Confirm Password: “*****”

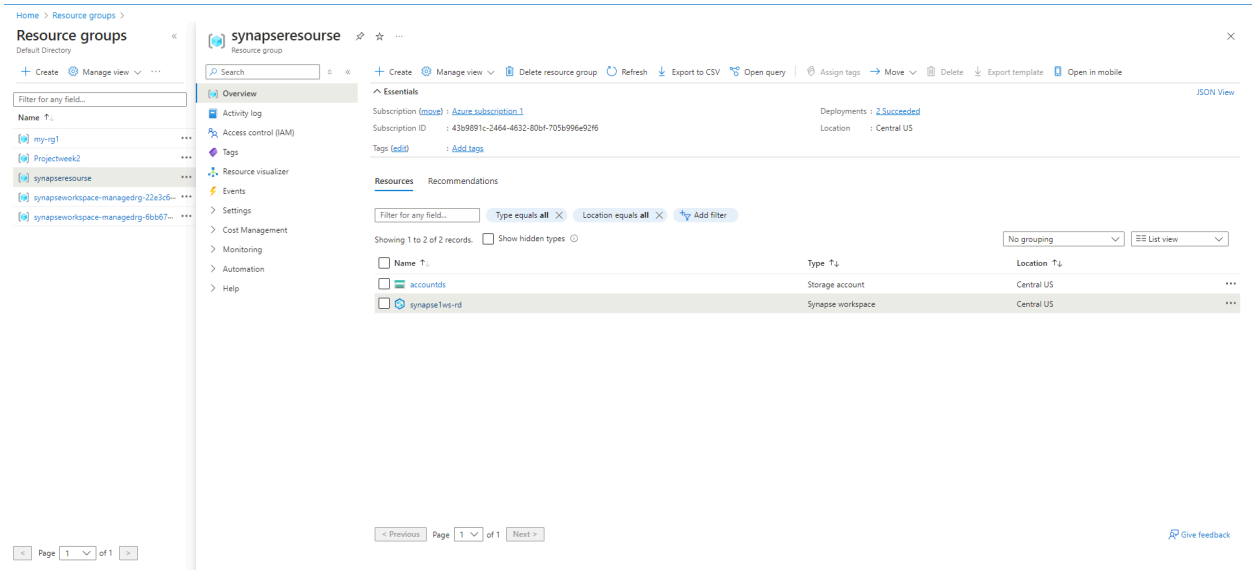


- Next Networking (no changes - keep it default)
- Next Tags (no changes - keep it default)
- Next Review + create
- Select “create”
- Deployment is in process.
- Deployment is successful.



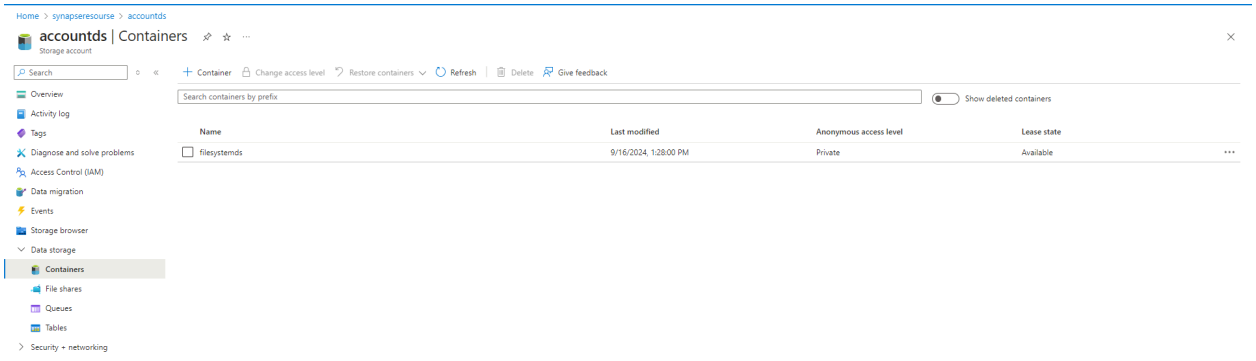
Step 5: Select “Go to Resource Group”

- You will be in “synapseresourse”
- Storage account: “accountds” and Synapse Workspace: “synapse1ws-rd” is created and will be seen.



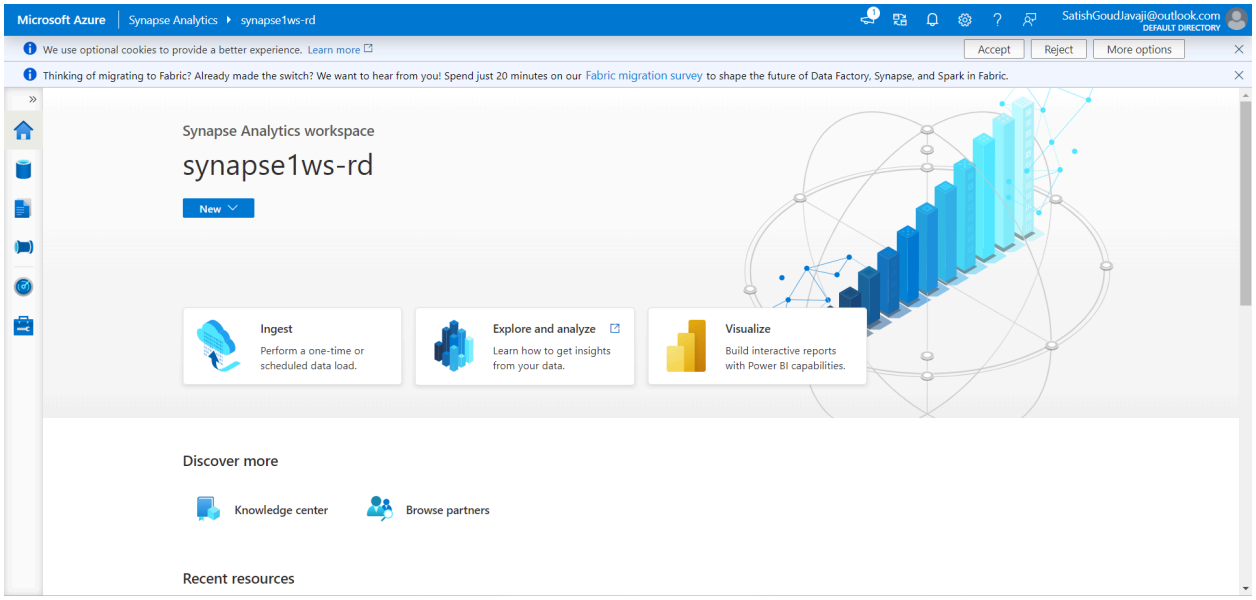
Step 6: Select - storage account - “accountds” (This step is to validate the file is created in the container or not)

- Select - Data Storage
- Select - Containers
- File is created in the container with the name “filesystemds” (The file name which we have given at step 3).



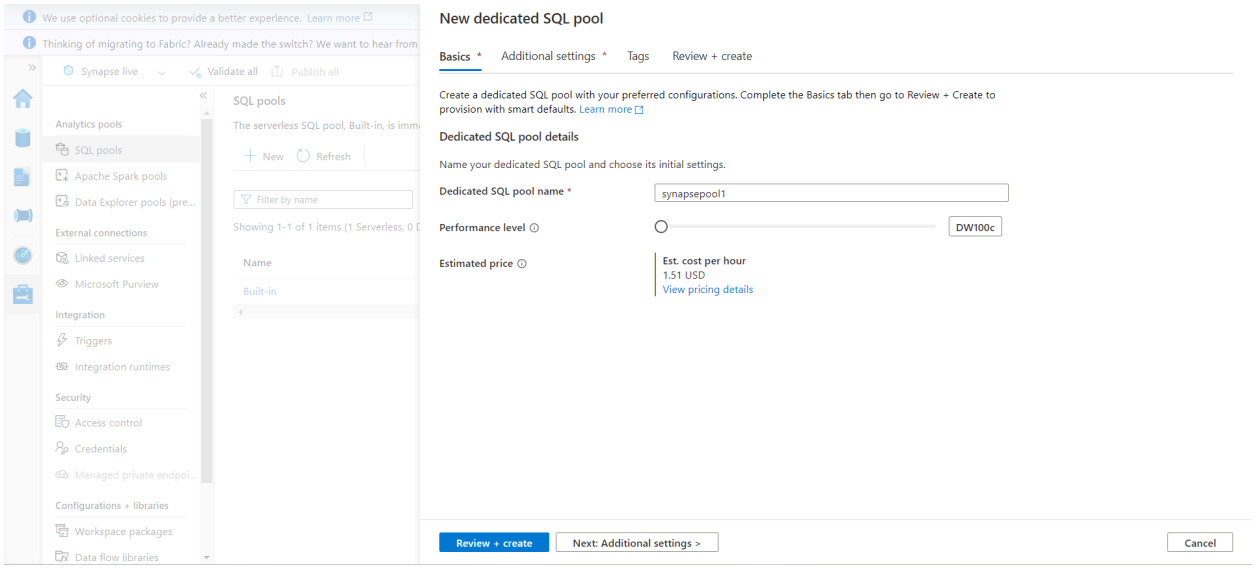
Step 7: Go to Resource group “synapseresourse”

- Select - Synapse workspace - “synapse1ws-rd”
- Select - “Open Synapse Studio”
- It will redirect to Synapse studio page as “synapse1ws-rd”



Step 8: Select - Manage - SQL pools

- In the SQL Pools - Select - “New”
- Dedicated SQL pool name: “synapsepool1”
- Performance level: Drag to min



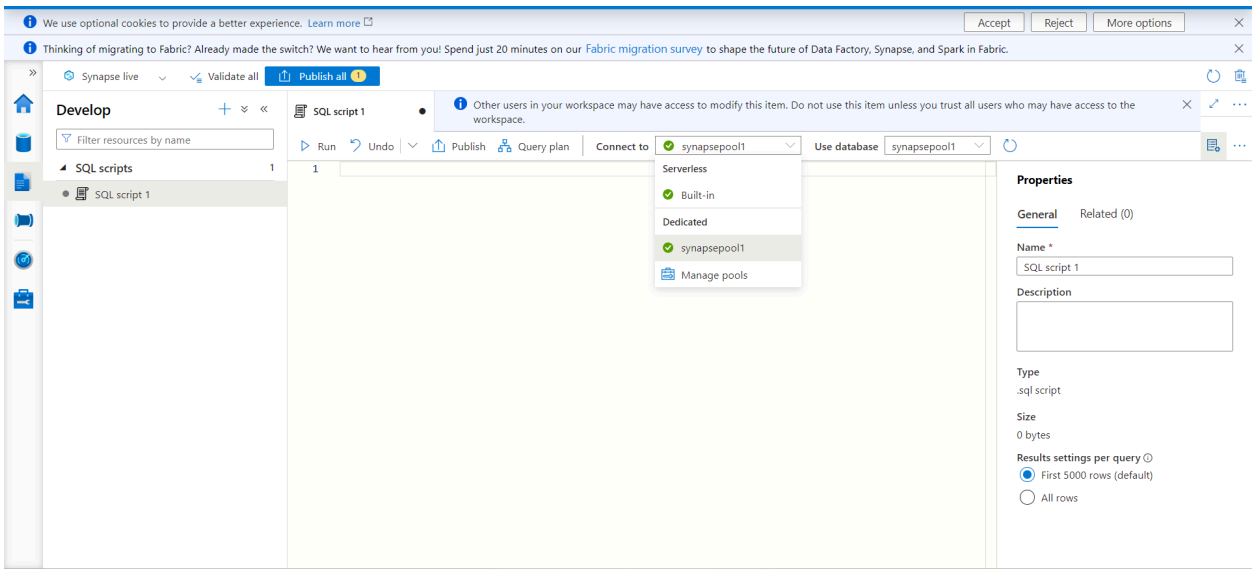
- Select “Review+Create” - Select “Create”.
- Deploying in process
- Dedicated SQL pool is created “synapsepool1”

Step 9: Select - Develop

- + select: “SQL script”
- Connect to: “synapsepool1”
- `CREATE TABLE Accounts(`

```
customer_id INT,
first_name VARCHAR(50),
last_name VARCHAR(50),
city VARCHAR(50)
);
```

- Select "Run"



Step 10: Select - Manage

- Select - Linked Services

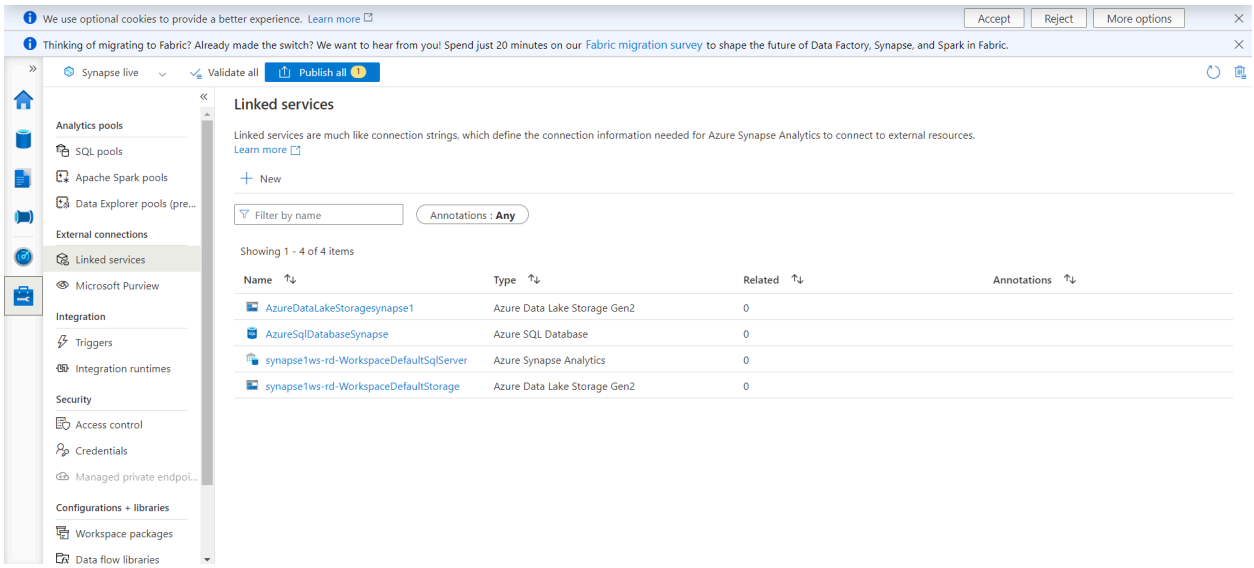
Azure SQL DataBase

- Select - "New"
- In search select "Azure SQL DataBase"
- Name: "AzureSqlDatabaseSynapse"
- Server name: "synapse1ws-rd"
- Database Name: "master"
- Authentication type: "SQL Authentication"
- User Name: "sqladminuser"
- Password: *****
- Select "Test Connection"
- Select - "Create"
- New SQL database is created "AzureSqlDatabaseSynapse"

Azure Data Lake Storage Gen2

- Select - "New"
- In search select "Azure Data Lake Storage Gen2"
- Name: "AzureDataLakeStoragesynapse1"
- Storage account name: "accountds"
- Select "Test Connection"

- Select - “Create”
- New Azure Data Lake Storage Gen2 is created “AzureDataLakeStoragesynapse1”



Step additional:

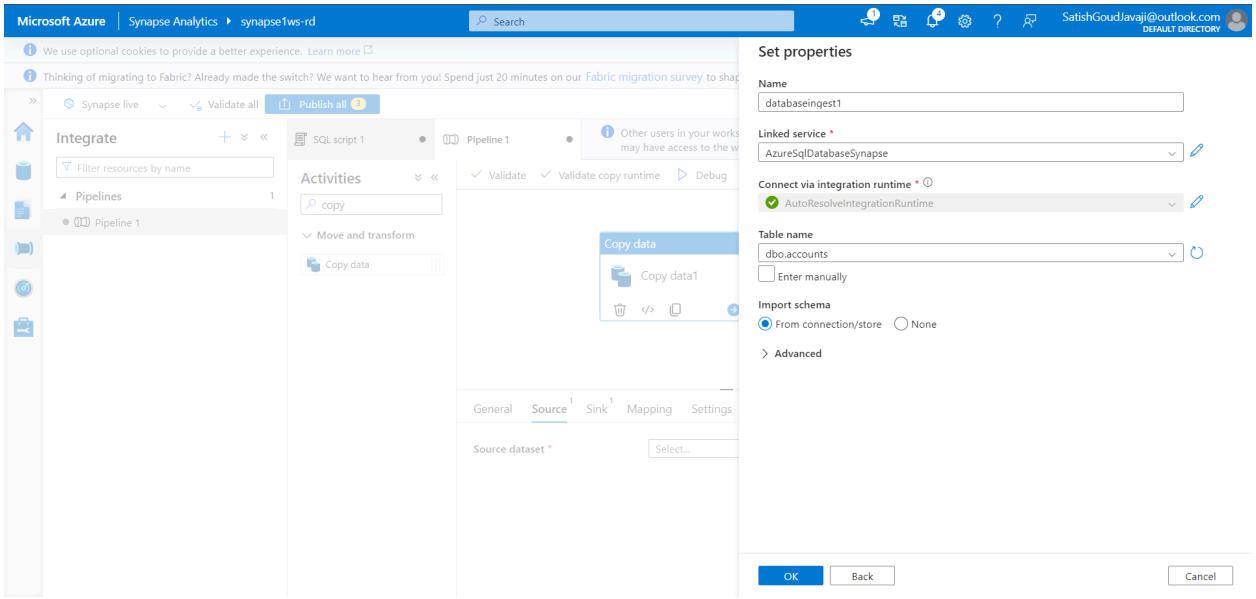
- Manage - Linked services
- Select - “AzureSQLDatabaseSynapse”
- Edit linked services.
- Account selection method: “From Azure subscription”
- Server name: “shreeserverproject2”
- Database name: “CustomerAccountLoanDB”
- Authentication type: “SQL Authentication”
- User name: “shreeserverproject2”
- Password: “*********”
- Select - “Test connection”

- Select - Manage - linked services
- Select - Azure Synapse analytics
- Name: “AzureSynapseAnalyticsingest”
- Account selection method: “From Azure subscription”
- Server name: “synapse1ws-rd (Synapse workspace)”
- Database name: “synapsepool1”
- User name: “sqladminuser”
- Password: “********”
- Select - “Test connection”
- Select - Create.

Step 11: Select - Integrate

- Select + Select “Pipeline”
- Search Activity: “copy data” (type)
- Drag copy data to pipeline page.

- Select - Source
- Select - New
- Select - “Azure SQL database”
- Name: “databaseingest1”
- Linked service: “AzureSqlDatabaseSynapse”
- Table name: “dbo.Accounts”
- Select - ok



- Select - Sink
- Select - New
- Search “Azure Data Lake Gen 2”
- Select - Parquet
- Name: “datalakeingest1”
- Linked service: “AzureDataLakeStoragesynapse1”
- File path - Select file icon
- Select - “filesystemds” Select “OK”

Step 12: Drag copy data to pipeline page.

- Select - New
- Search “Azure Data Lake Gen 2”
- Select - Parquet
- Name: “datalakeingest3”
- Linked service: “AzureDataLakeStoragesynapse1”
- File path - Select file icon
- Select - “filesystemds” Select
- Select - “dbo.accounts.parquet” and select “OK”

- Select - Sink
- Select - New

- Search “Azure Synapse Analytics dedicated pool”
- Name: “Synapseanalyticsing”
- SQL Pool: “synapsepool1”
- Table name: “dbo.accounts.
- Select - apply
- Drag copy data 1 tick mark to copy data 2
- Select copy data 2 - Select mapping - Select import schema.
- Publish all
- Validate
- Copy data 1 tick mark - Drag it to copy data 2.
- Publish all
- Debug
- Pipeline status Success