

## **TASK 2**

**Configure FTP/SFTP Server and Create an ADF Pipeline for Data Extraction:** Set up access to FTP/SFTP servers and build a pipeline in Azure Data Factory to extract files or datasets from these sources. This facilitates integration with external systems or partners.

(A sample SFTP is used using: <https://sftpcloud.io/tools/free-sftp-server>, it allows, 1GB free storage of cloud based SFTP server for one hour)

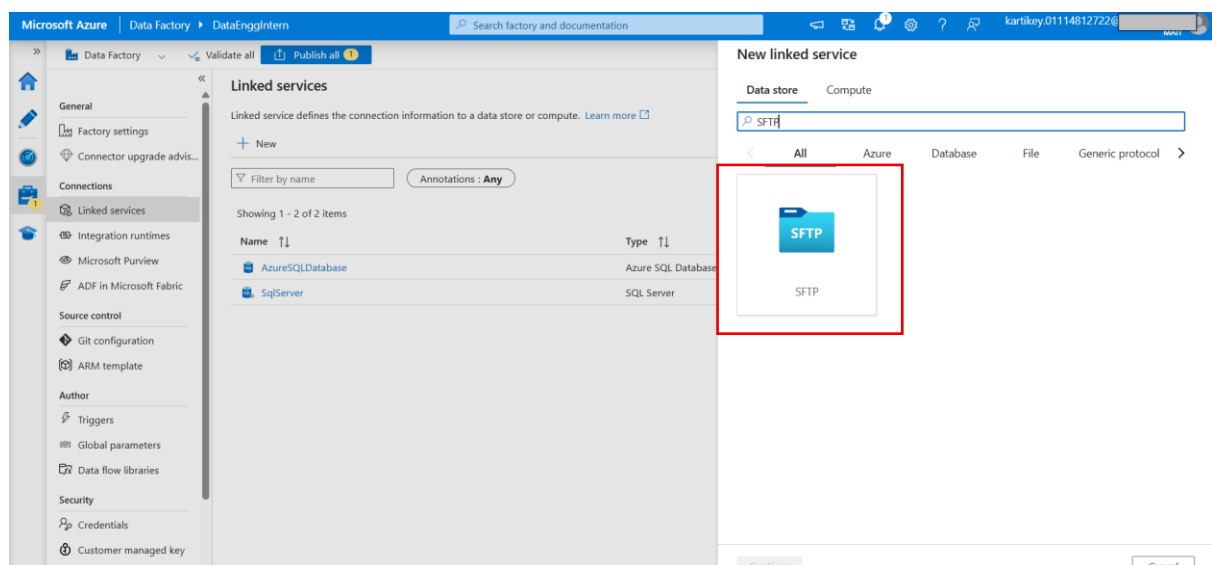
### **Step 1: Create Linked Services (Connections)**

We'll create two linked services:

- One for your **SFTP**
- One for **Azure SQL Database**

#### **A. Creating Linked Service for SFTP**

1. Go to **Manage > Linked Services > Click + New**.
2. In the search box, type **SFTP**, then select it and click **Continue**.



3. Fill in the details:
  - **Name:** <name>
  - **Host:** sftp.xxxxx.com
  - **Database name:** <db\_name>
  - **Authentication:** **SSH or Basic authentication**(enter username/password)

4. Click **Test Connection** (make sure it's green).
5. Click **Create**.

**New linked service**  
SFTP [Learn more](#)

**Name \***  
SFTP

**Description**

**Connect via integration runtime \***  
✓ AutoResolveIntegrationRuntime

**Host \***  
eu-central-1.sftpcloud.io

**Port**  
22

**SSH host key validation**  
Disable SSH host key validation

**Authentication type \***  
Basic

**User name \***  
5931b2ac6844441299cda33f638be78a

**Create** **Back**

✓ Connection successful  
[Test connection](#) **Cancel**

## B. Creating Linked Service for Azure SQL Database

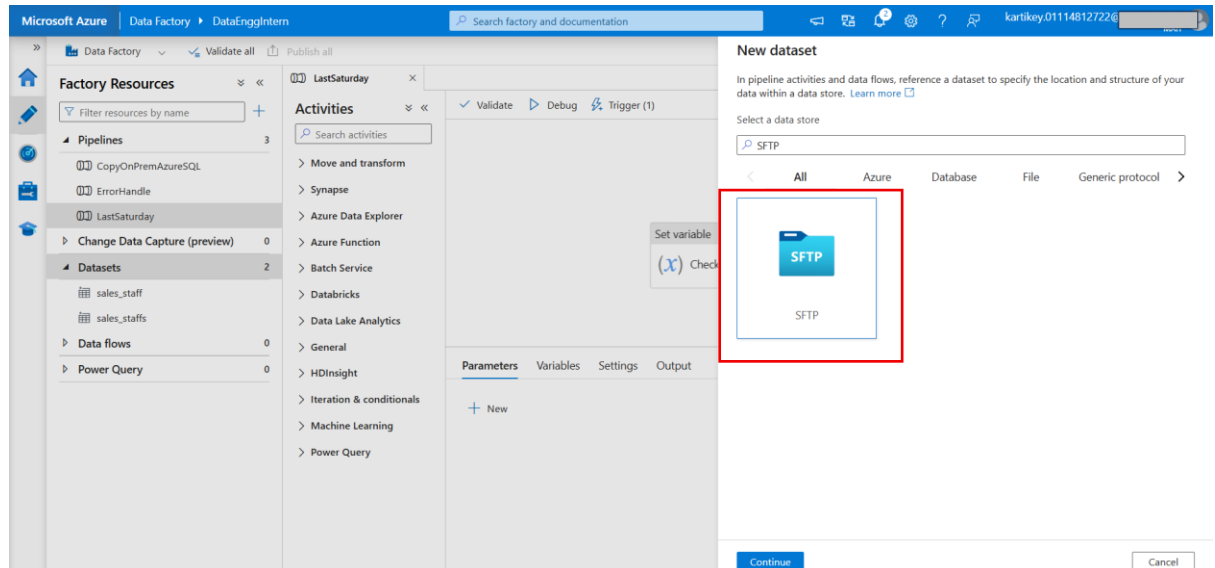
1. We created Azure SQL Database in Task 1

## Step 2: Create Datasets

These datasets define the structure and source/target of the data being moved.

### A. Create SFTP Dataset

1. Go to **Author** > **Datasets** > **+ New dataset**.



2. Choose:
  - **SFTP**
3. Choose file format
4. Choose:
  - **File name**
  - **Linked Service**
  - **File Path**
5. Click **OK**.

## Set properties

Name

DelimitedText2

Linked service \*

SFTP



File path

eu-central-1.sftpcloud.io /

sales

/

customers



First row as header



Import schema



From connection/store



From sample file



None

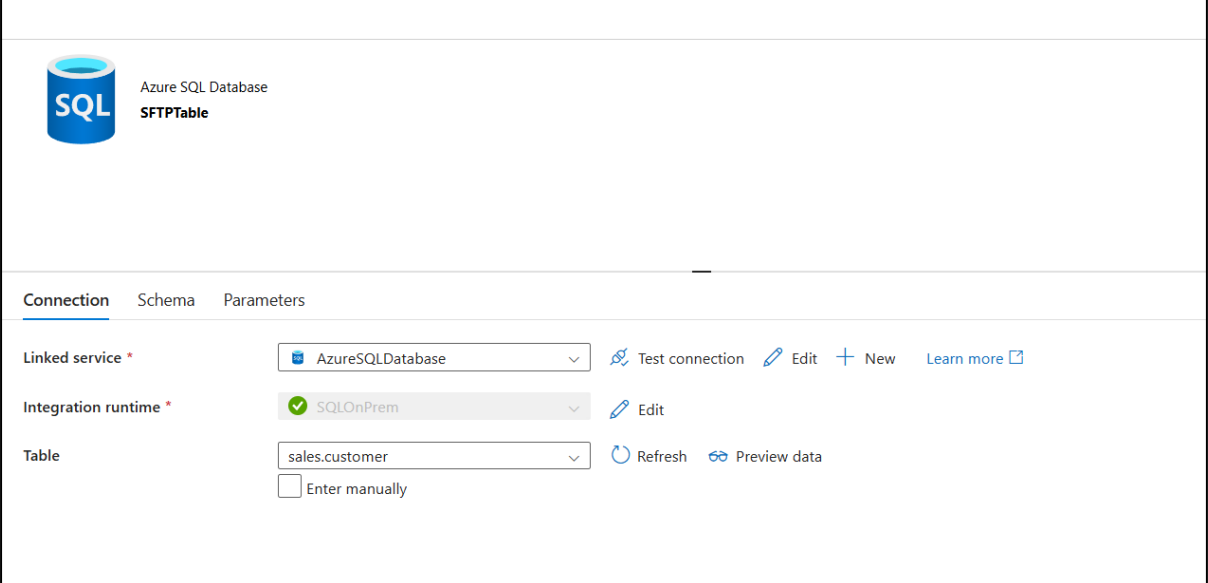
OK

Back

Cancel

## B. Create Sink Dataset (Azure SQL Database)

1. Click + **New dataset**.
2. Choose:
  - **Azure SQL Database**
  - Linked Service: <service\_name>
3. Name it: <table\_name>
4. Select the target table name (can be pre-created or we can let ADF create it).
5. Click **OK**.



The screenshot shows the 'New Dataset' configuration window in Azure Data Factory. At the top, there is a header with the 'SQL' icon, 'Azure SQL Database', and 'SFTPTable'. Below this is a tabbed interface with 'Connection', 'Schema', and 'Parameters' tabs. The 'Connection' tab is active and contains the following configuration:

- Linked service \***: A dropdown menu showing 'AzureSQLDatabase' with a blue icon. To the right are links for 'Test connection', 'Edit', '+ New', and 'Learn more'.
- Integration runtime \***: A dropdown menu showing 'SQLOnPrem' with a green checkmark icon. To the right is a link for 'Edit'.
- Table**: A dropdown menu showing 'sales.customer'. Below it is a checkbox labeled 'Enter manually'. To the right are links for 'Refresh' and 'Preview data'.

## Step 3: Create a Copy Pipeline

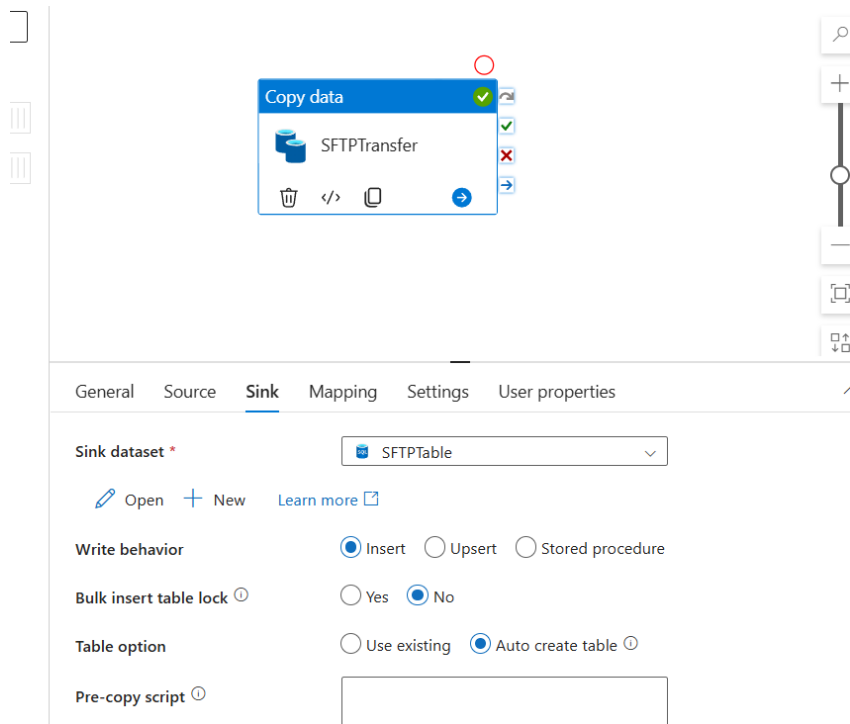
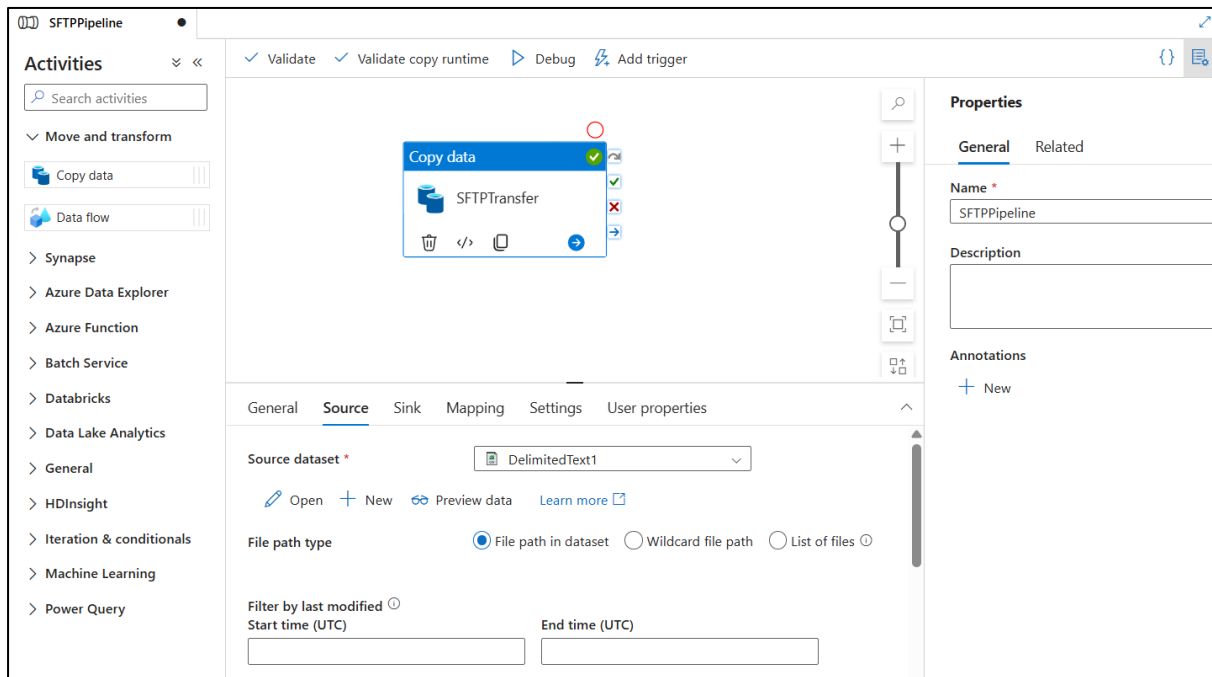
### A. Creating the Pipeline

1. In **Author**, click **Pipeline**.
2. Name the pipeline: <pipeline\_name>.

### B. Add a Copy Data Activity

1. Drag **Copy Data** from the Activities pane to the pipeline canvas.
2. Select it and configure:
  - **Source tab:**
    - Dataset: <table\_name>
    - Query/table: double-check settings
  - **Sink tab:**

- Dataset: <table\_name>
- We can auto-create the table or map columns manually in the **Mapping** tab



## TESTING THE PIPELINE

The screenshot displays the Azure Data Studio interface for the SFTPPipeline. The left sidebar shows the 'Activities' pane with a search bar and a list of activity categories: Move and transform, Synapse, Azure Data Explorer, Azure Function, Batch Service, Databricks, Data Lake Analytics, General, HDInsight, Iteration & conditionals, Machine Learning, and Power Query. The main canvas shows a 'Copy data' activity using the 'SFTPTransfer' connector, marked with a green checkmark. The bottom pane shows the 'Output' tab with the following details:

- Pipeline run ID: e251f48d-6c81-4474-998b-d8e2638dc324
- Pipeline status: Succeeded
- Activity name: SFTPTransfer
- Activity status: Succeeded
- Run start: 7/14/2025, 2:12:13 PM
- Duration: 30s

The right sidebar shows the 'Properties' pane with the 'General' tab selected, displaying the pipeline's name and description.

## DATA COPIED SUCCESSFULLY IN AZURE SQL

The screenshot displays the Azure Data Studio interface for a SQL query. The left sidebar shows the 'Tables' pane with a search bar and a list of tables: sales.customer, sales.customers, and sales.staff. The main canvas shows a SQL query in 'Query 2':

```
1 SELECT TOP (1000) * FROM [sales].[customer]
```

The bottom pane shows the 'Results' tab with the following data:

customer_id	first_name	last_name	phone	email
1	Debra	Burks		debra.burks@
2	Kasha	Todd		kasha.todd@
3	Tameka	Fisher		tameka.fishe
4	Danul	Spence		danul.spence