

Azure Data Factory

Lab: Ingest Data using copy data tool from Storage Blob to SQL Database.

Pre-requisites:

- Azure Pass subscription
- Azure Data Lake Storage Gen2 storage account
- Azure SQL Database

Lab Objective:

After completing this lab, you will be able to:

- Create pipeline using Copy Data tool to ingest the data.

Exercise: Ingest Data using Copy Data Tool.

Create a source blob

1. Launch **Notepad**. Copy the following text and save it in a file named **inputEmp.txt** on your disk:

```
FirstName|LastName  
John|Doe  
Jane|Doe
```

2. Create a container named **data** if not exists and upload the inputEmp.txt file to the container. You can use the Azure portal or various tools like [Azure Storage Explorer](#) to perform these tasks.

Create a sink SQL table

1. Use the following SQL script to create a table named **dbo.emp** in your SQL Database:

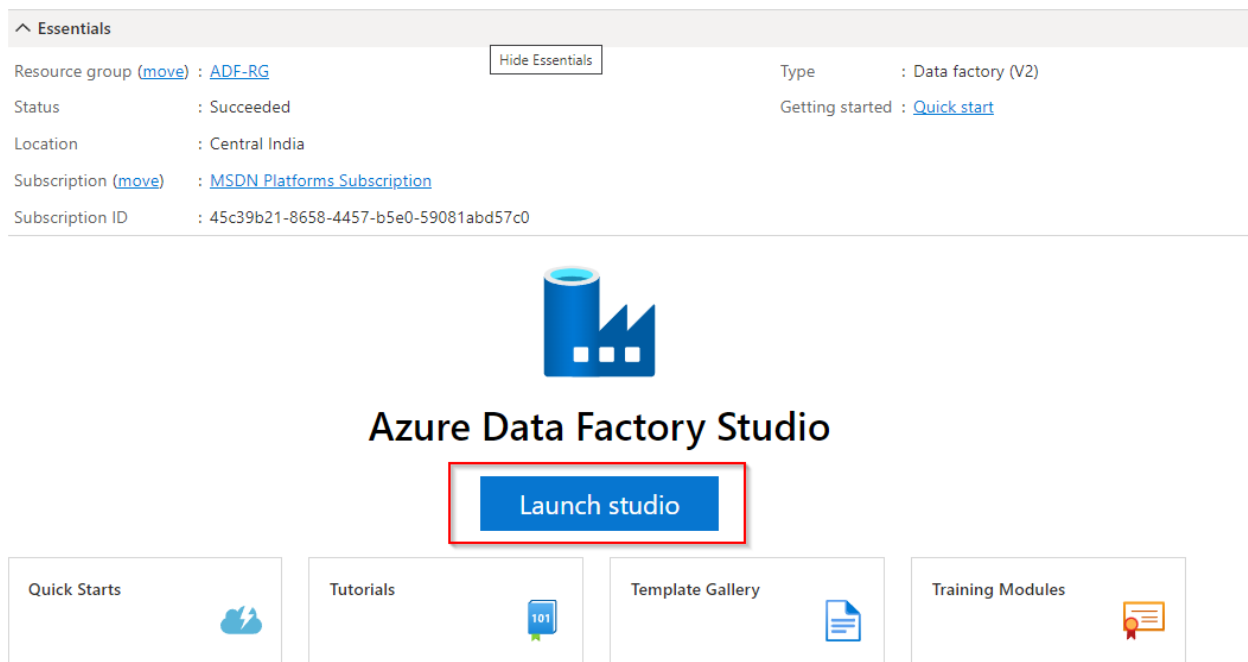
```
CREATE TABLE dbo.emp  
(  
    ID int IDENTITY(1,1) NOT NULL,  
    FirstName varchar(50),  
    LastName varchar(50)  
)  
GO
```

2. Allow Azure services to access SQL Server. Verify that the setting **Allow Azure services and resources to access this server** is enabled for your server that's running SQL Database. This setting lets Data Factory write data to your database instance.

To verify and turn on this setting, go to logical SQL server > Security > Firewalls and virtual networks > set the **Allow Azure services and resources to access this server** option to **ON**.

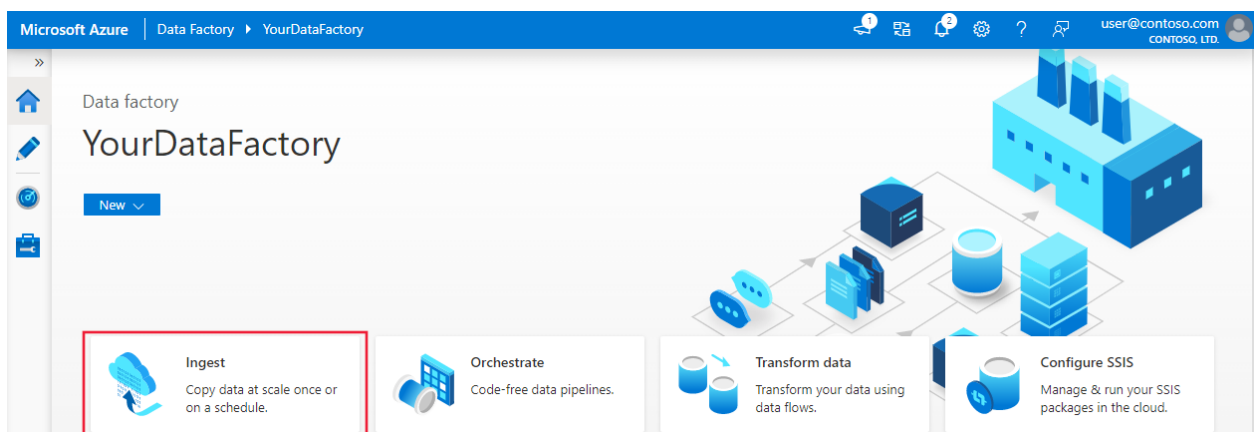
Open you Azure Data Factory workspace

1. To launch the Azure Data Factory user interface (UI) in a separate tab, select **Open** on the **Open Azure Data Factory Studio** tile.



Use the Copy Data tool to create a pipeline

1. On the home page of Azure Data Factory, select the **Ingest** tile to launch the Copy Data tool.



2. On the **Properties** page of the Copy Data tool, choose **Built-in copy task** under **Task type**, then select **Next**.

Copy Data tool

1 Properties

2 Source

3 Target

4 Settings


5 Review and finish

Use Copy Data Tool to perform a one-time or scheduled data load from 90+ data sources. Follow the wizard experience to specify your data loading settings, and let the Copy Data Tool generate the artifacts for you, including pipelines, datasets, and linked services. [Learn more](#)

Properties


Select copy data task type and configure task schedule

Task type



Built-in copy task

You will get single pipeline to copy data from 90+ data source easily.



Metadata-driven copy task (Preview)

Metadata is required to be stored in external control tables to load data at large-scale.

You will get single pipeline to quickly copy objects from data source store to destination in a very intuitive manner.

Task cadence or task schedule *

☒ Run once now
 ☐ Schedule
 ☐ Tumbling window

< Previous

Next >

Cancel

3. On the **Source data store** page, complete the following steps:
 - a. Select **+ Create new connection** to add a connection.
 - b. Select **Azure Data Lake Storage** from the gallery, and then select **Continue**.
 - c. On the **New connection (Azure Data Lake Storage)** page, select your Azure subscription from the **Azure subscription** list, and select your storage account from the **Storage account name** list. Test connection and then select **Create**.
 - d. Select the newly created linked service as source in the **Connection** block.
 - e. In the **File or folder** section, select **Browse** to navigate to the **data** folder, select the **inputEmp.txt** file, then select **OK**.
 - f. Select **Next** to move to next step.

Copy Data tool

✓ Properties

2 Source

• Dataset

○ Configuration

3 Target

4 Settings

5 Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type

Azure Blob Storage

Connection *

AzureBlobStorage

Edit

+ Create new connection

File or folder *

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

adfv2tutorial/inputEmp.txt

Browse

Options

☐ Binary copy ⓘ

☒ Recursively ⓘ

☐ Enable partition discovery ⓘ

Max concurrent connections ⓘ

Filter by last modified

Start time (UTC)End time (UTC) ⓘ

< Previous

Next >

- On the **File format settings** page, enable the checkbox for *First row as header*. Notice that the tool automatically detects the column and row delimiters, and you can preview data and view the schema of the input data by selecting **Preview data** button on this page. Then select **Next**.

Copy Data tool

Properties

2 Source

Dataset

Configuration

3 Target

4 Settings

5 Review and finish

File format settings

File format ⓘ
Text format

Detect text format **Preview data**

Column delimiter
Pipe (|)

☐ Edit

Row delimiter
Default (\r\n, or \r\n)

☐ Edit

☒ First row as header ⓘ

Advanced

Compression type
None

Additional columns ⓘ
[+ New](#)

Preview data

Linked service: AzureBlobStorage
Object: adfv2tutorial/inputEmp.txt

Preview Schema

FirstName	LastName
John	Doe
Jane	Doe

< Previous **Next >** Cancel

5. On the **Destination data store** page, completes the following steps:
 - a. Select **+ Create new connection** to add a connection.
 - b. Select **Azure SQL Database** from the gallery, and then select **Continue**.
 - c. On the **New connection (Azure SQL Database)** page, select your Azure subscription, server name and database name from the dropdown list. Then select **SQL authentication** under **Authentication type**, specify the username and password. Test connection and select **Create**.

New connection (Azure SQL Database)

Name *

Description

Connect via integration runtime * ⓘ

☒ Connection string ☐ Azure Key Vault

Account selection method ⓘ
☒ From Azure subscription ☐ Enter manually

Azure subscription

Server name *
 ⓘ

Database name *
 ⓘ

Authentication type *

User name *

☒ Password ☐ Azure Key Vault

Password *

Always encrypted ⓘ ☐

✓ Connection successful

d. Select the newly created linked service as sink, then select **Next**.

- On the **Destination data store** page, select **Use existing table** and select the **dbo.emp** table. Then select **Next**.
- On the **Column mapping** page, notice that the second and the third columns in the input file are mapped to the **FirstName** and **LastName** columns of the **emp** table. Adjust the mapping to make sure that there is no error, and then select **Next**.

Column mapping

Choose how source and destination columns are mapped

Table mappings (1)

- ☒ Source
Azure Blob Storage file
- ☐ Target
dbo.emp

Column mappings

Type conversion settings

+ New mapping Clear Reset Delete

Source	Type	Destination
FirstName	String	FirstName
LastName	String	LastName

Azure SQL Database sink properties

Pre-copy script

< Previous **Next >** Cancel

- On the **Settings** page, under **Task name**, enter **CopyFromBlobToSqlPipeline**, and then select **Next**.

Copy Data tool

Properties
Source
Target
Settings
Review and finish

Settings

Enter name and description for the copy data task, more options for data movement

Task name * **CopyFromBlobToSqlPipeline**

Task description

Data consistency verification ☐

Fault tolerance

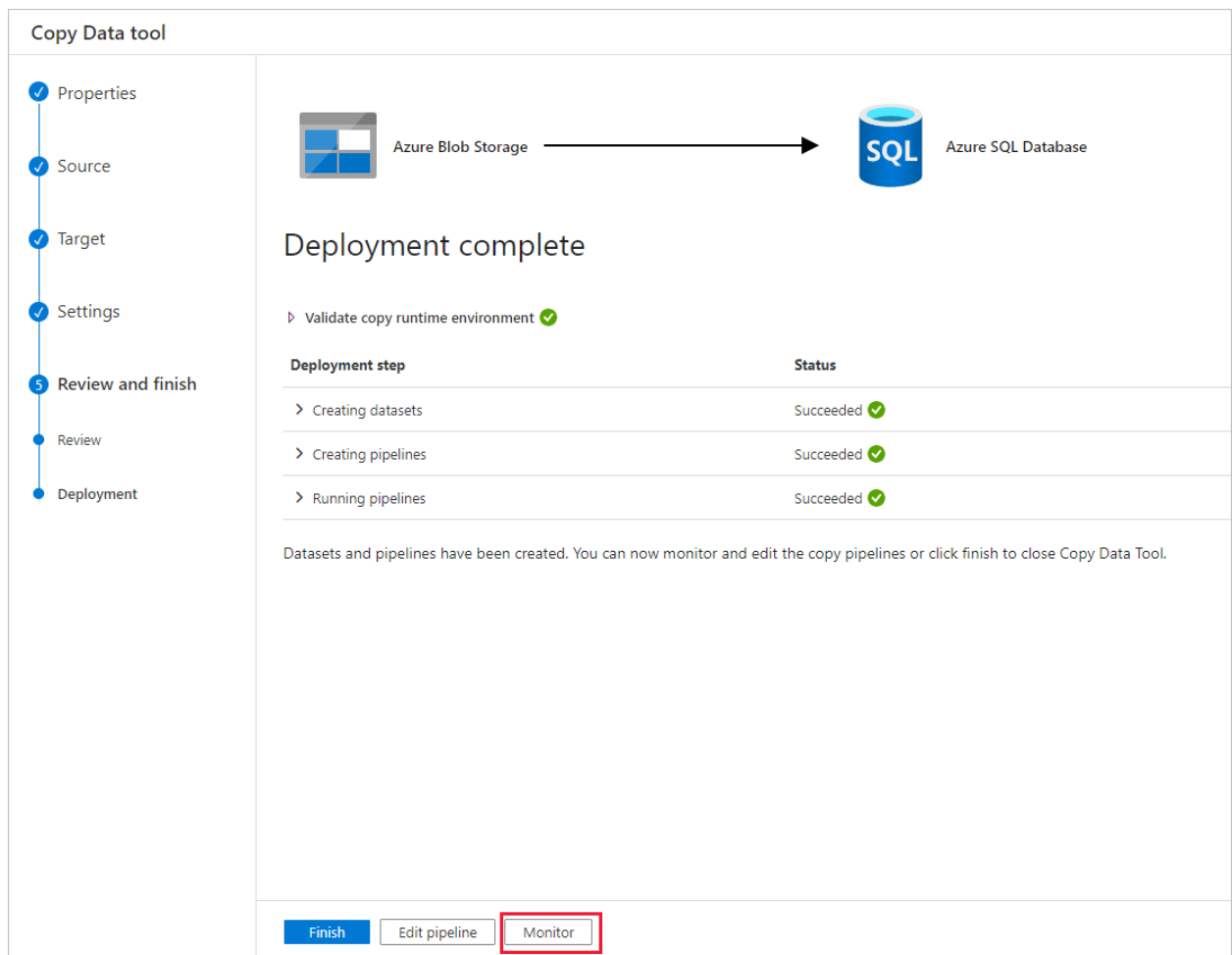
Enable logging ☐

Enable staging ☐

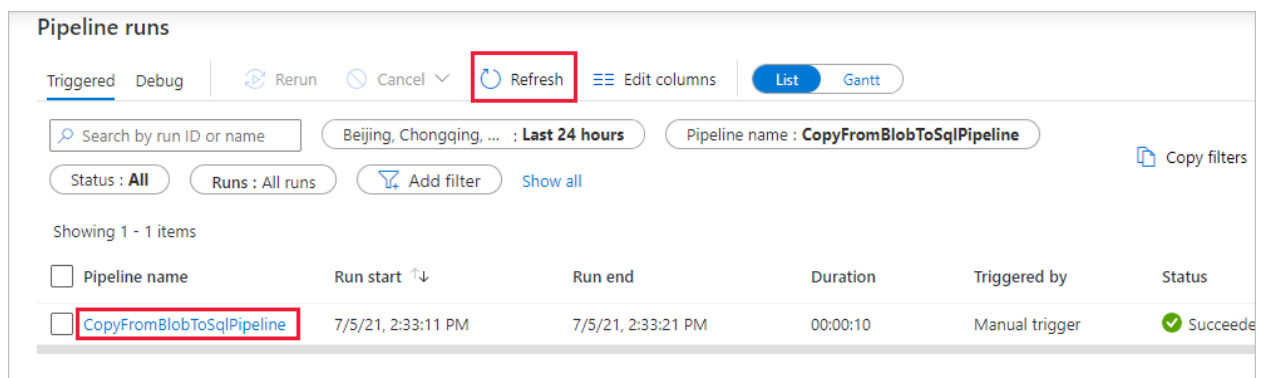
Advanced

< Previous **Next >** Cancel

9. On the **Summary** page, review the settings, and then select **Next**.
10. On the **Deployment** page, select **Monitor** to monitor the pipeline (task).



11. On the Pipeline runs page, select **Refresh** to refresh the list. Select the link under **Pipeline name** to view activity run details or rerun the pipeline.



12. On the "Activity runs" page, select the **Details** link (eyeglasses icon) under **Activity name** column for more details about copy operation. To go back to the "Pipeline runs" view, select the **All pipeline runs** link in the breadcrumb menu. To refresh the view, select **Refresh**.

All pipeline runs > CopyFromBlobToSqlPipeline - Activity runs

CopyFromBlobToSqlPipeline

List Gantt

Rerun Rerun from activity Rerun from failed activity Refresh Edit pipeline

Copy data ✓

Copy_vto

+ - [100%] []

Activity runs

Pipeline run ID e79cfba6-412d-412a-8129-ae43e67caaa8

All status ▾

Showing 1 - 1 of 1 items

Activity name	Activity type	Run start ↑↓	Duration	Status	Error	Integration runtime
Copy_vto	Copy data	7/5/21, 2:33:12 PM	00:00:09	✓ Succeeded		DefaultIntegrationRuntime (Eas

13. Verify that the data is inserted into the **dbo.emp** table in your SQL Database.
14. Select the **Author** tab on the left to switch to the editor mode. You can update the linked services, datasets, and pipelines that were created via the tool by using the editor.

>> <<

All pipeline runs > CopyFromBlobToSqlPipeline - Activity runs

CopyFromBlobToSqlPipeline

List Gantt

Rerun Rerun from activity Rerun from failed a

Copy data ✓

Copy_vto

Home Dashboards

Runs

Author Pipeline runs

Trigger runs

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics