

LAB - Copy data from Azure Blob storage to a SQL Database by using the Copy Data tool

In this tutorial, you use the Azure portal to create a data factory. Then you use the Copy Data tool to create a pipeline that copies data from Azure Blob storage to a SQL Database.

In this tutorial, you perform the following steps:

- Create a data factory.
- Use the Copy Data tool to create a pipeline.
- Monitor the pipeline and activity runs.

Prerequisites

- **Azure subscription:** If you don't have an Azure subscription, create a [free account](#) before you begin.
- **Azure Storage account:** Use Blob storage as the *source* data store. If you don't have an Azure Storage account, see the instructions in [Create a storage account](#).
- **Azure SQL Database:** Use a SQL Database as the *sink* data store. If you don't have a SQL Database, see the instructions in [Create a SQL Database](#).

Create a blob and a SQL table

Prepare your Blob storage and your SQL Database for the tutorial by performing these steps.

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 **adfvtutorial** 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

CREATE CLUSTERED INDEX IX_emp_ID ON dbo.emp (ID);
```

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**.

Note

The option to **Allow Azure services and resources to access this server** enables network access to your SQL Server from any Azure resource, not just those in your subscription. For more information, see [Azure SQL Server Firewall rules](#). Instead, you can use [Private endpoints](#) to connect to Azure PaaS services without using public IPs.

Create a data factory

1. On the left menu, select **Create a resource > Integration > Data Factory**:

[Home](#) >

New

 Search the Marketplace

Azure Marketplace [See all](#)

Featured [See all](#)

Get started

Recently created

AI + Machine Learning

Analytics

Blockchain

Compute

Containers

Databases

Developer Tools

DevOps

Identity

Integration

Internet of Things

IT & Management Tools

Media

Migration

Mixed Reality

Monitoring & Diagnostics

Networking

Security

Software as a Service (SaaS)

Storage

Web



Logic App

[Quickstarts + tutorials](#)



API Management

[Quickstarts + tutorials](#)



Service Bus

[Quickstarts + tutorials](#)



Integration Account

[Quickstarts + tutorials](#)



Integration Service Environment

[Learn more](#)



Logic Apps Custom Connector

[Learn more](#)



Data Factory

[Quickstarts + tutorials](#)



Data Catalog

[Learn more](#)



Apache Kafka® on Confluent Cloud™
for Azure (preview)

[Learn more](#)



Dell Boomi Atom (Windows)
(preview)

[Learn more](#)

2. On the **New data factory** page, under **Name**, enter **ADFTutorialDataFactory**.

The name for your data factory must be *globally unique*. You might receive the following error message:

Create Data Factory ...

Basics Git configuration Networking Advanced Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *	<input type="text" value="<your Azure subscription selection>"/>
Resource group *	<input type="text" value="YourResourceGroup"/>
	Create new

Instance details

Region *	<input type="text" value="South Central US"/>
Name *	<input type="text" value="ADFTutorialDataFactory"/>
	✖ The Data Factory name is already taken. Choose a different name.
Version *	<input type="text" value="V2"/>

If you receive an error message about the name value, enter a different name for the data factory. For example, use the name ***yourname*ADFTutorialDataFactory**. For the naming rules for Data Factory artifacts, see [Data Factory naming rules](#).

3. Select the Azure **subscription** in which to create the new data factory.
4. For **Resource Group**, take one of the following steps:
 - a. Select **Use existing**, and select an existing resource group from the drop-down list.
 - b. Select **Create new**, and enter the name of a resource group.To learn about resource groups, see [Use resource groups to manage your Azure resources](#).
5. Under **version**, select **V2** for the version.
6. Under **location**, select the location for the data factory. Only supported locations are displayed in the drop-down list. The data stores (for example, Azure Storage and SQL Database) and computes (for example, Azure HDInsight) that are used by your data factory can be in other locations and regions.
7. Select **Create**.
8. After creation is finished, the **Data Factory** home page is displayed.

Home >

ADFTutorialDataFactory

Data factory (V2)

Search (Ctrl+/) << Delete

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Networking
- Properties
- Locks

Getting started

- Quick start

Monitoring

- Alerts
- Metrics
- Diagnostic settings
- Logs

Automation

- Tasks (preview)

Essentials

Resource group (change)
< your resource group >

Status
Succeeded


Location
East US


Subscription (change)
< your Azure subscription >

Subscription ID
< your Azure subscription ID >

Type
Data factory (V2)

Getting started
Quick start

 **Open Azure Data Factory Studio**
Start authoring and monitoring your data pipelines and data flows.
[Open](#)

 **Read documentation**
Learn how to be productive quickly. Explore concepts, tutorials, and samples.
[Learn more](#)

Monitoring

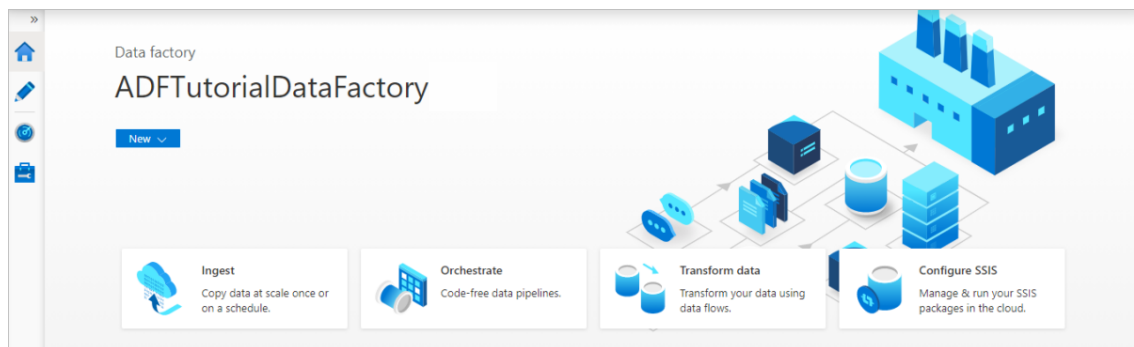
PipelineRuns

ActivityRuns

9. 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 Blob Storage** from the gallery, and then select **Continue**.
 - c. On the **New connection (Azure Blob 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 **adfv2tutorial** folder, select the **inputEmp.txt** file, then select **OK**.
 - f. Select **Next** to move to next step.

Copy Data tool

1 Properties

2 **Source**

3 Dataset

4 Configuration

5 Target

6 Settings

7 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 *: adfv2tutorial/inputEmp.txt [Browse](#)

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.

Options

☐ Binary copy ⓘ

☒ Recursively ⓘ

☐ Enable partition discovery ⓘ

Max concurrent connections ⓘ

Filter by last modified

Start time (UTC) End time (UTC) ⓘ

[< Previous](#) [Next >](#)

4. 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

1 Properties

2 **Source**

3 Dataset

4 Configuration

5 Target

6 Settings

7 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 \n\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:
- 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

Test connection

Cancel

Create

Back

- d. Select the newly created linked service as sink, then select **Next**.
6. On the **Destination data store** page, select **Use existing table** and select the **dbo.emp** table. Then select **Next**.
7. 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

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

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Target
- 4 Settings**
- 5 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 ☐

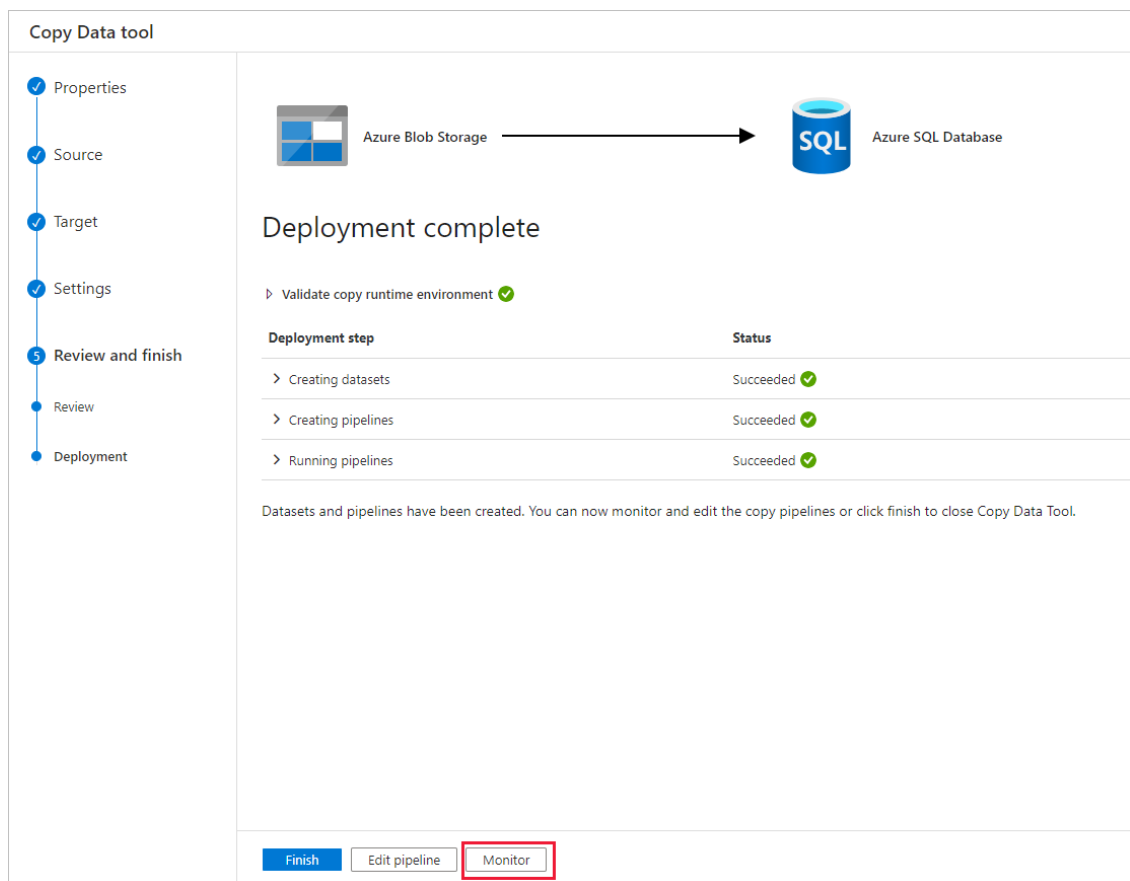
Advanced

< Previous

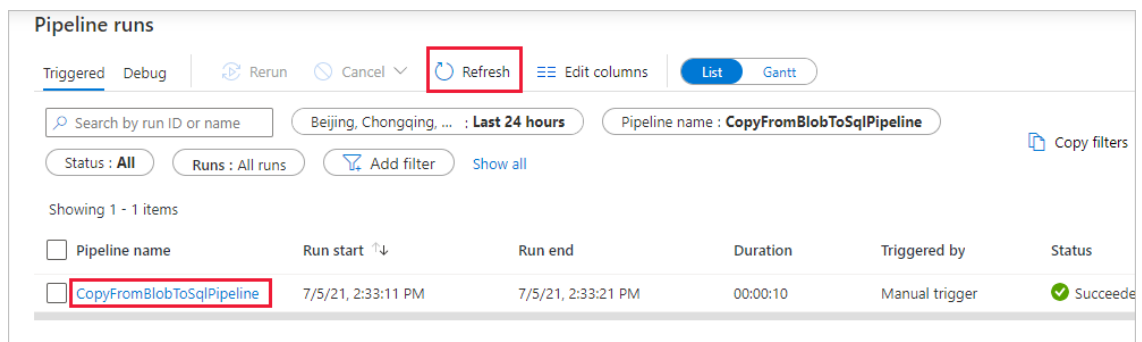
Next >

9. On the **Summary** page, review the settings, and then select **Next**.

10. On the **Deployment** page, select **Monitor** to monitor the pipeline (task).



- 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.



- 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

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 (East US)

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 activity

Copy data

Copy_vto

Left sidebar menu:

- Home
- Dashboards
- Runs
- Pipeline runs
- Trigger runs
- Runtimes & sessions
- Integration runtimes
- Data flow debug
- Notifications
- Alerts & metrics

Conclusion

The pipeline in this sample copies data from Blob storage to a SQL Database. You learned how to:

- Create a data factory.
- Use the Copy Data tool to create a pipeline.
- Monitor the pipeline and activity runs.

