Quickstart: Create a data factory by using the Azure Data Factory UI

9/25/2020 • 11 minutes to read • Edit Online

APPLIES TO: Azure Data Factory Azure Synapse Analytics (Preview)

This quickstart describes how to use the Azure Data Factory UI to create and monitor a data factory. The pipeline that you create in this data factory *copies* data from one folder to another folder in Azure Blob storage. To *transform* data by using Azure Data Factory, see Mapping data flow and Wrangling data flow (Preview).

> **NOTE**
> If you are new to Azure Data Factory, see Introduction to Azure Data Factory before doing this quickstart.

Prerequisites

Azure subscription

If you don't have an Azure subscription, create a free account before you begin.

Azure roles

To create Data Factory instances, the user account that you use to sign in to Azure must be a member of the *contributor* or *owner* role, or an *administrator* of the Azure subscription. To view the permissions that you have in the subscription, go to the Azure portal, select your username in the upper-right corner, select "..." icon for more options, and then select **My permissions**. If you have access to multiple subscriptions, select the appropriate subscription.

To create and manage child resources for Data Factory - including datasets, linked services, pipelines, triggers, and integration runtimes - the following requirements are applicable:

- To create and manage child resources in the Azure portal, you must belong to the **Data Factory Contributor** role at the resource group level or above.
- To create and manage child resources with PowerShell or the SDK, the **contributor** role at the resource level or above is sufficient.

For sample instructions about how to add a user to a role, see the Add roles article.

For more info, see the following articles:

- Data Factory Contributor role
- Roles and permissions for Azure Data Factory

Azure Storage account

You use a general-purpose Azure Storage account (specifically Blob storage) as both *source* and *destination* data stores in this quickstart. If you don't have a general-purpose Azure Storage account, see Create a storage account to create one.

Get the storage account name

You need the name of your Azure Storage account for this quickstart. The following procedure provides steps to get the name of your storage account:

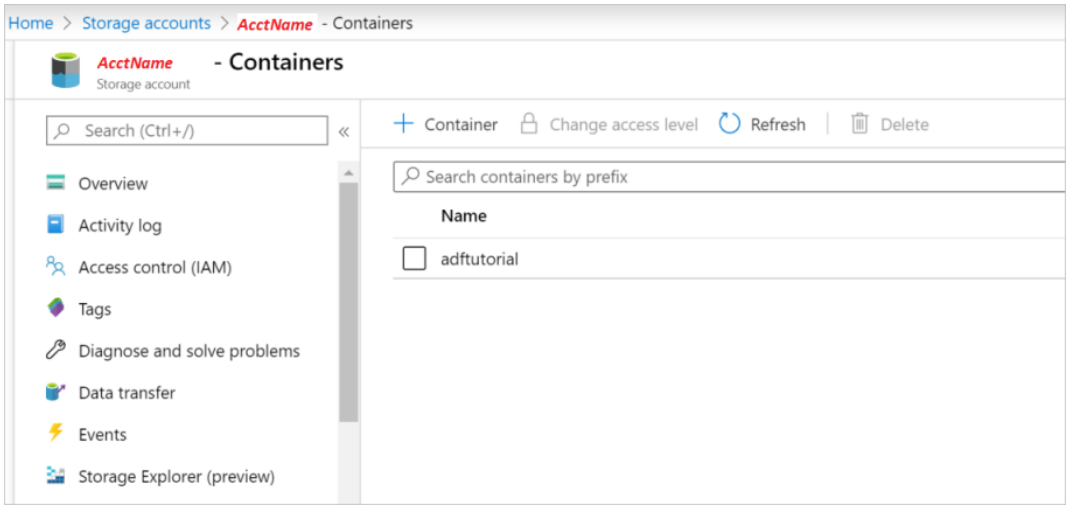
1. In a web browser, go to the Azure portal and sign in using your Azure username and password.
2. From the Azure portal menu, select **All services**, then select **Storage** > **Storage accounts**. You can also search for and select *Storage accounts* from any page.
3. In the **Storage accounts** page, filter for your storage account (if needed), and then select your storage account.

You can also search for and select *Storage accounts* from any page.

Create a blob container

In this section, you create a blob container named **adftutorial** in Azure Blob storage.

1. From the storage account page, select **Overview** > **Containers**.
2. On the *<Account name>* - **Containers** page's toolbar, select **Container**.
3. In the **New container** dialog box, enter **adftutorial** for the name, and then select **OK**. The *<Account name>* - **Containers** page is updated to include **adftutorial** in the list of containers.



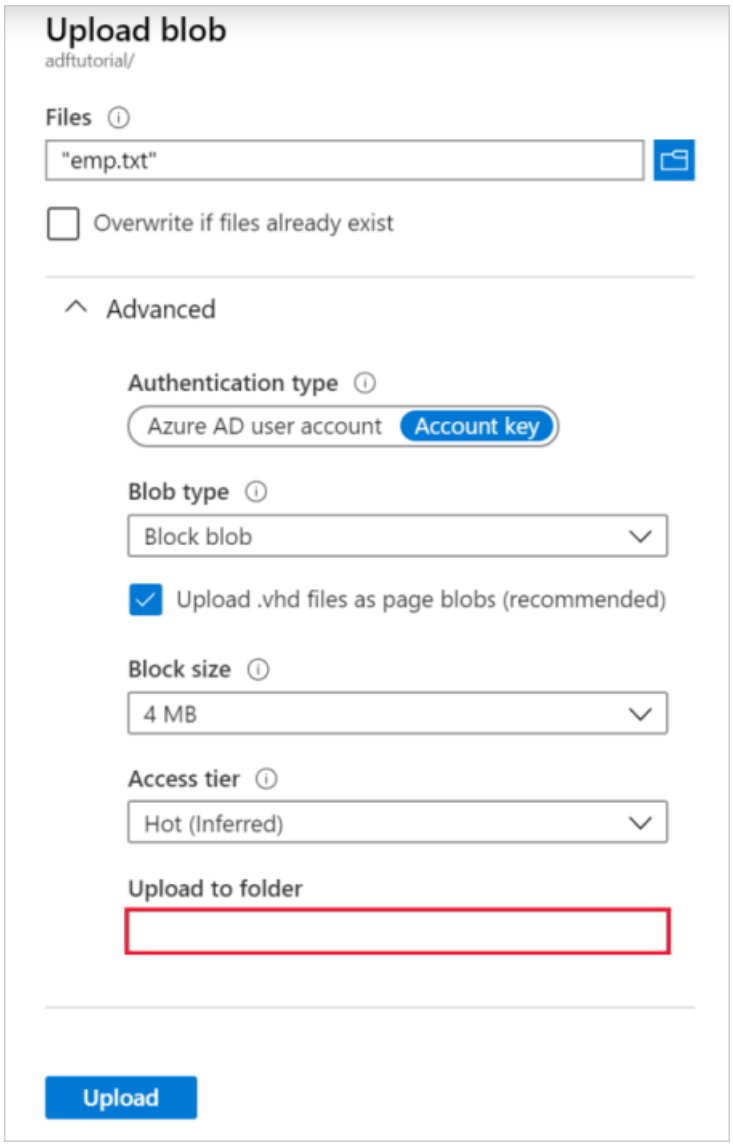
Add an input folder and file for the blob container

In this section, you create a folder named **input** in the container you created, and then upload a sample file to the input folder. Before you begin, open a text editor such as **Notepad**, and create a file named **emp.txt** with the following content:

```
John, Doe
Jane, Doe
```

Save the file in the **C:\ADFv2QuickStartPSH** folder. (If the folder doesn't already exist, create it.) Then return to the Azure portal and follow these steps:

1. In the *<Account name>* - **Containers** page where you left off, select **adftutorial** from the updated list of containers.
   a. If you closed the window or went to another page, sign in to the Azure portal again.
   b. From the Azure portal menu, select **All services**, then select **Storage** > **Storage accounts**. You can also search for and select *Storage accounts* from any page.
   c. Select your storage account, and then select **Containers** > **adftutorial**.
2. On the **adftutorial** container page's toolbar, select **Upload**.
3. In the **Upload blob** page, select the **Files** box, and then browse to and select the **emp.txt** file.
4. Expand the **Advanced** heading. The page now displays as shown:



5. In the **Upload to folder** box, enter **input**.
6. Select the **Upload** button. You should see the **emp.txt** file and the status of the upload in the list.
7. Select the **Close** icon (an **X**) to close the **Upload blob** page.

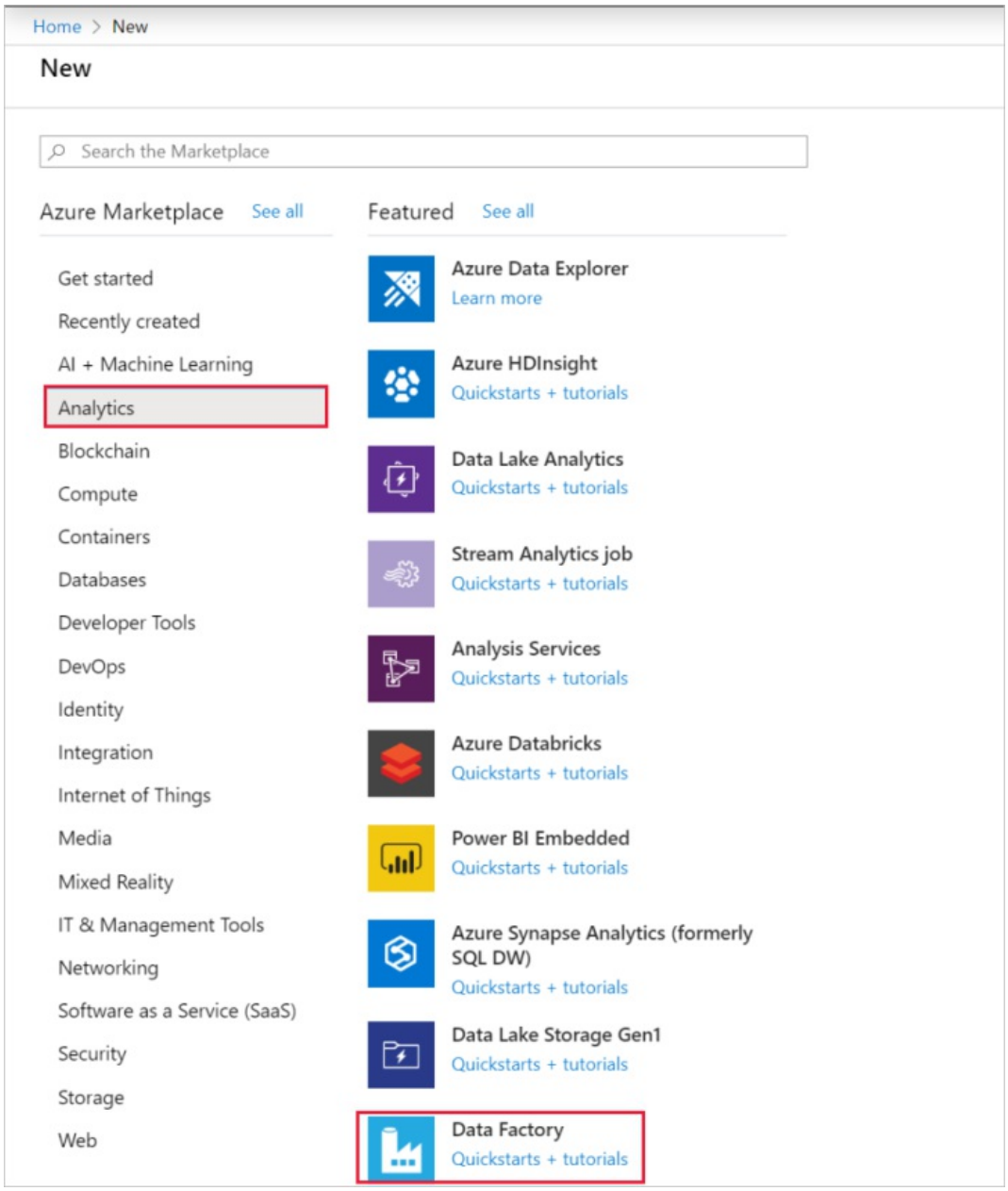
Keep the **adftutorial** container page open. You use it to verify the output at the end of this quickstart.

Video

Watching this video helps you understand the Data Factory UI:

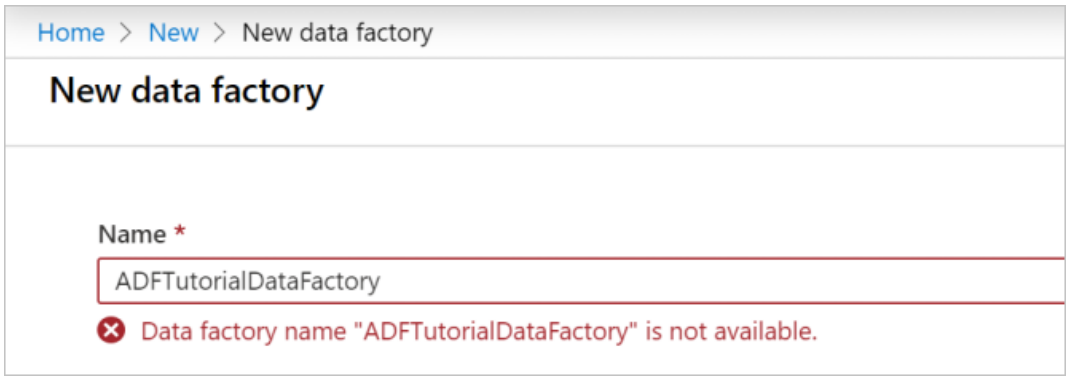
Create a data factory

1. Launch **Microsoft Edge** or **Google Chrome** web browser. Currently, Data Factory UI is supported only in Microsoft Edge and Google Chrome web browsers.
2. Go to the Azure portal.
3. From the Azure portal menu, select **Create a resource**.
4. Select **Analytics**, and then select **Data Factory**.



5. On the **New data factory** page, enter **ADFTutorialDataFactory** for **Name**.

   The name of the Azure data factory must be *globally unique*. If you see the following error, change the name of the data factory (for example, **<yourname>ADFTutorialDataFactory**) and try creating again. For naming rules for Data Factory artifacts, see the Data Factory - naming rules article.



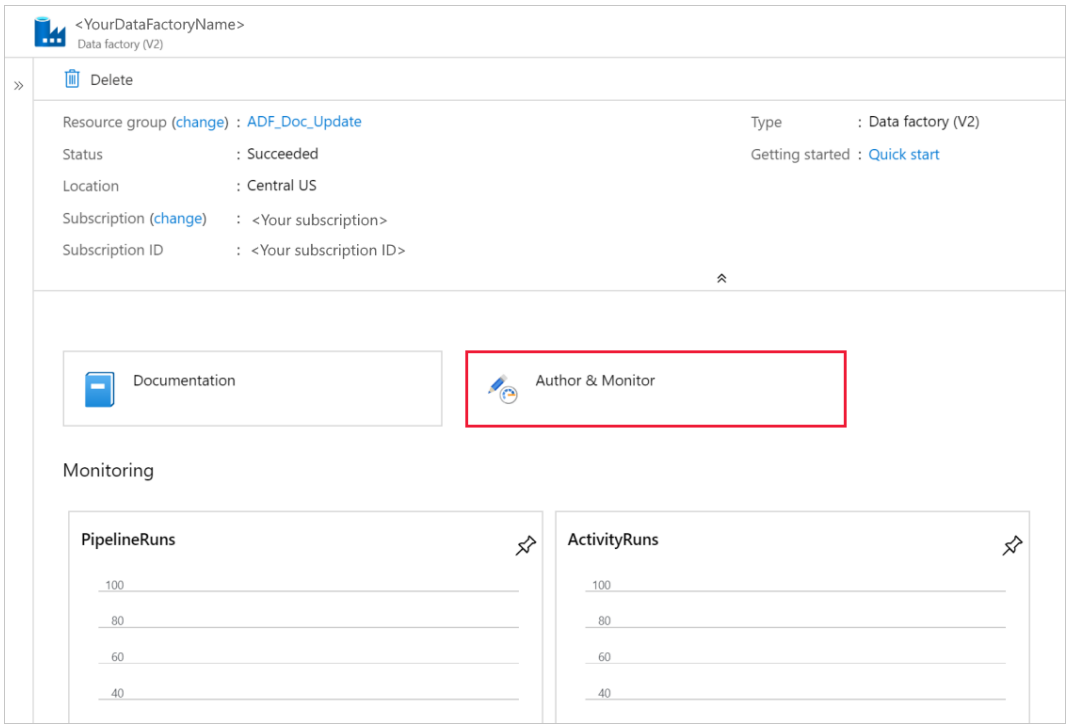
6. For **Subscription**, select your Azure subscription in which you want to create the data factory.
7. For **Resource Group**, use one of the following steps:
   - Select **Use existing**, and select an existing resource group from the list.

- Select **Create new**, and enter the name of a resource group.

  To learn about resource groups, see Using resource groups to manage your Azure resources.

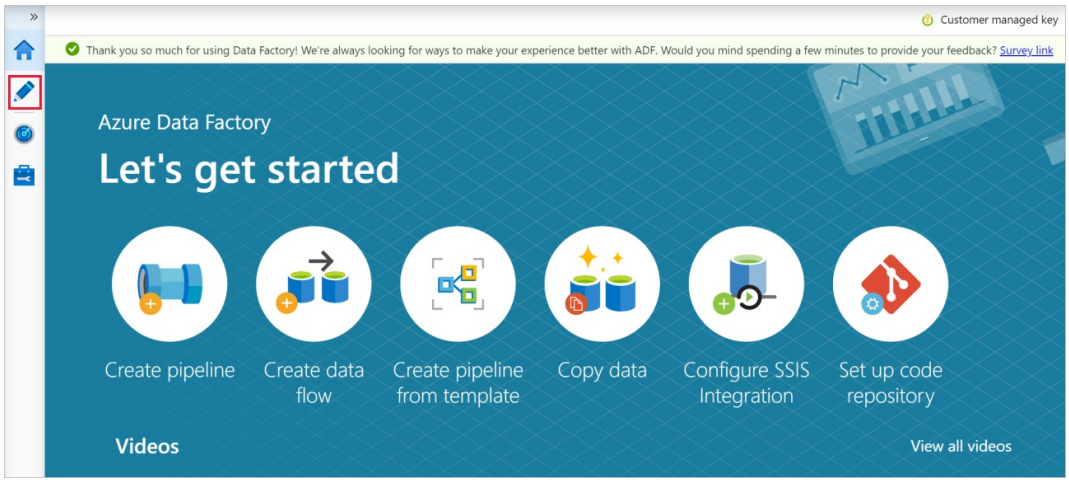
8. For **Version**, select **V2**.
9. For **Location**, select the location for the data factory.

   The list shows only locations that Data Factory supports, and where your Azure Data Factory meta data will be stored. The associated data stores (like Azure Storage and Azure SQL Database) and computes (like Azure HDInsight) that Data Factory uses can run in other regions.
10. Select **Create**. After the creation is complete, select **Go to resource** to navigate to the **Data Factory** page.
11. Select the **Author & Monitor** tile to start the Azure Data Factory user interface (UI) application on a separate tab.



> **NOTE**
>
> If you see that the web browser is stuck at "Authorizing", clear the **Block third-party cookies and site data** check box. Or keep it selected, create an exception for **login.microsoftonline.com**, and then try to open the app again.

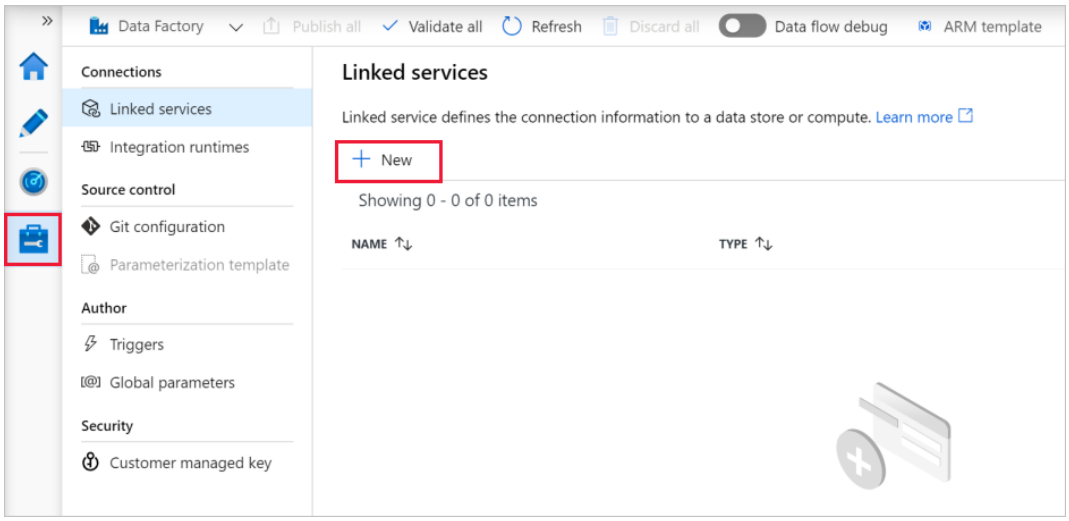
12. On the **Let's get started** page, switch to the **Author** tab in the left panel.



Create a linked service

In this procedure, you create a linked service to link your Azure Storage account to the data factory. The linked service has the connection information that the Data Factory service uses at runtime to connect to it.

1. Open Manage tab from the left pane.
2. On the Linked services page, select **+ New** to create a new linked service.



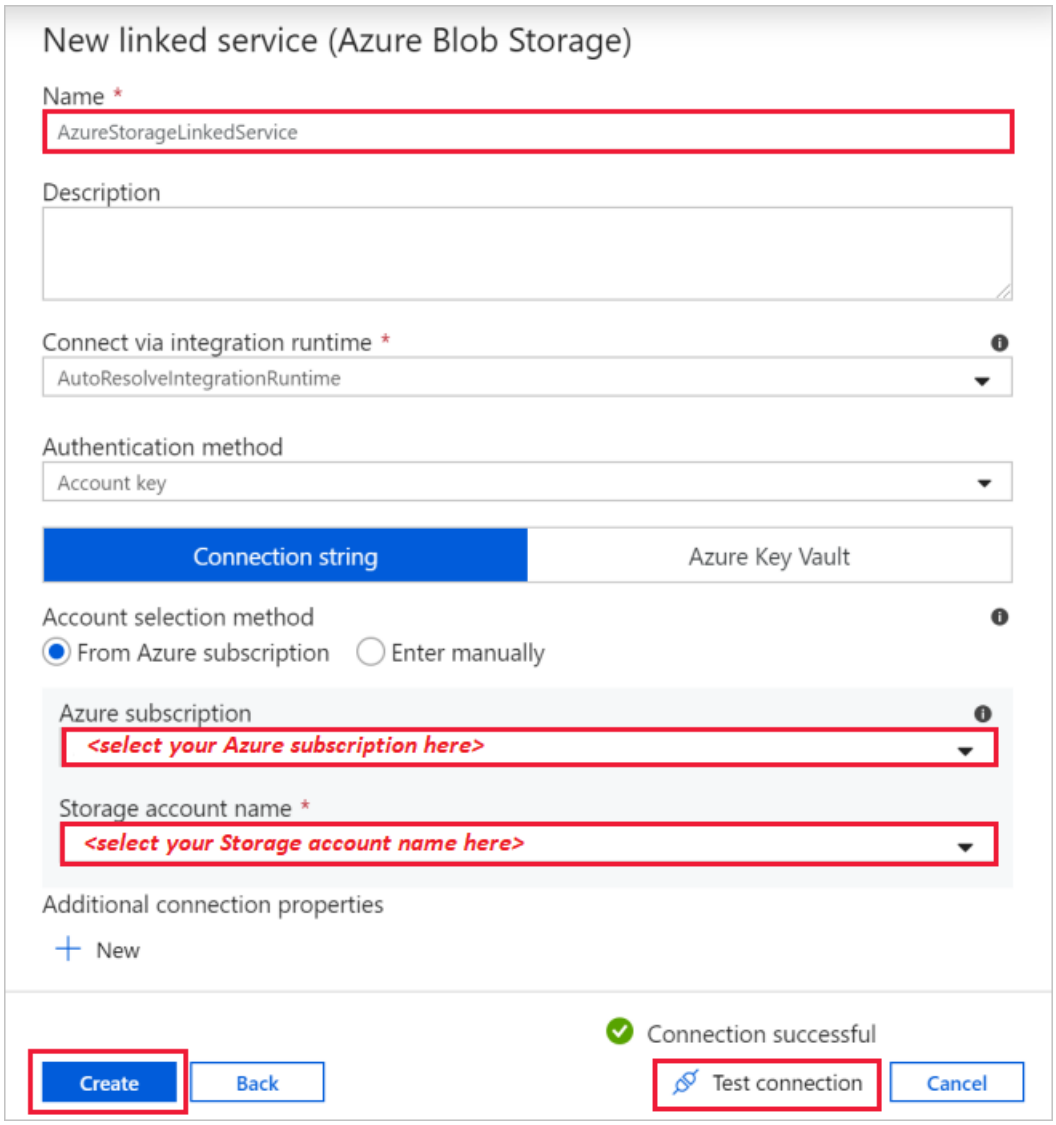
3. On the **New Linked Service** page, select **Azure Blob Storage**, and then select **Continue**.
4. On the New Linked Service (Azure Blob Storage) page, complete the following steps:

   a. For **Name**, enter **AzureStorageLinkedService**.

   b. For **Storage account name**, select the name of your Azure Storage account.

   c. Select **Test connection** to confirm that the Data Factory service can connect to the storage account.

   d. Select **Create** to save the linked service.



Create datasets

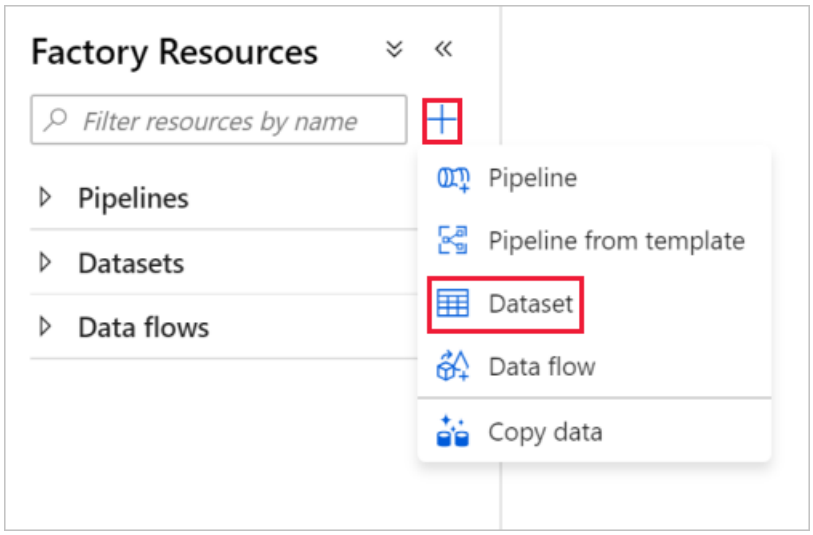
In this procedure, you create two datasets: **InputDataset** and **OutputDataset**. These datasets are of type **AzureBlob**. They refer to the Azure Storage linked service that you created in the previous section.

The input dataset represents the source data in the input folder. In the input dataset definition, you specify the blob container (**adftutorial**), the folder (**input**), and the file (**emp.txt**) that contain the source data.

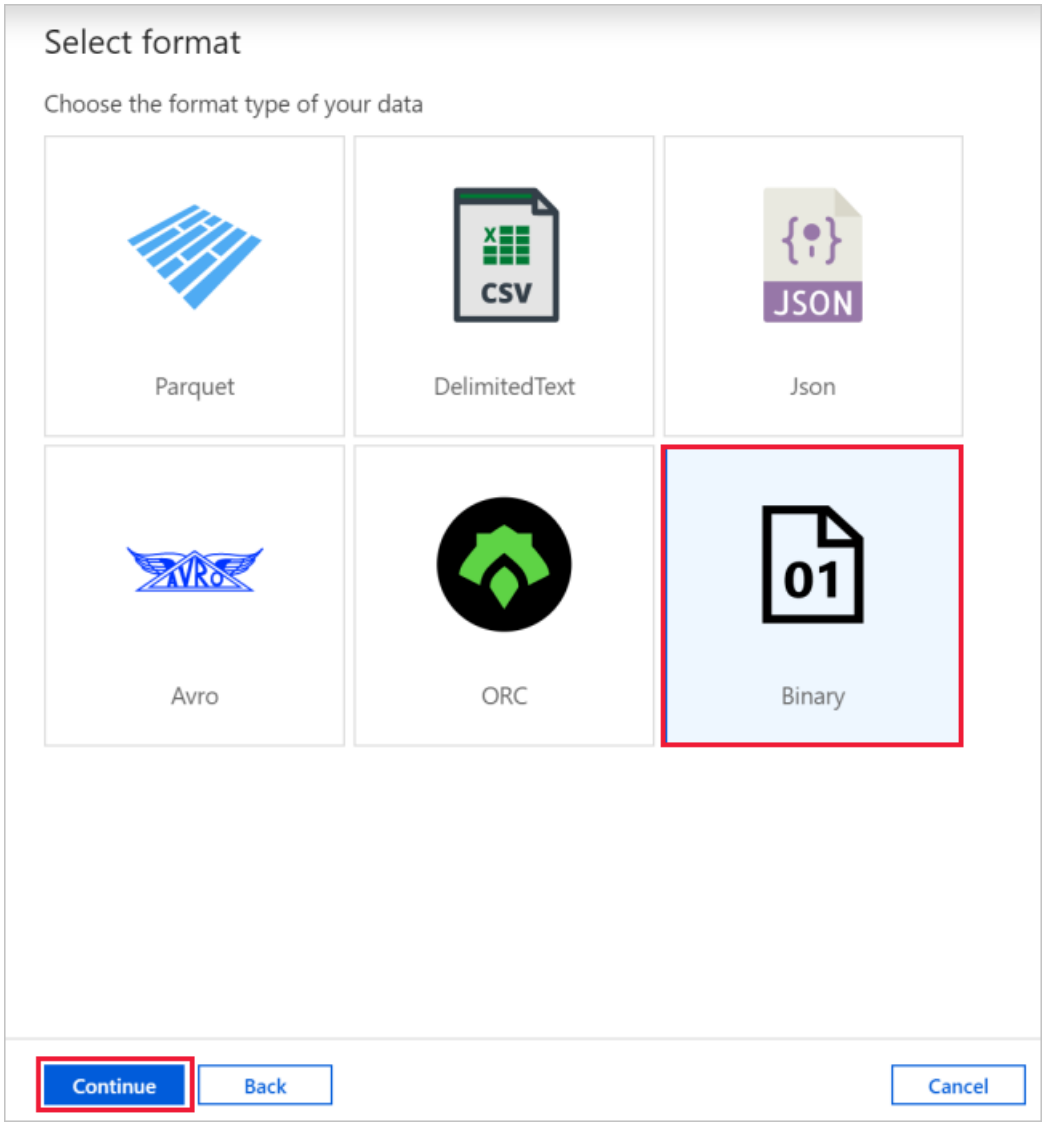
The output dataset represents the data that's copied to the destination. In the output dataset definition, you specify the blob container (**adftutorial**), the folder (**output**), and the file to which the data is copied. Each run of a pipeline has a unique ID associated with it. You can access this ID by using the system variable **RunId**. The name of the output file is dynamically evaluated based on the run ID of the pipeline.

In the linked service settings, you specified the Azure Storage account that contains the source data. In the source dataset settings, you specify where exactly the source data resides (blob container, folder, and file). In the sink dataset settings, you specify where the data is copied to (blob container, folder, and file).

1. Select the **+** (plus) button, and then select **Dataset**.



2. On the **New Dataset** page, select **Azure Blob Storage**, and then select **Continue**.
3. On the **Select Format** page, choose the format type of your data, and then select **Continue**. In this case, select **Binary** when copy files as-is without parsing the content.



4. On the **Set Properties** page, complete following steps:

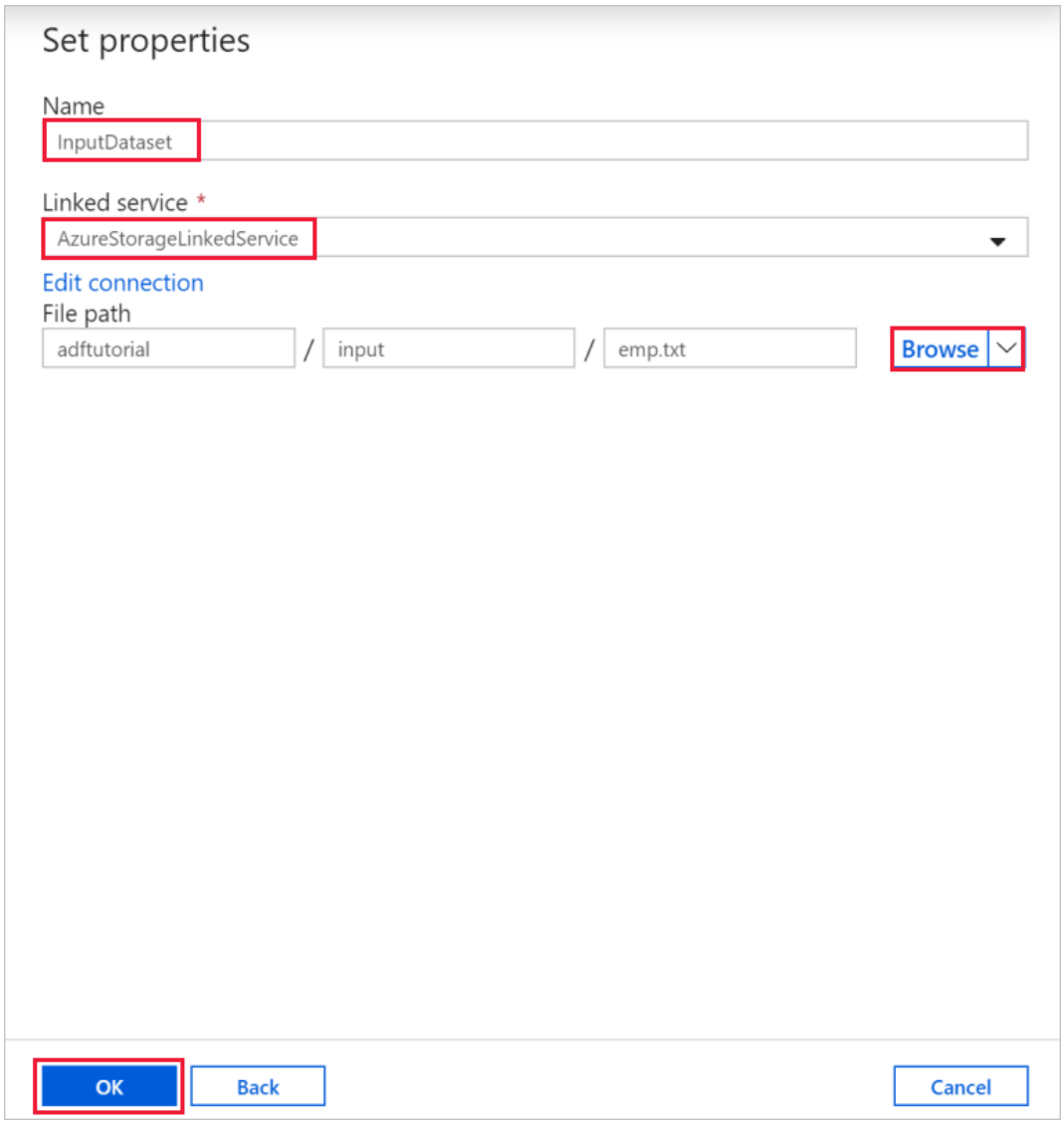
   a. Under **Name**, enter **InputDataset**.

   b. For **Linked service**, select **AzureStorageLinkedService**.

c. For **File path**, select the **Browse** button.

d. In the **Choose a file or folder** window, browse to the **input** folder in the **adftutorial** container, select the **emp.txt** file, and then select **OK**.

e. Select **OK**.



5. Repeat the steps to create the output dataset:

   a. Select the **+** (plus) button, and then select **Dataset**.

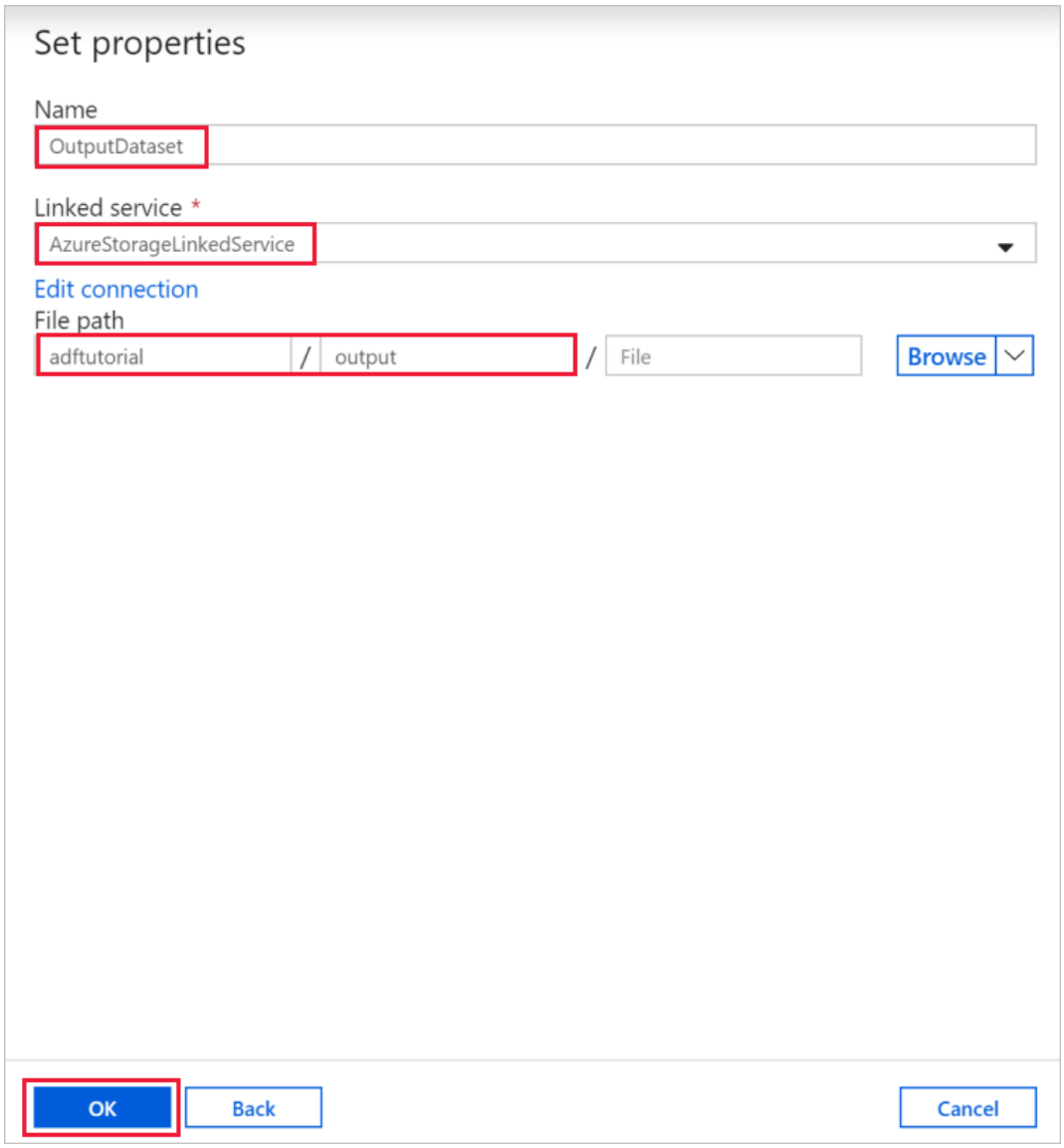
   b. On the **New Dataset** page, select **Azure Blob Storage**, and then select **Continue**.

   c. On the **Select Format** page, choose the format type of your data, and then select **Continue**.

   d. On the **Set Properties** page, specify **OutputDataset** for the name. Select **AzureStorageLinkedService** as linked service.

   e. Under **File path**, enter **adftutorial/output**. If the **output** folder doesn't exist, the copy activity creates it at runtime.

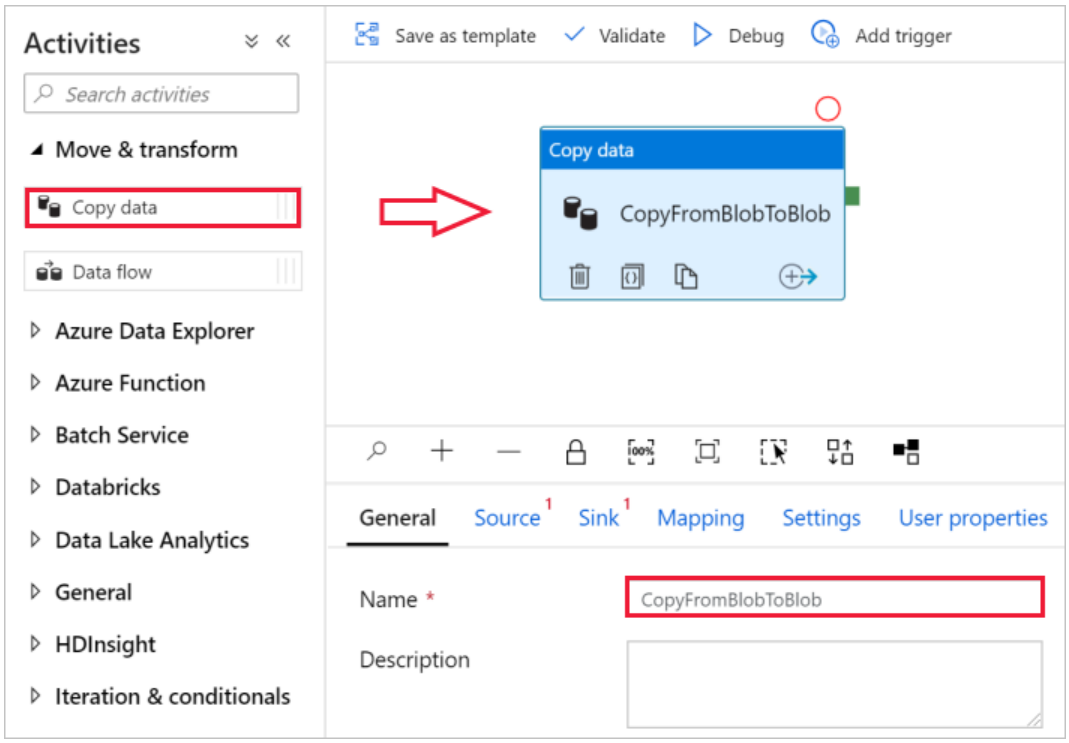
   f. Select **OK**.



Create a pipeline

In this procedure, you create and validate a pipeline with a copy activity that uses the input and output datasets. The copy activity copies data from the file you specified in the input dataset settings to the file you specified in the output dataset settings. If the input dataset specifies only a folder (not the file name), the copy activity copies all the files in the source folder to the destination.

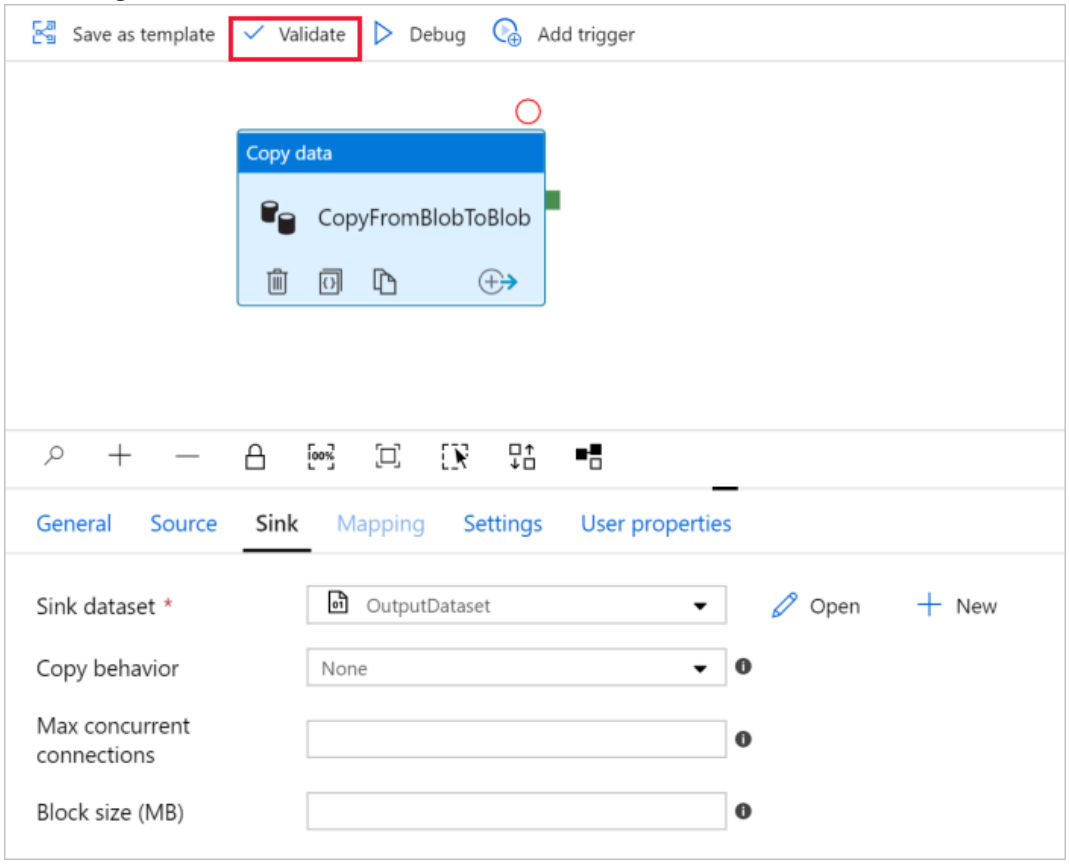
1. Select the **+** (plus) button, and then select **Pipeline**.
2. In the General panel under **Properties**, specify **CopyPipeline** for **Name**. Then collapse the panel by clicking the Properties icon in the top-right corner.
3. In the **Activities** toolbox, expand **Move & Transform**. Drag the **Copy Data** activity from the **Activities** toolbox to the pipeline designer surface. You can also search for activities in the **Activities** toolbox. Specify **CopyFromBlobToBlob** for **Name**.



4. Switch to the **Source** tab in the copy activity settings, and select **InputDataset** for **Source Dataset**.

5. Switch to the **Sink** tab in the copy activity settings, and select **OutputDataset** for **Sink Dataset**.

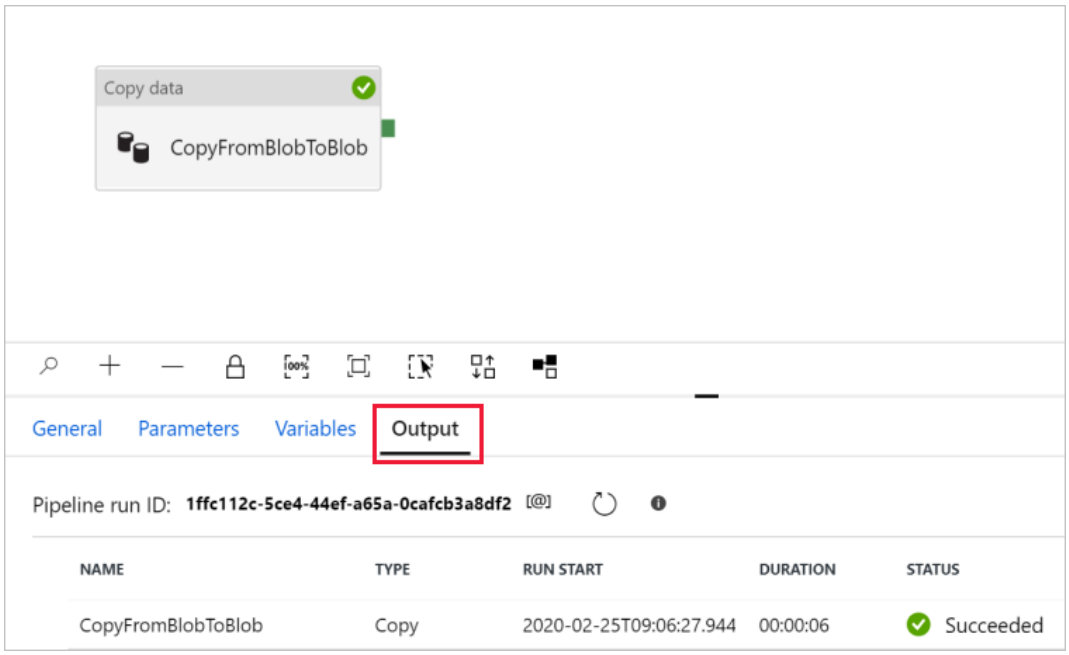
6. Click **Validate** on the pipeline toolbar above the canvas to validate the pipeline settings. Confirm that the pipeline has been successfully validated. To close the validation output, select the >> (right arrow) button.



Debug the pipeline

In this step, you debug the pipeline before deploying it to Data Factory.

1. On the pipeline toolbar above the canvas, click **Debug** to trigger a test run.
2. Confirm that you see the status of the pipeline run on the **Output** tab of the pipeline settings at the bottom.

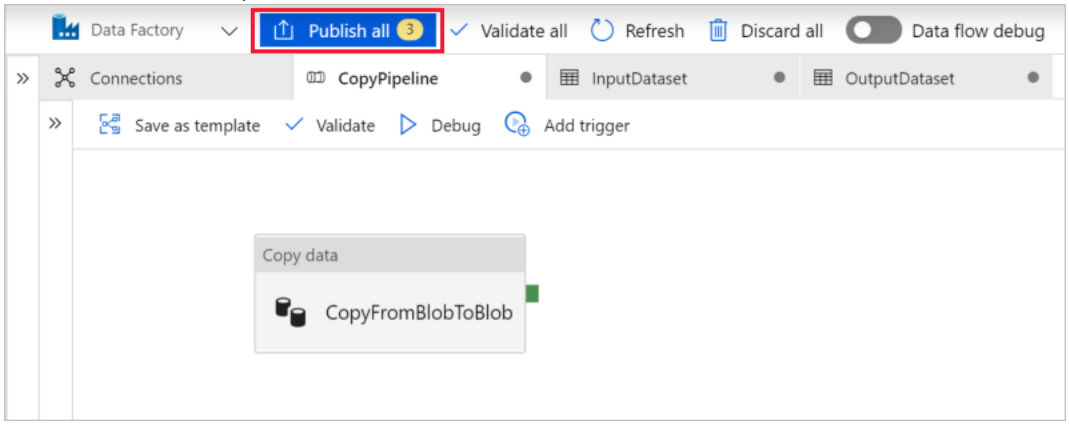


3. Confirm that you see an output file in the **output** folder of the **adftutorial** container. If the output folder doesn't exist, the Data Factory service automatically creates it.

Trigger the pipeline manually

In this procedure, you deploy entities (linked services, datasets, pipelines) to Azure Data Factory. Then, you manually trigger a pipeline run.

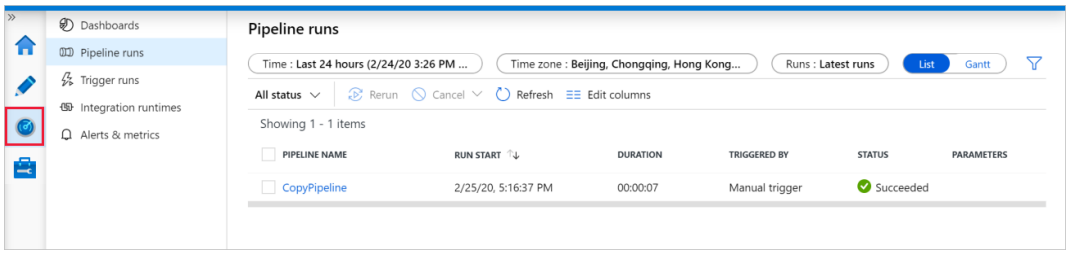
1. Before you trigger a pipeline, you must publish entities to Data Factory. To publish, select **Publish all** on the top.



2. To trigger the pipeline manually, select **Add Trigger** on the pipeline toolbar, and then select **Trigger Now**. On the **Pipeline run** page, select **Finish**.

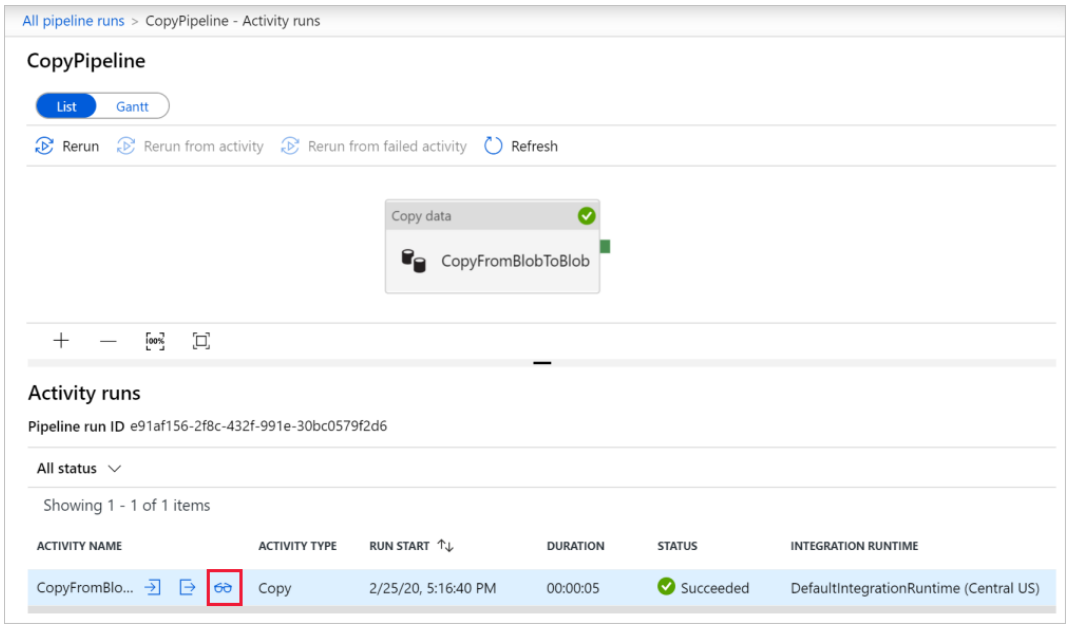
Monitor the pipeline

1. Switch to the **Monitor** tab on the left. Use the **Refresh** button to refresh the list.



2. Select the **CopyPipeline** link, you'll see the status of the copy activity run on this page.

3. To view details about the copy operation, select the **Details** (eyeglasses image) link. For details about the properties, see Copy Activity overview.



4. Confirm that you see a new file in the **output** folder.

5. You can switch back to the **Pipeline runs** view from the **Activity runs** view by selecting the **All pipeline runs** link.

Trigger the pipeline on a schedule

This procedure is optional in this tutorial. You can create a *scheduler trigger* to schedule the pipeline to run periodically (hourly, daily, and so on). In this procedure, you create a trigger to run every minute until the end date and time that you specify.

1. Switch to the **Author** tab.

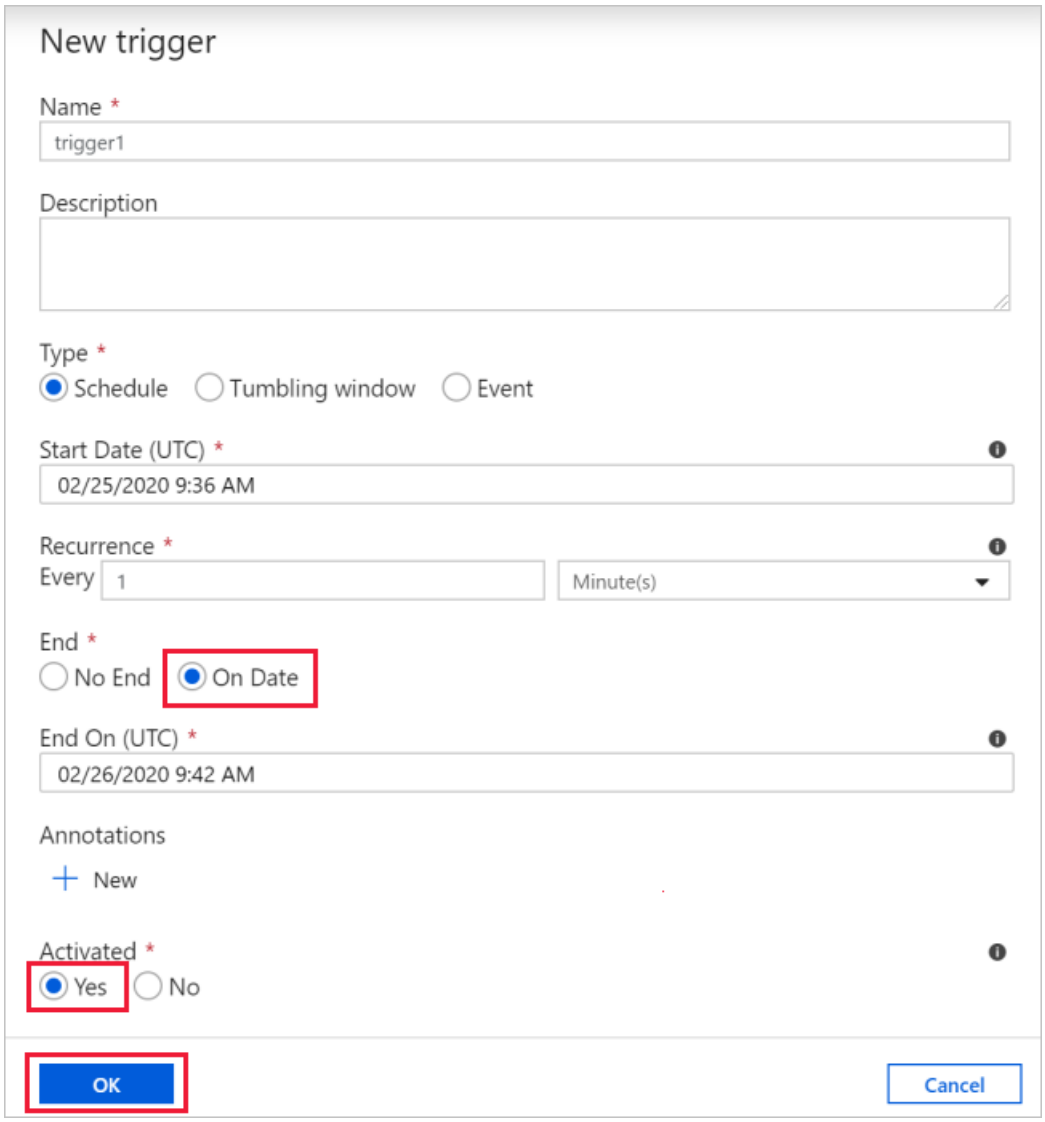
2. Go to your pipeline, select **Add Trigger** on the pipeline toolbar, and then select **New/Edit**.

3. On the **Add Triggers** page, select **Choose trigger**, and then select **New**.

4. On the **New Trigger** page, under **End**, select **On Date**, specify an end time a few minutes after the current time, and then select **OK**.

   A cost is associated with each pipeline run, so specify the end time only minutes apart from the start time. Ensure that it's the same day. However, ensure that there's enough time for the pipeline to run between the publish time and the end time. The trigger comes into effect only after you publish the solution to Data Factory, not when you save the trigger in the UI.

5. On the **New Trigger** page, select the **Activated** check box, and then select **OK**.



6. Review the warning message, and select **OK**.
7. Select **Publish all** to publish changes to Data Factory.
8. Switch to the **Monitor** tab on the left. Select **Refresh** to refresh the list. You see that the pipeline runs once every minute from the publish time to the end time.

   Notice the values in the **TRIGGERED BY** column. The manual trigger run was from the step (**Trigger Now**) that you did earlier.
9. Switch to the **Trigger runs** view.
10. Confirm that an output file is created for every pipeline run until the specified end date and time in the **output** folder.

Next steps

The pipeline in this sample copies data from one location to another location in Azure Blob storage. To learn about using Data Factory in more scenarios, go through the tutorials.

Quickstart: Use the Copy Data tool to copy data

9/25/2020 • 7 minutes to read • Edit Online

APPLIES TO: Azure Data Factory Azure Synapse Analytics (Preview)

In this quickstart, 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 a folder in Azure Blob storage to another folder.

> **NOTE**
> If you are new to Azure Data Factory, see Introduction to Azure Data Factory before doing this quickstart.

Prerequisites

Azure subscription

If you don't have an Azure subscription, create a free account before you begin.

Azure roles

To create Data Factory instances, the user account that you use to sign in to Azure must be a member of the *contributor* or *owner* role, or an *administrator* of the Azure subscription. To view the permissions that you have in the subscription, go to the Azure portal, select your username in the upper-right corner, select "..." icon for more options, and then select **My permissions**. If you have access to multiple subscriptions, select the appropriate subscription.

To create and manage child resources for Data Factory - including datasets, linked services, pipelines, triggers, and integration runtimes - the following requirements are applicable:

- To create and manage child resources in the Azure portal, you must belong to the **Data Factory Contributor** role at the resource group level or above.
- To create and manage child resources with PowerShell or the SDK, the **contributor** role at the resource level or above is sufficient.

For sample instructions about how to add a user to a role, see the Add roles article.

For more info, see the following articles:

- Data Factory Contributor role
- Roles and permissions for Azure Data Factory

Azure Storage account

You use a general-purpose Azure Storage account (specifically Blob storage) as both *source* and *destination* data stores in this quickstart. If you don't have a general-purpose Azure Storage account, see Create a storage account to create one.

Get the storage account name

You need the name of your Azure Storage account for this quickstart. The following procedure provides steps to get the name of your storage account:

1. In a web browser, go to the Azure portal and sign in using your Azure username and password.
2. From the Azure portal menu, select **All services**, then select **Storage** > **Storage accounts**. You can

also search for and select *Storage accounts* from any page.

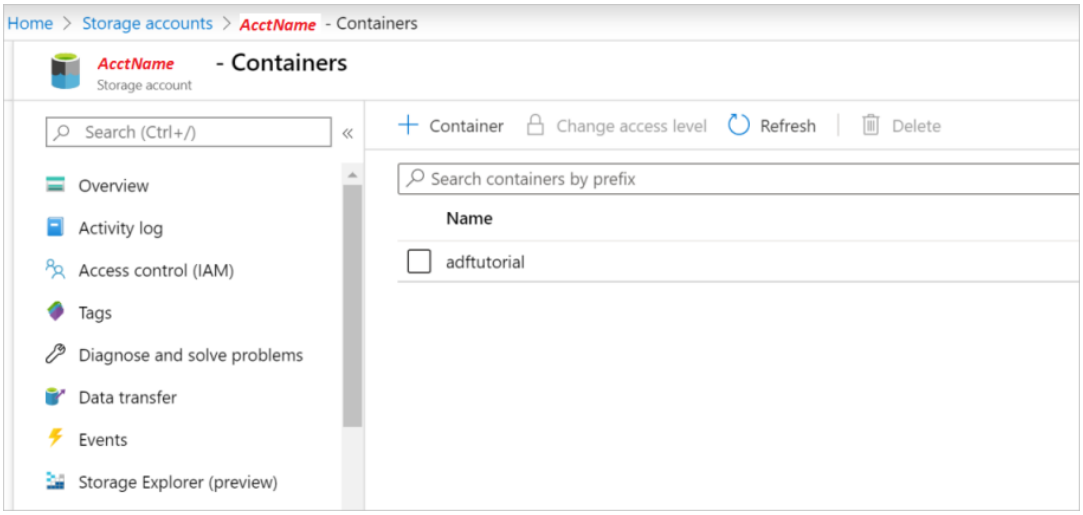
3. In the **Storage accounts** page, filter for your storage account (if needed), and then select your storage account.

You can also search for and select *Storage accounts* from any page.

Create a blob container

In this section, you create a blob container named **adftutorial** in Azure Blob storage.

1. From the storage account page, select **Overview** > **Containers**.
2. On the *<Account name>* - **Containers** page's toolbar, select **Container**.
3. In the **New container** dialog box, enter **adftutorial** for the name, and then select **OK**. The *<Account name>* - **Containers** page is updated to include **adftutorial** in the list of containers.



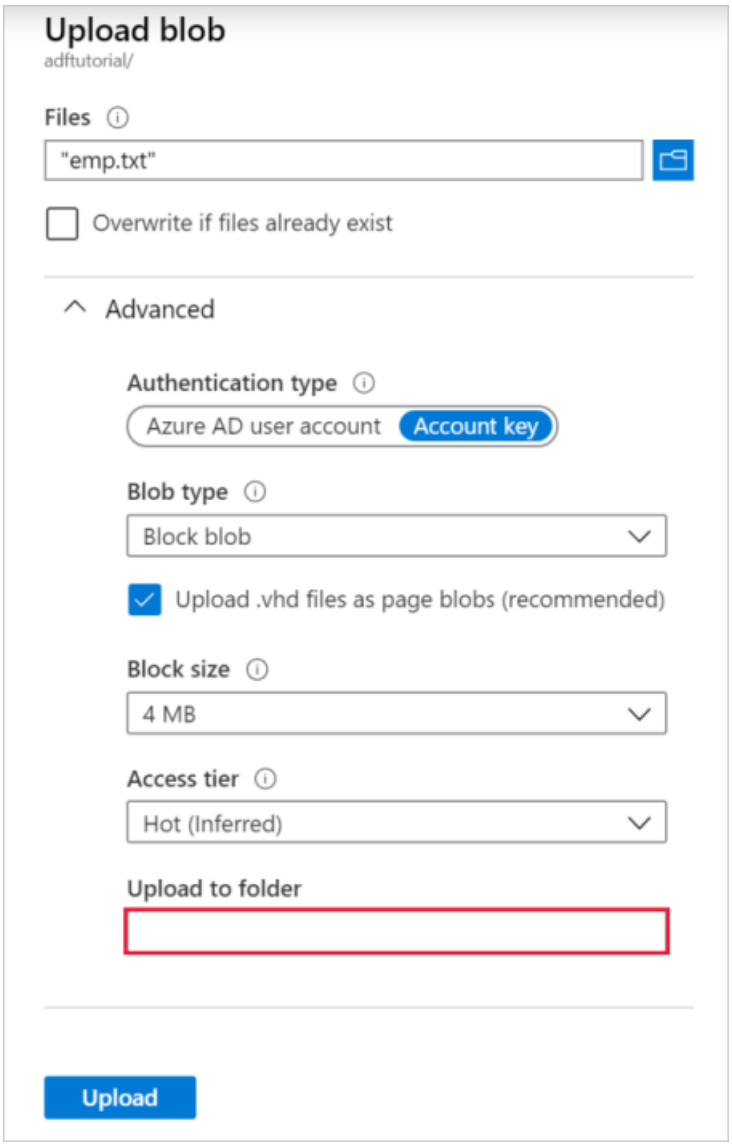
Add an input folder and file for the blob container

In this section, you create a folder named **input** in the container you created, and then upload a sample file to the input folder. Before you begin, open a text editor such as **Notepad**, and create a file named **emp.txt** with the following content:

```
John, Doe
Jane, Doe
```

Save the file in the **C:\ADFv2QuickStartPSH** folder. (If the folder doesn't already exist, create it.) Then return to the Azure portal and follow these steps:

1. In the *<Account name>* - **Containers** page where you left off, select **adftutorial** from the updated list of containers.
   a. If you closed the window or went to another page, sign in to the Azure portal again.
   b. From the Azure portal menu, select **All services**, then select **Storage** > **Storage accounts**. You can also search for and select *Storage accounts* from any page.
   c. Select your storage account, and then select **Containers** > **adftutorial**.
2. On the **adftutorial** container page's toolbar, select **Upload**.
3. In the **Upload blob** page, select the **Files** box, and then browse to and select the **emp.txt** file.
4. Expand the **Advanced** heading. The page now displays as shown:

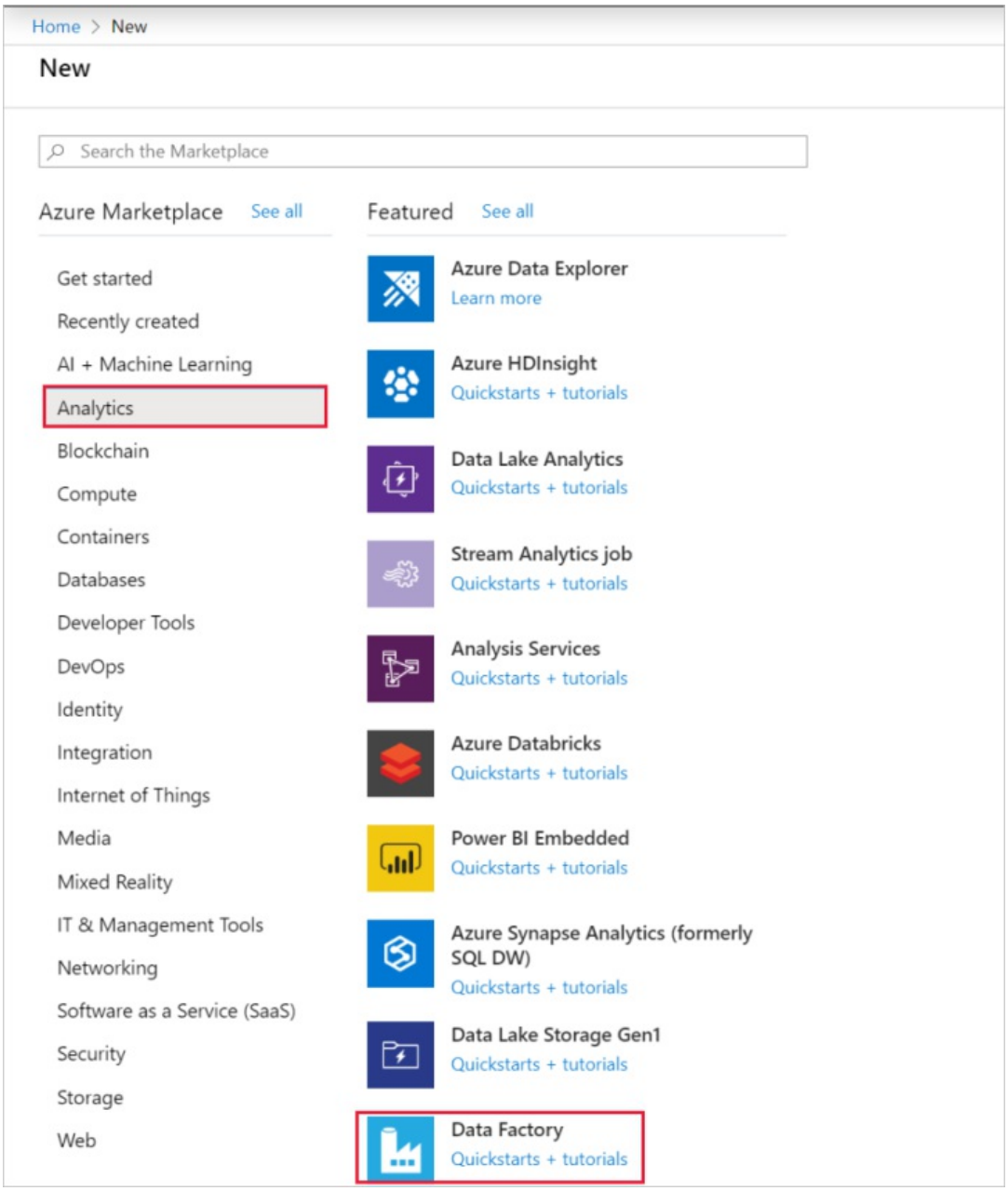


5. In the **Upload to folder** box, enter **input**.
6. Select the **Upload** button. You should see the **emp.txt** file and the status of the upload in the list.
7. Select the **Close** icon (an **X**) to close the **Upload blob** page.

Keep the **adftutorial** container page open. You use it to verify the output at the end of this quickstart.

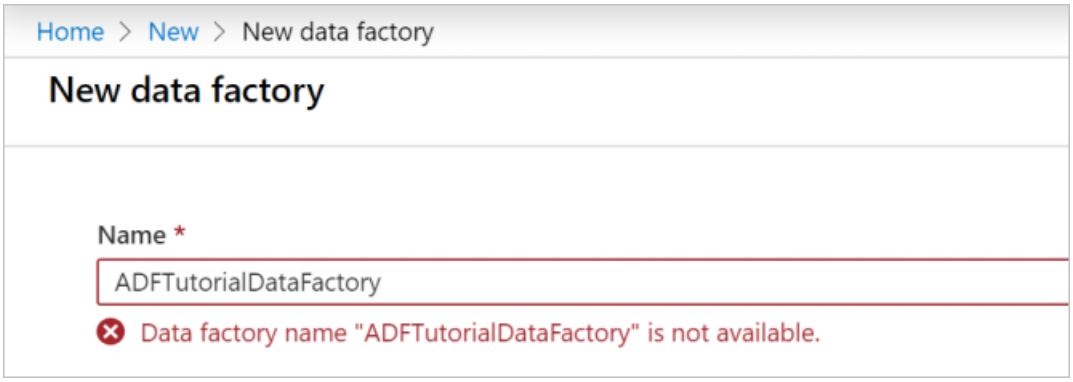
Create a data factory

1. Launch **Microsoft Edge** or **Google Chrome** web browser. Currently, Data Factory UI is supported only in Microsoft Edge and Google Chrome web browsers.
2. Go to the Azure portal.
3. From the Azure portal menu, select **Create a resource** > **Analytics** > **Data Factory**:



4. On the **New data factory** page, enter **ADFTutorialDataFactory** for **Name**.

   The name of the Azure data factory must be *globally unique*. If you see the following error, change the name of the data factory (for example, **<yourname>ADFTutorialDataFactory**) and try creating again. For naming rules for Data Factory artifacts, see the Data Factory - naming rules article.



5. For **Subscription**, select your Azure subscription in which you want to create the data factory.

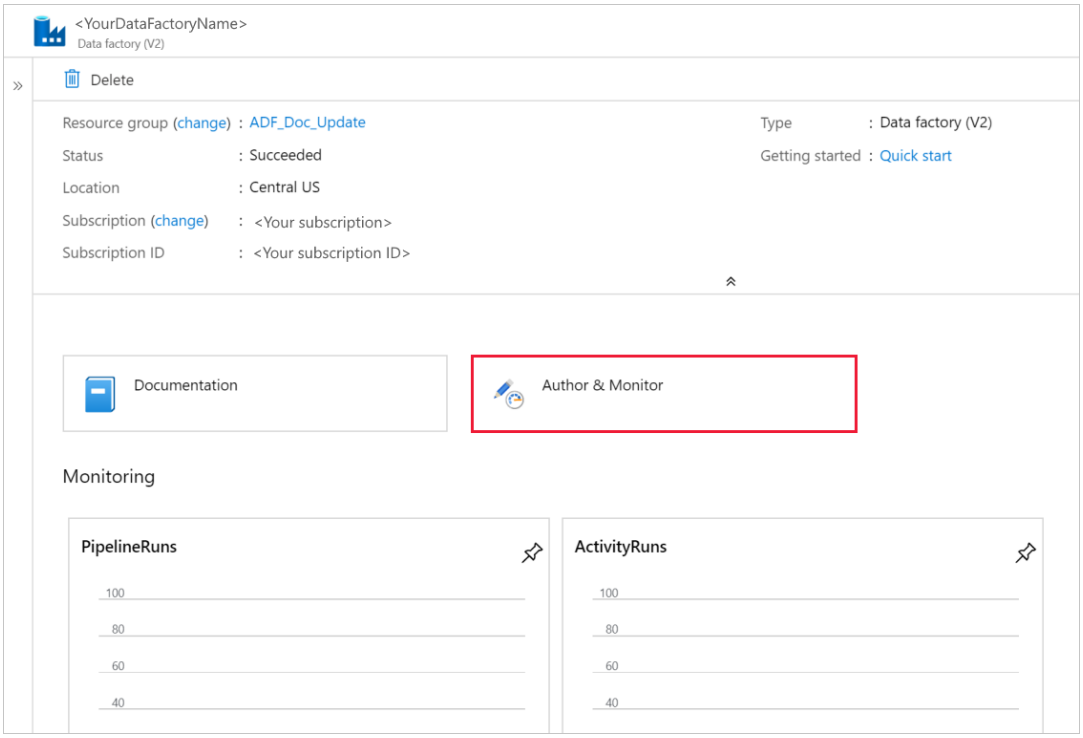
6. For **Resource Group**, use one of the following steps:

- Select **Use existing**, and select an existing resource group from the list.
- Select **Create new**, and enter the name of a resource group.

To learn about resource groups, see Using resource groups to manage your Azure resources.

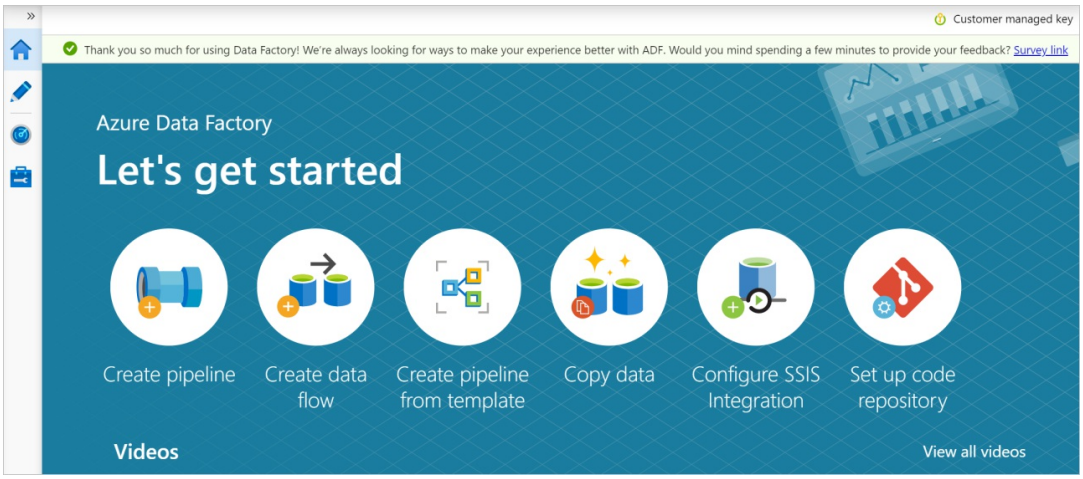
7. For **Version**, select **V2**.
8. For **Location**, select the location for the data factory.

   The list shows only locations that Data Factory supports, and where your Azure Data Factory meta data will be stored. The associated data stores (like Azure Storage and Azure SQL Database) and computes (like Azure HDInsight) that Data Factory uses can run in other regions.
9. Select **Create**.
10. After the creation is complete, you see the **Data Factory** page. Select the **Author & Monitor** tile to start the Azure Data Factory user interface (UI) application on a separate tab.

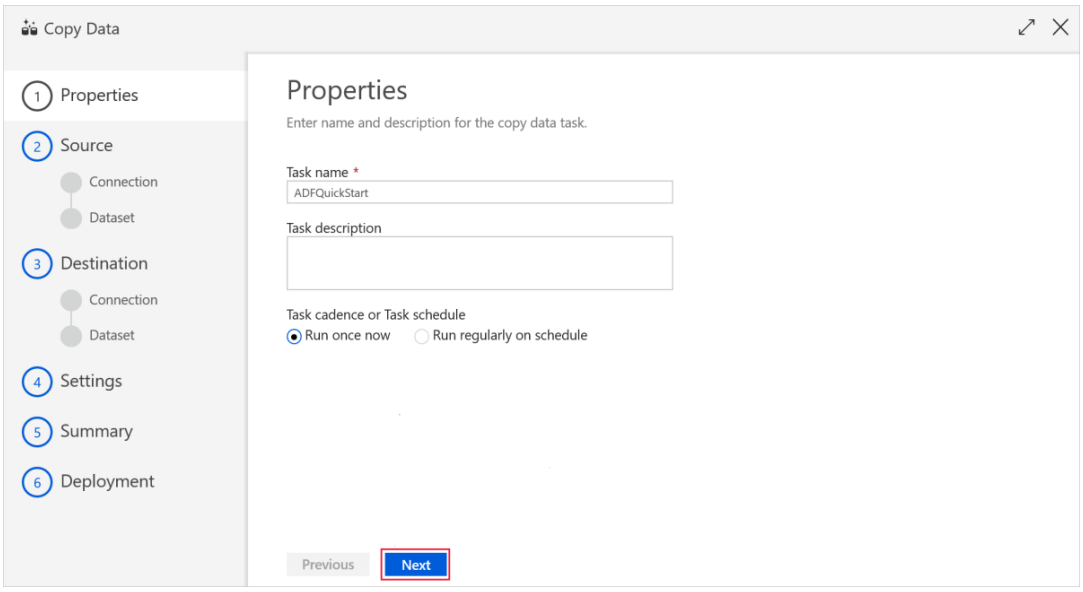


Start the Copy Data tool

1. On the **Let's get started** page, select the **Copy Data** tile to start the Copy Data tool.



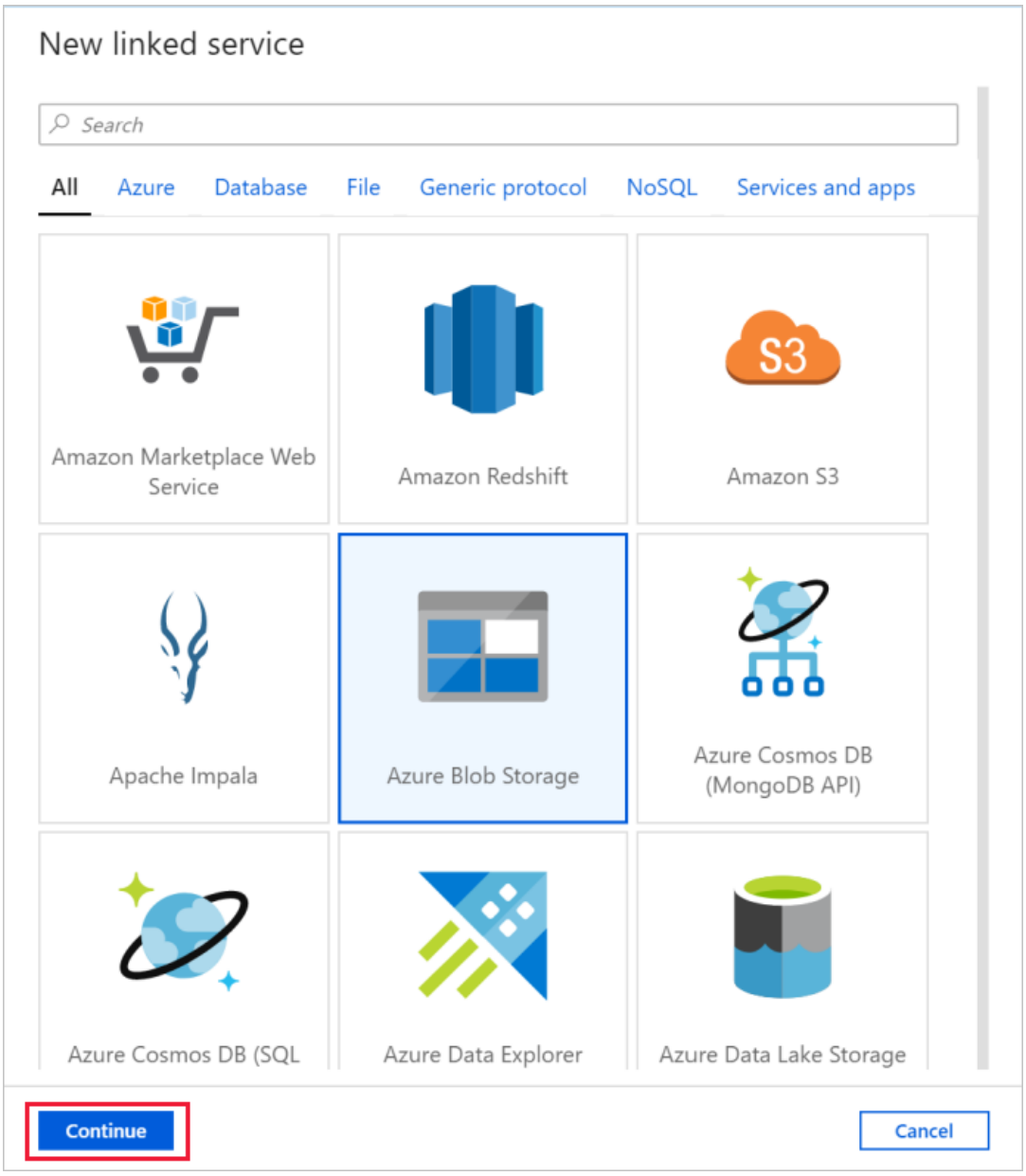
2. On the **Properties** page of the Copy Data tool, you can specify a name for the pipeline and its description, then select **Next**.



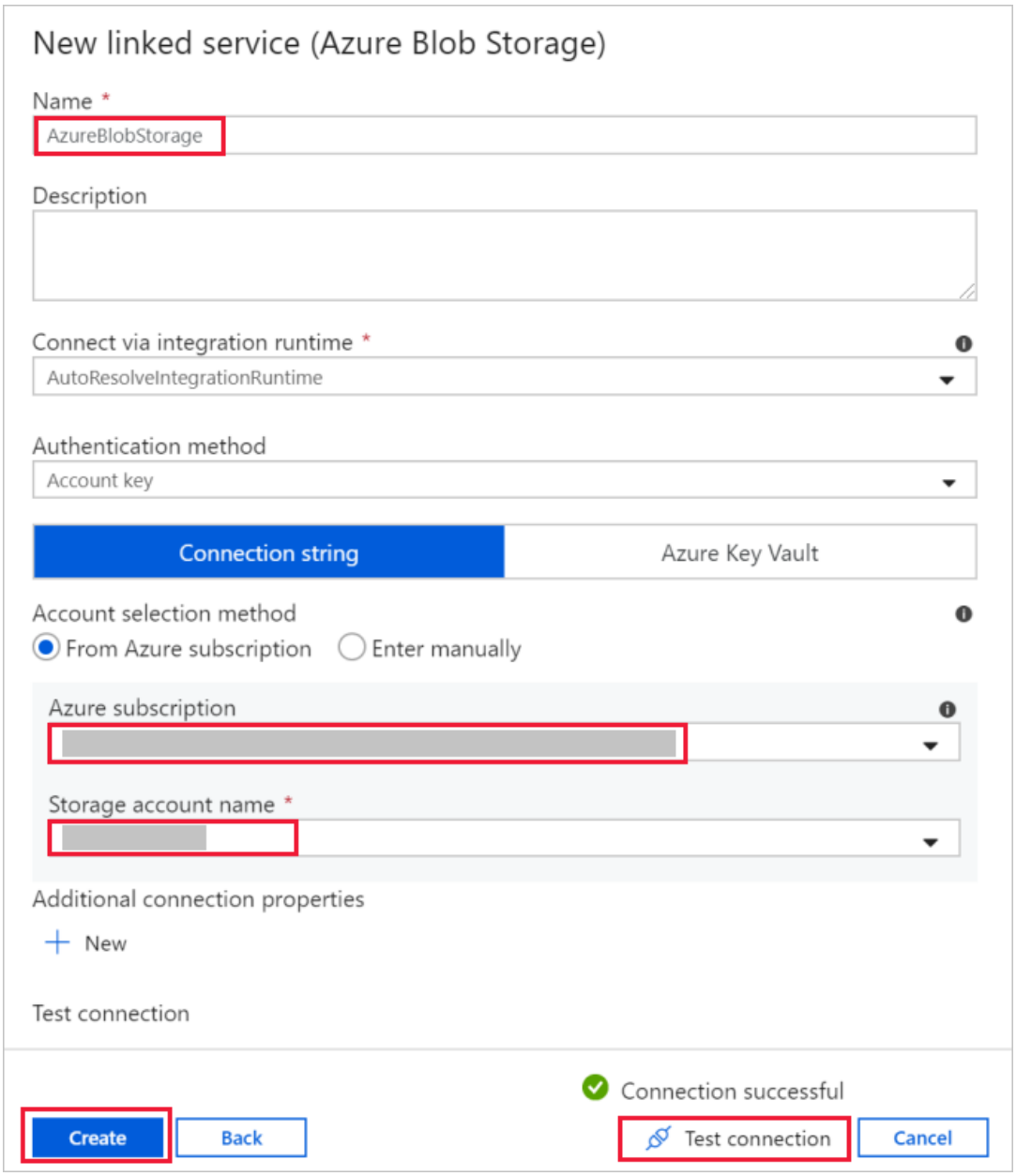
3. On the **Source data store** page, complete the following steps:

   a. Click **+ Create new connection** to add a connection.

   b. Select the linked service type that you want to create for the source connection. In this tutorial, we use **Azure Blob Storage**. Select it from the gallery, and then select **Continue**.



c. On the **New Linked Service (Azure Blob Storage)** page, specify a name for your linked service. Select your storage account from the **Storage account name** list, test connection, and then select **Create**.

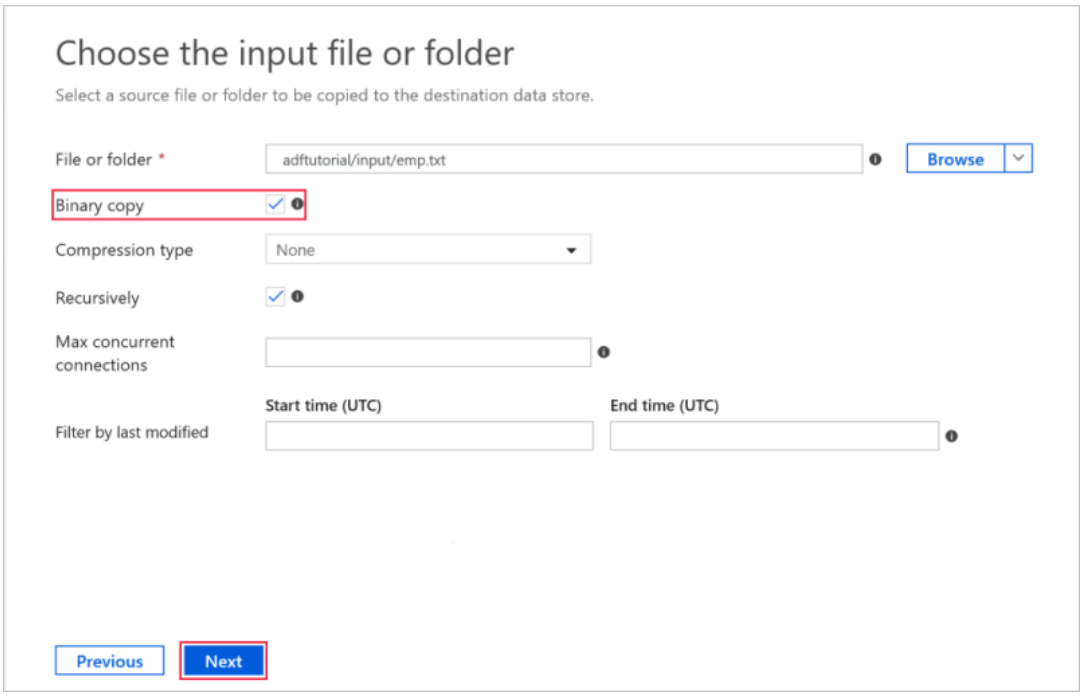


d. Select the newly created linked service as source, and then click **Next**.

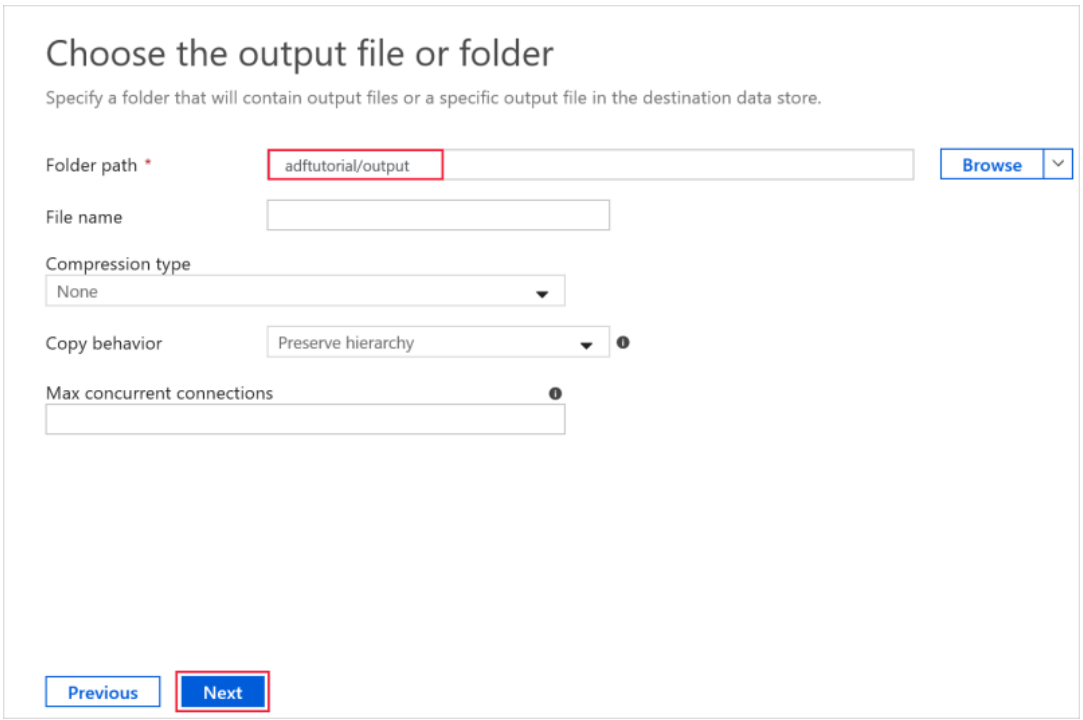
4. On the **Choose the input file or folder** page, complete the following steps:

   a. Click **Browse** to navigate to the **adftutorial/input** folder, select the **emp.txt** file, and then click **Choose**.

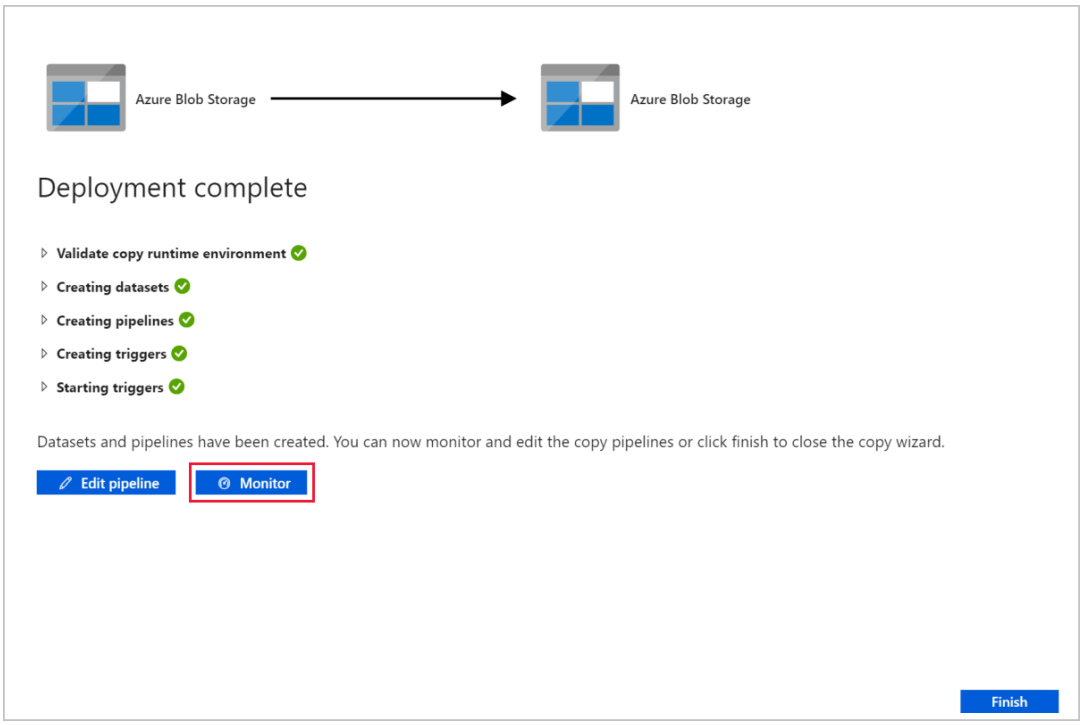
   d. Select the **Binary copy** checkbox to copy file as-is, and then select **Next**.



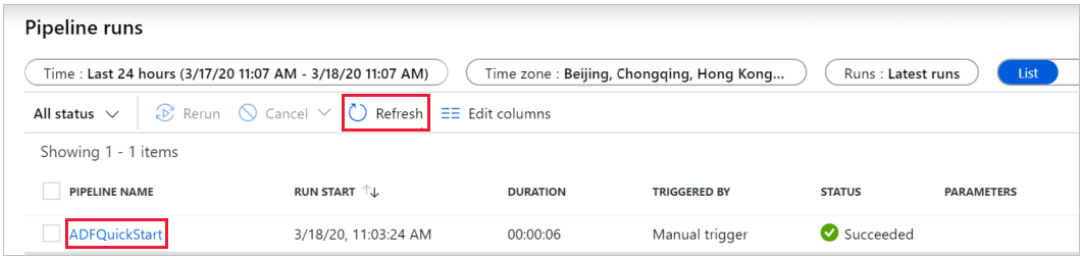
5. On the **Destination data store** page, select the **Azure Blob Storage** linked service you created, and then select **Next**.
6. On the **Choose the output file or folder** page, enter **adftutorial/output** for the folder path, and then select **Next**.



7. On the **Settings** page, select **Next** to use the default configurations.
8. On the **Summary** page, review all settings, and select **Next**.
9. On the **Deployment complete** page, select **Monitor** to monitor the pipeline that you created.



10. The application switches to the **Monitor** tab. You see the status of the pipeline on this tab. Select **Refresh** to refresh the list. Click the link under **PIPELINE NAME** to view activity run details or rerun the pipeline.

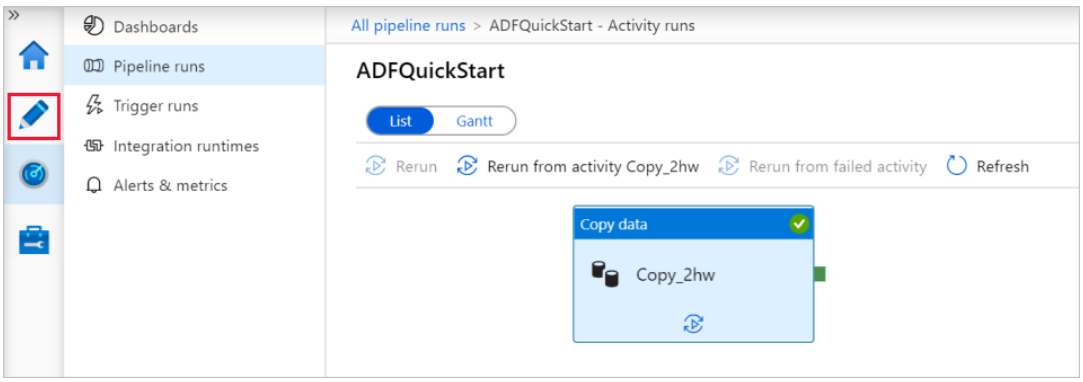


11. On the Activity runs page, select the **Details** link (eyeglasses icon) under the **ACTIVITY NAME** column for more details about copy operation. For details about the properties, see Copy Activity overview.

12. To go back to the Pipeline Runs view, select the **ALL pipeline runs** link in the breadcrumb menu. To refresh the view, select **Refresh**.

13. Verify that the **emp.txt** file is created in the **output** folder of the **adftutorial** container. If the output folder doesn't exist, the Data Factory service automatically creates it.

14. Switch to the **Author** tab above the **Monitor** tab on the left panel so that you can edit linked services, datasets, and pipelines. To learn about editing them in the Data Factory UI, see Create a data factory by using the Azure portal.



Next steps

The pipeline in this sample copies data from one location to another location in Azure Blob storage. To learn about using Data Factory in more scenarios, go through the tutorials.