

Challenge 3: Setup Azure Data Factory

Duration: 20 minutes

In this exercise, you will create a baseline environment for Azure Data Factory development to further operationalize data movement and processing. You will create a Data Factory service and then install the Data Management Gateway, which is the agent that facilitates data movement from on-premises to Microsoft Azure.

Task 1: Download and stage data to be processed

1. Open a web browser.
2. Download the AdventureWorks sample data from <http://bit.ly/2zi4Sqa>. If you are having trouble downloading the file, a zip file called FlightsAndWeather.zip is included in the lab-files folders.

Note: If you are using the optional VM provisioned in the Before the HOL document, ensure that you download and extract the data on the VM.

3. Extract it to a new folder called **C:\Data**.

Task 2: Install and configure Azure Data Factory Integration Runtime on your machine

1. To download the latest version of Azure Data Factory Integration Runtime, go to <https://www.microsoft.com/en-us/download/details.aspx?id=39717>.

Note: If you are using the optional VM provisioned in the Before the HOL document, ensure that you install the IR on the VM.

2. Select Download, then choose the download you want from the next screen.

Choose the download you want

<input type="checkbox"/> File Name	Size
<input checked="" type="checkbox"/> IntegrationRuntime_5.10.7949.4.msi	871.8 MB
<input type="checkbox"/> IntegrationRuntime_5.9.7949.2.msi	871.5 MB
<input type="checkbox"/> Release Notes.doc	233 KB

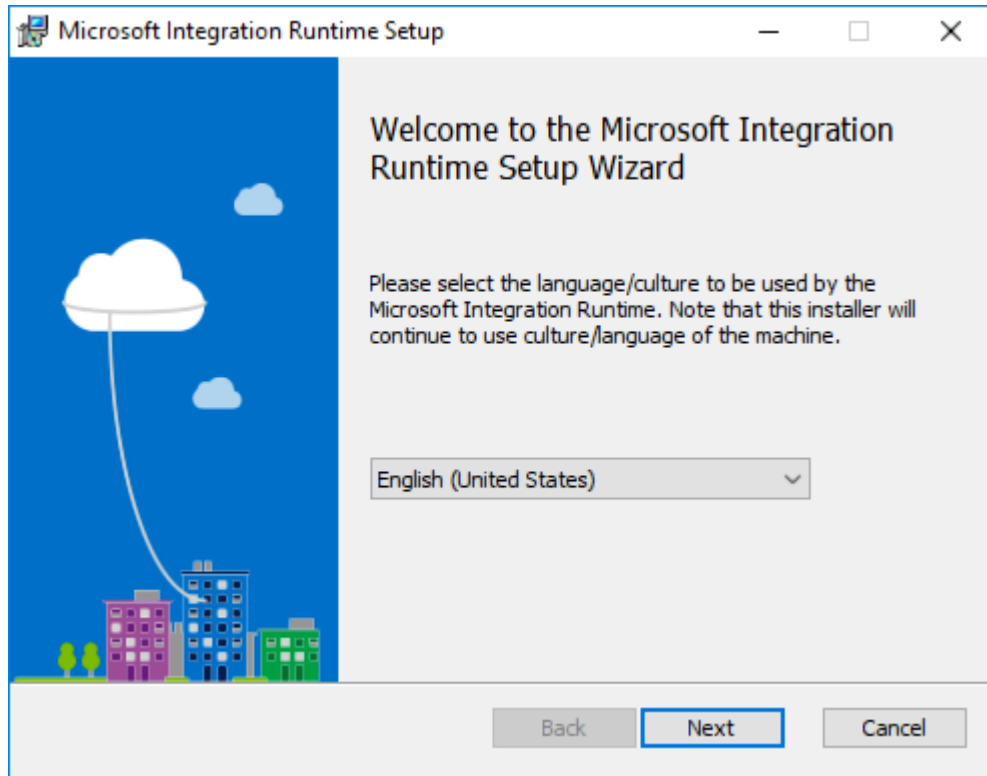
Download Summary:
KBMBGB

1. IntegrationRuntime_5.10.7949.4.msi

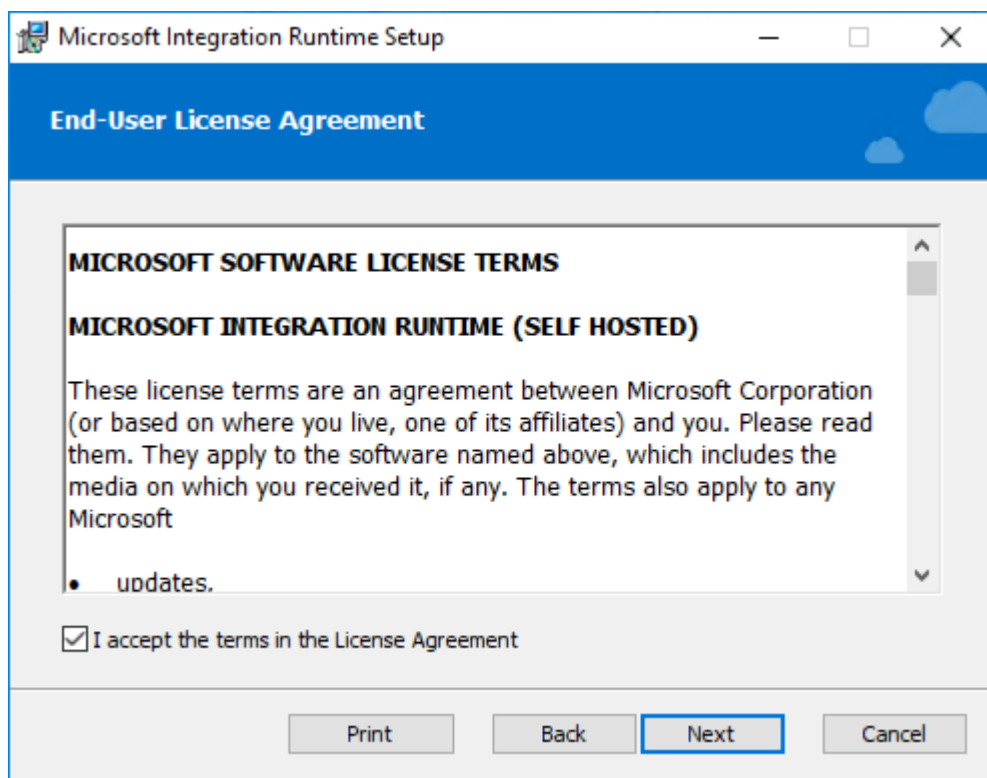
Total Size: 871.8 MB

Next

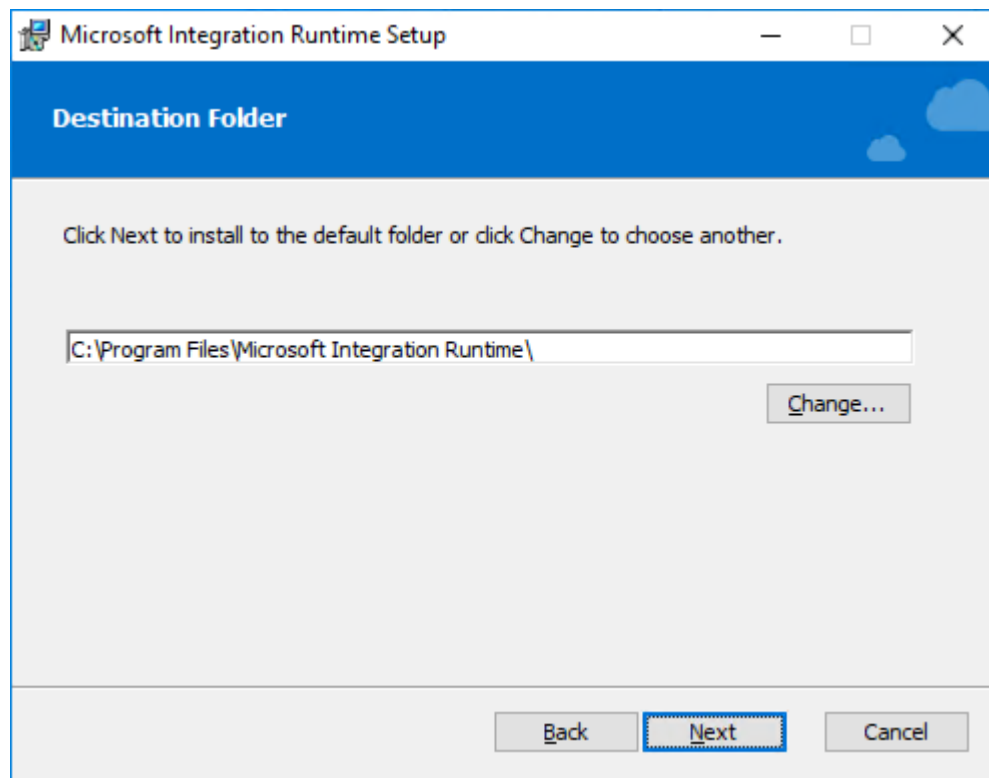
3. Run the installer once downloaded.
4. When you see the following screen, select Next.



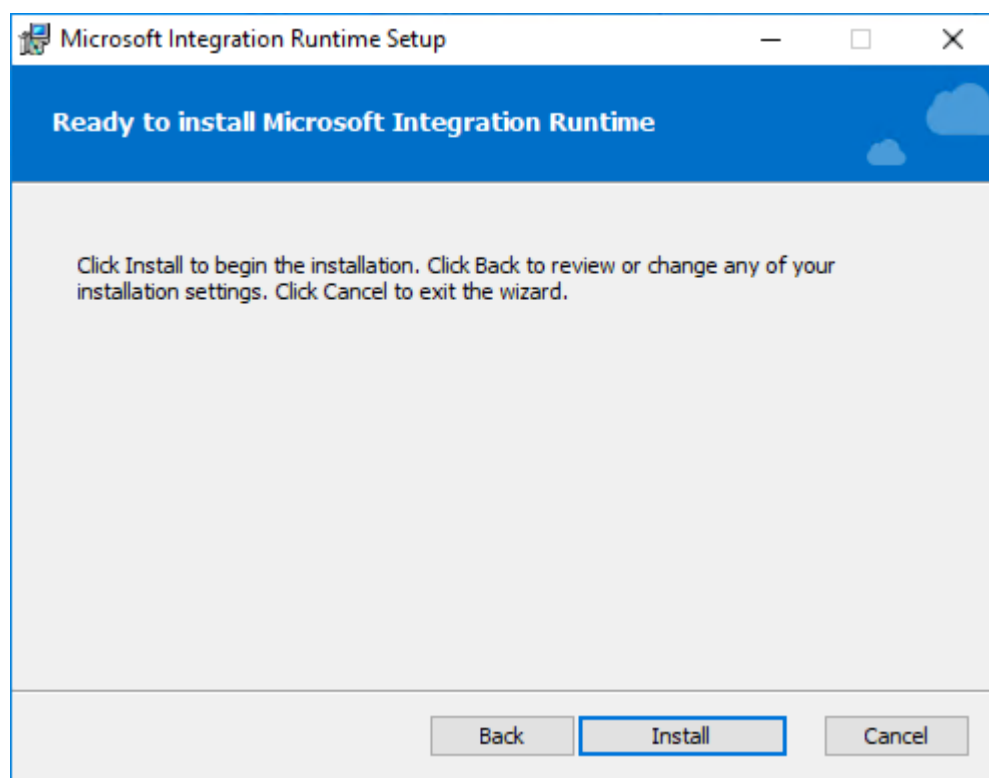
5. Check the box to accept the terms and select Next.



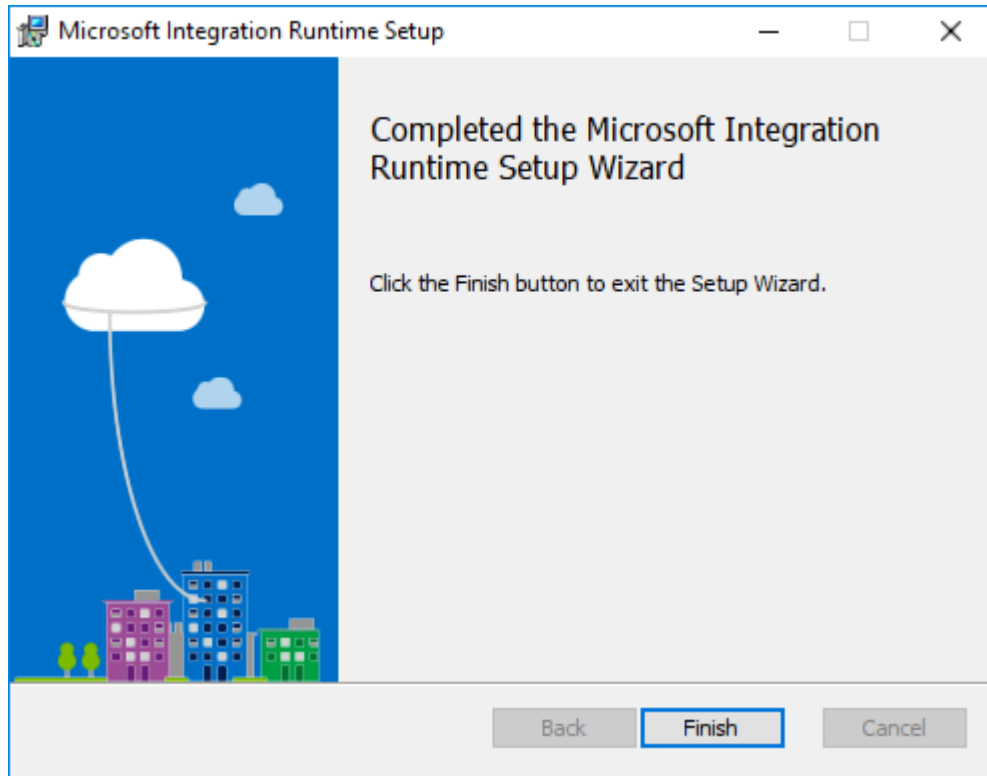
6. Accept the default Destination Folder and select Next.



7. Choose Install to complete the installation.



8. Select Finish once the installation has been completed.



9. After selecting Finish, the following screen will appear. Keep it open for now. You will come back to this screen once the Data Factory in Azure has been provisioned and obtain the gateway key to connect Data Factory to this "on-premises" server.

Microsoft Integration Runtime Configuration Manager

Register Integration Runtime (Self-hosted)

Welcome to Microsoft Integration Runtime Configuration Manager. Before you start, register your Integration Runtime (Self-hosted) node using a valid Authentication Key.

☐ Show Authentication Key [Learn how to find the Authentication Key](#)










HTTP Proxy

Current Proxy: No proxy Change


Register Cancel

Task 3: Configure Azure Data Factory

1. Launch a new browser window, and navigate to the Azure portal (<https://portal.azure.com>). Once prompted, log in with your Microsoft Azure credentials. If prompted, choose whether your account is an organization account or a Microsoft account. This will be based on which account was used to provision your Azure subscription used for this lab.
2. From the side menu in the Azure portal, choose **Resource groups**, then enter your resource group name into the filter box, and select it from the list.
3. Next, select your Azure Data Factory service from the list.

<input type="checkbox"/> Name ↑↓	Type ↑↓	Location ↑↓
<input type="checkbox"/>  adfmcwdatafactoryklf20211031	Data factory (V2)	West US 2
<input type="checkbox"/>  asastoremcwklf20211031	Storage account	West US 2
<input type="checkbox"/>  hands-on-lab-bigdata-vnet	Virtual network	West US 2
<input type="checkbox"/>  integrationruntime	Virtual machine	West US 2
<input type="checkbox"/>  integrationruntime-ip	Public IP address	West US 2
<input type="checkbox"/>  integrationruntime-nsg	Network security group	West US 2
<input type="checkbox"/>  integrationruntime812	Network interface	West US 2
<input type="checkbox"/>  integrationruntime_disk1_35bb5eef1bb944ba98c0d96055550d5e	Disk	West US 2
<input type="checkbox"/>  mcw-bdv-databricks-klf20211031	Azure Databricks Service	West US 2

4. On the Data Factory Overview screen, select **Open Azure Data Factory Studio**.

 Delete

Essentials

Resource group (change) : sai-big-data-viz

Type : Data factory (V2)

Status : Succeeded


Getting started : [Quick start](#)


Location : West US

Subscription (change) : [Client Development](#)

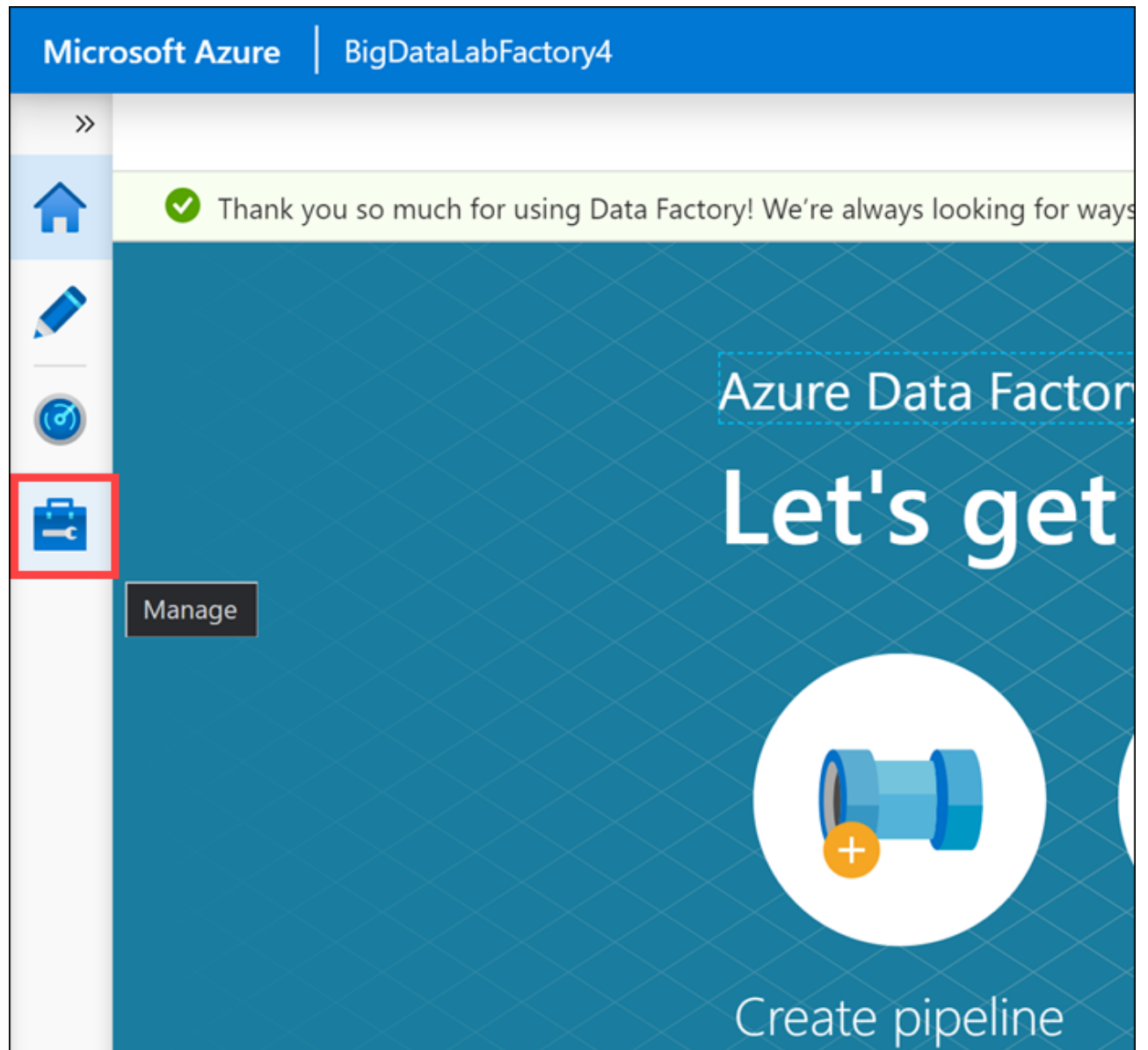
Subscription ID : e433f371-e5e9-4238-abc2-7c38aa596a18

Getting started


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](#)

5. A new page will open in another tab or new window. Within the Azure Data Factory site, select **Manage** on the menu.



6. Now, select **Integration runtimes (1)** in the menu, then select **+ New (2)**.

Microsoft Azure | bigdatalabdatafactory10

Data Factory | Validate all | Publish all | ARM template

Connections

- Linked services
- Integration runtimes** (1)
- Azure Purview (Preview)

Source control

- Git configuration
- ARM template
- Parameterization template

Author

- Triggers
- Global parameters

Integration runtimes

The integration runtime (IR) is the compute infrastructure to p

+ New (2) Refresh


Filter by name


Showing 1 - 1 of 1 items

Name ↑↓	Type ↑↓
AutoResolveIntegrationRuntime	Azure

7. In the Integration Runtime Setup blade that appears, select **Azure, Self-Hosted (1)**, then select **Continue (2)**.


Integration runtime setup

Integration Runtime is the native compute used to execute or dispatch activities. Choose what integration runtime to create based on required capabilities. [Learn more](#) 



Azure, Self-Hosted
Perform data flows, data movement and dispatch activities to external compute.

1



Azure-SSIS
Lift-and-shift existing SSIS packages to execute in Azure.

2

Continue


Cancel


8. Select **Self-Hosted (1)** then select **Continue (2)**.

Integration runtime setup

Network environment:

Choose the network environment of the data source / destination or external compute to which the integration runtime will connect to for data flows, data movement or dispatch activities:



**Azure**
Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.

**Self-Hosted**
Use this for running activities in an on-premise / private network
[View more](#) ▾

1

External Resources:

You can use an existing self-hosted integration runtime that exists in another resource. This way you can reuse your existing infrastructure where self-hosted integration runtime is setup.

**Linked Self-Hosted**
[Learn more](#) 

2

Continue

Back

Cancel

9. Enter a **Name (1)**, such as bigdatagateway-[initials], and select **Create (2)**.

Integration runtime setup

Private network support is realized by installing integration runtime to machines in the same on-premises network/VNET as the resource the integration runtime is connecting to. Follow below steps to register and install integration runtime on your self-hosted machines.



Name * ⓘ

bigdatagateway-drm

Description

Enter description here...

Type

Self-Hosted

Create Back Cancel

10. Under Option 2: Manual setup, copy the Key1 authentication key value by selecting the Copy button, then select **Close**.

Integration runtime setup

Settings Nodes Auto update Sharing

Install integration runtime on Windows machine or add further nodes using the Authentication Key.

Name ⓘ

bigdatagateway-drn





Option 1: Express setup

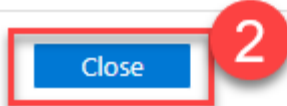
[Click here to launch the express setup for this computer](#)

Option 2: Manual setup

Step 1: [Download and install integration runtime](#)

Step 2: Use this key to register your integration runtime

Name	Authentication key
Key1	IR@0a24ae8a-c27c-414b-a3b7-9710e0a97ffc@bigdatalabdatafactory10€  
Key2	IR@0a24ae8a-c27c-414b-a3b7-9710e0a97ffc@bigdatalabdatafactory10€  




WARNING: Don't close the current screen or browser session.

11. Paste the **Key1 (1)** value into the box in the middle of the Microsoft Integration Runtime Configuration Manager screen.

Microsoft Integration Runtime Configuration Manager

Register Integration Runtime (Self-hosted)

Welcome to Microsoft Integration Runtime Configuration Manager. Before you start, register your Integration Runtime (Self-hosted) node using a valid Authentication Key.



☐ Show Authentication Key [Learn how to find the Authentication Key](#)

HTTP Proxy

Current Proxy: No proxy

2

12. Select **Register (2)**.

13. It can take up to a minute or two to register. If it takes more than a couple of minutes, and the screen does not respond or returns an error message, close the screen by selecting the **Cancel** button.

14. The following screen will be New Integration Runtime (Self-hosted) Node. Select **Finish**.

New Integration Runtime (Self-hosted) Node

Integration Runtime (Self-hosted) node name: ⓘ

integrationrunt



Below is the list of Integration Runtime (Self-hosted) Nodes:

bigdatagateway-drm
integrationrunt : Current New Node

☐ Enable remote access from intranet ⓘ

Finish

Cancel

15. Depending on your version of the Integration Runtime, you will be asked to specify a backup file. Select **Skip** and **OK**.

Specify backup file

Select the backup file for the Integration Runtime (Self-hosted) node: **bigdatagateway-skm** to restore or recover it. You can generate the backup file in the settings page of your previous Integration Runtime (Self-hosted) node. [Where can I find the backup file?](#)

Backup file (.irbackup/.dmgbbackup):

Browse


Password:


Skip

Import

Cancel

16. You will then get a screen with a confirmation message. Select the **Launch Configuration Manager** button to view the connection details.

 Microsoft Integration Runtime Configuration Manager ×



Register Integration Runtime (Self-hosted)

Welcome to Microsoft Integration Runtime Configuration Manager. Before you start, register your Integration Runtime (Self-hosted) node using a valid Authentication Key.

✓

☐ Show Authentication Key [Learn how to find the Authentication Key](#)

HTTP Proxy


Current Proxy: No proxy Change

✓

 Integration Runtime (Self-hosted) node has been registered successfully.

Note: You can associate up to 4 physical nodes with a Self-hosted Integration Runtime. This enables high availability and scalability for the Self-hosted Integration Runtime.
We recommend you setup at least 2 nodes for higher availability. [See Integration Runtime \(Self-hosted\) article for details.](#)

Launch Configuration Manager Close

 Microsoft Integration Runtime Configuration Manager — □ ×

Home Settings Diagnostics Update Help

✓

 Self-hosted node is connected to the cloud service

Data Factory:	bigdatalabdatafactory10
Integration Runtime:	bigdatagateway-drn
Node:	integrationrunt


Stop Service

Data Source Credential ⓘ

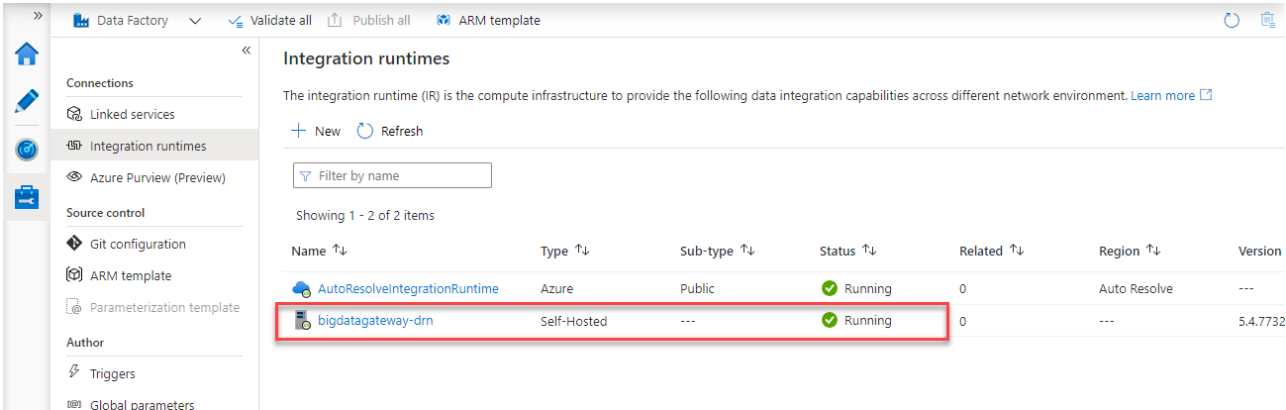
Credential store:	On-premises
Credential status:	In sync
Last backup time:	N/A

Generate Backup Import Backup

✓

 Connected to the cloud service (Data Factory V2) 

17. You can now return to the Azure Data Factory page and view the Integration Runtime you just configured. You may need to select **Refresh** to view the Running status for the IR.



18. Select the Azure Data Factory Overview button on the menu. Leave this open for the next exercise.

