

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 Data Factory
- Windows 10 Virtual Machine

Lab Objective:

After completing this lab, you will be able to:

- Create Integration Runtime (Azure & Self Hosted)

Exercise: Create Self-Hosted Integration Runtime and copy the data from virtual machine to storage account.

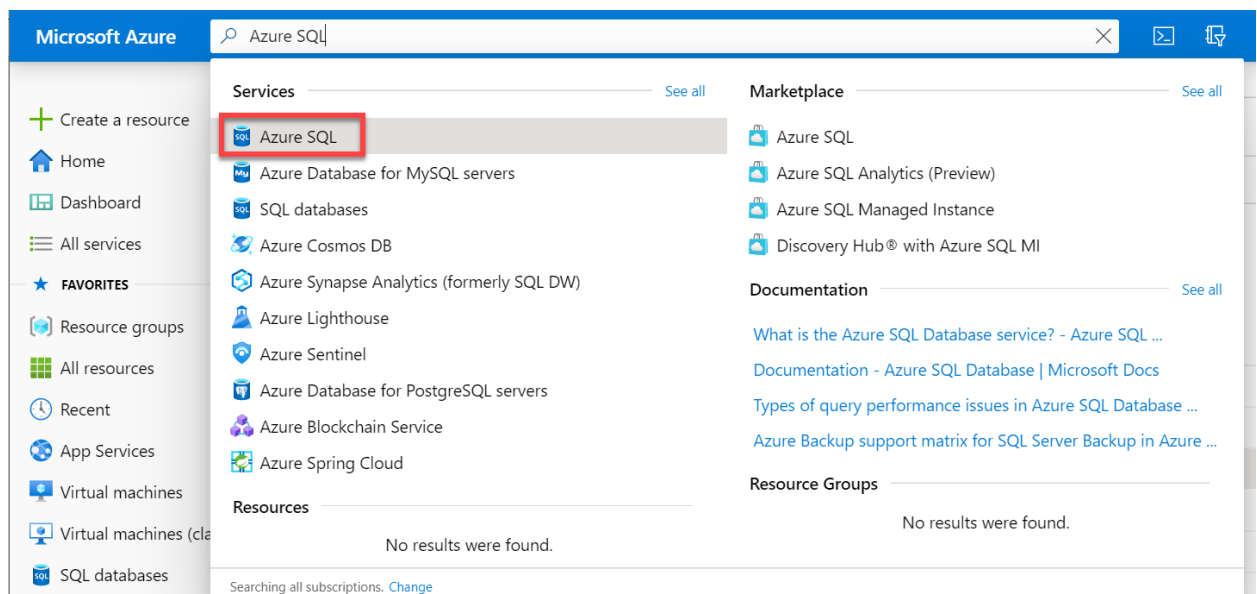
In this exercise you will be completing following task:

1. Provision a SQL Server on an Azure Virtual Machine
2. Connect to SQL Server on an Azure Virtual Machine
3. Create Self Hosted Integration Runtime in ADF
4. Copy data from SQL Server Database to Data Lake Storage

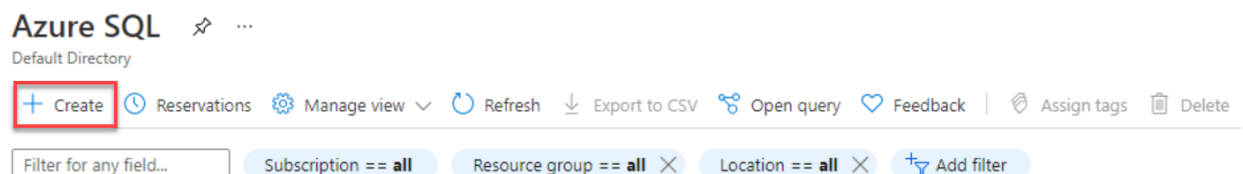
Task 1: Provision a SQL Server on an Azure Virtual Machine

Deploy a SQL Server on an Azure Virtual Machine

1. Locate the search bar at the top of the page. Search for **Azure SQL**. Select the search result for **Azure SQL** that appears in the results under **Services**.



2. On the **Azure SQL** blade, select **Create**.



3. On the **Select SQL deployment option** blade, click on the drop-down box under **SQL virtual machines**. Select the option labeled **Free SQL Server License: SQL 2019 Developer on Windows Server 2022**. Then select **Create**.

How do you plan to use the service?

SQL databases
Best for modern cloud applications. Hyperscale and serverless options are available.
Resource type: Single database
[Create](#) [Hide details](#)

SQL managed instances
Best for most migrations to the cloud. Lift-and-shift ready.
Resource type: Single instance
[Create](#) [Show details](#)

SQL virtual machines
Best for migrations and applications requiring OS-level access. Lift-and-shift ready.
Image: Free SQL Server License: SQL Server 2019 D...
[Please select an offer](#)

- SQL Server 2022 CTP on Windows Server 2022
- Free SQL Server License: SQL Server 2022 CTP Developer on Windows Server 2...
- SQL Server 2019 on Windows Server 2022**
- Free SQL Server License: SQL Server 2019 Developer on Windows Server 2022
- SQL Server 2019 Enterprise on Windows Server 2022
- SQL Server 2019 Standard on Windows Server 2022
- SQL Server 2019 Web on Windows Server 2022
- SQL Server 2019 on Windows Server 2019
- Free SQL Server License: SQL 2019 Developer on Windows Server 2019
- SQL Server 2019 Enterprise Windows Server 2019
- SQL Server 2019 Standard on Windows Server 2019
- SQL Server 2019 Web on Windows Server 2019
- (BYOL) SQL Server 2019 on Windows Server 2019
- (BYOL) SQL Server 2019 Enterprise Windows Server 2019

Single database
Single databases are a great fit for modern, cloud-born applications that need a fully managed database with predictable performance.
Featured capabilities:
✓ Hyperscale storage (up to 100TB)
✓ Serverless compute
✓ Easy management

Elastic pool
Elastic pools provide a cost-effective way of managing the performance of multiple databases with variable usage patterns.
Featured capabilities:
✓ Resource sharing for cost optimization
✓ Simplified performance management

4. On the **Create a virtual machine** page, enter the following information:
- **Subscription:** <Your subscription>
 - **Resource group:** SQL-VM-RG
 - **Virtual machine name:** sqlvm
 - **Region:** <your local region, same as the selected region for your resource group>
 - **Availability Options:** No infrastructure redundancy required
 - **Image:** Free SQL Server License: SQL 2019 Developer on Windows Server 2022 - Gen1
 - **Azure spot instance:** No (unchecked)
 - **Size:** Standard D2s_v3 (2 vCPUs, 8 GiB memory). You may need to select the "See all sizes" link to see this option)
 - **Administrator account username:** sqladmin
 - **Administrator account password:** pwd!DP300lab01 (or your own password that meets the criteria)
 - **Select inbound ports:** RDP (3389)
 - **Would you like to use an existing Windows Server license?:** No (unchecked)

Make note of the username and password for later use.

Project details

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

Subscription *	WWL Technical Content Development
Resource group *	(New) azureSQLServerVM_group

[Create new](#)

Instance details

Virtual machine name *	azureSQLServerVM
Region *	(US) East US
Availability options	No infrastructure redundancy required
Security type	Standard
Image *	Free SQL Server License: SQL Server 2019 Developer on Windows Server 2019
Run with Azure Spot discount	<input type="checkbox"/>
Size *	Standard_D2s_v3 - 2 vcpus, 8 GiB memory (\$137.24/month)

[See all images](#) | [Configure VM generation](#)
[See all sizes](#)

Administrator account

Username *	sqladmin
Password *
Confirm password *

5. Navigate to the **Management** tab and review the configuration.

Verify that **Enable auto_shutdown** is unchecked.

6. Navigate to the **SQL Server settings** tab and review the configuration.

Enable SQL Authentication

7. Select the **Review + create** button. Then select **Create**.
8. On the deployment blade, wait until the deployment is complete. The VM will take approximate 5-10 minutes to deploy. After the deployment is complete, select **Go to resource**.

Note: Your deployment may take several minutes to complete.

Task 2: Connect to SQL Server on an Azure Virtual Machine

1. On the **Overview** page for the virtual machine, select the **Connect** button and choose RDP.

Home > CreateVm-microsoftsqlserver.sql(2019-ws2019-sqldev-20220310081951 >

azureSQLServerVM Virtual machine

Search (Ctrl+/) << Connect Start Restart Stop Capture Delete Refresh Open in mobile CLI / PS Feedback

Overview

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

Settings

- Networking
- Connect
- Windows Admin Center (preview)
- Disks
- Size
- Security
- Advisor recommendations
- Extensions + applications
- Continuous delivery
- Availability + scaling

Connect

SSH **RDP** Bastion

SSH: [rglod22111985](#)

Operating system: Windows (Windows Server 2019 Datacenter)
Size: Standard DS2 v2 (2 vcpus, 7 GiB memory)
Public IP address: [104.44.142.211](#)
Virtual network/subnet: [contoso-rglod22111985-vnet/default](#)
DNS name: [Not configured](#)

Location: South Central US
Subscription (move): [GTL_DEV_CSR_2](#)
Subscription ID: b11bdf21-794f-487e-bb3c-1580be62e438
Tags (edit): [Click here to add tags](#)

Properties Monitoring Capabilities (8) Recommendations Tutorials

Virtual machine

Computer name	azureSQLServerV
Health state	-
Operating system	Windows (Windows Server 2019 Datacenter)
Publisher	microsoftsqlserver
Offer	sql2019-ws2019
Plan	sqldev-gen2
VM generation	V2
Agent status	Ready
Agent version	2.7.41491.1044
Host name	None

Networking

Public IP address	104.44.142.211
Public IP address (IPv6)	-
Private IP address	10.0.0.4
Private IP address (IPv6)	-
Virtual network/subnet	contoso-rglod22111985-vnet/default
DNS name	Configure

Size

Size	Standard DS2 v2
vCPUS	2

2. On the RDP tab, select the **Download RDP File** button.

RDP SSH Bastion

Connect with RDP

^ Suggested method for connecting

Azure has checked the status for the most common prerequisites when connecting using this method.

- Checking network security group for inbound access from your client's IP address. [Learn more](#)
- ✓ The VM's network interface has a Public IP address. [Learn more](#)
- ✓ The VM is running.

To connect to your virtual machine via RDP, select an IP address, optionally change the port number, and download the RDP file.

IP address *

Public IP address ()

Port number *

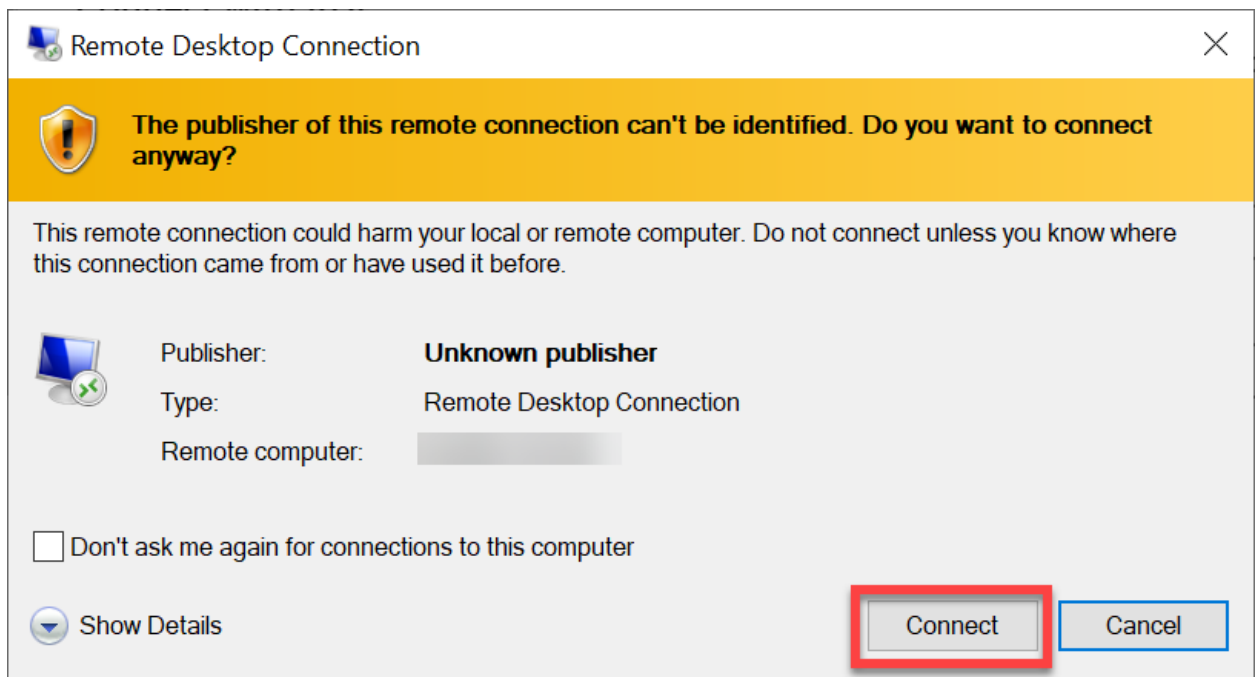
3389

Download RDP File

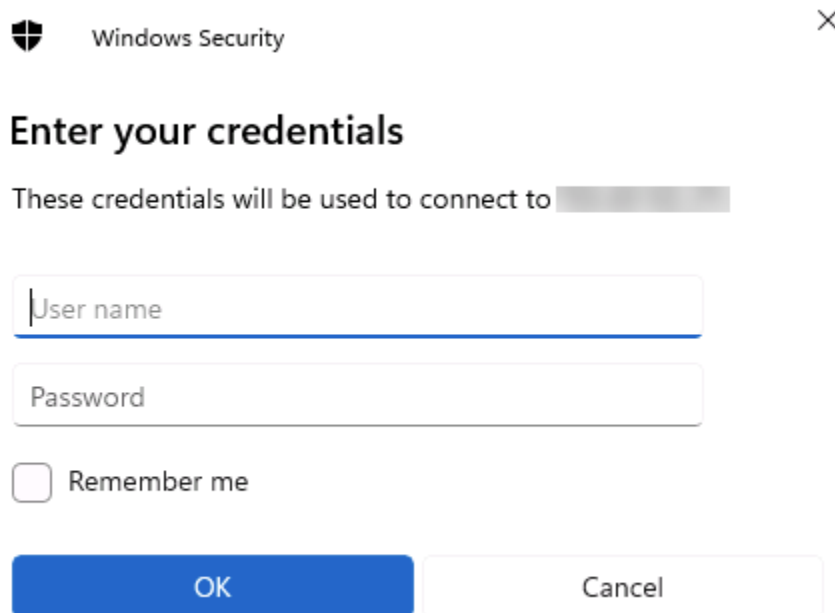
Can't connect?

- Test your connection
- Troubleshoot RDP connectivity issues

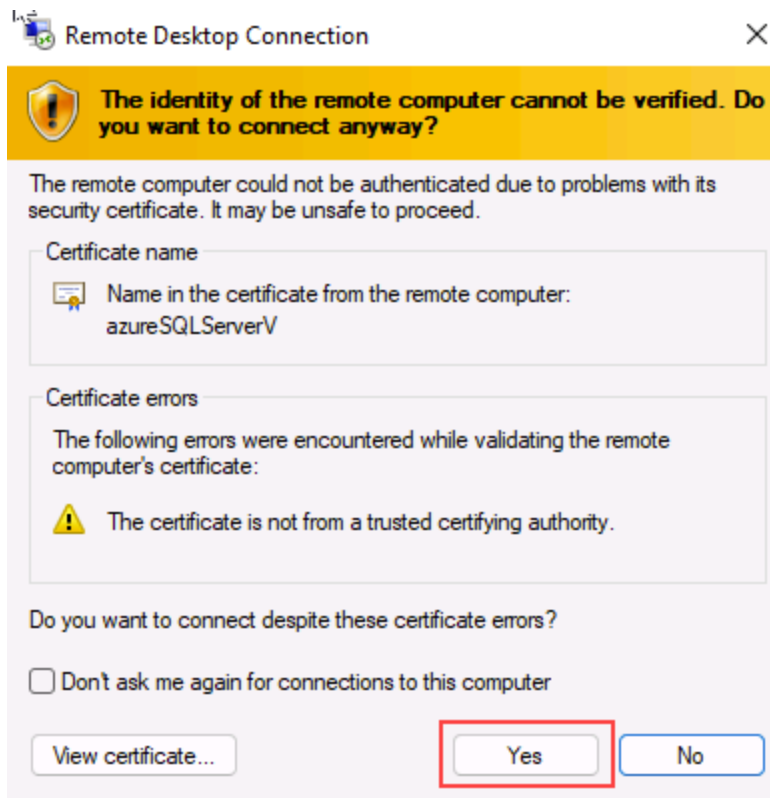
3. Open the RDP file that was just downloaded. When a dialog appears asking if you want to connect, select **Connect**.



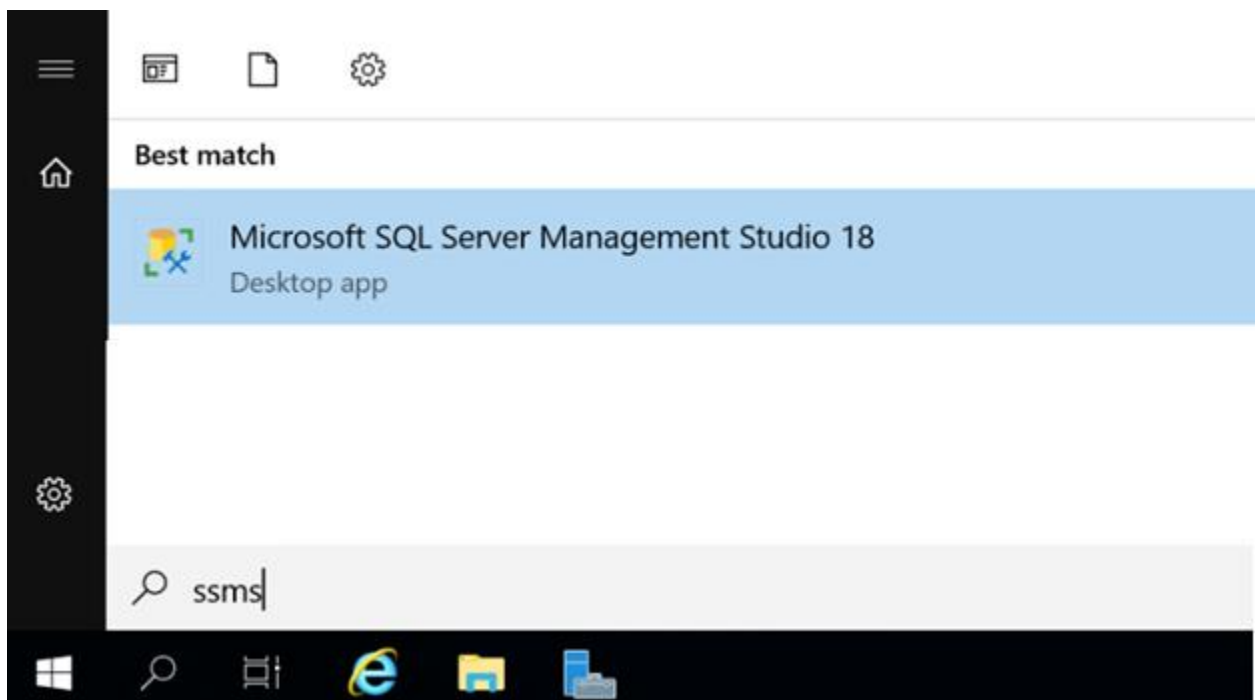
4. Enter the username and password selected during the virtual machine provisioning process. Then select **OK**.



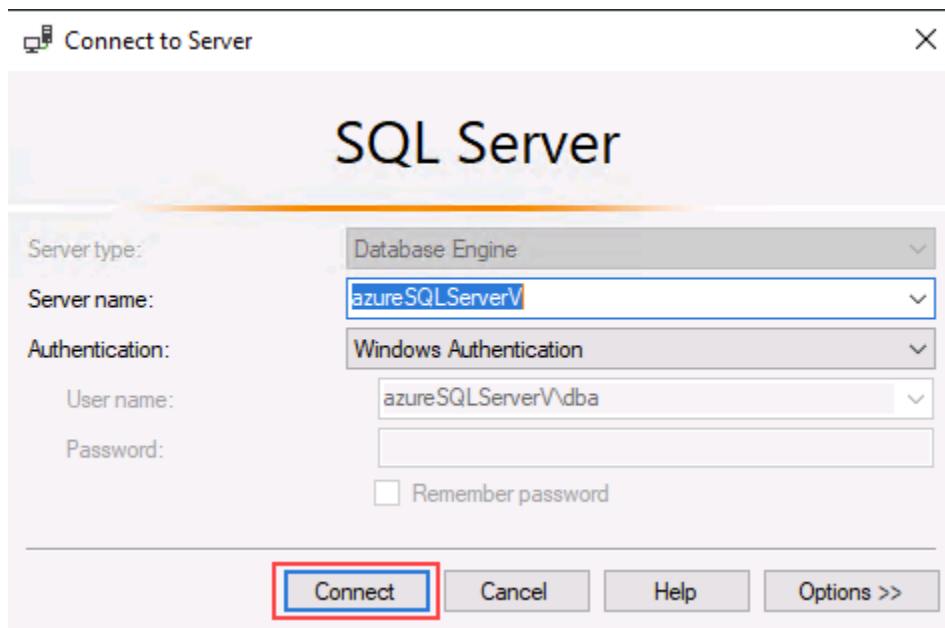
5. When the **Remote Desktop Connection** dialog appears asking if you want to connect, select **Yes**.



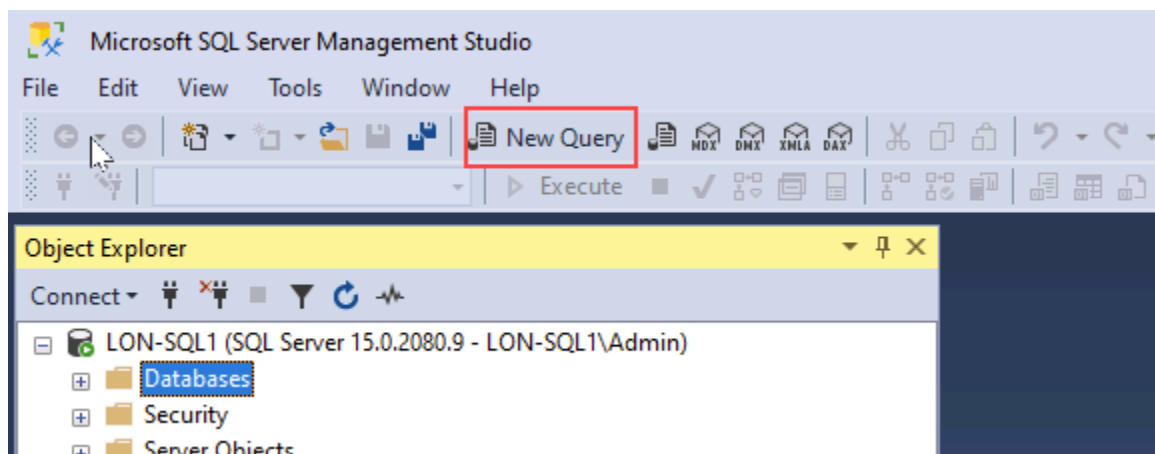
6. Select the Windows Start button and type SSMS. Select **Microsoft SQL Server Management Studio 18** from the list.



- When SSMS opens, notice that the **Connect to Server** dialog will be pre-populated with the default instance name. Select **Connect**.



- Select the **Databases** folder, and then **New Query**.



- Copy the AdventureWorks2017 file from the URL:

<https://github.com/MicrosoftLearning/dp-300-database-administrator/blob/master/Instructions/Templates/AdventureWorks2017.bak>

and store it at C:\LabFiles Location. (Create Directory LabFiles in c: drive)

10. In the new query window, copy and paste the below T-SQL into it. Execute the query to restore the database.

```
RESTORE DATABASE AdventureWorks2017
FROM DISK = 'C:\LabFiles\AdventureWorks2017.bak'
WITH RECOVERY,
    MOVE 'AdventureWorks2017'
    TO 'C:\LabFiles\AdventureWorks2017.mdf',
    MOVE 'AdventureWorks2017_log'
    TO 'C:\LabFiles\AdventureWorks2017_log.ldf';
```

Note: The database backup file name and path should match with what you've downloaded on step 1, otherwise the command will fail.

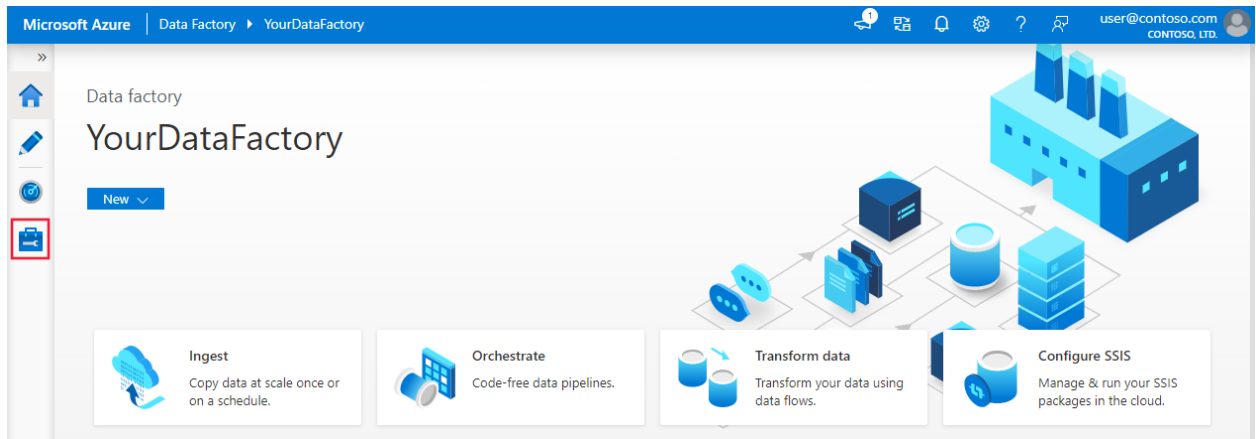
11. You should see a successful message after the restore is complete.

```
Database 'AdventureWorks2017' running the upgrade step from version 899 to version 900.
Database 'AdventureWorks2017' running the upgrade step from version 900 to version 901.
Database 'AdventureWorks2017' running the upgrade step from version 901 to version 902.
Database 'AdventureWorks2017' running the upgrade step from version 902 to version 903.
Database 'AdventureWorks2017' running the upgrade step from version 903 to version 904.
RESTORE DATABASE successfully processed 26283 pages in 0.619 seconds (331.712 MB/sec).
Completion time: 2022-04-13T12:30:50.0058660-07:00
```

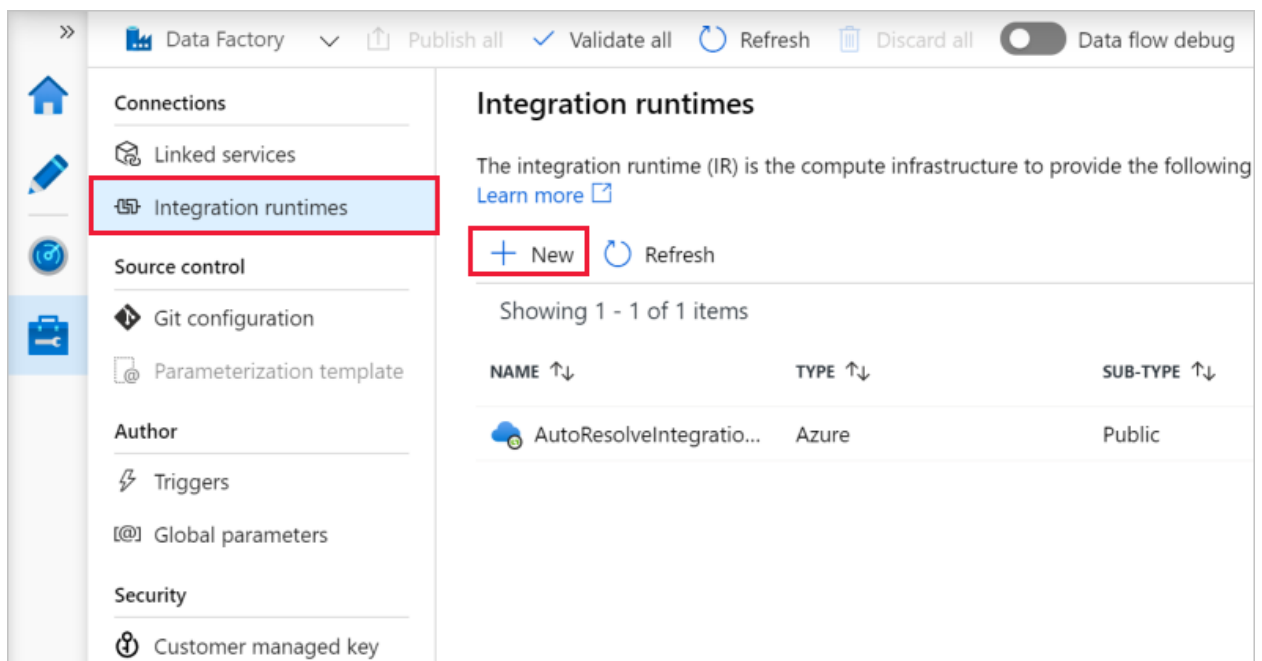
Task 3: Create Self Hosted Integration Runtime in ADF

Use the following steps to create a self-hosted IR using the Azure Data Factory UI.

1. On the home page of the Azure Data Factory UI, select the [Manage tab](#) from the leftmost pane.



2. Select **Integration runtimes** on the left pane, and then select **+New**.



3. On the **Integration runtime setup** page, select **Azure, Self-Hosted**, and then select **Continue**.
4. On the following page, select **Self-Hosted** to create a Self-Hosted IR, and then Select **Continue**.

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](#) ▾

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](#) 

Continue

Back

Cancel

5. Enter a name for your IR (SH-IR), and select **Create**.
6. On the **Integration runtime setup** page, select **Option 2** to set up manually.
7. The following instructions are based on manual setup:

Integration runtime setup

SettingsNodesAuto updateSharing

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

Name i

MySelfHostedIR





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	<div></div> <div></div>
Key2	<div></div> <div></div>

Close

- a. Copy and paste the authentication key. Select **Download and install integration runtime**.
- b. Download the self-hosted integration runtime on a **SQL Server Virtual Machine**. Run the installer.

Or Copy this URL and download in the SQL Server VM:

<https://www.microsoft.com/en-us/download/details.aspx?id=39717>

Choose the download you want

<input type="checkbox"/> File Name	Size
<input checked="" type="checkbox"/> IntegrationRuntime_5.24.8369.1.msi	1,001.4 MB
<input type="checkbox"/> IntegrationRuntime_5.23.8355.1.msi	978.5 MB
<input type="checkbox"/> Release Notes.doc	251 KB

- c. On the **Register Integration Runtime (Self-hosted)** page, paste the key you saved earlier, and select **Register**.

Microsoft Integration Runtime Configuration Manager

Register Integration Runtime (Self-hosted)

1 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

Diagnostic Tool

Troubleshoot problems (preview)

2 Register Cancel

- d. On the **New Integration Runtime (Self-hosted) Node** page, select **Finish**.

8. After the self-hosted integration runtime is registered successfully, you see the following window:

Microsoft Azure

Data Factory ▶ adf24x

Search

»

Data Factory

Validate all

Publish all

«

General

Factory settings

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration capabilities across d

+

New

↺

Refresh

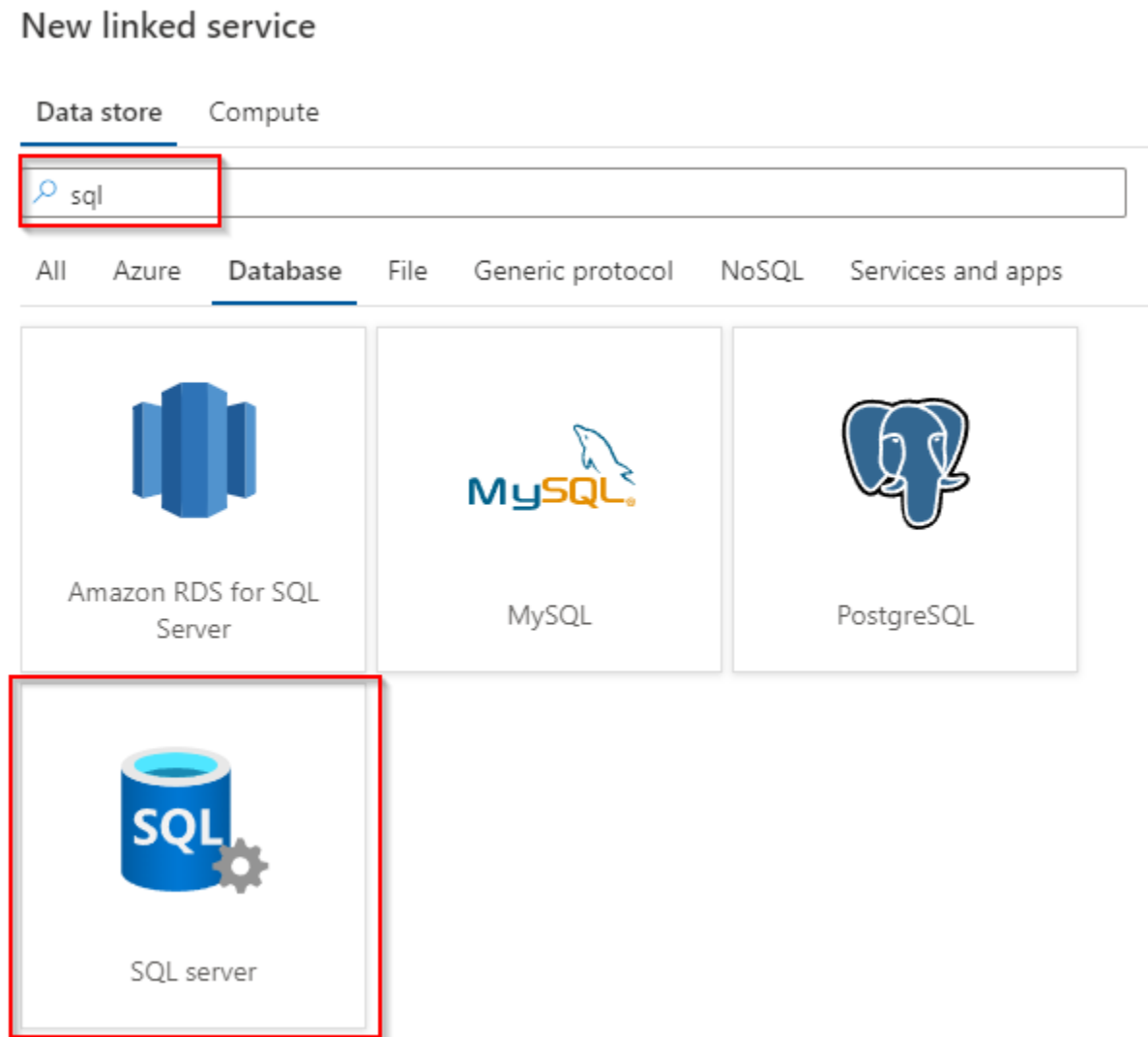
Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status
AutoResolveIntegrationR...	Azure	Public	Running
SH-IR	Self-Hosted	---	Running

Task 4: Copy data from SQL Server Database Table to Data Lake Storage


1. On the home page of the Azure Data Factory UI, select the [Manage tab](#) from the leftmost pane and select **Linked Service**.
2. Click + New to create a linked services and search sql to select the SQL Server and click continue.



3. Fill the following details
 - a. Name: SqlServerOnPremises
 - b. Connect via integration runtime: SH-IR
 - c. Server name: sqlvm
 - d. Database name: AdventureWorks2017

- e. Authentication Type: SQL Authentication
- f. Username: sqladmin
- g. Password: pwd!DP300lab01 (or your own password that you setup at the time of VM creation)

New linked service


 SQL server [Learn more](#) 


Name *

SqlServerOnPremisis

Description

Connect via integration runtime * 

 SH-IR

 The credentials are stored in the machines of self-hosted integration runtime if you don't choose to store them in Azure Key Vault.

Connection string

Azure Key Vault

Server name *

sqlvm

Database name *

AdventureWorks2017

Authentication type

SQL authentication

User name *

sqladmin

Password

Azure Key Vault

Password *

Create

Back



Connection successful

Test connection

Test connection

Cancel

4. Create a new pipeline, Author Tab -> + -> Pipeline

5. Drag copy activity from Activities canvas under move and transform section to pipeline surface area.
6. In source tab of copy activity create new dataset as shown.

Name: SqlServerTable

Linked service: SqlServerOnPremises

Table Name: Sales.SalesPerson

Set properties

Name

SqlServerTable

Linked service *

SqlServerOnPremis

Connect via integration runtime * ⓘ

✓ SH-IR

Table name

Sales.SalesPerson

Edit

Import schema

☒ From connection/store ☐ None

> Advanced

7. Switch to Sink tab and create a new dataset for data lake gen 2 storage account and follow the below steps:
 - a. Select format: DelimitedText
 - b. Set Properties

Name: storage_person

Linked Service: AzureDataLakeStorage

File Path: data/person/person.csv



First row as header: checked

Import Schema: None

Set properties

1 Name

Linked service * **2**

3 File path / /  | 

First row as header **4** ☒

Import schema ☐ From connection/store ☐ From sample file **5** ☒ None




> Advanced

8. Click Publish all.

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (3)

NAME	CHANGE	EXISTING
▼ Pipelines		
 pipeline2	(New)	-
▼ Datasets		
 SqlServerTable	(New)	-
 storage_person	(New)	-

9. Run the pipeline. After execution finish verify the file in the storage account.