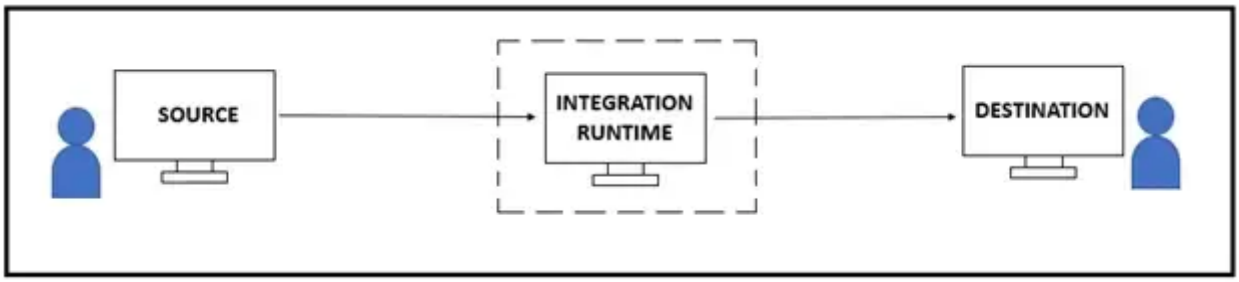
Amarpreet Singh Follow
Jun 8, 2020 · 5 min read · Listen

Self-Hosted Integration Runtime Setup



Integration Runtime

Integration Runtime(IR) is the compute infrastructure i.e. the pathway provided by Azure Data Factory to manage movement of data across the various on-premises systems (private virtual networks), as well as different cloud sources (public networks) in Azure environment(such as Azure Blobs, Data Lake, SQL DB/DWH, Cosmos DB etc.).

The Integration runtime are used in various pipeline activities such as Copy, Lookup, Get Metadata, Delete etc. As seen in above image, Integration Runtime is primarily installed within a Virtual network in same or different location based on the availability, it handles the integration between the different source and destination systems & services.

There are 3 types of Integration Runtime:

1. Azure IR

- Azure IR is a default Integration Runtime, generally this instance is known as “AutoResolveIntegrationRuntime” that is created by default, during creation of Azure Data factory. One can also setup this IR if default option

region.

- Basically, one need not worry about the infrastructure provisioning, Azure takes care of finding the IR for enabling high performance data transfer ensuring the safety and reliability.
- This IR can only be provisioned in Public network means this can be utilized, if one has to transfer data between Public network and Azure environment.

2. Self — Hosted IR

- Self-hosted IR can be provisioned for data transfer in both Public and Private network.
- If you have a use case, where you need to transfer data from on premises systems and from the public cloud sources, this IR suits best if you want to perform data integration securely in a private virtual network.
- One has to setup the Self-Hosted IR in order to use as it is not available by default. The steps for installation are given below.
- It is advisable to setup this IR on private networks (such as VMs) in the same location of your data factory and other resources, in order to ensure the smooth data transfer, keeping it in different location (until there is no other option available for particular resource) can experience high latency and network issues.

3. Azure — SSIS IR

- Like Self-hosted, Azure-SSIS IR can also be provisioned in both Private and Public network.
- Azure-SSIS IR is a fully managed instance dedicated to run you SSIS packages. One has the flexibility to scale compute capabilities up/down by specifying the number of nodes in the cluster.
- For migration of data from on-premises environment, one must join the Azure-SSIS IR to the Virtual Network that is connected to on-premises network

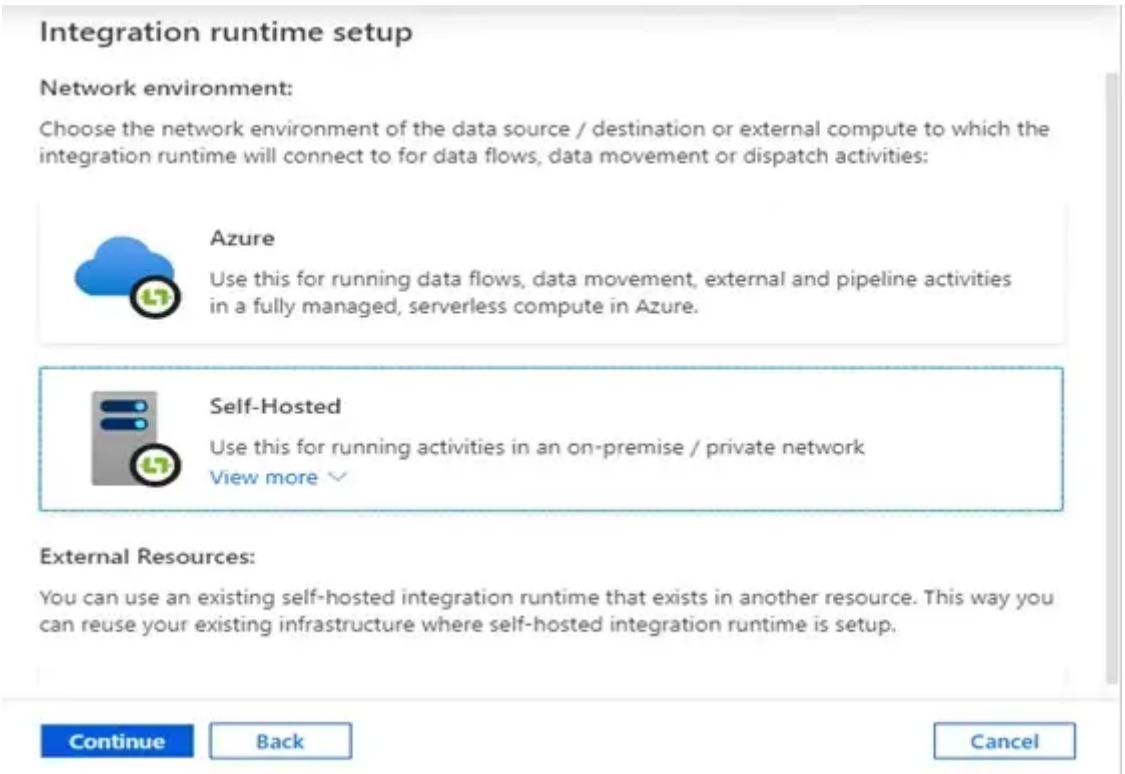

order to ensure high performance.

Following are the steps to setup self-hosted Integration Runtime:

Step1: Go to Data Factory instance in azure portal and open ***connections.*** Choose ***Integration Runtime.*** A new window containing an Azure IR will appear. Select ***"+ New"*** option.

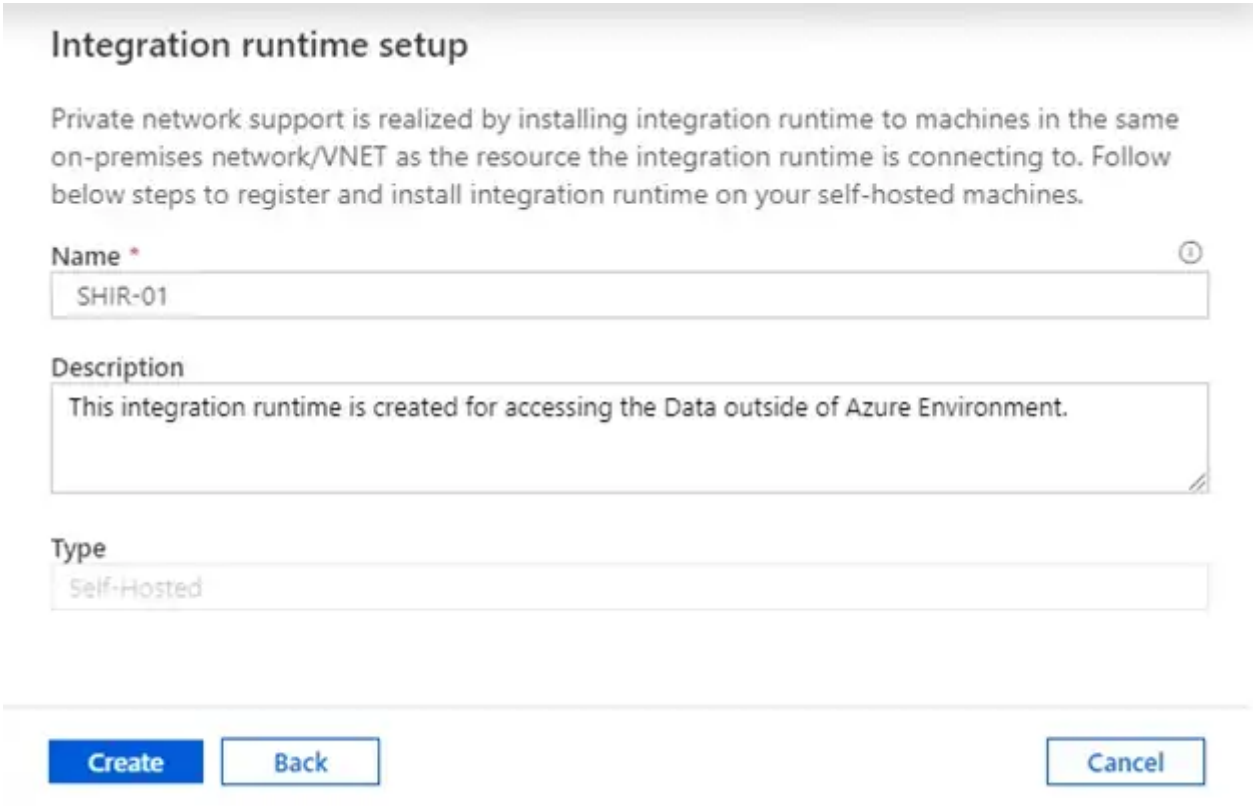
Connection window in Azure Data Factory

Choose ***Self-Hosted Integration Runtime*** option and click ***continue.***



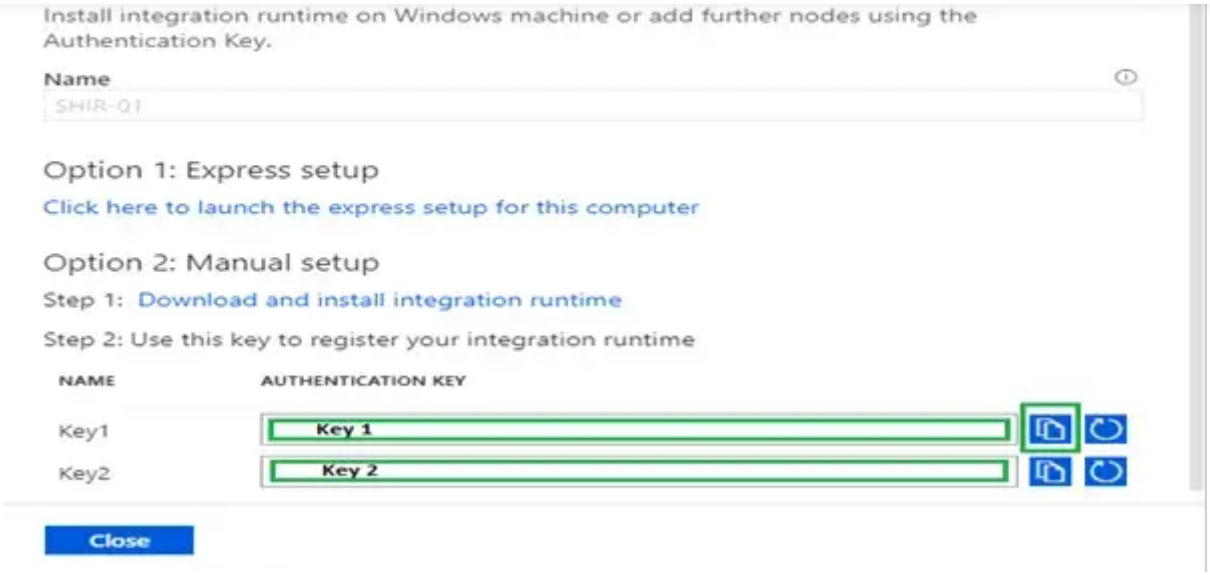
Integration runtime setup window

> *Note: Give a meaningful name to IR, as it would be used in many of the activities while setting up the data pipelines in Azure Data Factory.*



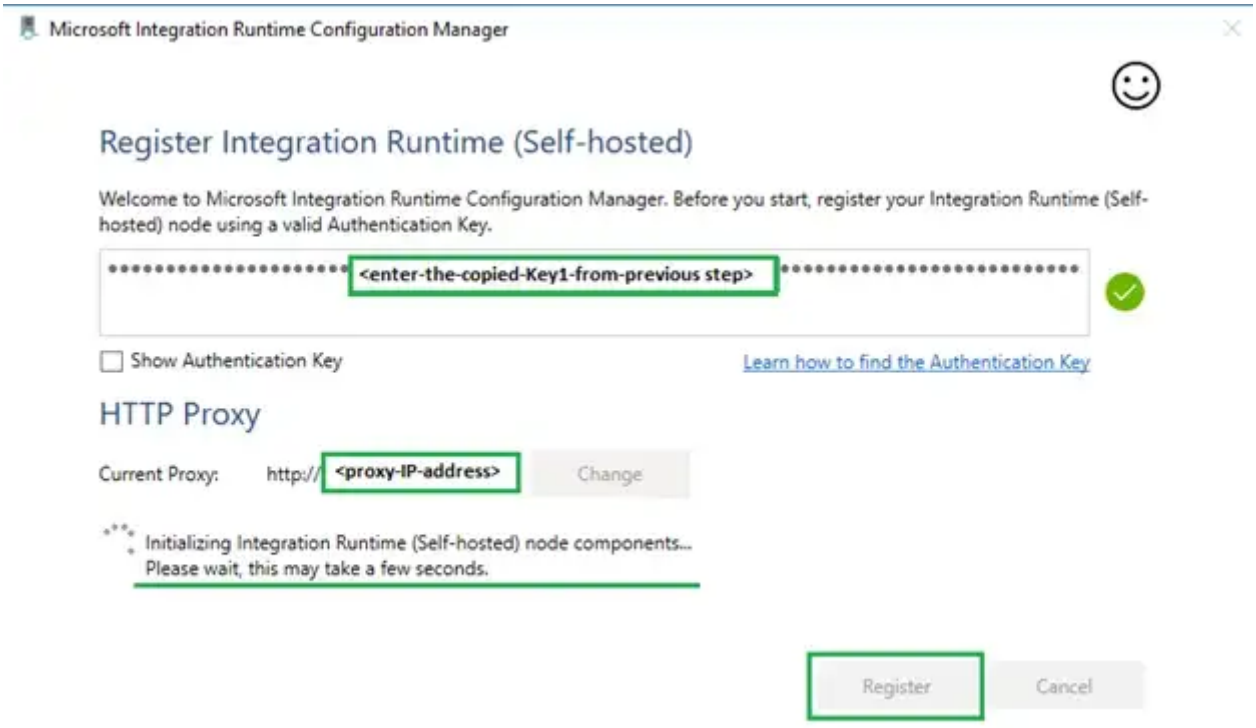
Enter Name and Description for Self-Hosted IR

Select the option 2**, Manual setup.** Firstly*, download and install integration runtime* on your virtual private network (like Virtual Machine) and secondly, copy <***Key1***> to clipboard for further use. Click ***Close.***



Option 2: Manual setup of Integration Runtime

Step 3: Enter the <*key 1*> copied ir shown in image below. Click ***Register***, it may take few seconds t etup on your private virtual network.



Register Integration Runtime (Self-hosted)

After few seconds you will receive the confirmation of setup completed successfully in same window.

up credentials to a file. You can use this file to restore or recover the integration runtime (Self-hosted) in case of a failure.
See Integration Runtime (Self-hosted) article for details.

Launch Configuration Manager Close

Self-Hosted IR Registered successfully

Step 4: Post the setup is complete, switch back to Azure portal to see the "***SHIR-01***" installed in the Integration runtime window.

SHIR-01 in Integration runtime window

This completes the setup of the Self-Hosted Integration Runtime.

> ***#Important Points to be noted while working with Parquet format:***
>
> *1. It is necessary to install* ***64-bit (minimum requirement for IR) JRE 8*** *(Java Runtime Environment) or* ***OpenJDK*** *and* ***Microsoft Visual C++ 2010 Re-distributable*** *Package on your IR machine, as it is not available by default*
>
> *2. After installing JRE, the path must be setup in System environment variable "****JAVA_HOME****" accordingly.*
>
> *3. If you face error such as "****java.lang.OutOfMemoryError****:Java heap space", you need to add "****_JAVA_OPTIONS****" in your IR machine to adjust the min/max heap size for JVM with value* ***-Xms256m -Xmx16g****.*
>
> *4. The flag* ***Xms*** *specifies the initial memory allocation pool for a Java Virtual Machine (JVM), while* ***Xmx*** *specifies the maximum memory allocation. This means that JVM will*

I hope this article was useful! ✌ Please post your comments if any questions on above setup and do extend your appreciation with a clap and share, if you feel this article was helpful.

I would like to take a moment, to thank everyone for the appreciation you all have showed on my previous articles. See you all with more articles on Azure. 😊