

Lab 2 - Integration Runtimes

Requirements

To be able to start the lab, it is important that Lab1 has been completed.

Objective

Azure Data Factory uses an *Integration Runtime (IR)*. In fact, this is where your pipeline is "executed": for example, in a specific Azure region or on-premises.

During this lab, you will create several types of IRs. You will also configure, install, and "link" them. Follow the assignments step by step.

Assignment 1 - Creating Azure Integration Runtime

Within ADF, you always get an Integration Runtime "for free": a so-called "Auto Resolve" IR. This means that Azure globally checks which place would be convenient to execute your pipeline and picks one for you. This can be helpful to make the first steps, but it is advisable to be more explicit about where your data is processed. Partly from performance and cost considerations, but also from a security and compliance perspective (knowing where your data is processed).

In this first assignment, we are going to create an Integration Runtime where you specify the location yourself. In your environment, you have two Data Factories, which will later both use the one IR we create.

1. Go to the [Azure portal](#). Click on the Data Factory that does **not** contain the word Linked. A new page will load with a **Launch Studio** button, click on it. You are now directed to the page of the ADF.
2. Click on the **toolbox** (Manage) on the left. Then click on **Integration runtimes** on the left. You will see that an **AutoResolveIntegrationRuntime** already exists. This is the standard IR within Azure that you can use to move data between different services like SQL Database, Azure Data Lake Store, etc.
3. Click on **New**, click on **Azure, Self-Hosted**, click on **continue**, then click on **Azure** and **continue** again. You can now create a new **Azure Integration Runtime**, where you can also specifically indicate the **Region**. This way, you can be sure that the data always remains within this region.
4. Give the IR a name and choose the **Region** where the rest of your resources are located. For naming, it is handy to start with **IR-** followed by what kind of IR it is, the -region, and the -environment.
 - Practical example: **IR-AzureResolve-EastUS-TST**
 - Training example: **IR-AzureResolve-WestEurope-Training**
5. Click on **Create**

Your IR is now being created and should be ready within a few seconds.

Assignment 2 - Self-Hosted Integration Runtime

Data from an on-premises environment or a shielded network is not directly accessible for Azure Data Factory. With a *self-hosted* IR, you can also connect these environments: you then install a *self-hosted Integration Runtime* on a VM in your shielded network (e.g. on-premises). This then establishes the connection with Azure Data Factory.

1. Go back to your Virtual Machine if you have it open, otherwise, you follow the steps from Lab1, assignment 2 again.
2. Go to the Azure portal (portal.azure.com) and log in with the account with which you perform the training.
3. Go back to the ADF (without linked in the name) via the resource group as explained in assignment 1.
4. Once you arrive at the screen of the IRs, click on **New**, click on **Azure, Self-Hosted**, click on **continue**, then click on **Self-Hosted** and **continue** again.

A Self-Hosted IR usually runs on an on-premises Virtual Machine with ports 443 and 1433 open so it can access Azure for both the Database (1433) and a storage account (443). The minimum requirements to run a Self-Hosted IR properly are:

- 4 CPU cores of 2GHz or faster
- 8GB RAM
- 80GB disk space

In practice, it turns out that 16GB RAM and an SSD can help to move the data better, where only the CPU could be the limiting factor.

The Integration Runtime can also be downloaded directly from <https://www.microsoft.com/en-us/download/details.aspx?id=39717>

5. You only need to give the IR a name. The region depends on where the Virtual Machine is located and click on **Create**.
6. When the IR has been created, you immediately get a screen with 2 options for installation. Since you are already on the Virtual Machine, choose Option 1 (Express setup). Click on **Click here to launch the express setup for this computer**. And then start the installation by clicking on the downloaded file. This can be done via the popup in the upper right corner or through **File Explorer** (The yellow folder) to go to **Downloads**.

If you can't access the Azure portal of your organization from the VM, **Option 2** is an option: you then manually download the IR, and use one of the keys shown during the installation.

7. When you start the installation, everything else happens automatically, wait patiently for the installation.
8. When the installation is complete, click on **Close**. In the ADF, you should now see a **Green circle** next to the Self-Hosted IR. If this is not the case, it should be after a **refresh**.

You have now created a *self-hosted* IR, allowing Azure Data Factory to move data from and to an on-premises environment.

Assignment 3 - Linking Integration Runtimes

If you use multiple Data Factories, you can choose to "link" an existing IR in multiple Data Factories. This prevents you from having to install an IR multiple times.

1. Click on the **Self-Hosted IR** you just created. Go to the **Sharing** tab and copy the resourceID by clicking on the **blue box** or selecting and copying it.
2. Click on the **Blue plus sign** with the text **Grant permissions to another Data Factory or user-assigned managed identity**
3. The adf-linked should automatically be among them, check it and then click on **Add**
4. Close the tab by clicking on **Apply**.
5. In the top right corner, you see a row of icons. Click on the 2nd from the left, the icon with **the 2 screens and arrows** (Switch Data Factory).
6. A new screen will appear, and most of it will be filled in beforehand. Choose the adf-linked at **Data Factory Name** and then click on **OK**.
7. The other ADF is now being loaded. Once this is completed, click on the **toolbox** (manage) on the left and then on **Integration runtimes** again.
8. Once you arrive at the screen of the IRs, click on **New**, click on **Azure, Self-Hosted**, click on **continue**, then click on **Linked Self-Hosted** and **continue** again.
9. Paste the copied resourceID

into the **Resource ID** field and give the IR a correct name.

10. Click on **Create** and the linked Self-Hosted IR should appear with a **Green circle**.

While linking self-hosted IRs is very useful, it's not always desirable that, for example, a development environment can extract data via the same IR as a production environment. It often happens that a Development and Test environment use the same Self-Hosted IR, just like for Acceptance and Production if there is a specific case for it.

The linking for a Self-Hosted IR is often used when there are multiple teams using the same data source. Since you only need one IR (and thus Virtual Machine), you save costs and resources.

Table of Contents

1. [Preparing the Azure environment](#)
2. [Integration Runtimes](#)
3. [Linked Services](#)
4. [Datasets](#)
5. [Pipelines](#)
6. [Triggers](#)
7. [Global Parameters](#)
8. [Activities](#)
9. [Batching and DIUs](#)