Integration Runtime is the Compute infrastructure used by Data Factory.

There are three types of integration runtimes.

Azure integration runtime,

Self-hosted integration runtime and,

Azure SSIS integration runtime

First, let's talk about Azure and the Self-hosted integration runtime.

Azure integration runtime can copy data between cloud data stores, and it can dispatch the activity to a variety of Compute services such as Azure HDInsight or SQL Server where the transformation takes place.

The Self-hosted Integration Runtime is a software with essentially the same code as the Azure IR, but you install it on an on-premises machine or a virtual machine (VM) in a virtual network. A self-hosted IR can run copy activities between a public cloud data store and a data store in a private network, and it can dispatch transform activities against compute resources in a private network.

But you might ask why Self-Hosted IR? Why can’t I run all activities on an Azure IR? The most popular answer is that Data Factory will not be able to directly access on-premises data sources, as they sit behind a firewall. It’s sometimes possible to establish a direct connection between Azure and on-premises data sources by configuring your firewall in a specific way—& if you do, you don’t need to use the Self-hosted IR—but this is something that admins aren’t always happy to do.

Let's take a scenario to understand how integration runtime works with data factory to orchestrate data movement.

So I have a Data Factory pipeline, with two activities. The first is a Copy activity to move the data from SQL server to Azure data Lake Store.

And the second activity processes the data with Data Lake analytics and stores it in the Data Lake Store.

So you have the on-prem environment with SQL Server, sitting behind a Firewall

And the Data Lake Store within the Azure cloud.

When the first activity begins, the Data Factory directs the Self-hosted IR to copy the data from SQL server to data Lake Store.

If you are wondering where this Self-hosted IR is defined, it is defined in the Linked service that connects to the SQL server. So in this case, the self-hosted integration runtime performs the copy activity.

Once the data is copied, the second Activity begins and Data Factory sends a command to the Azure integration runtime, which then dispatches the activity to Data Lake Analytics to transform the data and store in data lake store.

Note that your Data Factory instance and the Integration Runtime can be in different regions. The integration runtime is generally closer to the data source.

With Azure SSIS IR, you can natively execute SSIS packages in a managed environment. So, When we lift and shift the SSIS packages to Data Factory, we use Azure SSIS IR.

So, first, you create Azure SSIS IR which creates Integration Services Catalog in Azure SQL DB for your SSIS Packages to live. When the Pipeline activity begins, Data Factory sends a command to the Azure SSIS IR, which then executes the SSIS Packages