My Notes on this section

Exam Perspective

Data Factory Integration runtime is important topic, you should know difference between diff runtime, and in which scenario which runtime should be used.

Self-hosted runtime is more important.

My Notes

Integration Runtime
The integration runtime is the execution environment that provides the compute infrastructure for Data Factory.

Further study: Integration runtime in Azure Data Factory

Self-hosted runtime
When you use the Copy activity to copy data between Azure and a private network, you must use the self-hosted integration runtime.

Further study: Create and configure a self-hosted integration runtime
Azure integration runtime.
This is required when you need to copy data between Azure and public cloud services.

Further study: How to create and configure Azure Integration Runtime
Azure-SSIS integration runtime.
This is required when you want to run existing SSIS packages natively.

Further study: Create Azure-SSIS Integration Runtime in Azure Data Factory

Linked Service
A linked service stores the connection information from the source dataset, like user credentials, server address and database name.

Linked service will be used by the dataset.

Linked services in Azure Data Factory

Activity

An activity is the task that is executed, like copying data or performing a lookup. Activities use datasets to read or write data as the result of a pipeline.

Pipeline

A pipeline is a group of activities linked together to form a data pipeline.

Pipelines and activities in Azure Data Factory
Datasets in Azure Data Factory

Triggers

A tumbling window can define the starting time in the WindowStart setting and the ending time in the WindowEnd setting, defining a time frame to run the data pipeline.

Manual trigger – allow you to manually start pipelines

Schedule trigger – schedule execution of pipeline

Pipeline execution and triggers in Azure Data Factory
Create a trigger that runs a pipeline in response to an event
Create a trigger that runs a pipeline on a schedule
Create a trigger that runs a pipeline on a tumbling window

Databricks

This is an Apache Spark-based technology that allows you to run code in notebooks.

Code can be written in SQL, Python, Scala, and R.

You can have data automatically generate pie charts and bar charts when you run a notebook.

You can override the default language by specifying the language magic command `%<language>` at the beginning of a cell.