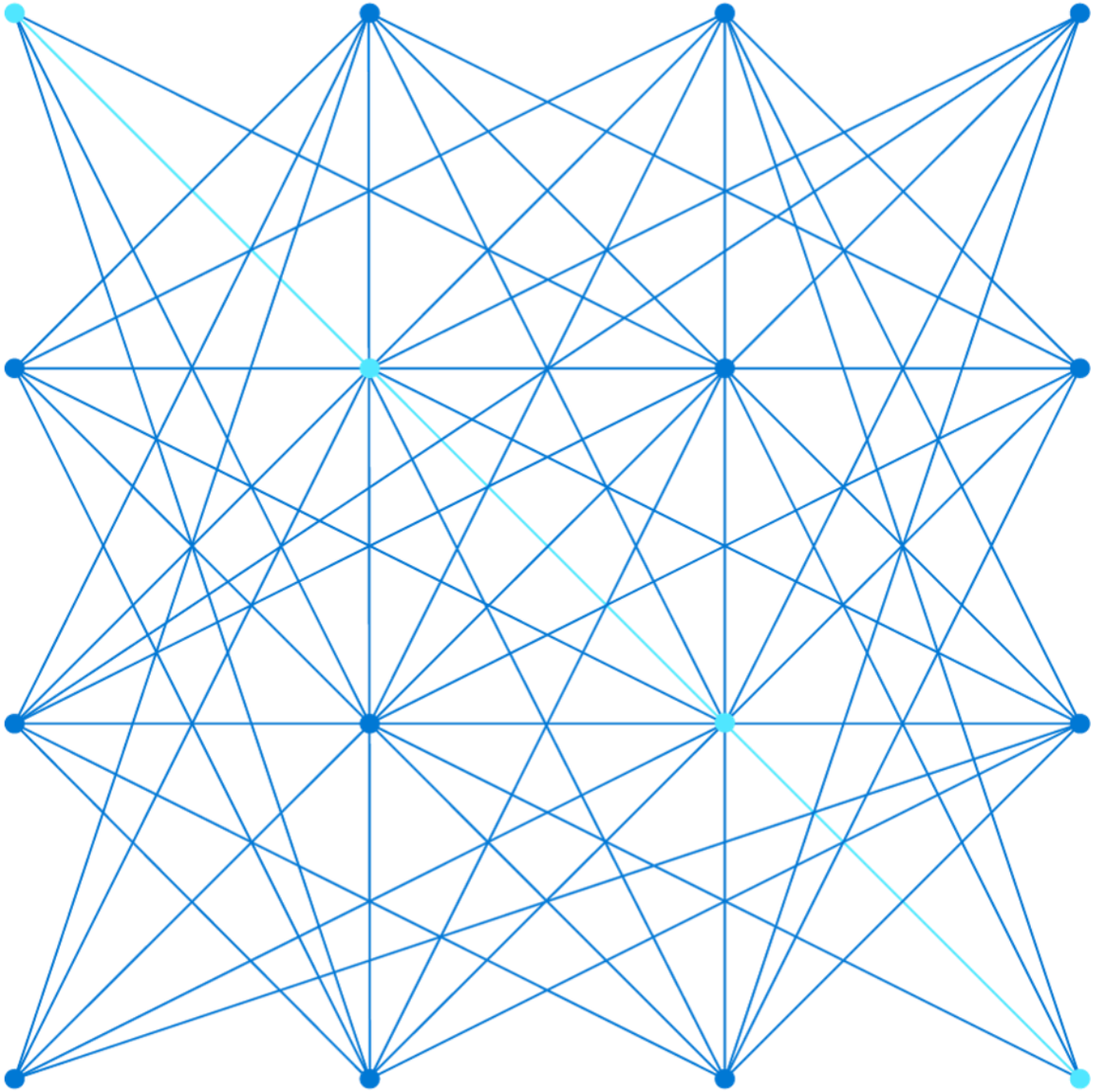
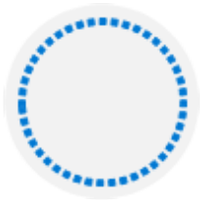
Microsoft Azure

DP-203T00: Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines



Agenda



Lesson 01 – Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

Lesson 01: Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

Integrating data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

1. Use Storage Account

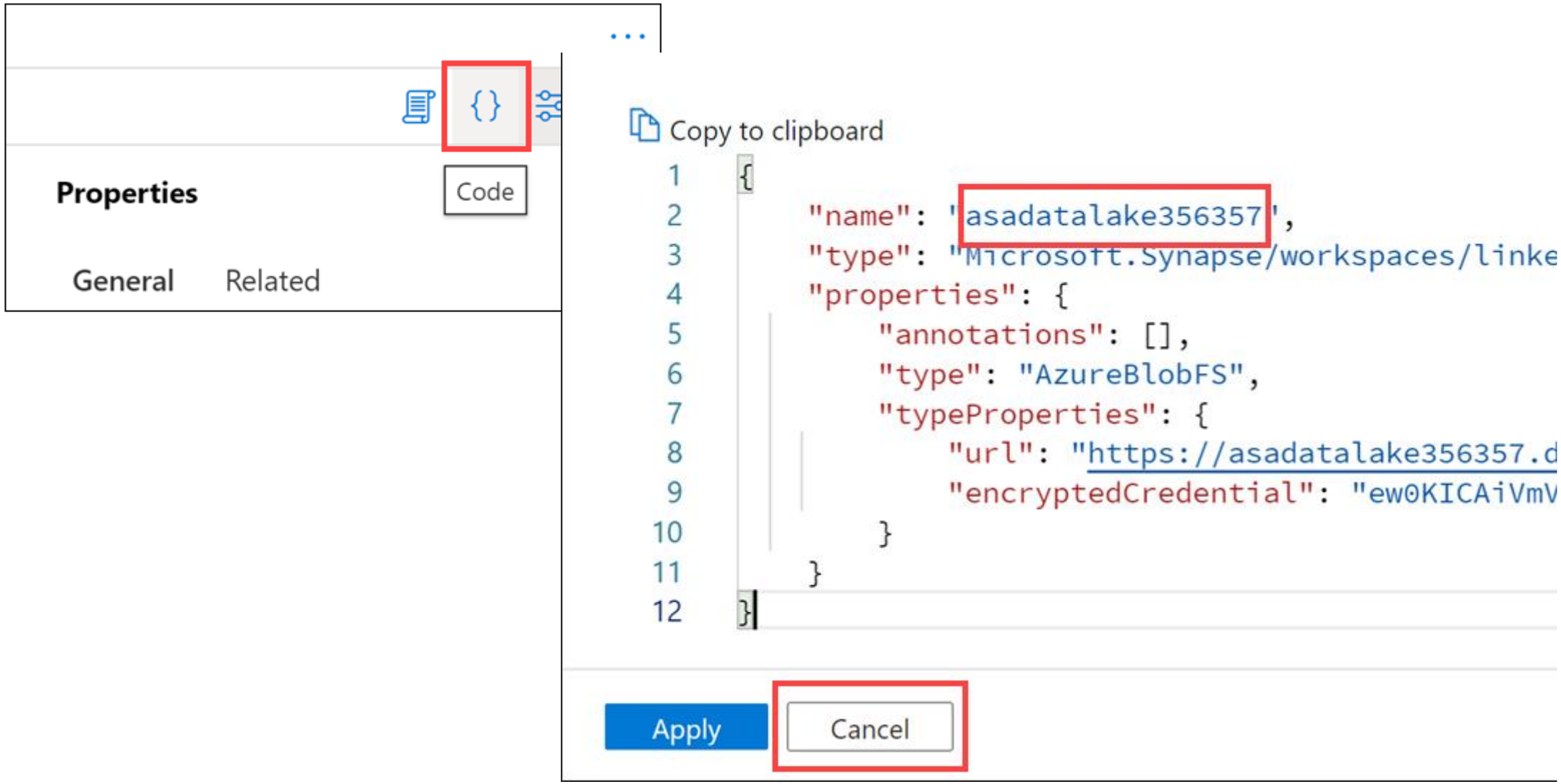
2. Use Azure Synapse Pipeline

3. Create data workflow pipeline

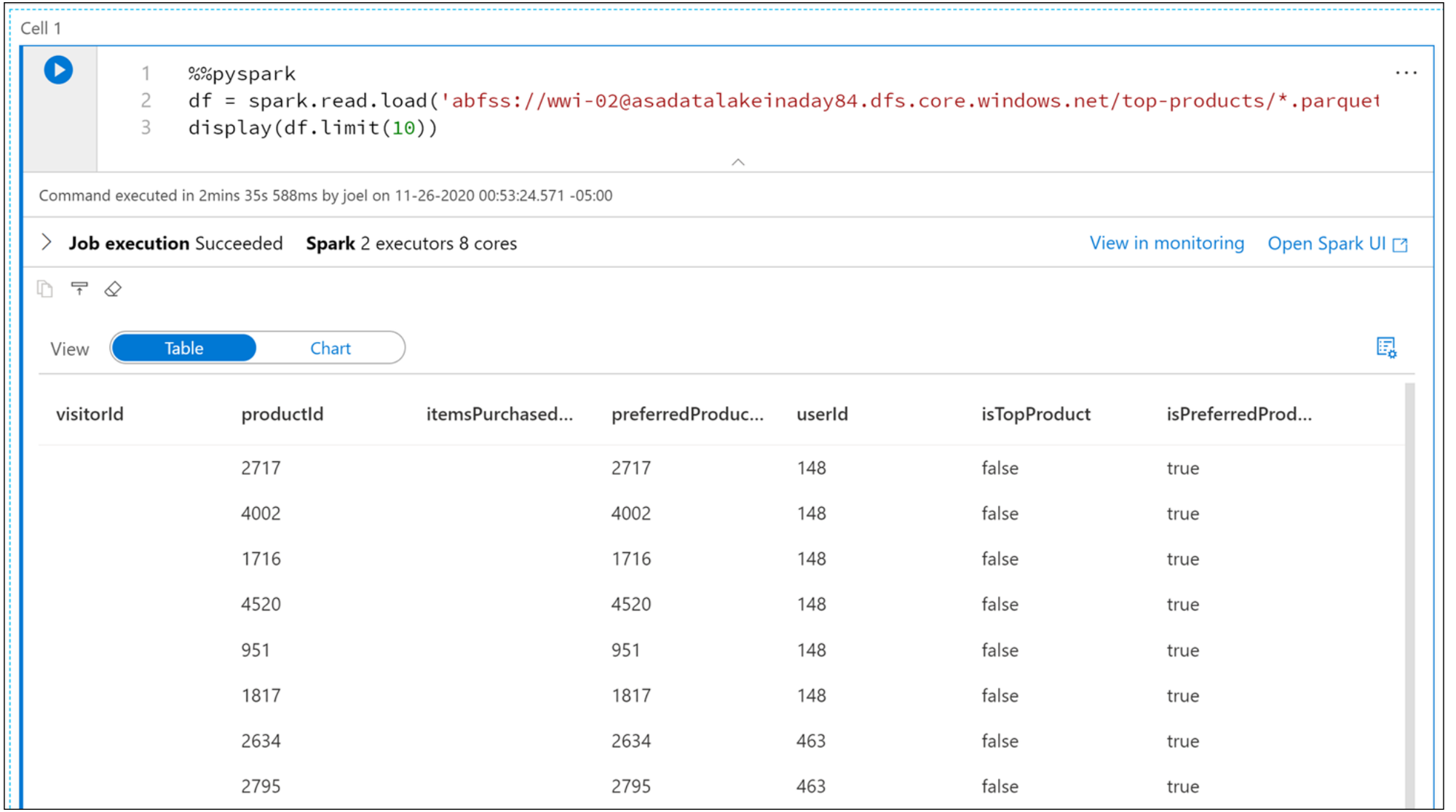
4. Add a Notebook Activity to the pipeline

5. Set parameters & dependency conditions

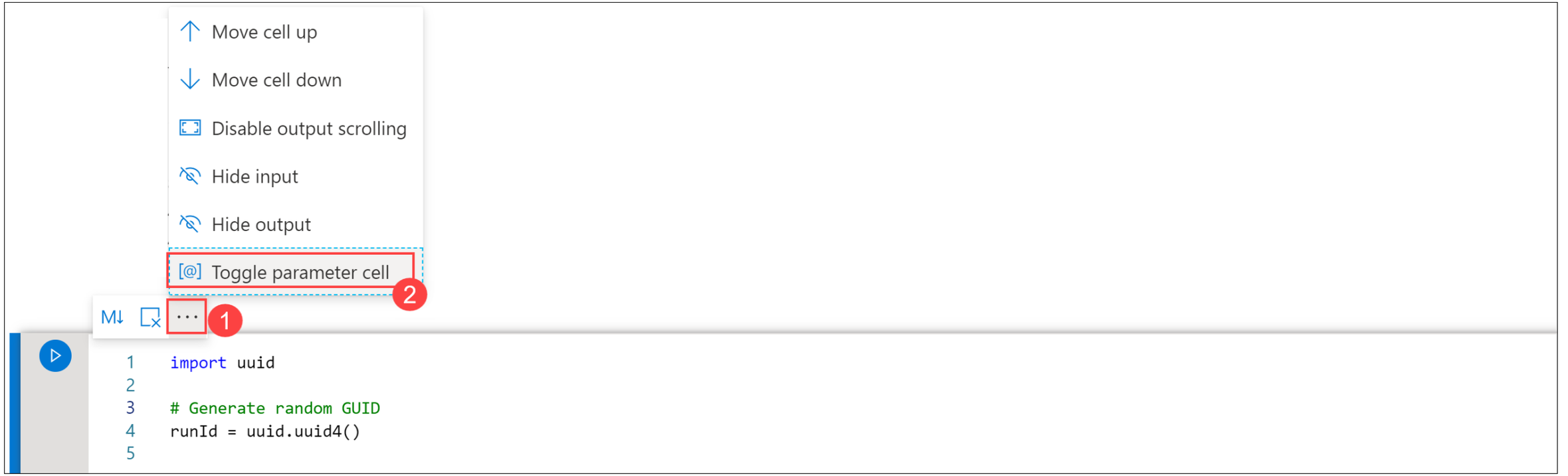
Programmatically creating Azure Data Factory or Azure Synapse Pipelines

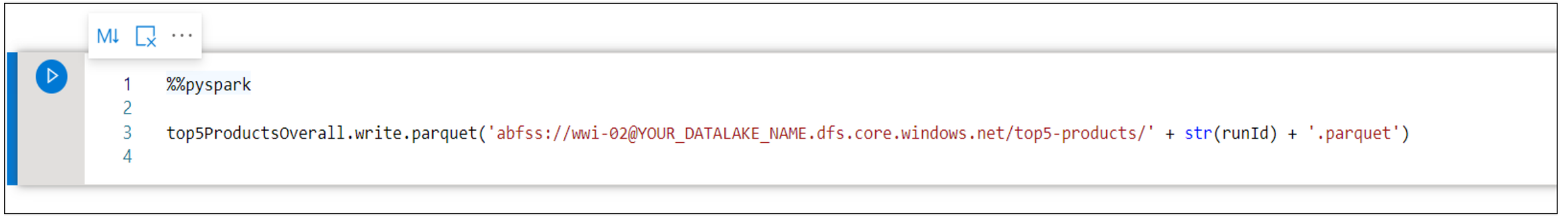


Create a notebook

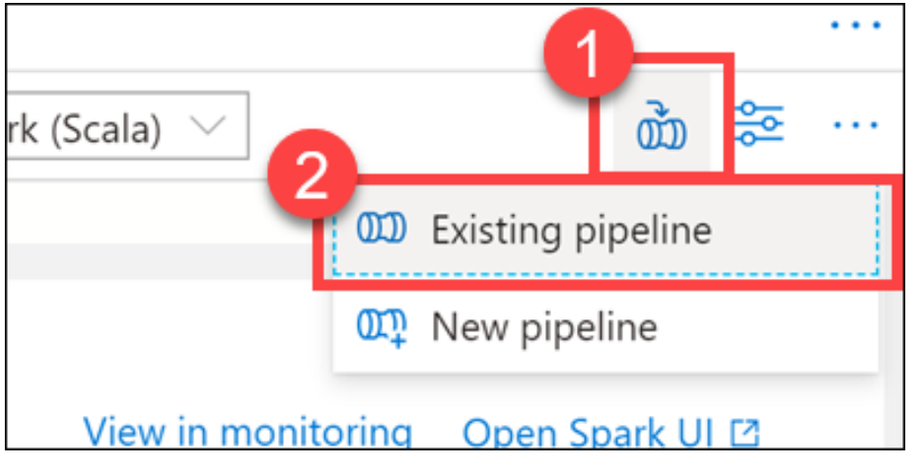


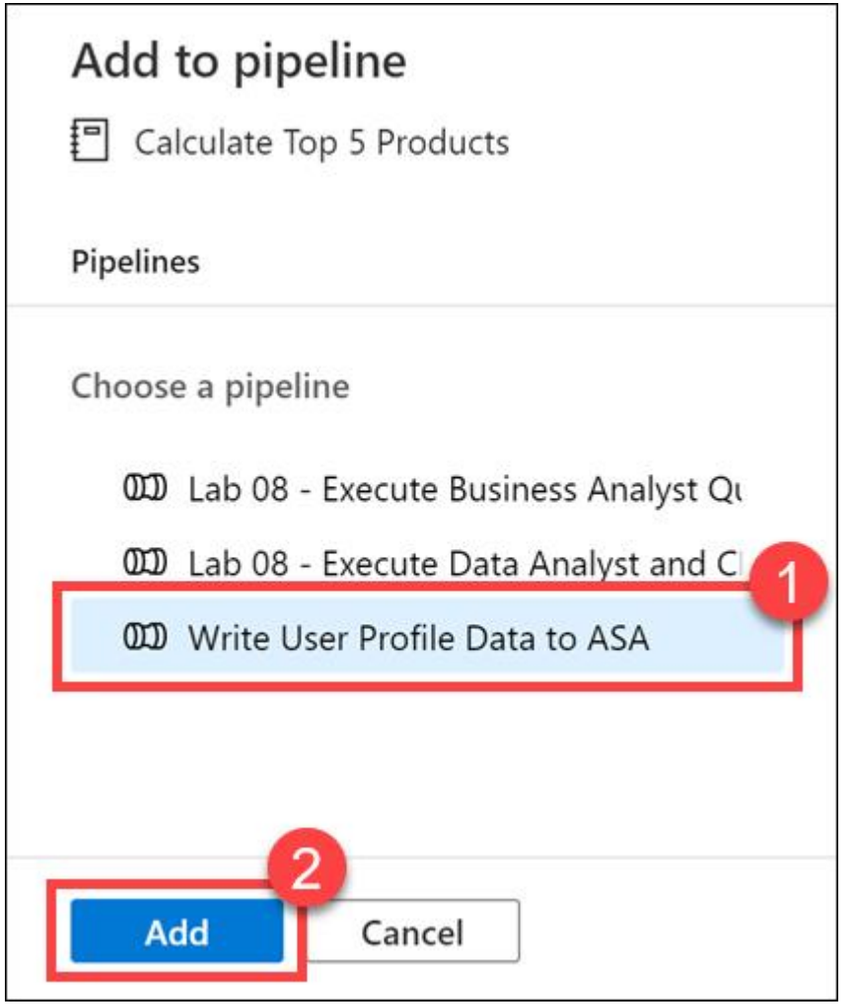
Create notebook parameters



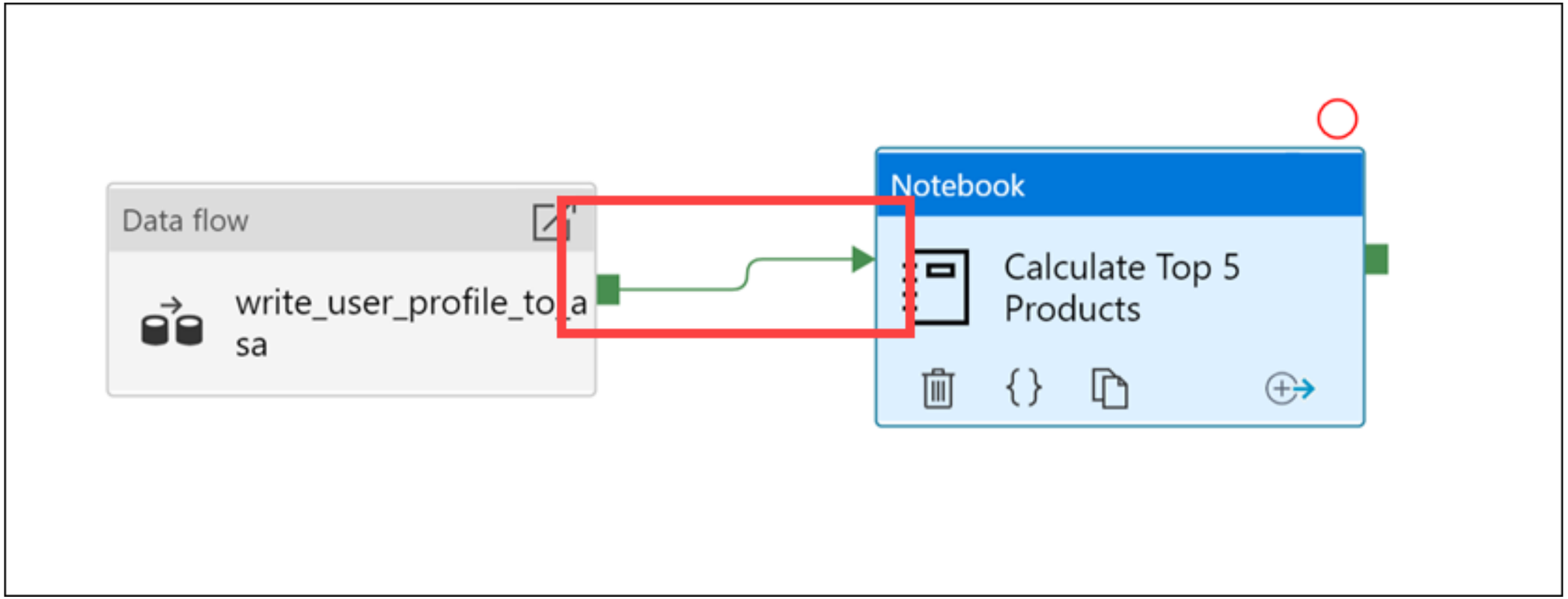


Integrating a notebook into a pipeline





Controlling notebook execution in a pipeline



Review questions

Q01 – Which control flow setting allows you to pass arguments while you're invoking the pipeline?

A01 – Parameter

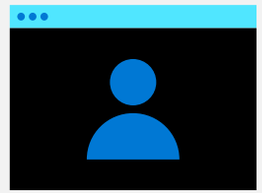
Q02 – There are four dependency conditions that link control flow activities together. Success, Failed, Completed. What is the fourth?

A02 – Skipped

Q03 – What type of parameter should you use when you have multiple pipelines where the parameters names and values are identical?

A03 – Global parameters

Lab: Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines



http://fmdk.io/dp2037

Lab overview

In the lab, the students will create a notebook to query user activity and purchases that they have been made in the past 12 months. They will then add the notebook to a pipeline using the new Notebook activity and execute this notebook after the Mapping Data Flow as part of their orchestration process. While configuring this, the students will implement parameters to add dynamic content in the control flow and validate how the parameters can be used.

Lab objectives

After completing this lab, you will be able to:

Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

Lab review

Question 1 – Which Spark library is used to generate a random GUID?

Question 2 – What color is the Success dependency condition?

Question 3 – Which tab enables you to add dynamic content?

Question 4 – Where can you view information about a pipeline execution?

Module summary

In this module, you have learned about:

Integrate data from Notebooks with Azure Data Factory or Azure Synapse Pipelines

Next steps

After the course, consider watching the video on channel 9 that provides a summary of performing [Iterative development and debugging with Azure Data Factory]. Note that the content in this video also applies to Azure Synapse Pipelines too.

Microsoft Azure