Name : Vishal Jha

Batch : Data Engineering

Date : 21-02-2024

Topic – Azure Databricks Coding Challenge – 03

3. Execute & explain, Azure Datafactory and its copy activity.

Explanation –

Azure Data Factory (ADF) is a cloud-based data integration service that allows you to create, schedule, and manage data pipelines to move and transform data from various sources to various destinations. It provides a way to create workflows (pipelines) that orchestrate and automate the movement and transformation of data.

Copy Activity is one of the core activities in Azure Data Factory, used to copy data from a source to a destination. It supports a wide range of source and destination data stores, including Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage, Amazon S3, Google Cloud Storage, and many others.

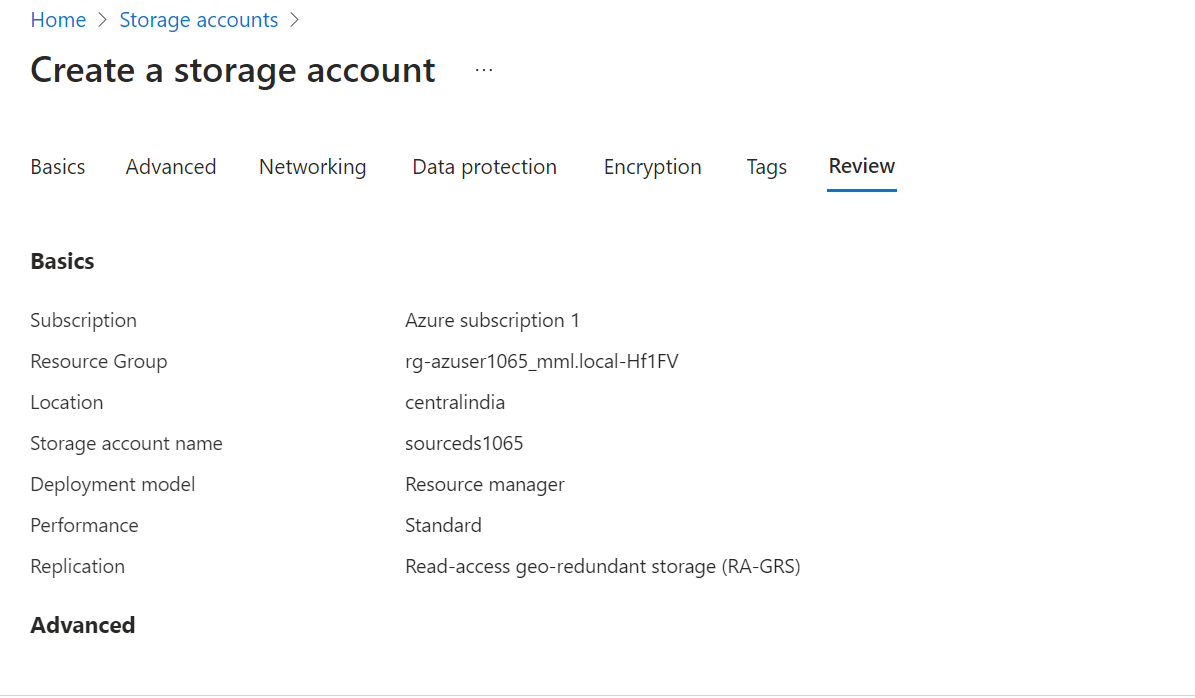
Here's an example to demonstrate how to use Copy Activity in Azure Data Factory:

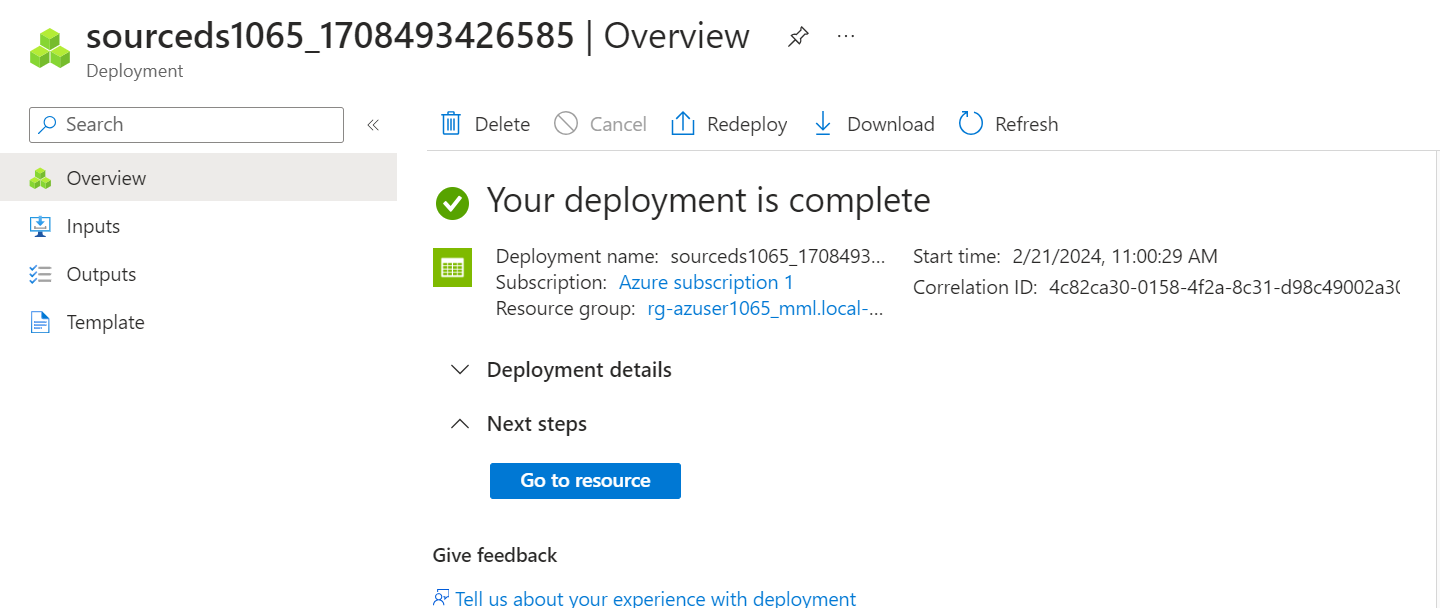
1. Create a Data Factory: First, create an Azure Data Factory instance in the Azure portal.
2. Create a Pipeline: Inside your Data Factory, create a new pipeline. A pipeline is a logical grouping of activities that together perform a task.
3. Add Copy Activity: Add a Copy Activity to your pipeline. In the Copy Activity settings, you will configure the source dataset, the destination dataset, and the mapping of columns between the source and destination.
4. Configure Source and Destination Datasets: Define the source dataset, specifying the connection information and format of the source data. Similarly, define the destination dataset, specifying the connection information and format of the destination data.
5. Mapping: If the schema of the source and destination datasets is different, you can define mappings to map columns from the source to the destination.
6. Trigger and Monitor: Trigger the pipeline to run manually or set up a schedule for it to run automatically. You can monitor the progress and status of your pipeline runs in the Azure Data Factory portal.
7. Monitor Pipeline Runs: Monitor the pipeline runs to ensure that the data copy operation is successful. You can view details such as the number of rows copied, the duration of the operation, and any errors encountered.

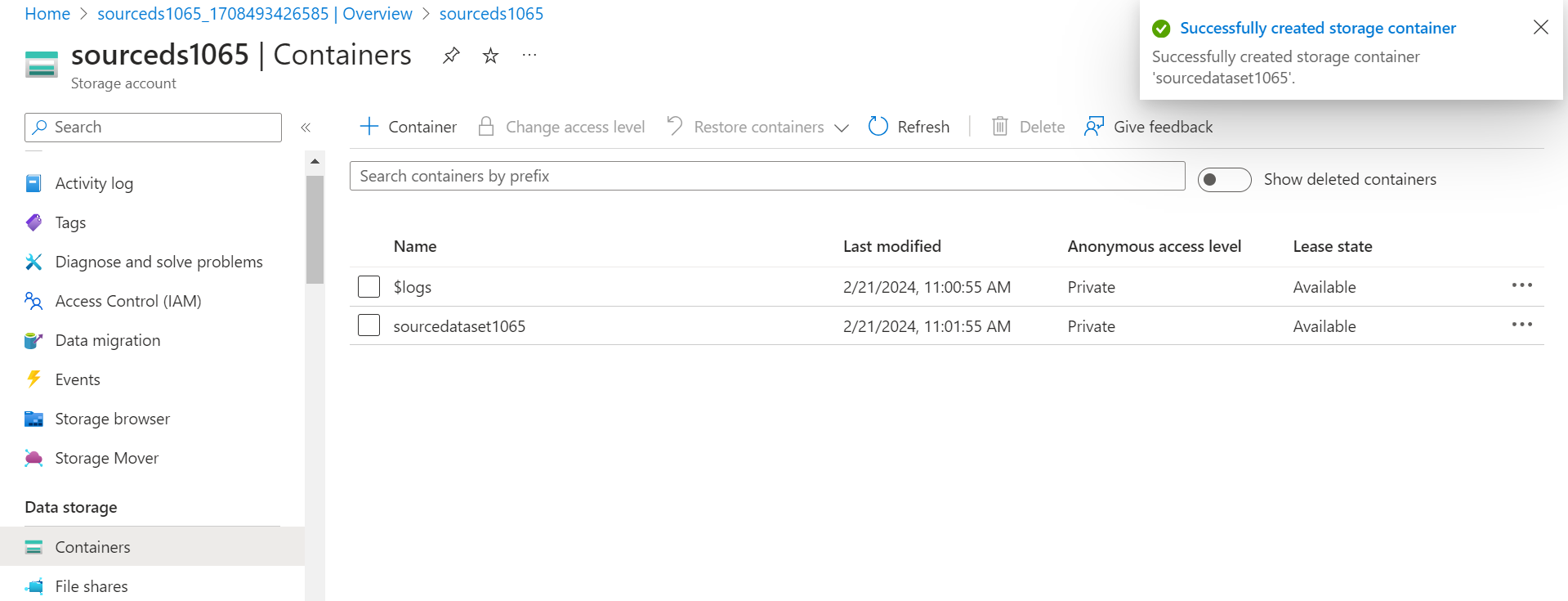
Overall, Azure Data Factory and its Copy Activity provide a powerful way to move and transform data between different data sources and destinations, helping organizations to build scalable and efficient data integration workflows in the cloud.

Execution –

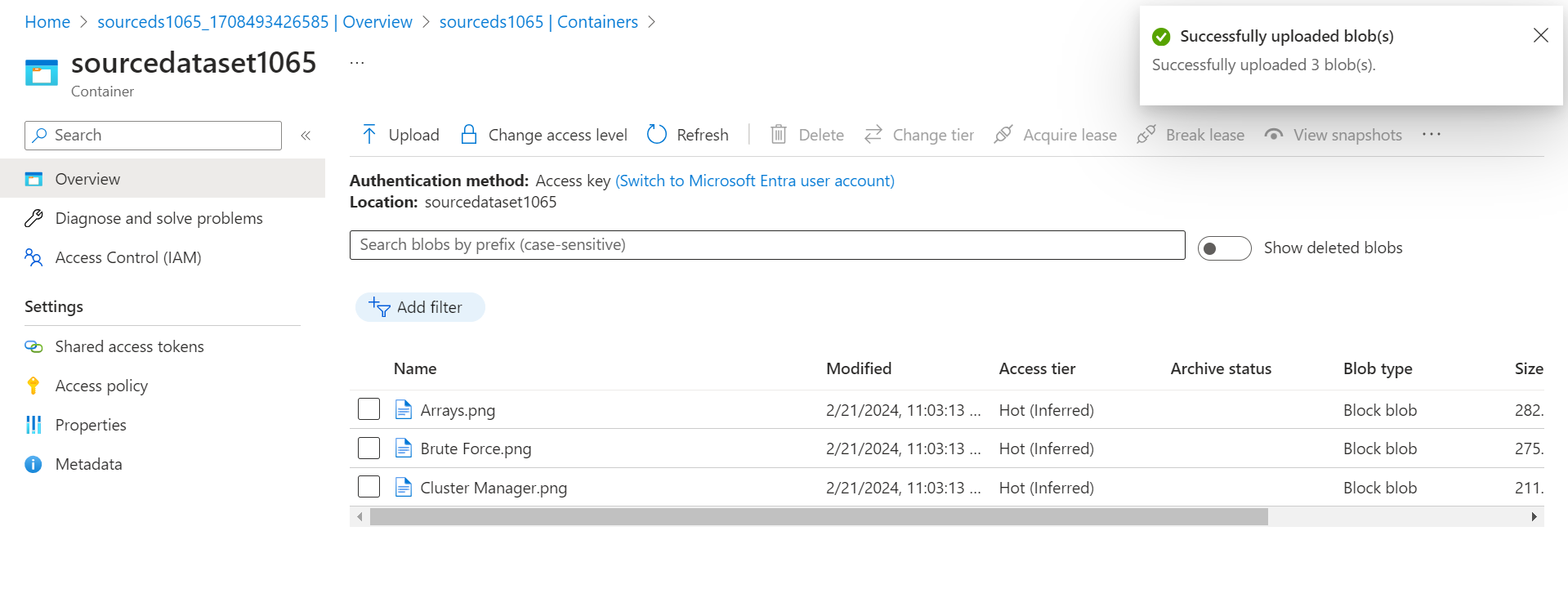
1. Created Source Storage Account and Container in it.



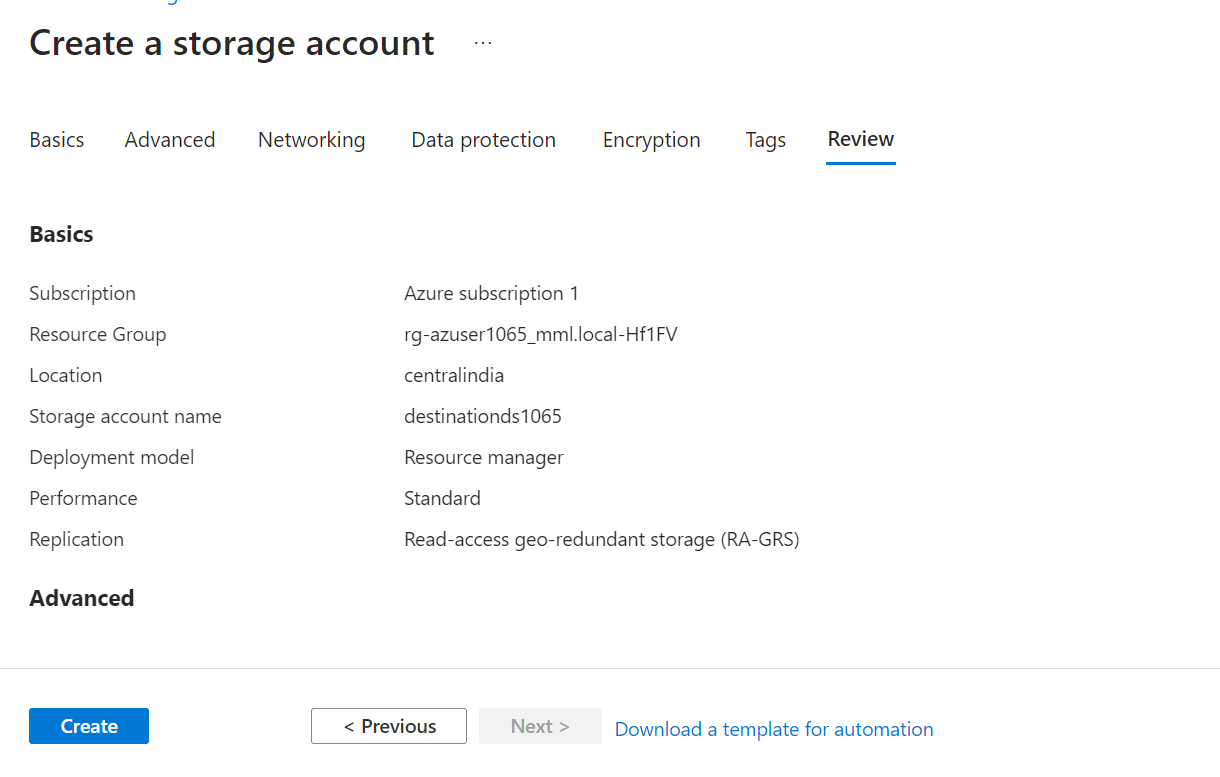


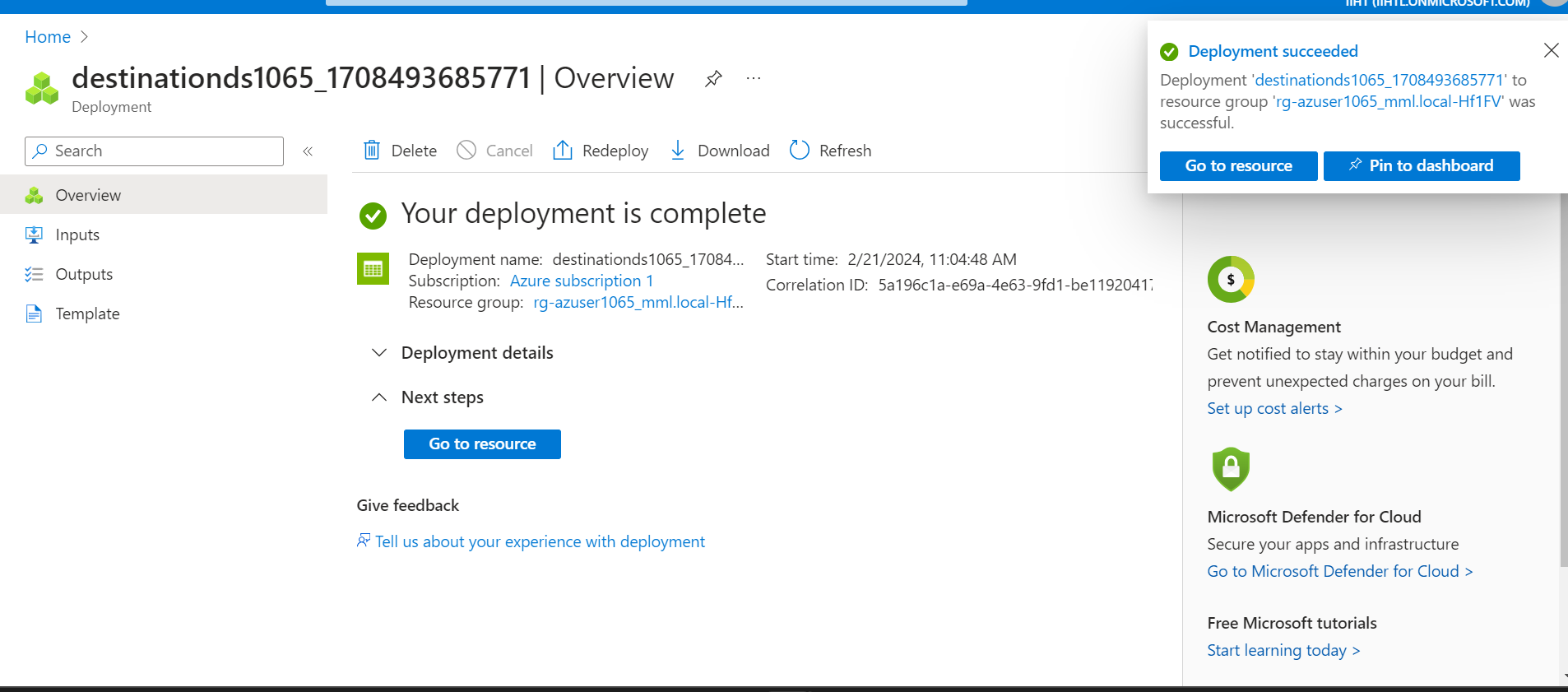


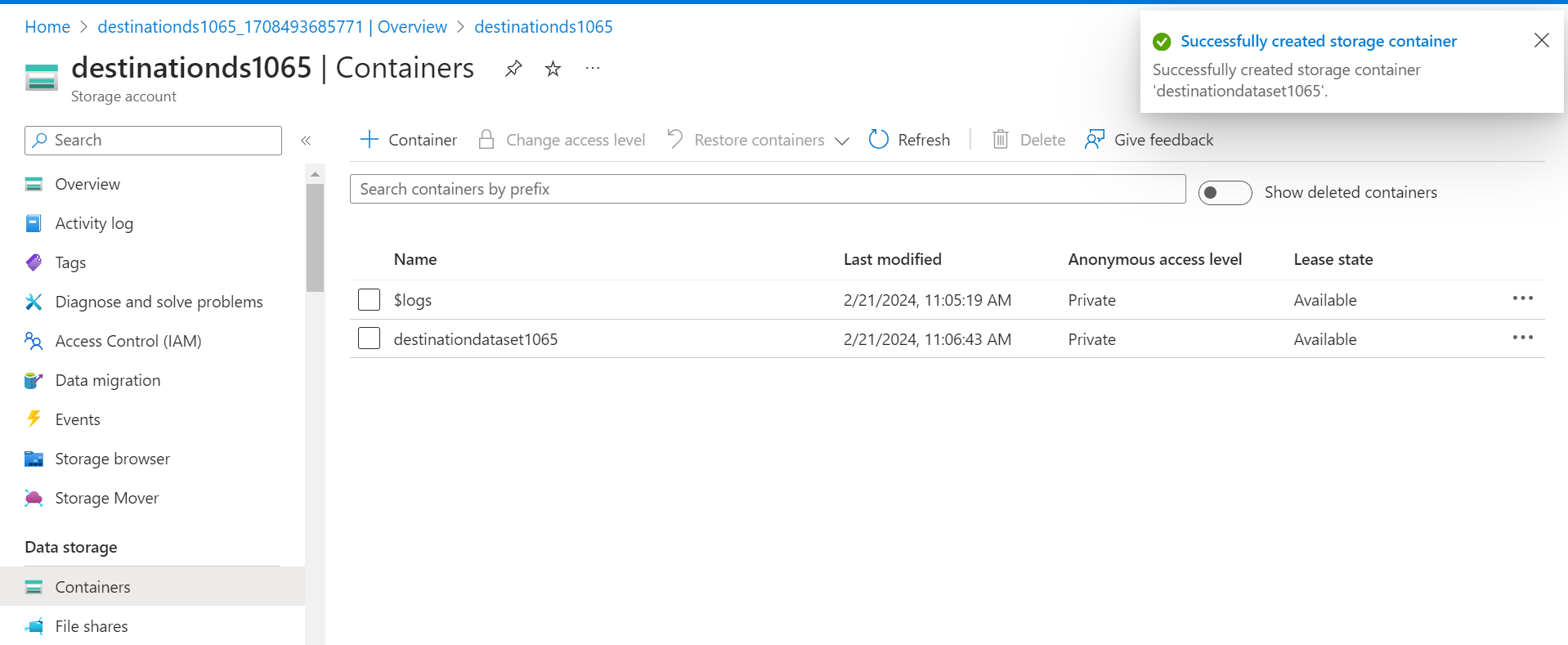
1. Uploaded few files in Source Data Storage Container.



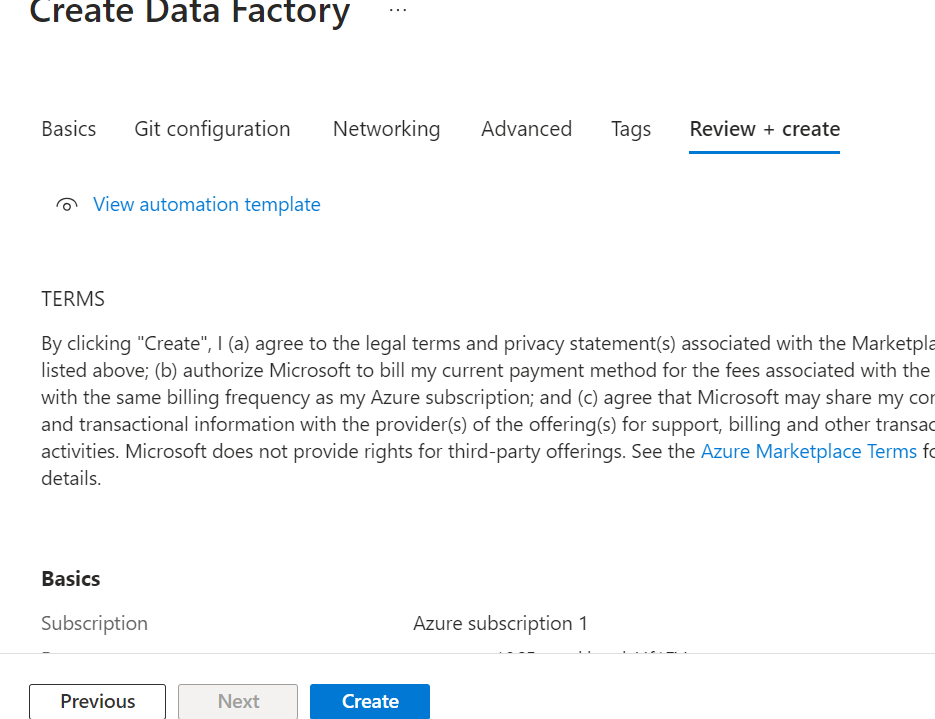
1. Creating another Storage Account for Destination and Container in it.

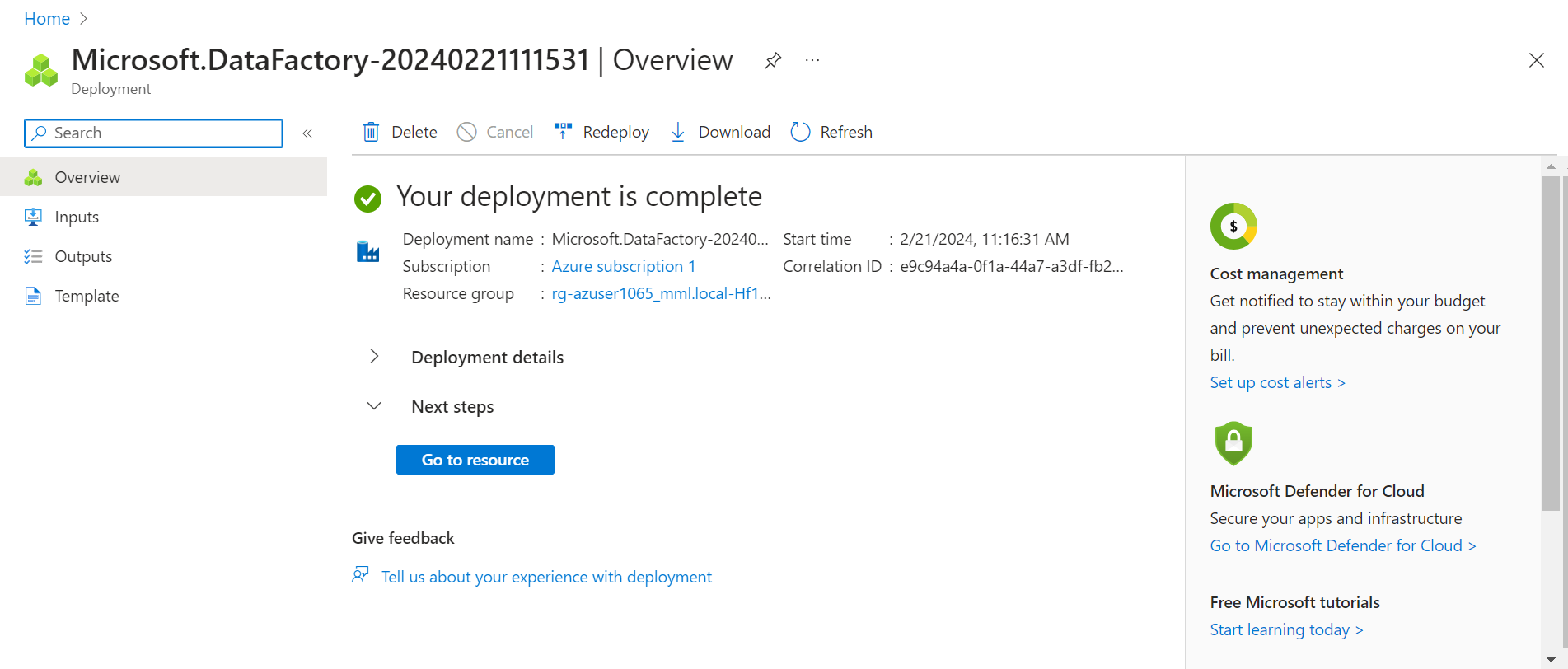


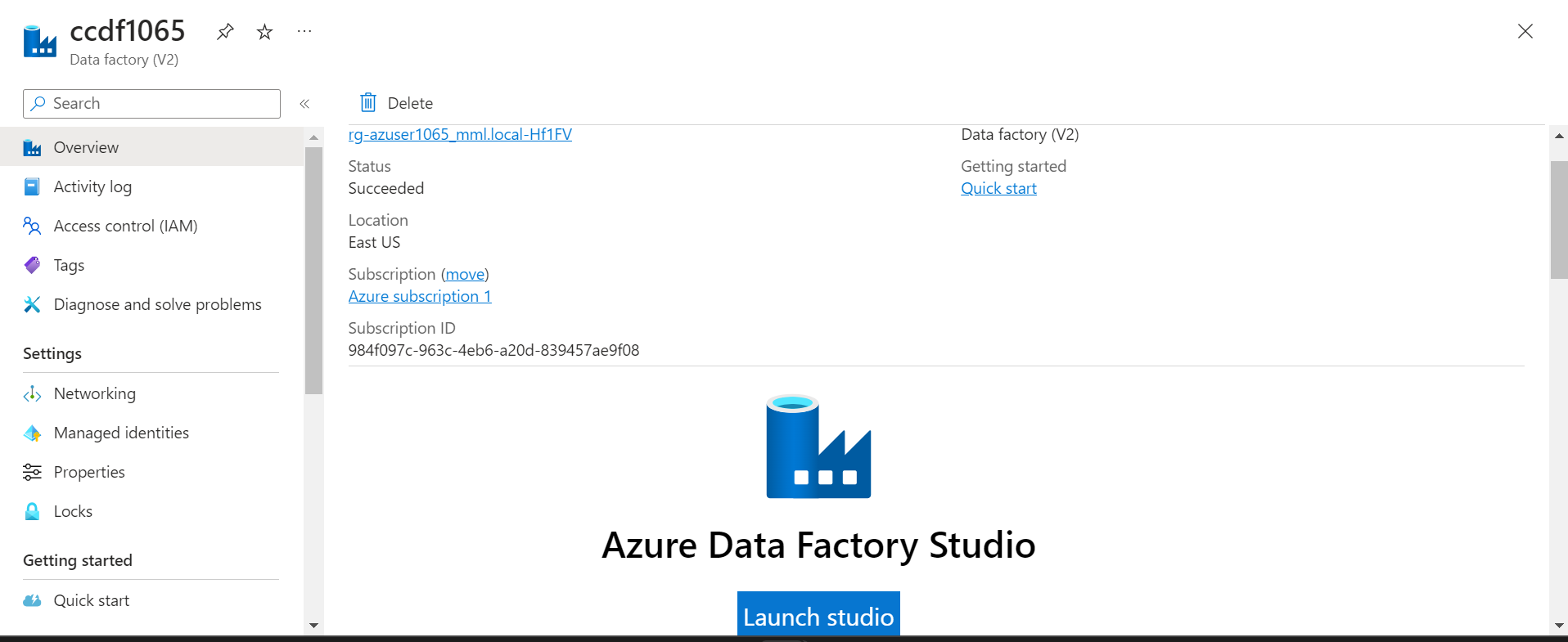




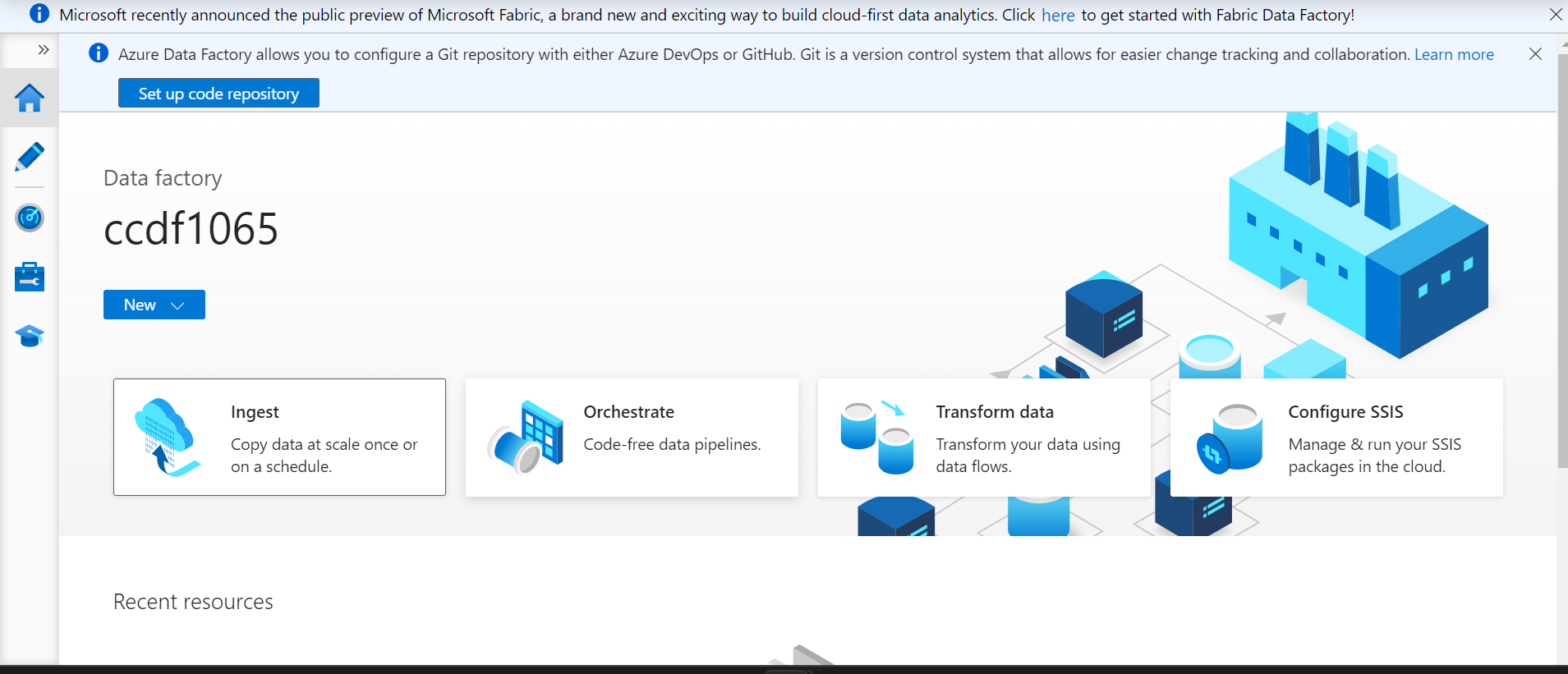
1. Create Data Factory.



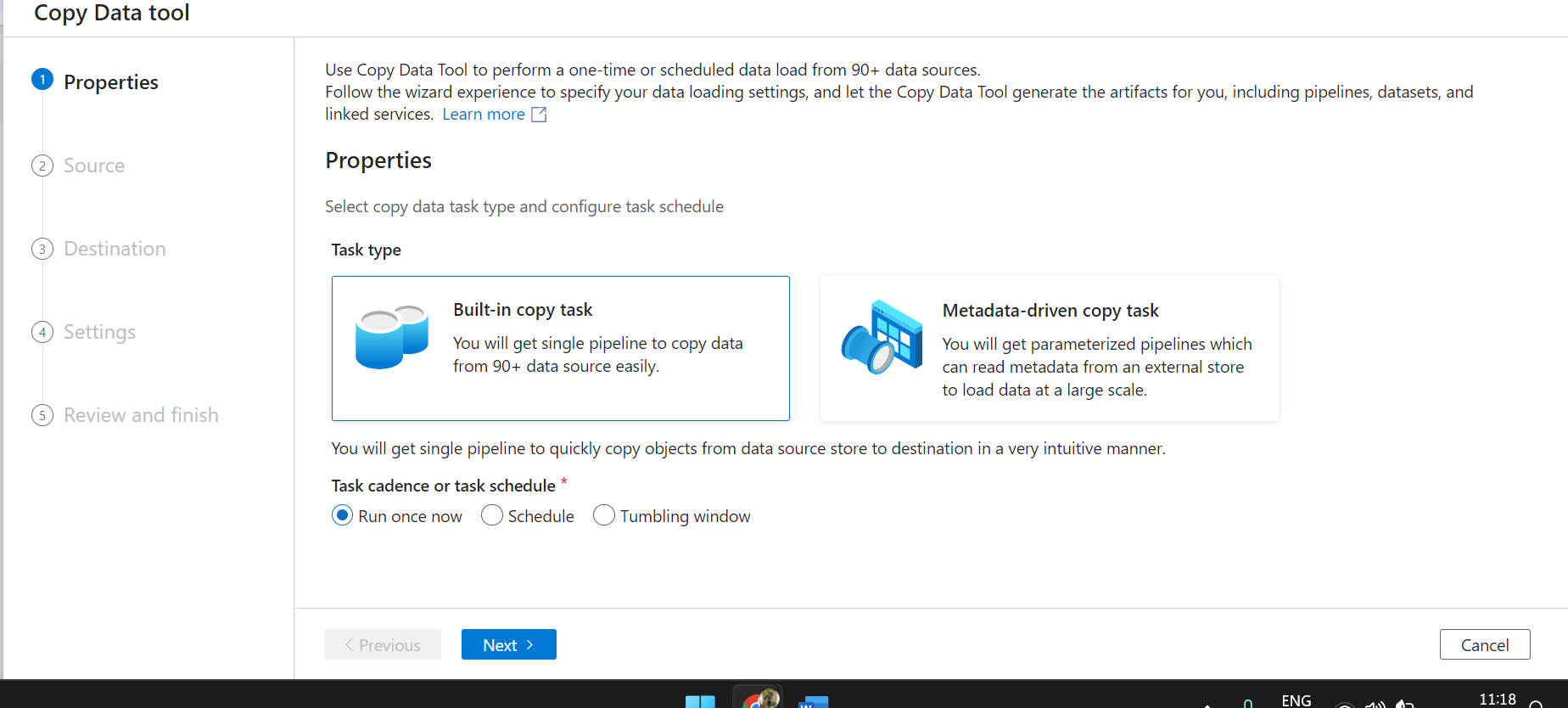




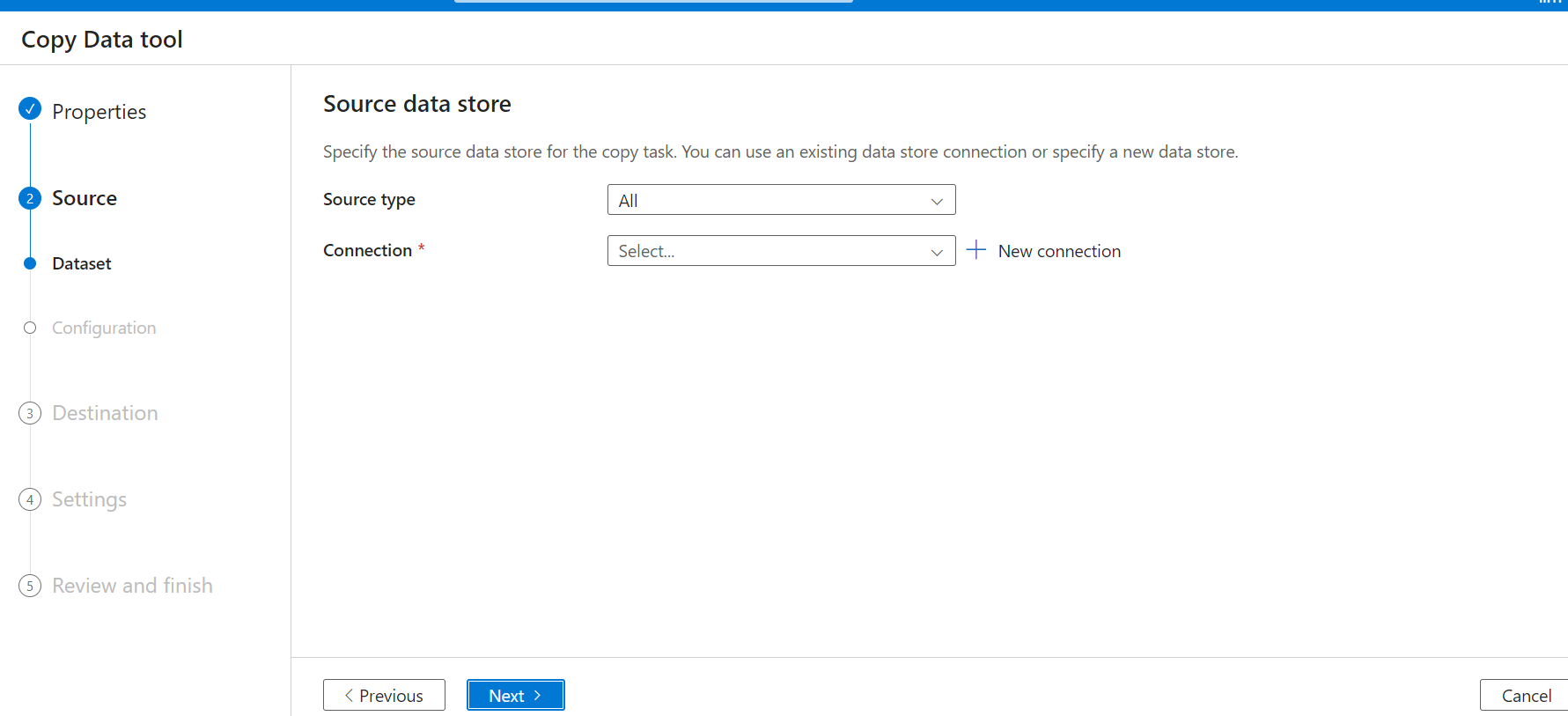
1. ADF Studio.



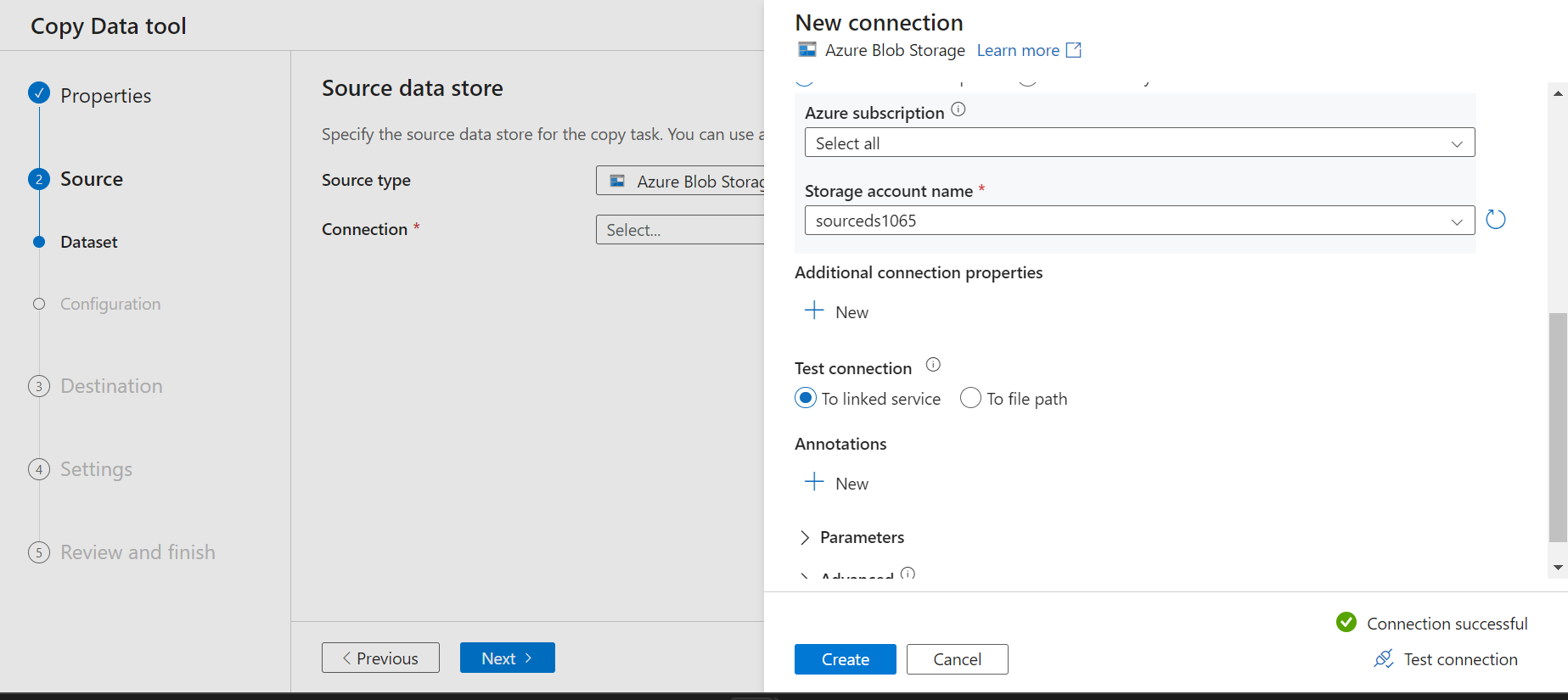
1. Click on Injest.



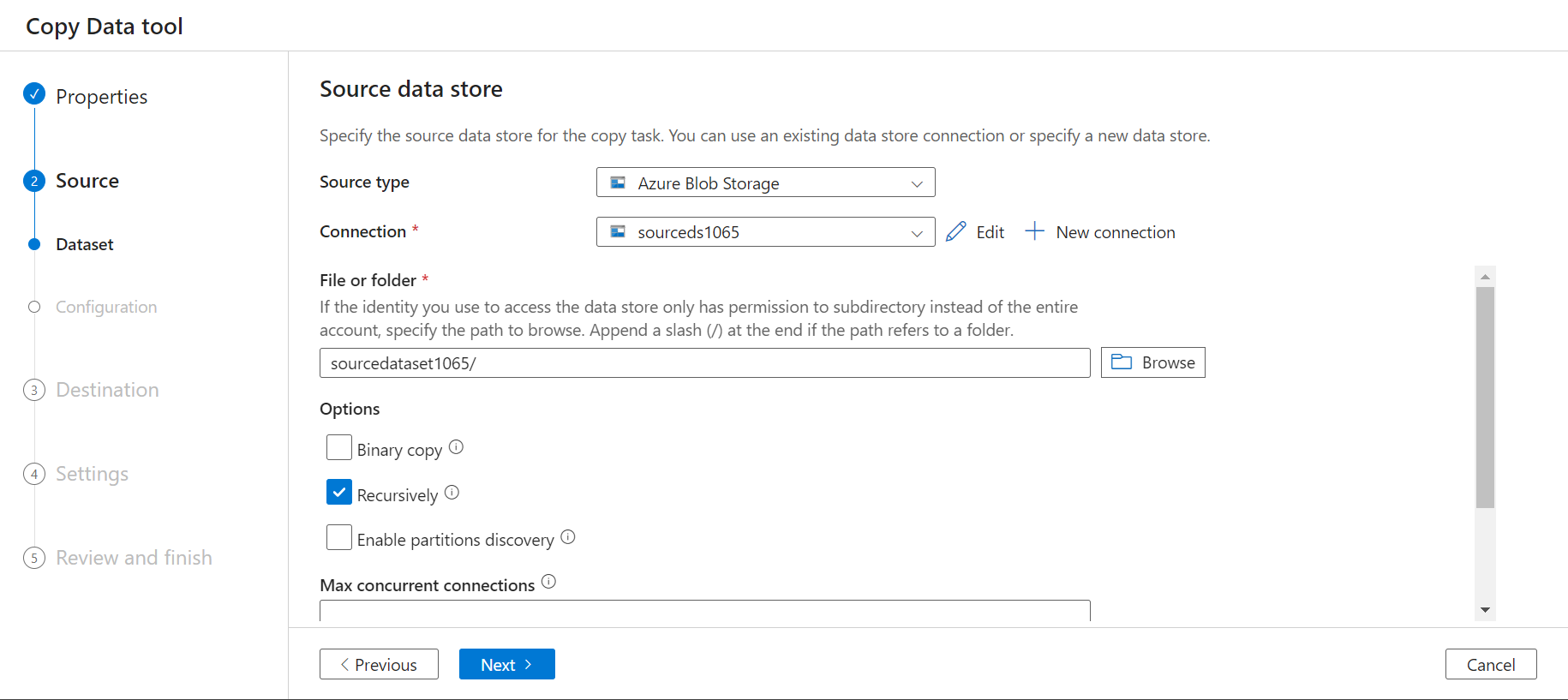
1. Select Built-in copy task and next.



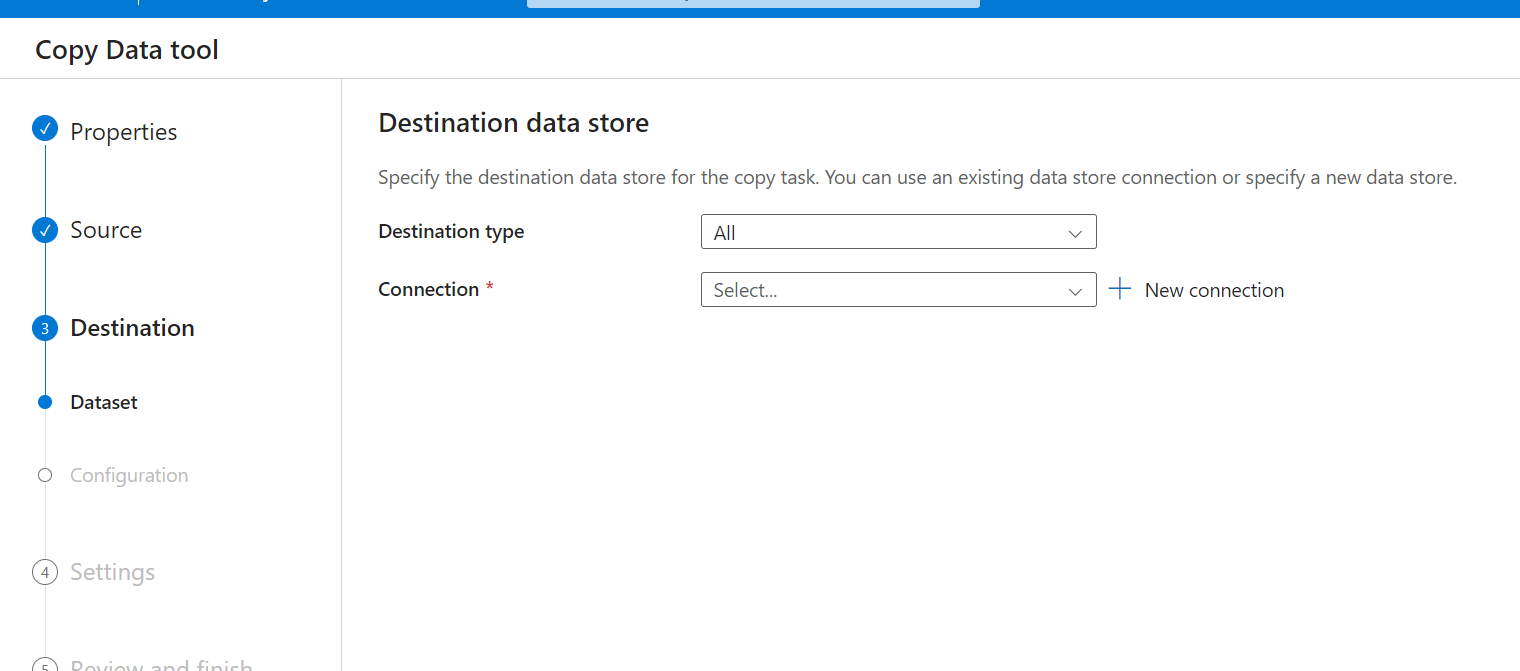
1. Now Give Source Dataset detail.

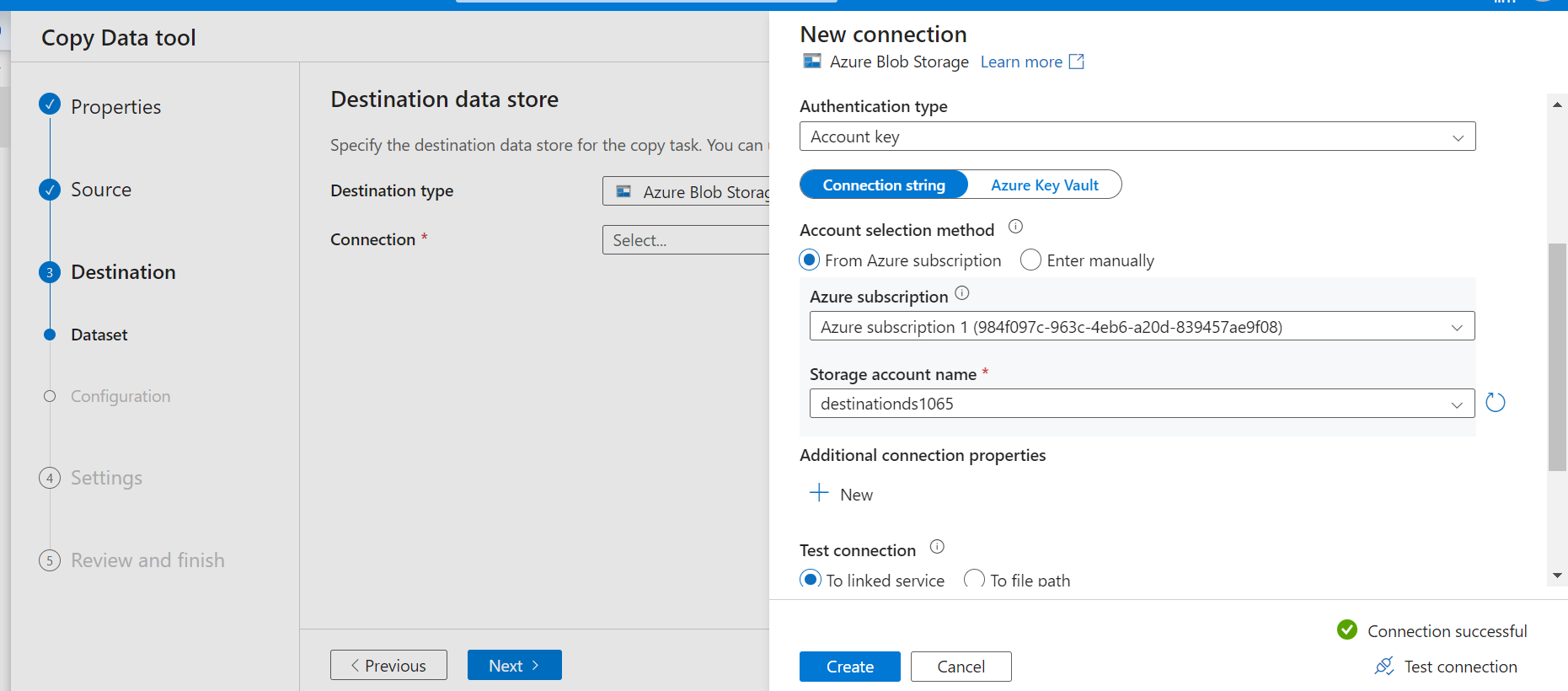


1. Select Source Data Storage Container.

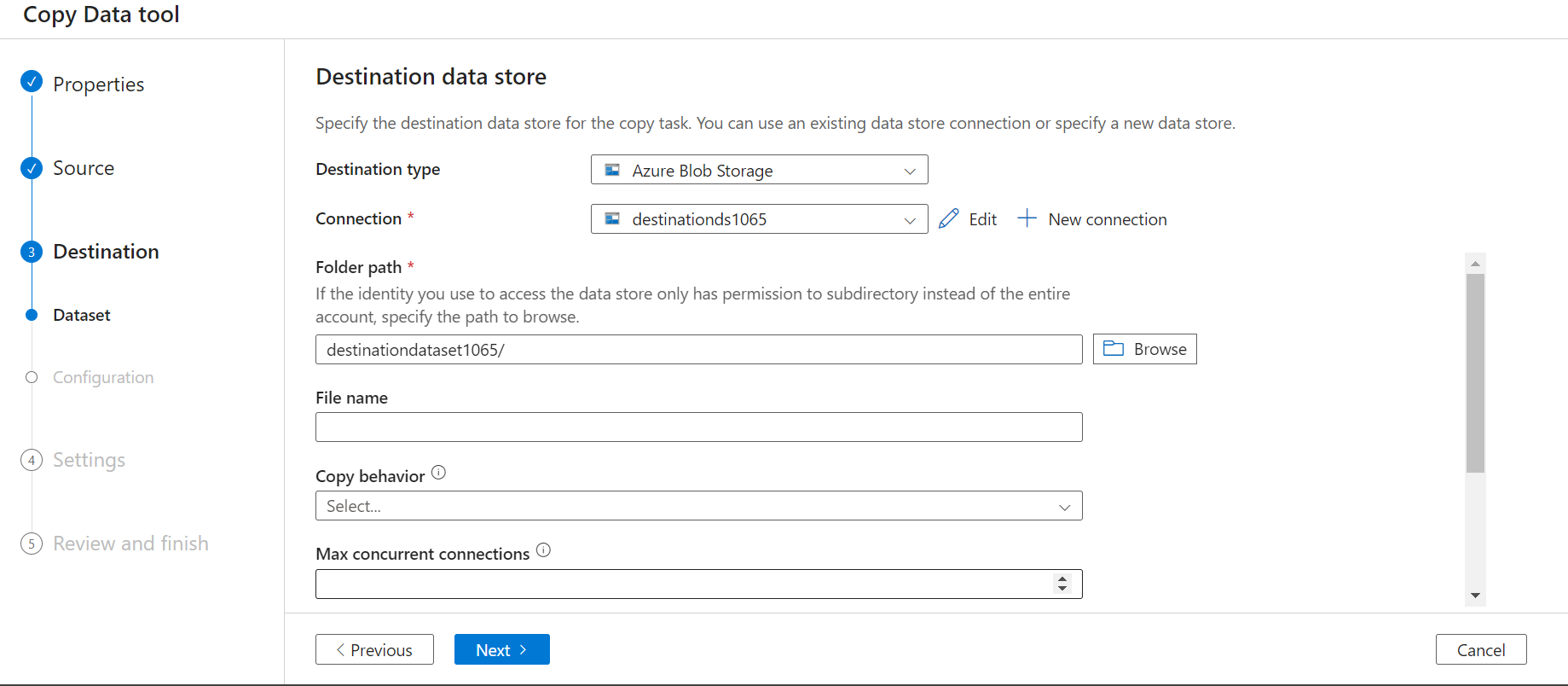


1. Select Destination Data Storage.

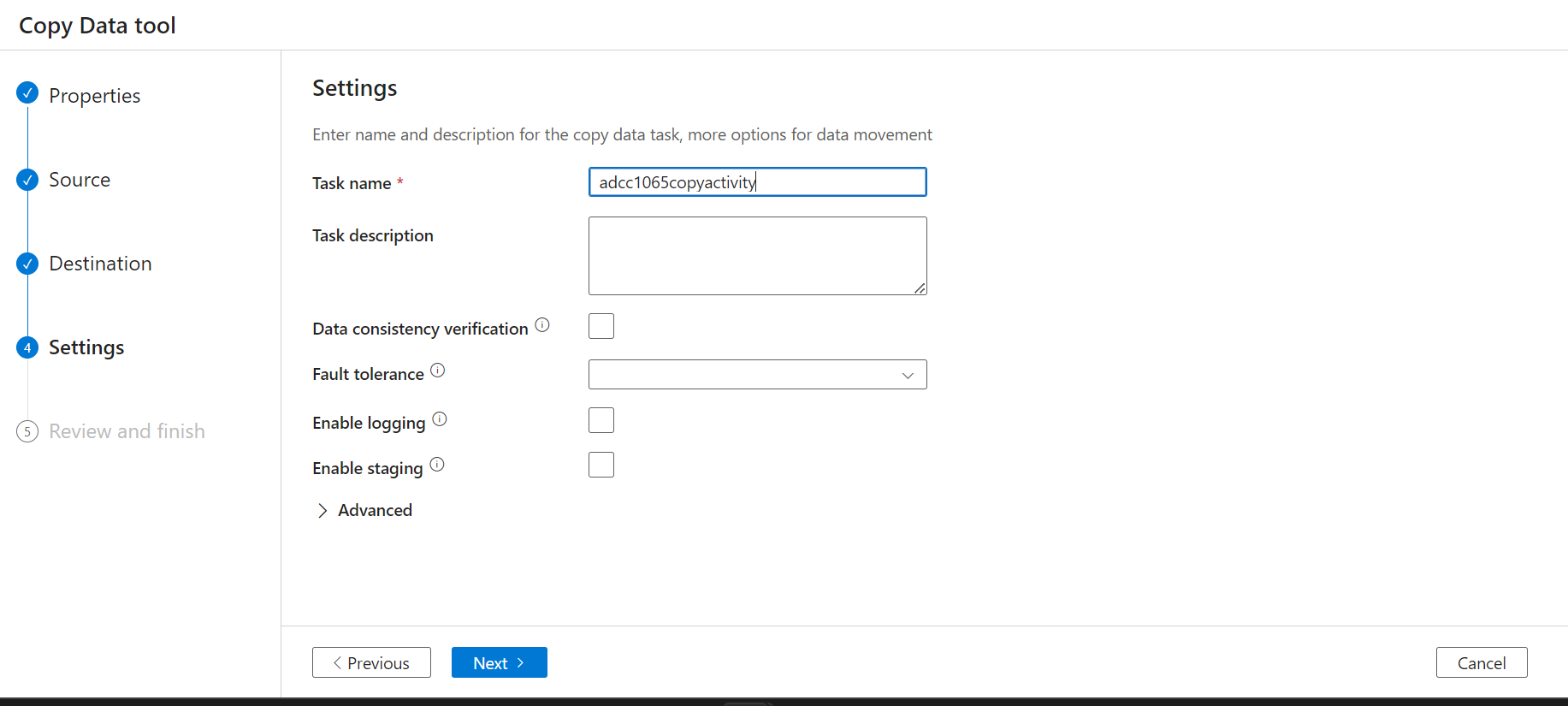




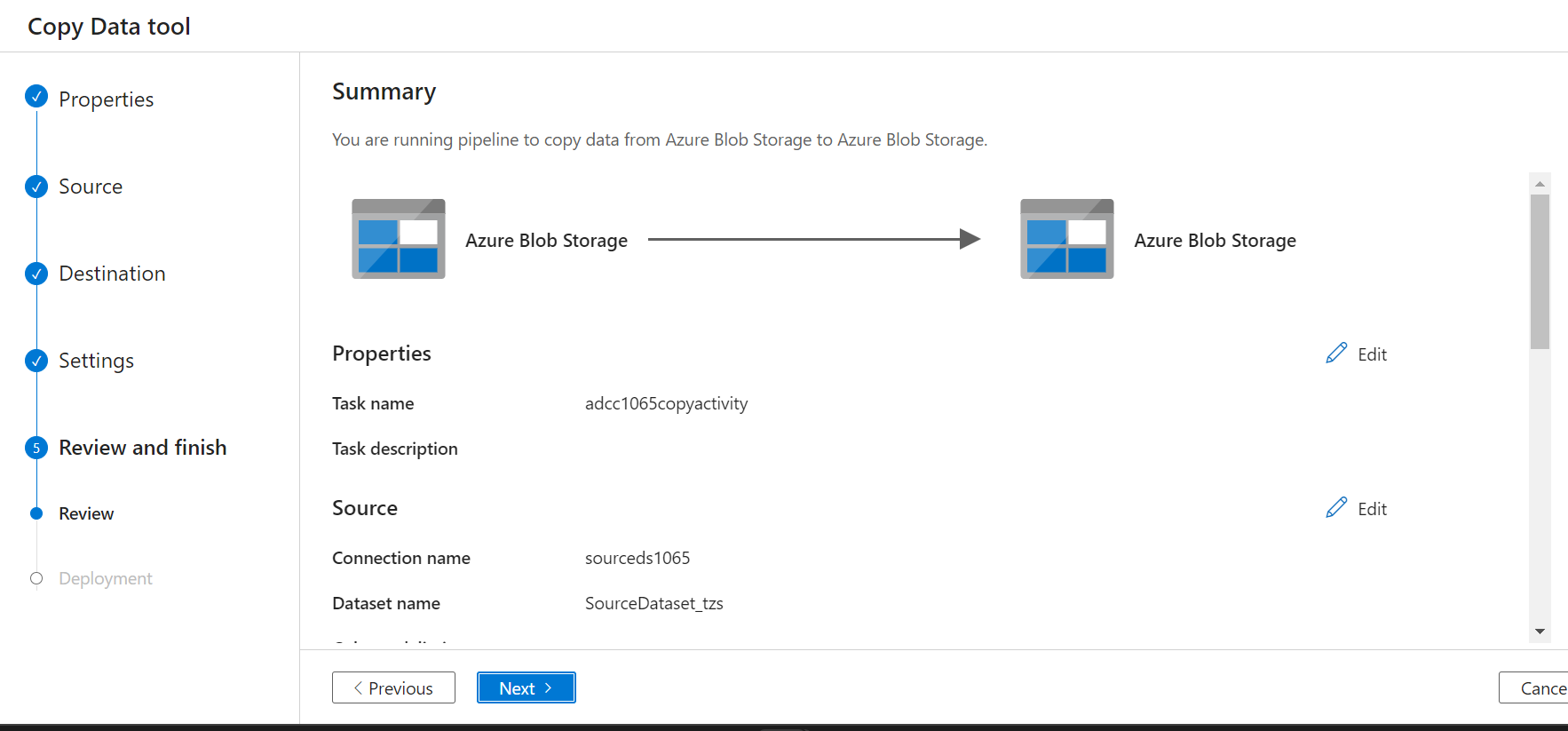
1. Select Destination Data Storage Container.



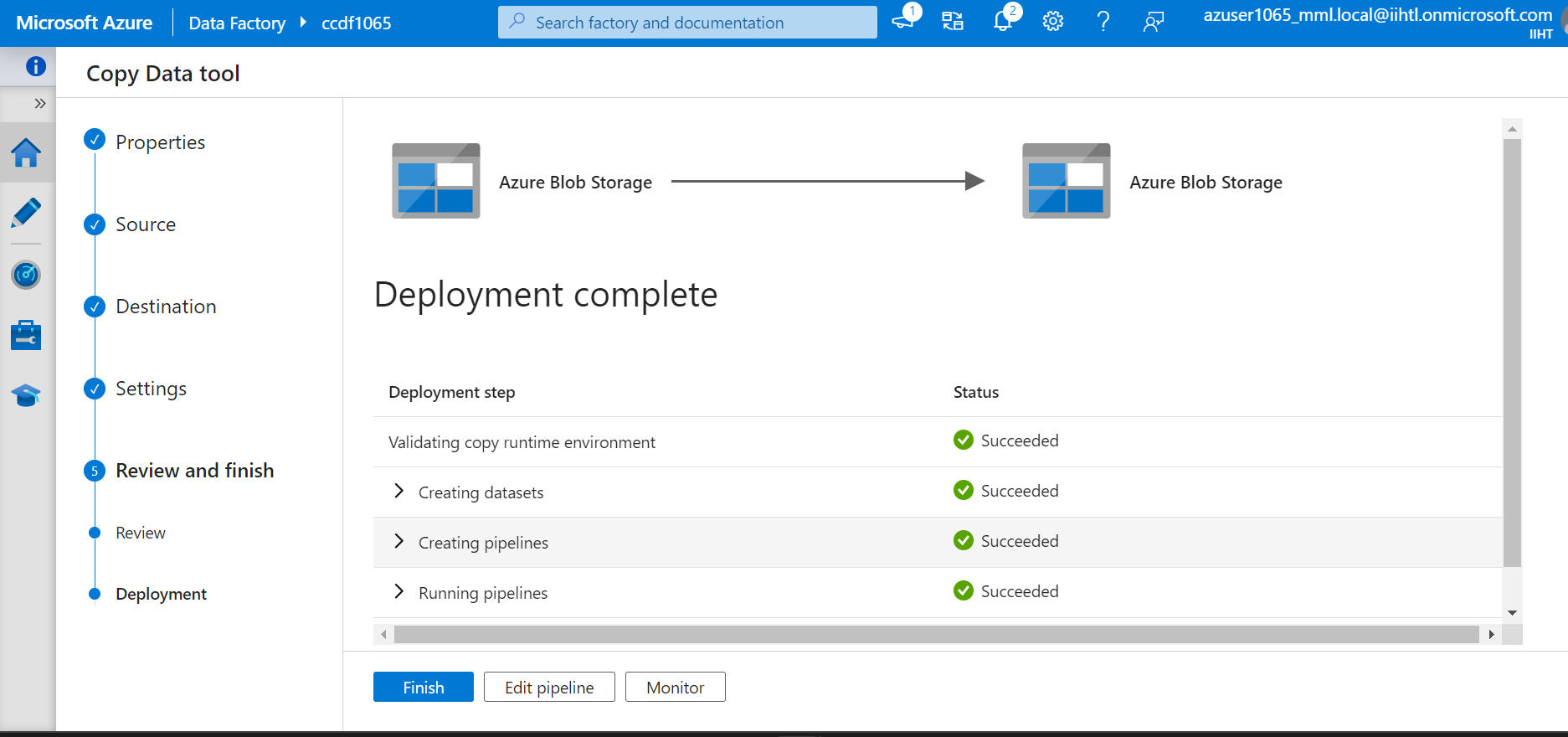
1. Give pipeline a name.



1. Review



1. Deployment Completed.



1. Files get Copied From Source Data Storage Container to Destination Data Storage Container.

