**Coding Challenge-3**

**Question 3: Execute & explain, Azure datafactory and its copy activity.**

**Solution:**

**Azure Datafactory:** It is a cloud based data integration service that slows you to create data driven workflows in the cloud for orchestrating and automating data movement and data transformation. Basically, it helps in transferring data from source to destination and it doesn’t store the data itself. For example, we have a bus which transfers students from their home to school. This bus here acts like an azure datafactory.

Three steps of Azure Datafactory:

* Connect and collect
* Transform and Enrich
* Publish

**Copy activity in azure datafactory:**

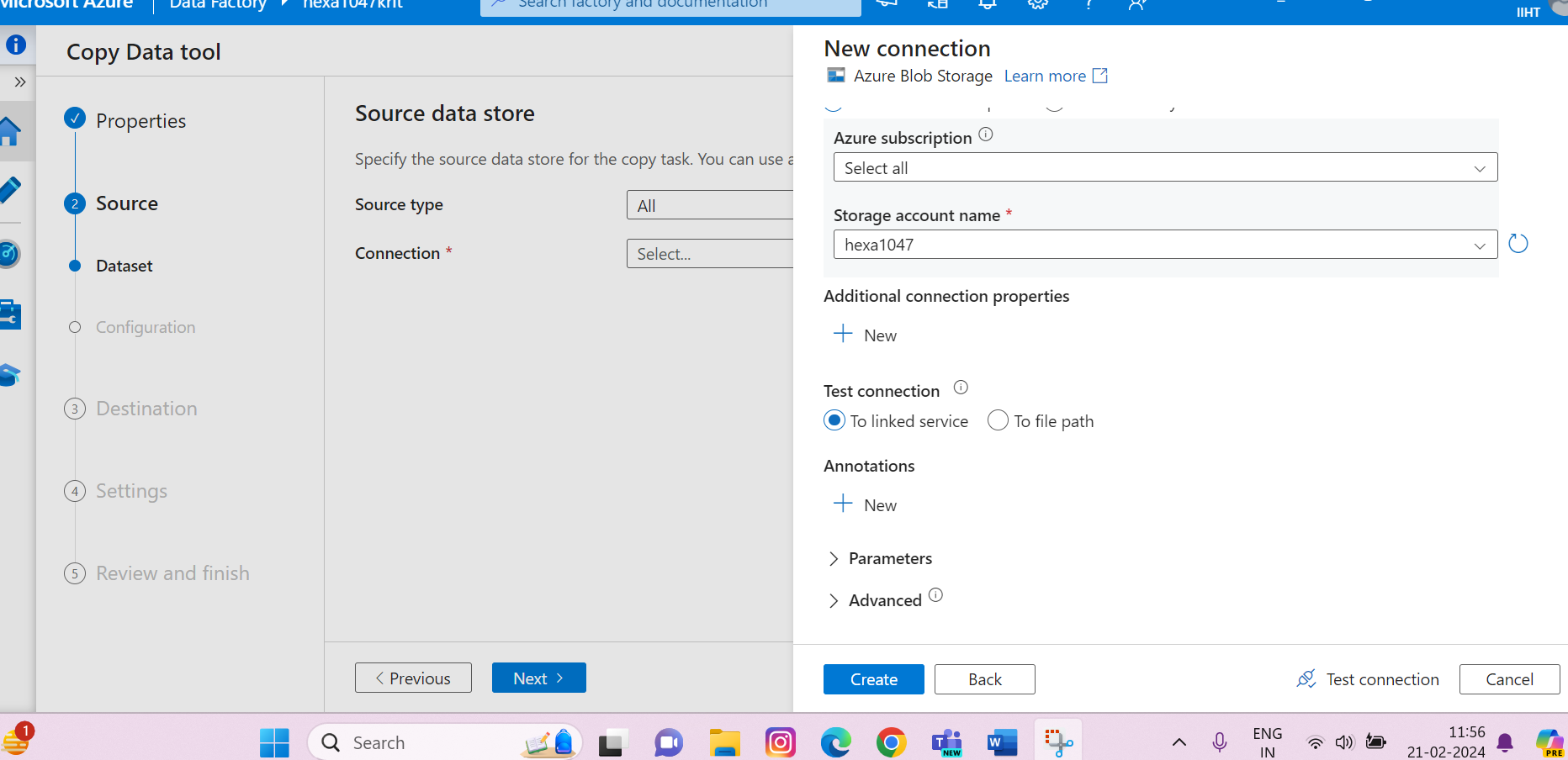
In data pipeline to move data from both on premise and cloud source data stores to a centralization data store in the cloud for further analysis copy activity is performed. Basically, it copies data from a source data store to a sink data store.

**Performing copy activity in azure databricks**

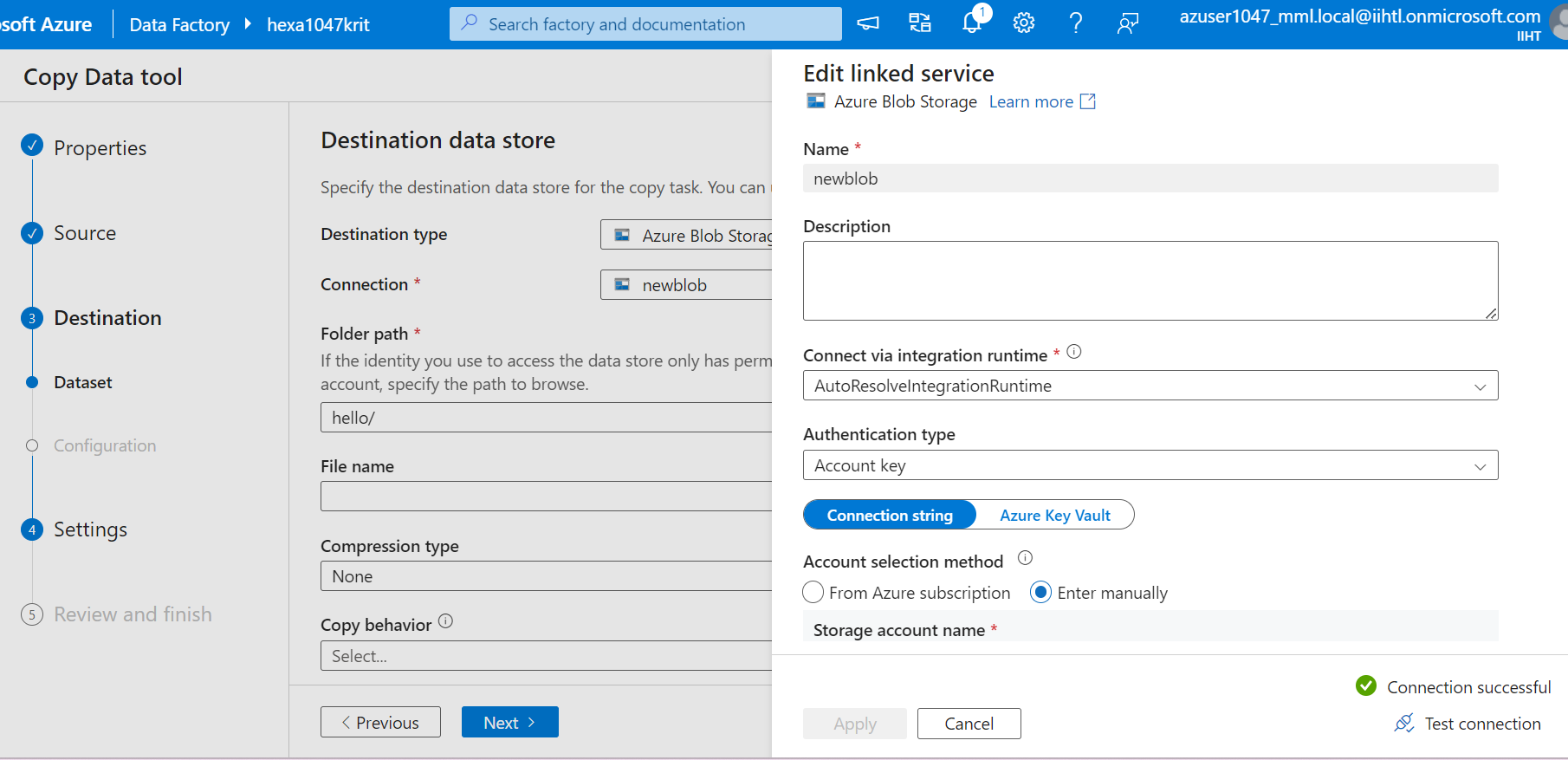
**Step1: Create a datafactory**

**Step 2: Launch datafactory and click on ingest.**

**Step 3: Create a dataset by connecting to your storage account where the file is stored.**

****

**Step 4: Connect to the destination storage account and test connection:**

****

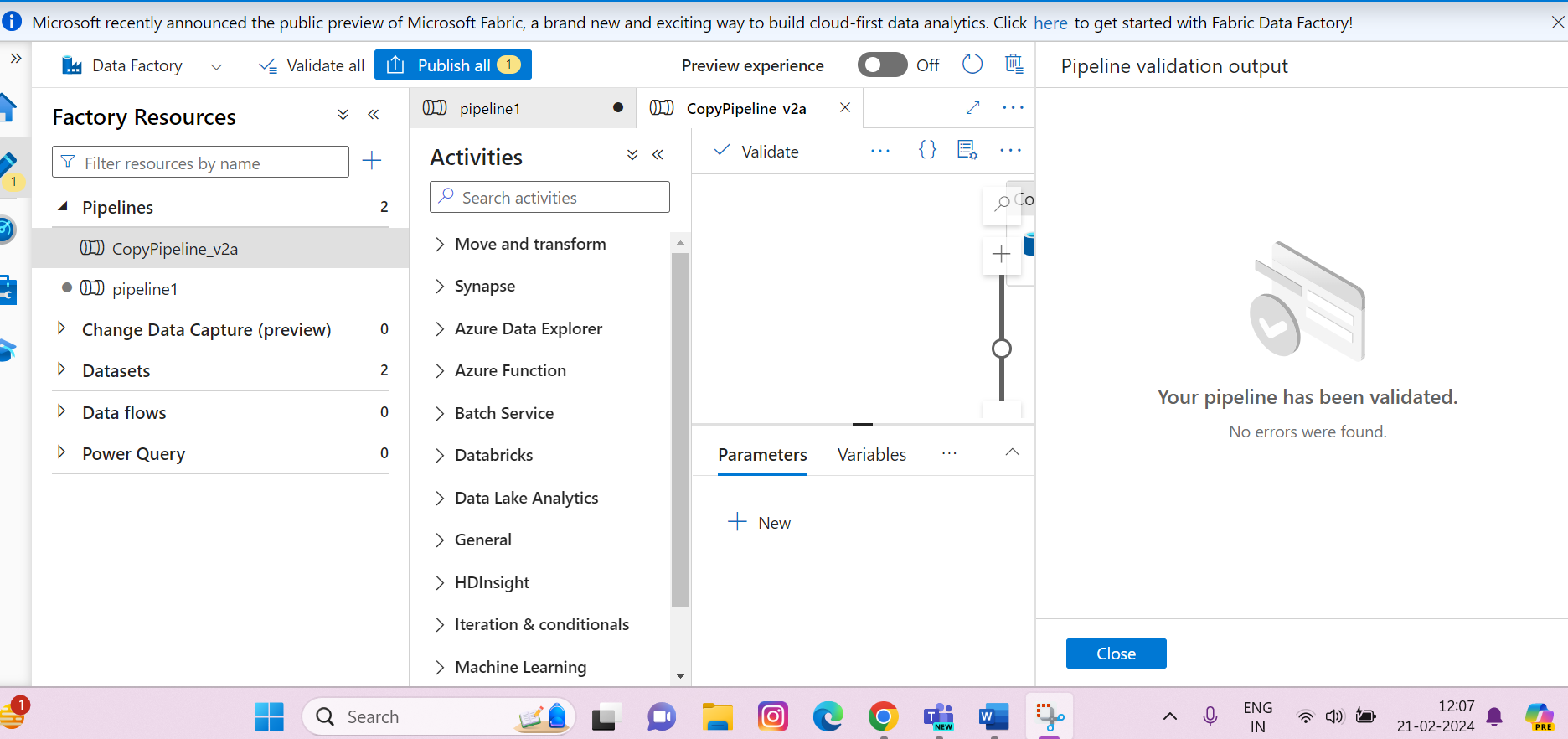
**Step 5: Create a pipeline and then check for all the details.**

**Step 6: Review all the details and click finish to create the data set and pipelines.**

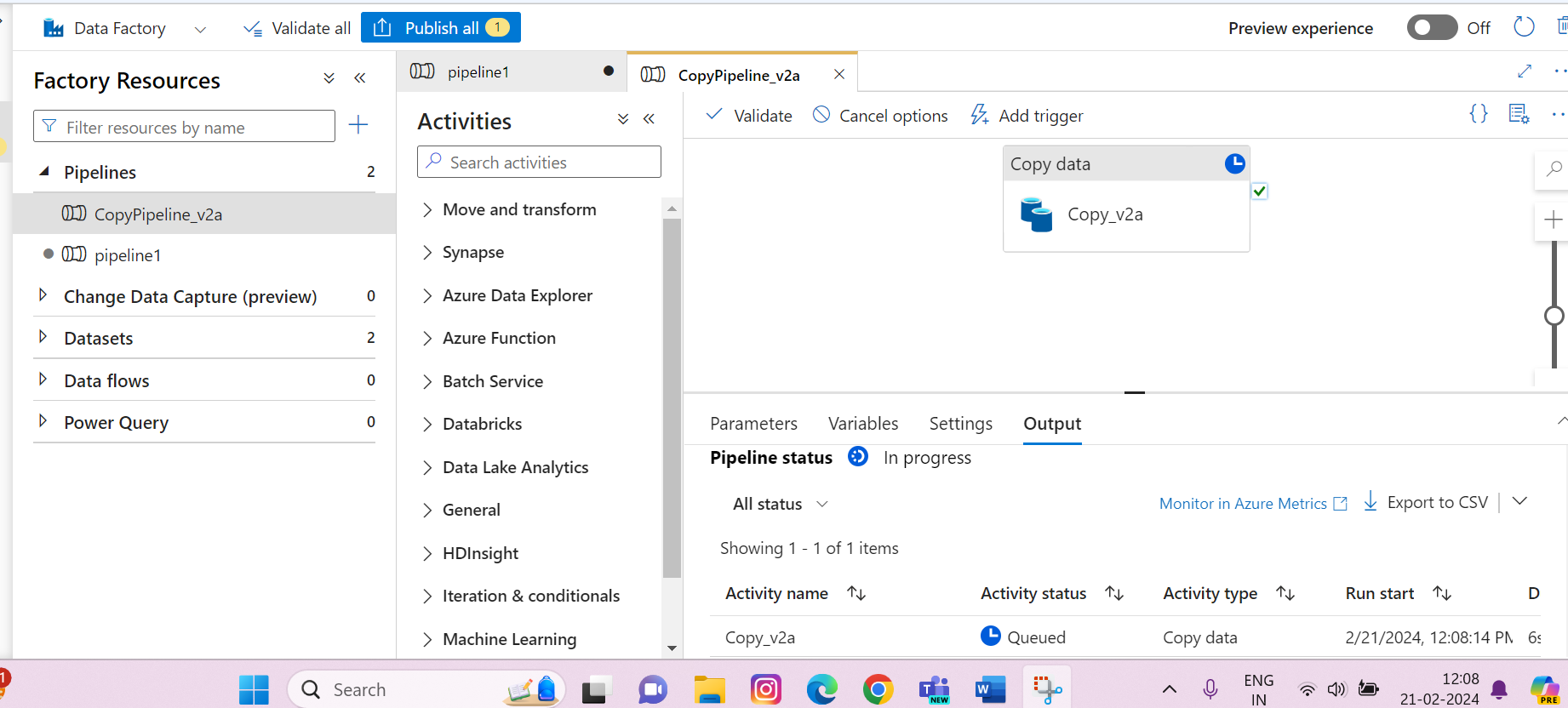
**A screenshot of a computer

Description automatically generated**

**Step 6: go to author and select the pipeline you created and performing validation.**

****

**Step 7: perform debugging.**

****

**Step 8: Output should be succeeded that means successfully copy activity has been performed.**

**A computer screen with a blue background

Description automatically generated**