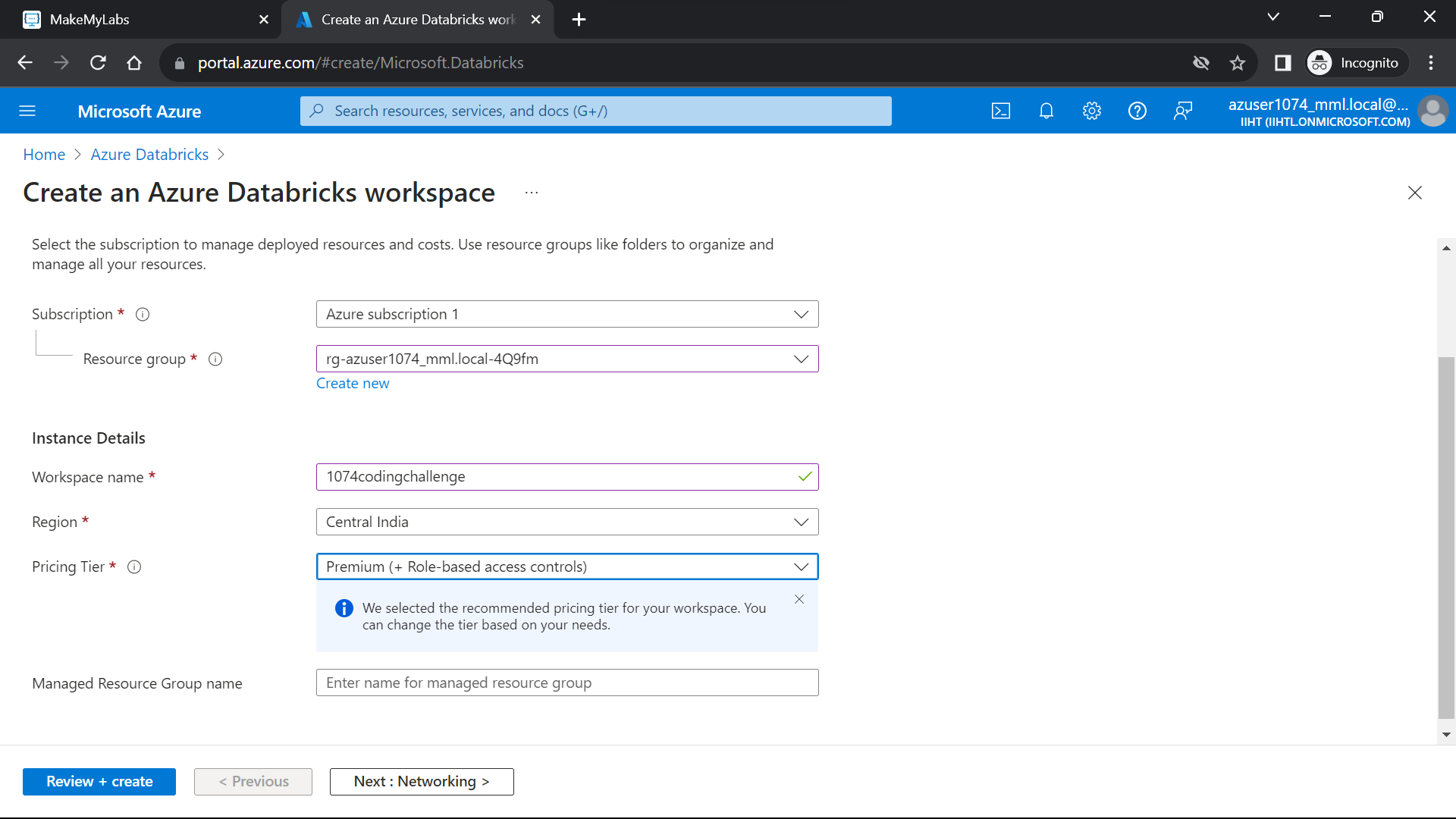
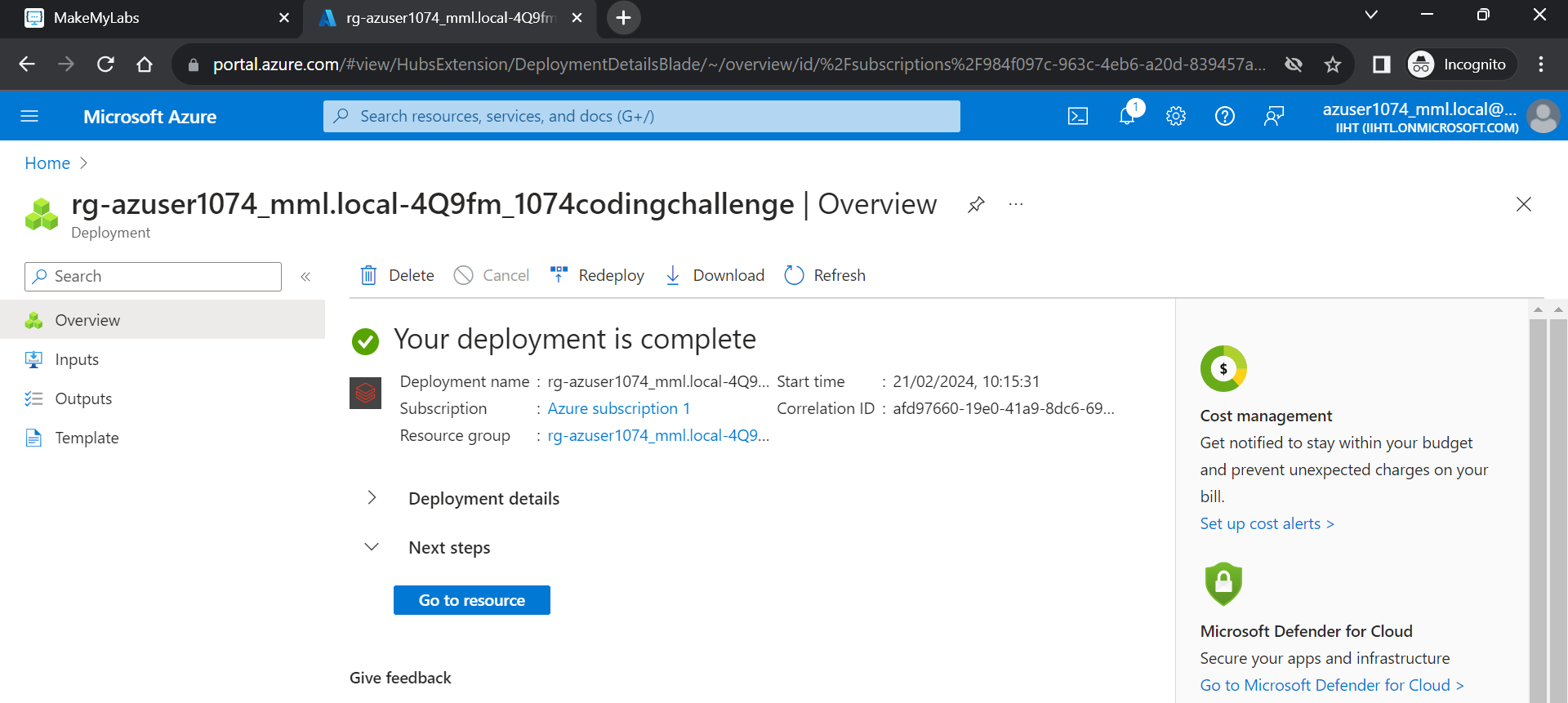
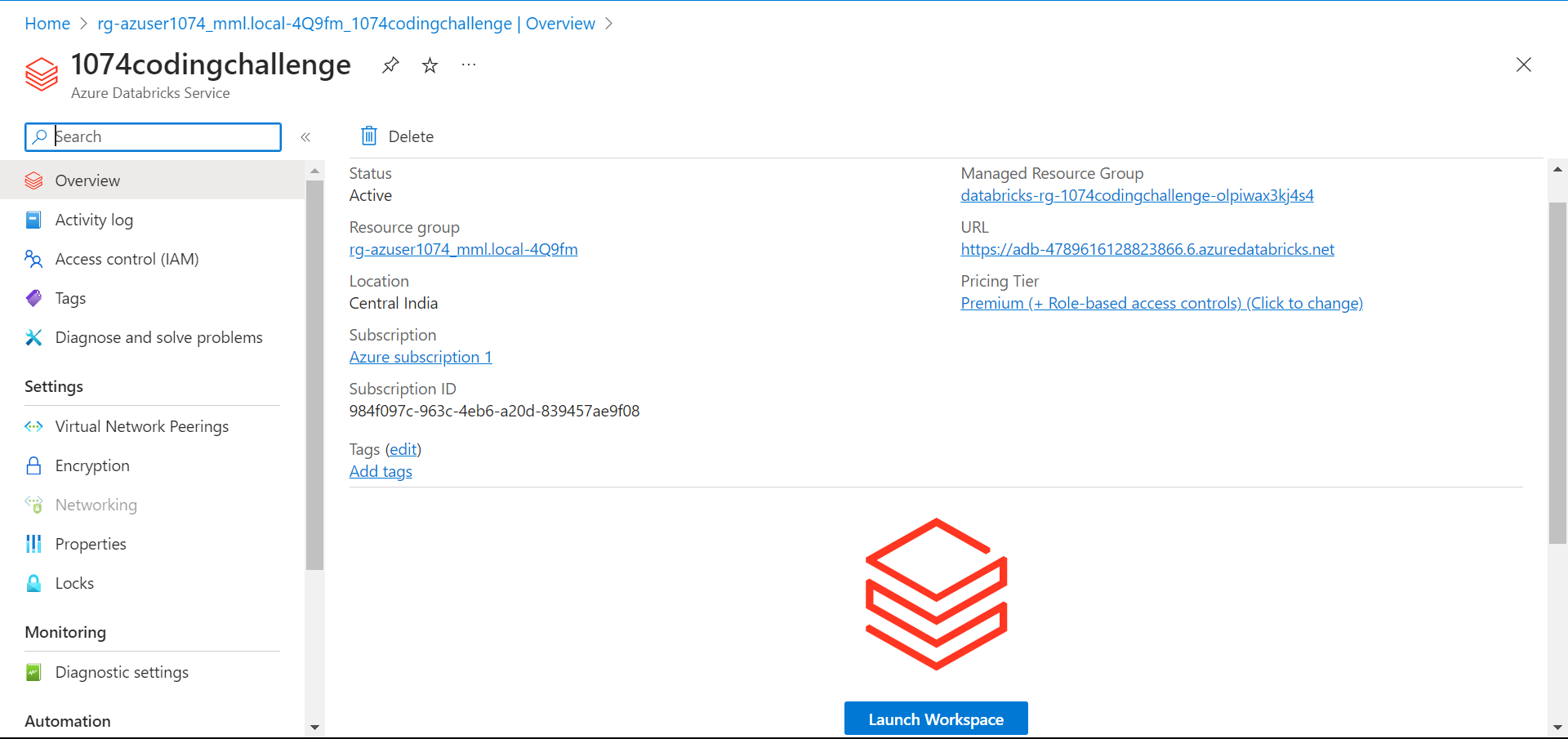
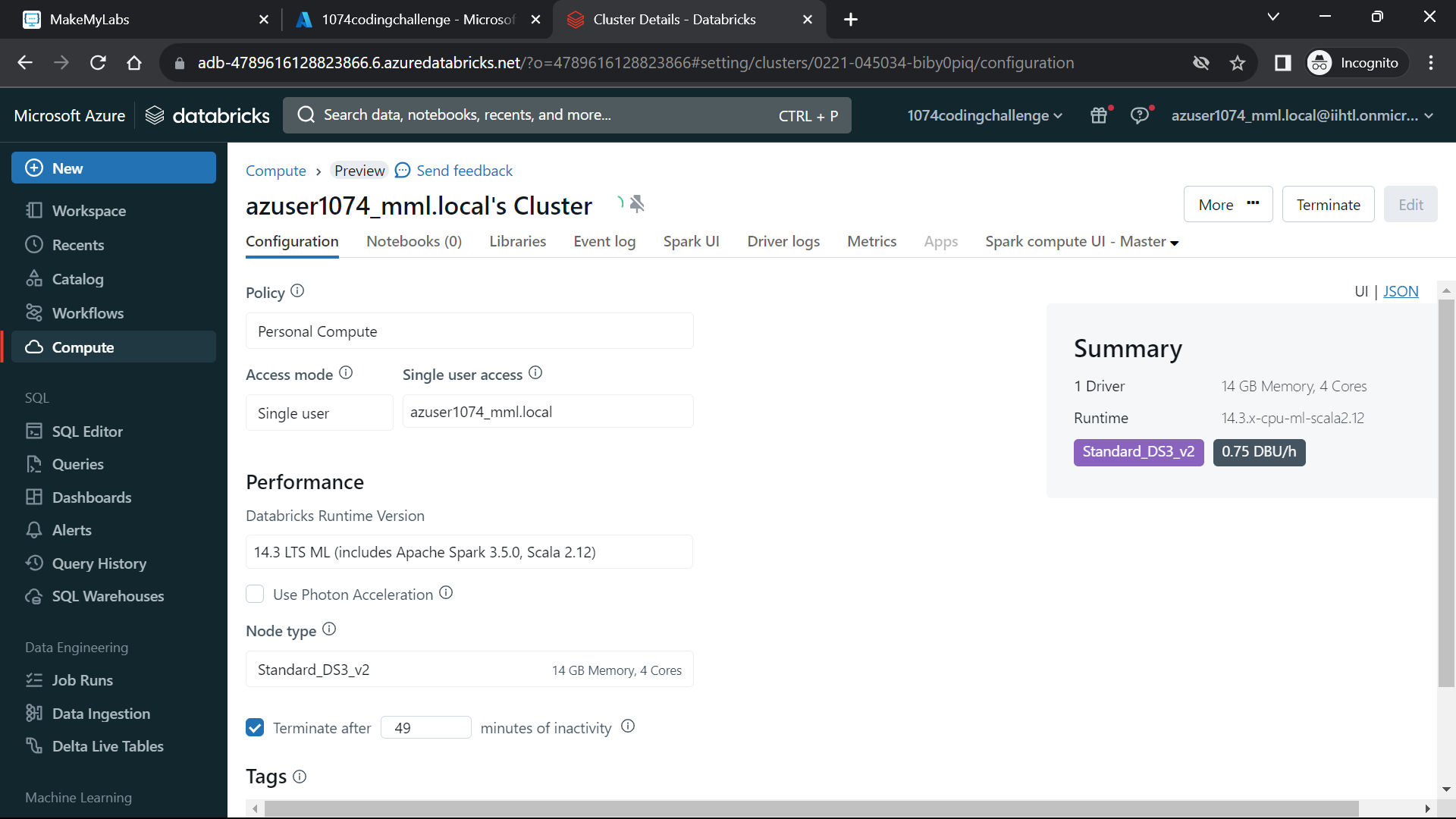
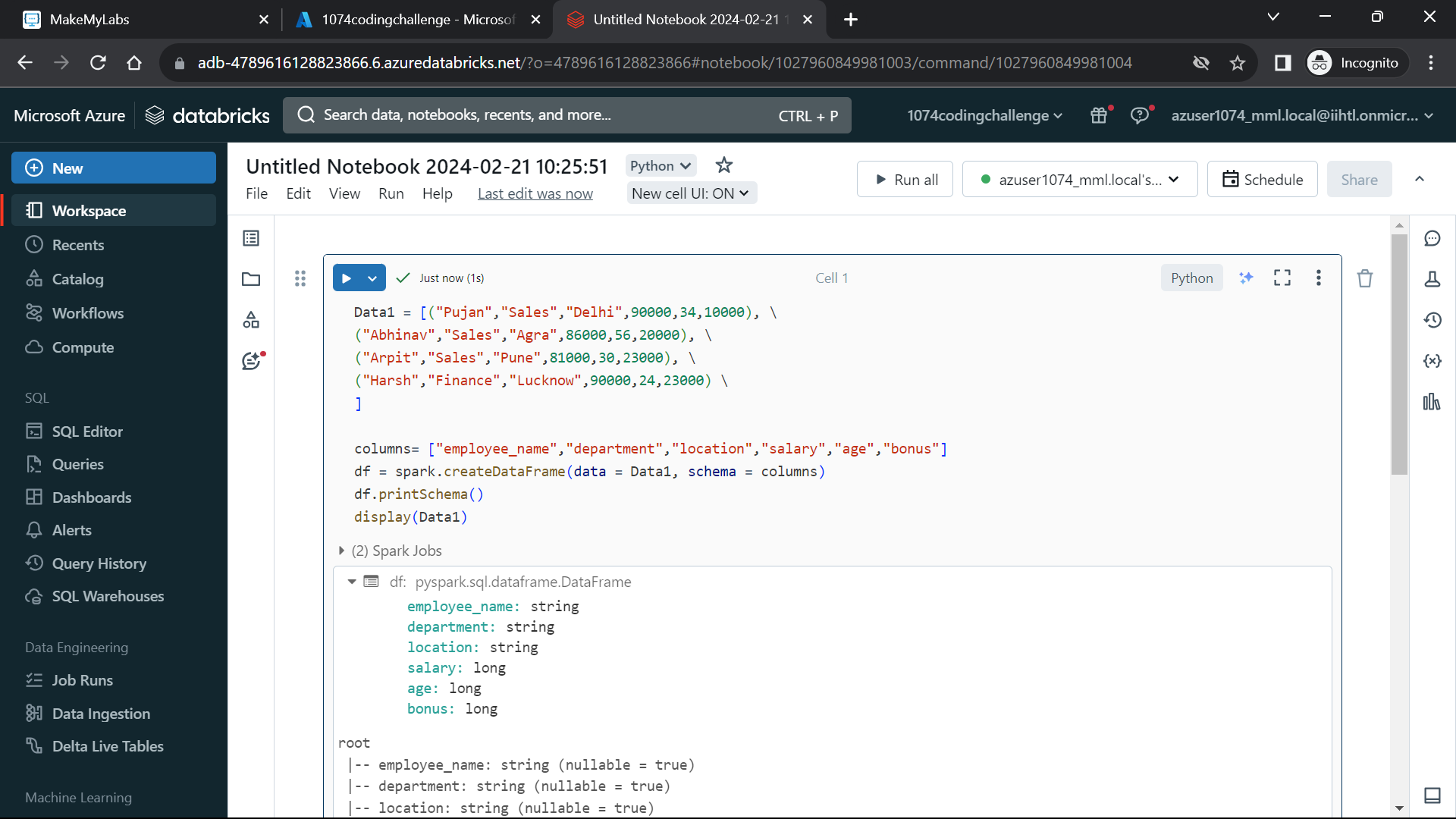
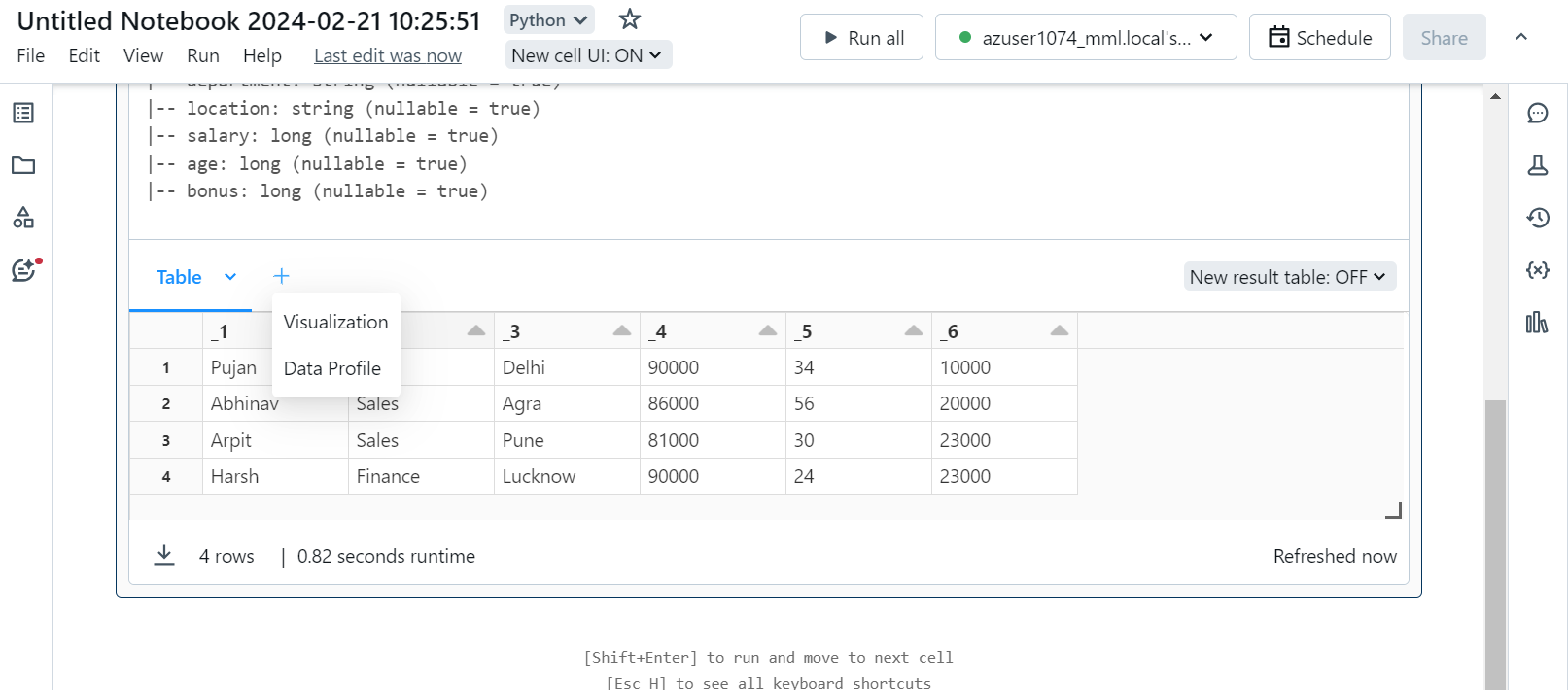
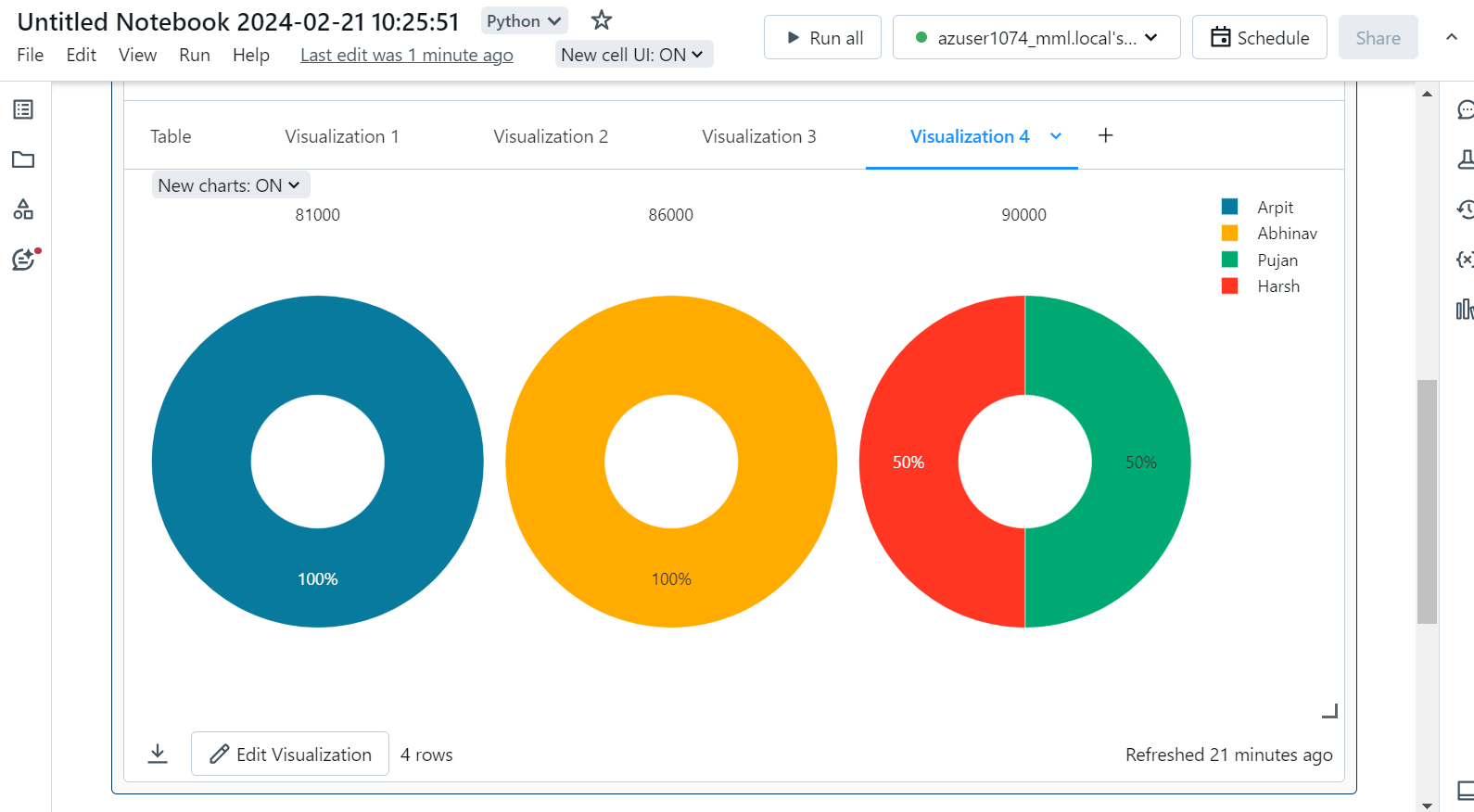
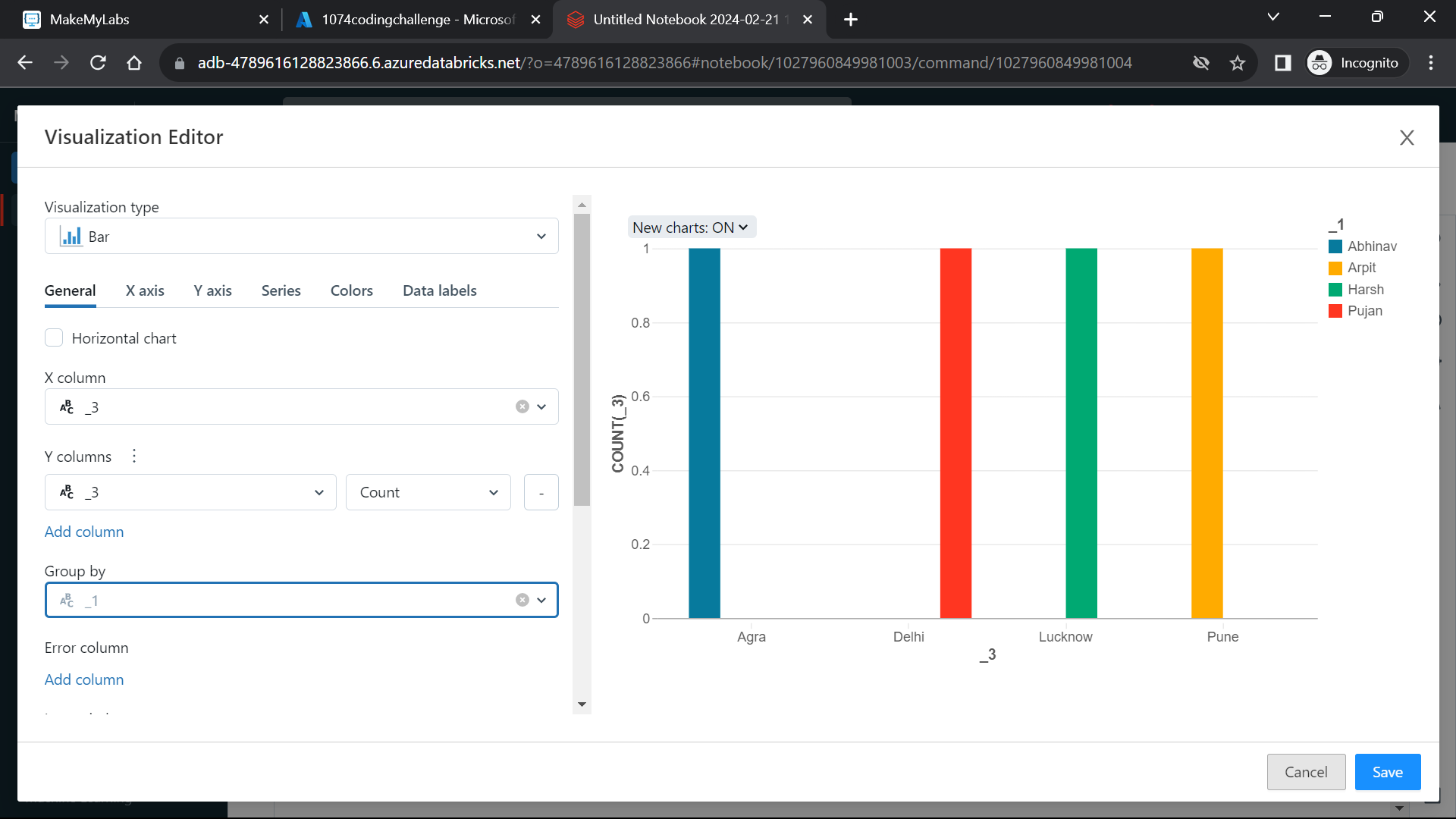
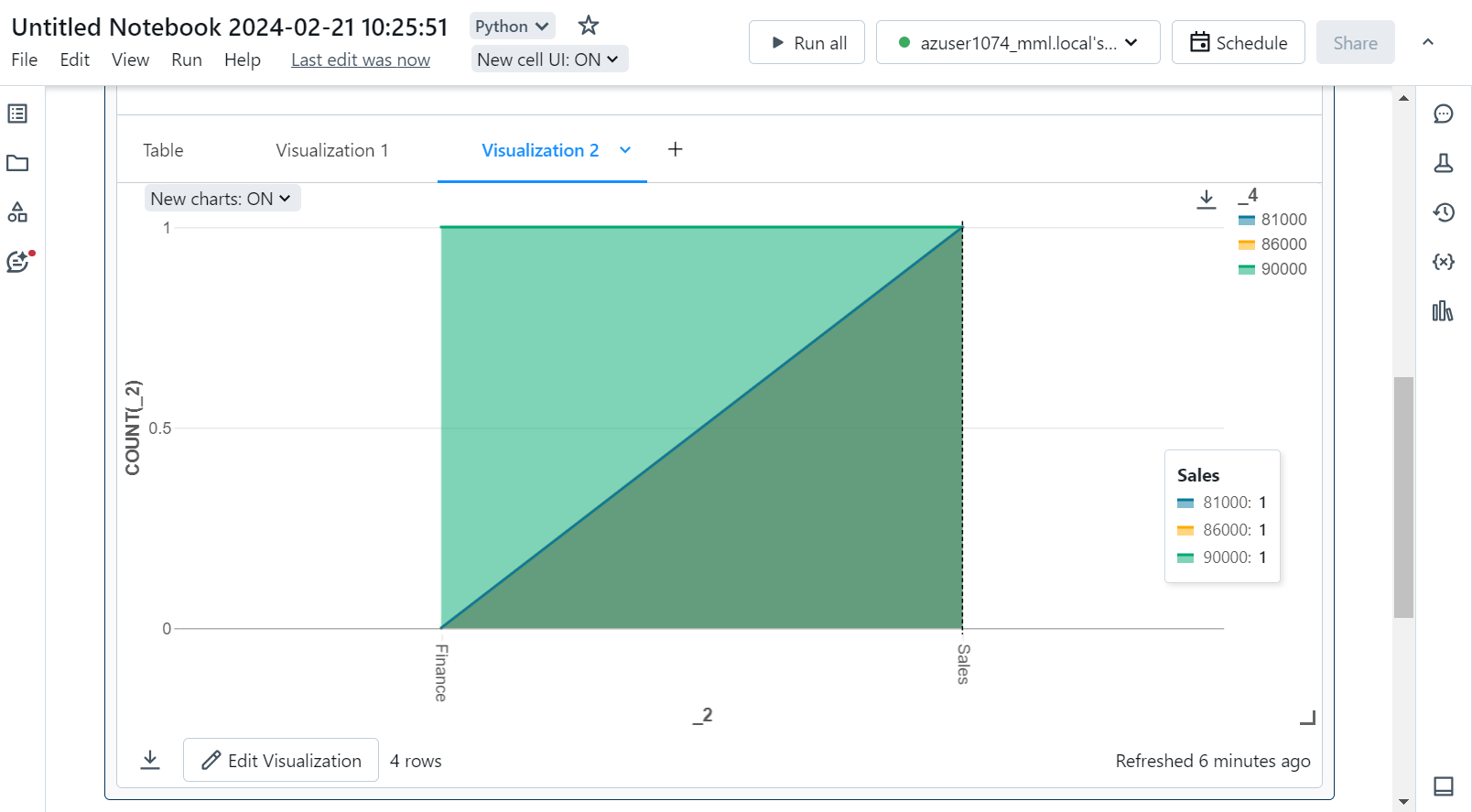
**PUJAN RASTOGI**

**AZURE DATABRICKS CODING CHALLENGE 21 February 2024**

**Exploratory data analysis (EDA) in Databricks &Visualizing data in Databricks**

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process that involves summarizing, visualizing, and understanding the main characteristics of a dataset. Databricks, especially in Azure Databricks, provides a powerful platform for performing EDA and visualizing data.

**** ****

****       

**Explain Overview of 3 level namespace and creating Unity Catalog objects.**

A three-level namespace in Unity Catalogue consists of three levels of hierarchy — catalog, schema, and objects. Catalogs are like databases, schemas are like folders or directories, and objects can be tables, views, functions, or other entities.

In Unity Catalog, the hierarchy of primary data objects flows from metastore to table or volume:

Metastore: The top-level container for metadata. Each metastore exposes a three-level namespace (catalog.schema.table) that organizes your data.

Catalog: The first layer of the object hierarchy, used to organize your data assets.

Schema: Also known as databases, schemas are the second layer of the object hierarchy and contain tables and views.

Tables, views, and volumes: At the lowest level in the data object hierarchy are tables, views, and volumes. Volumes provide governance for non-tabular data.

Models: Although they are not, strictly speaking, data assets, registered models can also be managed in Unity Catalog and reside at the lowest level in the object hierarchy.

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

Azure Data Factory is a fully managed, serverless data integration service that enables you to create, schedule, and manage data pipelines. Data pipelines are workflows that can move data from various source systems to different destination systems, transforming and processing the data along the way.

ADF can be used for:

* Supporting data migrations
* Getting data from a client's server or online data to an Azure Data Lake
* Carrying out various data integratid processes
* Integrating data from different ERP systems and loading it into Azure
* Synapse for reporting

ADF perform three steps:

* Connect and Collect
* Transform and Enrich
* Publish

**Copy Activity:**

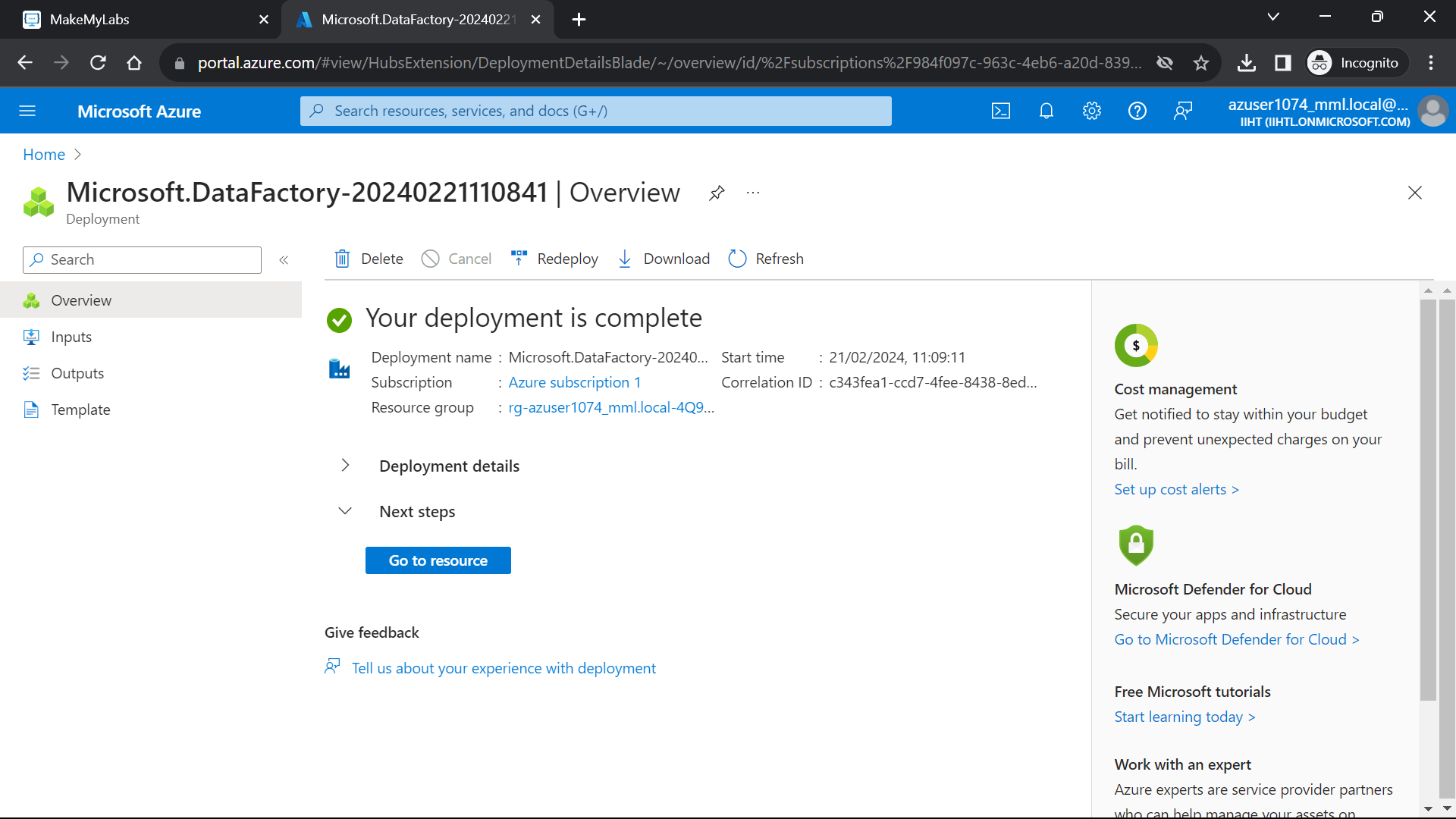
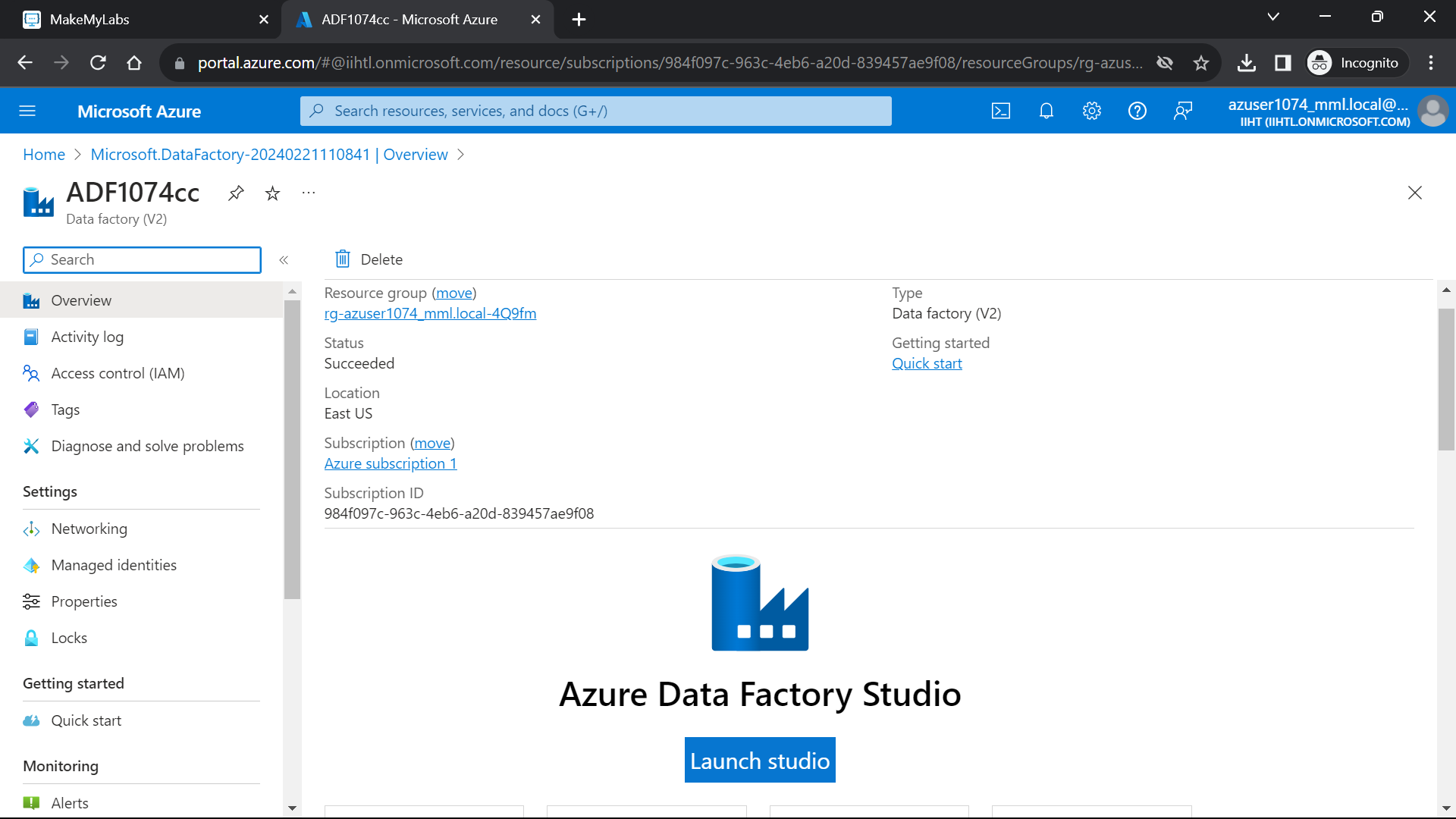
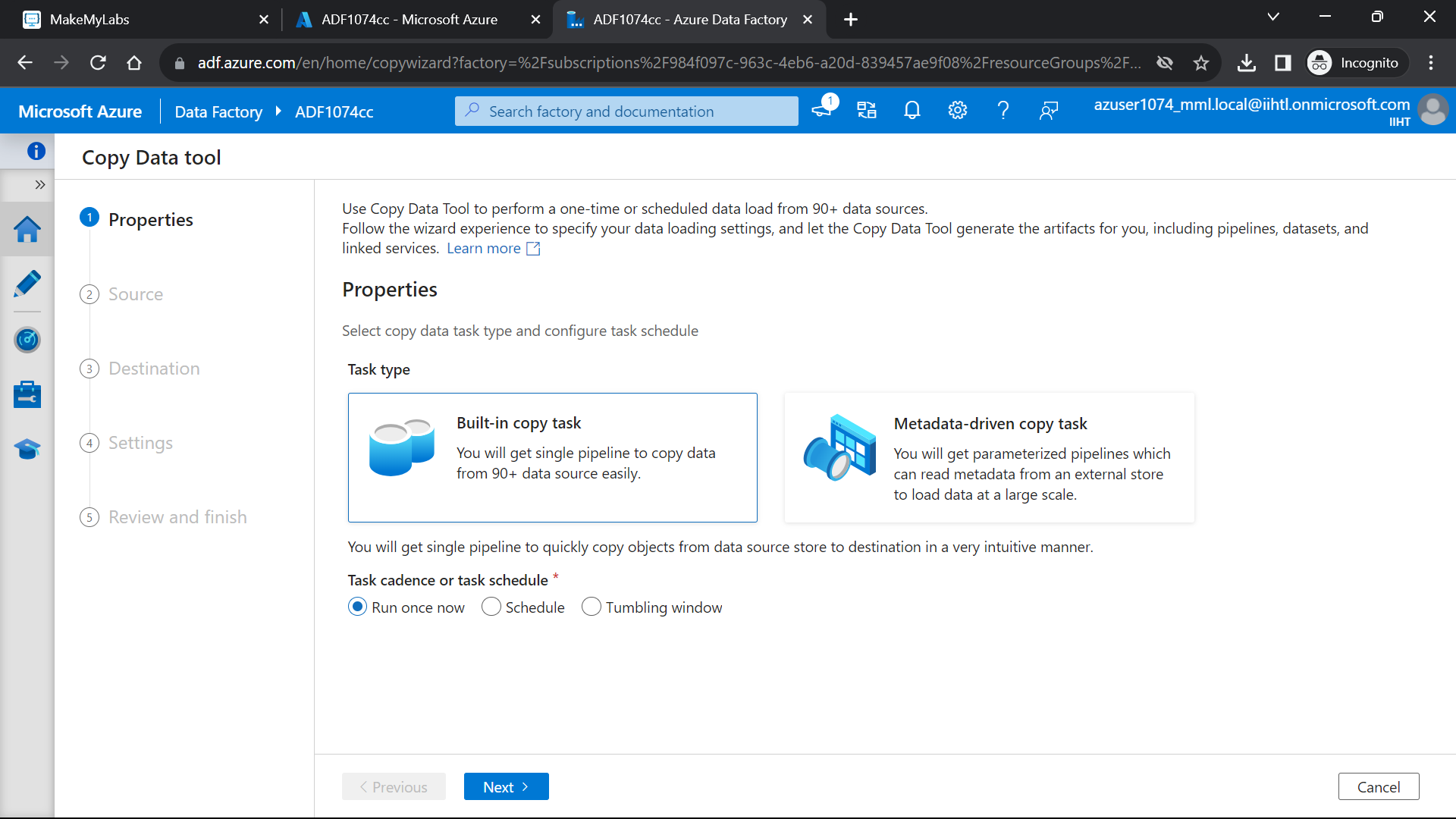
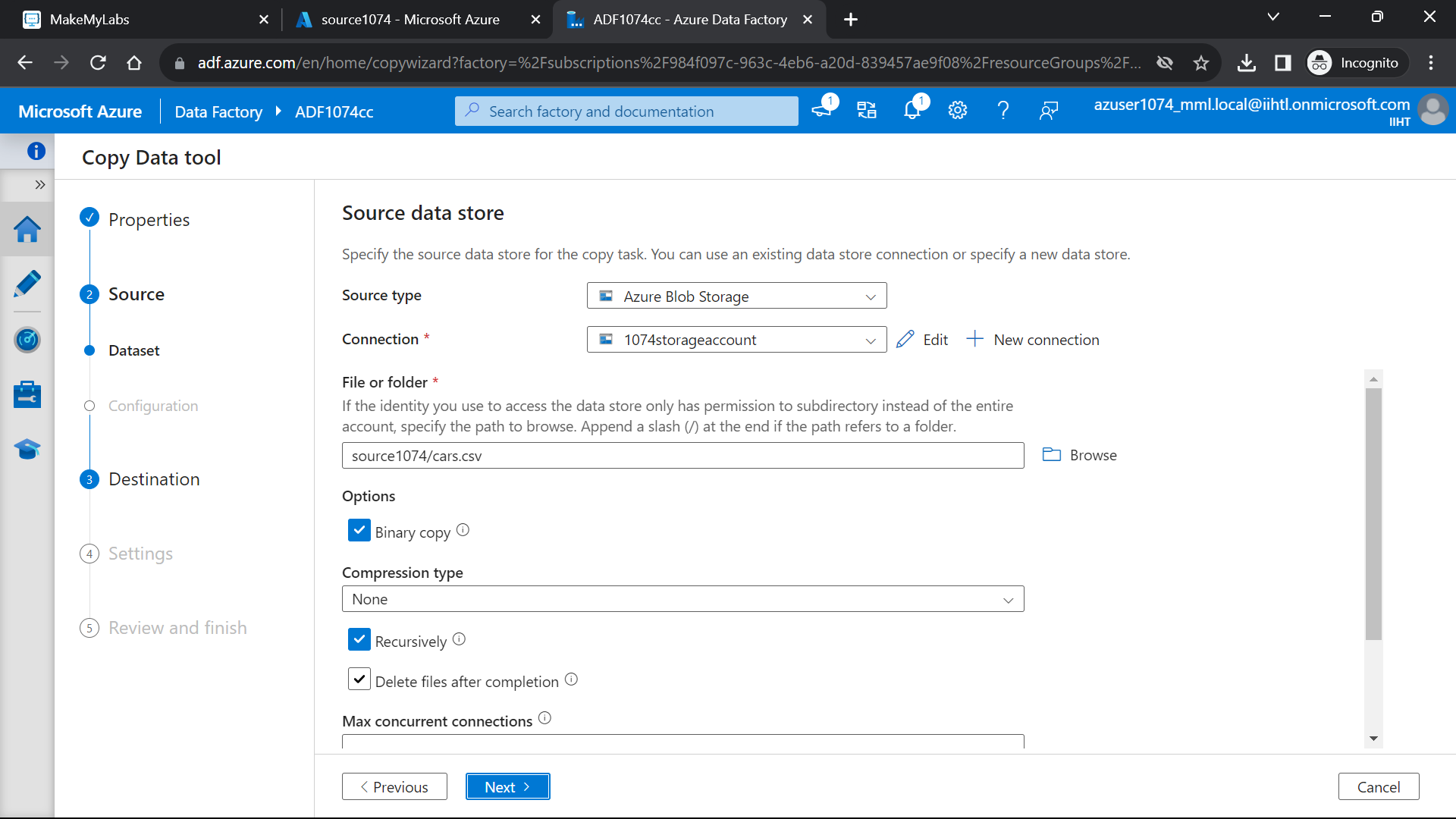
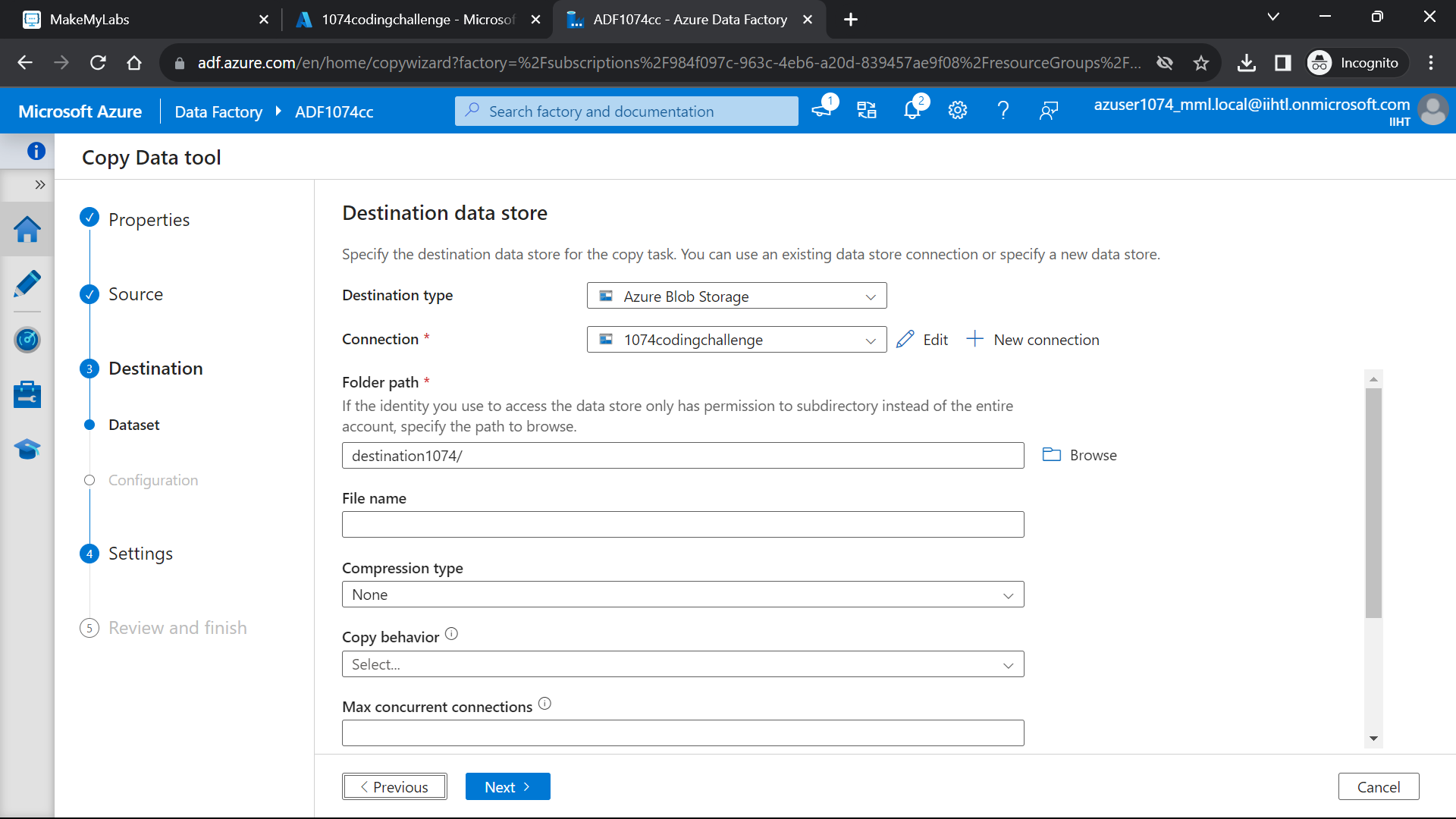
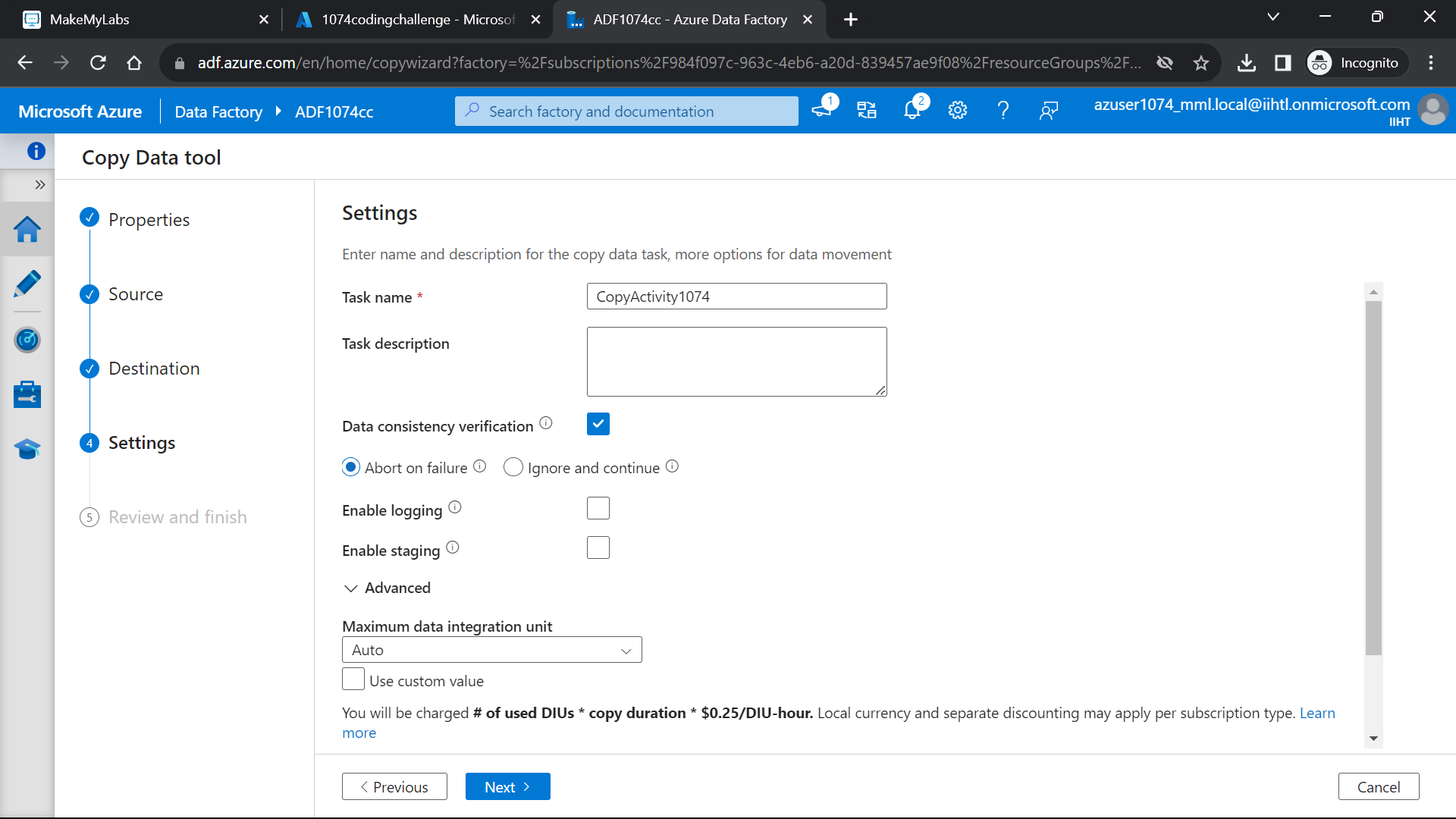
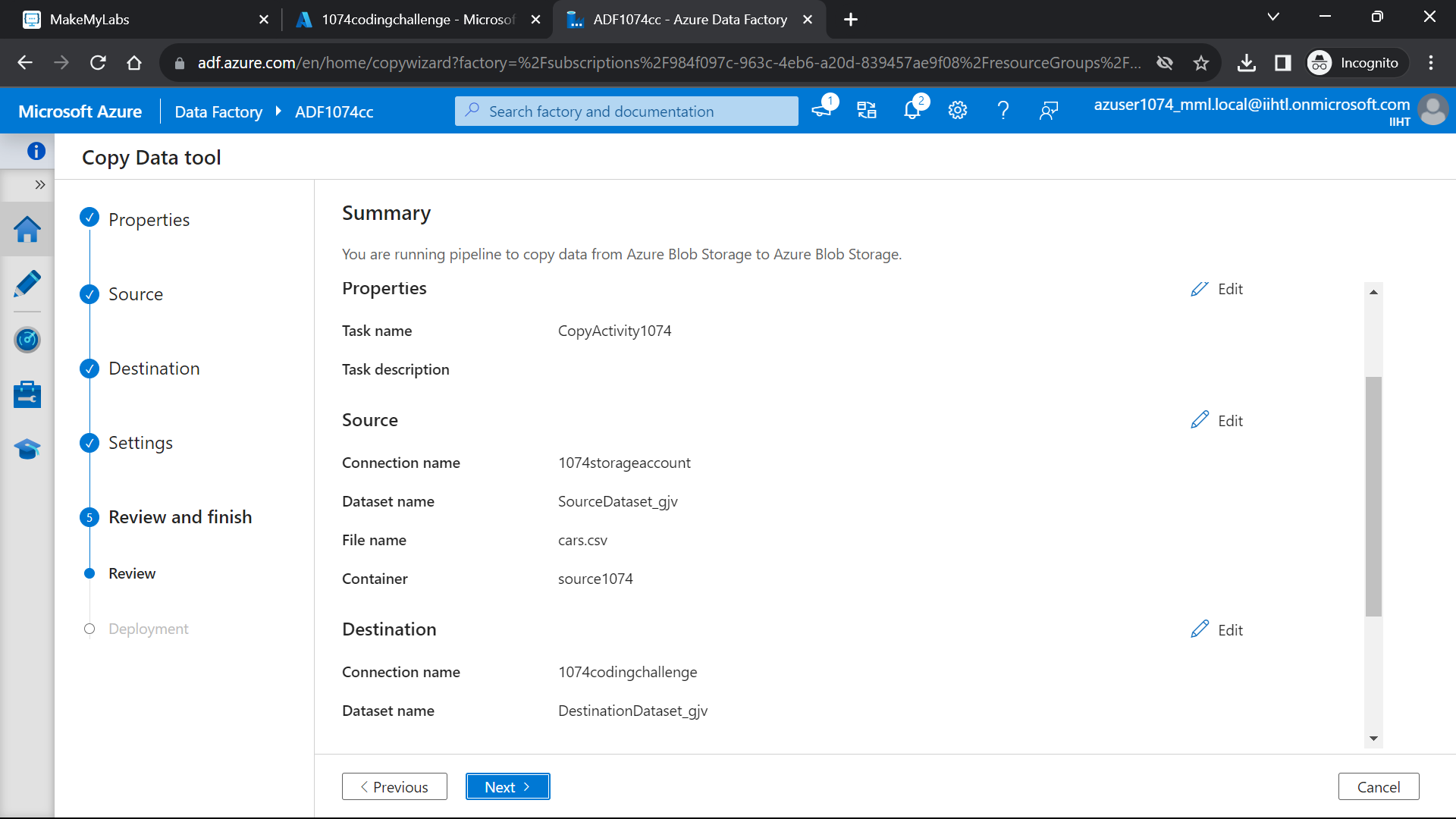
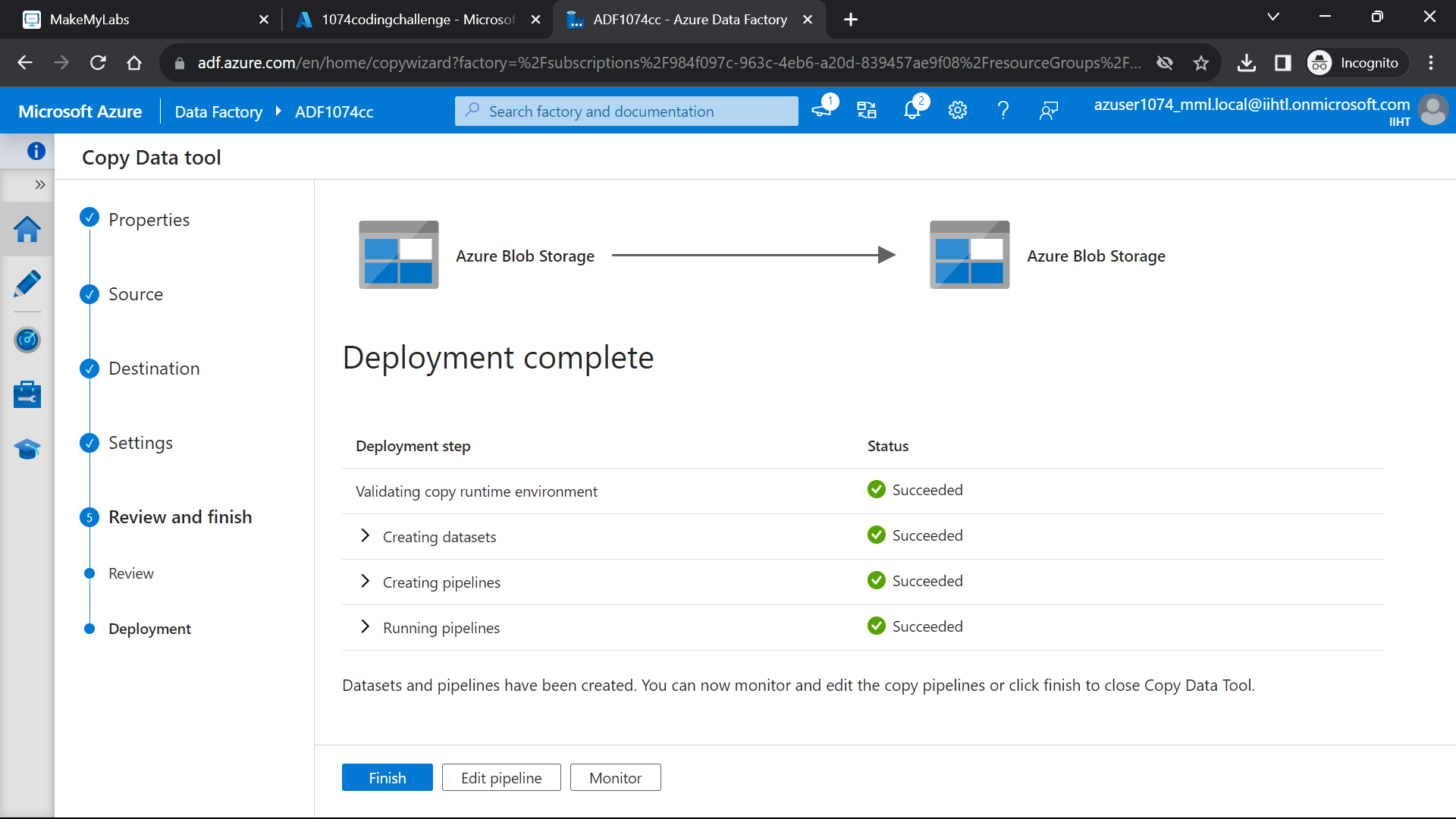
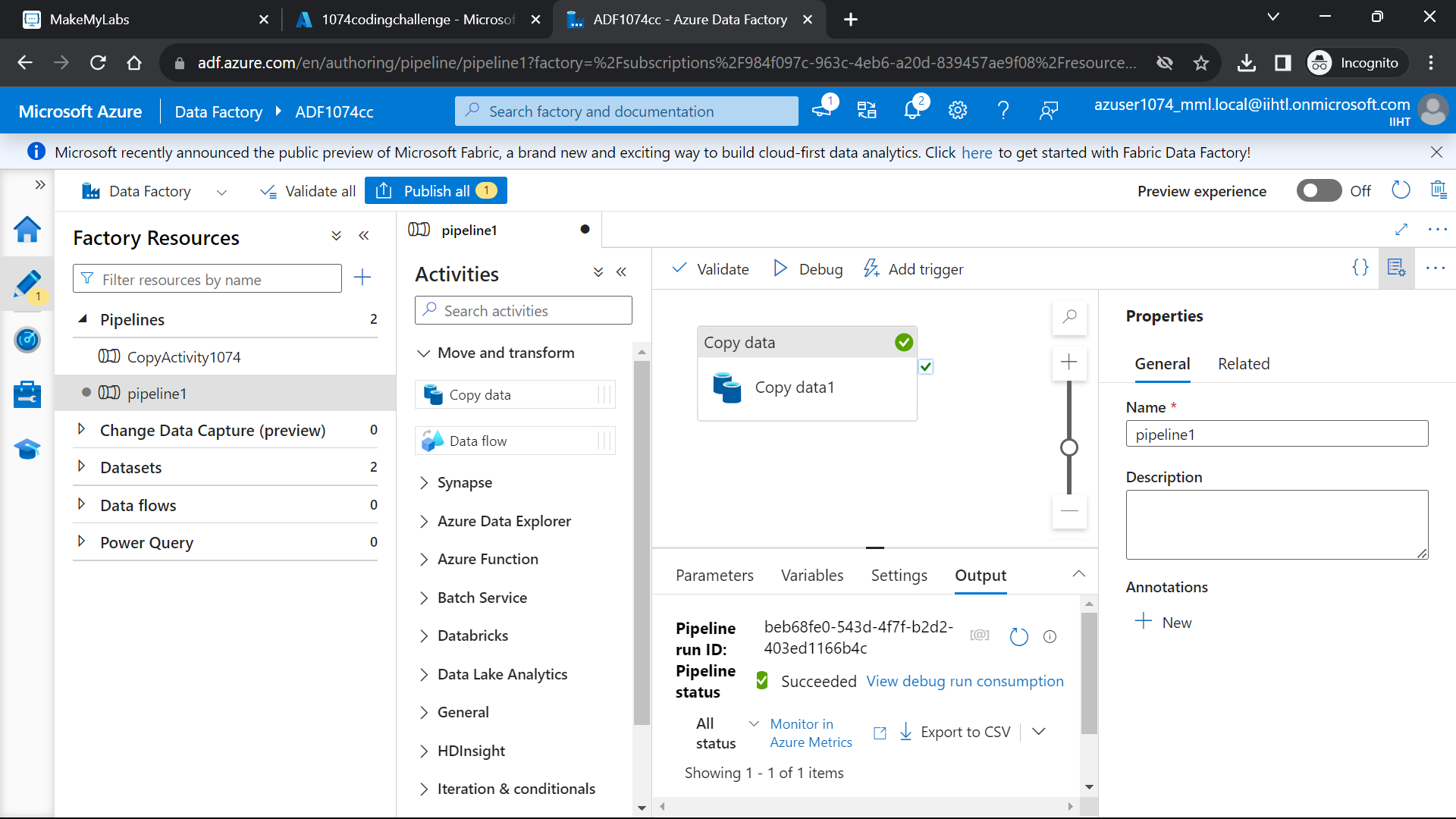
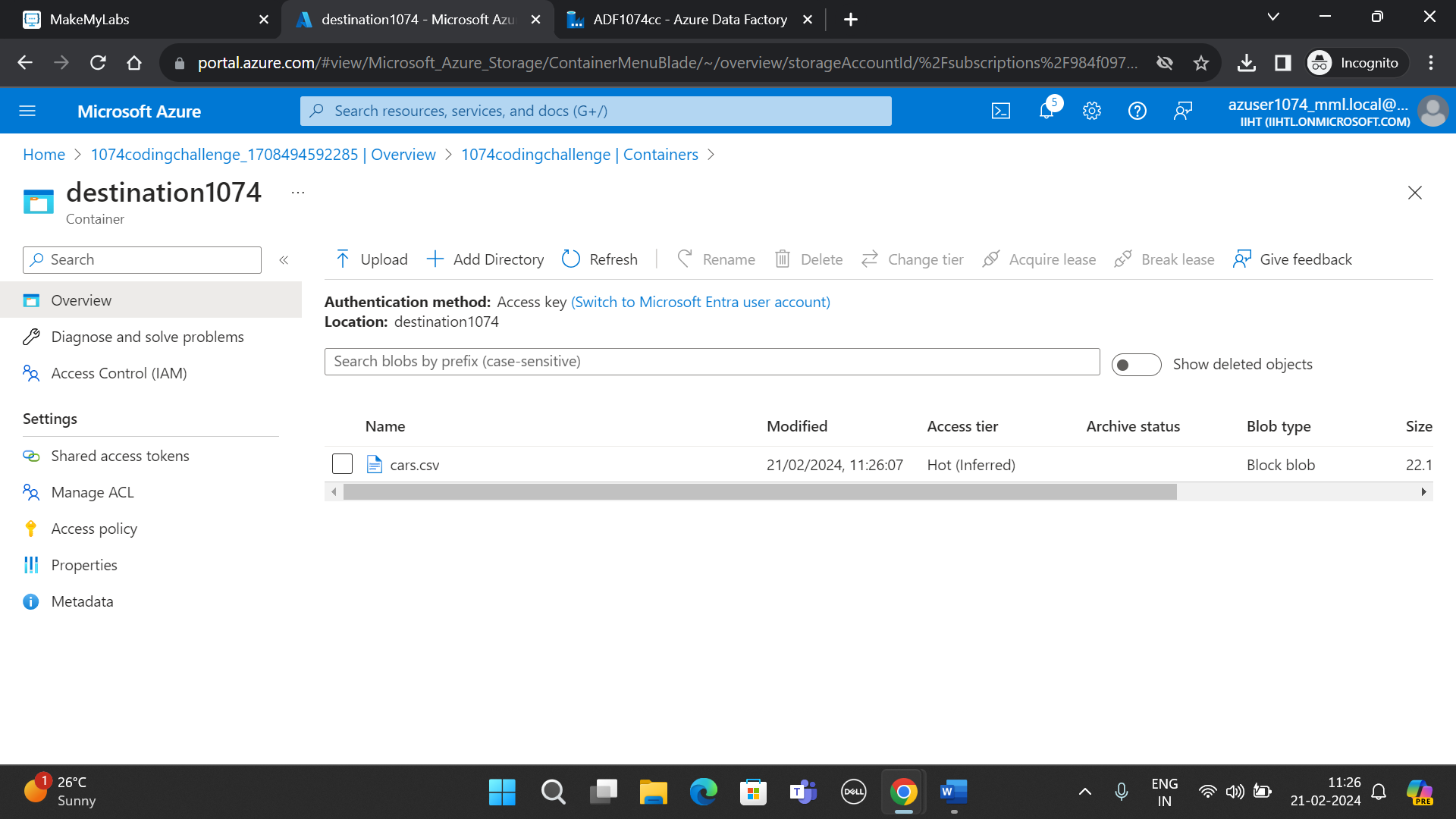
Copy Activity is a specific activity type within Azure Data Factory that is used for moving data from a source to a destination. It supports a variety of source and sink (destination) data stores. Some key points about Copy Activity:

Source and Destination: Copy Activity can copy data between different types of data stores, including Azure Blob Storage, Azure SQL Database, Azure Data Lake Storage etc.

Data Movement: It supports efficient data movement using parallel copy, making it suitable for handling large volumes of data.

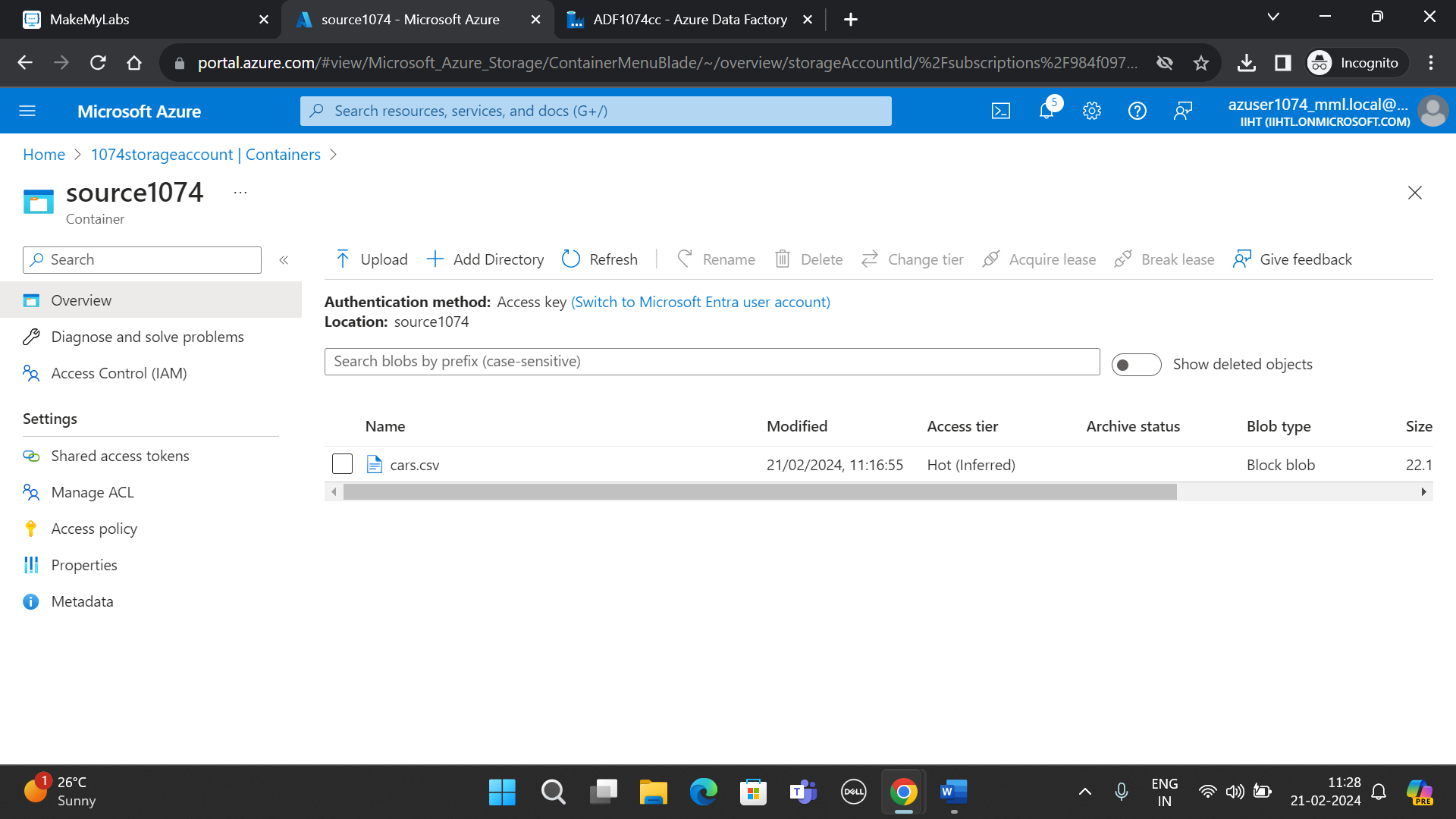
Transformation: Copy Activity also allows simple data transformations during the copy process, such as column mapping and data type conversions.

Monitoring: You can monitor the progress of Copy Activity runs, view detailed logs, and troubleshoot any issues that may arise during the data movement process.

YOU CAN SEE THE TIME IN WHICH DATA IS COPIED FROM SOURCE TO DESTINATION. 11.26 AM THE DATA IS COPIED TO DESTINATION CONTAINER.

BELOW IS THE SOURCE CONTAINER:



YOU CAN CHECK THE TIME OF FILE UPLOAD FOR CONFIRMATION.