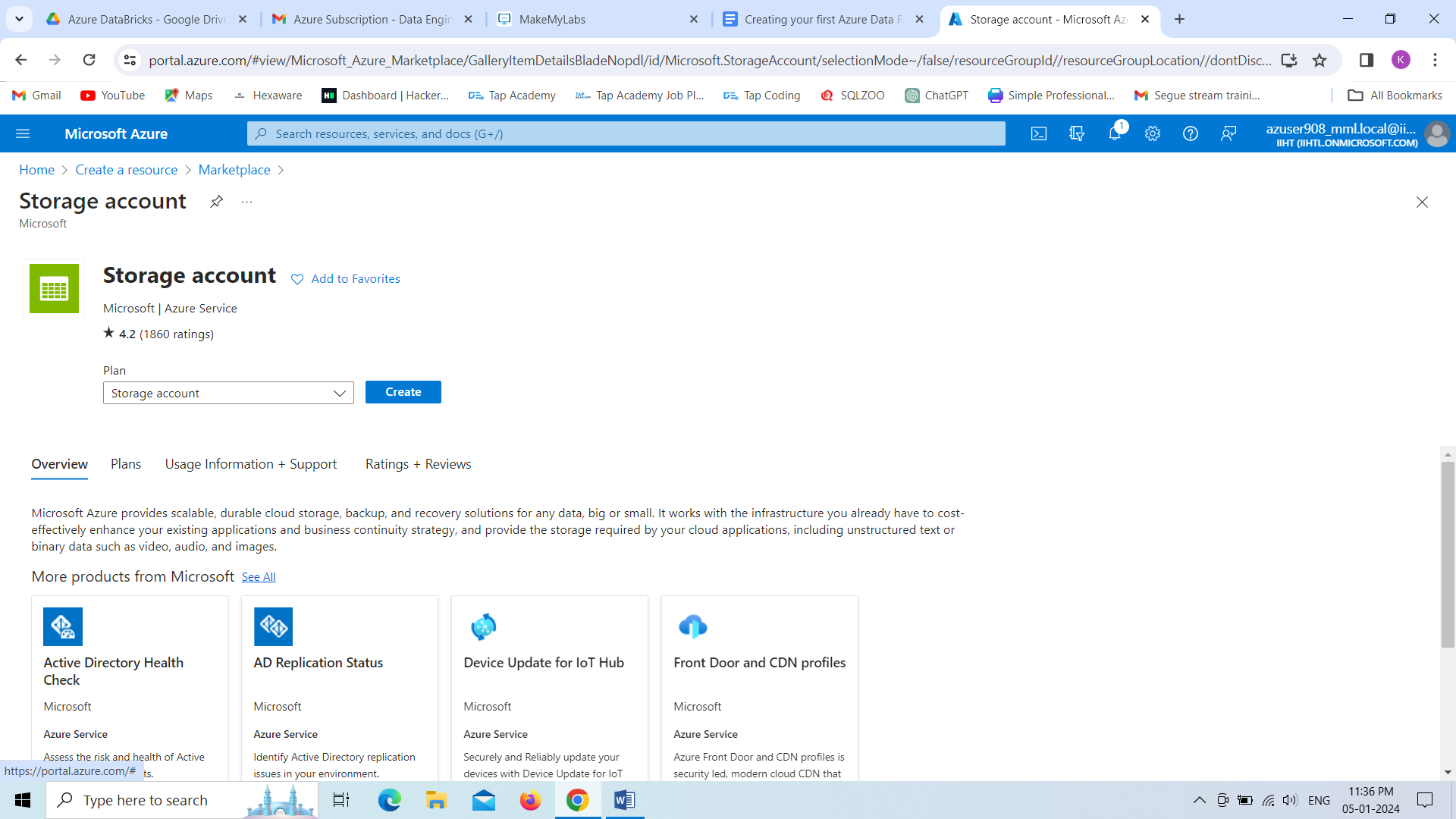
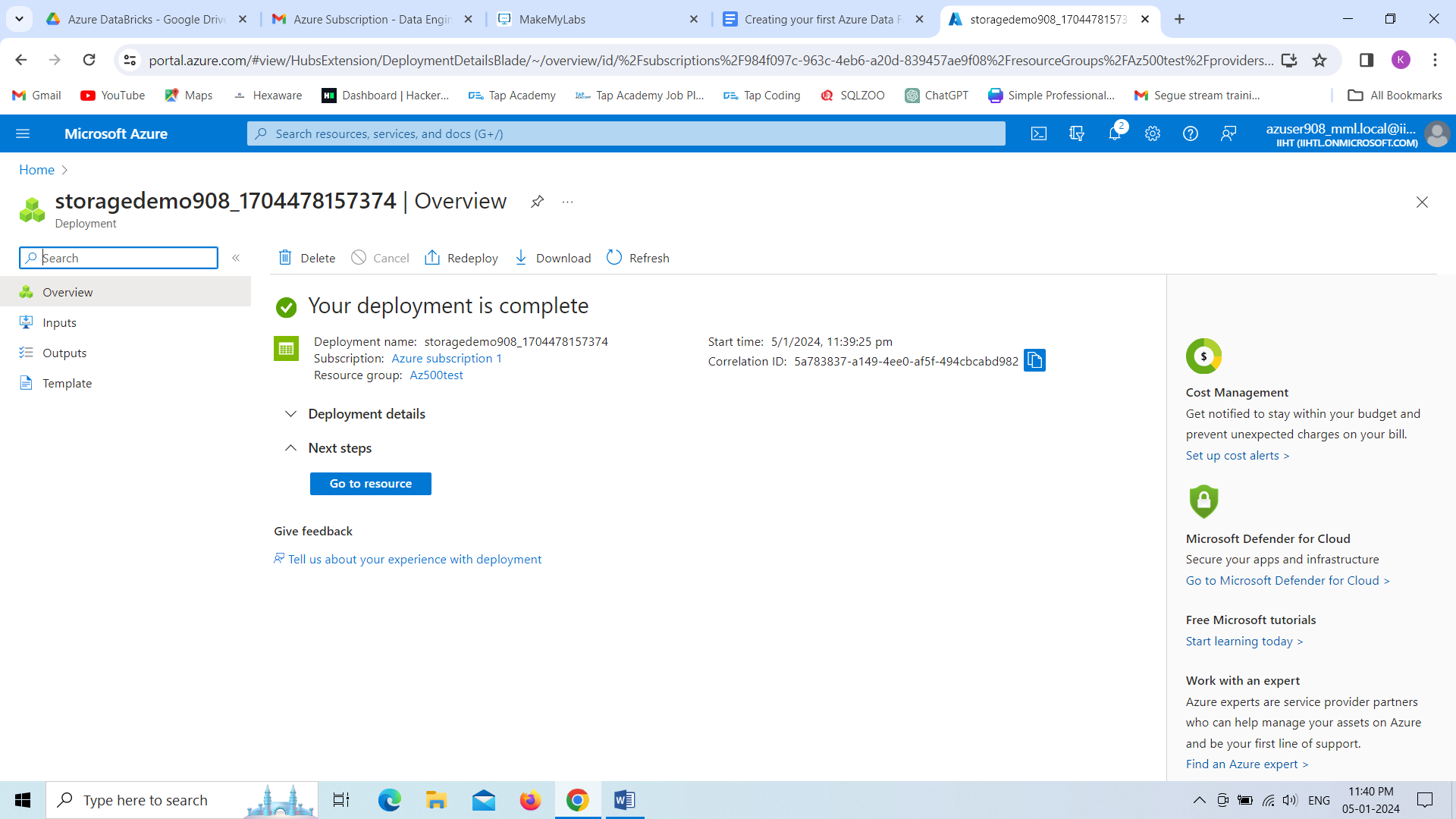
**Assignment(05-01-2024)**

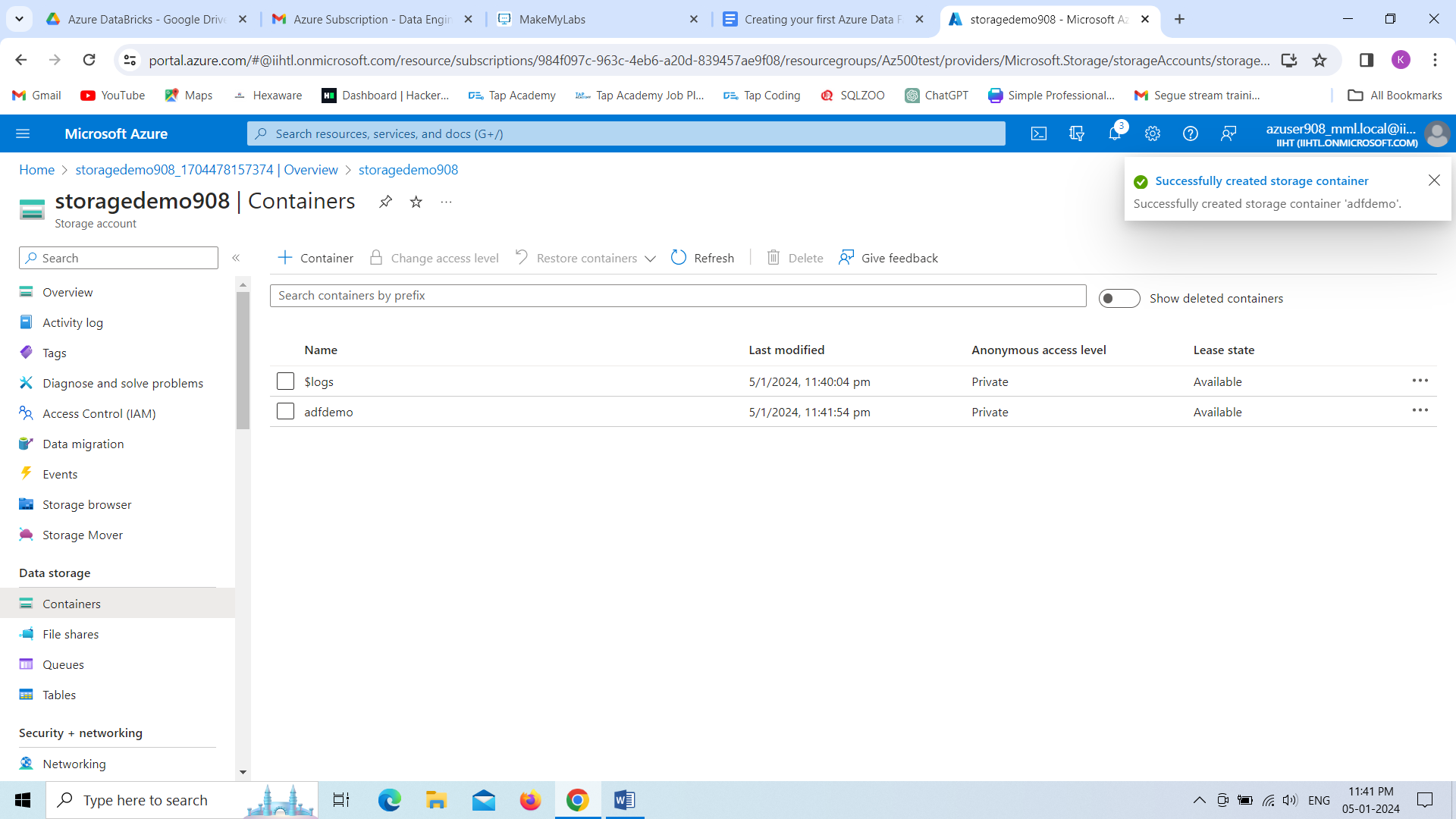
**Creating your first Azure Data Factory pipeline**

First we need to create the storage account,let’s see how can we do this.

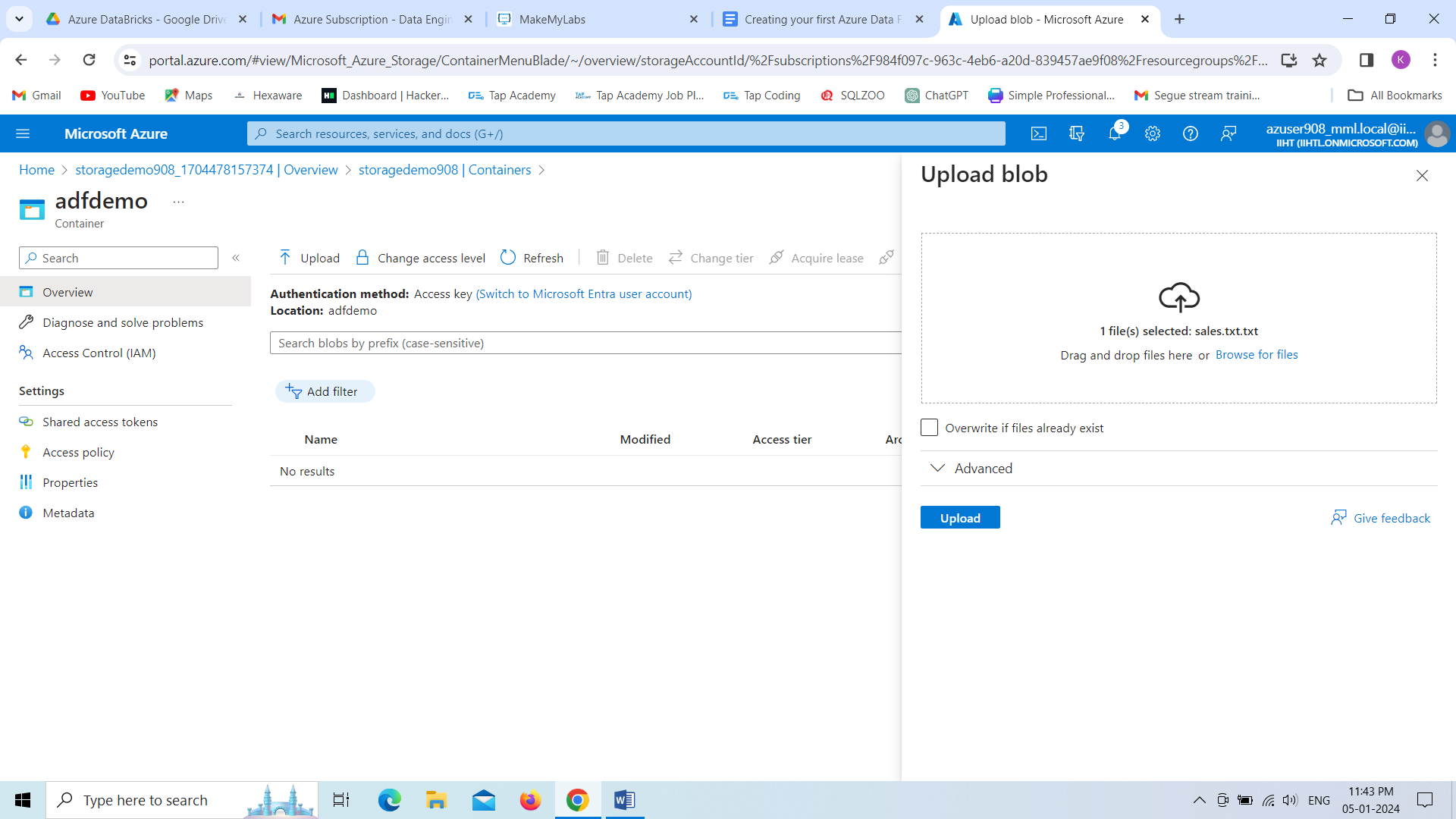


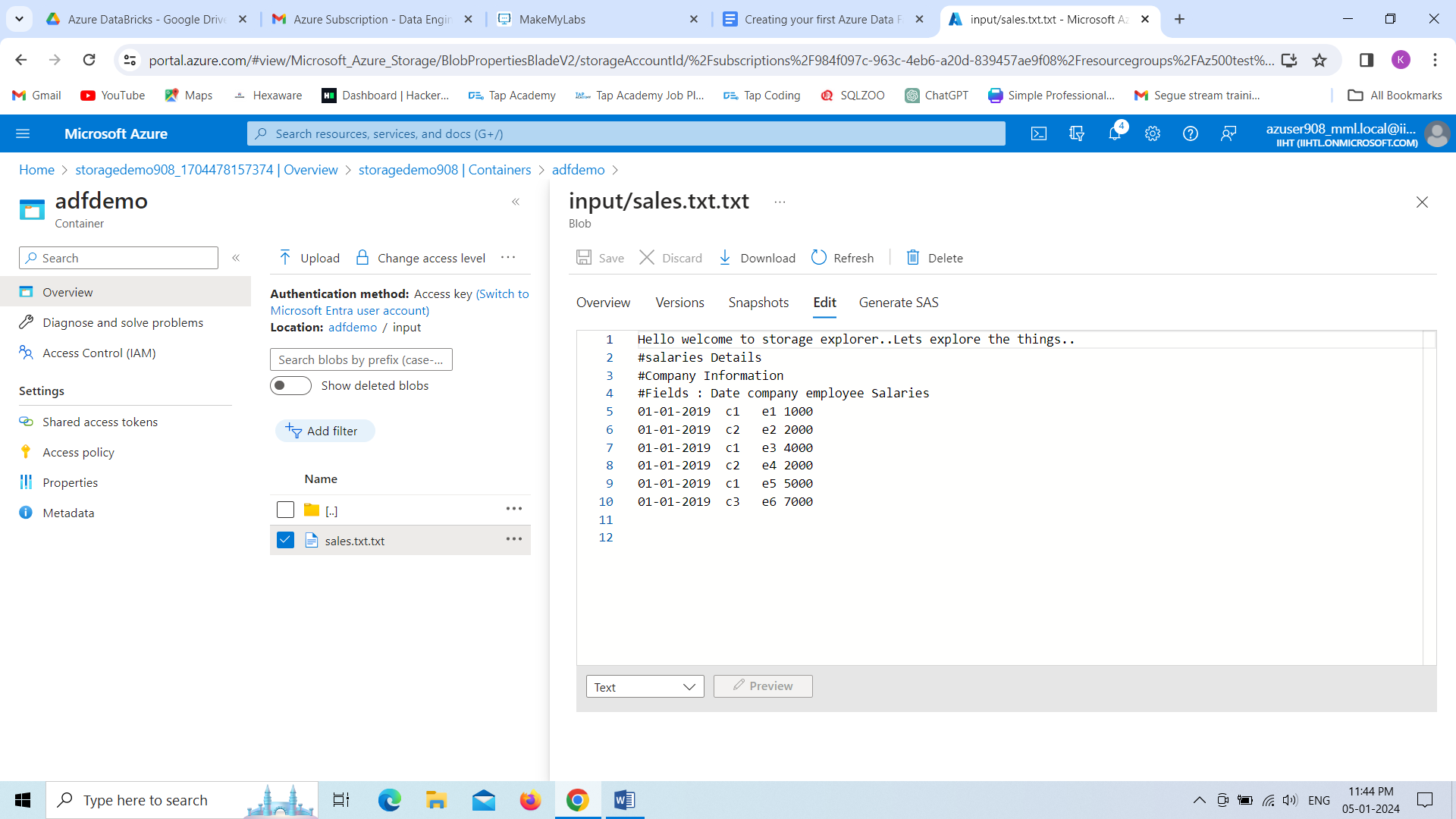


**G**o to the resource,creating the container in order to create a file.



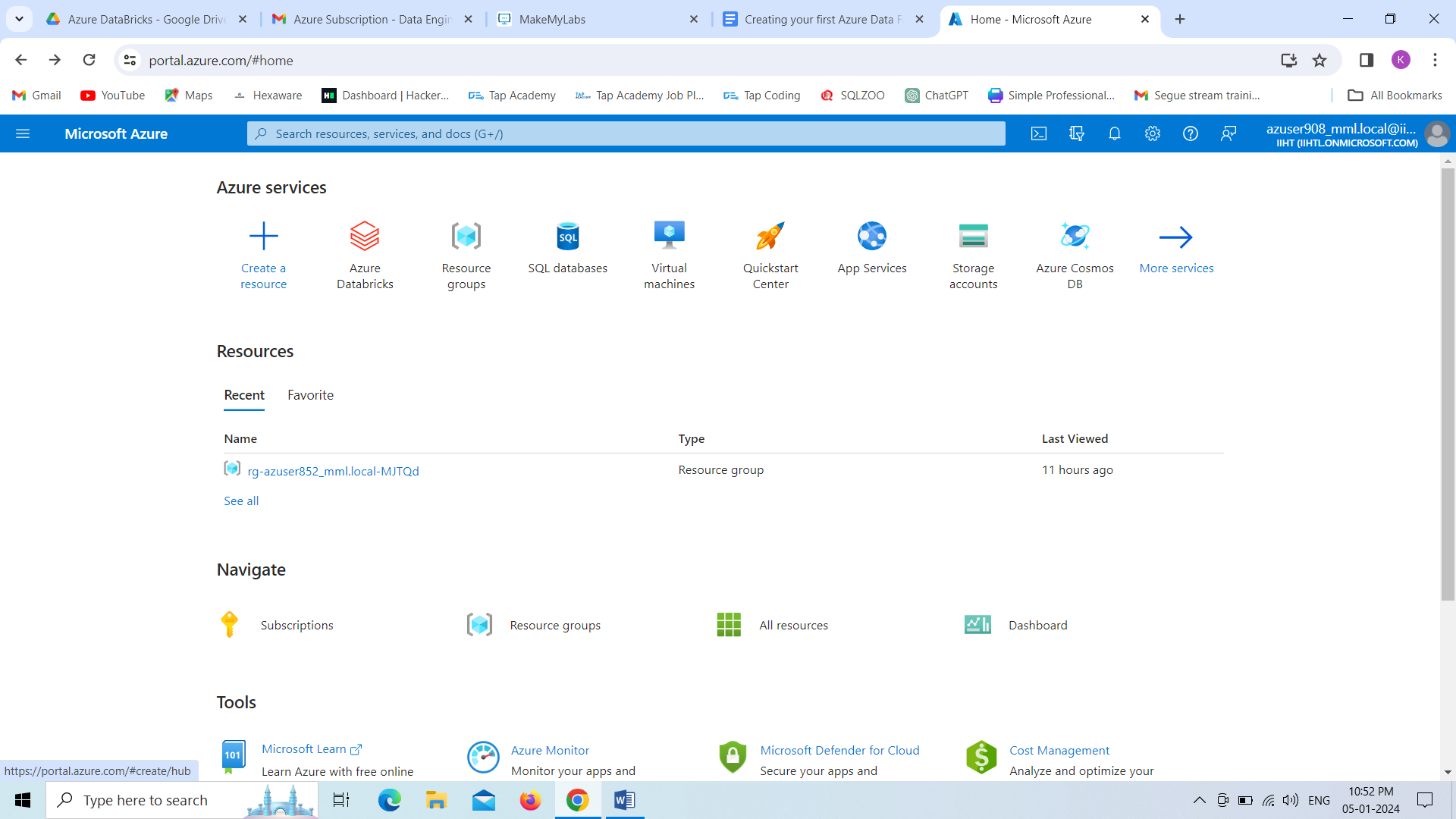
Here we are uploading the data nothing but file from **upload** option.



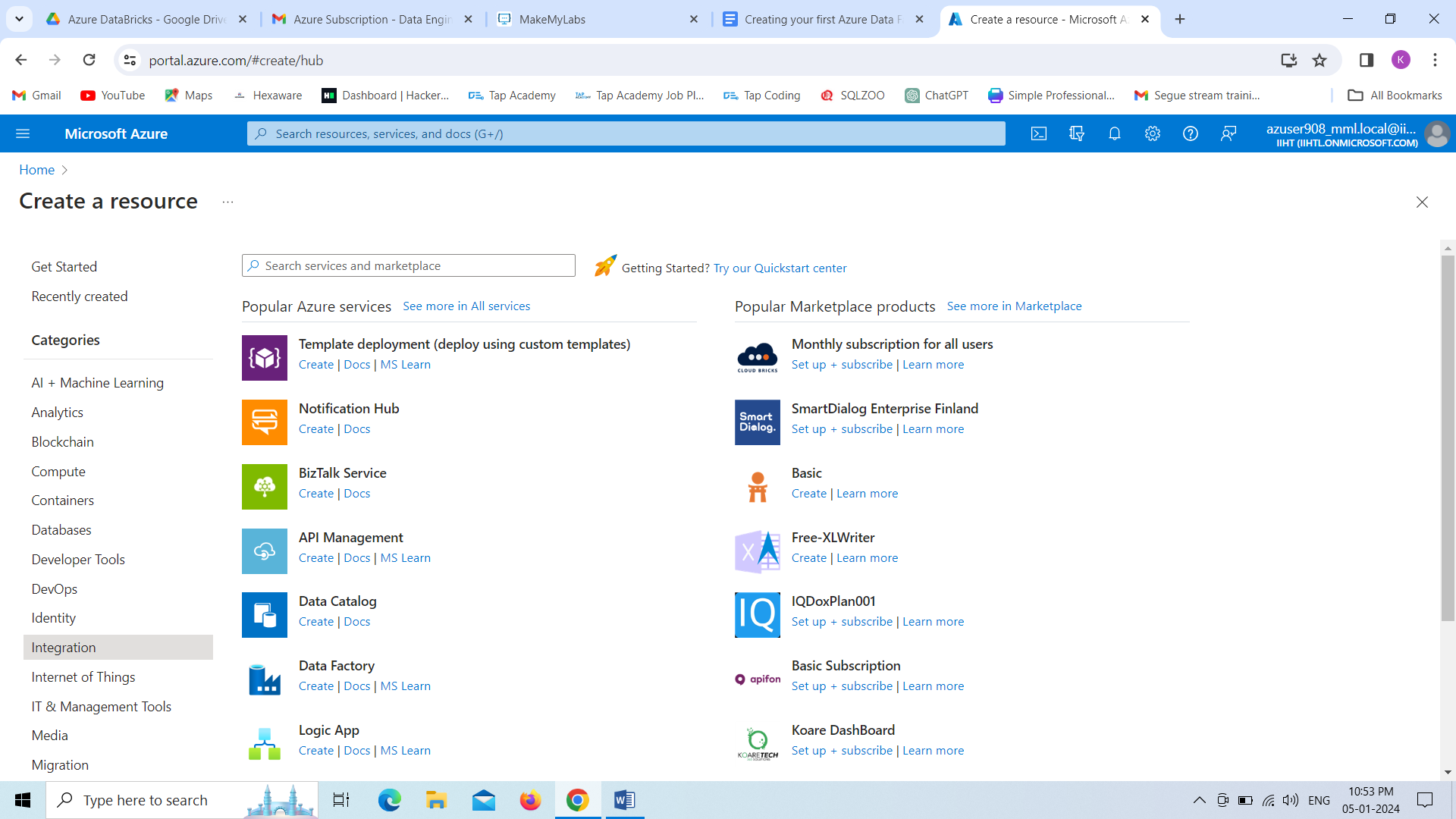
Whatever the data that we have uploaded,just we are seeing the data from **input** folder ,file named as **sales.txt**

**Create Azure Data Factory:**

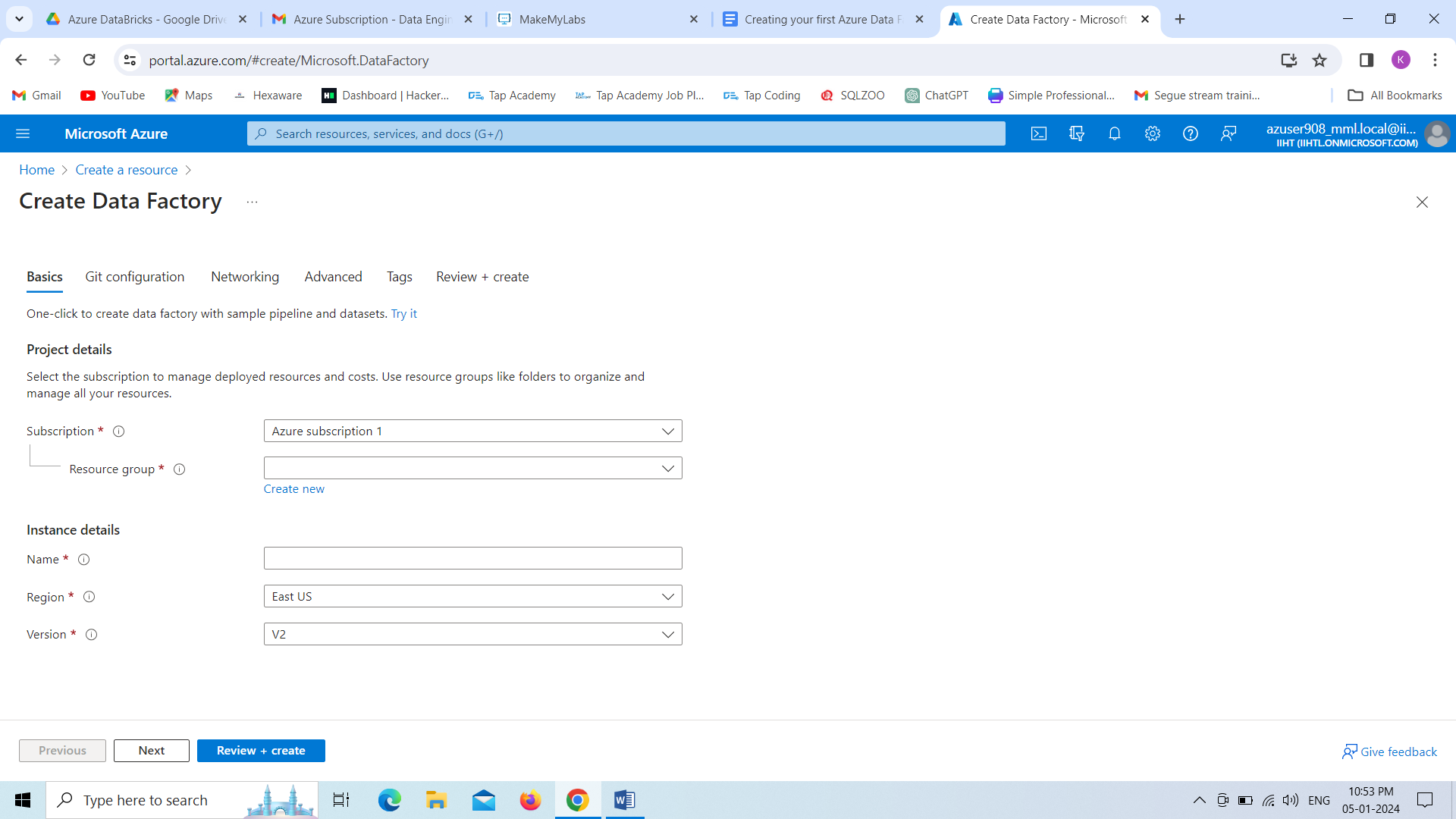
First step, log into the portal and click the **Create a resource** button.

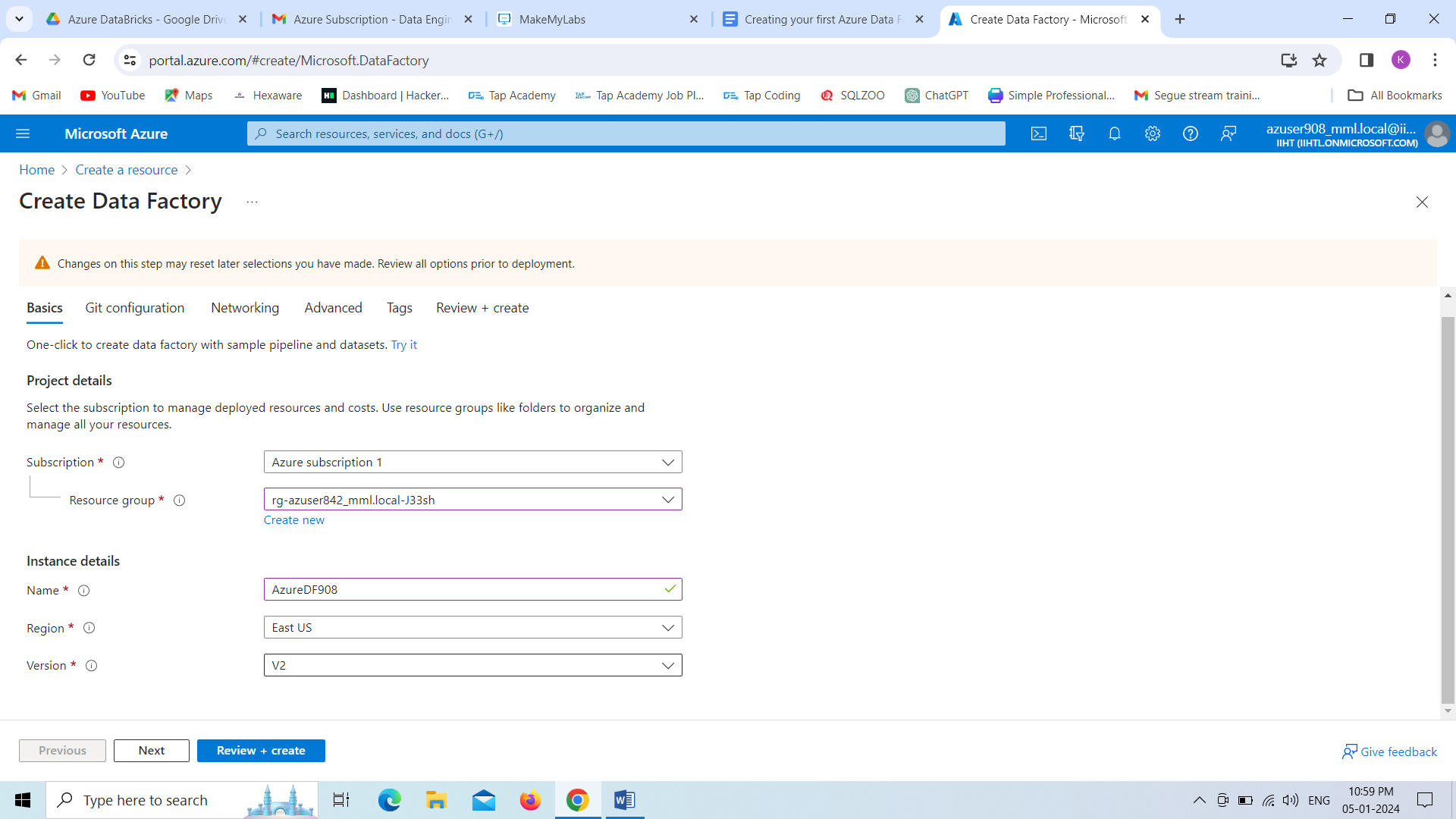


Next, select the **Integration** option and then click **Data Factory.**

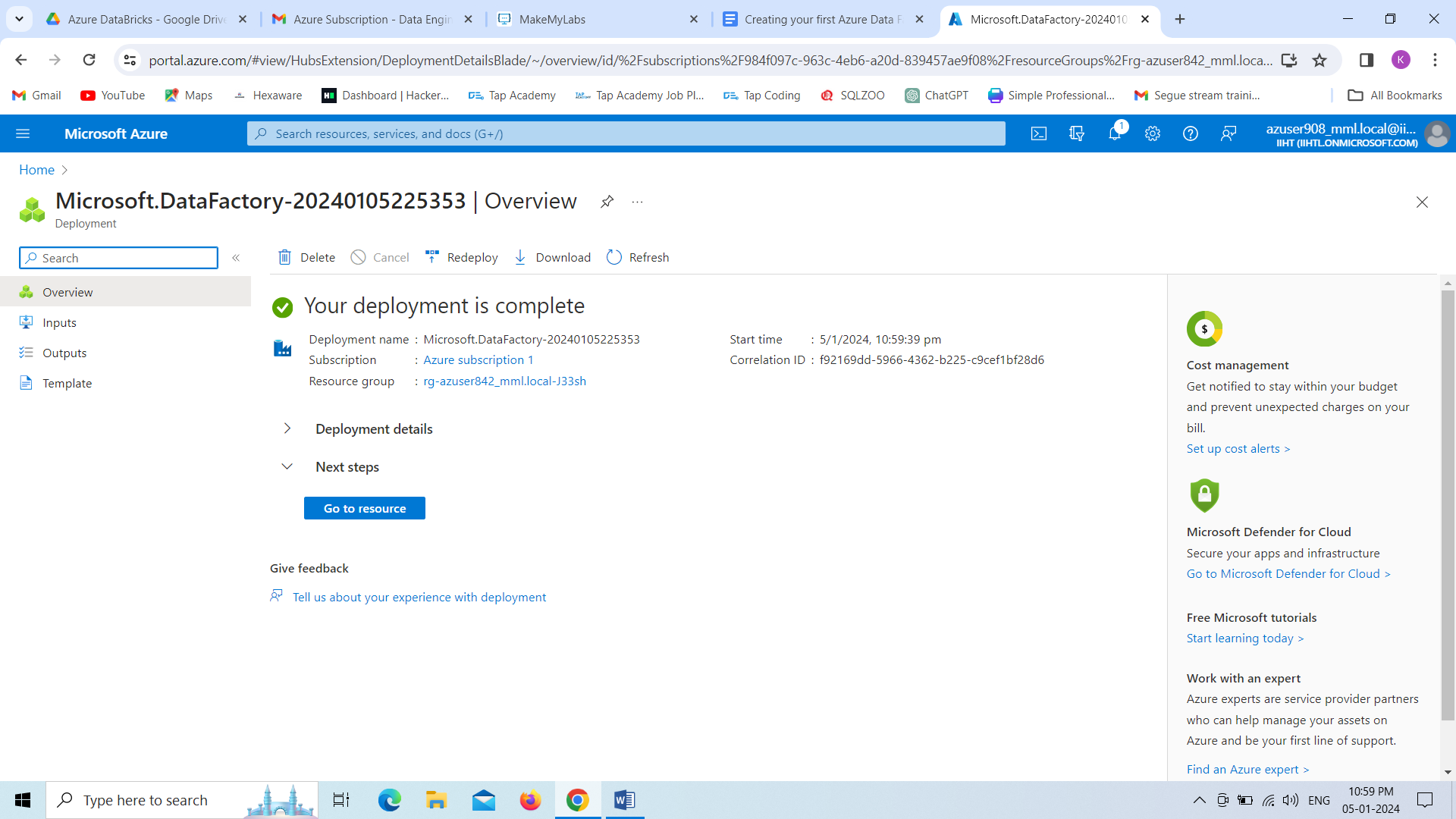


We are creating the data factory ,with the existing resource group.

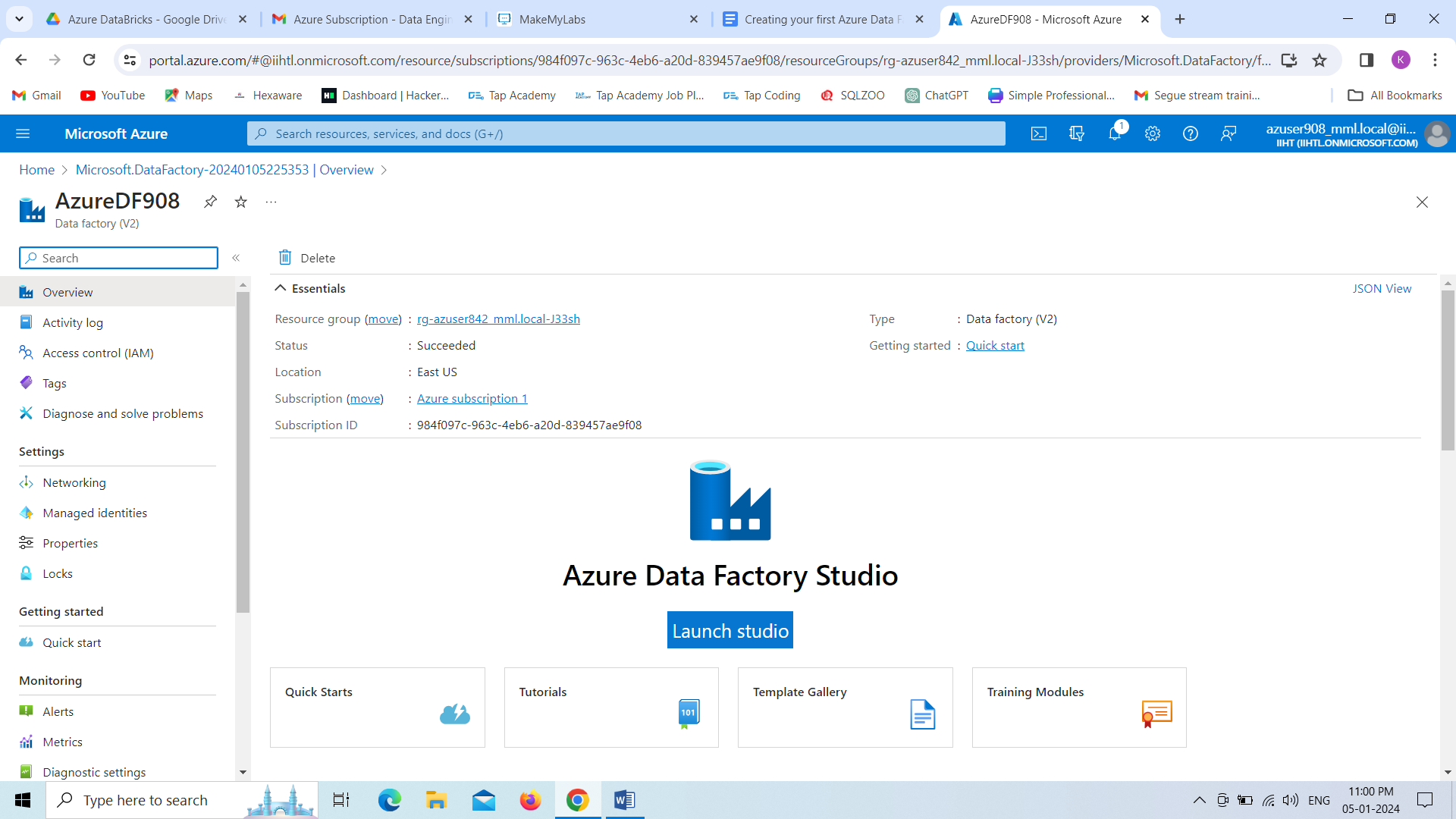




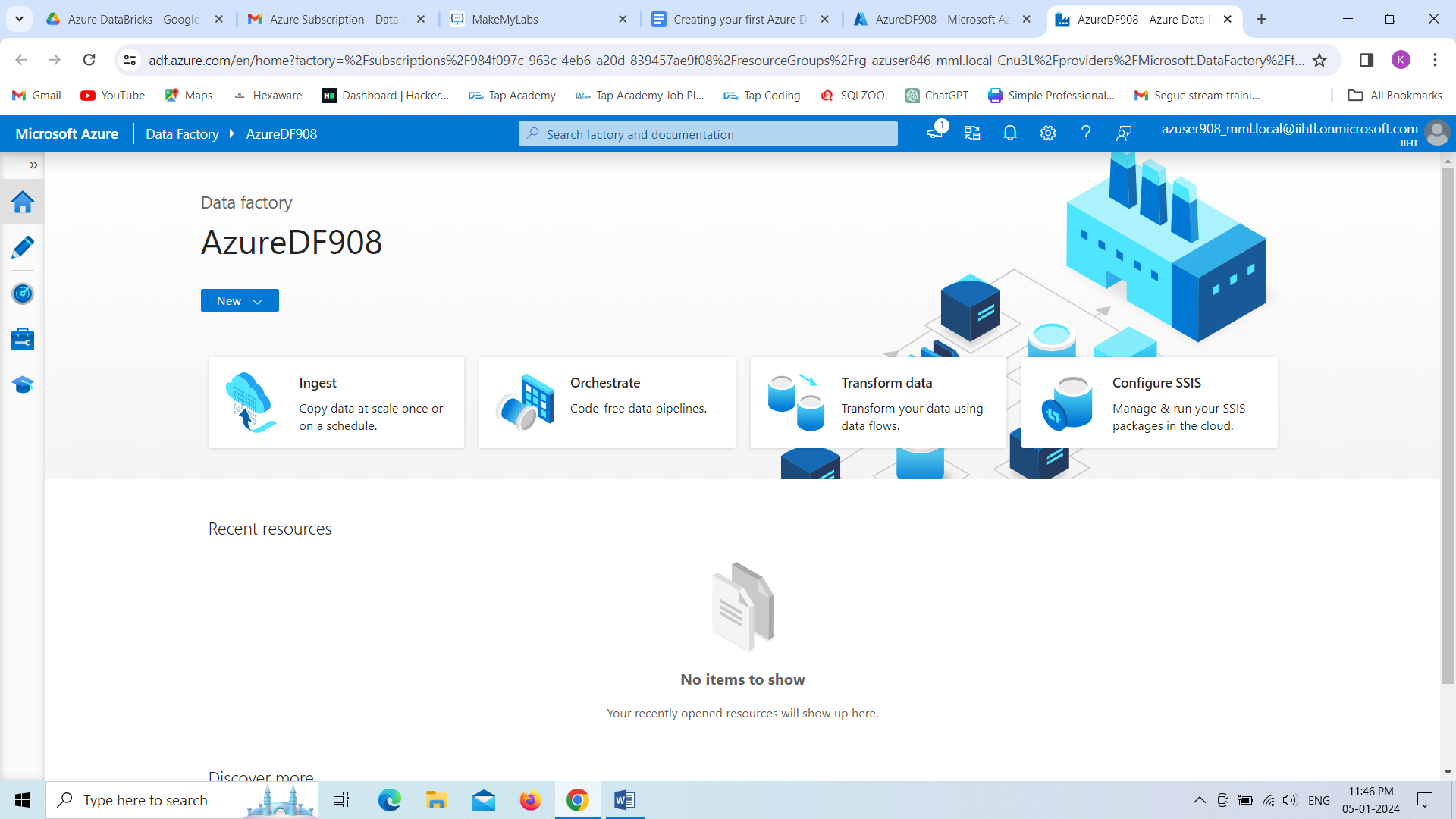
Azure Data Factory, Deployment is successfully completed,Click on **Go to Resource.**



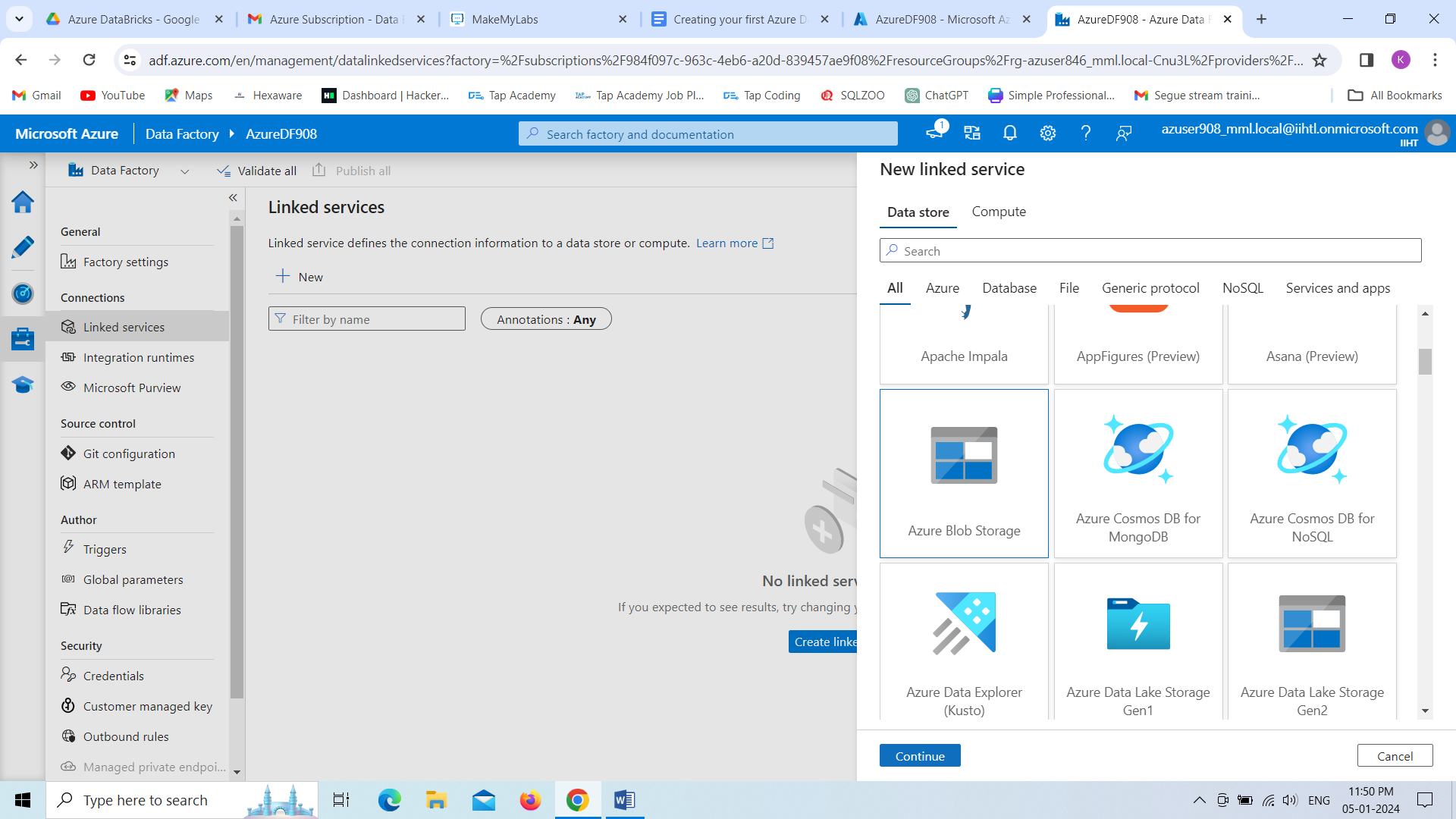
Now we need to click on **Lunch Studio**,it will redirect to new page.

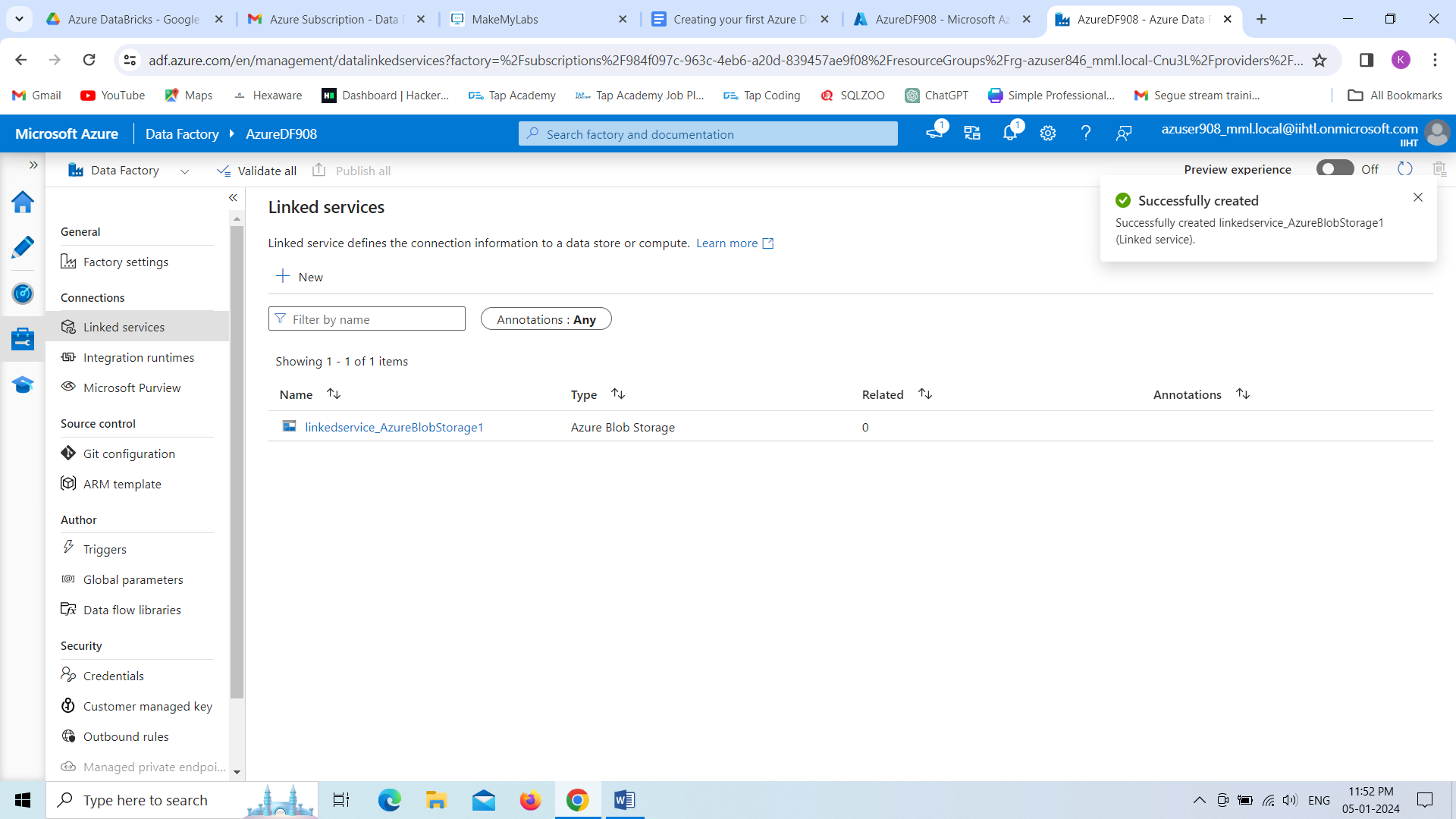


Now we are Creating the Azure Data Factory Pipeline

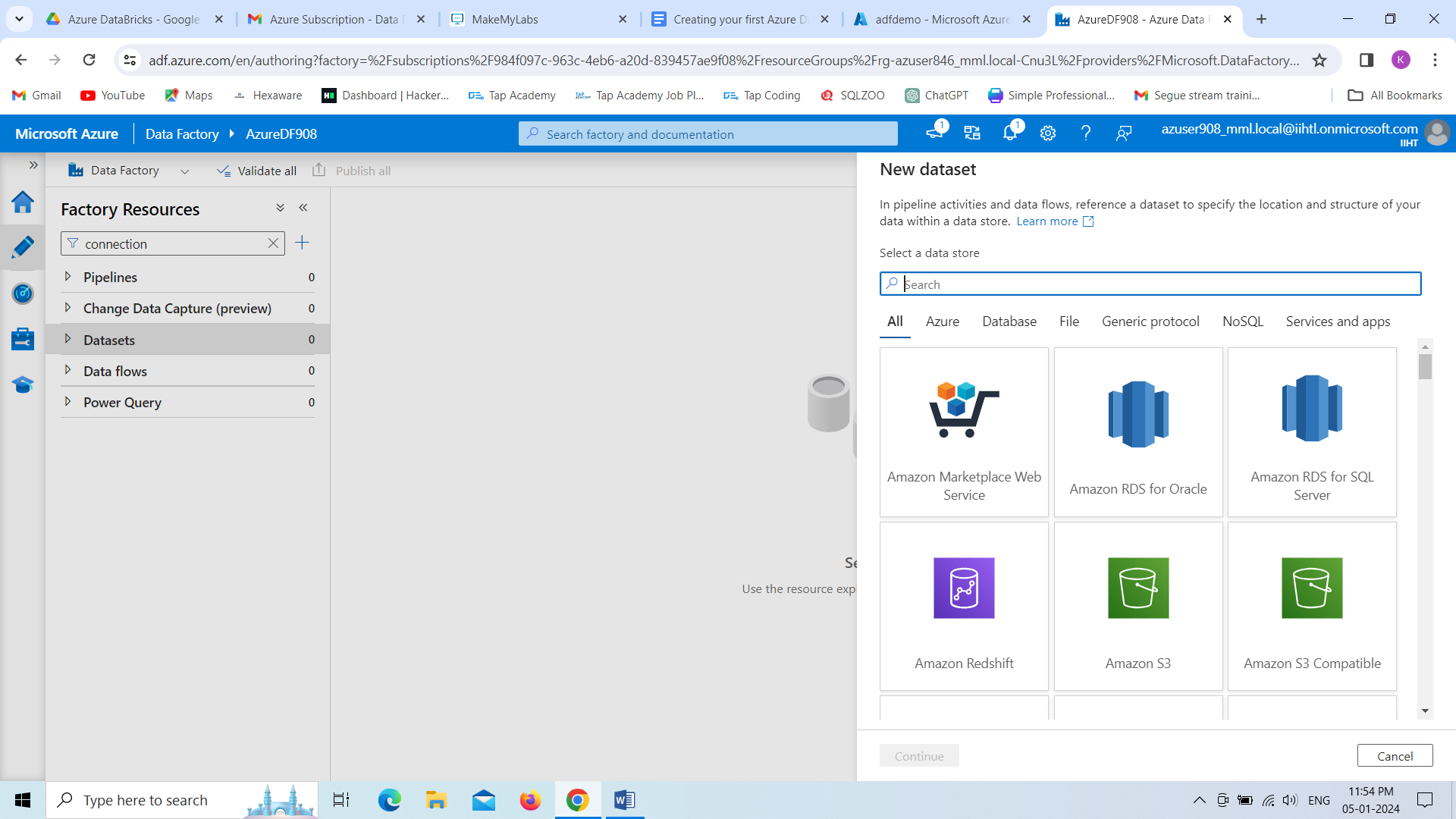


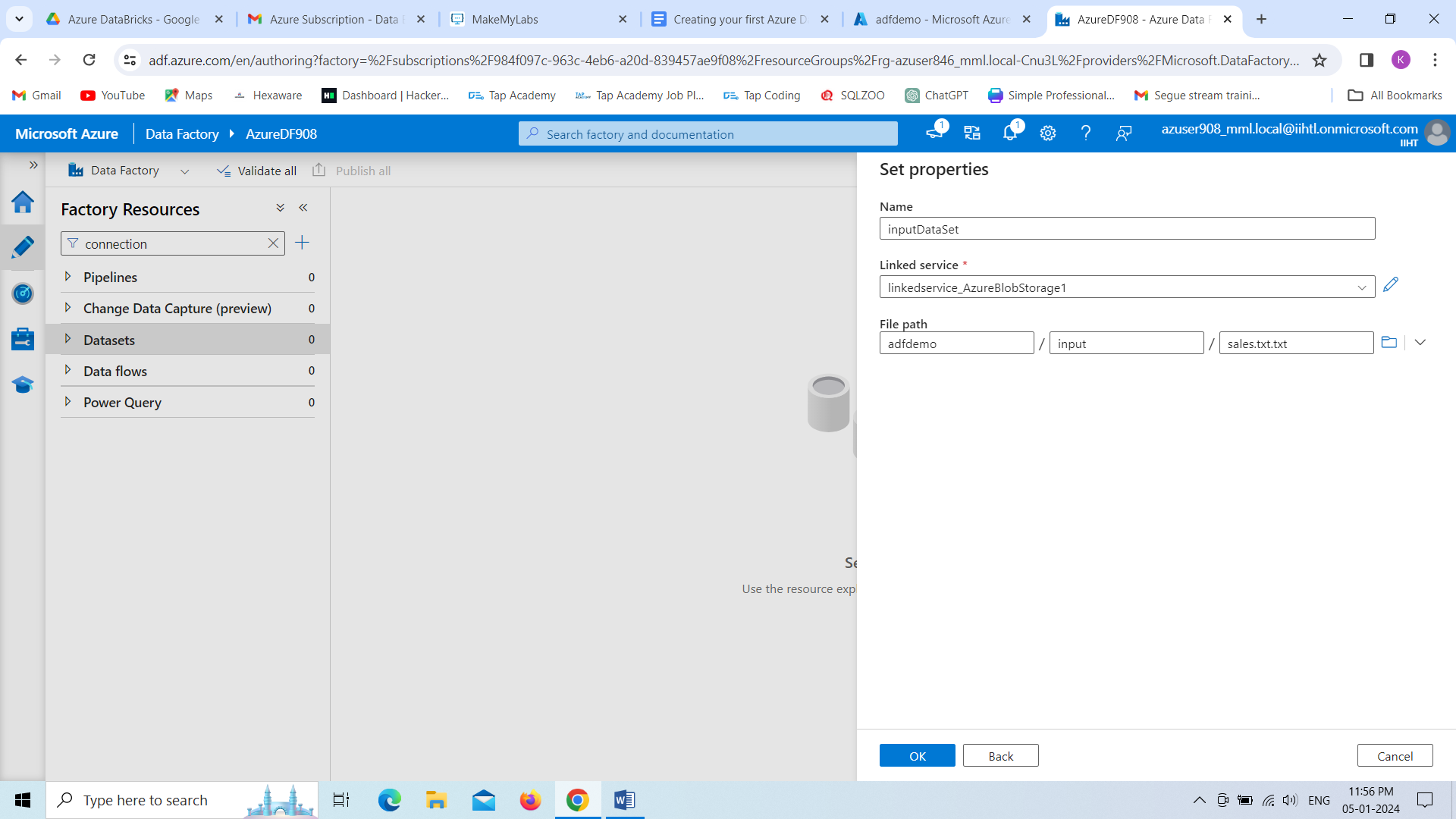
First we have to create the linked services for copying data.



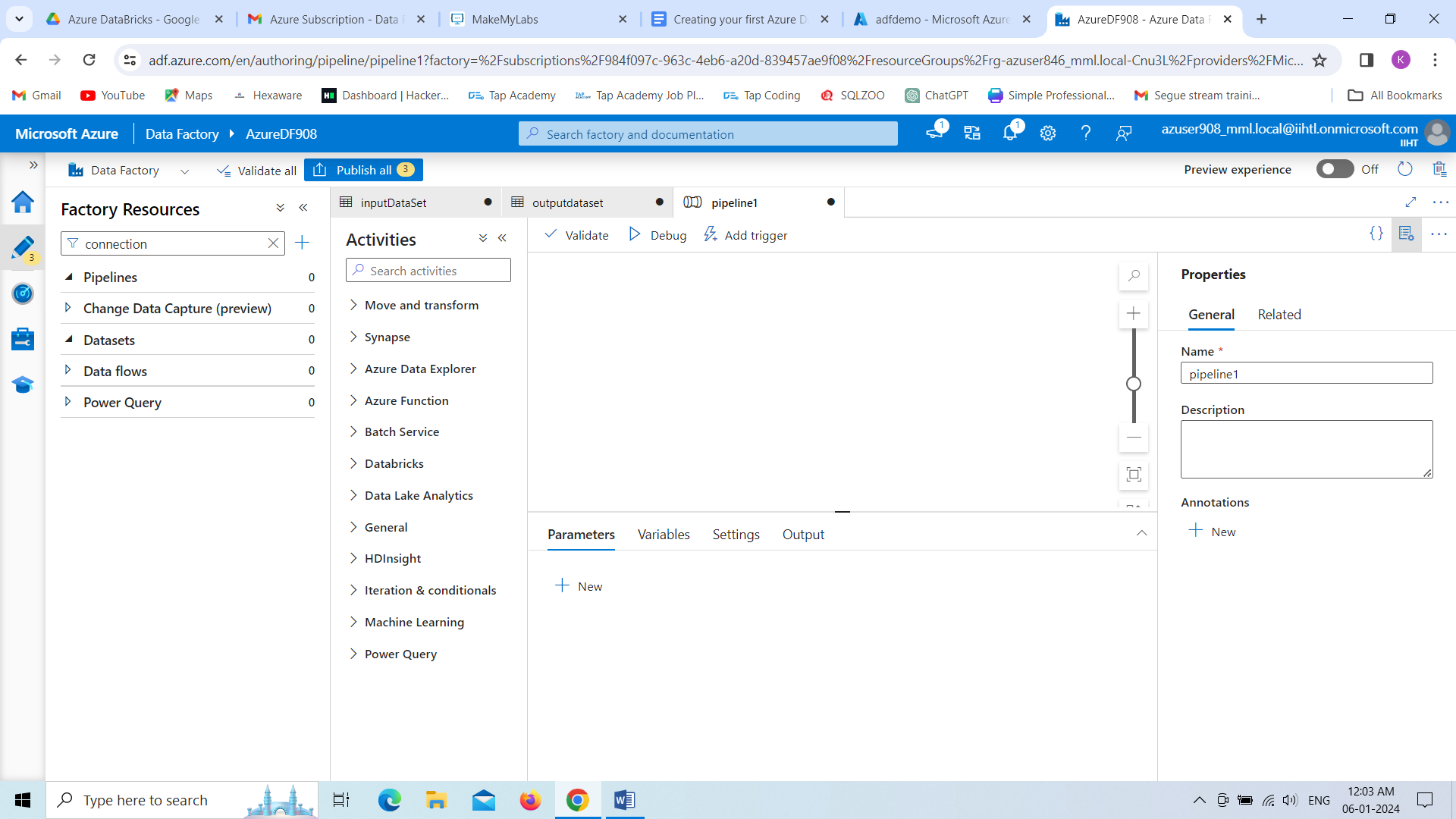


Now we are creating the datasets for input and output data.

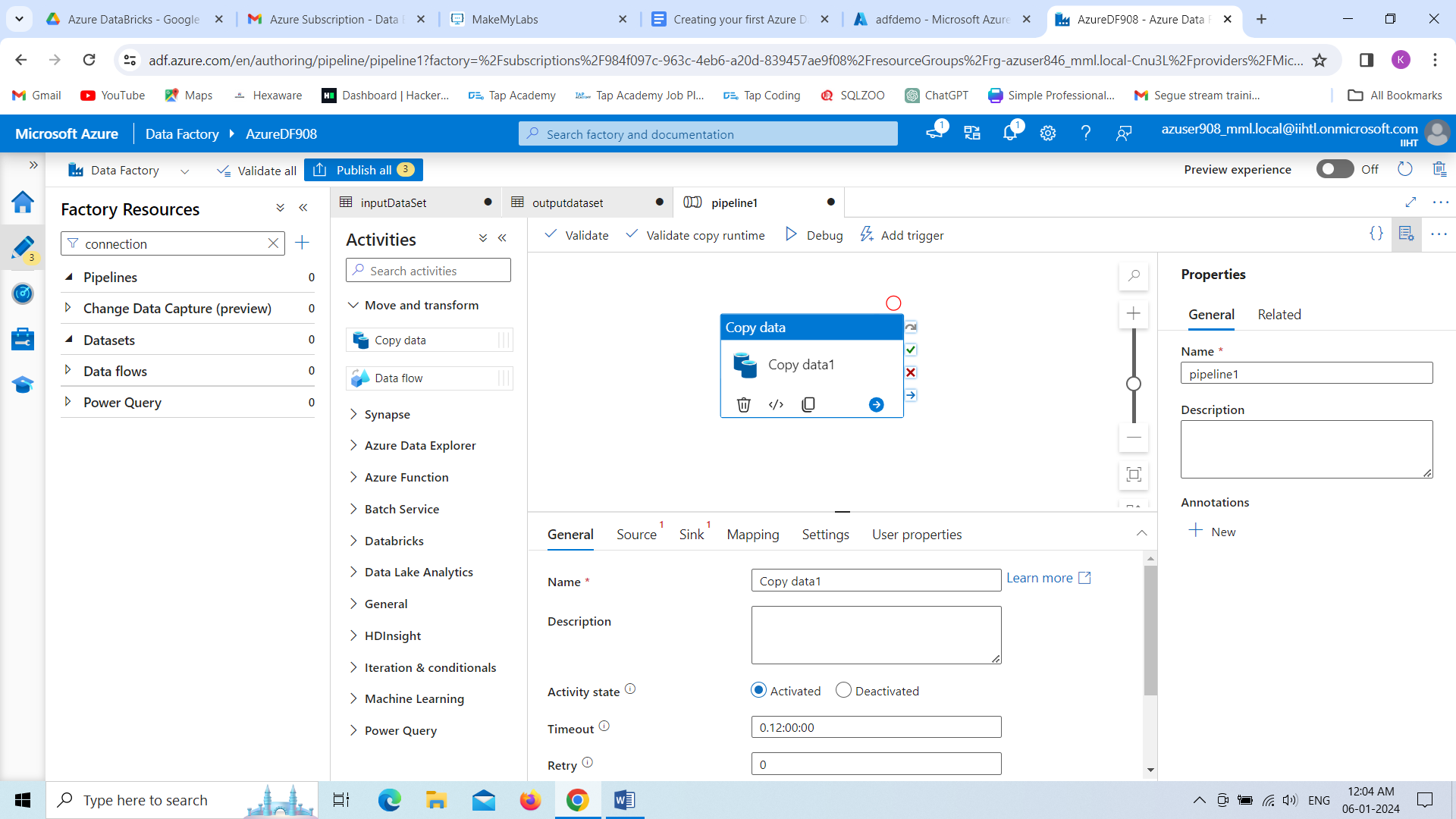




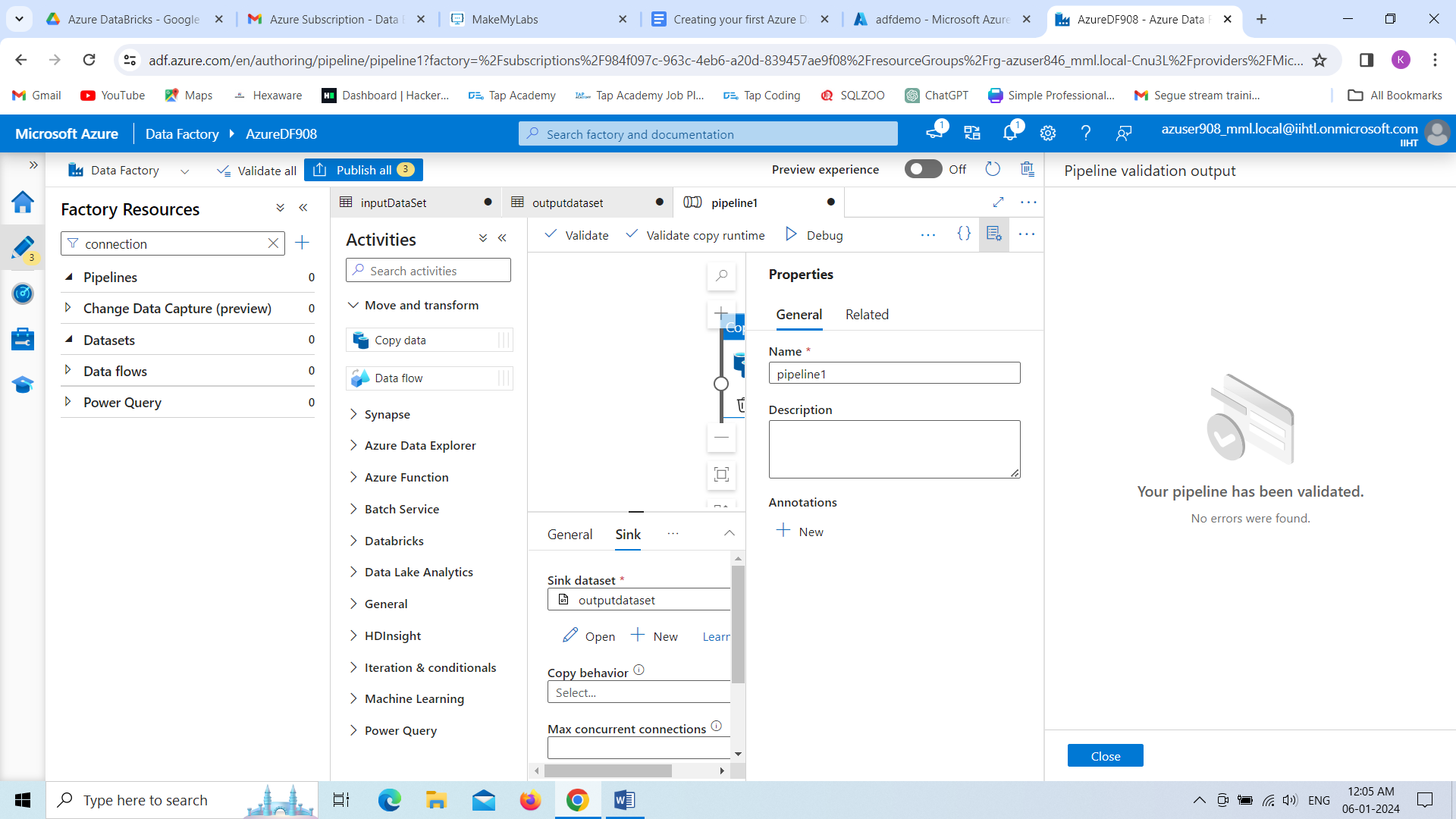
Now we are creating the pipeline



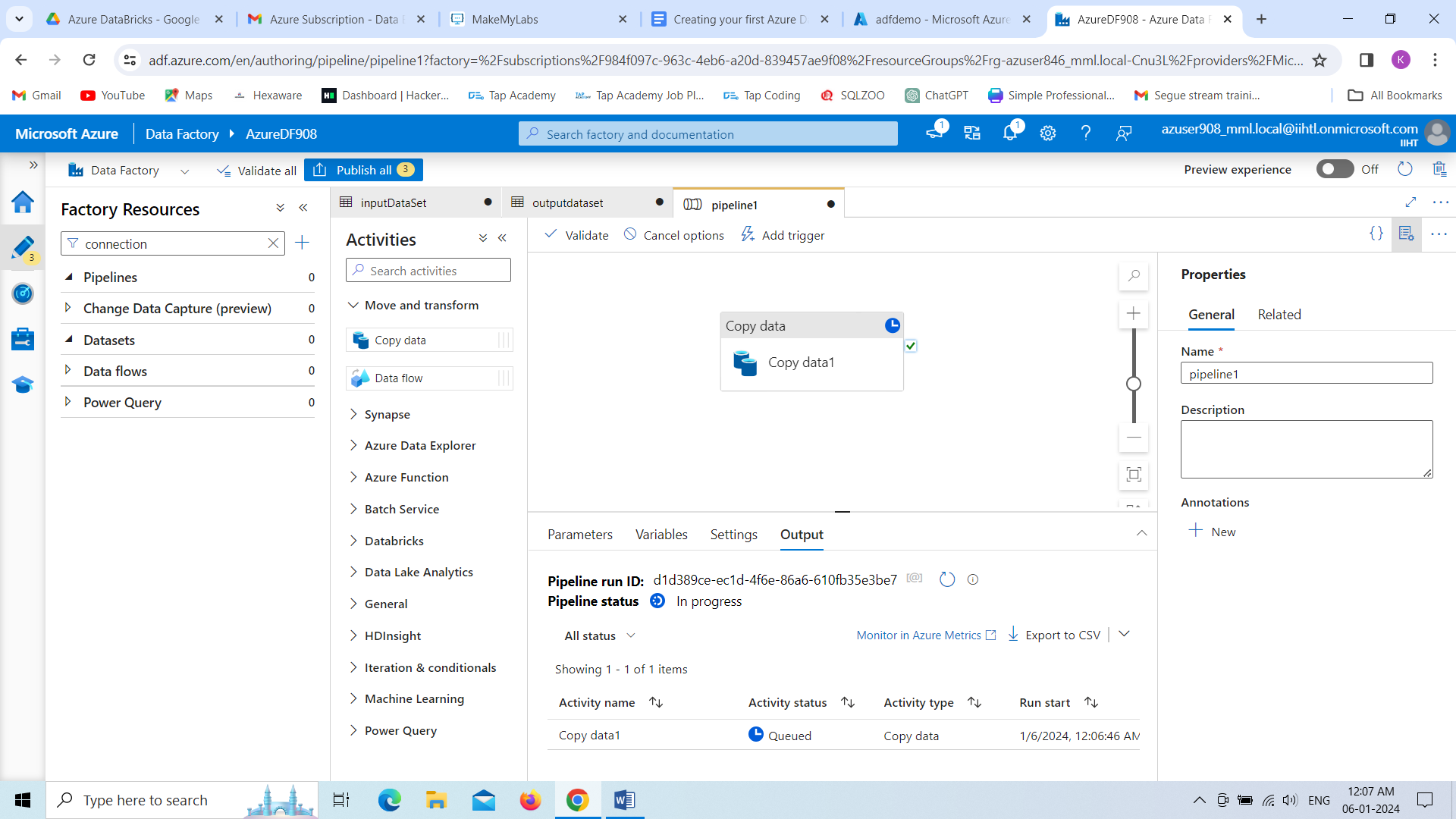
From move and transform ,we need to drag copy data



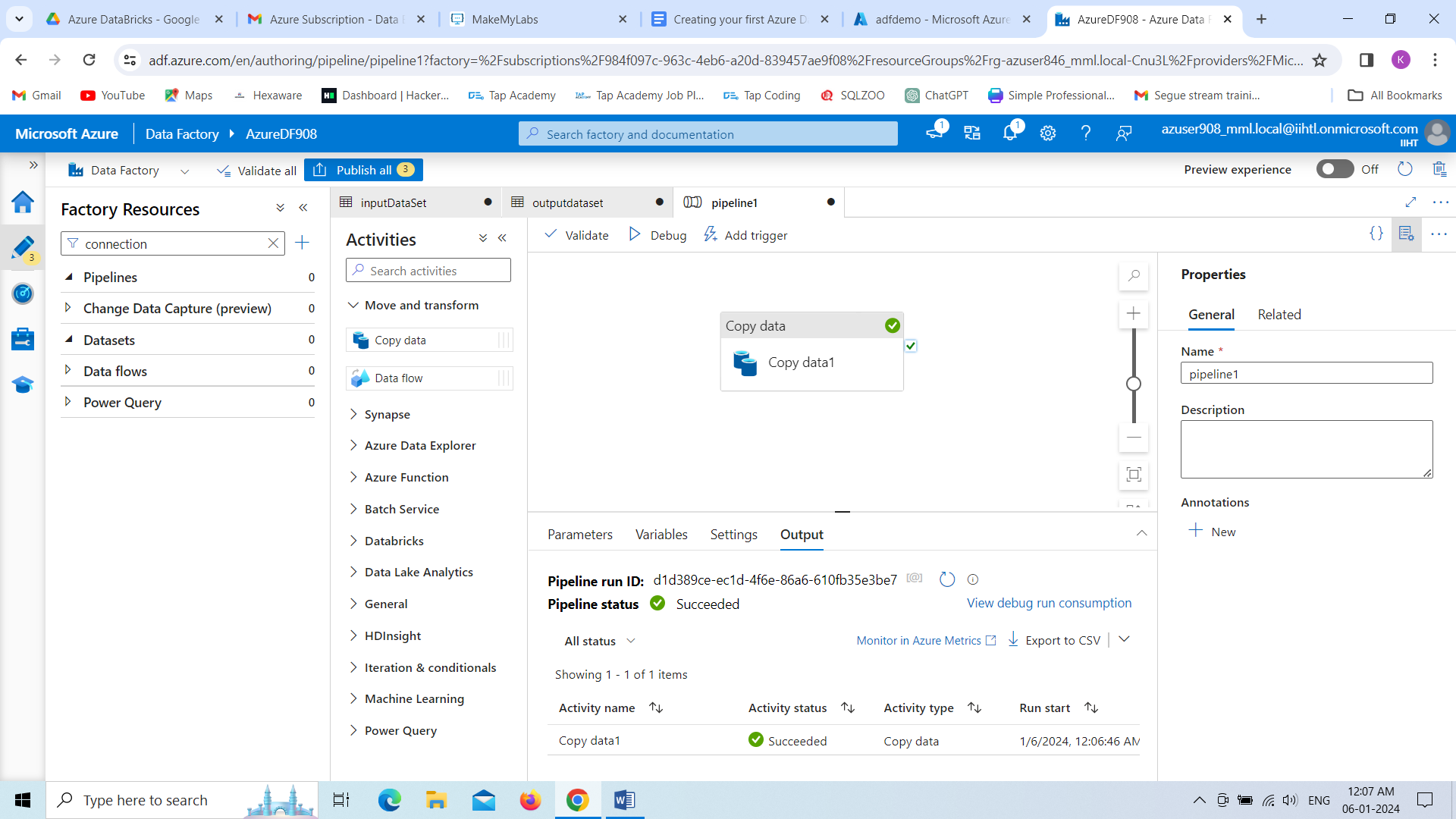
If we want to check whether our pipeline is created or not, we can click on **validate** option, then it is showing like as below.



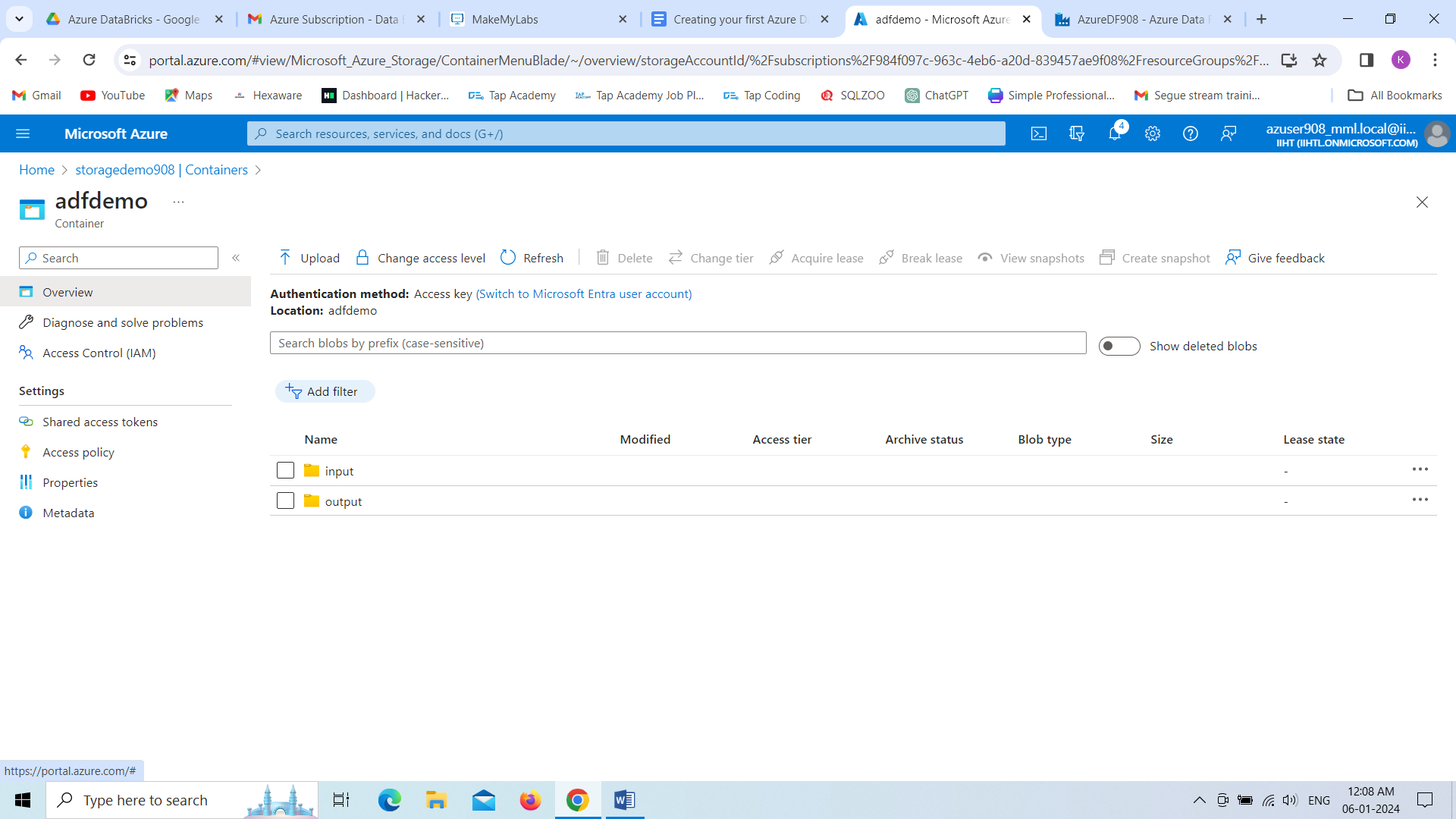
Here we can see the data is Copying



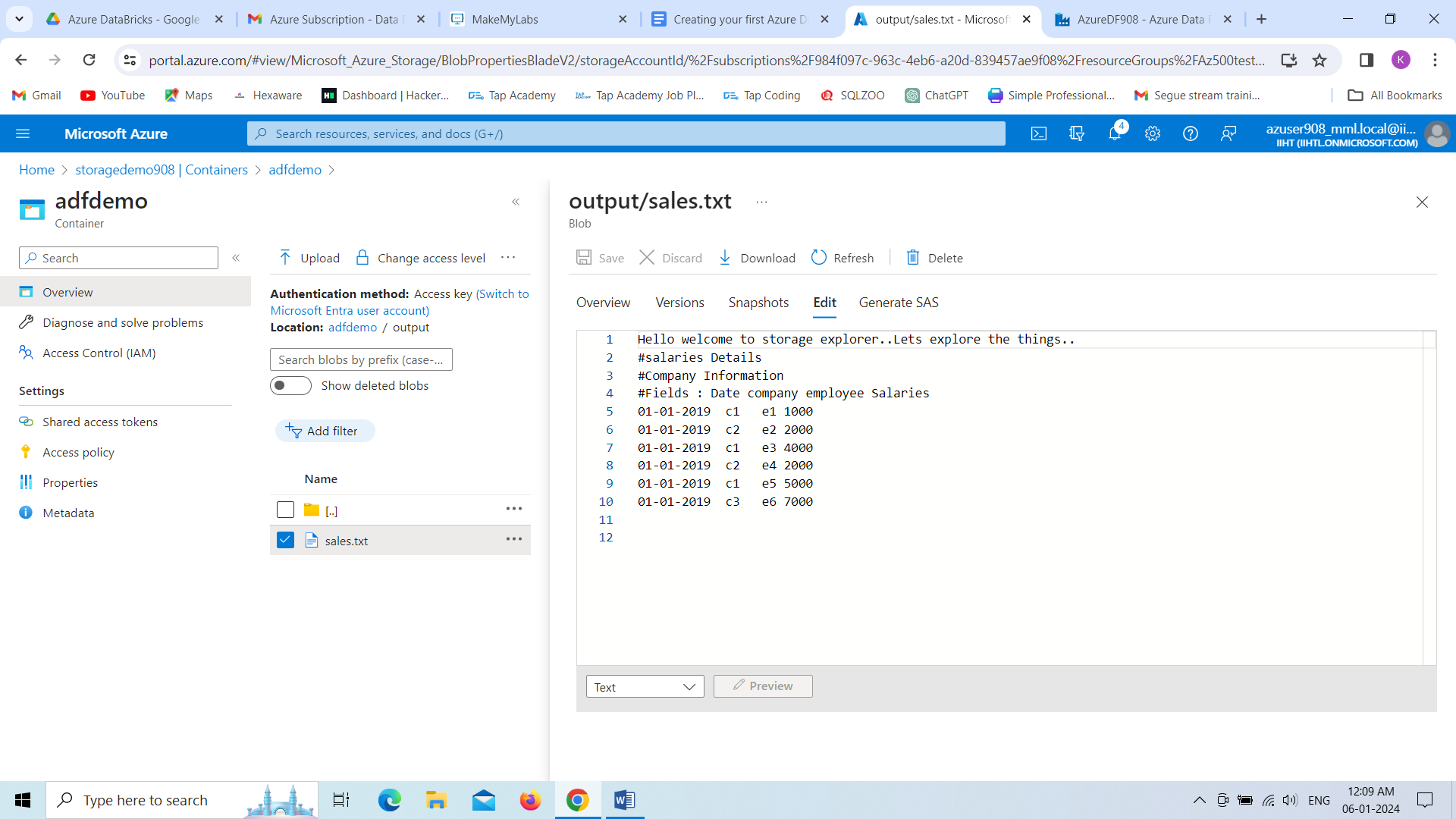
That green symbol show that, data is copied successfully.



Now we can see, Data is successfully copied which shows new folder as **output** in that we have sales.txt



All the data is copied, we can see below.



**Introduction to Delta Lakehouse**

* Delta Lakehouse is an open-source storage framework that brings the best of data warehouses and data lakes together.
* creating a powerful hybrid architecture called the lake house.
* It sits on top of your existing data lake storage (like S3 or ADLS) and brings several key features to the table.

**Key Features:**

**ACID Transactions**: It will ensures data consistency and reliability with serializable transactions, even for concurrent operations. No more worrying about data corruption.

**Scalable Metadata:** Handles petabyte-scale tables with ease, efficiently managing billions of partitions and files without performance hiccups.

**Time Travel:** Access any previous version of your data at any point in time, perfect for audits, rollbacks, or reproducing analyses.

**Open Source:** Community-driven development, open standards, open protocol, and open discussions, allowing for flexibility and customization.

**Benefits:**

**Simplified Data Pipelines:** ACID transactions and schema enforcement reduce errors and reprocessing, making pipelines more reliable and efficient.

**Faster Analytics:** Optimized data layout and efficient querying techniques lead to faster queries and improved performance for analysts.

**Cost Optimization:** Delta Lake leverages the existing data lake storage, avoiding the need for expensive data duplication.