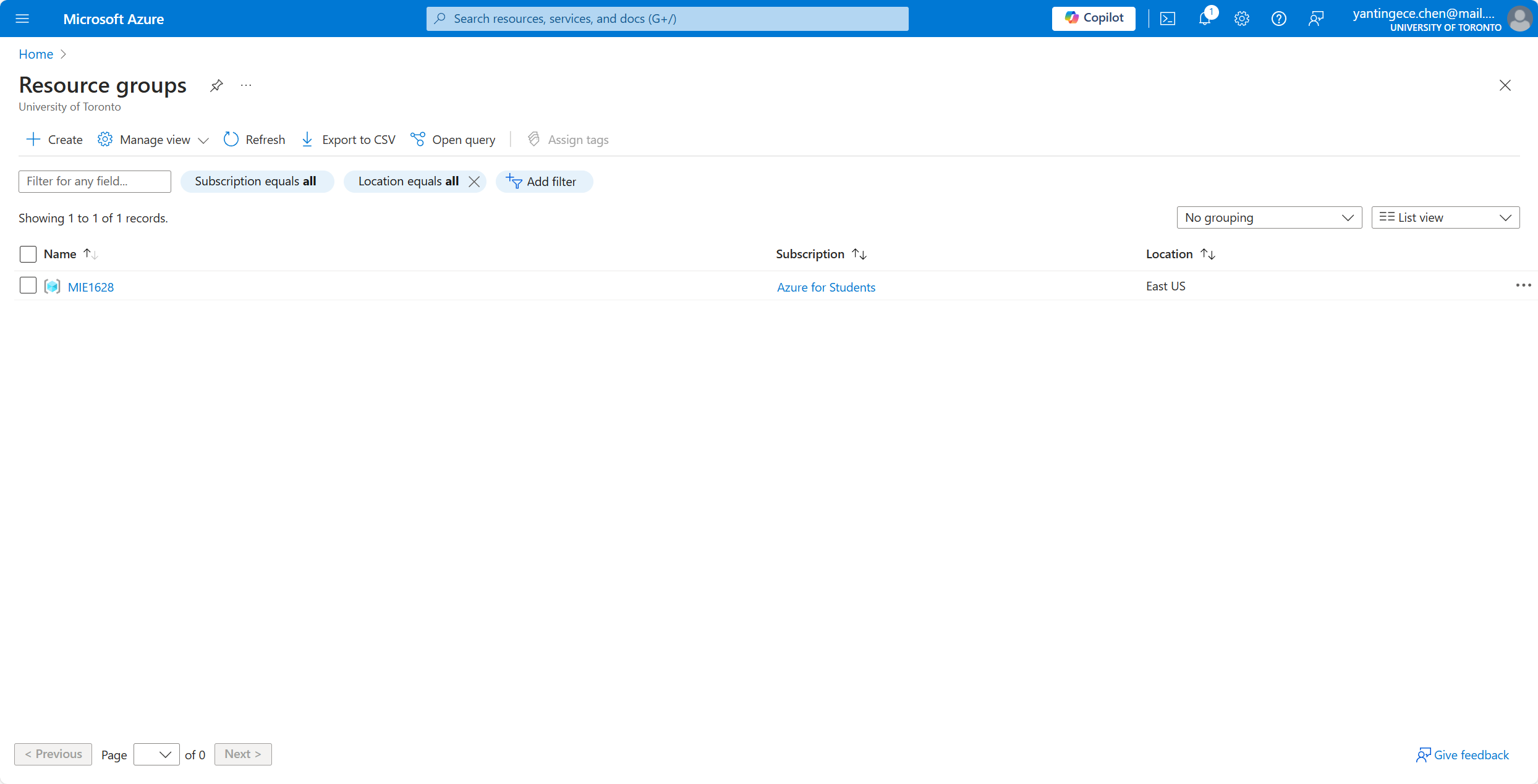
## Assignment 4

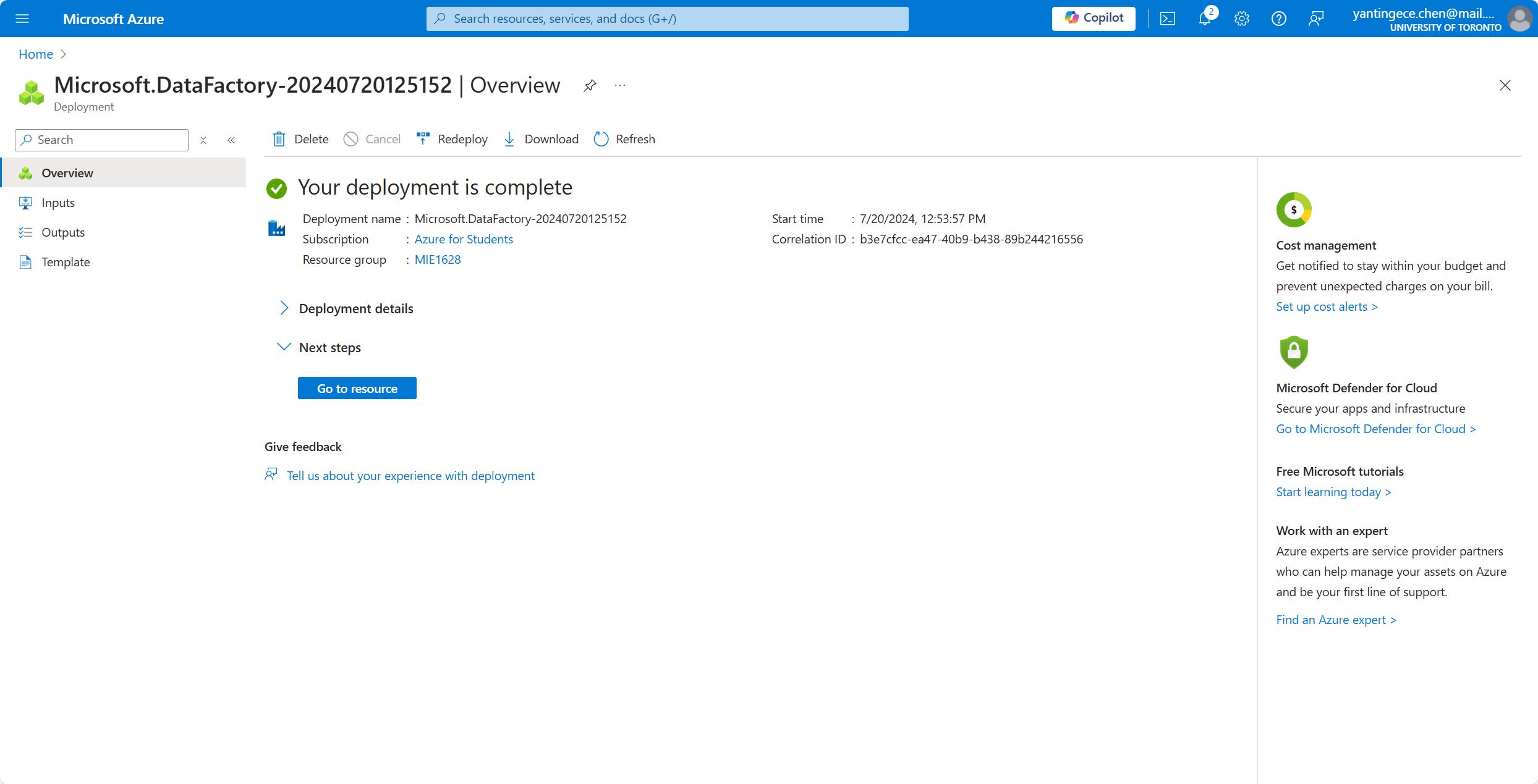
## Assignment on Azure Cloud Platform

### Part A:

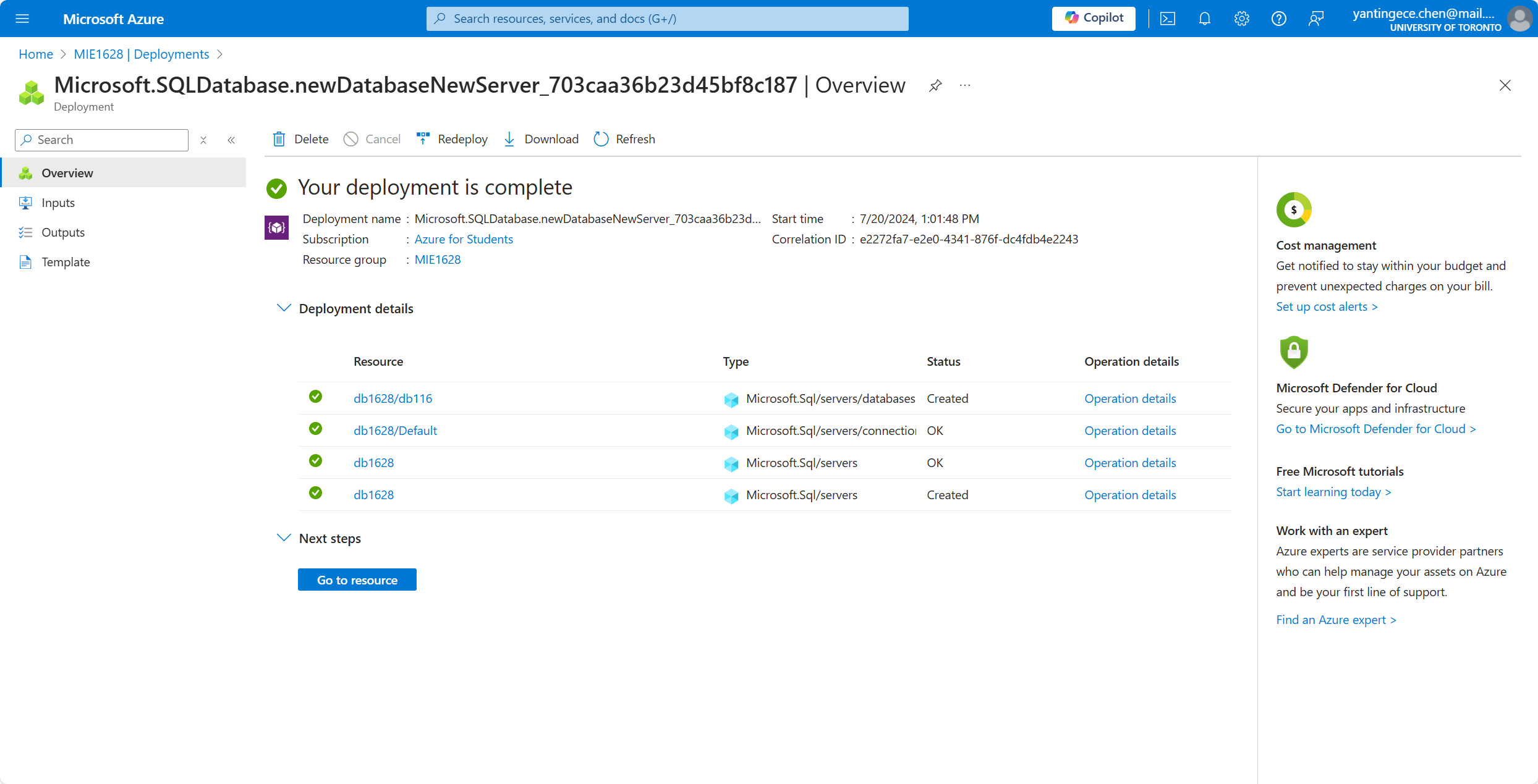
1. [Marks: 5] Create a resource group in your Azure portal and deploy three resources. Azure Data Factory, Azure SQL DB and Blob storage account.

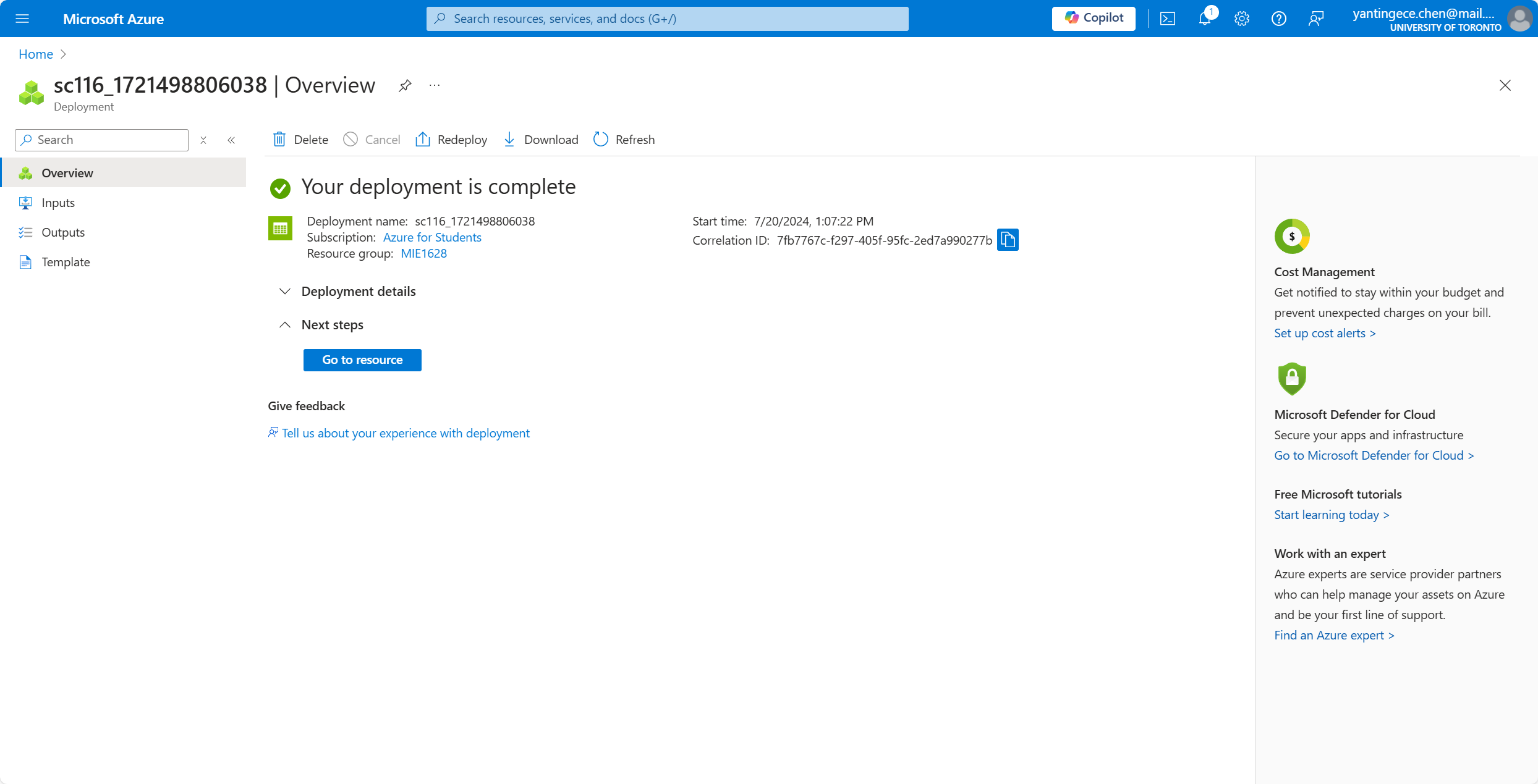
Resource group:

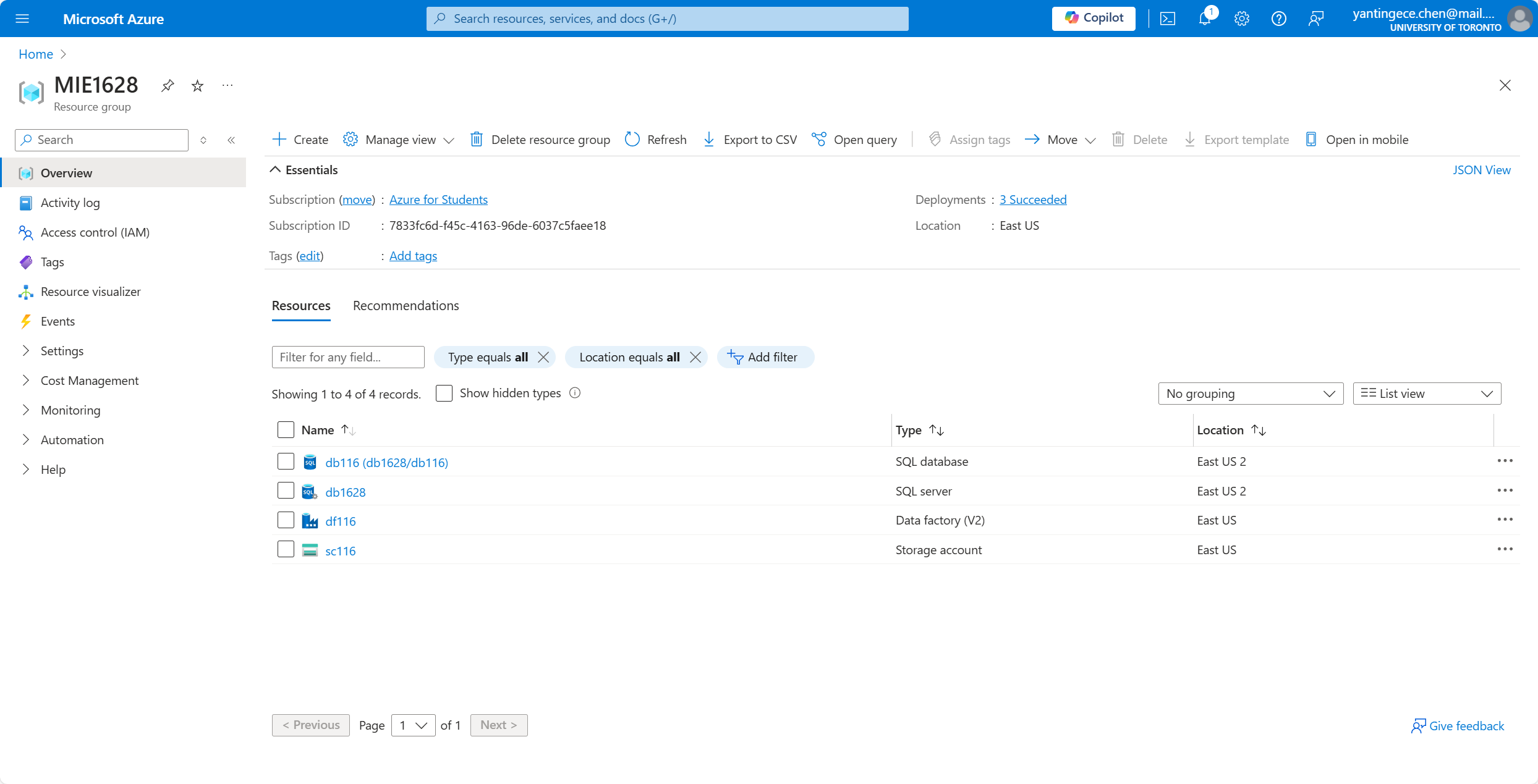


Azure Data Factory:

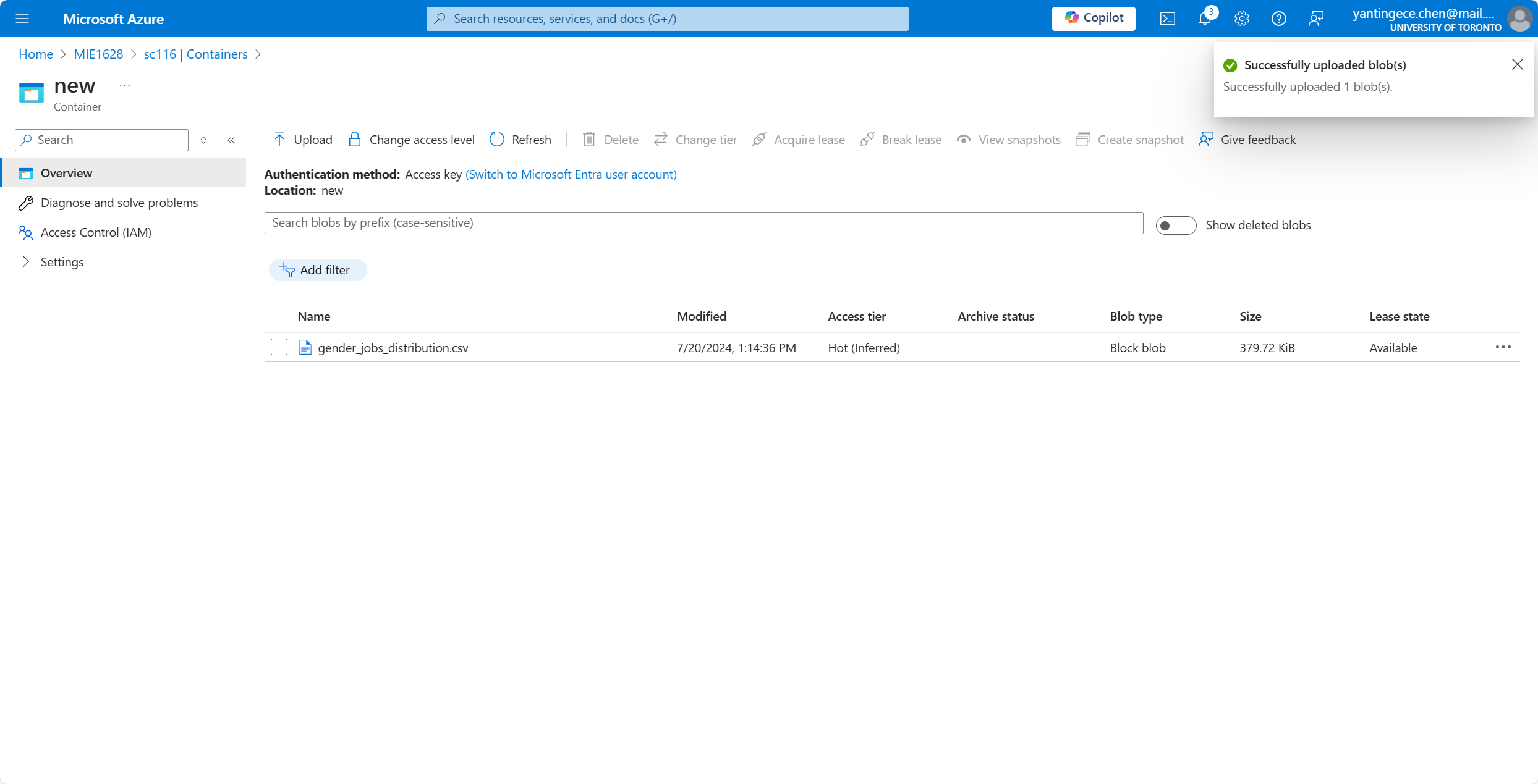
Azure SQL DB:

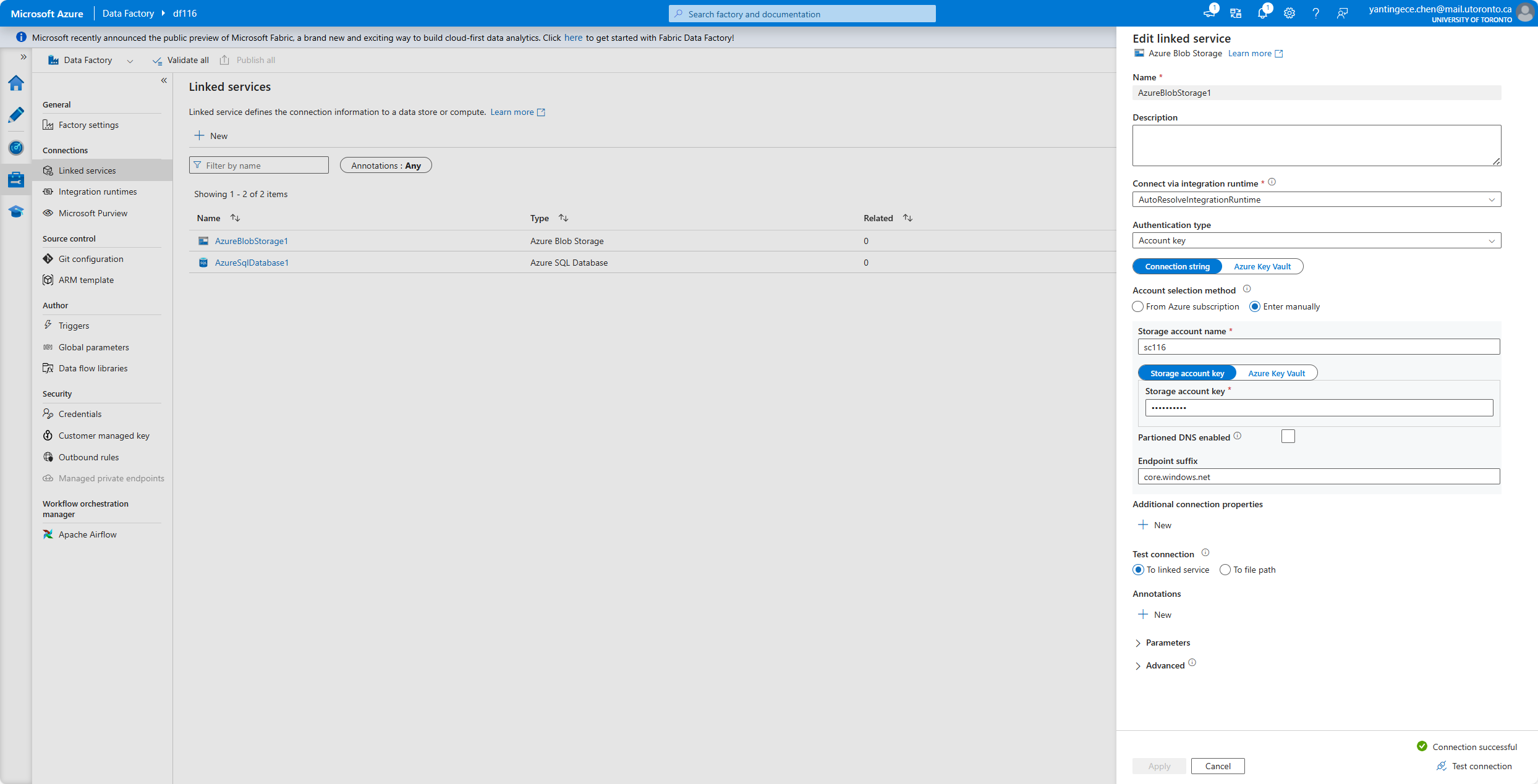
Blob storage account:

Overall:

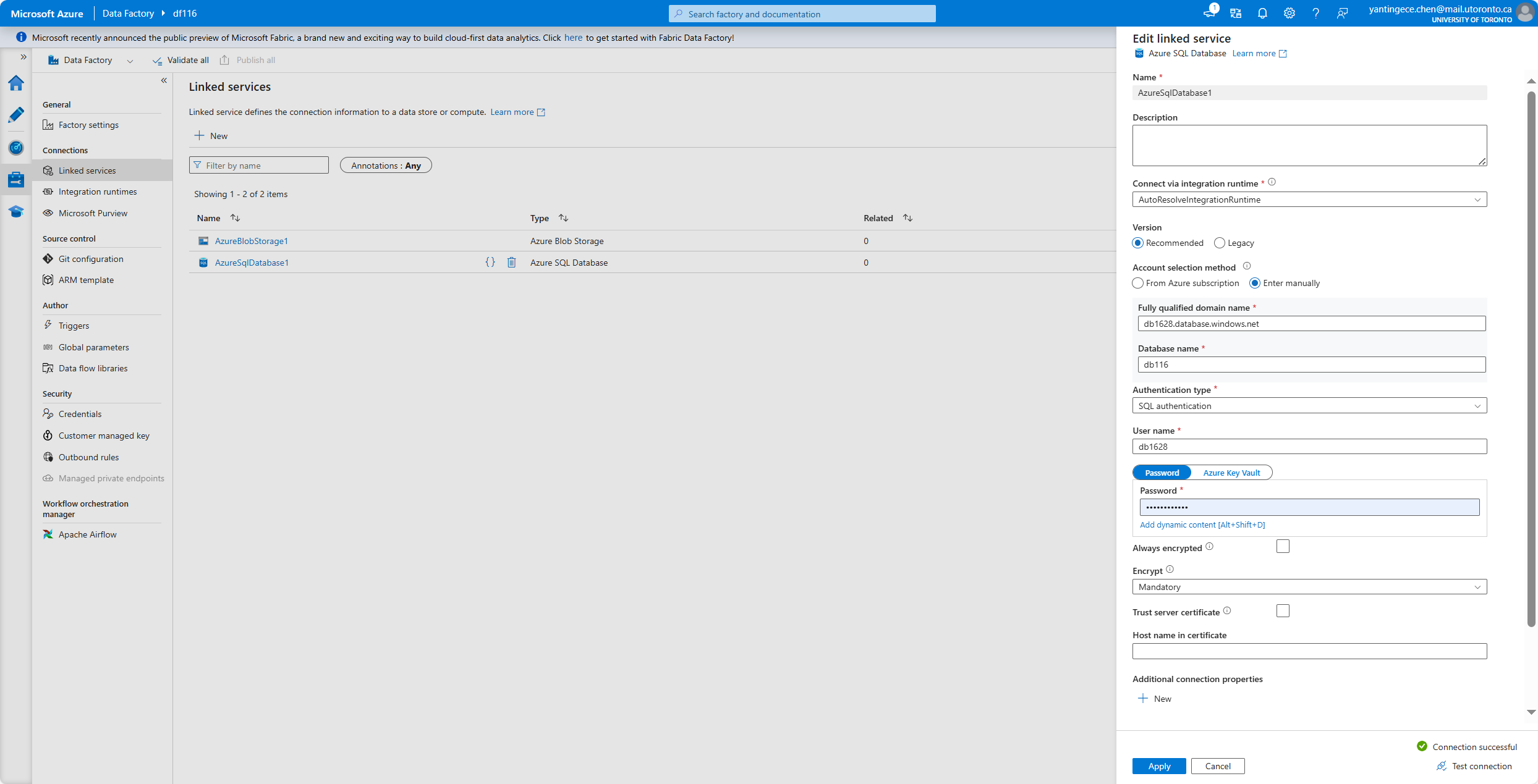


1. [Marks: 15] Now create a pipeline in Azure Data Factory and copy gender\_jobs\_data.csv file from the Blob storage account to Azure SQL DB. (First copy this file from your local machine to Blob Storage).

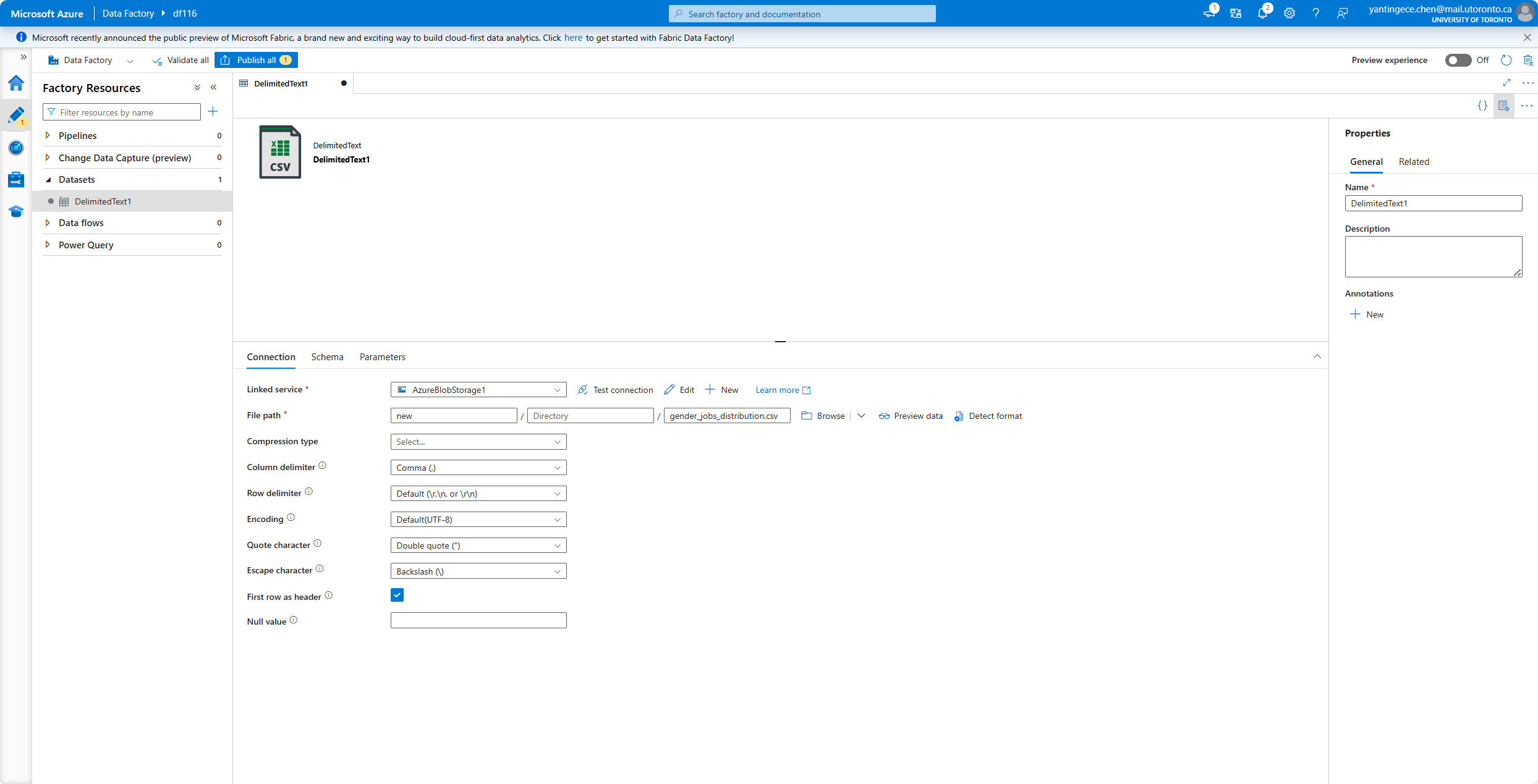
* Upload the CSV File to Blob Storage:
* Linked Service to Blob Storage



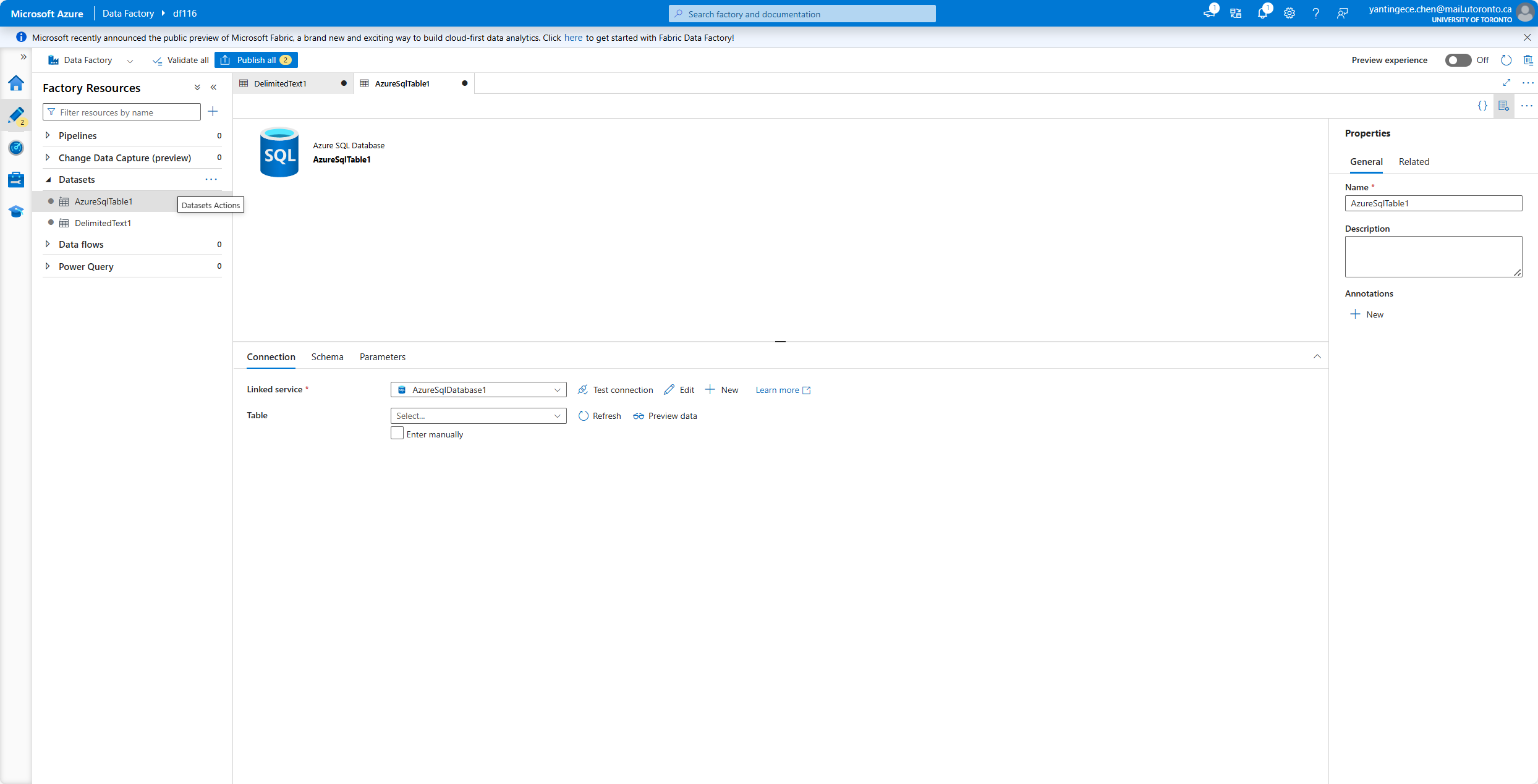
* Linked Service to Azure SQL Database



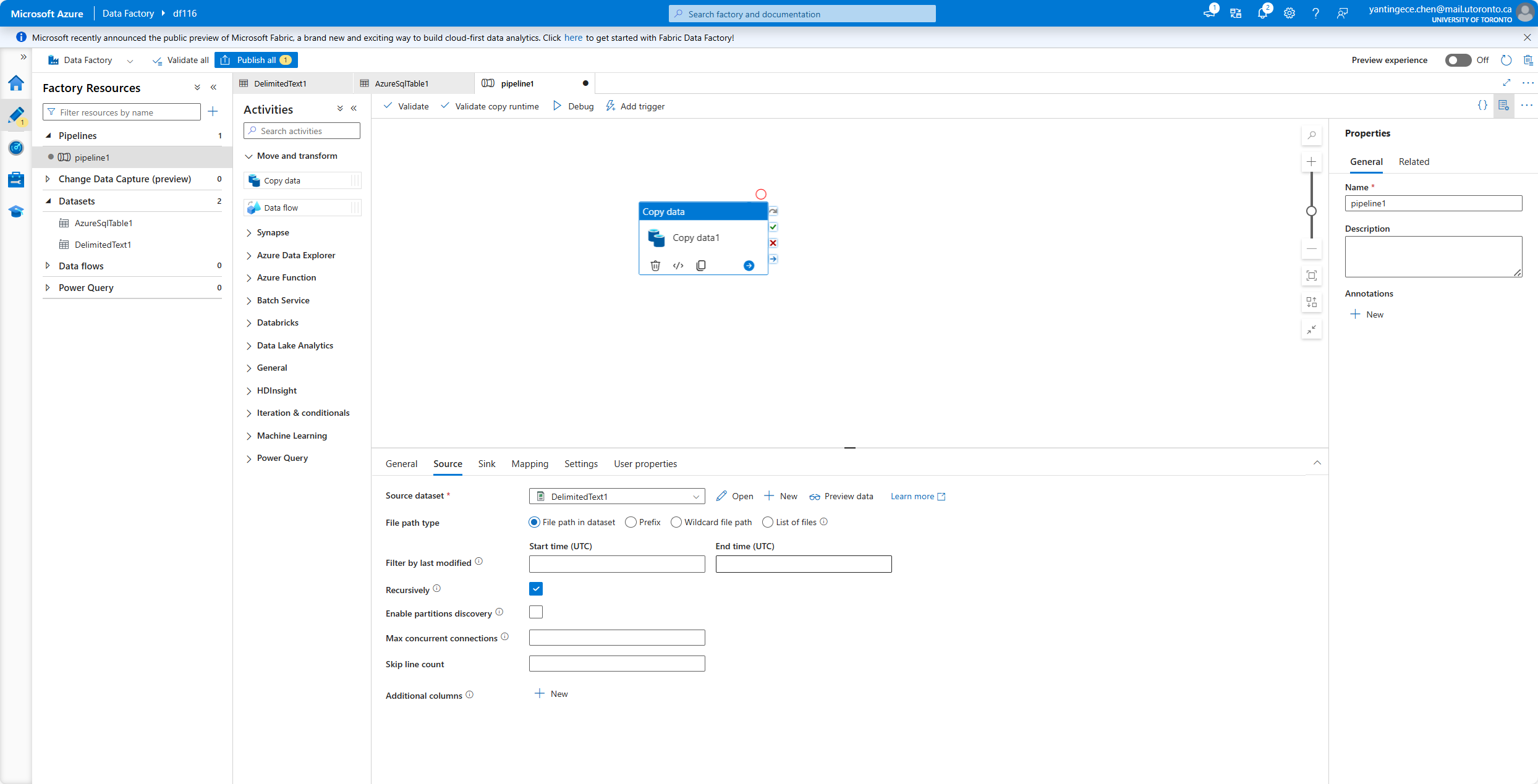
* Dataset for Blob Storage

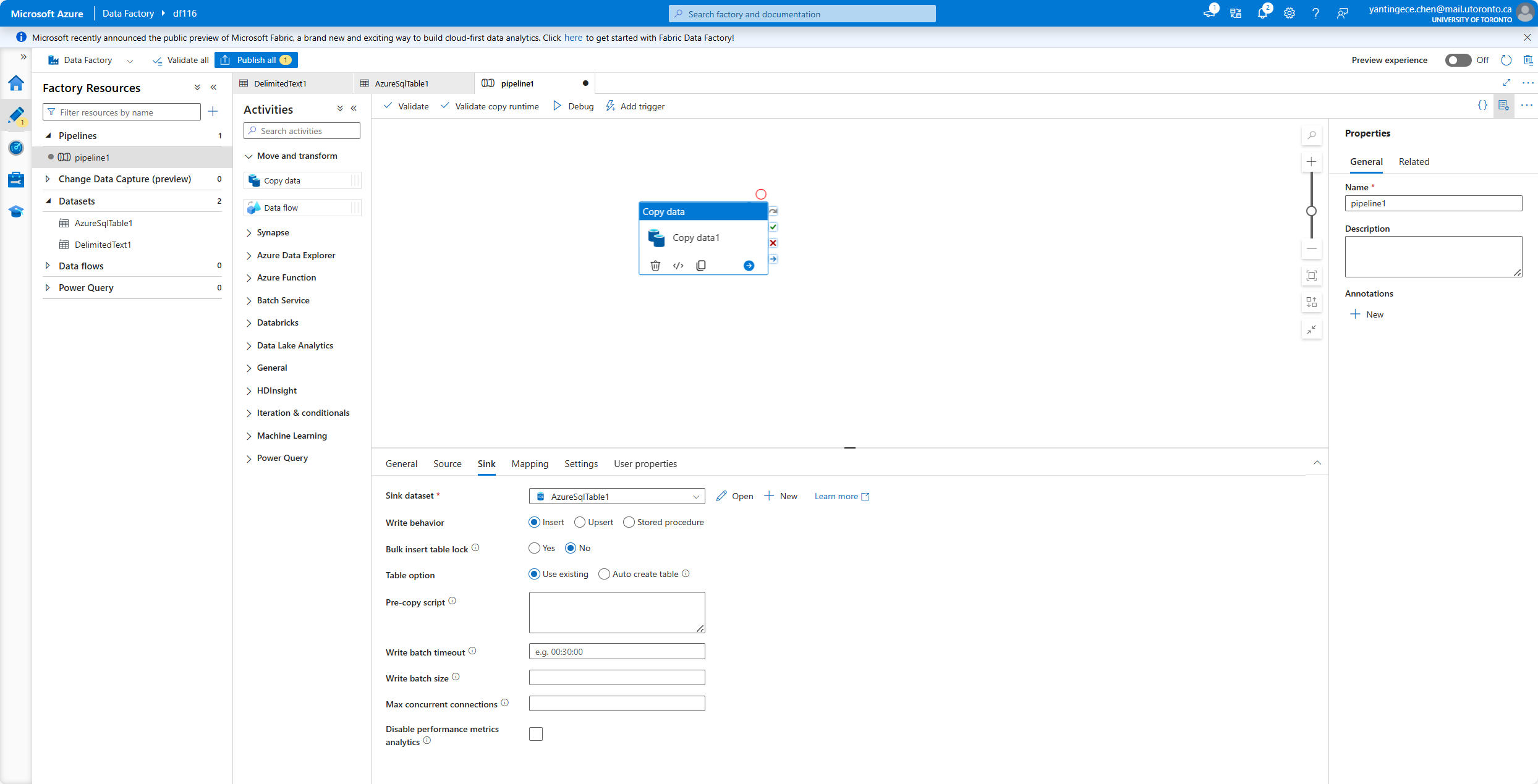


* Dataset for Azure SQL Database

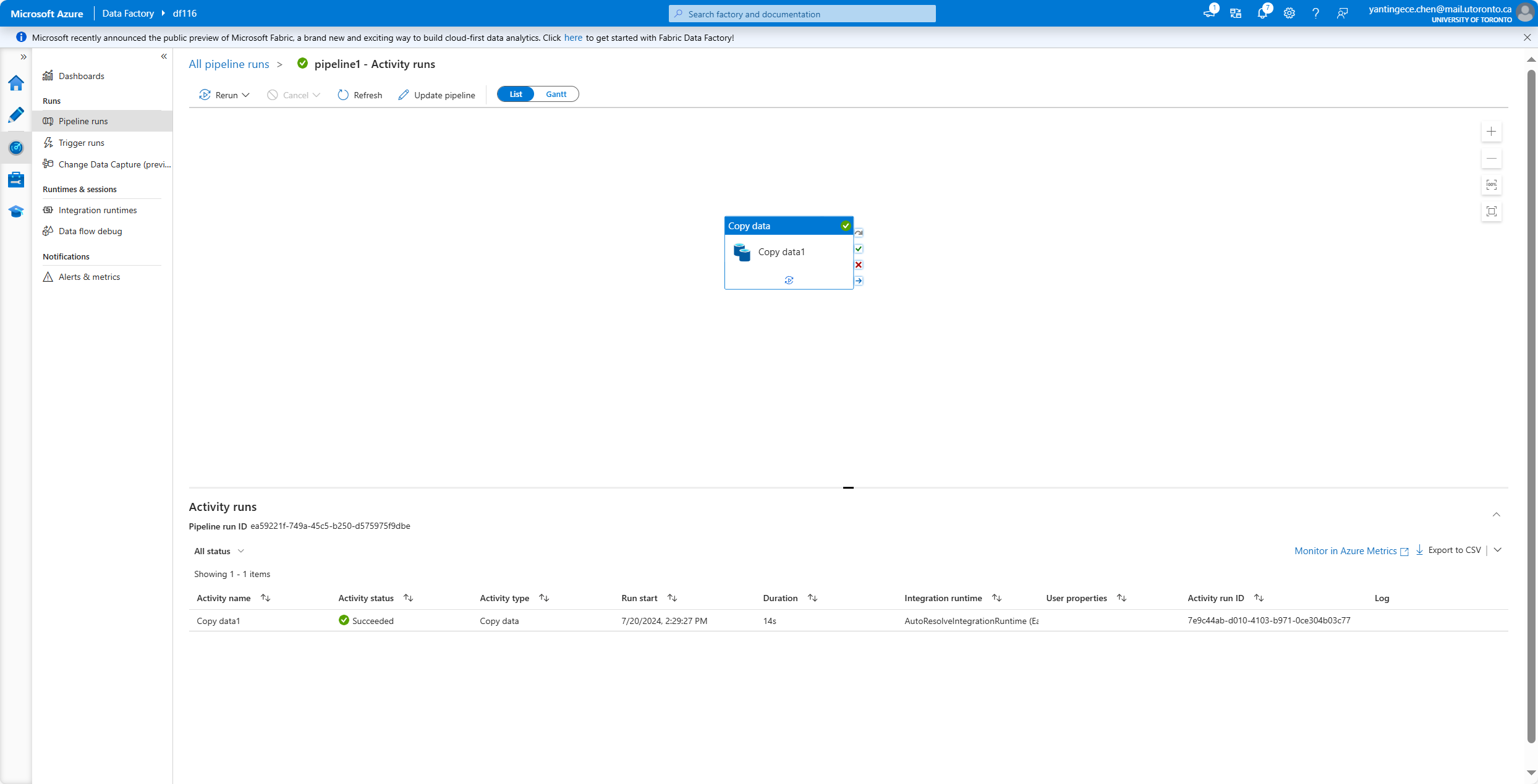


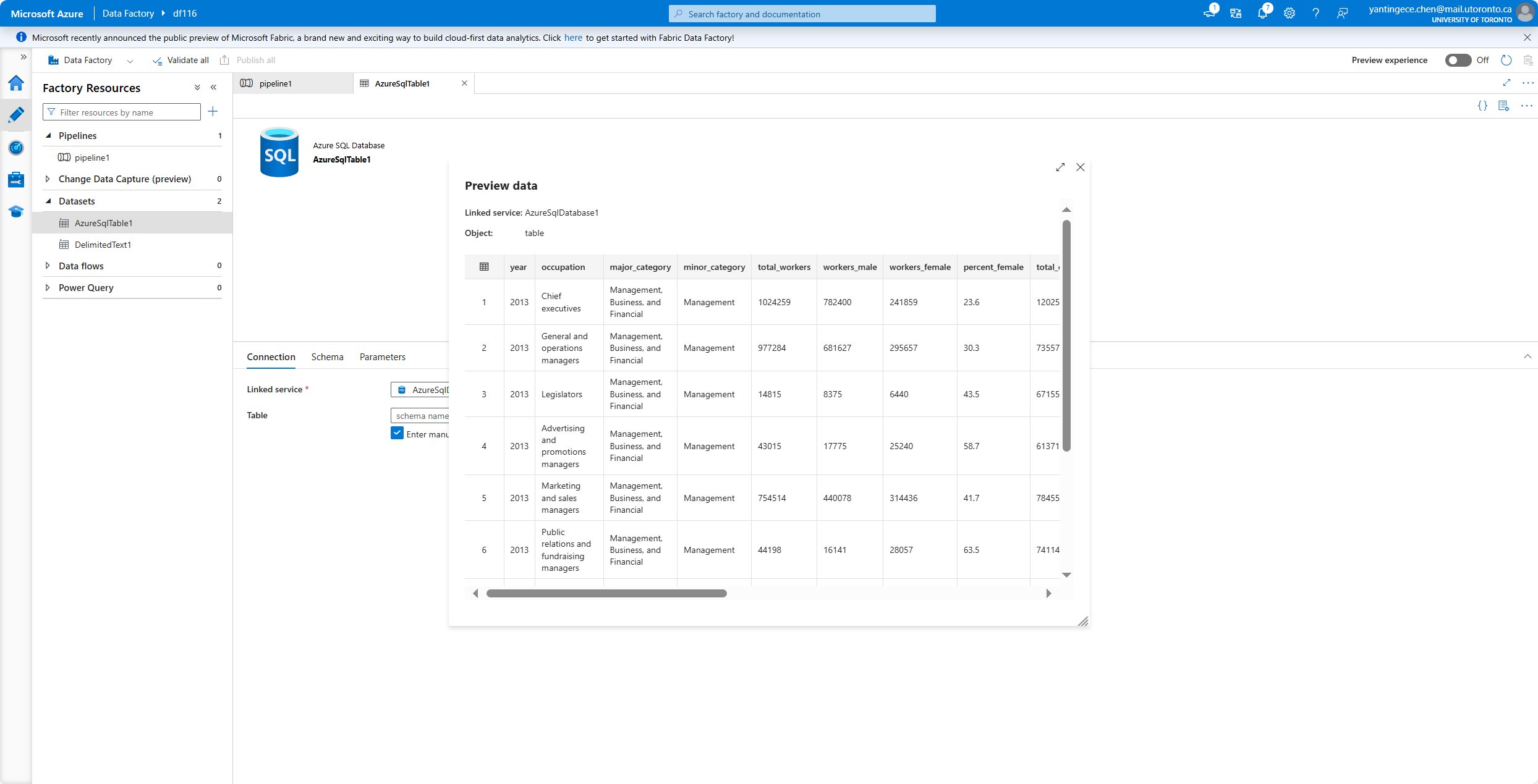
* Create a Pipeline





* Publish and Trigger the Pipeline





1. Explain the different types of triggers available in ADF. Now create a schedule trigger and run your pipeline every 3 minutes. Show 5 successful runs.

* Schedule Trigger:

Executes a pipeline on a specified schedule.

Useful for running pipelines at regular intervals, such as hourly, daily, or weekly.

* Tumbling Window Trigger:

Executes a pipeline in a series of fixed-size, non-overlapping time intervals (windows). Useful for scenarios where you need to process data in chunks, such as hourly data processing or daily batch jobs.

* Event-based Trigger:

Executes a pipeline based on events, such as the arrival or deletion of a file in Azure Blob Storage or Azure Data Lake Storage.

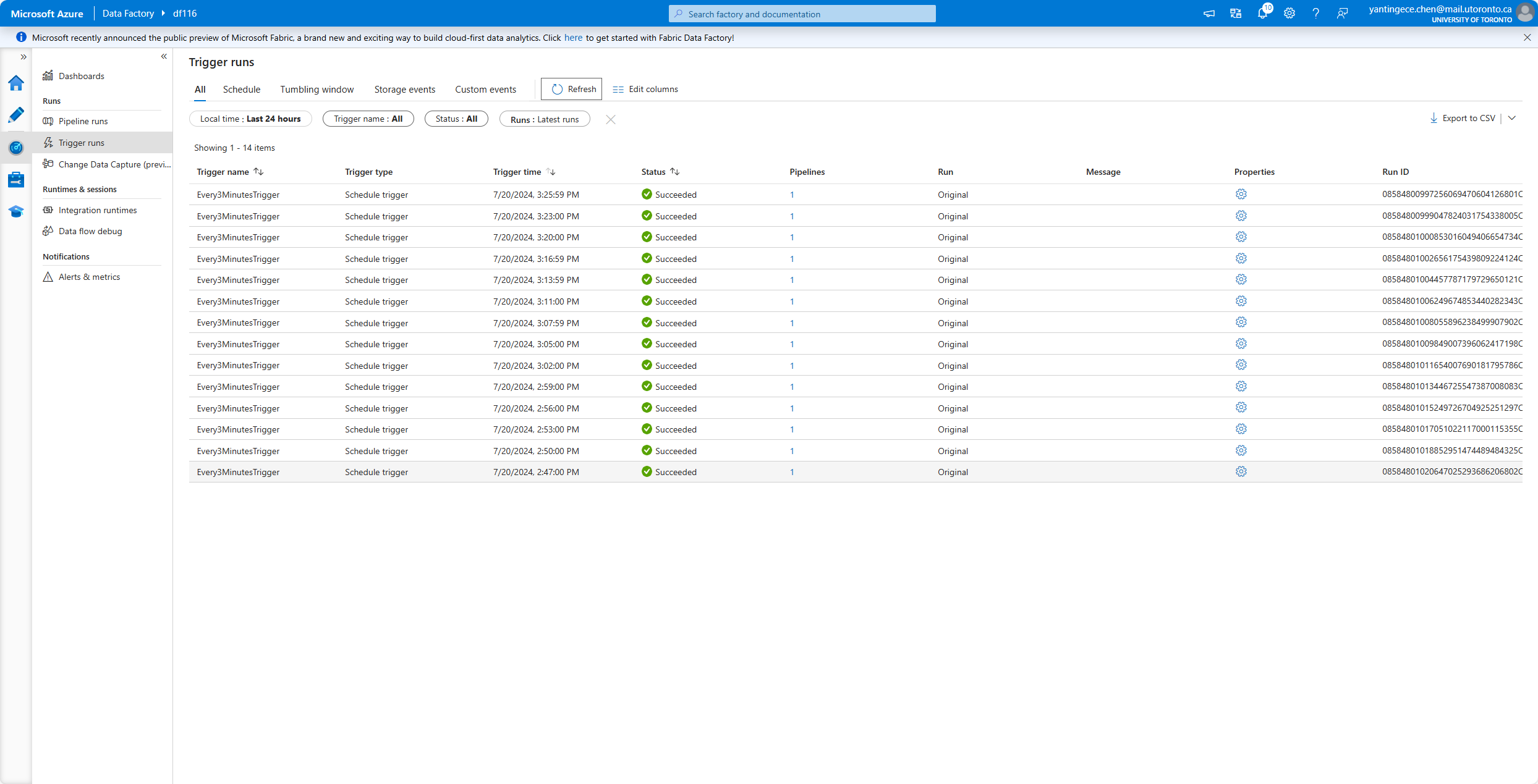
Useful for triggering pipelines when new data arrives.

* Manual Trigger:

Allows manual execution of a pipeline.

Useful for on-demand executions or for testing purposes.

* Show 5 successful runs.



1. A client needs to replicate objects from ADLS Gen 2 in Canada Central to ADLS Gen 2 in West Europe. Let’s say they want to do this in a bi-directional way. How can you set this up?

To set up bi-directional replication between Azure Data Lake Storage (ADLS) Gen2 accounts in Azure, we can use Azure Data Factory (ADF) and event triggers. Below is a summary of the steps to implement this:

* Create Linked Services

Create two linked services in ADF:

1. ADLS Canada Central: connects to the ADLS Gen2 account located in Central Canada.

2. ADLS West Europe: connects to an ADLS Gen2 account located in Western Europe.

* Create datasets

Create two datasets, both binary, for each ADLS account:

1. DatasetADLSCanada: points to the ADLS in Central Canada.

2. DatasetADLSEurope: points to the ADLS for Western Europe.

* Create a Data Factory Pipeline

1. Copy pipeline from Canada to Europe:

- Create a pipeline named `CopyCanadaToEurope`.

- Add a `Copy Data` activity and set the source dataset to `DatasetADLSCanada` and the destination dataset to `DatasetADLSEurope`.

2. Copy pipeline from Europe to Canada:

- Create a pipeline named `CopyEuropeToCanada`.

- Add the `Copy Data` activity and set the source dataset to `DatasetADLSEurope` and the destination dataset to `DatasetADLSCanada`.

* Create event triggers

1. Canada to Europe Event Trigger:

- Create `EventTriggerCanadaToEurope` to monitor the New Blob event in the ADLS for Central Canada.

- Attach the trigger to the `CopyCanadaToEurope` pipeline.

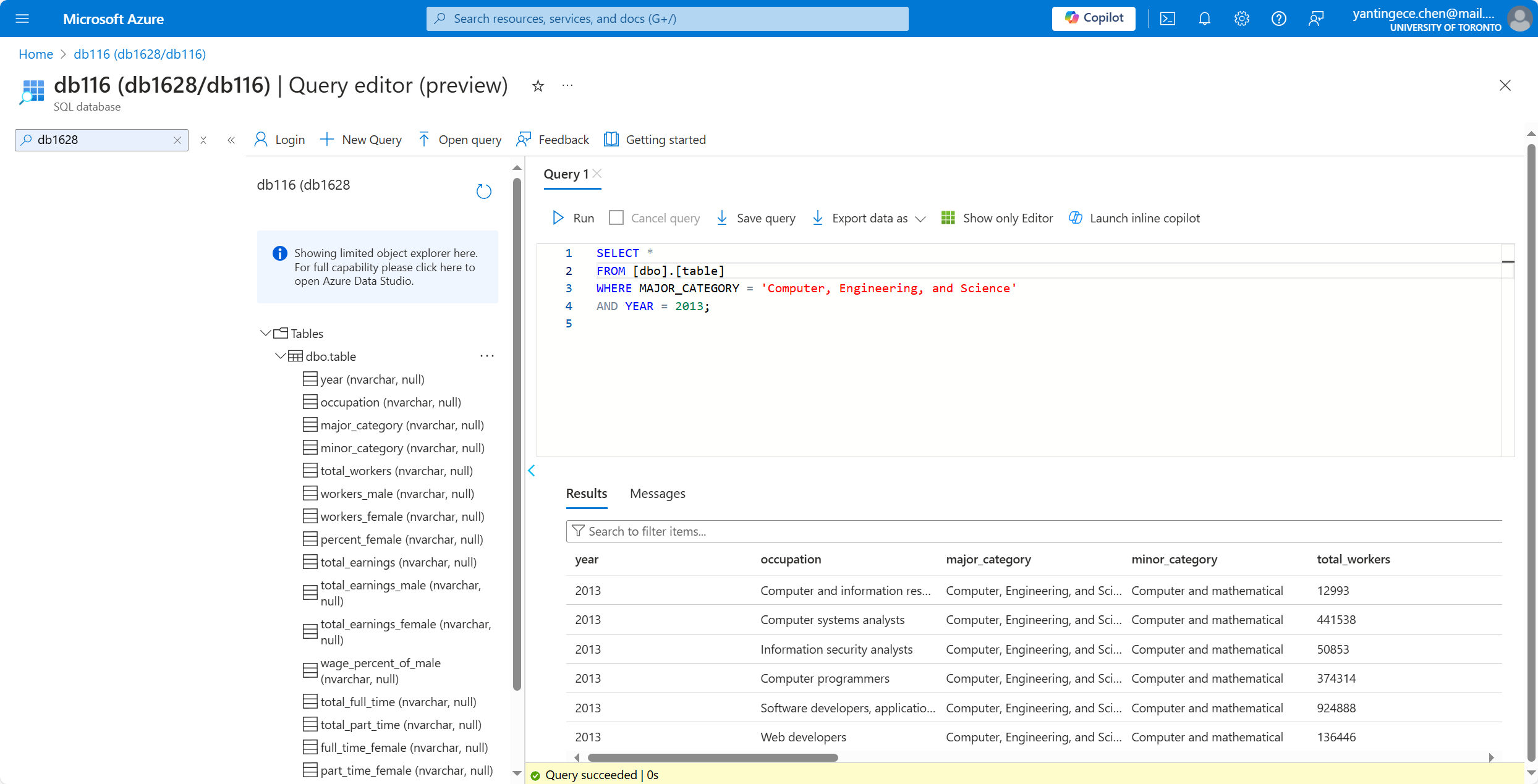
2. Event Trigger Europe to Canada:

- Create `EventTriggerEuropeToCanada` to monitor the New Blob event in the ADLS for Western Europe.

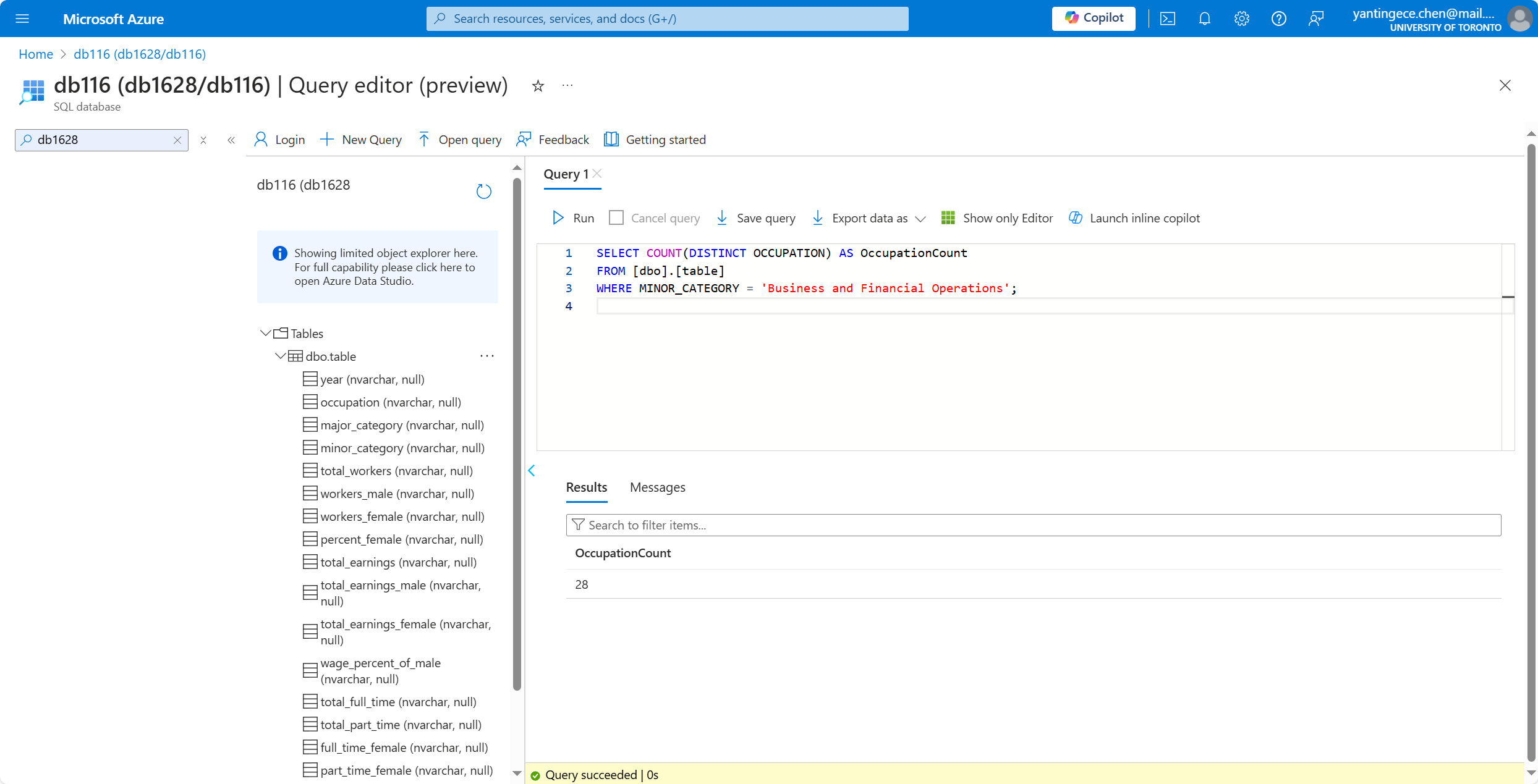
- Attach the trigger to the `CopyEuropeToCanada` pipeline.

### PART B:

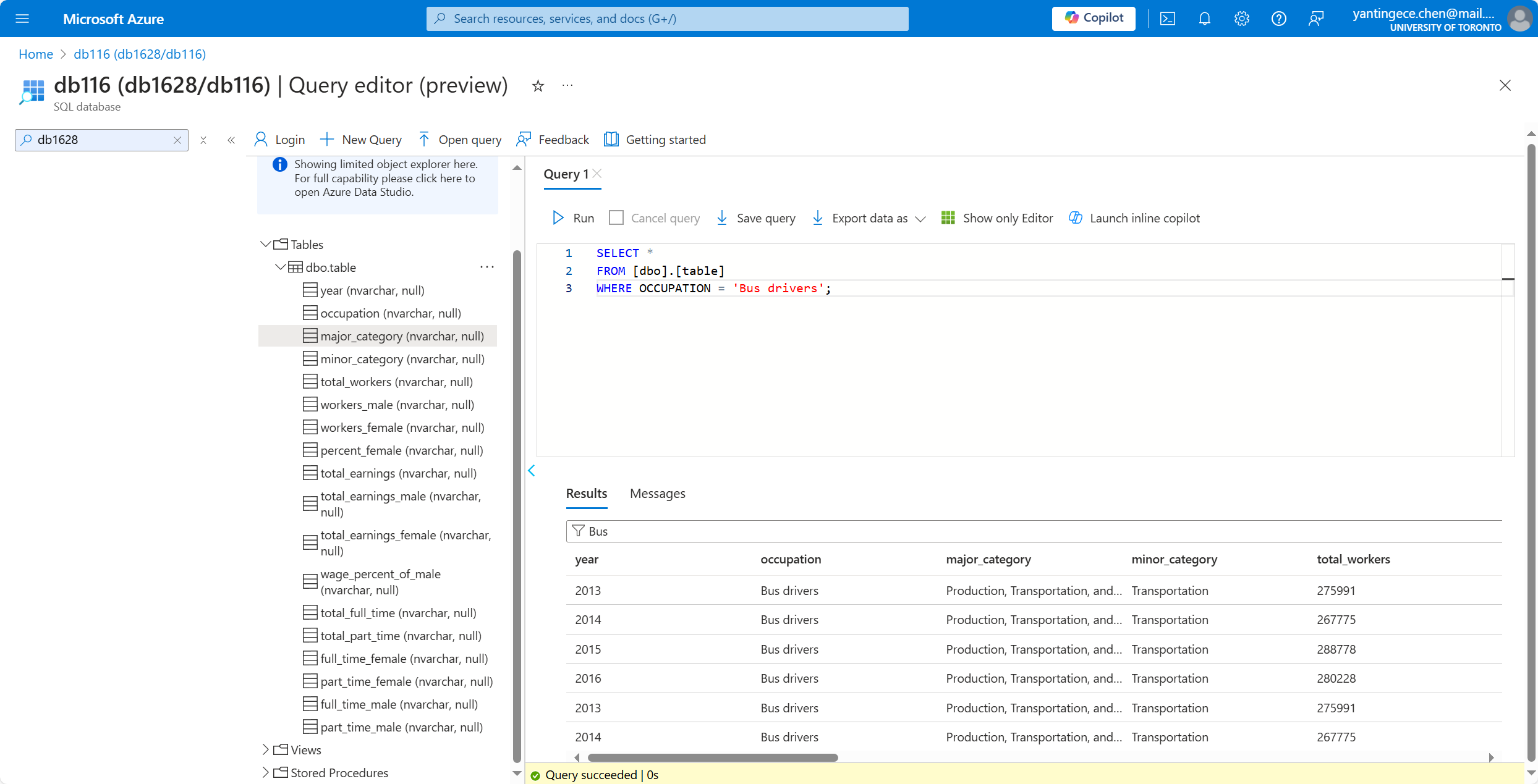
1. In the gender\_jobs\_data table - Filter all the OCCUPATIONS in MAJOR\_CATEGORY of Computer, Engineering, and Science for the YEAR 2013



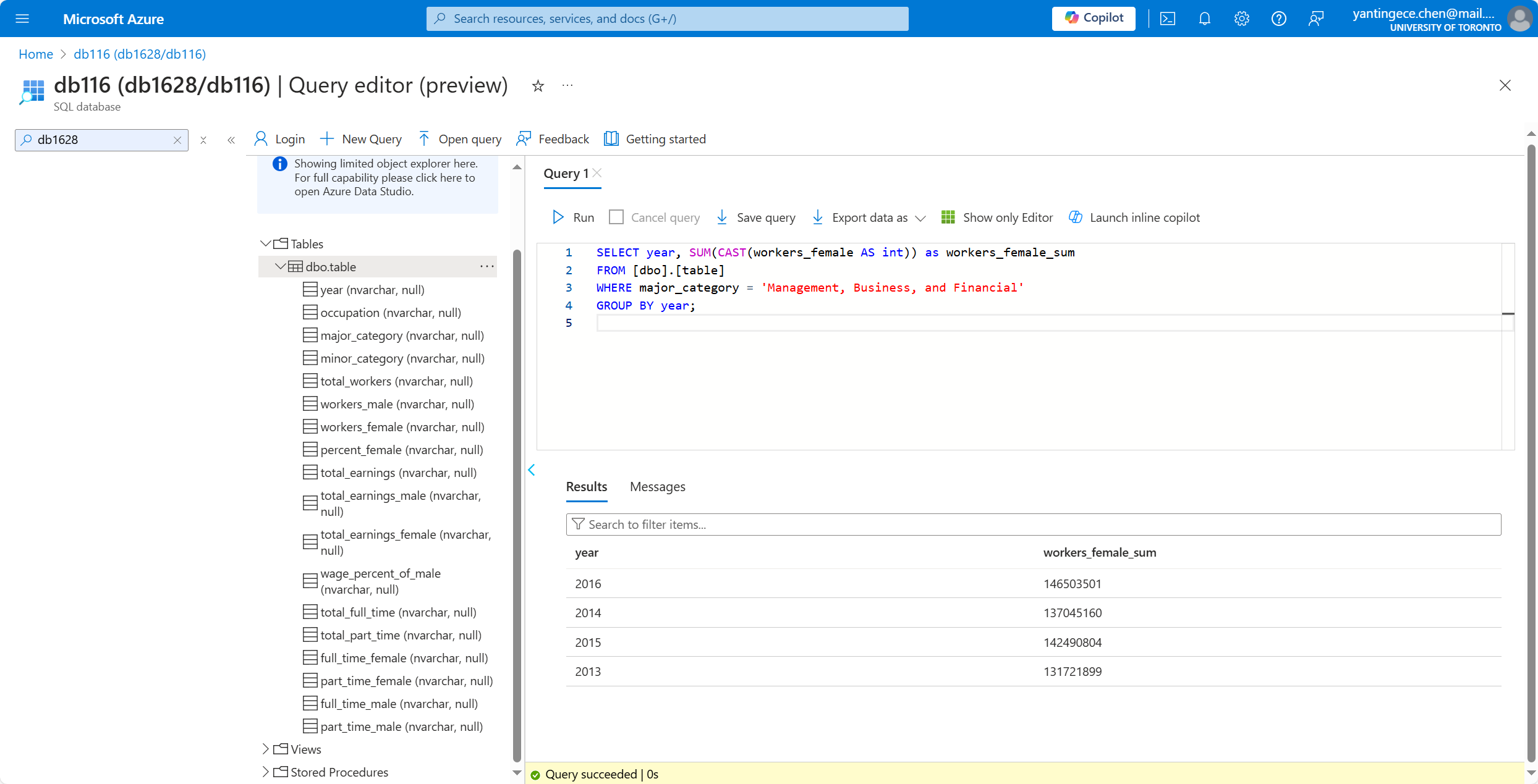
2. In the gender\_jobs\_data table - How many OCCUPATIONS exist in the MINOR\_CATEGORY of Business and Financial Operations overall?



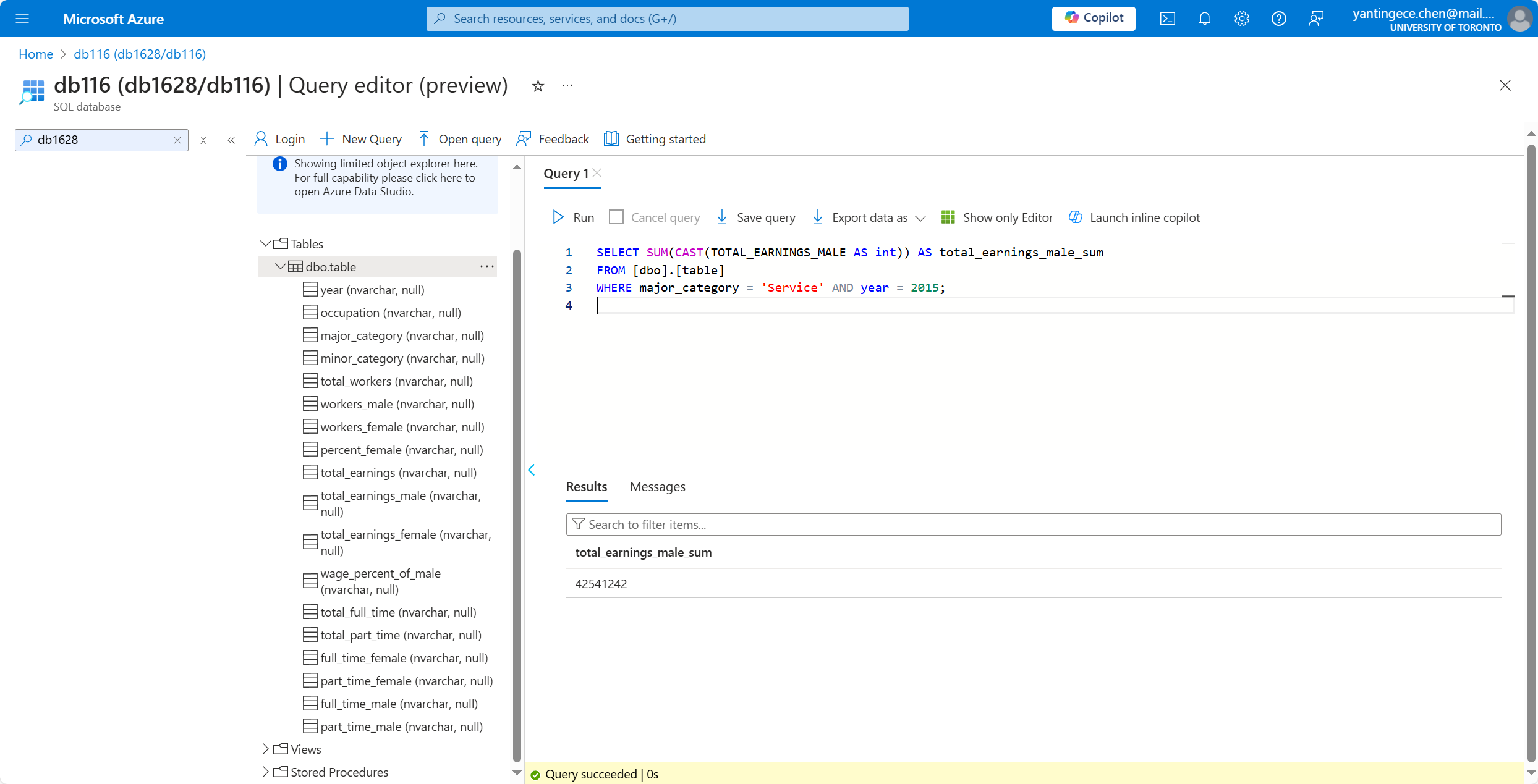
1. In the gender\_jobs\_data table - Get all relevant information for bus drivers across all years



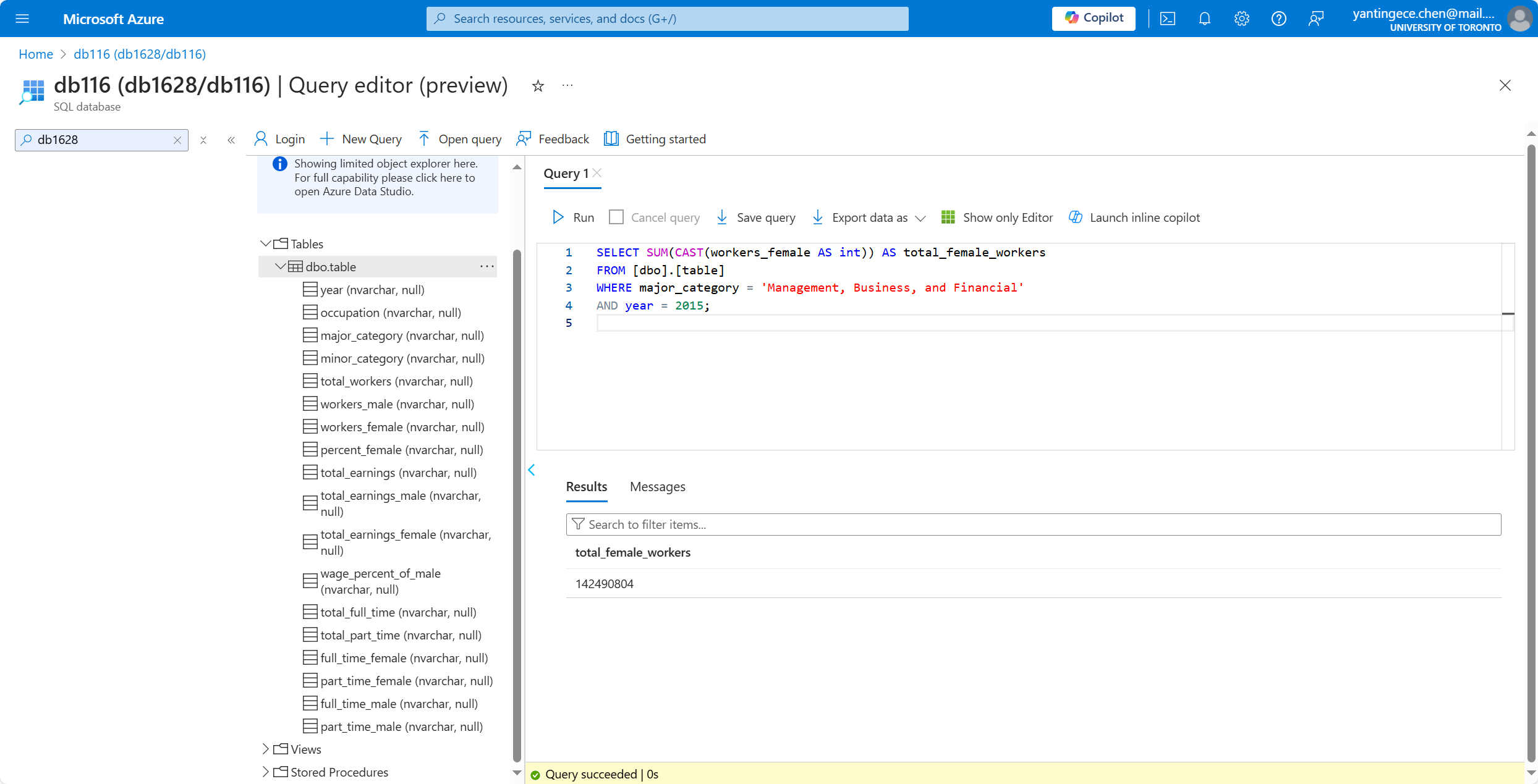
1. In the gender\_jobs\_data table - Summarize the total number of WORKERS\_FEMALE in the MAJOR\_CATEGORY of Management, Business, and Financial by each year.



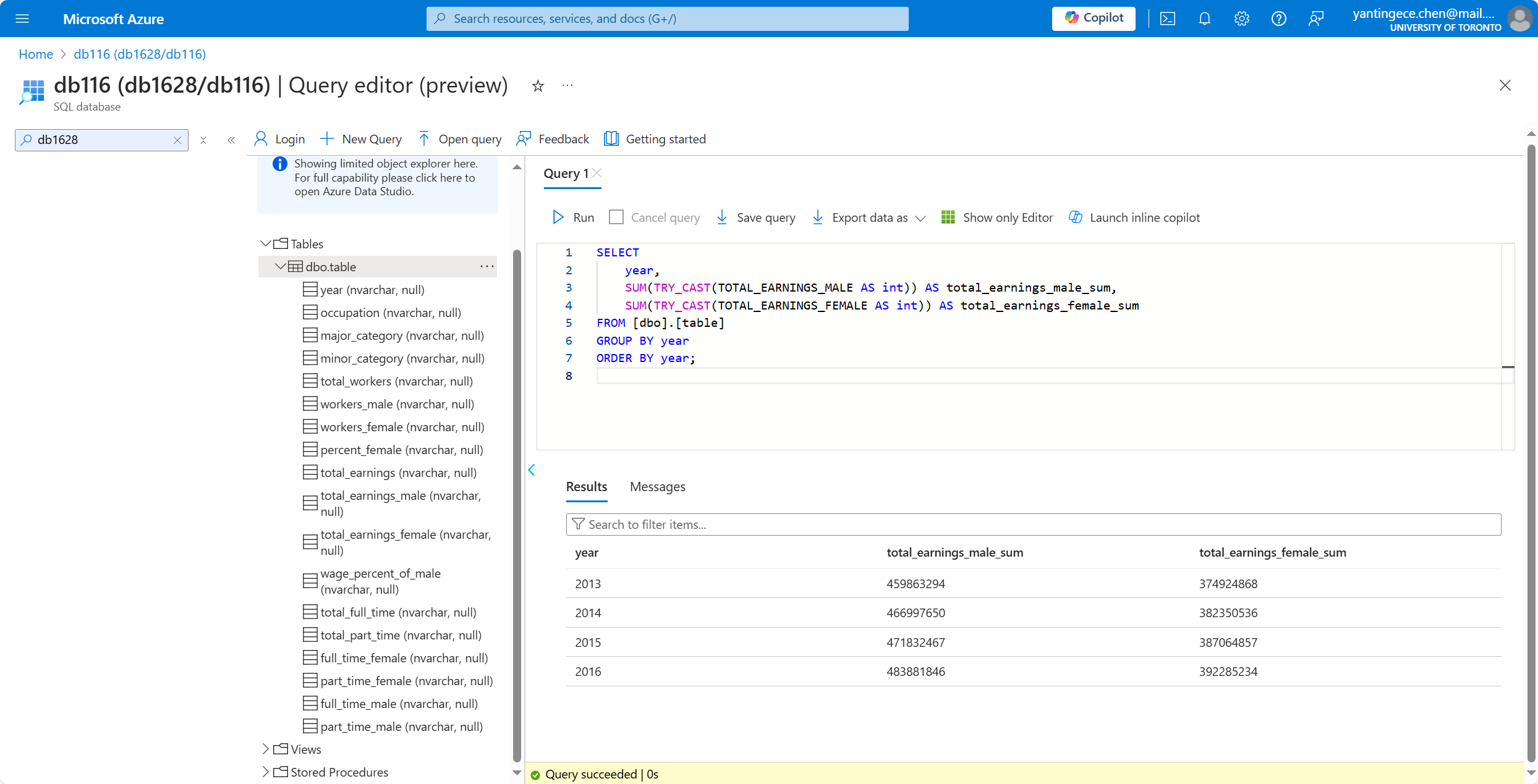
1. In the gender\_jobs\_data table - What were the total earnings of male (TOTAL\_EARNINGS\_MALE) employees in the Service MAJOR\_CATEGORY for the year 2015?



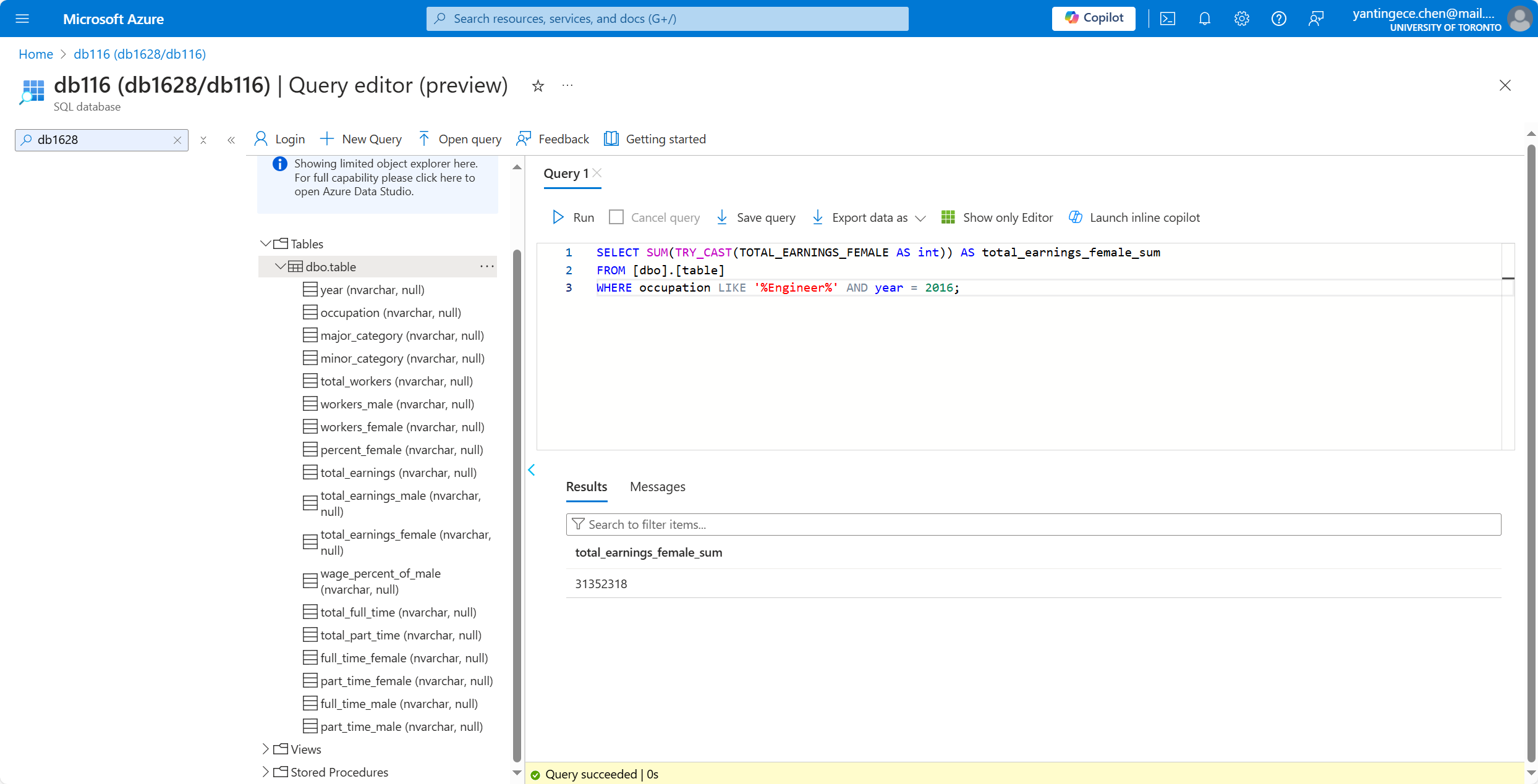
1. How many female workers were in management roles in the year 2015?



1. Compare the TOTAL\_EARNINGS\_MALE and TOTAL\_EARNINGS\_FEMALE earnings irrespective of occupation by each year



1. How much money (TOTAL\_EARNINGS\_FEMALE) did female workers make as engineers in 2016?



1. What is the total number of full-time and part-time female workers versus male workers year over year?

