1628 Assignment 4

Assignment on Azure Cloud Platform

Due by Jul 23, 2023

Ziruo Song

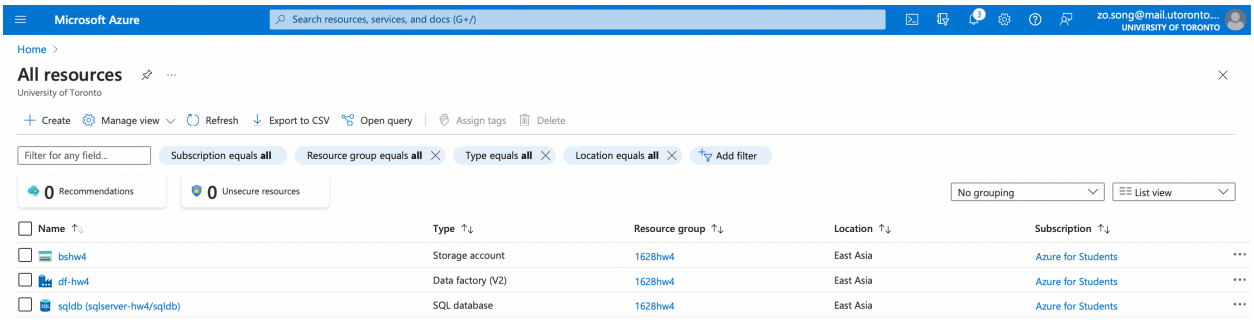
Part A:

For Part A: Submit all screenshots showing deployed resources in your Azure portal, Azure Blob Storage, Azure Data Factory, ADLS Gen 2 and Azure SQL DB including your account information at the top right corner of the webpage. Include the successful pipeline runs screenshots with triggers.

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.

   ANS:

   The resource group I created was named '1628hw4', and the 3 deployed resources were grouped as shown below.

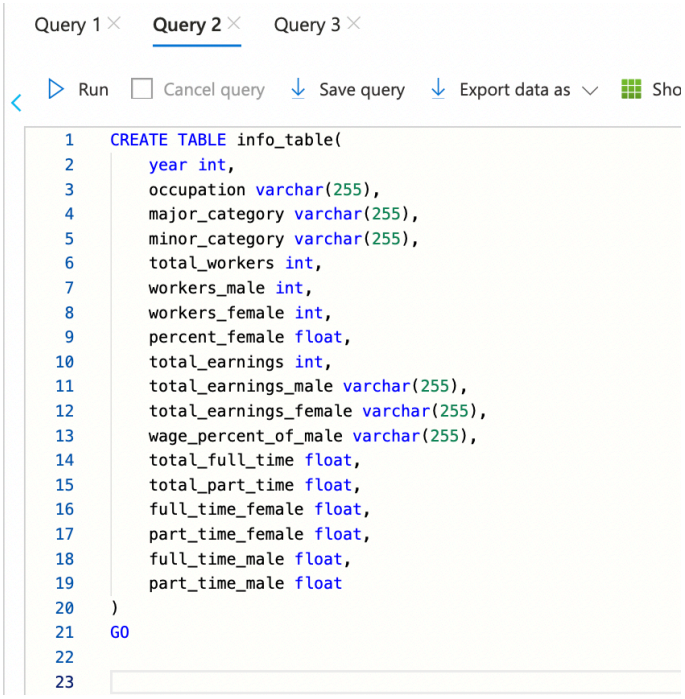


2. [Marks: 15] Now create a pipeline in Azure Data Factory and copy *gender_jobs_data.csv* file from the Blob storage account to Azure SQL DB. (First copy this file from your local machine to Blob Storage). See this https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-portal for reference.)

ANS:

Code for Sink table:

(*total_earnings_male*, *total_earnings_male*, and *total_earnings_male* contain NAs. Here I choose the method 2 according to the methods on Piazza.)



Preview of output data:

Preview data

Linked service: AzureSqlDatabase1

Object: dbo.info_table

	year	occupation	major_category	minor_category	total_workers	workers_male	workers_female	p
1	2013	Chief executives	Management, Business, and Financial	Management	1024259	782400	241859	2
2	2013	General and operations managers	Management, Business, and Financial	Management	977284	681627	295657	3
3	2013	Legislators	Management, Business, and Financial	Management	14815	8375	6440	4
4	2013	Advertising and promotions managers	Management, Business, and Financial	Management	43015	17775	25240	5
5	2013	Marketing and sales managers	Management, Business, and Financial	Management	754514	440078	314436	4
		Public relations and	Management,					

Success of pipeline:

Parameters Variables Settings **Output**

Pipeline run ID: 98cb167b-88f3-4b05-91f6-f9b11f313bfb

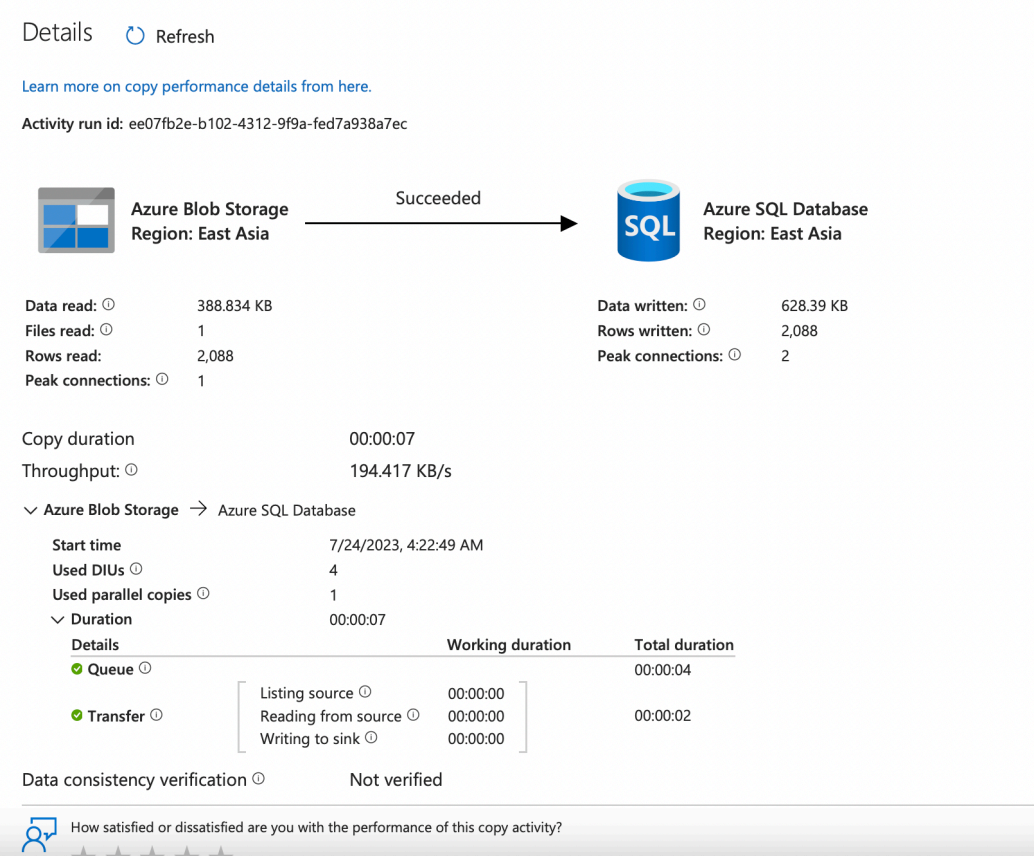
View debug run

All status

Export to CSV

Showing 1 - 1 of 1 items

Activity name	Status	Activity type	Run start
CopyFromBlobToSql	Succeeded	Copy data	7/24/2023, 4:22:48 AM



3. [Marks: 10] 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.

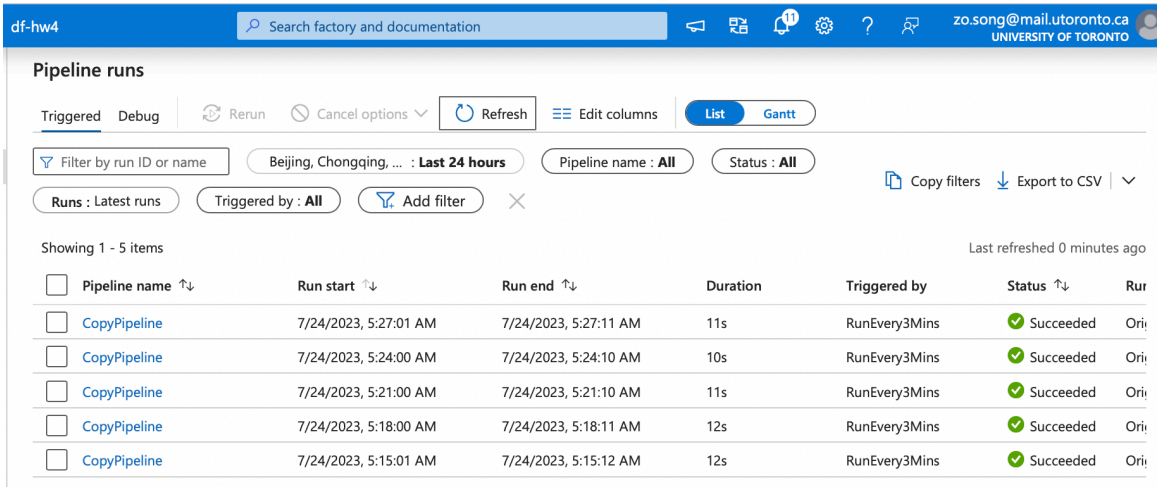
ANS:

Explanation: Based on the official document of Azure and other resource online, there are three types of trigger offered by Azure Data Factory (ADF) right now:

- Schedule trigger, it runs pipelines on a predefined schedule with flexible time intervals, like daily, or every 5 mins.
- Tumbling Window Trigger: Executes Data Pipelines at specific time slices or intervals, which mean this is quite useful for historical data scenarios.Compared with schedule trigger above, Tumbling Window Triggers support Backfill scenarios, Retry capability, concurrency, and but only one pipeline can be triggered.

- Event-based Trigger: It responds to blob-related events, triggering pipelines based on changes in data sources.

5 successful runs:



4. [Marks: 20] 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?

ANS:

By hint, to establish bi-directional replication between ADLS Gen 2 storage accounts located in Canada Central and West Europe, we can utilize Azure Data Factory with Event Triggers. Firstly, create ADLS Gen 2 storage accounts in both Canada Central and West Europe regions, and then set up Linked Services and defining datasets, like what we did for the above sections. However, we can choose to build two separate pipelines, instead of one, to do it bi-directionally. Also configure Copy Activities to effectively transfer the data between the two storage accounts, and we can create Event Triggers to trigger the respective pipelines based on changes in data sources.

PART B:

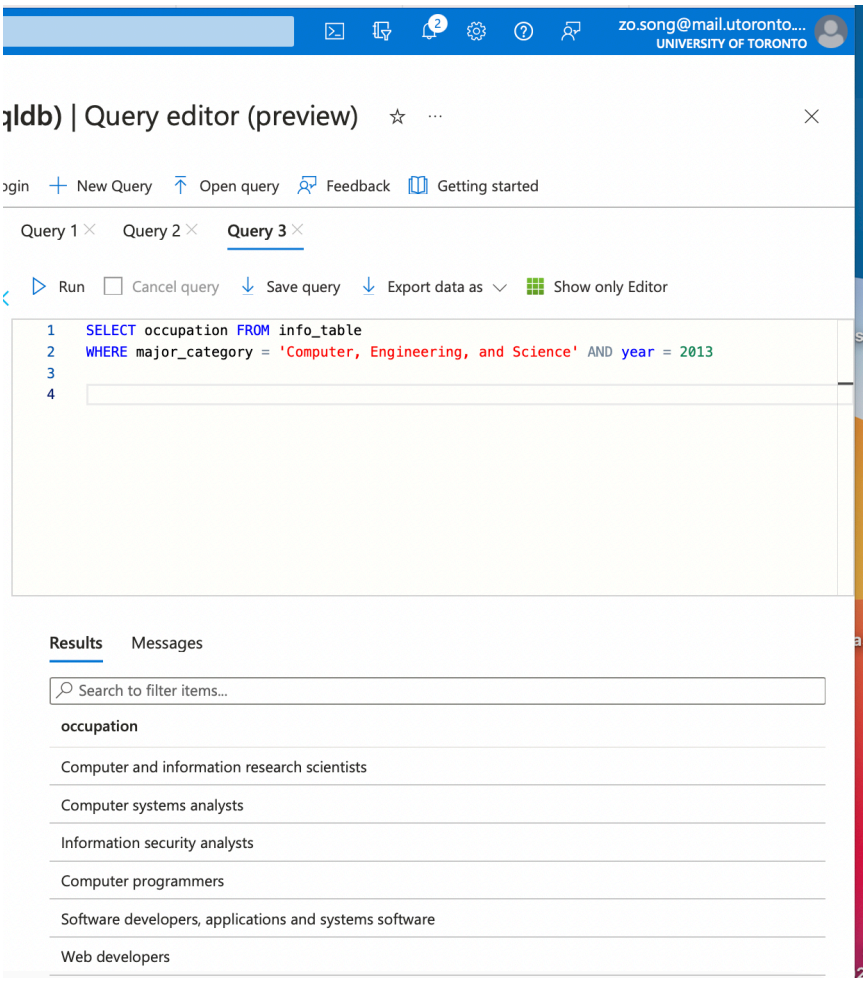
For Part B: Please submit all screenshots showing deployed resources in your Azure portal, Azure SQL DB and Query Editor screenshots where you run your code with output.

In this part, you will use Query Editor in Azure SQL DB and use the *gender_jobs_data.csv* table to perform the below queries.

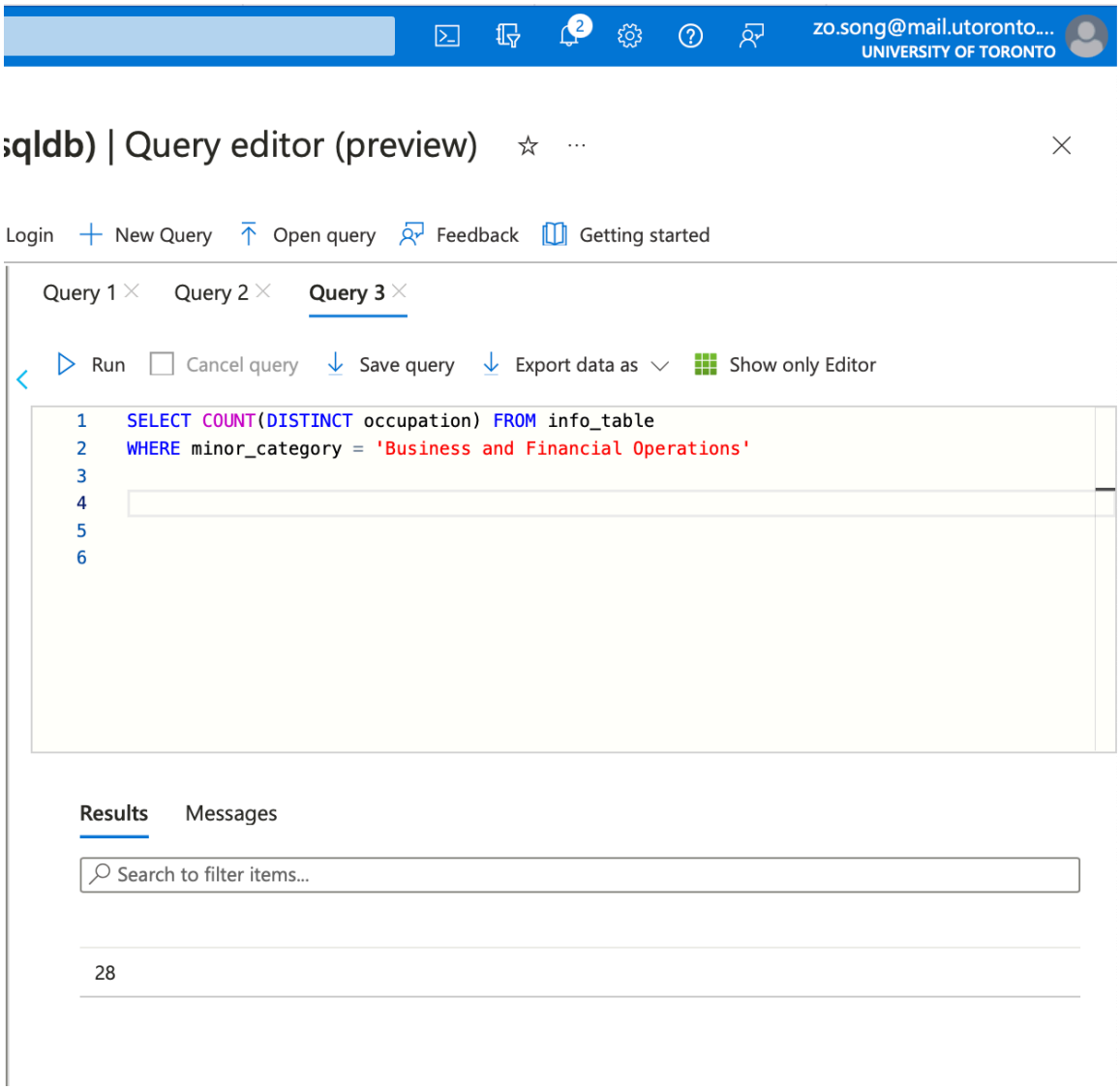
Implementation

You need to use Azure SQL Database for this part.

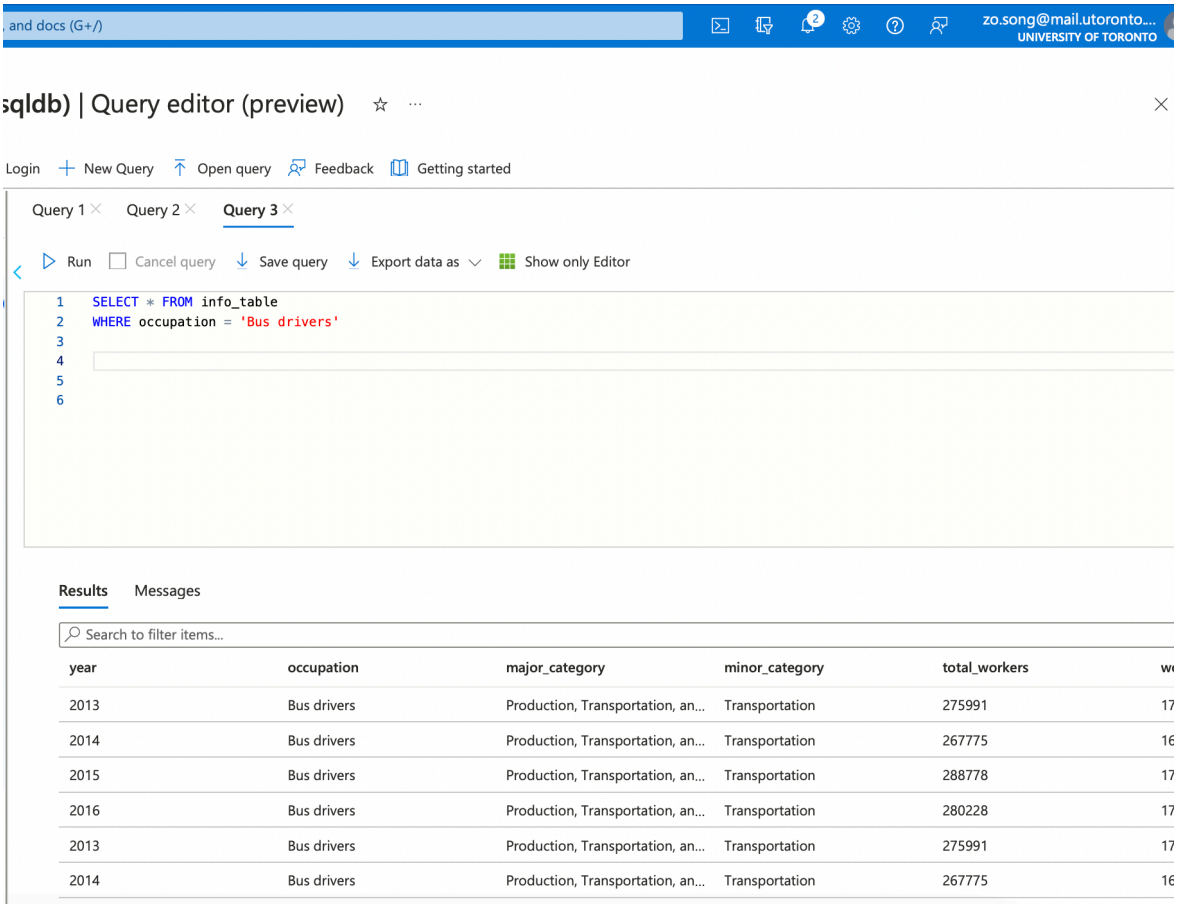
1. [Marks:5] In the *gender_jobs_data* table - Filter all the OCCUPATIONS in MAJOR_CATEGORY of Computer, Engineering, and Science for the YEAR 2013



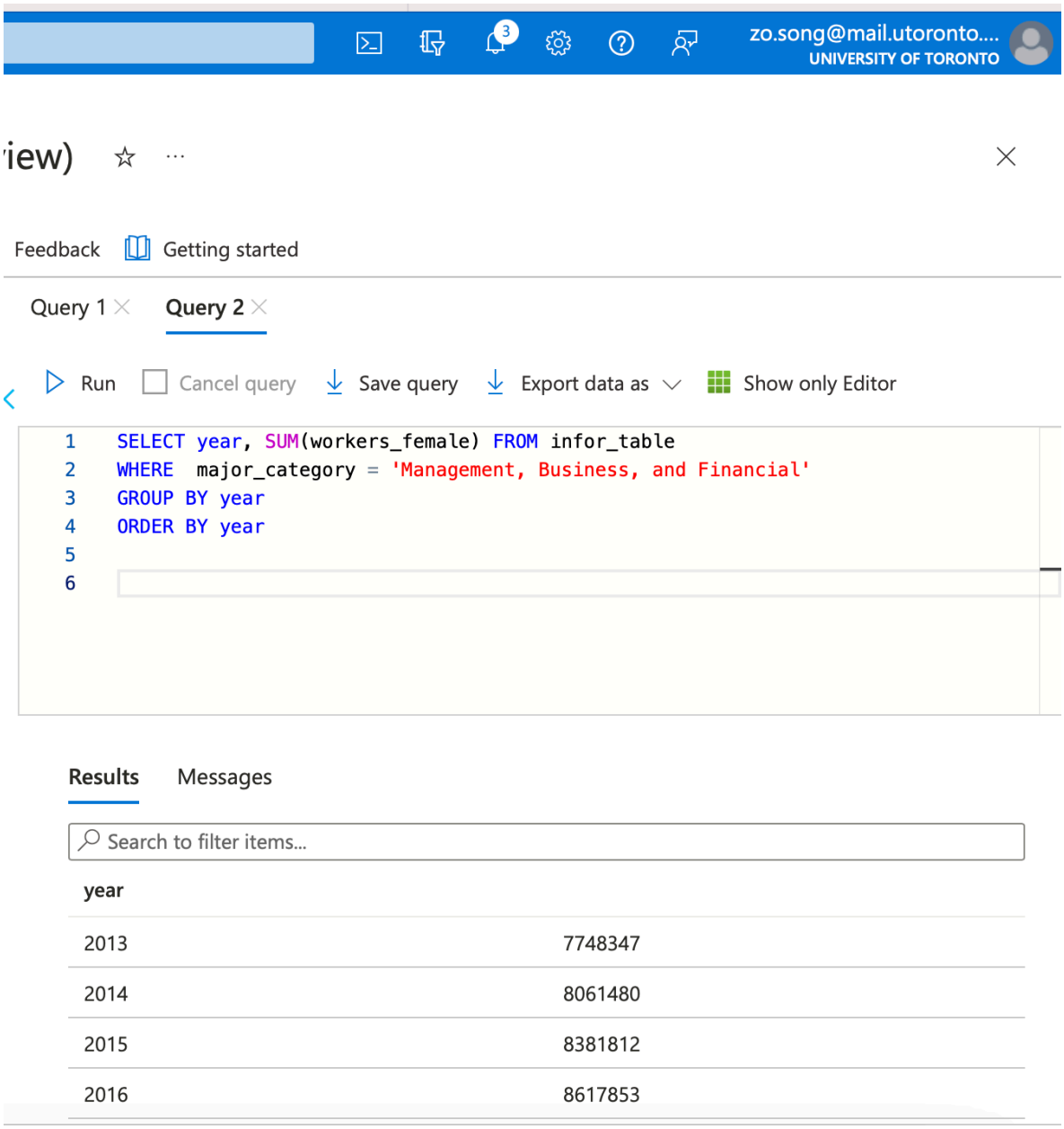
2. [Marks:5] In the *gender_jobs_data* table - How many OCCUPATIONS exist in the MINOR_CATEGORY of Business and Financial Operations overall?



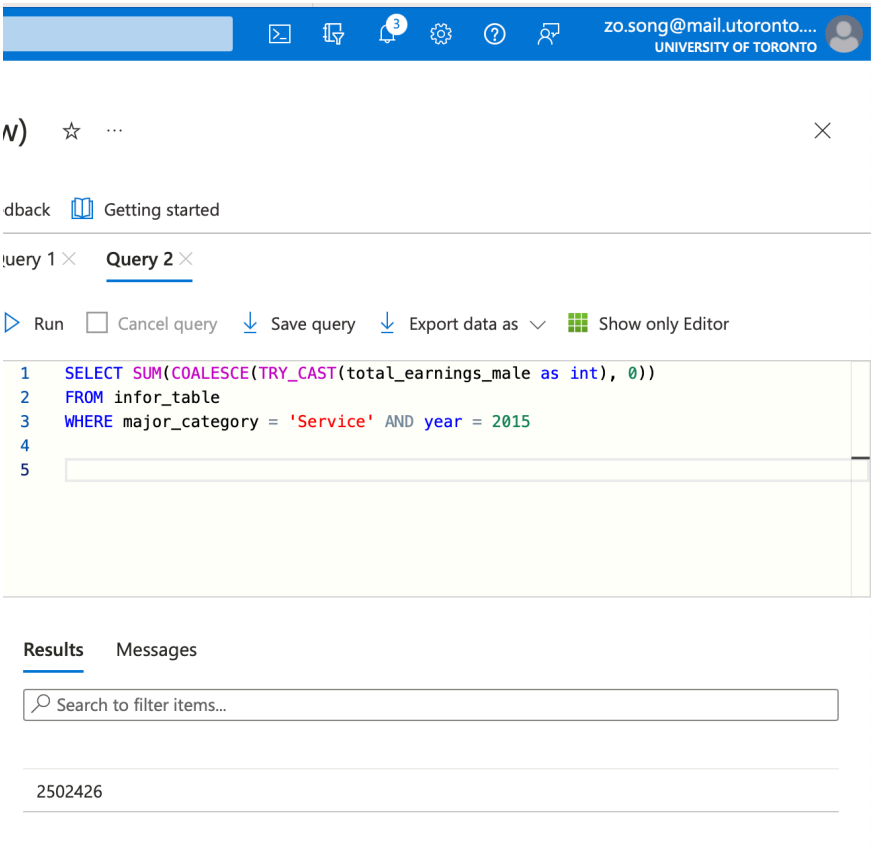
3. [Marks:5] In the *gender_jobs_data* table - Get all relevant information for bus drivers across all years



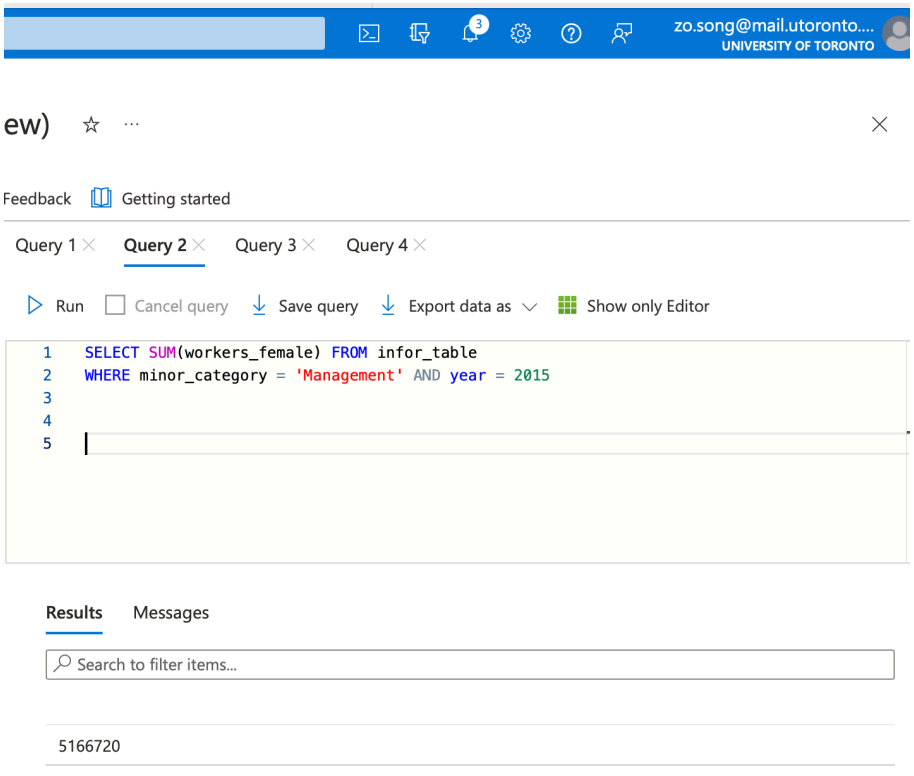
4. [Marks:5] In the *gender_jobs_data* table - Summarize the total number of WORKERS_FEMALE in the MAJOR_CATEGORY of Management, Business, and Financial by each year.



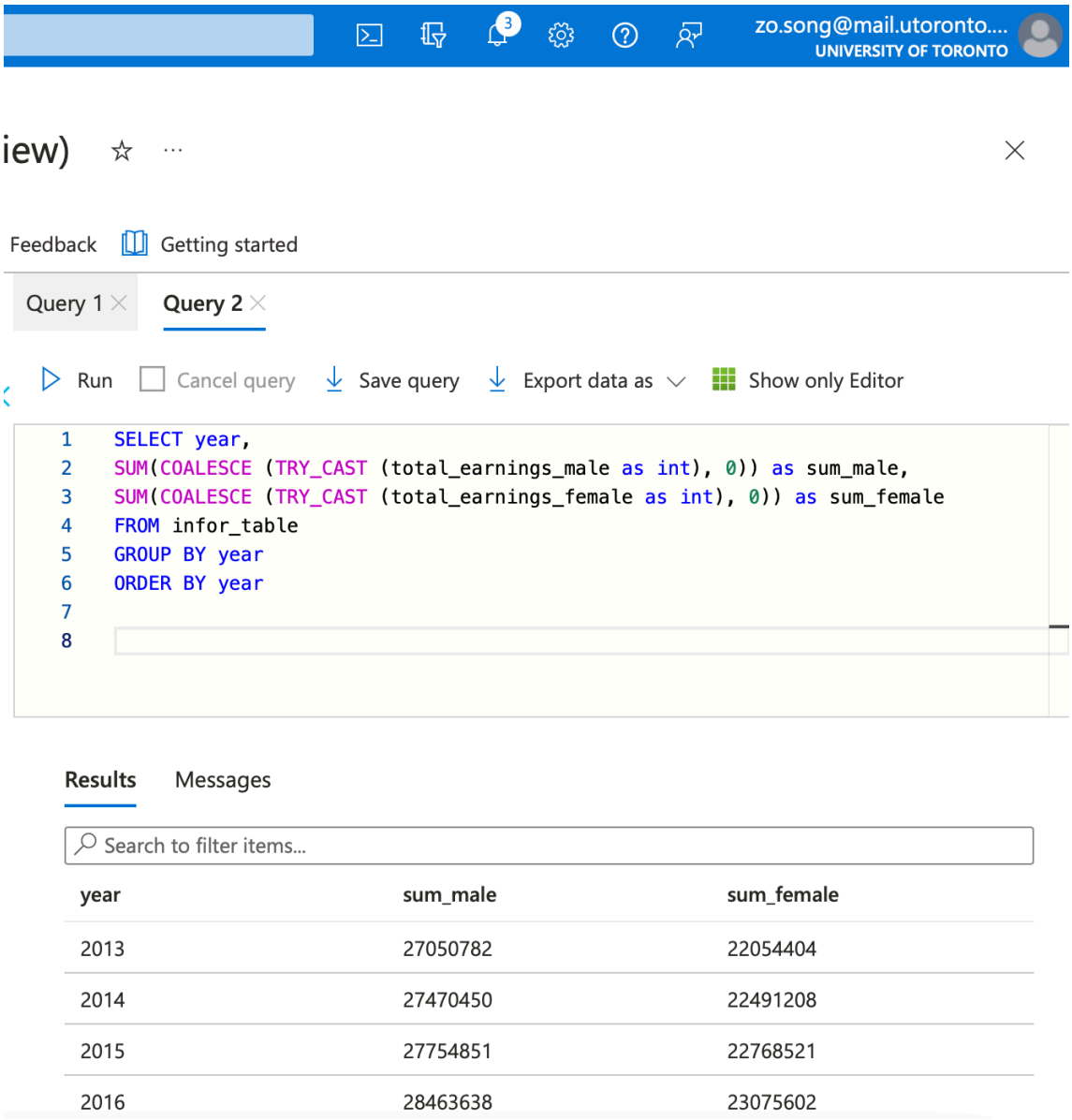
5. [Marks:5] In the *gender_jobs_data* table - What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?



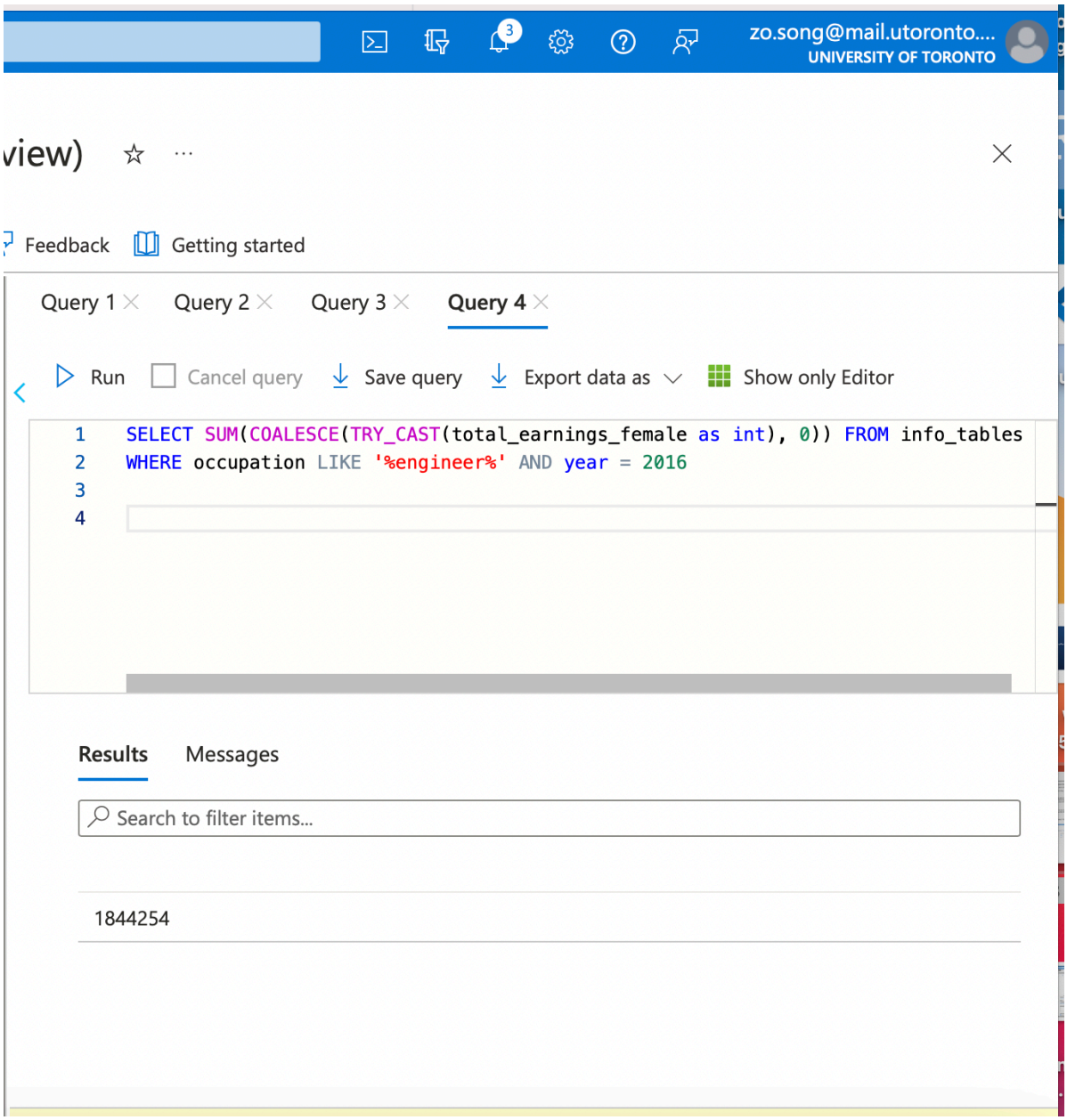
6. [Marks:5] In the *gender_jobs_data* table - How many female workers were in management roles in the year 2015?



7. [Marks:5] In the *gender_jobs_data* table - Compare the TOTAL_EARNINGS_MALE and TOTAL_EARNINGS_FEMALE earnings irrespective of occupation by each year

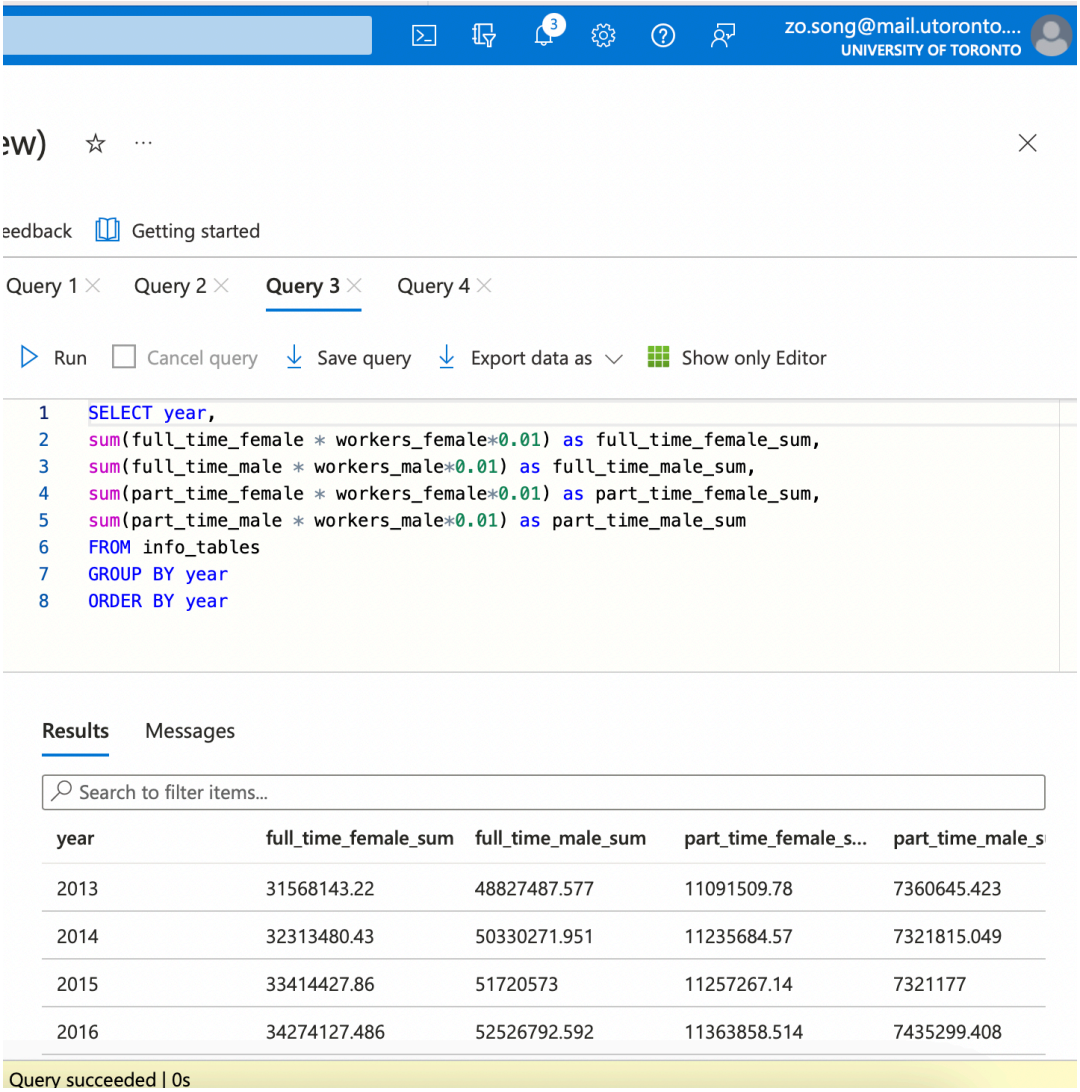


8. [Marks:5] In the *gender_jobs_data* table - How much money (TOTAL_EARNINGS_FEMALE) did female workers make as engineers in 2016?



9. [Marks:10] What is the total number of full-time and part-time female workers versus male workers year over year?

   - calculate by multiply each rate with the number.


zo.song@mail.utoronto....
UNIVERSITY OF TORONTO
eedback
Getting started
Query 1
Query 2
Query 3
Query 4
Run
Cancel query
Save query
Export data as
Show only Editor
1 SELECT year,
2 sum(full_time_female * workers_female*0.01) as full_time_female_sum,
3 sum(full_time_male * workers_male*0.01) as full_time_male_sum,
4 sum(part_time_female * workers_female*0.01) as part_time_female_sum,
5 sum(part_time_male * workers_male*0.01) as part_time_male_sum
6 FROM info_tables
7 GROUP BY year
8 ORDER BY year
Results
Messages
Search to filter items...
year
full_time_female_sum
full_time_male_sum
part_time_female_s...
part_time_male_s
2013
31568143.22
48827487.577
11091509.78
7360645.423
2014
32313480.43
50330271.951
11235684.57
7321815.049
2015
33414427.86
51720573
11257267.14
7321177
2016
34274127.486
52526792.592
11363858.514
7435299.408
Query succeeded | 0s