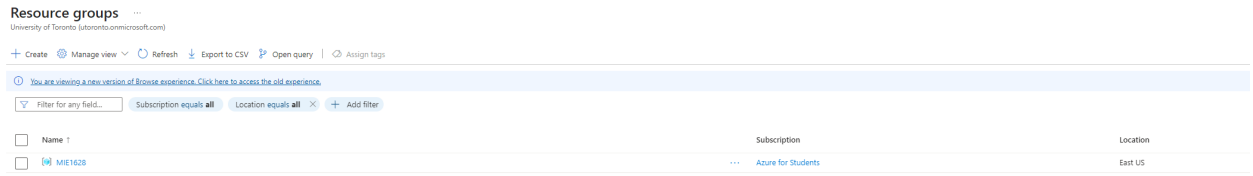
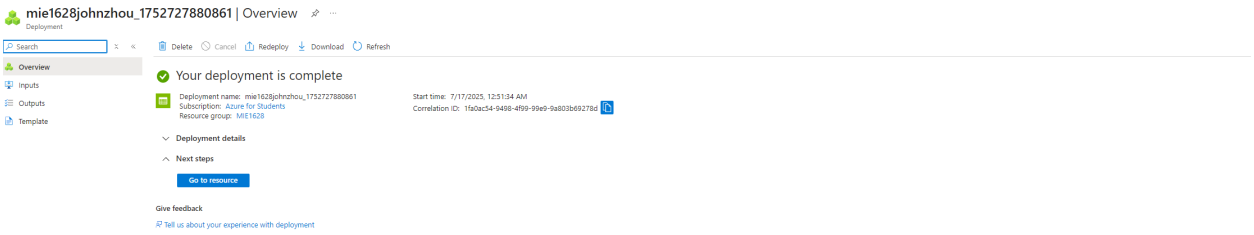
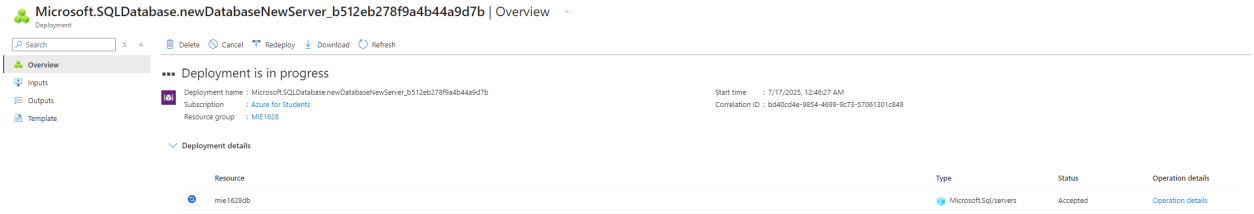
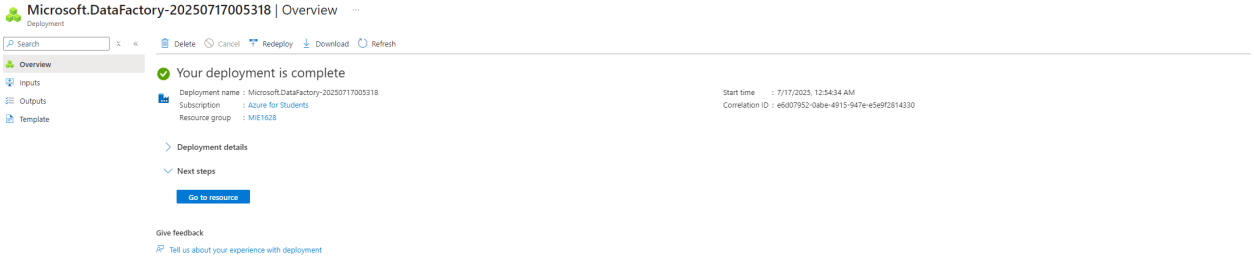
Part A: [Marks 60]

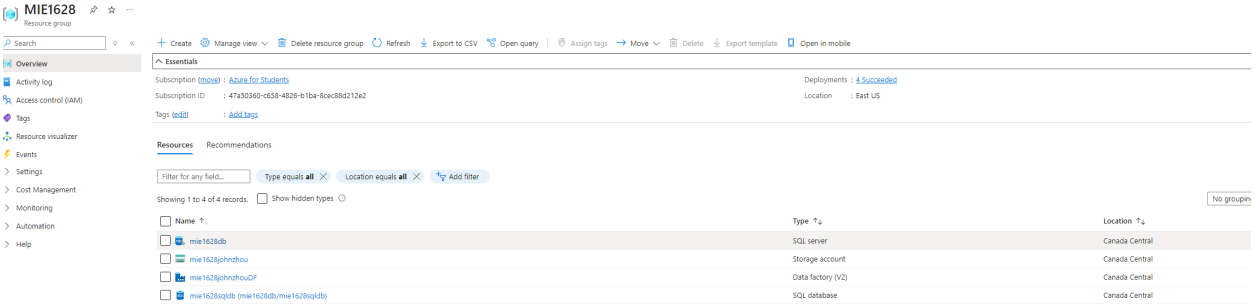
1. [Marks: 5] Create a resource group in your Azure portal and deploy three resources. Azure Data Factory, Azure SQL DB and Blob storage account.





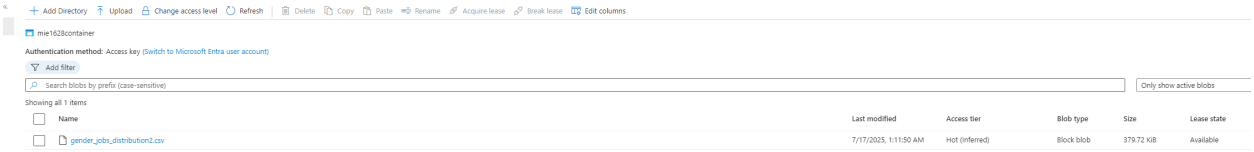




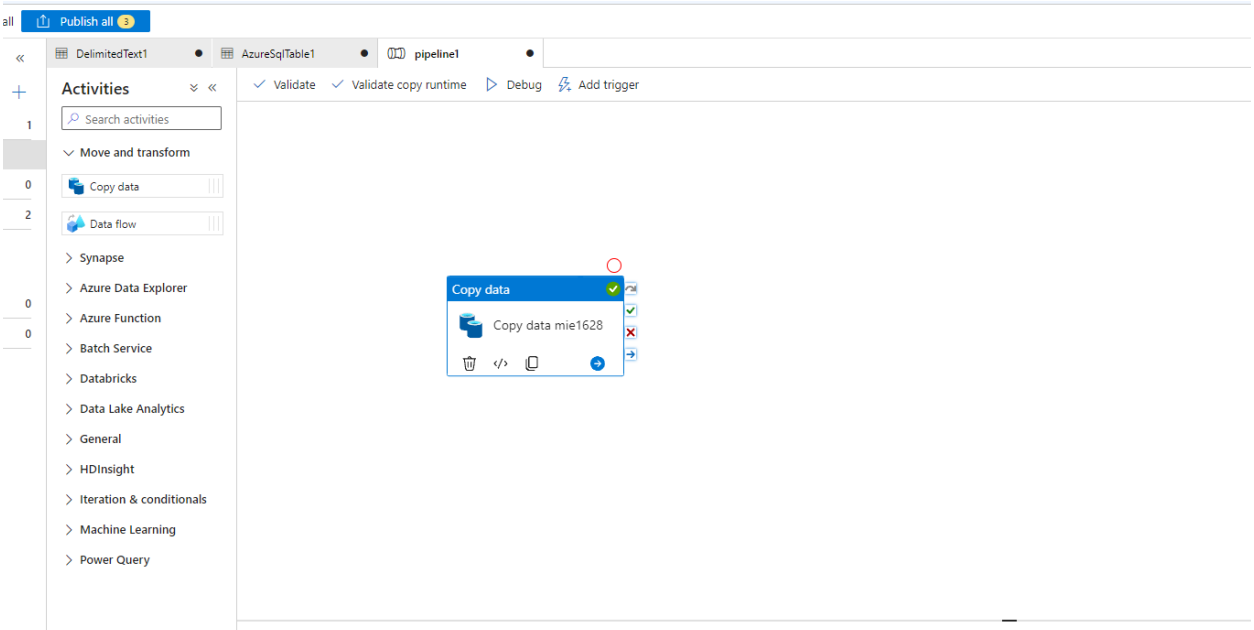


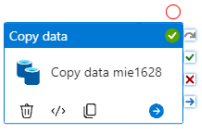
2. [Marks: 15] Now create a pipeline in Azure Data Factory and copy gender_jobs_data.csv file from the Blob storage account to Azure SQL DB. (First copy this file from your local machine to Blob Storage). See this https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copydata-portal for reference.

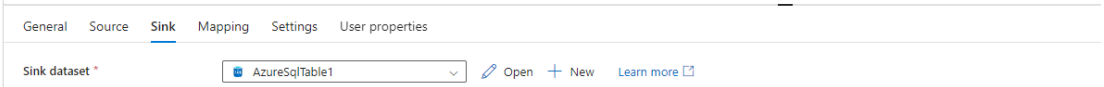
Copy from local machine to Blob storage:

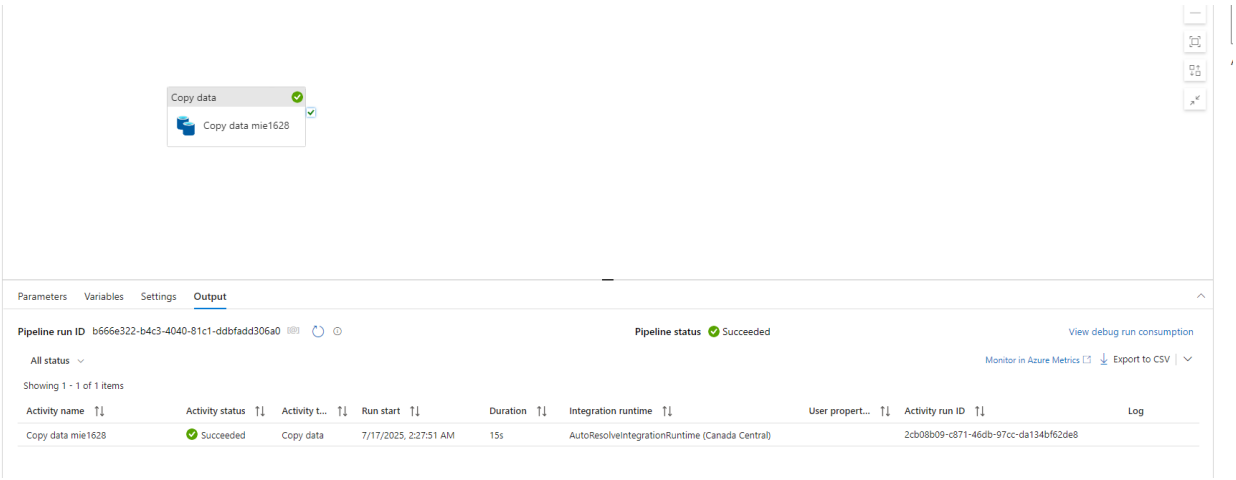


setup sink and source

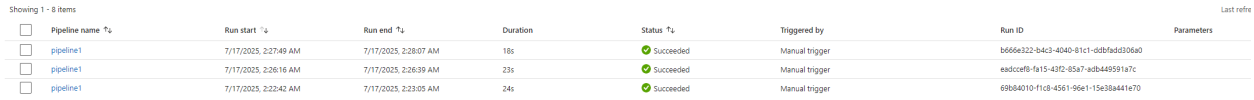


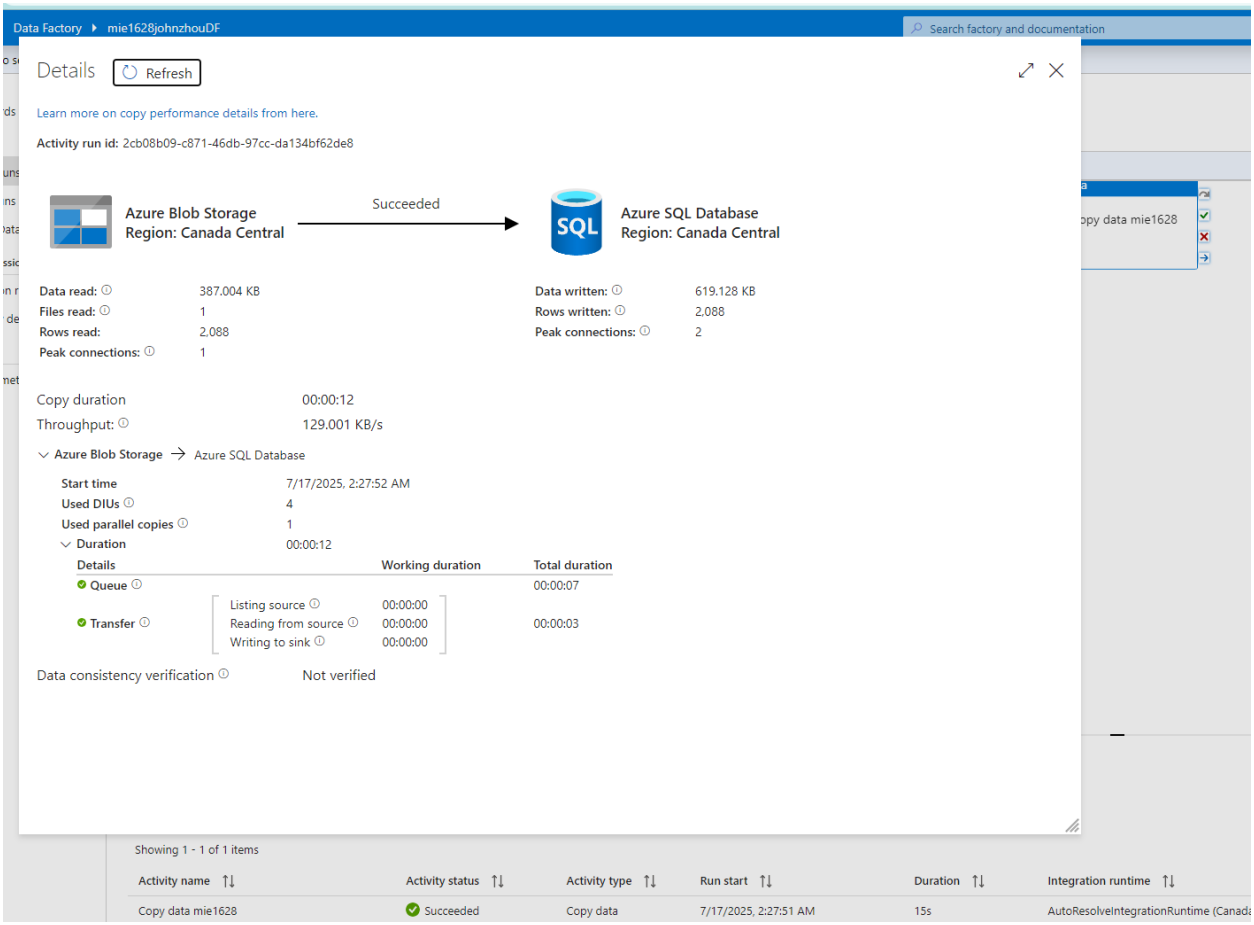






Detail of debug running:

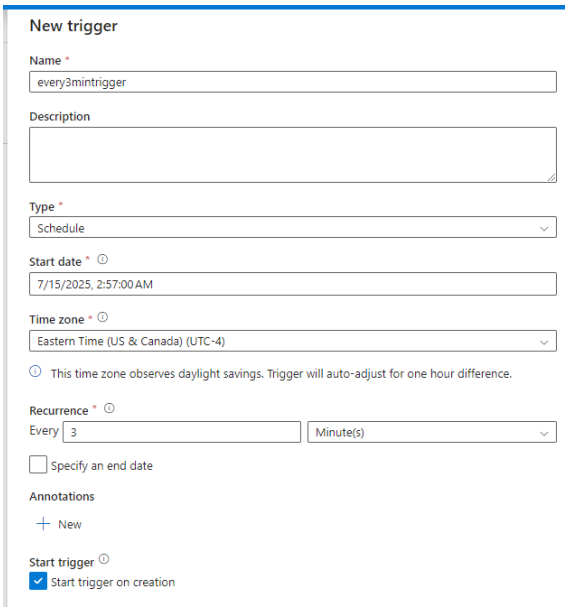


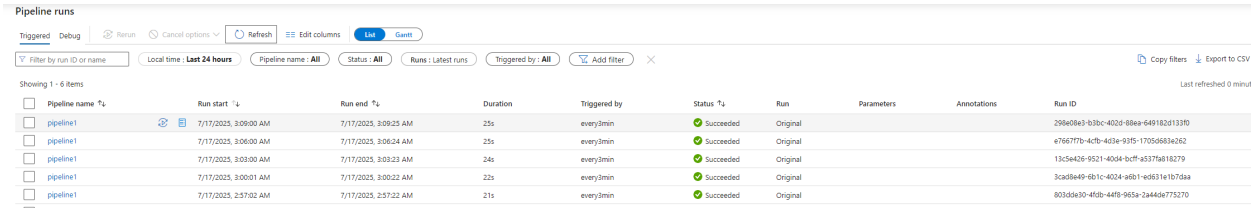


3. [Marks: 10] Explain the different types of triggers available in ADF. Now create a schedule trigger and run your pipeline every 3 minutes. Show 5 successful runs.

- Tumbling windows trigger: trigger the pipeline running in fixed-size, non-overlapping time windows. It is used for time-based incremental data loads, and it can ensure that processing only happen once.
- Schedule trigger: trigger the pipeline running based on a pre-specified schedule and configure different specs like start-end time, frequency etc. It is used for running task in a regular basis.
- Event-based trigger: trigger the pipeline in response to certain event happened within Azure resources
- Manual via UI trigger: manually trigger the pipeline from ADF user interface (UI) for on-demand basis.

Scheduled 5 successful run every 3 min:



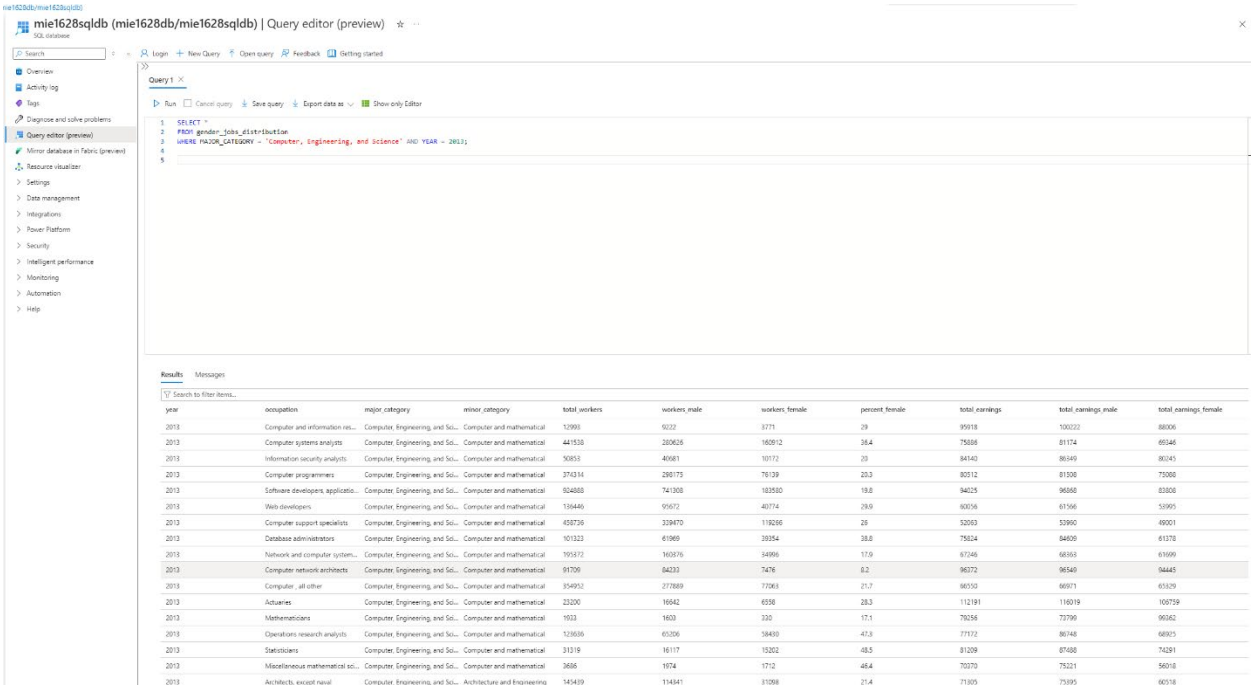


4. [Marks: 20] A client needs to replicate objects from ADLS Gen 2 in Canada Central to ADLS Gen2 in West Europe. Let's say they want to do this in a bi-directional way. How can you set this up? Explain in words.

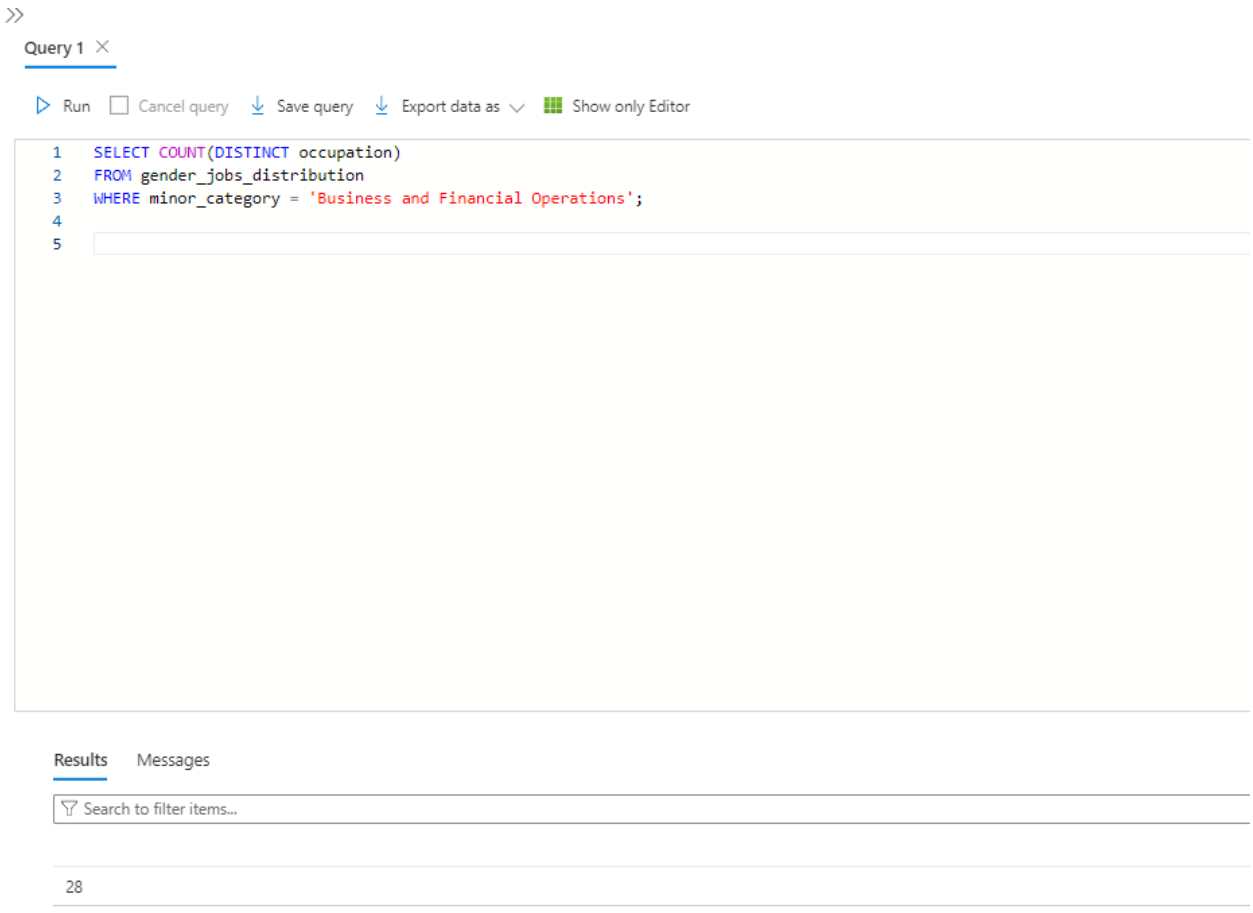
- To replicate the objected in from ADLS Gen 2 in Canada Central to ADLS Gen2 in West Europe, we could use ADF to setup two pipelines: one transfer and copy from Canada to Europe and the other in reverse as a bi-directional pipeline.
- For one pipeline, we could set up ADLS Gen 2 in Canada Central and ADLS Gen2 in West Europe as source and sink and the other pipeline set up in the reversed way.
- We could set up an event-based trigger to respond to any storage event, for example, the modification happened in one of the storage, and trigger the pipeline to synchronize and replicate the modification to the other instance.

Part B: [Marks: 20]

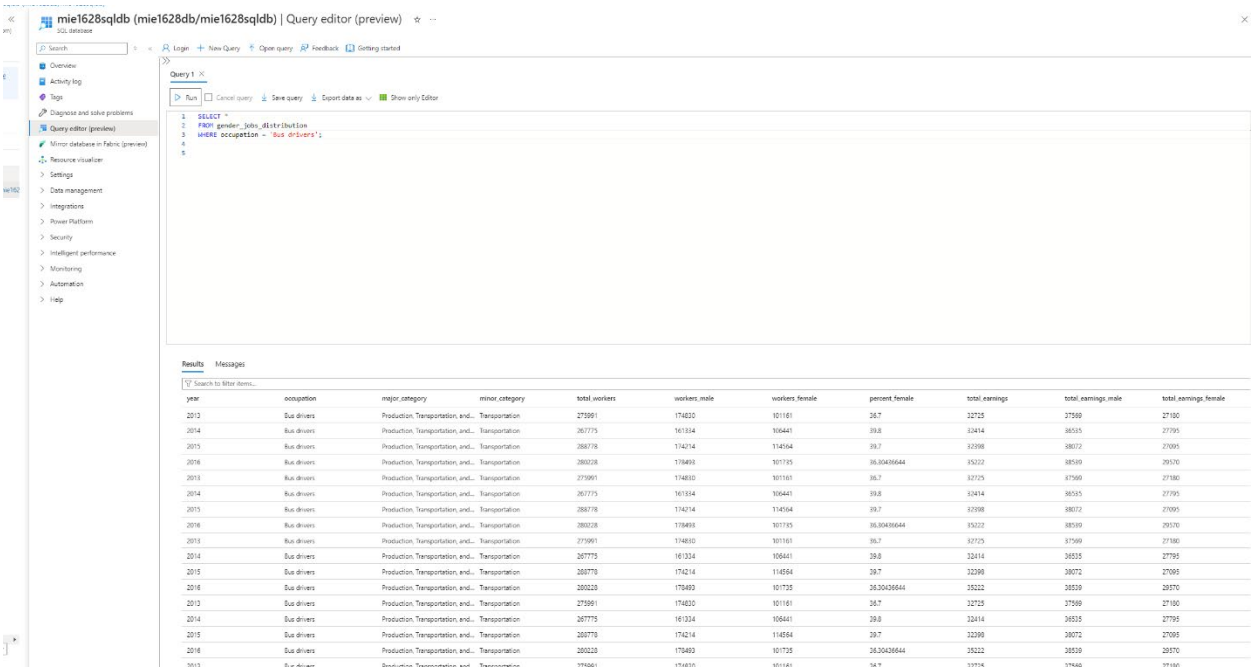
1. [Marks:5] In the gender_jobs_data table - Filter all the OCCUPATIONS in MAJOR_CATEGORY of Computer, Engineering, and Science for the YEAR 2013



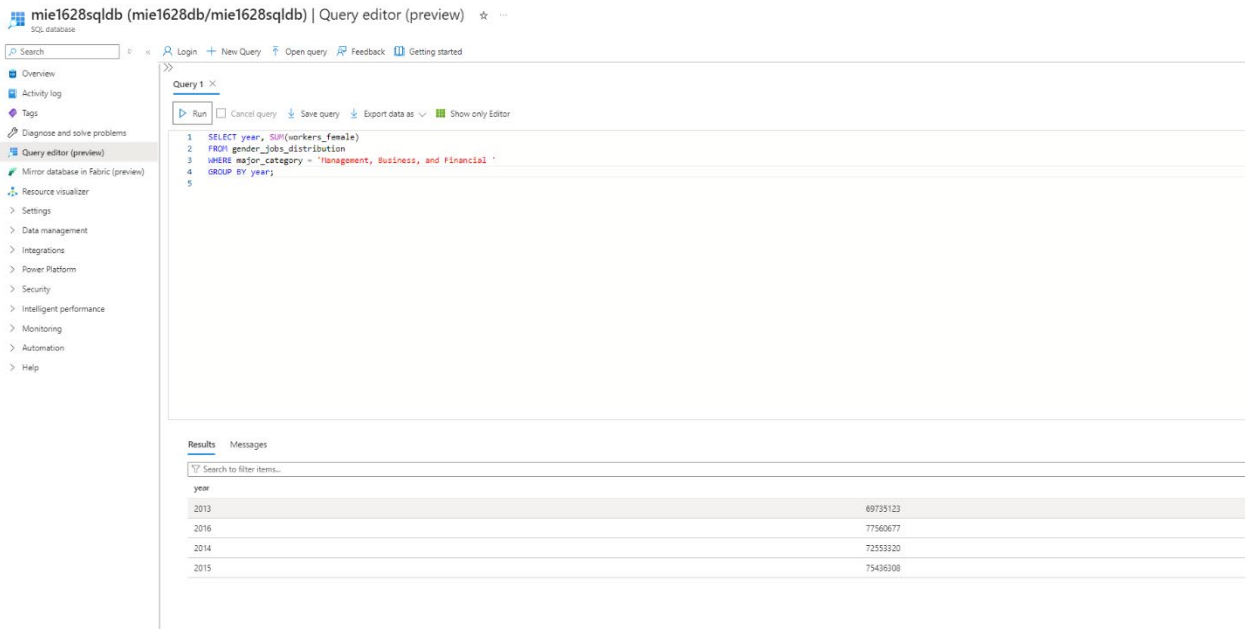
2. [Marks:5] In the gender_jobs_data table - How many OCCUPATIONS exist in the MINOR_CATEGORY of Business and Financial Operations overall?



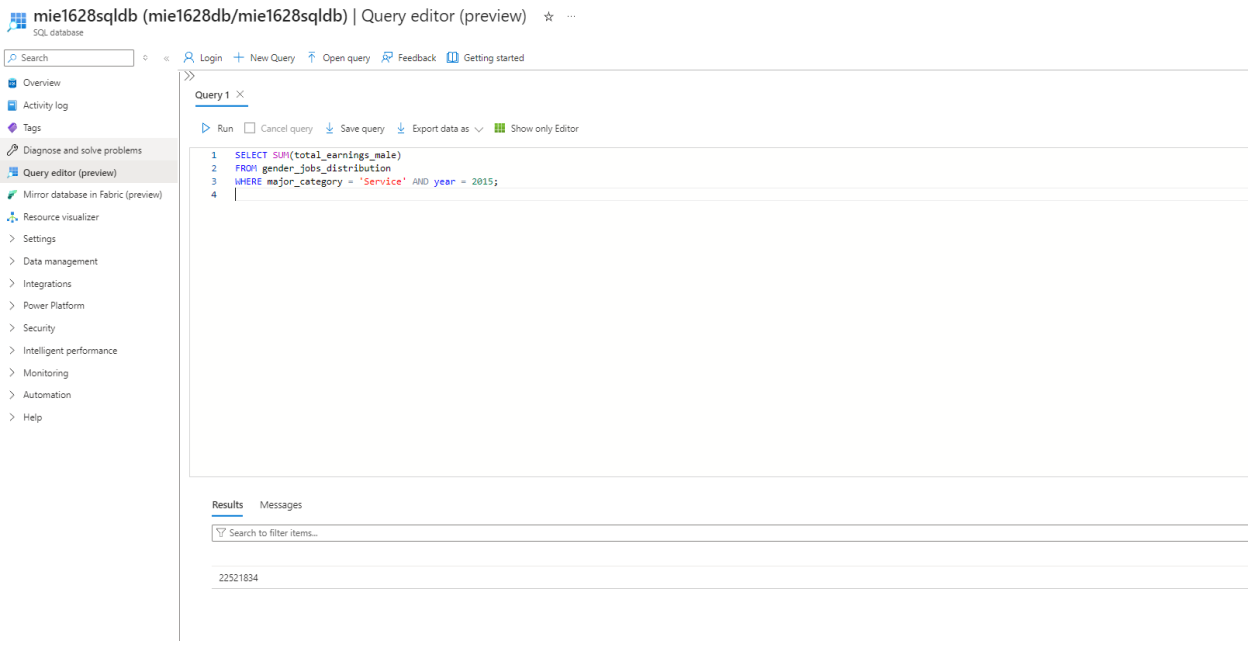
3. [Marks:5] In the gender_jobs_data table - Get all relevant information for bus drivers across all years



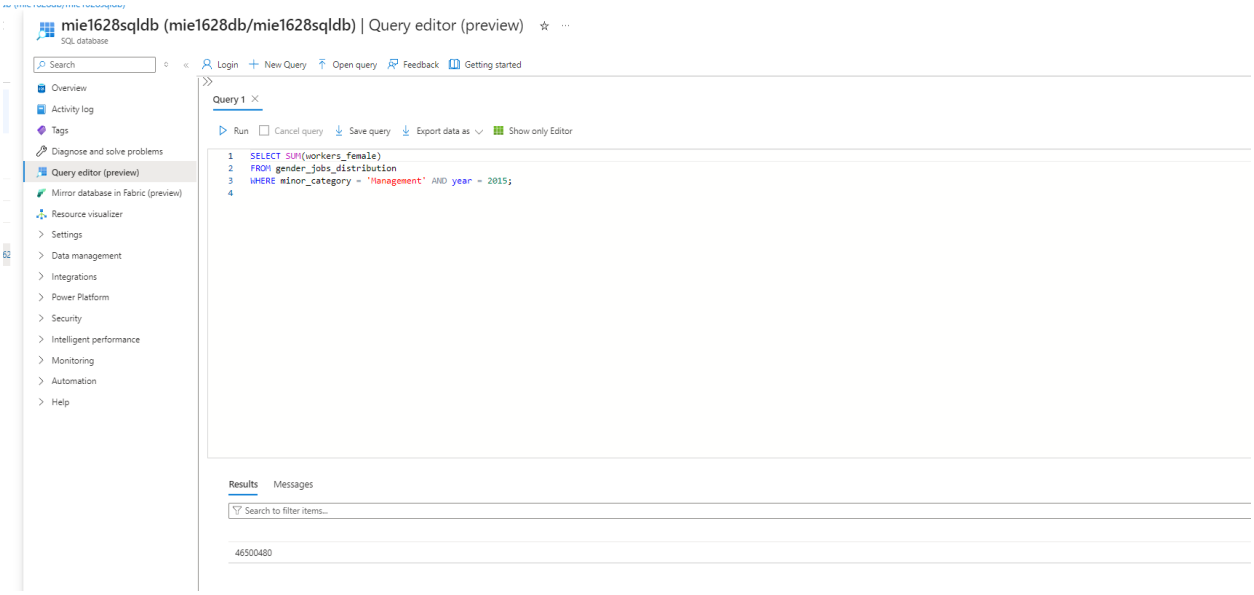
4. [Marks:5] In the gender_jobs_data table - Summarize the total number of WORKERS_FEMALE in the MAJOR_CATEGORY of Management, Business, and Financial by each year.



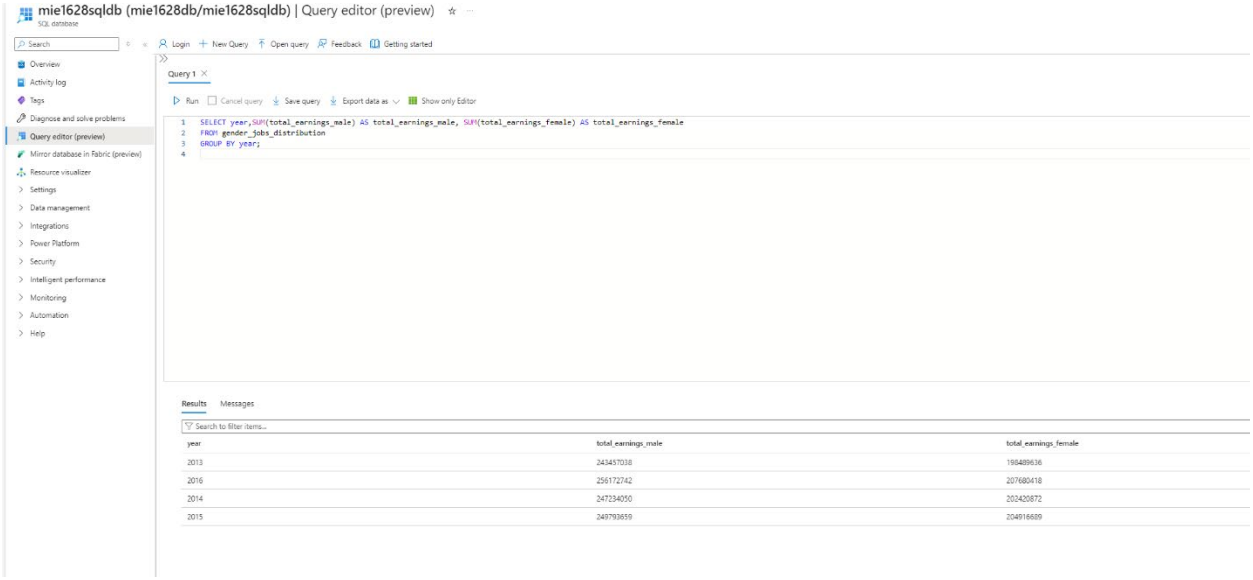
5. [Marks:5] In the gender_jobs_data table - What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?



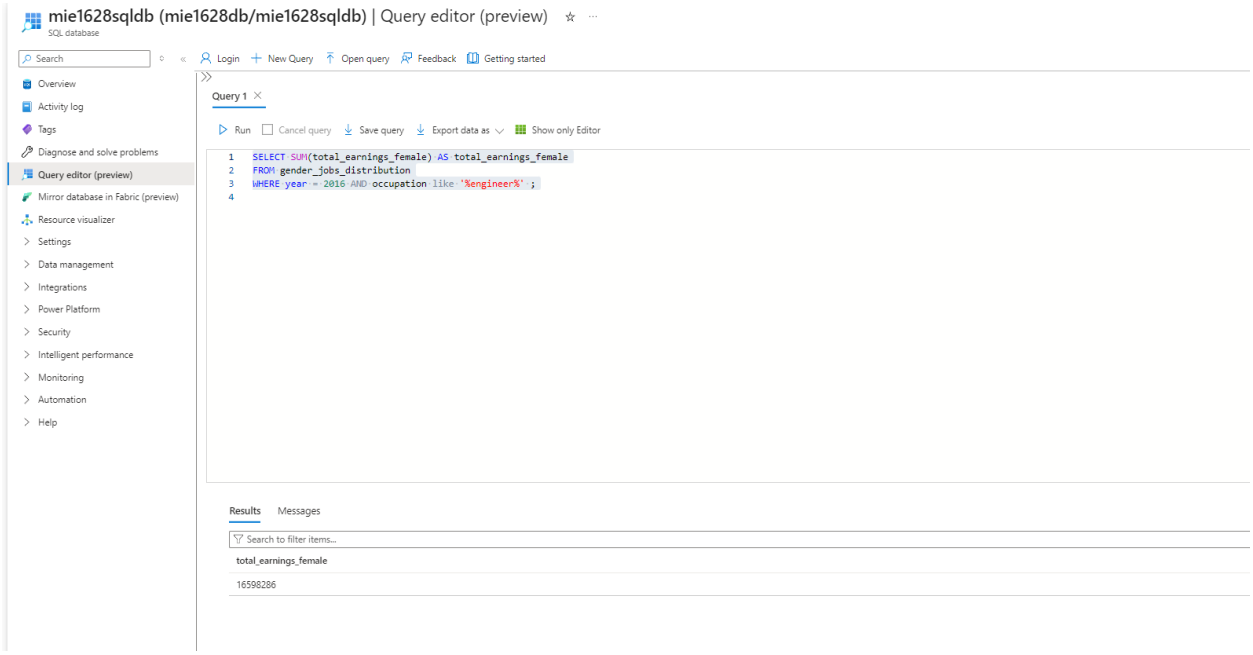
6. [Marks:5] In the gender_jobs_data table - How many female workers were in management roles in the year 2015?



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?