

# Assignment 4

## Assignment on Azure Cloud Platform

### Part A:

- [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:

The screenshot shows the Microsoft Azure Resource groups page. At the top, there is a search bar and a Copilot button. The main area displays a single resource group named "MIE1628". The details for this group are as follows:

Name	Subscription	Location
MIE1628	Azure for Students	East US

At the bottom of the page, there are navigation links for "Page 1 of 1" and a "Give feedback" button.

Azure Data Factory:

The screenshot shows the Microsoft Azure Microsoft.DataFactory-20240720125152 | Overview page. The deployment status is shown as "Your deployment is complete". Deployment details include:

Deployment name	Start time
Microsoft.DataFactory-20240720125152	7/20/2024, 12:53:57 PM

Other sections visible on the page include "Inputs", "Outputs", "Template", "Deployment details", "Next steps", "Go to resource", "Give feedback", "Tell us about your experience with deployment", "Cost management", "Microsoft Defender for Cloud", "Free Microsoft tutorials", "Work with an expert", and "Find an Azure expert".

## Azure SQL DB:

**Deployment details**

Resource	Type	Status	Operation details
db1628/db116	Microsoft.Sql/servers/databases	Created	Operation details
db1628/Default	Microsoft.Sql/servers/connections	OK	Operation details
db1628	Microsoft.Sql/servers	OK	Operation details
db1628	Microsoft.Sql/servers	Created	Operation details

## Blob storage account:

**Deployment details**

Resource	Type	Status	Operation details
sc116_1721498806038	Microsoft.Storage/storageAccounts	Created	Operation details

## Overall:

**Resources**

Name	Type	Location
db116 (db1628/db116)	SQL database	East US 2
db1628	SQL server	East US 2
db116	Data factory (V2)	East US
uc116	Storage account	East US

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).

- Upload the CSV File to Blob Storage:

The screenshot shows the Microsoft Azure Blob Storage interface. At the top, there's a success message: "Successfully uploaded blob(s) Successfully uploaded 1 blob(s)." Below the header, there are tabs for Overview, Diagnose and solve problems, Access Control (IAM), and Settings. The Overview tab is selected. It displays a table with one row for the uploaded file:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
gender_jobs_distribution.csv	7/20/2024, 1:14:36 PM	Hot (Inferred)		Block blob	379.72 KiB	Available

- Linked Service to Blob Storage

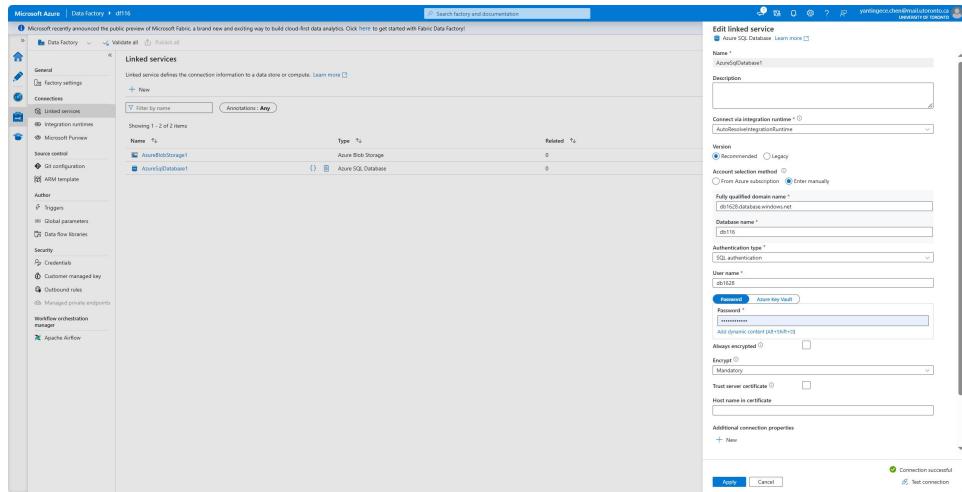
The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation menu includes Data Factory, General, Connectors, Linked services, and more. The main area shows a list of linked services:

Name	Type	Related
AzureBlobStorage1	Azure Blob Storage	0
AzureSqlDatabase1	Azure SQL Database	0

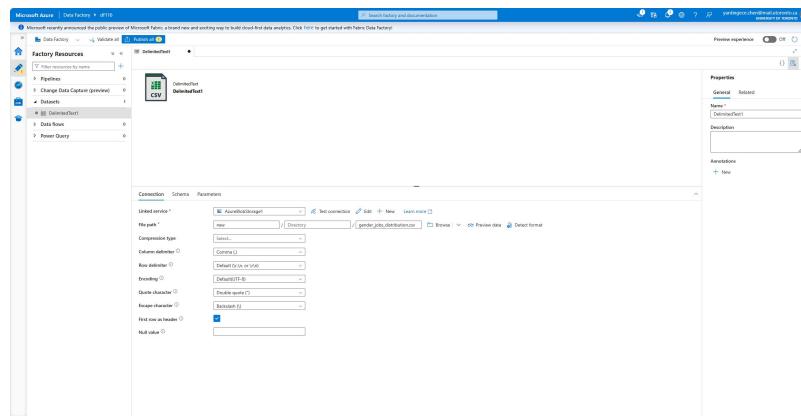
To the right, a detailed configuration pane is open for the "AzureBlobStorage1" linked service. It shows the following settings:

- Edit linked service**: Set to "Azure Blob Storage".
- Name**: "Azureblobstorage1".
- Description**: (empty).
- Connect via integration runtime**: "Autofailover/integrationRuntime".
- Authentication type**: "Account key".
- Connection using**: "Azure Key Vault".
- Connection selected method**: "From Azure subscription".
- Storage account name**: "actf16".
- Storage account key**: "Azure Key Vault".
- Storage account key**: (redacted).
- Partitioned DNS enabled**: (unchecked).
- Endpoint suffix**: "core.windows.net".
- Additional connection properties**:
  - + New
  - Test connection: "To linked service" (selected).
  - Annotations: "+ New", "> Parameters", "> Advanced".
- Connection successful**: "Connection successful".

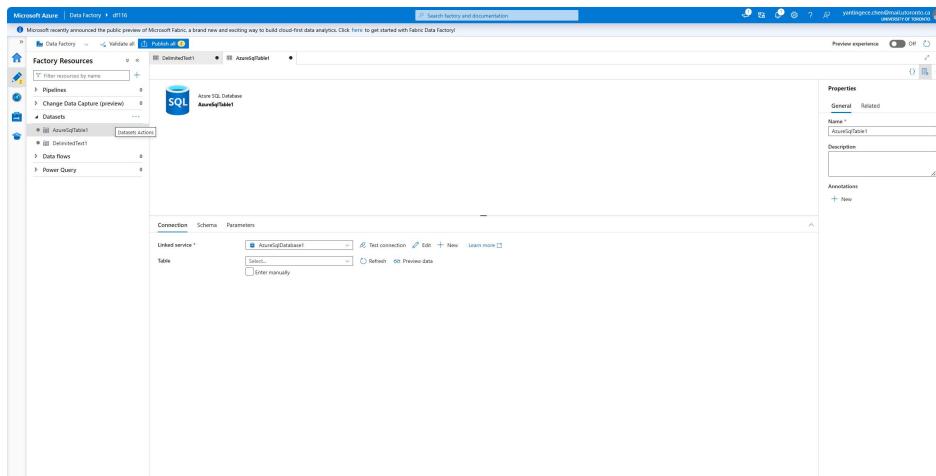
## ● Linked Service to Azure SQL Database



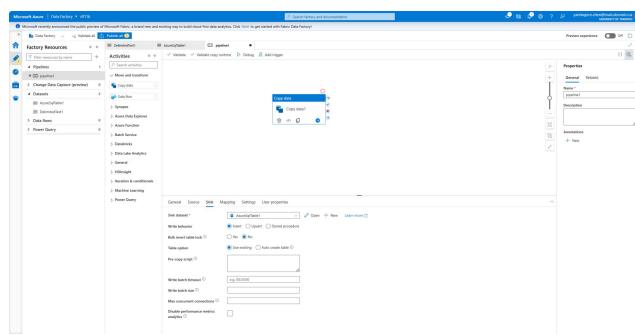
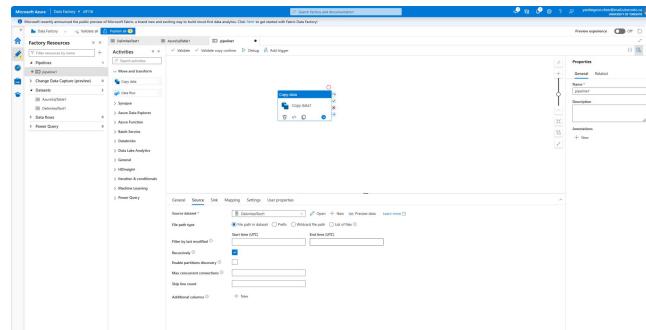
## ● Dataset for Blob Storage



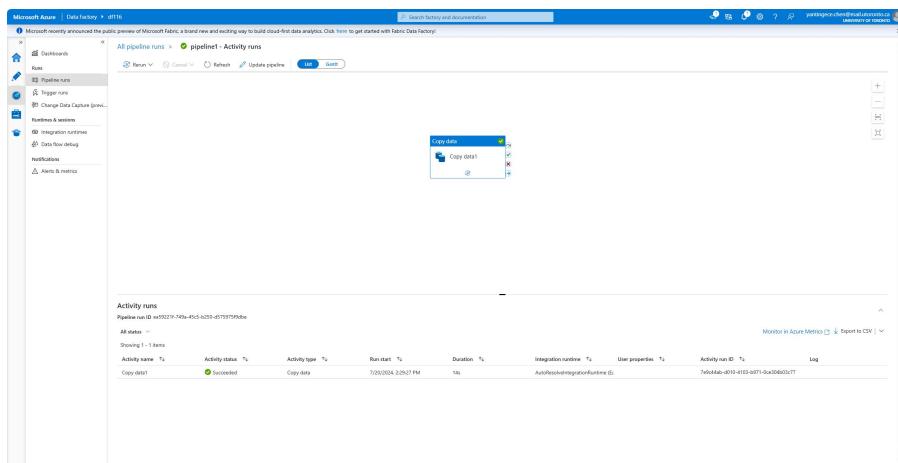
## ● Dataset for Azure SQL Database



## ● Create a Pipeline



## ● Publish and Trigger the Pipeline



The screenshot shows the Azure Data Factory interface. On the left, the navigation pane includes 'Dashboards', 'Runs', 'Pipeline runs' (selected), 'Trigger runs', 'Change Data Capture (preview)', 'Runtimes & sessions', 'Integration runtimes', 'Data flow debug', and 'Notifications'. The main area displays 'All pipeline runs > pipeline1 - Activity runs'. A specific activity run is selected, showing a copy operation from 'Azure Blob Storage' to 'Azure SQL Database'. The status is 'Succeeded'. Key metrics shown include: Data read: 388.834 KB, Files read: 1, Rows read: 2,098, Peak connections: 2; Data written: 687.964 KB, Rows written: 2,098, Peak connections: 2. The copy duration was 00:00:11 with a throughput of 129.611 KB/s. The activity run ID is 7e9c4ab-d010-4103-b971-0e30ab03c77.

The screenshot shows the Azure Data Factory interface with the pipeline 'pipeline1' selected. In the center, the 'Preview data' section is open for the dataset 'AzureSqlTable1'. It shows a preview of the data from the 'AzureSqlDatabase' linked service. The table structure includes columns: year, occupation, major\_category, minor\_category, total\_workers, workers\_male, workers\_female, percent\_female, and total. The data preview shows 6 rows of data from 2013, categorized by occupation and industry.

	year	occupation	major_category	minor_category	total_workers	workers_male	workers_female	percent_female	total
1	2013	Chef executives	Management, Business, and Financial	Management	1004259	782400	241859	23.6	12025
2	2013	General and operations managers	Management, Business, and Financial	Management	977284	681627	295657	30.3	73557
3	2013	Legislators	Management, Business, and Financial	Management	14015	9375	6480	43.5	67155
4	2013	Advertising and promotions managers	Management, Business, and Financial	Management	43015	17775	25240	58.7	61371
5	2013	Marketing and sales managers	Management, Business, and Financial	Management	754514	440078	314436	41.7	78455
6	2013	Public relations and fundraising managers	Management, Business, and Financial	Management	44190	16141	28057	63.5	76114

3. 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.**

Trigger name	Trigger type	Trigger time	Status	Pipelines	Run	Message	Properties	Run ID
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:25:59 PM	Succeeded	1	Original			065480097256069470654126801C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:26:00 PM	Succeeded	1	Original			0654800994094782303174305C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:26:00 PM	Succeeded	1	Original			0654800100853016249406654734C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:16:59 PM	Succeeded	1	Original			0654801002651751549892324124C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:15:59 PM	Succeeded	1	Original			0654801004517781797296050121C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:11:00 PM	Succeeded	1	Original			065480100624974833402834C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:07:59 PM	Succeeded	1	Original			0654801008535961384990790C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:05:00 PM	Succeeded	1	Original			0654801009400079062417198C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 3:02:00 PM	Succeeded	1	Original			065480101165400079010179578C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 2:59:00 PM	Succeeded	1	Original			0654801013446725147367008003C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 2:56:00 PM	Succeeded	1	Original			06548010134472619495251297C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 2:53:00 PM	Succeeded	1	Original			06548010170520217000011355C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 2:50:00 PM	Succeeded	1	Original			06548010188529514744984325C
Every1MinutesTrigger	Schedule trigger	7/20/2024, 2:47:00 PM	Succeeded	1	Original			06548010206470253938620802C

4. 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

```

SELECT *
FROM [dbo].[table]
WHERE MAJOR_CATEGORY = 'Computer, Engineering, and Science'
AND YEAR = 2013;

```

The screenshot shows the Microsoft Azure portal interface with the query editor open. The query editor window has a title bar "db116 (db1628/db116) | Query editor (preview)" and a toolbar with "Run", "Cancel query", "Save query", "Export data as", "Show only Editor", and "Launch inline clipboard". Below the toolbar is a "Query 1" section containing the T-SQL code provided above. To the left of the editor is a sidebar titled "Tables" which lists various columns of the "table" (likely gender\_jobs\_data). To the right of the editor is a "Results" pane showing the query results as a table:

year	occupation	major_category	minor_category	total_workers
2013	Computer and information re...	Computer, Engineering, and Sc...	Computer and mathematical	12993
2013	Computer systems analysts,...	Computer, Engineering, and Sc...	Computer and mathematical	44538
2013	Information security analysts...	Computer, Engineering, and Sc...	Computer and mathematical	5083
2013	Computer programmers...	Computer, Engineering, and Sc...	Computer and mathematical	37434
2013	Software developers, applicati...	Computer, Engineering, and Sc...	Computer and mathematical	92488
2013	Web developers	Computer, Engineering, and Sc...	Computer and mathematical	136446

The status bar at the bottom of the editor says "Query succeeded | 0s".

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

```

SELECT COUNT(DISTINCT OCCUPATION) AS OccupationCount
FROM [dbo].[table]
WHERE MINOR_CATEGORY = 'Business and Financial Operations';

```

The screenshot shows the Microsoft Azure portal interface with the query editor open. The query editor window has a title bar "db116 (db1628/db116) | Query editor (preview)" and a toolbar with "Run", "Cancel query", "Save query", "Export data as", "Show only Editor", and "Launch inline clipboard". Below the toolbar is a "Query 1" section containing the T-SQL code provided above. To the left of the editor is a sidebar titled "Tables" which lists various columns of the "table" (likely gender\_jobs\_data). To the right of the editor is a "Results" pane showing the query results as a table:

OccupationCount
28

The status bar at the bottom of the editor says "Query succeeded | 0s".

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

```

SELECT *
FROM [dbo].[table]
WHERE OCCUPATION = 'bus drivers';
    
```

The screenshot shows the Microsoft Azure portal interface with the query editor open. The query editor title bar says "db116 (db1628/db116) | Query editor (preview)". The left sidebar shows the database structure with tables like "year", "occupation", and "bus drivers". The main area shows the query "SELECT \* FROM [dbo].[table] WHERE OCCUPATION = 'bus drivers';" and its results. The results table has columns: year, occupation, major\_category, minor\_category, and total\_workers. The data shows bus drivers from 2013 to 2014 across different categories.

year	occupation	major_category	minor_category	total_workers
2013	Bus drivers	Production, Transportation, and...	Transportation	273991
2014	Bus drivers	Production, Transportation, and...	Transportation	267775
2015	Bus drivers	Production, Transportation, and...	Transportation	288778
2016	Bus drivers	Production, Transportation, and...	Transportation	280228
2013	Bus drivers	Production, Transportation, and...	Transportation	273991
2014	Bus drivers	Production, Transportation, and...	Transportation	267775

4. 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.

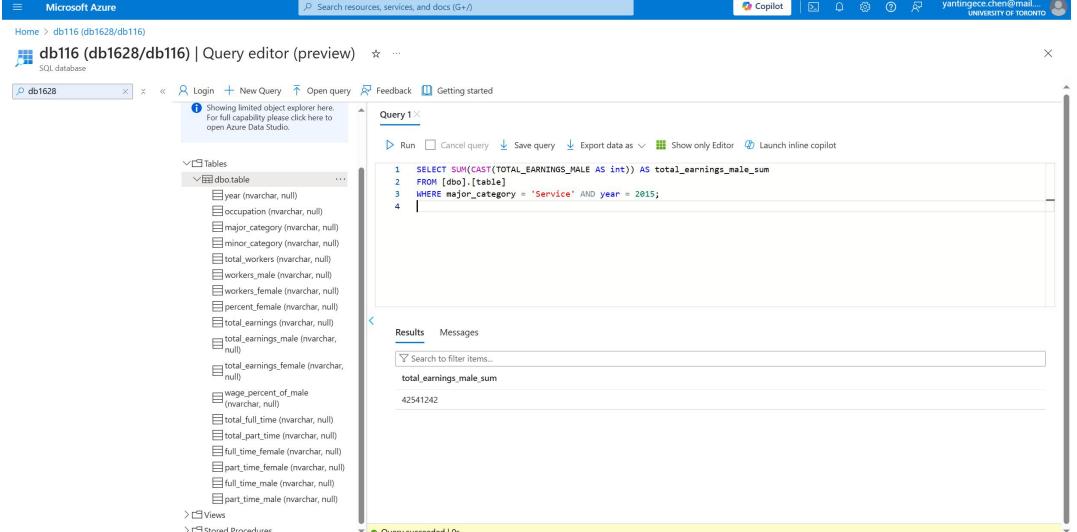
```

SELECT year, SUM(CAST(workers_female AS int)) as workers_female_sum
FROM [dbo].[table]
WHERE major_category = 'Management, Business, and Financial'
GROUP BY year;
    
```

The screenshot shows the Microsoft Azure portal interface with the query editor open. The query editor title bar says "db116 (db1628/db116) | Query editor (preview)". The left sidebar shows the database structure with tables like "year", "occupation", and "bus drivers". The main area shows the query "SELECT year, SUM(CAST(workers\_female AS int)) as workers\_female\_sum FROM [dbo].[table] WHERE major\_category = 'Management, Business, and Financial' GROUP BY year;" and its results. The results table has columns: year and workers\_female\_sum. The data shows the total number of female workers in management, business, and financial categories for the years 2013, 2014, 2015, and 2016.

year	workers_female_sum
2016	146503501
2014	137045160
2015	142490804
2013	131721899

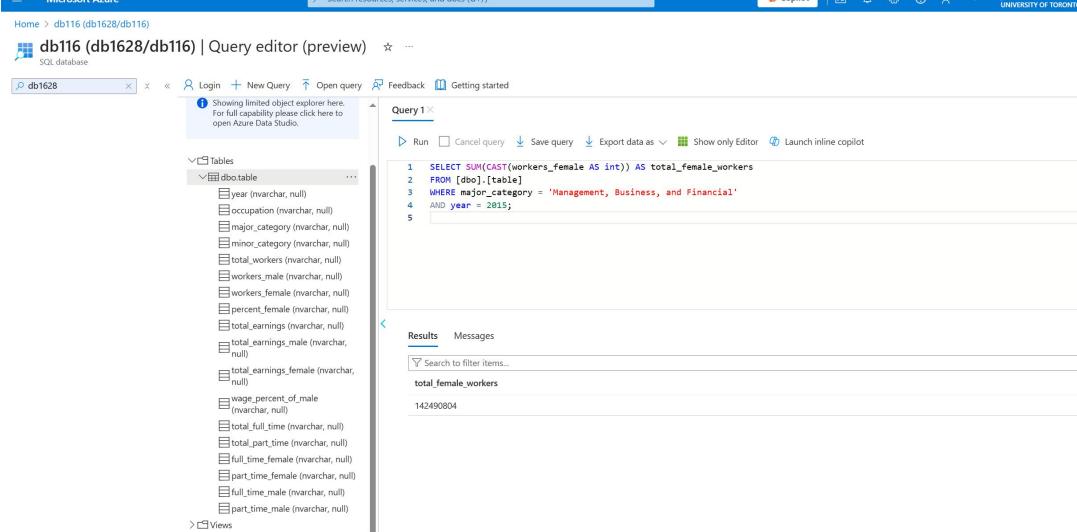
5. In the gender\_jobs\_data table - What were the total earnings of male (TOTAL\_EARNINGS\_MALE) employees in the Service MAJOR\_CATEGORY for the year 2015?



```
SELECT SUM(CAST(TOTAL_EARNINGS_MALE AS int)) AS total_earnings_male_sum
FROM [dbo].[table]
WHERE major_category = 'Service' AND year = 2015;
```

The screenshot shows the Microsoft Azure SQL Database Query Editor interface. The query editor window has a title bar "db116 (db1628/db116) | Query editor (preview)" and a status bar "yattinge.chen@mail.utoronto.ca". The main area contains a "Query 1" section with the provided SQL code. Below the code, the results pane shows the output: "total\_earnings\_male.sum" and the value "42541242". A message at the bottom says "Query succeeded | 0s".

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



```
SELECT SUM(CAST(workers_female AS int)) AS total_female_workers
FROM [dbo].[table]
WHERE major_category = 'Management, Business, and Financial'
AND year = 2015;
```

The screenshot shows the Microsoft Azure SQL Database Query Editor interface. The query editor window has a title bar "db116 (db1628/db116) | Query editor (preview)" and a status bar "yattinge.chen@mail.utoronto.ca". The main area contains a "Query 1" section with the provided SQL code. Below the code, the results pane shows the output: "total\_female\_workers" and the value "142490804". A message at the bottom says "Query succeeded | 0s".

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

The screenshot shows the Microsoft Azure portal with the query editor open for the db116 database. The query is:

```
1 SELECT
2     year,
3     SUM(TRY_CAST(TOTAL_EARNINGS_MALE AS int)) AS total_earnings_male_sum,
4     SUM(TRY_CAST(TOTAL_EARNINGS_FEMALE AS int)) AS total_earnings_female_sum
5 FROM [dbo].[table]
6 GROUP BY year
7 ORDER BY year;
```

The results table shows the total earnings for males and females from 2013 to 2016.

year	total_earnings_male_sum	total_earnings_female_sum
2013	459863294	374924868
2014	466997650	382350536
2015	471832467	387064857
2016	483881846	392285234

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

The screenshot shows the Microsoft Azure portal with the query editor open for the db116 database. The query is:

```
1 SELECT SUM(TRY_CAST(TOTAL_EARNINGS_FEMALE AS int)) AS total_earnings_female_sum
2 FROM [dbo].[table]
3 WHERE occupation LIKE '%Engineer%' AND year = 2016;
```

The results table shows the total earnings for female engineers in 2016.

total_earnings_female_sum
31352318

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

The screenshot shows the Microsoft Azure SQL Database Query editor interface. The query window displays a T-SQL script to calculate the total number of full-time and part-time female and male workers by year. The results tab shows the output for the years 2013 through 2016.

```
1 SELECT
2     year,
3     SUM(CAST(full_time_female AS float) + CAST(part_time_female AS float)) AS total_females_workers,
4     SUM(CAST(full_time_male AS float) + CAST(part_time_male AS float)) AS total_males_workers
5
6     FROM [dbo].[table]
7
8 WHERE TRY_CAST(full_time_female AS float) IS NOT NULL
9 AND TRY_CAST(part_time_female AS float) IS NOT NULL
10 AND TRY_CAST(full_time_male AS float) IS NOT NULL
11 AND TRY_CAST(part_time_male AS float) IS NOT NULL
12
13 GROUP BY year
14
15 ORDER BY year;
```

year	total_female_workers	total_male_workers
2013	887400	887400
2014	887400	887400
2015	887400	887400
2016	887400	887400