

Part A:

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

Resource Groups:

[Home](#) >

Resource groups

University of Toronto (utoronto.onmicrosoft.com)

[+](#) Create [⚙️](#) Manage view [↺](#) Refresh [↓](#) Export to CSV [🔗](#) Open query [🏷️](#) Assign tags

Filter for any field... Subscription equals all Location equals all [+🔗 Add filter](#)

[🛡️ 0](#) Unsecure resources [🔗 0](#) Recommendations [No grouping](#) [⌵](#)

[☰](#) List view [⌵](#)

<input type="checkbox"/> Name ↑↓	Subscription ↑↓	Location ↑↓	
<input type="checkbox"/> 🔗 Steven	Azure subscription 1	Canada East	⋮

Inside Resource Group:

[Home](#) > [Resource groups](#) >

Steven

Resource group

[«](#) [+](#) Create [⚙️](#) Manage view [🗑️](#) Delete resource group [↺](#) Refresh [↓](#) Export to CSV [⋮](#)

[🔗 Overview](#) [📅 Activity log](#) [👤 Access control \(IAM\)](#) [🏷️ Tags](#) [📊 Resource visualizer](#) [⚡ Events](#)

Settings

- [📁 Deployments](#)
- [🛡️ Security](#)
- [📄 Policies](#)
- [📑 Properties](#)
- [🔒 Locks](#)

Cost Management

- [💰 Cost analysis](#)
- [📢 Cost alerts \(preview\)](#)
- [💵 Budgets](#)

Essentials

[Subscription \(move\)](#) [Deployments](#)
[Azure subscription 1](#) [4 Succeeded](#)

Subscription ID: 7aaa6527-b424-41f1-8dbc-a12b2e314f00 Location: Canada East

Tags [\(edit\)](#)
[Click here to add tags](#)

Resources

[Recommendations](#)

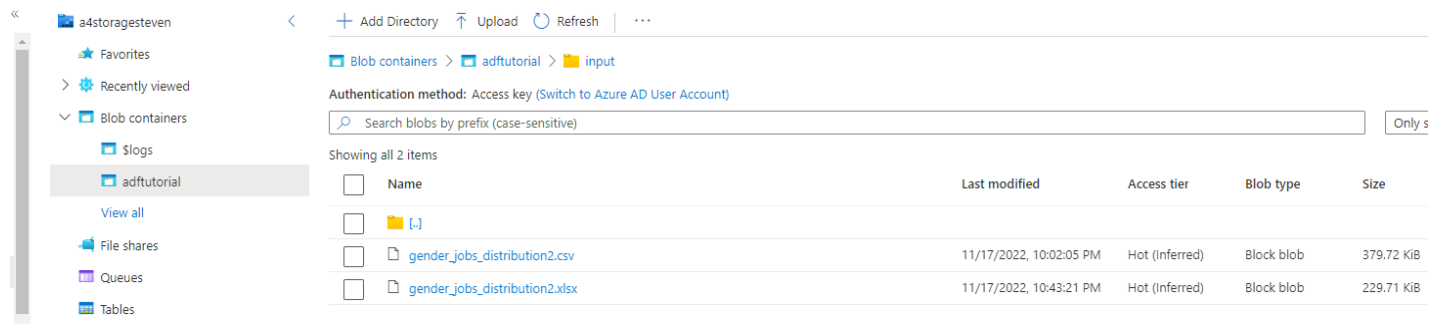
Filter for any field... Type equals all [⌵](#) [+🔗 Add filter](#) [📄 More \(1\)](#)

Showing 1 to 4 of 4 records. ☐ Show hidden types ⓘ [No grouping](#) [⌵](#)

[☰](#) List view [⌵](#)

<input type="checkbox"/> Name ↑↓	Type ↑↓	Location ↑↓	
<input type="checkbox"/> 🔗 a4sqlldb (a4sqlserver/a4sqlldb)	SQL database	Canada Central	⋮
<input type="checkbox"/> 🔗 a4sqlserver	SQL server	Canada Central	⋮
<input type="checkbox"/> 🔗 a4steven	Data factory (V2)	Canada East	⋮
<input type="checkbox"/> 🔗 a4storagesteven	Storage account	Canada Central	⋮

2. Upload gender_jobs_distribution2.csv to blob storage, in the storage browser. It seems csv file is encrypted so another xlsx is created and uploaded to the adftutorial folder.

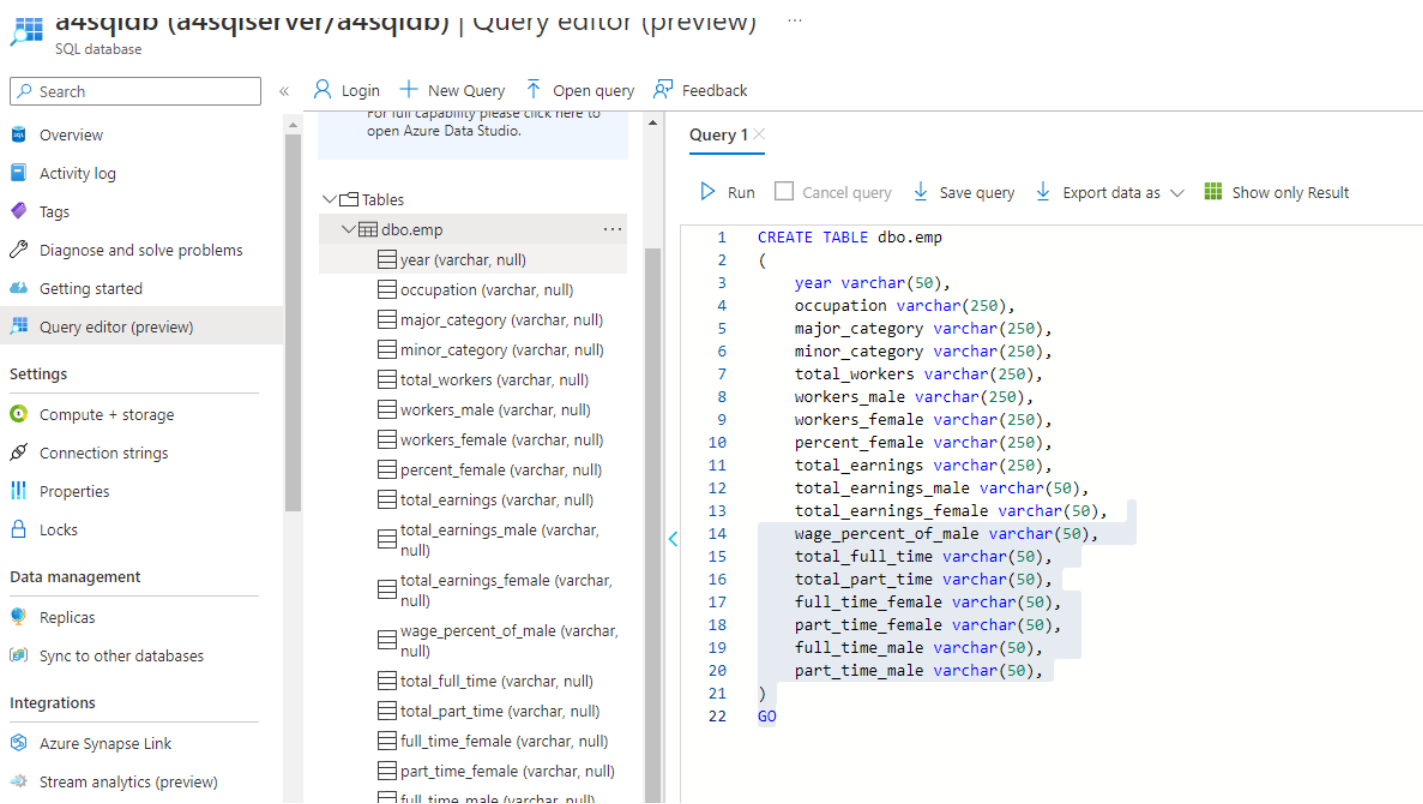


Authentication method: Access key (Switch to Azure AD User Account)

Search blobs by prefix (case-sensitive)

Showing all 2 items

	Name	Last modified	Access tier	Blob type	Size
<input type="checkbox"/>	[-]				
<input type="checkbox"/>	gender_jobs_distribution2.csv	11/17/2022, 10:02:05 PM	Hot (Inferred)	Block blob	379.72 KiB
<input type="checkbox"/>	gender_jobs_distribution2.xlsx	11/17/2022, 10:43:21 PM	Hot (Inferred)	Block blob	229.71 KiB



SQL database

Search

Login + New Query Open query Feedback

For full capability please click here to open Azure Data Studio.

Tables

dbo.emp

- year (varchar, null)
- occupation (varchar, null)
- major_category (varchar, null)
- minor_category (varchar, null)
- total_workers (varchar, null)
- workers_male (varchar, null)
- workers_female (varchar, null)
- percent_female (varchar, null)
- total_earnings (varchar, null)
- total_earnings_male (varchar, null)
- total_earnings_female (varchar, null)
- wage_percent_of_male (varchar, null)
- total_full_time (varchar, null)
- total_part_time (varchar, null)
- full_time_female (varchar, null)
- part_time_female (varchar, null)
- full_time_male (varchar, null)

Query 1 X

Run Cancel query Save query Export data as Show only Result

```
1 CREATE TABLE dbo.emp
2 (
3     year varchar(50),
4     occupation varchar(250),
5     major_category varchar(250),
6     minor_category varchar(250),
7     total_workers varchar(250),
8     workers_male varchar(250),
9     workers_female varchar(250),
10    percent_female varchar(250),
11    total_earnings varchar(250),
12    total_earnings_male varchar(50),
13    total_earnings_female varchar(50),
14    wage_percent_of_male varchar(50),
15    total_full_time varchar(50),
16    total_part_time varchar(50),
17    full_time_female varchar(50),
18    part_time_female varchar(50),
19    full_time_male varchar(50),
20    part_time_male varchar(50),
21 )
22 GO
```



Excel
SourceBlobData

Preview data

Linked service: AzureStorageLinkedService

Object: gender_jobs_distribution2.xlsx

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

Connection Schema Parameters

Linked service *

File path *

Compression type

Worksheet mode

Sheet name * ⓘ













Range ⓘ

Null value

First row as header

Notifications

Dismiss all

-  **Publishing completed** 
Successfully published
10 hours ago
-  **Publishing error** 
Validation of model(s) failed. Fix it before publishing
10 hours ago
-  **Publishing completed** 
Successfully published
10 hours ago
-  **Publishing error** 
Validation of model(s) failed. Fix it before publishing
10 hours ago
-  **Successfully created** 
Successfully created AzureSqlDatabaseLinkedService (Linked service).
11 hours ago
-  **Successfully created** 
Successfully created AzureStorageLinkedService (Linked service).
11 hours ago

Notifications shows successful publications:

3. Explain the different types of triggers available in ADF.

3 types of triggers are Schedule Trigger, Tumbling Window Trigger, and Event-based Trigger. For Schedule Trigger, it offer more versatility by supporting various scheduling intervals, such as minutes, hours, days, weeks, or months. For Tumbling Window Trigger, it is a trigger that goes off periodically, beginning at a predetermined start time, and keeps state. For Event-based Trigger, it enables you to determine pipeline execution based on the dependence of file deposit or removal events.

Setup a scheduled trigger every 3 minutes.

Edit trigger

Name *

trigger1

Description

Type *

ScheduleTrigger

Start date * ⓘ

11/18/22 16:21:00

Time zone * ⓘ

Coordinated Universal Time (UTC)

Recurrence * ⓘ

Every

3

Minute(s)

☒ Specify an end date

End On * ⓘ

11/19/22 00:21:00

Annotations

+ New

Status ⓘ

☒ Started ☐ Stopped

Trigger runs

All Schedule Tumbling window Storage events Custom events Refresh Edit columns

Local time : Last 24 hours Trigger name : All Status : All Runs : Latest runs X

Showing 1 - 9 items


Trigger name ↑↓	Trigger type	Trigger time ↑↓	Status ↑↓	Pipelines	Run	Message	Propertie
trigger1	Schedule trigger	Nov 18, 2022, 11:48:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:45:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:41:59 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:39:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:36:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:33:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:29:59 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:27:00 am	✔ Succeeded	1	Original		⚙
trigger1	Schedule trigger	Nov 18, 2022, 11:24:01 am	✔ Succeeded	1	Original		⚙





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?

Firstly, create 2 pipelines for replicating data in the data factory. 1 pipeline in Canada Central, the other one in West Europe. Secondly, set up the configuration of the source for 2 pipelines, select Azure Data Lake Gen2 and completing the account information to setup the linked service, then provide the file path for the material to be transferred. Thirdly, setup the sink for 2 pipelines. Then debug and publish the pipelines. And lastly, set up the event-based Triggers for the 2 pipelines.

Part B

1. Filter all the OCCUPATIONS in MAJOR_CATEGORY of Computer, Engineering, and Science for the YEAR 2013. Results saved in query1result.csv.

Query 1 

 Run ☐ Cancel query  Save query  Export data as  Show only Editor

```
1 SELECT DISTINCT occupation From [dbo].[emp]
2 WHERE major_category = 'Computer, Engineering, and Science' AND year = '2013'
```

Results Messages

 Search to filter items...

occupation

Actuaries

Aerospace engineers

Agricultural and food science technicians

Agricultural and food scientists

Agricultural engineers

Architects, except naval

Astronomers and physicists

Atmospheric and space scientists

2. How many OCCUPATIONS exist in the MINOR_CATEGORY of Business and Financial Operations overall?

```
1 SELECT COUNT(DISTINCT occupation) From [dbo].[emp]
2 WHERE minor_category = 'Business and Financial Operations'
```



Results Messages

Search to filter items...

28

3. Get all relevant information for bus drivers across all years.

The output is also saved in query3result.csv.

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT * From [dbo].[emp]
2 WHERE occupation = 'bus drivers'
3 ORDER BY year
```






Results Messages

Search to filter items...

year	occupation	major_category	minor_category	total_workers	workers_male	workers_fem
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161
2013	Bus drivers	Production, Transportation, and...	Transportation	275991	174830	101161

4. Summarize the total number of WORKERS_FEMALE in the MAJOR_CATEGORY of Management, Business, and Financial by each year.

Run ☐ Cancel query  Save query  Export data as  Show only Editor

```
1 SELECT year, SUM(Try_cast(workers_female as float)) AS total_num FROM [dbo].[emp]
2 WHERE major_category = 'Management, Business, and Financial'
3 GROUP BY year
```

Results Messages

 Search to filter items...

year	total_num
2015	100581744
2014	96737760
2013	92980164
2016	103414236

5. What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?





```
1 SELECT SUM(Try_cast(total_earnings_male as float)) AS total_num FROM [dbo].[emp]
2 WHERE major_category = 'Service' AND year = '2015'
```

Results Messages

 Search to filter items...

total_num
30029112

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

 Run ☐ Cancel query  Save query  Export data as  Show only Editor

```
1 SELECT SUM(Try_cast(workers_female as float)) AS total_num FROM [dbo].[emp]
2 WHERE minor_category = 'Management' AND year = '2015'
```





Results Messages

 Search to filter items...

total_num

62000640

7. Compare the TOTAL_EARNINGS_MALE and TOTAL_EARNINGS_FEMALE earnings irrespective of occupation by each year

 Run ☐ Cancel query  Save query  Export data as  Show only Editor

```
1 SELECT year,
2 SUM(Try_cast(total_earnings_female as float)) AS total_num_female,
3 SUM(Try_cast(total_earnings_male as float)) AS total_num_male
4 FROM [dbo].[emp]
5 GROUP BY year
6 ORDER BY year
```

Results Messages

 Search to filter items...

year	total_num_female	total_num_male
2013	264652848	324609384
2014	269894496	329645400
2015	273222252	333058212
2016	276907224	341563656

8. How much money (TOTAL_EARNINGS_FEMALE) did female workers make as engineers in 2016?

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT year,
2 SUM(Try_cast(total_earnings_female as float)) AS total_earnings
3 FROM [dbo].[emp]
4 WHERE (occupation LIKE '%engineer%' OR occupation LIKE '%engineers%') AND year = '2016'
5 GROUP BY year
```

Results Messages

Search to filter items...

year	total_earnings
2016	22131048

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

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT year,
2 ROUND(SUM((Try_cast(full_time_male as float))*(Try_cast(workers_male as float))/100),0) AS Full_Male,
3 ROUND(SUM((Try_cast(part_time_male as float))*(Try_cast(workers_male as float))/100),0) AS Part_Male,
4 ROUND(SUM((Try_cast(full_time_female as float))*(Try_cast(workers_female as float))/100),0) AS Full_Female,
5 ROUND(SUM((Try_cast(part_time_female as float))*(Try_cast(workers_female as float))/100),0) AS Part_Female
6 FROM [dbo].[emp]
7 GROUP BY year
8 ORDER BY year
```

Results Messages

Search to filter items...

year	Full_Male	Part_Male	Full_Female	Part_Female
2013	585929851	88327745	378817719	133098117
2014	603963263	87861781	387761765	134828215
2015	620646876	87854124	400973134	135087206
2016	630321511	89223593	411289530	136366302