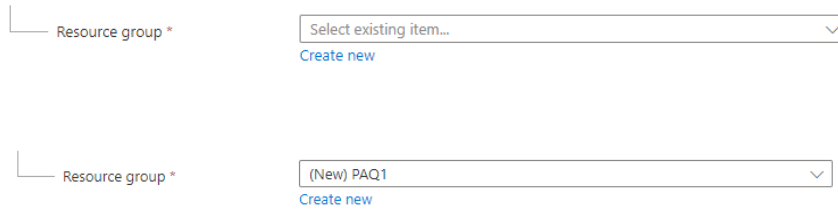


## Part A:

### 1. Resources

- a. Creating a resource group: click create any resource and then create new under resource group.



Resource group \*

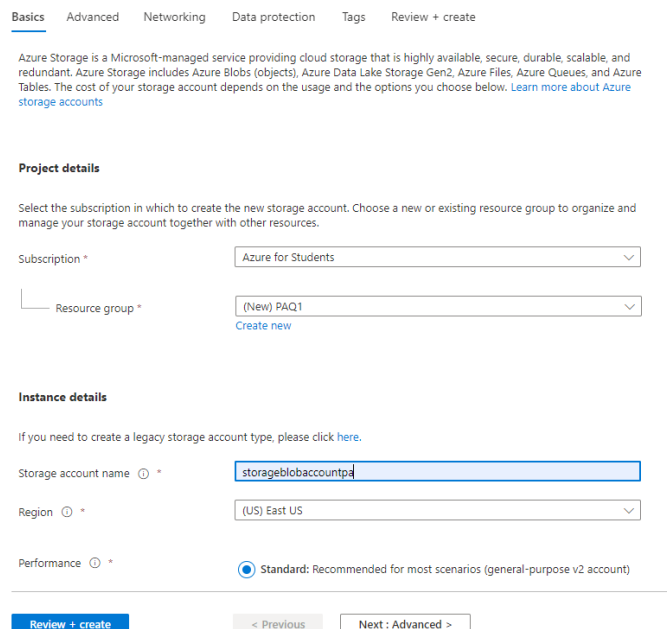
Select existing item... [Create new](#)

(New) PAQ1 [Create new](#)

### b. Creating Blob Storage Account + container:

- i. Click create a resource
- ii. Create name
- iii. Click Review + Create
- iv. Wait for Validation to Pass
- v. Click Create
- vi. Deploy Resource
- vii. Click on resource
- viii. Click Blob Service
- ix. Create Container
- x. Name input
- xi. Upload .csv file to input container

### Create a storage account ...



Basics Advanced Networking Data protection Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

**Project details**

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription \* Azure for Students

Resource group \* (New) PAQ1 [Create new](#)

**Instance details**

If you need to create a legacy storage account type, please click [here](#).

Storage account name ⓘ \* storageblobaccountpa

Region ⓘ \* (US) East US

Performance ⓘ \* ☒ Standard: Recommended for most scenarios (general-purpose v2 account)

[Review + create](#) < Previous Next: Advanced >

Microsoft Azure

Home > storageblobaccountpa\_1637013388055 | Overview

Deployment

Search (Ctrl+F)

Delete Cancel Redeploy Refresh

Overview

Inputs

Outputs

Template

Deployment is in progress

Deployment name: storageblobaccountpa\_1637013388055  
Subscription: Azure for Students  
Resource group: PAQ1

Start time: 11/15/2021, 4:56:30 PM  
Correlation ID: 3e24baef-f9b1-4638-893b-b1748fec3a78

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			

Notifications

More events in the activity log → Dismiss all

Upload Completed for gender\_jobs\_distribution2.csv  
379.72 KIB | storageblobaccountpa  
a few seconds ago

Successfully created storage container  
Successfully created storage container 'input'.  
a few seconds ago

### c. Create Data Factory

- Click create a resource
- Search for data factory
- Click create
- Select resource group
- Create name
- Click configure git later on 'git configuration'
- Click Review + Create
- Wait for validation to pass
- Click Create
- Deploy Resource

Home > Create a resource > Marketplace

### Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

#### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*

Resource group \*   
[Create new](#)

#### Instance details

Region \*

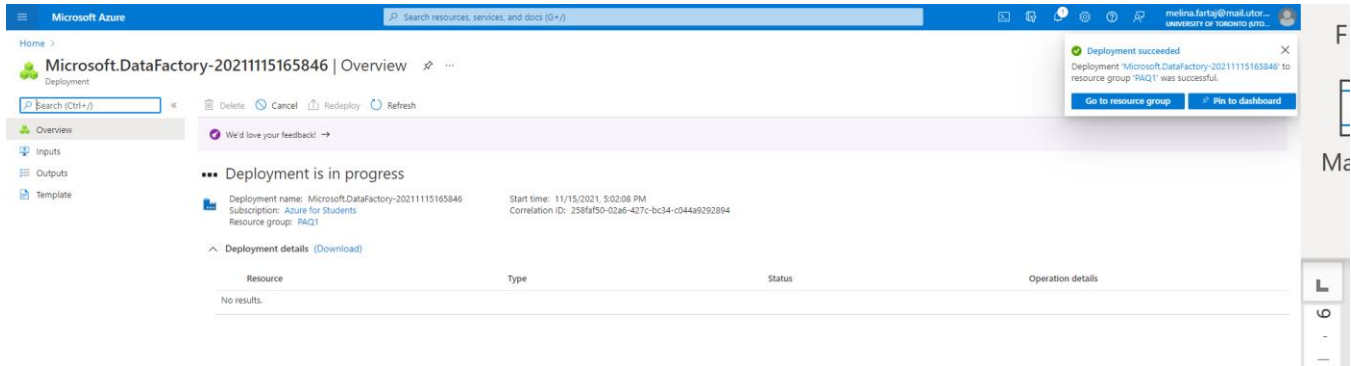
Name \*

Version \*

Review + create

< Previous

Next : Git configuration >



- d. Create SQL Database
  - i. Click create a resource
  - ii. Find SQL database
  - iii. Click create
  - iv. Select resource group
  - v. Create name
  - vi. Create server
  - vii. Click review + create
  - viii. Wait for validation to pass
  - ix. Click create
  - x. Deploy resource

Microsoft

Server details

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name \*

Location \*

Authentication

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Azure AD authentication [Learn more](#) or using an existing Azure AD user, group, or application as Azure AD admin [Learn more](#), or select both SQL and Azure AD authentication.

Authentication method

☒ Use SQL authentication

☐ Use only Azure Active Directory (Azure AD) authentication

☐ Use both SQL and Azure AD authentication

Server admin login \*

Password \*

Confirm password \*

Microsoft

Create SQL Database

Basics Networking Security Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more](#) ⓘ

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \*

Resource group \*

[Create new](#)

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name \*

Server \*

[Create new](#)

Want to use SQL elastic pool? \* ☐ Yes ☒ No

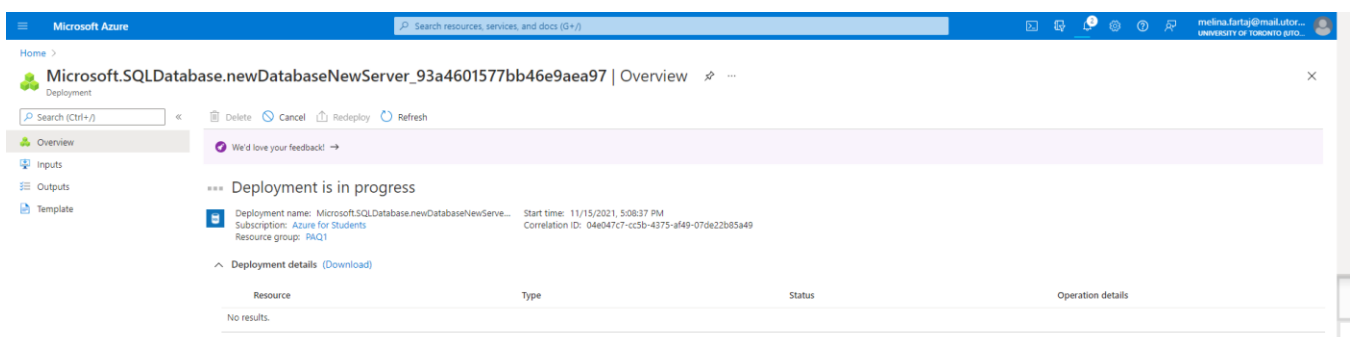
Compute + storage \*

**General Purpose**

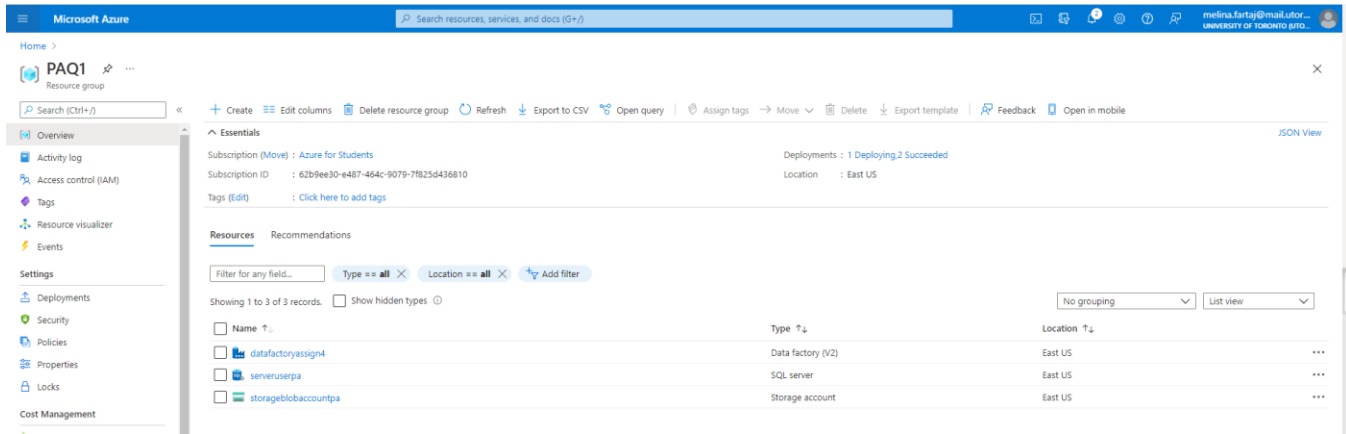
Gen5, 2 vCores, 32 GB storage, zone redundant disabled

[Configure database](#)

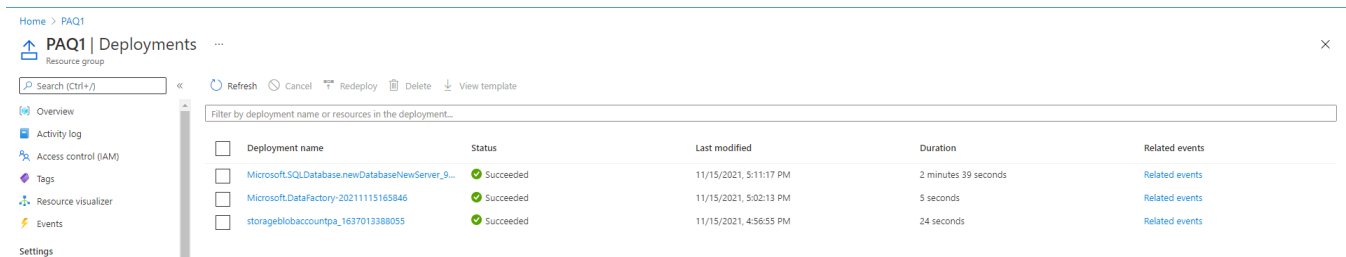
Backup storage redundancy



- e. Overview of Resource group:



## f. Deployed resources



## 2. Pipeline

- a. Click on datafactory (datafactoryassign4) in resource group
- b. Click open Azure Data Factory Studio
- c. Orchestrate
- d. Move copy data to pane
- e. Change name under general
- f. Create new source dataset
  - i. Click New
  - ii. Azure Blob Storage
  - iii. DelimitedText
  - iv. Name Source and create Linked service
- g. Create Sink dataset
  - i. Create sql table in sql database
  - ii. Click New
  - iii. Azure SQL Database
  - iv. Name Sink and create Linked service
- h. Validate Pipeline
- i. Debug Pipeline
- j. Publish





DelimitedText  
SourceDataset\_20w

Connection Schema Parameters

Linked service \* AzureBlobStorageConnection Test connection Edit + New Learn more

File path \* input / Directory / gender\_jobs\_distribut Browse Preview data

Compression type None

Column delimiter Comma (,) Edit

Row delimiter Default (\r\n, or \n\n) Edit

Encoding Default(UTF-8)

Escape character Backslash (\) Edit

Quote character Double quote (") Edit

First row as header ☒

New linked service (Azure SQL Database)

Name \* AzureSqlDatabaseConnection

Description

Connect via integration runtime \* AutoResolveIntegrationRuntime

Connection string Azure Key Vault

Account selection method From Azure subscription Enter manually

Azure subscription Select all

Server name \* No server

Database name \* No database

Authentication type \* SQL authentication

User name \*

Password Azure Key Vault

Always encrypted ☐

Additional connection properties

Create Cancel Test connection



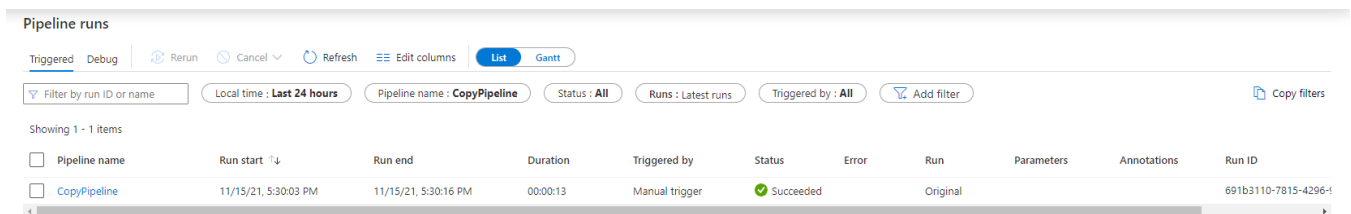
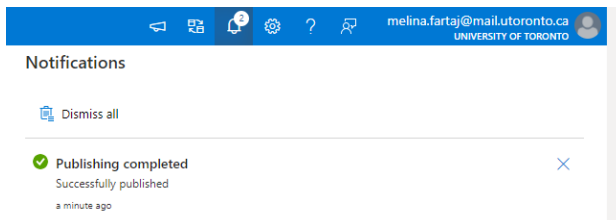
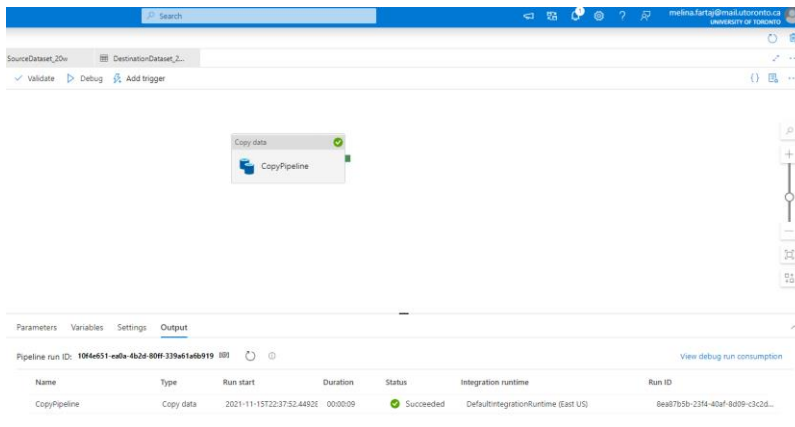
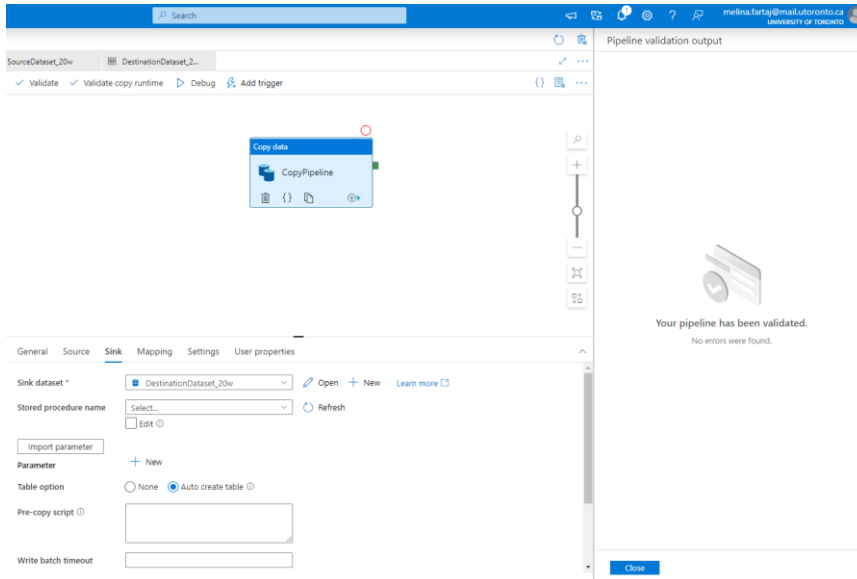
Azure SQL Database  
DestinationDataset\_20w

Connection Schema Parameters

Linked service \* AzureSqlDatabaseConnection Test connection Edit + New Learn more

Table gender\_jobs\_data.Azure Blob Storage Refresh Preview data

Edit



### 3. Triggers (screenshot + explain)

#### a. Different Types of Triggers

- Manual – This is on demand execution. The user goes into data factory designer studio and start this pipeline manually.
- Schedule – Scheduling the time your pipeline gets triggered based on a wall clock time.
- Tumbling Window- User defines a time and says that each time interval becomes a slice and that 1 slice will execute the data at that point of time and so on. This goes well with time series data.

- iv. Event-based – Trigger based on a new file arrival or file deletion. Trigger is based on an event.
- v. Logic App Calls – Actions within our logic apps to create and run pipelines.
- b. Schedule Trigger every 3 min
  - i. Click on Author tab
  - ii. Click add trigger
  - iii. New/edit trigger
  - iv. Schedule for type
  - v. Every 3 min
  - vi. OK
  - vii. Publish

melina.fartaj@mail.utoronto.ca  
UNIVERSITY OF TORONTO

### New trigger

Name \*

Description

Type \*

Schedule

Start date \* ⓘ

Time zone \* ⓘ

Coordinated Universal Time (UTC)

Recurrence \* ⓘ

Minute(s)

☐ Specify an end date

Annotations

[+ New](#)

Start trigger ⓘ

☒ Start trigger on creation

OK

Cancel

datafactoryassign4

melina.fartaj@mail.utoronto.ca  
UNIVERSITY OF TORONTO

cookies, this can happen if you are in incognito mode or have an ad blocker enabled. Please go to your browser privacy settings to verify that it is not blocking third party cookies. [Learn more](#)

### Pipeline runs

Triggered

Debug

Run

Cancel

Refresh

Edit columns

List

Gantt

Local time: Last 24 hours

Pipeline name: All

Status: All

Runs: Latest runs

Triggered by: All

Add filter

Copy filters

Showing 1 - 13 items

<input type="checkbox"/>	Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run	Parameters	Annotations	Run ID
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:22:00 PM	11/15/21, 6:22:10 PM	00:00:10	TriggerEvery3Min	Succeeded		Original			3a30f639-c371-4f32-b...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:19:03 PM	11/15/21, 6:19:12 PM	00:00:09	TriggerEvery3Min	Succeeded		Original			945a1941-abee-4d04-f...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:16:00 PM	11/15/21, 6:16:12 PM	00:00:11	TriggerEvery3Min	Succeeded		Original			90f69470-8128-47c2-b...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:13:00 PM	11/15/21, 6:13:14 PM	00:00:13	TriggerEvery3Min	Succeeded		Original			70aefc93-8f37-47b5-a...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:10:00 PM	11/15/21, 6:10:14 PM	00:00:13	TriggerEvery3Min	Succeeded		Original			48006c3b-121c-498f-b...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:07:02 PM	11/15/21, 6:07:12 PM	00:00:10	TriggerEvery3Min	Succeeded		Original			5cccd805-ccac-4d4c-b...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:04:00 PM	11/15/21, 6:04:10 PM	00:00:10	TriggerEvery3Min	Succeeded		Original			6e6ca478-190b-43db-i...
<input type="checkbox"/>	CopyPipeline	11/15/21, 6:01:00 PM	11/15/21, 6:01:13 PM	00:00:13	TriggerEvery3Min	Succeeded		Original			8aac85fc-b414-4114-a...
<input type="checkbox"/>	CopyPipeline	11/15/21, 5:58:00 PM	11/15/21, 5:58:12 PM	00:00:11	TriggerEvery3Min	Succeeded		Original			70deb7e1-2d8f-49ff-a...
<input type="checkbox"/>	CopyPipeline	11/15/21, 5:55:00 PM	11/15/21, 5:55:12 PM	00:00:12	TriggerEvery3Min	Succeeded		Original			2b78547f-3ab0-41dc-t...
<input type="checkbox"/>	CopyPipeline	11/15/21, 5:52:00 PM	11/15/21, 5:52:11 PM	00:00:10	TriggerEvery3Min	Succeeded		Original			4e6c9d4a-8476-4520-t...
<input type="checkbox"/>	CopyPipeline	11/15/21, 5:49:00 PM	11/15/21, 5:49:14 PM	00:00:13	TriggerEvery3Min	Succeeded		Original			9398eaff-84fb-4318-b...
<input type="checkbox"/>	CopyPipeline	11/15/21, 5:30:03 PM	11/15/21, 5:30:16 PM	00:00:13	Manual trigger	Succeeded		Original			691b3110-7815-4296-t...



Trigger name	Trigger type	Trigger time	Status	Pipelines	Run	Message	Properties	Run ID
TriggerEvery3Min	Schedule trigger	11/15/21, 6:25:00 PM	Succeeded	1	Original			08585445881852542178388278357C160
TriggerEvery3Min	Schedule trigger	11/15/21, 6:22:00 PM	Succeeded	1	Original			0858544588365366336667862804C169
TriggerEvery3Min	Schedule trigger	11/15/21, 6:19:02 PM	Succeeded	1	Original			08585445885425378880725229487C191
TriggerEvery3Min	Schedule trigger	11/15/21, 6:16:00 PM	Succeeded	1	Original			08585445887250390554133785403C195
TriggerEvery3Min	Schedule trigger	11/15/21, 6:13:00 PM	Succeeded	1	Original			08585445889051231938442382824C192
TriggerEvery3Min	Schedule trigger	11/15/21, 6:10:00 PM	Succeeded	1	Original			0858544589094099457048037790C166
TriggerEvery3Min	Schedule trigger	11/15/21, 6:07:02 PM	Succeeded	1	Original			08585445892633946381742182242C170
TriggerEvery3Min	Schedule trigger	11/15/21, 6:04:00 PM	Succeeded	1	Original			08585445894451245001730896762C147
TriggerEvery3Min	Schedule trigger	11/15/21, 6:00:59 PM	Succeeded	1	Original			085854458962548370812417123900C168
TriggerEvery3Min	Schedule trigger	11/15/21, 5:58:00 PM	Succeeded	1	Original			08585445898049942430866366749C190
TriggerEvery3Min	Schedule trigger	11/15/21, 5:55:00 PM	Succeeded	1	Original			08585445899851834905980733129C142
TriggerEvery3Min	Schedule trigger	11/15/21, 5:52:00 PM	Succeeded	1	Original			08585445901654415478643826713C165
TriggerEvery3Min	Schedule trigger	11/15/21, 5:48:59 PM	Succeeded	1	Original			0858544590345558712967482748C158

#### 4. Replicate objects

##### a. Create two ADLS Gen 2 Storage accounts

- Click create resource → storage account
- Put in resource group, storage name and location (Canada central for one and west Europe for another)
- Go to advanced tab
- Click 'Enable hierarchical namespace' for Data Lake Storage Gen2
- Review + Create
- Create
- Create Containers in both storage account

##### b. Use AzCopy to copy objects

- AzCopy is a command-line utility that can be used to copy data to and from storage accounts.
- Open cloud shell
- Type `azcopy - - version` to check downloaded (it should be already there)
- Receive the SAS token
  - Click on storage resource
  - Shared access signature under security + networking
  - Check mark container and object under allowed resource types
  - Click Generate SAS and connection string
  - Copy the SAS token
  - Repeat for source and destination storage accounts
- Use format below command to copy all storage account contents to another storage account.
- Fill in the required names and paste in the SAS tokens and enter into the bash command

```
azcopy copy 'https://<source-storage-account-name>.file.core.windows.net/<SAS-token>' 'https://<destination-storage-account-name>.file.core.windows.net/<SAS-token>' --recursive'
```

## Part B:

### 1. Occupations

Search resources, services, and docs (G+)

melina.fartaj@mailutor...  
UNIVERSITY OF TORONTO

ry editor (preview) ...

Feedback

Query 3

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 SELECT DISTINCT occupation
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE major_category = 'Computer, Engineering, and science' AND year = '2013'
4 ;
```

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. Count of Occupations

Search resources, services, and docs (G+)

melina.fartaj@mailutor...  
UNIVERSITY OF TORONTO

ry editor (preview) ...

Feedback

Query 3 Query 5

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 SELECT COUNT(DISTINCT occupation)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE minor_category = 'Business and Financial Operations'
4 ;
```

Results Messages

Search to filter items...

28
----

### 3. Relevant Information

Search resources, services, and docs (G+)

melina.fartaj@mail.utor...

UNIVERSITY OF TORONTO

ery editor (preview) ...

Feedback

Query 3 × Query 5 × Query 6 ×

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 SELECT *
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE occupation = 'bus drivers'
4 ;
```

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
2014	Bus drivers	Production, Transportation, and...	Transportation	267775	161334	106441
2015	Bus drivers	Production, Transportation, and...	Transportation	288778	174214	114564
2016	Bus drivers	Production, Transportation, and...	Transportation	280228	178493	101735

### 4. Number of Female Workers

Search resources, services, and docs (G+)

melina.fartaj@mail.utor...

UNIVERSITY OF TORONTO

ery editor (preview) ...

Feedback

Query 7 × Query 8 × Query 9 ×

Run

Cancel query

Save query

Export data as

Show only Editor

```
1 SELECT year, SUM(workers_female)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE major_category = 'Management, Business, and Financial'
4 GROUP BY year
5 ;
```

Results Messages

Search to filter items...

year	
2013	7748347
2014	8061480
2015	8381812
2016	8617853

## 5. Total earnings in Service

Search resources, services, and docs (G+)

melina.fartaj@mail.utor...  
UNIVERSITY OF TORONTO

ery editor (preview) ...

Feedback

Query 7 × Query 8 × Query 9 × Query 10 ×

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT SUM(total_earnings_male)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE major_category = 'Service' and year = '2015'
4 ;
```

Results Messages

Search to filter items...

2502426

## 6. Management roles

Search resources, services, and docs (G+)

melina.fartaj@mail.utor...  
UNIVERSITY OF TORONTO

ry editor (preview) ...

Feedback

Query 7 × Query 8 × Query 9 × Query 10 × Query 11 ×

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT SUM(workers_female)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 WHERE minor_category = 'Management' and year = '2015'
4 ;
```

Results Messages

Search to filter items...

5166720

## 7. Total earnings

Search resources, services, and docs (G+)

melina.fartaj@mail.utor...  
UNIVERSITY OF TORONTO

ery editor (preview) ...

Feedback

Query 7 × Query 8 × Query 9 × Query 10 × Query 11 × Query 12 ×

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT year, SUM(total_earnings_male), SUM(total_earnings_female)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 GROUP BY year
4 ;
```

Results Messages

Search to filter items...

year		
2016	28463638	23075602
2014	27470450	22491208
2015	27754851	22768521
2013	27050782	22054404

## 8. Total earnings as Engineers

Search resources, services, and docs (G+)

melina.fartaj@mailutor...  
UNIVERSITY OF TORONTO

ry editor (preview) ...

Feedback

Query 7 ×

Query 8 ×

Query 9 ×

Query 10 ×

Query 11 ×

Query 12 ×

Query 13 ×

Run

Cancel query

Save query

Export data as

Show only Editor

```

1  --assuming management engineers count also
2
3  SELECT SUM(total_earnings_female)
4  FROM [gender_job_data].[Azure Blob Storage file]
5  WHERE occupation LIKE '%engineer%' and year = '2016'
6  ;

```

Results

Messages

Search to filter items...

1844254

## 9. Number of Workers

**Query editor (preview)**

Feedback

Query 7 × Query 8 × Query 9 × Query 10 × Query 11 × Query 12 × Query 13 × **Query 14 ×**

Run ☐ Cancel query  Export data as Show only Editor

```

1 SELECT year, SUM(full_time_female), SUM(full_time_male), SUM(part_time_female), SUM(part_time_male)
2 FROM [gender_job_data].[Azure Blob Storage file]
3 GROUP BY year
4 ;

```

**Results** Messages

year	SUM(full_time_female)	SUM(full_time_male)	SUM(part_time_female)	SUM(part_time_male)
2016	39202.199999999996	45727.199999999996	12997.799999999999	6472.7999999999995
2014	38732.4	45570.600000000002	13467.599999999999	6629.3999999999994
2015	39045.6	45727.199999999996	13154.400000000001	6472.7999999999995
2013	38628	45361.800000000004	13572	6838.2000000000006