

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.

Resource groups ...

University of Toronto (utoronto.onmicrosoft.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags

You are viewing a new version of Browse experience. Click here to access the old experience.

Filter for any field... Subscription equals all Location equals all Add filter

Name	Subscription	Location
MIE1628	Azure for Students	East US

mie1628johnzhou_1752727880861 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Inputs Outputs Template

Your deployment is complete

Deployment name: mie1628johnzhou_1752727880861
Subscription: Azure for Students
Resource group: MIE1628

Deployment details

Next steps

Go to resource Give feedback Tell us about your experience with deployment

Microsoft.SQLDatabase.newDatabaseNewServer_b512eb278f9a4b44a9d7b | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Inputs Outputs Template

Deployment is in progress

Deployment name: Microsoft.SQLDatabase.newDatabaseNewServer_b512eb278f9a4b44a9d7b
Subscription: Azure for Students
Resource group: MIE1628

Deployment details

Resource	Type	Status	Operation details
mie1628db	Microsoft.Sql/servers	Accepted	Operation details

Microsoft.DataFactory-20250717005318 | Overview

Deployment

Search Delete Cancel Redeploy Download Refresh

Overview

Inputs Outputs Template

Your deployment is complete

Deployment name: Microsoft.DataFactory-20250717005318
Subscription: Azure for Students
Resource group: MIE1628

Deployment details

Next steps

Go to resource Give feedback Tell us about your experience with deployment

MIE1628

Resource group

Search Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template Open in mobile

Overview

Activity log Access control (IAM) Tags Resource visualizer Events Settings Cost Management Monitoring Automation Help

Essentials

Subscription (move) : Azure for Students
Subscription ID : 47a50360-c65b-4826-b1ba-8ec39d212e2
Tags (edit) Add tags

Deployments : 4 Succeeded
Location : East US

Resources Recommendations

Showing 1 to 4 of 4 records. □ Show hidden types Add filter

Name	Type	Location
mie1628db	SQL server	Canada Central
mie1628johnzhou	Storage account	Canada Central
mie1628johnzhouDF	Data factory (V2)	Canada Central
mie1628sqlDB (mie1628db/mie1628sqlDB)	SQL database	Canada Central

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:

A screenshot of the Azure Blob Storage interface. At the top, there are navigation buttons: '+ Add Directory', 'Upload', 'Change access level', 'Refresh', 'Delete', 'Copy', 'Paste', 'Rename', 'Acquire lease', 'Break lease', and 'Edit columns'. Below this, a container named 'mie1628container' is selected. The authentication method is set to 'Access key (Switch to Microsoft Entra user account)'. A search bar at the top right contains the placeholder 'Search blobs by prefix (case-sensitive)' and a filter button 'Add filter'. The results table shows one item: 'gender_jobs_distribution2.csv'. Columns include 'Name', 'Last modified', 'Access tier', 'Blob type', 'Size', and 'Lease state'. The blob was last modified on 7/17/2023, 1:11:50 AM, has a 'Hot (inferred)' access tier, is a 'Block blob', and is 379.72 KB in size, with an 'Available' lease state.

setup sink and source

A screenshot of the Azure Data Factory Pipeline designer. The top navigation bar includes 'all', 'Publish all', and a three-dot menu. The main area is titled 'Activities' and shows a list of activities: 'DelimitedText1', 'AzureSqlTable1', and 'pipeline1'. Under 'Activities', there are sections for 'Move and transform' (containing 'Copy data' and 'Data flow'), 'Synapse', 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Databricks', 'Data Lake Analytics', 'General', 'HDInsight', 'Iteration & conditionals', 'Machine Learning', and 'Power Query'. In the center, a 'Copy data' activity is selected and previewed. The preview window shows the configuration: 'Copy data mie1628' with a green checkmark. Below the preview are icons for trash, copy, and more options. The bottom of the screen shows tabs for 'General', 'Source', 'Sink', 'Mapping', 'Settings', and 'User properties', with 'Sink' currently selected. The 'Sink dataset' dropdown is set to 'AzureSqlTable1'. There are also buttons for 'Open', 'New', and 'Learn more'.

Copy data

Copy data mie1628

Parameters Variables Settings Output

Pipeline run ID: b666e322-b4c3-4040-81c1-ddbfadd306a0 Pipeline status: Succeeded

All status: Succeeded

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User properties	Activity run ID	Log
Copy data mie1628	Succeeded	Copy data	7/17/2025, 2:27:51 AM	15s	AutoResolveIntegrationRuntime (Canada Central)		2cb08b09-c871-46db-97cc-da134bf62de8	

Detail of debug running:

Showing 1 - 8 items

Pipeline name	Run start	Run end	Duration	Status	Triggered by	Run ID	Last refre
pipeline1	7/17/2025, 2:27:49 AM	7/17/2025, 2:28:07 AM	18s	Succeeded	Manual trigger	b666e322-b4c3-4040-81c1-ddbfadd306a0	
pipeline1	7/17/2025, 2:26:16 AM	7/17/2025, 2:26:39 AM	23s	Succeeded	Manual trigger	eadcef8-fa15-43f0-857-4db449591a7c	
pipeline1	7/17/2025, 2:22:42 AM	7/17/2025, 2:23:05 AM	24s	Succeeded	Manual trigger	69b84010-f1c8-4561-96e1-15e38a441e70	

Data Factory > mie1628johnzhouDF

Details Refresh

Learn more on copy performance details from here.

Activity run id: 2cb08b09-c871-46db-97cc-da134bf62de8

Azure Blob Storage → Azure SQL Database

Succeeded

Azure Blob Storage Region: Canada Central **Azure SQL Database** Region: Canada Central

Data read: 387,004 KB **Data written:** 619,128 KB

Files read: 1 **Rows written:** 2,088

Rows read: 2,088 **Peak connections:** 2

Peak connections: 1

Copy duration: 00:00:12 **Throughput:** 129,001 KB/s

Start time: 7/17/2025, 2:27:52 AM **Used DLU:** 4

Used parallel copies: 1 **Duration:** 00:00:12

Details **Working duration** **Total duration**

Queue 00:00:07

Transfer [Listing source 00:00:00 Reading from source 00:00:00 Writing to sink 00:00:00] 00:00:03

Data consistency verification: Not verified

Showing 1 - 1 of 1 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime
Copy data mie1628	Succeeded	Copy data	7/17/2025, 2:27:51 AM	15s	AutoResolveIntegrationRuntime (Canada Central)

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 happens 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 tasks in a regular basis.
- Event-based trigger: trigger the pipeline in response to certain events happening within Azure resources
- Manual via UI trigger: manually trigger the pipeline from ADF user interface (UI) for on-demand basis.

Scheduled 5 successful runs every 3 min:

New trigger

Name *
every3mintrigger

Description

Type *
Schedule

Start date *
7/15/2025, 2:57:00 AM

Time zone *
Eastern Time (US & Canada) (UTC-4)

This time zone observes daylight savings. Trigger will auto-adjust for one hour difference.

Recurrence *
Every 3 Minute(s)

Specify an end date

Annotations
+ New

Start trigger
 Start trigger on creation

Pipeline runs

Showing 1 - 6 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run	Parameters	Annotations	Run ID
pipeline1	7/17/2025, 3:09:00 AM	7/17/2025, 3:09:23 AM	23s	every3min	Succeeded	Original			29be08e3-b3bc402d-88ea-e49182d133f0
pipeline1	7/17/2025, 3:09:00 AM	7/17/2025, 3:09:24 AM	23s	every3min	Succeeded	Original			e76677b-4cfb-4d3e-93f5-17056d33e262
pipeline1	7/17/2025, 3:09:00 AM	7/17/2025, 3:09:23 AM	24s	every3min	Succeeded	Original			13c5a426-9521-40d4-8cf4-a377a818279
pipeline1	7/17/2025, 3:09:01 AM	7/17/2025, 3:09:22 AM	21s	every3min	Succeeded	Original			3ca3d449-61c1-4024-a9b1-e4631e107da
pipeline1	7/17/2025, 2:57:02 AM	7/17/2025, 2:57:22 AM	20s	every3min	Succeeded	Original			803dd630-4fb0-44f9-965a-2a44de775270

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 objects 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**

The screenshot shows a Microsoft Azure Data Studio interface. On the left, there's a sidebar with various navigation links like Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview), Mirror database in Fabric (preview), Resource visualizer, Settings, Data management, Integrations, Power Platform, Security, Intelligent performance, Monitoring, Automation, and Help. The 'Query editor (preview)' link is currently selected.

The main area is titled 'Query1' and contains the following SQL code:

```
1 SELECT *
2 FROM gender_jobs_distribution
3 WHERE MAJOR_CATEGORY = 'Computer, Engineering, and Science' AND YEAR = 2013;
```

Below the code, there are buttons for Run, Cancel query, Save query, Export data as, and Show only Editor.

The results section shows a table with the following data:

year	occupation	major_category	minor_category	total_workers	workers_male	workers_female	percent_female	total_earnings	total_earnings_male	total_earnings_female
2013	Computer and information re...	Computer, Engineering, and Sci...	Computer and mathematical	129993	9222	3771	29	95918	100222	88006
2013	Computer systems analysts	Computer, Engineering, and Sci...	Computer and mathematical	441533	280636	160912	38.4	75888	81174	65146
2013	Information security analysts	Computer, Engineering, and Sci...	Computer and mathematical	50953	40681	10172	20	84140	86349	8045
2013	Computer programmers	Computer, Engineering, and Sci...	Computer and mathematical	374314	280175	76139	20.3	80312	81308	73000
2013	Software developers, applicati...	Computer, Engineering, and Sci...	Computer and mathematical	924888	741908	183880	19.8	94225	96686	83808
2013	Web developers	Computer, Engineering, and Sci...	Computer and mathematical	136446	95072	40174	28.9	80376	81566	5995
2013	Computer support specialists	Computer, Engineering, and Sci...	Computer and mathematical	456736	329470	119296	26	52053	52990	49001
2013	Database administrators	Computer, Engineering, and Sci...	Computer and mathematical	101323	61969	39354	38.8	75024	84405	61378
2013	Network and computer systems...	Computer, Engineering, and Sci...	Computer and mathematical	169372	103976	34998	17.9	67246	68383	61669
2013	Computer network architects	Computer, Engineering, and Sci...	Computer and mathematical	91799	64233	7476	0.2	96372	96540	94445
2013	Computer, all other	Computer, Engineering, and Sci...	Computer and mathematical	354952	277889	77061	21.7	86550	66971	61519
2013	Actuaries	Computer, Engineering, and Sci...	Computer and mathematical	23200	16642	6558	28.3	112191	116019	105759
2013	Mathematicians	Computer, Engineering, and Sci...	Computer and mathematical	1633	1603	320	17.1	76256	77796	69362
2013	Operations research analysts	Computer, Engineering, and Sci...	Computer and mathematical	123658	65206	58430	47.3	77172	80748	68975
2013	Statisticians	Computer, Engineering, and Sci...	Computer and mathematical	31319	16117	15202	48.5	81209	87488	74211
2013	Miscellaneous mathematical sci...	Computer, Engineering, and Sci...	Computer and mathematical	3668	1974	1712	46.4	70370	75321	56118
2013	Architects, except naval	Computer, Engineering, and Sci...	Architecture and Engineering	145439	114341	31598	21.4	71305	75395	60318

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

>>

Query 1 X

▶ Run Cancel query ⬇ Save query ⬇ Export data as ▼ >Show only Editor

```
1  SELECT COUNT(DISTINCT occupation)
2  FROM gender_jobs_distribution
3  WHERE minor_category = 'Business and Financial Operations';
4
5
```

Results Messages

Search to filter items...

28

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

The screenshot shows the Microsoft Power BI Query editor interface. The left sidebar contains navigation links for Overview, Activity log, Tags, Diagnose and solve problems, and a section for Query editor (preview) which is currently selected. Under Query editor (preview), there are options for Migrating database in Fabric (preview), Resource visualizer, Settings, Data management, Integration, Power Platform, Security, Intelligent performance, Monitoring, and Automation.

The main area displays a query titled "Query 1" with the following DAX code:

```
1 SELECT
2     SUM([order_lines])
3     WHERE occupation = 'Bus drivers';
4
5
```

Below the query, there are buttons for Run, Cancel query, Save query, Export data as, and Show only Editor.

At the bottom, a table titled "Results" shows data for bus drivers across various years. The columns include year, occupation, major_category, minor_category, total_workers, workers_male, workers_female, percent_female, total_earnings, total_earnings_male, and total_earnings_female. The data is as follows:

year	occupation	major_category	minor_category	total_workers	workers_male	workers_female	percent_female	total_earnings	total_earnings_male	total_earnings_female
2013	Bus drivers	Production	Transportation and... Transportation	279991	174820	101161	36.7	32725	27599	27180
2014	Bus drivers	Production	Transportation and... Transportation	267775	161324	106441	39.8	32414	28555	27795
2015	Bus drivers	Production	Transportation and... Transportation	268778	174214	114564	39.7	32398	30072	20995
2016	Bus drivers	Production	Transportation and... Transportation	280228	178498	107735	36.30436644	35222	38399	29570
2013	Bus drivers	Production	Transportation and... Transportation	279991	174820	101161	36.7	32725	47569	27180
2014	Bus drivers	Production	Transportation and... Transportation	267775	161324	106441	39.8	32414	38055	27795
2015	Bus drivers	Production	Transportation and... Transportation	268778	174214	114564	39.7	32398	38072	27095
2016	Bus drivers	Production	Transportation and... Transportation	280228	178498	107735	36.30436644	35222	38399	29570
2013	Bus drivers	Production	Transportation and... Transportation	279991	174820	101161	36.7	32725	37599	27180
2014	Bus drivers	Production	Transportation and... Transportation	267775	161324	106441	39.8	32414	36555	27795
2015	Bus drivers	Production	Transportation and... Transportation	268778	174214	114564	39.7	32398	30072	27095
2016	Bus drivers	Production	Transportation and... Transportation	280228	178498	107735	36.30436644	35222	38399	29570
2013	Bus drivers	Production	Transportation and... Transportation	279991	174820	101161	36.7	32725	37599	27180
2014	Bus drivers	Production	Transportation and... Transportation	267775	161324	106441	39.8	32414	36555	27795
2015	Bus drivers	Production	Transportation and... Transportation	268778	174214	114564	39.7	32398	30072	27095
2016	Bus drivers	Production	Transportation and... Transportation	280228	178498	107735	36.30436644	35222	38399	29570
2013	Bus drivers	Business services and... Transportation	194261	194261	101161	34.7	33514	174460	33165	

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.

The screenshot shows the Azure SQL Database interface. On the left, there's a sidebar with various navigation links like Overview, Activity log, Tags, and Resource visualizer. The main area is the Query editor (preview) where a query is being run. The query itself is:

```
1 SELECT year, SUM(workers_female)
2 FROM gender_jobs_distribution
3 WHERE major_category = 'Management, Business, and Financial'
4 GROUP BY year;
```

Below the query, the Results tab is selected, showing the following data:

year	workers_female
2013	69735123
2016	77560677
2014	72553320
2015	75436308

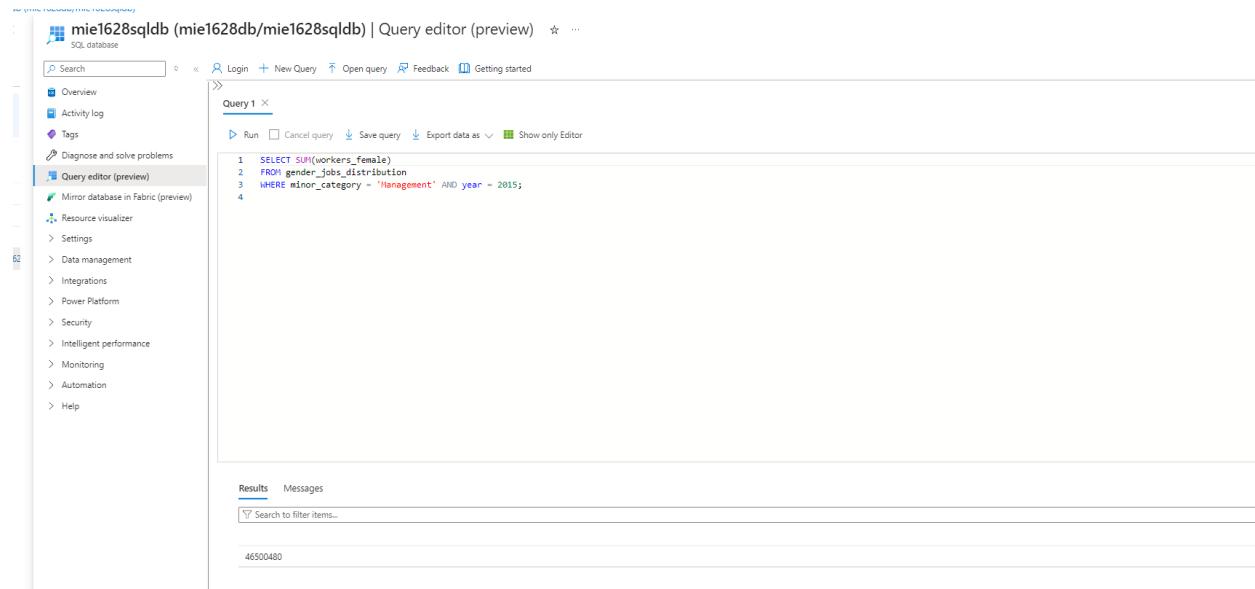
5. [Marks:5] In the gender_jobs_data table - What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?

The screenshot shows the Power BI Query Editor interface. On the left, there's a sidebar with various options like Overview, Activity log, Tags, and Resource visualizer. The main area has a search bar and navigation links for Login, New Query, Open query, Feedback, and Getting started. Below that is the Query Editor itself, which contains a tab labeled 'Query 1'. Inside 'Query 1', there's a code editor with the following SQL query:

```
1 SELECT SUM(total_earnings_male)
2 FROM gender_jobs_distribution
3 WHERE major_category = 'Service' AND year = 2015;
4
```

Below the code editor are buttons for Run, Cancel query, Save query, Export data as, and Show only Editor. At the bottom of the editor, there are tabs for Results and Messages, and a search bar for filtering items.

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



The screenshot shows the Azure portal interface for a SQL database named 'mie1628sqlldb'. The left sidebar contains navigation links for Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview) (which is selected), Mirror database in Fabric (preview), Resource visualizer, Settings, Data management, Integration, Power Platform, Security, Intelligent performance, Monitoring, Automation, and Help. The main area is titled 'Query 1' and contains the following SQL code:

```
1 SELECT SUM(workers_female)
2 FROM gender_jobs_distribution
3 WHERE minor_category = 'Management' AND year = 2015;
4
```

Below the code, there are tabs for 'Results' and 'Messages', and a search bar. The results section displays the output of the query: '46500480'.

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

The screenshot shows the Power BI Query Editor interface. On the left, there's a navigation pane with various options like Overview, Activity log, Tags, and Diagnose and solve problems. The main area has a search bar and tabs for Login, New Query, Open query, Feedback, and Getting started. Below that is the Query Editor with a toolbar for Run, Cancel query, Save query, Export data as, and Show only Editor. The query itself is:

```
1 SELECT year, SUM(total_earnings_male) AS total_earnings_male, SUM(total_earnings_female) AS total_earnings_female
2 FROM gender_jobs_distribution
3 GROUP BY year;
```

At the bottom, there are Results and Messages tabs, and a search bar for filtering items. The results table shows data for four years:

year	total_earnings_male	total_earnings_female
2013	243457038	195489636
2016	256172742	207680418
2014	247234050	202420872
2015	240793959	204916689

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

The screenshot shows the Azure SQL Database Query Editor interface. On the left, there's a sidebar with various navigation links like Overview, Activity log, Tags, and Query editor (preview). The main area has a search bar and tabs for Login, New Query, Open query, Feedback, and Getting started. A 'Query 1' tab is open, showing a SQL script:

```
1 SELECT SUM(total_earnings_female) AS total_earnings_female
2 FROM gender_jobs_distribution
3 WHERE year = 2016 AND occupation like '%engineer%';
4
```

Below the query, there are 'Run', 'Cancel query', 'Save query', 'Export data as', and 'Show only Editor' buttons. The results section shows a single row:

total_earnings_female
16598286

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

The screenshot shows the Microsoft Data Studio interface. On the left is a sidebar with navigation links: Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview) (which is selected), Mirror database in Fabric (preview), Resource visualizer, Settings, Data management, Integrations, Power Platform, Security, Intelligent performance, Monitoring, Automation, and Help.

The main area has a title bar "mie1628sqldb (mie1628db/mie1628sqldb) | Query editor (preview)" and a toolbar with Search, Login, New Query, Open query, Feedback, Getting started, Run, Cancel query, Save query, Export data as, and Show only Editor.

The "Query 1" tab is open, containing the following SQL code:

```
-- What is the total number of full-time and part-time female workers versus male workers year over year?  
SELECT year, SUM(full_time_male) AS total_full_time_male_workers, SUM(part_time_male) AS total_part_time_male_workers, SUM(full_time_female) AS total_full_time_female_workers, SUM(part_time_female) AS total_part_time_female_workers  
FROM gender_jobs_distribution  
GROUP BY year
```

The "Results" tab is selected, showing a table with the following data:

year	total_full_time_male_workers	total_part_time_male_workers	total_full_time_female_workers	total_part_time_female_workers
2013	403256	61543	347652	122146
2014	410133	39964	342391	121206
2015	411544	56205	351410	118369
2016	411544	56205	352819	116900