

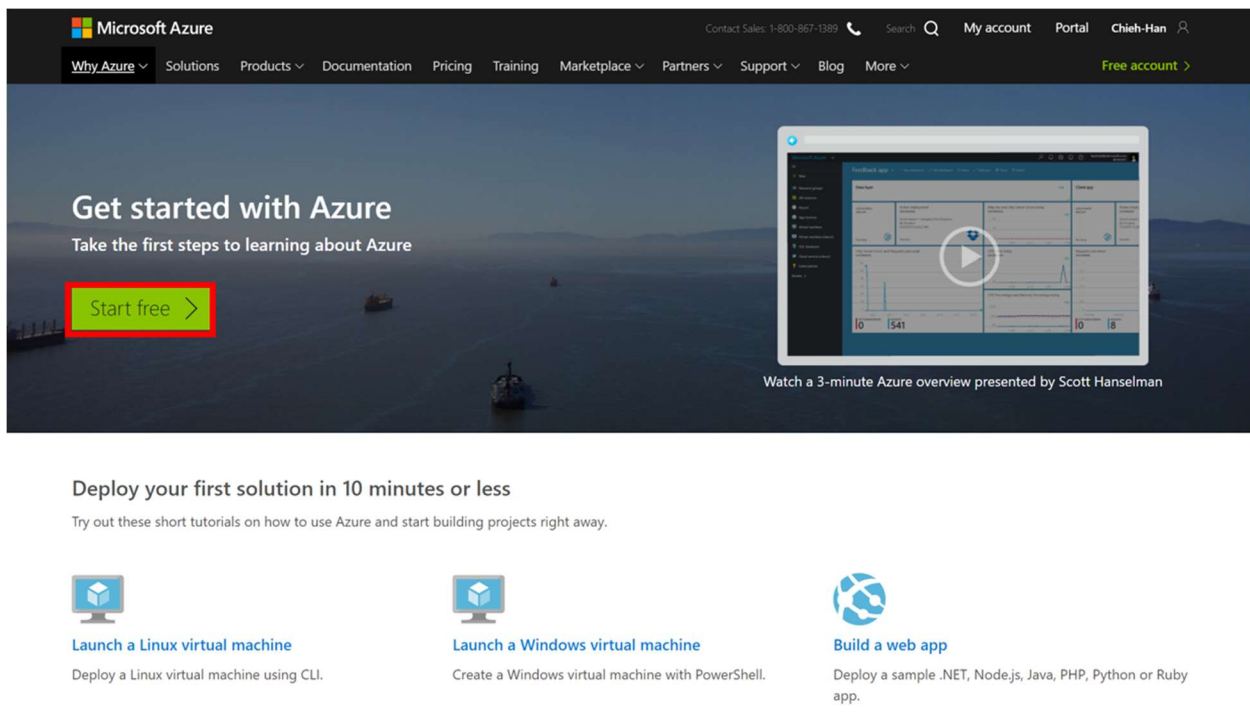
Microsoft – DAT278X

From Graph to Knowledge Graph – Algorithms and Applications

Challenge Lab Environment Setup Guide

1. Create a free Azure Account

<https://azure.microsoft.com/en-us/get-started>



Microsoft Azure

Contact Sales: 1-800-867-1389 Search My account Portal Chief-Han

Why Azure Solutions Products Documentation Pricing Training Marketplace Partners Support Blog More

Get started with Azure

Take the first steps to learning about Azure


Start free >

Watch a 3-minute Azure overview presented by Scott Hanselman

Deploy your first solution in 10 minutes or less

Try out these short tutorials on how to use Azure and start building projects right away.

- Launch a Linux virtual machine**
Deploy a Linux virtual machine using CLI.
- Launch a Windows virtual machine**
Create a Windows virtual machine with PowerShell.
- Build a web app**
Deploy a sample .NET, Node.js, Java, PHP, Python or Ruby app.



Contact Sales 1-800-867-1389

Search

My account

Portal

Chieh-Han

Why Azure

Solutions

Products

Documentation

Pricing

Training

Marketplace

Partners

Support

Blog

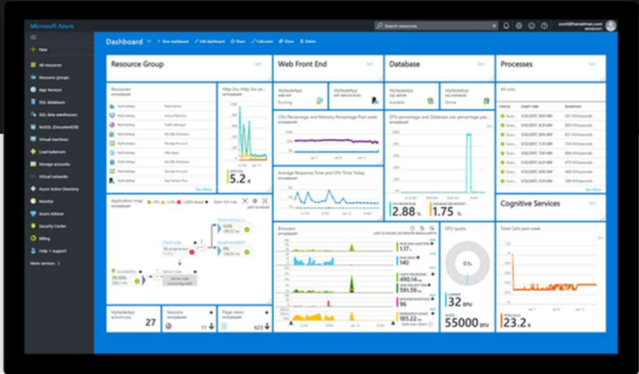
More

Create your Azure free account today

Get started building your next great idea with Azure

Start free

Or buy now




Chat live with an agent

Microsoft Azure

chiewu@microsoft.com Sign out

Azure free account sign up

Start with a \$200 credit for 30 days, and keep going for free



About you

Country/Region

United States

First name

Chieh-Han

Last name

Wu

Email address

Phone

(425) 538-5745

By proceeding you acknowledge the [privacy statement](#) and [subscription agreement](#)


Next

Microsoft Azure

chiewu@microsoft.comSign out

Azure free account sign up

Start with a \$200 credit for 30 days, and keep going for free



1 About you

2 Agreement

☒ I agree to the [subscription agreement](#), [offer details](#), [privacy statement](#), and [communications policy](#).

I will receive information, tips, and offers about Azure, including Azure Newsletter and pricing updates, and other Microsoft products and services. [privacy statement](#)

Sign up

English

Privacy & Cookies

Trademarks

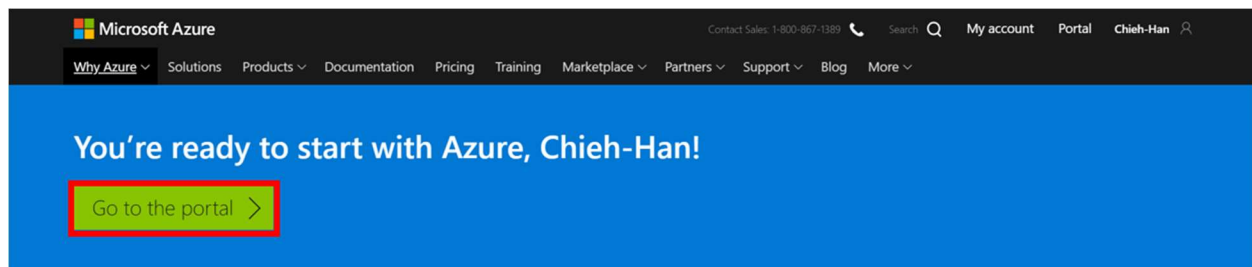
Legal

Support

Give us feedback

© 2018 Microsoft

2. Create a Data Lake Analytics and a Data Lake Store Account
 - a. Sign on to the Azure portal.



Join the demo to see Azure in action

Learn how to use the portal to create and manage resources and get your questions answered by Azure technical experts during a live Q&A. This demo covers:

- Building a virtual machine.
- Creating a web app.
- Deploying a SQL database.
- Customizing your DevOps dashboard.



Join a live demo or watch on demand

[Change time zone +](#)

- 7/12/2018, 9:00 AM PDT
- 7/19/2018, 9:00 AM PDT
- 7/26/2018, 9:00 AM PDT
- Watch on demand

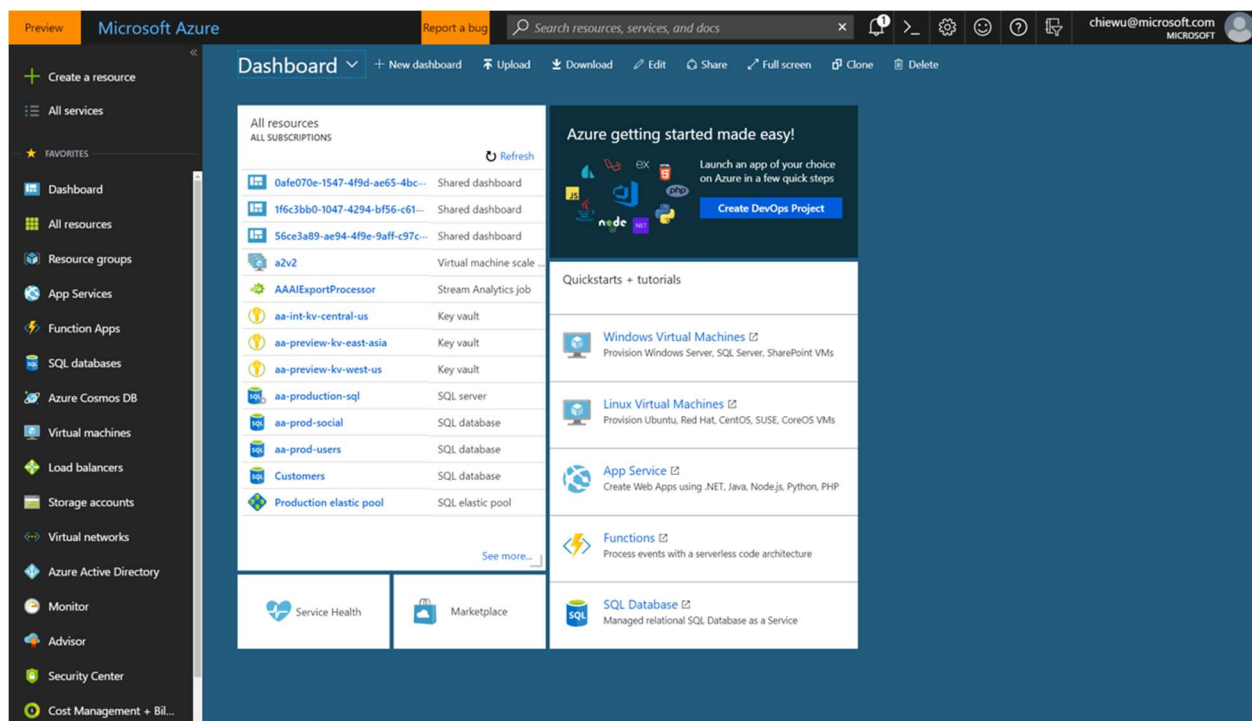
Hi, Chieh-Han.

We need a few more details to register you for the demo.

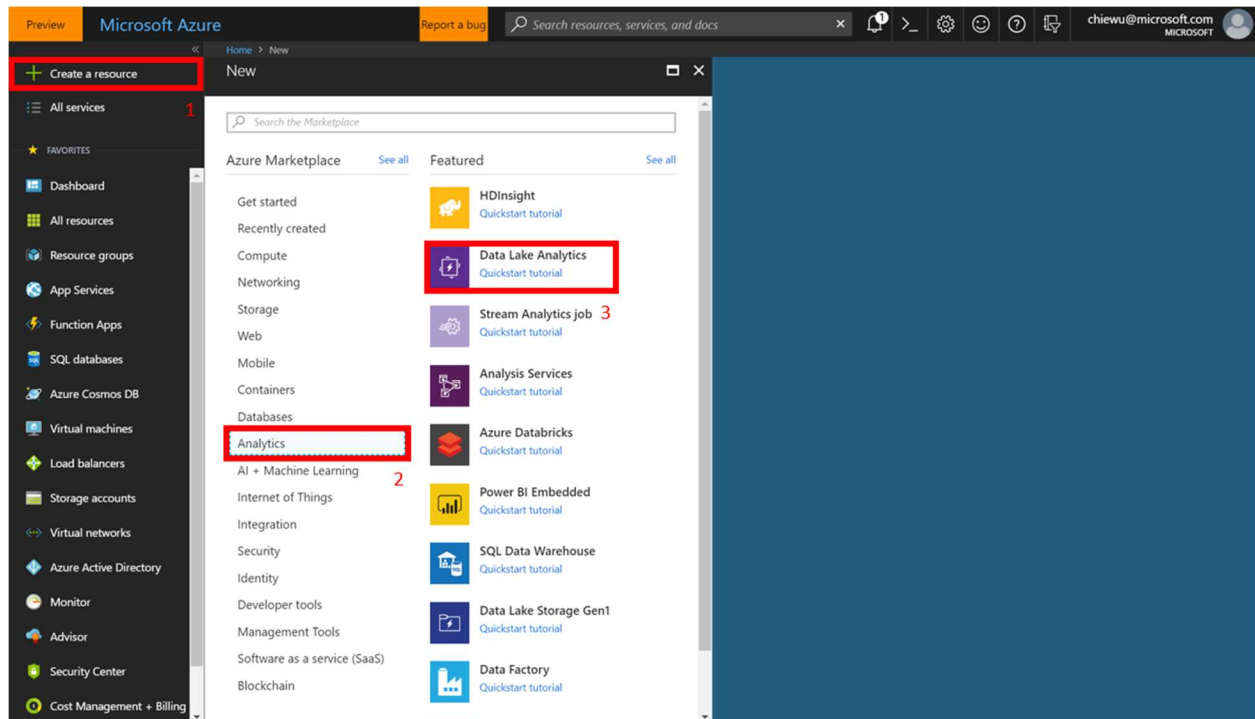
Country/region

Microsoft may use your contact information to provide und

[Chat live with an agent](#)

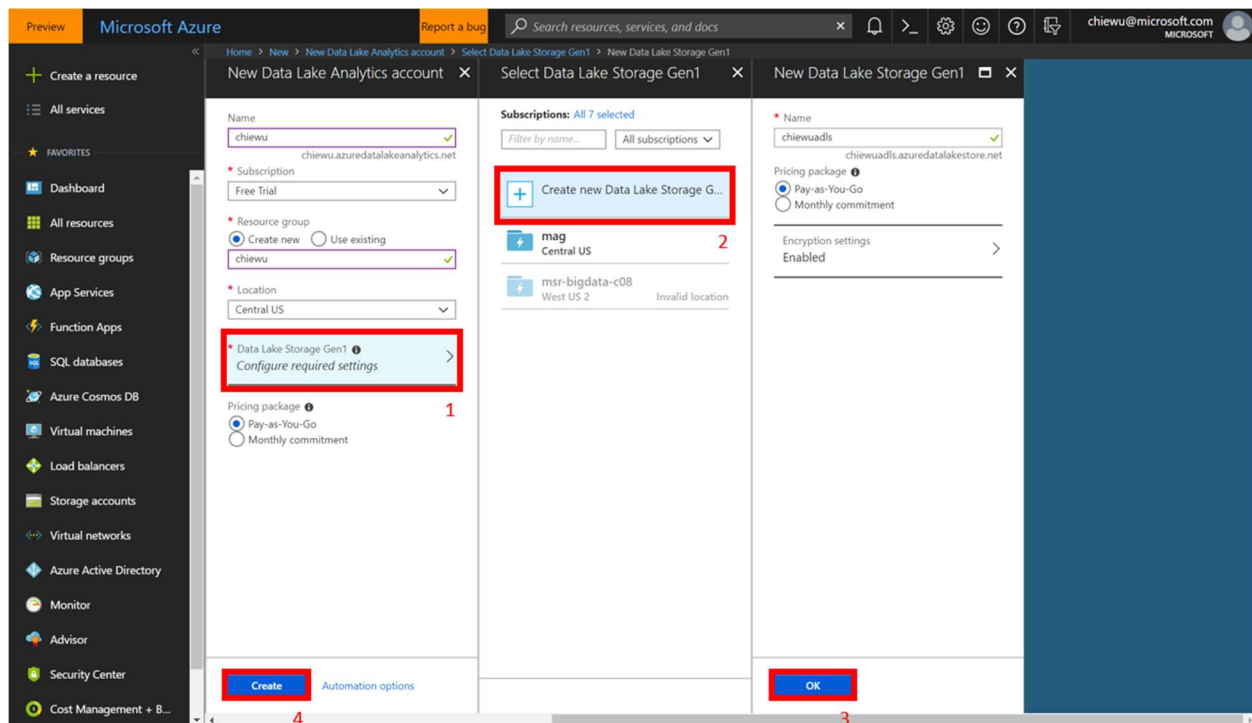


b. Click **Create a resource** > **Analytics** > **Data Lake Analytics**.

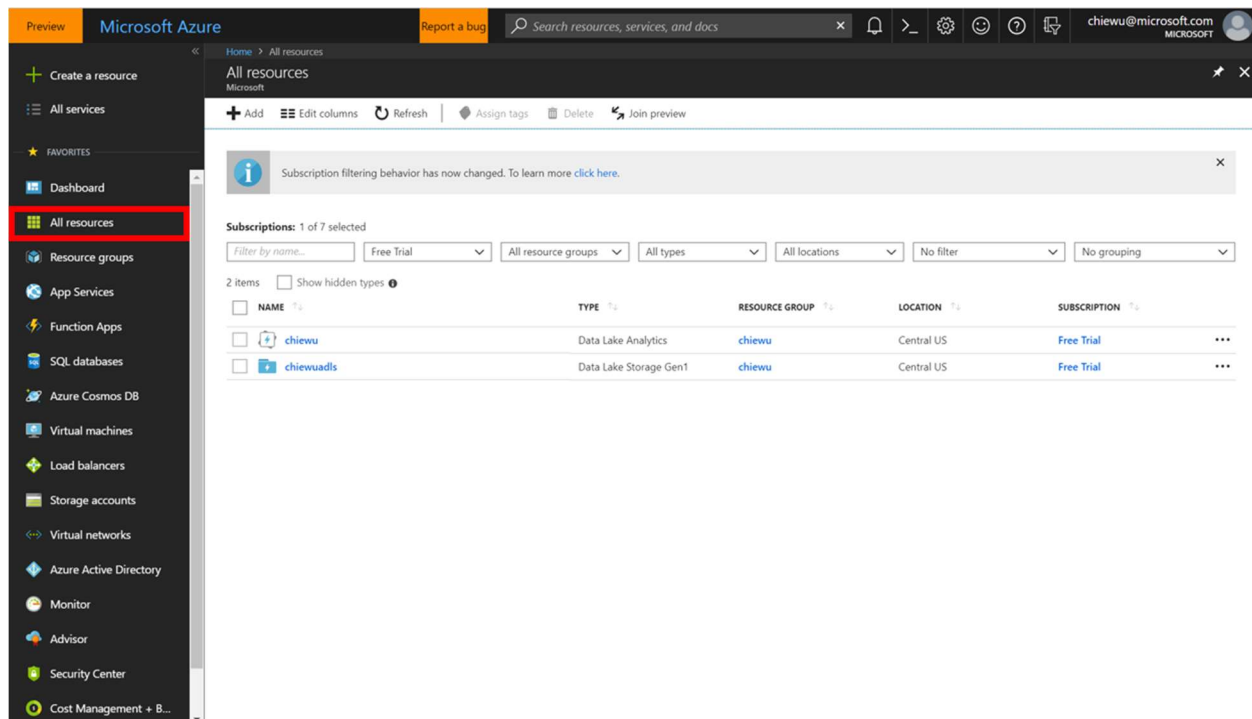


c. Select values for the following items:

- Name: Name your Data Lake Analytics account (Only lower case letters and numbers allowed).
- Subscription: Choose the Azure subscription used for the Analytics account.
- Resource Group: Select an existing Azure Resource Group or create a new one.
- Location: Select an Azure data center for the Data Lake Analytics account.
- Data Lake Store: Follow the instruction to create a new Data Lake Store account, or select an existing one.
- Optionally, select a pricing tier for your Data Lake Analytics account.



d. View your Data Lake Analytics and Data Lake Store Account in **All resources**.



3. Upload the Lab data to Azure Data Lake Store account
 - a. Go to Azure Data Lake Analytics account and click **Data explorer**.

The screenshot shows the Microsoft Azure portal interface for the 'chiewu' Data Lake Analytics account. The left sidebar contains navigation options like 'Create a resource', 'All services', and 'FAVORITES'. The top navigation bar includes 'Home > All resources > chiewu'. The 'Data explorer' button is highlighted in the top bar. The main content area shows account details: 'chiewu', 'Pay-as-You-Go', 'September 2018', and 'Pricing tier info'. It also displays 'Jobs and AU-hours consumed' with a bar chart and a table of job status counts.

Job Status	Count
Running	0
Queued	0
Pipelines	0
Submitters	0

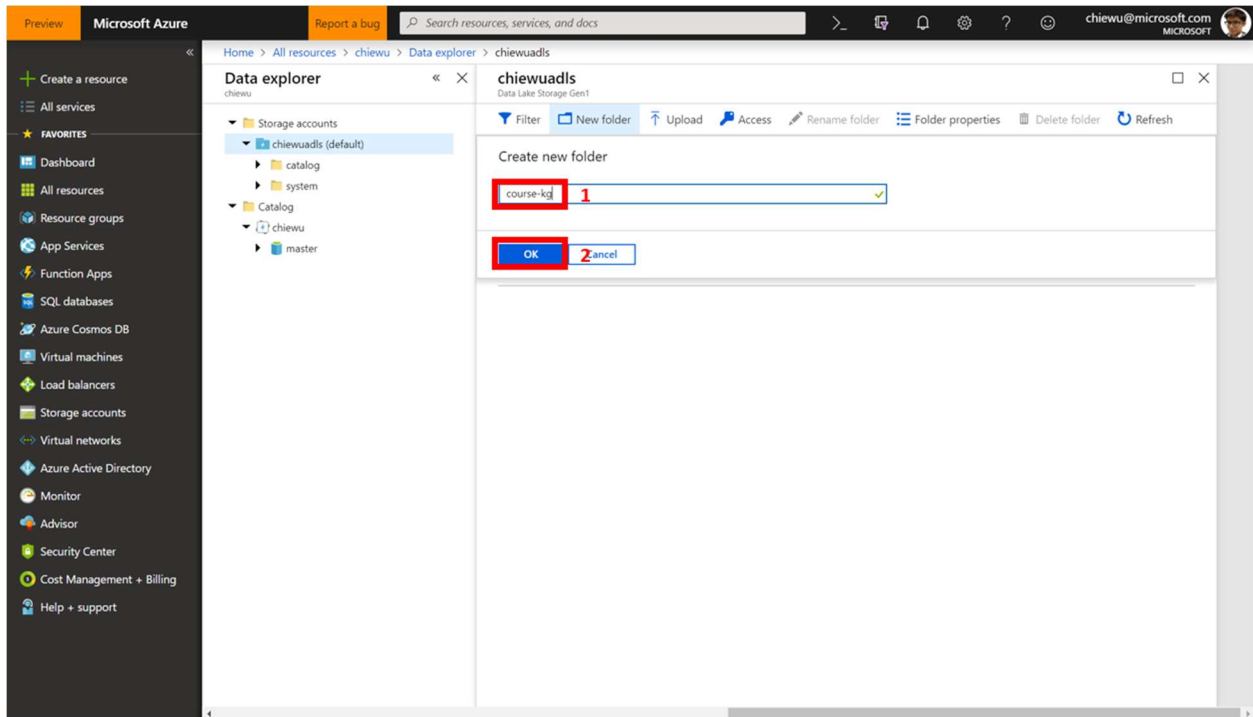
AU Status	Count
0 in use	16 available
My usage: 0	Other usage: 0
	16 total

- b. Click **New folder** to create a new folder under Azure Data Lake Store account.

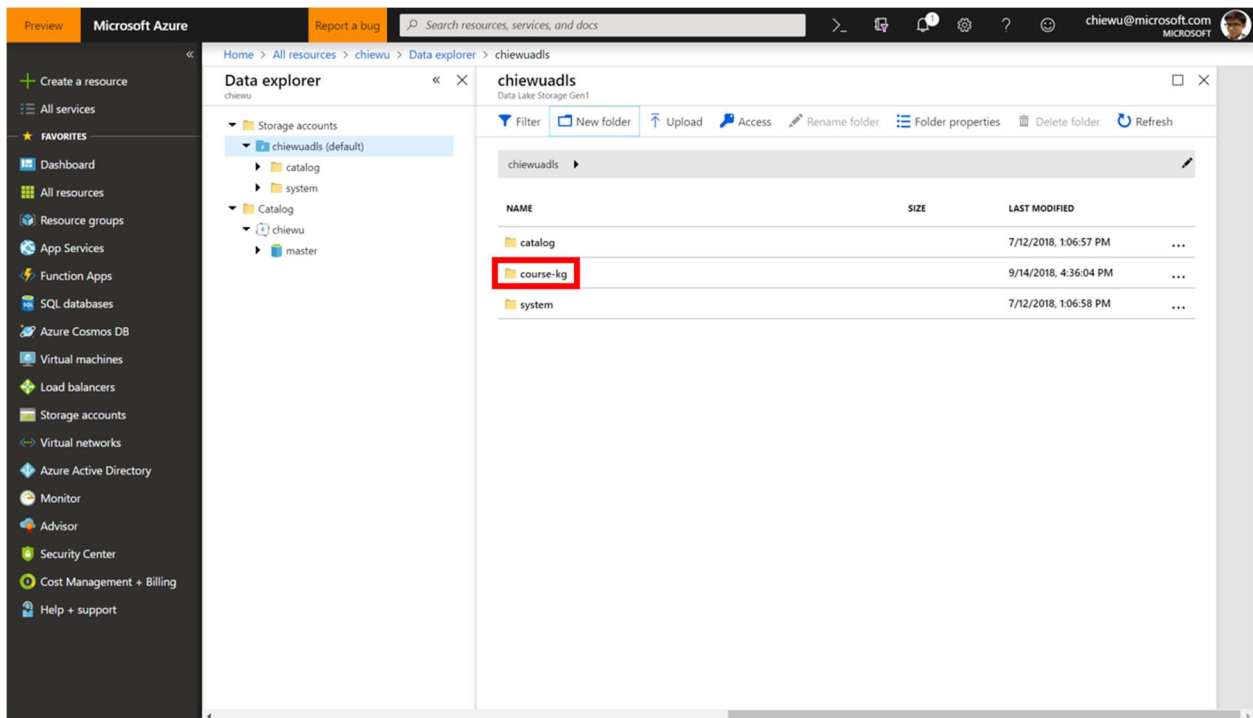
The screenshot shows the Microsoft Azure portal interface for the 'chiewuadls' Data Lake Storage Gen1 account. The left sidebar contains navigation options like 'Create a resource', 'All services', and 'FAVORITES'. The top navigation bar includes 'Home > All resources > chiewu > Data explorer > chiewuadls'. The 'New folder' button is highlighted in the top bar. The main content area shows a list of folders: 'catalog' and 'system'.

NAME	SIZE	LAST MODIFIED
catalog		7/12/2018, 1:06:57 PM
system		7/12/2018, 1:06:58 PM

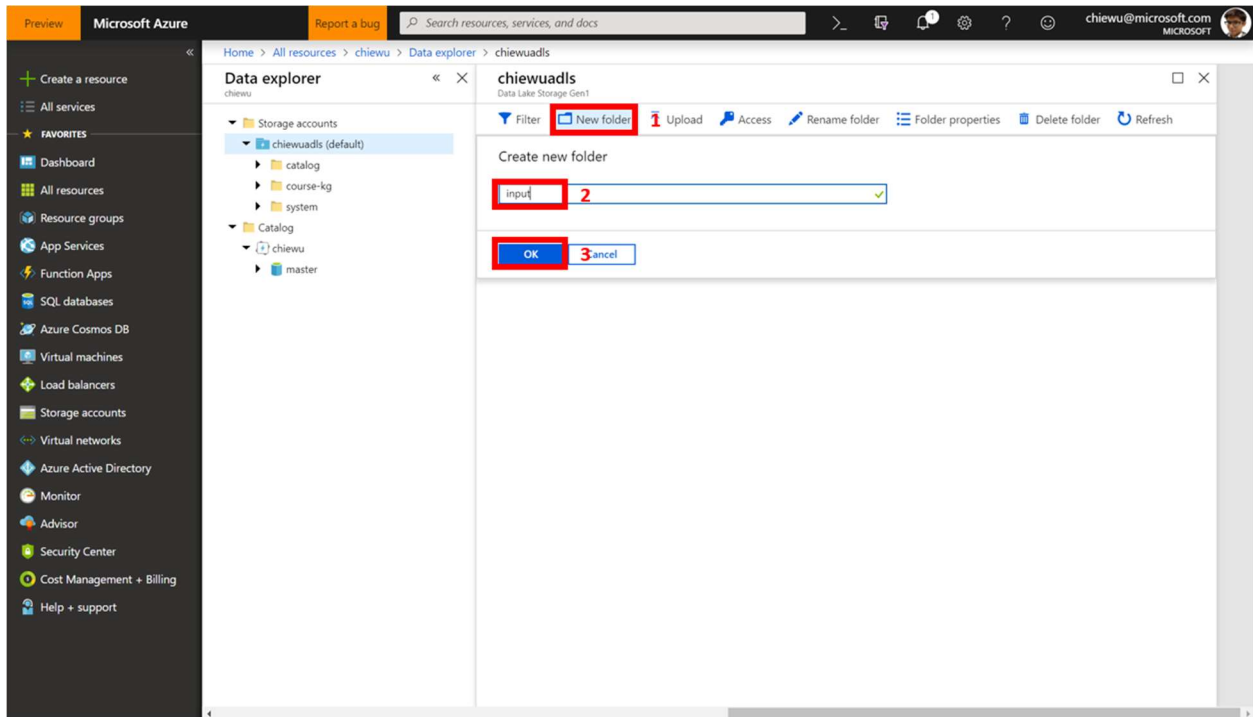
c. Use “**course-kg**” as the folder name and click **OK**.



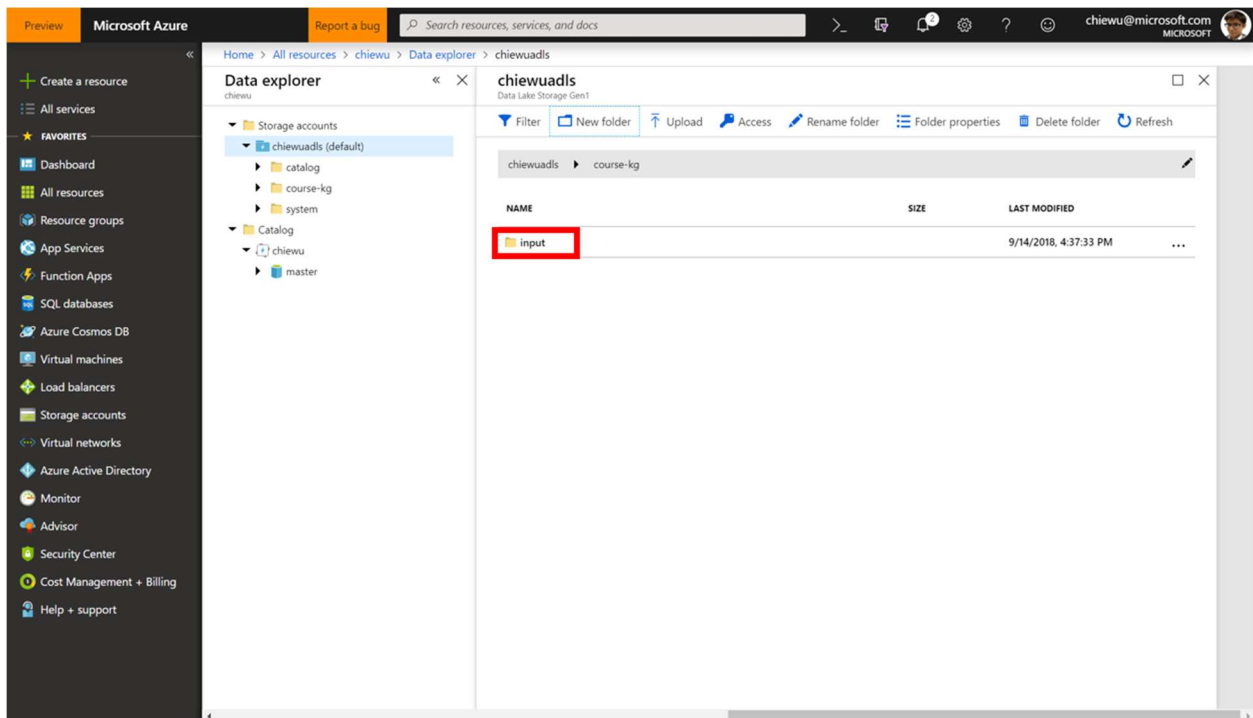
d. Click **course-kg** to go into this folder.



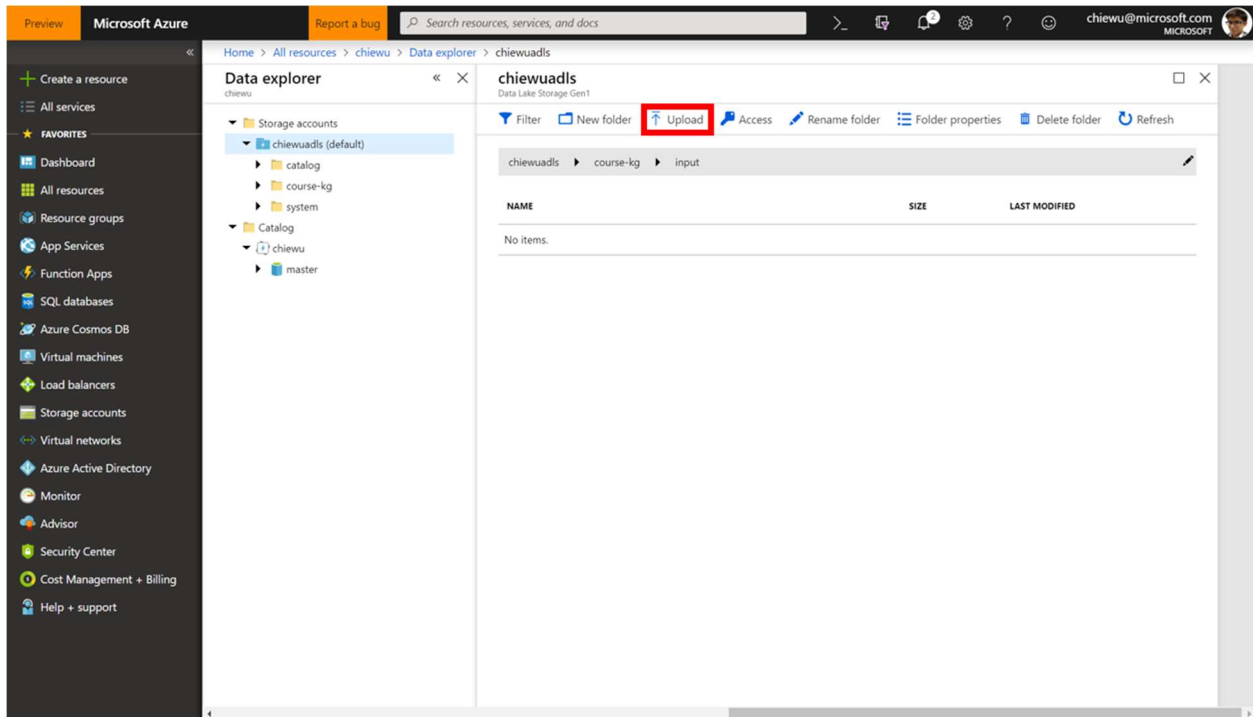
e. Create a new folder “input” under “course-kg” folder.



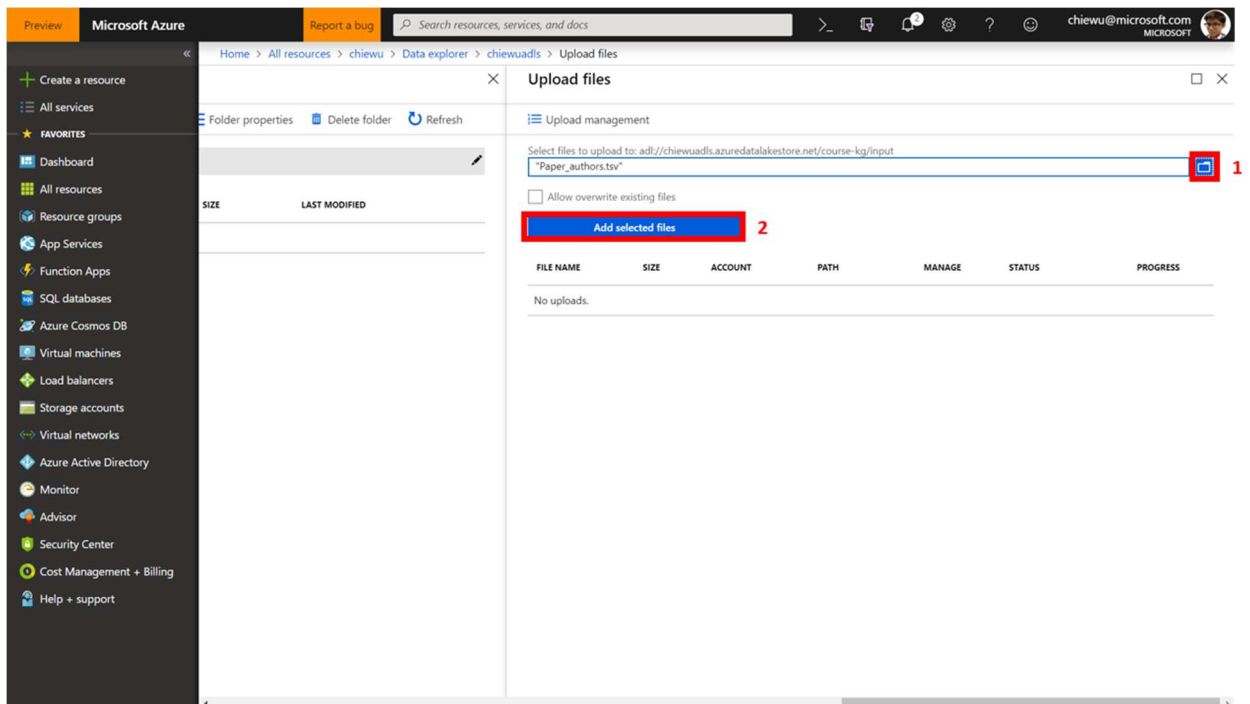
f. Click input to go into this folder.



g. Click **Upload**.



h. Upload **"Paper_authors.tsv"** to the **"course-kq/input"** folder.



- i. Upload “Paper_venue.tsv” to the “course-kg/input” folder.

Microsoft Azure portal interface showing the 'Upload files' dialog. The dialog is open for the 'chiewuadls' storage account. The 'Select files to upload to' field is set to 'adl://chiewuadls.azuredatastore.net/course-kg/input'. A red box labeled '1' highlights the 'Add selected files' button. Another red box labeled '2' highlights the 'Add selected files' button. The 'Upload management' table shows the upload progress for 'Paper_authors.tsv' (59.38 MB) and 'Paper_venue.tsv' (14.8 MB).

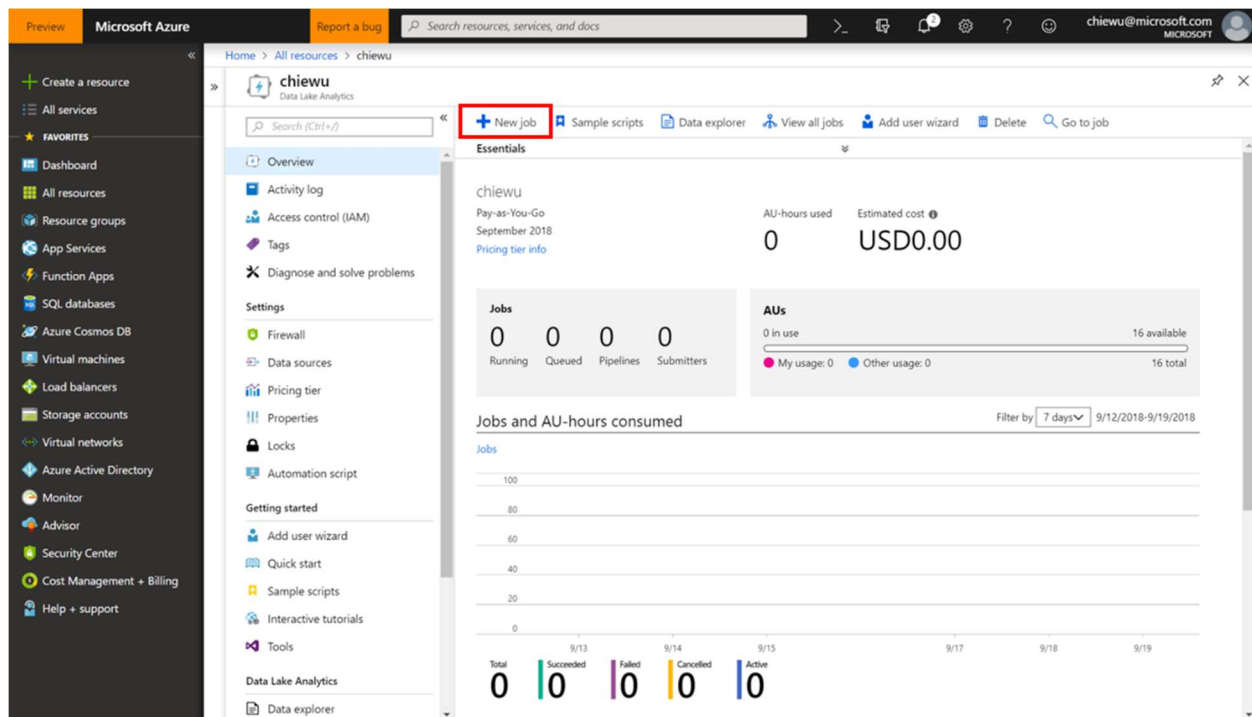
FILE NAME	SIZE	ACCOUNT	PATH	MANAGE	STATUS	PROGRESS
Paper_authors.tsv	59.38 MB	chiewuadls	/course-kg/input		Completed	100%

- j. View uploaded files under “course-kg/input” folder.

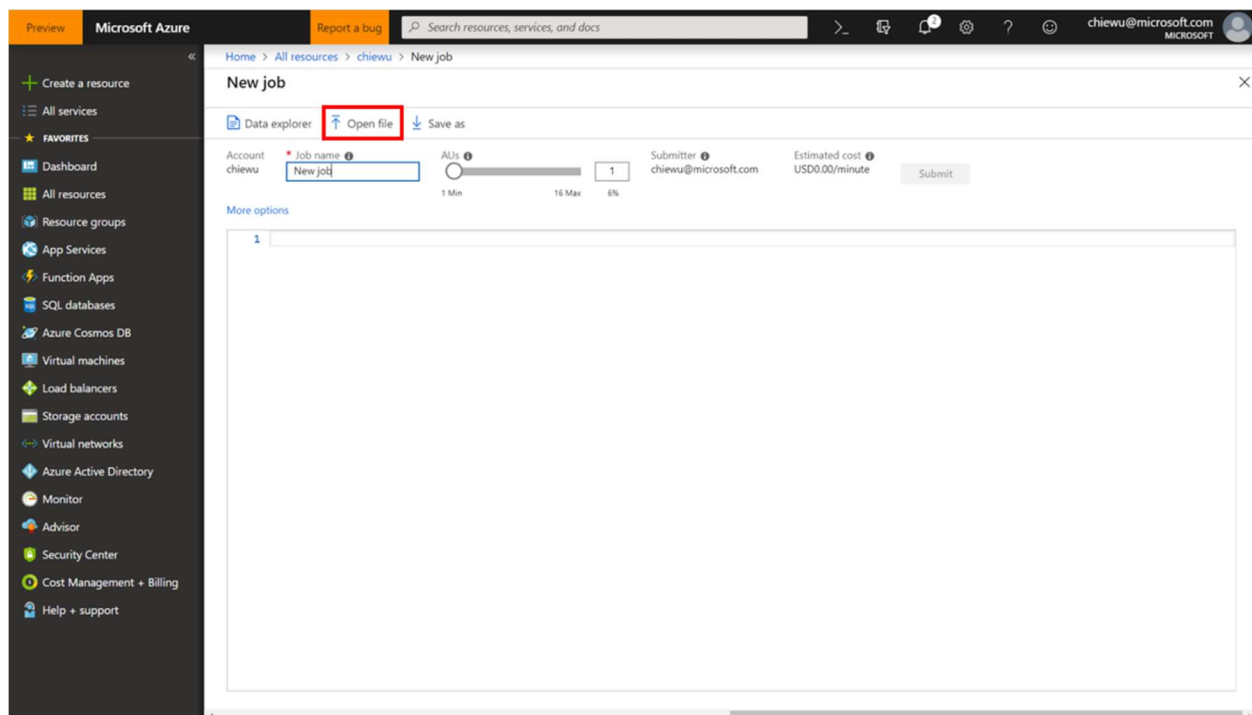
Microsoft Azure portal interface showing the 'Data explorer' view for the 'chiewuadls' storage account. The 'Data explorer' view shows the hierarchy of folders: 'chiewuadls (default)' > 'catalog' > 'course-kg' > 'input'. The 'input' folder is selected, and the 'Data explorer' table shows the files 'Paper_authors.tsv' (59.4 MB) and 'Paper_venue.tsv' (14.8 MB). A red box highlights the 'Paper_venue.tsv' file.

NAME	SIZE	LAST MODIFIED
Paper_authors.tsv	59.4 MB	9/14/2018, 4:40:38 PM
Paper_venue.tsv	14.8 MB	9/14/2018, 4:41:55 PM

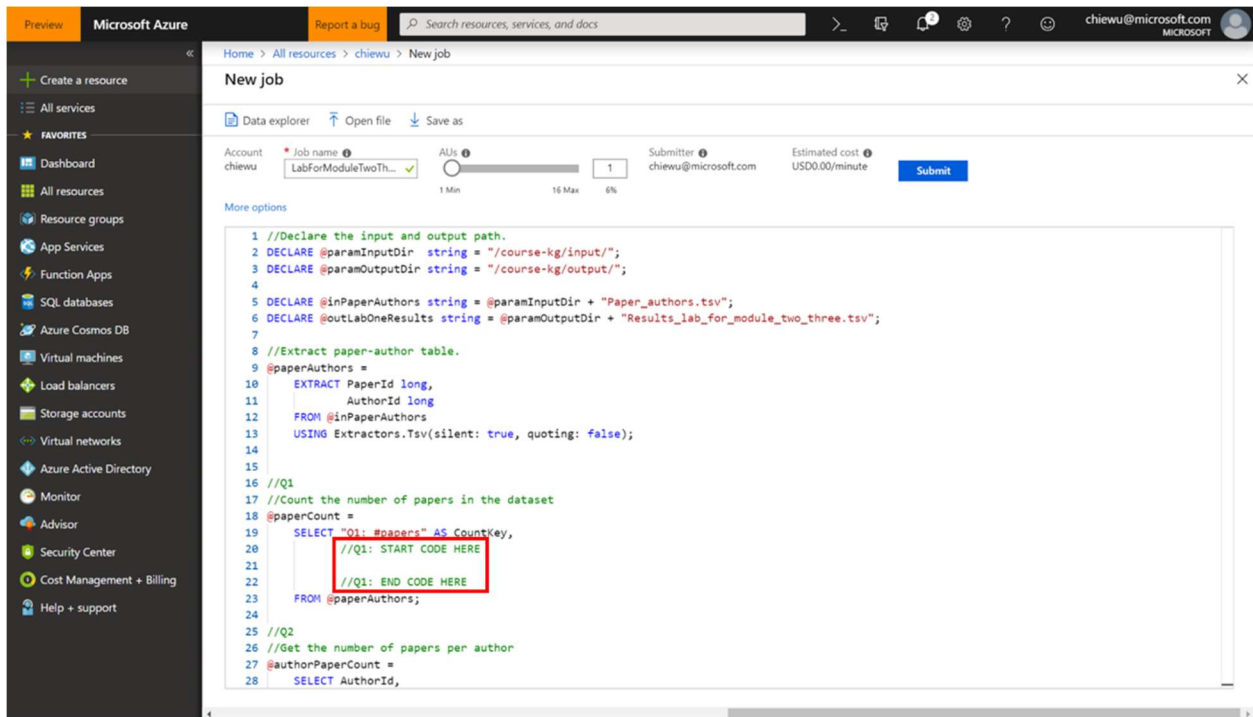
4. Upload the Lab data to Azure Data Lake Store account
- a. Go to Azure Data Lake Analytics account and click **New job** on the top menu.



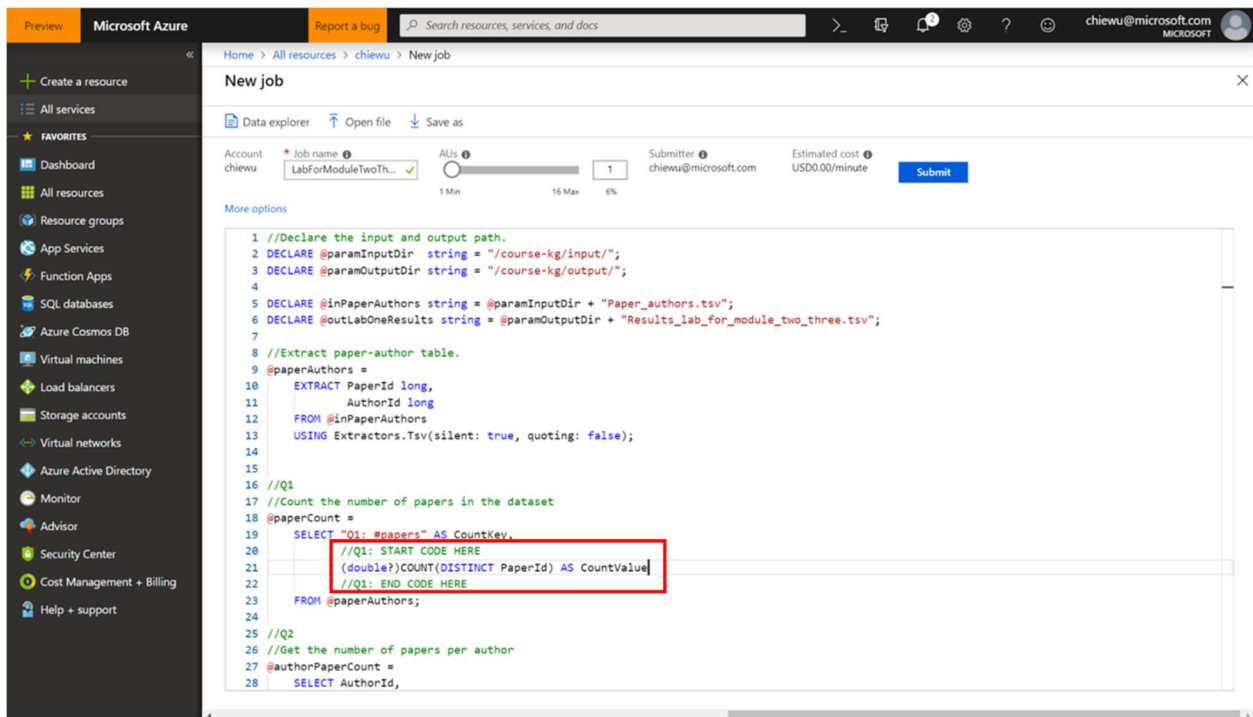
- b. Click **Open file** to open the Lab script written in U-SQL, for example, the first lab's script "DAT278x_Lab1_Graph.usql".



- c. Fill in the missing line of code for each question (see details in Lab instructions).



d. For the first question Q1 in the first lab, the correct code is filled below.



e. After completing all the missing code for the lab, click **Submit** to submit the job. The computing cost will be covered by [your Azure credit](#) that you received when you registered for the Azure free account.

Microsoft Azure | Report a bug | Search resources, services, and docs | chiewu@microsoft.com

Home > All resources > chiewu > New job

New job

Data explorer | Open file | Save as

Account: chiewu | Job name: LabForModuleTwoTh... | ALU: 1 (1 Min to 16 Max, 6%) | Submitter: chiewu@microsoft.com | Estimated cost: USD0.00/minute | Submit

More options

```
1 //Declare the input and output path.
2 DECLARE @paramInputDir string = "/course-kg/input/";
3 DECLARE @paramOutputDir string = "/course-kg/output/";
4
5 DECLARE @inPaperAuthors string = @paramInputDir + "Paper_authors.tsv";
6 DECLARE @outLabOneResults string = @paramOutputDir + "Results_lab_for_module_two_three.tsv";
7
8 //Extract paper-author table.
9 @paperAuthors =
10     EXTRACT PaperId long,
11           AuthorId long
12     FROM @inPaperAuthors
13     USING Extractors.Tsv(silent: true, quoting: false);
14
15
16 //Q1
17 //Count the number of papers in the dataset
18 @paperCount =
19     SELECT "Q1: #papers" AS CountKey,
20           //Q1: START CODE HERE
21           (double?)COUNT(DISTINCT PaperId) AS CountValue
22           //Q1: END CODE HERE
23     FROM @paperAuthors;
24
25 //Q2
26 //Get the number of papers per author
27 @authorPaperCount =
28     SELECT AuthorId,
```

- f. After running the script successfully, the results will be output to the “/course-kg/output/” folder.