

Microsoft - DAT278x

## From Graph to Knowledge Graph

### – Algorithms and Applications

## Challenge Lab Environment Setup Guide

The challenge labs for DAT278x will run on Microsoft Azure Data Lake Analytics environment. You can sign-up with an Azure free account, upload the data and finish the challenge lab assignments.

Lab assignments are running on U-SQL – a data processing language that unifies a declarative SQL-like syntax with C# programming. U-SQL can be used to process both structured and unstructured data in big data environments.

Prior basic knowledge and experience with SQL, U-SQL, Azure Data Lake Storage (ADLS), and Azure Data Lake Analytics (ADLA) would be required to complete the challenge labs.

You can refer to below resources to fill the knowledge gaps if any:

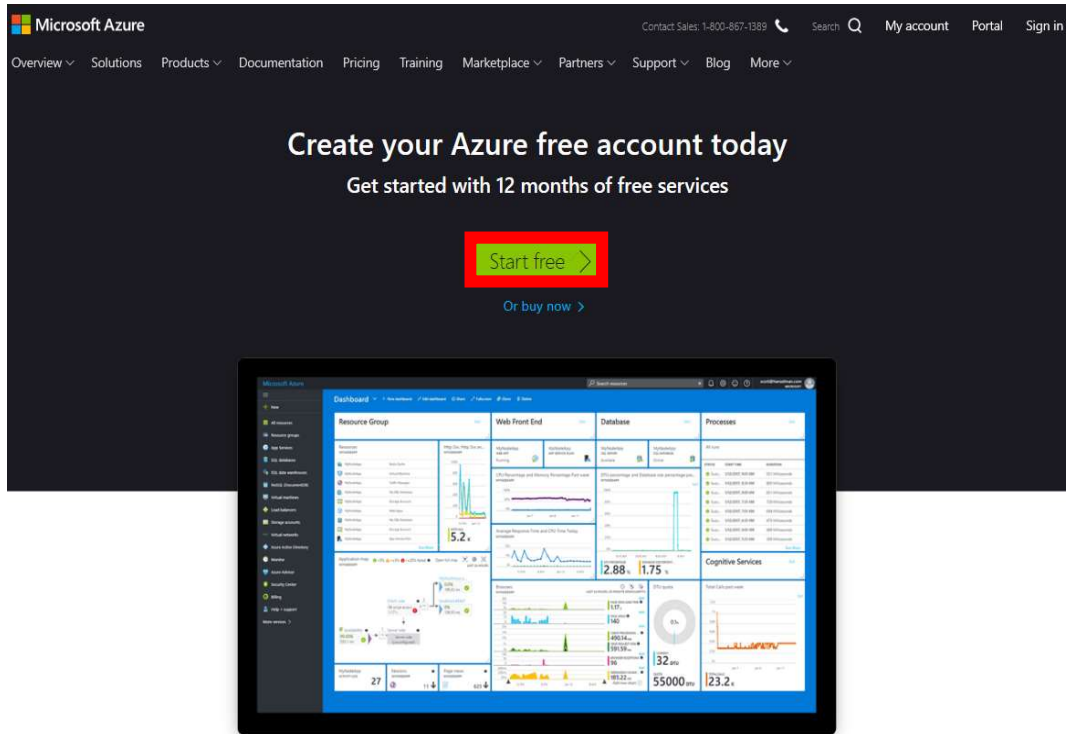
- EdX course on [Processing Big Data with Azure Data Lake Analytics](#)
- U-SQL resource site: [usql.io](https://usql.io)

In the remaining of this document, we include step-by-step instructions to:

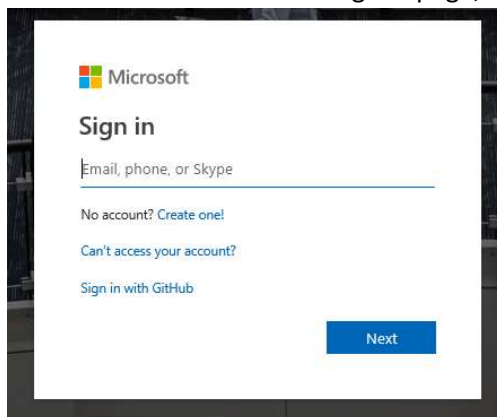
1. Create an Azure free account.
2. Create an Azure Data Lake Analytics (ADLA) resource and an Azure Data Lake Storage (ADLS) resource.
3. Create folder structure in Azure Data Lake Storage (ADLS) to prepare for smoke testing and challenge lab runs.
4. Upload data files to Azure Data Lake Storage (ADLS).
5. Run a smoke test U-SQL script to verify that the environment is setup properly.

## 1. Create a free Azure account:

1.1 Go to this page: <https://azure.microsoft.com/free> ; click on “start-free”.



1.2 Land on the Microsoft Sign in page, sign-in with your Microsoft Account.



1.3 Land on Microsoft Azure sign-up page, “about you” section, fill in your information and click “Next”.

The screenshot shows the Microsoft Azure sign-up page. At the top, there's a header with the Microsoft Azure logo, the email 'EdX.DAT278x.Staff@outlook.com', and a 'Sign out' link. Below the header is a blue banner with the text 'Azure free account sign up' and 'Start with a \$200 credit for 30 days, and keep going for free'. The main section is titled '1 About you' and contains several form fields: 'Country/Region' (dropdown menu showing 'United States'), 'First name' (text box with 'Staff'), 'Last name' (text box with 'EdX.DAT278x'), 'Email address' (text box with 'EdX.DAT278x.Staff@outlook.com'), and 'Phone' (text box with 'Example: (425) 555-0100'). Below these fields is a link to the 'privacy statement' and 'subscription agreement'. At the bottom of the form is a red 'Next' button.

1.4 Provide phone information and credit card information for identity verification purpose.  
(**PLEASE NOTE:** Your credit card will **NOT** be charged unless you upgrade your subscription.  
For EdX DAT278x course, the “free trial” credit would be enough to finish all labs.)

**For students only:** if you are a student with a valid school email address, you can sign up Azure without providing credit card information. For more info, please refer below links.  
<https://azure.microsoft.com/free/free-account-students-faq/>

The screenshot shows the Microsoft Azure sign-up page, specifically the 'Identity verification by phone' section. The header is the same as the previous screenshot. The main section is titled '2 Identity verification by phone' and contains a text box for 'Country code' (dropdown menu showing 'United States (+1)'), a text box for 'Phone number', and two buttons: 'Text me' and 'Call me'. Below these buttons is a text box for 'Verification code'. At the bottom of the form are two buttons: 'Verify code' and 'I did not receive a code'.

### 3 Identity verification by card ^

We ask for your credit card number to verify your identity and to keep out spam and bots.

**You won't be charged unless you upgrade.**



1.5 To accept the “Agreement” and click the “Sign up” to proceed.

Microsoft Azure EdX.DAT278x.Staff@outlook.com Sign out

## Azure free account sign up

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

- About you
- Identity verification by phone
- Identity verification by card
- Agreement**

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

I will receive information, tips, and offers from Microsoft or selected partners about Azure, including Azure Newsletter, Pricing updates, and other Microsoft products and services.

[Sign up](#)

1.6 You can see below page after successfully sign up with Azure account, click on “Go to the portal”

Microsoft Azure Contact Sales: 1-800-857-1889 Search My account Portal Sign in

Overview Solutions Products Documentation Pricing Training Marketplace Partners Support Blog More

## You're ready to start with Azure

[Go to the 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.

### Register for a live demo or watch on demand

Select a time or watch on demand

- ☒ 1/17/2019, 9:00 AM PST (Change time zone)
- ☐ 1/24/2019, 9:00 AM PST
- ☐ 1/31/2019, 9:00 AM PST
- ☐ Watch on demand

Sign in to complete registration

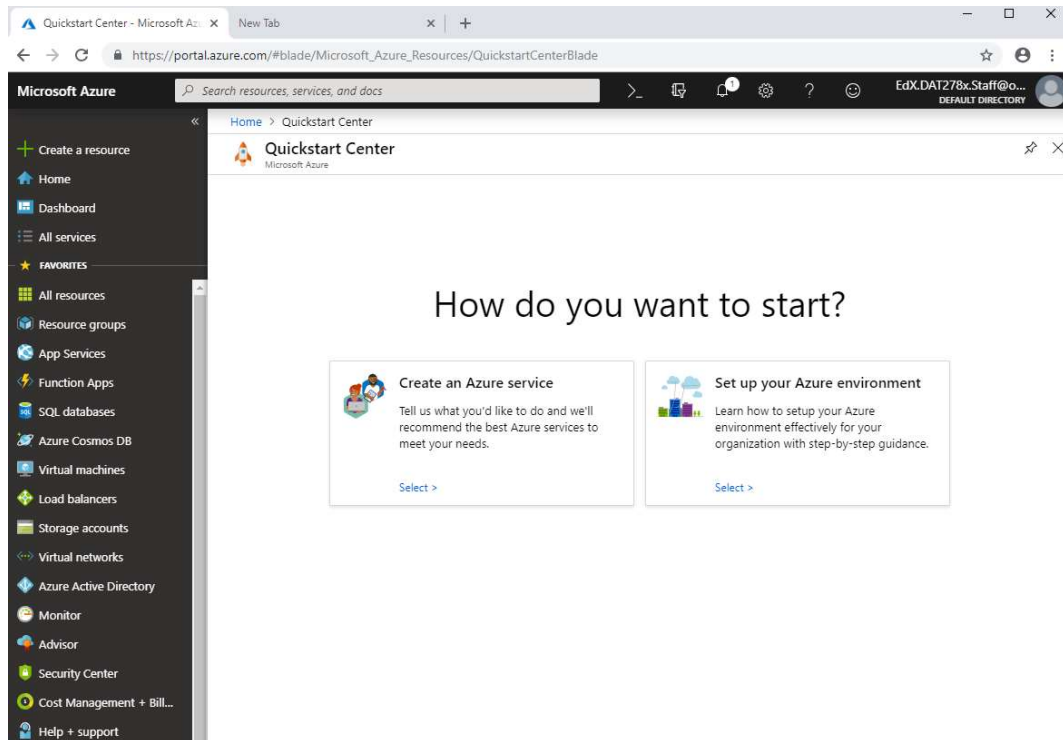
Microsoft

Or, please tell us a little about yourself to register for the demo.

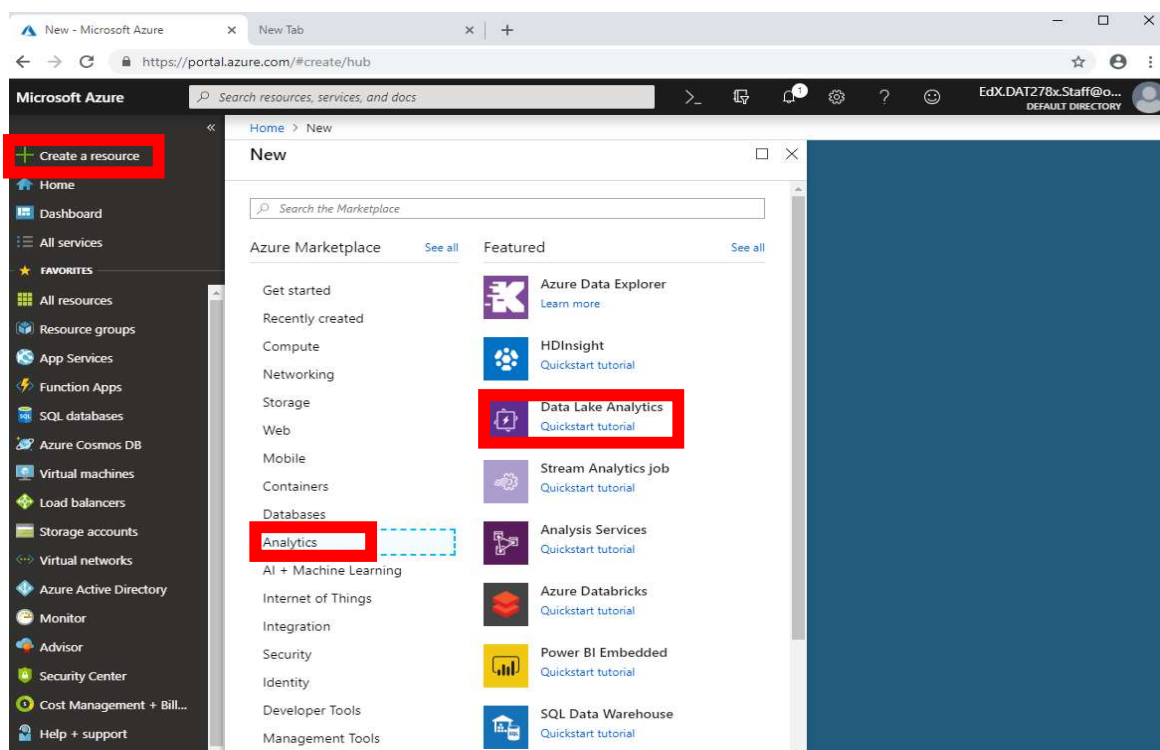
First name

## 2 On Azure Portal, create an Azure Data Lake Analytics (ADLA) resource and an Azure Data Lake Storage (ADLS) resource

2.1 Sign on to the Azure Portal to see below page:



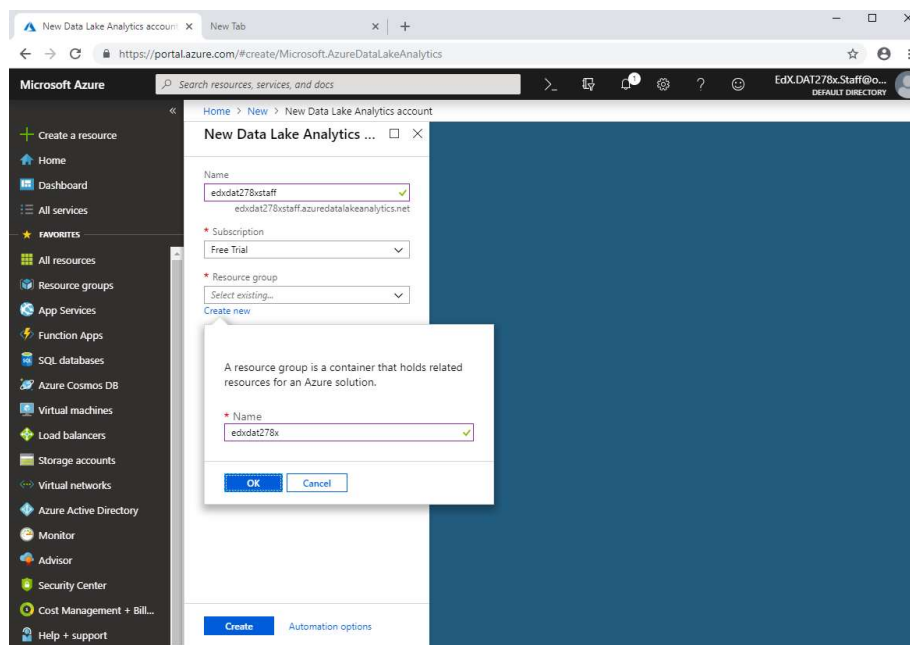
2.2 Click on “Create a resource” -> “Analytics” -> “Data Lake Analytics”



## 2.3 Fill in values for following items:

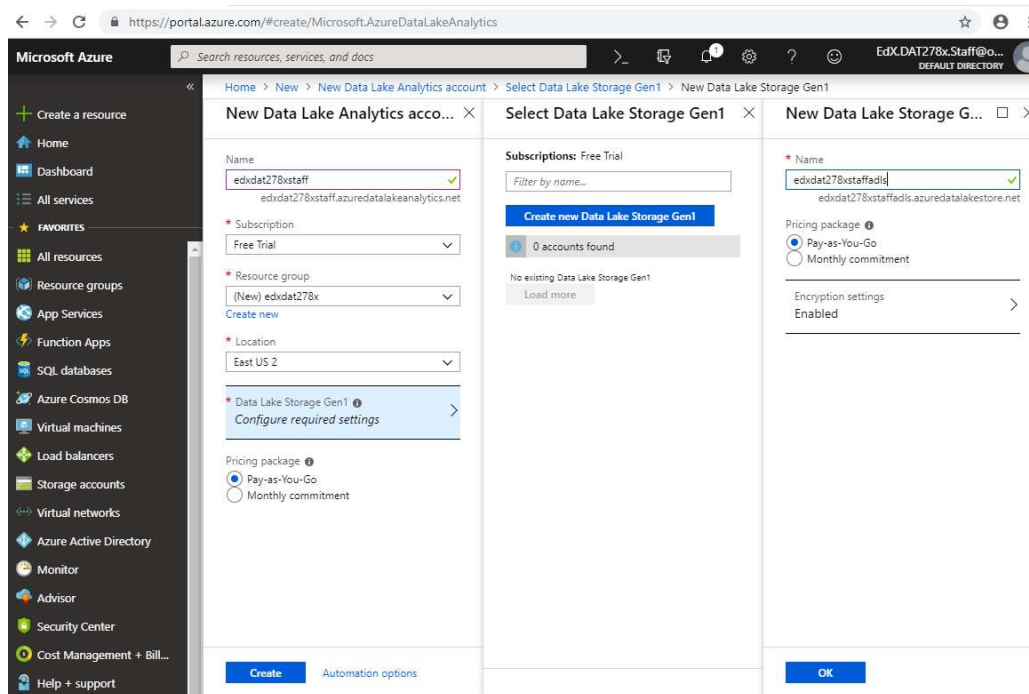
- Name:** Name your Data Lake Analytics account (Only lower-case letters and numbers allowed).
- Subscription:** Choose the Azure subscription (select “free trial” for DAT278x course unless you have other subscriptions) used for the Analytics account.
- Resource Group:** Select an existing Azure Resource Group or create a new one. You shall create a new one if this is the first time you use Azure or you want a separate Resource Group for DAT278x course.
- Location:** Select an Azure data center for the Data Lake Analytics account (using default value is fine).
- 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.

### 2.3.1 Pick a new name for Resource Group

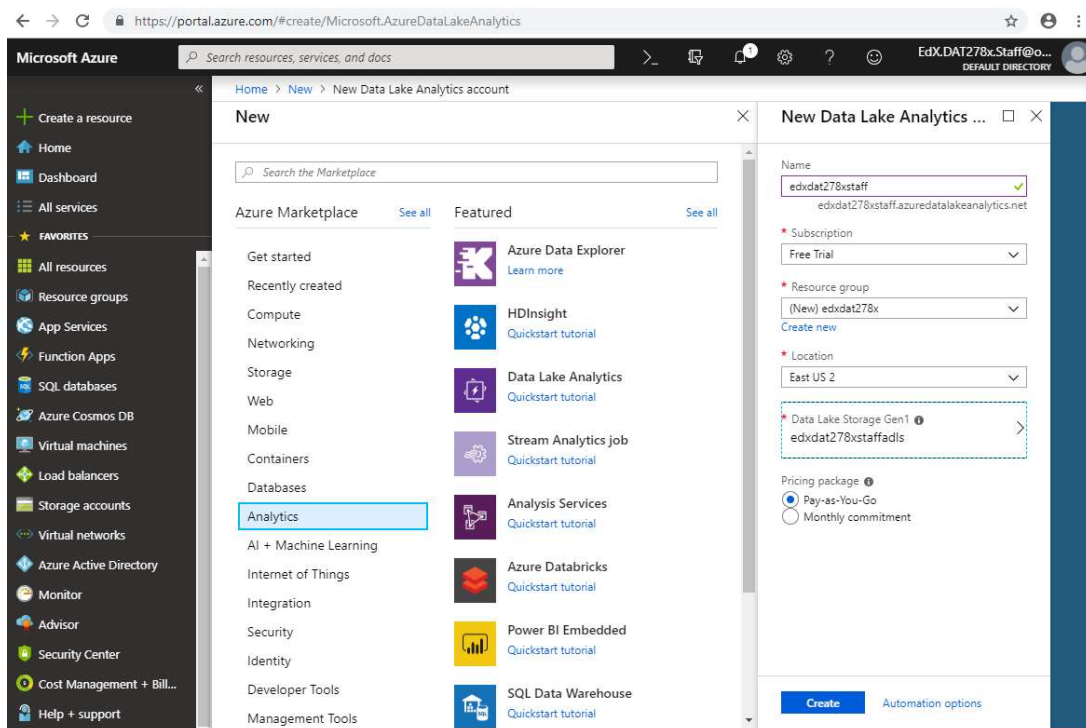


### 2.3.2 Create a new Azure Data Lake Storage (ADLS) account to associate with the Azure Data Lake Analytics (ADLA) account.

Use the default “Pay-as-You-Go” option for “free trial” subscription, pick a different package if you have other subscriptions.



2.3.3 Validate all the fill-in values are correct and then click “Create”.



## 2.4 View the created Data Lake Analytics and Data Lake Storage resources in “All resources”.

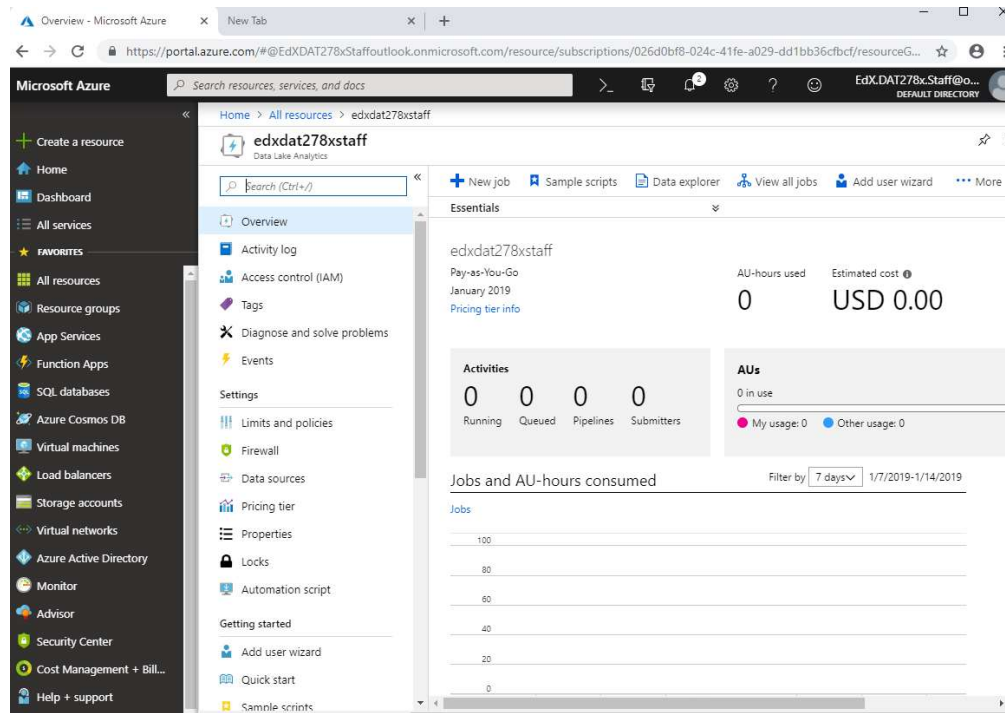
The screenshot shows the Microsoft Azure portal interface. The left-hand navigation pane is visible, with the 'All resources' link highlighted by a red rectangular box. The main content area displays the 'All resources' view, showing a table of resources. The table has columns for NAME, TYPE, RESOURCE GROUP, LOCATION, and SUBSCRIPTION. Two resources are listed:

NAME	TYPE	RESOURCE GROUP	LOCATION	SUBSCRIPTION
edxdat278xstaff	Data Lake Analytics	edxdat278x	East US 2	Free Trial
edxdat278xstaffadls	Data Lake Storage G...	edxdat278x	East US 2	Free Trial

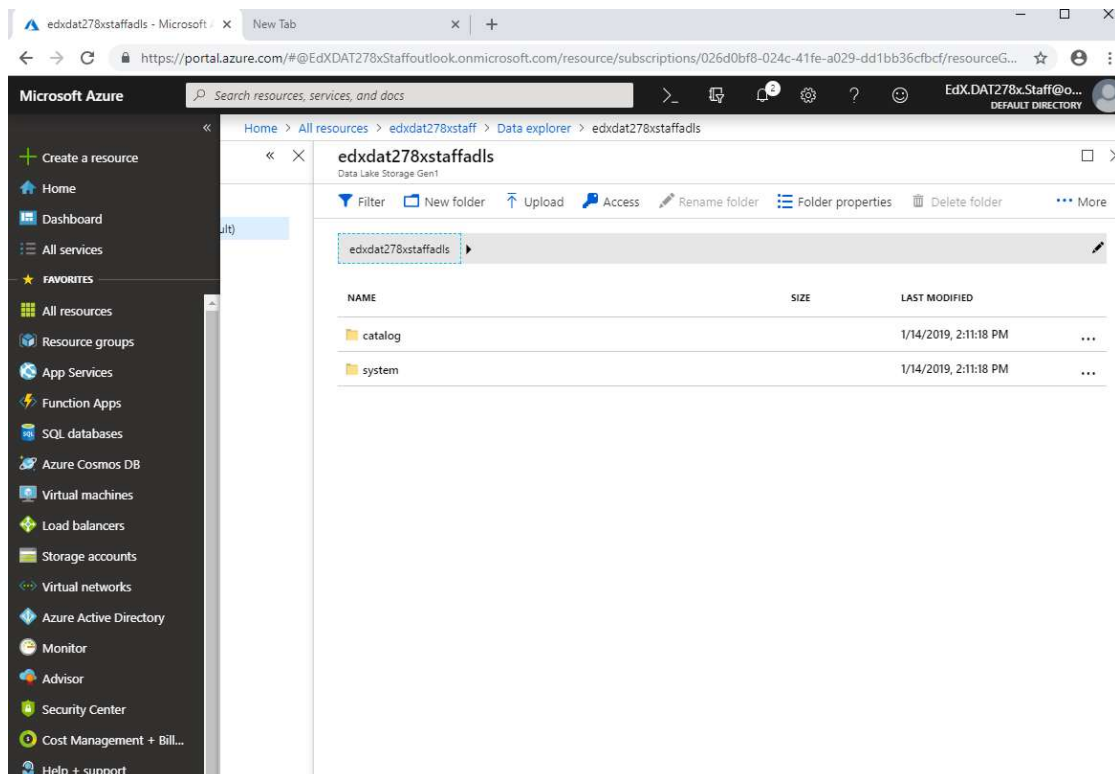


### 3 Create folder structure in Azure Data Lake Storage (ADLS) to prepare for challenge lab runs.

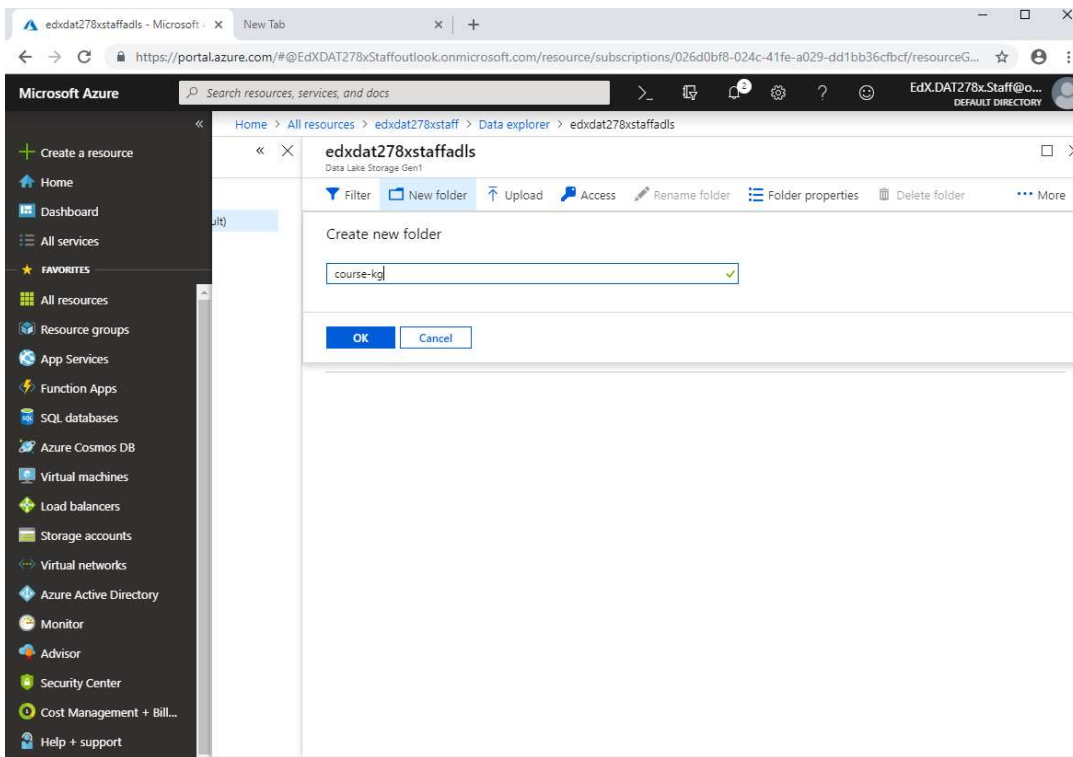
3.1 Go to Azure Data Lake Analytics account and click **Data explorer**.



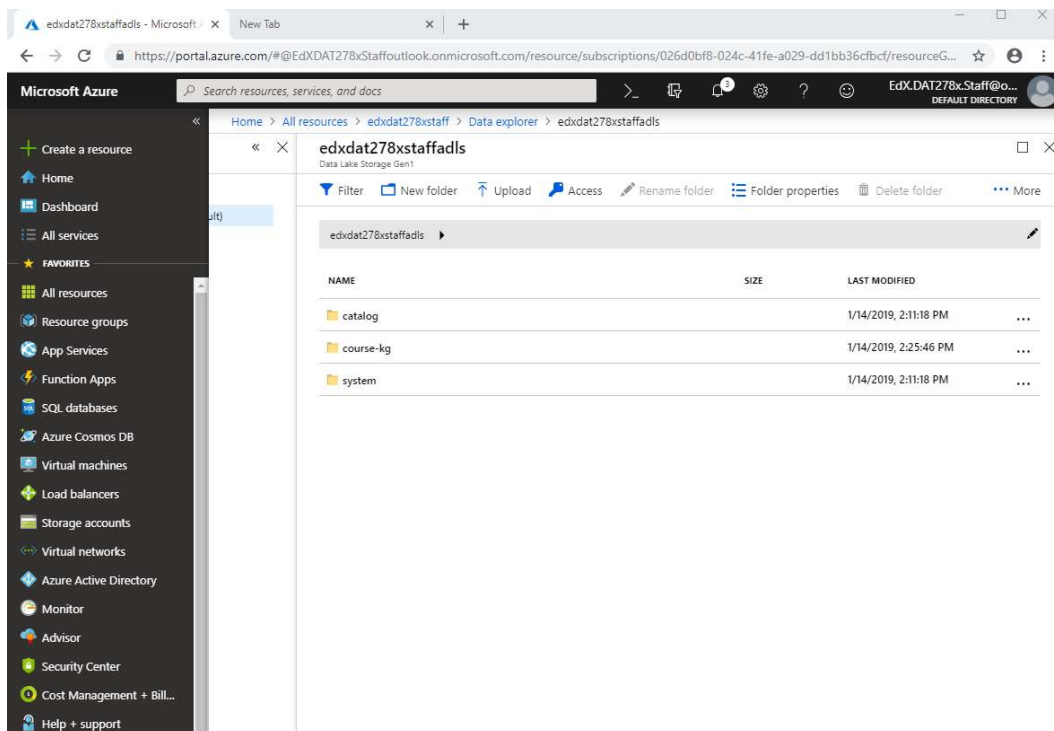
3.2 Click **New folder** to create a new folder under Azure Data Lake Storage account.

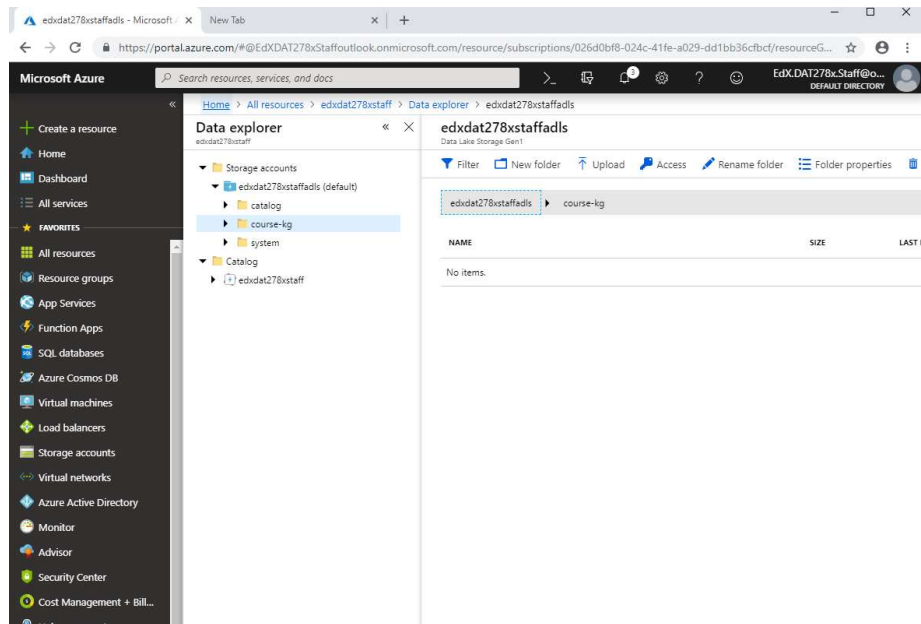


3.3 Use “course-kg” as the folder name and click **OK**.

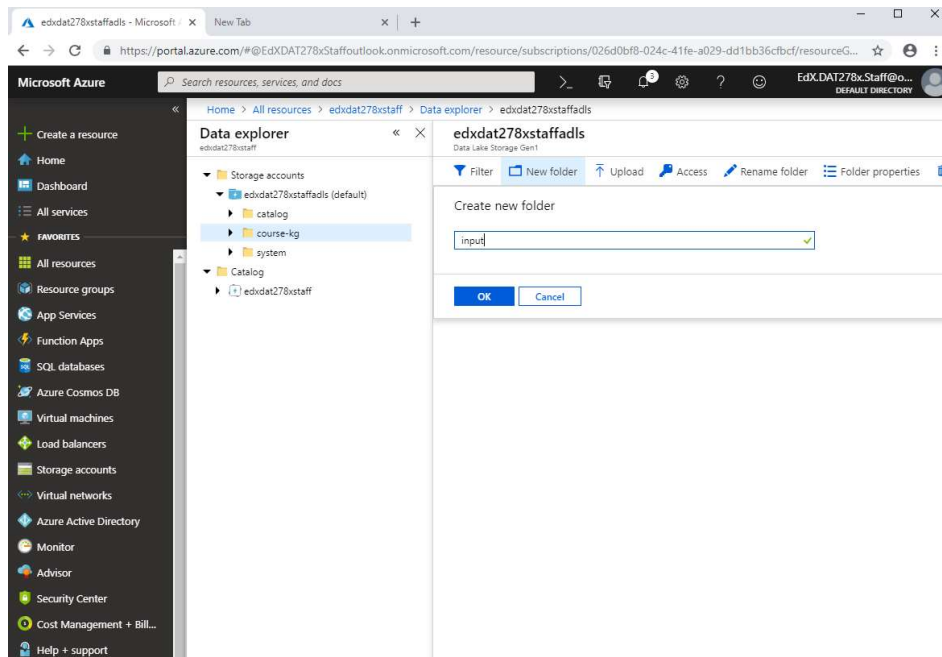


3.4 Click **course-kg** to go into this folder





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



3.6 Click **input** to go into this folder.

Microsoft Azure portal interface showing the Data Explorer for the storage account **edxdatt278xstaffadls**.

**Left Navigation Panel:**

- Create a resource
- Home
- Dashboard
- All services
- FAVORITES
- All resources
- Resource groups
- App Services
- Function Apps
- SQL databases
- Azure Cosmos DB
- Virtual machines
- Load balancers
- Storage accounts
- Virtual networks
- Azure Active Directory
- Monitor
- Advisor
- Security Center
- Cost Management + Bill...
- Help + support

**Breadcrumbs:** Home > All resources > edxdatt278xstaff > Data explorer > edxdatt278xstaffadls

**Data Explorer (Left Pane):**

- Storage accounts
  - edxdatt278xstaffadls (default)
    - catalog
      - course-kg
        - input
    - system
  - Catalog
    - edxdatt278xstaff

**edxdatt278xstaffadls (Right Pane):**

Data Lake Storage Gen1

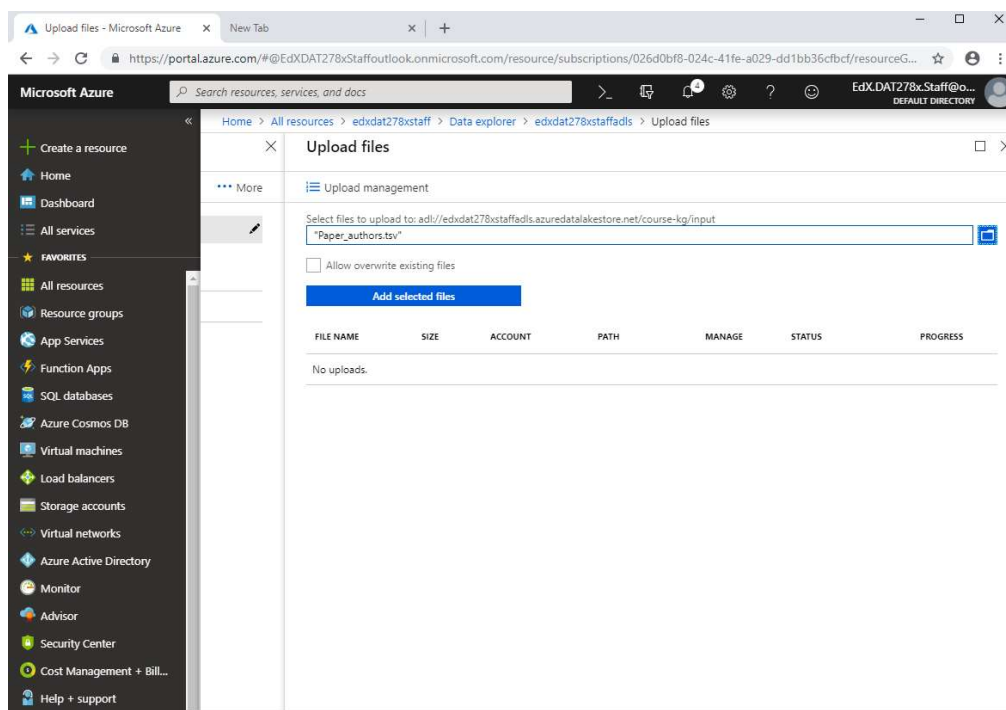
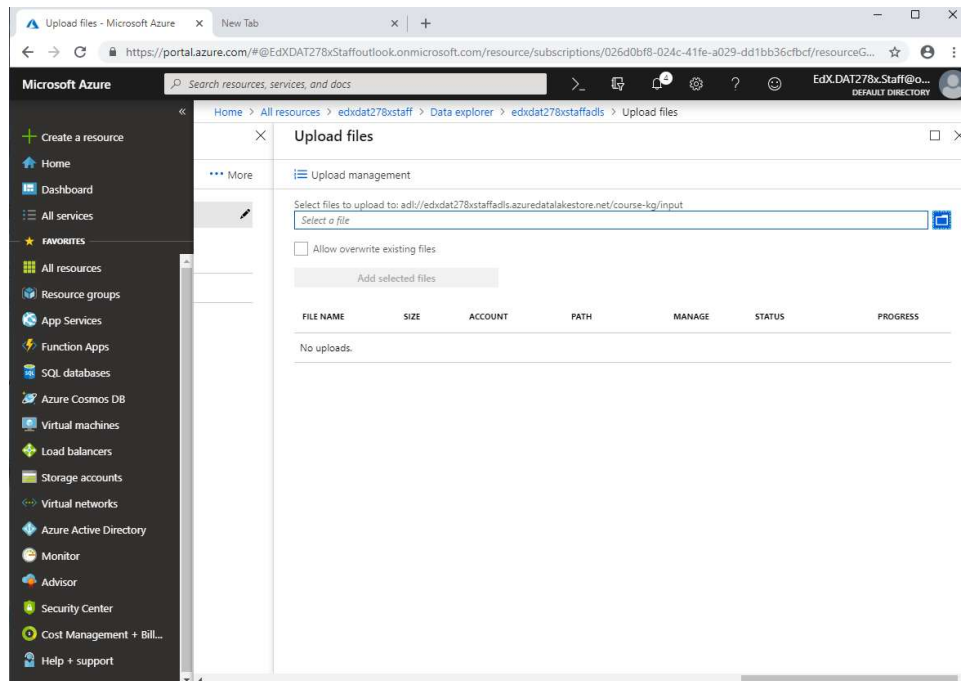
Filter | New folder | Upload | Access | Rename folder | Folder properties

edxdatt278xstaffadls > course-kg > input

NAME	SIZE	LAST MODIFIED
No items.		

## 4 Upload data files to Azure Data Lake Storage (ADLS) to prepare for smoke testing and Challenge Lab runs:

### 4.1 Upload Paper\_authors.tsv and Paper\_venue.tsv to the created “/course-kq/input/”.



Microsoft Azure

Home > All resources > edxd278xstaff > Data explorer > edxd278xstaffadls > Upload files

Upload files

Upload management

Select files to upload to: adl://edxd278xstaffadls.azuredatalakestore.net/course-kg/input

"Paper\_authors.tsv"

☐ Allow overwrite existing files

Add selected files

FILE NAME	SIZE	ACCOUNT	PATH	MANAGE	STATUS	PROGRESS
Paper_authors.tsv	59.38 MiB	edxd278xstaffadls	/course-kg/input		Completed	100%

#### Upload management

Select files to upload to: adl://edxd278xstaffadls.azuredatalakestore.net/course-kg/input

"Paper\_venue.tsv"

☐ Allow overwrite existing files

Add selected files

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

Microsoft Azure

Home > All resources > edxd278xstaff > Data explorer > edxd278xstaffadls > Upload files

Upload files

File upload completed 2:41 PM  
Successfully uploaded Paper\_venue.tsv to edxd278xstaffadls.

Upload management

Select files to upload to: adl://edxd278xstaffadls.azuredatalakestore.net/course-kg/input

"Paper\_venue.tsv"

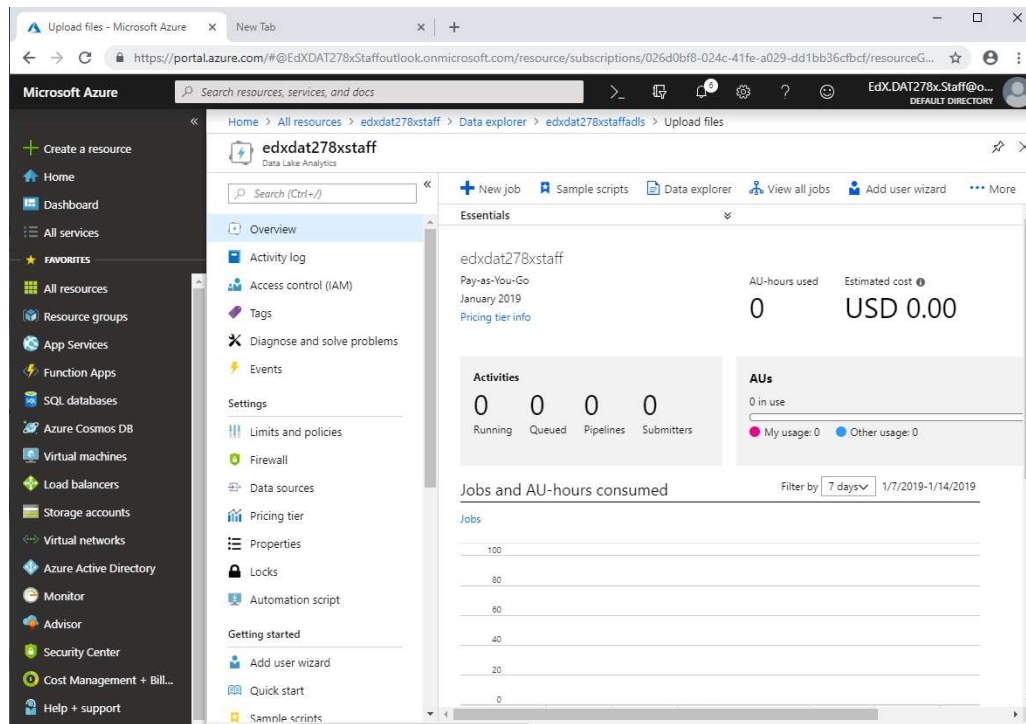
☐ Allow overwrite existing files

Add selected files

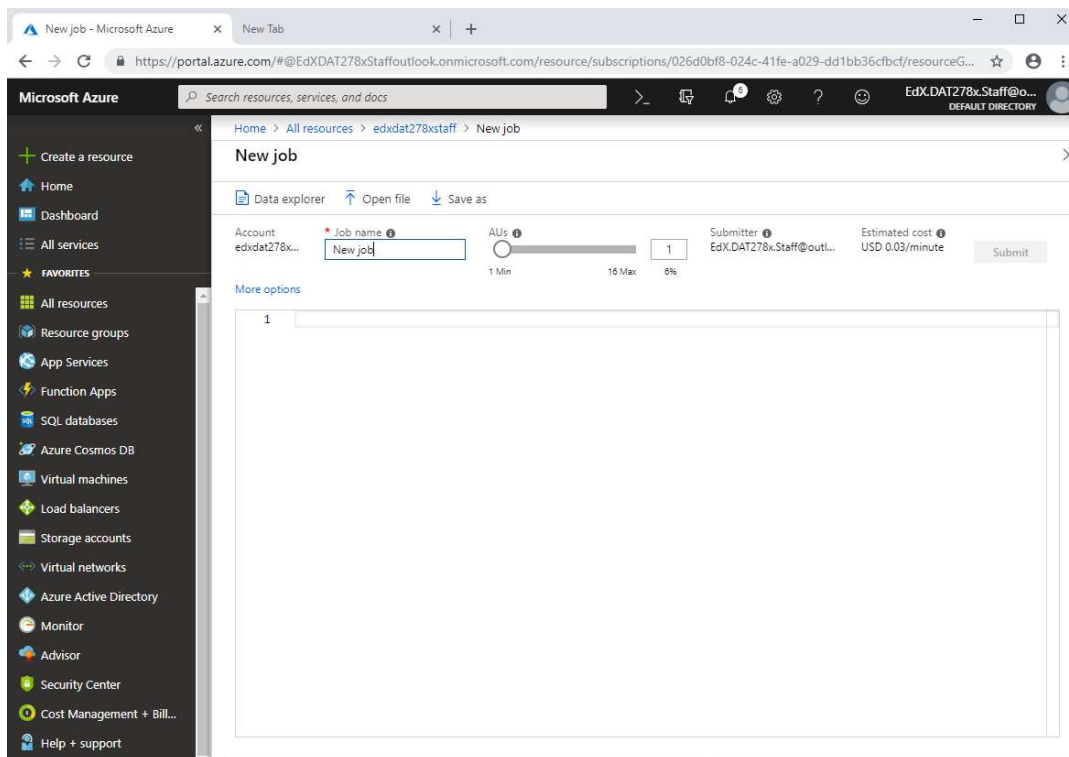
FILE NAME	SIZE	ACCOUNT	PATH	MANAGE	STATUS	PROGRESS
Paper_authors.tsv	59.38 MiB	edxd278xstaffadls	/course-kg/input		Completed	100%
Paper_venue.tsv	14.78 MiB	edxd278xstaffadls	/course-kg/input		Completed	100%

## 5 Run a smoke test U-SQL script to verify that the environment is setup properly.

5.1 Go to Azure Data Lake Analytics account and click **New job** on the top menu.

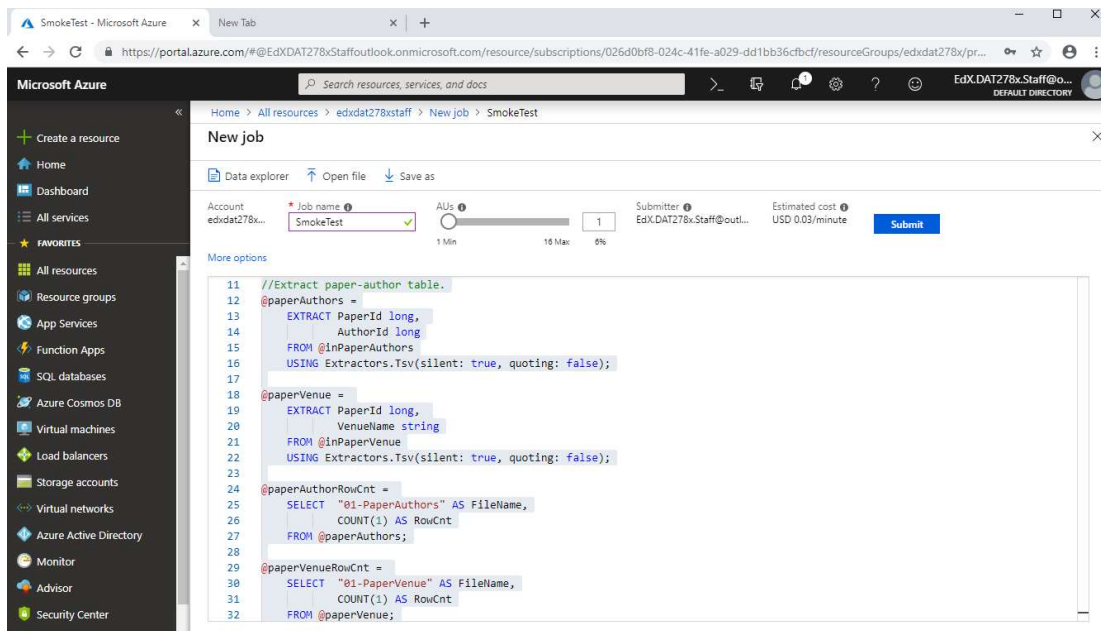


5.2 Click **Open file** to open the downloaded U-SQL script “DAT278x\_SmokeTest.usql”.

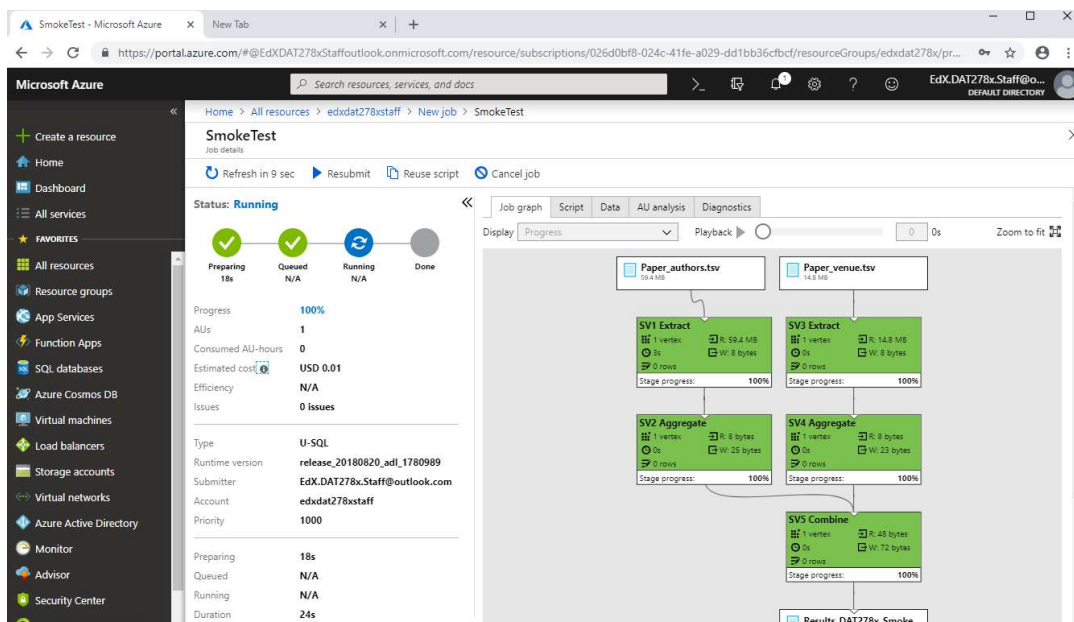




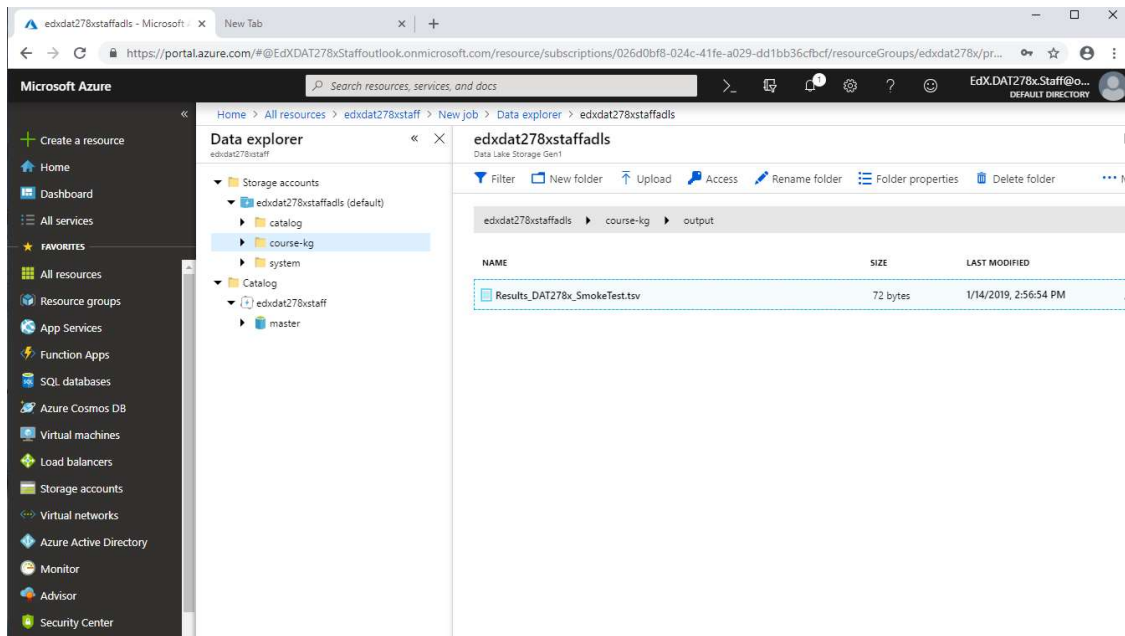
5.3 Click **Submit** to submit the job. The computing cost will be covered by [Azure credit](#) which you received when registering the free trial account.



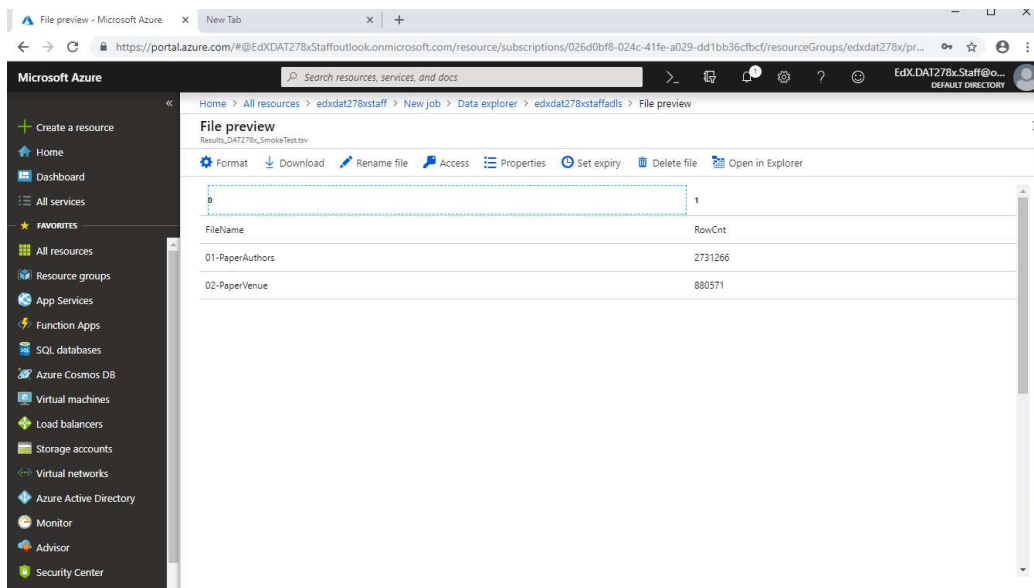
5.4 View the job progress as follow. After running the script successfully, the result file will be outputted to **"/course-kg/output/"** folder.







5.5 Click the output file “Results\_DAT278x\_SmokeTest.tsv”. Verify it contains the same contents as the following snapshot.



If you can follow above steps without issues and generate the same resulting contents for the smoke testing script, **Congratulations!** You have set up the lab environment successfully and you are ready for the challenge labs!

If you have any issues or questions on above steps, please use the discussion forum of DAT278x on the EdX platform for feedbacks, our course staff would monitor and answer questions you raise.