School of Science, Computing and Emerging Technologies

COS80023 Big Data

Pass Task 2: Storage

Overview

Consult this week's videos and possibly other materials to answer questions about the options and considerations for data storage.

Purpose

Demonstrate awareness of the considerations for choosing storage options for Big Data sets.

Task

Answer the questions below.

Time

This task should be completed in your second lab class or the week after and submitted to Canvas for feedback. It should be discussed and signed off in tutorial 3 or 4.

This task should take no more than 2 – 2 ½ hours to complete (including online work on Azure).

Resources

- Lecture
- MS Azure tutorial for relational Database using SSMS: https://docs.microsoft.com/en-us/azure/sql-database/sql-database-design-first-database
- Any other online material
- genAI – Allowed for research. Must be able to explain

Feedback

Discuss your answers with the tutorial instructor.

Next

Get started on module 3.

Pass Task 2 — Submission Details and Assessment Criteria

Write down the questions and answers in a text or Word document and upload to Canvas. Your tutor will give online feedback and discuss the tasks with you in the lab when they are complete.

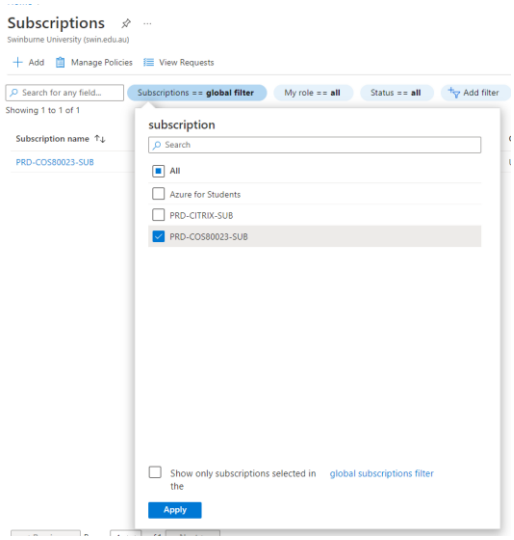
Subtask 2.1

Overview

Log into MS Azure, create an SQL database and upload the data provided. Write a query that shows the suppliers and their products (Supplier name, country, product name, unit and price). Run the query in the Azure Query editor. Take a screenshot of the result, including the entire Azure page that shows your database and the query editor. Upload the screenshot to Canvas (either within a pdf or as a picture. Demonstrate your solution to the tutor.

1. Logging into Azure and Selecting the Correct Subscription

Log into the Azure Portal (https://portal.azure.com) with your Swinburne student email. Click on Subscriptions (key shaped icon) under Azure services. You can also type Subscriptions in the search bar. Make sure to choose the correct subscription (PRD-COS80023-SUB). You might have to toggle “Show only subscription in global filter” option to see this subscription.



After selecting the subscription, you can move to next step.

2. Discovering Your Resource Group

Throughout this semester you will be creating and interacting with many Azure resources. All of them will be placed inside your resource group. Your tutor will be creating this resource group at the start of your tutorial with appropriate access rights for yourself. After the allotted time has lapsed (it will vary for Azure tasks) your access will be cut off and your resource group will be deleted.

You can look for your resource group under Resource groups icon or you can bring up the resource groups by searching for them in search bar.

After clicking, you should be able to see a resource group which has your student number as its name.

3. Creating Relational Database

Go to your Home. Type SQL Databases under Azure Services (top middle). Click +Create above the empty listing, or the 'Create SQL database' in the middle of the screen.

Select the correct subscription (containing COS80023) and your resource group.

Name your database <yourstudentnumber>rdb.

Create a new server <yourstudentnumber>server. Choose Australia East as a location.

Select "Use SQL Authentication"

Choose user<yourstudentnumber>, e.g. user12345678 as you admin user. (This cannot start with a number).

Choose a strong password that is more than 8 characters long, has special characters and upper and lower case letters. (The system will tell you immediately whether a password is acceptable).

Click Ok to create the server.

Choose this <yourstudentnumber>server for your database.

Leave tick no to using an Elastic pool.

Observe the pricing that appears in a square on top of the screen. By default the Workload Environment is set to 'Production'. This means the database is running in production mode and serving clients. Change the setting to 'Development'. Observe how this changes the pricing. Development means, unless you as developer connect to the database, it will switch itself off after a set period and stop incurring costs.

Leave the other settings as they are.

Click on 'Review+Create', then 'Create'. The overview page tells you that the deployment is in progress.

Hint: Pay attention to the bell-shaped notification icon:

It provides more detail about the progress of a deployment with any resource you create.

Click on <yourstudentnumber>rdb. Then click on 'Query editor (preview)' in the menu on the left.

4. Accessing the Relational Database

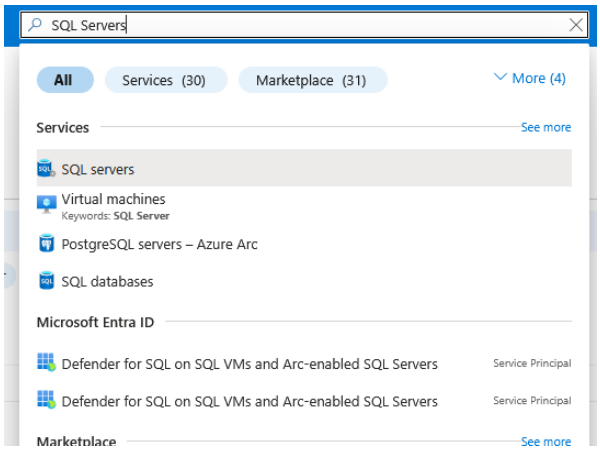
You are prompted for the password you specified. Even if you remember it correctly, you will be prevented from accessing the database.

Read the information provided carefully.

> **Information**: It is normal for online databases to be accessed from an online application only, e.g. a web site that shows information from the database. It is not normal, and usually dangerous, for databases to be accessed directly, using SQL or another suitable query language, from a remote desktop. By default, this is prevented by a firewall. The authentication response explains how to create a secure endpoint. This is worth studying if your database goes into production.

For now, we just want to create our first database, and this means exposing the database port to the outside world. To do this, we have to find the server we created earlier.

Go to Azure Home and type 'SQL Servers' in the search bar. Click on the corresponding entry in the list.



A listing with <yourstudentnumber>server appears. Click on it and find Security on the navigation pane on your left. Click to expand the options.

Click on Networking. On the Public access tab, switch the toggle to 'Selected networks'. This is not very safe and should not be used in production mode, but private endpoints require virtual networks, which provides encryption but requires more work to set up. For the development phase, we can open a port.

Choosing Selected networks opens the Firewall rules section. Click on 'Add your client IPv4 address …' .

5. Adding Data to the Database

After this step you should be able to log in using the user account you created earlier.

After logging in, run the Create Table statements and data Insert statements provided. Expand the 'Tables' item to the left of the query windows. This should show the tables created. Run the query that shows the suppliers and their products (Supplier name, country, product name, unit and price).

Take a screenshot and submit it on Canvas.

Important: You **MUST** delete your database server at the end of this exercise. If you do not, it will keep running and use resources.