MICROSOFT AZURE TUTORIAL:

The purpose of this tutorial is to instruct a new user who has never used Azure services before what they need to do in order to import their databases onto the Azure cloud, move them onto the Azure-hosted MySQL instance, and then load those databases onto their local machine’s instance of MySQL Workbench for further analysis. We assume that the user already has an Azure account.

The general idea is that you need a cloud storage account, where you will store your CSV files to Azure’s Blobs. You then will install a Data Factory service which will serve as the pipeline to get your data from the blob into MySQL. Finally once your data is piped into Azure MySQL, you will be able to query your databases on your local machine in Workbench!

Why do we do it this way? Why can we not just load things into the cloud directly from MySQL? It’s slow! Even if you have a great computer with 64gb of ram and a GTX 2080 gpu with a Ryzen 7 CPU, does not matter. And Azure disables LOAD DATA LOCAL INFILE SQL commands so you don’t have that option either. Thus you have to do it this way.

**NOTE: *Ensure that after each step you go into Access Control (IAM) setting on left side of the screen, and add all your classmates to be able to access the project! Otherwise your team will not be able to help you!!!***  
  
**STORAGE ACCOUNT**

1. Create a Resource => Storage Account
   1. Name it whatever you like
   2. Be sure to createa new Resource Group – this will be used by all applications
2. Next, go to Advanced tab. Change Access Tier to cool to save money.
3. Next, go to Enable public access from all networks
4. Hit Review and Create, then Create.
5. Congratulations you now have your storage unit! It will takea few minutes to generate everything. But once that is done, navigate to Containers underneat the Data Storage option on left column
6. Then, at the top of the screen click the blue plus sign that says Container. Assign it a name and hit create. You will want one folder per census year.
7. Once you create your blob, click on it, and there you can easily upload all of your data sets for that particular year. You will need to repeat this for every census year.

**AZURE DATABASE FOR MYSQL**

1. Click on Create a Resource, then search for MySQL. Click on Azure Database for MySQL.
2. Click the blue Create button next to the Plan menu
3. Under Resource Type, select Flexible Server (recommended) and hit Create
4. Assign the server to the same resource grou and subscription as you did Storage.
5. Enter a server name, select “For Development or Hobby Projects”.
6. Click on Compute + Storage, and select one to fit your needs. For the census project we recommend Standard\_B2MS with 50gb total storage size. You do not need high availability or backsup.
7. Click next. Under connectivity method choose “Public Access”. Check the box under Firewall Rules. Then click the blue hyperlink that says Add 0.0.0.0+255.255.255.255
   1. What this does is allow every IP address to be able to added to your firewalls approve list. For production settings this is very bad but for a class room project, its fine.
8. Click Review + Create at bottom of screen and wait. Congratulations!

**DATA FACTORY**

1. Click Create a Resource, and search for Data Factory. Click the blue Create button.
2. Select the same subscription and resource group as the others. Asssign it a name. Click Review + Create on bottom of the screen.
3. Data Factory will take longer than the others to create itself. Itll take between 10-20 minutes depending. After its loaded, on your dashboard click on the data factory you created.
4. Once you are on the Azure Data Factory webpage, click the blue pencil icon on the left bar of the screen.
5. Now select the Datasets option. Click on the blue plus sign and select Dataset.
6. In the New Dataset window that pops up, type Blob and click Azure Blob Storage. Press continue.
7. Select Delimited Text as format (***NOT EXCEL FILE!!!!)***. Click continue. Assign it a name, link it to the Azure Blob Storage you created earlier, and don’t forget to check First Row as a Header box. Click OK.
8. In the connection box, click “Browse” to the right of the File Path. Select the folder you want to load in bulk. Leave the rest of those options alone.
9. You must repeat this process for each folder you created in the blob. IE three times.
10. Click the blue arrow on the left of the screen by Factory Resources again, same as before.
11. This time, type in MySQL in the search bar and click Azure Database for MySQL. Click continue.
12. Link it to your Azure MySQL you created earlier. Select the specific table you wish to copy. Click the import schema box. Hit OK.
13. The rest works the same way as for the excel files. Look at the Schema tab and take note of the number of columns, column names, as well as data types. ***These must exactly match all of those conditions for both your MySQL dataset as well as the CSV you created earlier. If they are not, you will get errors!***
14. Once you’ve verified that, click on the blue plus again under Factory Resources. This time you will be created a pipeline.
15. On the Activities part, under Move & Transform, click on it and then drag Copy Data into the white space. You should see a blue box called Copy Data with a garbage icon, curly brackets and the copy symbol below. Assign it a name and then click Source tab.
16. For your source tab, you will select the CSV dataset you created earlier from the blob. Then, you will click on the Sink tab, and then you will select the corresponding MySQL dataset. For example, you have one pipeline: its input is the 2000 census CSV and its output is the 2000 census MySQL dataset.
17. Next click on the Validate option towards the top of the screen to the right of the Activities box. It has a blue check mark. If you have correctly configured everything, it will say “Your Pipeline Is Validated” on the left. This means you are good to go! ***Be mindful you will need one pipeline for each census data set, so three total.***
18. Click the blue House button on the left to go back to the main data factory page. Then click the Ingest option, it should have a blue cloud with an arrow pointing up.
19. Select Built-In Copy Task and select Run Once Now, then hit next. For source, select Azure Blob Storage and then the corresponding blob. Then select Browse and find the corresponding folder for the year you want to pipe. Click Next at the bottom of the screen then next again.
20. On the Target screen repeat the same process, but with your corresponding MySQL database. Click next. Double check your mappings then hit next, then next again.
21. Congratulations! Give it about 20-30 minutes and then you’ll be good!