CS 348 Fall 2014

Assignment Zero

Due Date: NA

Background

This assignment is not marked and has no deliverables. However it is an extremely important learning exercise using which you will be able to complete other assignments effectively.

The purpose of the assignment is to assist you in doing the following tasks:

a) Setup a database server
b) Create a database
c) Load a database with a reasonable amount of data
d) Begin writing queries against the database
e) Begin writing programs that utilize the database

There are two parts to this assignments of which you can do either or both if you like. However, it is highly recommended that you complete Part A before proceeding to Part B.

Part A: The easy way to get set up for CS348

For this part you have to complete the following tasks

a) If you are unfamiliar with the term *virtual machine* you must do some independent reading on the topic. Wikipedia is a good start. In addition, there are many tutorials and videos on youtube that you can consider watching.
b) Install the recommended Virtual Machine software for your system: We recommend that you install VMware Player for Windows or Oracle's Virtual Box if you have a Mac. Both of these software are free for non-commercial use and very highly supported by their communities. You will have to troubleshoot any installation problems yourself.
c) Download the provided virtual machine image for CS348 from the course webpage. You need all three files to be in the same folder/directory to be able to run the virtual machine. The large file is the virtual disk for the machine, the ovf

(open virtualization format) file contains the configuration/specification of the virtual machine hardware and .mf file contains hash checks to ensure the disk and machine configuration have not been tampered.

d) Open/Import the virtual machine and then start it. Wait for it to boot and then login. Once the machine boots fully, you'll be taken to a graphical interface. The machine comes pre-installed with everything you need for assignments 1 and 3. The machine has the Ubuntu 14.04 operating system installed and you can customize the user interface (look and feel), install other programs in the virtual machine as you deem fit. Refer to the Ubuntu user manual if you are unfamiliar with linux or have not used Ubuntu before. You can login using "cs348" as both the username and password. Note: In the current machine configuration it is allowed to use 1GB of RAM from your host computer's RAM. You may need to reduce it to 512MB if you are running the virtual machine on a memory limited computer.

e) Open the MySQL workbench, connect to the database installed locally on your virtual machine and then execute the query "SELECT COUNT(*) FROM LINEITEM" against the TPC-H dataset. The icon shown above starts the MySQL workbench which allows you to query the pre-installed database server (MySQL). The answer you should get from the above query should be 600,572.

f) Finally, compile and execute the Test.java program using your favourite java IDE or using the console. The icon shown above will open the a console window. Change directory to /home/cs348/Desktop/A0, and then enter the command "javac Test.java" Make sure it compiles without errors. Then execute the program using the command "java Test" and it should print the result of the query shown in part (e).

If you have completed all the steps above then you are all set for CS348. You are now running a virtual machine which is configured to be a MySQL based database server. In part (e) you opened a connection to the database and ran a query against a pre-existing database. In part (f) you compiled and ran your own java program that connected to the database and ran a query.

Part B: The hard way to get setup for CS348

For this part you are to do 5 tasks (a-e) mentioned below and have absolute freedom in how you do them. Some recommendations and details of each task are given below

a) Download and Install a SQL compatible relational database server of your choosing on your computer (virtual or otherwise).

You can choose any of the popular free DBMS such as PostgreSQL, MySQL, DB2 Express, Microsoft SQL Server Express or Oracle Express Edition depending on your operating system and computing resources. Make sure to read the limitations of these database products, and provided documentation on how to get them setup.

b) Create two databases, i.e., the TPC-H database and the Chinook database on the database server product you installed in part (a)

The schemas to the databas is provided on the course website. You can use the graphical table editors provided by the DBMS, or use SQL statements to create tables.

c) Load the sample data provided on the course webpage into the appropriate tables in the databases you created in step (b)

You will have to explore different techniques (loading and importing data in bulk) available in your selected DBMS to see how best to do this. Most RDBMs have built in commands using which you can load data in files directly into your tables.

d) Access the database server and execute the following query “SELECT COUNT(*) FROM LINEITEM” against the TPC-H dataset

Most database systems come with a graphical querying environment in which you can type, execute, and debug queries. However regardless of your chosen system the result of the above query should be the same (i.e., number of rows in the LINEITEM table). Find the best environment to execute and debug queries and run the above query to ensure that it is working properly.

e) Compile and execute the Test.java program provided on the course webpage against the TPC-H dataset

This task is slightly more difficult and involves setting up your own database (server and dataset) so that programs can access it. It may require setting up a user in the server and giving it sufficient permissions to query the TPC-H database that you just

created. Depending on your choice of database server you will need to install the appropriate JDBC (Java DB Connectivity drivers) that allow java programs to communicate with the server.