**MSCD664**

**Lab 2 – Cassandra NoSQL Database**

**Name: Rohan Purohit**

**Date: Sept 11/2016**

----------------------------------------------------------------------------------

The purpose of this lab is to become familiar with the Cassandra Virtual environment. Please respond to the instructions/questions highlighted in yellow.

***It is critical that you watch and complete all of the steps in the two FTE (From the Experts) Cassandra Installation and Configuration videos (e.g. Parts 1 & 2).***

You will perform the following tasks in the lab:

* Starting your Cassandra database within the Ubuntu Virtual Desktop environment.
* Start the CQL shell (similar to SQL)
* Create a few Cassandra column families (table) and then add data.
* Query data from the Cassandra database.
* Drop the Cassandra families (tables).
* Shutdown Cassandra gracefully and exit the virtual box.

**Read Chapter 10 in the NoSQL Distilled text.**

To get an “A” grade on this lab:

* Answer all questions.
* Provide output of steps that require you to perform a task. The output can be a cut and paste from the screen to a word document or screen shots; the windows snipping tool works well. I need to see something to verify that you ran the labs. I’m expecting you to turn in a word document that contains each question and also the corresponding answers to questions asked and output from Cassandra statements. It is okay to paste your results into this word document and then submit the completed document.
* Tasks that require an output from you will be indicated clearly with yellow.

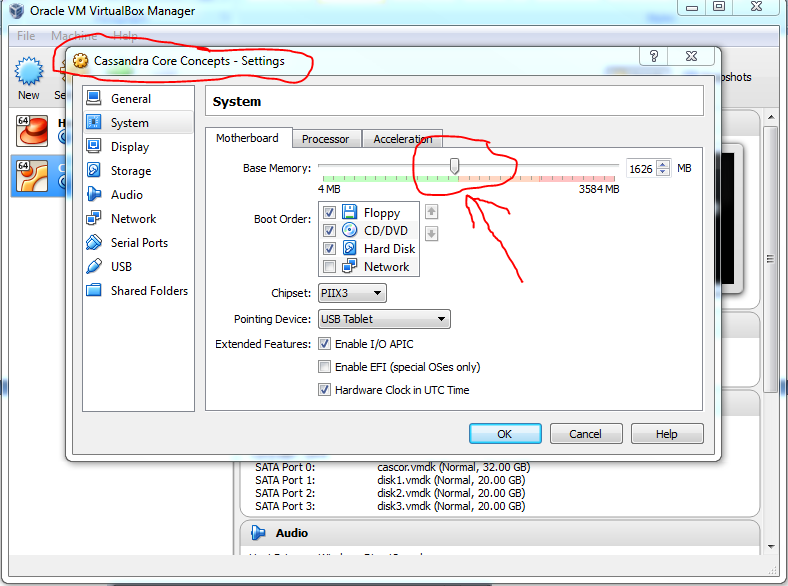
**----------------------------------------------------------------------------------**

**Part 1: Understanding Your Cassandra Environment**

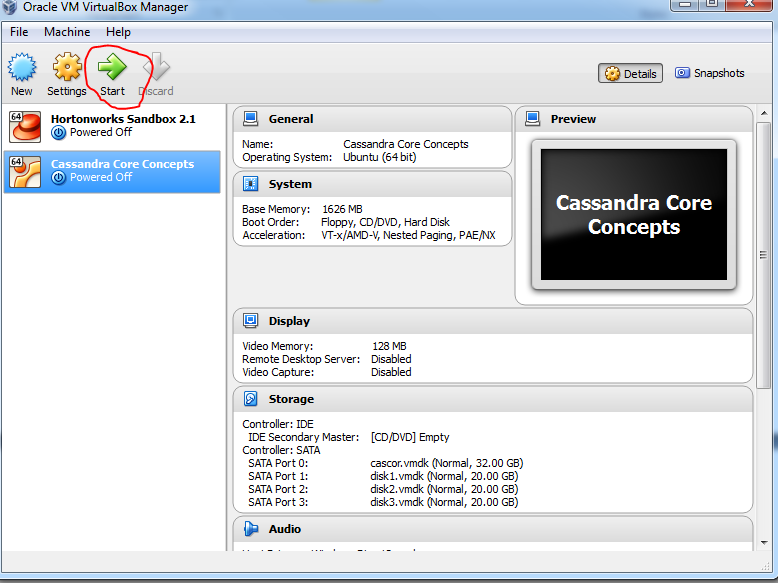
**-----------------------------------------------------------------------------------**

1. Start the Oracle virtual machine and check the amount of memory for Cassandra by looking at the system setting. If you have a small amount of memory on your PC (such as 4 Gig of RAM), adjust the Cassandra Core Concepts System settings as shown below. If you use up all of the memory on your PC, it will lock up and you’ll have to reboot your PC. I moved the slider to the edge of the pink area still within the green area of memory and Cassandra worked fine.

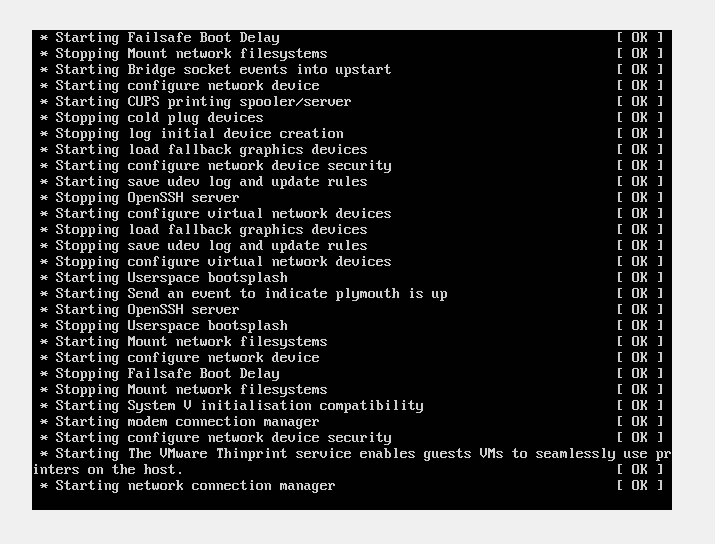
**Note: If you have problems with getting Cassandra to work in the VM environment on your PC, then please send an email to your instructor with the details. Do not spend hours suffering in silence.**



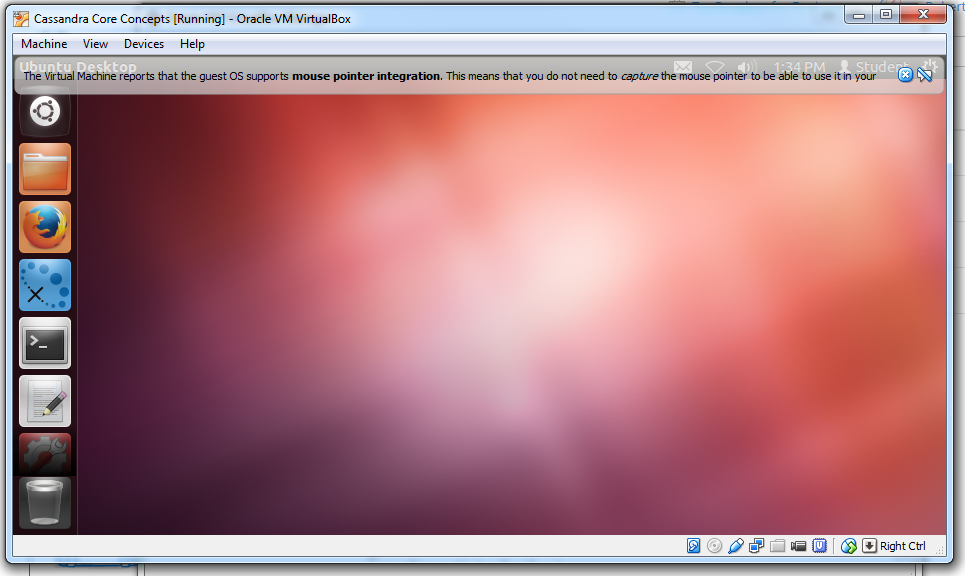
Start Cassandra by clicking on the green start button:



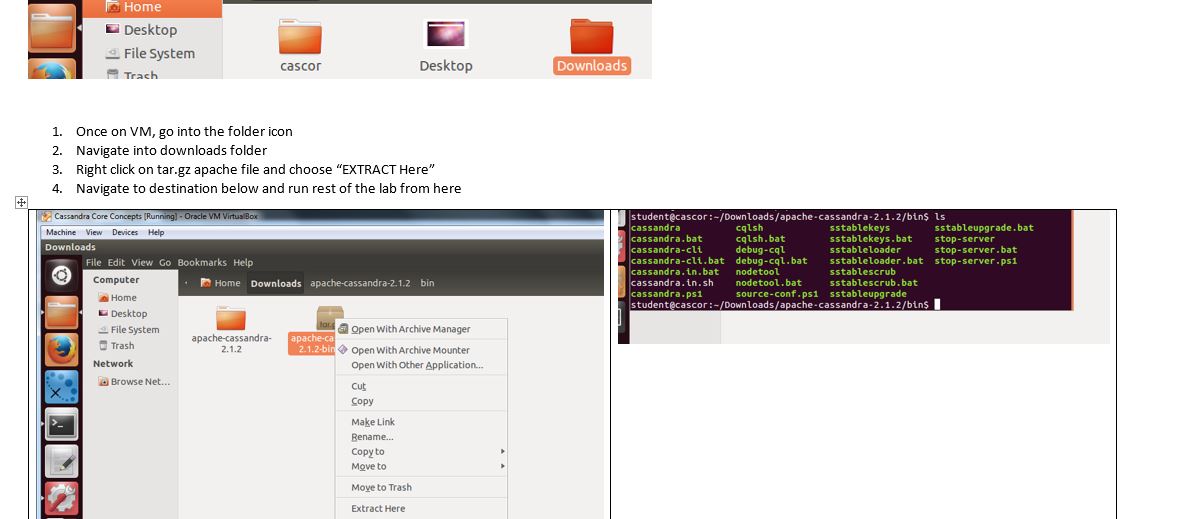
As the startup progresses, your window will contain status items as shown below.



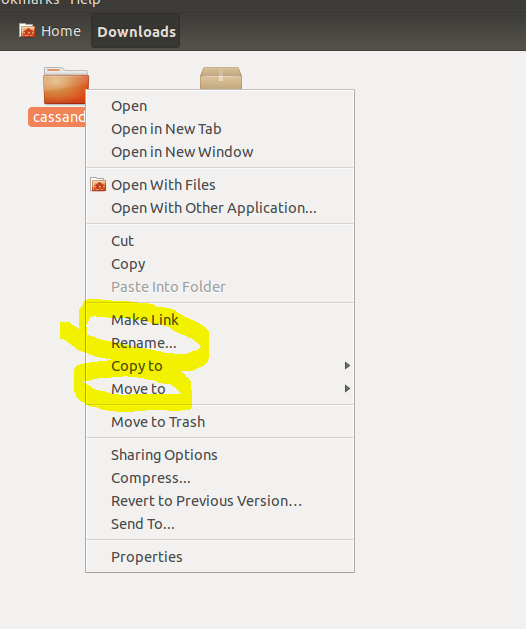
1. When the Cassandra Ubuntu Desktop is ready to use, your window will appear as shown below.



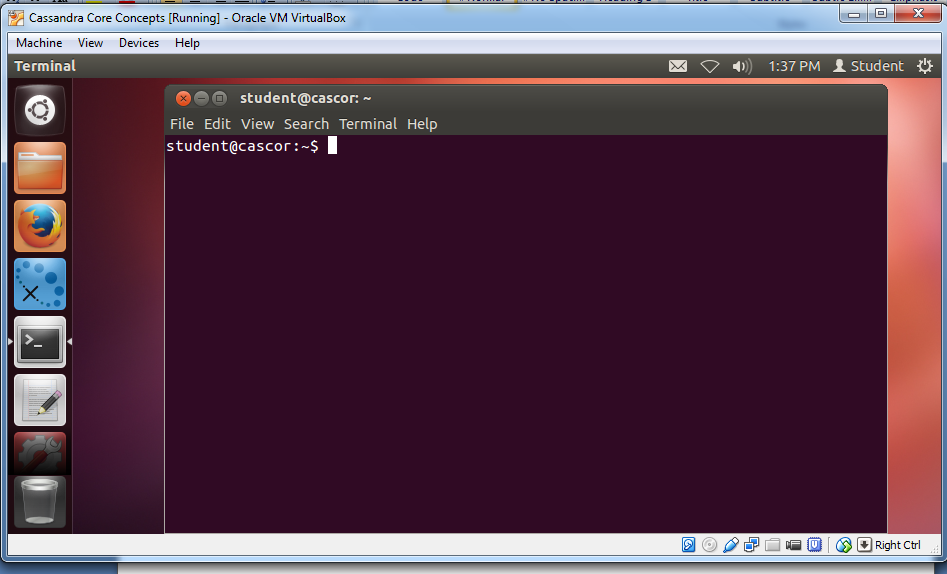
1. Prior to starting Cassandra, you need to unzip the binaries that are included with the VM files. Follow the instructions below to unzip the binary. You can run Cassandra from the new folder that is created in the Downloads directory or you can move and rename the folder to a shorter name such as Cassandra (see below).



To move and rename the folder, right click on the name of the new directory that was created after the extract. Select rename from the drop down list and then shorten the name to Cassandra (as shown below). To move the directory to the student directory, right click and use the “Move to” option.

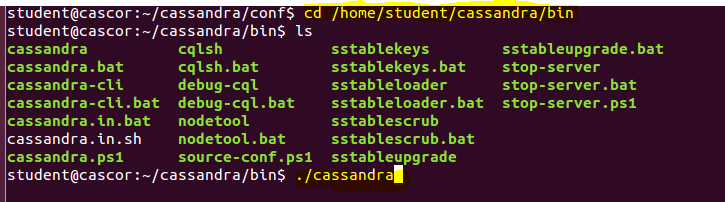


Open a terminal session as you did during the installation process. Use the Linux pwd command to show your current directory (e.g. pwd).



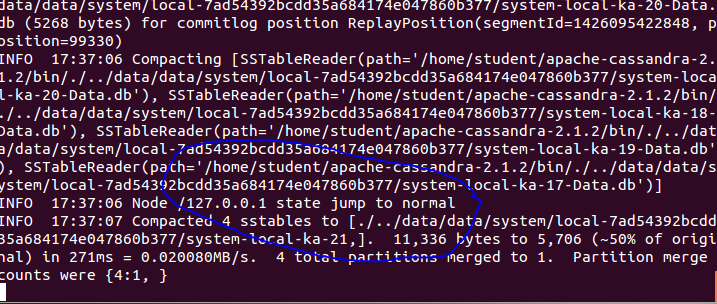
Using the CD command, change directories to the /home/student/cassandra/bin directory or the directory where you extracted Cassandra which might still be in the Downloads directory. Either way, navigate to the bin directory, then enter the ls Linix command to list the files in the directory as shown below. You will see the binaries and other executable files listed. Start the cassandra binary with the following command:

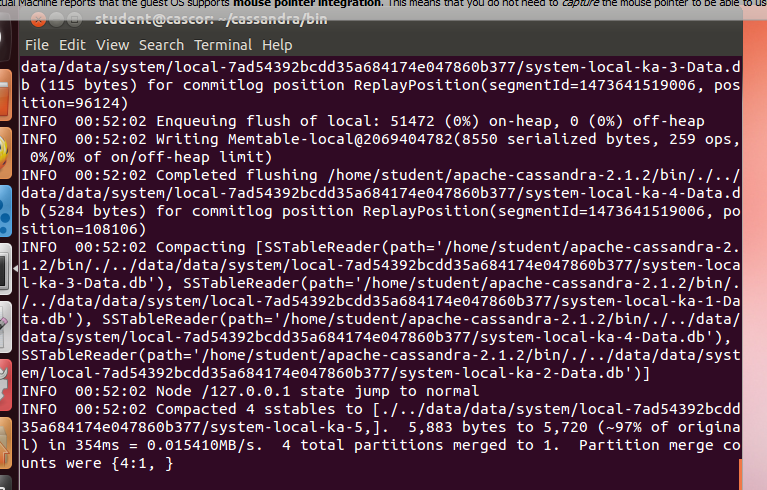
**./cassandra**

****

You will see a lot of startup text during the startup process and there will be several iterations of text. Wait until you see the “127.0.0.1 state jump to normal” which will indicate that Cassandra has started. Press the enter (return) key to get a shell script prompt.

**Note: If you wait along time (5 minutes) and cassandra does not reach this point, then you may have an issue with your PC and please contact your instructor via email.**



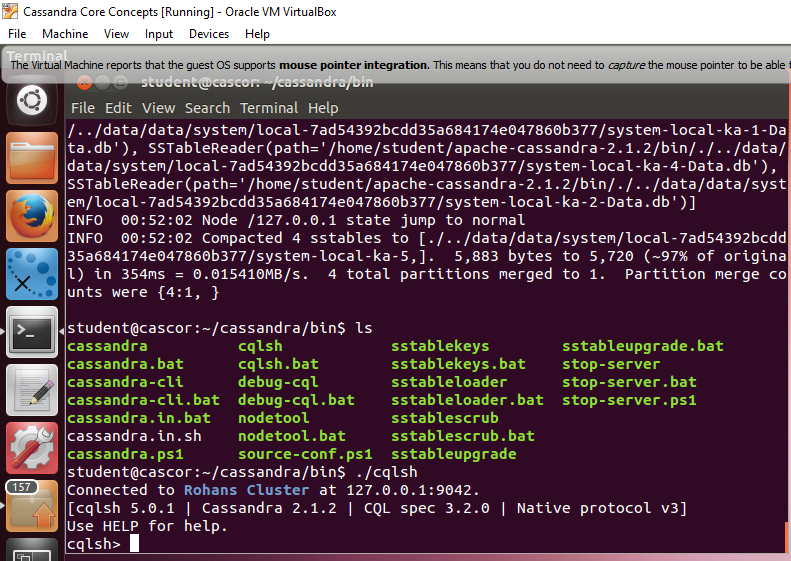


At the command prompt (remember to press the enter key to make it appear), you will start the cqlsh script. This is a similar environment to sqlplus and it allows you to enter commands that are similar to ANSI SQL. First enter the ls command to see the contents of the bin directory.

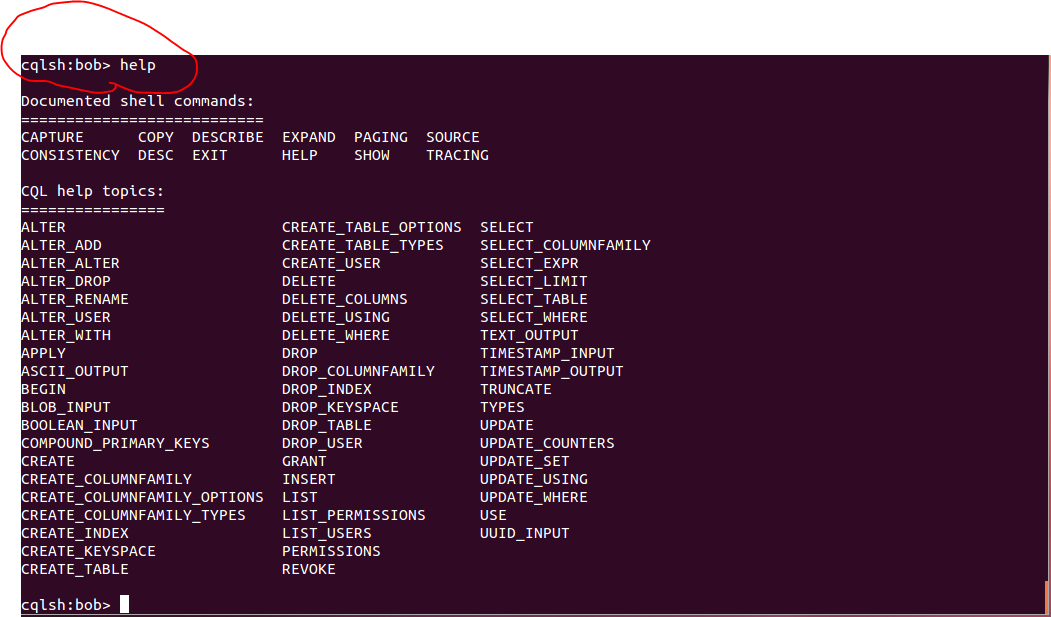
Then start the cqlsh script via (dot / cqlsh) >> ./cqlsh.

The prompt will change to have cqlsh>.

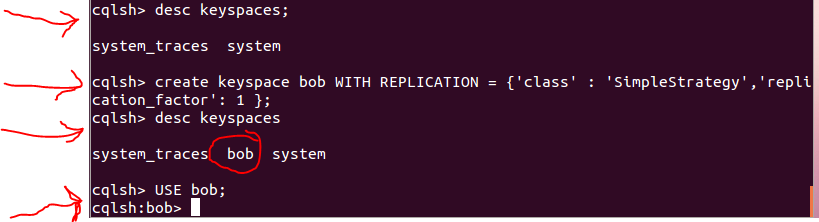
Provide a screenshot of your CQL session to show that the environment is *running*:

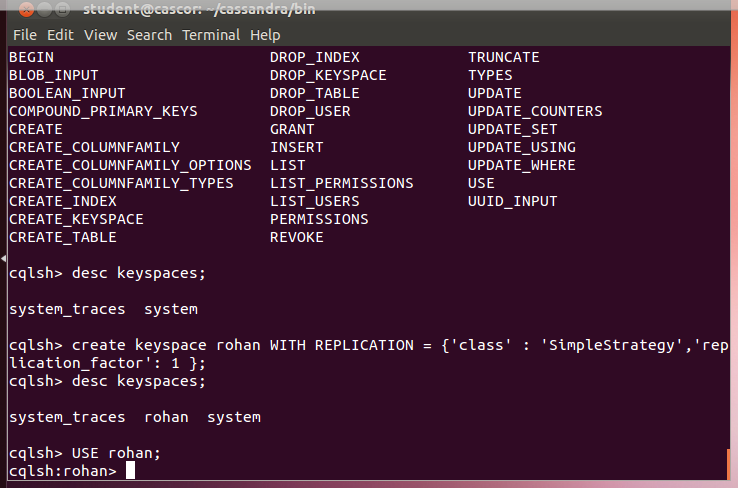


1. Help is a useful feature in this environment. Type help and press return to see help on various commands. You can then type help followed by the command name, such as help create\_table.

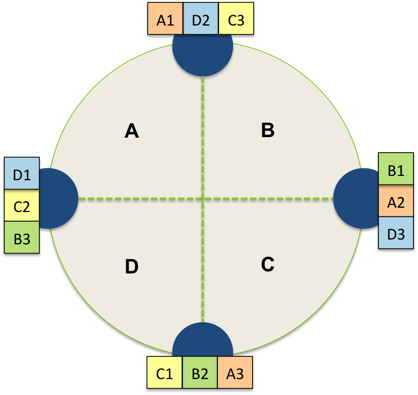


1. The next step is to create a keyspace. In the example below, first describe the current keyspaces. System and system\_traces were automatically created for you. Create a keyspace using your name, in this case I used my name ‘bob’. We will use a SimpleStrategy for replication (see below) and a replication factor of 1 since we only have 1 node (server) setup in this VM environment. After creating the keyspace, enter the USE command with the name of your new keyspace. Notice that the prompt has change. In terms of RDBMS, a keyspace is a schema, similar to what you would find in an Oracle Database.

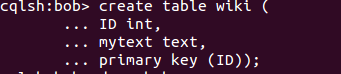




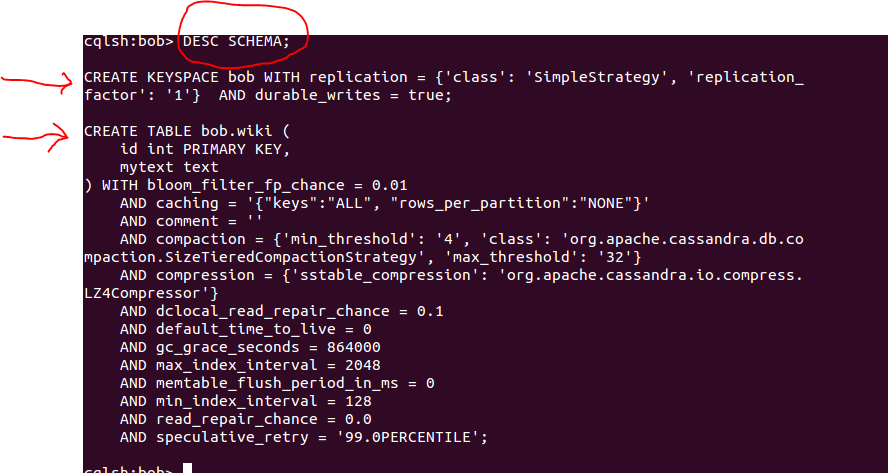
SimpleStrategy places the first replica on a node determined by the [partitioner](http://www.datastax.com/docs/1.0/cluster_architecture/partitioning#partitioning). Additional replicas are placed on the next nodes clockwise in the ring without considering rack or data center location. The following graphic shows three replicas of three rows placed across four nodes:

[](http://www.datastax.com/docs/_images/simple_strategy_replication.pn)

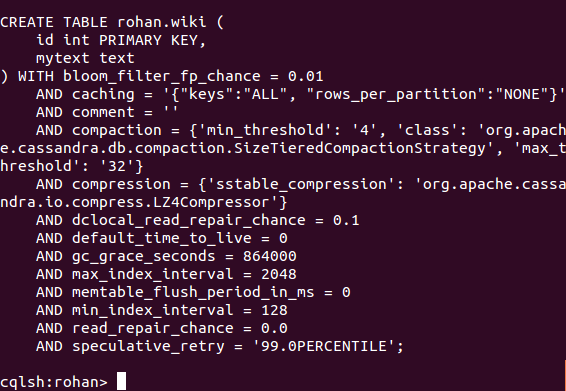
1. Within the schema, create a table called wiki. Keep in mind that tables in Cassandra CQL are equivalent to the column families that we created in the HBase lab 1. Notice the three (**…)** dots on the second line. This indicates that the command is continued on the next line. If you ever make a mistake you can press **Control-c** to cancel a command. In this example, when you reach the end of the command, you will enter a semi-colon to end the command. You will specify the primary key field at the bottom of the CQL statement for this table.



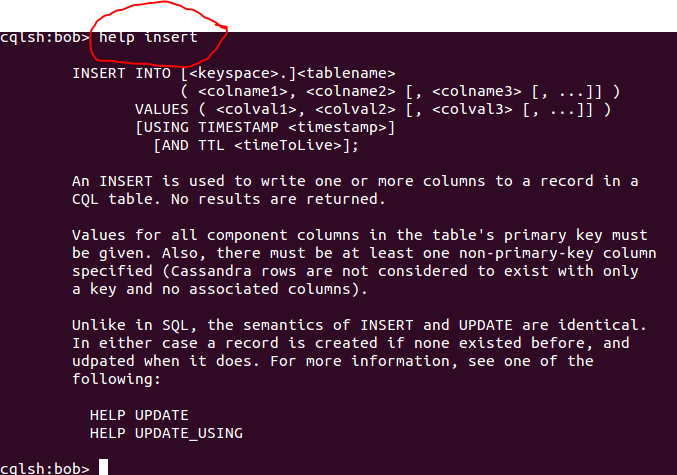
1. At this point, you want to check to see if the table was created. Use the **DESC SCHEMA;** command to view the contents of the bob schema (keyspace). Notice that the keyspace bob and the wiki metadata is shown below.



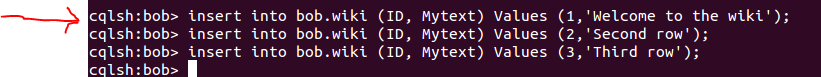
Provide a screenshot of your CQL session to show your new keyspace and table:



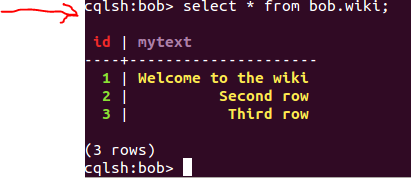
1. Before inserting data, you will use the help command to get information about the INSERT syntax within the CQL environment. You can see that the syntax is similar to ANSI SQL.

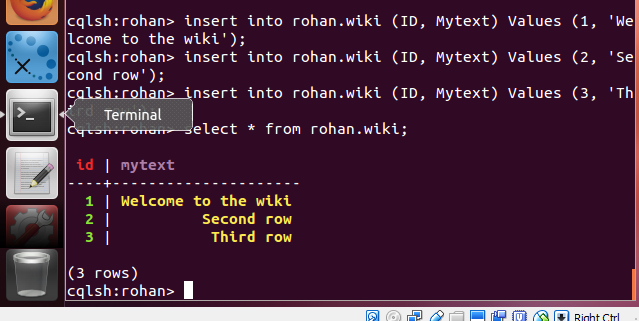


1. Insert three rows of data into the new wiki table (as shown below).



Then verify that the data was inserted to the table using **select \* from bob.wiki;**





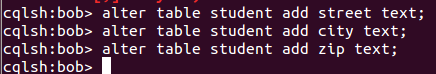
1. Create a table for student (similar to what you did in the HBase lab 1). Describe the schema the schema to verify the new table was created.



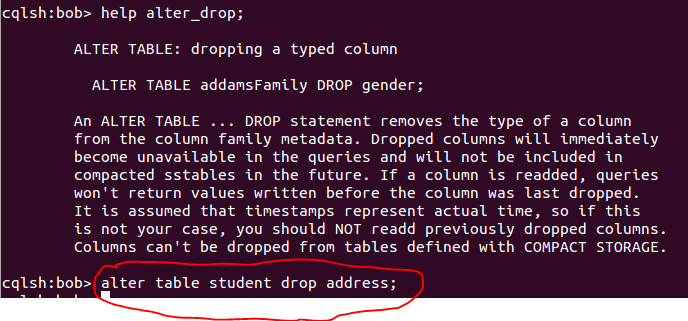
**DESC SCHEMA;**



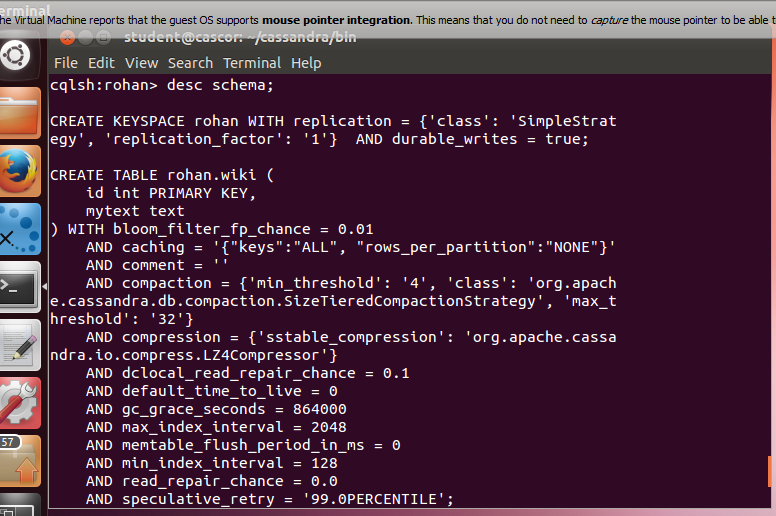
1. You will now break the address column into components. You will define three new columns for the column-family called street, city and zip. You will then remove the address column from the table. Prior to dropping the column, request help for the syntax to drop a column.



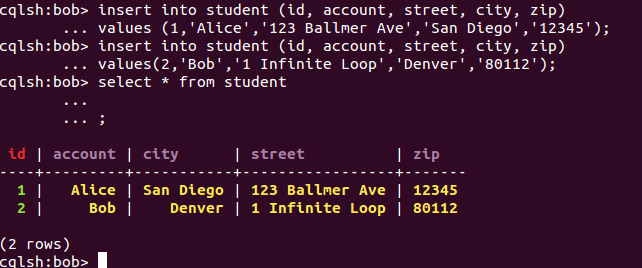
Get help for the alter drop syntax and then drop the address column.

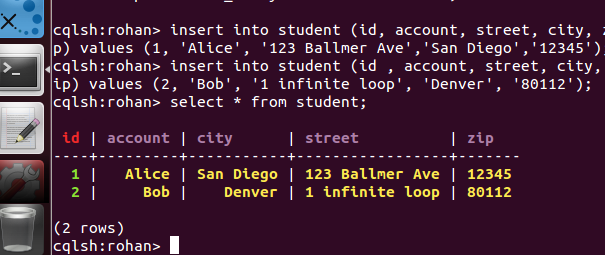


Describe the schema again and then provide a screenshot of your CQL session to show your revised table structure:

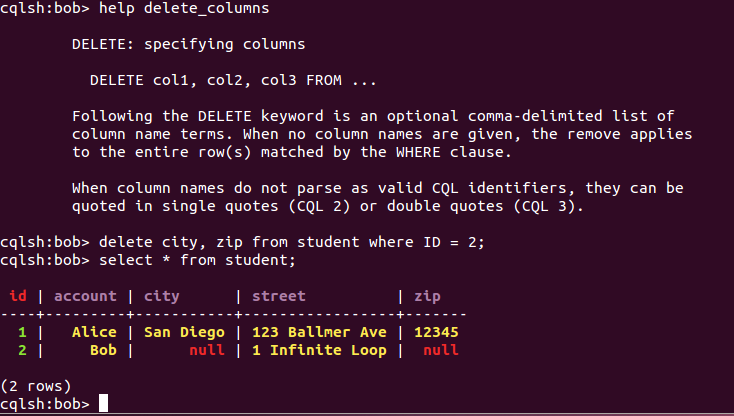


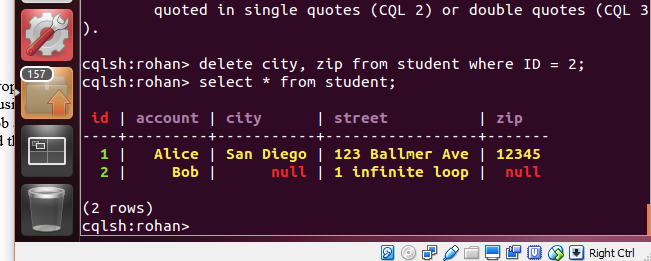
1. Insert three rows into the student table and then **select \* from student;** to verify the data. In the example below, I didn’t enter the semi-colon right away, so it did not run my command until I entered the semi-colon two lines down.



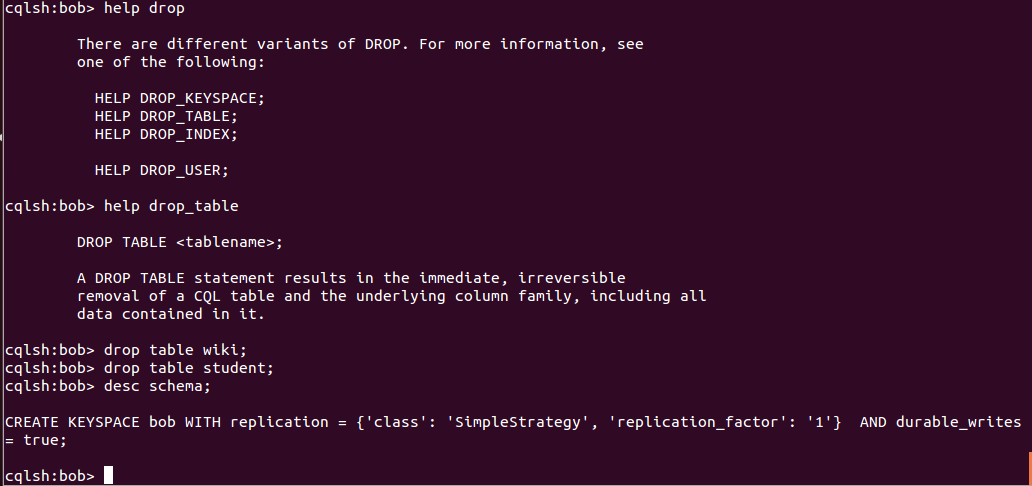


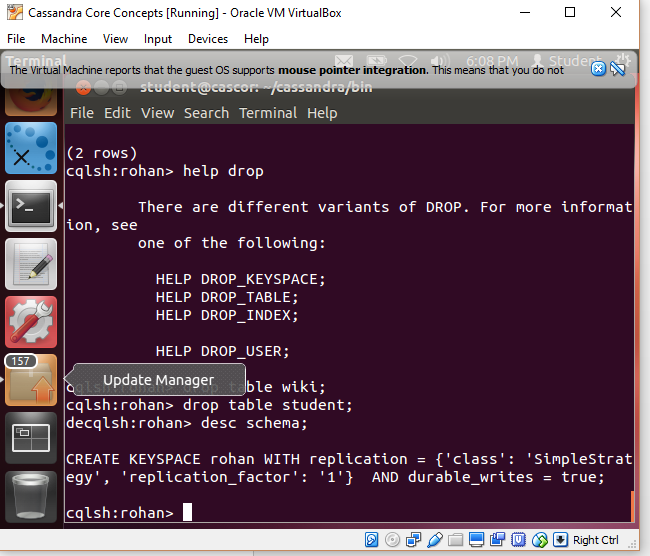
You will delete some of the values in the columns by using the delete command. Notice that I specify the two column values that need to be deleted and specify the row number using the where clause. The database does not store actual nulls, however CQL shows them as null values. Similar to the HBase Lab 1, the values are not kept in the table for row 2.



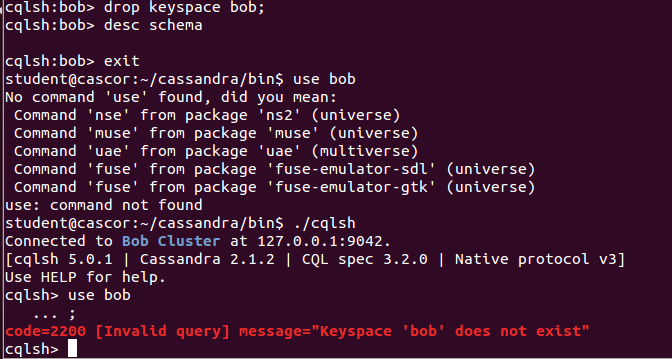


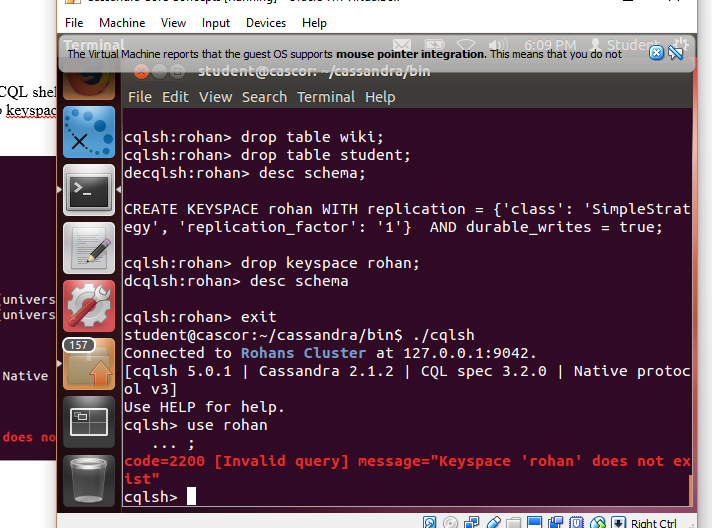
1. You will now remove the tables that you created and drop the keyspace (schema). First drop the two tables (wiki and student) using the drop command (as shown below). Describe the keyspace bob and you will see that the tables are gone. There is not an undo command and the tables and data are permanently deleted.





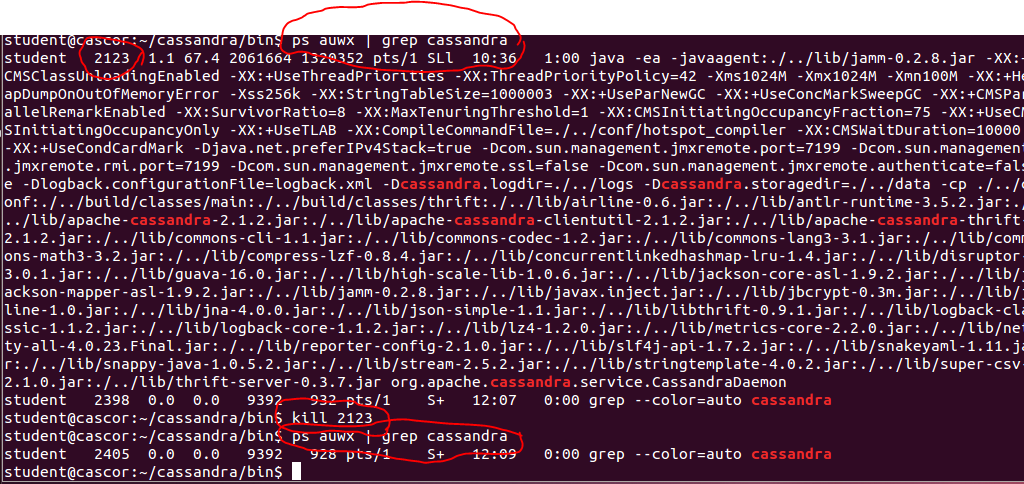
1. Now drop the keyspace bob. I exited out of the CQL shell and then I started it up again using ./cqlsh. If you try to USE the bob keyspace now, you will get an error that it does not exist.

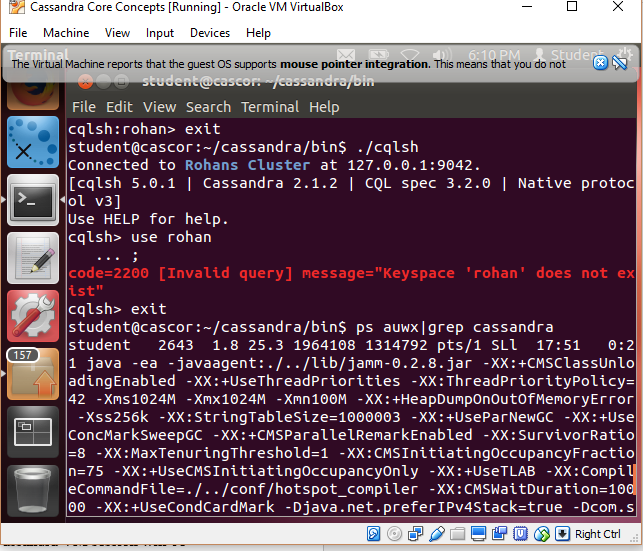




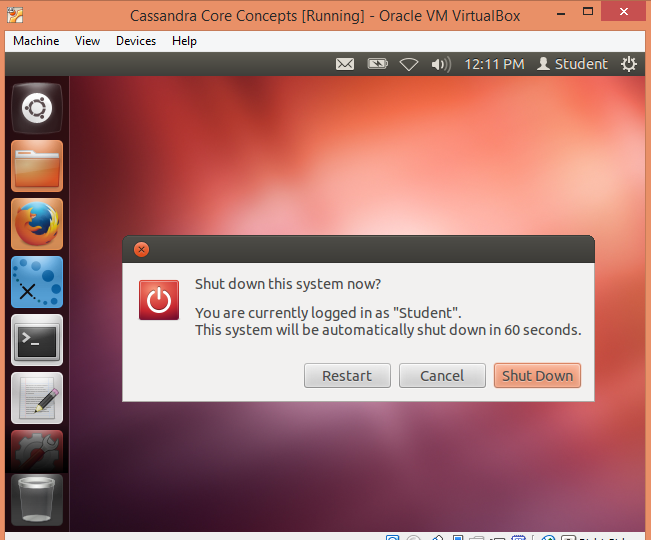
1. Type exit to end the CQL session. Then, enter: **ps auwx | grep cassandra** command to show the PID for the Cassandra session. Take note of the number next to the word student which is 2123 in this case. On your PC it could be a different number.

Use that number with the **kill** command to shutdown the Cassandra database (gracefully) as shown below. Run the ps auwx command again and you will see that Cassandra has shut down.





1. The last thing is to-do is to shutdown the Cassandra VM session. Click on Machine menu item and then click on the ACPI shutdown. Click Shut Down when prompted. After a minute or two, your Cassandra VM session will be closed. At this point, exit from the Oracle VBOX environment.



**We hope that you enjoyed this brief lab using Cassandra. If you run the same lab on a Cassandra cluster with many nodes (server machines), then the results will be the similar.**