

CS595—Big Data Technologies

Assignment #12

Worth: 12 points

Due by the start of the next class period

Assignments should be uploaded via the Blackboard portal.

Exercise 1) (4 points)

Read the article “A Big Data Modeling Methodology for Apache Cassandra” available on the blackboard in the ‘Articles’ section. Provide a ½ page summary including your comments and impressions.

Exercise 2) (2 points)

For this and the following exercises you will use an instance of the Cassandra database that I have set up for you in the Azure cloud. Note, as I am paying for this myself, I will only keep the database available until next Thursday.

Note, sometimes you pick up invisible characters if you cut and paste the below commands, so type command sequences in manually.

To access the database do the following:

- a) Access the database VM via ssh...
ssh -p 22 cass1@13.90.97.198
- b) When prompted for a password use...
Unix79127912
- c) Once logged in to the Cassandra VM create a working directory for yourself.
The name of the directory should be your IIT id (mine is A20155104)
- d) This will be your working directory into which you will place your CQL command files
- e) Open a second terminal window and ssh into the VM again as above.
- f) Change to the working directory you created previously.
- g) Start the Cassandra shell by entering the following...
cqlsh -u cassandra
- h) When prompted for a password, enter the following...
PCAIGPNTuin2 <- the 4th letter is the capital letter I,
- i) Now create a file in your working directory called init.cql and enter the following commands. Use your IIT id as the name of your keyspace...

```
CREATE KEYSPACE <IIT id> WITH REPLICATION = { 'class' : 'SimpleStrategy',  
'replication_factor' : 1 };
```

- j) Then execute this file in the CQL shell as follows...
source './init.cql'
- k) At this point you have created a keyspace unique to you. So make that
keyspace the default by entering:
USE KEYSPACE <IIT id>;

Now create a file in your working directory called ex2.cql. In this file write the command to create a table named 'Music' with the following characteristics:

| Attribute Name | Attribute Type | Primary Key / Cluster Key |
|-------------------|----------------|---------------------------|
| artistName | text | Primary Key |
| albumName | text | Cluster Key |
| numberSold | int | Non Key Column |
| cost | int | Non Key Column |

Execute ex2.cql. Then execute the shell command 'DESCRIBE TABLE Music' and include the output as the result of this exercise.

Exercise 3) (2 points)

Now create a file in your working directory called ex3.cql. In this file write the commands to insert the following records into table 'Music'...

| artistName | albumName | numberSold | cost |
|----------------------|---------------|------------|------|
| Mozart | Greatest Hits | 100000 | 10 |
| Taylor Swift | Fearless | 2300000 | 15 |
| Black Sabbath | Paranoid | 534000 | 12 |
| Katy Perry | Prism | 800000 | 16 |
| Katy Perry | Teenage Dream | 750000 | 14 |

- a) Execute ex3.cql. Provide the content of this file as the result of this exercise.
- b) Execute the command 'SELECT * FROM Music;' and provide the output of this command as another result of the exercise.

Exercise 4) (2 points)

Now create a file in your working directory called ex4.cql. In this file write the commands to query only Katy Perry songs. Execute ex4.cql. Provide the content of this file and result of executing this file as the result of this exercise.

Exercise 5) (2 points)

Now create a file in your working directory called `ex5.cql`. In this file write the commands to query only albums that have sold 700000 copies or more. Execute `ex5.cql`. Provide the content of this file and the result of executing this file as the result of this exercise.