CS595—Big Data Technologies

Assignment #9

Worth: 5 points (1 point for each problem)

Due by the start of the next class period

Assignments should be uploaded via the Blackboard portal.

Readings

NoSQL Distilled: Chapters 8 and 10

Starting HBase

To start HBase first establish ssh tunneling as usual. Then access the Ambari Hadoop admin console by pointing your browser to localhost:8080. Log on Ambari with username and password maria_dev. In the upper right you should see a selection 'Services.' Click on it and choose 'HBase.' Then, from the HBase page choose 'Service Actions' and select start. Then log on to maria_dev and enter 'hbase shell.'

There are two documents about the HBase shell on the blackboard in the "Free Books and Chapters" section.

Exercises

Exercise 1)

Create an HBase table with the following characteristics

Table Name: cs595Tbl

First column family: cf1

Second column family: cf2

Then execute the DESCRIBE command on the table and return command you wrote and the output as the results of this exercise.

```
create 'cs595Tb1','cf1','cf2'
```

```
hbase(main):002:0> describe 'cs595Tb1'
Table cs595Tb1 is ENABLED
cs595Tb1
COLUMN FAMILIES DESCRIPTION
{NAME => 'cf1', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEE
P_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COM
PRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '655
36', REPLICATION_SCOPE => '0'}
{NAME => 'cf2', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEE
P_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COM
PRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '655
36', REPLICATION_SCOPE => '0'}
2 row(s) in 0.0890 seconds
```

Exercise 2)

Put the following data into the table created in exercise 1:

Row Key	Column Family	Column (Qualifier)	Value
Row1	cf1	name	Sam
Row2	cf1	name	Ahmed
Row1	cf2	job	Pilot
Row2	cf2	job	Doctor
Row1	cf2	level	LZ3
Row2	cf2	level	AR7

Execute the SCAN command on this table returning all rows, column families and columns as the result of the exercise.

```
hbase(main):004:0> put 'cs595Tb1','1','cf1:name','Sam'
hbase(main):005:0> put 'cs595Tb1','2','cf1:name','Ahmed'
hbase(main):006:0> put 'cs595Tb1','1','cf2:job','Pilot'
hbase(main):007:0> put 'cs595Tb1','2','cf2:job','Doctor'
hbase(main):008:0> put 'cs595Tb1','1','cf2:level','LZ3'
hbase(main):009:0> put 'cs595Tb1','2','cf2:level','AR7'
```

```
hbase(main):010:0> scan 'cs595Tb1'
ROW                   COLUMN+CELL
 1                    column=cf1:name, timestamp=1510086067993, value=Sam
 1                    column=cf2:job, timestamp=1510086129879, value=Pilot
 1                    column=cf2:level, timestamp=1510086175729, value=LZ3
 2                    column=cf1:name, timestamp=1510086100419, value=Ahmed
 2                    column=cf2:job, timestamp=1510086145672, value=Doctor
 2                    column=cf2:level, timestamp=1510086248734, value=AR7
2 row(s) in 0.0210 seconds
```

Exercise 3)

Using the above table write command that will get the value associated with row (Row1), column family (cf2) and column/qualifier (level). Provide the command and its result as the output of this exercise.

```
hbase(main):012:0> get 'cs595Tb1','1',{COLUMN => 'cf2:level'}
```

```
hbase(main):012:0> get 'cs595Tb1','1',{COLUMN => 'cf2:level'}
COLUMN                CELL
 cf2:level            timestamp=1510086175729, value=LZ3
1 row(s) in 0.0070 seconds
```

Exercise 4)

Using the above table write command that will get the value associated with row (Row2), column family (cf1) and column/qualifier (name). Provide the command and its result as the output of this exercise.

```
hbase(main):013:0> get 'cs595Tb1','2',{COLUMN => 'cf1:name'}
```

```
hbase(main):013:0> get 'cs595Tb1','2',{COLUMN => 'cf1:name'}
COLUMN                CELL
 cf1:name             timestamp=1510086100419, value=Ahmed
1 row(s) in 0.0060 seconds
```

Exercise 5)

Using the above table write a SCAN command that will return information about only one row using the LIMIT modifier. Provide the command and its result as the output of this exercise.

```
hbase(main):014:0> scan 'cs595Tb1',{LIMIT=>1}
```

```
hbase(main):014:0> scan 'cs595Tb1',{LIMIT=>1}
ROW                   COLUMN+CELL
 1                    column=cf1:name, timestamp=1510086067993, value=Sam
 1                    column=cf2:job, timestamp=1510086129879, value=Pilot
 1                    column=cf2:level, timestamp=1510086175729, value=LZ3
1 row(s) in 0.0440 seconds
```