CSP554—Big Data Technologies

Assignment #11

Worth: 5 points (1 point for each problem)

Exercises

Exercise 1)

Create an HBase table with the following characteristics

Table Name: csp554Tbl

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.

```
hbase(main):027:0> create 'csp554Tbl', {NAME => 'cf1'}, {NAME => 'cf2'}
0 row(s) in 1.2580 seconds

=> Hbase::Table - csp554Tbl
hbase(main):028:0> describe 'csp554Tbl'
Table csp554Tbl is ENABLED
csp554Tbl
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 0.0170 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. Provide the command and its result as the output of this exercise.

```
hbase(main):029:0> put 'csp554Tbl', 'Row1', 'cf1:name', 'Sam'
0 row(s) in 0.1190 seconds

hbase(main):030:0> put 'csp554Tbl', 'Row2', 'cf1:name', 'Ahmed'
0 row(s) in 0.0110 seconds

hbase(main):031:0> put 'csp554Tbl', 'Row1', 'cf2:job', 'Pilot'
0 row(s) in 0.0080 seconds

hbase(main):032:0> put 'csp554Tbl', 'Row2', 'cf2:job', 'Doctor'
0 row(s) in 0.0210 seconds

hbase(main):033:0> put 'csp554Tbl', 'Row1', 'cf2:level', 'LZ3'
0 row(s) in 0.0060 seconds

hbase(main):034:0> put 'csp554Tbl', 'Row2', 'cf2:level', 'AR7'
0 row(s) in 0.0050 seconds

hbase(main):035:0> scan 'csp554Tbl'
ROW                    COLUMN+CELL
 Row1                  column=cf1:name, timestamp=1618619433764, value=Sam
 Row1                  column=cf2:job, timestamp=1618619494799, value=Pilot
 Row1                  column=cf2:level, timestamp=1618619531046, value=LZ3
 Row2                  column=cf1:name, timestamp=1618619466609, value=Ahmed
 Row2                  column=cf2:job, timestamp=1618619510470, value=Doctor
 Row2                  column=cf2:level, timestamp=1618619544949, value=AR7
2 row(s) in 0.0270 seconds
```

Exercise 3)

Using the above table write a 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):036:0> get 'csp554Tbl', 'Row1', {COLUMN => ['cf2', 'cf2:level']}
COLUMN                    CELL
 cf2:level                timestamp=1618619531046, value=LZ3
1 row(s) in 0.0430 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):043:0> get 'csp554Tbl', 'Row2', {COLUMN => ['cf1', 'cf1:name']}
COLUMN                    CELL
 cf1:name                 timestamp=1618619466609, value=Ahmed
1 row(s) in 0.0130 seconds
```

Exercise 5)

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

```
hbase(main):011:0> scan 'csp554Tbl', {LIMIT => 2}
ROW                       COLUMN+CELL
 Row1                     column=cf1:name, timestamp=1618632887082, value=Sam
 Row1                     column=cf2:job, timestamp=1618632919431, value=Pilot
 Row1                     column=cf2:level, timestamp=1618632952640, value=LZ3
 Row2                     column=cf1:name, timestamp=1618632898886, value=Ahmed
 Row2                     column=cf2:job, timestamp=1618632931866, value=Doctor
 Row2                     column=cf2:level, timestamp=1618632970100, value=AR7
2 row(s) in 0.0190 seconds
```