

BIG DATA APPLICATION AND ANALYTICS PRACTICAL

Activity: HBase Table Creation

Name: Muhamad Rafiq Iqbal Bin Samsudin (17206926)

Objective: Create an HBase table with the following column families: personal, name, age, and gender.

Execute five HBase commands to manage the table.

1. Open your HBase shell or client.

The screenshot shows a terminal window titled 'cloudera@quickstart:~/Desktop'. The terminal output is as follows:

```
[cloudera@quickstart Desktop]$ sudo jps
5400 NameNode
6739 RunJar
5709 JobHistoryServer
5792 NodeManager
7978 Bootstrap
5315 JournalNode
8300
5530 SecondaryNameNode
6898 RunJar
6629 ThriftServer
5219 DataNode
23161 Jps
19045 Main
5147 QuorumPeerMain
5659 Bootstrap
7372 HistoryServer
6494 RESTServer
6091 ResourceManager
8216 Bootstrap
7342 Bootstrap
8253
[cloudera@quickstart Desktop]$ sudo service hbase-master start
starting master, logging to /var/log/hbase/hbase-master-quickstart.cloudera.out
Started HBase master daemon (hbase-master): [ OK ]
[cloudera@quickstart Desktop]$ sudo jps
5400 NameNode
23433 Jps
6739 RunJar
5709 JobHistoryServer
5792 NodeManager
7978 Bootstrap
5315 JournalNode
8300
5530 SecondaryNameNode
6898 RunJar
6629 ThriftServer
5219 DataNode
19045 Main
5147 QuorumPeerMain
5659 Bootstrap
7372 HistoryServer
6494 RESTServer
6091 ResourceManager
8216 Bootstrap
7342 Bootstrap
8253
23233 HMaster
[cloudera@quickstart Desktop]$ hbase shell
2024-10-27 22:55:45,273 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017
```

Annotations on the screenshot:

- A box labeled "Check the list of java running process" points to the `sudo jps` command.
- A box labeled "Starting hbase master" points to the `sudo service hbase-master start` command.
- A box labeled "Hmaster Running" points to the `HMaster` process in the second `sudo jps` output.
- A box labeled "Start the Hbase" points to the `hbase shell` command.

2. Create an HBase table named "student_info" with the specified column families: personal, name, age, and gender.

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017

hbase(main):001:0> list
TABLE
hbasetable123test
1 row(s) in 0.2200 seconds

=> ["hbasetable123test"]
hbase(main):002:0> create 'Studentinfo','Name','Age','Gender'
0 row(s) in 1.2280 seconds

=> Hbase::Table - Studentinfo
hbase(main):003:0> put 'Studentinfo','r1','Name:First','Muhamad'
0 row(s) in 0.1130 seconds

hbase(main):004:0> put 'Studentinfo','r1','Name:Middle','Rafiq'
0 row(s) in 0.0050 seconds

hbase(main):005:0> put 'Studentinfo','r1','Age:C1','24'
0 row(s) in 0.0050 seconds

hbase(main):006:0> put 'Studentinfo','r1','Gender:C1','Male'
0 row(s) in 0.0030 seconds

hbase(main):007:0> put 'Studentinfo','r2','Name:First','Ahmad'
0 row(s) in 0.0030 seconds

hbase(main):008:0> put 'Studentinfo','r2','Name:Middle','Daniel'
0 row(s) in 0.0050 seconds

hbase(main):009:0> put 'Studentinfo','r2','Gender:C1','Male'
0 row(s) in 0.0050 seconds

hbase(main):010:0> put 'Studentinfo','r2','Age:C1','25'
0 row(s) in 0.0030 seconds

hbase(main):011:0> scan 'Studentinfo'
ROW COLUMN+CELL
r1 column=Age:C1, timestamp=1730097591988, value=24
r1 column=Gender:C1, timestamp=1730097605978, value=Male
r1 column=Name:First, timestamp=1730097552350, value=Muhamad
r1 column=Name:Middle, timestamp=1730097571562, value=Rafiq
r2 column=Age:C1, timestamp=1730097672302, value=25
r2 column=Gender:C1, timestamp=1730097652941, value=Male
r2 column=Name:First, timestamp=1730097630493, value=Ahmad
r2 column=Name:Middle, timestamp=1730097643264, value=Daniel
2 row(s) in 0.0360 seconds

hbase(main):012:0> get 'Studentinfo','r1'
COLUMN CELL
Age:C1 timestamp=1730097591988, value=24
Gender:C1 timestamp=1730097605978, value=Male
Name:First timestamp=1730097552350, value=Muhamad
Name:Middle timestamp=1730097571562, value=Rafiq
4 row(s) in 0.0100 seconds

hbase(main):013:0> get 'Studentinfo','r2'
COLUMN CELL
Age:C1 timestamp=1730097672302, value=25
Gender:C1 timestamp=1730097652941, value=Male
Name:First timestamp=1730097630493, value=Ahmad
Name:Middle timestamp=1730097643264, value=Daniel
4 row(s) in 0.0100 seconds
```

1. Check the list of current existing table

2. Create table with Table name of 'Studentinfo' and Column families of 'Name','Age', and 'Gender'

3. Inputting data to every column

4. Display and retrieve data from the entire table

5. Display the specific row of student in R1

6. Display the specific row of student in R2

3. Update Records

- Update Age for both students.

```
hbase(main):014:0> put 'Studentinfo','r1','Age:C1','26'
0 row(s) in 0.0050 seconds

hbase(main):015:0> put 'Studentinfo','r2','Age:C1','26'
0 row(s) in 0.0040 seconds

hbase(main):016:0> get 'Studentinfo','r1'
COLUMN                                CELL
Age:C1                                timestamp=1730098264638, value=26
Gender:C1                             timestamp=1730097605978, value=Male
Name:First                            timestamp=1730097552350, value=Muhamad
Name:Middle                           timestamp=1730097571562, value=Rafiq
4 row(s) in 0.0290 seconds

hbase(main):017:0> get 'Studentinfo','r2'
COLUMN                                CELL
Age:C1                                timestamp=1730098273195, value=26
Gender:C1                             timestamp=1730097652941, value=Male
Name:First                            timestamp=1730097630493, value=Ahmad
Name:Middle                           timestamp=1730097643264, value=Daniel
4 row(s) in 0.0070 seconds
```

7. Update age using same command 'put'

8. Both age change to '26'

4. Delete Records

- For example, we want to delete the first name

```
hbase(main):018:0> delete 'Studentinfo','r1','Name:First'
0 row(s) in 0.0170 seconds

hbase(main):019:0> delete 'Studentinfo','r2','Name:First'
0 row(s) in 0.0020 seconds

hbase(main):020:0> get 'Studentinfo','r1'
COLUMN                                CELL
Age:C1                                timestamp=1730098264638, value=26
Gender:C1                             timestamp=1730097605978, value=Male
Name:Middle                           timestamp=1730097571562, value=Rafiq
3 row(s) in 0.0050 seconds

hbase(main):021:0> get 'Studentinfo','r2'
COLUMN                                CELL
Age:C1                                timestamp=1730098273195, value=26
Gender:C1                             timestamp=1730097652941, value=Male
Name:Middle                           timestamp=1730097643264, value=Daniel
3 row(s) in 0.0060 seconds
```

9. Delete 'First name' from the table

10. First name column no longer existed

5. Insert New Data

- Add 1 more student to the table.

```
hbase(main):022:0> put 'Studentinfo','r3','Name:Middle','Khairah'
0 row(s) in 0.0060 seconds

hbase(main):023:0> put 'Studentinfo','r3','Age:C1','24'
0 row(s) in 0.0030 seconds

hbase(main):024:0> put 'Studentinfo','r3','Gender:C1','Female'
0 row(s) in 0.0030 seconds

hbase(main):025:0> scan 'Studentinfo'
ROW      COLUMN+CELL
r1       column=Age:C1, timestamp=1730098264638, value=26
r1       column=Gender:C1, timestamp=1730097605978, value=Male
r1       column=Name:Middle, timestamp=1730097571562, value=Rafiq
r2       column=Age:C1, timestamp=1730098273195, value=26
r2       column=Gender:C1, timestamp=1730097652941, value=Male
r2       column=Name:Middle, timestamp=1730097643264, value=Daniel
r3       column=Age:C1, timestamp=1730098792879, value=24
r3       column=Gender:C1, timestamp=1730098812003, value=Female
r3       column=Name:Middle, timestamp=1730098780204, value=Khairah
3 row(s) in 0.0150 seconds
```

11. Repeat same command 'put' to add new data

12. Data added to the same table

6. Drop a table (Additional)

```
hbase(main):026:0> list
TABLE
Studentinfo
hbasetable123test
2 row(s) in 0.0120 seconds
```

13. List all the available table. In this case, there are 2 table which is 'Studentinfo' and 'hbasetable123test', We want to delete the second table

```
=> ["Studentinfo", "hbasetable123test"]
hbase(main):027:0> disable 'hbasetable123test'
0 row(s) in 2.2550 seconds
```

14. First disable the table

```
hbase(main):028:0> drop 'hbasetable123test'
0 row(s) in 1.2450 seconds
```

15. Delete table using 'drop' command

```
hbase(main):029:0> list
TABLE
Studentinfo
1 row(s) in 0.0030 seconds
```

16. 'hbasetable123test' table no longer existed