

Name: Vedant Jolly
Roll No: 2019130026
Class: BE COMPS
Batch: A

EXPERIMENT NO. 7 HBASE

AIM: Carry out CRUD operations using HBase.

Download HBase (v1.4.9): hbase-1.4.9-bin.tar.gz
(Link: <https://archive.apache.org/dist/hbase/1.4.9/>)
Extract the zipped file to your C:\ drive.

Environment Variables:

HBASE_HOME → C:\hbase-1.4.9
User Variables and System Variables → Path → C:\hbase-1.4.9\bin

Starting the nodes and yarn:

```
PS C:\WINDOWS\system32> cd F:\Hadoop\hadoop-3.2.2\sbin
```

```
PS F:\Hadoop\hadoop-3.2.2\sbin> .\start-all.cmd
```

This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

```
PS F:\Hadoop\hadoop-3.2.2\sbin> jps
```

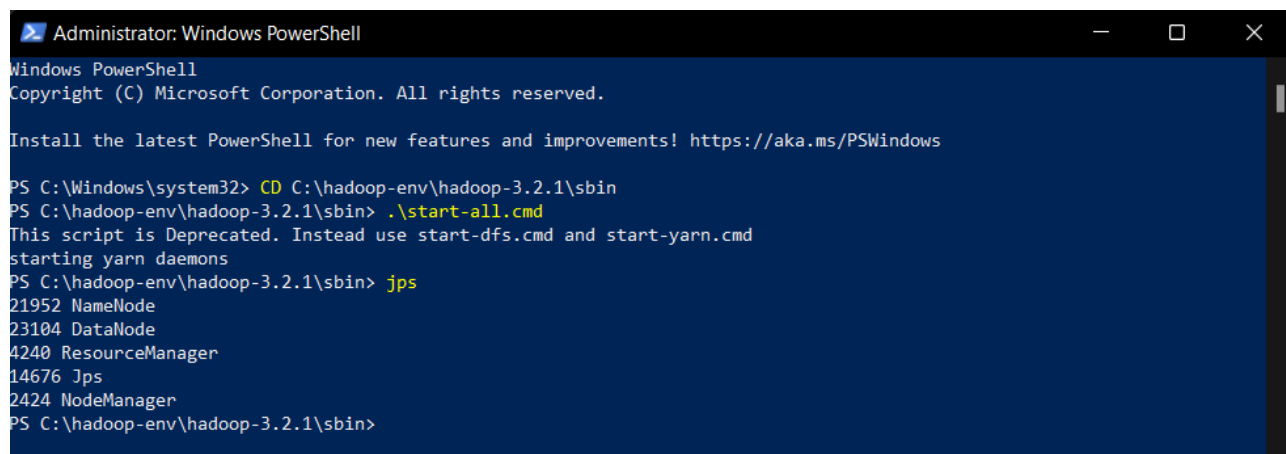
20804 DataNode

22296 NameNode

14892 ResourceManager

16204 Jps

17020 NodeManager



```
Administrator: Windows PowerShell
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Windows\system32> CD C:\hadoop-env\hadoop-3.2.1\sbin
PS C:\hadoop-env\hadoop-3.2.1\sbin> .\start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons
PS C:\hadoop-env\hadoop-3.2.1\sbin> jps
21952 NameNode
23104 DataNode
4240 ResourceManager
14676 Jps
2424 NodeManager
PS C:\hadoop-env\hadoop-3.2.1\sbin>
```

```
PS C:\WINDOWS\system32> cd C:\hbase-1.4.9\bin
```

```
PS C:\hbase-1.4.9\bin> hbase
```

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\vedan> cd C:\hbase\hbase-1.4.9\bin
PS C:\hbase\hbase-1.4.9\bin> hbase
Usage: hbase [<options>] [<command>] [<args>]
where <command> an option from one of these categories::
Options:
  --config DIR      Configuration direction to use. Default: ../conf

Commands:
Some commands take arguments. Pass no args or -h for usage."
  shell             Run the HBase shell
  hbck              Run the hbase 'fsck' tool
  wal              Write-ahead-log analyzer
  hfile            Store file analyzer
  zkcli            Run the ZooKeeper shell
  upgrade           Upgrade hbase
  master           Run an HBase HMaster node
  regionserver     Run an HBase HRegionServer node
  zookeeper        Run a Zookeeper server
  rest             Run an HBase REST server
  thrift           Run the HBase Thrift server
  thrift2          Run the HBase Thrift2 server
  classpath        Dump hbase CLASSPATH
  mapredcp         Dump CLASSPATH entries required by mapreduce
  version          Print the version
  CLASSNAME        Run the class named CLASSNAME
PS C:\hbase\hbase-1.4.9\bin> |
```

```
PS C:\hbase-1.4.9\bin> hbase version
```

```
PS C:\hbase-1.4.9\bin> hbase shell
```

```
Select Administrator: Windows PowerShell

PS C:\hbase-1.4.9\bin> hbase version
HBase 1.4.9
Source code repository git://apurtell-1tm4.internal.salesforce.com/Users/apurtell/src/hbase revision=d625b212e46d01cb17db9ac2e9e927fdb201afa1
Compiled by apurtell on Wed Dec  5 11:54:10 PST 2018
From source with checksum a7716fc1849b07ea6dd830a08291e754
PS C:\hbase-1.4.9\bin> hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/C:/hbase-1.4.9/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/F:/Hadoop/hadoop-3.2.2/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.4.9, rd625b212e46d01cb17db9ac2e9e927fdb201afa1, Wed Dec  5 11:54:10 PST 2018
```

CREATE:

It is used to create a table.

create '<table name>', '<column family>'

hbase(main):001:0> create 'student', 'personalinfo', 'academics'

LIST:

It is used to list all the tables.

hbase(main):002:0> list

```
hbase(main):001:0> create 'student', 'personalinfo', 'academics'
0 row(s) in 1.7670 seconds

=> Hbase::Table - student
hbase(main):002:0> list
TABLE

student

1 row(s) in 0.2230 seconds

=> ["student"]
```

DESCRIBE:

It gives the description of a table.

describe 'table name'

hbase(main):003:0> describe 'student'

```
Select Administrator: Windows PowerShell
hbase(main):003:0> describe 'student'
Table student is ENABLED

student, {TABLE_ATTRIBUTES => {METADATA => {'hbase.store.file-tracker.impl' => 'DEFAULT'}}}

COLUMN FAMILIES DESCRIPTION

{NAME => 'academics', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0', METADATA => {'INDEX_BLOCK_ENCODING' => 'NONE'}}

{NAME => 'personalinfo', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COMPRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0', METADATA => {'INDEX_BLOCK_ENCODING' => 'NONE'}}

2 row(s) in 0.1930 seconds
```

PUT:

It is used to put a cell value at a specified column in a specified row in a particular table. Using this command, we can insert rows into a table. Its syntax is as follows:

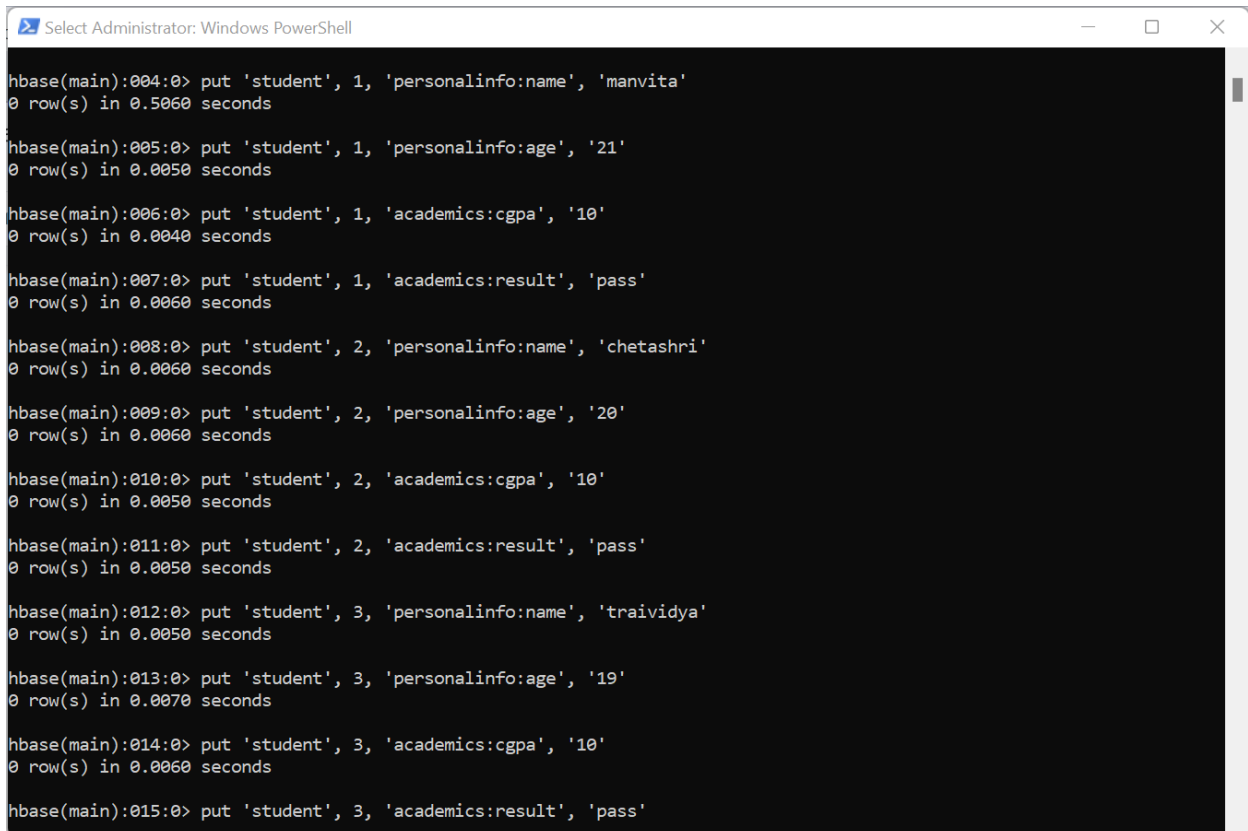
put '<table name>', 'row', '<columnfamily:columnname>', '<value>'

```
hbase(main):004:0> put 'student', 1, 'personalinfo:name', 'vedant'  
0 row(s) in 0.5060 seconds
```

```
hbase(main):005:0> put 'student', 1, 'personalinfo:age', '21'  
0 row(s) in 0.0050 seconds
```

```
hbase(main):006:0> put 'student', 1, 'academics:cgpa', '10'  
0 row(s) in 0.0040 seconds
```

```
hbase(main):007:0> put 'student', 1, 'academics:result', 'pass'  
0 row(s) in 0.0060 seconds
```



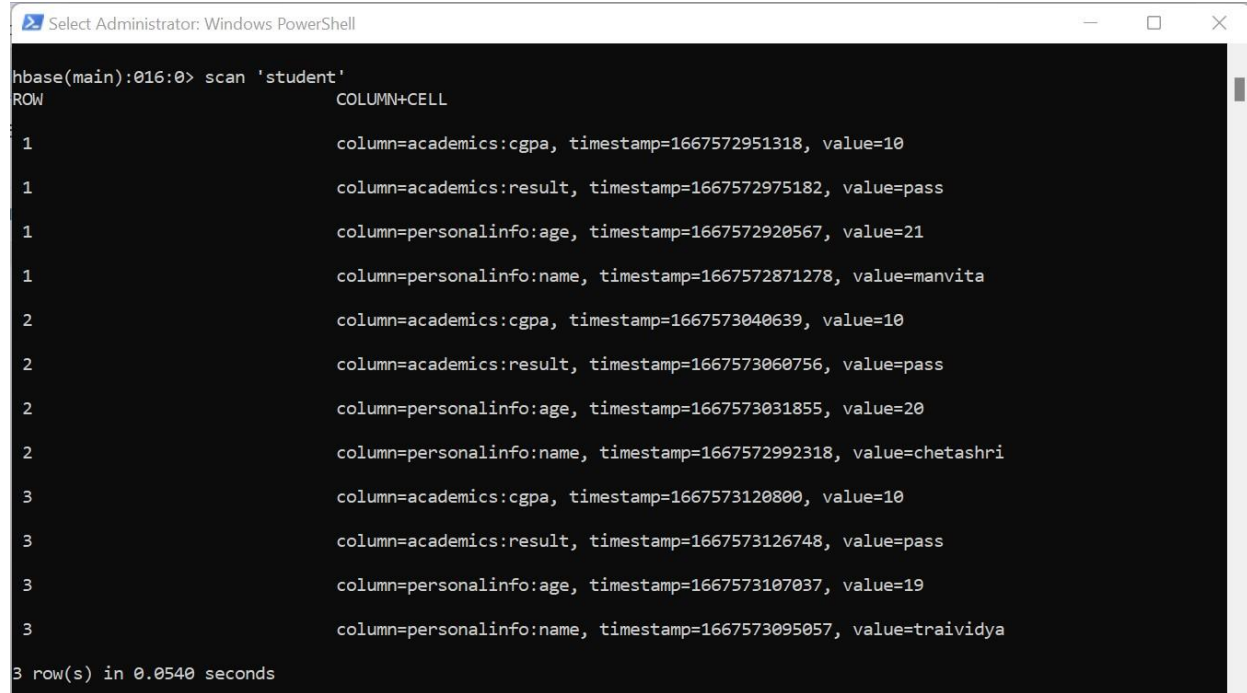
```
Select Administrator: Windows PowerShell  
  
hbase(main):004:0> put 'student', 1, 'personalinfo:name', 'manvita'  
0 row(s) in 0.5060 seconds  
  
hbase(main):005:0> put 'student', 1, 'personalinfo:age', '21'  
0 row(s) in 0.0050 seconds  
  
hbase(main):006:0> put 'student', 1, 'academics:cgpa', '10'  
0 row(s) in 0.0040 seconds  
  
hbase(main):007:0> put 'student', 1, 'academics:result', 'pass'  
0 row(s) in 0.0060 seconds  
  
hbase(main):008:0> put 'student', 2, 'personalinfo:name', 'chetashri'  
0 row(s) in 0.0060 seconds  
  
hbase(main):009:0> put 'student', 2, 'personalinfo:age', '20'  
0 row(s) in 0.0060 seconds  
  
hbase(main):010:0> put 'student', 2, 'academics:cgpa', '10'  
0 row(s) in 0.0050 seconds  
  
hbase(main):011:0> put 'student', 2, 'academics:result', 'pass'  
0 row(s) in 0.0050 seconds  
  
hbase(main):012:0> put 'student', 3, 'personalinfo:name', 'traividya'  
0 row(s) in 0.0050 seconds  
  
hbase(main):013:0> put 'student', 3, 'personalinfo:age', '19'  
0 row(s) in 0.0070 seconds  
  
hbase(main):014:0> put 'student', 3, 'academics:cgpa', '10'  
0 row(s) in 0.0060 seconds  
  
hbase(main):015:0> put 'student', 3, 'academics:result', 'pass'
```

SCAN:

It is used to scan and then return the table data.

scan '<table name>'

hbase(main):016:0> scan 'student'



```
Select Administrator: Windows PowerShell

hbase(main):016:0> scan 'student'
ROW          COLUMN+CELL
1            column=academics:cgpa, timestamp=1667572951318, value=10
1            column=academics:result, timestamp=1667572975182, value=pass
1            column=personalinfo:age, timestamp=1667572920567, value=21
1            column=personalinfo:name, timestamp=1667572871278, value=manvita
2            column=academics:cgpa, timestamp=1667573040639, value=10
2            column=academics:result, timestamp=1667573060756, value=pass
2            column=personalinfo:age, timestamp=1667573031855, value=20
2            column=personalinfo:name, timestamp=1667572992318, value=chetashri
3            column=academics:cgpa, timestamp=1667573120800, value=10
3            column=academics:result, timestamp=1667573126748, value=pass
3            column=personalinfo:age, timestamp=1667573107037, value=19
3            column=personalinfo:name, timestamp=1667573095057, value=traividya

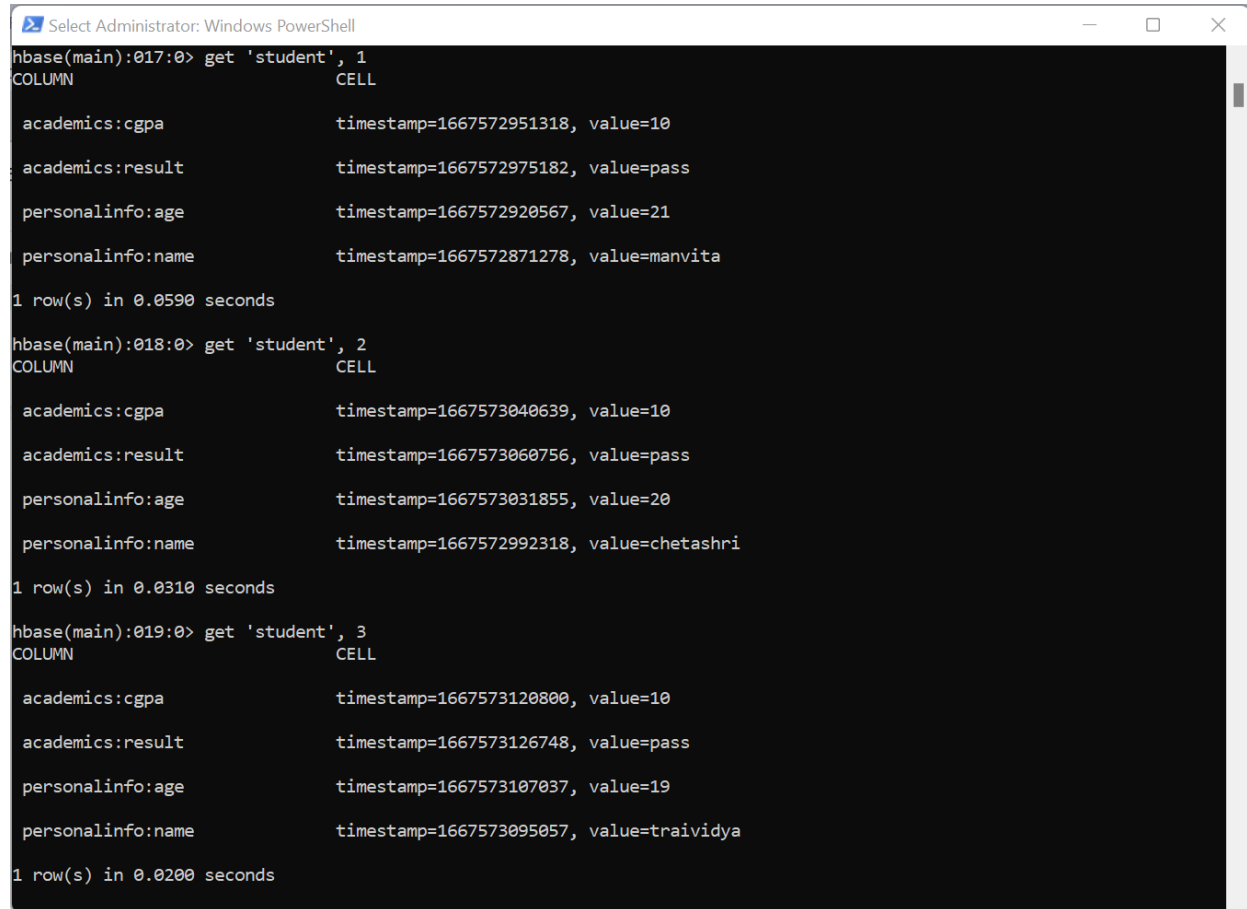
3 row(s) in 0.0540 seconds
```

GET (READ):

It is used to read the contents of row or a cell.

get '<table name>', 'row'

hbase(main):017:0> get 'student', 1



```
Select Administrator: Windows PowerShell

hbase(main):017:0> get 'student', 1
COLUMN                                CELL
academics:cgpa                        timestamp=1667572951318, value=10
academics:result                      timestamp=1667572975182, value=pass
personalinfo:age                     timestamp=1667572920567, value=21
personalinfo:name                    timestamp=1667572871278, value=manvita
1 row(s) in 0.0590 seconds

hbase(main):018:0> get 'student', 2
COLUMN                                CELL
academics:cgpa                        timestamp=1667573040639, value=10
academics:result                      timestamp=1667573060756, value=pass
personalinfo:age                     timestamp=1667573031855, value=20
personalinfo:name                    timestamp=1667572992318, value=chetashri
1 row(s) in 0.0310 seconds

hbase(main):019:0> get 'student', 3
COLUMN                                CELL
academics:cgpa                        timestamp=1667573120800, value=10
academics:result                      timestamp=1667573126748, value=pass
personalinfo:age                     timestamp=1667573107037, value=19
personalinfo:name                    timestamp=1667573095057, value=traividya
1 row(s) in 0.0200 seconds
```

UPDATE:

Execute the 'put' command again to overwrite the previous value.

Select Administrator: Windows PowerShell

```
hbase(main):019:0> get 'student', 3
COLUMN                                CELL

academics:cgpa                        timestamp=1667573120800, value=10
academics:result                      timestamp=1667573126748, value=pass
personalinfo:age                     timestamp=1667573107037, value=19
personalinfo:name                    timestamp=1667573095057, value=traividya

1 row(s) in 0.0200 seconds

hbase(main):020:0> put 'student', 3, 'personalinfo:age', '20'
0 row(s) in 0.0050 seconds

hbase(main):021:0> get 'student', 3
COLUMN                                CELL

academics:cgpa                        timestamp=1667573120800, value=10
academics:result                      timestamp=1667573126748, value=pass
personalinfo:age                     timestamp=1667573282599, value=20
personalinfo:name                    timestamp=1667573095057, value=traividya

1 row(s) in 0.0180 seconds
```

DELETE:

It is used to delete a cell value in a table.

delete '<table name>', '<row>', '<column name >'

hbase(main):023:0> delete 'student', 1, 'academics:result'

Select Administrator: Windows PowerShell

```
hbase(main):022:0> get 'student', 1
COLUMN                                CELL
academics:cgpa                        timestamp=1667572951318, value=10
academics:result                      timestamp=1667572975182, value=pass
personalinfo:age                      timestamp=1667572920567, value=21
personalinfo:name                    timestamp=1667572871278, value=manvita

1 row(s) in 0.0200 seconds

hbase(main):023:0> delete 'student', 1, 'academics:result'
0 row(s) in 0.0770 seconds

hbase(main):024:0> get 'student', 1
COLUMN                                CELL
academics:cgpa                        timestamp=1667572951318, value=10
personalinfo:age                      timestamp=1667572920567, value=21
personalinfo:name                    timestamp=1667572871278, value=manvita

1 row(s) in 0.0150 seconds
```



```
Select Administrator: Windows PowerShell
hbase(main):025:0> scan 'student'
ROW COLUMN+CELL
1      column=academics:cgpa, timestamp=1667572951318, value=10
1      column=personalinfo:age, timestamp=1667572920567, value=21
1      column=personalinfo:name, timestamp=1667572871278, value=manvita
2      column=academics:cgpa, timestamp=1667573040639, value=10
2      column=academics:result, timestamp=1667573060756, value=pass
2      column=personalinfo:age, timestamp=1667573031855, value=20
2      column=personalinfo:name, timestamp=1667572992318, value=chetashri
3      column=academics:cgpa, timestamp=1667573120800, value=10
3      column=academics:result, timestamp=1667573126748, value=pass
3      column=personalinfo:age, timestamp=1667573282599, value=20
3      column=personalinfo:name, timestamp=1667573095057, value=traividya
3 row(s) in 0.0630 seconds
```

CONCLUSION: In this experiment I learnt how to perform CRUD operations on HBase.