

EXPERIMENT 5

AIM: To execute HBase commands to load, insert, retrieve, update, or delete data.

THEORY:

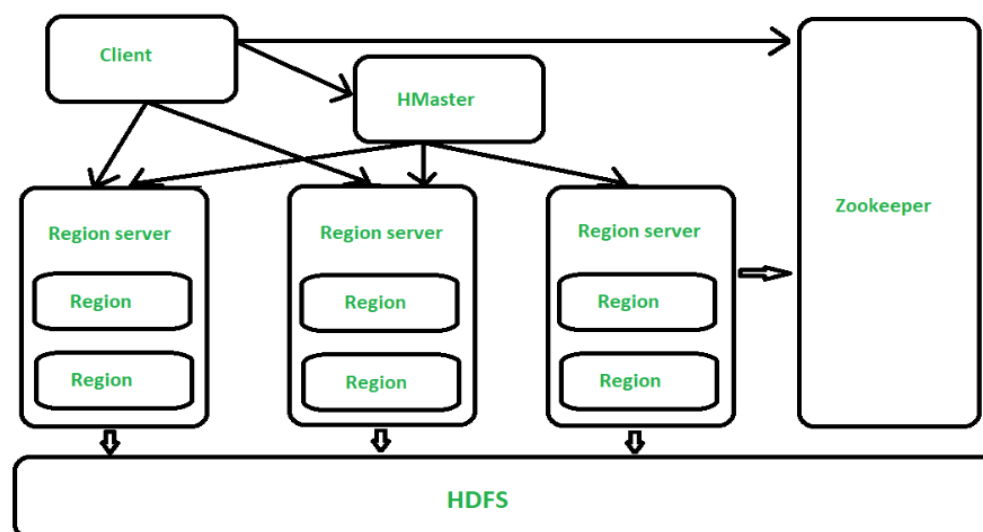
HBase is a data model that is similar to Google's big table. It is an open source, distributed database developed by *Apache* software foundation written in Java. HBase is an essential part of our Hadoop ecosystem. HBase runs on top of HDFS (Hadoop Distributed File System). It can store massive amounts of data from terabytes to petabytes. It is column oriented and horizontally scalable.

Features of HBase:

1. It is linearly scalable across various nodes as well as modularly scalable, as it divided across various nodes.
2. HBase provides consistent read and writes.
3. It provides atomic read and write means during one read or write process, all other processes are prevented from performing any read or write operations.
4. It provides easy to use Java API for client access.
5. It supports Thrift and REST API for non-Java front ends which supports XML, Protobuf and binary data encoding options.
6. It supports a Block Cache and Bloom Filters for real-time queries and for high volume query optimization.
7. HBase provides automatic failure support between Region Servers.
8. It support for exporting metrics with the Hadoop metrics subsystem to files.
9. It doesn't enforce relationship within your data.
10. It is a platform for storing and retrieving data with random access

HBase architecture has 3 main components:

- HMaster
- Region Server
- Zookeeper.



Architecture of HBase :

All the 3 components are described below:

1. HMaster:

The implementation of Master Server in HBase is HMaster. It is a process in which regions are assigned to region server as well as DDL (create, delete table) operations. It monitor all Region Server instances present in the cluster. In a distributed environment, Master runs several background threads. HMaster has many features like controlling load balancing, failover etc.

2. Region Server:

HBase Tables are divided horizontally by row key range into Regions. Regions are the basic building elements of HBase cluster that consists of the distribution of tables and are comprised of Column families. Region Server runs on HDFS DataNode which is present in Hadoop cluster. Regions of Region Server are responsible for several things, like handling, managing, executing as well as reads and writes HBase operations on that set of regions. The default size of a region is 256 MB.

3. Zookeeper:

It is like a coordinator in HBase. It provides services like maintaining configuration information, naming, providing distributed synchronization, server failure notification etc. Clients communicate with region servers via zookeeper.

Advantages of HBase :

1. Can store large data sets
2. Database can be shared
3. Cost-effective from gigabytes to petabytes
4. High availability through failover and replication

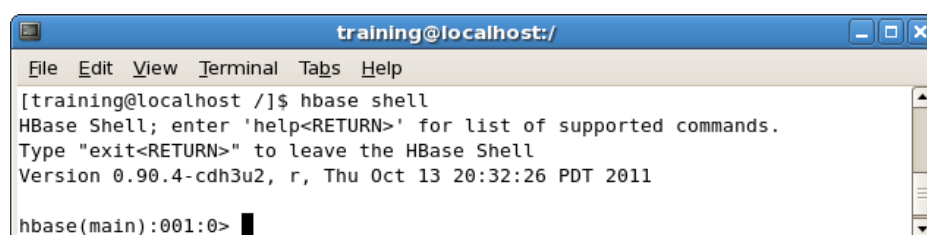
Disadvantages of HBase:

1. No support SQL structure
2. No transaction support
3. Sorted only on key
4. Memory issues on the cluster

IMPLEMENTATION:

1) start

hbase shell

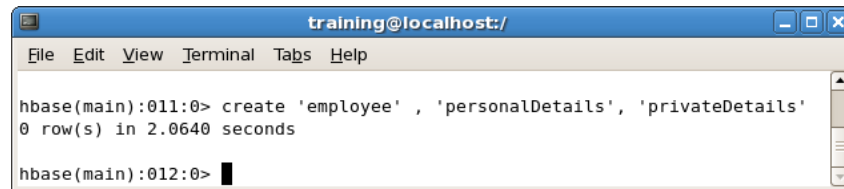


```
training@localhost:/  
File Edit View Terminal Tabs Help  
[training@localhost ~]$ hbase shell  
HBase Shell; enter 'help<RETURN>' for list of supported commands.  
Type "exit<RETURN>" to leave the HBase Shell  
Version 0.90.4-cdh3u2, r, Thu Oct 13 20:32:26 PDT 2011  
hbase(main):001:0>
```

2) Create

Syntax: create <tablename>, <columnfamilyname>

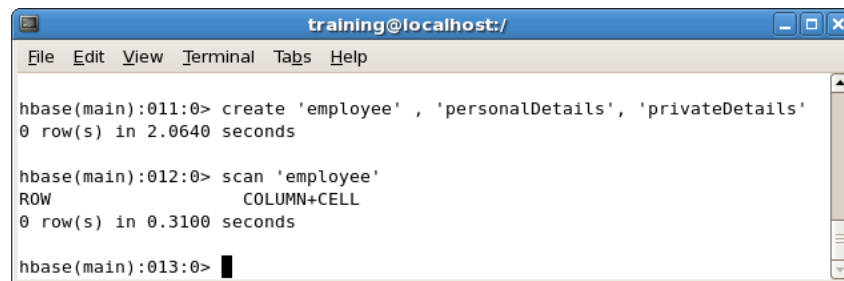
create 'employee' , 'personalDetails', 'privateDetails'



```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):011:0> create 'employee' , 'personalDetails', 'privateDetails'  
0 row(s) in 2.0640 seconds  
  
hbase(main):012:0> █
```

3) scan

scan 'employee'

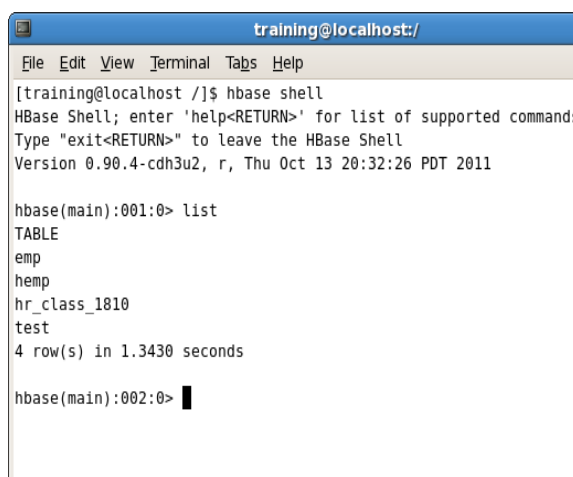


```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):011:0> create 'employee' , 'personalDetails', 'privateDetails'  
0 row(s) in 2.0640 seconds  
  
hbase(main):012:0> scan 'employee'  
ROW          COLUMN+CELL  
0 row(s) in 0.3100 seconds  
  
hbase(main):013:0> █
```

4) check for table

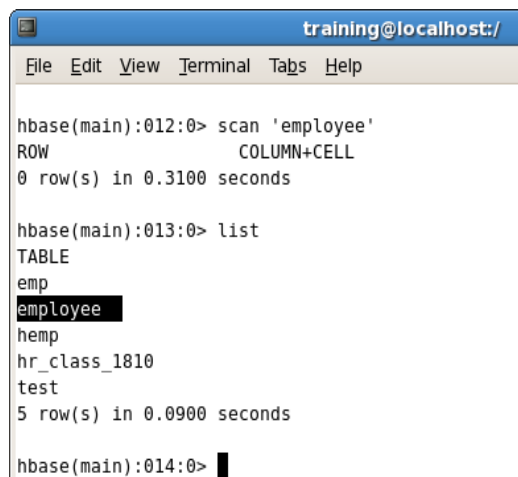
Syntax: list

list



```
training@localhost:/  
File Edit View Terminal Tabs Help  
[training@localhost ~]$ hbase shell  
HBase Shell; enter 'help<RETURN>' for list of supported commands  
Type "exit<RETURN>" to leave the HBase Shell  
Version 0.90.4-cdh3u2, r, Thu Oct 13 20:32:26 PDT 2011  
  
hbase(main):001:0> list  
TABLE  
emp  
hemp  
hr_class_1810  
test  
4 row(s) in 1.3430 seconds  
  
hbase(main):002:0> █
```

Before

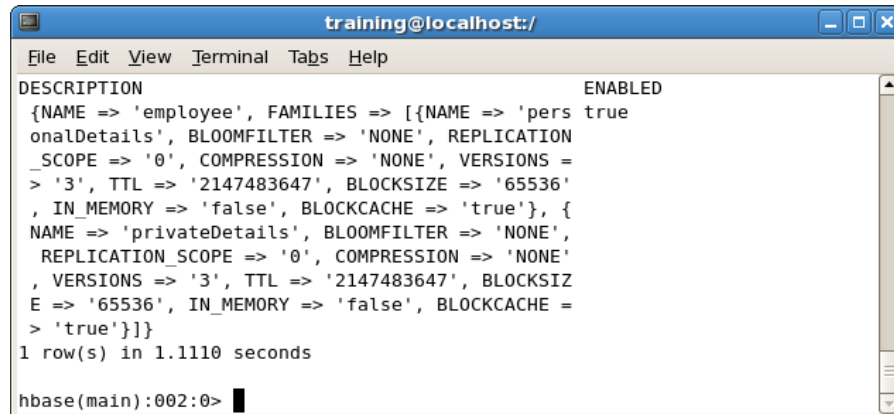


```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):012:0> scan 'employee'  
ROW          COLUMN+CELL  
0 row(s) in 0.3100 seconds  
  
hbase(main):013:0> list  
TABLE  
emp  
employee  
hemp  
hr_class_1810  
test  
5 row(s) in 0.0900 seconds  
  
hbase(main):014:0> █
```

After

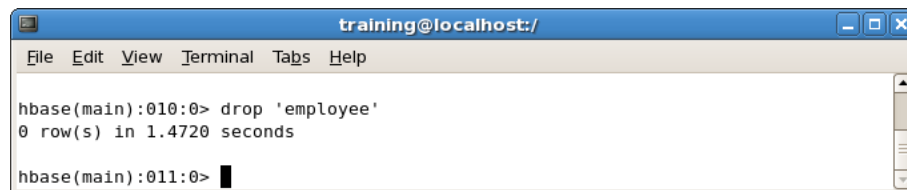
5) Syntax:describe <table name>

describe 'employee'



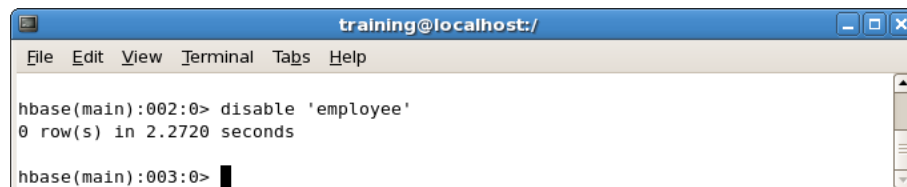
```
training@localhost:/  
File Edit View Terminal Tabs Help  
DESCRIPTION  
ENABLED  
{NAME => 'employee', FAMILIES => [{NAME => 'pers true  
onalDetails', BLOOMFILTER => 'NONE', REPLICATION  
_SCOPE => '0', COMPRESSION => 'NONE', VERSIONS =  
> '3', TTL => '2147483647', BLOCKSIZE => '65536'  
, IN_MEMORY => 'false', BLOCKCACHE => 'true'}, {  
NAME => 'privateDetails', BLOOMFILTER => 'NONE',  
REPLICATION_SCOPE => '0', COMPRESSION => 'NONE'  
, VERSIONS => '3', TTL => '2147483647', BLOCKSIZ  
E => '65536', IN_MEMORY => 'false', BLOCKCACHE =  
> 'true'}}}  
1 row(s) in 1.1110 seconds  
hbase(main):002:0>
```

6) Syntax:drop <table name>



```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):010:0> drop 'employee'  
0 row(s) in 1.4720 seconds  
hbase(main):011:0>
```

7) disable 'employee'



```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):002:0> disable 'employee'  
0 row(s) in 2.2720 seconds  
hbase(main):003:0>
```

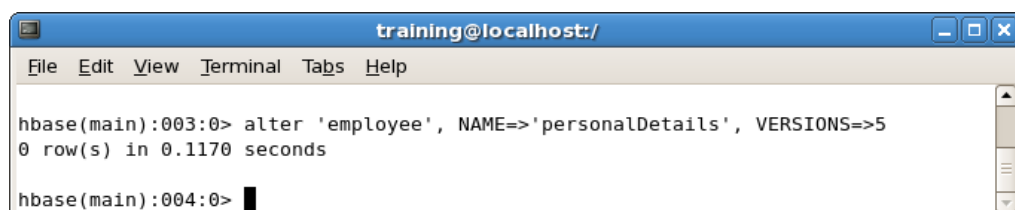
8) Alter

Syntax: alter <tablename>, NAME=><column familyname>, VERSIONS=>5

I] Change the max cell number.

i) single column family

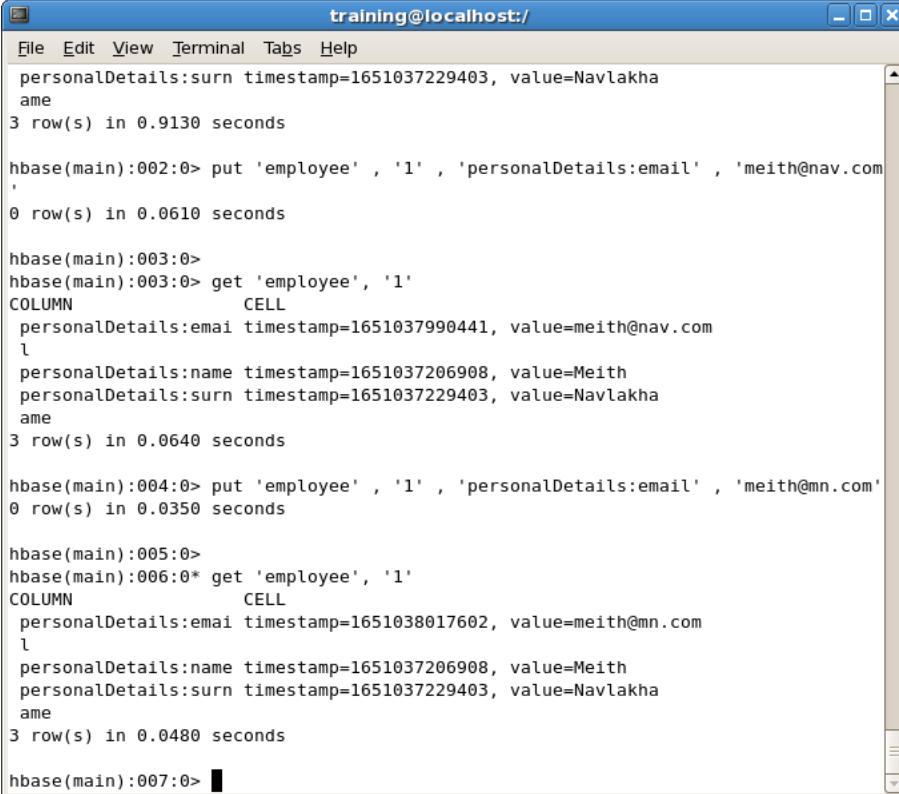
alter 'employee', NAME=>'personalDetails', VERSIONS=>5



```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):003:0> alter 'employee', NAME=>'personalDetails', VERSIONS=>5  
0 row(s) in 0.1170 seconds  
hbase(main):004:0>
```

II] Update readonly


alter 'employee', READONLY



```
training@localhost:/  
File Edit View Terminal Tabs Help  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.9130 seconds  
  
hbase(main):002:0> put 'employee' , '1' , 'personalDetails:email' , 'meith@nav.com'  
'  
0 row(s) in 0.0610 seconds  
  
hbase(main):003:0>  
hbase(main):003:0> get 'employee', '1'  
COLUMN CELL  
personalDetails:emai timestamp=1651037990441, value=meith@nav.com  
l  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.0640 seconds  
  
hbase(main):004:0> put 'employee' , '1' , 'personalDetails:email' , 'meith@mn.com'  
0 row(s) in 0.0350 seconds  
  
hbase(main):005:0>  
hbase(main):006:0* get 'employee', '1'  
COLUMN CELL  
personalDetails:emai timestamp=1651038017602, value=meith@mn.com  
l  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.0480 seconds  
  
hbase(main):007:0> █
```

III] Delete Col Family

alter 'employee', 'delete'=>'personalDetails'



```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):005:0> alter 'employee', 'delete'=>'personalDetails'  
0 row(s) in 0.1010 seconds  
  
hbase(main):006:0> █
```

9) Insert

put 'employee' , '1' , 'personalDetails:name' , 'Meith'

put 'employee' , '2' , 'personalDetails:name' , 'Mukesh'

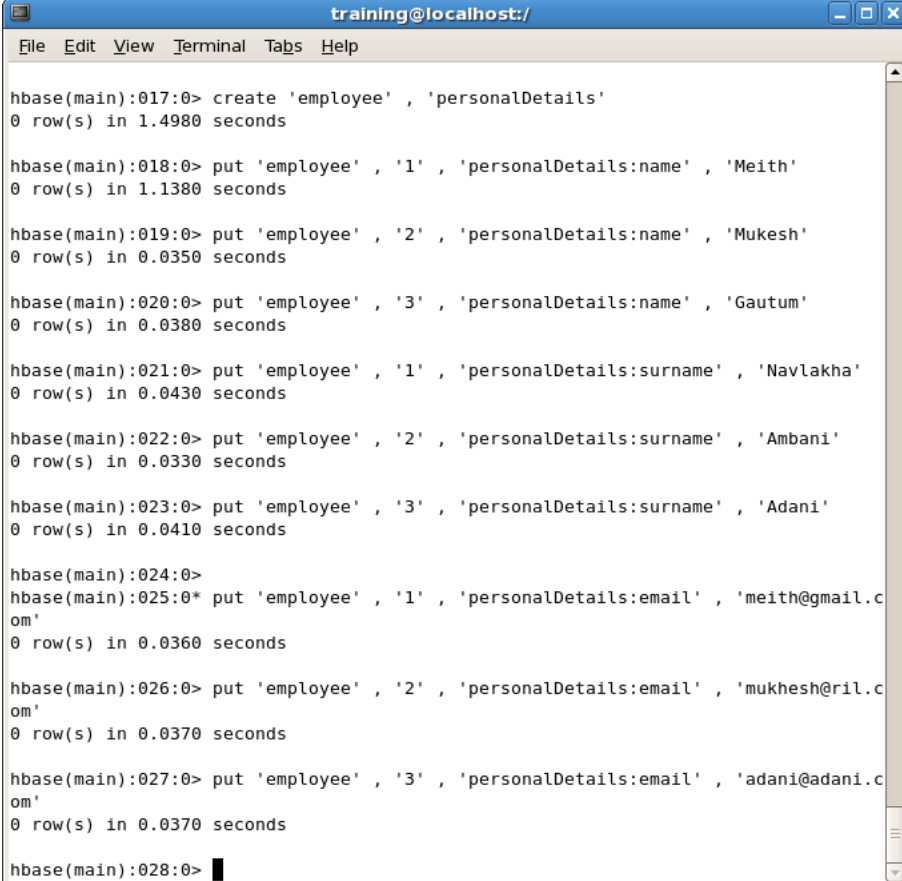
put 'employee' , '3' , 'personalDetails:name' , 'Gautum'

put 'employee' , '1' , 'personalDetails:surname' , 'Navlakha'

put 'employee' , '2' , 'personalDetails:surname' , 'Ambani'

put 'employee' , '3' , 'personalDetails:surname' , 'Adani'

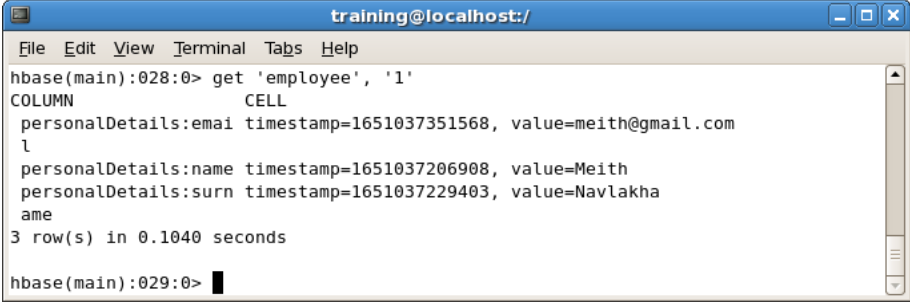
```
put 'employee' , '1' , 'personalDetails:email' , 'meith@gmail.com'  
put 'employee' , '2' , 'personalDetails:email' , 'mukhesh@ril.com'  
put 'employee' , '3' , 'personalDetails:email' , 'adani@adani.com'
```



```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):017:0> create 'employee' , 'personalDetails'  
0 row(s) in 1.4980 seconds  
  
hbase(main):018:0> put 'employee' , '1' , 'personalDetails:name' , 'Meith'  
0 row(s) in 1.1380 seconds  
  
hbase(main):019:0> put 'employee' , '2' , 'personalDetails:name' , 'Mukesh'  
0 row(s) in 0.0350 seconds  
  
hbase(main):020:0> put 'employee' , '3' , 'personalDetails:name' , 'Gautum'  
0 row(s) in 0.0380 seconds  
  
hbase(main):021:0> put 'employee' , '1' , 'personalDetails:surname' , 'Navlakha'  
0 row(s) in 0.0430 seconds  
  
hbase(main):022:0> put 'employee' , '2' , 'personalDetails:surname' , 'Ambani'  
0 row(s) in 0.0330 seconds  
  
hbase(main):023:0> put 'employee' , '3' , 'personalDetails:surname' , 'Adani'  
0 row(s) in 0.0410 seconds  
  
hbase(main):024:0>  
hbase(main):025:0* put 'employee' , '1' , 'personalDetails:email' , 'meith@gmail.c  
om'  
0 row(s) in 0.0360 seconds  
  
hbase(main):026:0> put 'employee' , '2' , 'personalDetails:email' , 'mukhesh@ril.c  
om'  
0 row(s) in 0.0370 seconds  
  
hbase(main):027:0> put 'employee' , '3' , 'personalDetails:email' , 'adani@adani.c  
om'  
0 row(s) in 0.0370 seconds  
  
hbase(main):028:0>
```

10) get data

```
get 'employee', '1'
```



```
training@localhost:/  
File Edit View Terminal Tabs Help  
  
hbase(main):028:0> get 'employee', '1'  
COLUMN CELL  
personalDetails:email timestamp=1651037351568, value=meith@gmail.com  
1  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.1040 seconds  
  
hbase(main):029:0>
```

get 'employee', '1' , {COLUMN => 'personalDetails:name'}

```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):009:0> get 'employee', '1' , {COLUMN => 'personalDetails:name'}  
COLUMN          CELL  
personalDetails:name timestamp=1651037206908, value=Meith  
1 row(s) in 0.0460 seconds  
  
hbase(main):010:0> █
```

11) update

put 'employee', '1' , 'personalDetails:email' , 'meith@mn.com'

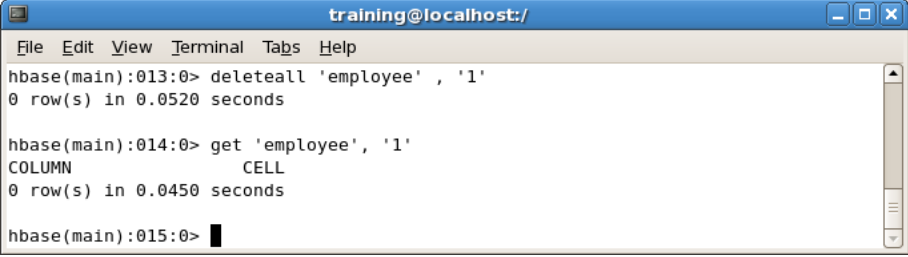
```
training@localhost:/  
File Edit View Terminal Tabs Help  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.9130 seconds  
  
hbase(main):002:0> put 'employee' , '1' , 'personalDetails:email' , 'meith@nav.com'  
'  
0 row(s) in 0.0610 seconds  
  
hbase(main):003:0>  
hbase(main):003:0> get 'employee', '1'  
COLUMN          CELL  
personalDetails:emai timestamp=1651037990441, value=meith@nav.com  
l  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.0640 seconds  
  
hbase(main):004:0> put 'employee' , '1' , 'personalDetails:email' , 'meith@mn.com'  
0 row(s) in 0.0350 seconds  
  
hbase(main):005:0>  
hbase(main):006:0* get 'employee', '1'  
COLUMN          CELL  
personalDetails:emai timestamp=1651038017602, value=meith@mn.com  
l  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
3 row(s) in 0.0480 seconds  
  
hbase(main):007:0> █
```

12) Delete

delete 'employee' , '1' , 'personalDetails:email'

```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):010:0> delete 'employee' , '1' , 'personalDetails:email'  
0 row(s) in 0.0580 seconds  
  
hbase(main):011:0> get 'employee', '1'  
COLUMN          CELL  
personalDetails:name timestamp=1651037206908, value=Meith  
personalDetails:surn timestamp=1651037229403, value=Navlakha  
ame  
2 row(s) in 0.0570 seconds
```

deleteall 'employee' , '1'



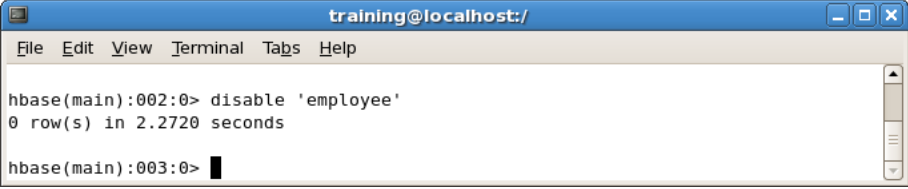
```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):013:0> deleteall 'employee' , '1'  
0 row(s) in 0.0520 seconds  
  
hbase(main):014:0> get 'employee', '1'  
COLUMN          CELL  
0 row(s) in 0.0450 seconds  
hbase(main):015:0>
```

13) Count No. of Rows

count <'tablename'>, CACHE =>1000

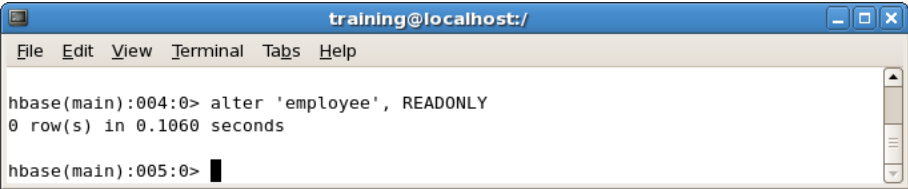
count 'employee', CACHE =>1000

14) disable



```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):002:0> disable 'employee'  
0 row(s) in 2.2720 seconds  
hbase(main):003:0>
```

15) readonly



```
training@localhost:/  
File Edit View Terminal Tabs Help  
hbase(main):004:0> alter 'employee', READONLY  
0 row(s) in 0.1060 seconds  
hbase(main):005:0>
```

CONCLUSION:

In this experiment, I explored and learnt about HBase, a database build on top of HDFS for faster execution of data over distributed system. We first learnt about what is HBase and it's application followed by executing it's queries on Cloudera platform. The output is observed and attached. Thus, by the virtue of the experiment, we have successfully understood and implemented the HBase commands.