

HBASE (Module – 5)

- HBase is a distributed column-oriented database built on top of the Hadoop file system.
- HBase sits on top of the Hadoop File System and provides read and write access.
- HBase is used whenever we need to provide fast random access to available data.

Storage Mechanism in HBase:

- Table is a collection of rows.
- Row is a collection of column families.
- Column family is a collection of columns.
- Column is a collection of key value pairs.

Given below is an example schema of table in HBase.

Rowid	Column Family			Column Family			Column Family			Column Family		
	col1	col2	col3	col1	col2	col3	col1	col2	col3	col1	col2	col3
1												
2												
3												

- Syntax to create table

create '<table name>', '<column family1>', '<column family2>',...
- Syntax to insert data

put '<table name>', 'row1', '<column family:colname>', '<value>'

Check if HBase is in good health or not through 'cloudera manager'

To check if HBase is active or not through terminal

```
[cloudera@localhost ~]$ sudo jps
```

```
7666 HRegionServer
```

```
7676 HMaster
```

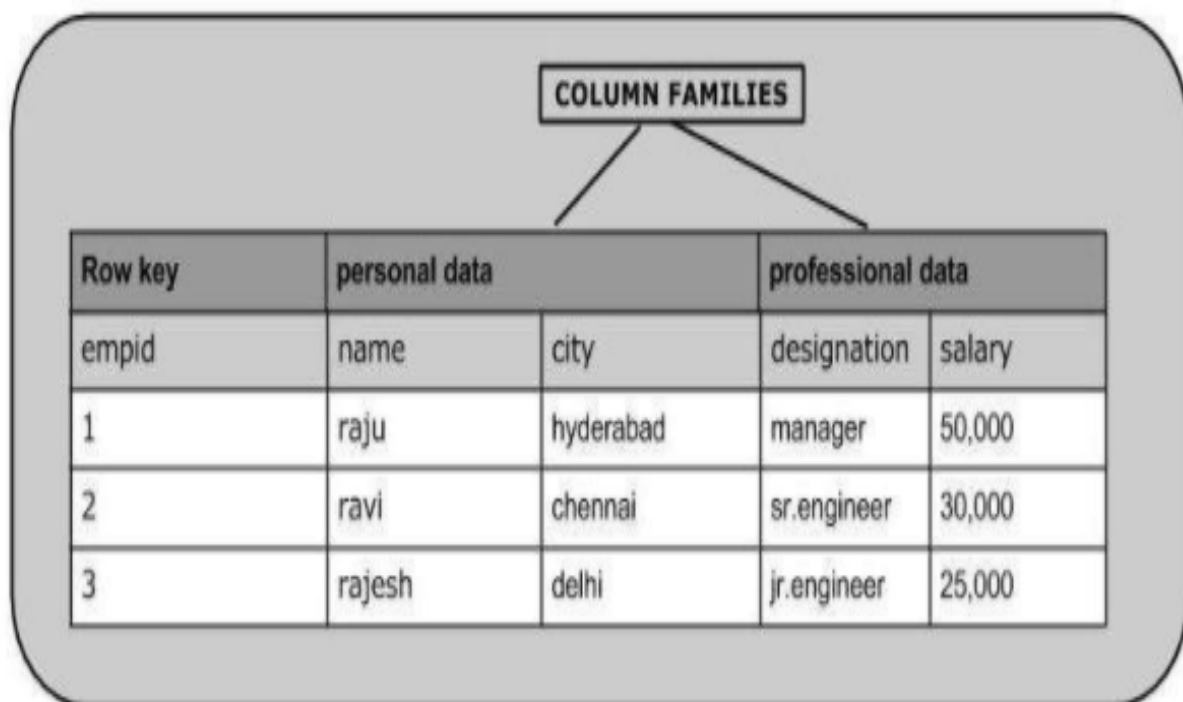
To start HBase

```
[cloudera@localhost ~]$ hbase shell
```

To show the HBase version

```
hbase(main):002:0> version
```

The following image shows column families in a column-oriented database:



To create a table

```
hbase(main):002:0> create 'emp1','personal data','professional data'
```

```
hbase(main):001:0> describe 'emp1'
```

To display list of tables

```
hbase(main):002:0> list
```

To insert data into the table

First row:

```
hbase(main):007:0> put 'emp1','row1','personal data:name','raju'
```

```
hbase(main):008:0> put 'emp1','row1','personal data:city','hyd'
```

```
hbase(main):009:0> put 'emp1','row1','professional data:designation','manager'
```

```
hbase(main):010:0> put 'emp1','row1','professional data:salary','50000'
```

To display table contents

```
hbase(main):011:0> scan 'emp1'
```

Second row:

```
hbase(main):012:0> put 'emp1','row2','personal data:name','ravi'
```

```
hbase(main):013:0> put 'emp1','row2','personal data:city','chennai'
```

```
hbase(main):017:0> put 'emp1','row2','professional data:designation','sr. engineer'
```

```
hbase(main):018:0> put 'emp1','row2','professional data:salary','30000'
```

To display table contents

```
hbase(main):011:0> scan 'emp1'
```

Third row:

```
hbase(main):012:0> put 'emp1','row3','personal data:name','rajesh'
```

```
hbase(main):013:0> put 'emp1','row3','personal data:city','delhi'
```

```
hbase(main):017:0> put 'emp1','row3','professional data:designation','jr. engineer'
```

```
hbase(main):018:0> put 'emp1','row3','professional data:salary','25000'
```

To display table contents

```
hbase(main):011:0> scan 'emp1'
```

To read data of a particular row or specific column

```
hbase(main):027:0> get 'emp1','row3'
```

```
hbase(main):004:0> get 'emp1','row3',{COLUMN=>'personal data:city'}
```

To update the existing data

```
hbase(main):008:0> put 'emp1','row3','personal data:city','New Delhi'
```

To disable the table

```
hbase(main):024:0> disable 'emp1'
```

Note: After disable, list will display the table name but scan command will show error

```
hbase(main):020:0> list
```

```
hbase(main):027:0> scan 'emp1'
```

```
hbase(main):020:0> enable 'emp1'
```

```
hbase(main):027:0> scan 'emp1'
```

To alter the table (adding new column to the table)

```
hbase(main):017:0> disable 'emp1'  
hbase(main):018:0> alter 'emp1',NAME=>'Employee history'  
hbase(main):020:0> enable 'emp1'  
hbase(main):022:0> describe 'emp1'
```

To delete a particular column family

```
hbase(main):037:0> disable 'emp1'  
hbase(main):038:0> alter 'emp1','delete'=>'Employee history'  
hbase(main):039:0> enable 'emp1'  
hbase(main):022:0> describe 'emp1'
```

To delete a specific cell in a table

```
hbase(main):045:0> delete 'emp1','row3','personal data:name'  
hbase(main):046:0> get 'emp1','row3'
```

To delete complete row in a table

```
hbase(main):049:0> deleteall 'emp1','row3'  
hbase(main):050:0> get 'emp1','row3'
```

To drop the table

```
hbase(main):024:0> disable 'emp1'  
hbase(main):001:0> drop 'emp1'  
hbase(main):002:0> list
```

MODULE 5

HBASE

TO START HBASE

```
[cloudera@localhost ~]$ hbase shell
```

GENERAL SHELL COMMANDS

```
hbase(main):001:0> status
```

```
1 servers, 0 dead, 2.0000 average load
```

```
hbase(main):002:0> version
```

```
0.94.6-cdh4.4.0, rUnknown, Tue Sep 3 20:09:51 PDT 2013
```

```
hbase(main):003:0> whoami
```

```
cloudera (auth:SIMPLE)
```

```
hbase(main):004:0> table_help
```

```
Help for table-reference commands.
```

EXP 25: HBase- CREATE TABLE & VERIFY

```
hbase(main):007:0> create 'emp','personal data','professional data'
```

```
hbase(main):008:0> describe 'emp'
```

To Verify table

```
hbase(main):008:0> list
```

```
TABLE
```

```
emp
```

```
hbase(main):014:0> exists 'emp'
```

```
Table emp does exist
```

EXP 26: HBase- STORE DATA IN TABLE

Insertion of data to first row

```
hbase(main):009:0> put 'emp','row1','personal data:name','raju'
```

```
hbase(main):010:0> put 'emp','row1','personal data:city','hyd'
```

```
hbase(main):011:0> put 'emp','row1','professional data:designation','manager'
```

```
hbase(main):012:0> put 'emp','row1','professional data:salary','50000'
```

```
hbase(main):013:0> scan 'emp'
```

```
ROW COLUMN+CELL
```

```
row1 column=personal data:city, timestamp=1523425226089, value=hyd
```

```
row1 column=personal data:name, timestamp=1523425200247, value=raju
```

```
row1 column=professional data:designation, timestamp=1523425255478,
```

```
value=manager
```

```
row1 column=professional data:salary, timestamp=1523425270378,
```

```
value=50000
```

Insertion of data to second row

```
hbase(main):013:0> put 'emp','row2','personal data:name','milind'
```

```
hbase(main):018:0> put 'emp','row2','personal data:city','chennai'
```

```
hbase(main):020:0> put 'emp','row2','professional data:designation','soft Engineer'
hbase(main):015:0> put 'emp','row2','professional data:salary','30000'
```

Insertion of data to third row

```
hbase(main):022:0> put 'emp','row3','personal data:name','anita'
hbase(main):023:0> put 'emp','row3','personal data:city','delhi'
hbase(main):024:0> put 'emp','row3','professional data:designation','tester'
hbase(main):025:0> put 'emp','row3','professional data:salary','40000'
```

EXP 27: TO READ THE DATA WHICH HAS BEEN ENTERED

➤ SCAN command reads all rows of that table

```
hbase(main):026:0> scan 'emp'
ROW                                COLUMN+CELL
row1                               column=personal data:city, timestamp=1523426980414, value=hyd
row1                               column=personal data:name, timestamp=1523426797711, value=raju
row1                               column=professional data:designation, timestamp=1523426835511,
value=manager
row1                               column=professional data:salary, timestamp=1523426852045,
value=50000
row2                               column=personal data:city, timestamp=1523427018284,
value=chennai
row2                               column=personal data:name, timestamp=1523426910414,
value=milind
row2                               column=professional data:designation, timestamp=1523427061723,
value=soft Engineer
row2                               column=professional data:salary, timestamp=1523426965356,
value=30000
row3                               column=personal data:city, timestamp=1523427129733, value=delhi
row3                               column=personal data:name, timestamp=1523427115249, value=anita
row3                               column=professional data:designation, timestamp=1523427162000,
value=tester
row3                               column=professional data:salary, timestamp=1523427187838,
value=40000
3 row(s) in 0.0190 seconds
```

EXP 28: GET command is used to read data of a particular row or specific column

```
hbase(main):027:0> get 'emp','row3'
COLUMN                                CELL
personal data:city                    timestamp=1523427129733, value=delhi
personal data:name                    timestamp=1523427115249, value=anita
professional data:designation          timestamp=1523427162000, value=tester
```


professional data:salary timestamp=1523427187838, value=40000

hbase(main):004:0> get 'emp','row3',{COLUMN=>'personal data:city'}

COLUMN	CELL
personal data:city	timestamp=1523427129733, value=delhi

NOTE!!! COLUMN must be in CAPITAL letters

hbase(main):020:0> list

TABLE
emp
example
expert

TO UPDATE/MODIFY THE EXISTING DATA

Updating the row3 personal data column family's city column with different values

hbase(main):008:0> put 'emp','row3','personal data:city','2bang'

hbase(main):009:0> put 'emp','row3','personal data:city','3kochi'

hbase(main):010:0> get 'emp','row3',{COLUMN=>'personal data:city',VERSIONS=>3}

COLUMN	CELL
personal data:city	timestamp=1523428351922, value=3kochi
personal data:city	timestamp=1523428337355, value=2bang
personal data:city	timestamp=1523427129733, value=delhi

hbase(main):011:0> put 'emp','row3','personal data:city','4dehradun'

hbase(main):012:0> get 'emp','row3',{COLUMN=>'personal data:city',VERSIONS=>3}

COLUMN	CELL
personal data:city	timestamp=1523428391129, value=4dehradun
personal data:city	timestamp=1523428351922, value=3kochi
personal data:city	timestamp=1523428337355, value=2bang

hbase(main):013:0> get 'emp','row3',{COLUMN=>'personal data:city',VERSIONS=>2}

COLUMN	CELL
personal data:city	timestamp=1523428391129, value=4dehradun
personal data:city	timestamp=1523428351922, value=3kochi

TO DISABLE THE TABLE

AFTER disable: "list" command will display the table name but "scan" command will show error

hbase(main):024:0> disable 'expert'

hbase(main):020:0> list

TABLE

emp
example
expert

```
hbase(main):027:0> scan 'expert'
ROW                                COLUMN+CELL
ERROR: org.apache.hadoop.hbase.DoNotRetryIOException: expert is disabled.
```

TO ALTER TABLE to add additional column family 'xyz with version 5'

```
hbase(main):017:0> disable 'emp'
hbase(main):018:0> alter 'emp',NAME=>'xyz',VERSIONS=>5
hbase(main):020:0> enable 'emp'
hbase(main):022:0> describe 'emp'
DESCRIPTION                                ENABLED
{NAME => 'personal data', DATA_BLOCK_ENCODING => 'NONE', BLOOMFI
LTER => 'NONE', REPLICATION_SCOPE => '0', VERSIONS => '3', COMPRESSION =>
'NONE', MIN_VERSIONS => '0
', TTL => '2147483647', KEEP_DELETED_CELLS => 'false', BLOCKSIZE => '65536',
IN_MEMORY => 'false', E
NCODE_ON_DISK => 'true', BLOCKCACHE => 'true'}, {NAME => 'professional data',
DATA_BLOCK_ENCODING =>
'NONE', BLOOMFILTER => 'NONE', REPLICATION_SCOPE => '0', VERSIONS => '3',
COMPRESSION => 'NONE', MI
N_VERSIONS => '0', TTL => '2147483647', KEEP_DELETED_CELLS => 'false', BLOCKSIZE =>
'65536', IN_MEMO
RY => 'false', ENCODE_ON_DISK => 'true', BLOCKCACHE => 'true'}, {NAME => 'emp',
FAMILIES => [{NAME => 'xyz', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER =>
'NON true
E', REPLICATION_SCOPE => '0', COMPRESSION => 'NONE', VERSIONS => '5', TTL =>
'2147483647', MIN_VERSI
ONS => '0', KEEP_DELETED_CELLS => 'false', BLOCKSIZE => '65536', ENCODE_ON_DISK
=> 'true', IN_MEMORY
=> 'false', BLOCKCACHE => 'true'},]}
```

```
hbase(main):023:0> put 'emp','row3','xyz:city','5mumbai'
hbase(main):024:0> put 'emp','row3','xyz:city','6jodhpur'
hbase(main):028:0> put 'emp','row3','xyz:city','7hubli'
hbase(main):029:0> put 'emp','row3','xyz:city','8london'
hbase(main):030:0> put 'emp','row3','xyz:city','9shimla'
hbase(main):031:0> put 'emp','row3','xyz:city','10kanpur'
hbase(main):032:0> put 'emp','row3','xyz:city','11nagpur'
```

```
hbase(main):033:0> get 'emp','row3',{COLUMN=>'xyz:city',VERSIONS=>5}
COLUMN                                CELL
xyz:city                                timestamp=1523431975773, value=11nagpur
```

```

xyz:city          timestamp=1523431968306, value=10kanpur
xyz:city          timestamp=1523431961690, value=9shimla
xyz:city          timestamp=1523431953548, value=8london
xyz:city          timestamp=1523431947056, value=7hubli
5 row(s) in 0.0080 seconds
hbase(main):034:0> scan 'emp'
ROW              COLUMN+CELL
row1             column=personal data:city, timestamp=1523426980414, value=hyd
row1             column=personal data:name, timestamp=1523426797711, value=raju
row1             column=professional data:designation, timestamp=1523426835511,
value=manager
row1             column=professional data:salary, timestamp=1523426852045,
value=50000
row2             column=personal data:city, timestamp=1523427018284,
value=chennai
row2             column=personal data:name, timestamp=1523426910414,
value=milind
row2             column=professional data:designation, timestamp=1523427061723,
value=soft Engineer
row2             column=professional data:salary, timestamp=1523426965356,
value=30000
row3             column=personal data:city, timestamp=1523428391129,
value=4dehradun
row3             column=personal data:name, timestamp=1523427115249, value=anita
row3             column=professional data:designation, timestamp=1523427162000,
value=tester
row3             column=professional data:salary, timestamp=1523427187838,
value=40000
row3             column=xyz:city, timestamp=1523431975773, value=11nagpur
3 row(s) in 0.0620 seconds

```

EXP 29: TO DELETE

A PARTICULAR COLUMN FAMILY

```

hbase(main):037:0> disable 'emp'
hbase(main):038:0> alter 'emp','delete'=>'xyz'
hbase(main):039:0> enable 'emp'
hbase(main):042:0> get 'emp','row3'

```

```

COLUMN          CELL
personal data:city      timestamp=1523428391129, value=4dehradun
personal data:name      timestamp=1523427115249, value=anita
professional data:designation timestamp=1523427162000, value=tester
professional data:salary timestamp=1523427187838, value=40000
4 row(s) in 0.0080 seconds

```

A SPECIFIC CELL IN TABLE

hbase(main):045:0> delete 'emp','row3','personal data:name'

hbase(main):046:0> get 'emp','row3'

COLUMN

CELL

personal data:city timestamp=1523428391129, value=4dehradun

professional data:designation timestamp=1523427162000, value=tester

professional data:salary timestamp=1523427187838, value=40000

3 row(s) in 0.1870 seconds

DELETE COMPLETE ROW IN TABLE

hbase(main):049:0> deleteall 'emp','row3'

hbase(main):050:0> get 'emp','row3'

COLUMN

CELL

0 row(s) in 0.0340 seconds

hbase(main):051:0> scan 'emp'

ROW

COLUMN+CELL

row1 column=personal data:city, timestamp=1523426980414,

value=hyd

row1 column=personal data:name, timestamp=1523426797711,

value=raju

row1 column=professional data:designation,

timestamp=1523426835511, value=manager

row1 column=professional data:salary, timestamp=1523426852045,

value=50000

row2 column=personal data:city, timestamp=1523427018284,

value=chennai

row2 column=personal data:name, timestamp=1523426910414,

value=milind

row2 column=professional data:designation,

timestamp=1523427061723, value=soft Engineer

row2 column=professional data:salary, timestamp=1523426965356,

value=30000

2 row(s) in 0.0390 seconds

EXP 30: TO DROP THE TABLE

To drop , first table has to be disabled.

hbase(main):005:0> list

TABLE

emp

example

exp

expert

hbase(main):024:0> disable 'expert'

hbase(main):001:0> drop 'expert'

```
hbase(main):002:0> list
```

```
TABLE
```

```
emp
```

```
example
```

To drop all three table which starts with name 'ex'

```
hbase(main):006:0> disable_all 'ex.*'
```

```
example
```

```
exp
```

```
expert
```

```
Disable the above 3 tables (y/n)?
```

```
y
```

```
3 tables successfully disabled
```

```
hbase(main):007:0> drop_all 'ex.*'
```

```
example
```

```
exp
```

```
expert
```

```
Drop the above 3 tables (y/n)?
```

```
y
```

```
3 tables successfully dropped
```

```
hbase(main):052:0> count 'emp'
```

```
2 row(s) in 0.7080 seconds
```