



BIG DATA  
DEVELOPMENT

ACADGILD

## Session 09: HBase

---

### Assignment 2

**Problem 1:** Create an HBase table named 'clicks' with a column family 'hits' such that it should be able to store last 5 values of qualifiers inside 'hits' column family.

## Big Data and Hadoop Development

```
hbase(main):005:0> create 'clicks','hits'
```

```
0 row(s) in 1.2580 seconds
```

```
hbase(main):005:0> create 'clicks','hits'
0 row(s) in 1.2580 seconds

=> Hbase::Table - clicks
```

```
hbase(main):027:0> put 'clicks','192.168.1.1','hits:site','tpoint'
```

```
0 row(s) in 0.0090 seconds
```

```
hbase(main):028:0> put 'clicks','192.122.1.12','hits:site','acadgild'
```

```
0 row(s) in 0.0050 seconds
```

```
hbase(main):029:0> put 'clicks','192.122.1.0','hits:site','google'
```

```
0 row(s) in 0.0050 seconds
```

## *Big Data and Hadoop Development*

```
hbase(main):030:0> put 'clicks','192.122.1.19','hits:site','wipro'
```

```
0 row(s) in 0.0050 seconds
```

```
hbase(main):031:0> put 'clicks','192.126.1.19','hits:site','IBM'
```

```
0 row(s) in 0.0090 seconds
```

```
hbase(main):032:0> put 'clicks','192.126.1.19','hits:count','10'
```

```
0 row(s) in 0.0090 seconds
```

```
hbase(main):033:0> put 'clicks','192.126.1.19','hits:count','12'
```

```
0 row(s) in 0.0040 seconds
```

```
hbase(main):034:0> put 'clicks','192.122.1.19','hits:count','11'
```

```
0 row(s) in 0.0090 seconds
```

```
hbase(main):035:0> put 'clicks','192.122.1.0','hits:count','1200'
```

```
0 row(s) in 0.0060 seconds
```

## *Big Data and Hadoop Development*

```
hbase(main):036:0> put 'clicks','192.122.1.12','hits:count','100'
```

```
0 row(s) in 0.0050 seconds
```

```
hbase(main):037:0> put 'clicks','192.168.1.1','hits:count','200'
```

```
0 row(s) in 0.0080 seconds
```

```
hbase(main):027:0> put 'clicks','192.168.1.1','hits:site','tpoint'
0 row(s) in 0.0090 seconds

hbase(main):028:0> put 'clicks','192.122.1.12','hits:site','acadgild'
0 row(s) in 0.0050 seconds

hbase(main):029:0> put 'clicks','192.122.1.0','hits:site','google'
0 row(s) in 0.0050 seconds

hbase(main):030:0> put 'clicks','192.122.1.19','hits:site','wipro'
0 row(s) in 0.0050 seconds

hbase(main):031:0> put 'clicks','192.126.1.19','hits:site','IBM'
0 row(s) in 0.0090 seconds

hbase(main):032:0> put 'clicks','192.126.1.19','hits:count','10'
0 row(s) in 0.0090 seconds

hbase(main):033:0> put 'clicks','192.126.1.19','hits:count','12'
0 row(s) in 0.0040 seconds

hbase(main):034:0> put 'clicks','192.122.1.19','hits:count','11'
0 row(s) in 0.0090 seconds

hbase(main):035:0> put 'clicks','192.122.1.0','hits:count','1200'
0 row(s) in 0.0060 seconds

hbase(main):036:0> put 'clicks','192.122.1.12','hits:count','100'
0 row(s) in 0.0050 seconds

hbase(main):037:0> put 'clicks','192.168.1.1','hits:count','200'
0 row(s) in 0.0080 seconds
```

## Big Data and Hadoop Development

**Problem 2:** Add few records in the table and update some of them. Use IP Address as row-key. Scan the table to view if all the previous versions are getting displayed.

```
hbase(main):038:0> scan 'clicks'
```

ROW	COLUMN+CELL
192.122.1.0	column=hits:count, timestamp=1508747749204, value=1200
192.122.1.0	column=hits:site, timestamp=1508747610628, value=google
192.122.1.12	column=hits:count, timestamp=1508747773638, value=100
192.122.1.12	column=hits:site, timestamp=1508747588477, value=acadgild
192.122.1.19	column=hits:count, timestamp=1508747730331, value=11
192.122.1.19	column=hits:site, timestamp=1508747636995, value=wipro
192.126.1.19	column=hits:count, timestamp=1508747701290, value=12
192.126.1.19	column=hits:site, timestamp=1508747658235, value=IBM
192.168.1.1	column=hits:count, timestamp=1508747796110, value=200
192.168.1.1	column=hits:site, timestamp=1508747561473, value=tpoint

5 row(s) in 0.0210 seconds

## Big Data and Hadoop Development

```
hbase(main):038:0> scan 'clicks'
ROW                                COLUMN+CELL
192.122.1.0                        column=hits:count, timestamp=1508747749204, value=1200
192.122.1.0                        column=hits:site, timestamp=1508747610628, value=google
192.122.1.12                       column=hits:count, timestamp=1508747773638, value=100
192.122.1.12                       column=hits:site, timestamp=1508747588477, value=acadgild
192.122.1.19                       column=hits:count, timestamp=1508747730331, value=11
192.122.1.19                       column=hits:site, timestamp=1508747636995, value=wipro
192.126.1.19                       column=hits:count, timestamp=1508747701290, value=12
192.126.1.19                       column=hits:site, timestamp=1508747658235, value=IBM
192.168.1.1                        column=hits:count, timestamp=1508747796110, value=200
192.168.1.1                        column=hits:site, timestamp=1508747561473, value=tpoint
5 row(s) in 0.0210 seconds
hbase(main):039:0> █
```

```
hbase(main):005:0> create 'clicks','hits'
0 row(s) in 1.2580 seconds

=> Hbase::Table - clicks
hbase(main):006:0> put 'clicks','1','hits:ip','192.1.1.1'
0 row(s) in 0.1010 seconds

hbase(main):007:0> put 'clicks','1','hits:site','quora'
0 row(s) in 0.0110 seconds

hbase(main):008:0> put 'clicks','2','hits:ip','192.1.1.2'
0 row(s) in 0.0060 seconds

hbase(main):009:0> put 'clicks','3','hits:ip','192.1.1.2'
0 row(s) in 0.0050 seconds

hbase(main):010:0> put 'clicks','4','hits:ip','192.1.1.2'
0 row(s) in 0.0090 seconds

hbase(main):011:0> put 'clicks','5','hits:ip','192.1.1.2'
0 row(s) in 0.0080 seconds

hbase(main):012:0> put 'clicks','2','hits:site','google'
0 row(s) in 0.0080 seconds

hbase(main):013:0> put 'clicks','3','hits:site','yahoo'
0 row(s) in 0.0080 seconds

hbase(main):014:0> put 'clicks','4','hits:site','IBM'
0 row(s) in 0.0050 seconds

hbase(main):015:0> put 'clicks','5','hits:site','wipro'
0 row(s) in 0.0070 seconds
```



## Big Data and Hadoop Development

```
hbase(main):016:0> scan 'clicks'
ROW                                COLUMN+CELL
1      column=hits:ip, timestamp=1508747019445, value=192.1.1.1
1      column=hits:site, timestamp=1508747053594, value=quora
2      column=hits:ip, timestamp=1508747132749, value=192.1.1.2
2      column=hits:site, timestamp=1508747186680, value=google
3      column=hits:ip, timestamp=1508747142513, value=192.1.1.2
3      column=hits:site, timestamp=1508747203804, value=yahoo
4      column=hits:ip, timestamp=1508747151422, value=192.1.1.2
4      column=hits:site, timestamp=1508747218388, value=IBM
5      column=hits:ip, timestamp=1508747159989, value=192.1.1.2
5      column=hits:site, timestamp=1508747237796, value=wipro
5 row(s) in 0.0380 seconds
hbase(main):017:0> █
```