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.

Solution-

We will create a HBase table first with name “clicks” and column family “hits” and will specify the maximum version of its columns to 5 using below command-

- Create ‘clicks’, {NAME=> ‘hits’, VERSIONS=>5}

```
hbase(main):001:0> create 'clicks', {NAME=> 'hits', VERSIONS=> 5}
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an ex
2017-11-23 16:55:01,134 WARN  [main] util.NativeCodeLoader: Unable to
0 row(s) in 14.7010 seconds

=> Hbase::Table - clicks
hbase(main):002:0> list
TABLE
clicks
customer
2 row(s) in 0.1750 seconds

=> ["clicks", "customer"]
hbase(main):003:0>
```

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.

Now we will insert data to the table using put with IP addresses as ROW KEYS.
Here we have used 2 IP addresses- **192.168.1.4** and **192.168.1.3** as ROW KEYS for column family **'hits'**.
We are inserting columns Country, State, selfie and groupfie for this column family and also their corresponding value-

```
hbase(main):003:0> put 'clicks','192.168.1.4','hits:Country','USA'
0 row(s) in 0.9960 seconds

hbase(main):004:0> put 'clicks','192.168.1.4','hits:State','New Jersey'
0 row(s) in 0.0520 seconds

hbase(main):005:0> put 'clicks','192.168.1.4','hits:selfie','180'
0 row(s) in 0.0390 seconds

hbase(main):006:0> put 'clicks','192.168.1.4','hits:groupfie','100'
0 row(s) in 0.1150 seconds

hbase(main):007:0>
hbase(main):008:0* put 'clicks','192.168.1.3','hits:Country','India'
0 row(s) in 0.0280 seconds

hbase(main):009:0> put 'clicks','192.168.1.3','hits:State','Rajasthan'
0 row(s) in 0.0390 seconds

hbase(main):010:0> put 'clicks','192.168.1.3','hits:selfie','199'
0 row(s) in 0.0270 seconds

hbase(main):011:0> put 'clicks','192.168.1.3','hits:groupfie','95'
0 row(s) in 0.0690 seconds
```

Below is the result after inserting data-

```
hbase(main):012:0> scan 'clicks'
ROW                                   COLUMN+CELL
 192.168.1.3                          column=hits:Country, timestamp=1511437715210, value=India
 192.168.1.3                          column=hits:State, timestamp=1511437715389, value=Rajasthan
 192.168.1.3                          column=hits:groupfie, timestamp=1511437715737, value=95
 192.168.1.3                          column=hits:selfie, timestamp=1511437715572, value=199
 192.168.1.4                          column=hits:Country, timestamp=1511437713572, value=USA
 192.168.1.4                          column=hits:State, timestamp=1511437714024, value=New Jersey
 192.168.1.4                          column=hits:groupfie, timestamp=1511437714774, value=100
 192.168.1.4                          column=hits:selfie, timestamp=1511437714454, value=180
2 row(s) in 0.2620 seconds
```

Now we will upsert 3 more data in the table with same row keys and will try to update column for selfie for both row keys-

```
hbase(main):017:0> put 'clicks','192.168.1.4','hits:selfie','200'
0 row(s) in 0.0580 seconds

hbase(main):018:0> put 'clicks','192.168.1.4','hits:selfie','210'
0 row(s) in 0.0290 seconds

hbase(main):019:0> put 'clicks','192.168.1.4','hits:selfie','250'
0 row(s) in 0.0190 seconds

hbase(main):020:0> put 'clicks','192.168.1.3','hits:selfie','205'
0 row(s) in 0.0180 seconds

hbase(main):021:0> put 'clicks','192.168.1.3','hits:selfie','250'
0 row(s) in 0.0210 seconds

hbase(main):022:0> put 'clicks','192.168.1.3','hits:selfie','299'
0 row(s) in 0.0610 seconds
```

In below screenshot using **scan** command we can see all the previous versions of column “selfie” which we upserted. Currently it has 4 versions per ROW KEY

```
hbase(main):023:0> scan 'clicks', {COLUMN=>'hits:selfie',VERSIONS=>5}
ROW                                   COLUMN+CELL
 192.168.1.3                          column=hits:selfie, timestamp=1511438934257, value=299
 192.168.1.3                          column=hits:selfie, timestamp=1511438933137, value=250
 192.168.1.3                          column=hits:selfie, timestamp=1511438933041, value=205
 192.168.1.3                          column=hits:selfie, timestamp=1511437715572, value=199
 192.168.1.4                          column=hits:selfie, timestamp=1511438932955, value=250
 192.168.1.4                          column=hits:selfie, timestamp=1511438932842, value=210
 192.168.1.4                          column=hits:selfie, timestamp=1511438932648, value=200
 192.168.1.4                          column=hits:selfie, timestamp=1511437714454, value=180
2 row(s) in 0.1280 seconds

hbase(main):024:0>
```

Now again we are upserting some data for column “selfie” and thus making a total of 6 versions for column selfie.

Then we are doing a scan on the table for column “selfie” and we can see that there are only 5 version available for column selfie.

The earliest one version is not being displayed.

```
hbase(main):024:0> put 'clicks','192.168.1.4','hits:selfie','280'
0 row(s) in 0.0440 seconds

hbase(main):025:0> put 'clicks','192.168.1.3','hits:selfie','305'
0 row(s) in 0.0270 seconds

hbase(main):026:0> put 'clicks','192.168.1.4','hits:selfie','285'
0 row(s) in 0.0370 seconds

hbase(main):027:0> put 'clicks','192.168.1.3','hits:selfie','310'
0 row(s) in 0.0490 seconds

hbase(main):028:0>
hbase(main):029:0* scan 'clicks', {COLUMN=>'hits:selfie',VERSIONS=>5}
ROW                                   COLUMN+CELL
 192.168.1.3                          column=hits:selfie, timestamp=1511439134636, value=310
 192.168.1.3                          column=hits:selfie, timestamp=1511439119598, value=305
 192.168.1.3                          column=hits:selfie, timestamp=1511438934257, value=299
 192.168.1.3                          column=hits:selfie, timestamp=1511438933137, value=250
 192.168.1.3                          column=hits:selfie, timestamp=1511438933041, value=205
 192.168.1.4                          column=hits:selfie, timestamp=1511439126751, value=285
 192.168.1.4                          column=hits:selfie, timestamp=1511439118799, value=280
 192.168.1.4                          column=hits:selfie, timestamp=1511438932955, value=250
 192.168.1.4                          column=hits:selfie, timestamp=1511438932842, value=210
 192.168.1.4                          column=hits:selfie, timestamp=1511438932648, value=200
2 row(s) in 0.1050 seconds
```

Even if we try scanning table with VERSIONS=>7 we will get only 5 VERSIONS as table has been restricted to store only 5 versions

```
hbase(main):002:0> scan 'clicks', {COLUMN=>'hits:selfie',VERSIONS=>7}
ROW                                   COLUMN+CELL
 192.168.1.3                          column=hits:selfie, timestamp=1511439134636, value=310
 192.168.1.3                          column=hits:selfie, timestamp=1511439119598, value=305
 192.168.1.3                          column=hits:selfie, timestamp=1511438934257, value=299
 192.168.1.3                          column=hits:selfie, timestamp=1511438933137, value=250
 192.168.1.3                          column=hits:selfie, timestamp=1511438933041, value=205
 192.168.1.4                          column=hits:selfie, timestamp=1511439126751, value=285
 192.168.1.4                          column=hits:selfie, timestamp=1511439118799, value=280
 192.168.1.4                          column=hits:selfie, timestamp=1511438932955, value=250
 192.168.1.4                          column=hits:selfie, timestamp=1511438932842, value=210
 192.168.1.4                          column=hits:selfie, timestamp=1511438932648, value=200
2 row(s) in 0.9260 seconds

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