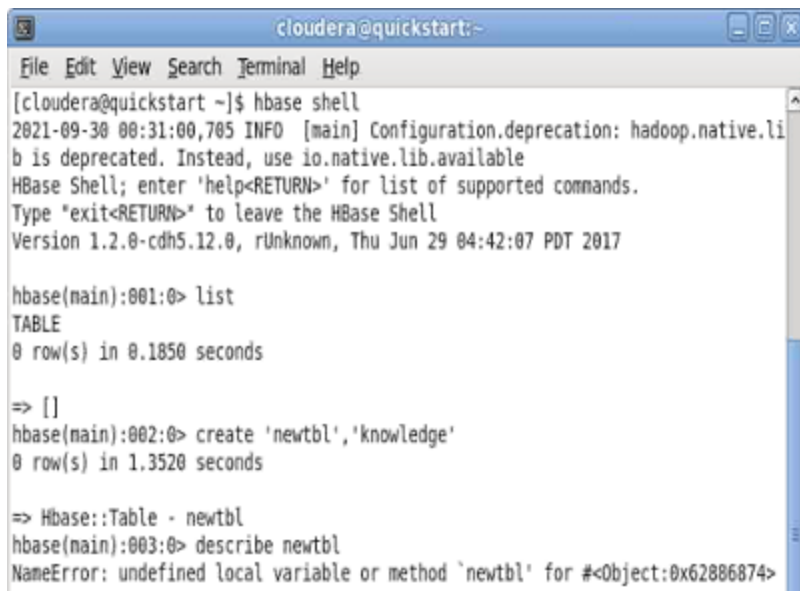19162121031

SMIT R PATEL

SME 5

HBASE PRACTICAL

BDA

AIM- To understand basic commands in HBase.

Exercise: In order to work with an 'unstructured data' containing file, your sibling wants you to help her work in Hadoop. She realises that she might need to modify data and have random reads/writes, which traditional HDFS would not help with. Give her a solution and provide a way for her to work through this problem while teaching her how to work with that tool.

Tasks:-

1. Open the HBase shell and create a table with a column family.

2. Work inside that column family and create columns with values.

```
[cloudera@quickstart Desktop]$ hbase shell
2021-09-30 00:45:33,239 INFO  [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.12.0, rUnknown, Thu Jun 29 04:42:07 PDT 2017

hbase(main):001:0> status 'summary'
1 active master, 0 backup masters, 1 servers, 0 dead, 3.0000 average load

hbase(main):002:0> list
TABLE
newtbl
1 row(s) in 0.0340 seconds

=> ["newtbl"]
hbase(main):003:0> put 'newtbl','r1','knowledge:sports','chess'
0 row(s) in 0.1510 seconds

hbase(main):004:0> put 'newtbl','r1','knowledge:science','Maths'
0 row(s) in 0.0090 seconds

hbase(main):005:0> put 'newtbl','r1','knowledge:Music','Spiritual'
0 row(s) in 0.0140 seconds

hbase(main):006:0> put 'newtbl','r1','knowledge:Economics','Macro'
0 row(s) in 0.0050 seconds

hbase(main):007:0> put 'newtbl','r2','knowledge:Economics','Micro'
0 row(s) in 0.0060 seconds

hbase(main):008:0> put 'newtbl','r2','knowledge:Music','sleep'
0 row(s) in 0.0080 seconds

hbase(main):009:0> scan 'newtbl'
```

3. Query data from the above-mentioned table.

```
hbase(main):017:0> get 'newtbl', 'r1'
COLUMN                    CELL
 knowledge:Economics        timestamp=1632988279919, value=Macro
 knowledge:Music            timestamp=1632988245409, value=Spiritual
 knowledge:science          timestamp=1632988230605, value=Maths
 knowledge:sports           timestamp=1632988177749, value=chess
4 row(s) in 0.0320 seconds

hbase(main):018:0> get 'newtbl', 'r2'
COLUMN                    CELL
 knowledge:Economics        timestamp=1632988309609, value=Micro
 knowledge:Music            timestamp=1632988326358, value=sleep
2 row(s) in 0.0160 seconds

hbase(main):019:0> list
TABLE
newtbl
1 row(s) in 0.0100 seconds

=> ["newtbl"]
```

4. Give commands for enabling and disabling the table, and also alter data values inside it.

```
File  Edit  View  Search  Terminal  Help
hbase(main):009:0> scan 'newtbl'
ROW                 COLUMN+CELL
 r1                 column=knowledge:Economics, timestamp=1632988279919, value
                    =Macro
 r1                 column=knowledge:Music, timestamp=1632988245409, value=Spi
                    ritual
 r1                 column=knowledge:science, timestamp=1632988230605, value=M
                    aths
 r1                 column=knowledge:sports, timestamp=1632988177749, value=ch
                    ess
 r2                 column=knowledge:Economics, timestamp=1632988309609, value
                    =Micro
 r2                 column=knowledge:Music, timestamp=1632988326358, value=sle
                    ep
2 row(s) in 0.0750 seconds

hbase(main):010:0> is_enabled 'newtbl'
true
0 row(s) in 0.0380 seconds

hbase(main):011:0> disable 'newtbl'
0 row(s) in 2.3270 seconds
```

```
hbase(main):012:0> scan 'newtbl'
ROW                          COLUMN+CELL

ERROR: newtbl is disabled.

Scan a table; pass table name and optionally a dictionary of scanner
specifications.  Scanner specifications may include one or more of:
TIMERANGE, FILTER, LIMIT, STARTROW, STOPROW, ROWPREFIXFILTER, TIMESTAMP,
MAXLENGTH or COLUMNS, CACHE or RAW, VERSIONS, ALL_METRICS or METRICS

If no columns are specified, all columns will be scanned.
To scan all members of a column family, leave the qualifier empty as in
'col_family'.

The filter can be specified in two ways:
1. Using a filterString - more information on this is available in the
Filter Language document attached to the HBASE-4176 JIRA
2. Using the entire package name of the filter.

If you wish to see metrics regarding the execution of the scan, the
ALL_METRICS boolean should be set to true. Alternatively, if you would
prefer to see only a subset of the metrics, the METRICS array can be
defined to include the names of only the metrics you care about.

Some examples:

  hbase> scan 'hbase:meta'
  hbase> scan 'hbase:meta', {COLUMNS => 'info:regioninfo'}
  hbase> scan 'ns1:t1', {COLUMNS => ['c1', 'c2'], LIMIT => 10, STARTROW => 'xyz'}
  hbase> scan 't1', {COLUMNS => ['c1', 'c2'], LIMIT => 10, STARTROW => 'xyz'}
  hbase> scan 't1', {COLUMNS => 'c1', TIMERANGE => [1303668804, 1303668904]}
  hbase> scan 't1', {REVERSED => true}
  hbase> scan 't1', {ALL_METRICS => true}
  hbase> scan 't1', {METRICS => ['RPC_RETRIES', 'ROWS_FILTERED']}
  hbase> scan 't1', {ROWPREFIXFILTER => 'row2', FILTER => "
    (QualifierFilter (>=, 'binary:xyz')) AND (TimestampsFilter ( 123, 456))"}
  hbase> scan 't1', {FILTER =>
    org.apache.hadoop.hbase.filter.ColumnPaginationFilter.new(1, 0)}
  hbase> scan 't1', {CONSISTENCY => 'TIMELINE'}
```

```
hbase(main):013:0> scan 'newtbl'alter 'newtbl', 'info'
SyntaxError: (hbase):13: syntax error, unexpected tIDENTIFIER

scan 'newtbl'alter 'newtbl', 'info'
                ^

hbase(main):014:0> alter 'newtbl', 'info'
Updating all regions with the new schema...
1/1 regions updated.
Done.
0 row(s) in 1.9910 seconds

hbase(main):015:0> enable 'newtbl'
0 row(s) in 1.3090 seconds

hbase(main):016:0> describe 'newtbl'
Table newtbl is ENABLED
newtbl
COLUMN FAMILIES DESCRIPTION
{NAME => 'info', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', COMPRESSION => 'N
ONE', VERSIONS => '1', TTL => 'FOREVER', MIN_VERSIONS => '0', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536'
, IN_MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'knowledge', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERSIONS =>
'1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FALSE', BLOCKSIZE => '6
5536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
2 row(s) in 0.0340 seconds
```

Extra ;-


```
                              cloudera@quickstart:~/Desktop

File  Edit  View  Search  Terminal  Help

hbase(main):023:0> is_enabled 'newtbl'
true
0 row(s) in 0.0330 seconds

hbase(main):024:0> scan 'newtbl'
ROW                         COLUMN+CELL
 r1                         column=knowledge:Economics, timestamp=1632988279919, value=Macro
 r1                         column=knowledge:Music, timestamp=1632988245409, value=Spiritual
 r1                         column=knowledge:science, timestamp=1632988230605, value=Maths
 r1                         column=knowledge:sports, timestamp=1632988177749, value=chess
 r2                         column=knowledge:Economics, timestamp=1632988309609, value=Micro
 r2                         column=knowledge:Music, timestamp=1632988326358, value=sleep
2 row(s) in 0.0470 seconds

hbase(main):025:0> enable 'newtbl'
0 row(s) in 0.0450 seconds

hbase(main):026:0> scan 'newtbl'
ROW                         COLUMN+CELL
 r1                         column=knowledge:Economics, timestamp=1632988279919, value=Macro
 r1                         column=knowledge:Music, timestamp=1632988245409, value=Spiritual
 r1                         column=knowledge:science, timestamp=1632988230605, value=Maths
 r1                         column=knowledge:sports, timestamp=1632988177749, value=chess
 r2                         column=knowledge:Economics, timestamp=1632988309609, value=Micro
 r2                         column=knowledge:Music, timestamp=1632988326358, value=sleep
2 row(s) in 0.0230 seconds
```

File   Edit   View   Search   Terminal   Help

```
hbase(main):027:0> is_disabled 'newtbl'
false
0 row(s) in 0.0360 seconds

hbase(main):028:0> scan 'newtbl'
ROW                              COLUMN+CELL
 r1                              column=knowledge:Economics, timestamp=1632988279919, value=Macro
 r1                              column=knowledge:Music, timestamp=1632988245409, value=Spiritual
 r1                              column=knowledge:science, timestamp=1632988230605, value=Maths
 r1                              column=knowledge:sports, timestamp=1632988177749, value=chess
 r2                              column=knowledge:Economics, timestamp=1632988309609, value=Micro
 r2                              column=knowledge:Music, timestamp=1632988326358, value=sleep
2 row(s) in 0.0270 seconds

hbase(main):029:0> disable 'newtbl'
0 row(s) in 2.3000 seconds

hbase(main):030:0> scan 'newtbl'
ROW                              COLUMN+CELL

ERROR: newtbl is disabled.

Scan a table; pass table name and optionally a dictionary of scanner
specifications.  Scanner specifications may include one or more of:
TIMERANGE, FILTER, LIMIT, STARTROW, STOPROW, ROWPREFIXFILTER, TIMESTAMP,
MAXLENGTH or COLUMNS, CACHE or RAW, VERSIONS, ALL_METRICS or METRICS

If no columns are specified, all columns will be scanned.
To scan all members of a column family, leave the qualifier empty as in
'col_family'.

The filter can be specified in two ways:
1. Using a filterString - more information on this is available in the
Filter Language document attached to the HBASE-4176 JIRA
2. Using the entire package name of the filter.

If you wish to see metrics regarding the execution of the scan, the
ALL_METRICS boolean should be set to true. Alternatively, if you would
```