Assignment 10

Task1

a. What is a NoSQL database?

NoSQL means Not Only SQL.. Nowadays, with the rise of unstructured or semi-structered data, the storage or management of data is quite cumbersome for the realtional databases. Such data are hard to be handled on the cluster. A NoSQL database solves this problem. A NoSQL database:

- Runs well on clusters
- Is an open-source
- Is schema-less
- Can support large volumes of data by running on clusters.
- File-based database and is not based on relational model of storing data.
- Is highly distributable.

b. How does data get stored in NoSQL database?

In a NoSQL database, the data is store in single tables rather than joining of multiple tables. It is stored on economical servers, which allow it to be handled more efficiently.

c. What is a column family in HBase?

A column family is a group of related data that can be accessed together. In this concept, a column can be added to any row instead of adding it to for all the other rows.

Example:

```
{
        “Data”:{
                        “employee”: {
```

```
                    “Name”: “Avani”
                    “Id”: “1234”
                    }
            “student”: {
                    “Name”: Abc”
                    “fees”:”123”
                }
        }
}
```

Here, table is “Data”. Column families are “employee” and “student” with the columns Name, Id and Name, fees respectively.

d. How many maximum number of columns can be added to HBase table?

There is no limit for the maximum number of columns to be added to a column family in HBase table.

e. Why columns are not defined at the time of table creation in HBase?

Columns are not defined at the time of table creation because in HBase, it is not necessary for every row to have same columns. And if a column is defined at the time of table creation then that will bound all the rows to have the same columns, even if it is not needed. This, will, in turn, occupy the unnecessary space in the table.

f. How does data get managed in HBase?

Data in Hbase is organized into tables. Any characters that are legal in file paths are used to name tables. Tables are further organized into rows that store data. Each row is identified by a unique row key which does not belong to any data type but is stored as a bytearray. Column families are further used to group data in rows. Column families define the physical structure of data so they are defined upfront and their modification is difficult. Each row in a table

has same column families. Data in a column family is addressed using a column qualifier.

g. What happens internally when new data gets inserted into HBase table?

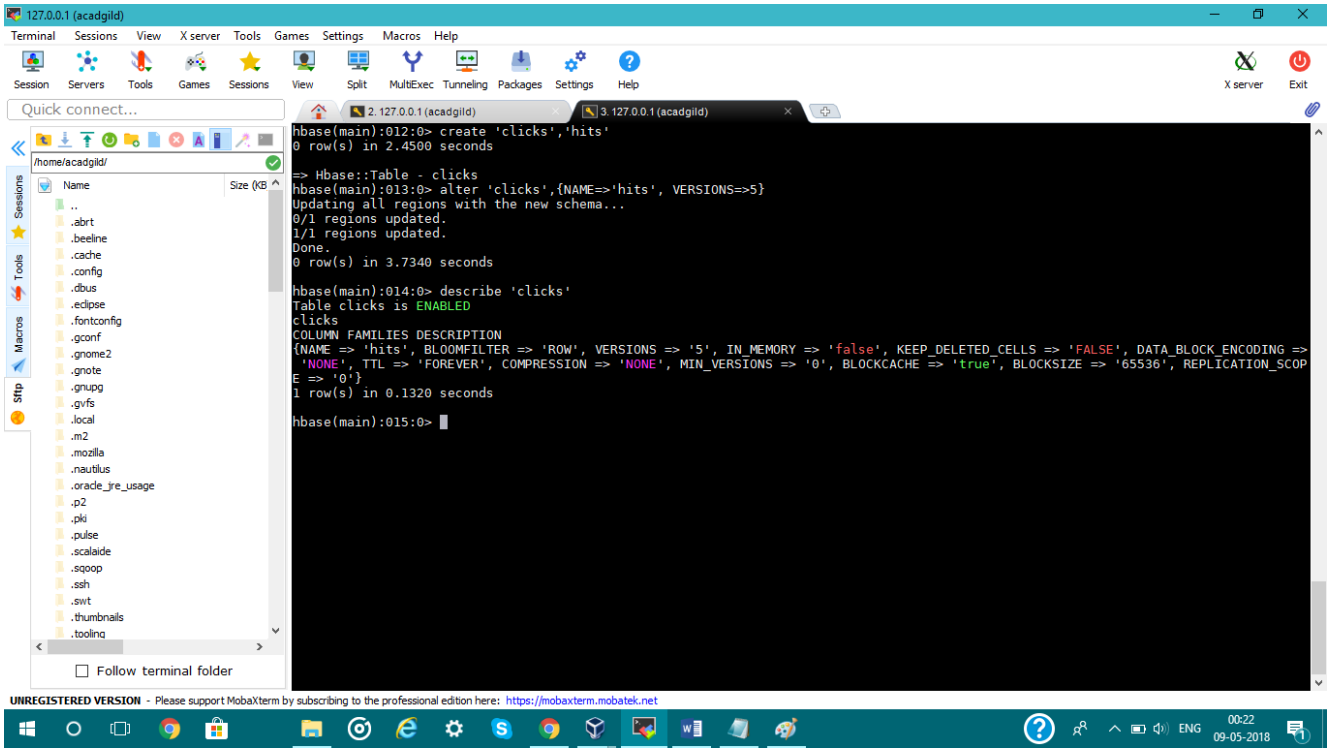
- The client gets the Region server that hosts the META table from ZooKeeper.
- The client will query the .META. server to get the region server corresponding to the row key it wants to access. The client caches this information along with the META table location.
- It will get the Row from the corresponding Region Server.
- For future reads, the client uses the cache to retrieve the META location and previously read row keys.

Task2

a. 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.

Commands used:

- create ‘clicks’,’hits’
- alter ‘clicks’,{NAME=>’hits’, VERSION=>5}



b. 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.

Commands used

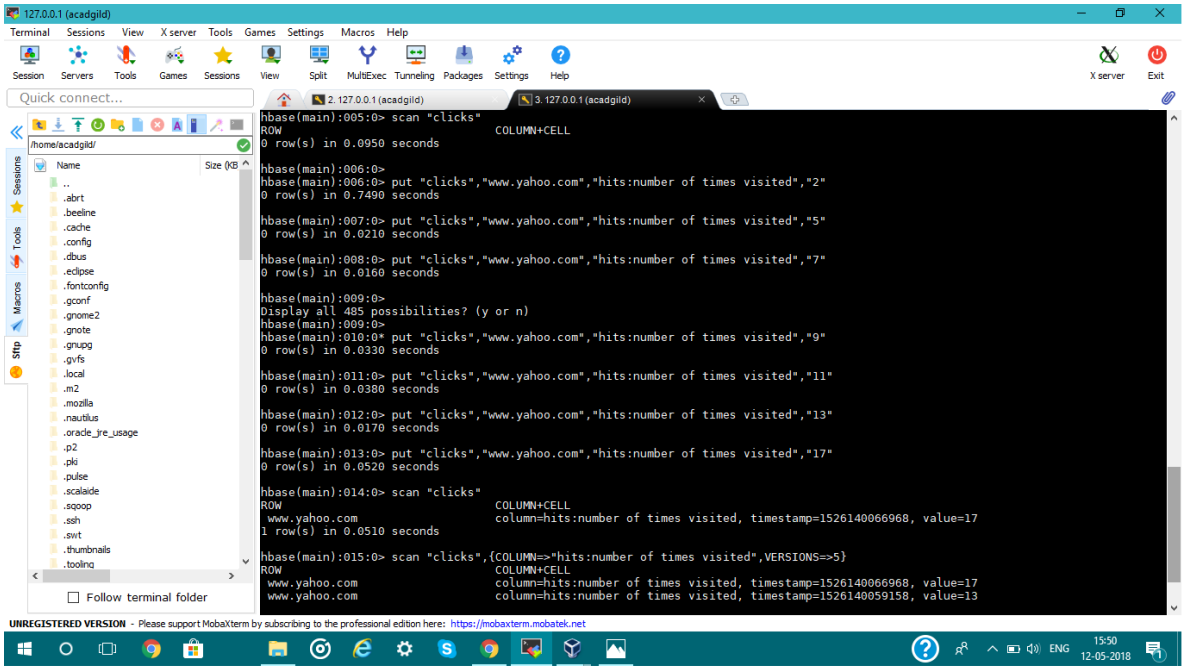
//addition of records

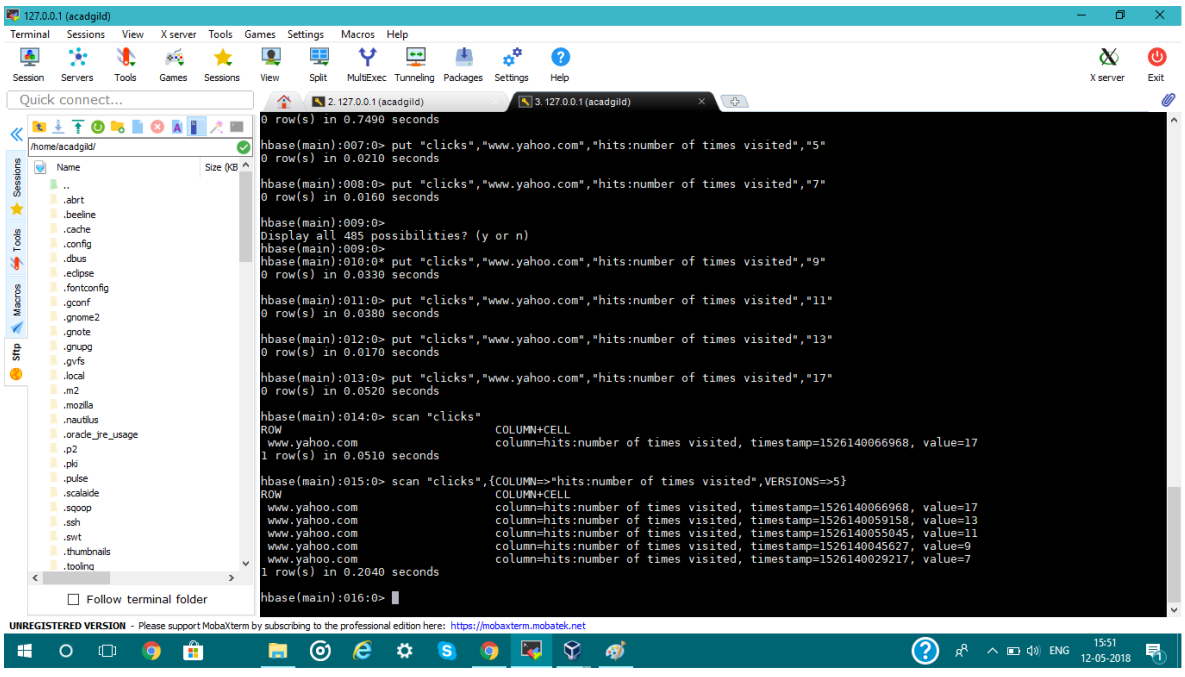
- put "clicks","www.yahoo.com","hits:number of times visited","2"
- put "clicks","www.yahoo.com","hits:number of times visited","5"
- put "clicks","www.yahoo.com","hits:number of times visited","7"
- put "clicks","www.yahoo.com","hits:number of times visited","9"
- put "clicks","www.yahoo.com","hits:number of times visited","11"
- put "clicks","www.yahoo.com","hits:number of times visited","13"
- put "clicks","www.yahoo.com","hits:number of times visited","17"

//Scanning of table to get the previous versions

scan "clicks",{COLUMN=>"hits:number of times visited",VERSIONS=>5}

Output


127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 127.0.0.1 (acadgild)
3. 127.0.0.1 (acadgild)
/home/acadgild/
Name Size (KB
Follow terminal folder
hbase(main):005:0> scan "clicks"
ROW COLUMN+CELL
0 row(s) in 0.0950 seconds
hbase(main):006:0>
hbase(main):006:0> put "clicks","www.yahoo.com","hits:number of times visited","2"
0 row(s) in 0.7490 seconds
hbase(main):007:0> put "clicks","www.yahoo.com","hits:number of times visited","5"
0 row(s) in 0.0210 seconds
hbase(main):008:0> put "clicks","www.yahoo.com","hits:number of times visited","7"
0 row(s) in 0.0160 seconds
hbase(main):009:0>
Display all 485 possibilities? (y or n)
hbase(main):009:0>
hbase(main):010:0* put "clicks","www.yahoo.com","hits:number of times visited","9"
0 row(s) in 0.0330 seconds
hbase(main):011:0> put "clicks","www.yahoo.com","hits:number of times visited","11"
0 row(s) in 0.0380 seconds
hbase(main):012:0> put "clicks","www.yahoo.com","hits:number of times visited","13"
0 row(s) in 0.0170 seconds
hbase(main):013:0> put "clicks","www.yahoo.com","hits:number of times visited","17"
0 row(s) in 0.0520 seconds
hbase(main):014:0> scan "clicks"
ROW COLUMN+CELL
www.yahoo.com column=hits:number of times visited, timestamp=1526140066968, value=17
1 row(s) in 0.0510 seconds
hbase(main):015:0> scan "clicks",{COLUMN=>"hits:number of times visited",VERSIONS=>5}
ROW COLUMN+CELL
www.yahoo.com column=hits:number of times visited, timestamp=1526140066968, value=17
www.yahoo.com column=hits:number of times visited, timestamp=1526140059158, value=13
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
ENG 15:50 12-05-2018


127.0.0.1 (acadgild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split MultiExec Tunneling Packages Settings Help X server Exit
Quick connect...
2. 127.0.0.1 (acadgild)
3. 127.0.0.1 (acadgild)
/home/acadgild/
Name Size (KB
Follow terminal folder
0 row(s) in 0.7490 seconds
hbase(main):007:0> put "clicks","www.yahoo.com","hits:number of times visited","5"
0 row(s) in 0.0210 seconds
hbase(main):008:0> put "clicks","www.yahoo.com","hits:number of times visited","7"
0 row(s) in 0.0160 seconds
hbase(main):009:0>
Display all 485 possibilities? (y or n)
hbase(main):009:0>
hbase(main):010:0* put "clicks","www.yahoo.com","hits:number of times visited","9"
0 row(s) in 0.0330 seconds
hbase(main):011:0> put "clicks","www.yahoo.com","hits:number of times visited","11"
0 row(s) in 0.0380 seconds
hbase(main):012:0> put "clicks","www.yahoo.com","hits:number of times visited","13"
0 row(s) in 0.0170 seconds
hbase(main):013:0> put "clicks","www.yahoo.com","hits:number of times visited","17"
0 row(s) in 0.0520 seconds
hbase(main):014:0> scan "clicks"
ROW COLUMN+CELL
www.yahoo.com column=hits:number of times visited, timestamp=1526140066968, value=17
1 row(s) in 0.0510 seconds
hbase(main):015:0> scan "clicks",{COLUMN=>"hits:number of times visited",VERSIONS=>5}
ROW COLUMN+CELL
www.yahoo.com column=hits:number of times visited, timestamp=1526140066968, value=17
www.yahoo.com column=hits:number of times visited, timestamp=1526140059158, value=13
www.yahoo.com column=hits:number of times visited, timestamp=1526140055045, value=11
www.yahoo.com column=hits:number of times visited, timestamp=1526140045627, value=9
www.yahoo.com column=hits:number of times visited, timestamp=1526140029217, value=7
1 row(s) in 0.2040 seconds
hbase(main):016:0>
UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
ENG 15:51 12-05-2018