

## Big Data And Hadoop

### Session 18 – Assignment 1

#### Problem Statement:

- Create a HBase table 'customer' with one column family 'details'.
- Write a shells script that loads the content of the file customers.dat in the HBase table.
- The structure of file looks like:  
id, name, location, age

#### Solution:

##### A. Place input file in HDFS

Step 1: The input file is present on local file system at **/home/acadgild/Abhilasha/hbase** as shown below.

```
acadgild@localhost:~/Abhilasha/hbase
File Edit View Search Terminal Help
[acadgild@localhost hbase]$ pwd
/home/acadgild/Abhilasha/hbase
[acadgild@localhost hbase]$ ls -l
total 12
-rw-rw-r-- 1 acadgild acadgild 833 Sep 24 18:27 commands
-rw-rw-r-- 1 acadgild acadgild 257 Sep 24 17:55 commands~
-rw-rw-r-- 1 acadgild acadgild 60 Oct 17 22:09 customers.dat
[acadgild@localhost hbase]$
```

Name of the file is customers.dat

Step 2: We now place the file on HDFS using the **put** command at location **/abhilasha** as follows

```
acadgild@localhost:~/Abhilasha/hbase
File Edit View Search Terminal Help
[acadgild@localhost hbase]$ hadoop fs -put /home/acadgild/Abhilasha/hbase/customers.dat /abhilasha
17/10/20 15:42:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
[acadgild@localhost hbase]$
```

Step 3: On listing the directory **/abhilasha**, we find the file with name **customers.dat** placed.

```
acadgild@localhost:~/Abhilasha/hbase
File Edit View Search Terminal Help
[acadgild@localhost hbase]$ hadoop fs -ls /abhilasha
17/10/20 15:43:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 5 items
drwxr-xr-x - acadgild supergroup 0 2017-08-29 21:58 /abhilasha/FlumeTitanicData
-rw-r--r-- 1 acadgild supergroup 60 2017-10-20 15:43 /abhilasha/customers.dat
drwxr-xr-x - acadgild supergroup 0 2017-09-05 16:52 /abhilasha/flume
drwxr-xr-x - acadgild supergroup 0 2017-09-17 15:18 /abhilasha/hive
drwxr-xr-x - acadgild supergroup 0 2017-08-29 22:53 /abhilasha/project12
[acadgild@localhost hbase]$
```

Step 4: We can view the content of this file using **cat** command as follows:

```
acadgild@localhost:~/Abhilasha/hbase
File Edit View Search Terminal Help
[acadgild@localhost hbase]$ hadoop fs -cat /abhilasha/customers.dat
17/10/20 15:44:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1,Amit,IND,18
2,Sumit,PAK,20
3,Rohit,AUS,26
4,Namit,UK,24[acadgild@localhost hbase]$
```

## B. Start HBase

Step 1: We first start HBase using the command **start-hbase.sh** as follows:

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ start-hbase.sh
starting master, logging to /usr/local/hbase/logs/hbase-acadgild-master-localhost.localdomain.out
[acadgild@localhost ~]$
```

Step 2: Next we start the hbase shell using the command **hbase shell** as follows:

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hbase shell
2017-09-24 17:49:57,986 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.a
vailble
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 22:35:44 PDT 2015
hbase(main):001:0>
```

## C. Create table

Step 1: Now we create the table named **customer** using the command **create 'customer', 'details'** as follows. Here, **details** is the name of the column family of the table **customer**.

```
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hbase shell
2017-10-20 15:46:48,865 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.a
vailble
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 22:35:44 PDT 2015
hbase(main):001:0> create 'customer','details';
hbase(main):002:0>
```

Step 2: Using **list** command, we can see the created table appear in the list.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
hbase(main):002:0* list  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-10-20 15:47:33,760 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
0 row(s) in 2.4930 seconds  
  
TABLE  
clicks  
customer  
htest  
3 row(s) in 0.0190 seconds  
  
=> ["clicks", "customer", "htest"]  
hbase(main):003:0>
```

#### D. Load data from file into HBase table

We use the below command to read data from file **customers.dat** and load it in hbase table name **customer**

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -  
Dimporttsv.columns=HBASE_ROW_KEY,details:name, details:location, details:age -  
Dimporttsv.separator=', '  
customer /abhilasha/customers.dat
```

Here, we specify the delimiter to be ',' in the file. This is the field separator for the records in the input file.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.columns=HBASE_ROW_KEY,details:name,details:location,details:age -Dimporttsv.separator=', ' customer /abhilasha/customers.dat;  
SLF4J: Class path contains multiple SLF4J bindings.  
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.  
2017-10-20 16:28:51,556 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

The data in the table after execution of the above command can be viewed using the **scan** command as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
  
hbase(main):040:0> scan 'customer'  
ROW COLUMN+CELL  
1 column=details:age, timestamp=1508497130956, value=18  
1 column=details:location, timestamp=1508497130956, value=IND  
1 column=details:name, timestamp=1508497130956, value=Amit  
2 column=details:age, timestamp=1508497130956, value=20  
2 column=details:location, timestamp=1508497130956, value=PAK  
2 column=details:name, timestamp=1508497130956, value=Sumit  
3 column=details:age, timestamp=1508497130956, value=26  
3 column=details:location, timestamp=1508497130956, value=AUS  
3 column=details:name, timestamp=1508497130956, value=Rohit  
4 column=details:age, timestamp=1508497130956, value=24  
4 column=details:location, timestamp=1508497130956, value=UK  
4 column=details:name, timestamp=1508497130956, value=Namit  
4 row(s) in 0.0430 seconds  
  
hbase(main):041:0> █
```

acacgild@localhost:~/... hbase - File Browser \*commands (~/Abhila... acacgild@localhost:~