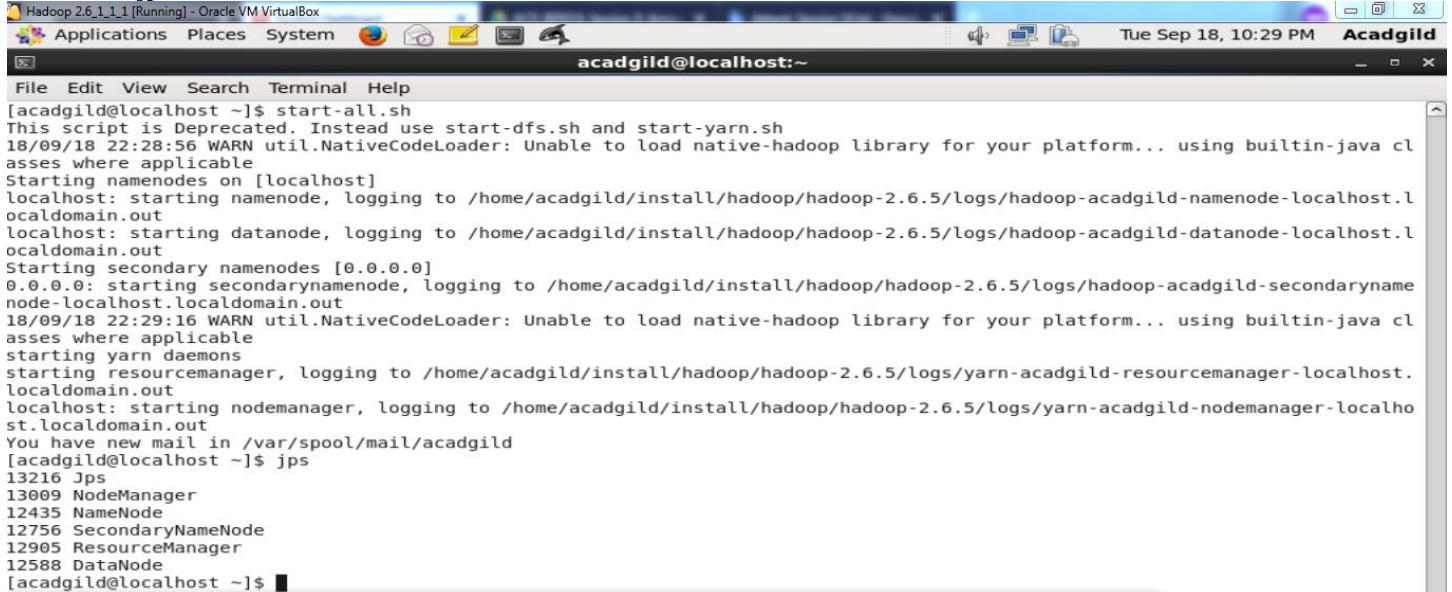


Session 8 – HIVE Basics Assignments 1.

1. Start the hadoop in VM and check all daemons are running. This can be done by following commands:

\$start-all.sh

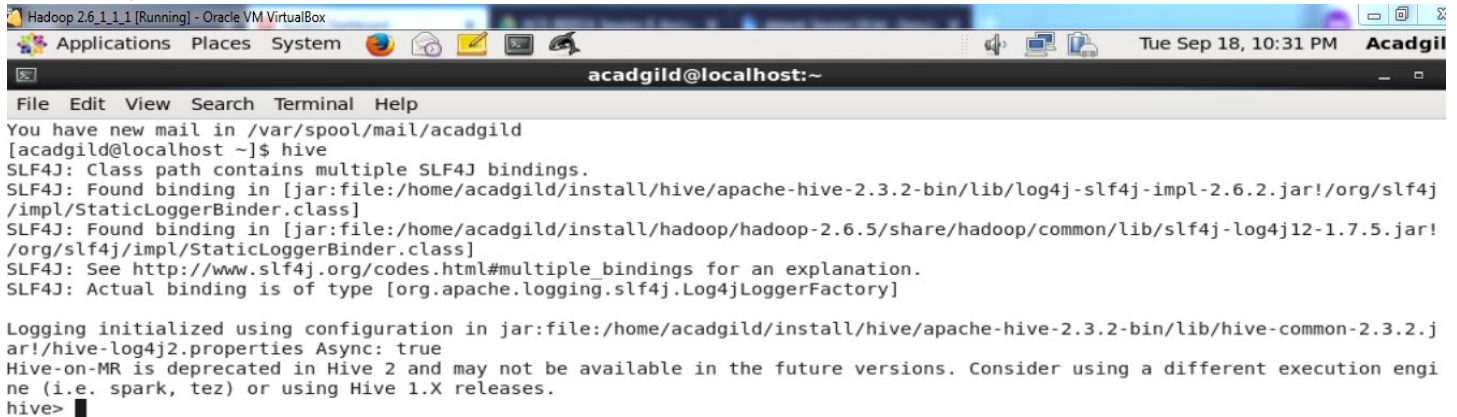
\$jps



```
Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
18/09/18 22:28:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.l
ocaldomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.l
ocaldomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondaryname
node-localhost.localdomain.out
18/09/18 22:29:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.
localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.
localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ jps
13216 Jps
13009 NodeManager
12435 NameNode
12756 SecondaryNameNode
12905 ResourceManager
12588 DataNode
[acadgild@localhost ~]$
```

2. Now start the hive CLI with the following command:

\$hive



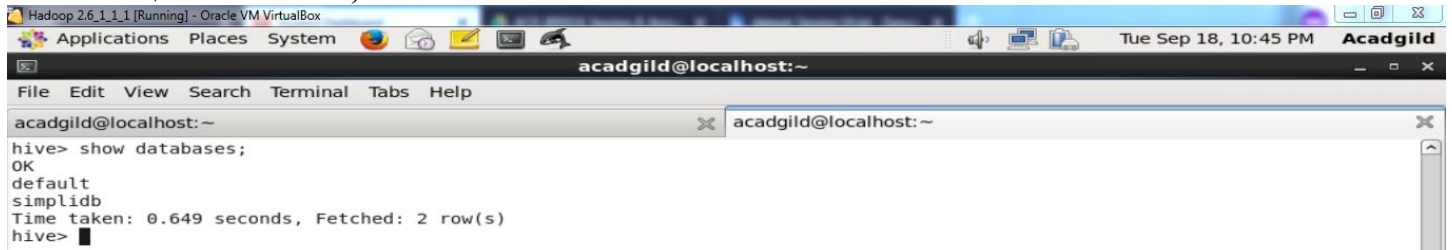
```
Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j
/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/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.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.j
ar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engi
ne (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

Task 1:

1. Check what are the databases available with ***show databases*** command:

\$show databases;



```
Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
hive> show databases;
OK
default
simplitdb
Time taken: 0.649 seconds, Fetched: 2 row(s)
hive>
```

2. Now we will create a new data base with name ***custom*** with the command ***create database custom*** as follows:

hive>create database custom;

3. The database created can be observed with the command ***show databases*** as follows:

hive> show databases;

```

acadgild@localhost:~
hive> show databases;
OK
default
simplidb
Time taken: 0.649 seconds, Fetched: 2 row(s)
hive> create database custom;
OK
Time taken: 0.339 seconds
hive> show databases;
OK
custom
default
simplidb
Time taken: 0.06 seconds, Fetched: 3 row(s)
hive> █

```

4. We can observe that the database *custom* is created.
5. Now create the table inside the database *custom* as follows:

```

Use custom;
CREATE TABLE temperature_data (
➤ Dt STRING,
➤ Zip_code INT,
➤ Temp INT
)
➤ Row format delimited fields terminated by ',';

```

6. Check the table status:

hive>show tables;

```

hive> CREATE TABLE temperature_data(
> dt date,
> zip_code INT,
> temp INT
> )
> row format delimited fields terminated by ',';
OK
Time taken: 0.202 seconds
hive> show tables;
OK
temperature_data
Time taken: 0.091 seconds, Fetched: 1 row(s)
hive> select * from temperature_data;
OK
Time taken: 3.05 seconds
hive> █

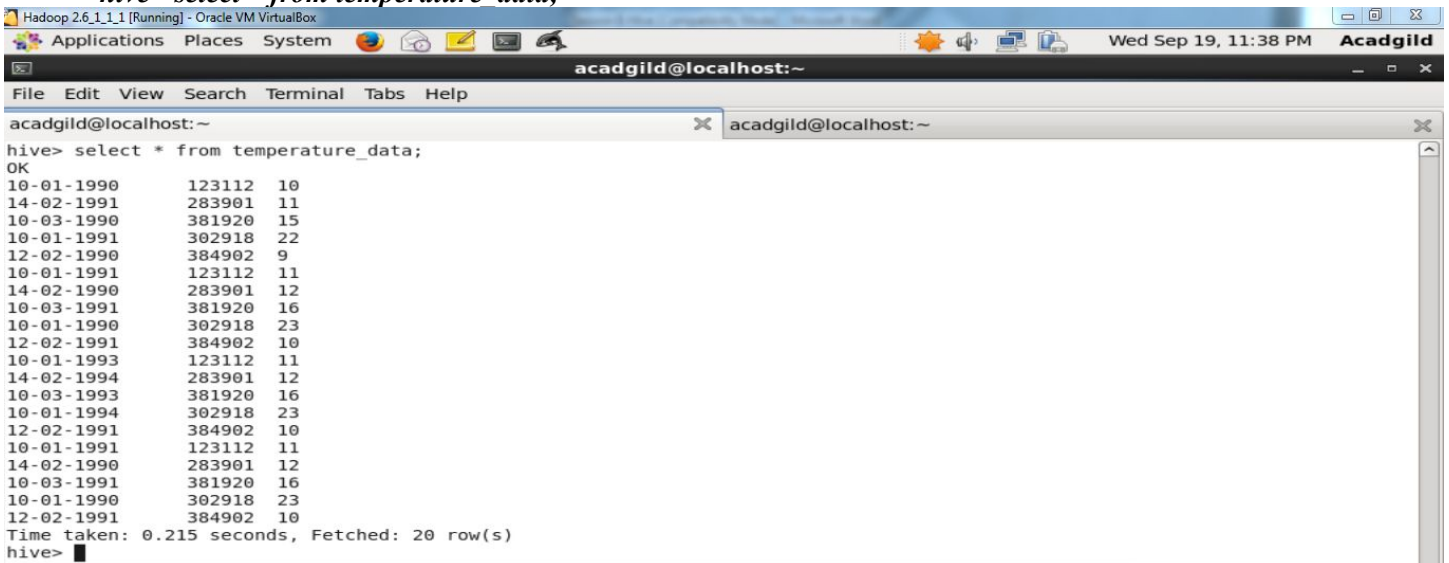
```

7. Load the data into the table *temperature_data* as follows and use select to display all the data from the table.

```

hive> LOAD DATA LOCAL INPATH '/home/acadgild/dataset_session14.txt' INTO temperature_data;
hive> select * from temperature_data;

```



```

Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
hive> select * from temperature_data;
OK
10-01-1990      123112  10
14-02-1991      283901  11
10-03-1990      381920  15
10-01-1991      302918  22
12-02-1990      384902   9
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
10-01-1993      123112  11
14-02-1994      283901  12
10-03-1993      381920  16
10-01-1994      302918  23
12-02-1991      384902  10
10-01-1991      123112  11
14-02-1990      283901  12
10-03-1991      381920  16
10-01-1990      302918  23
12-02-1991      384902  10
Time taken: 0.215 seconds, Fetched: 20 row(s)
hive> █

```

Task 2:

- Fetch date and temperature from *temperature_data* where zip code is greater than 300000 and less than 399999.

1. Write query for selecting date and temperature from table as follows:

\$select dt, temp from temperature_data where (zip_Code > 300000 AND zip_Code < 399999);

```

Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Sep 20, 12:41 AM Acadgild
acadmild@localhost:~
File Edit View Search Terminal Tabs Help
acadmild@localhost:~
hive> select dt, temp from temperature_data where (zip_Code > 300000 AND zip_Code < 399999);
OK
10-03-1990      15
10-01-1991      22
12-02-1990       9
10-03-1991      16
10-01-1990      23
12-02-1991      10
10-03-1993      16
10-01-1994      23
12-02-1991      10
10-03-1991      16
10-01-1990      23
12-02-1991      10
Time taken: 0.196 seconds, Fetched: 12 row(s)
hive>

```

2. We can see that all the temperature values and corresponding date where the zip code is above 300000 and 399999.

- Calculate maximum temperature corresponding to every year from temperature_data table.

```

Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Sep 20, 12:48 AM Acadgild
acadmild@localhost:~
File Edit View Search Terminal Tabs Help
acadmild@localhost:~
hive> select temp, dt from temperature_data where (year(to_date(from_unixtime(unix_timestamp(dt, 'mm-dd-yyyy')))) == 1990);
OK
10      10-01-1990
15      10-03-1990
9        12-02-1990
12      14-02-1990
23      10-01-1990
12      14-02-1990
23      10-01-1990
Time taken: 0.28 seconds, Fetched: 7 row(s)
hive> select temp, dt from temperature_data where (year(to_date(from_unixtime(unix_timestamp(dt, 'mm-dd-yyyy')))) == 1991);
OK
11      14-02-1991
22      10-01-1991
11      10-01-1991
16      10-03-1991
10      12-02-1991
10      12-02-1991
11      10-01-1991
16      10-03-1991
10      12-02-1991
Time taken: 0.166 seconds, Fetched: 9 row(s)
hive> select temp, dt from temperature_data where (year(to_date(from_unixtime(unix_timestamp(dt, 'mm-dd-yyyy')))) == 1993);
OK
11      10-01-1993
16      10-03-1993
Time taken: 0.162 seconds, Fetched: 2 row(s)
hive> select temp, dt from temperature_data where (year(to_date(from_unixtime(unix_timestamp(dt, 'mm-dd-yyyy')))) == 1994);
OK
12      14-02-1994
23      10-01-1994
Time taken: 0.187 seconds, Fetched: 2 row(s)
hive>

```

We can see that each with temperature values and we can observe that the maximum values present in each year.

- Calculate maximum temperature from temperature_data corresponding to those years which have at least 2 entries in the table.

```

Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System Thu Sep 20, 1:01 AM Acadgild
acadmild@localhost:~
File Edit View Search Terminal Tabs Help
acadmild@localhost:~
hive> select dt, temp from temperature_data where (temp > 22);
OK
10-01-1990      23
10-01-1994      23
10-01-1990      23
Time taken: 0.17 seconds, Fetched: 3 row(s)
hive>

```

We can observe that the maximum temperature 23 is appeared in two dates 10-01-1990 and 10-01-1994.

- Create a view on the top of last query, name it temperature_data_yw.

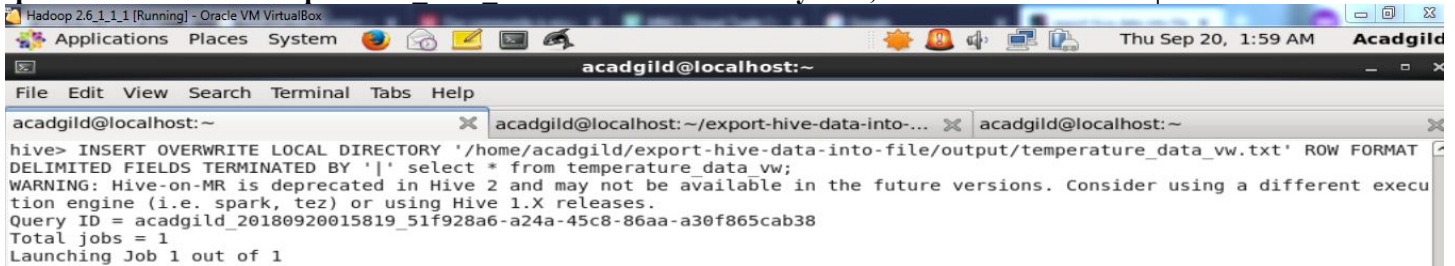
```

hive> create view temperature_data_vw select dt, temp from temperature_data where (temp > 22);
FAILED: ParseException line 1:32 missing AS at 'select' near '<EOF>'
hive> create view temperature_data_vw as select dt, temp from temperature_data where (temp > 22);
OK
Time taken: 1.826 seconds
hive> select * from temperature_data_vw;
OK
10-01-1990      23
10-01-1994      23
10-01-1990      23
Time taken: 0.196 seconds, Fetched: 3 row(s)
hive> █

```

We can observe that creation of view **temperature_data_vw** and we displayed records from the view with *select * from temperature_data_vw;*

- Export contents from **temperature_data_vw** to a file in local file system, such that each file is '|' delimited.



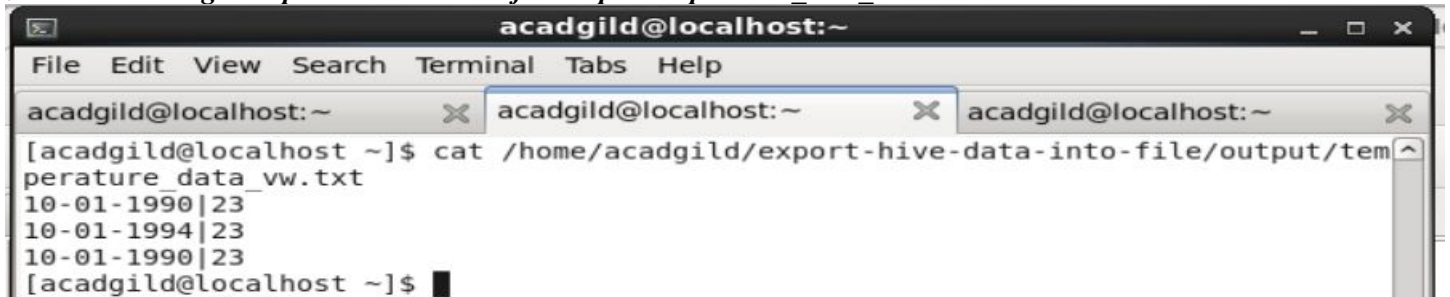
```

Hadoop 2.6.1_1.1 [Running] - Oracle VM VirtualBox
Applications Places System
acadgild@localhost:~
File Edit View Search Terminal Tabs Help
acadgild@localhost:~
hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/acadgild/export-hive-data-into-file/output/temperature_data_vw.txt' ROW FORMAT
DELIMITED FIELDS TERMINATED BY '|' select * from temperature_data_vw;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20180920015819_51f928a6-a24a-45c8-86aa-a30f865cab38
Total jobs = 1
Launching Job 1 out of 1

```

The output can be printed from the file **temperature_data_vw.txt** file as follows:

\$cat /home/acadgild/export-hive-daa-into-file/output/temperature_data_vw.txt



```

acacgild@localhost:~
File Edit View Search Terminal Tabs Help
acacgild@localhost:~
[acacgild@localhost ~]$ cat /home/acadgild/export-hive-data-into-file/output/tem
perature_data_vw.txt
10-01-1990|23
10-01-1994|23
10-01-1990|23
[acacgild@localhost ~]$ █

```