SPARK SQL HANDSON LAB HIVE SETUP, & USING SPARKSQL CREATE A HIVE TABLE & LOAD DATA

By www.HadoopExam.com

1. As we will be creating tables in Apache Hive, so need to do basic configuration setup on QuickStart VM. So that Spark can connect with Apache Hive.

2. Location of the Installed Apache Spark in QuickVM is

```
/usr/lib/spark
```

3. Location where the hive is installed

```
/etc/hive
```

4. Open a new terminal and change the user id to root as below.

```
Sudo su root
```

5. Now copy the "hive-site.xml""from hive/conf directory to Spark conf directory

```
cp /etc/hive/conf/hive-site.xml /usr/lib/spark/conf
```

6. Open the terminal and then start spark-shell

```
spark-shell
```

7. Create an instance of HiveContext:

```
val hc = new org.apache.spark.sql.hive.HiveContext(sc)
```

8. Create a Hive table HadoopExamEmployee with first_name, last_name, and age as columns

```
hc.sql("create table if not exists HadoopExamEmployee(first_name string, last_name string, age int) row format delimited fields terminated by ','")
```

9. Now check in Hive, whether HadoopExamEmployee table is created or not. If table was not already there than it will be created.

```
Select * from HadoopExamEmployee
```

10. Now we will load data in this table from a csv file.
11. Create a text file in a directory(e.g. HadoopExam) in HDFS and copy the below content in Employee.txt file.

```
Anita,Sharma,32
Lokesh,Teja,39
Vishnu,Kumar,40
Venkat,Theeru,56
Jatin,Pandit,59
Peeyush,Chawla,43
```

12. Now load the above data in table, which we have created.

```
hc.sql("load data inpath \"HadoopExam/Employee.txt\" into table HadoopExamEmployee")
```

13. Now check whether data is loaded in the table or not.

```
Select * from HadoopExamEmployee
```

14. Read the data using HiveQL in Spark

val Employee = hc.sql("from HadoopExamEmployee select first_name,last_name,age")

15. Now look the content of DataFrame.

Employee.collect.foreach(println)

16. Now create another table using the output of Query only

hc.sql("create table Employee2 as select first_name, last_name from HadoopExamEmployee")

17. Check the table, whether it has been created or not.

Select * from Employee2

Mandatory: For regular Updates on Products you have to Subscribe Please subscribe