

## Task 1: Spark Hive Integration

Copy the hive-site.xml file from to \$SPARK\_HOME/conf

Add the following properties to hive-site.xml on spark side :

- <property>
- <name>hive.metastore.uris</name>
- <value>thrift://localhost:9083</value>
- <description>password for connecting to mysql server</description>
- </property>

Write the code in Scala IDE to list the Databases in the hive. Source code is uploaded separately.

Keep your hadoop started

Start hive metastore by executing hive —service metastore command.

Run the code from the IDE. You should be able to see all the hive databases.

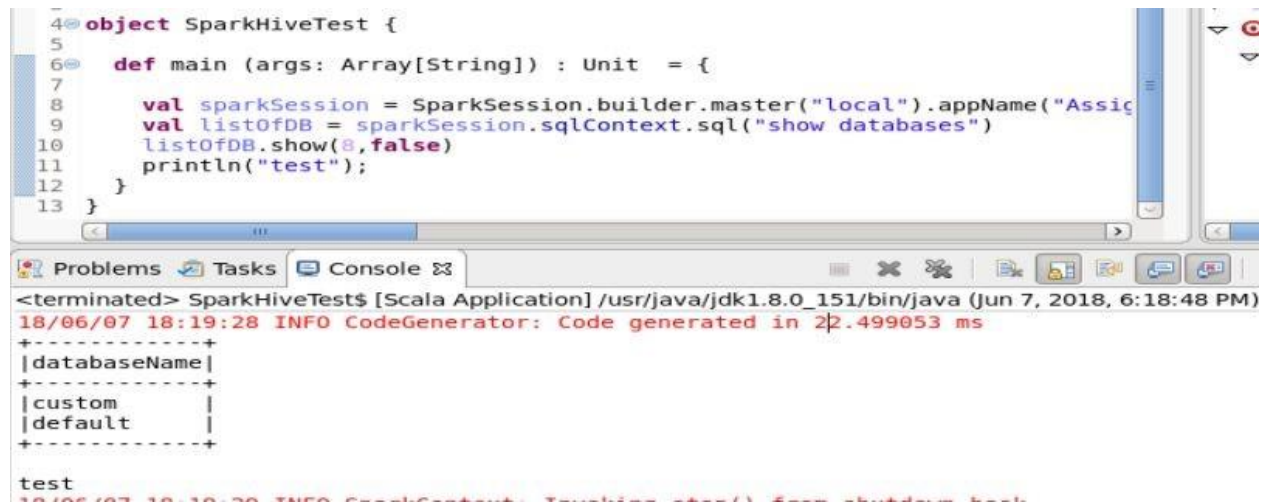
## Starting metastore:

```
STARTUP_MSG: build = git://stakiar-MBP.local/Users/stakiar/Desktop/scratch-space/apache-hive -r 857a9fd8ad725a53bd95c1b2d6612f9
01155f44d; compiled by 'stakiar' on Thu Nov 9 09:11:39 PST 2017
*****
2018-06-07T18:16:45,098 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Starting hive metastore on port 9083
2018-06-07T18:16:45,853 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - 0: Opening raw store with implementation cla
ss:org.apache.hadoop.hive.metastore.ObjectStore
2018-06-07T18:16:53,806 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Added admin role in metastore
2018-06-07T18:16:53,812 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Added public role in metastore
2018-06-07T18:16:53,852 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - No user is added in admin role, since config
is empty
2018-06-07T18:16:54,184 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Starting DB backed MetaStore Server with Set
ting JGI enabled
2018-06-07T18:16:54,219 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Started the new metaserver on port [9083]...
2018-06-07T18:16:54,219 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Options.minWorkerThreads = 200
2018-06-07T18:16:54,219 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - Options.maxWorkerThreads = 1000
2018-06-07T18:16:54,220 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - TCP keepalive = true
```

## Hive database in the terminal:

```
hive> show databases;
OK
custom
default
Time taken: 0.04 seconds, Fetched: 2 row(s)
hive> █
```

## Output after running the code in Scala Ide



```
4 object SparkHiveTest {
5
6 def main (args: Array[String]) : Unit = {
7
8     val sparkSession = SparkSession.builder.master("local").appName("Assig
9     val listofDB = sparkSession.sqlContext.sql("show databases")
10    listofDB.show(8, false)
11    println("test");
12 }
13 }
```

Problems Tasks Console

<terminated> SparkHiveTest\$ [Scala Application] /usr/java/jdk1.8.0\_151/bin/java (Jun 7, 2018, 6:18:48 PM)

18/06/07 18:19:28 INFO CodeGenerator: Code generated in 22.499053 ms

databaseName
custom
default

test

## Task 2: Spark Hbase Integration

API code in scala ide to create a new table in hbase is uploaded separately.

Run the code in scala ide and check hbase for new table

Start the hbase shell using the below commands

Start-hbase.sh

Hbase shell

List of tables before running the code:

```
hbase(main):001:0> list
TABLE
SparkHBasesTable
TRANSACTIONS
bulktable
clicks
4 row(s) in 0.3620 seconds

=> ["SparkHBasesTable", "TRANSACTIONS", "bulktable", "clicks"]
```

List of tables after running the code:

```
hbase(main):001:0> list
TABLE
SparkHBasesTable
TRANSACTIONS
bulktable
clicks
4 row(s) in 0.3620 seconds

=> ["SparkHBasesTable", "TRANSACTIONS", "bulktable", "clicks"]
hbase(main):002:0> list
TABLE
SparkHBasesTable
SparkHBasesTable1
TRANSACTIONS
bulktable
clicks
5 row(s) in 0.0160 seconds

=> ["SparkHBasesTable", "SparkHBasesTable1", "TRANSACTIONS", "bulktable", "clicks"]
```

Newly created Hbase table:

```
hbase(main):004:0> scan 'SparkHBasesTable1'
ROW COLUMN+CELL
row1      column=cf:column, timestamp=1528376937645, value=value1
row10     column=cf:column, timestamp=1528376937716, value=value10
row2      column=cf:column, timestamp=1528376937673, value=value2
row3      column=cf:column, timestamp=1528376937679, value=value3
row4      column=cf:column, timestamp=1528376937683, value=value4
row5      column=cf:column, timestamp=1528376937688, value=value5
row6      column=cf:column, timestamp=1528376937694, value=value6
row7      column=cf:column, timestamp=1528376937701, value=value7
row8      column=cf:column, timestamp=1528376937705, value=value8
row9      column=cf:column, timestamp=1528376937711, value=value9
10 row(s) in 0.4000 seconds
```

### Task 3 : Spark Kafka Integration:

After creating the scala object write the consumer code for counting the words entered in the producer and run as scala application

Before running the consumer code, start the zookeeper server , kafka server, create the topic and start the producer to enter the message that needs to sent to the consumer using below commands:

Starting zookeeper

```
cd $KAFKA_HOME
```

```
#!/bin/zookeeper-server-start ./etc/kafka/zookeeper.properties
```

```
./bin/zookeeper-server-start.sh ./config/zookeeper.properties
```

Starting broker

```
#!/bin/kafka-server-start./etc/kafka/server.properties
```

```
./bin/kafka-server-start.sh ./config/server.properties
```

Creating topic

```
./bin/kafka-topics.sh --create --topic Mytopic10 --zookeeper localhost:2181
```

```
--partitions 1 --replication-factor 1
```

Starting the producer:

```
./bin/kafka-console-producer.sh --broker-list localhost:9092 --topic
```

Mytopic10

Input:

```
[acadgild@localhost kafka_2.12-0.10.1.1]$ ./bin/kafka-console-producer.sh --brok
Hello Everyone, This is khafka and Spark Integration session.
This example is a word count program to count the words using khafka and spark I
```

Output:

Wordcount screenshot

```
Time: 1529576170000 ms
-----
(null,Hello Everyone, This is khafka and Spark Integration session.)
(null,This example is a word count program to count the words using khafka and spark Integration.)
T
```

```
Problems Tasks Console
<terminated> WordCounts [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jun 21, 2018, 3:45:27 PM)
Time: 1529576170000 ms
-----
(example,1)
(Spark,1)
(session.,1)
(This,2)
(Integration.,1)
(word,1)
(Integration,1)
(the,1)
(is,2)
(a,1)
```