ASSIGNMENT

Task1: Spark-Hive Integration

Sol: Step1: copy the hive-site.xml file from $HIVE-HOME conf folder to $SPARK-HOME conf folder

➔ cp /home/acadgild/install/hive/apache-hive-2.3.2-bin/conf/hive-site.xml /home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/conf

Step2: edit the hive-site.xml file

➔ vi hive-site.xml

➔ enter the following property tags in configuration

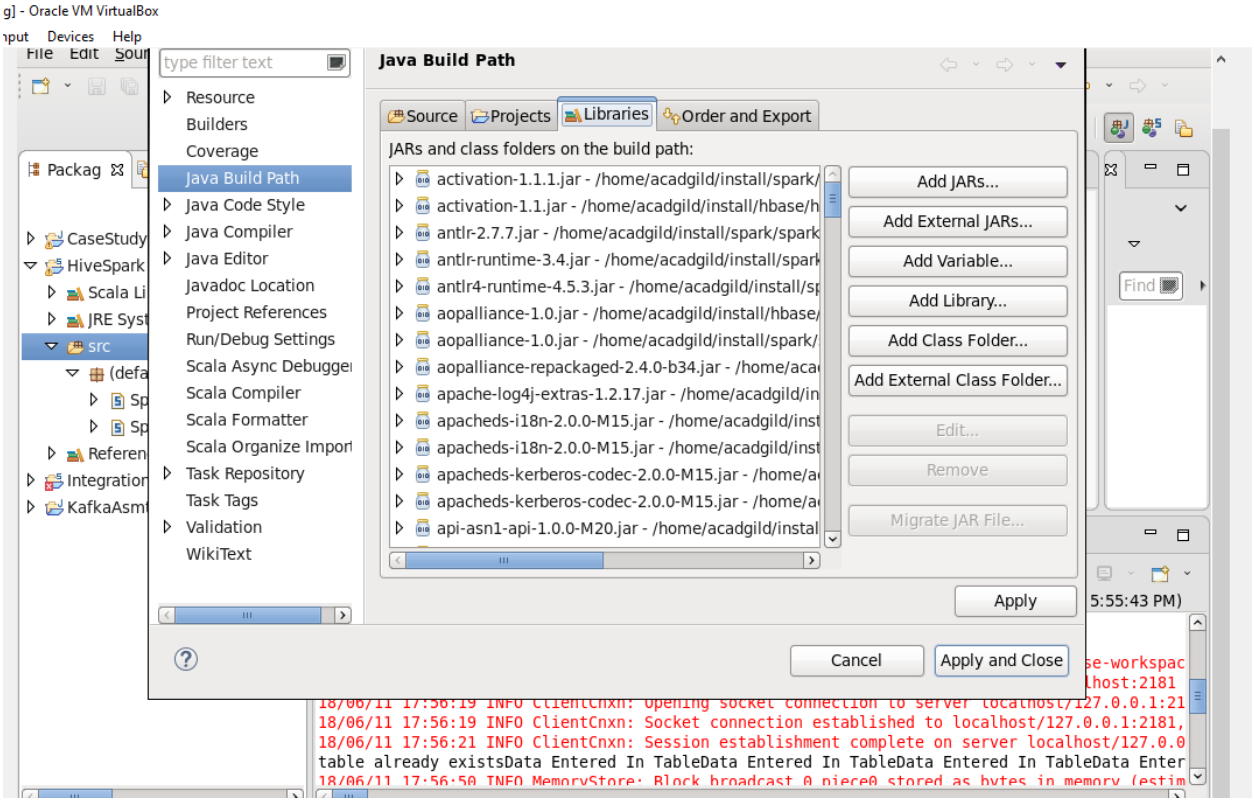
➔ <property>
<name>hive.metastore.uris</name>
<value>thrift://localhost:9083</value>
<description>password for connecting to mysql server</description>
</property>

acadgild@localhost:~/install/spark/spark-2.2.1-bin-hadoop2.7/conf

```
</property>

<property>
  <name>javax.jdo.option.ConnectionPassword</name>
  <value>Root@123</value>
</property>

<property>
  <name>datanucleus.autoCreateSchema</name>
  <value>true</value>
</property>

<property>
  <name>datanucleus.fixedDatastore</name>
  <value>true</value>
</property>

<property>
 <name>datanucleus.autoCreateTables</name>
 <value>True</value>
 </property>

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

Step3: copy the code to eclipse IDE (SparkHiveTest.scala)

Step4: Add the jar files from spark and hbase folders

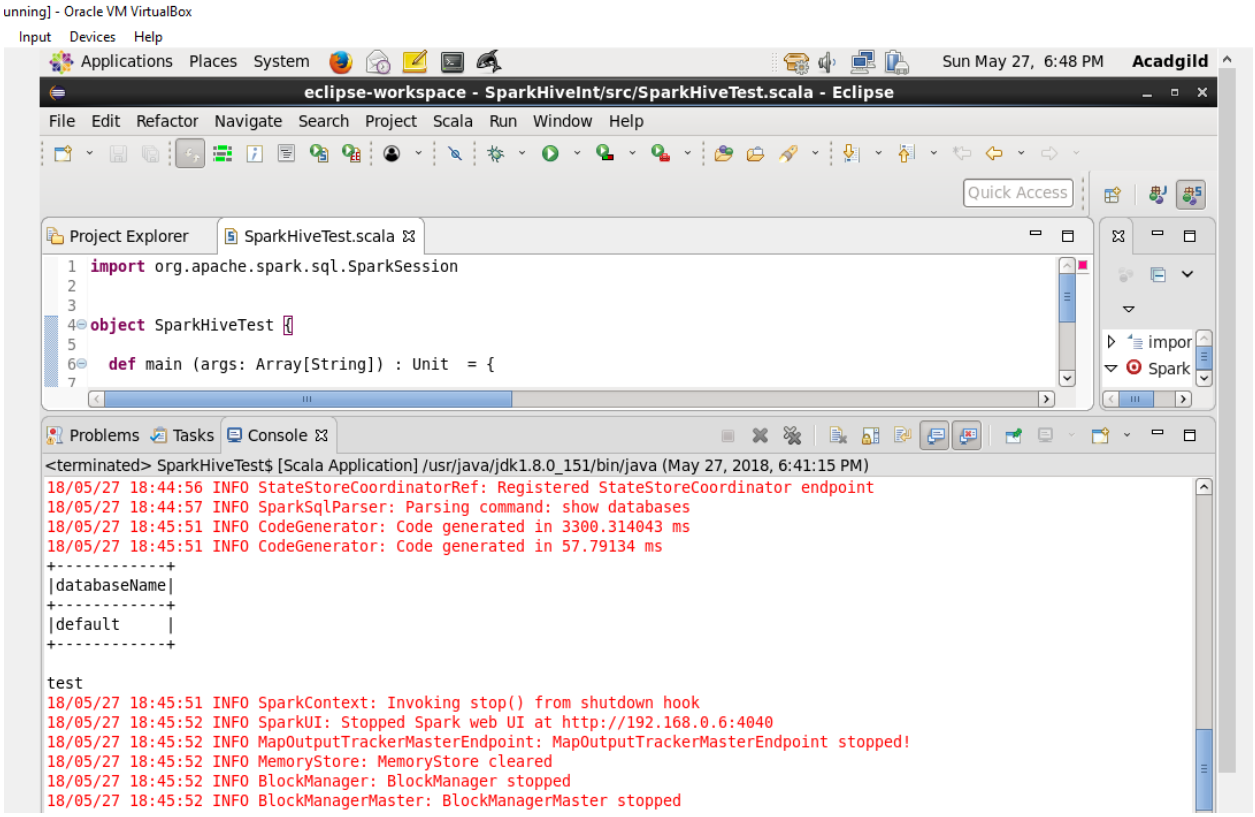


Step5: start hive server2

➔ Hive –service metastore

```
acadgild@localhost:~
[acadgild@localhost ~]$ hive --service metastore
2018-06-11 17:49:45: Starting Hive Metastore Server
/home/acadgild/install/hive/apache-hive-2.3.2-bin/bin/ext/metastore.sh: line 29:
 export: ` -Dproc_metastore  -Dlog4j.configurationFile=hive-log4j2.properties  -
Djava.util.logging.config.file=/home/acadgild/install/hive/apache-hive-2.3.2-bin
/conf/parquet-logging.properties  ': not a valid identifier
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/sha
re/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]
2018-06-11T17:49:54,399 INFO [main] org.apache.hadoop.hive.conf.HiveConf - Found
 configuration file file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/conf/
hive-site.xml
2018-06-11T17:50:01,387 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaSto
re - STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting HiveMetaStore
STARTUP_MSG:   host = localhost/127.0.0.1
STARTUP_MSG:   args = []
STARTUP_MSG:   version = 2.3.2
STARTUP_MSG:   classpath = /home/acadgild/install/hive/apache-hive-2.3.2-bin/con
f:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-core-1.6.0.jar:
/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-fate-1.6.0.jar:/h
ome/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-start-1.6.0.jar:/ho
me/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-trace-1.6.0.jar:/hom
e/acadgild/install/hive/apache-hive-2.3.2-bin/lib/activation-1.1.jar:/home/acadg
ild/install/hive/apache-hive-2.3.2-bin/lib/aether-api-0.9.0.M2.jar:/home/acadgil
d/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-file-0.9.0.M2.jar:/hom
e/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-okhttp-0.0.9.
jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-impl-0.9.0.M2.j
ar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-spi-0.9.0.M2.jar
```

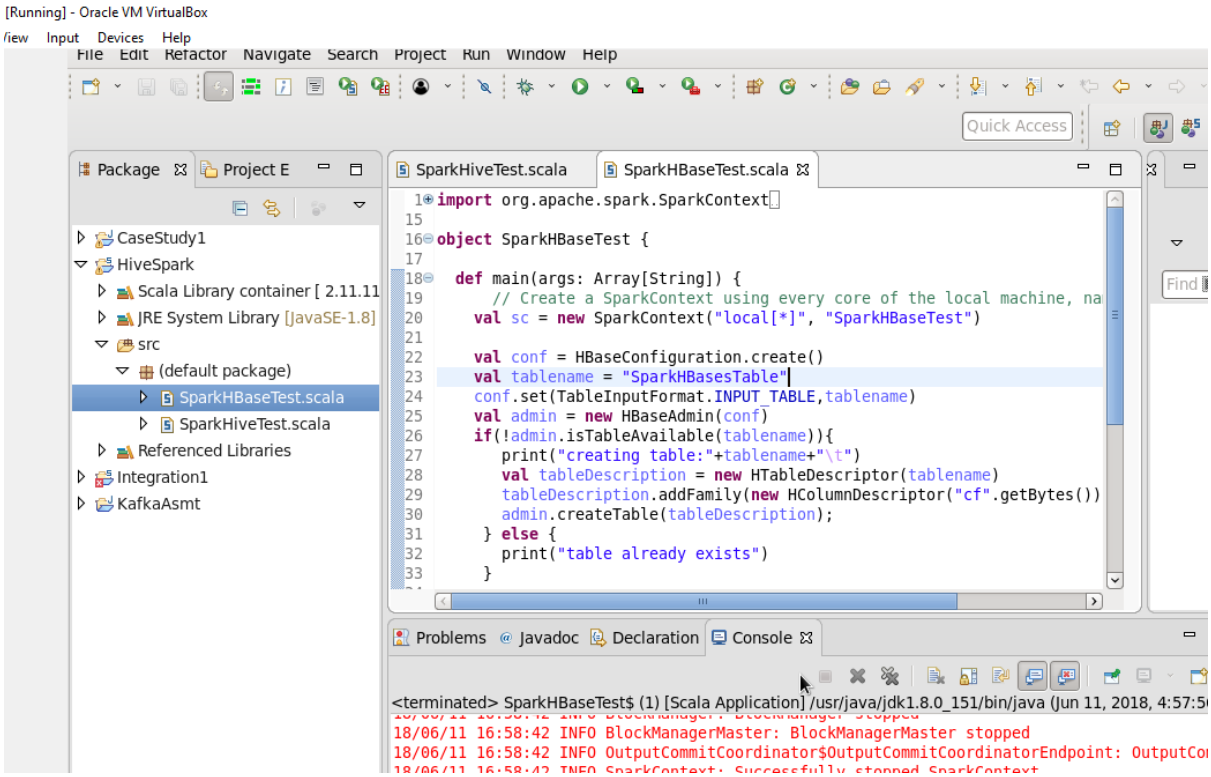
Step6: Run the program



Task2: Spark-Hbase Integration

Sol: Step1: Add the code to same scala project.

➔ SparkHbaseTest.scala // program file



Step2: Run the program.

Step3: In Putty -> cd install/hbase/hbase-1.2.6/bin

./start-hbase.sh // start hbase

hbase shell // start hbase shell

scan “SparkHbasesTable”

```
hbase(main):001:0> scan "SparkHBasesTable"
ROW                   COLUMN+CELL
 row1                 column=cf:column1, timestamp=1528716512101, value=value1
 row10                column=cf:column1, timestamp=1528716512269, value=value10
 row2                 column=cf:column1, timestamp=1528716512180, value=value2
 row3                 column=cf:column1, timestamp=1528716512194, value=value3
 row4                 column=cf:column1, timestamp=1528716512204, value=value4
 row5                 column=cf:column1, timestamp=1528716512209, value=value5
 row6                 column=cf:column1, timestamp=1528716512214, value=value6
 row7                 column=cf:column1, timestamp=1528716512227, value=value7
 row8                 column=cf:column1, timestamp=1528716512242, value=value8
 row9                 column=cf:column1, timestamp=1528716512259, value=value9
10 row(s) in 1.3850 seconds

hbase(main):002:0>
```