

## Assignment 25

### Task1

#### Integrate Spark Hive

```
import org.apache.spark.sql.SparkSession

object SparkHiveTest {

  def main (args: Array[String]) : Unit = {

    val sparkSession = SparkSession.builder.master("local").appName("spark session
example").config("spark.sql.warehouse.dir", "/user/hive/warehouse").config("hive.metastor
e.uris", "thrift://localhost:9083").enableHiveSupport().getOrCreate()
    val listOfDB = sparkSession.sqlContext.sql("show databases")    //command for showing
databases in hive
    listOfDB.show(8,false)
    println("test");
  }
}
```

### Task2

#### Integrate Spark HBase

```
import org.apache.spark.SparkContext

import org.apache.hadoop.hbase.HBaseConfiguration
import org.apache.hadoop.hbase.mapreduce.TableInputFormat
import org.apache.hadoop.hbase.client.HBaseAdmin
import org.apache.hadoop.hbase.{HTableDescriptor,HColumnDescriptor}
import org.apache.hadoop.hbase.util.Bytes
import org.apache.hadoop.hbase.client.{Put,HTable}
import org.apache.log4j._
import org.apache.hadoop.hbase.io.ImmutableBytesWritable
import org.apache.hadoop.hbase.client.Result

object SparkHBaseTest {

  def main(args: Array[String]) {
    // Create a SparkContext using every core of the local machine, named RatingsCounter
    val sc = new SparkContext("local[*]", "SparkHBaseTest")

    println("hello spark hbase ---> 1")

    val conf = HBaseConfiguration.create()
```

```

val tablename = "SparkHBaseTable"
conf.set(TableInputFormat.INPUT_TABLE,tablename)
val admin = new HBaseAdmin(conf)
if(!admin.isTableAvailable(tablename)){ //checking the existence of table
    print("creating table:"+tablename+"\t")
    val tableDescription = new HTableDescriptor(tablename)
    tableDescription.addFamily(new HColumnDescriptor("cf".getBytes())); //adding the
column family
    admin.createTable(tableDescription);
} else {
    print("table already exists")
}

val table = new HTable(conf,tablename);
for(x <- 1 to 10){
    var p = new Put(new String("row" + x).getBytes());
    p.add("cf".getBytes(),"column1".getBytes(),new String("value" + x).getBytes());
    table.put(p);
    print("Data Entered In Table")
}
val hBaseRDD = sc.newAPIHadoopRDD(conf, classOf[TableInputFormat],
classOf[ImmutableBytesWritable],classOf[Result])
print("RecordCount->>" + hBaseRDD.count())
sc.stop()
}
}

```

## Screenshots

### SparkHive Integration

```
hive> show databases;  
OK  
custom  
default  
simplicdb  
test  
Time taken: 33.61 seconds, Fetched: 4 row(s)  
hive> █
```

```
2. 127.0.0.1 (acadgild) 3. 127.0.0.1 (acadgild) 4. 127.0.0.1 (acadgild) 5. 127.0.0.1 (acadgild)  
Last login: Fri Jun 8 03:42:57 2018 from 10.0.2.2  
[acadgild@localhost ~]$ hive --service metastore  
2018-06-08 03:59:32: 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/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]  
2018-06-08T03:59:40,602 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  
2018-06-08T03:59:49,616 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - 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/conf:/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:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-start-1.6.0.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-trace-1.6.0.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/activation-1.1.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-api-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-file-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-okhttp-0.9.0.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-impl-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-spi-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/
```

```
2.127.0.0.1 (acadgild) 3.127.0.0.1 (acadgild) 4.127.0.0.1 (acadgild) 5.127.0.0.1 (acadgild)
Last login: Fri Jun 8 03:42:57 2018 from 10.0.2.2
[acadgild@localhost ~]$ hive --service metastore
2018-06-08 03:59:32: 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/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]
2018-06-08T03:59:40,602 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
2018-06-08T03:59:49,616 INFO [main] org.apache.hadoop.hive.metastore.HiveMetaStore - 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/conf:/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:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-start-1.6.0.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/accumulo-trace-1.6.0.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/activation-1.1.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-api-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-file-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-connector-okhttp-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-impl-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/aether-spi-0.9.0.M2.jar:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/
```

## Output

```
eclipse-workspace - Eclipse
File Edit Refactor Navigate Search Project Scala Run Window Help
Problems Tasks Console Progress
<terminated> SparkHiveTest$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jun 8, 2018, 4:03:19 AM)
18/06/08 04:04:49 INFO SessionState: Created HDFS directory: /tmp/hive/acadgild/41865e02-4183-4dec-9697-01f4bffaeadc
18/06/08 04:04:49 INFO SessionState: Created local directory: /tmp/acadgild/41865e02-4183-4dec-9697-01f4bffaeadc
18/06/08 04:04:49 INFO SessionState: Created HDFS directory: /tmp/hive/acadgild/41865e02-4183-4dec-9697-01f4bffaeadc
18/06/08 04:04:49 INFO HiveClientImpl: Warehouse location for Hive client (version 1.2.1) is /user/hive/warehouse
18/06/08 04:04:50 INFO StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
18/06/08 04:04:50 INFO SparkSQLParser: Parsing command: show databases
18/06/08 04:05:12 INFO CodeGenerator: Code generated in 1773.963415 ms
18/06/08 04:05:12 INFO CodeGenerator: Code generated in 221.485762 ms
18/06/08 04:05:13 INFO SparkContext: Invoking stop() from shutdown hook
18/06/08 04:05:13 INFO SparkUI: Stopped Spark web UI at http://192.168.5.112:4040
+-----+
|databaseName|
+-----+
|custom      |
|default     |
|simplydb    |
|test        |
+-----+
test
18/06/08 04:05:14 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/06/08 04:05:15 INFO MemoryStore: MemoryStore cleared
18/06/08 04:05:15 INFO BlockManager: BlockManager stopped
18/06/08 04:05:15 INFO BlockManagerMaster: BlockManagerMaster stopped
18/06/08 04:05:15 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/06/08 04:05:15 INFO SparkContext: Successfully stopped SparkContext
18/06/08 04:05:15 INFO ShutdownHookManager: Shutdown hook called
18/06/08 04:05:15 INFO ShutdownHookManager: Deleting directory /tmp/spark-b4022c78-5dfd-46f1-9343-d2e548044069
```

## SparkHBase Integration

```
hbase(main):002:0> list
TABLE
SparkHBaseTable
bulktable
clicks
employee
htest
student
transactions
transactions_counts
8 row(s) in 0.0750 seconds

=> ["SparkHBaseTable", "bulktable", "clicks", "employee", "htest", "student", "transactions", "transactions_counts"]
hbase(main):003:0> list
TABLE
SparkHBaseTable
SparkHBaseTable_new
bulktable
clicks
employee
htest
student
transactions
transactions_counts
9 row(s) in 0.0730 seconds

=> ["SparkHBaseTable", "SparkHBaseTable_new", "bulktable", "clicks", "employee", "htest", "student", "transactions", "transactions_counts"]
hbase(main):004:0>
```

## Output

```
eclipse-workspace - TestScalaProject/src/SparkHBaseTest.scala - Eclipse
File Edit Refactor Navigate Search Project Scala Run Window Help

SparkHBaseTest$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jun 8, 2018, 5:20:57 AM)
18/06/08 05:21:24 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication
18/06/08 05:21:24 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.5.102, 34969, No
18/06/08 05:21:24 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.5.102:34969 with 260.8 MB RAM, No
18/06/08 05:21:24 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.5.102, 34969, No
18/06/08 05:21:24 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.5.102, 34969, None)
hello spark hbase --> 1
18/06/08 05:21:27 INFO RecoverableZooKeeper: Process identifier=hconnection-0x2af616d3 connecting to ZooKeeper ens
18/06/08 05:21:28 INFO ZooKeeper: Client environment:zookeeper.version=3.4.6-1569965, built on 02/20/2014 09:09 GMT
18/06/08 05:21:28 INFO ZooKeeper: Client environment:host.name=localhost
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.version=1.8.0_151
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.vendor=Oracle Corporation
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.home=/usr/java/jdk1.8.0_151/jre
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.class.path=/home/acadgild/.p2/pool/plugins/org.scala-ide.
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.io.tmpdir=/tmp
18/06/08 05:21:28 INFO ZooKeeper: Client environment:java.compiler=<NA>
18/06/08 05:21:28 INFO ZooKeeper: Client environment:os.name=Linux
18/06/08 05:21:28 INFO ZooKeeper: Client environment:os.arch=amd64
18/06/08 05:21:28 INFO ZooKeeper: Client environment:os.version=2.6.32-696.18.7.el6.x86_64
18/06/08 05:21:28 INFO ZooKeeper: Client environment:user.name=acadgild
18/06/08 05:21:28 INFO ZooKeeper: Client environment:user.home=/home/acadgild
18/06/08 05:21:28 INFO ZooKeeper: Client environment:user.dir=/home/acadgild/eclipse-workspace/TestScalaProject
18/06/08 05:21:28 INFO ZooKeeper: Initiating client connection, connectString=localhost:2181 sessionTimeout=90000 v
18/06/08 05:21:28 INFO ClientCnxn: Opening socket connection to server localhost/0:0:0:0:0:0:1:2181. Will not at
18/06/08 05:21:28 INFO ClientCnxn: Socket connection established to localhost/0:0:0:0:0:0:1:2181, initiating sess
18/06/08 05:21:28 INFO ClientCnxn: Session establishment complete on server localhost/0:0:0:0:0:0:1:2181, sessio
creating table:SparkHBaseTable new 18/06/08 05:21:43 INFO HBaseAdmin: Created SparkHBaseTable_new
Data Entered In TableData Entered In TableData Entered In TableData Entered In TableData Entered In TableData Enter
```

```
eclipse-workspace - TestScalaProject/src/SparkHBaseTest.scala - Eclipse
File Edit Refactor Navigate Search Project Scala Run Window Help

Problems Tasks Console Progress

<terminated> SparkHBaseTest$ [Scala Application] /usr/java/jdk1.8.0_151/bin/java (Jun 8, 2018, 5:20:57 AM)
18/06/08 05:21:53 INFO DAGScheduler: Submitting 1 missing tasks from ResultStage 0 (NewHadoopRDD[0] at newAPIHadoop
18/06/08 05:21:53 INFO TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
18/06/08 05:21:54 INFO TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition
18/06/08 05:21:54 INFO Executor: Running task 0.0 in stage 0.0 (TID 0)
18/06/08 05:21:55 INFO NewHadoopRDD: Input split: HBase table split(table name: SparkHBaseTable_new, scan: , start
18/06/08 05:21:55 INFO RecoverableZooKeeper: Process identifier=hconnection-0x48cd986e connecting to ZooKeeper ense
18/06/08 05:21:55 INFO ZooKeeper: Initiating client connection, connectString=localhost:2181 sessionTimeout=90000 v
18/06/08 05:21:55 INFO ClientCnxn: Opening socket connection to server localhost/0:0:0:0:0:0:0:1:2181. Will not att
18/06/08 05:21:55 INFO ClientCnxn: Socket connection established to localhost/0:0:0:0:0:0:0:1:2181, initiating sess
18/06/08 05:21:55 INFO ClientCnxn: Session establishment complete on server localhost/0:0:0:0:0:0:0:1:2181, sessio
18/06/08 05:21:55 INFO TableInputFormatBase: Input split length: 0 bytes.
18/06/08 05:21:56 INFO ConnectionManager$HConnectionImplementation: Closing zookeeper sessionId=0x163dc9d13e20009
18/06/08 05:21:56 INFO ClientCnxn: EventThread shut down
18/06/08 05:21:56 INFO ZooKeeper: Session: 0x163dc9d13e20009 closed
18/06/08 05:21:57 INFO Executor: Finished task 0.0 in stage 0.0 (TID 0). 875 bytes result sent to driver
18/06/08 05:21:57 INFO TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 3055 ms on localhost (executor dri
18/06/08 05:21:57 INFO TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/06/08 05:21:57 INFO DAGScheduler: ResultStage 0 (count at SparkHBaseTest.scala:41) finished in 3.382 s
18/06/08 05:21:57 INFO DAGScheduler: Job 0 finished: count at SparkHBaseTest.scala:41, took 5.682785 s
RecordCount-->1018/06/08 05:21:57 INFO SparkUI: Stopped Spark web UI at http://192.168.5.102:4040
18/06/08 05:21:58 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/06/08 05:21:58 INFO MemoryStore: MemoryStore cleared
18/06/08 05:21:58 INFO BlockManager: BlockManager stopped
18/06/08 05:21:58 INFO BlockManagerMaster: BlockManagerMaster stopped
18/06/08 05:21:58 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/06/08 05:21:58 INFO SparkContext: Successfully stopped SparkContext
18/06/08 05:21:58 INFO ShutdownHookManager: Shutdown hook called
18/06/08 05:21:58 INFO ShutdownHookManager: Deleting directory /tmp/spark-c3adc151-1941-4c34-a9c5-dcfea7677486

eclipse-workspace - Te...
```