Java-Success.com

Prepare to fast-track, choose & go places with 800+ Java & Big Data Q&As with lots of code & diagrams.

05: Setting up & getting started with Spark local mode with Sbt & Scala

Posted on November 7, 2018

This extends Setting up & getting started with sbt.

Spark 2.3.2 requires Scala 2.11.x. You need to use the right Scala version.

Step 1: On Eclipse set the Scala compiler to 2.11. Right mouse click on the project “sbt-tutorial” –> Properties –> Scala Compiler –> Tick “Use Project Settings” and select “latest **2.11** bundle (dynamic)”.

300+ Java Interview FAQs

300+ Java FAQs

16+ Java Key Areas Q&As

150+ Java Architect FAQs

80+ Java Code Quality Q&As

150+ Java Coding Q&As

300+ Big Data Interview FAQs

300+ Big Data FAQs

Tutorials - Big Data

TUT - Starting Big Data

TUT - Starting Spark & Scala

TOP

Step 2: Add the Spark dependency to "**build.sbt**" file.

```
organization := "com.sbt-tut"

version:= "0.1"

scalaVersion := "2.11.12"

libraryDependencies += "org.apache.spark" %% "sparl

```

Step 3: From the command-line run the following command, and it may take some time to download the dependencies.

```
~/projects/sbt-tutorial]$ sbt eclipse

```

Step 4: Write a very simple Spark code in Scala as shown below within a package named "com.scalaproject" and the Scala file is named "**SimpleSpark.scala**".

```
package com.scalaproject

import org.apache.spark.sql.SparkSession;
import org.apache.spark.sql.Dataset;

object SimpleSpark {

  def main(args: Array[String]): Unit = {
     val spark = SparkSession.builder.appName("Sir
                              .config("spark.maste
                             .getOrCreate()

     val list = List(1,2,3,4,5)
     val rdd = spark.sparkContext.parallelize(list

     val multipleBy2Rdd = rdd.map(_ * 2)
     println(multipleBy2Rdd.collect().toList)

  }
}

```

TUT - Starting with Python
TUT - Kafka
TUT - Pig
TUT - Apache Storm
TUT - Spark Scala on Zeppelin
TUT - Cloudera
TUT - Cloudera on Docker
TUT - File Formats
TUT - Spark on Docker
TUT - Flume
TUT - Hadoop (HDFS)
TUT - HBase (NoSQL)
TUT - Hive (SQL)
TUT - Hadoop & Spark
TUT - MapReduce
TUT - Spark and Scala
TUT - Spark & Java
TUT - PySpark on Databricks
TUT - Zookeeper

Step 5: Right mouse click on the file **SimpleSpark.scala** and "Run as" –> "Scala Application" within eclipse.

Output:

```
Using Spark's default log4j profile: org/apache/sp
18/11/07 23:01:44 INFO SparkContext: Running Spark
18/11/07 23:01:44 WARN NativeCodeLoader: Unable to
18/11/07 23:01:44 INFO SparkContext: Submitted app
18/11/07 23:01:45 INFO SecurityManager: Changing v
18/11/07 23:01:45 INFO SecurityManager: Changing m
18/11/07 23:01:45 INFO SecurityManager: Changing v
18/11/07 23:01:45 INFO SecurityManager: Changing m
18/11/07 23:01:45 INFO SecurityManager: SecurityMa
18/11/07 23:01:45 INFO Utils: Successfully started
18/11/07 23:01:45 INFO SparkEnv: Registering MapOu
18/11/07 23:01:45 INFO SparkEnv: Registering Block
18/11/07 23:01:45 INFO BlockManagerMasterEndpoint
18/11/07 23:01:45 INFO BlockManagerMasterEndpoint
18/11/07 23:01:45 INFO DiskBlockManager: Created l
18/11/07 23:01:45 INFO MemoryStore: MemoryStore st
18/11/07 23:01:45 INFO SparkEnv: Registering Outpu
18/11/07 23:01:45 INFO Utils: Successfully started
18/11/07 23:01:45 INFO SparkUI: Bound SparkUI to 0
18/11/07 23:01:45 INFO Executor: Starting executor
18/11/07 23:01:45 INFO Utils: Successfully started
18/11/07 23:01:45 INFO NettyBlockTransferService:
18/11/07 23:01:45 INFO BlockManager: Using org.apa
18/11/07 23:01:45 INFO BlockManagerMaster: Registe
18/11/07 23:01:45 INFO BlockManagerMasterEndpoint
18/11/07 23:01:45 INFO BlockManagerMaster: Registe
18/11/07 23:01:45 INFO BlockManager: Initialized B
18/11/07 23:01:46 INFO SparkContext: Starting job
18/11/07 23:01:46 INFO DAGScheduler: Got job 0 (co
18/11/07 23:01:46 INFO DAGScheduler: Final stage:
18/11/07 23:01:46 INFO DAGScheduler: Parents of fi
18/11/07 23:01:46 INFO DAGScheduler: Missing paren
18/11/07 23:01:46 INFO DAGScheduler: Submitting Re
18/11/07 23:01:46 INFO MemoryStore: Block broadcas
18/11/07 23:01:46 INFO MemoryStore: Block broadcas
18/11/07 23:01:46 INFO BlockManagerInfo: Added bro
18/11/07 23:01:46 INFO SparkContext: Created broad
18/11/07 23:01:46 INFO DAGScheduler: Submitting 2
18/11/07 23:01:46 INFO TaskSchedulerImpl: Adding t
18/11/07 23:01:46 INFO TaskSetManager: Starting ta
18/11/07 23:01:46 INFO Executor: Running task 0.0
18/11/07 23:01:46 INFO Executor: Finished task 0.0
18/11/07 23:01:46 INFO TaskSetManager: Starting ta
18/11/07 23:01:46 INFO Executor: Running task 1.0
18/11/07 23:01:46 INFO TaskSetManager: Finished ta
```

```
18/11/07 23:01:46 INFO Executor: Finished task 1.0
18/11/07 23:01:46 INFO TaskSetManager: Finished ta
18/11/07 23:01:46 INFO TaskSchedulerImpl: Removed
18/11/07 23:01:46 INFO DAGScheduler: ResultStage 0
18/11/07 23:01:46 INFO DAGScheduler: Job 0 finishe
List(2, 4, 6, 8, 10)
18/11/07 23:01:46 INFO SparkContext: Invoking stop
18/11/07 23:01:46 INFO SparkUI: Stopped Spark web
18/11/07 23:01:46 INFO ContextCleaner: Cleaned acc
18/11/07 23:01:46 INFO MapOutputTrackerMasterEndpo
18/11/07 23:01:46 INFO MemoryStore: MemoryStore cl
18/11/07 23:01:46 INFO BlockManager: BlockManager
18/11/07 23:01:46 INFO BlockManagerMaster: BlockMa
18/11/07 23:01:46 INFO OutputCommitCoordinator$Out
18/11/07 23:01:46 INFO SparkContext: Successfully
18/11/07 23:01:46 INFO ShutdownHookManager: Shutdo
18/11/07 23:01:46 INFO ShutdownHookManager: Delet

```

Note: In this tutorial we run it in local mode, which is good for development & debugging. When you run in cluster mode, you need to package your code as a jar file, and run it with the spark-submit command against a Spark cluster.

Step 6: You can package it as a jar file in the command-line.

```
~/projects/sbt-tutorial]$ sbt package

```

projects/sbt-tutorial/target/scala-2.11/**sbt-tutorial_2.11-0.1.jar** will be built.

‹ 21: Q121 – Q124 Scala Extractor Interview Q&As

06: Setting up Spark-shell on Mac & getting started ›

Disclaimer

Top