

Setting UP Spark 2.0 environment on intellij community edition version 2016.2.2

23/08/2016

Laser Nahoom-Kabakov

Contents:

Download and install latest intellij IDEA community edition.....	2
Download and install the latest Oracle/Sun JDK.....	2
Set up your JAVA_HOME variable	2
Download and configure winutil (windows only).....	3
Start intellij IDEA.....	3
Install Scala and SBT plugins for IntelliJ	5
Create and configure a new Scala and SBT project.....	8
Configure libraries for the newly created project.....	10
Validate your configuration:.....	11
Test your setup.....	12

Download and install latest intelliJ IDEA community edition

Click the link below for the download:

For windows:

<https://www.jetbrains.com/idea/download/download-thanks.html?platform=windows&code=IIC>

for linux:

<https://www.jetbrains.com/idea/download/#section=linux>

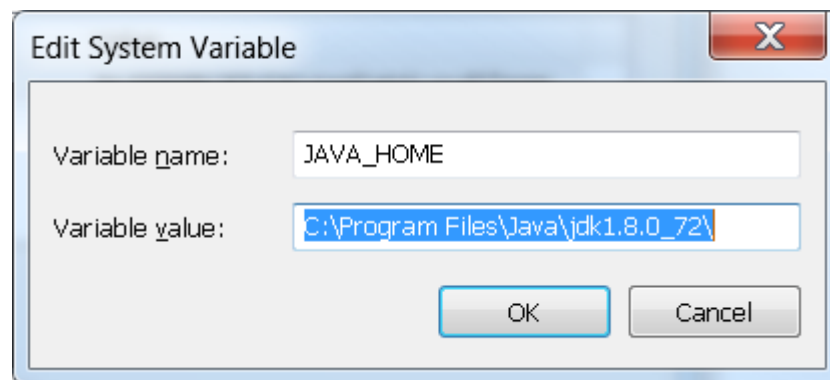
Download and install the latest Oracle/Sun JDK

Click the link below for the download:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Set up your JAVA_HOME variable

Make sure to set up your JAVA_HOME to: "C:\Program Files\Java\<YOUR_JVM_LOCATION>", pictured below:



Download and configure winutil (windows only)

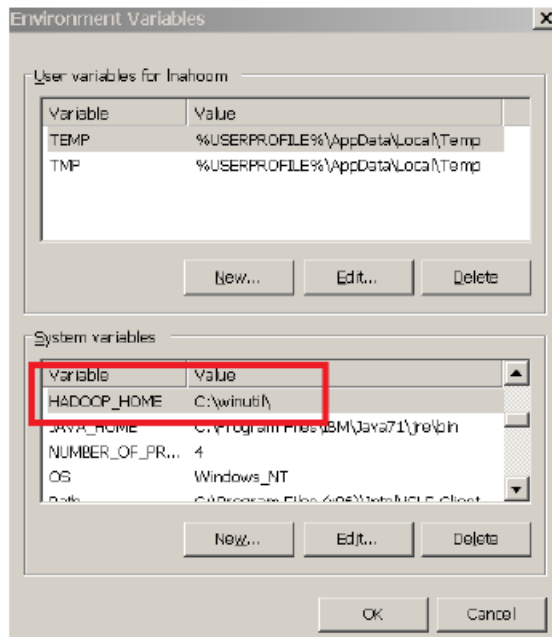
Download winutil.exe from here:

<http://public-repo-1.hortonworks.com/hdp-win-alpha/winutils.exe>

Make the following path on your file system and add the winutils.exe there

`C:\winutil\bin\winutils.exe`

Create a new Environment variable named HADOOP_HOME and point it to C:\winutil\




Start IntelliJ IDEA


Skip all configurations at startup and don't install anything. Until you get to the "create new project screen", pictured below:





IntelliJ IDEA

Version 2016.2.2

 Create New Project

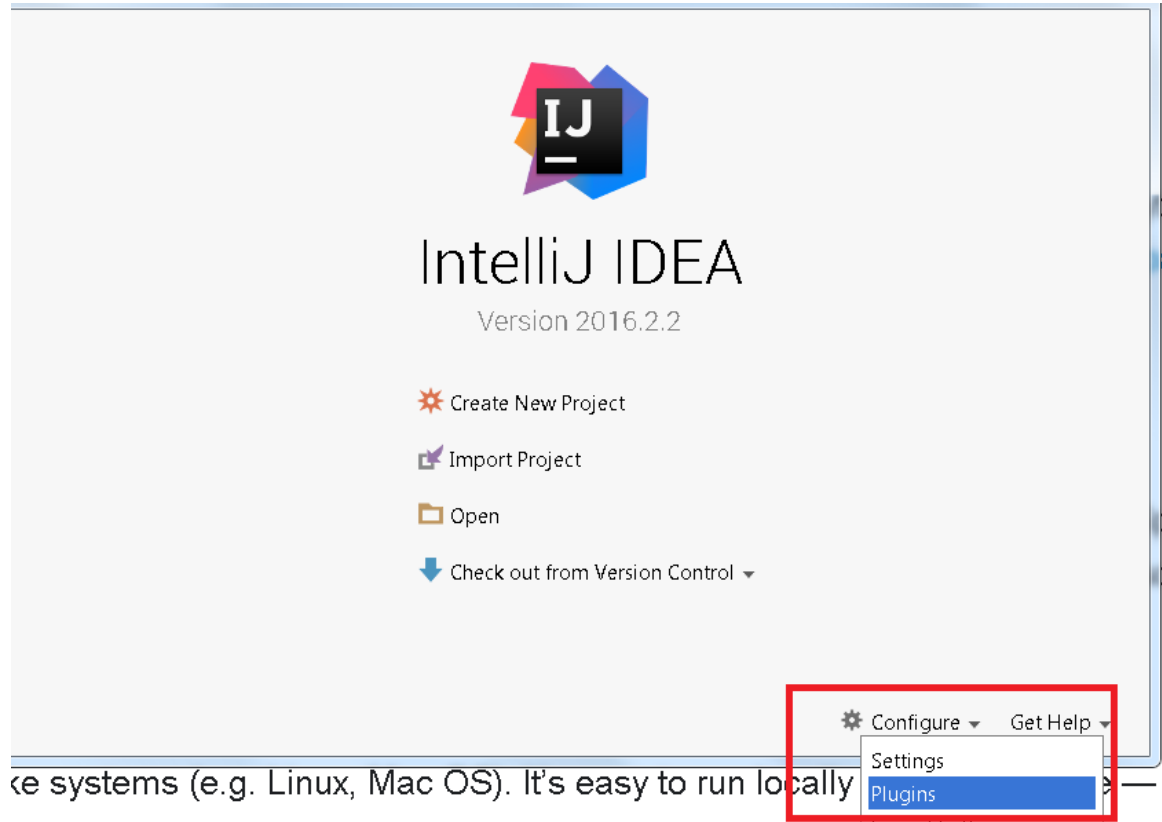
 Import Project

 Open

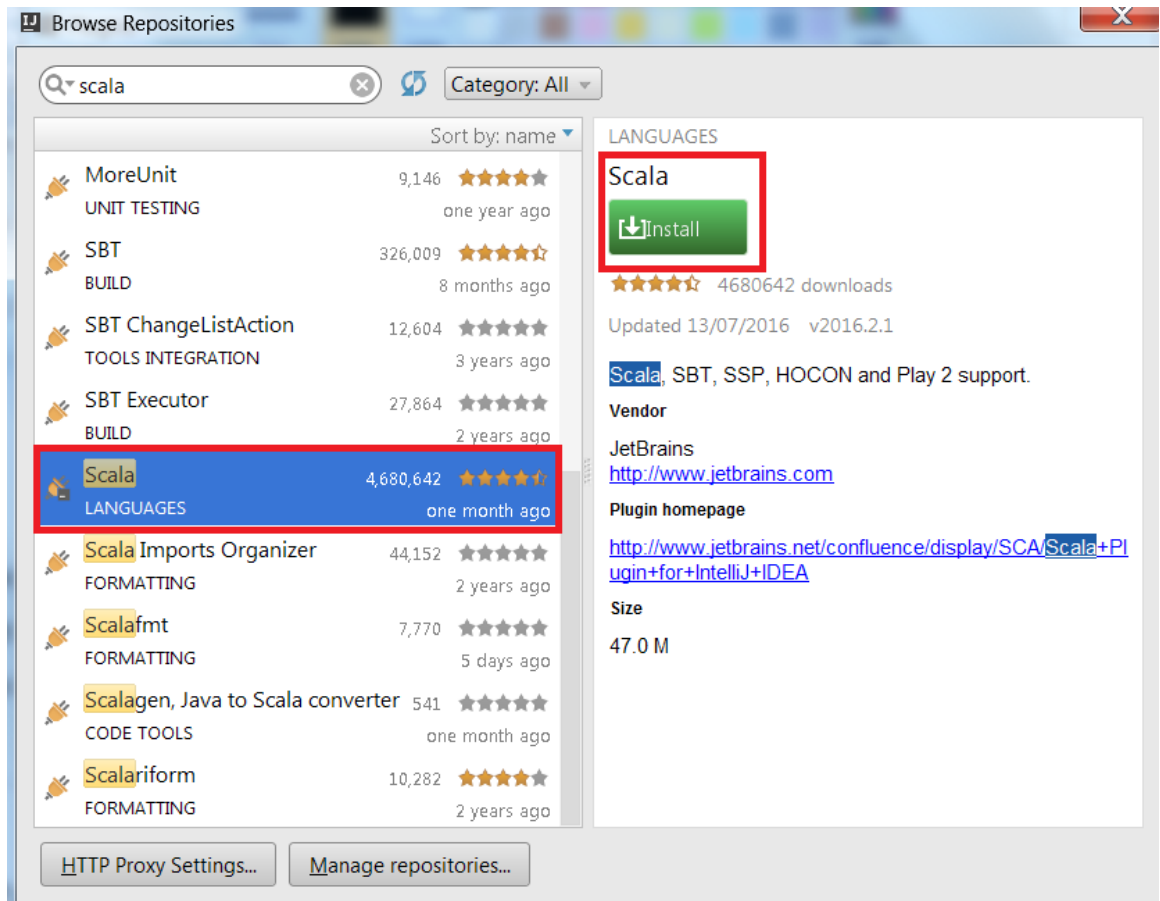
 Check out from Version Control ▼

Install Scala and SBT plugins for IntelliJ

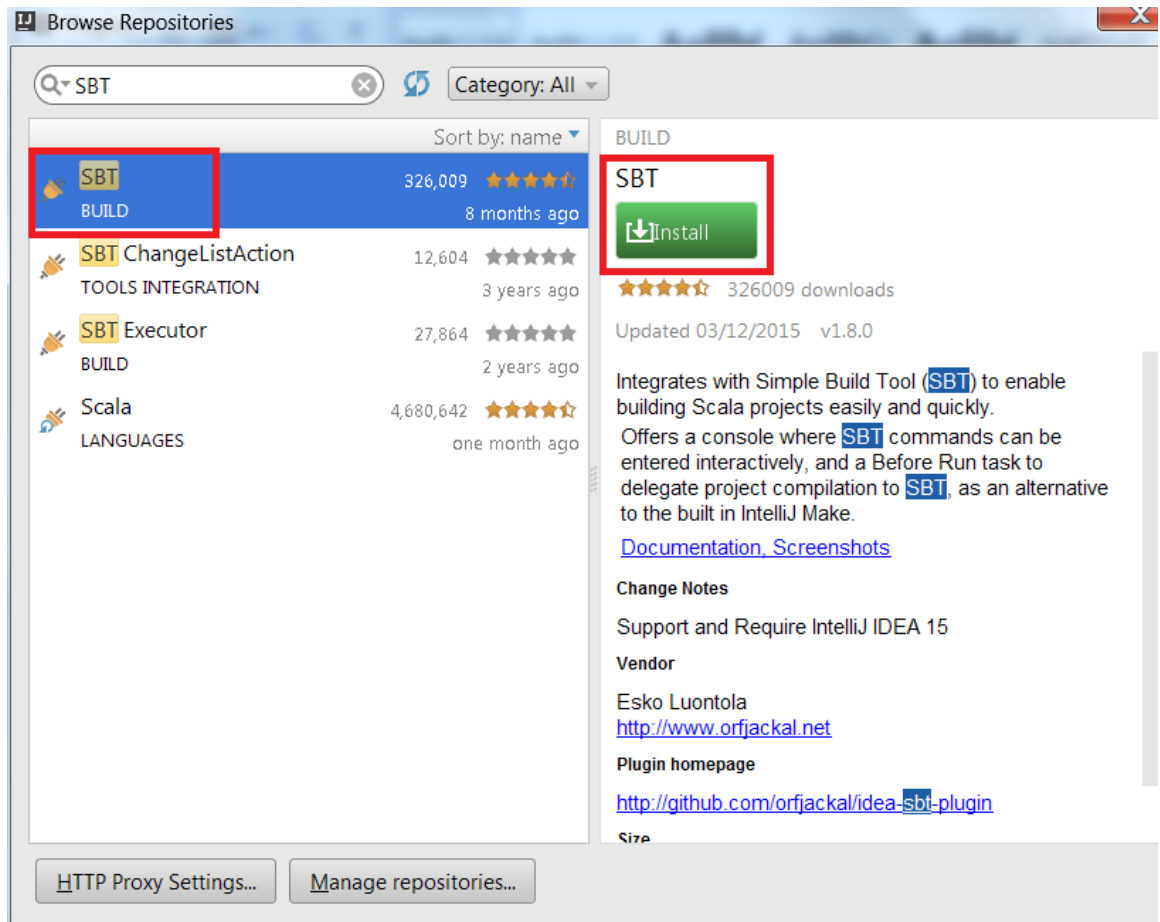
1. On this page go to the configure→plugins→ Browse Repositories



2. Install the Scala plugin listed below:



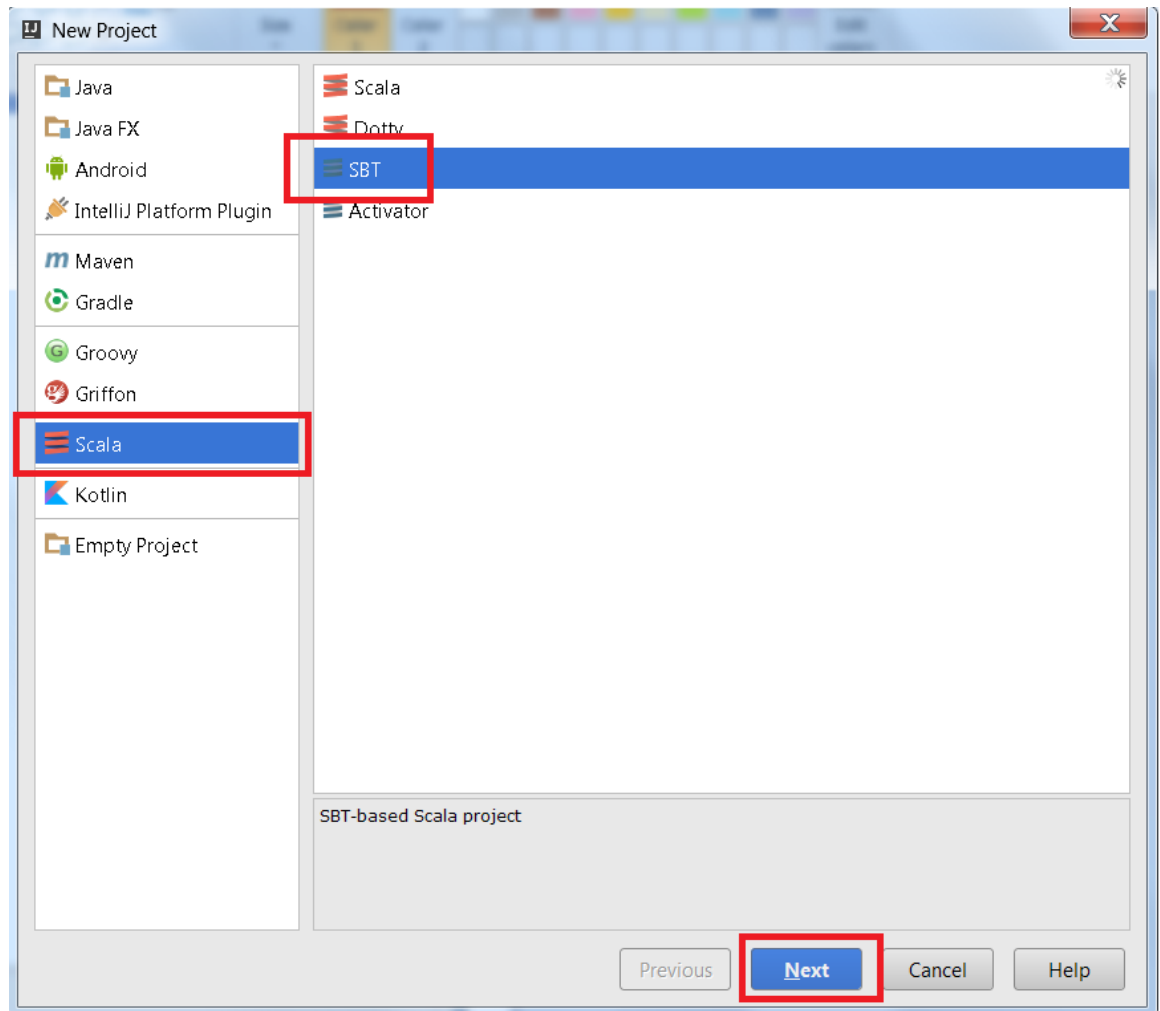
3. Install the SBT plugin listed below



4. Restart IntelliJ IDEA as requested

Create and configure a new Scala and SBT project

1. Create a new Scala SBT project



2. Create the project with the following configuration settings

New Project

Project name: SparkClass_2.0

Project location: C:\Data\IDEA_Workspaces\SparkClass_2.0

Project SDK: 1.8 (java version "1.8.0_72-ea")

SBT version: 0.13.8

Scala version: 2.11.8

☒ Use auto-import

☒ Create directories for empty content roots automatically

Download: ☒ Sources ☐ Javadocs ☒ Sources for SBT and plugins

▼ More Settings

Module name: SparkClass_2.0

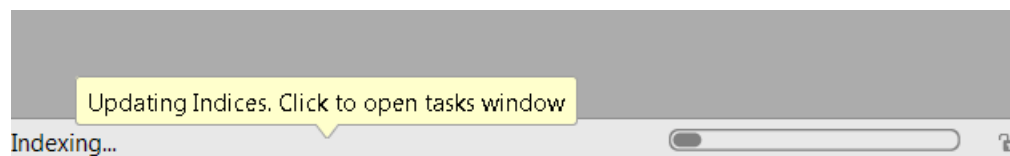
Content root: C:\Data\IDEA_Workspaces\SparkClass_2.0

Module file location: C:\Data\IDEA_Workspaces\SparkClass_2.0

Project format: .idea (directory based)

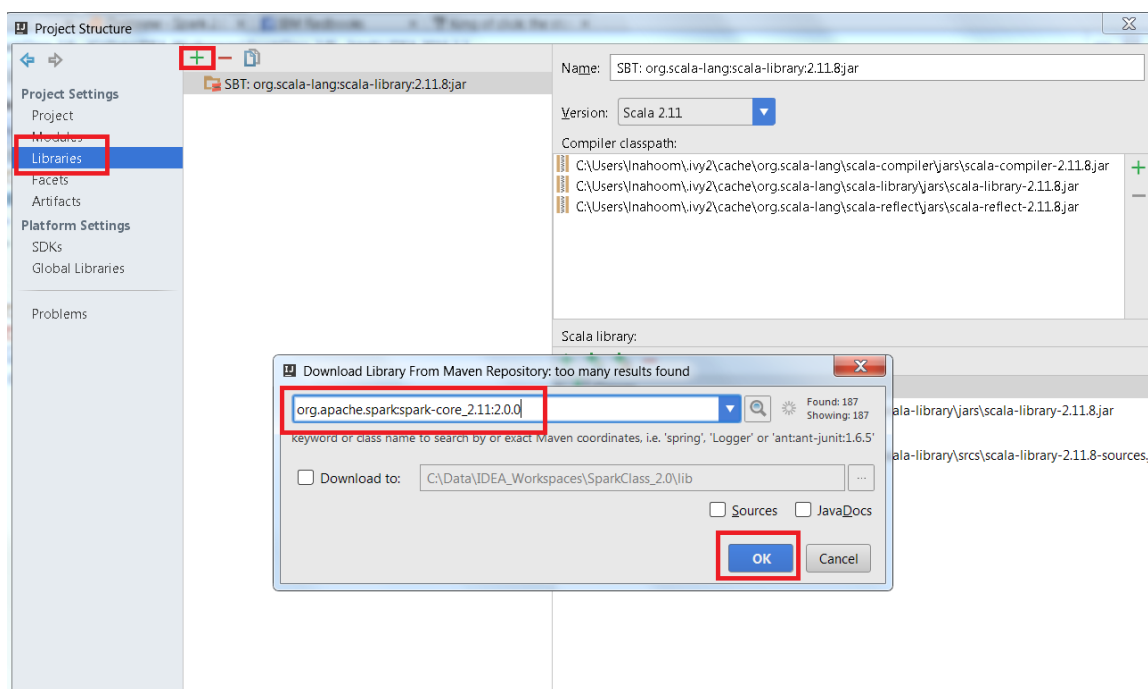
Previous **Finish** Cancel Help

3. Wait for the indexing completion!!



Configure libraries for the newly created project

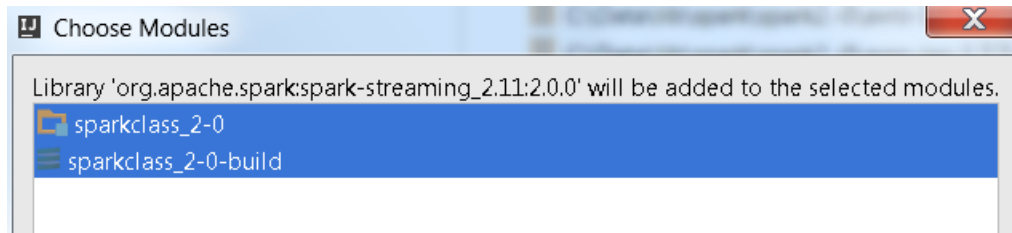
1. Open the module setting and go to the Libraries and add the following libraries from maven as shown in the screenshot below:



Libraries to add:

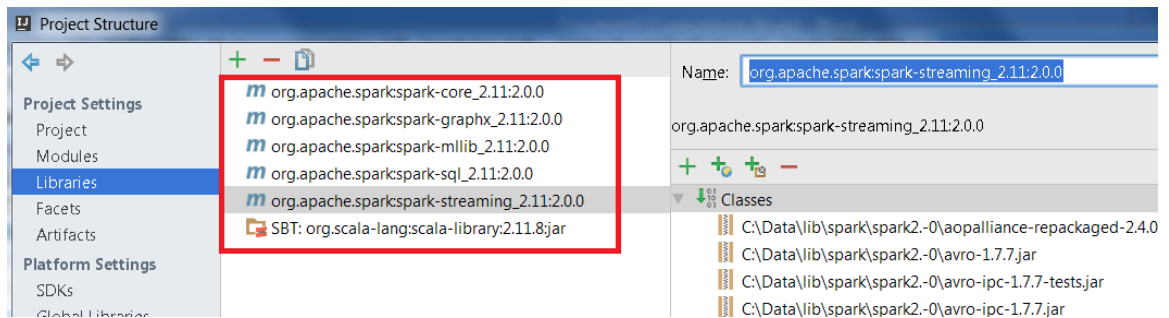
Package	Artifact ID
org.apache.spark:spark-core_2.11:2.0.0	spark-core_2.11
org.apache.spark:spark-sql_2.11:2.0.0	spark-sql_2.11
org.apache.spark:spark-mllib_2.11:2.0.0	spark-mllib_2.11
org.apache.spark:spark-streaming_2.11:2.0.0	spark-streaming_2.11
org.apache.spark:spark-graphx_2.11:2.0.0	spark-graphx_2.11

2. When adding each library add the libraries to the following projects:
choose to add support for the following modules:



Validate your configuration:

After the addition of all of the libraries, your project libraries should look like this:



Test your setup

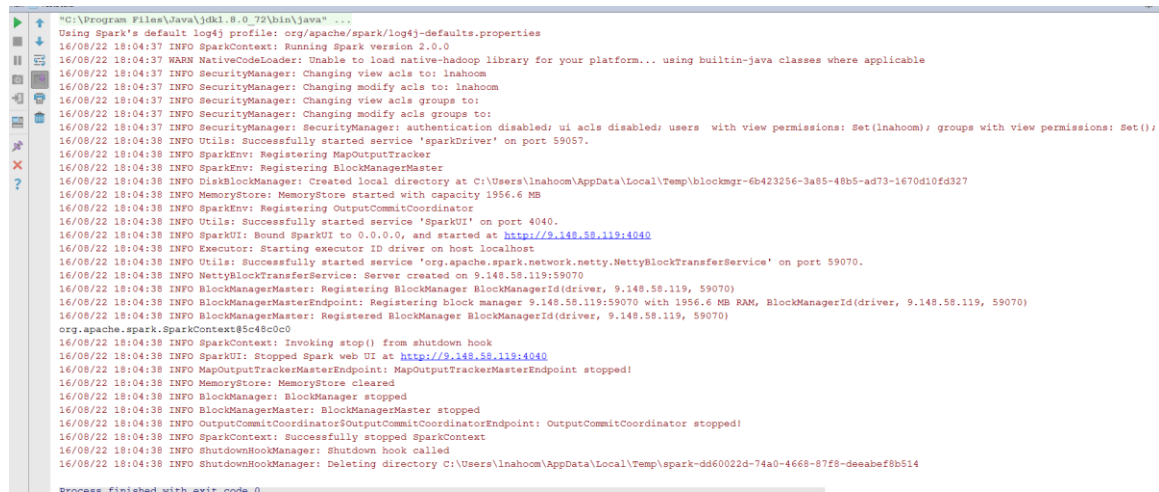
Code and run the following Scala test object to validate your configuration works

```
import org.apache.spark.{SparkConf, SparkContext}

/**
 * Created by lnahoom on 22/08/2016.
 */
object TestScala {

  def main(args: Array[String]): Unit = {
    val conf = new SparkConf()
    conf.setAppName("Datasets Test")
    conf.setMaster("local[2]")
    val sc = new SparkContext(conf)
    println(sc)
  }
}
```

Correct output should look like this:

A screenshot of the IntelliJ IDEA console window showing the output of running the TestScala application. The console displays a series of log messages from Spark 2.0, including information about the SparkContext, SecurityManager, SparkEnv, and the SparkUI. The logs indicate that the application was successfully started and is running on a local host. The console output is as follows:

```
"C:\Program Files\Java\jdk1.8.0_72\bin\java" ...
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
16/08/22 18:04:37 INFO SparkContext: Running Spark version 2.0.0
16/08/22 18:04:37 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
16/08/22 18:04:37 INFO SecurityManager: Changing view acls to: lnahoom
16/08/22 18:04:37 INFO SecurityManager: Changing modify acls to: lnahoom
16/08/22 18:04:37 INFO SecurityManager: Changing view acls groups to:
16/08/22 18:04:37 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(lnahoom); groups with view permissions: Set();
16/08/22 18:04:38 INFO Utils: Successfully started service 'sparkDriver' on port 59057.
16/08/22 18:04:38 INFO SparkEnv: Registering MapOutputTracker
16/08/22 18:04:38 INFO SparkEnv: Registering BlockManagerMaster
16/08/22 18:04:38 INFO DiskBlockManager: Created local directory at C:\Users\lnahoom\AppData\Local\Temp\blockmgr-6b423256-3a85-48b5-ad73-1670d10fd327
16/08/22 18:04:38 INFO MemoryStore: MemoryStore started with capacity 1956.6 MB
16/08/22 18:04:38 INFO SparkEnv: Registering OutputCommitCoordinator
16/08/22 18:04:38 INFO Utils: Successfully started service 'SparkUI' on port 4040.
16/08/22 18:04:38 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://9.148.58.119:4040
16/08/22 18:04:38 INFO Executor: Starting executor ID driver on host localhost
16/08/22 18:04:38 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 59070.
16/08/22 18:04:38 INFO NettyBlockTransferService: Server created on 9.148.58.119:59070
16/08/22 18:04:38 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 9.148.58.119, 59070)
16/08/22 18:04:38 INFO BlockManagerMasterEndpoint: Registering block manager 9.148.58.119:59070 with 1956.6 MB RAM, BlockManagerId(driver, 9.148.58.119, 59070)
16/08/22 18:04:38 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 9.148.58.119, 59070)
org.apache.spark.SparkContext@5c48c0c0
16/08/22 18:04:38 INFO SparkContext: Invoking stop() from shutdown hook
16/08/22 18:04:38 INFO SparkUI: Stopped Spark web UI at http://9.148.58.119:4040
16/08/22 18:04:38 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
16/08/22 18:04:38 INFO MemoryStore: MemoryStore cleared
16/08/22 18:04:38 INFO BlockManager: BlockManager stopped
16/08/22 18:04:38 INFO BlockManagerMaster: BlockManagerMaster stopped
16/08/22 18:04:38 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
16/08/22 18:04:38 INFO SparkContext: Successfully stopped SparkContext
16/08/22 18:04:38 INFO ShutdownHookManager: Shutdown hook called
16/08/22 18:04:38 INFO ShutdownHookManager: Deleting directory C:\Users\lnahoom\AppData\Local\Temp\spark-dd60022d-74a0-4668-87f8-deaebf8b514
Process finished with exit code 0
```

Congratulations, you have been able to set up the IntelliJ with your Spark 2.0 support.