Streaming Big Data with Spark Streaming and Scala - Hands On

Getting Started with IntelliJ IDEA

Install the Course Materials

The scripts and data for this course can be downloaded from

https://github.com/packtpublishing/streaming-big-data-with-spark-streaming-scala-and-spark-3-

Download and un-zip this file and move the SparkStreaming folder (which contains another SparkStreaming folder) to a path you'll remember.

Install IntelliJ IDEA Community Edition

Make sure you have **JDK 8 or 11** installed. **Apache Spark is not compatible with newer versions of Java**. Enter

    java -version

from a command or terminal prompt to see what version, if any, you have installed already. If you don't already have JDK 8 or 11 installed, you can download it while setting up IntelliJ.

Next, install IntelliJ IDEA Community Edition, after selecting your platform (Windows, Mac, or Linux). After installation, select "plugins" from the start-up screen and **install the Scala plugin** for IntelliJ, using **Scala 2.12**.

WINDOWS ONLY: Copy the **hadoop** folder inside your SparkStreaming folder onto the top level of your C: drive. Create a new environment variable (enter "environment variables" in the Windows search bar, click on "Add Environment Variables," and add a new system variable) named HADOOP_HOME with a value of C:\Hadoop. Next select the PATH environment variable, and APPEND a new entry, separated by a semi-colon, of %HADOOP_HOME%\bin Now, restart IntelliJ to make sure the new environment variables are picked up.

This is necessary because Spark assumes HDFS exists, but HDFS does not exist on Windows. The contents of the hadoop folder implements a wrapper for HDFS on Windows.

Import the Course Project

From the IntelliJ welcome screen, select "**Open or Import**".

Select your **SparkStreaming/SparkStreaming** folder.

Try it Out

Expand the project's tree view to show the SparkStreaming/src/main/scala/com.sundogsoftware.spark folder.

Right click on "HelloWorld" and select "Run HelloWorld"

You should see a message like:

```
Hello world! The u.data file has 100000 lines.
```

But you might see a "class not found" error. If so, just quit IntelliJ, restart it, and try again. It's just a bug in IntelliJ.

Once you see the "Hello World" message, everything is set up successfully! If not, go back and look for a step you may have missed. Sometimes IntelliJ just gets confused – you might need to refresh the SBT configuration as shown in the setup video, or even re-add the dstream-twitter library. If you're stuck, we're here to help – use the Q&A or comments feature on the site you're taking this course on.