

Introduction to Apache Spark preparations

- Download and install [Scala 2.11.x](#). Try running some commands in the Scala REPL just to make sure your installation is working properly. (This also requires a Java Runtime with a version higher than 1.6, preferably 1.8 just so you have the same thing as I do)
- Also download and install [SBT](#). Also, make sure to boot the sbt console (`sbt` on the command line), at least once before the introduction, as this will take quite some time the first time.
- Download [Apache Spark 2.0.1](#) The Hadoop version doesn't matter, but choose one that is pre-built for some specific Hadoop version. Personally, I'm running the version built for Hadoop 2.7, so if you want to make sure you're running the same stuff as I am, use that one. Unpack the binaries in some preferred directory. Note that we're going to use the binaries quite often so you may want to put the `bin` folder on your `PATH` for easy access.
- Now, when you're all set up, it should be possible to follow along the following set of commands without getting any errors:
 - First, from a terminal run `$SPARK_DIR/bin/spark-shell`
 - This should start up a Scala REPL with the necessary Spark dependencies on your classpath, and a `SparkSession` already instantiated for you (you will get more familiar with this later :))
 - In the Scala/Spark REPL, run

```
spark.createDataset(List(1,2,3)).count()
```

This should print some logging along with the result `res0: Long = 3`

- (Optional, but it will save some time downloading some stuff) you can find a small template for a Spark project on my [GitHub](#). Clone this repo and run `sbt compile` in the project's root.

If this all worked out, you're good to go :)

During the introduction, we will mostly be working from the Spark REPL, but we'll also cover building some 'larger' projects. For this, you may want to use a nice editor or IDE (hint: [IntelliJ](#)). Packaging the larger projects such that Spark can work with them is done via SBT, which explains why you also installed that.

If anything isn't working out for you, please just contact me, as I may have missed something silly because I've been running this installation for some time now.

