Compiling own Scala Programs on Spark framework

**Installing and linking SBT**

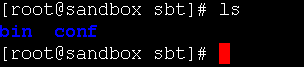
* To compile Scala programs on Spark we need the use of SBT or Maven
* Download SBT into the sandbox using the below command

wget https://dl.bintray.com/sbt/native-packages/sbt/0.13.8/sbt-0.13.8.tgz

* Unzip the tgz file using the below command.

tar zxvf sbt-0.13.8.tgz

* On unzipping we will get 2 directories under the directory **sbt**



* In the dir bin we can find a jar file named “**sbt-launch.jar**”
* Copy the jar file to **$SPARK\_HOME/bin** using the below command.

cp sbt-launch.jar $SPARK\_HOME/bin

* Create a file sbt in $SPARK\_HOME/bin and add the below lines.

SBT\_OPTS="-Xms512M -Xmx1536M -Xss1M -XX:+CMSClassUnloadingEnabled -XX:MaxPermSize=256M"

java $SBT\_OPTS -jar `dirname $0`/sbt-launch.jar "$@"

* Provide the executable permissions to the file.

chmod u+x sbt

* Now to compile a file first create a directory under which create the below directory structure.

/<new-dir>/src/main/scala

* mkdir to create a directory
* mkdir –p /dir1/dir2/dir3 to create a directory structure
* Place the scala file in src/main/scala location
* Under the new-dir create a sbt file and add the below details.

name := "Simple Project

version := "1.0"

scalaVersion := "2.10.4"

libraryDependencies += "org.apache.spark" %% "spark-core" % "1.1.0"

* libraryDependencies will change depending on the programs.
* for streaming we have to add the below dependencies.

libraryDependencies += "org.apache.spark" % "spark-streaming\_2.10" % "1.1.0"

* once SBT file is in place run the below command to compile the scala programs and to form the jar file and classes file.
* Example of directory structure.

./simple.sbt

./src

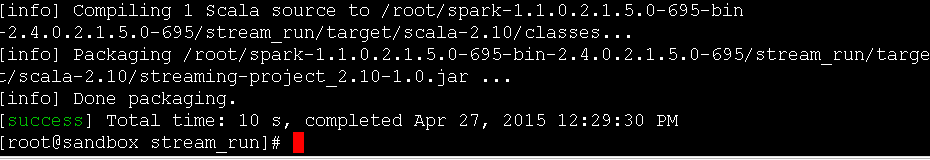
./src/main

./src/main/scala

./src/main/scala/<program-name>.scala

* To change the source directory please fallow the details as mentioned in link : <http://www.scala-sbt.org/0.13/docs/Howto-Customizing-Paths.html>
* To compile the program from the location where we have the \*.sbt file run the below command.

sbt package



* After compilation a jar file and classes will be created.
* To run the jar file on spark use the below command.

Spark-submit \

--class “<class-file-name>” \

Jar file location.

Refrences:

[1]: <https://spark.apache.org/docs/1.1.0/quick-start.html>

[2]: <http://www.scala-sbt.org/0.13/tutorial/Manual-Installation.html>