# How to run Scala SBT application of Spark in Cloudx lab

**Problem statement:** Count the no of number of lines with "a" and "b" in the file stored in HDFS.

**Steps to be followed:**

```
mkdir spark_testing
cd spark_testing/
```

create a sbt file with the command :   touch simple.sbt

```
src/main/scala
mkdir src
cd src/

mkdir main
cd main/

mkdir scala
cd scala/

vi SimpleApp.scala
```

```
import org.apache.spark.SparkContext
import org.apache.spark.SparkContext._
import org.apache.spark.SparkConf

object SimpleApp {
  def main(args: Array[String]) {
    val logFile = "YOUR_SPARK_HOME/README.md" // file stored on your HDFS
    val conf = new SparkConf().setAppName("Simple Application")
         val sc = new SparkContext(conf)
    val logData = sc.textFile(logFile, 2).cache()
    val numAs = logData.filter(line => line.contains("a")).count()
    val numBs = logData.filter(line => line.contains("b")).count()
    println(s"Lines with a: $numAs, Lines with b: $numBs")

  }
}
```

```
cd.. cd.. cd..
vi simple.sbt
name := "Simple Project"
version := "1.0"
scalaVersion := "2.11.8"
libraryDependencies += "org.apache.spark" %% "spark-sql" % "2.2.0" // specify ur dependencies in ur sbt file
```

```
sbt package // package a jar containing ur application
```

```
spark-submit \
  --class "SimpleApp" \
  --master local[4] \
  target/scala-2.11/simple-project_2.11-1.0.jar
```