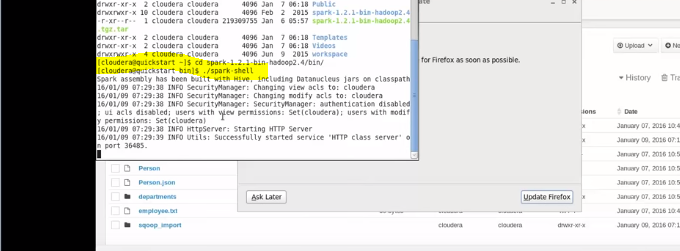
When you enter command

**Spark-shell it will open spark shell in local mode. Its by default, you need to change if required by giving –master command.**

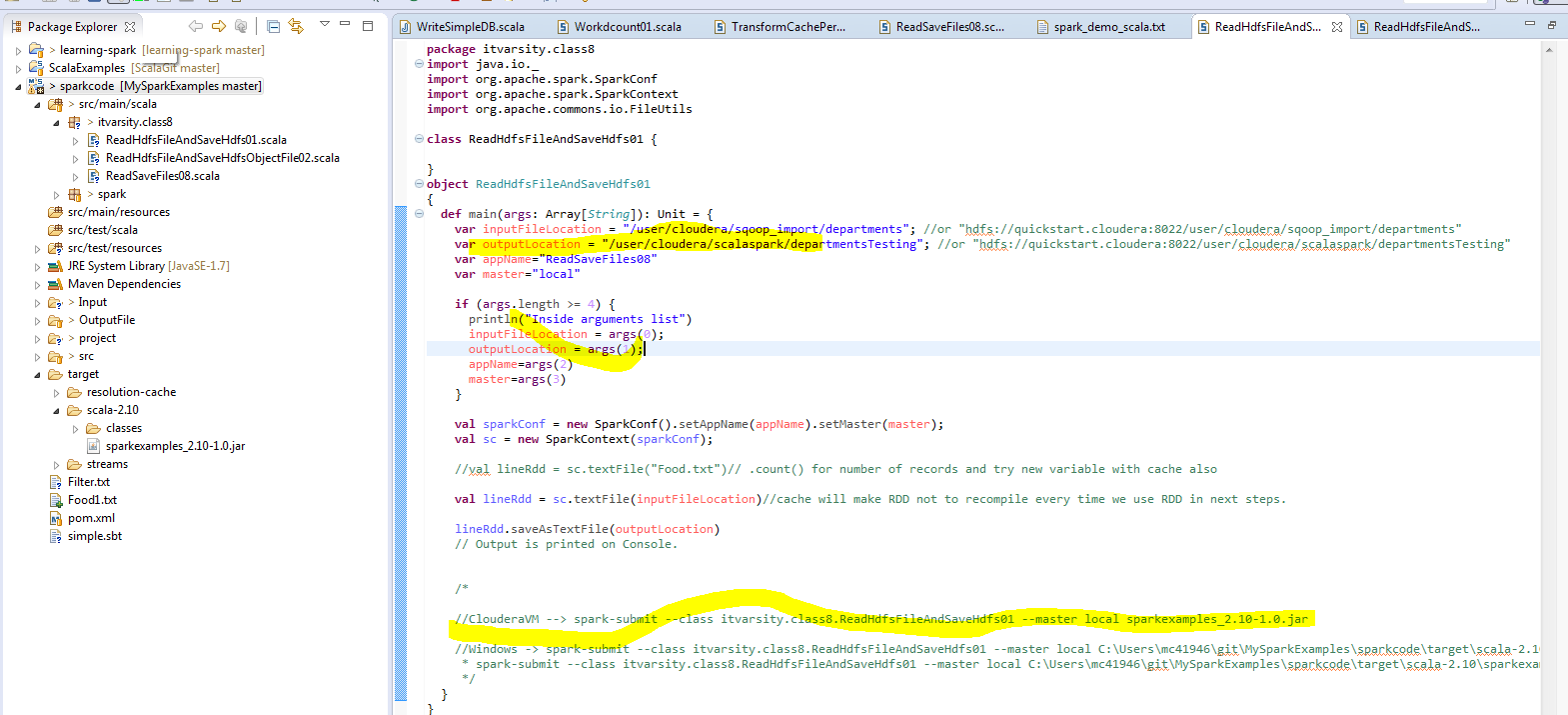


RDD is cacheable, resilient, scalable, lazy evaluated and type inferenced (This is Scala)

lineRdd.saveAsObjectFile(outputLocation)

lineRdd.saveAsTextFile(outputLocation)

see the output location can be hdfs as /user or hdfs:// .. as user



**Submitting application:**

//ClouderaVM --> spark-submit --class itvarsity.class8.ReadHdfsFileAndSaveHdfs01 --master local sparkexamples\_2.10-1.0.jar

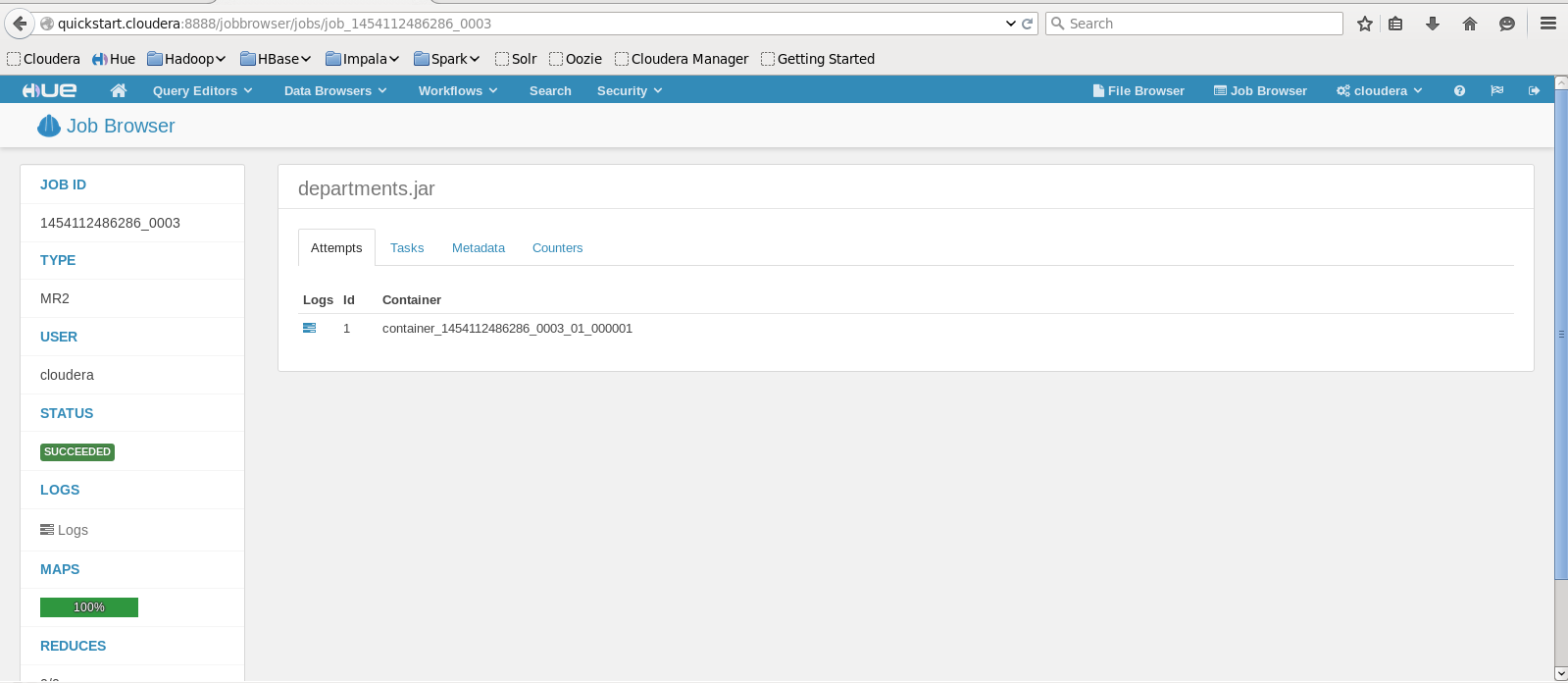
//Windows -> spark-submit --class itvarsity.class8.ReadHdfsFileAndSaveHdfs01 --master local C:\Users\mc41946\git\MySparkExamples\sparkcode\target\scala-2.10\sparkexamples\_2.10-1.0.jar

\* spark-submit --class itvarsity.class8.ReadHdfsFileAndSaveHdfs01 --master local C:\Users\mc41946\git\MySparkExamples\sparkcode\target\scala-2.10\sparkexamples\_2.10-1.0.jar

\*/

**Verifying it in the logs:**

Not sure why the logs not visible as the spark web ui is killed after running job. You can access it by Job brower and then go to Spark UI if it is still running.



Spark Web UI if it is still running.