# Goal 1: Create Fat-JAR on Laptop and run JavaSparkPi on Cluster

## Fat-JAR creation:

gradle fatJarWithTests

task fatJarWithTests(type: Jar) {

baseName = project.name + '-all-with-tests'

from { configurations.testCompile.collect { it.isDirectory() ? it : zipTree(it) } }

from sourceSets.test.output

with jar

}

## Command from example working

dcos spark run --submit-args='-Dspark.mesos.coarse=true --driver-cores 1 --driver-memory 1024M --class org.apache.spark.examples.SparkPi hdfs://hdfs/spark-examples/spark-examples\_2.10-1.4.0-SNAPSHOT.jar 30'

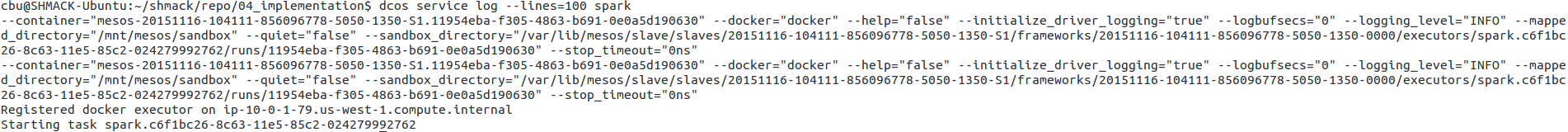


## Command from Fat-JAR not working

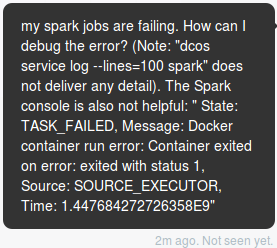
dcos spark run --submit-args='-Dspark.mesos.coarse=true --driver-cores 1 --driver-memory 1024M --class com.zuehlke.shmack.sparkjobs.tutorials.JavaSparkPi hdfs://hdfs/libs/sparkjobs-all-with-tests-1.0-SNAPSHOT.jar 30'



### Where is the “Why”? Where is the Log?

This is all ;-(   


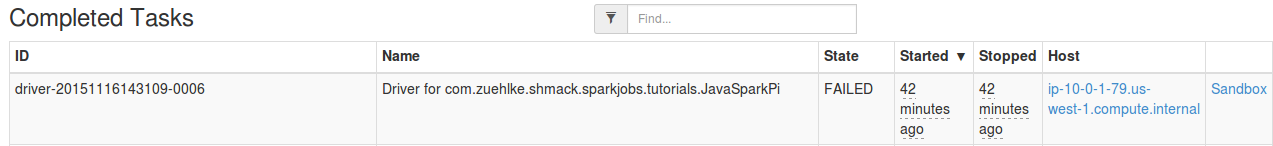
Request by chat:



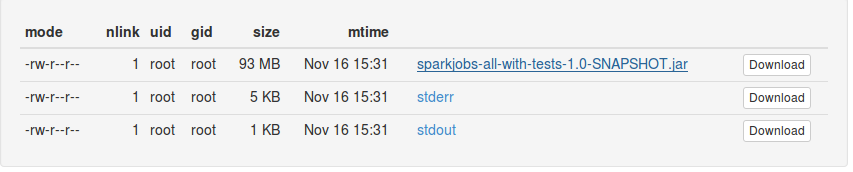
Solution:



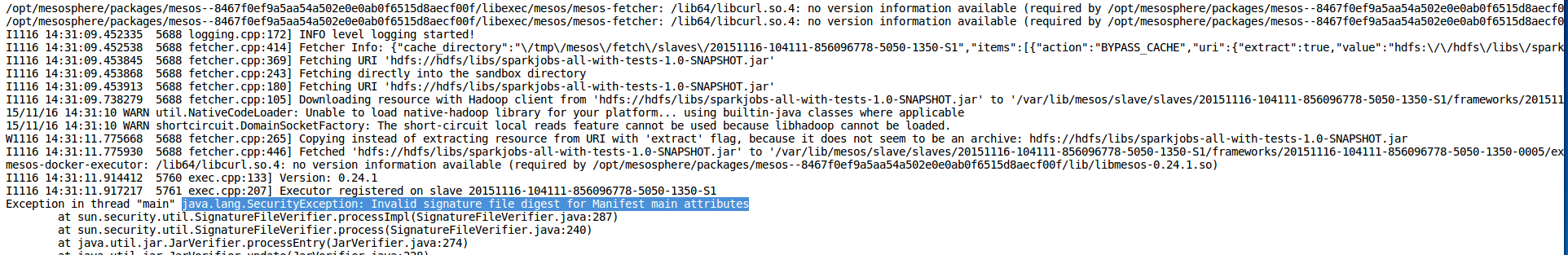
🡪



* Click on “sandbox”



Stderr shows the problem:



java.lang.SecurityException: Invalid signature file digest for Manifest main attributes

### Why is the manifest defective?

### Numerous (1022) duplicate files (but not the manifest)

unzip -l ../sparkjobs-all-with-tests-1.0-SNAPSHOT.jar | awk '{print $4;}' | sort | wc

58484 58481 3616989

unzip -l ../sparkjobs-all-with-tests-1.0-SNAPSHOT.jar | awk '{print $4;}' | sort -u | wc

57462 57461 3564247

* 58484 - 57462 = 1022

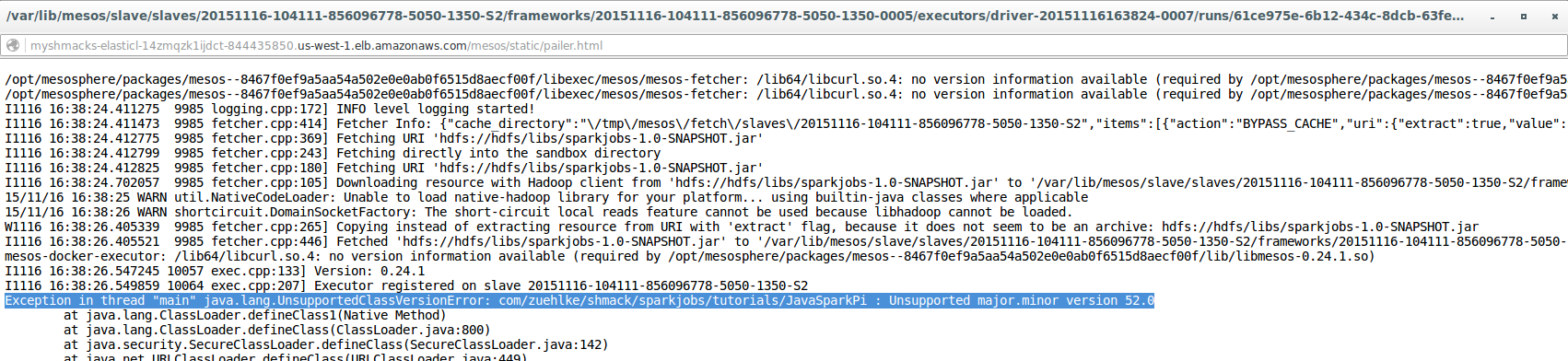
Which files:

unzip -l ../sparkjobs-all-with-tests-1.0-SNAPSHOT.jar | awk '{print $4;}' | sort | awk '{if (x[$1]) { x\_count[$1]++; print $0; if (x\_count[$1] == 1) { print x[$1] } } x[$1] = $0}'

* See <https://github.com/Zuehlke/SHMACK/blob/master/03_analysis_design/Issues/Issue-7%20Spark%20Word%20Count/duplicate-classes-in-fat-jar.txt>

## Try “Slim-JAR”

dcos spark run --submit-args='-Dspark.mesos.coarse=true --driver-cores 1 --driver-memory 1024M --class com.zuehlke.shmack.sparkjobs.tutorials.JavaSparkPi hdfs://hdfs/libs/sparkjobs-1.0-SNAPSHOT.jar 30'



* Need to downgrade to Java 7

# Goal 2: Run WordCount on Cluster