



Apache Spark

Running in Spark Cluster

Spark Cluster

Hadoop Cluster Setup References

<https://www.cloudera.com/tutorials/sandbox-deployment-and-install-guide/3.html>

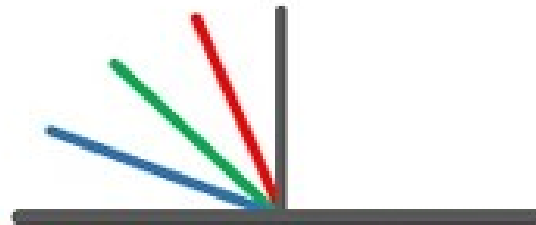
<https://www.cloudera.com/tutorials/learning-the-ropes-of-the-hdp-sandbox.html>

Copy from here (No Line Breaks)

```
spark-submit --verbose --class guru.learningjournal.spark.examples.HelloSpark --  
files log4j.properties,spark.conf --conf 'spark.driver.extraJavaOptions=-  
Dlog4j.configuration=log4j.properties -Dlogfile.name=hello-spark-driver' --  
conf 'spark.executor.extraJavaOptions=-Dlog4j.configuration=log4j.properties -  
Dlogfile.name=hello-spark-executor' --master yarn --deploy-mode cluster hellospark_2.11-  
0.1.jar /user/root/data/sample.csv
```

Formatted

```
spark-submit
--verbose
--class guru.learningjournal.spark.examples.HelloSpark
--files log4j.properties,spark.conf
--conf 'spark.driver.extraJavaOptions=-Dlog4j.configuration=log4j.properties -Dlogfile.name=hello-spark-driver'
--conf 'spark.executor.extraJavaOptions=-Dlog4j.configuration=log4j.properties -Dlogfile.name=hello-spark-executor'
--master yarn
--deploy-mode cluster
hellospark_2.11-0.1.jar /user/root/data/sample.csv
```



LEARNING JOURNAL

www.learningjournal.guru