**Spark**

**Use the same environment as Hadoop**  
  
**Install the spark&python package**

wget http://download.nextag.com/apache/spark/spark-2.2.0/spark-2.2.0-bin-hadoop2.7.tgz

tar xvzf spark-2.2.0-bin-hadoop2.7.tgz

mv spark-2.2.0-bin-hadoop2.7 spark

rm spark-2.2.0-bin-hadoop2.7.tgz

**Set environment variables**

1.~/.bashrc

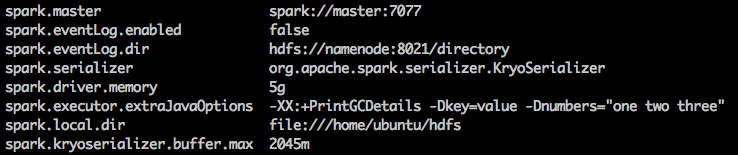
export SPARK\_HOME=/home/ubuntu/sparksample

export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

PATH=$PATH:$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin::$SPARK\_HOME/bin

source ~/.bashrc

2.spark/conf/spark-defaults.conf

****

**Check Spark version**

spark-submit --version



**Compile and run file**

spark-submit --executor-cores 5 --num-executors 17 --executor-memory 8G ~/spark\_sort.py