Step 1:

Download both scala and spark .tar files and unzip them to any location.

my versions:

🡪 spark-1.5.1-bin-hadoop2.6.tgz

🡪 scala-2.10.4.tgz

Step 2:

open bashrc “ gedit ~/.bashrc “ and add spark and scala classpaths.

🡪

export SPARK\_HOME=/home/msruser1/hadoopcluster/spark-1.5.1-bin-hadoop2.6  
export PATH=$SPARK\_HOME/bin:$PATH

export SCALA\_HOME= /home/msruser1/hadoopcluster/scala-2.10.4

export PATH=$SCALA\_HOME/bin:$PATH

🡪 source .bashrc

Upto here spark is ready. if you want advanced configurations go to below steps. But for single node there is no need to go to next steps.

Step 3:

create empty directory ( like spark\_data in any location ) and edit spark-env.sh .

rename $SPARK\_HOME/conf/ spark-env.sh.template to spark-env.sh and add these properties.

🡪

export SCALA\_HOME/home/msruser1/hadoopcluster/scala-2.10.4  
export SPARK\_WORKER\_MEMORY=1g  
export SPARK\_WORKER\_INSTANCES=2  
export SPARK\_WORKER\_DIR=/home/msruser1/hadoopcluster/spark\_data   
export SPARK\_MASTER\_IP=msrcluster1

**Note :**

export SPARK\_WORKER\_INSTANCES = “ number worker instances to run in same machine )

Step 5:

Rename slaves.template file into slaves file. This file is used to mention the slaves address.

in this scenarios I am also install workser node in master node also.

🡪 vi slaves

msrcluster1

Step 6:

**Rename “/conf/ spark-defaults.conf.template “ to “ spark-defaults.conf “ and mention master address.**

**🡪** spark.master spark://msrcluster1:7077

Step 7:

Start spark cluster.

$ cd $SPARK\_HOME

$ sbin/start-master.sh

$ sbin/start-slaves.sh ( not start-slave.sh)

**jps:**

master

worker

worker