|  |
| --- |
| yum install wget |

下载scalc-2.11.8

|  |
| --- |
| wget https://downloads.lightbend.com/scala/2.11.8/scala-2.11.8.tgz |

解压安装包

Tar -zxvf path -C -topath

配置环境变量

|  |
| --- |
| #scala  export SCALA\_HOME=/module/scala-2.11.8  export PATH=$PATH:$SCALA\_HOME/bin |

然后刷新source .bash\_profile

# scala -version 查看是否安装成功

Scala code runner version 2.11.12 -- Copyright 2002-2017, LAMP/EPFL

下载spark2.1.0

wget http://d3kbcqa49mib13.cloudfront.net/spark-2.1.0-bin-hadoop2.6.tgz

解压安装

文件名太长改为mv /spark-2.1.0-bin-hadoop2.6.tgz spark-2.1.0

配置环境变量

#spark

export SPARK\_HOME=/module/spark-2.1.0

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

输入：source .bash\_profile

使环境变量快速生效

将文件spark中conf文件

复制一个文件 cp log4j.properties.template  log4j.properties

然后进入log4j.properties 并将log4j.logger.org.apache.spark.repl.Main=INFO  (把原来的WARN改为INFO,保存即可)

yan

内存资源给的过小，yarn直接kill掉进程，则报rpc连接失败、ClosedChannelException等错误

需要在hadoop的yarn文件中配置

|  |
| --- |
| <property>  <name>yarn.nodemanager.pmem-check-enabled</name>  <value>false</value>  </property>  <!--是否启动一个线程检查每个任务正使用的虚拟内存量，如果任务超出分配值，则直接将其  杀掉，默认是true -->  <property>  <name>yarn.nodemanager.vmem-check-enabled</name>  <value>false</value>  </property> |

配置spark conf中这个

spark-defaults.conf

|  |
| --- |
| spark.yarn.historyServer.address=hadoop15:18080  spark.history.ui.port=18080 |

配置spark-env.sh

|  |
| --- |
| export JAVA\_HOME=/module/jdk1.8.0\_181  export SCALA\_HOME=/module/scala-2.11.8  export HADOOP\_HOME=/module/hadoop-2.7.2  export HADOOP\_CONF\_DIR=/module/hadoop-2.7.2/etc/hadoop  export SPARK\_MASTER\_IP=192.168.242.10  export SPARK\_MASTER\_PORT=7077  export YARN\_CONF\_DIR=/module/hadoop-2.7.2/etc/hadoop  export SPARK\_MASTER\_HOST=192.168.242.10  export SPARK\_LOCAL\_IP=192.168.242.10  export SPARK\_EXECUTOR\_MEMORY=1024m |
| 启动hadoop和spark进程  spark-shell启动进程 |

Mac 修改spark目录用户组

sudo chown -R liuqing:staff spark

