Spark on Yarn

**1.解压spark安装包**

$>tar -xvf spark-2.1.0-bin-hadoop2.7.tgz -C /opt/client/Spark/

**2.创建软连接**

$>ln -s spark-2.1.0-bin-hadoop2.7.tgz spark

**3.修改conf目录下的spark-env.sh文件，添加环境变量**

export JAVA\_HOME=/opt/client/JDK/jdk

export SCALA\_HOME=/opt/client/Scala/scala

export HADOOP\_HOME=/opt/client/Hadoop/hadoop

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

export SPARK\_MASTER\_IP=192.168.2.145

export SPARK\_MASTER\_PORT=7077

export SPARK\_MASTER\_WEBUI\_PORT=7070

export SPARK\_LOCAL\_DIR=/opt/client/Spark/tmp

export SPARK\_HISTORY\_OPTS="-Dspark.history.ui.port=7777 -Dspark.history.fs.logDirectory=hdfs://hacluster/sparkJobHistory"

其中-Dspark.history.ui.port为指定启动historyserver时的端口

-Dspark.history.fs.logDirectory为指定historyserver读取日志时路径，与spark-default.xml 文件中的配置对应

**4.修改slaves文件（spark on yarn模式下可以不配，不启动master-worker就行）**

添加worker节点的ip

**5.修改spark-default.xml文件**

spark.eventLog.enabled=true

spark.eventLog.dir=hdfs://hacluster/sparkJobHistory

开启日志功能，并写入指定的hdfs路径下

6.通过spark-submit测试是否安装成功

$>spark-submit --master yarn-client --class org.apache.spark.examples.SparkPi /opt/client/Spark/spark/examples/jars/spark-examples\_2.11-2.1.0.jar

