**Hadoop安装教程**

# 解压JDK

$ tar -zxf jdk-7u67-linux-x64.tar.gz -C /opt/modules/

## 配置环境变量

#JAVA\_HOME

export JAVA\_HOME=/opt/modules/jdk1.7.0\_67

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

# 解压Hadoop

$ tar -zxf hadoop-2.5.0.tar.gz -C /opt/modules/

建议：删除rm -rf doc/ 里面是英文文档，没有作用

# 配置过程

## 配置.sh文件

分别为/home/hadoop/hadoop-2.5.0/etc/hadoop目录下hadoop-env.sh、mapred-env.sh、yarn-env.sh三个文件指定Java的安装路径

export JAVA\_HOME=/opt/modules/jdk1.7.0\_67

# 配置xml文件

配置/home/hadoop/hadoop-2.5.0/etc/hadoop目录下xml文件。

## 配置core-site.xml

指定namenode所在机器位置及交互端口号，另外也是指定使用的是HDFS文件系统

<property>

<name>fs.defaultFS</name>

<value>hdfs://MyDream:8020</value>

</property>

更改Hadoop默认的临时目录位置，用于存放块文件和元数据文件

<property>

<name>hadoop.tmp.dir</name>

<value>/home/hadoop/hadoop-2.5.0/tmp</value>

</property>

## 配置hdfs-site.xml

指定副本数个数

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

## 配置slaves

指定从节点位置，包括datanode和nodemanager，将主机名配置进去即可。

## 格式化namenode

对于元数据进行初始化，否则无法读取元数据

bin/hdfs namenode –format

## 启动namenode、datanode

sbin/hadoop-daemon.sh strat namenode

sbin/hadoop-daemon.sh strat datanode

-》更改mapred-site.xml文件名称，指定MR运行在yarn上

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

-》yarn-site.xml

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>hadoop-senior01.ibeifeng.com</value>

</property>

yarn外部管理界面端口号：8088

HDFS外部管理界面端口号：50070

secondarynamenode外部管理界面端口号：50090

$ bin/yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.0.jar wordcount /user/beifeng/wordcount/input /user/beifeng/wordcount/output2

出错一定要学会看错误信息，看日志，查看日志必须查看.log后缀的