# Hadoop集群搭建文档

## 伪分布环境集群搭建

准备机器：hadoop100

Linux基础环境准备：java、ip、hostname、hosts、iptables、chkconfig、ssh、免密码登录

注意：Jdk建议使用1.8

hadoop安装包下载地址：

官网下载地址：https://archive.apache.org/dist/hadoop/common/hadoop-2.7.5/ 【版本最全，国内下载速度慢】

镜像站下载地址：https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop 【只有最近的一些版本，国内下载速度快】

具体操作步骤：

注意：在word中拷贝配置文件信息的时候注意空格之类的问题，建议先把配置内容拷贝到notepad++之类的文本中，再从这些文本中把配置信息拷贝到linux系统的配置文件中。

|  |
| --- |
| # 建议把hadoop安装包放在/data/soft目录下  cd /data/soft  tar -zxvf hadoop-2.7.5.tar.gz  cd hadoop-2.7.5/etc/hadoop  # 开始修改配置文件  vi hadoop-env.sh  export JAVA\_HOME=/data/soft/jdk1.8  export HADOOP\_LOG\_DIR=/data/hadoop\_repo/logs/hadoop  vi yarn-env.sh export JAVA\_HOME=/data/soft/jdk1.8  export YARN\_LOG\_DIR=/data/hadoop\_repo/logs/yarn  vi core-site.xml  <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://hadoop100:9000</value>  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>/data/hadoop\_repo</value>  </property>  </configuration>  vi hdfs-site.xml  <configuration>  <property>  <name>dfs.replication</name>  <value>1</value>  </property>  </configuration>  vi yarn-site.xml  <configuration>  <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  </configuration>  mv mapred-site.xml.template mapred-site.xml  vi mapred-site.xml  <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  </configuration>  vi slaves  localhost  # 至此，配置文件修改完毕  # 下面开始格式化hdfs文件系统  cd /data/soft/hadoop-2.7.5  bin/hdfs namenode -format    看到里面的successfully就说明文件系统格式化成功了。  启动hadoop  cd /data/soft/hadoop-2.7.5  sbin/start-all.sh  验证，执行jps，能看到下面进程就说明启动成功    还可以通过hdfs和yarn的web界面验证  http://hadoop100:50070  http://hadoop100:8088  停止hadoop  cd /data/soft/hadoop-2.7.5  sbin/stop-all.sh |

## 分布式环境集群搭建

准备机器：hadoop100、hadoop101、hadoop102

Linux基础环境准备：java、ip、hostname、hosts、iptables、chkconfig、ssh、免密码登录

注意：Jdk建议使用1.8

注意：针对免密码登录功能需要配置主节点可以免密登录到从节点【还需要保证主节点能够免密码登录自己】

hadoop100是主节点

ssh-copy-id -i hadoop101

ssh-copy-id -i hadoop102

这样，就可以实现主节点到从节点的 SSH 免密码登录了

注意：需要保证集群的各个节点时间同步 ntpdate -u ntp.sjtu.edu.cn

集群规划

主节点：hadoop100 从节点：hadoop101、hadoop102

hadoop100：namenode、resourcemanager、secondarynamenode

hadoop101：datanode、nodemanager

hadoop102：datanode、nodemanager

具体操作步骤：

|  |
| --- |
| # 注意，先操作hadoop100这个节点  # 建议把hadoop安装包放在/data/soft目录下  cd /data/soft  tar -zxvf hadoop-2.7.5.tar.gz  cd hadoop-2.7.5/etc/hadoop  # 开始修改配置文件  vi hadoop-env.sh  export JAVA\_HOME=/data/soft/jdk1.8  export HADOOP\_LOG\_DIR=/data/hadoop\_repo/logs/hadoop  vi yarn-env.sh export JAVA\_HOME=/data/soft/jdk1.8  export YARN\_LOG\_DIR=/data/hadoop\_repo/logs/yarn  vi core-site.xml  <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://hadoop100:9000</value>  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>/data/hadoop\_repo</value>  </property>  </configuration>  vi hdfs-site.xml  <configuration>  <property>  <name>dfs.replication</name>  <value>2</value>  </property>  <property>  <name>dfs.namenode.secondary.http-address</name>  <value>hadoop100:50090</value>  </property>  </configuration>  vi yarn-site.xml  <configuration>  <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.resourcemanager.hostname</name>  <value>hadoop100</value>  </property>  </configuration>  mv mapred-site.xml.template mapred-site.xml  vi mapred-site.xml  <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  </configuration>  vi slaves  hadoop101  hadoop102  # 至此，配置文件修改完毕  # 把hadoop100这个节点上配置好的hadoop的安装包拷贝到其他两个节点中  # 注意：需要保证这两个节点中的/data目录是存在的  # 注意：保证每个节点的/etc/hosts文件中都有三个节点的ip和主机名映射  scp -rq hadoop-2.7.5 hadoop101:/data/soft  scp -rq hadoop-2.7.5 hadoop102:/data/soft  # 下面开始在hadoop100节点上格式化hdfs文件系统  cd /data/soft/hadoop-2.7.5  bin/hdfs namenode -format    看到里面的successfully就说明文件系统格式化成功了。  启动hadoop，有两种方案  第一种：在hadoop100节点上执行下面命令  cd /data/soft/hadoop-2.7.5  sbin/start-all.sh  第二种：在hadoop100节点上执行下面命令  cd /data/soft/hadoop-2.7.5  sbin/start-dfs.sh  sbin/start-yarn.sh  验证，执行jps，看到下面信息说明启动成功  在hadoop100上看到如下进程信息：    在hadoop101上看到如下进程信息：    在hadoop102上看到如下进程信息： |