centos7 安装 hadoop 2.7.3集群 步骤：

**一、            配置hosts文件，建立hadoop运行帐号**

**二、            配置ssh免密码连入**

**三、            安装配置jdk**

**四、            下载并解压hadoop安装包**

**五、            配置namenode,修改site文件**

**六、            配置hadoop-env.sh文件**

**七、            配置masters和slaves文件**

**八、          克隆  hadoop集群**

**九、            格式化namenode**

**十、            启动hadoop**

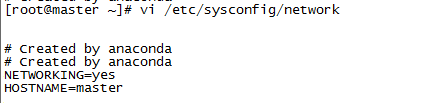
**十一、      用jps检验各后台进程是否成功启动**

**十二、      通过网站查看集群情况**

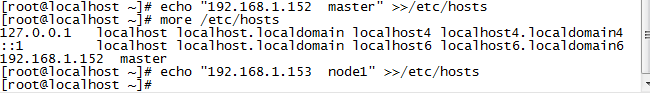
**一、            配置hosts文件，建立hadoop运行帐号**

修改主机名称 为master

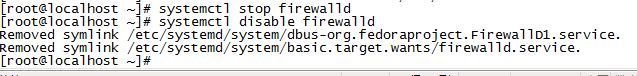




修改ip映射关系， 添加master 和 node 节点的ip映射关系



关闭防护墙



创建用户和用户组

创建hadoop群组和hadoop用户

      groupadd hadoop

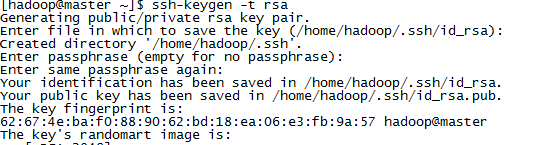
      useradd hadoop -g hadoop

passwd hadoop

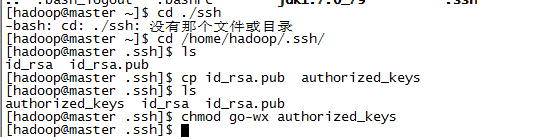
      更改hadoop 用户的密码

**二、            配置ssh免密码连入**

使用 hadoop账号登录 ，生成无密码登录秘钥对 （冒号处 回车）



拷贝和权限修改



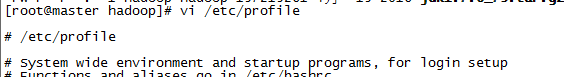
**三、            安装配置jdk**

切换到root用户

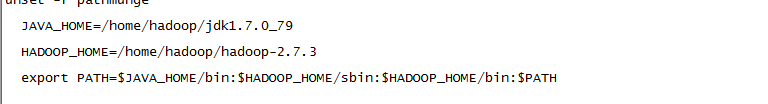
解压jdk



配置jdk



添加如下配置



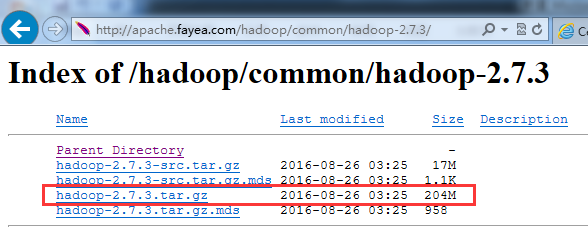
文件生效



**四、            下载并解压hadoop安装包**

下载hadoop

http://apache.fayea.com/hadoop/common/hadoop-2.7.3/



上传并解压文件 到 /home/hadoop/

  tar -zxvf hadoop-2.7.3.tar.gz

  chown -R hadoop:hadoop /home/hadoop/hadoop-2.7.3

创建相关目录：

  mkdir -p /home/hadoop/data/hadoop/ #创建 hadoop 数据目录结构

  mkdir -p /home/hadoop/data/hadoop/tmp/   #创建 tmp

  mkdir -p /home/hadoop/data/hadoop/hdfs/   #创建hdfs

  mkdir -p /home/hadoop/data/hadoop/hdfs/data  #创建datanode目录

  mkdir -p /home/hadoop/data/hadoop/hdfs/name  #创建namenode目录

  mkdir -p /home/hadoop/data/hadoop/hdfs/namesecondary

chown -R hadoop:hadoop  /home/hadoop/data/

**五、            配置namenode,修改site文件**

**六、            配置hadoop-env.sh文件**

**七、            配置masters和slaves文件**

使用hadoop账号登录修改

 vi core-site.xml  #全局配置

|  |
| --- |
| <configuration>  <property>  <name>fs.defaultFS</name>  <value>hdfs://master:9000</value>  <!--hadoop namenode 服务器地址和端口，以域名形式-->  </property>  <property>  <name>dfs.namenode.checkpoint.period</name>  <value>1800</value>  <!-- editlog每隔30分钟触发一次合并，默认为60分钟 -->  </property>  <property>    <name>fs.checkpoint.size</name>    <value>67108864</value>  </property>  <property>  <name>fs.trash.interval</name>  <value>1440</value>  <!-- Hadoop文件回收站,自动回收时间,单位分钟,这里设置是1天，默认为0。-->  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>/home/hadoop/data/hadoop/tmp</value>  <!-- Hadoop的默认临时路径，这个最好配置，如果在新增节点或者其他情况下莫名其妙的DataNode启动不了，就删除此文件中的tmp目录即可。不过如果删除了NameNode机器的此目录，那么就需要重新执行NameNode格式化的命令。/data/hadoop/tmp这里给的路径不需要创建会自动生成.-->  </property>  <property>  <name>io.file.buffer.size</name>  <value>131702</value>  <!-- 流文件的缓冲区-->  </property>  </configuration> |

vi hdfs-site.xml  #hdfs中NameNode,DataNode局部配置

|  |
| --- |
| <configuration>  <property>  <name>dfs.namenode.name.dir</name>  <value>/home/hadoop/data/hadoop/hdfs/name</value>  <!--HDFS namenode数据镜象目录-->  <description>  </description>  </property>  <property>  <name>dfs.datanode.data.dir</name>  <value>/home/hadoop/data/hadoop/hdfs/data</value>  <!-- HDFS datanode数据镜象存储路径,可以配置多个不同的分区和磁盘中,使用,号分隔 -->  <description> </description>  </property>  <property>  <name>dfs.namenode.http-address</name>  <value>master:50070</value>  <!---HDFS Web查看主机和端口-->  </property>  <property>  <name>dfs.namenode.secondary.http-address</name>  <value>node1:50090</value>  <!--辅控HDFS web查看主机和端口-->  </property>  <property>  <name>dfs.webhdfs.enabled</name>  <value>true</value>  </property>  <property>  <name>dfs.replication</name>  <value>3</value>  <!--HDFS数据保存份数，通常是3-->  </property>  <property>  <name>dfs.datanode.du.reserved</name>  <value>1073741824</value>  <!-- datanode 写磁盘会预留 1G空间 给其他程序使用,而非写满,单位 bytes-->  </property>  <property>  <name>dfs.block.size</name>  <value>134217728</value>  <!--HDFS数据块大小，当前设置为128M/Blocka-->  </property>  <property>  <name>dfs.permissions.enabled</name>  <value>false</value>  <!-- HDFS 关闭文件权限 -->  </property>  </configuration>  执行复制配置文件：  cp etc/hadoop/mapred-site.xml.template etc/hadoop/mapred-site.xml |

vi etc/hadoop/mapred-site.xml  #配置MapReduce，使用yarn框架、jobhistory使用地址以及web地址

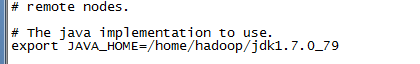
|  |
| --- |
| <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>mapreduce.jobtracker.http.address</name>  <value>master:50030</value>  </property>  <property>  <name>mapred.job.tracker</name>  <value>http://master:9001</value>  </property>  <property>  <name>mapreduce.jobhistory.address</name>  <value>master:10020</value>  </property>  <property>  <name>mapreduce.jobhistory.webapp.address</name>  <value>master:19888</value>  </property>  </configuration> |

vi etc/hadoop/yarn-site.xml  配置yarn-site.xml文件

|  |
| --- |
| <configuration>  <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property>  <property>  <name>yarn.resourcemanager.address</name>  <value>master:8032</value>  </property>  <property>  <name>yarn.resourcemanager.scheduler.address</name>  <value>master:8030</value>  </property>  <property>  <name>yarn.resourcemanager.resource-tracker.address</name>  <value>master:8031</value>  </property>  <property>  <name>yarn.resourcemanager.admin.address</name>  <value>master:8033</value>  </property>  <property>  <name>yarn.resourcemanager.webapp.address</name>  <value>master:8088</value>  </property>  </configuration> |

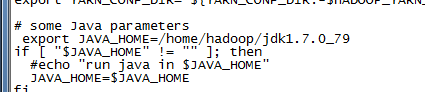
修改文件 vi hadoop-env.sh

将其中的${JAVA\_HOME}用 /home/hadoop/jdk1.7.0\_79 代替



修改文件 vi yarn-env.sh

将其中的${JAVA\_HOME}用 /home/hadoop/jdk1.7.0\_79 代替



**八、          克隆  hadoop集群**

通过虚拟机克隆多个node节点机器（如果物理机部署，则按照上面的步骤 安装其他机器，保证文件，账号，目录的相同性，避免出现问题）

克隆的虚机 按照第一步修改主机信息名称 ip地址

**九、            格式化namenode**

在master上进行配置

    vi /home/hadoop/hadoop2.7.3/etc/hadoop/slaves

    删除localhost

    添加

    node1

    vi /home/hadoop/hadoop2.7.3/etc/hadoop/masters

    删除localhost

    添加

    node1   #目的是为了在node1节点上存放secondnamenode

执行格式化：

/home/hadoop/hadoop2.7.3/bin/hdfs namenode -format

**十、            启动hadoop**

master上 /home/hadoop/hadoop2.7.3/sbin 下执行

开启所有节点：start-all.sh

关闭所有节点：stop-all.sh

**十一、      用jps检验各后台进程是否成功启动**

各节点执行 jps

|  |
| --- |
| master：  98956 JobHistoryServer  98820 NodeManager  118806 Jps  118176 NameNode  118540 ResourceManager |
| node1：  106408 SecondaryNameNode  106602 Jps  106301 DataNode  106496 NodeManager | |

**十二、      通过网站查看集群情况**

