**安装Hadoop**

解压

tar -zxvf hadoop-2.6.5.tar.gz /usr/local/apps/hadoop

**修改环境变量**

# 修改配置文件

vi /etc/profile

# 在最后下添加

export HADOOP\_HOME=/usr/local/apps/hadoop

export PATH=$PATH:/usr/local/apps/hadoop/bin:/usr/local/apps/hadoop/sbin

# 刷新配置文件

source /etc/profile

**修改配置文件**

这些配置文件全部位于 /usr/local/apps/hadoop /etc/hadoop 文件夹下

**hadoop-env.sh**

vi /usr/local/hadoop/etc/hadoop/hadoop-env.sh  
修改export JAVA\_HOME 语句为 export JAVA\_HOME=/usr/local/apps/

jdk

**core-site.xml**

<configuration>

<!-- 指定HDFS（namenode）的通信地址 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

</property>

<!-- 指定hadoop运行时产生文件的存储路径 -->

<property>

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

<value>/usr/local/apps/tmp</value>

</property>

</configuration>

**hdfs-site.xml**

<configuration>

<!-- 设置namenode的http通讯地址 -->

<property>

<name>dfs.namenode.http-address</name>

<value>master:50070</value>

</property>

<!-- 设置secondarynamenode的http通讯地址 -->

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>slave1:50090</value>

</property>

<!-- 设置namenode存放的路径 -->

<property>

<name>dfs.namenode.name.dir</name>

<value>/usr/local/apps/Hadoop/name</value>

</property>

<!-- 设置hdfs副本数量 -->

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

<!-- 设置datanode存放的路径 -->

<property>

<name>dfs.datanode.data.dir</name>

<value>/usr/local/apps/Hadoop/data</value>

</property>

</configuration>

**mapred-site.xml**

**必须先**

**mv mapred-site.xml.template mapred-site.xml**

<configuration>

<!-- 通知框架MR使用YARN -->

<property>

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

<value>yarn</value>

</property>

</configuration>

**yarn-site.xml**

<configuration>

<!-- 设置 resourcemanager 在哪个节点-->

<property>

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

<value>master</value>

</property>

<!-- reducer取数据的方式是mapreduce\_shuffle -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

</configuration>

遇到的问题：datanode启动不了或者启动后自动关闭，namenode 启动不了或者启动后自动关闭，原因是，频繁的format使得文件丢失（pid文件），解决方法，删除所有节点上的data文件夹、name文件夹、tmp文件夹，然后format，重新启动