# 安装配置Hadoop环境

## hadoop环境变量

命令：gedit /etc/profile

加入：

export HADOOP\_HOME=/home/hadoop/app/hadoop-2.6.0-cdh5.7.0

export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

命令：source /etc/profile

## Hadoop目录介绍

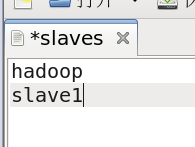
hadoop-2.6.0-cdh5.7.0/etc/hadoop是Hadoop配置目录

## Hadoop核心配置

1.配置hadoop-env.sh文件修改java\_home环境变量

2.配置yarn-env.sh文件修改java\_home环境变量（1.x版本没有）

3.配置slaves文件添加slave节点（删除原来的localhost）



4.配置core-site.xml文件新增hadoop核心配置

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop:9000/</value>

<description>默认的HDFS端口，用于NameNode与DataNode之间到的通讯，IP为NameNode的地址</description>

</property>

<property>

<name>io.file.buffer.size </name>

<value>131072 </value>

<description>文件缓存大小 </description>

</property>

<property>

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

<value>file:/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/tmp </value>

<description>存放hadoop文件系统依赖的基本配置</description>

</property>

</configuration>

5.配置hdfs-site.xml文件新增hdfs配置信息

<configuration>

<property>

<name>dfs.replication</name>

<value>2</value>

<description>设置副本数</description>

</property>

<property>

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

<value>file:/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/hdfs/name</value>

<description>设置存放NameNode的文件路径</description>

</property>

<property>

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

<value>file: /home/hadoop/app/hadoop-2.6.0-cdh5.7.0/hdfs/data</value>

<description>设置存放DataNode的文件路径</description>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

<description>权限还是不要的好</description>

</property>

<property>

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

<value>slave1:9001 </value>

<description>SNN路径</description>

</property>

</configuration>

6.配置mapred-site.xml文件新增mapredure配置

（配置作业运行的方式，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>

<description>设置资源调度</description>

</property>

<property>

<name>mapreduce.jobhistory.address</name>

<value>hadoop:10020</value>

<description>指定历史服务器的地址和端口</description>

</propery>

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>hadoop:19888</value>

<description>历史服务器的web地址</description>

</propery>

<property>

<name>mapreduce.application.classpath</name>

<value>

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/etc/hadoop,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hadoop/common/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hadoop/common/lib/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hadoop/hdfs/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hdfs/lib/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/mapreduce/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/mapreduce/lib/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hadoop/yarn/\*,

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/share/hadoop/yarn/lib/\*

</value>

<description>设置MapReduce资源调度类路径，如果不设置可能会报错</description>

</property>

</configuration>

7.配置yarn-site.xml新增yarn功能

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>mapreduce\_shuffle</value>

<description>NodeManager上运行的附属服务。需配置成mapreduce\_shuffle，才可运行MapReduce程序</description>

</property>

<property>

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

<value>hadoop</value>

<description>resourcemanager</description>

</property>

<property>

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

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

<description>NodeManager上运行的附属服务。需配置成mapreduce\_shuffle，才可运行MapReduce程序</description>

</property>

<property>

<name>yarn.nodemanager.vmem-check-enabled</name>

<value>false</value>

<description>由于我的测试机内存少，所以就关闭虚拟内存检测s</description>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>hadoop:8035</value>

<description>The hostname of the RM.</description>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>hadoop:8030</value>

<description>The hostname of the RM.</description>

</property>

<property>

<name>yarn.resourcemanager.address</name>

<value>hadoop:8032</value>

<description>The hostname of the RM.</description>

</property>

<property>

<name>yarn.resourcemanager.admin.address</name>

<value>hadoop:8033</value>

<description>The hostname of the RM.</description>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>hadoop:8088</value>

<description>The hostname of the RM.</description>

</property>

</configuration>

## SSH连接

1.配置/etc/hosts，映射其它机器IP

192.168.88.131 slave1

192.168.88.129 hadoop

2.将配置好的hadoop主机中的/Hadoop/etc/hadoop文件夹复制到slave1对应的位置

scp -r

/home/hadoop/app/hadoop-2.6.0-cdh5.7.0/etc/hadoop hadoop@slave1:/home/hadoop/etc/

# 验证Hadoop环境

## 验证

### 格式化namenode

1. 进入hadoop文件根路径
2. ./bin/hdfs namenode -format