## 1,集群部署规划

|  |  |  |  |
| --- | --- | --- | --- |
|  | 9.30.217.126 | 9.30.189.182 | 9.30.211.96 |
| HDFS | NameNode  DataNode | DataNode | SecondaryNameNode  DataNode |
| YARN | NodeManager | ResourceManager  NodeManager | NodeManager |

## 2,配置集群

1. 核心配置文件

配置core-site.xml

<!-- 指定HDFS中NameNode的地址 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs:// 9.30.217.126:9000</value>

</property>

<!-- 指定Hadoop运行时产生文件的存储目录 -->

<property>

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

<value>/opt/software/hadoop/data/tmp</value>

</property>

1. HDFS配置文件

配置hadoop-env.sh

JAVA\_HOME=/opt/software/jdk

配置hdfs-site.xml

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

<!-- 指定Hadoop辅助名称节点主机配置 -->

<property>

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

<value>9.30.211.96:50090</value>

</property>

1. YARN配置文件

配置yarn-env.sh

export JAVA\_HOME=/opt/software/jdk

配置yarn-site.xml

<!-- Reducer获取数据的方式 -->

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<!-- 指定YARN的ResourceManager的地址 -->

<property>

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

<value>9.30.189.182</value>

</property>

1. MapReduce配置文件

配置mapred-env.sh

export JAVA\_HOME=/opt/software/jdk

配置mapred-site.xml

<!-- 指定MR运行在Yarn上 -->

<property>

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

<value>yarn</value>

</property>

1. 在集群上分发配置好的Hadoop配置文件

配置slaves

cd /opt/software/hadoop/etc/Hadoop

vi workers(3.0之前叫slaves)

9.30.217.126

9.30.189.182

9.30.211.96

注意：该文件中添加的内容结尾不允许有空格，文件中不允许有空行。

## 3,启动集群

（1）如果集群是第一次启动，需要格式化NameNode（注意格式化之前，一定要先停止上次启动的所有namenode和datanode进程，然后再删除data和log数据）

bin/hdfs namenode -format

（2）启动HDFS

sbin/start-dfs.sh

（3）启动YARN

sbin/start-yarn.sh

注意：NameNode和ResourceManger如果不是同一台机器，不能在NameNode上启动 YARN，应该在ResouceManager所在的机器上启动YARN。

Web端查看SecondaryNameNode

<http://9.30.211.96:50090/status.html>

Web端查看NameNode

<http://9.30.217.126:9870/>

Web端查看ResourceManager

<http://9.30.189.182:8088/cluster>