

02-Hadoop (完全分布式)

笔记本: 模块一配置笔记

创建时间: 2020/10/3 16:11

更新时间: 2020/10/22 16:19

作者: 封

二、Hadoop配置 (完全分布式)

0.软件包

涉及到的软件包, 都存放在/opt/soft/

1.解压hadoop包

在master中, 把hadoop包解压到指定位置,按照要求是否需要重命名

```
tar -zxvf /opt/soft/hadoop-2.7.7.tar.gz -C /opt/soft/
```

```
重命名 mv /opt/soft/hadoop2.7.7 /opt/soft/hadoop
```

2.配置环境变量

在master中, vi /etc/profile

输入:

```
export HADOOP_HOME=/opt/soft/hadoop
```

```
export PATH=$HADOOP_HOME/sbin:$HADOOP_HOME/bin:$PATH
```

```
source /etc/profile 更新环境变量
```

输入 `hadoop version`, 看有无版本信息输出

3.配置hadoop文件

文件名称	配置项	配置值	含义
hadoop-env.sh	JAVA_HOME	jdk的配置路径	JAVA_HOME
core-site.xml	fs.defaultFS	hdfs://master:9000	配置NameNode地址,9000是RPC通信端口
core-site.xml	hadoop.tmp.dir	/opt/soft/hadoop/data/tmp	临时文件存储
hdfs-site.xml	dfs.replication	3	副本数，默认是3
hdfs-site.xml	dfs.namenode.name.dir	/opt/soft/hadoop/data/name	namenode数据存储地址
hdfs-site.xml	dfs.datanode.data.dir	/opt/soft/hadoop/data/data	datanode数据存储地址
mapred-site.xml	mapreduce.framework.name	yarn	配置yarn表示集群模式，配置local为本地模式
yarn-site.xml	yarn.resourcemanager.hostname	master	ResourceManager的主机名
yarn-site.xml	yarn.nodemanager.aux-services	mapreduce_shuffle	NodeManager上运行的附属服务
slaves	DataNode的地址	从节点1，从节点2	DataNode的主机名

4.网络拷贝

在master中，把配置好的hadoop的文件复制到其他服务器中。指令：

```
scp /opt/soft/hadoop root@slave1:/opt/soft/
scp /opt/soft/hadoop root@slave2:/opt/soft/
```

在master中，把配置好的环境变量的文件复制到其他服务器中。指令：

```
scp /etc/profile root@slave1:/etc/
scp /etc/profile root@slave2:/etc/
```

在其余服务器执行 source /etc/profile 更新环境变量

4.格式化hadoop

在master运行hdfs namenode -format 如果提示格式化成功，则不需要再格式化。

```
20/04/28 18:12:21 INFO namenode.FSImage: Allocated new BlockPoolId: BP-703524719-189.80.
20/04/28 18:12:21 INFO common.Storage: Storage directory /home/hadoopdir/dfs/name has be
20/04/28 18:12:21 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with
20/04/28 18:12:21 INFO util.ExitUtil: Exiting with status 0
20/04/28 18:12:21 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at masterNode1/189.80.63.8
*****/
```

5.启动hadoop

在master，执行start-all.sh

如果没有配置环境变量则进入hadoop下的sbin目录，输入 start-all.sh，输入yes即可启动

6.jps查看3台服务器进程

master节点有4个进程，使用指令jps

NameNode

SecondaryNameNode

ResourceManager

Jps

其他节点（slave）有3个进程，使用指令jps

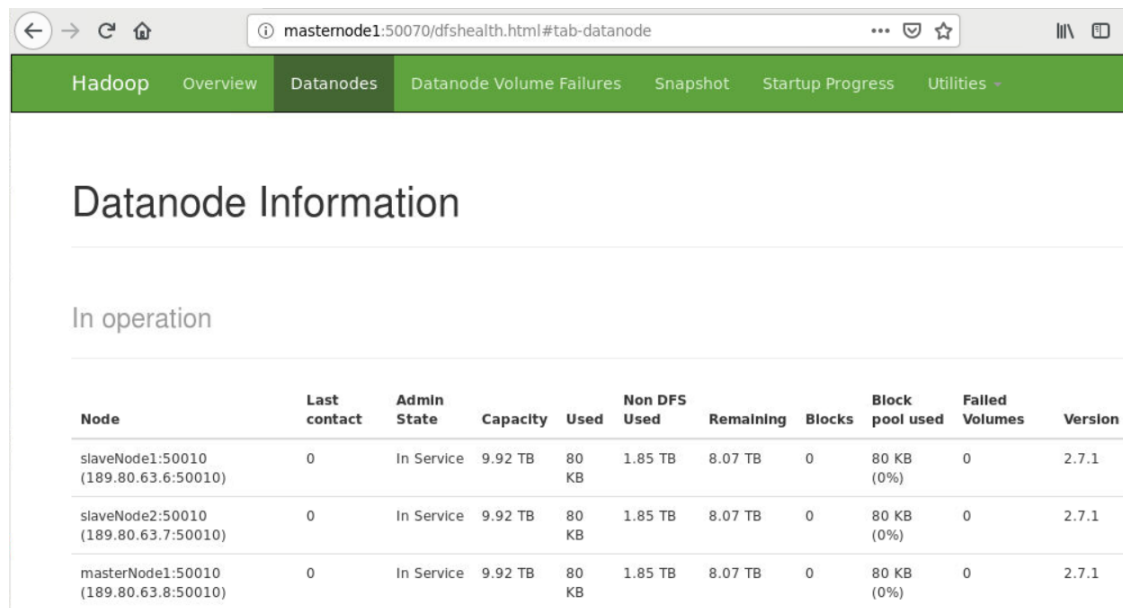
DataNode

NodeManager

Jps

7.用web端查看节点信息

在浏览器输入master的IP地址:50070



The screenshot shows the Hadoop DFS Health page in a web browser. The address bar displays 'masternode1:50070/dfshealth.html#tab-datanode'. The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes (selected), Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Datanode Information' and shows 'In operation' status. Below this is a table with 11 columns: Node, Last contact, Admin State, Capacity, Used, Non DFS Used, Remaining, Blocks, Block pool used, Failed Volumes, and Version. The table lists three nodes: slaveNode1:50010, slaveNode2:50010, and masterNode1:50010, all in 'In Service' state with 9.92 TB capacity and 80 KB used.

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
slaveNode1:50010 (189.80.63.6:50010)	0	In Service	9.92 TB	80 KB	1.85 TB	8.07 TB	0	80 KB (0%)	0	2.7.1
slaveNode2:50010 (189.80.63.7:50010)	0	In Service	9.92 TB	80 KB	1.85 TB	8.07 TB	0	80 KB (0%)	0	2.7.1
masterNode1:50010 (189.80.63.8:50010)	0	In Service	9.92 TB	80 KB	1.85 TB	8.07 TB	0	80 KB (0%)	0	2.7.1

8.结束hadoop

在master输入stop-all.sh停止集群

9.附录（对第3点源码）

9.1core-site.xml

```
<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://master:9000</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/opt/soft/hadoop/data</value>
</property>
</configuration>
```

9.2hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/opt/soft/hadoop/data/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/opt/soft/hadoop/data/data</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
</configuration>
```

9.3mapred-site.xml

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
    <description>指定mapreduce使用yarn框架</description>
  </property>
</configuration>
```

9.4yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master</value>
    <description>指定resourcemanager所在的hostname</description>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

9.5slaves

```
slaves1  
slaves2
```

9.6hadoop-env.sh

```
export JAVA_HOME=/opt/soft/jdk
```