

## 10-hdfs配置-集群（两个人一组）

-----成都尚学堂-mr-zeng-----

-----两个人一组-----

**hdfs集群：多个计算机共同存储数据，利于扩展，存储量更大**

**hdfs各守护进程**

- a) namenode（主-管理者）：访问hdfs的入口节点，存储文件的元数据（名字，大小，分了那些块，块的位置...）
- b) secondarynamenode（备用者）：namenode的备份，重复记录了namenode的数据，防止丢失（高可用-》冷备）
- c) datanode（从-工作者）：存储的文件的块数据。（hadoop1的默认64m，hadoop2默认128m）

假设3个计算机节点（实际工作中更多，几百个节点都可以，多的节点主要用于datanode）

node1：namenode datanode

node2：secondarynamenode datanode

node3：datanode

### a) 各节点启动运行-配置好【网络和域名映射】

通过ping 域名；命令检查是否配置成功

### b) 各节点间配置ssh免密码登录

ps》不配置在启动hadoop的时候需要输入用到的节点密码-比较繁琐

### c) hadoop环境配置-----【在hdfs单节点的基础上更改】-----【配置如下】-----

1) 配置 etc/hadoop/core-site.xml

【namenode的地址】

vi core-site.xml

```
<property>
  <name>fs.defaultFS</name>
  <value>node1的域名:9000</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/node1/hadoop-2.7.2/tmp</value>
</property>
```

2) 配置etc/hadoop/hdfs-site.xml:

【secondarynamenode 的地址】

```
<property>
  <name>dfs.namenode.secondary.http-address</name>
  <value>node2的域名:50090</value>
</property>
```

3) 配置slaves 用以配置(datanode和nodemanager的地址)

【datanode 的地址】

etc/hadoop/slaves

按行分割配置各地址

node1的域名

node2的域名

node3的域名

### d) hadoop启动（在主节点（namenode的节点-node1操作））

1) 格式化hdfs文件系统-（类似于硬盘格式化-这里-会把namenode需要的元数据基础信息写到对应的文件夹（hadoop.tmp.dir））

ps---》只需要进行一次-格式化会删除当前hdfs所有的数据

hdfs namenode -format

2) 启动hadoop的hdfs（文件系统）

home/hadoop-2.7.2/sbin/ 下

./start-dfs.sh

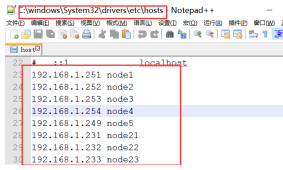
### 3 ) 防火墙关闭(为了方便其他节点的系统访问hdfs)

`service iptables stop` 关闭防火墙

`service iptables status` 查看当前状态

如果没有关闭,则需要对防火墙进行设置---》建议关闭,这样可以在其他客户端访问当前hdfs系统

### 4 ) window 里修改hosts-加入所有节点的ip和域名的映射



### 5 ) 客户端系统 ( window , mac , linux等 ) 网页进入hadoop的hdfs文件系统管理目录

A ) ip访问 : <http://namenode的ip:50070/> 注意浏览器拦截问题 建议使用google

B ) 域名访问 : <http://namenode的域名:50070/>

### 6 ) 测试hdfs文件系统正常 : 在hdfs上传文件一个文件

`hadoop fs -put /root/jdk-7u79-linux-x64.tar.gz hdfs://node1:9000/jdk`

### 7 ) 查看文件是否上传成功-再次重复 ( 4 ) 访问网页进入hadoop的hdfs文件系统管理目录

ps-》如果查看文件保存在多个节点上,并且下载-可以从多个节点下载-集群搭建成功