Hadoop安装（centos7下配置Hadoop3.2.2）

笔记本:	Hadoop		
创建时间:	2021/9/20 11:08	**更新时间:**	2021/9/21 10:52
作者:	邹家豪		
URL:	https://blog.csdn.net/qq_41515513/article/details/101873098		

Hadoop安装（centos7下配置Hadoop3.2.2）

一、安装centos7

二、下载和解压jak hadoop（克隆之前进行）

1.建立安装目录

```
   cd /
   mkdir software
```

2.下载

```
   wget https://dlcdn.apache.org/hadoop/common/hadoop-3.2.2/hadoop-3.2.2.tar.gz   (hadoop-3.2.2)
   wget https://repo.huaweicloud.com/java/jdk/8u202-b08/jdk-8u202-linux-x64.tar.gz   (jdk.1.8.0_202)
```

3.解压

```
   tar -zxvf hadoop-3.2.2.tar.gz -C /software/
   tar -zxvf jdk-8u202-linux-x64.tar.gz -C /software/
```

三、配置环境变量(1和2可以；3和4可以；1234都可以设置)

1.vi /etc/profiel

```
   export JAVA_HOME=/software/jdk1.8.0_202
   export PATH=$JAVA_HOME/bin:$PATH
   export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
   export JAVA_HOME PATH CLASSPATH

   export HADOOP_HOME=/software/hadoop-3.2.2
   export PATH=$PATH:$HADOOP_HOME/bin
   export PATH=$PATH:$HADOOP_HOME/sbin
```

2.source /etc/profile

3.vi ~/.bash_profile

```
   export JAVA_HOME=/software/jdk1.8.0_202
   export JAVA_BIN=$JAVA_HOME/bin
   export JAVA_LIB=$JAVA_HOME/lib
   export CLASSPATH=.:$JAVA_LIB/tools.jar:$JAVA_LIB/dt.jar
```

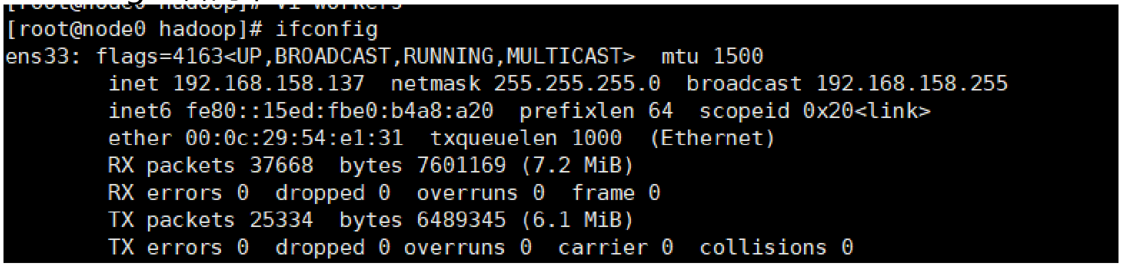
```
    export HADOOP_HOME=/software/hadoop-3.2.2

 PATH=$PATH:$JAVA_BIN:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

    export PATH
4.source ~/.bash_profile
```

四、克隆主机（主机node）

克隆机node1,node2,node3

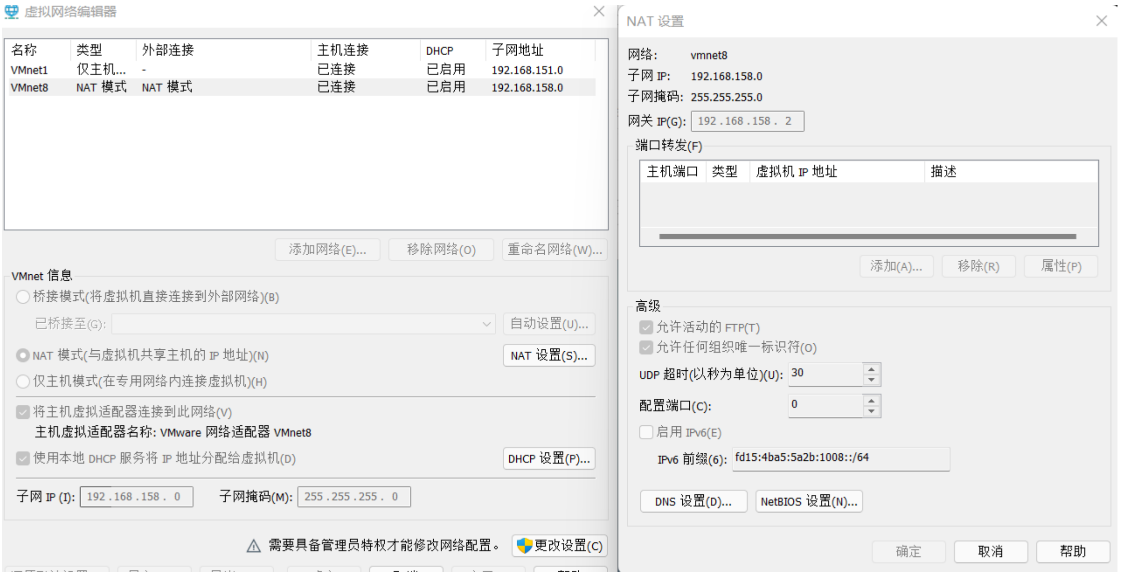
五、修改固定IP（所有机器都建议修改，修改方式相同）

1.ifconfig查看网卡



```
[root@node0 hadoop]# vi workers
[root@node0 hadoop]# ifconfig
ens33: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.158.137  netmask 255.255.255.0  broadcast 192.168.158.255
        inet6 fe80::15ed:fbe0:b4a8:a20  prefixlen 64  scopeid 0x20<link>
        ether 00:0c:29:54:e1:31  txqueuelen 1000  (Ethernet)
        RX packets 37668  bytes 7601169 (7.2 MiB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 25334  bytes 6489345 (6.1 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

```
2.vi /etc/sysconfig/network-scripts/ifcfg-ens33
    BOOTPROTO="static"(将原来的修改为static)

    IPADDR=192.168.158.137（自定义）
    GATEWAY=192.168.158.2(和虚拟网卡设置有关 -->网关IP的值)
    DNS1=192.168.158.2
```



虚拟网络编辑器

名称	类型	外部连接	主机连接	DHCP	子网地址
VMnet1	仅主机...	-	已连接	已启用	192.168.151.0
VMnet8	NAT 模式	NAT 模式	已连接	已启用	192.168.158.0

VMnet 信息

○ 桥接模式(将虚拟机直接连接到外部网络)(B)

已桥接至(G):

● NAT 模式(与虚拟机共享主机的 IP 地址)(N)

○ 仅主机模式(在专用网络内连接虚拟机)(H)

☑ 将主机虚拟适配器连接到此网络(V)

主机虚拟适配器名称: VMware 网络适配器 VMnet8

☑ 使用本地 DHCP 服务将 IP 地址分配给虚拟机(D)

子网 IP (I): 192 .168 .158 . 0　子网掩码(M): 255 .255 .255 . 0

⚠ 需要具备管理员特权才能修改网络配置。

NAT 设置

网络: vmnet8

子网 IP: 192.168.158.0

子网掩码: 255.255.255.0

网关 IP(G): 192 .168 .158 . 2

端口转发(F)

主机端口	类型	虚拟机 IP 地址	描述

高级

☑ 允许活动的 FTP(T)

☑ 允许任何组织唯一标识符(O)

UDP 超时(以秒为单位)(U): 30

配置端口(C): 0

□ 启用 IPv6(E)

IPv6 前缀(6): fd15:4ba5:5a2b:1008::/64

```
    重启网络
    service network restart
3.关闭防火墙
    systemctl disable firewalld
    重启后查看状态
    systemctl status firewalld
```

六、修改hostname和hosts(多台机器同时进行)

1.vi /etc/hostname

删除所有然后写上自己的命名 我的命名node0(不同主机不同命名)

2.vi /etc/hosts

```
192.168.158.137 node0
192.168.158.138 node1
192.168.158.139 node2
192.168.158.140 node3
```

七、免密登录（所有主机同时进行）

1.在家目录里建立一个1.sh,复制下面的东西进入1.sh

```
cd ~
vi 1.sh
```

----------------------复制进去

```
ssh-keygen -t rsa
ssh-copy-id -i ~/.ssh/id_rsa.pub node0
ssh-copy-id -i ~/.ssh/id_rsa.pub node1
ssh-copy-id -i ~/.ssh/id_rsa.pub node2
ssh-copy-id -i ~/.ssh/id_rsa.pub node3
```

2.执行1.sh

```
cd ~
bash 1.sh
```

八、建立自己需要的文件夹

1.建立logs文件夹（也可以不建立 后面运行时会自己creating,并且所有机器都要建立)

```
cd /software/hadoop-3.2.2/
mkdir logs
```

2.建立配置文件需要的文件夹(同样所有主机都要建立相同的)

```
cd /
mkdir data
cd data
mkdir hadoop
cd hadoop
mkdir hdfs tmp
cd hdfs
mkdir name data
```

九、修改hadoop配置文件

1.hadoop-env.sh

```
export JAVA_HOME=/software/jdk1.8.0_202
export HADOOP_HOME=/software/hadoop-3.2.2
export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

export HDFS_NAMENODE_USER=root
export HDFS_DATANODE_USER=root
export HDFS_SECONDARYNAMENODE_USER=root
```

```
    export YARN_RESOURCEMANAGER_USER=root
    export YARN_NODEMANAGER_USER=root
```

2.core-site.xml

```
    <configuration>
            <property>
                    <!-- 必须设置:默认文件系统（存储层和运算层解耦 -->
                    <!-- 此处值为uri结构: 使用内置的hdfs系统 端口号一般都是9000 -->
                    <name>fs.defaultFS</name>
                    <value>hdfs://node0:9000</value>
            </property>
            <property>
                    <!-- 必须设置：hadoop在本地的工作目录，用于放hadoop进程的临时数据，可以自己指定 -->
                    <name>hadoop.tmp.dir</name>
                    <value>/data/hadoop/tmp</value>
            </property>
    </configuration>
```

3.hdfs-site.xml

(需要自己建立文件夹)

```
    <configuration>
            <!-- hdfs存储数据的副本数量（避免一台宕机），可以不设置，默认值是3-->
            <property>
                    <name>dfs.replication</name>
                    <value>2</value>
           </property>

            <!--hdfs 监听namenode的web的地址，默认就是9870端口，如果不改端口也可以不设置 -->
            <property>
                    <name>dfs.namenode.http-address</name>
                    <value>node0:9870</value>
            </property>

            <!-- hdfs保存datanode当前数据的路径，默认值需要配环境变量，建议使用自己创建的路径，方便管理-->
            <property>
                    <name>dfs.datanode.data.dir</name>
                    <value>/data/hadoop/hdfs/data</value>
            </property>
    <!-- hdfs保存namenode当前数据的路径，默认值需要配环境变量，建议使用自己创建的路径，方便管理-->
    <property>
      <name>dfs.namenode.name.dir</name>
      <value>/data/hadoop/hdfs/name</value>
    </property>
  </configuration>
```

4.mapred-site.xml

```
  <configuration>
    <!-- 必须设置，mapreduce程序使用的资源调度平台，默认值是local，若不改就只能单机运行，不会到集群上了 -->
    <property>
      <name>mapreduce.framework.name</name>
      <value>yarn</value>
    </property>
    <!-- 这是3.2以上版本需要增加配置的，不配置运行mapreduce任务可能会有问题，记得使用自己的路径 -->
    <property>
      <name>mapreduce.application.classpath</name>
      <value>
        /software/hadoop-3.2.2/etc/hadoop,
        /software/hadoop-3.2.2/share/hadoop/common/*,
        /software/hadoop-3.2.2/share/hadoop/common/lib/*,
        /software/hadoop-3.2.2/hadoop/hdfs/*,
        /software/hadoop-3.2.2/share/hadoop/hdfs/lib/*,
        /software/hadoop-3.2.2/share/hadoop/mapreduce/*,
        /software/hadoop-3.2.2/share/hadoop/mapreduce/lib/*,
        /software/hadoop-3.2.2/share/hadoop/yarn/*,
        /software/hadoop-3.2.2/share/hadoop/yarn/lib/*
      </value>
    </property>
  </configuration>
```

5.yarn-site.xml

```
  <configuration>
    <!-- Site specific YARN configuration properties -->
    <!-- 必须配置 指定YARN的老大（ResourceManager）在哪一台主机 -->
    <property>
      <name>yarn.resourcemanager.hostname</name>
      <value>node0</value>
    </property>

    <!-- 必须配置 提供mapreduce程序获取数据的方式 默认为空 -->
```

```
        <property>
              <name>yarn.nodemanager.aux-services</name>
              <value>mapreduce_shuffle</value>
        </property>
  </configuration>
```

6.workers

```
node0
node1
node2
node3
```

十、发送配置文件

```
vi /software/hadoop-3.2.2/etc
scp -r hadoop root@node1:/software/hadoop-3.2.2/etc/
scp -r hadoop root@node2:/software/hadoop-3.2.2/etc/
scp -r hadoop root@node3:/software/hadoop-3.2.2/etc/
```

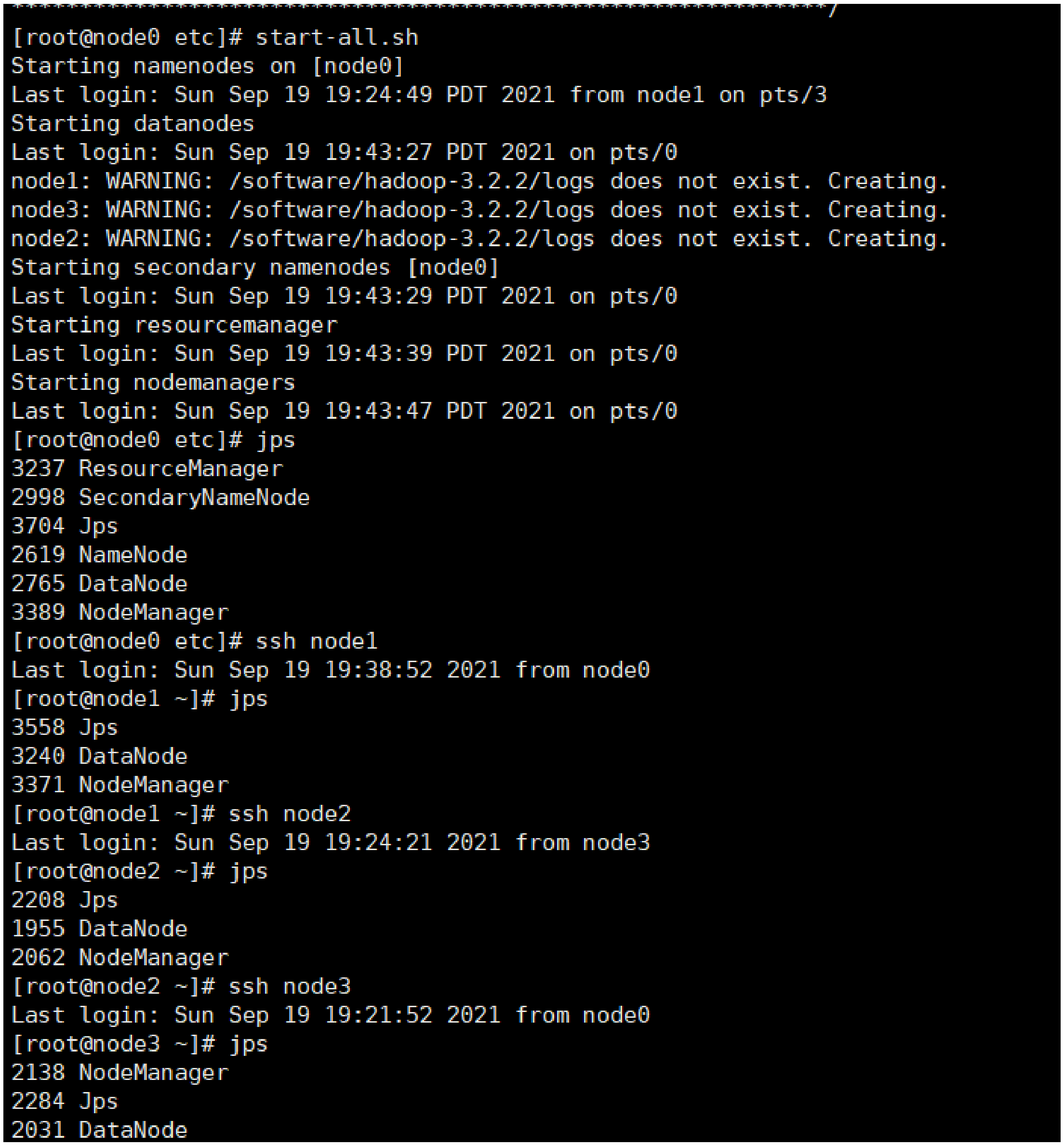
十一、格式化namenode

```
 hadoop namenode -format
```

十二、运行查看

start-all.sh(运行)

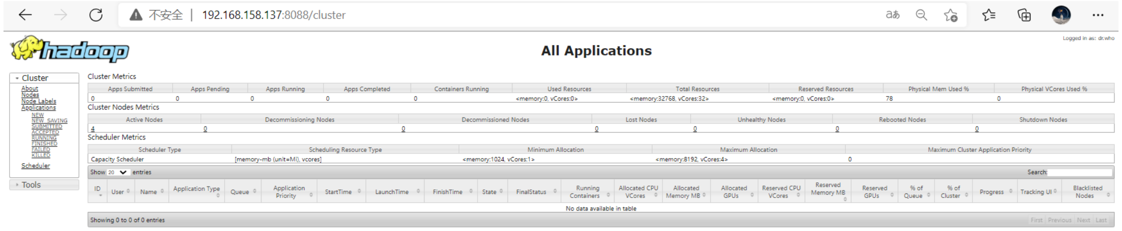
jps(查看)



```
************************************************************/
[root@node0 etc]# start-all.sh
Starting namenodes on [node0]
Last login: Sun Sep 19 19:24:49 PDT 2021 from node1 on pts/3
Starting datanodes
Last login: Sun Sep 19 19:43:27 PDT 2021 on pts/0
node1: WARNING: /software/hadoop-3.2.2/logs does not exist. Creating.
node3: WARNING: /software/hadoop-3.2.2/logs does not exist. Creating.
node2: WARNING: /software/hadoop-3.2.2/logs does not exist. Creating.
Starting secondary namenodes [node0]
Last login: Sun Sep 19 19:43:29 PDT 2021 on pts/0
Starting resourcemanager
Last login: Sun Sep 19 19:43:39 PDT 2021 on pts/0
Starting nodemanagers
Last login: Sun Sep 19 19:43:47 PDT 2021 on pts/0
[root@node0 etc]# jps
3237 ResourceManager
2998 SecondaryNameNode
3704 Jps
2619 NameNode
2765 DataNode
3389 NodeManager
[root@node0 etc]# ssh node1
Last login: Sun Sep 19 19:38:52 2021 from node0
[root@node1 ~]# jps
3558 Jps
3240 DataNode
3371 NodeManager
[root@node1 ~]# ssh node2
Last login: Sun Sep 19 19:24:21 2021 from node3
[root@node2 ~]# jps
2208 Jps
1955 DataNode
2062 NodeManager
[root@node2 ~]# ssh node3
Last login: Sun Sep 19 19:21:52 2021 from node0
[root@node3 ~]# jps
2138 NodeManager
2284 Jps
2031 DataNode
```

十三、windos通过web访问

1.http://192.168.158.137:8088



All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
0	0	0	0	0	<memory:0, vCores:0>	<memory:32768, vCores:32>	<memory:0, vCores:0>	78	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
4	0	0	0	0	0	0

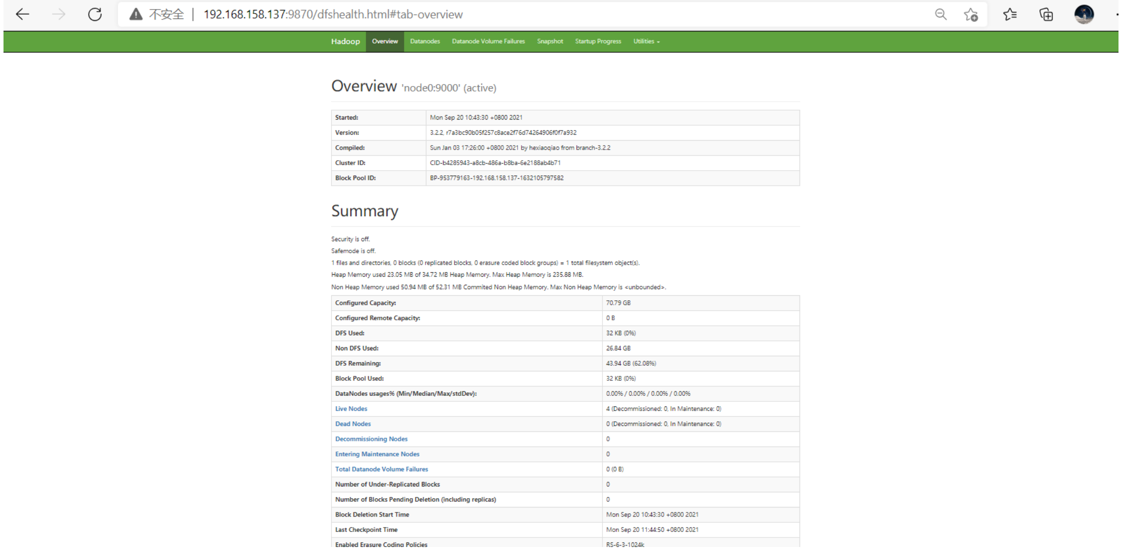
Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	Reserved CPU VCores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																						

Showing 0 to 0 of 0 entries

2.http://192.168.158.137:9870



Overview 'node0:9000' (active)

Started:	Mon Sep 20 10:43:30 +0800 2021
Version:	3.2.2, r7a3bc90b05f257c8ace2f76d74264906f0f7a932
Compiled:	Sun Jan 03 17:26:00 +0800 2021 by hexiaoqiao from branch-3.2.2
Cluster ID:	CID-b4285943-a8cb-486a-b8ba-6e2188ab4b71
Block Pool ID:	BP-953779163-192.168.158.137-1632105797582

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 23.05 MB of 34.72 MB Heap Memory. Max Heap Memory is 235.88 MB.

Non Heap Memory used 50.94 MB of 52.31 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	70.79 GB
Configured Remote Capacity:	0 B
DFS Used:	32 KB (0%)
Non DFS Used:	26.84 GB
DFS Remaining:	43.94 GB (62.08%)
Block Pool Used:	32 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	4 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Mon Sep 20 10:43:30 +0800 2021
Last Checkpoint Time	Mon Sep 20 11:44:50 +0800 2021
Enabled Erasure Coding Policies	RS-6-3-1024k

3.http://192.168.158.137:9868

无法访问，二进制编译问题 如果自己编译就不会出现问题