HDFS HA+Federation+YARN 部署

目录

1. 集群环境的节点分布

JournalNode：chinahadoop2　chinahadoop3　chinahadoop4

Cluster1 HA： chinahadoop1(Active NameNode) chinahadoop2(Standby NameNode)
Cluster2 HA： chinahadoop3(Active NameNode) chinahadoop4(Standby NameNode)
DataNode: chinahadoop1 chinahadoop2 chinahadoop3 chinahadoop4
NodeManager：chinahadoop1 chinahadoop2 chinahadoop3 chinahadoop4
ResourceManager：chinahadoop1

2. 搭建集群

2.1. 修改配置文件

hadoop-env.sh、core-site.xml、mapred-site.xml、yarn-site.xml、salves
这五个文件在每个节点上的内容是相同的。
注意在配置文件 hdfs-site.xml 中有一个地方不同。
在 Cluster1 的所有节点（chinahadoop1 和 chinahadoop2）上，要配置成：

```
<property>
    <name>dfs.namenode.shared.edits.dir</name>
    <value>qjournal://chinahadoop2:8485;chinahadoop3:8485;chinahadoop4:8485/chinahadoop-cluster1</value>
</property>
```

在 Cluster2 的所有节点（chinahadoop3 和 chinahadoop4）上，要配置成：

```
<property>
    <name>dfs.namenode.shared.edits.dir</name>
    <value>qjournal://chinahadoop2:8485;chinahadoop3:8485;chinahadoop4:8485/chinahadoop-cluster2</value>
</property>
```

配置文件下载连接（仅供参考）
链接：http://pan.baidu.com/s/1kU5RtZ5 密码：ms4s

2.2. 启动 JournalNode

在 chinahadoop2、chinahadoop3 和 chinahadoop4 上，
分别执行命令 sbin/hadoop-daemon.sh start journalnode

2.3. 在 Cluster1 上操作

在 chinahadoop1 上格式化 namenode
执行命令 bin/hdfs namenode -format -clusterId chinahadoop-cluster

```
16/01/31 17:51:06 INFO namenode.FSImage: Allocated new BlockPoolId: BP-120993581
4-192.168.1.108-1454233866943
16/01/31 17:51:06 INFO common.Storage: Storage directory /home/chinahadoop/hadoo
p/federation/hdfs/name has been successfully formatted.
16/01/31 17:51:07 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
16/01/31 17:51:07 INFO util.ExitUtil: Exiting with status 0
16/01/31 17:51:07 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at chinahadoop1/192.168.1.108
************************************************************/
```

已连接 chinahadoop1:22。 SSH2 xterm 80x28 28,42 7 会话 CAP NUM

格式化成功后，马上启动 namenode

执行命令 sbin/hadoop-daemon.sh start namenode

在 chinahadoop2 上格式化 namenode

执行命令 bin/hdfs namenode -bootstrapStandby

```
16/01/31 18:01:48 INFO namenode.TransferFsImage: Opening connection to http://ch
inahadoop1:50070/imagetransfer?getimage=1&txid=0&storageInfo=-57:1242172679:0:ch
inahadoop-cluster
16/01/31 18:01:48 INFO namenode.TransferFsImage: Image Transfer timeout configur
ed to 60000 milliseconds
16/01/31 18:01:48 INFO namenode.TransferFsImage: Transfer took 0.04s at 0.00 KB/
s
16/01/31 18:01:48 INFO namenode.TransferFsImage: Downloaded file fsimage.ckpt_00
0000000000000000 size 358 bytes.
16/01/31 18:01:48 INFO util.ExitUtil: Exiting with status 0
16/01/31 18:01:48 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at chinahadoop2/192.168.1.109
************************************************************/
[chinahadoop@chinahadoop1 custom-shell]$
```

已连接 chinahadoop1:22。 SSH2 xterm 80x28 28,42 7 会话 CAP NUM

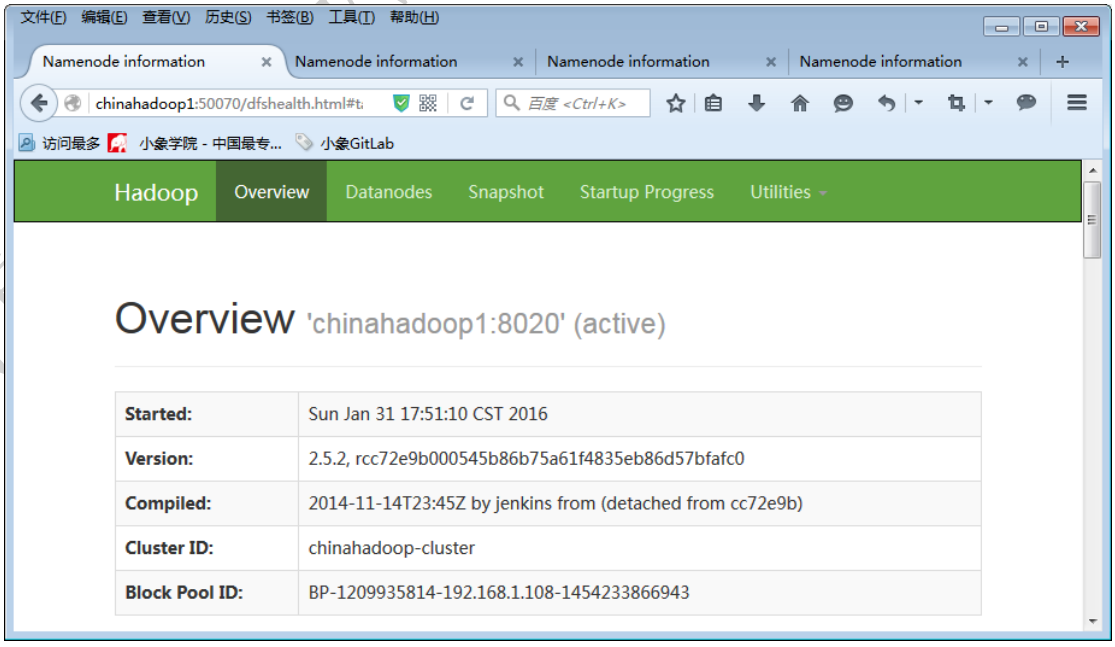
格式化成功后，马上启动 namenode

执行命令 sbin/hadoop-daemon.sh start namenode

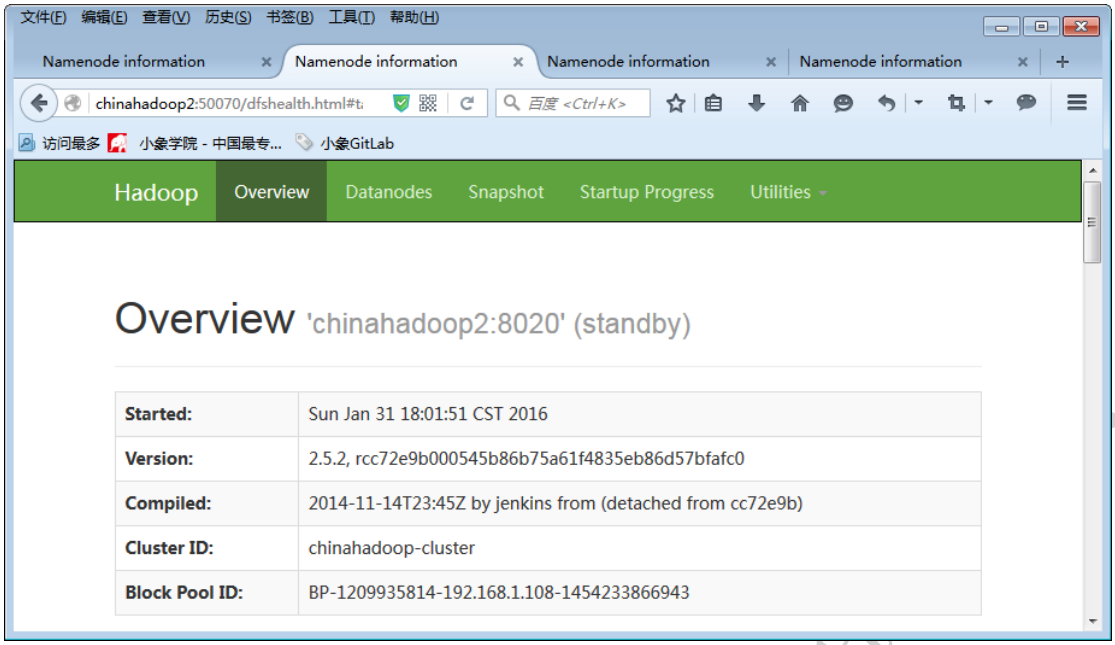
在 chinahadoop1 上切换 active namenode

执行命令 bin/hdfs haadmin -ns chinahadoop-cluster1 -transitionToActive nn1

在浏览器上访问 chinahadoop1:50070



在浏览器上访问 chinahadoop2:50070



2.4. 在 Cluster2 上操作

在 chinahadoop3 上格式化 namenode

执行命令 bin/hdfs namenode -format -clusterId chinahadoop-cluster

```
16/01/31 18:07:13 INFO namenode.FSImage: Allocated new BlockPoolId: BP-83990692-
192.168.1.110-1454234833946
16/01/31 18:07:13 INFO common.Storage: Storage directory /home/chinahadoop/hadoo
p/federation/hdfs/name has been successfully formatted.
16/01/31 18:07:14 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
16/01/31 18:07:14 INFO util.ExitUtil: Exiting with status 0
16/01/31 18:07:14 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at chinahadoop3/192.168.1.110
************************************************************/
```

格式化成功后，马上启动 namenode

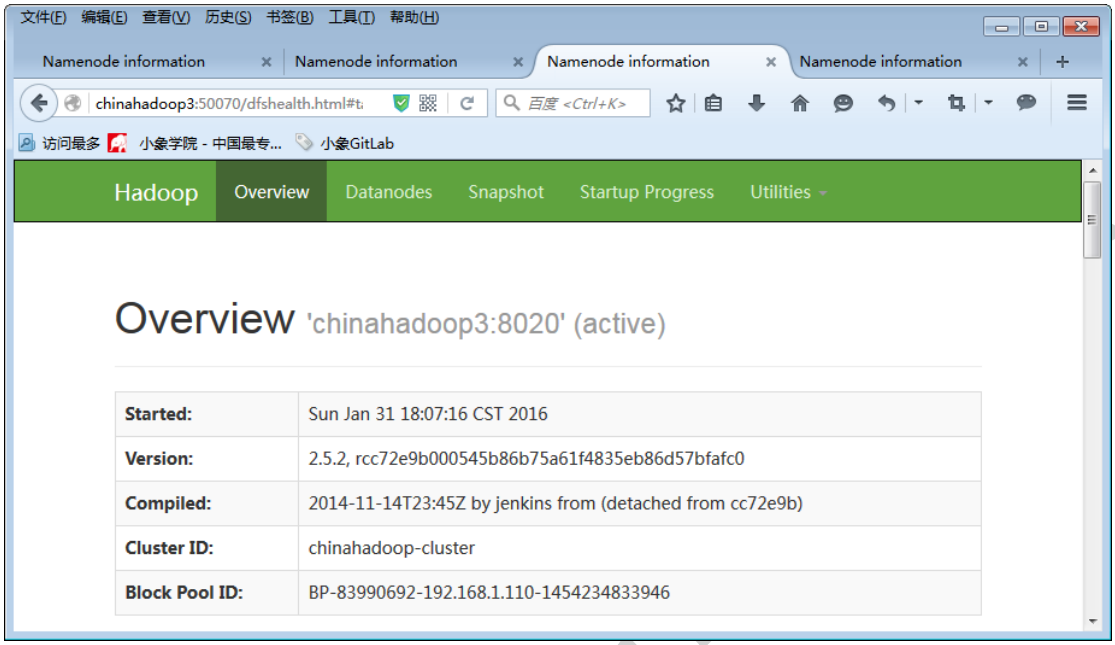
执行命令 sbin/hadoop-daemon.sh start namenode

在 chinahadoop4 上格式化 namenode

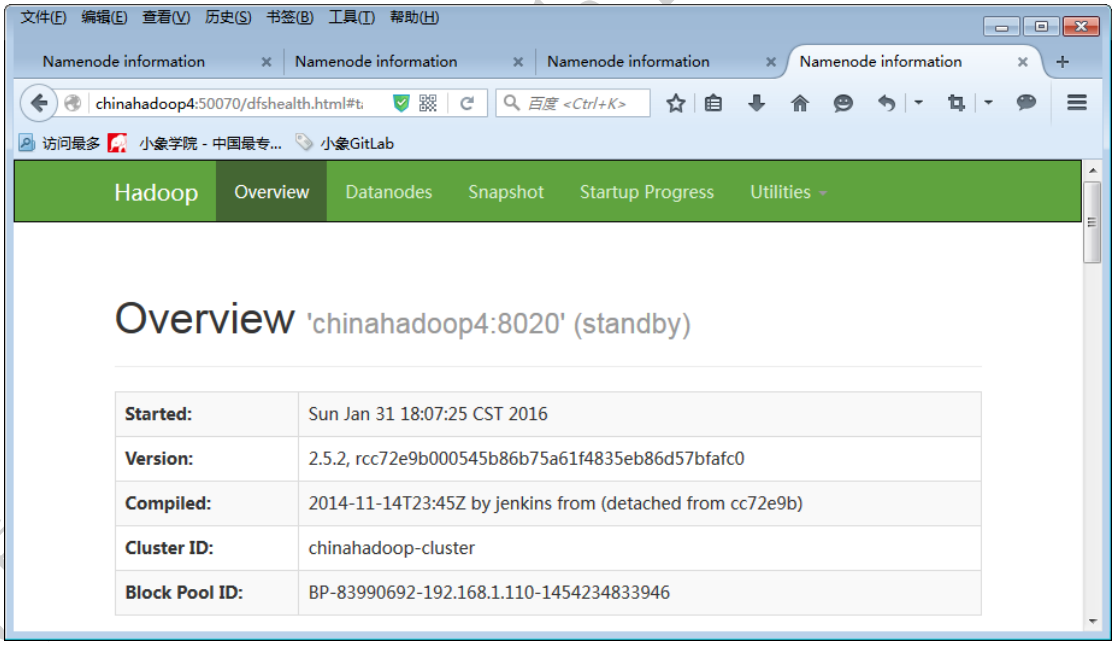
执行命令 bin/hdfs namenode -bootstrapStandby

```
16/01/31 18:07:22 INFO namenode.TransferFsImage: Opening connection to http://ch
inahadoop3:50070/imagetransfer?getimage=1&txid=0&storageInfo=-57:2133992084:0:ch
inahadoop-cluster
16/01/31 18:07:22 INFO namenode.TransferFsImage: Image Transfer timeout configur
ed to 60000 milliseconds
16/01/31 18:07:22 INFO namenode.TransferFsImage: Transfer took 0.05s at 0.00 KB/
s
16/01/31 18:07:22 INFO namenode.TransferFsImage: Downloaded file fsimage.ckpt_00
0000000000000000 size 358 bytes.
16/01/31 18:07:22 INFO util.ExitUtil: Exiting with status 0
16/01/31 18:07:22 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at chinahadoop4/192.168.1.111
************************************************************/
[chinahadoop@chinahadoop1 custom-shell]$
```

格式化成功后，马上启动 namenode
执行命令 sbin/hadoop-daemon.sh start namenode
在 chinahadoop3 上切换 active namenode
执行命令 bin/hdfs haadmin -ns chinahadoop-cluster2 -transitionToActive nn3
在浏览器上访问 chinahadoop3:50070



在浏览器上访问 chinahadoop4:50070



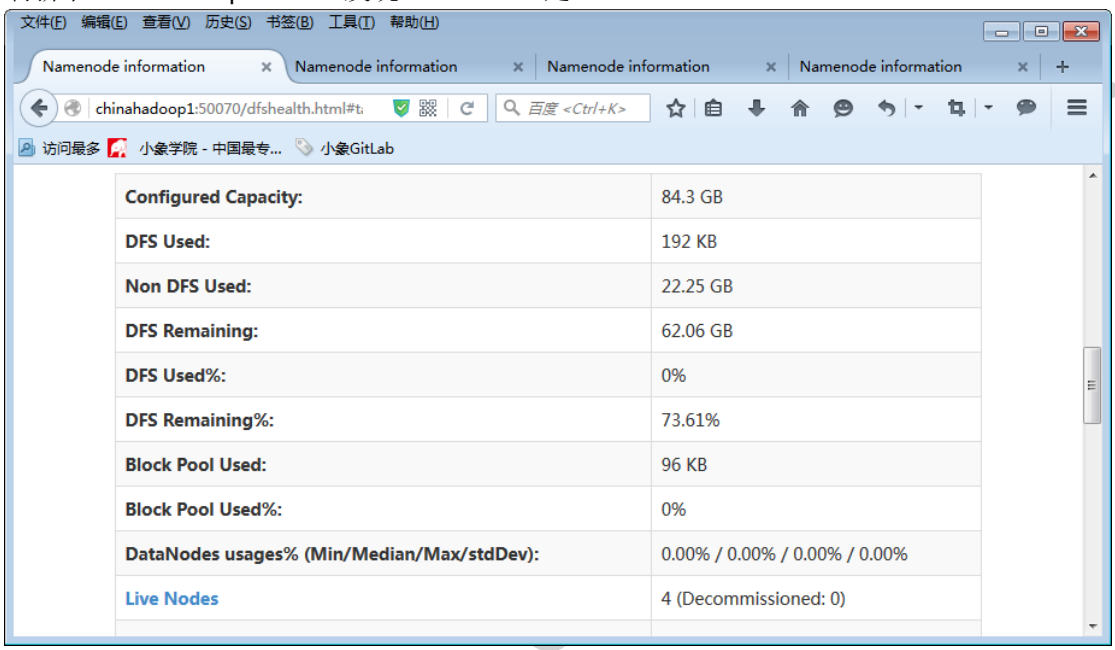
2.5. 启动 DataNode

在 chinahadoop1 上执行命令 sbin/hadoop-daemons.sh start datanode

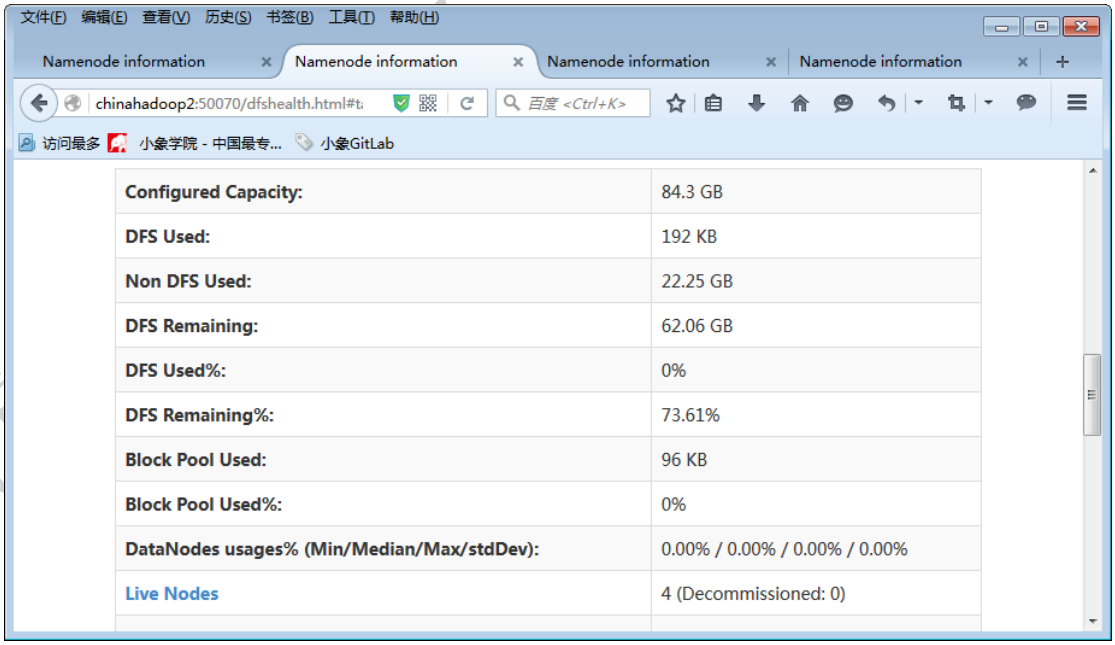
```
[chinahadoop@chinahadoop1 hadoop-2.5.2]$ sbin/hadoop-daemons.sh start datanode
chinahadoop4: starting datanode, logging to /home/chinahadoop/hadoop/federation/
hadoop-2.5.2/logs/hadoop-chinahadoop-datanode-chinahadoop4.out
chinahadoop2: starting datanode, logging to /home/chinahadoop/hadoop/federation/
hadoop-2.5.2/logs/hadoop-chinahadoop-datanode-chinahadoop2.out
chinahadoop3: starting datanode, logging to /home/chinahadoop/hadoop/federation/
hadoop-2.5.2/logs/hadoop-chinahadoop-datanode-chinahadoop3.out
chinahadoop1: starting datanode, logging to /home/chinahadoop/hadoop/federation/
hadoop-2.5.2/logs/hadoop-chinahadoop-datanode-chinahadoop1.out
[chinahadoop@chinahadoop1 hadoop-2.5.2]$
```

已连接 chinahadoop1:22。 SSH2 xterm 80x28 28,42 7 会话 CAP NUM

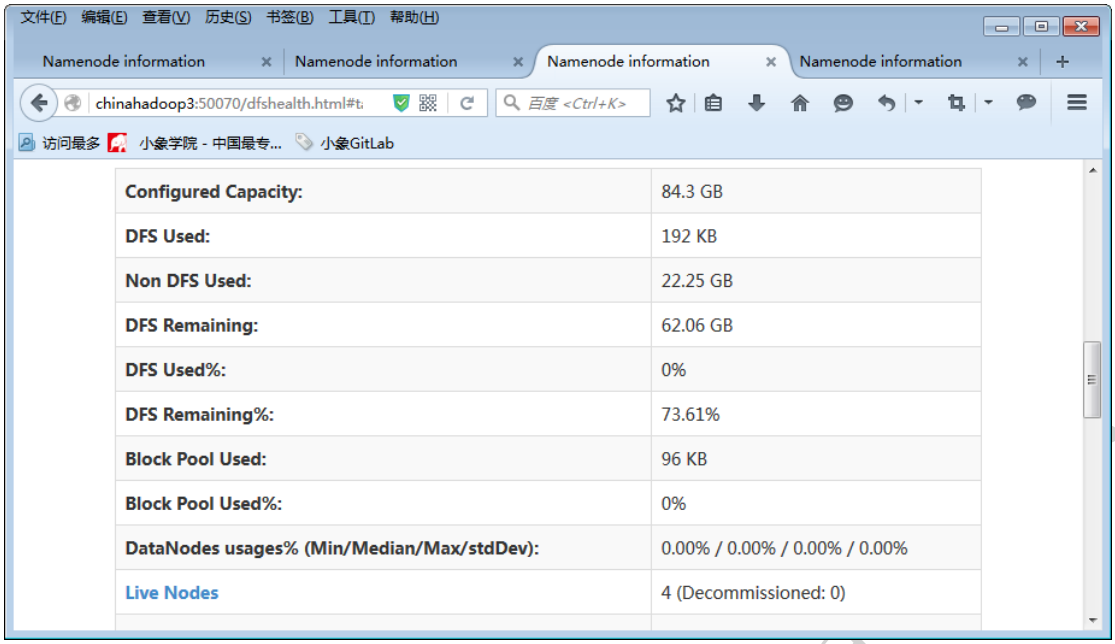
刷新下 chinahadoop1:50070 发现 Live Nodes 是 4



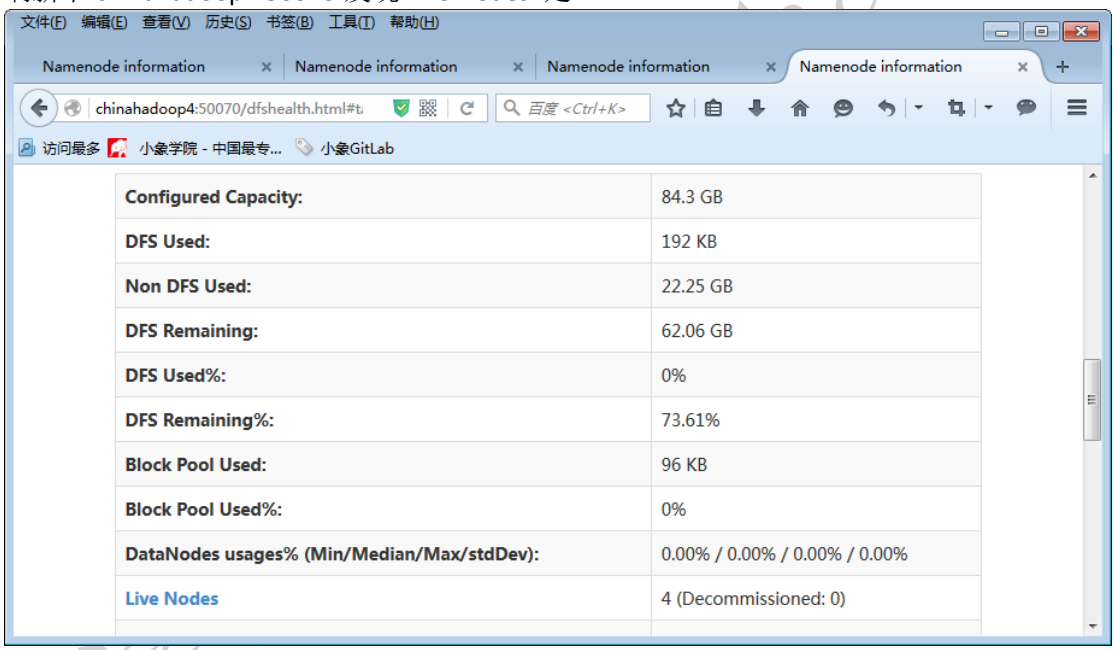
刷新下 chinahadoop2:50070 发现 Live Nodes 是 4



刷新下 chinahadoop3:50070 发现 Live Nodes 是 4



刷新下 chinahadoop4:50070 发现 Live Nodes 是 4



2.6. 启动 yarn

在 chinahadoop1 上执行命令 sbin/start-yarn.sh

```
[chinahadoop@chinahadoop1 hadoop-2.5.2]$ sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/chinahadoop/hadoop/federation/hadoop-
2.5.2/logs/yarn-chinahadoop-resourcemanager-chinahadoop1.out
chinahadoop2: starting nodemanager, logging to /home/chinahadoop/hadoop/federati
on/hadoop-2.5.2/logs/yarn-chinahadoop-nodemanager-chinahadoop2.out
chinahadoop4: starting nodemanager, logging to /home/chinahadoop/hadoop/federati
on/hadoop-2.5.2/logs/yarn-chinahadoop-nodemanager-chinahadoop4.out
chinahadoop3: starting nodemanager, logging to /home/chinahadoop/hadoop/federati
on/hadoop-2.5.2/logs/yarn-chinahadoop-nodemanager-chinahadoop3.out
chinahadoop1: starting nodemanager, logging to /home/chinahadoop/hadoop/federati
on/hadoop-2.5.2/logs/yarn-chinahadoop-nodemanager-chinahadoop1.out
[chinahadoop@chinahadoop1 hadoop-2.5.2]$
```

已连接 chinahadoop1:22。 SSH2 xterm 80x28 28,42 7 会话 CAP NUM

3. 执行一个 MapReduce 任务

执行命令

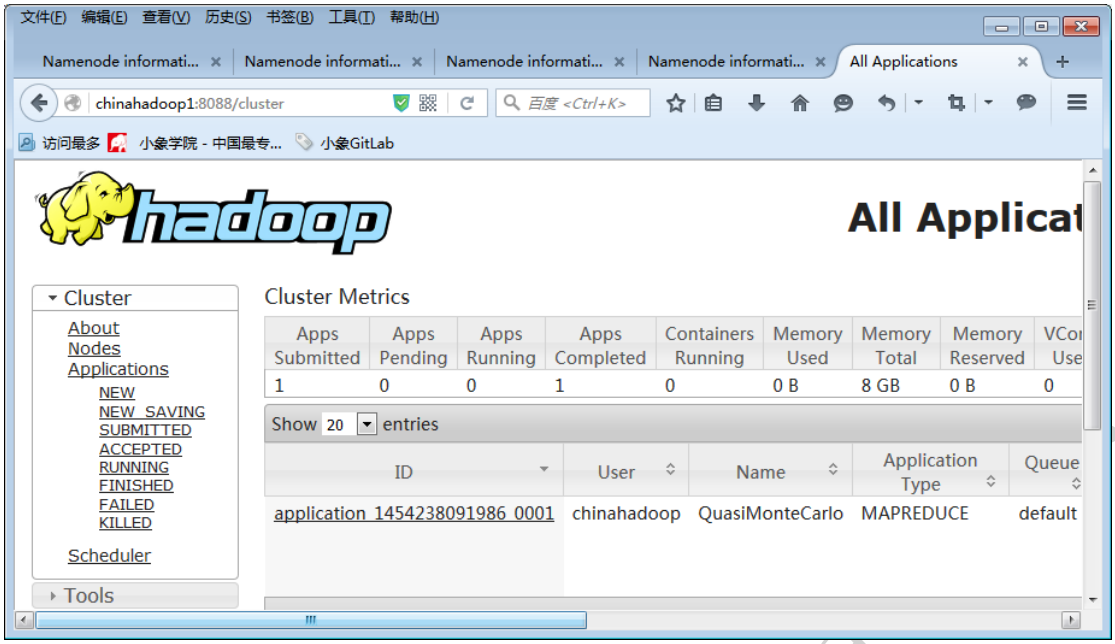
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.2.jar pi 2 10

```
[chinahadoop@chinahadoop1 hadoop-2.5.2]$ bin/hadoop jar share/hadoop/mapreduce/h
adoop-mapreduce-examples-2.5.2.jar pi 2 10
Number of Maps  = 2
Samples per Map = 10
16/01/31 19:02:45 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Wrote input for Map #0
Wrote input for Map #1
Starting Job
16/01/31 19:02:48 INFO client.RMProxy: Connecting to ResourceManager at chinahad
oop1/192.168.1.108:8032
16/01/31 19:02:50 INFO input.FileInputFormat: Total input paths to process : 2
16/01/31 19:02:51 INFO mapreduce.JobSubmitter: number of splits:2
16/01/31 19:02:53 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_14
54238091986_0001
16/01/31 19:02:54 INFO impl.YarnClientImpl: Submitted application application_14
54238091986_0001
16/01/31 19:02:54 INFO mapreduce.Job: The url to track the job: http://chinahado
op1:8088/proxy/application_1454238091986_0001/
16/01/31 19:02:54 INFO mapreduce.Job: Running job: job_1454238091986_0001
16/01/31 19:03:08 INFO mapreduce.Job: Job job_1454238091986_0001 running in uber
 mode : false
16/01/31 19:03:08 INFO mapreduce.Job:  map 0% reduce 0%
16/01/31 19:03:22 INFO mapreduce.Job:  map 100% reduce 0%
16/01/31 19:03:36 INFO mapreduce.Job:  map 100% reduce 100%
16/01/31 19:03:40 INFO mapreduce.Job: Job job_1454238091986_0001 completed succe
ssfully
16/01/31 19:03:40 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=50
                FILE: Number of bytes written=299316
                FILE: Number of read operations=0
```

```
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=546
                HDFS: Number of bytes written=215
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=3
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=23312
                Total time spent by all reduces in occupied slots (ms)=11054
                Total time spent by all map tasks (ms)=23312
                Total time spent by all reduce tasks (ms)=11054
                Total vcore-seconds taken by all map tasks=23312
                Total vcore-seconds taken by all reduce tasks=11054
                Total megabyte-seconds taken by all map tasks=23871488
                Total megabyte-seconds taken by all reduce tasks=11319296
        Map-Reduce Framework
                Map input records=2
                Map output records=4
                Map output bytes=36
                Map output materialized bytes=56
                Input split bytes=310
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=56
                Reduce input records=4
                Reduce output records=0
                Spilled Records=8
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=296
                CPU time spent (ms)=1860
                Physical memory (bytes) snapshot=515444736
                Virtual memory (bytes) snapshot=2921689088
                Total committed heap usage (bytes)=257433600
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=236
        File Output Format Counters
                Bytes Written=97
Job Finished in 52.129 seconds
Estimated value of Pi is 3.80000000000000000000
[chinahadoop@chinahadoop1 hadoop-2.5.2]$
已连接 chinahadoop1:22。   SSH2 xterm 80x32 32,42 7 会话   CAP NUM
```

在浏览器上访问 chinahadoop1:8088

可以看到任务 application_1454238091986_0001



4. 停止集群

使用自定义脚本停止集群 sh stop_chinahadoop.sh

```
[chinahadoop@chinahadoop1 custom-shell]$ sh stop_chinahadoop.sh
16/01/31 19:24:45 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Stopping namenodes on [chinahadoop1 chinahadoop2 chinahadoop3 chinahadoop4]
chinahadoop2: stopping namenode
chinahadoop4: stopping namenode
chinahadoop1: stopping namenode
chinahadoop3: stopping namenode
chinahadoop1: stopping datanode
chinahadoop2: stopping datanode
chinahadoop4: stopping datanode
chinahadoop3: stopping datanode
Stopping journal nodes [chinahadoop2 chinahadoop3 chinahadoop4]
chinahadoop4: stopping journalnode
chinahadoop2: stopping journalnode
chinahadoop3: stopping journalnode
16/01/31 19:25:08 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
stopping yarn daemons
stopping resourcemanager
chinahadoop4: stopping nodemanager
chinahadoop1: stopping nodemanager
chinahadoop2: stopping nodemanager
chinahadoop3: stopping nodemanager
no proxyserver to stop
[chinahadoop@chinahadoop1 custom-shell]$
```

5. 自定义脚本

下图显示的是这次搭建集群时写的一些自定义脚本。

```
[chinahadoop@chinahadoop1 custom-shell]$ pwd
/home/chinahadoop/hadoop/federation/custom-shell
[chinahadoop@chinahadoop1 custom-shell]$ ls
active_cluster1_master.sh  cluster2-master-conf       rsync_chinahadoop.sh
active_cluster2_master.sh  cluster2-standby-conf      slaves-conf
cluster1-conf              init_cluster1.sh           start_chinahadoop.sh
cluster1-master-conf       init_cluster2.sh           start_journalnode.sh
cluster1-standby-conf      journalnode-conf           stop_chinahadoop.sh
cluster2-conf              rsync_chinahadoop_file.sh  stop_journalnode.sh
[chinahadoop@chinahadoop1 custom-shell]$
```

已连接 chinahadoop1:22。 SSH2 xterm 80x27 27,42 1 会话 CAP NUM

自定义脚本文件下载连接（仅供参考）

链接：http://pan.baidu.com/s/1mi8hzLI 密码：hlen