

思考题：hadoop 完全分布式(3 台虚拟机) 步骤：

一 部署多节点

二 启动 hadoop 完全分布式

三 wordcount 测试

四 日志查看

以下详细说明：

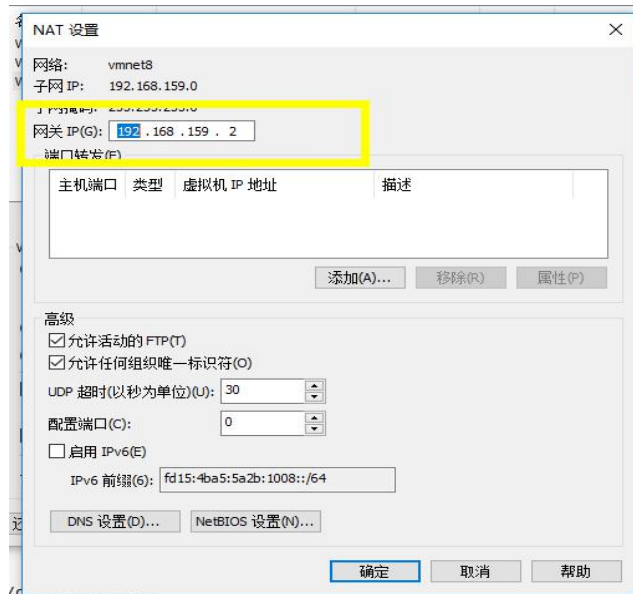
一、部署多节点

① NAT 静态 IP 配置：

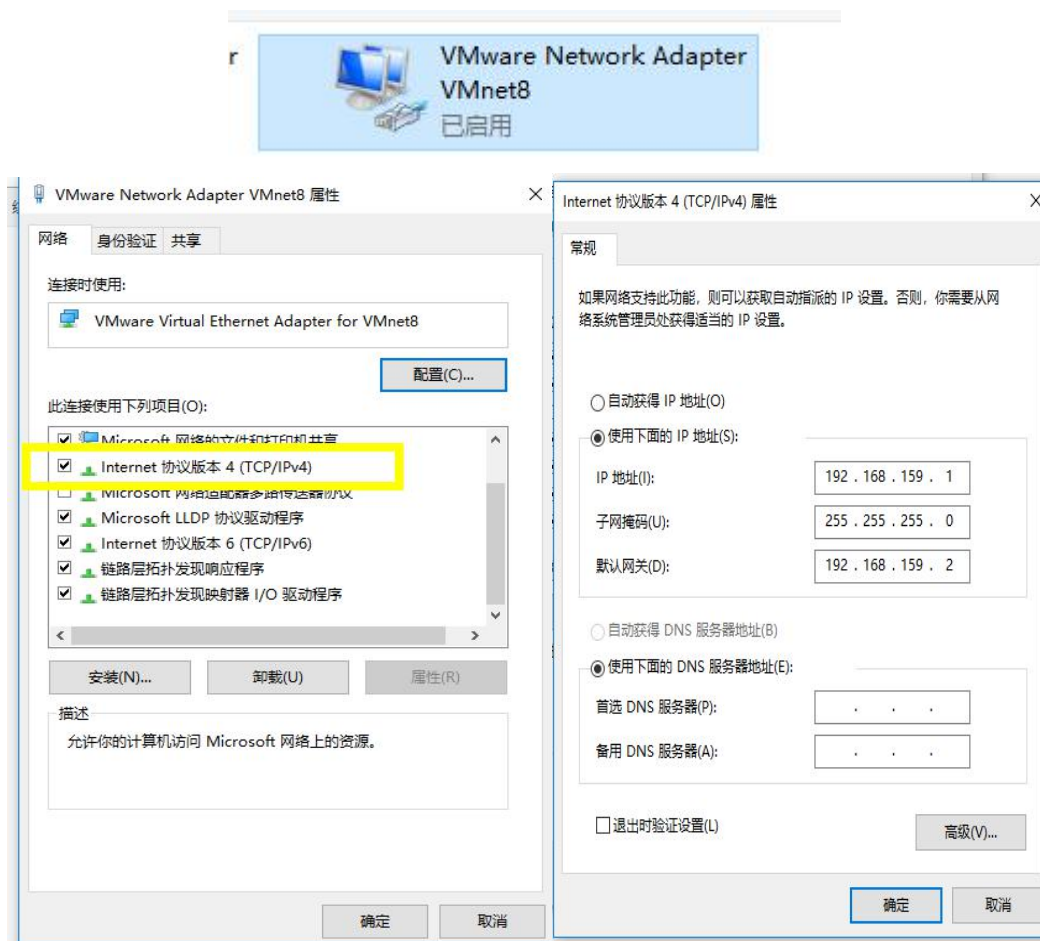
■ 点击编辑-虚拟网络编辑器，点击 VMnet8，子网 IP 设置成如下：192.168.159.0



■ 点击 Net 设置，网关 IP 设置成如下：192.168.159.2。 点击确定。



② 对主机的虚拟网卡进行设置:网络-属性-更改适配器设置,右键属性,选择 TCP/IPv4,配置如下:



③ 修改主机 hosts 文件:在路径框中输入:C:\Windows\System32\drivers\etc,将 hosts 文件复制到桌面(因为权限原因无法直接修改),打开在其中添加:

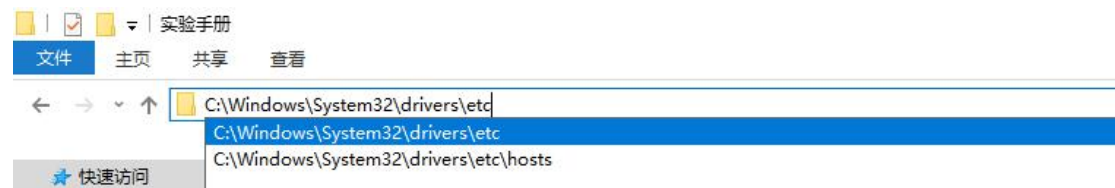
#apache hadoop

192.168.159.160 amaster.basichadoop.com amaster

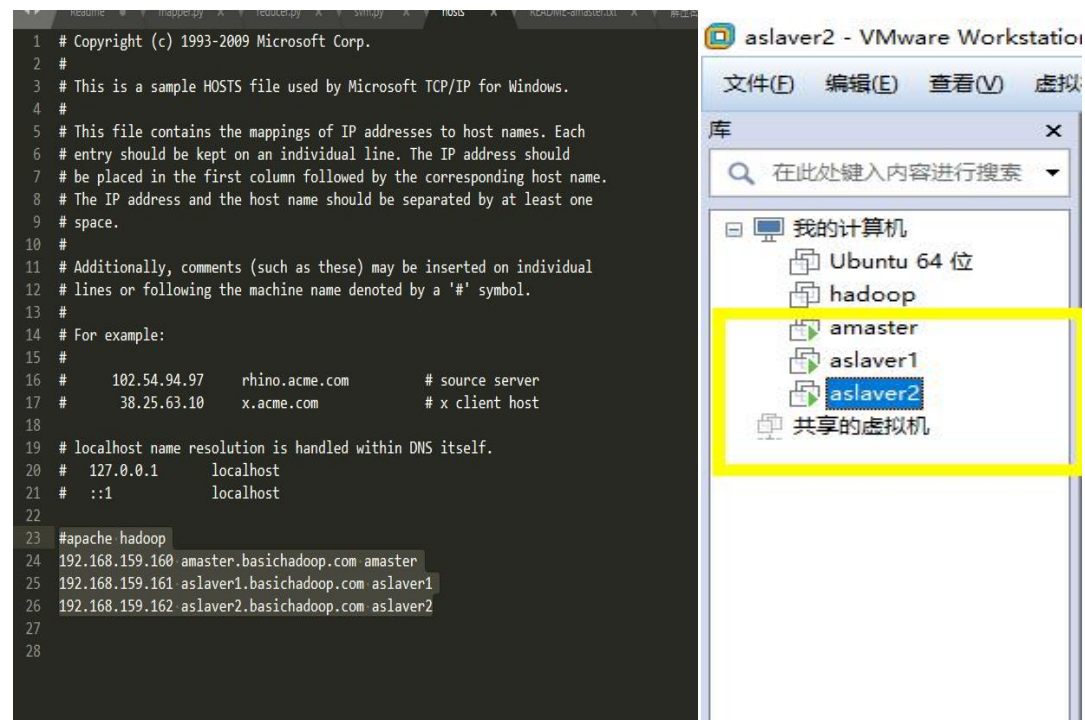
192.168.159.161 aslaver1.basichadoop.com aslaver1

192.168.159.162 aslaver2.basichadoop.com aslaver2

保存，复制回原路径覆盖即可。



名称	修改日期	类型	大小
hosts	2018/7/2 22:42	文件	1 KB
lmhosts.sam	2017/9/29 21:44	SAM 文件	4 KB
networks	2017/9/29 21:44	文件	1 KB
protocol	2017/9/29 21:44	文件	2 KB
services	2017/9/29 21:44	文件	18 KB



此时 IP 配置完成，开始启动 hadoop 三个节点。

二、启动 hadoop 完全分布式

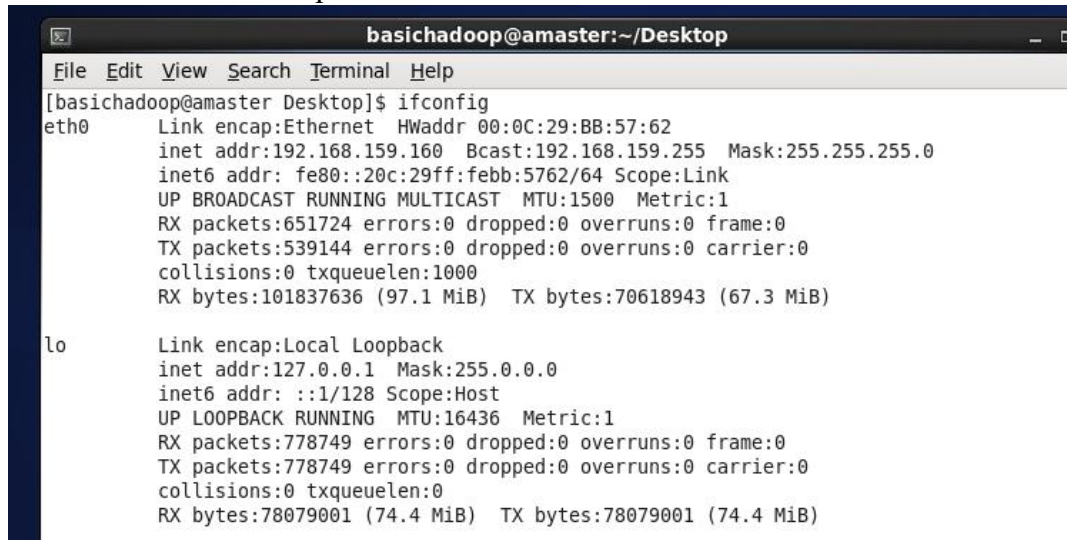
① 因为是三台虚拟机组成的完全分布式，我们第一步当然是测试相互间是否通信正常：

■ 查看三台虚拟机的 IP：\$ ifconfig

我们发现：amaster ip:192.168.159.160

aslaver1 ip:192.168.159.161

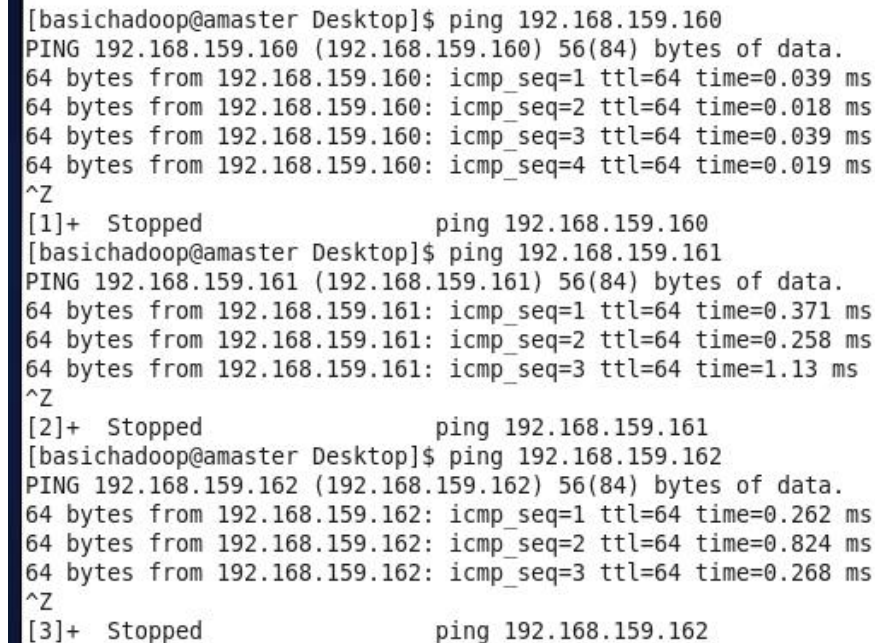
aslaver2 ip:192.168.159.162



```
basichadoop@amaster:~/Desktop
File Edit View Search Terminal Help
[basichadoop@amaster Desktop]$ ifconfig
eth0      Link encap:Ethernet  HWaddr 00:0C:29:BB:57:62
          inet addr:192.168.159.160  Bcast:192.168.159.255  Mask:255.255.255.0
          inet6 addr: fe80::20c:29ff:febb:5762/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:651724 errors:0 dropped:0 overruns:0 frame:0
          TX packets:539144 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:101837636 (97.1 MiB)  TX bytes:70618943 (67.3 MiB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          inet6 addr: ::1/128 Scope:Host
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
          RX packets:778749 errors:0 dropped:0 overruns:0 frame:0
          TX packets:778749 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:78079001 (74.4 MiB)  TX bytes:78079001 (74.4 MiB)
```

■ 我们在主节点 amaster 中分别 ping 一下这三个 IP：皆可 ping 通，相互通信正常。



```
[basichadoop@amaster Desktop]$ ping 192.168.159.160
PING 192.168.159.160 (192.168.159.160) 56(84) bytes of data.
64 bytes from 192.168.159.160: icmp_seq=1 ttl=64 time=0.039 ms
64 bytes from 192.168.159.160: icmp_seq=2 ttl=64 time=0.018 ms
64 bytes from 192.168.159.160: icmp_seq=3 ttl=64 time=0.039 ms
64 bytes from 192.168.159.160: icmp_seq=4 ttl=64 time=0.019 ms
^Z
[1]+  Stopped                  ping 192.168.159.160
[basichadoop@amaster Desktop]$ ping 192.168.159.161
PING 192.168.159.161 (192.168.159.161) 56(84) bytes of data.
64 bytes from 192.168.159.161: icmp_seq=1 ttl=64 time=0.371 ms
64 bytes from 192.168.159.161: icmp_seq=2 ttl=64 time=0.258 ms
64 bytes from 192.168.159.161: icmp_seq=3 ttl=64 time=1.13 ms
^Z
[2]+  Stopped                  ping 192.168.159.161
[basichadoop@amaster Desktop]$ ping 192.168.159.162
PING 192.168.159.162 (192.168.159.162) 56(84) bytes of data.
64 bytes from 192.168.159.162: icmp_seq=1 ttl=64 time=0.262 ms
64 bytes from 192.168.159.162: icmp_seq=2 ttl=64 time=0.824 ms
64 bytes from 192.168.159.162: icmp_seq=3 ttl=64 time=0.268 ms
^Z
[3]+  Stopped                  ping 192.168.159.162
```

■ 继续在本地命令行中 ping 一下，查看通信是否正常：

ping amaster.basichadoop.com

ping aslaver1.basichadoop.com

ping aslaver2.basichadoop.com

```
C:\Users\gsz>ping amaster.basichadoop.com

正在 Ping amaster.basichadoop.com [192.168.159.160] 具有 32 字节的数据:
来自 192.168.159.160 的回复: 字节=32 时间=1ms TTL=64
来自 192.168.159.160 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.160 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.160 的回复: 字节=32 时间<1ms TTL=64

192.168.159.160 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 0ms, 最长 = 1ms, 平均 = 0ms

C:\Users\gsz>ping aslaver1.basichadoop.com

正在 Ping aslaver1.basichadoop.com [192.168.159.161] 具有 32 字节的数据:
来自 192.168.159.161 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.161 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.161 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.161 的回复: 字节=32 时间<1ms TTL=64

192.168.159.161 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 0ms, 最长 = 0ms, 平均 = 0ms

C:\Users\gsz>ping aslaver2.basichadoop.com

正在 Ping aslaver2.basichadoop.com [192.168.159.162] 具有 32 字节的数据:
来自 192.168.159.162 的回复: 字节=32 时间=1ms TTL=64
来自 192.168.159.162 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.162 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.159.162 的回复: 字节=32 时间<1ms TTL=64

192.168.159.162 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
    最短 = 0ms, 最长 = 1ms, 平均 = 0ms
```

通信一切正常后，我们开始启动节点。

② 第一步，先启动 zookeeper。

■ 三个节点都执行一遍：

```
$ cd /opt/modules/zookeeper-3.4.5
```

```
$ bin/zkServer.sh start
```

```
[basichadoop@amaster Desktop]$
[basichadoop@amaster Desktop]$ cd /opt/modules/zookeeper-3.4.5
[basichadoop@amaster zookeeper-3.4.5]$
[basichadoop@amaster zookeeper-3.4.5]$

[basichadoop@aslaver1 Desktop]$ cd /opt/modules/zookeeper-3.4.5
[basichadoop@aslaver1 zookeeper-3.4.5]$
[basichadoop@aslaver1 zookeeper-3.4.5]$
[basichadoop@aslaver1 zookeeper-3.4.5]$ bin/zkServer.sh start
JMX enabled by default
Using config: /opt/modules/zookeeper-3.4.5/bin/../conf/zoo.cfg
Starting zookeeper ... STARTED
```

■ 三个节点都查看一下节点启动状态：

```
$ bin/zkServer.sh status
```

```
[basichadoop@aslaver1 zookeeper-3.4.5]$ bin/zkServer.sh status
JMX enabled by default
Using config: /opt/modules/zookeeper-3.4.5/bin/../conf/zoo.cfg
Mode: leader
```

zk 集群启动完毕.

③ 第二步, 启动 hadoop 的 hdfs 与 yarn:

- 节点 **amaster** 中, 进入 hadoop 目录, 启动 hdfs。注意每一个节点启动了不同任务。

```
$ cd /opt/modules/hadoop-2.6.5
```

```
$ sbin/start-dfs.sh
```

```
[basichadoop@amaster zookeeper-3.4.5]$ cd /opt/modules/hadoop-2.6.5
[basichadoop@amaster hadoop-2.6.5]$
[basichadoop@amaster hadoop-2.6.5]$
[basichadoop@amaster hadoop-2.6.5]$
[basichadoop@amaster hadoop-2.6.5]$ sbin/start-dfs.sh

Starting namenodes on [amaster aslaver1]
amaster: starting namenode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-bas
aslaver1: starting namenode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-ba
basichadoop namenode-aslaver1.basichadoop.com.out
aslaver1: starting datanode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-ba
basichadoop datanode-aslaver1.basichadoop.com.out
amaster: starting datanode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-bas
basichadoop datanode-amaster.basichadoop.com.out
aslaver2: starting datanode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-ba
basichadoop datanode-aslaver2.basichadoop.com.out
Starting journal nodes [amaster aslaver1 aslaver2]
amaster: starting journalnode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-
basichadoop-journalnode-amaster.basichadoop.com.out
aslaver2: starting journalnode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop
-basichadoop-journalnode-aslaver2.basichadoop.com.out
aslaver1: starting journalnode, logging to /opt/modules/hadoop-2.6.5/logs/hadoop
-basichadoop-journalnode-aslaver1.basichadoop.com.out
Starting Hadoop Failover Controllers on NN hosts [amaster aslaver1]
amaster: starting zkfc, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-basicha
dooop-zkfc-amaster.basichadoop.com.out
aslaver1: starting zkfc, logging to /opt/modules/hadoop-2.6.5/logs/hadoop-basich
adooop-zkfc-aslaver1.basichadoop.com.out
```

- 节点 **aslaver2** 中, 进入 hadoop 目录, 启动 MR 的 historyserver 进程和 yarn。

```
$ cd /opt/modules/hadoop-2.6.5
```

```
$ sbin/mr-jobhistory-daemon.sh start historyserver
```

```
$ sbin/start-yarn.sh
```

此时 hadoo 的历史日志进程和 yarn 资源节点管理都成功启动。

```
[basichadoop@aslaver2 zookeeper-3.4.5]$
[basichadoop@aslaver2 hadoop-2.6.5]$ sbin/mr-jobhistory-daemon.sh start historyserver
starting historyserver, logging to /opt/modules/hadoop-2.6.5/logs/mapred-basichadoop-historyserver-a
aslaver2.basichadoop.com.out
[basichadoop@aslaver2 hadoop-2.6.5]$ sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /opt/modules/hadoop-2.6.5/logs/yarn-basichadoop-resourcemanager-aslaver
2.basichadoop.com.out
aslaver1: starting nodemanager, logging to /opt/modules/hadoop-2.6.5/logs/yarn-basichadoop-nodemanager-aslav
er1.basichadoop.com.out
amaster: starting nodemanager, logging to /opt/modules/hadoop-2.6.5/logs/yarn-basichadoop-nodemanager-amaste
r.basichadoop.com.out
aslaver2: starting nodemanager, logging to /opt/modules/hadoop-2.6.5/logs/yarn-basichadoop-nodemanager-aslav
er2.basichadoop.com.out
```

④ 现在我们查看一下 jps, 看一下各节点启动情况。

- amaster 中: `$ jps`

```
[basicadoop@amaster hadoop-2.6.5]$ jps
3448 JournalNode
3160 NameNode
3917 Jps
3262 DataNode
3619 DFSZKFailoverController
2857 QuorumPeerMain
3768 NodeManager
```

■ aslaver1 中: \$jps

```
[basicadoop@aslaver1 zookeeper-3.4.5]$ jps
2656 QuorumPeerMain
2748 NameNode
2984 DFSZKFailoverController
2812 DataNode
2891 JournalNode
3185 NodeManager
3349 Jps
```

■ aslaver2 中: \$jps

```
[basicadoop@aslaver2 hadoop-2.6.5]$ jps
2624 QuorumPeerMain
2792 JournalNode
3133 NodeManager
2922 JobHistoryServer
3471 Jps
3030 ResourceManager
2722 DataNode
```

我们对照一下应该启动起来的说明，发现全部启动完成。

	192.168.159.160	192.168.159.161	192.168.159.162	
	amaster	aslaver1	aslaver2	
zookeeper	QuorumPeerMain	QuorumPeerMain	QuorumPeerMain	
hdfs	namenode	namenode		zk选举
hdfs	datanode	datanode	datanode	
hdfs		secondnamenode		
hdfs-HA-Log	JournalNode	JournalNode	JournalNode	
hdfs-HA-Controller	ZKFC	ZKFC		
mapreduce-Log			JobHistoryServer	
yarn			ResourceManager	
yarn	NodeManager	NodeManager	NodeManager	
hbase-master	master	backup		zk选举
hbase-regionserver	regionserver	regionserver	regionserver	
hive	metastore-server			
hive	hiveserver2-server			

⑤ 浏览器查看，打开主机（外部）浏览器：输入：
http://amaster.basichadoop.com:50070，可以查看 hadoop 集群信息。

← → 不安全 amaster.basichadoop.com:50070/dfshealth.html#tab-overview

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Overview 'amaster:8020' (active)

Started:	Mon Jul 02 23:09:12 CST 2018
Version:	2.6.5, re8c9fe0b4c252ca2ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-c6c3f840-4adf-4545-bf45-074aa36f9c36
Block Pool ID:	BP-242001306-192.168.159.160-1496053342010

Summary

Security is off.
Safemode is off.
208 files and directories, 90 blocks = 298 total filesystem object(s).
Heap Memory used 40.78 MB of 67.71 MB Heap Memory. Max Heap Memory is 966.69 MB.
Non Heap Memory used 47.03 MB of 47.25 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	105.71 GB
DFS Used:	3.6 MB
Non DFS Used:	18.19 GB
DFS Remaining:	87.52 GB

点击 Live Nodes:三个节点信息都在此:

Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	3 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
amaster.basichadoop.com (192.168.159.160:50010)	2	In Service	35.24 GB	1.2 MB	6.93 GB	28.31 GB	86	1.2 MB (0%)	0	2.6.5
aslaver2.basichadoop.com (192.168.159.162:50010)	0	In Service	35.24 GB	1.2 MB	5.61 GB	29.62 GB	86	1.2 MB (0%)	0	2.6.5
aslaver1.basichadoop.com (192.168.159.161:50010)	1	In Service	35.24 GB	1.2 MB	5.64 GB	29.59 GB	86	1.2 MB (0%)	0	2.6.5

Decomissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction
------	--------------	-------------------------	------------------------------	--

Hadoop, 2016.

Legacy UI

■ 输入: <http://aslaver1.basichadoop.com:50070>

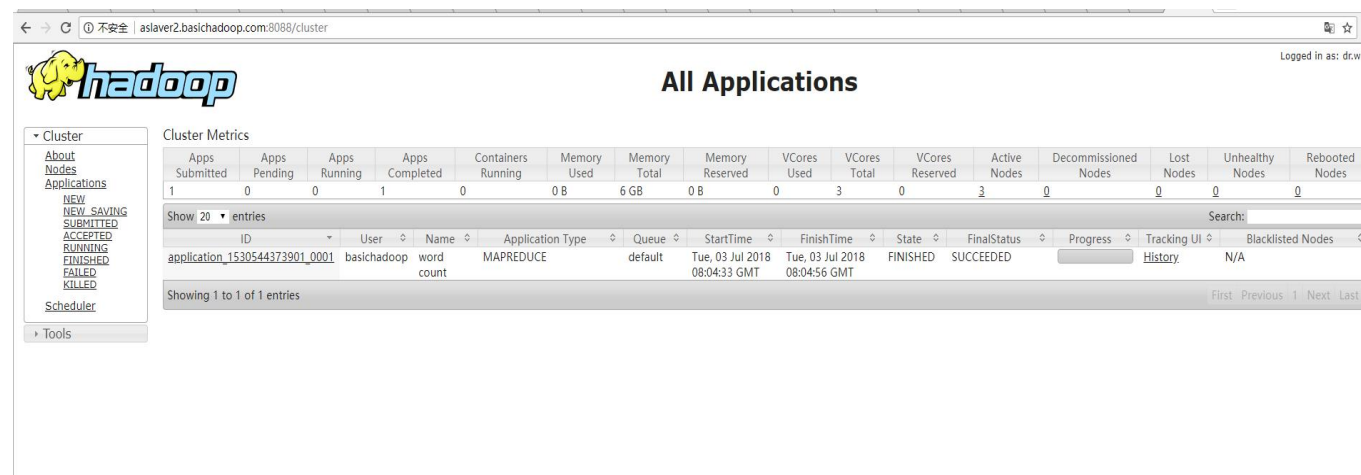
表示两个 namenode
amaster 启动着 (active)
aslaver1 等待状态 (standby)
一旦 amaster 出现故障, aslaver1 自动起来。

Overview 'aslaver1:8020' (standby)

Started:	Mon Jul 02 23:09:24 CST 2018
Version:	2.6.5, re8c9fe0b4c252caf2ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-c6c3f840-4adf-4545-bf45-074aa36f9c36
Block Pool ID:	BP-242001306-192.168.159.160-1496053342010

Summary

■ 查看 aslaver2 节点: yarn 资源管理, 监控 task



Cluster Metrics

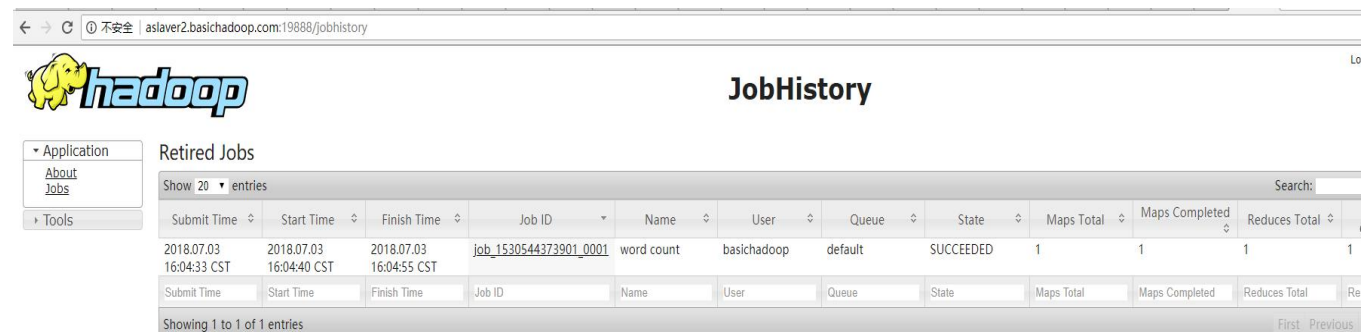
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	6 GB	0 B	0	3	0	2	0	0	0	0

Show: 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI	Blacklisted Nodes
application_1530544373901_0001	basichadoop	word count	MAPREDUCE	default	Tue, 03 Jul 2018 08:04:33 GMT	Tue, 03 Jul 2018 08:04:56 GMT	FINISHED	SUCCEEDED		History	N/A

Showing 1 to 1 of 1 entries

■ yarn 的 task 历史: <http://aslaver2.basichadoop.com:19888>



Retired Jobs

Show: 20 entries

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total
2018.07.03 16:04:33 CST	2018.07.03 16:04:40 CST	2018.07.03 16:04:55 CST	job_1530544373901_0001	word count	basichadoop	default	SUCCEEDED	1	1	1

Showing 1 to 1 of 1 entries

三、wordcount 测试

① 进入 hadoop 目录下：

```
$ cd /opt/modules/hadoop-2.6.5
```

② 与伪分布式思路一致，同样按照创建输入文件的文件夹，放入待处理的输入文件，执行运行命令，输出，查看结果的思路来：

■ 创建输入文件夹：

```
$ bin/hdfs dfs -mkdir -p /data/input
```

```
[basichadoop@amaster hive-1.1.0]$  
[basichadoop@amaster hive-1.1.0]$ cd /opt/modules/hadoop-2.6.5  
[basichadoop@amaster hadoop-2.6.5]$  
[basichadoop@amaster hadoop-2.6.5]$ bin/hdfs dfs -mkdir -p /data/input  
[basichadoop@amaster hadoop-2.6.5]$
```

■ 放入待处理文件：

```
$ bin/hdfs dfs -put README.txt /data/input
```

```
bash: /data/input: No such file or directory  
[basichadoop@amaster hadoop-2.6.5]$ bin/hdfs dfs -put README.txt /data/input  
[basichadoop@amaster hadoop-2.6.5]$
```

■ 查看输入文件夹：

```
bin/hdfs dfs -ls /data/input
```

```
[basichadoop@amaster hadoop-2.6.5]$  
[basichadoop@amaster hadoop-2.6.5]$ bin/hdfs dfs -ls /data/input  
Found 1 items  
-rw-r--r--  3 basichadoop supergroup      1366 2018-07-03 16:01 /data/input/README.txt  
[basichadoop@amaster hadoop-2.6.5]$  
[basichadoop@amaster hadoop-2.6.5]$
```

■ 在 /share/hadoop/mapreduce 文件夹下有测试用例用的 jar 包：
hadoop-mapreduce-examples-2.6.5.jar:

```
[basichadoop@amaster mapreduce]$  
[basichadoop@amaster mapreduce]$ ls  
hadoop-mapreduce-client-app-2.6.5.jar      hadoop-mapreduce-client-jobclient-2.6.5-tests.jar  
hadoop-mapreduce-client-common-2.6.5.jar   hadoop-mapreduce-client-shuffle-2.6.5.jar  
hadoop-mapreduce-client-core-2.6.5.jar     hadoop-mapreduce-examples-2.6.5.jar  
hadoop-mapreduce-client-hs-2.6.5.jar       lib  
hadoop-mapreduce-client-hs-plugins-2.6.5.jar lib-examples  
hadoop-mapreduce-client-jobclient-2.6.5.jar sources  
[basichadoop@amaster mapreduce]$
```

■ 我们输入指令进行运算：

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar  
wordcount /data/input /data/output/result
```

```

[basichadoop@master ~]$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.5.jar wordcount /data/input /data/output/result
18/07/03 16:04:32 INFO client.RMPProxy: Connecting to ResourceManager at aslaver2/192.168.159.162:8032
18/07/03 16:04:32 INFO mapreduce.JobSubmissionFiles: Permissions on staging directory /tmp/hadoop-yarn/staging/basichadoop/.stage-1530544373901 are incorrect: rwxrwxrwx. Fixing permissions to correct value rwx-----
18/07/03 16:04:33 INFO input.FileInputFormat: Total input paths to process : 1
18/07/03 16:04:33 INFO mapreduce.JobSubmitter: number of splits:1
18/07/03 16:04:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1530544373901_0001
18/07/03 16:04:33 INFO impl.YarnClientImpl: Submitted application application_1530544373901_0001
18/07/03 16:04:33 INFO mapreduce.Job: The url to track the job: http://aslaver2:8088/proxy/application_1530544373901_0001/
18/07/03 16:04:33 INFO mapreduce.Job: Running job: job_1530544373901_0001
18/07/03 16:04:42 INFO mapreduce.Job: Job job_1530544373901_0001 running in uber mode : false
18/07/03 16:04:42 INFO mapreduce.Job: map 0% reduce 0%
18/07/03 16:04:51 INFO mapreduce.Job: map 100% reduce 0%
18/07/03 16:04:57 INFO mapreduce.Job: map 100% reduce 100%
18/07/03 16:04:58 INFO mapreduce.Job: Job job_1530544373901_0001 completed successfully
18/07/03 16:04:58 INFO mapreduce.Job: Counters: 49
File System Counters
  FILE: Number of bytes read=1836
  FILE: Number of bytes written=1306
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1468
  HDFS: Number of bytes written=1306
  HDFS: Number of read operations=6
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Data-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=6501
  Total time spent by all reduces in occupied slots (ms)=3171
  Total time spent by all map tasks (ms)=6501
  Total time spent by all reduce tasks (ms)=3171
  Total vcore-milliseconds taken by all map tasks=6501
  Total vcore-milliseconds taken by all reduce tasks=3171
  Total megabyte-milliseconds taken by all map tasks=6657024
  Total megabyte-milliseconds taken by all reduce tasks=3247104
Map-Reduce Framework
  Total megabyte-milliseconds taken by all reduce tasks=3247104
Map-Reduce Framework
  Map input records=31
  Map output records=179
  Map output bytes=2055
  Map output materialized bytes=1836
  Input split bytes=102
  Combine input records=179
  Combine output records=131
  Reduce input groups=131
  Reduce shuffle bytes=1836
  Reduce input records=131
  Reduce output records=131
  Spilled Records=262
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=83
  CPU time spent (ms)=1510
  Physical memory (bytes) snapshot=329129984
  Virtual memory (bytes) snapshot=1687105536
  Total committed heap usage (bytes)=196153344
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=1366
File Output Format Counters
  Bytes Written=1306

```

③ 查看结果:

■ \$ bin/hdfs dfs -ls /data/output

```
/data/output: unknown command
[basichadoop@amaster hadoop-2.6.5]$ bin/hdfs dfs -ls /data/output
Found 1 items
drwxr-xr-x - basichadoop supergroup 0 2018-07-03 16:04 /data/output/result
[basichadoop@amaster hadoop-2.6.5]$ ■
```

■ \$ bin/hdfs dfs -cat /data/output/result/part-r-00000

```
[basichadoop@amaster hadoop-2.6.5]$ bin/hdfs dfs -cat /data/output/result/part-r-00000
(BIS), 1
(ECCN) 1
(TSU) 1
(see 1
5D002.C.1, 1
740.13) 1
<http://www.wassenaar.org/> 1
Administration 1
Apache 1
BEFORE 1
BIS 1
Bureau 1
Commerce, 1
Commodity 1
Control 1
Core 1
Department 1
ENC 1
Exception 1
Export 2
For 1
Foundation 1
Government 1
Hadoop 1
Hadoop, 1
Industry 1
Jetty 1
License 1
Number 1
Regulations, 1
SSL 1
Section 1
Security 1
See 1
Software 2
Technology 1
The 4
This 1
U.S. 1
Unrestricted 1
about 1
algorithms. 1
```


四、日志查看

① 此时在本地浏览器输入：<http://amaster.basichadoop.com:50070/>

The screenshot shows the Hadoop Overview page for the 'amaster:8020' node. The page has a green header with navigation tabs: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Overview 'amaster:8020' (active)'. Below the title is a table with the following information:

Started:	Mon Jul 02 23:09:12 CST 2018
Version:	2.6.5, re8c9fe0b4c252ca12ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-c6c3f840-4adf-4545-bf45-074aa36f9c36
Block Pool ID:	BP-242001306-192.168.159.160-1496053342010

Below the table is a 'Summary' section. It contains the following text:

Security is off.
Safemode is off.
212 files and directories, 94 blocks = 306 total filesystem object(s).
Heap Memory used 37.46 MB of 67.71 MB Heap Memory. Max Heap Memory is 966.69 MB.
Non Heap Memory used 46.85 MB of 47.06 MB Committed Non Heap Memory. Max Non Heap Memory is 130 MB.

Below the text is a table with the following information:

Configured Capacity:	105.71 GB
DFS Used:	3.69 MB
Non DFS Used:	18.18 GB
DFS Remaining:	87.53 GB
DFS Used%:	0%
DFS Remaining%:	82.8%
Block Pool Used:	3.69 MB
Block Pool Used%:	no

② 查看刚刚 wordcount 测试的日志：

The screenshot shows the Hadoop Browse Directory page. The page has a green header with navigation tabs: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Browse Directory'. Below the title is a search bar with the text '/' and a 'Go!' button. Below the search bar is a table with the following information:

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	basichadoop	supergroup	0 B	0	0 B	data
drwxrwxrwx	basichadoop	supergroup	0 B	0	0 B	datas
drwxrwxrwx	basichadoop	supergroup	0 B	0	0 B	hbase
drwxrwxrwx	basichadoop	supergroup	0 B	0	0 B	test
drwxrwxrwx	basichadoop	supergroup	0 B	0	0 B	tmp
drwxrwxrwx	basichadoop	supergroup	0 B	0	0 B	user

Below the table is the text 'Hadoop, 2016.'

- 点击”data”,因为我们刚刚新建的输入输出文件都在 data 文件夹下面。input 和 output 文件都在该目录下面

amaster.basichadoop.com:50070/explorer.html#/data

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Browse Directory

/dataGo!

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	basichadoop	supergroup	0 B	0	0 B	input
drwxr-xr-x	basichadoop	supergroup	0 B	0	0 B	output

Hadoop, 2016.

- 我们尝试点击”input”:发现我们刚刚提交的输入文件在这，继续点击，发现可以下载。因此这就涉及到了权限问题了，在企业级项目中，日志的查看只有部分权限。

Browse Directory

/data/inputGo!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	basichadoop	supergroup	1.33 KB	3	128 MB	README.txt

Hadoop, 2016.

File information - README.txt

Download

Block information -- Block 0

Block ID: 1073742050
Block Pool ID: BP-242001306-192.168.159.160-1496053342010
Generation Stamp: 1227
Size: 1366
Availability:

- aslaver2.basichadoop.com
- aslaver1.basichadoop.com
- amaster.basichadoop.com

Close

- 点击 output,发现输出结果：继续点击：

Browse Directory

/data/output						Go!
Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	basichadoop	supergroup	0 B	0	0 B	result

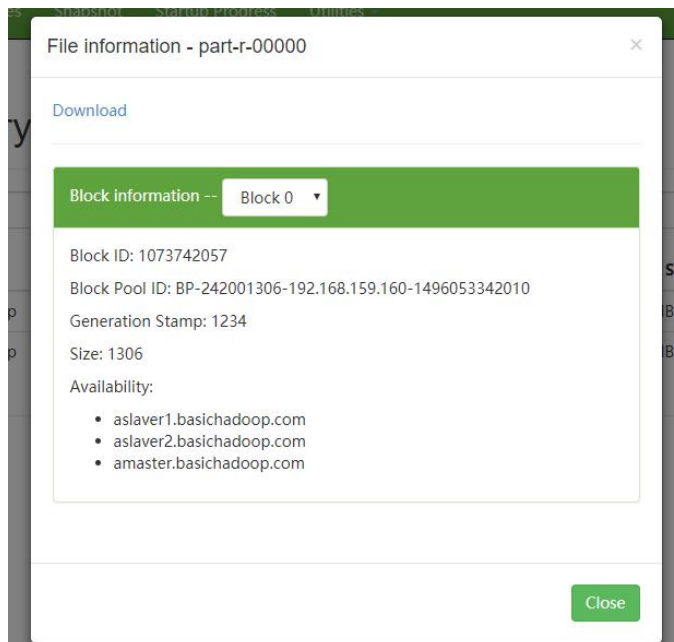
Hadoop, 2016.

Browse Directory

/data/output/result						Go!
Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	basichadoop	supergroup	0 B	3	128 MB	_SUCCESS
-rw-r--r--	basichadoop	supergroup	1.28 KB	3	128 MB	part-r-00000

Hadoop, 2016.

- 点击 [part-r-00000](#),发现结果可下载:



- 我们下载一下,打开: 和我们在命令行里所见一致:

```
File Edit Selection Find View Goto Tools Project
Readme mapper.py reducer.py
15 Control 1
16 Core 1
17 Department 1
18 ENC 1
19 Exception 1
20 Export 2
21 For 1
22 Foundation 1
23 Government 1
24 Hadoop 1
25 Hadoop, 1
26 Industry 1
27 Jetty 1
28 License 1
29 Number 1
30 Regulations, 1
31 SSL 1
32 Section 1
33 Security 1
34 See 1
35 Software 2
36 Technology 1
37 The 4
38 This 1
39 U.S. 1
```