Q0: 基础实验环境说明：

(1) 操作系统:CentOS6.5 VMWare虚拟机环境

```
➜  examUse  tree ./
./
├── elephaht_examhos
|   ├── Enmoedu_Hadoop_CentOS_6.5_CDH_5.6.vmdk
|   └── Enmoedu_Hadoop_CentOS_6.5_CDH_5.6.vmx
└── monkey_exam
    ├── Enmoedu_Hadoop_CentOS_6.5_CDH_5.6.vmdk
    └── Enmoedu_Hadoop_CentOS_6.5_CDH_5.6.vmx
```

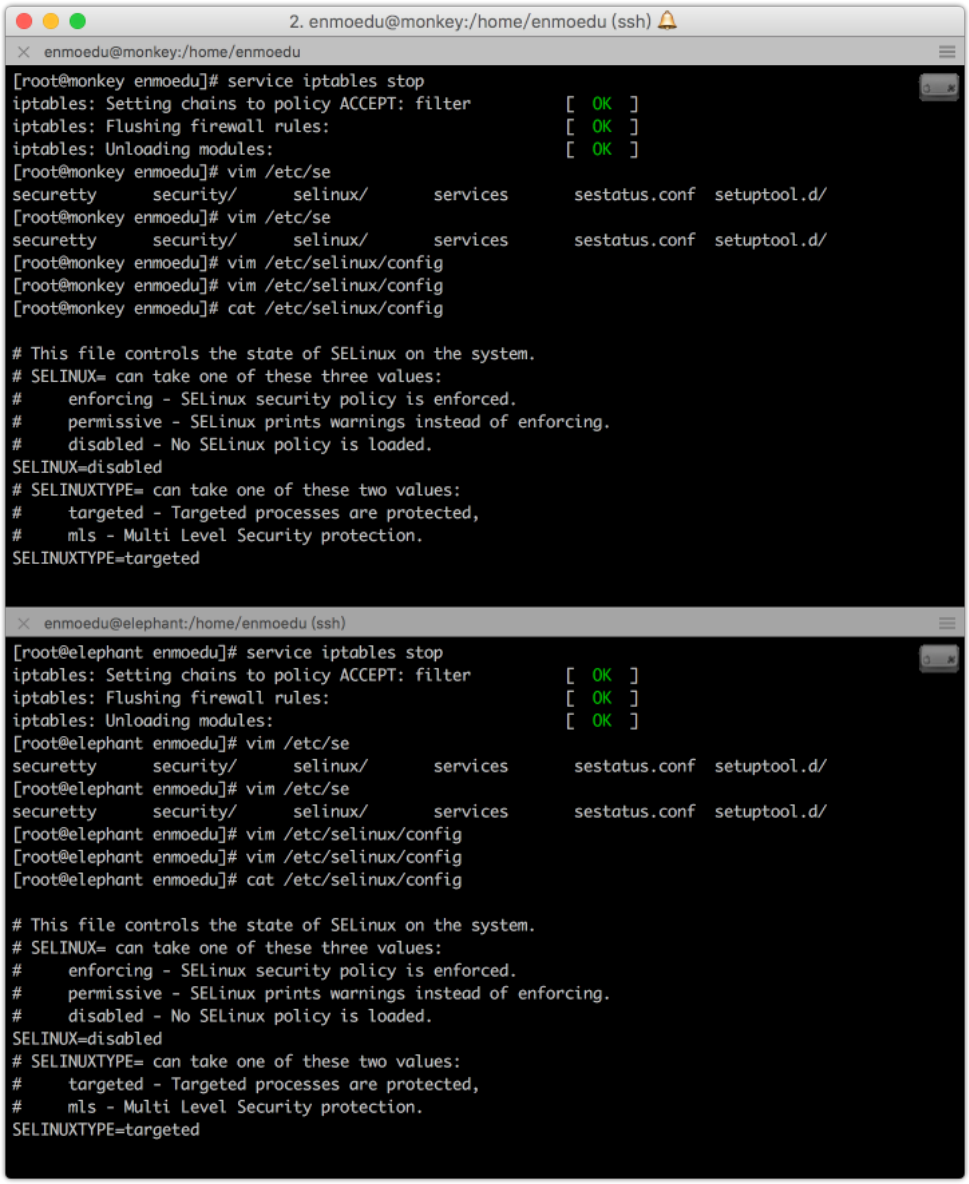
(2) 局域网环境，网卡设置为桥接模式，主从两节点(elephant/monkey)

```
[enmoedu@bogon ~]$ hostname
elephant
[enmoedu@bogon ~]$ ifconfig | grep 10
          inet addr:10.8.2.157  Bcast:10.8.2.255  Mask:255.255.255.0
          collisions:0 txqueuelen:1000
          RX packets:10 errors:0 dropped:0 overruns:0 frame:0
          TX packets:10 errors:0 dropped:0 overruns:0 carrier:0

[enmoedu@bogon ~]$ hostname
monkey
[enmoedu@bogon ~]$ ifconfig | grep 10
          inet addr:10.8.2.168  Bcast:10.8.2.255  Mask:255.255.255.0
          collisions:0 txqueuelen:1000
```

(3) 关闭主从节点的防火墙和 `SELinux`

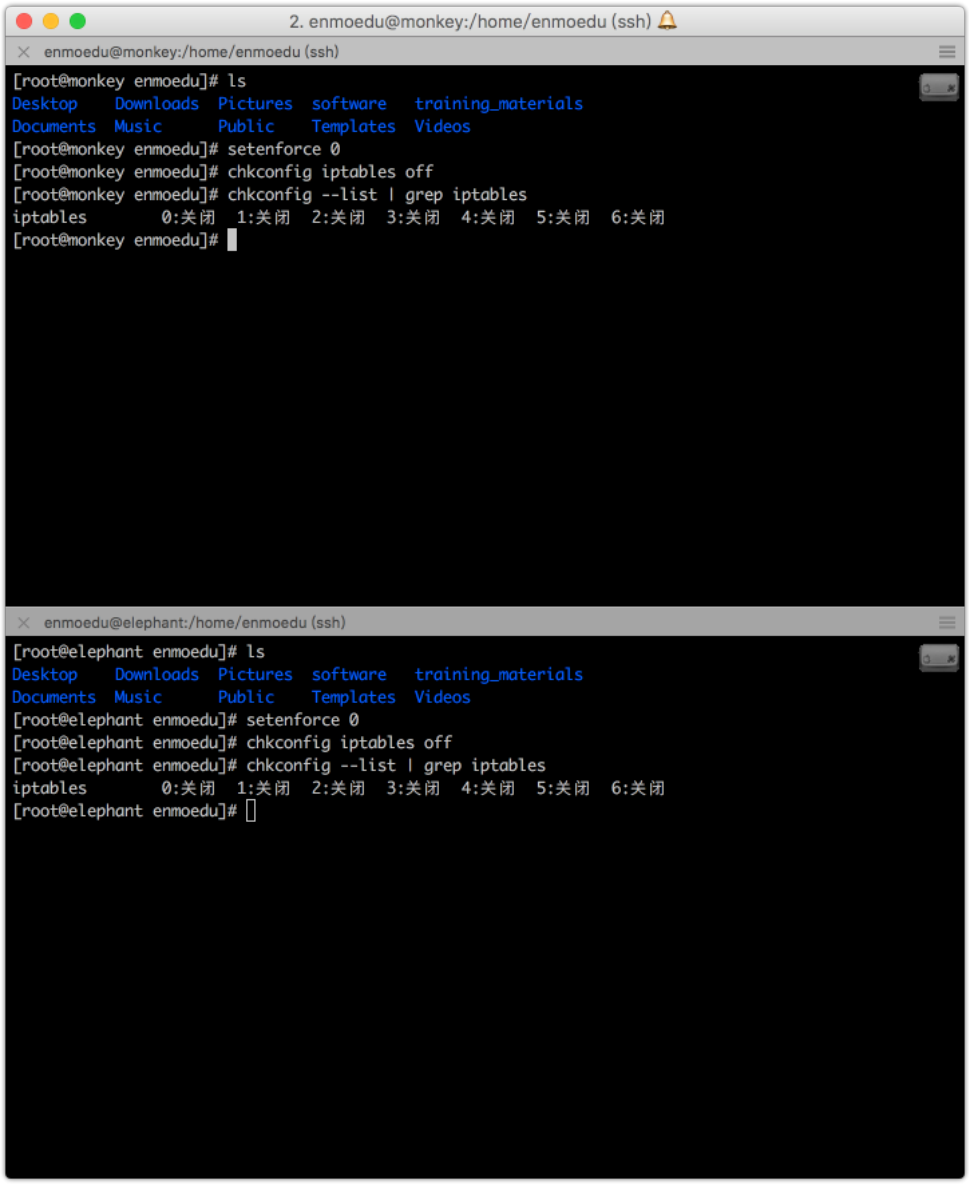
以下所有操作默认使用 `su` 切换到 `root` 用户

```
# change to root user
[root@elephant enmoedu]# service iptables stop
iptables: Setting chains to policy ACCEPT: filter          [  OK  ]
iptables: Flushing firewall rules:                         [  OK  ]
iptables: Unloading modules:                               [  OK  ]

# edit selinux configure file
[root@elephant enmoedu]# vim /etc/selinux/config

# change `enforcing` to `disabled`
SELINUX=disabled
```



设置关闭防火墙和SELinux的行为永久生效：

```
# shutdown selinux forever
[root@monkey enmoedu]# setenforce 0

# deny iptables restart when system reboot
[root@monkey enmoedu]# chkconfig iptables off

# check iptables setting
[root@monkey enmoedu]# chkconfig --list | grep iptables
iptables        0:关闭    1:关闭    2:关闭    3:关闭    4:关闭    5:关闭    6:关闭
[root@monkey enmoedu]#
```



Q1. 查看Linux操作系统kernel 版本参考命令

- `uname -a` 显示系统信息一般习惯使用 `uname` 指令，`2.6.32-431.el6.x86_64` 便是系统内核版本号

```
[enmoedu@monkey ~]$ uname -a
Linux monkey 2.6.32-431.el6.x86_64 #1 SMP Fri Nov 22 03:15:09 UTC 2013 x86_64 x86_64 x86_64 GNU/Linux
```

- `/proc/version` 文件里也可以查阅到内核版本号

```
[enmoedu@monkey ~]$ cat /proc/version
Linux version 2.6.32-431.el6.x86_64 (mockbuild@c6b8.bsys.dev.centos.org) (gcc version 4.4.7 20120313 (Red Hat 4.4.7-4) (GCC) ) #1 SMP Fri N
ov 22 03:15:09 UTC 2013
```

Q2. 在 Linux 操作系统上搭建NTP 时间服务器。能够利用 同步主机时间。

(1) 检查主从节点时间

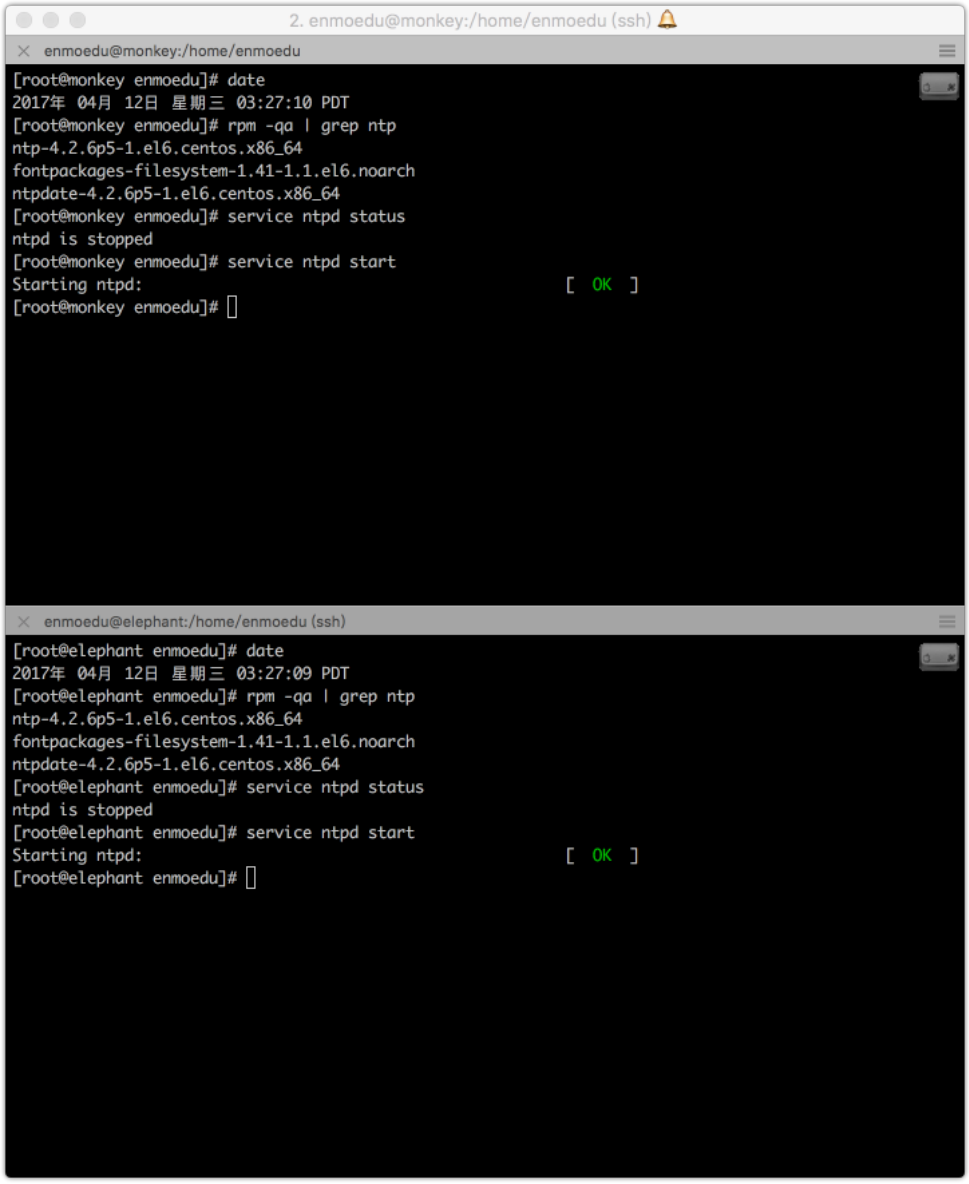
```
# date
2017年 04月 12日 星期三 03:27:10 PDT
```

(2) 检查主从节点是否安装 `NTP` 包

```
[root@monkey enmoedu]# rpm -qa | grep ntp
ntp-4.2.6p5-1.el6.centos.x86_64
fontpackages-filesystem-1.41-1.1.el6.noarch
ntpdate-4.2.6p5-1.el6.centos.x86_64

# 如果没有发现ntp包，可以尝试使用默认的yum源安装(推荐使用163源)
# yum install ntp
```

(3) 检查主从节点 `NTP` 服务状态，并启动

```
[root@monkey enmoedu]# service ntpd status
ntpd is stopped
[root@monkey enmoedu]# service ntpd start
Starting ntpd:                                             [  OK  ]
```



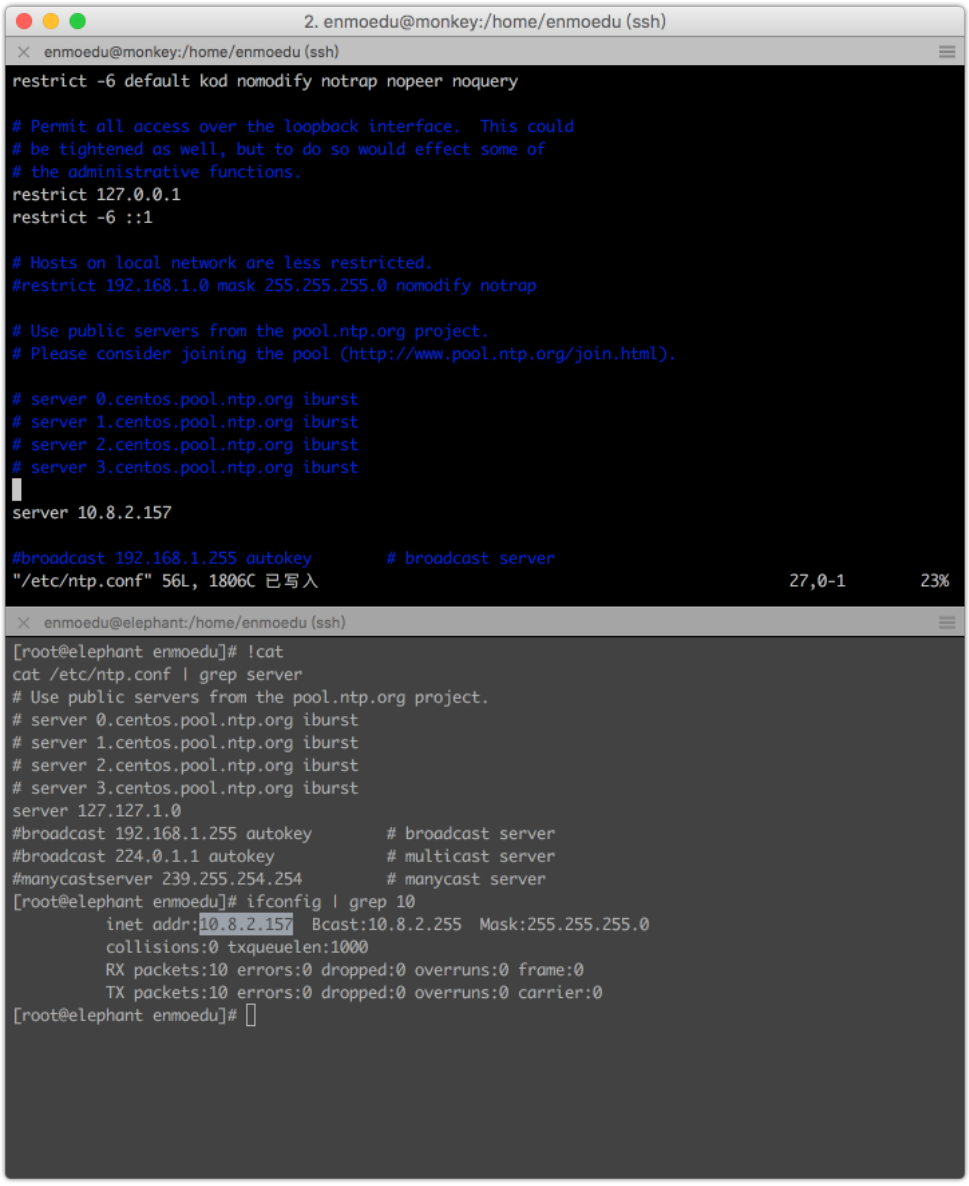
(4) 设计 `elephant` 为主节点，作为时间服务器， `monkey` 从主节点获取时间

将 `elephant` 作为ntp server，配置 `/etc/ntp.conf` 文件，将 `server` 设置为 `127.127.1.0` 本地物理时钟地址

```
[root@elephant enmoedu]# vim /etc/ntp.conf
server 127.127.1.0
```

相应的修改 `monkey` 的 `ntp.conf` ，将 `server` 设置成 `elephant` 的IP地址:

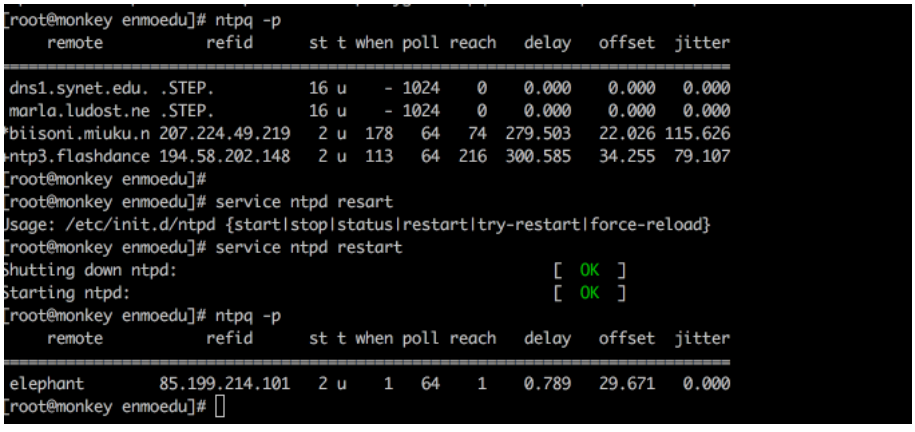
```
[root@monkey enmoedu]# vim /etc/ntp.conf
server 10.8.2.157 ## 主节点的IP地址
```



(5) 重启一下ntpd服务，从节点使用 `ntpd -p` 检查是否从主节点同步时间

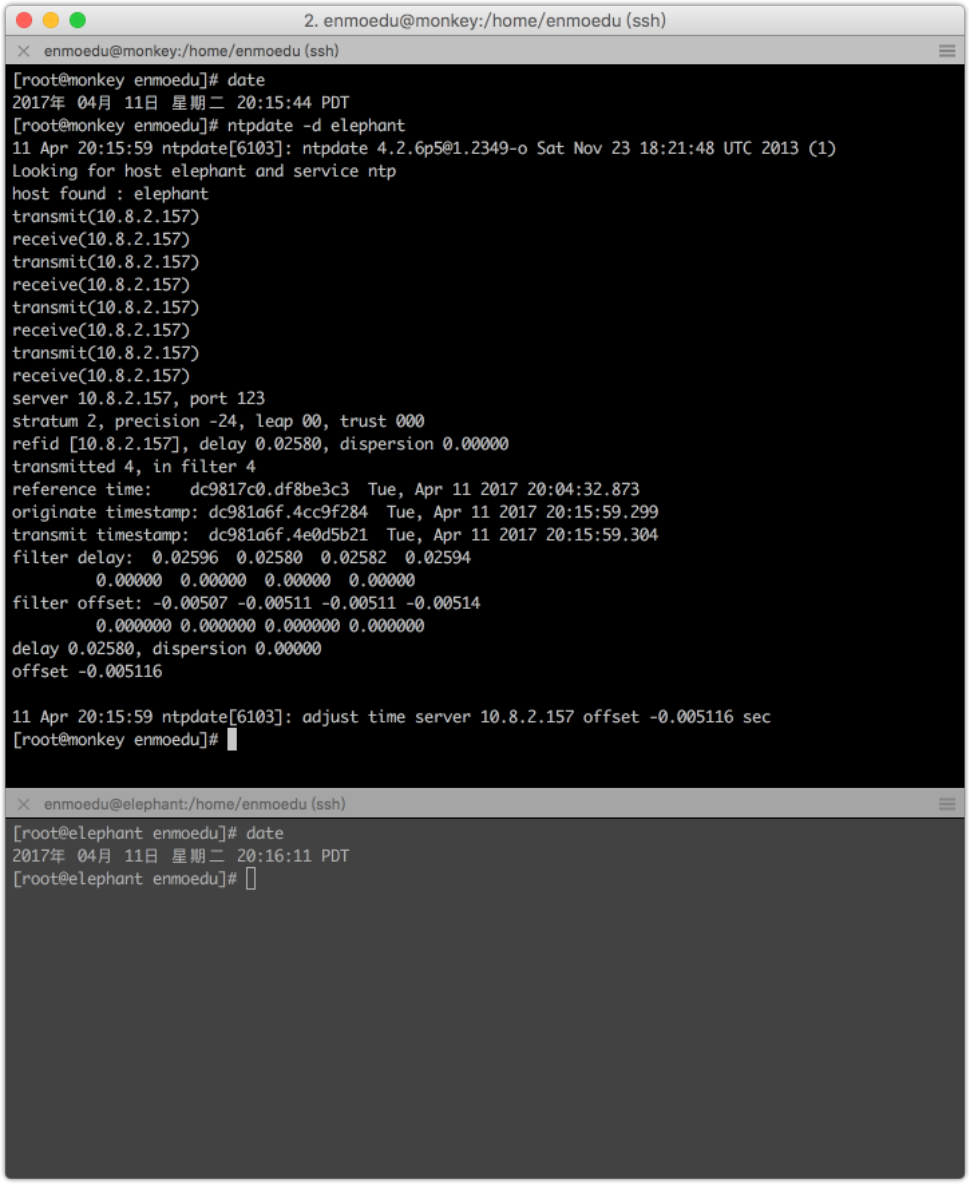
`remote` 显示成为主节点hostname，说明设置生效

```
[root@monkey enmoedu]# service ntpd restart
Shutting down ntpd:                                        [  OK  ]
Starting ntpd:                                             [  OK  ]
[root@monkey enmoedu]# ntpq -p
     remote           refid      st t when poll reach   delay   offset  jitter
==============================================================================
 elephant        85.199.214.101   2 u    1   64    1    0.789   29.671   0.000
```



(6) 使用 `ntpdate -d elephant` 同步主从节点时间

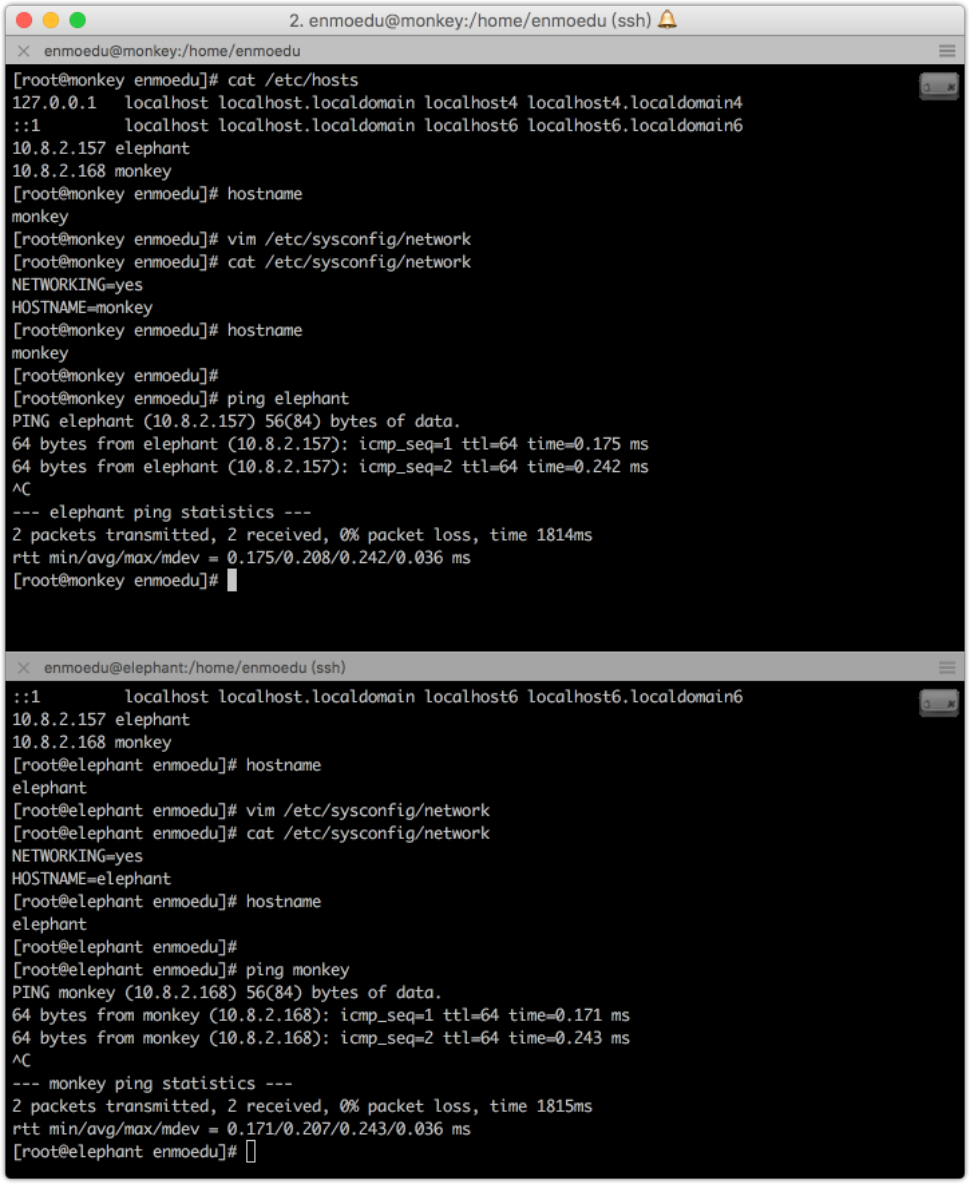
```
[root@monkey enmoedu]# ntpdate -d elephant
```



Q3. 在 Linux 操作系统上搭建安装服务器，能够利用 yum install命令从安装服务器上下载对应的软件包。

(0) 检查并配置主从hostname，检查网络是否连通

```
## 1. check hosts configure
[root@monkey enmoedu]# cat /etc/hosts
127.0.0.1   localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6
10.8.2.157 elephant
10.8.2.168 monkey

## 2. modify /etc/sysconfig/network configure change `HOSTNAME` fields
[root@monkey enmoedu]# cat /etc/sysconfig/network
NETWORKING=yes
HOSTNAME=monkey

[root@elephant enmoedu]# cat /etc/sysconfig/network
NETWORKING=yes
HOSTNAME=elephant

## 3. re-declare hostname (I've set it before)
[root@monkey enmoedu]# hostname monkey
monkey
[root@elephant enmoedu]# hostname elephant
elephant


## if reset network that need to restart network service
# service network restart
```



(1) 主节点检查并启动 `httpd` 服务

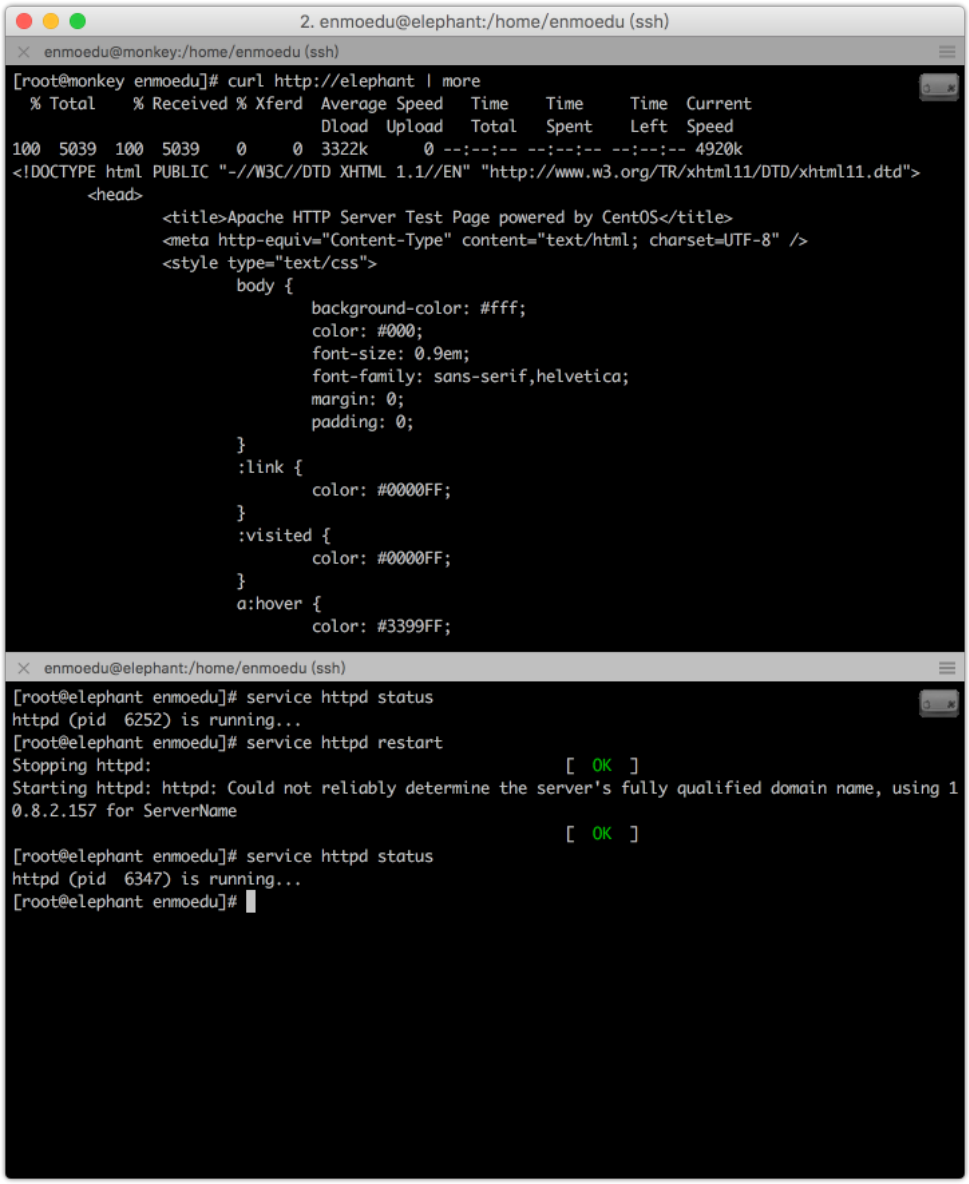
检查或安装 `httpd` 服务

```
## check httpd package is install or not
[root@elephant enmoedu]# rpm -qa | grep httpd
httpd-2.2.15-29.el6.centos.x86_64
httpd-tools-2.2.15-29.el6.centos.x86_64

## if not install use `yum` default soruce (recommand 163 mirrors)
[root@elephant enmoedu]# yum install httpd
```

启动 `httpd`

```
[root@elephant enmoedu]# service httpd status
httpd is stopped
[root@elephant enmoedu]# service httpd start
Starting httpd: httpd: Could not reliably determine the server's fully qualified domain name, using 10.8.2.157 for ServerName
                                                           [  OK  ]

设置开启自启动
[root@elephant enmoedu]# chkconfig httpd on
```

启动后从节点检查80端口是否可访问apache欢迎页



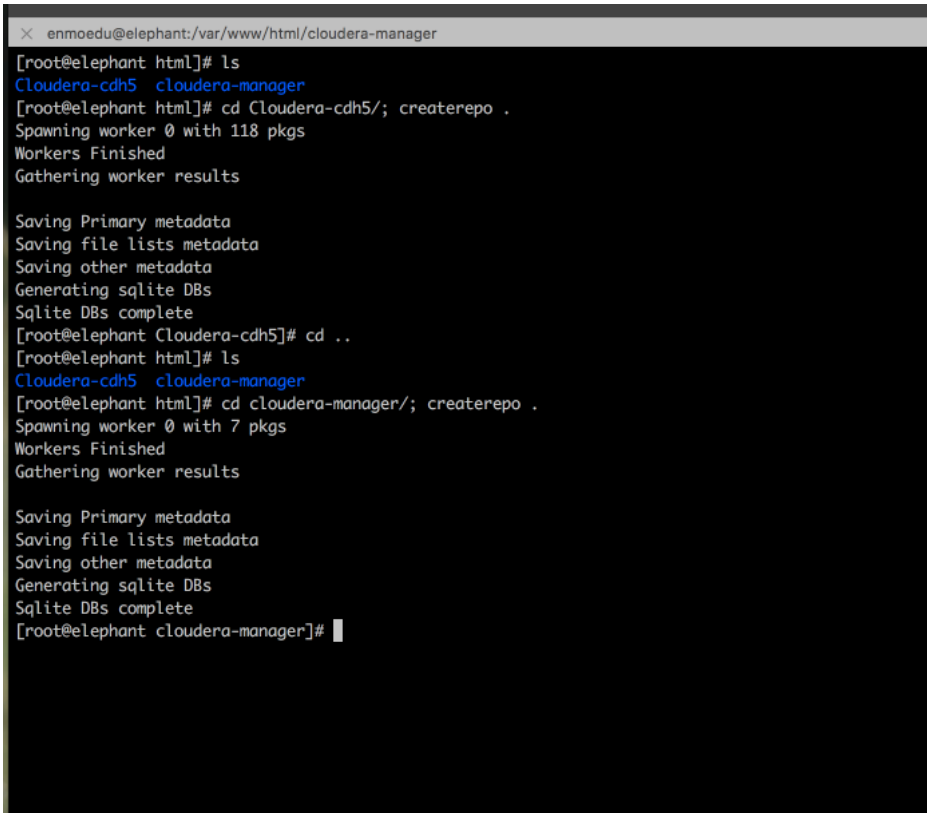
(3) 准备测试 `yum server` 用的软件包，以 `CDH` 套装为例,创建成 `repo`

将准备好的 `CDH` 包组挪到 `httpd` 默认的 `web` 目录:

```
[root@elephant software]# ls
Cloudera-cdh5  cloudera-manager
[root@elephant software]# mv Cloudera-cdh5/ /var/www/html/
[root@elephant software]# mv cloudera-manager/ /var/www/html/
[root@elephant software]# cd /var/www/html/
[root@elephant html]# pwd
/var/www/html
[root@elephant html]# ls
Cloudera-cdh5  cloudera-manager
```

创建 `repo`：

```
[root@elephant html]# pwd
/var/www/html
[root@elephant html]# cd Cloudera-cdh5/; createrepo .

Spawning worker 0 with 118 pkgs
Workers Finished
Gathering worker results

Saving Primary metadata
Saving file lists metadata
Saving other metadata
Generating sqlite DBs
Sqlite DBs complete

[root@elephant html]# pwd
/var/www/html
[root@elephant html]# cd cloudera-manager/; createrepo .
Spawning worker 0 with 7 pkgs
Workers Finished
Gathering worker results

Saving Primary metadata
Saving file lists metadata
Saving other metadata
Generating sqlite DBs
Sqlite DBs complete
```



(4) 创建yum repo config文件，分发到所有节点

yum的repo配置文件存放于 `/etc/yum.repos.d` ,在这里创建 `repo` 文件：

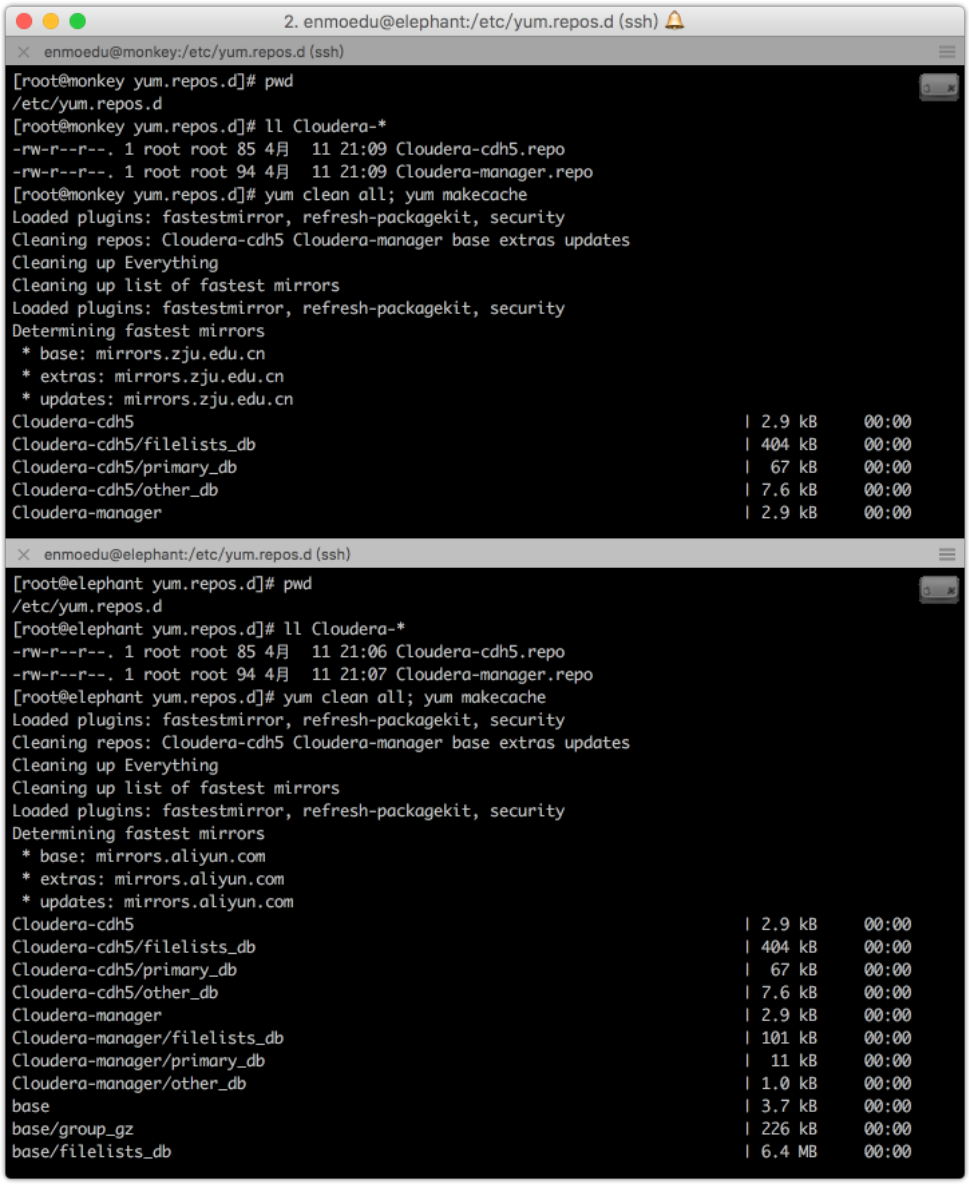
```
[root@elephant yum.repos.d]# pwd
/etc/yum.repos.d
[root@elephant yum.repos.d]# ll Cloudera-*
-rw-r--r--. 1 root root 85 4月  11 21:06 Cloudera-cdh5.repo
-rw-r--r--. 1 root root 94 4月  11 21:07 Cloudera-manager.repo
[root@elephant yum.repos.d]# cat Cloudera-*
[Cloudera-cdh5]
name=Cloudera-cdh5
baseurl=http://elephant/Cloudera-cdh5/
gpgcheck=0
[Cloudera-manager]
name=Cloudera-manager
baseurl=http://elephant/cloudera-manager/
gpgcheck=0
[root@elephant yum.repos.d]#
```

将创建的两个 `Cloudera Repo` 文件scp到从节点

```
[root@elephant yum.repos.d]# scp Cloudera-* root@monkey:/etc/yum.repos.d/
The authenticity of host 'monkey (10.8.2.168)' can't be established.
RSA key fingerprint is 46:69:2d:47:1f:7d:69:2b:a4:c9:93:7d:09:25:c3:62.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'monkey,10.8.2.168' (RSA) to the list of known hosts.
root@monkey's password:
Cloudera-cdh5.repo          100%   85     0.1KB/s   00:00
Cloudera-manager.repo       100%   94     0.1KB/s   00:00
[root@elephant yum.repos.d]#
```

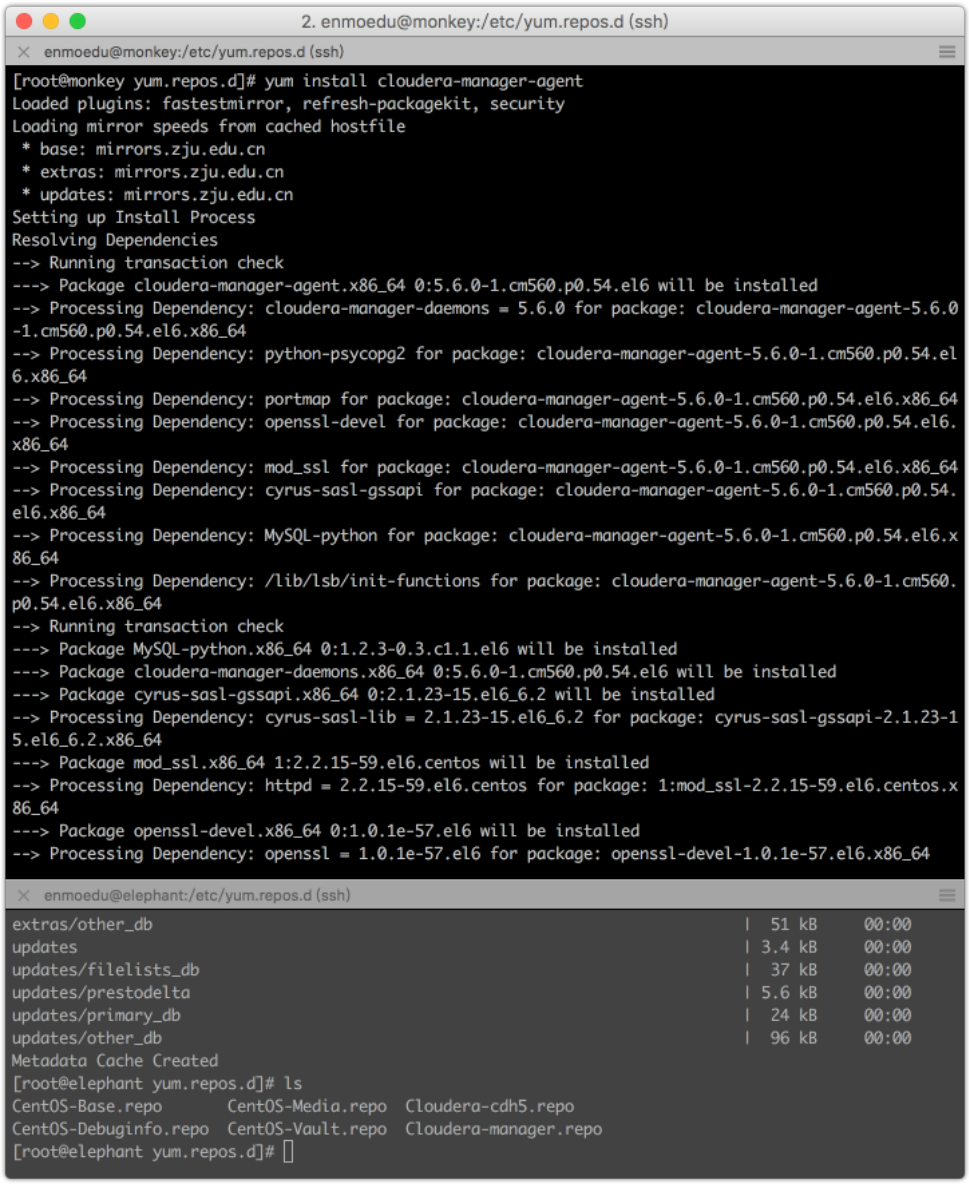
(5) 所有节点同时更新 `yum cache` 缓存

```
[root@elephant yum.repos.d]# pwd
/etc/yum.repos.d
[root@elephant yum.repos.d]# ll Cloudera-*
-rw-r--r--. 1 root root 85 4月  11 21:06 Cloudera-cdh5.repo
-rw-r--r--. 1 root root 94 4月  11 21:07 Cloudera-manager.repo
[root@elephant yum.repos.d]# yum clean all; yum makecache
```



(6) 从节点中，下载一个主节点有的软件包

```
[root@monkey yum.repos.d]# yum install cloudera-manager-agent
```



Q4. 学习搭建时间服务器 NTP 的目的是什么?

单个节点默认是从本地物理时钟时间读取时间信息（`BIOS`时间），与标准时间有一定时间差距，是不准的。当希望校对时间获取最标准时间，应该从网络同步获取**标准时间**，但即使从网络同步也需要预想到网络传输消耗的时间。

承上所述，多个节点在集群内工作时，是不一定能保证每台都是标准时间，即使都从网络获取标准时间，我们也不一定能保证各个节点的网络环境稳定。 如果希望多个节点之间能保证时间同步，那么这个时候的解决办法就是，让集群中多个节点从某个特定的节点获取时间，保证整个集群的时间同步。

作为NTP服务器的节点，负责统一所有节点时间，保障集群节点之间的协作不会受到时间不统一问题的影响。

Q5. 学习搭建安装服务器的目的是什么?

1. 整个CDH安装所需的包太多，并不适合直接网络传输。搭建安装服务器，有利于加快构建集群的效率，排除网络的干扰因素。
2. 解决包依赖问题。每个软件包依赖繁多，使用`yum`解决依赖的问题。

Q6. Linux 文件系统管理数据的思路是什么？我们学习这个知识点的目的是什么?

1. Linux文件系统，以ext系列为例，以inode为准存储，文件系统分为多块，由superblock为首，然后跟着inode节点位信息，存储inodetable信息，然后跟着各级间接指针，指向文件块。
2. 大数据存储技术构建于Linux文件系统之上，设计思路也与此一脉相承。了解`Linux`文件系统的设计思路，有利于了解大数据存储文件系统的设计思路。

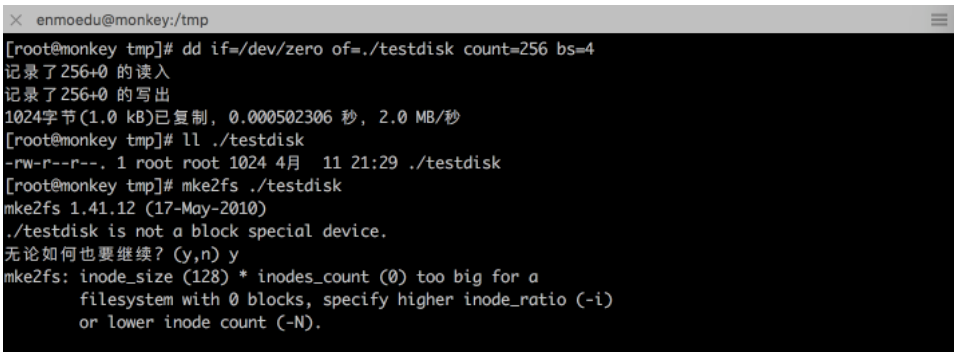
Q7. 利用dd 创建1M文件testdisk，并利用操作系统格式化的命令将testdisk 格式化成ext2文件系统，利用文件系统相关命令，找到inode count 信息（整型数字）

(1) 使用dd创建1M的 `testdisk`

`shell [root@monkey tmp]# dd if=/dev/zero of=./testdisk count=256 bs=4 记录了256+0 的读入 记录了256+0 的写出 1024字节(1.0 kB)已复制，0.00050230`

(2) 使用 `mke2fs` 将 `testdisk` 格式化为 `ext2` 文件系统

```
[root@monkey tmp]# mke2fs ./testdisk
mke2fs 1.41.12 (17-May-2010)
./testdisk is not a block special device.
无论如何也要继续? (y,n) y
mke2fs: inode_size (128) * inodes_count (0) too big for a
    filesystem with 0 blocks, specify higher inode_ratio (-i)
    or lower inode count (-N).
```



(3) 读取 `inode count` 信息

```
[root@elephant tmp]# tune2fs -l ./testfile | grep count
Inode count:              128
Block count:              1024
Reserved block count:     51
Mount count:              0
Maximum mount count:      35
```