# 1、安装虚拟机

## 1.1、准备软件：

1、VMware Workstation 12.zip 虚拟机软件

2、CentOS-6.5-x86\_64-bin-DVD1.iso 64位操作系统

## 1.2、安装虚拟机软件

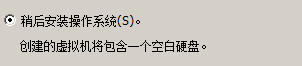
**要想成功安装，务必删除掉已有的虚拟机软件**

## 1.3、安装操作系统

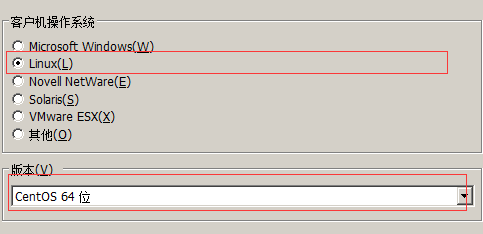
* 运行虚拟机软件，点击软件右上角的文件，选择新建虚拟机，会弹出新建虚拟机向导页面
* 选择典型安装，然后点击下一步



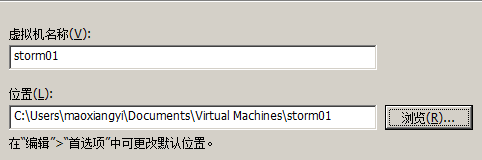
* 选择客户端操作系统，**务必选择“稍后安装操作系统”，然后点击下一步**



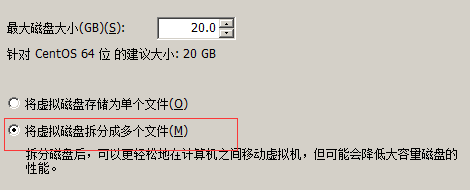
* 选择客户端操作系统的类型为linux、版本为CentOS 64位，然后点击下一步



* 输入虚拟机名称，并选择虚拟机镜像文件安装的目录，然后下一步



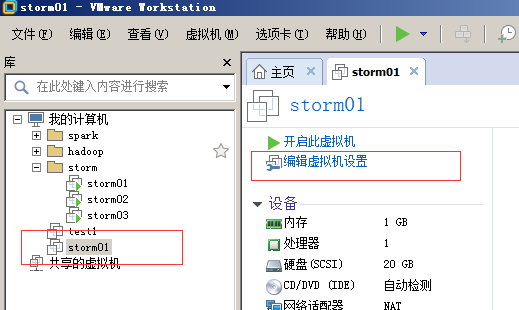
* 选择磁盘大小，并选择镜像文件存储的格式，然后点击下一步



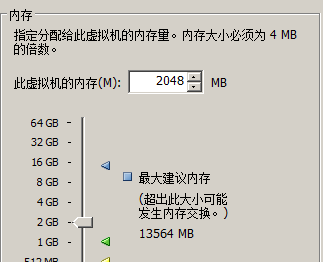
* 直接点击完成



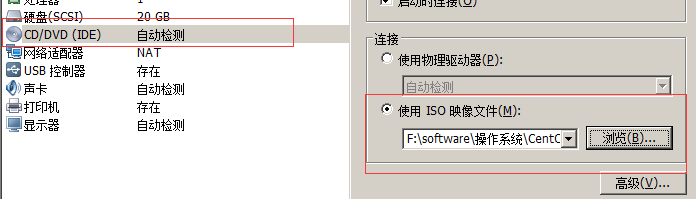
* 完成之后，虚拟机软件的左侧会出现你配置的虚拟机名称，点击**编辑虚拟机设置**



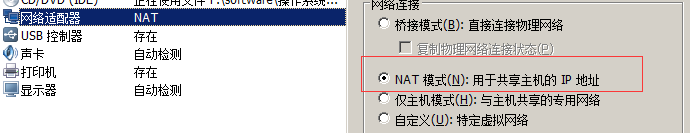
* 根据自己的机器配置信息设置虚拟机内存大小，一般建议给2G以上



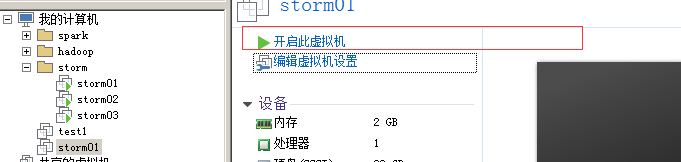
* 重要操作：指定虚拟机启动时的镜像文件



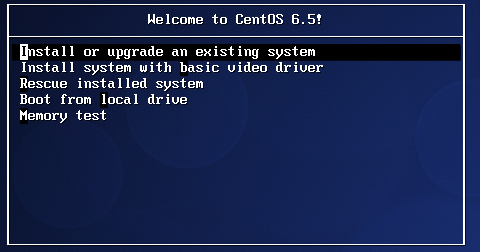
* 重要操作：指定虚拟机网络类型为NAT，如果已经NAT就忽略此步骤



* 配置完成之后，点击确定操作保存配置文件
* 启动虚拟机，选中你的虚拟机名称，在右侧的窗口点击 开启此虚拟机



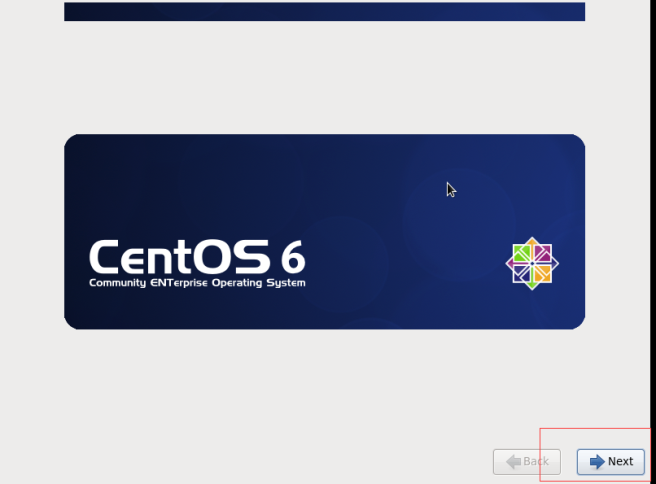
* 直接回车，安装操作系统



* 使用键盘上的Tab按键，选择Skip，然后点击回车



* 然后一路回车，进入以下的页面，并点击Next操作



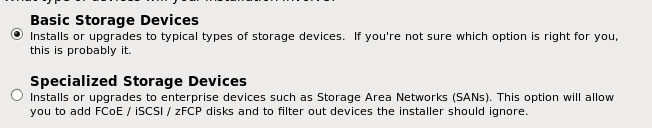
* 选择语言



* 选择键盘



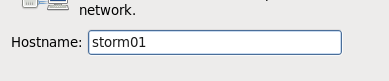
* 选择硬盘类型



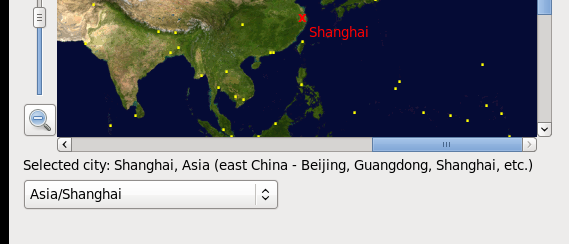
* 选择 yes discard any data 确定当前配置



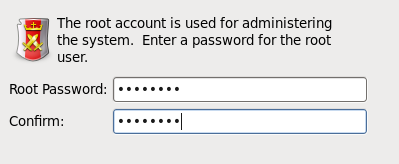
* 输入hostname 主机名称



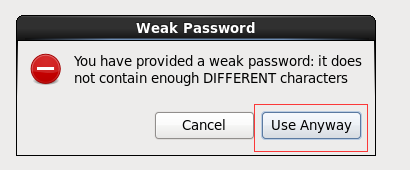
* 选择时区为上海



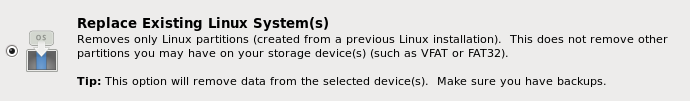
* 输入操作系统的密码为8个1



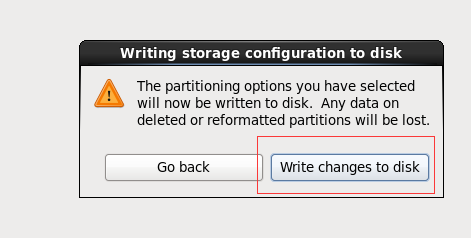
* 输入密码之后，点击下一步，会提示密码过于简单，是否使用，选择使用



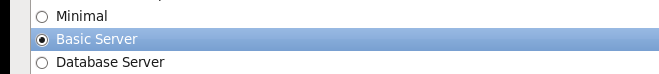
* 选择替换已经存在的系统



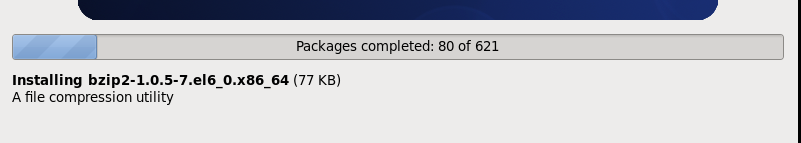
* 将以上配置信息写入到磁盘



* 选择服务器类型为Basic Server



* 服务器会自动安装，安装完成之后点击reboot按钮即可



## 1.4、配置linux的网卡

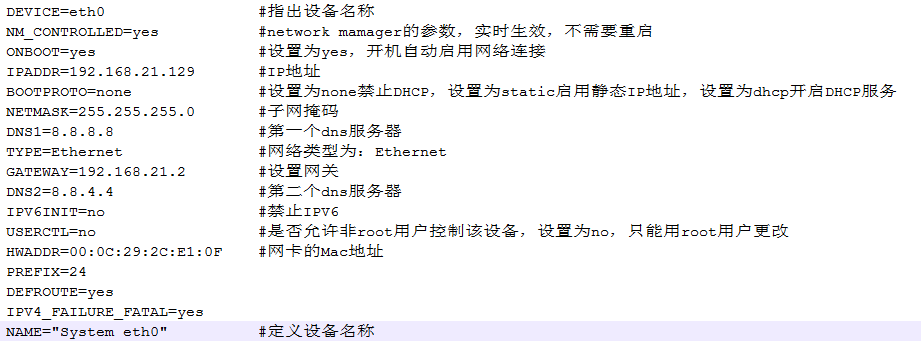
**新安装的操作系统，既是使用了桥接的方式，也不一定能够上网**，因为每个人的网络环境并非一样的。所以如果你在使用ping 命令，可能会碰到一下的错误：

    1）ping域名的时候出现ping:unknown host  xxx.xxx

    2）ping域名的时候出现connect: network is unreachable 的问题

    以上两种问题的出现，一般都是ip地址没有配置好。

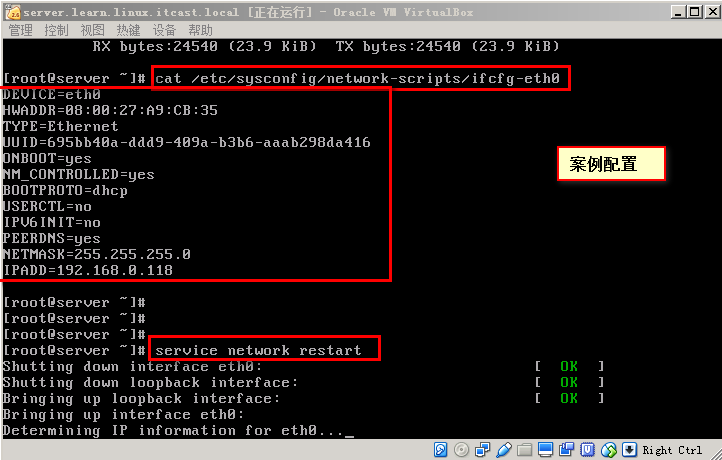
下图的是常用网卡配置的参数



    配置网卡有几种方式，可以自行百度下，本文是修改网卡的配置文件。

    配置文件名称：vi /etc/sysconfig/network-scripts/ifcfg-eth0

**配置完毕之后，重启网卡 service network restart**



## 1.5、准备三个虚拟机搭建Storm集群

**重复步骤1.3和1.4的操作，完成三个虚拟机的安装。**

## 1.6、配置集群Hosts

   所有的虚拟机上都需要配置hosts

         vi  /etc/hosts

            192.168.239.128 storm01 zk01 hadoop01

            192.168.239.129 storm02 zk02 hadoop02

            192.168.239.130 storm03 zk03 hadoop03

## 1.7、服务器间免密码登录[可选]

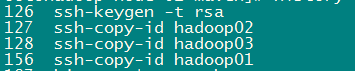
配置免密码登陆，在所有机器上执行以下命令

    ssh-keygen -t rsa （四个回车）

     执行完这个命令后，会生成两个文件id\_rsa（私钥）、id\_rsa.pub（公钥）

     将公钥拷贝到要免密登陆的目标机器上

     ssh-copy-id {其他机器的hostname}



## 1.8、关闭服务器的防火墙

（1） 重启后永久性生效：

开启：chkconfig iptables on

关闭：chkconfig iptables off

（2） 即时生效，重启后失效：

开启：service iptables start

关闭：service iptables stop

## 1.9、创建服务器上的统一工作目录

mkdir /export

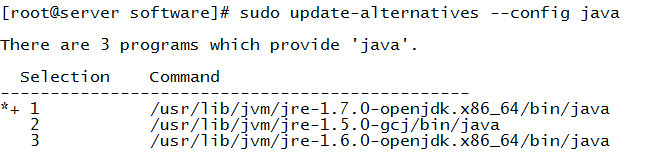
mkdir /export/servers

chmod 755 -R /export

# 2、配置JDK

## 2.1、查看当前系统上是否安装了JDK

    进入一个新的系统，一般来讲是没有安装过JDK的。但是有些linux的版本会默认安装openJDK。open JDK实在oracle公司开源的代码上进行同步开发并发布的版本。如果要查看当前系统是否已经安装了JDK，可以使用命令：sudo update-alternatives --config java



## 2.2、下载JDK

     由于当前系统中并没有我们需要的oracle JDK，我们需要自己安装一个。

     先在oracle的网站上寻找JDK的下载地址，然后使用wget命令进行下载。或者，你可以将数据下载到电脑上，然后上传到linux上。

wget <http://download.oracle.com/otn-pub/java/jdk/8u60-b27/jdk-8u60-linux-x64.tar.gz?AuthParam=1441700527_0f80e68acf711ce8da8af237439f406a>



## 2.3、解压安装包

       mv jdk-8u60-linux\*   jdk-8u60-linux-x64.tar.gz

       tar -zxvf jdk-8u60-linux-x64.tar.gz -C ../install/

## 2.4、配置JDK部署信息到环境变量

        #set java env

        export JAVA\_HOME=/export/servers/jdk

        export JRE\_HOME=${JAVA\_HOME}/jre

        export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib

        export PATH=${JAVA\_HOME}/bin:$PATH

## 2.5、让配置信息生效

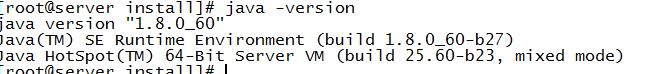
        source /etc/profile

## 2.6、将oracle JDK 配置成默认的JDK

update-alternatives --install /usr/bin/java java /export/servers/jdk/bin/java 300

update-alternatives --install /usr/bin/javac javac /export/servers/jdk//bin/javac 300

## 2.7、产看当前机器的JDK版本



# 3、安装zookeeper

## 3.1、下载zookeeper安装包，

建议下载3.4.5及以上的版本

<http://www.apache.org/dyn/closer.cgi/zookeeper/>

## 3.2、下载完毕之后，解压文件

    tar -zxvf zookeeper-3.4.7.tar.gz -C  /export/servers/

   cd /export/servers

    ln -s zookeeper-3.4.7 zookeeper

## 3.3、修改配置文件

     cd  /export/servers/zookeeper/conf

     mv zoo\_sample.cfg zoo.cfg

     vi zoo.cfg

  输入以下内容：

|  |
| --- |
| #基本事件单元，以毫秒为单位。它用来控制心跳和超时，默认情况下最小的会话超时时间为两倍的 tickTime  tickTime=2000  #此配置表示，允许 follower （相对于 leader 而言的“客户端”）连接并同步到 leader 的初始化连接时间，它以 tickTime 的倍数来表示。当超过设置倍数的 tickTime 时间，则连接失败  initLimit=10  #此配置表示， leader 与 follower 之间发送消息，请求和应答时间长度。如果 follower 在设置的时间内不能与 leader 进行通信，那么此 follower 将被丢弃  syncLimit=5  #数据目录. 可以是任意目录,其中的dataDir目录和dataLogDir需要提前建立好  #注意 应该谨慎地选择日志存放的位置，使用专用的日志存储设备能够大大地提高系统的性能，如果将日志存储在比较繁忙的存储设备上，那么将会在很大程度上影响系统的性能。  dataDir=/export/servers/data/zookeeper  #log目录, 同样可以是任意目录. 如果没有设置该参数, 将使用和dataDir相同的设置，其中的dataDir目录和dataLogDir需要提前建立好  #注意 应该谨慎地选择日志存放的位置，使用专用的日志存储设备能够大大地提高系统的性能，如果将日志存储在比较繁忙的存储设备上，那么将会在很大程度上影响系统的性能。  dataLogDir=/export/servers/logs/zookeeper  #监听client连接的端口号.  clientPort=2181  #这个操作将限制连接到 ZooKeeper 的客户端的数量，限制并发连接的数量，它通过 IP 来区分不同的客户端。此配置选项可以用来阻止某些类别的 Dos 攻击。将它设置为 0 或者忽略而不进行设置将会取消对并发连接的限制。  maxClientCnxns=0  #最小的会话超时时间以及最大的会话超时时间。  #其中，最小的会话超时时间默认情况下为 2 倍的 tickTme 时间  #最大的会话超时时间默认情况下为 20 倍的会话超时时间  minSessionTimeout=4000  maxSessionTimeout=10000  #server.X=A:B:C 其中X是一个数字, 表示这是第几号server. A是该server所在的IP地址. B配置该server和集群中的leader交换消息所使用的端口. C配置选举leader时所使用的端口.  #在之前设置的dataDir中新建myid文件, 写入一个数字, 该数字表示这是第几号server. 该数字必须和zoo.cfg文件中的server.X中的X一一对应.  server.1=192.168.52.106:2888:3888  server.2=192.168.52.107:2888:3888  server.3=192.168.52.108:2888:3888 |

## 3.4、创建zk的数据目录和日志目录

    mkdir -p /export/servers/data/zookeeper

    mkdir -p /export/servers/logs/zookeeper

## 3.5、在数据目录下创建zk节点的编号

    在上文中 /export/servers/data/zookeeper 的目录下，创建myid文件。

    myid文件的内容，根据所属主机编号来编写。

    解释：

    创建数据目录,也就是在你zoo.cfg配置文件里dataDir指定的那个目录下创建myid文件,并且指定id,改id为你zoo.cfg文件中server.1=localhost:2887:3887中的1.只要在myid头部写入1即可.

## 3.6、分发修改后的安装文件

   scp zookeeper-3.4.6 hadoop02:/export/servers/

   scp zookeeper-3.4.6 hadoop03:/export/servers/

## 3.7、在分发后的机器上，执行步骤5的操作。

    创建数据目录,也就是在你zoo.cfg配置文件里dataDir指定的那个目录下创建myid文件,并且指定id,改id为你zoo.cfg文件中server.1=localhost:2887:3887中的1.只要在myid头部写入1即可.

## 3.8、在所有 所有 所有机器上配置环境变量

       #set ZK env

        export ZK\_HOME=/export/servers/zk

        export PATH=${ZK\_HOME}/bin:$PATH

## 3.9、在所有 所有 所有机器上让配置文件生效

    source  /etc/profile

## 3.10、启动zk集群

   依次在不同的节点上，输入zkServers.sh start

   出现错误后，可参考：<http://maoxiangyi.cn/index.php/archives/121>

## 3.11、查看zk集群的状态

    依次在不同的节点上，输入zkServers.sh status

    只有一个主节点，leader  其他都是follow

# 4、安装storm

## 4.1、下载安装包

wget    <http://124.202.164.6/files/1139000006794ECA/apache.fayea.com/storm/apache-storm-0.9.5/apache-storm-0.9.5.tar.gz>

## 4.2、解压安装包

tar -zxvf apache-storm-0.9.5.tar.gz -C /export/servers/

cd /export/servers/

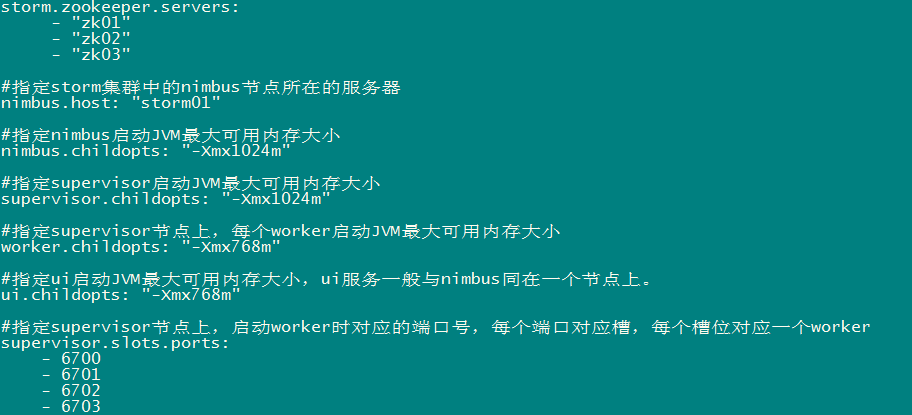
ln -s apache-storm-0.9.5 storm

## 4.3、修改配置文件

mv /export/servers/storm/conf/storm.yaml /export/servers/storm/conf/storm.yaml.bak

vi /export/servers/storm/conf/storm.yaml

输入以下内容：



|  |
| --- |
| #指定storm使用的zk集群  storm.zookeeper.servers:  - "zk01"  - "zk02"  - "zk03"  #指定storm集群中的nimbus节点所在的服务器  nimbus.host: "storm01"  #指定nimbus启动JVM最大可用内存大小  nimbus.childopts: "-Xmx1024m"  #指定supervisor启动JVM最大可用内存大小  supervisor.childopts: "-Xmx1024m"  #指定supervisor节点上，每个worker启动JVM最大可用内存大小  worker.childopts: "-Xmx768m"  #指定ui启动JVM最大可用内存大小，ui服务一般与nimbus同在一个节点上。  ui.childopts: "-Xmx768m"  #指定supervisor节点上，启动worker时对应的端口号，每个端口对应槽，每个槽位对应一个worker  supervisor.slots.ports:  - 6700  - 6701  - 6702  - 6703 |

## 4.4、分发安装包

scp -r /export/servers/apache-storm-0.9.5 storm02:/export/servers

然后分别在各机器上创建软连接

cd /export/servers/

ln -s apache-storm-0.9.5 storm

## 4.5、启动集群

* 在nimbus.host所属的机器上启动 nimbus服务

cd /export/servers/storm/bin/

nohup ./storm nimbus &

* 在nimbus.host所属的机器上启动ui服务

cd /export/servers/storm/bin/

nohup ./storm ui &

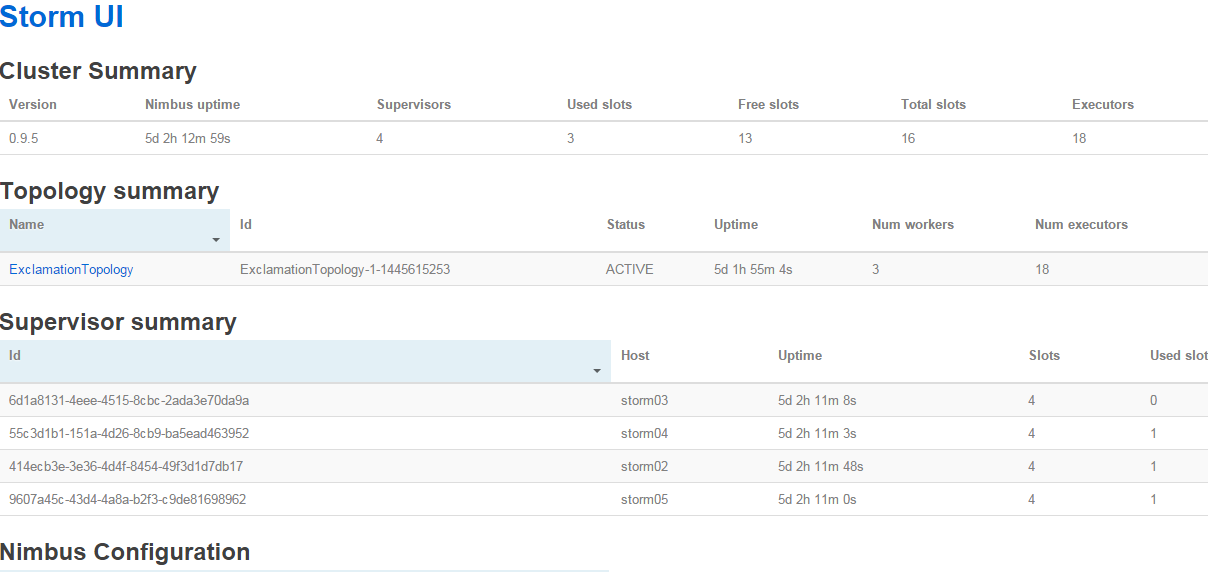
* 在其它个点击上启动supervisor服务

cd /export/servers/storm/bin/

nohup ./storm supervisor &

## 4.6、查看集群

访问nimbus.host:/8080，即可看到storm的ui界面。



## 4.7、Storm常用操作命令

有许多简单且有用的命令可以用来管理拓扑，它们可以提交、杀死、禁用、再平衡拓扑。

* 提交任务命令格式：storm jar 【jar路径】 【拓扑包名.拓扑类名】 【拓扑名称】

bin/storm jar examples/storm-starter/storm-starter-topologies-0.10.0.jar storm.starter.WordCountTopology wordcount

* 杀死任务命令格式：storm kill 【拓扑名称】 -w 10（执行kill命令时可以通过-w [等待秒数]指定拓扑停用以后的等待时间）

storm kill topology-name -w 10

* 停用任务命令格式：storm deactivte 【拓扑名称】

storm deactivte topology-name

我们能够挂起或停用运行中的拓扑。当停用拓扑时，所有已分发的元组都会得到处理，但是spouts的nextTuple方法不会被调用。销毁一个拓扑，可以使用kill命令。它会以一种安全的方式销毁一个拓扑，首先停用拓扑，在等待拓扑消息的时间段内允许拓扑完成当前的数据流。

* 启用任务命令格式：storm activate【拓扑名称】

storm activate topology-name

* 重新部署任务命令格式：storm rebalance 【拓扑名称】

storm rebalance topology-name

再平衡使你重分配集群任务。这是个很强大的命令。比如，你向一个运行中的集群增加了节点。再平衡命令将会停用拓扑，然后在相应超时时间之后重分配工人，并重启拓扑。