**HDP2.3大数据平台**

**安装文档**

# 版本历史

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **文档说明** | **修订人** | **修订日期** |
| 1.0 | 文档创建 | 谷海涛 | 2014 年 12 月 9 日 |
| 1.1 | 文档修改 | 罗春雨 | 2015 年 10 月 15日 |
| 1.2 | 文档修改 | 谌鹏川 | 2016年11月4日 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 目录

[版本历史 2](#_Toc432756067)

[目录 3](#_Toc432756068)

[1 Ambari架构 4](#_Toc432756069)

[2 集群配置 5](#_Toc432756070)

[2.1 集群组成 5](#_Toc432756071)

[2.2 基础软件要求 6](#_Toc432756072)

[2.3 修改主机名并设置DNS解析 6](#_Toc432756073)

[2.4 配置ssh无密码登录 7](#_Toc432756074)

[2.5 配置ntp服务 8](#_Toc432756075)

[2.6 关闭防火墙服务 8](#_Toc432756076)

[2.7 配置selinux和umask 9](#_Toc432756077)

[2.8 关闭THP（推荐） 9](#_Toc432756078)

[2.9 修改系统Linux值（推荐） 10](#_Toc432756079)

[3 部署软件 11](#_Toc432756080)

[3.1 JDK下载安装部署 11](#_Toc432756081)

[3.2 Ambari和HDP下载配置 11](#_Toc432756082)

[3.3 Ambari安装部署 13](#_Toc432756083)

[3.4 HDP安装部署 16](#_Toc432756084)

[4 部署注意事项 28](#_Toc432756085)

[4.1 重启注意检查防火墙 28](#_Toc432756086)

# 1 Ambari架构



Ambari代理：部署在集群中的每个节点上，负责采集所在节点的状态信息（如CPU负载、内存使用率等）和执行结果信息，并且汇总后通过心跳发送给服务端；此外，还处理服务端的执行请求。

Ambari服务端：汇总代理端发送的所有节点的信息；部分运行管理操作通过代理端完成，服务端则负责相应运行管理操作的下发，并将所有状态信息的维护和变更记录在数据库中；对于其他与大数据平台中组件交互的运行管理操作，需要调用大数据平台接口完成。包括以下组件：

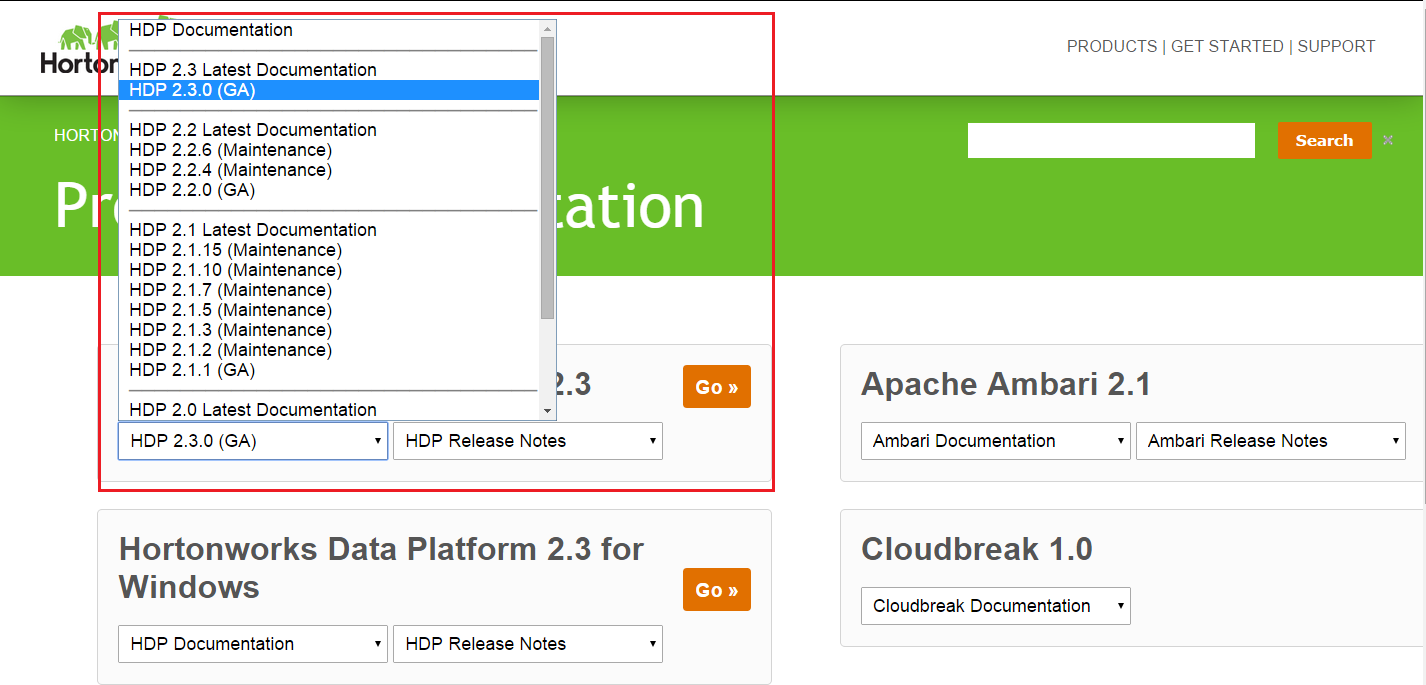
* Postgres关系型数据库：存储了集群配置信息；
* 授权提供者：与某个有组织的认证/授权提供者结合，例如LDAP服务；
* REST API：向Ambari Web前端页面开放API，此外REST API可以被定制应用使用。

Ambari Web：向用户提供的可视化界面。

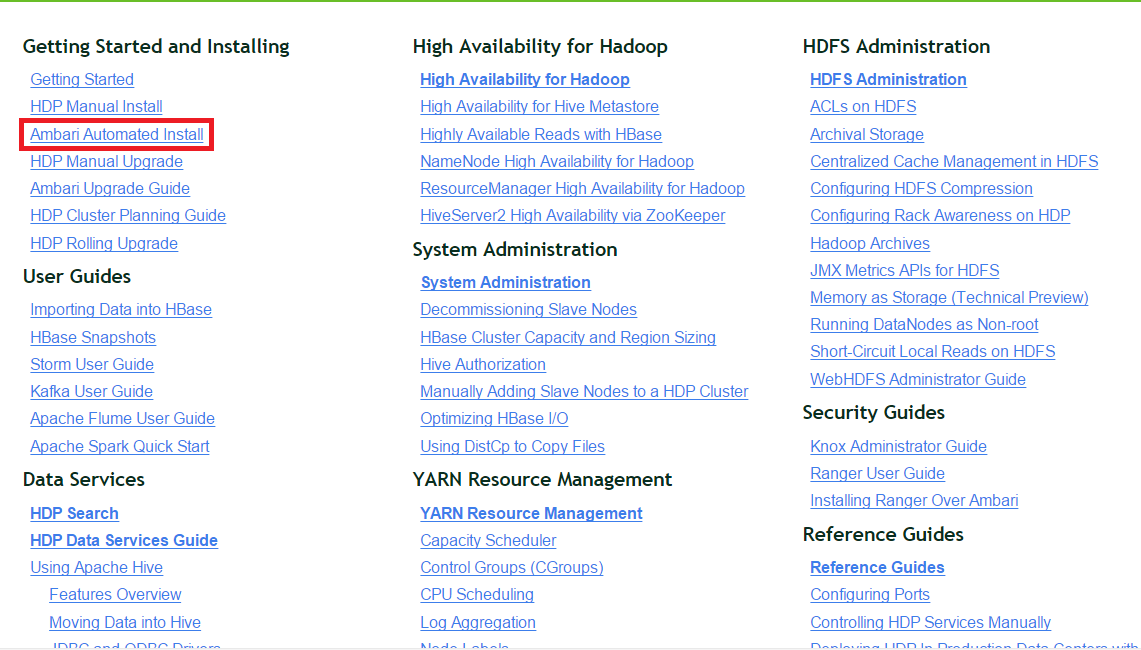
# 2 集群配置

本文使用的Ambari2.1和HDP2.3部署，如果想使用更新的版本，可以参考下面的网址部署

<http://docs.hortonworks.com/>



选择进入相应的版本后，选择进入“[Ambari Automated Install](http://docs.hortonworks.com/HDPDocuments/Ambari-2.1.1.0/bk_Installing_HDP_AMB/content/index.html)”，参考该文档进行部署



## 2.1 集群组成

本次部署总共有6个节点，每个节点都使用centos最小化安装（mini），如下所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 主机ip | 主机名 | 操作系统 | CPU | 内存 | 作用 |
| 192.1.18.239 | ambari.hdp.ap | Centos6.6 | Intel G640 双核 2.8GHz | 2g | 部署Ambari  Server |
| 192.1.18.240 | hdp1.hdp.ap | Centos6.6 | Intel i5-3470 双核四线程 3.2GHz | 4g | hdp部署节点1 |
| 192.1.18.241 | hdp2.hdp.ap | Centos6.6 | Intel i5-3470 双核四线程 3.2GHz | 4g | hdp部署节点2 |
| 192.1.18.242 | hdp3.hdp.ap | Centos6.6 | Intel i5-3470 双核四线程 3.2GHz | 4g | hdp部署节点3 |
| 192.1.18.243 | hdp4.hdp.ap | Centos6.6 | Intel i5-3470 双核四线程 3.2GHz | 4g | hdp部署节点4 |
| 192.1.18.244 | hdp5.hdp.ap | Centos6.6 | Intel G630 双核 2.7GHz | 2g | hdp部署节点5 |

## 2.2 基础软件要求

集群每个节点需要安装scp、curl（centos mini已安装）、unzip、、tar（centos mini已安装）、wget和python（2.6以上，centos 6.6默认就是 2.6 的），每个节点执行如下命令安装（需要Internet连接）

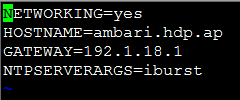
yum install openssh-clients -y

yum install unzip -y

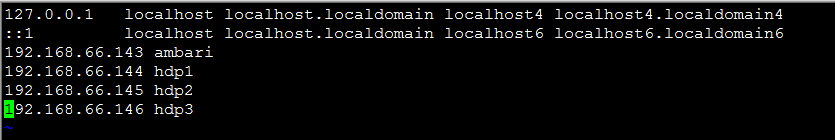
yum install wget –y

## 2.3 修改主机名并设置DNS解析

按照上述表中的主机名在对应的机器上修改它们的hostname，在/etc/sysconfig/network中修改hostname项，修改完后重启生效，以ambari.hdp.ap为例，修改/etc/sysconfig/network内容如下所示：



然后修改机器的/etc/hosts文件，将每个机器的hostname和ip对应起来，将修改后的/etc/hosts文件复制到集群中的每台机器，操作如下所示（按照实际情况填写）：



将hosts文件拷贝到其他五个节点

## 2.4 配置ssh无密码登录

因为在ambari和hdp部署过程中，ambari和hdp部署的节点都有可能互相访问，所以建议六个节点中的每个节点之间都可以ssh无密码登录，包括它们自己登录到自己。

（root用户）

在ambari.hdp.ap机器上执行如下操作，首先查看该用户根目录下的.ssh文件夹中是否有id\_rsa文件，如果没有执行ssh-keygen命令，一路回车，然后查看生成的文件，为了避免之后的麻烦，建议将该目录下所有文件的权限修改成600，本文档部署在ambari机器上的操作如下所示，其他机器类似：

ssh-keygen

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

chmod 700 ~/.ssh

chmod 600 ~/.ssh/authorized\_keys

在ambari.hdp.ap机器上执行如下将命令

for host in hdp{1,2,3,4,5}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

在hdp1机器上执行如下将命令

for host in ambari.hdp.ap hdp{2,3,4,5}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

在hdp2机器上执行如下将命令

for host in ambari.hdp.ap hdp{1,3,4,5}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

在hdp3机器上执行如下将命令

for host in ambari.hdp.ap hdp{1,2,4,5}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

在hdp4机器上执行如下将命令

for host in ambari.hdp.ap hdp{1,2,3,5}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

在hdp5机器上执行如下将命令

for host in ambari.hdp.ap hdp{1,2,3,4}.hdp.ap;do echo $host; ssh-copy-id -i /root/.ssh/id\_rsa root@$host;done

## 2.5 配置ntp服务

因为hdp中有服务需要集群的时间同步，所以集群中的每一个机器需要安装启动ntp服务保证集群时间的一致，安装启动如下所示：

for host in ambari.hdp.ap hdp{1,2,3,4,5}.hdp.ap;do ssh root@$host yum install ntp -y; ssh root@$host /etc/init.d/ntpd start; done

注意：我搭建了一版，虽然不算集群，但是与遇到了时间不同步的问题。

## 2.6 关闭防火墙服务

因为集群中的线程之间需要互相通信，所以需要合理配置防火墙，最简单的方式是直接关闭防火墙。操作如下所示：

for host in ambari.hdp.ap hdp{1,2,3,4,5}.hdp.ap;do ssh root@$host /etc/init.d/iptables stop; ssh root@$host chkconfig iptables off; done

或者：

service iptables stop

service ip6tables stop

设置成始终保持关闭：

chkconfig iptables off

chkconfig ip6tables off

## 2.7 配置selinux和umask

 为了避免引起不必要的麻烦，关闭集群节点上的selinux服务，然后重启，操作如下所示：

编辑/etc/sysconfig/selinux，设置

 SELINUX=disabled

如果机器安装了PackageKit，在中将其关闭，操作如下：

vi /etc/yum/pluginconf.d/refresh-packagekit.conf

[C:\Users\Administrator\AppData\Local\YNote\data\guhaitao_1986@126.com\bed8a27a95db42e3a78cbbe32af1a8a2\170959_nnpc_1169607.png](http://static.oschina.net/uploads/space/2014/1017/170959_NNpC_1169607.png)

 确保集群中节点的umask值为0022，操作如下：

umask

## 2.8 关闭THP（推荐）

在hdp1.hdp.ap、hdp2.hdp.ap、hdp3.hdp.ap、hdp4.hdp.ap、hdp5.hdp.ap机器上均执行如下命令

echo never > /sys/kernel/mm/redhat\_transparent\_hugepage/enabled

echo never > /sys/kernel/mm/redhat\_transparent\_hugepage/defrag

echo never > /sys/kernel/mm/transparent\_hugepage/enabled

echo never > /sys/kernel/mm/transparent\_hugepage/defrag

## 2.9 修改系统Linux值（推荐）

使用ulimit -a来查看所有限制值

ulimit -a

其中“open files (-n) 1024”，是Linux操作系统对一个进程打开的文件句柄数量的限制

修改/etc/security/limits.conf，添加如下

hadoop - nofile 32768

hadoop - nproc 32000

重启生效

## 2.10关闭yum自动更新(minimal安装不需要此步骤)

service yum-updatesd stop #关闭系统自动更新

chkconfig yum-updatesd off #禁止开启启动

## 2.11更新openssl

CentOS 6.5自带的open ssl 有bug, 需要更新到新版本才能开始安装ambari.

查看安装的openssl包

rpm -qa | grep openssl

安装：

yum install /tmp/soft /openssl-1.0.1e-16.el6\_5.x86\_64.rpm -y

## 2.12配置yum repo

### 关闭yum的fastestmiror plugin

cd /etc/yum/pluginconf.d

vi fastestmirror.conf

enabled=1 改为 enabled=0

# 3 部署软件

## 3.1 JDK下载安装部署

在ambari服务器上创建目录/usr/java

mkdir -p /usr/java

将下载的jdk-7u71-linux-x64.tar.gz（推荐使用jdk1.7）上传至/usr/java目录，在ambari节点上执行如下命令：

cd /usr/java/

tar -xzvf jdk-7u71-linux-x64.tar.gz

rm -rf jdk-7u71-linux-x64.tar.gz

编辑/etc/profile，添加

export JAVA\_HOME=/usr/java/jdk1.7.0\_71

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

然后执行source /etc/profile，使用java -version验证之，其他五个节点配置相同

## 3.2 Ambari和HDP下载配置

本文档部署所用的软件为：

1. ambari-2.1.1-centos6.tar.gz，下载地址为:

<http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.1.1/ambari-2.1.1-centos6.tar.gz>

http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.1.0/ambari-2.2.1.0-centos6.tar.gz

http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.2.0/ambari-2.2.2.0-centos6.tar.gz

1. HDP-2.3.0.0-centos6-rpm.tar.gz，下载地址为:

[http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/HDP-2.3.0.0-centos6-rpm.tar.gz](http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0/HDP-2.3.0.0-centos6-rpm.tar.gz" \t "_blank)

<http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.2.0/HDP-2.4.2.0-centos6-rpm.tar.gz>

1. HDP-UTILS-1.1.0.20-centos6.tar.gz，下载地址为:

http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6/HDP-UTILS-1.1.0.20-centos6.tar.gz。

手动下载所需的上述软件；将下载的ambari-2.1.1-centos6.tar.gz软件放到部署某一台的机器的/var/www/html/ambari目录下, HDP-2.3.0.0-centos6-rpm.tar.gz、HDP-UTILS-1.1.0.20-centos6.tar.gz软件放到该台的机器的/var/www/html/目录下（该机器必须先安装httpd服务，并且启动httpd服务），然后解压，使用如下命令开启httpd服务：

cd /var/www/html

service httpd start

chkconfig httpd on 把httpd设置为自动启动

 在该机器上添加/etc/yum.repos.d/ambari.repo文件，它的内容如下所示，其中的ip地址要换成存放ambari、HDP软件那台机器的ip：

**vi /etc/yum.repos.d/ambari.repo**

**注：其中2.x为版本的缩写，根据自己版本补充完整**

#VERSION\_NUMBER=2.4.0.1-460

[Updates-ambari-2.4.0.1]

name=ambari-2.4.0.1 - Updates

#baseurl=http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.2.2.0

baseurl=http://192.168.17.136/ambari/AMBARI-2.4.0.1/centos6/2.4.0.1-1

gpgcheck=1

#gpgkey=http://public-repo-1.hortonworks.com/ambari/centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

gpgkey=http://192.1.18.239/ambari/AMBARI-2.4.0.1/centos6/2.4.0.1-1/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

enabled=1

priority=1

然后用/var/www/html/AMBARI-2.2.2.0 下面有个setup\_repo.sh ，可以创建ambari.repo

配置HDP.repo文件：

**vi /etc/yum.repos.d/HDP.repo**

#VERSION\_NUMBER=2.5.0.0-258

[HDP-2.5.0.0]

name=HDP Version - HDP-2.5.0.0

#baseurl=http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.2.0

baseurl=http://192.168.2.253/HDP/centos6

gpgcheck=1

#gpgkey=http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.2.0/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

gpgkey=http://192.168.17.136/HDP/centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

enabled=1

priority=1

**vi /etc/yum.repos.d/HDP-UTILS.repo**

[HDP-UTILS-1.1.0.21]

name=HDP Utils Version - HDP-UTILS-1.1.0.21

#baseurl=http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6

baseurl=http://192.168.2.253/HDP-UTILS-1.1.0.21/repos/centos6

gpgcheck=1

#gpgkey=http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.4.2.0/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

gpgkey=http://192.168.17.136/HDP-UTILS-1.1.0.21/repos/centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

enabled=1

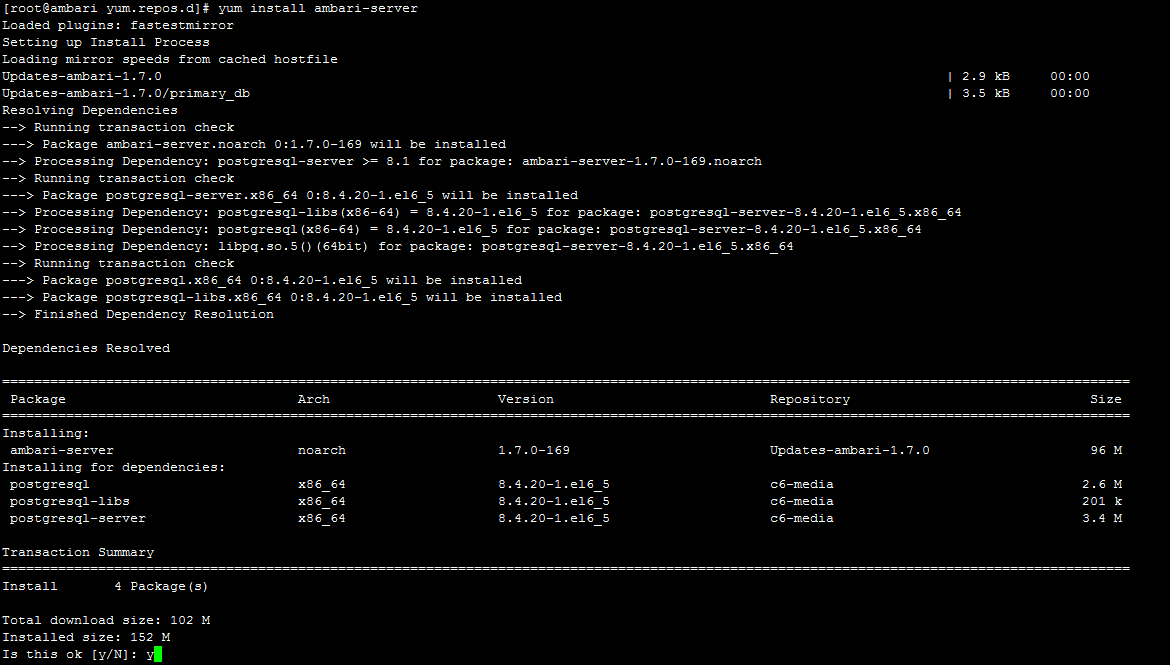
priority=1

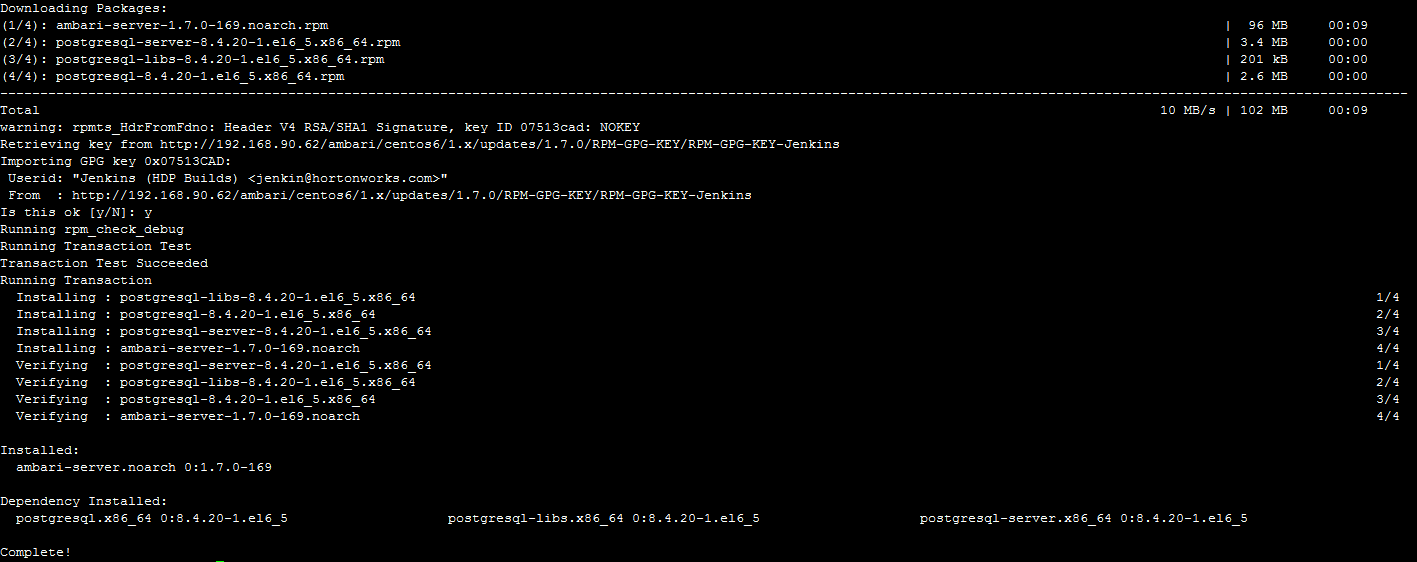
## 3.3 Ambari安装部署

在安装ambari的机器上执行yum install ambari-server命令，操作如下所示:

yum install ambari-server

注意：这里直接复制过去 不好使，需要手敲

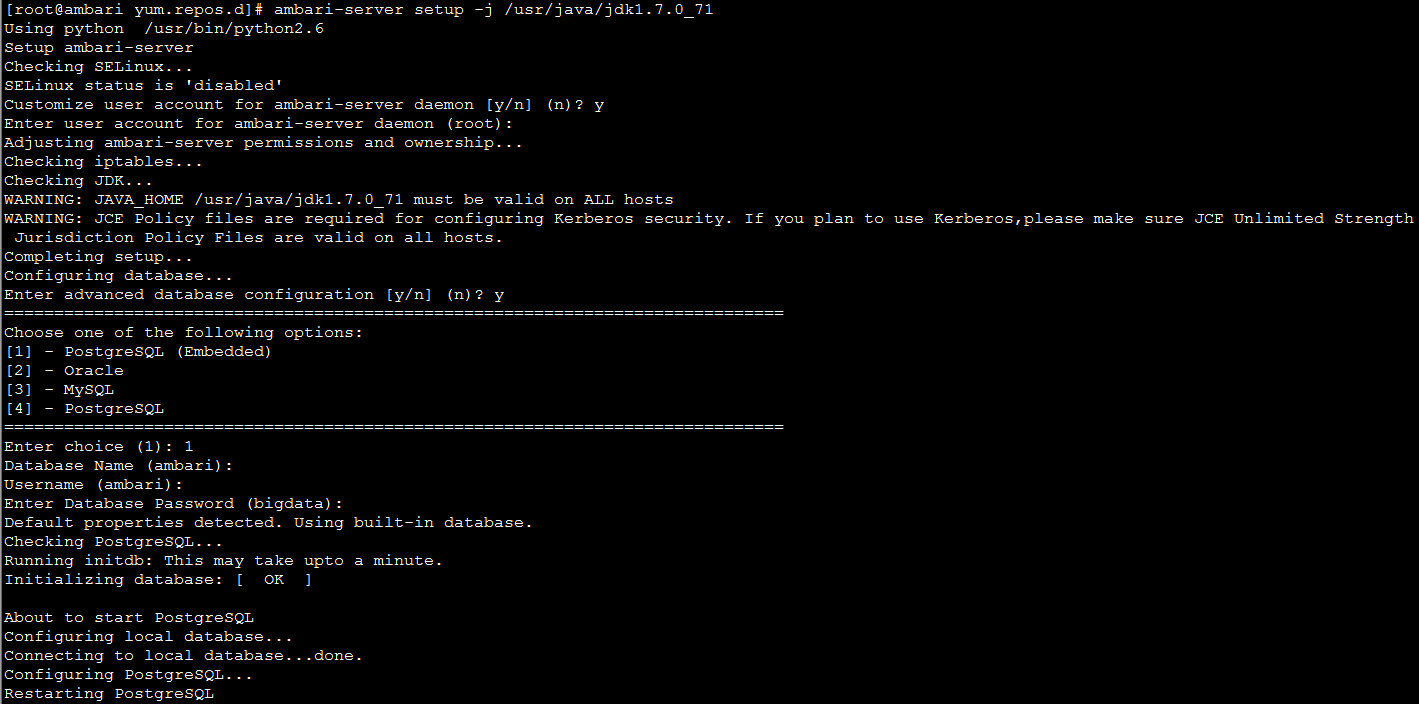




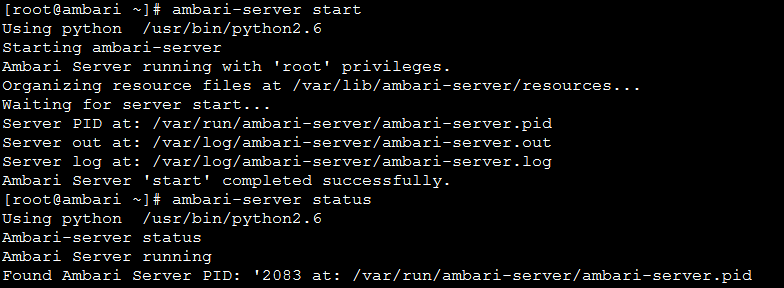
没有报错的话，ambari-server安装成功，接下来执行

ambari-server setup -j /usr/java/jdk1.7.0\_71

启动-j后面的操作指定jdk安装的位置，操作如下：



接下来用/usr/sbin/ambari-server start命令启动ambari-server，用ambari-server stop命令关闭ambari-server，用**ambari-server statu**s命令查看ambari-server的启动状态，操作如下：



http://192.168.2.253:8080/views/ADMIN\_VIEW/2.4.0.1/INSTANCE/#/

## 3.4 HDP安装部署

修改c:\windows\system32\drivers\etc\hosts，添加如下

192.1.18.239 ambari.hdp.ap

192.1.18.240 hdp1.hdp.ap

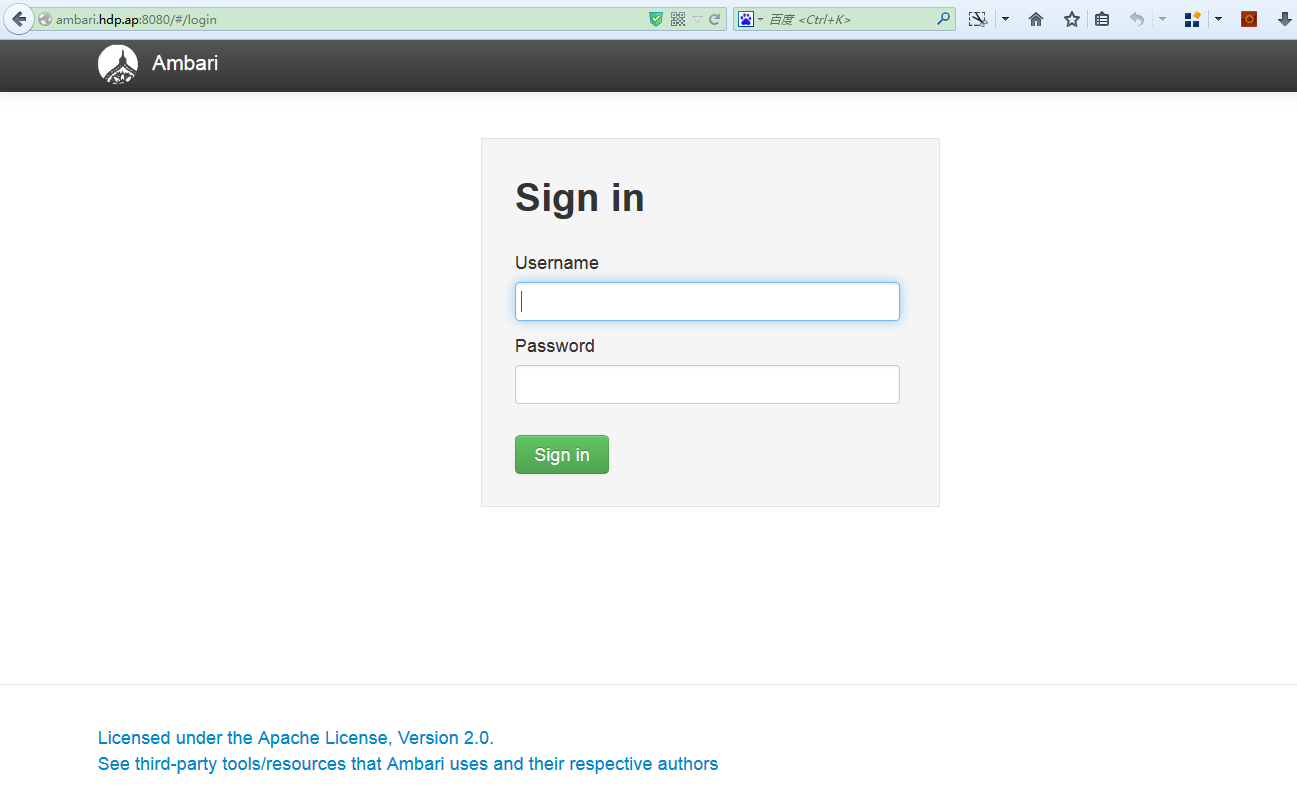
192.1.18.241 hdp2.hdp.ap

192.1.18.242 hdp3.hdp.ap

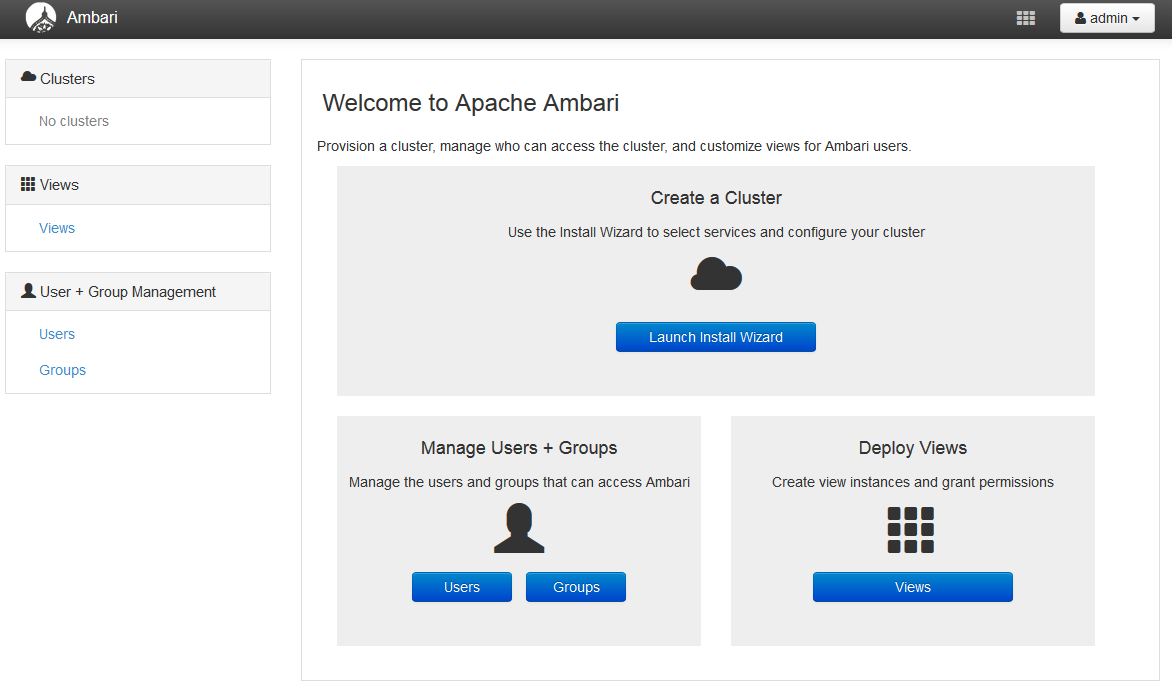
192.1.18.243 hdp4.hdp.ap

192.1.18.244 hdp5.hdp.ap

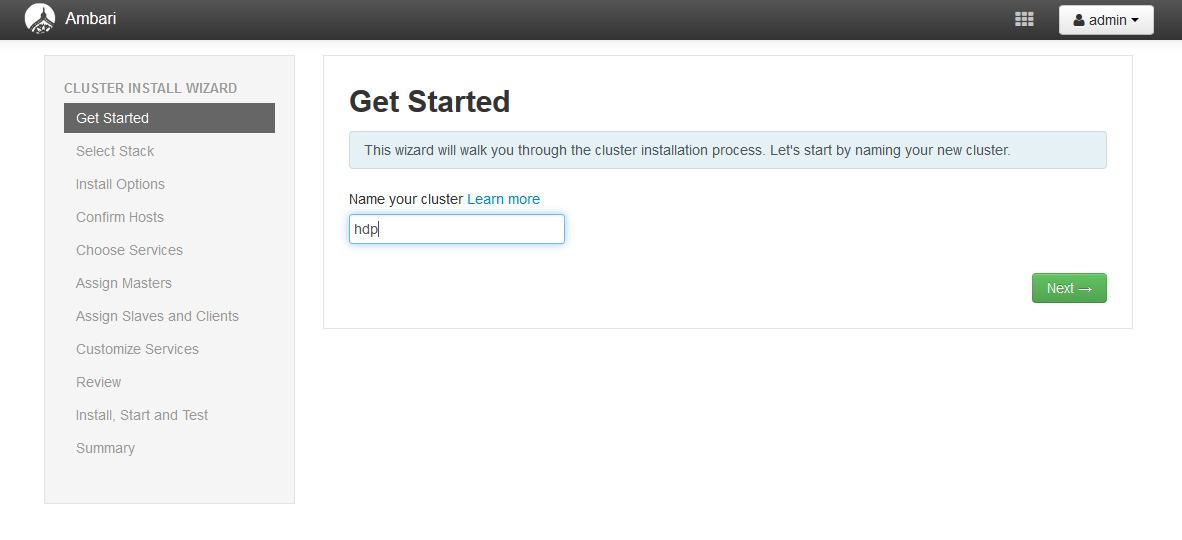
Ambari-server启动成功后，可以通过http://ambari.hdp.ap:8080访问，用户名和密码都是admin，访问界面如下所示：



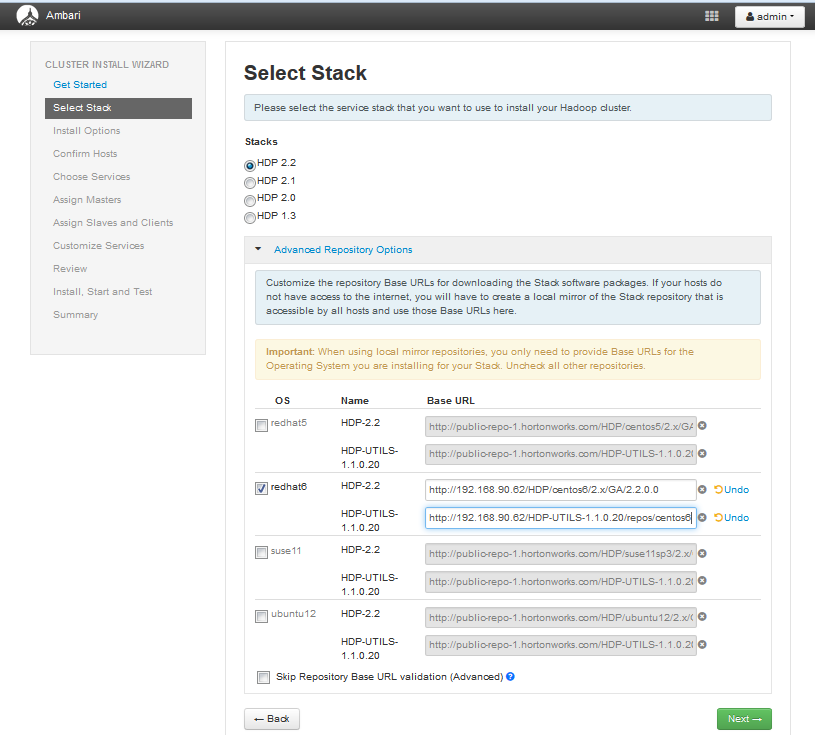
用admin登录，出现如下界面。



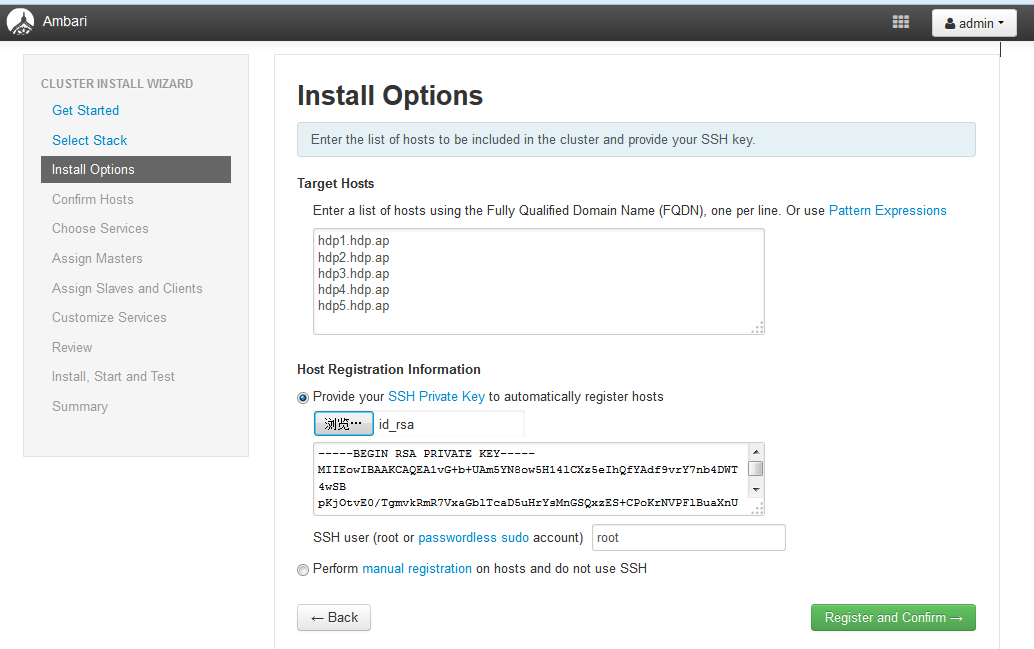
点击“Launch Install Wizard”，接下来可以安装hdp，首先命名集群，选择下一步，操作如下所示：



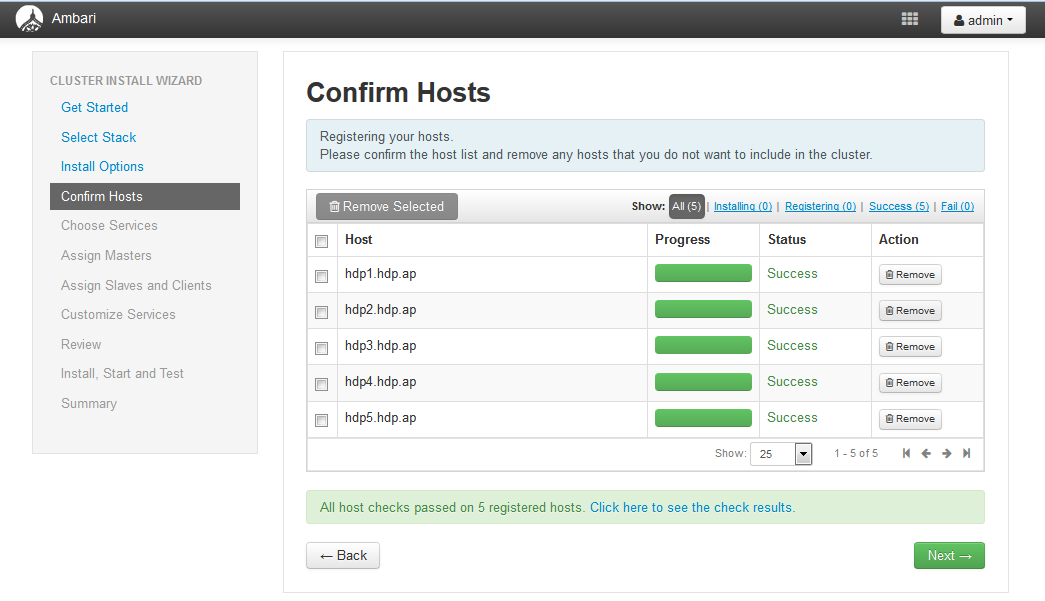
然后选择安装的hdp的版本，本次部署用HDP2.3，所以选择HDP2.3，点开下面的“Advanced Repository Options”，选择操作系统对应的选项，本次部署选择redhat6，修改其中“HDP-2.3”和“HDP-UTILS-1.1.0.20”指向的地址，修改成/etc/yum.repo.d/hdp.repo中它们指向的地址，本次部署“HDP-2.3”指向“http:// 192.168.90.62/HDP/centos6/2.x/updates/2.3.0.0”，“HDP-UTILS-1.1.0.20”指向“http://192.168.90.62/HDP-UTILS-1.1.0.20/repos/centos6”，然后点击下一步：



然后在“Target Hosts”方框中填写部署hdp的集群，本次部署填写如下；还需要指定部署ambari机器的“SSH Private Key”，点击“选择文件”按钮，找到部署ambari.hdp.ap机器上的“~/.ssh/id\_rsa”文件，或者将该文件的内容复制到对应的方框中，然后点击“Register and Confirm”按钮：



接下来会进行集群的注册和验证，出现下述画面代表成功，可以点击“Status”栏中的“Success”或者“Failed”查看日志，如下图所示：



 接下来下一步，显示“选择服务”界面，默认是全选，可以根据自己机器的配置和需要选择服务，这里我没有安装如下服务，以后还可以安装：

Falcon——Data management and processing platform

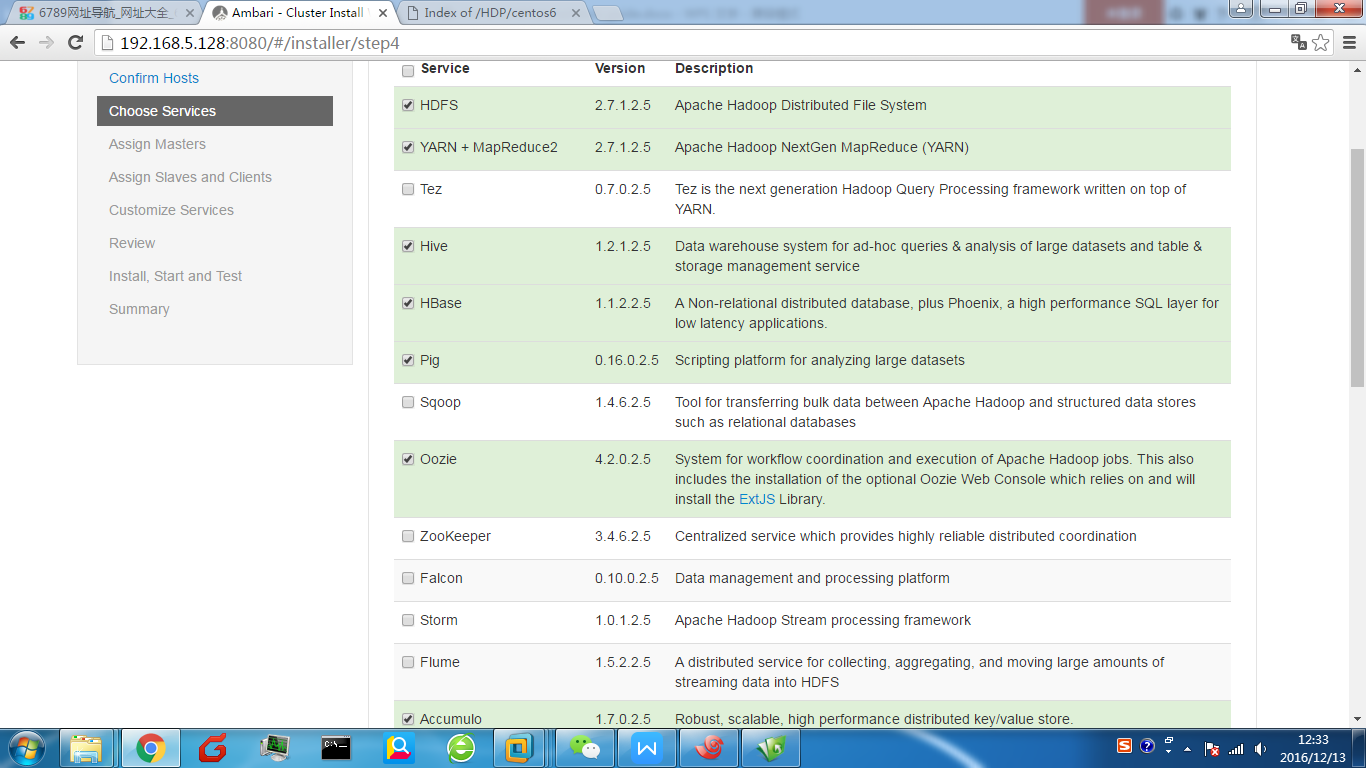
Storm——Apache Hadoop Stream processing framework

Flume——A distributed service for collecting, aggregating, and moving large amounts of streaming data into HDFS

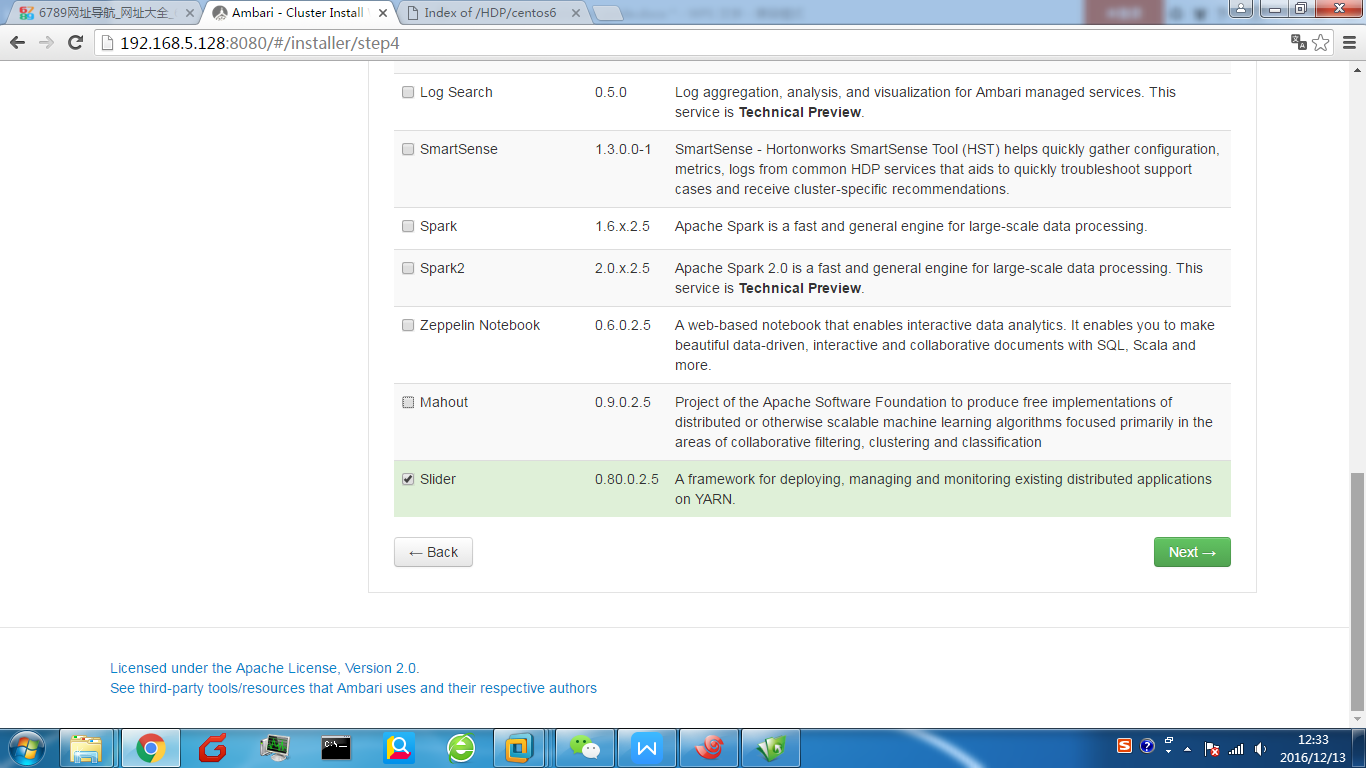
Kafka——A high-throughput distributed messaging system

Knox——Provides a single point of authentication and access for Apache Hadoop services in a cluster

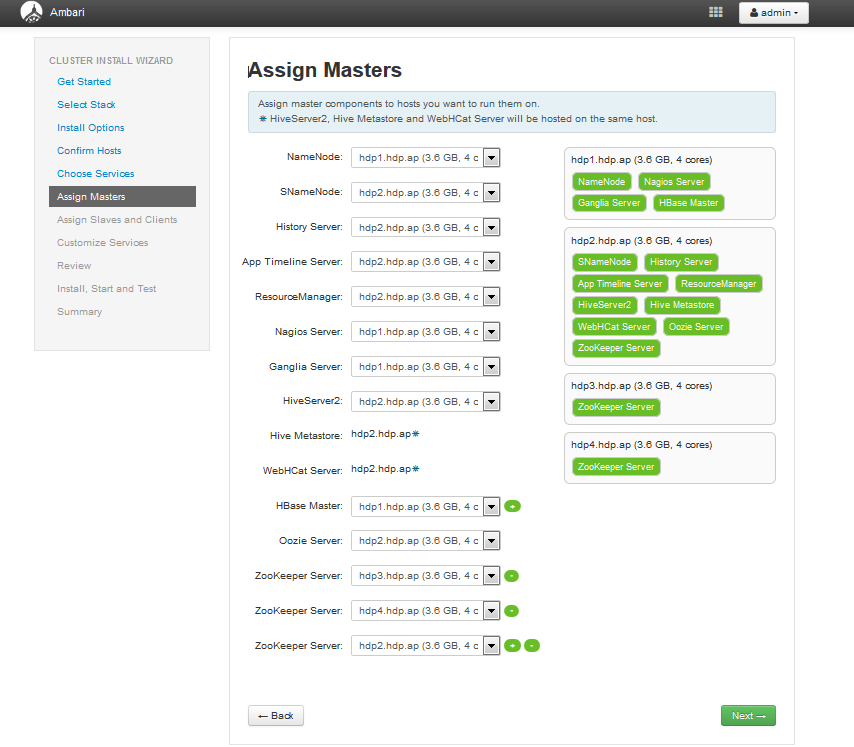
|  |  |  |
| --- | --- | --- |
|  |  | |
|  | |



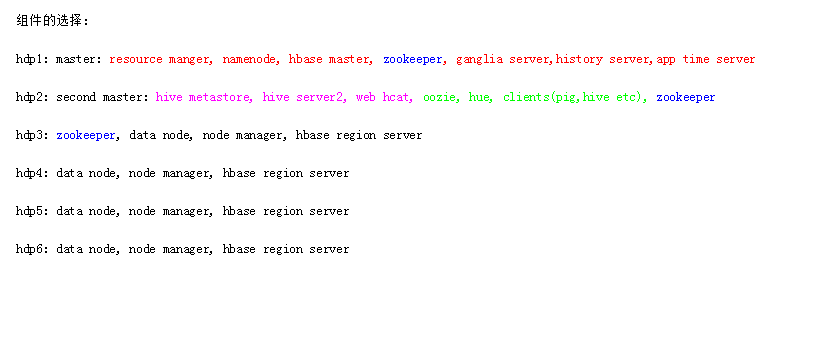




然后点击下一步，接下来显示的是“指定主节点”的界面，这个根据实际情况在下拉菜单中选择，尽可能均匀分配，如下图所示：



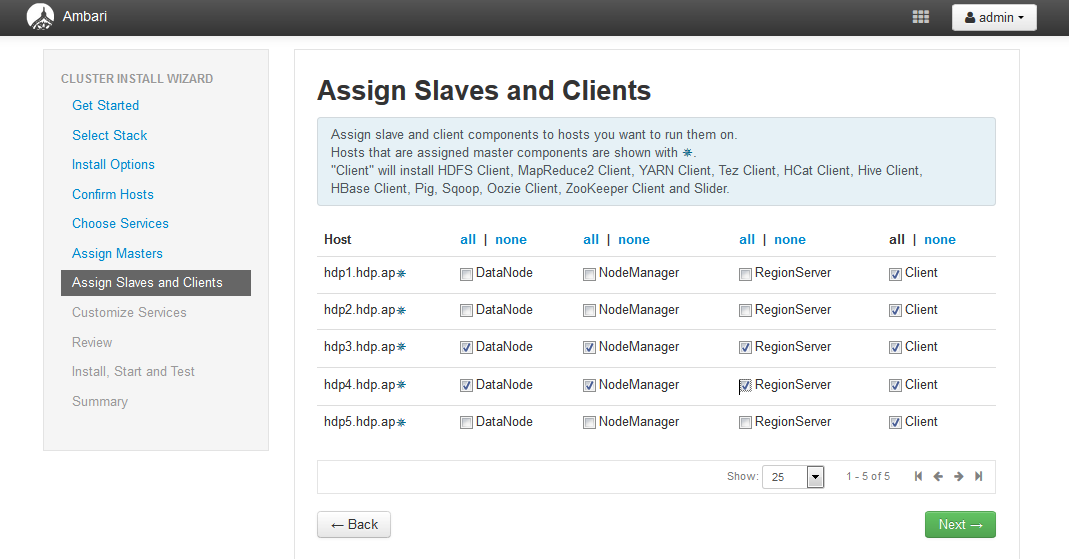
各节点的配置样例（实际配置情况以自己实际需要为准）：



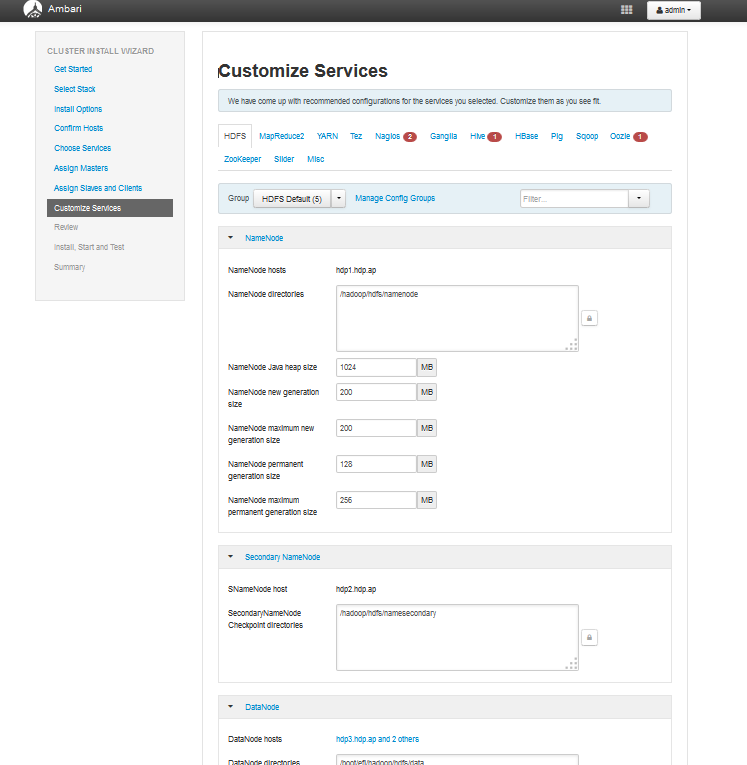
样例中hdp1是NameNode，hdp2是SecondNameNode，hdp3-6是datanode。

咱们的carposerver可以部署到second master上。

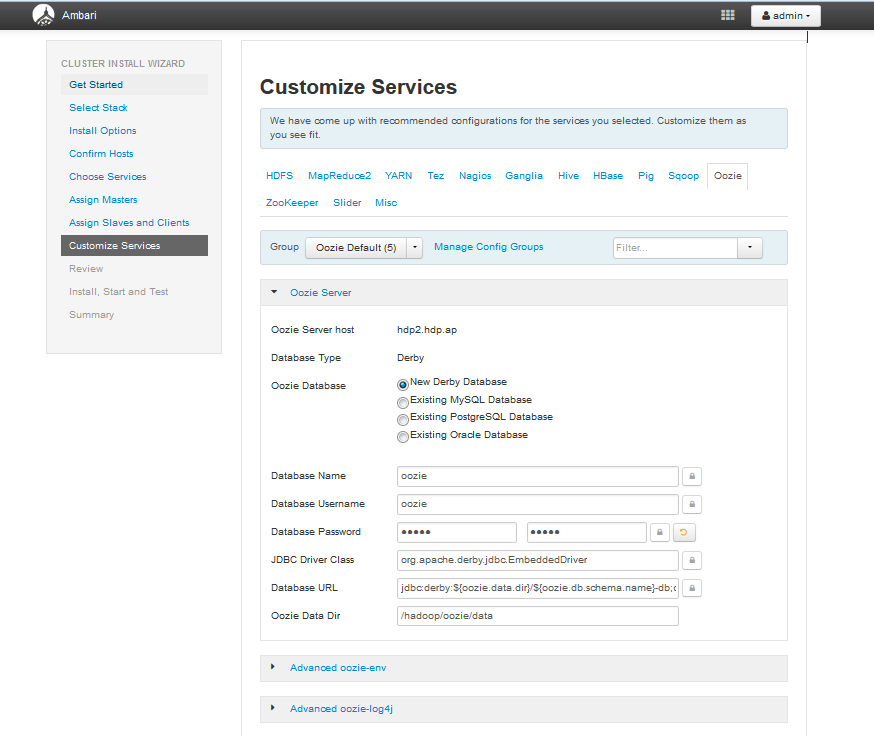
然后点击下一步，接下来显示的是“指定从节点和客户端”的界面，这个根据实际情况选择服务，尽可能均匀分配，如下图所示：



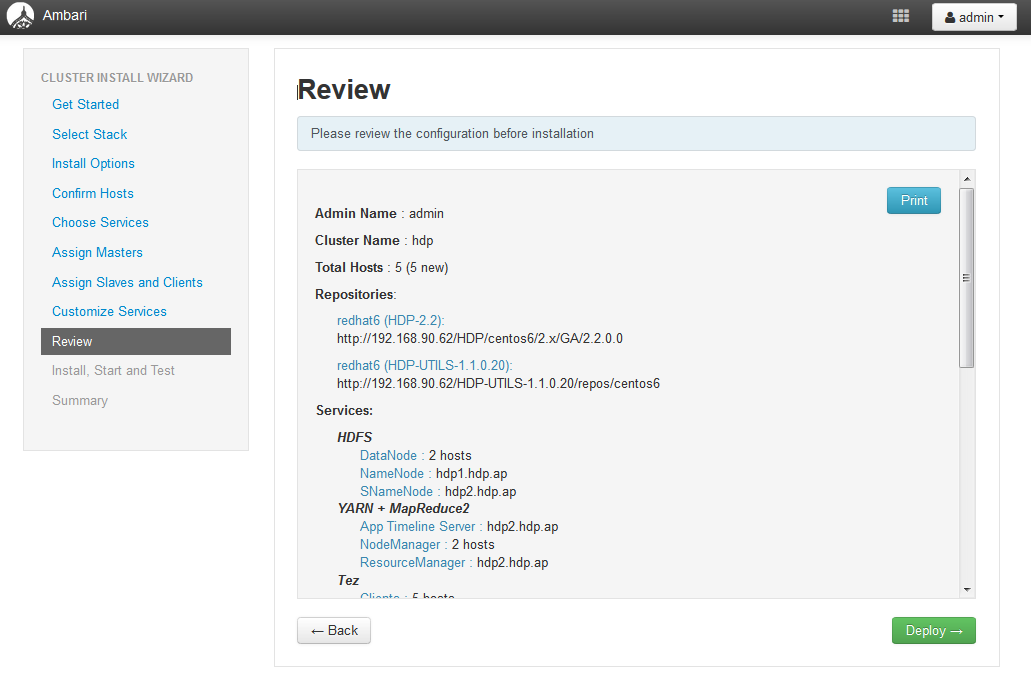
然后点击下一步，接下来显示的是“配置服务选项”的界面，这个根据机器性能配置，如下图所示：



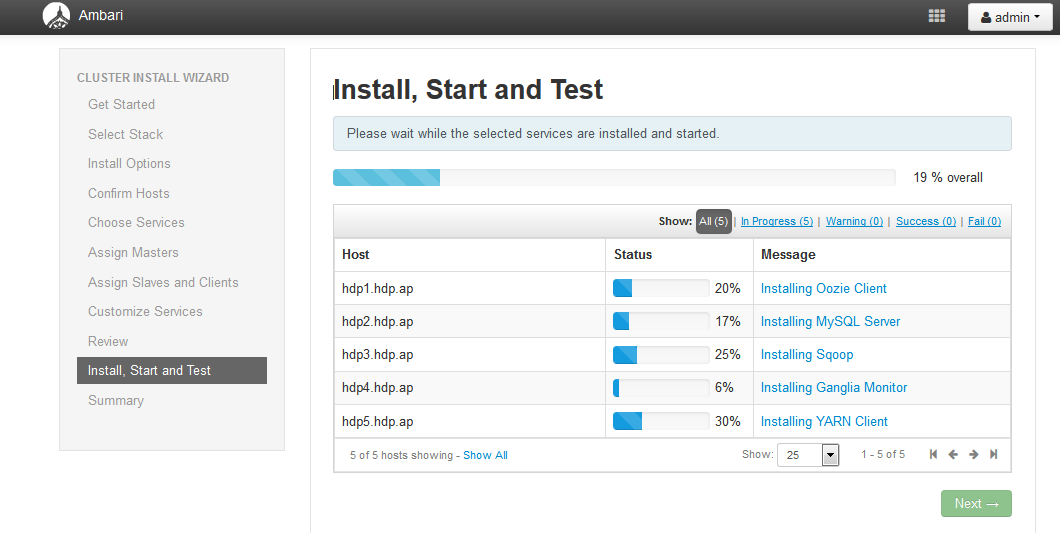
其中旁边显示红色小圆圈带数字的表示此项服务中的某些项必须配置，其中的“Hive”、“Oozie”和“Nagios”项中需要设置密码和email，点开进行配置，操作如下：



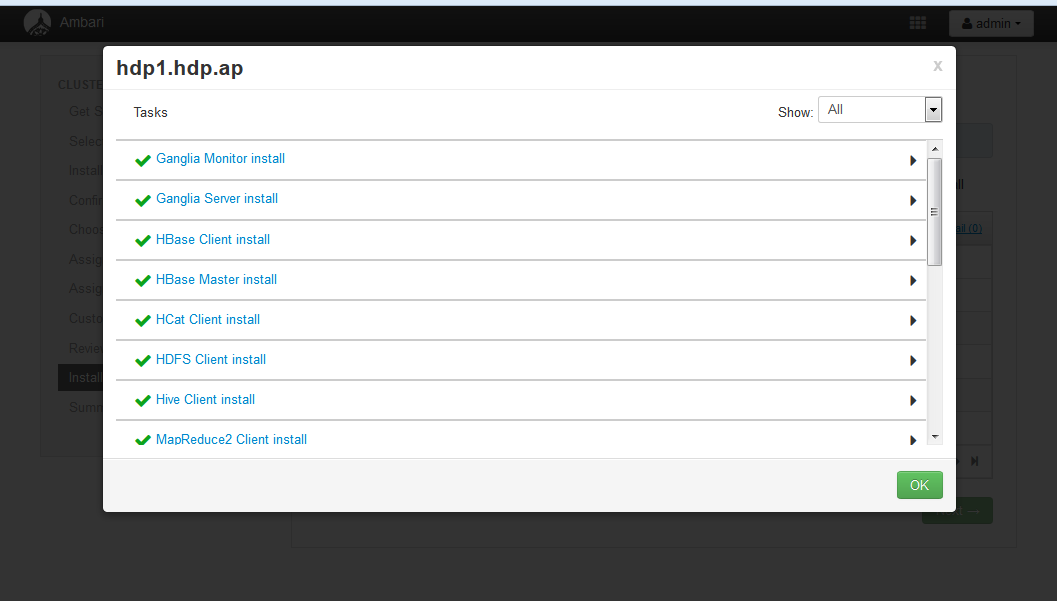
然后点击下一步，接下来显示的是“复查”界面，显示了集群配置服务的情况，没有问题，点击“部署”按钮，如下图所示：



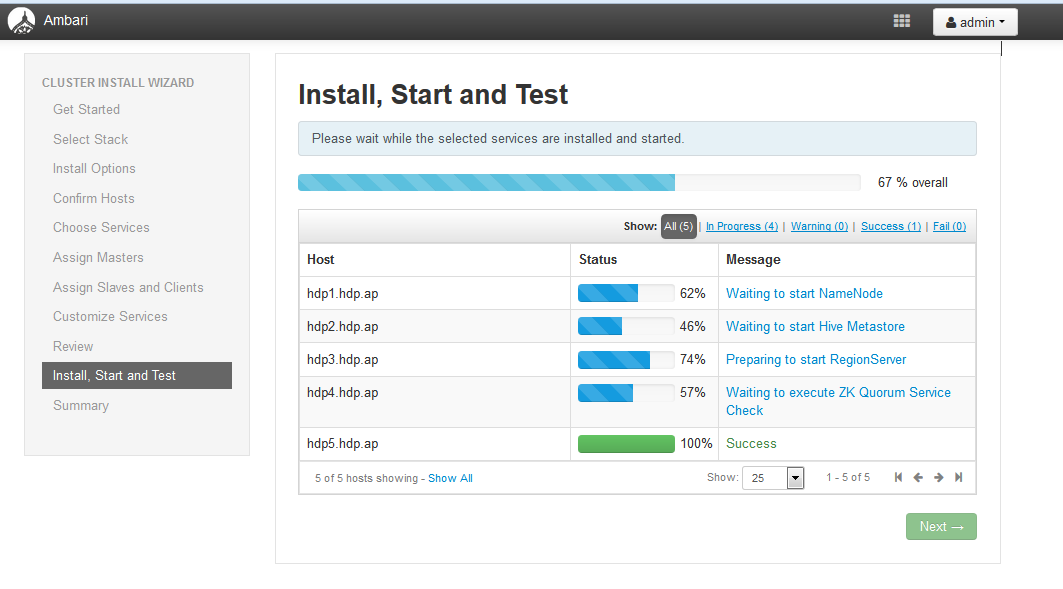
接下来进入“安装、启动、测试”界面，如下图所示：

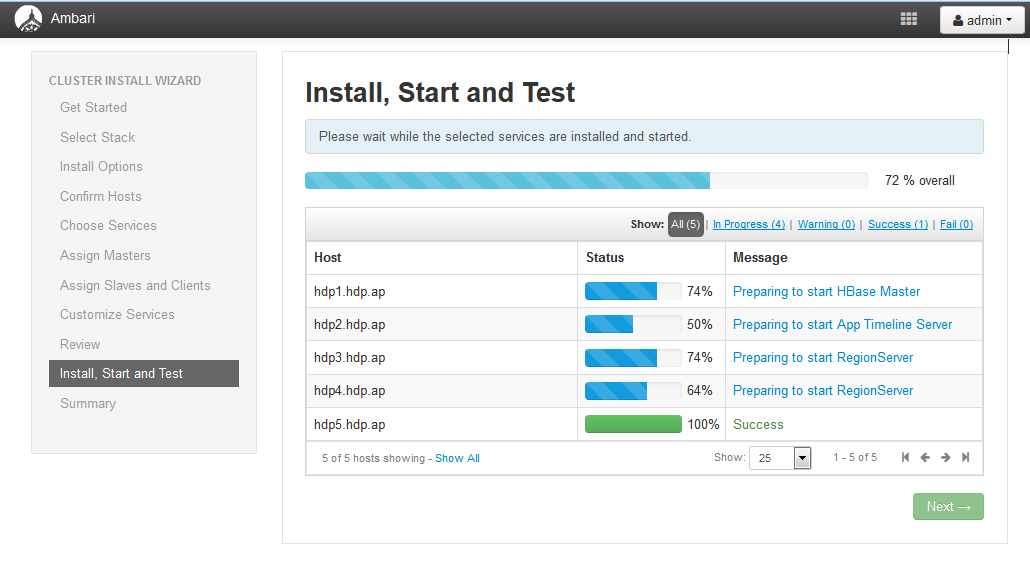


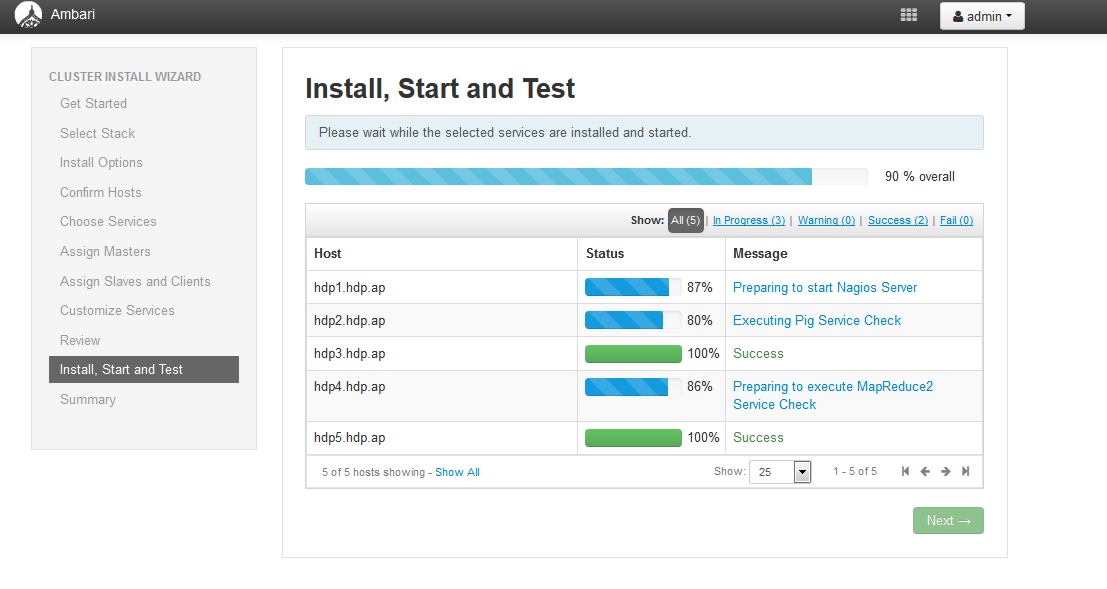
在“安装、启动、测试”过程中，随时可以点击“Message”栏中的项查看日志，日志显示了部署的进展情况，如下图所示：



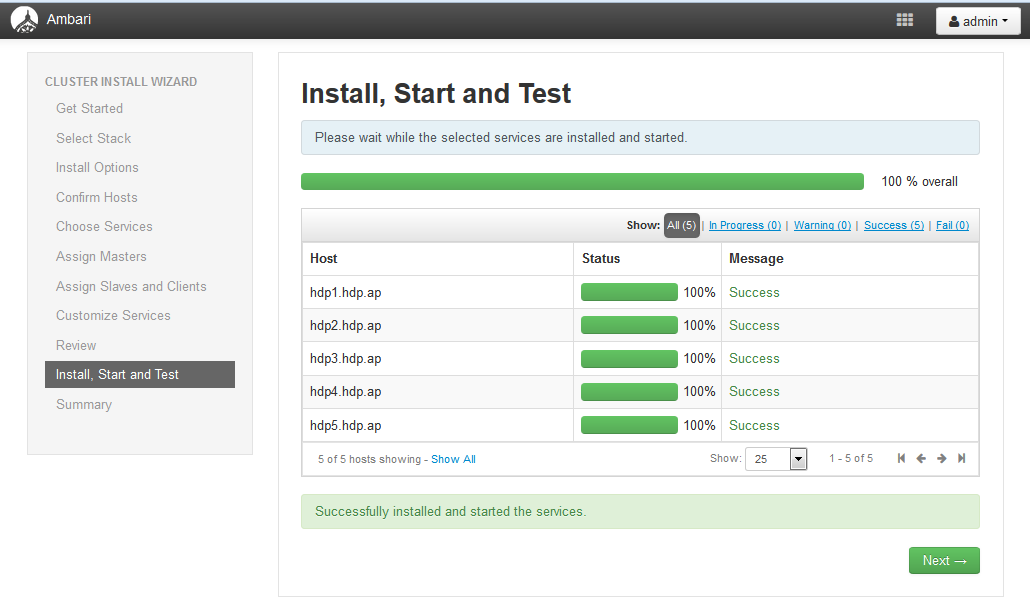
下图截取了几张部署过程中的图，仅供参考：



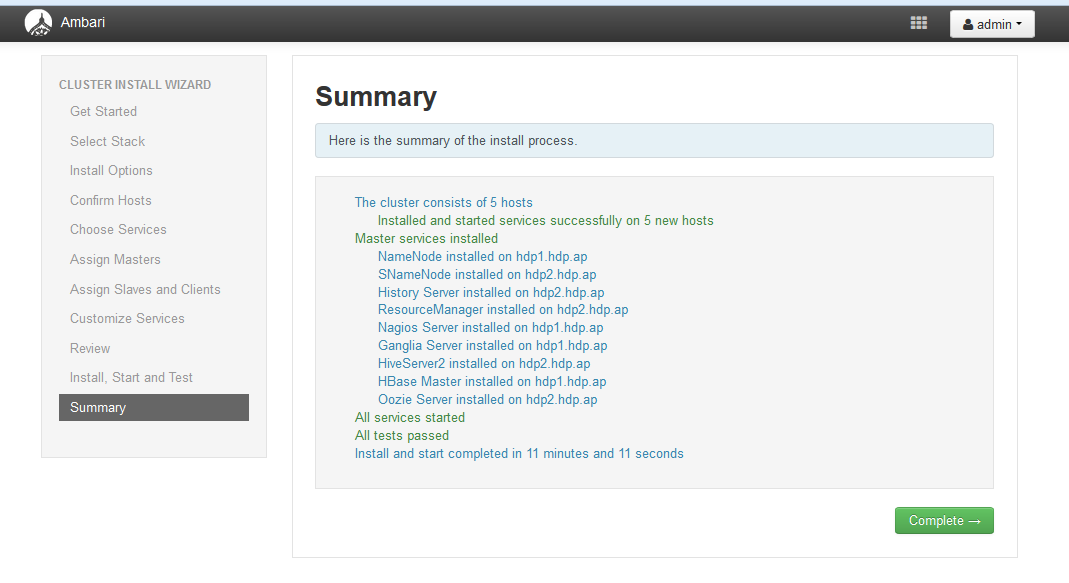




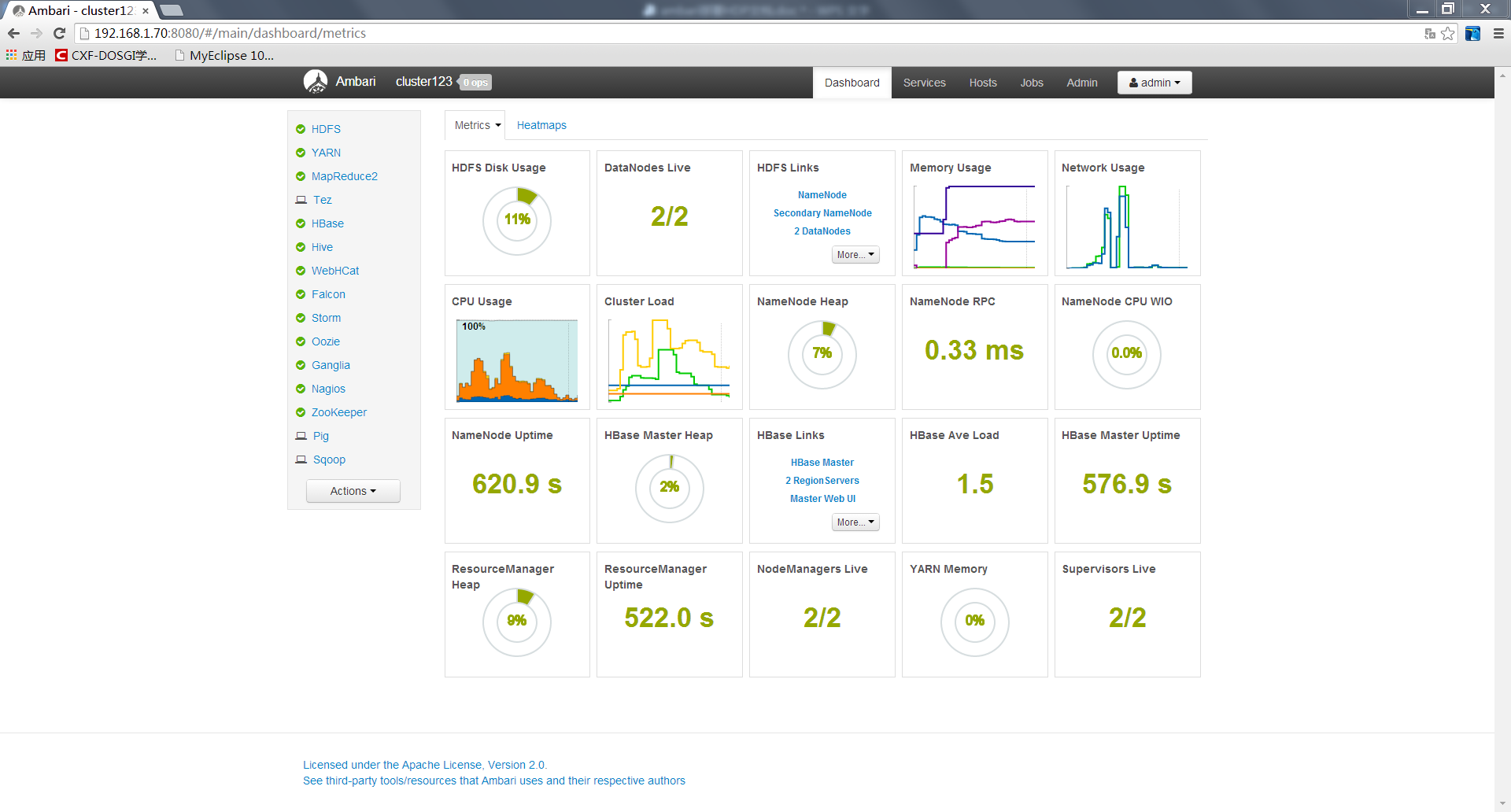
出现下面画面表示“安装、启动、测试”成功：



然后点击下一步，显示的是“部署情况”界面，如下图所示：



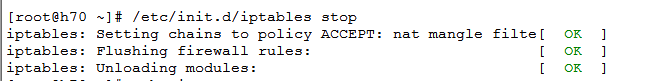
然后点击“完成”按钮，进入ambari管理界面，如下图所示：

[](http://static.oschina.net/uploads/space/2014/1017/173016_7qPW_1169607.png)

# 4 部署注意事项

## 4.1 重启注意检查防火墙

    每次机器重启后，检查下防火墙的状态，如果防火墙开着，一定要执行关闭命令，操作如下：

[](http://static.oschina.net/uploads/space/2014/1017/173143_EojJ_1169607.png)