**大数据平台一键部署工具Ambari**

**安装部署手册**

目录

[一、背景 3](#_Toc467484141)

[二、安装前准备 3](#_Toc467484142)

[三、安装 4](#_Toc467484143)

[2.1 配置本地源 4](#_Toc467484144)

[2.1.1 安装配置本地源需要的组件 4](#_Toc467484145)

[2.1.2 安装httpd服务 4](#_Toc467484146)

[2.1.3 下载、配置Ambari本地源ambari.repo 5](#_Toc467484147)

[2.1.2 配置HDP本地源 6](#_Toc467484148)

[2.1.3 配置HDP-UTILS 7](#_Toc467484149)

[2.1.4 生成本地源 7](#_Toc467484150)

[2.2 关闭Selinux和THP 8](#_Toc467484151)

[2.2.1 关闭Selinux 8](#_Toc467484152)

[2.2.2 关闭THP 8](#_Toc467484153)

[三、Ambari安装 8](#_Toc467484154)

[四、部署大数据平台 11](#_Toc467484155)

**说明：**

**红色字体： 提醒或注意**

**绿色字体： 注释**

**蓝色字体： linux或hadoop命令**

# 一、背景

虽然大数据越来越流行，但其学习的门槛却一直阻碍着很多的分布式应用初学者或者大数据的业务应用开发者。多个产品之间的不兼容问题，快速集成和维护也显得比较困难。不管是 Hadoop V1 或者 V2 的安装，又或者 Spark/YARN 等的集成，都不是几行简单的命令可以完成的，而是需要手工修改很多的集群配置，这进一步增加了业务开发者的学习和使用难度。有了 Ambari，这些都不再是难题。

# 二、安装前准备

先明确几个概念：

1. Ambari只能安装Hortonworks Data Platform，即Hortonworks的开源Hadoop，不支持Apach的Hadoop平台；

2. 对于已经安装了Apach Hadoop或者其他Hadoop平台的，不能使用Ambari来管理；

再说几个注意事项：

1. Ambari默认的安装方式是使用yum，从远程下载HDP组件安装，而HDP平台安装包都非常大（本例中使用的HDP-2.4.0安装包为6G），Ambari平台有又30分钟的Timeout限制，如果在30分钟内下载不完HDP，就会造成安装失败。建议修改下载源文件，配置为本地源；

2. Ambari安装过程为自动安装，自动安装脚本会创建很多用户和其他组件（如系统自带的Java和数据库），建议使用一套干净的环境来安装。

3. 如第2点所示，脚本会自动创建用户和安装组件，建议使用root用户；

4. 系统请关闭Selinux、防火墙和THP；

5. 集群机器请事先配置ssh互信；还是建议使用root用户来互信；

# 三、安装

操作系统：CentOS 6.5

集群机器情况：

|  |  |  |  |
| --- | --- | --- | --- |
| **机器名** | **IP** | **功能** | **备注** |
| Master | 192.168.101.41 | NameNode and JobTracker | 以下可能简称为主机 |
| Master2 | 192.168.101.42 | Second Namenode | 以下可能简称为节点 |
| Slave1 | 192.168.101.43 | DataNode and TaskTracker | 以下可能简称为节点 |
| Slave2 | 192.168.101.44 | DataNode and TaskTracker | 以下可能简称为节点 |

## 2.1 配置本地源

上文提到，为了加快安装速度和防止超时错误，建议为HDP配置本地源，请在事先在网上下载HDP、HDP-UTILS和Ambari，本例中几个组件版本为HDP-2.4.0，HDP-UTILS-1.1.0.20和Ambari-2.2.1.0。我们把master服务器做为源服务器。以下操作在master机器运行。

### 2.1.1 安装配置本地源需要的组件

yum install httpd

yum install yum-utils

yum repolist

yum install createrepo

### 2.1.2 安装httpd服务

1. 直接使用命令：**yum install httpd**；安装完成后，会生成 /var/www/html 目录。

2. 在/var/www/html目录下，分别建立ambari和hdp目录

**cd /var/www/html**

**mkdir ambari**

**mkdir hdp**

3. 启动httpd服务

**Service httpd start**

4. 设置httpd服务开机自动启动

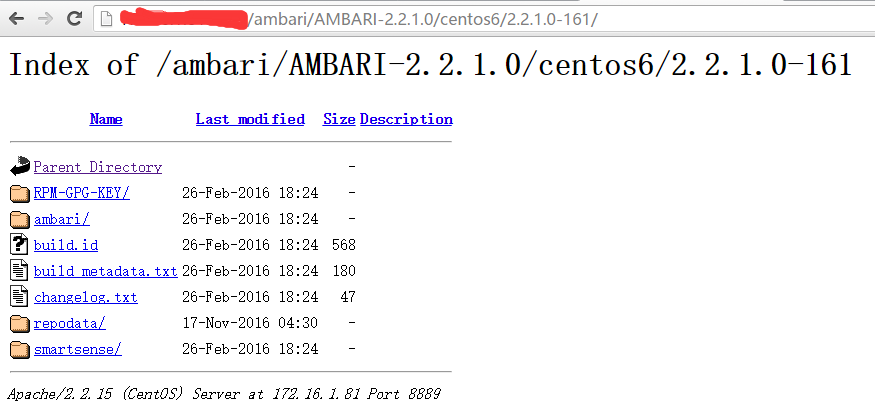
**Chkconfig httpd on**

### 2.1.3 下载、配置Ambari本地源ambari.repo

1. 把下载的Ambari tar包解压后拷贝到刚才建立的/var/www/html/ambary/目录中，在浏览器中输入地址，就可以看到ambari目录结构

http://master:8889/ambari/AMBARI-2.2.1.0/centos6/2.2.1.0-161/

域名后8889为httpd服务配置的端口



这个地址，就是本地源的地址

2. 使用wget命令：

**wget**[**http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.4.1.25/ambari.repo**](http://public-repo-1.hortonworks.com/ambari/centos6/1.x/updates/1.4.1.25/ambari.repo)

命令完成后，wget命令会默认下载到本目录，把下载后的ambari.repo文件拷贝到/etc/ yum.repos.d目录下。

修改ambari.repo文件

[Updates-ambari-2.0.1]

name=ambari-2.2.1 - Updates

baseurl=http://master:8889/ambari/AMBARI-2.2.1.0/centos6/2.2.1.0-161/

gpgcheck=0

enabled=1

priority=1

红色字体为修改部分，把baseurl换成本地的url。

### 2.1.2 配置HDP本地源

把下载的HDP-2.4.0.0-centos6-rpm.tar.gz包拷贝到/var/www/html/hdp/目录下，用浏览器输入网址http://master:8889/hdp/HDP-2.4.0/centos6/2.x/updates/2.4.0.0/ ，查看HDP目录结构



在/etc/yum.repo.d目录中，修改HDP.repo文件

[HDP-2.4]

name=HDP-2.4

baseurl=http://master:8889/hdp/HDP-2.4.0/centos6/2.x/updates/2.4.0.0/

path=/

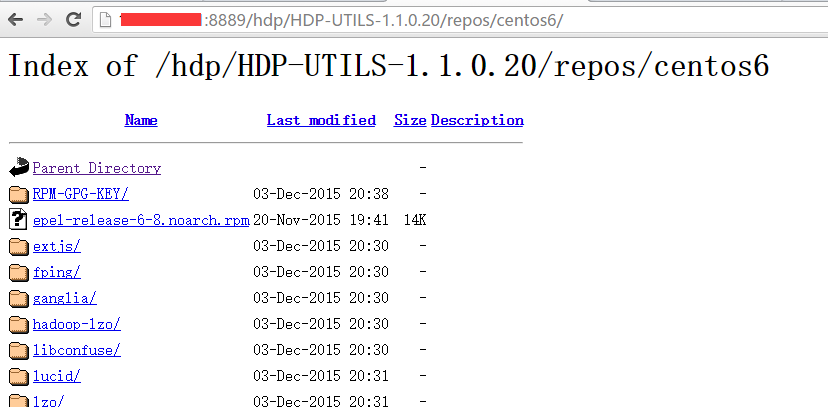
enabled=1

gpgcheck=0

红色字体为修改部分，把baseurl换成本地的url。

### 2.1.3 配置HDP-UTILS

同样，把下载的HDP-UTILS-1.1.0.20-centos6.tar.gz包拷贝到/var/www/html/hdp/目录下，用浏览器输入网址http://master:8889/hdp/HDP-UTILS-1.1.0.20/repos/centos6/ ，查看HDP-UTILS目录结构



在/etc/yum.repo.d目录中，修改HDP.repo文件

[HDP-UTILS-1.1.0.20]

name=HDP-UTILS-1.1.0.20

baseurl=http://master:8889/hdp/HDP-UTILS-1.1.0.20/repos/centos6/

path=/

enabled=1

gpgcheck=0

红色字体为修改部分，把baseurl换成本地的url。

### 2.1.4 生成本地源

使用createrepo命令，生成本地源

**createrepo /var/www/html/hdp/HDP-2.4.0/centos6/2.x/updates/2.4.0.0**

**createrepo /var/www/html/hdp/HDP-UTILS-1.1.0.20/repos/centos6**

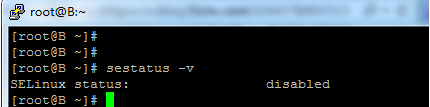
**createrepo /var/www/html/ambari/AMBARI-2.2.1.0/centos6/2.2.1.0-161**

## 2.2 关闭Selinux和THP

### 2.2.1 关闭Selinux

注意，在集群的每个节点，都要关闭Selinux。

使用**sestatus -v**命令，查看Selinux状态。



如果不是disable状态，编辑/etc/sysconfig/selinux 文件

**vi /etc/sysconfig/selinux**

把里边的一行改为

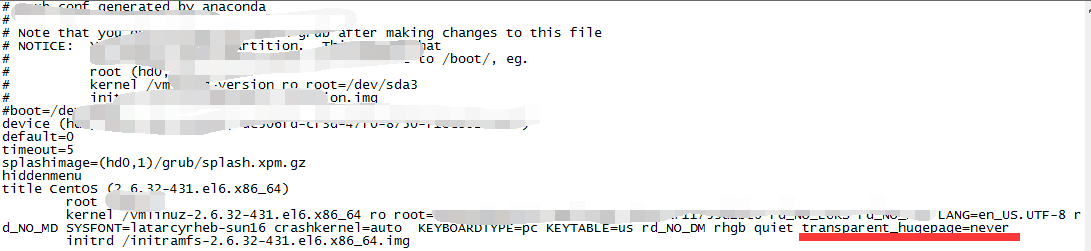
SELINUX=disabled

保存，然后重启机器。

### 2.2.2 关闭THP

在集群的每个节点，都要关闭。编辑/etc/grub.conf文件，在kernel 行后面加入

transparent\_hugepage=never，保存退出，重启机器。



# 三、Ambari安装

使用命令查看案例列表：

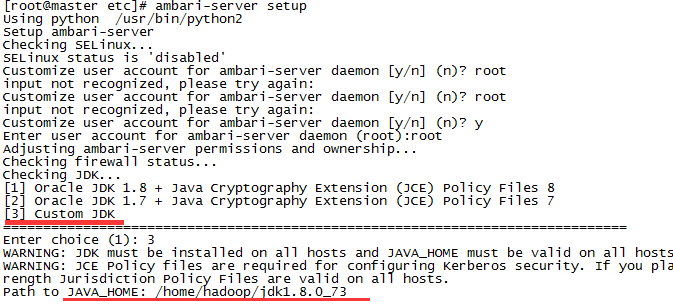
**yum clean all**

**yum list|grep ambary**

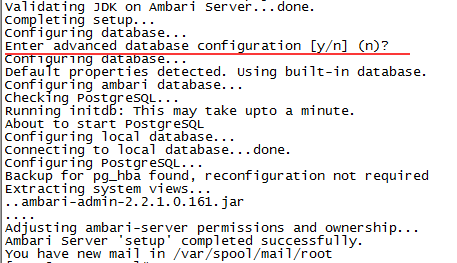
直接使用命令**yum install ambari-server**安装即可，由于配置了本地源，安装过程非常快，可比网络资源节省一半以上的时间。

安装完成后，使用命令， **amari-server setup** 设置Ambari，基本都可以一路回车使用默认设置。为了演示完整的安装过程，本例中做了一些特殊设置。请同学们根据环境的实际情况自行选择配置。

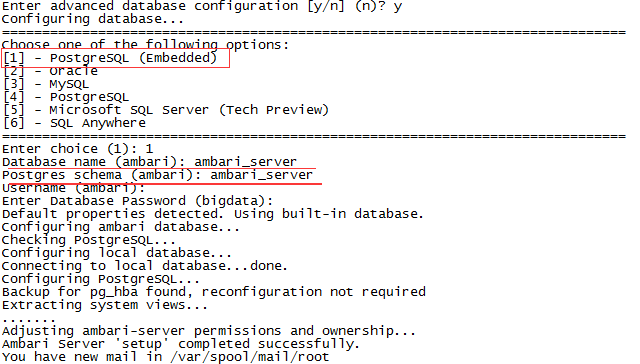
1.设置JDK



2.设置数据库，Ambari默认使用的是PostgreSQL，也可指定其他数据库



为了达到演示效果，本例就不使用其他数据库了，简单改变PostgreSQL中的几项设置即可。

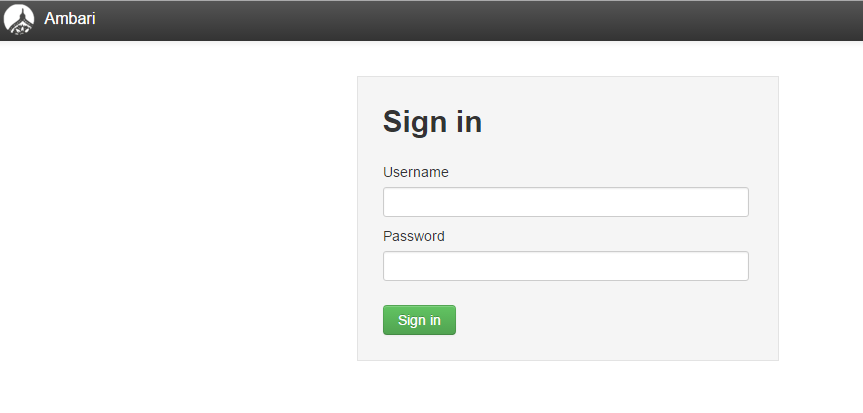


Ambari默认使用的是8080端口，如果端口被占用，可修改配置文件/etc/ambari-server/conf/ambari.properties，在文件中增加 client.api.port=<port\_number>

配置完成后，使用命令 **ambari-server start** 启动Ambari。

启动成功后，在浏览器输入网址<http://master:8080> ,看到如下界面，就说明安装成功了。

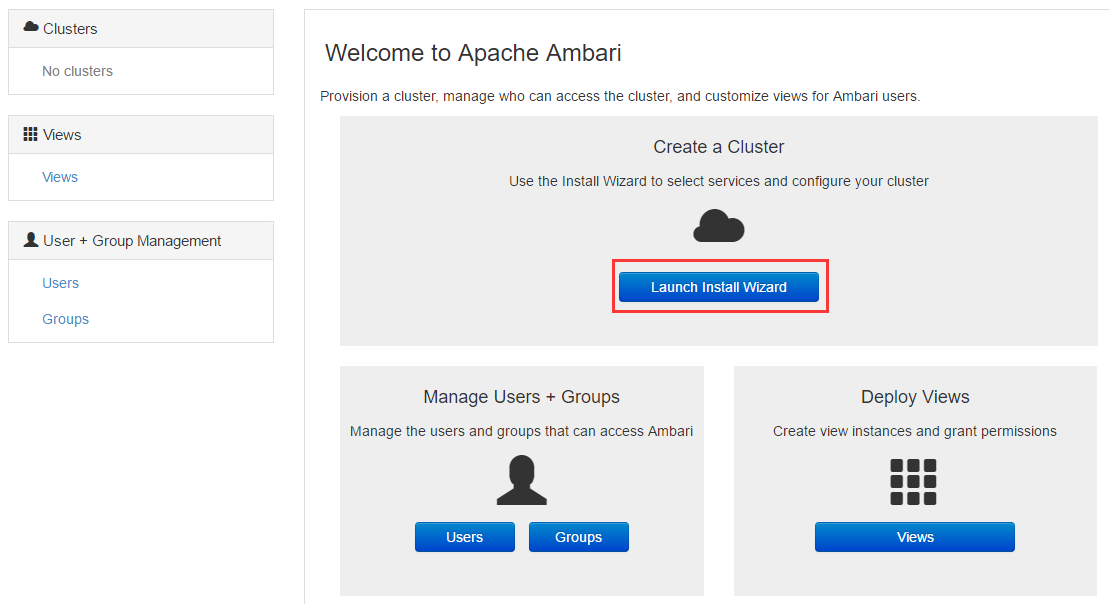
Ambari默认用户名/密码是：admin/admin



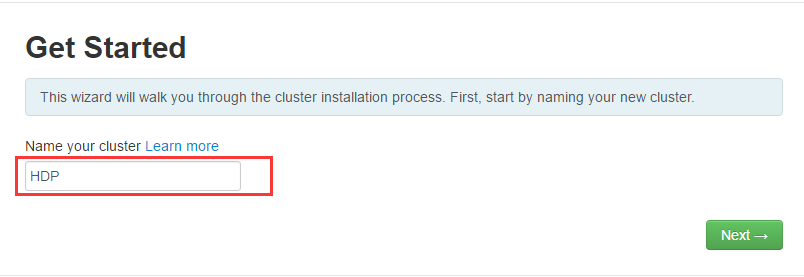
# 四、部署大数据平台

图文并茂

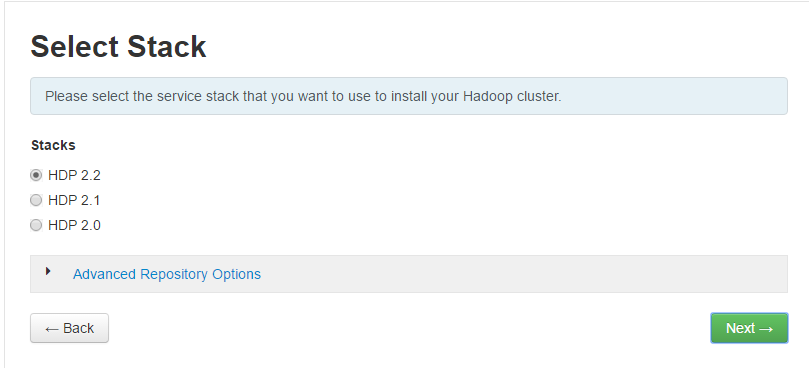
步骤一，点击运行安装向导



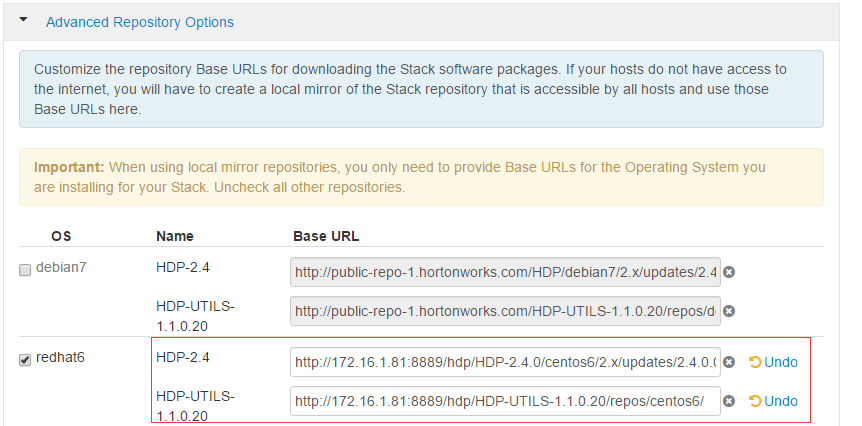
步骤二，输入你的集群名称



步骤三，选择HDP版本

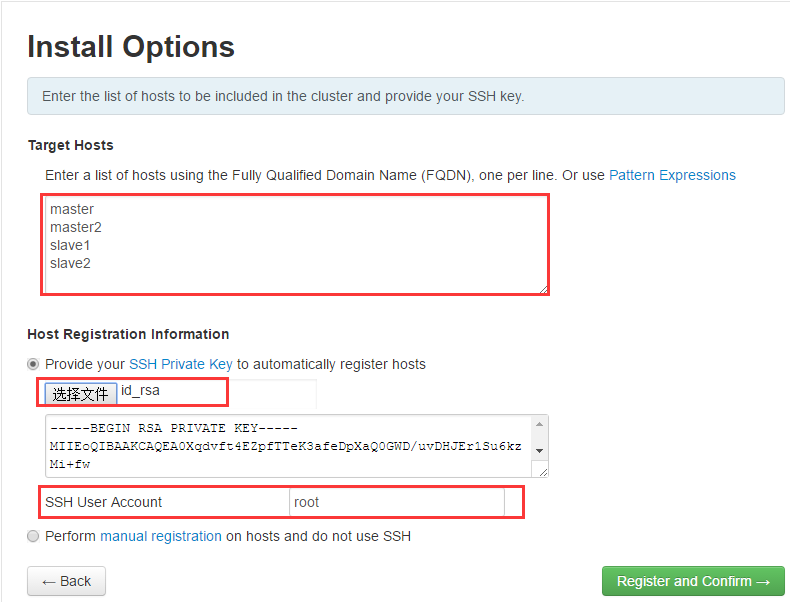


**注意： HDP资源默认是从网络下载，请点击下方的Advance Repository Options，把HDP和HDP-UTILS地址替换成我们上方配置的本地源，否则容易发生超时错误。**

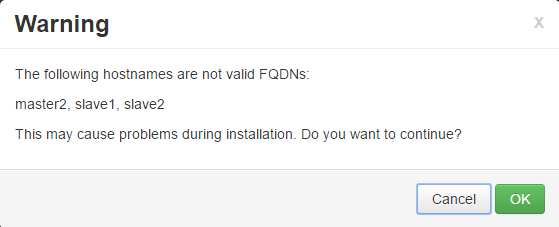


步骤四，输入安装选项，数据集群的机器host name，ssh互信的私钥和ssh用户；

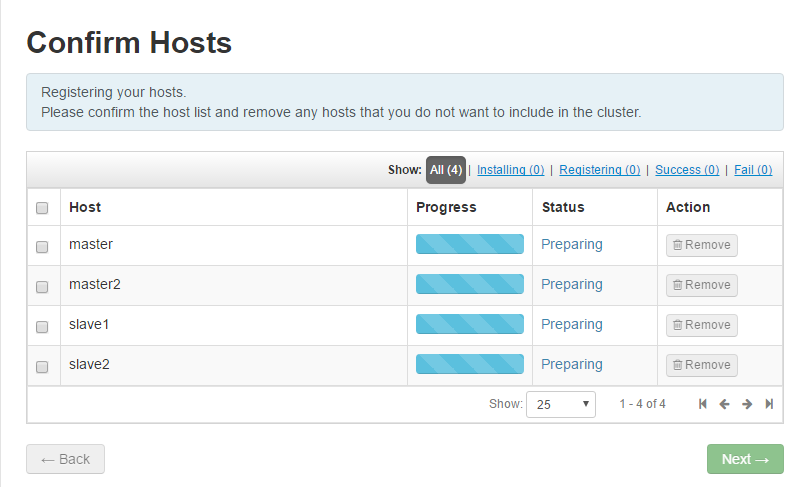
**请注意：在ssh user accout的地方，为了安装成功，我使用的是root用户**



由于我的机器不在域中，系统会提示不是一个完整的FQDN(完整的FQDN类似于master.example.com这种格式)，不用管他，直接点OK按钮。



步骤五，选择完机器后，Ambari会自动在其他的机器上安装ambari-agent代理



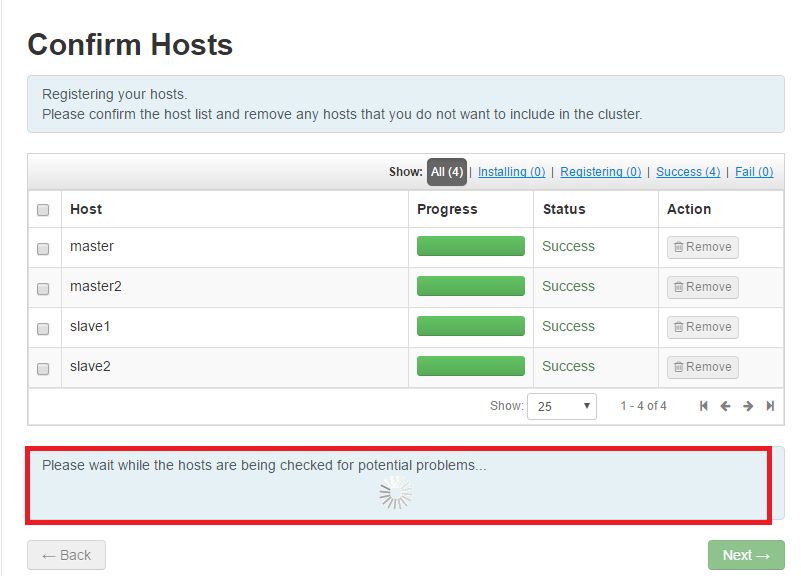
**可能会出现的问题：**

**ambari-agent注册失败，日志显示端口被占用，网上查询结果是Agent使用的8670端口**

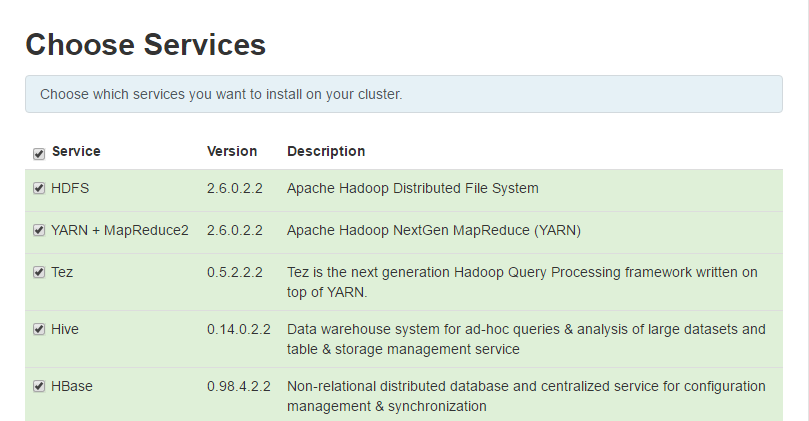
**手工连接到失败的服务器上，netstat -anp|grep 8670，发现果然在监听状态，然后输入命令kill -9 [进程id]，再Retry便可成功**

安装成功后，界面如下：

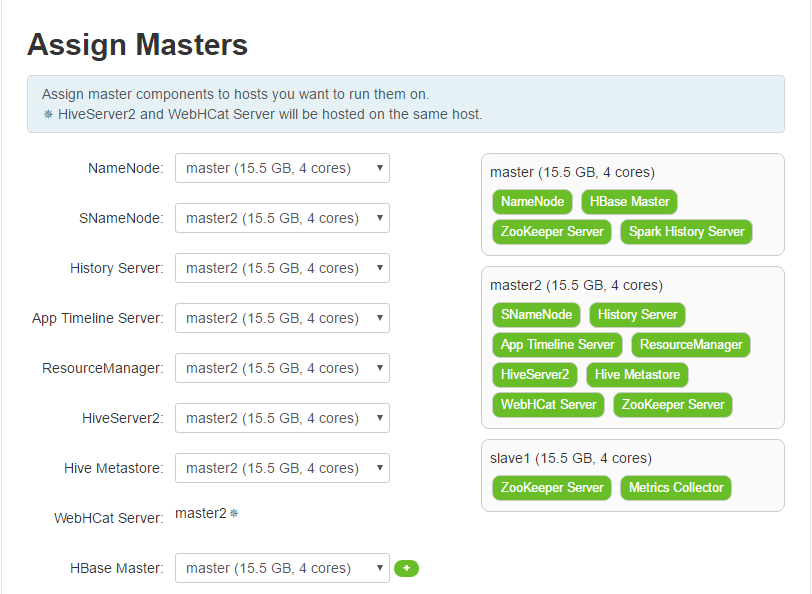
**请注意：画面下方的potential problems，这里系统在坚持潜在的问题，完成后，会有显示当前问题列表，检查完成后，系统会给出问题解决方法，请把列出的警告的problems也一一解决掉，否则安装可能会不成功**



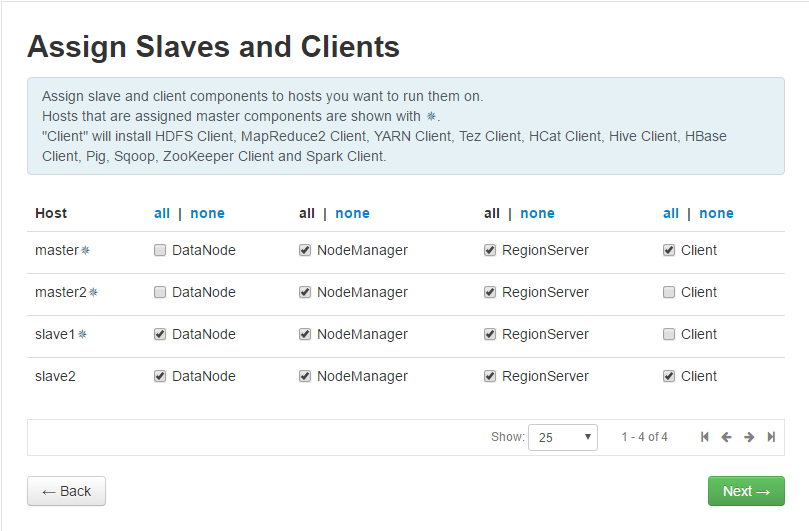
步骤六，选择需要安装的服务组件，已经包括了HDFS，Yarn，Hbase，Hive，Spark等等，的确非常方便，不用在一一部署了



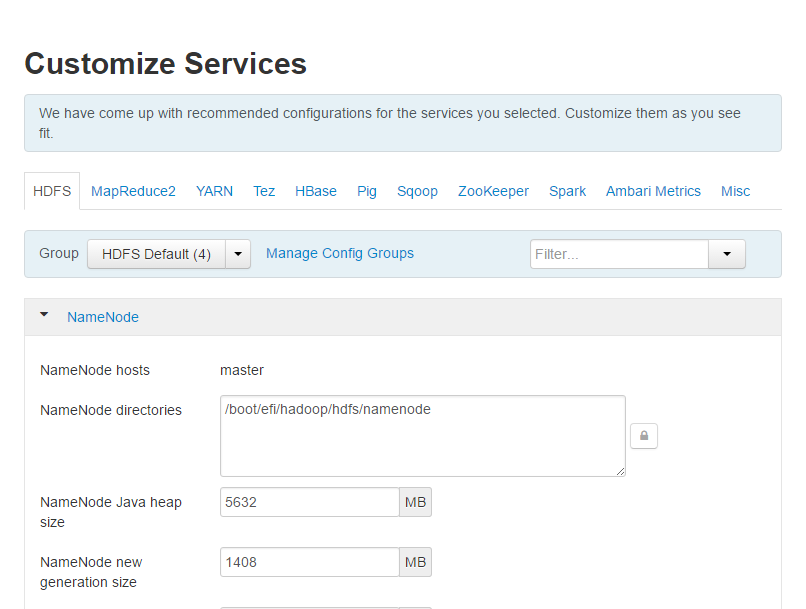
步骤七，分配Master服务器，可以看到能做NameNode和2nd NameNode高可用，Hbase也是



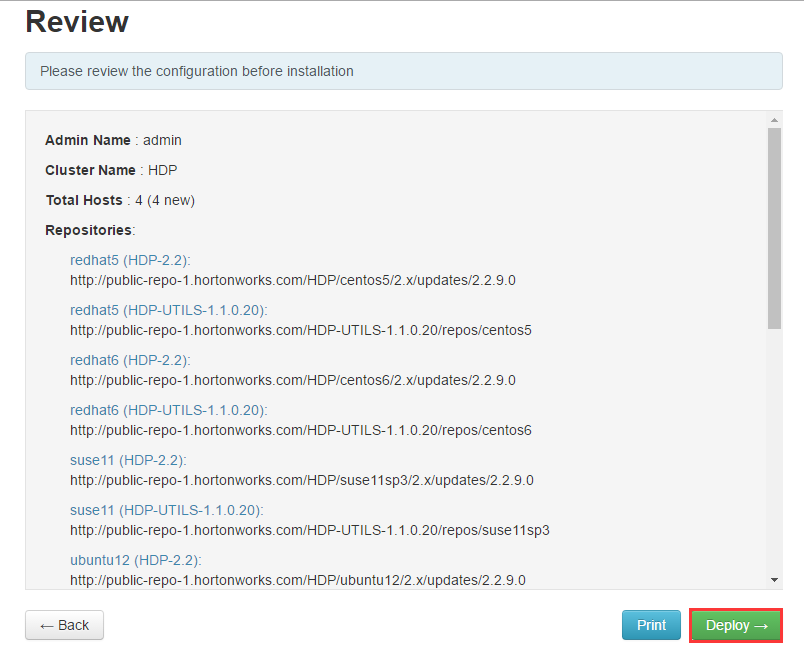
步骤八，选择Slave节点，Client随便选一个Slave安装就可以了。



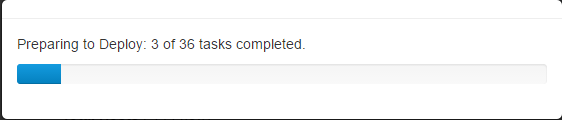
步骤九，配置文件，可图形化配置core-site.xml，hdfs-site.xml等文件，有时候系统出现标记，如Hive组件，系统会要求你输入元数据库的用户名，密码和地址，输入完成后，点击下一步



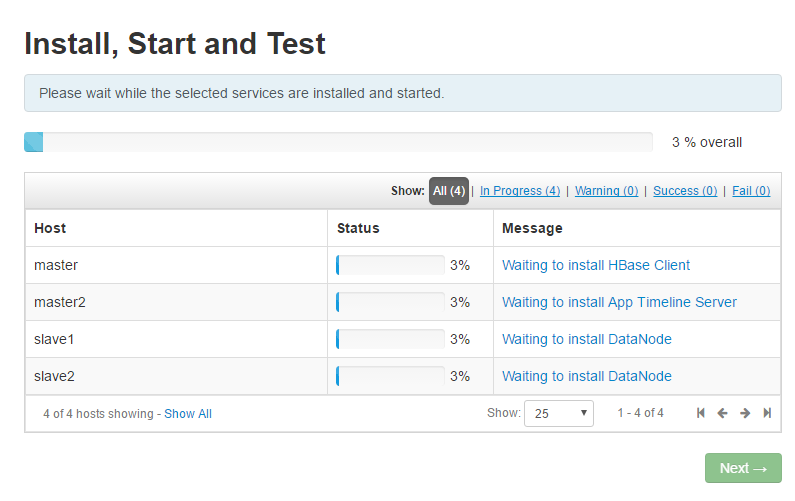
点击下一步后，检查配置



步骤十，系统根据配置，开始部署



等待安装完成，因为我们配置了本地源，安装非常迅速



**可能会出现的问题：**

**Too many levels of symbolic links**

安装HDFS和HBASE的时候出现/usr/hdp/current/hadoop-client/conf doesn't exist

是由于/etc/hadoop/conf和/usr/hdp/current/hadoop-client/conf目录互相链接，造成死循环，所以要改变一个的链接

需要在失败的节点运营命令：

**cd /etc/hadoop**

**rm -rf conf**

**ln -s /etc/hadoop/conf.backup /etc/hadoop/conf**

**HBASE也会遇到同样的问题，解决方式同上**

**cd /etc/hbase**

**rm -rf conf**

**ln -s /etc/hbase/conf.backup /etc/hbase/conf**

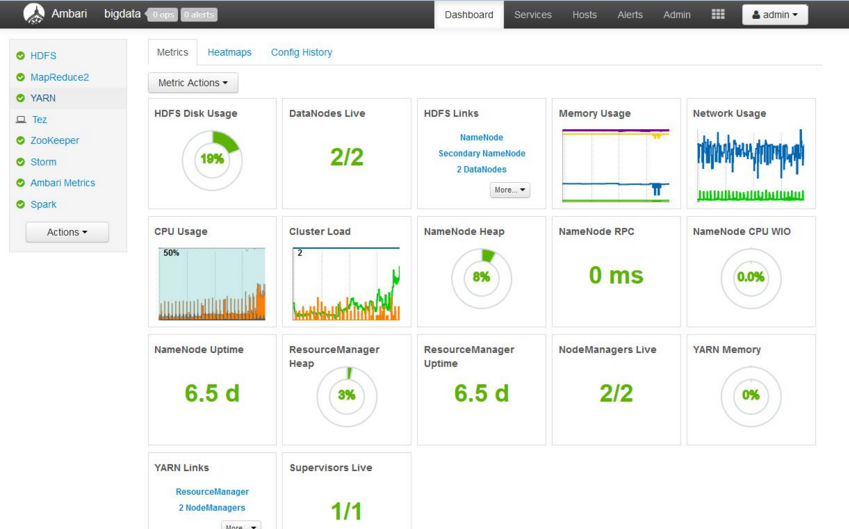
**ZooKeeper也会遇到同样的问题，解决方式同上**

**cd /etc/zookeeper**

**rm -rf conf**

**ln -s /etc/zookeeper/conf.backup /etc/zookeeper/conf**

步骤十一、安装完成后，系统会自动系统大数据平台



大功告成。