实训报告

小组人员

组长：陈泳桥 学号：20100120020012

姓名：何金蔚 学号：20100120020028

姓名：曹移 学号：20100120020011

一.环境准备

1.1安装jdk

解压jdk-8u112-linux-x64.tar.gz

```
tar -xvf jdk-8u112-linux-x64.tar.gz -C /opt
ln -s /opt/jdk1.8.0_271 /opt/jdk
```

修改环境变量：

```
vim /etc/profile
```

```
export JAVA_HOME=/opt/jdk
export PATH=${JAVA_HOME}/bin:$PATH
```

执行生效命令：

```
source /etc/profile
```

测试：

```
java -version
```

2.2关闭防火墙

```
systemctl stop firewalld.service #关闭防火墙
systemctl disable firewalld #关闭开机启动
systemctl status firewalld #查看防火墙状态
```

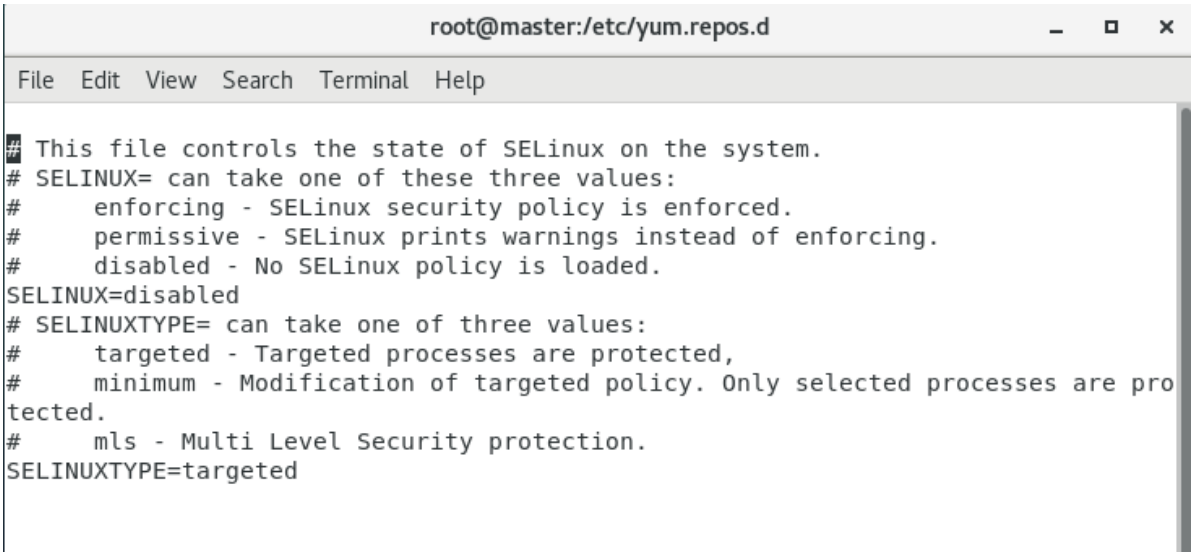
2.3关闭SELinux、PackageKit、检测umask值

使用命令关闭SELinux：

```
setenforce 0
```

修改配置文件进行配置： vim /etc/selinux/config

```
# This file controls the state of SELinux on the system.
# SELINUX= can take one of these three values:
#     enforcing - SELinux security policy is enforced.
#     permissive - SELinux prints warnings instead of enforcing.
#     disabled - No SELinux policy is loaded.
SELINUX=disabled
# SELINUXTYPE= can take one of three values:
#     targeted - Targeted processes are protected,
#     minimum - Modification of targeted policy. Only selected processes are
protected.
#     mls - Multi Level Security protection.
SELINUXTYPE=targeted
```

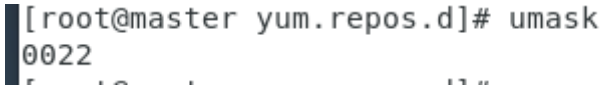


如果系统有安装 PackageKit，需要修改配置文件：/etc/yum/pluginconf.d/refresh-packagekit.conf

```
enabled=0
```

确保umask值为0022

```
#查看
umask
#修改为0022
umask 0022
```



针对所有交互用户进行永久性修改：

```
echo umask 0022 >> /etc/profile
```

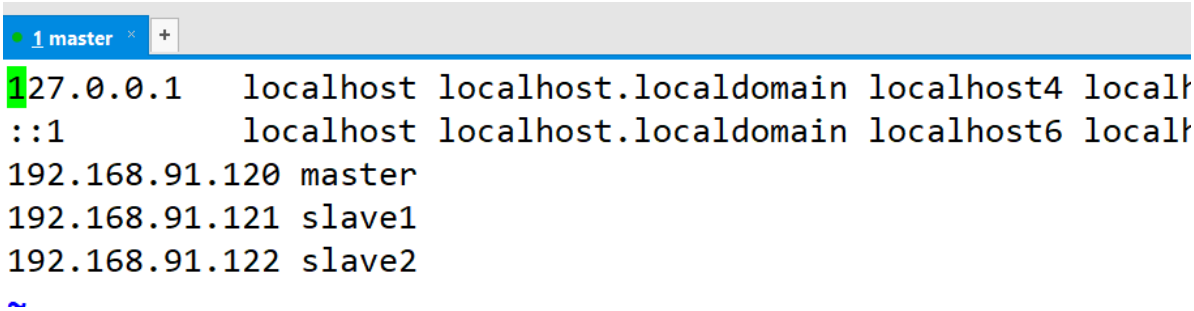
2.4最大文件描述

二.集群准备

3.1hosts配置

在主机master上修改/etc/hosts文件

```
vim /etc/hosts
```



使用scp同步到salve1，slave2

```
scp /etc/hosts root@slave1:/etc/
scp /etc/hosts root@slave2:/etc/
```

3.2同步时钟

配置/etc/ntp.conf

```
vi /etc/ntp.conf
```

```
systemctl start ntpd.service #启动服务
systemctl enable ntpd.service #开机启动
```

在slave1和slave2上运行：

```
ntpdate master
```

将时间同步设定为定时任务：

```
crontab -e
```

```
*/10 * * * * /usr/sbin/ntpdate master
```

3.3免密登录

配置master节点无密码登录到其他节点

```
ssh-keygen -t rsa #一直Enter键
ssh-copy-id master
ssh-copy-id slave1
ssh-copy-id slave2
```

测试：

```
[root@master Desktop]# ssh slave1 date; ssh slave2 date; ssh master date;
Wed Jun  1 04:05:48 PDT 2022
Wed Jun  1 04:05:48 PDT 2022
Wed Jun  1 19:05:50 CST 2022
[root@master Desktop]#
```

三.创建本地源

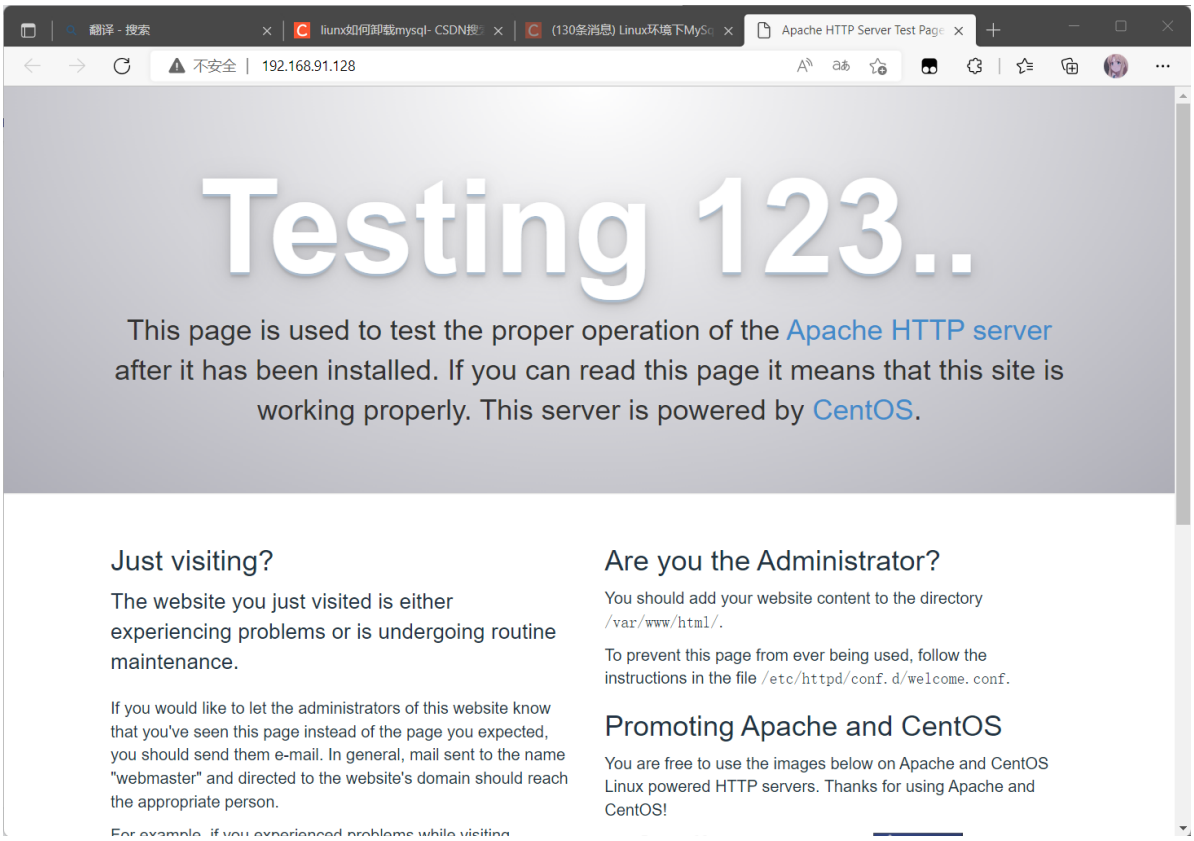
4.1安装httpd服务

```
yum -y install httpd #安装httpd
systemctl restart httpd.service #开启httpd服务
systemctl enable httpd.service #开机启动
```

4.2将包放到/var/www/html目录

```
tar -zxvf /root/ambari-2.6.0.0-centos7.tar.gz -C /var/www/html/
tar -zxvf /root/HDP-2.6.3.0-centos7-rpm.tar.gz -C /var/www/html/
mkdir /var/www/html/HDP-UTILS
tar -zxvf /root/HDP-UTILS-1.1.0.21-centos7.tar.gz -C /var/www/html/HDP-UTILS
```

测试：



4.3制作本地源

3.1安装本地源制作相关工具

```
yum install yum-utils createrepo yum-plugin-priorities repolist -y createrepo
/var/www/html
```

3.2修改源地址

```
vi /var/www/html/ambari/centos7/2.6.0.0-267/ambari.repo
```

修改内容:

```
#VERSION_NUMBER=2.6.0.0-267
[ambari-2.6.0.0]
name=ambari Version - ambari-2.6.0.0
baseurl=http://master/ambari/centos7/2.6.0.0-267
gpgcheck=1
gpgkey=http://master/ambari/centos7/2.6.0.0-267/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

拷贝文件:

```
cp /var/www/html/ambari/centos7/2.6.0.0-267/ambari.repo /etc/yum.repos.d/
```

修改内容:

```
#VERSION_NUMBER=2.6.3.0-235
[HDP-2.6.3.0]
name=HDP Version - HDP-2.6.3.0
baseurl=http://master/HDP/centos7/2.6.3.0-235
gpgcheck=1
gpgkey=http://master/HDP/centos7/2.6.3.0-235/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1

[HDP-UTILS-1.1.0.21]
name=HDP-UTILS Version - HDP-UTILS-1.1.0.21
baseurl=http://master/HDP-UTILS
gpgcheck=1
gpgkey=http://master/HDP-UTILS/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

拷贝文件:

```
cp /var/www/html/HDP/centos7/2.6.3.0-235/hdp.repo /etc/yum.repos.d/
```

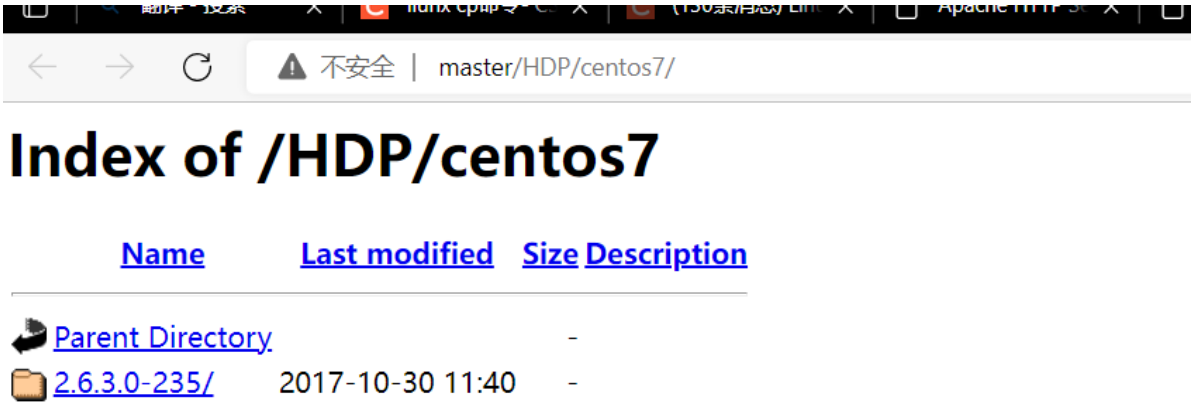
清除yum缓存:

```
yum clean all
yum makecache
yum repolist
```

测试：

http://master/ambari/centos7/

http://master/HDP/centos7/



拷贝源文件到子节点：

```
cd /etc/yum.repos.d
scp ambari.repo slave1:/etc/yum.repos.d/ambari.repo
scp ambari.repo slave2:/etc/yum.repos.d/ambari.repo
scp hdp.repo slave1:/etc/yum.repos.d/hdp.repo
scp hdp.repo slave2:/etc/yum.repos.d/hdp.repo
```

四.安装Ambari

4.1安装Ambari服务

```
yum -y install ambari-server
```

4.2配置Ambari

```
ambari-server setup
```

4.2使用MySQL作为元数据库

4.2.1卸载MariaDB已有MySQL

```
rpm -qa | grep -i mariadb
```

```
[root@master yum.repos.d]# rpm -qa |grep -i mariadb
mariadb-libs-5.5.68-1.el7.x86_64
[root@master yum.repos.d]#
```

```
rpm -e --nodeps mariadb-libs-5.5.84-1.el7.x86_64
```

删除服务：

```
systemctl list-unit-files --type=service | grep -i mysql
```

```
chkconfig --del mysql
```

4.2.2安装MySQL数据库

1.离线安装：

安装依赖

```
yum -y install perl autoconf
```

解压并安装

```
tar -xvf MySQL-5.6.46-1.el7.x86_64.rpm-bundle.tar
rpm -ivh MySQL-client-5.6.46-1.el7.x86_64.rpm
rpm -ivh MySQL-server-5.6.46-1.el7.x86_64.rpm
#查看安装时产生的随机密码
cat /root/.mysql_secret #启动mysql
service mysql start mysql -uroot -p #登入mysql
Enter password: #输入刚才查看的随机密码
mysql> set password for 'root'@'localhost' = password('123456'); #修改密码
mysql> grant all privileges on *.* to 'root'@'%' identified by '123456';
mysql> flush privileges; #刷新
mysql> exit
#MySQL服务自启动
chkconfig mysql on
```

2.MySQL中user表中主机的配置

```
mysql -uroot -p123456
mysql> show databases; #显示数据库
mysql> use mysql; #使用mysql数据库
mysql> show tables; #显示数据库mysql中的所有表
mysql> desc user; #显示user表的结构
mysql> select User,Host,Password from user; #查看User表
mysql> update user set host='%' where host='localhost'; #修改 user 表，把 Host 表内
容修改为% # 删除 root 用户的其他 host
mysql> delete from user where Host='master';
delete from user where Host='127.0.0.1';
delete from user where Host='::1';
mysql> flush privileges; #刷新
mysql> quit;
```

3.配置MySQL驱动

```
mkdir /usr/share/java  ##创建配置目录文件夹
##拷贝文件
cp /root/mysql-connector-java-5.1.40.jar /usr/share/java/mysql-connector-
java.jar
```

修改ambari.properties，添加mysql驱动路径

```
vi /etc/ambari-server/conf/ambari.properties
```

在最后面添加以下内容

```
server.jdbc.driver.path=/usr/share/java/mysql-connector-java.jar
```

4.在MySQL中创建数据库

```
mysql -uroot -p123456
mysql> CREATE DATABASE ambari;
mysql> use ambari;
mysql> CREATE USER 'ambari'@'%' IDENTIFIED BY 'ambarizk123';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'ambari'@'%';
mysql> CREATE USER 'ambari'@'localhost' IDENTIFIED BY 'ambarizk123';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'ambari'@'localhost';
mysql> CREATE USER 'ambari'@'master' IDENTIFIED BY 'ambarizk123';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'ambari'@'master';
mysql> FLUSH PRIVILEGES;
mysql> source /var/lib/ambari-server/resources/Ambari-DDL-MySQL-CREATE.sql
mysql> show tables;
mysql> use mysql;
mysql> select Host,User,Password from user where user='ambari';
mysql> CREATE DATABASE hive;
mysql> use hive;
mysql> CREATE USER 'hive'@'%' IDENTIFIED BY 'hive';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hive'@'%';
mysql> CREATE USER 'hive'@'localhost' IDENTIFIED BY 'hive';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hive'@'localhost';
mysql> CREATE USER 'hive'@'master' IDENTIFIED BY 'hive';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'hive'@'master';
```

```
mysql> FLUSH PRIVILEGES;
mysql> CREATE DATABASE oozie;
mysql> use oozie;
mysql> CREATE USER 'oozie'@'%' IDENTIFIED BY 'oozie';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'oozie'@'%';
mysql> CREATE USER 'oozie'@'localhost' IDENTIFIED BY 'oozie';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'oozie'@'localhost';
mysql> CREATE USER 'oozie'@'master' IDENTIFIED BY 'oozie';
mysql> GRANT ALL PRIVILEGES ON *.* TO 'oozie'@'master ';
mysql> FLUSH PRIVILEGES;
```

5.配置Ambari

执行服务

```
ambari-server setup
```

配置执行流程：

(1)提示是否自定义设置输入 y

```
Customize user account for ambari-server daemon [y/n] (n)? y
```

(2) ambari-server 账号, 如果直接回车就是默认选择 root 用户

```
Enter user account for ambari-server daemon (root):
```

(3)设置jdk，输入3

```
Checking JDK...
Do you want to change Oracle JDK [y/n] (n)? y
[1] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8
[2] Oracle JDK 1.7 + Java Cryptography Extension (JCE) Policy Files 7
[3] Custom JDK
==============================================================================
Enter choice (1): 3
```

如果上面选择3自定义jdk，则需要设置JAVA_HOME, 输入：/opt/jdk

```
WARNING: JDK must be installed on all hosts and JAVA_HOME must be valid on all
hosts.
WARNING: JCE Policy files are required for configuring Kerberos security. If you
plan to use
Kerberos,please make sure JCE Unlimited Strength Jurisdiction Policy Files are
valid on all
hosts.
Path to JAVA_HOME: /opt/jdk Validating JDK on Ambari Server...done.
Completing setup...
```

(4)数据库配置，选择y

```
Configuring database...
Enter advanced database configuration [y/n] (n)? y
```

(5)选择数据库类型，输入：3

```
Configuring database...
========================================================================= ============
Choose one of the following options:
[1] - PostgreSQL (Embedded)
[2] - Oracle
[3] - MySQL
[4] - PostgreSQL
[5] - Microsoft SQL Server (Tech Preview)
[6] - SQL Anywhere
========================================================================= ============
Enter choice (3): 3
```

(6)设置数据库的具体配置信息，根据实际情况输入，如果和括号内相同，则可以直接回车。

如果想重命名，就输入。

```
Hostname (localhost): master
Port (3306):
Database name :ambari
Username :ambari
Enter Database Password (bigdata):ambarizk123 (这里输入时不会显示) Re-enter
password:ambarizk123 (这里输入时不会显示)
```

(7)将Ambari数据库脚本导入到数据库

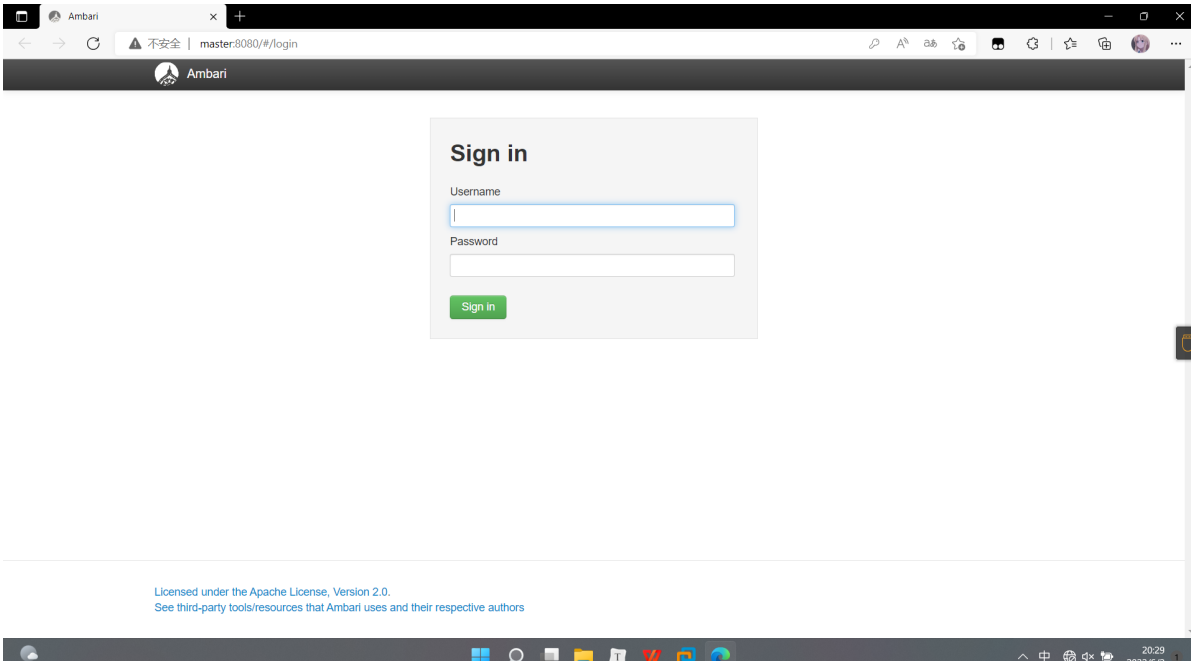
```
WARNING: Before starting Ambari Server, you must run the following DDL against
the
database to create the schema: /var/lib/ambari-server/resources/Ambari-DDL-
MySQL- CREATE.sql Proceed with configuring remote database connection properties
[y/n] (y)? y
```

4.3启动Ambari

```
ambari-server start
```

测试：： http://master:8080 (需要配置hosts, 没配置时可以直接访

问 192.168.91.120:8080)

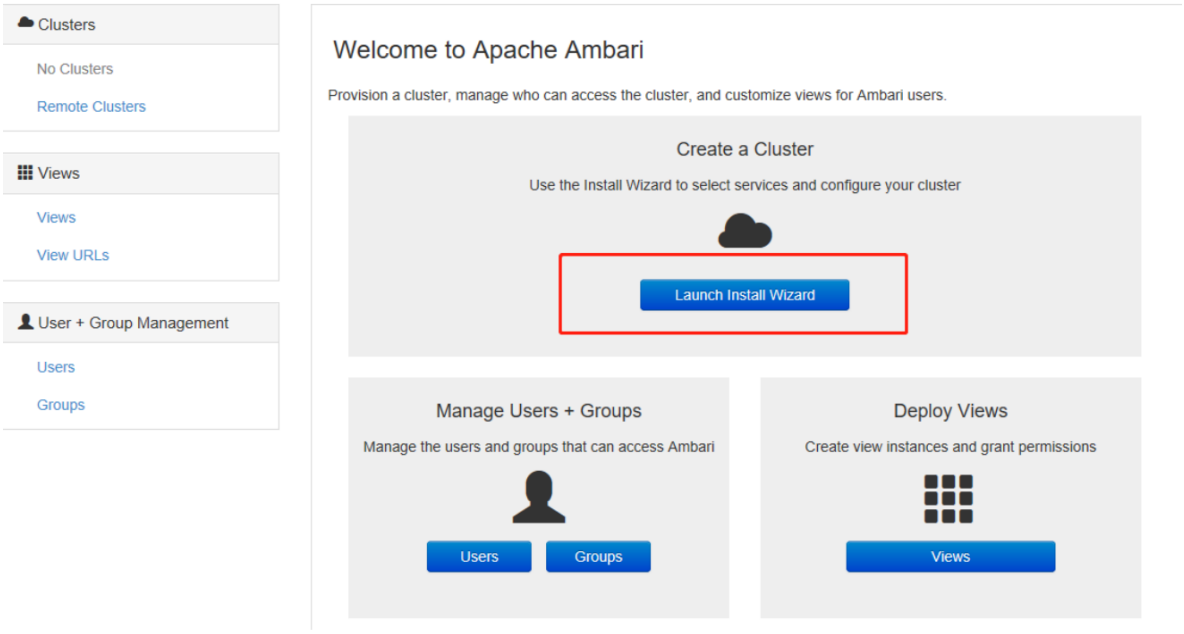


五.搭建集群

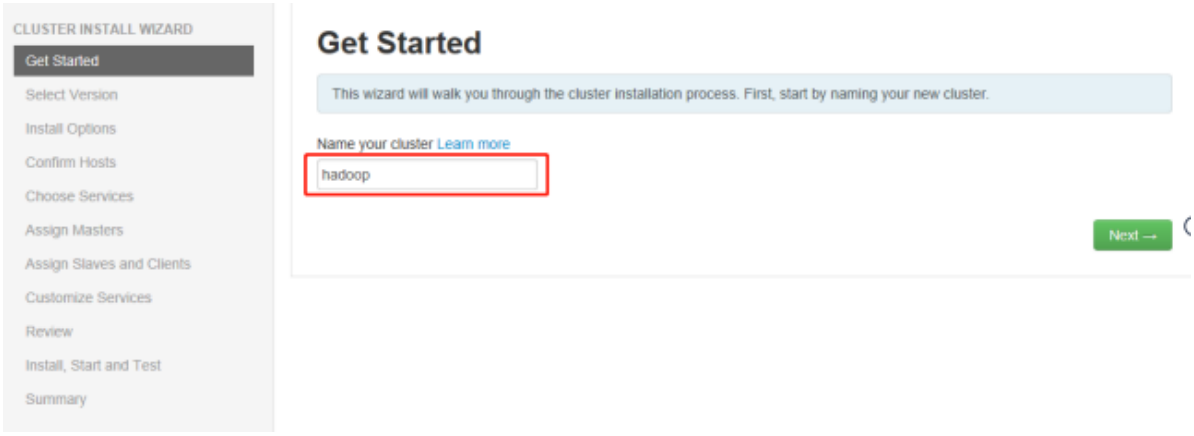
1.登录

登录界面，默认管理员账户登录http://master:8080/ (需要配置hosts, 没配置时可以访问192.168.91.128:8080)， 账户：admin 密码：admin

2.安装向导



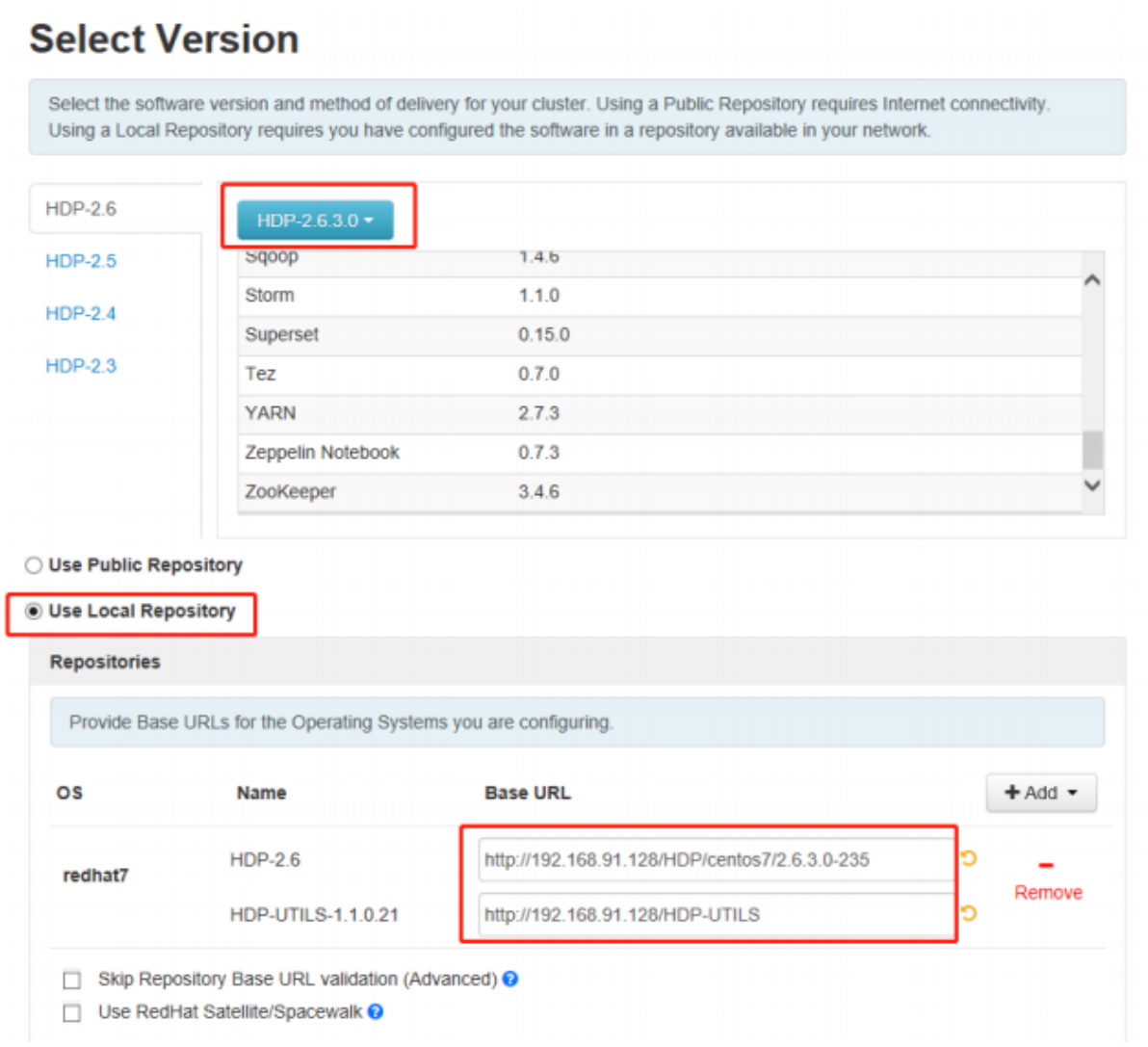
2.1配置集群的名字为Hadoop



2.2选择版本和本地库

在redhat7后面填写：

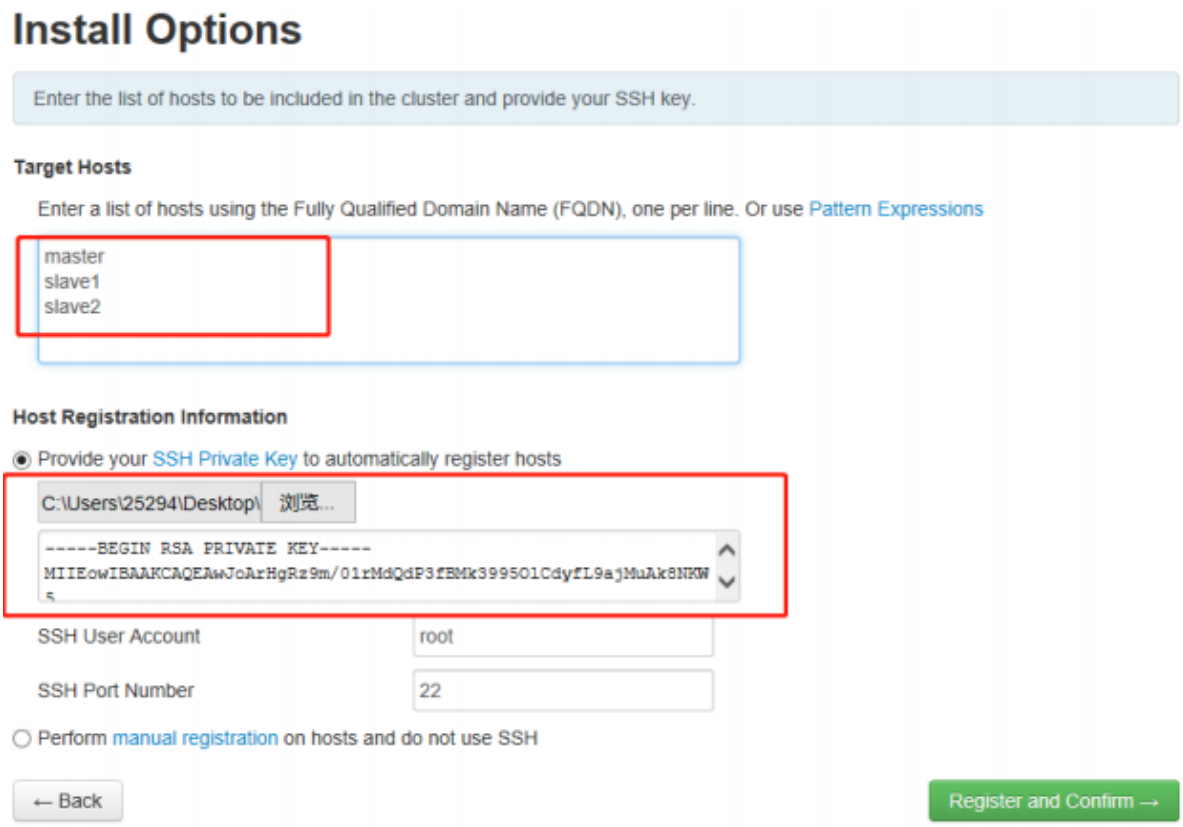
```
http://192.168.91.128/HDP/centos7/2.6.3.0-235 http://192.168.91.128/HDP-UTILS
```



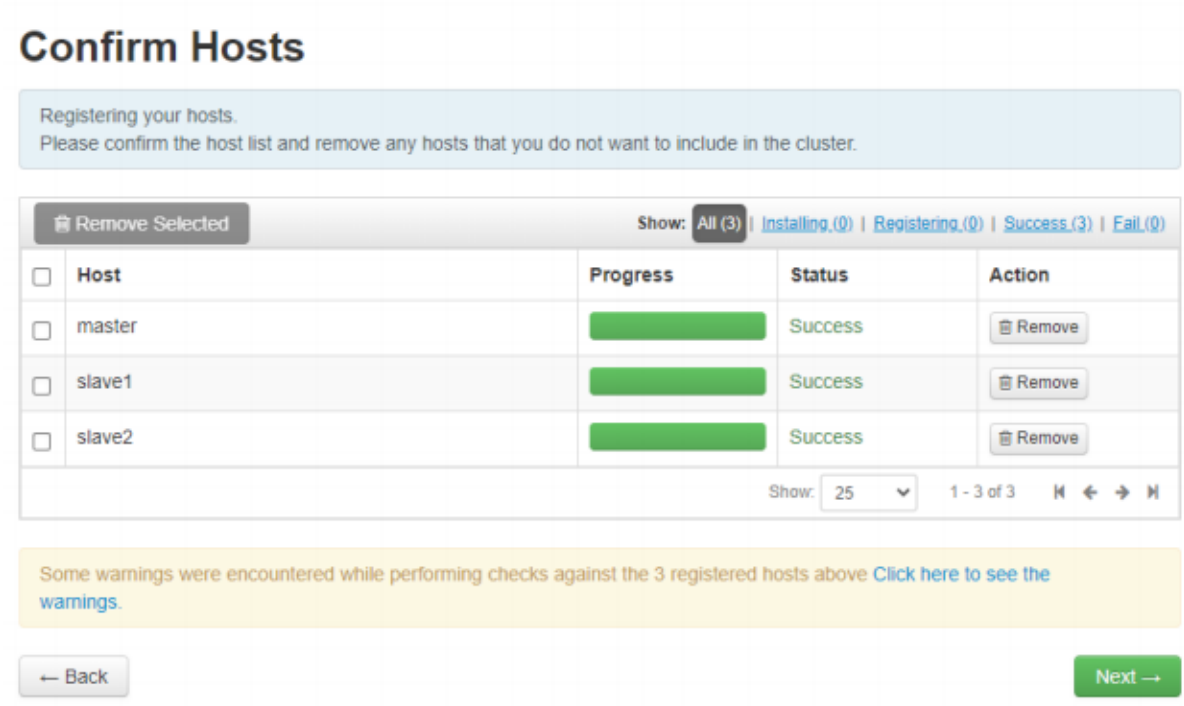
2.3安装配置

填写主机地址以及主节点的id.rsa文件

需要首先把master节点为的私密（/root/.ssh/id.rsa）拷贝到windows：



2.4 安装ambari的agent，同时检查系统问题



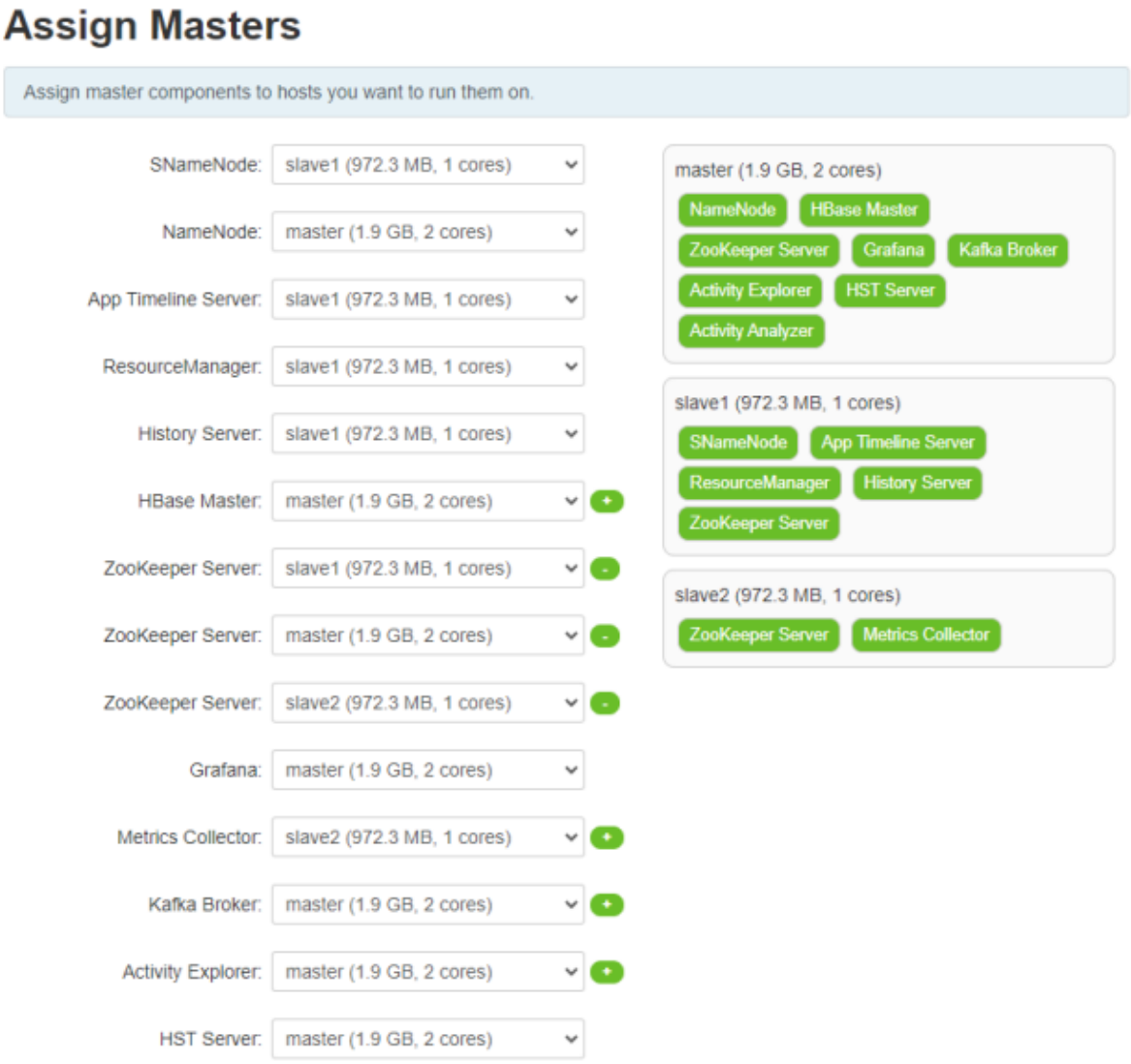
2.5选择要安装的服务

Choose Services

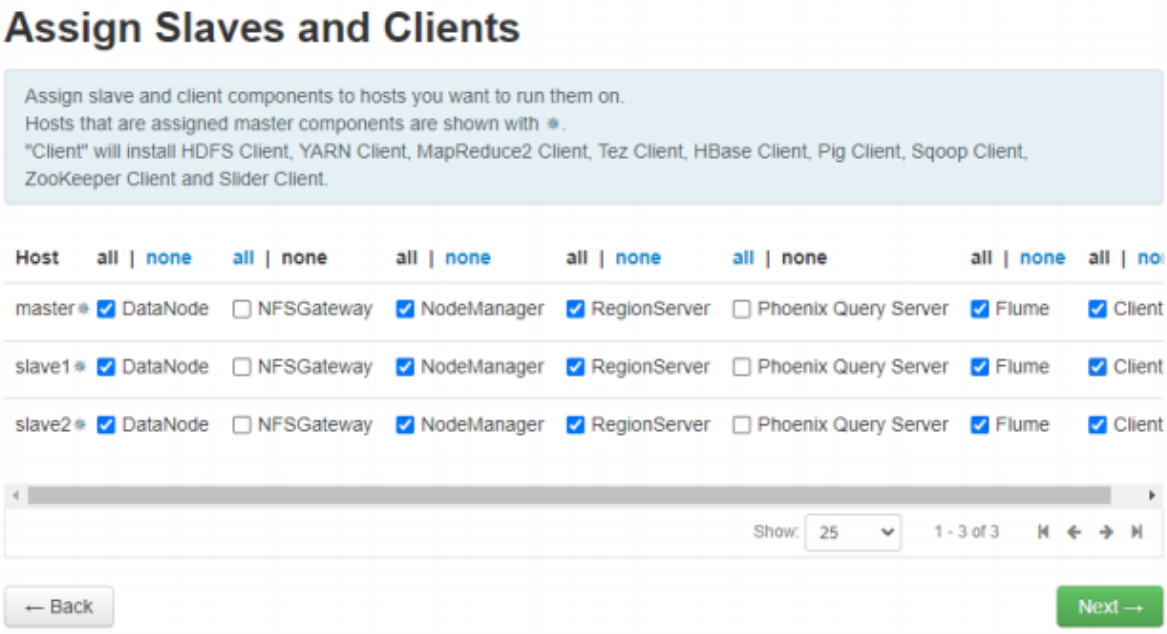
Choose which services you want to install on your cluster.

Service	Version	Description
☑ HDFS	2.7.3	Apache Hadoop Distributed File System
☑ YARN + MapReduce2	2.7.3	Apache Hadoop NextGen MapReduce (YARN)
☑ Tez	0.7.0	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
☐ Hive	1.2.1000	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
☑ HBase	1.1.2	A Non-relational distributed database, plus Phoenix, a high performance SQL layer for low latency applications.
☑ Pig	0.16.0	Scripting platform for analyzing large datasets
☑ Sqoop	1.4.6	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
☐ Oozie	4.2.0	System for workflow coordination and execution of Apache Hadoop jobs. This also includes the installation of the optional Oozie Web Console which relies on and will install the ExtJS Library.
☑ ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
☐ Falcon	0.10.0	Data management and processing platform
☐ Storm	1.1.0	Apache Hadoop Stream processing framework
☑ Flume	1.5.2	A distributed service for collecting, aggregating, and moving large amounts of streaming data into HDFS

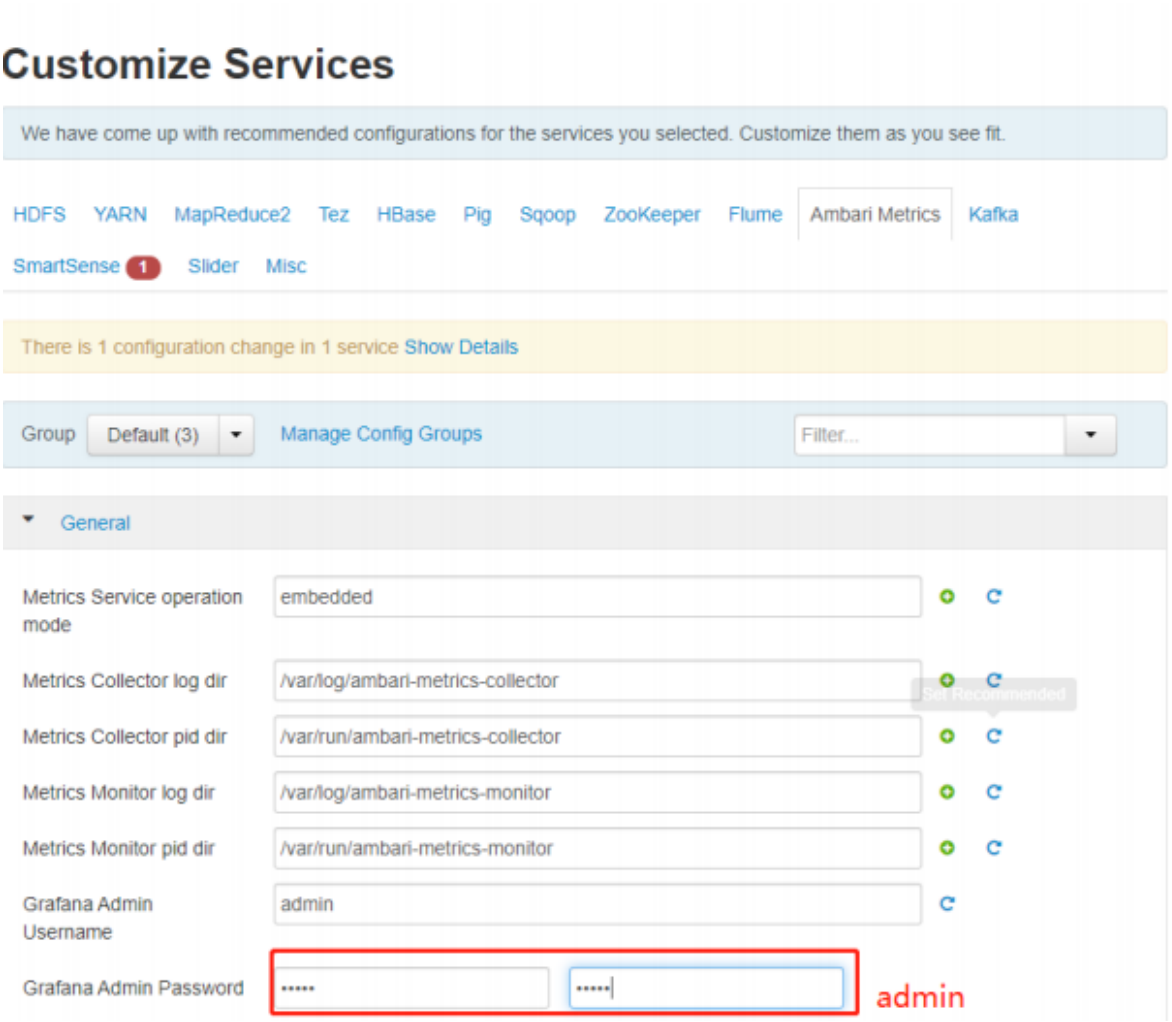
2.6分配主节点



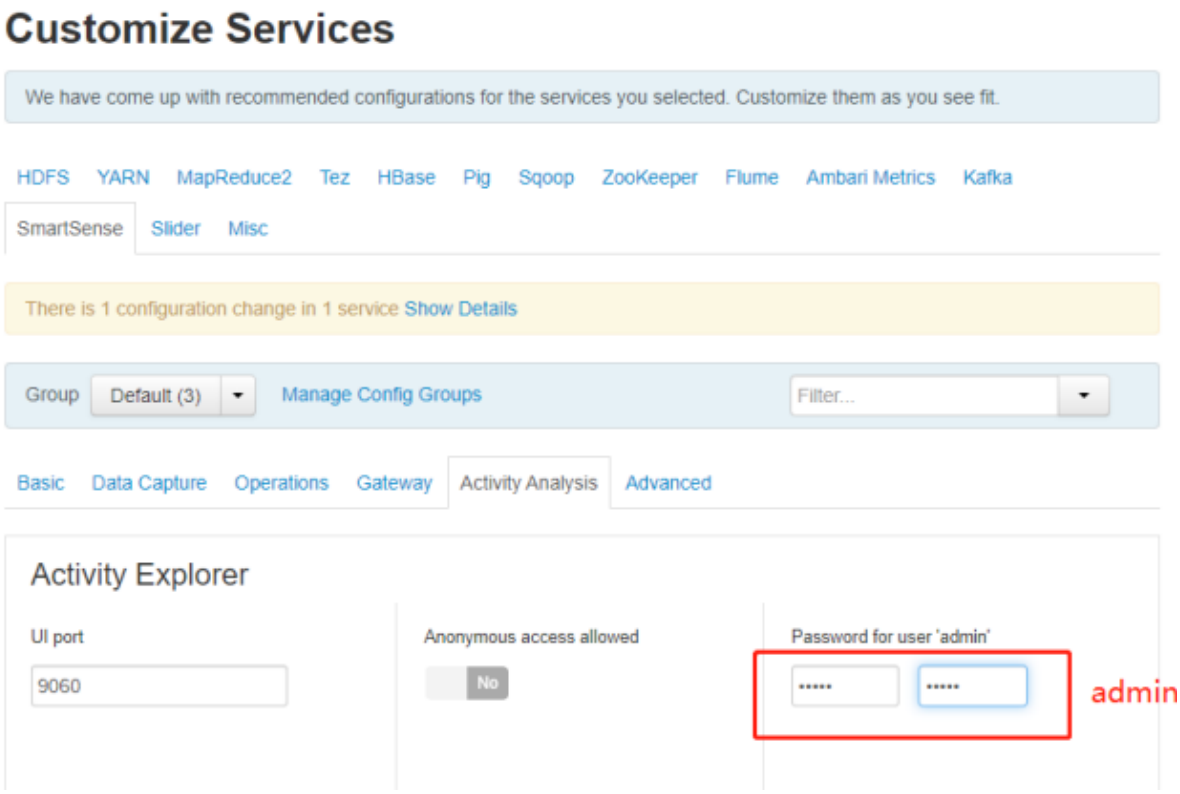
2.7分配从节点



2.8自定义安装



账号密码都是admin



2.9检查

```
Admin Name : admin
Cluster Name : hadoop
```

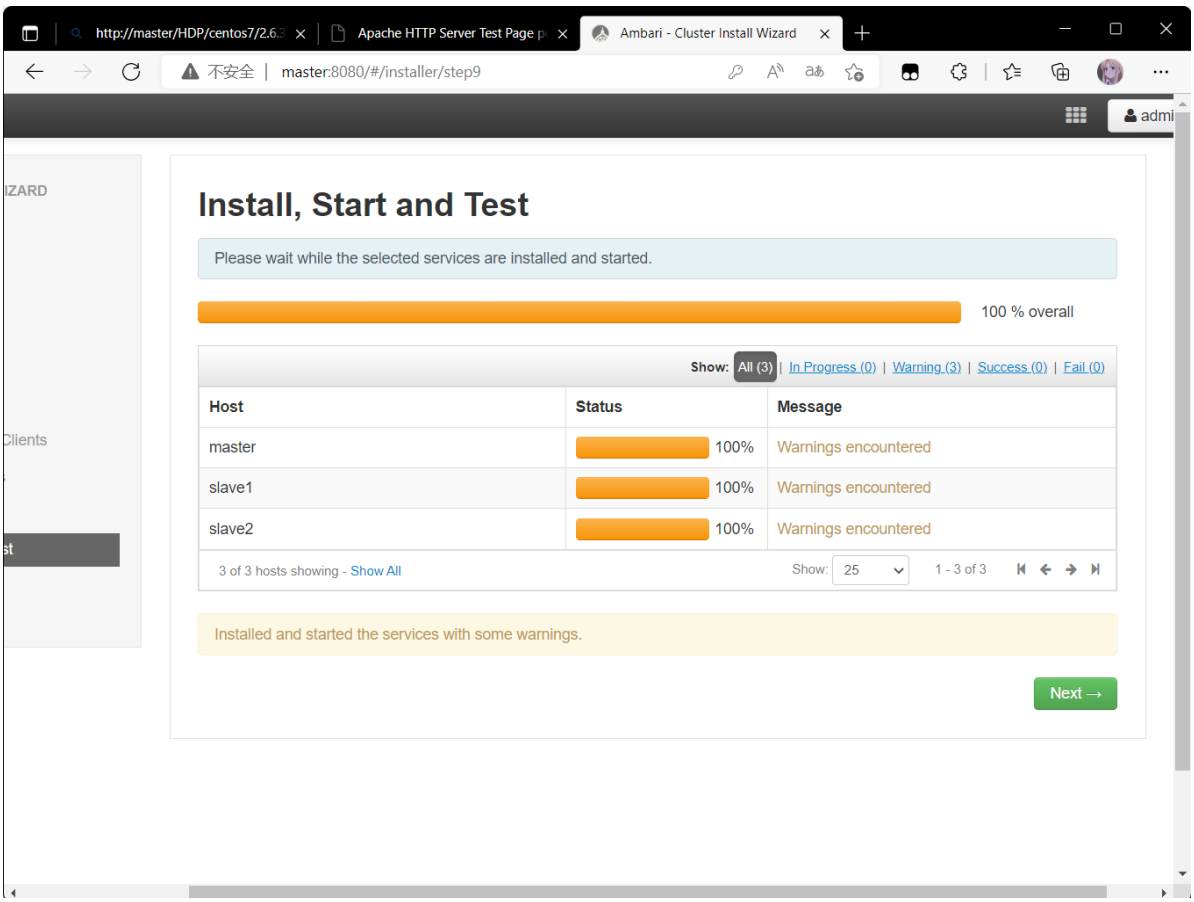
```
Total Hosts : 3 (3 new)
Repositories: redhat7 (HDP-2.6): http://192.168.91.128/HDP/centos7/2.6.3.0-235
redhat7 (HDP-UTILS-1.1.0.21): http://192.168.91.128/HDP-UTILS Services:

HDFS
DataNode : 3 hosts
NameNode : master
NFSGateway : 0 host
SNameNode : slave1
YARN + MapReduce2
App Timeline Server : slave1
NodeManager : 3 hosts
ResourceManager : slave1
Tez
Clients : 3 hosts
HBase
Master : master
RegionServer : 3 hosts
Phoenix Query Server : 0 host
Pig Clients : 3 hosts S
qoop Clients : 3 hosts
ZooKeeper Server : 3 hosts
Flume
Flume : 3 hosts
Ambari Metrics
Metrics Collector : slave2
Grafana : master
Kafka Broker : master
SmartSense
Activity Analyzer : master
Activity Explorer : master
HST Server : master
Slider
Clients : 3 hosts
```

2.10安装

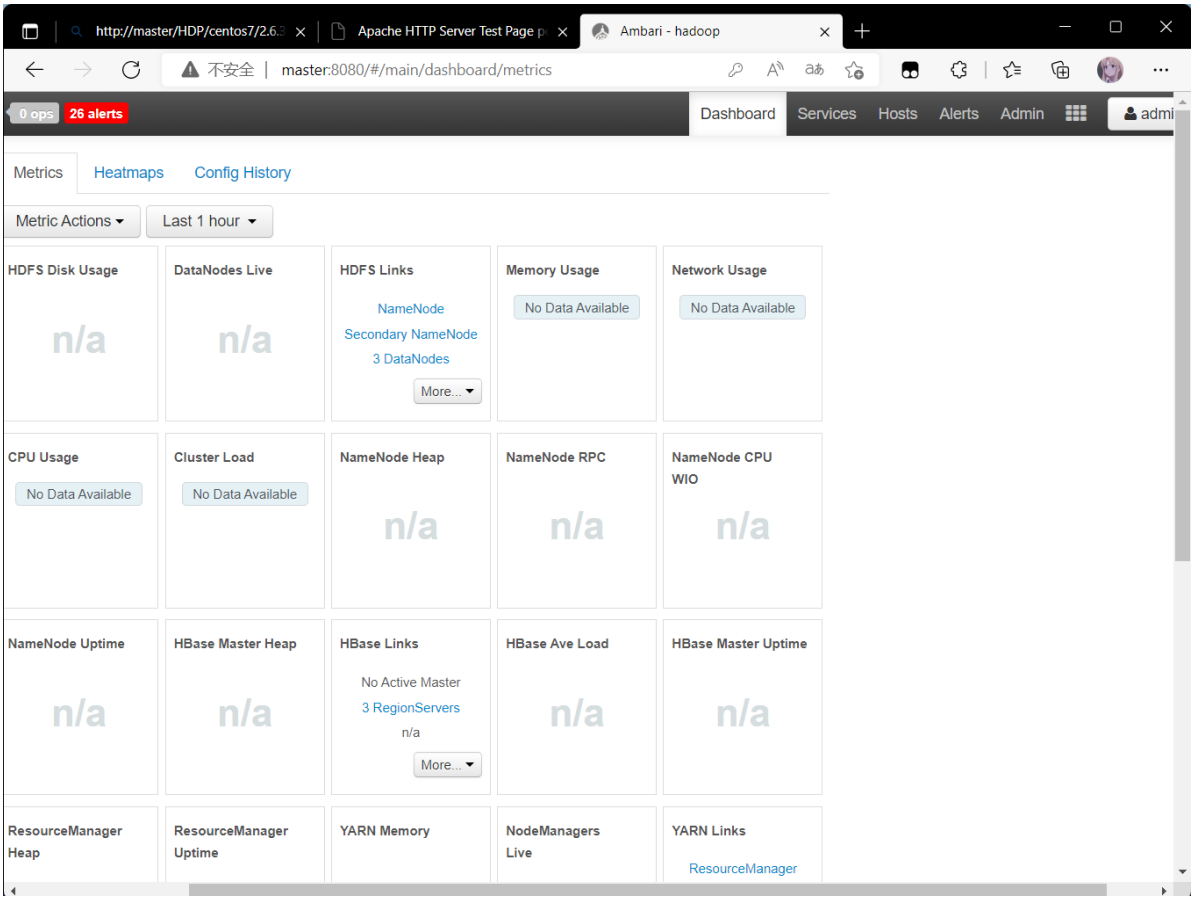
注释掉 /usr/lib/python2.6/site-packages/resource_management/libraries/script/script.py文件的

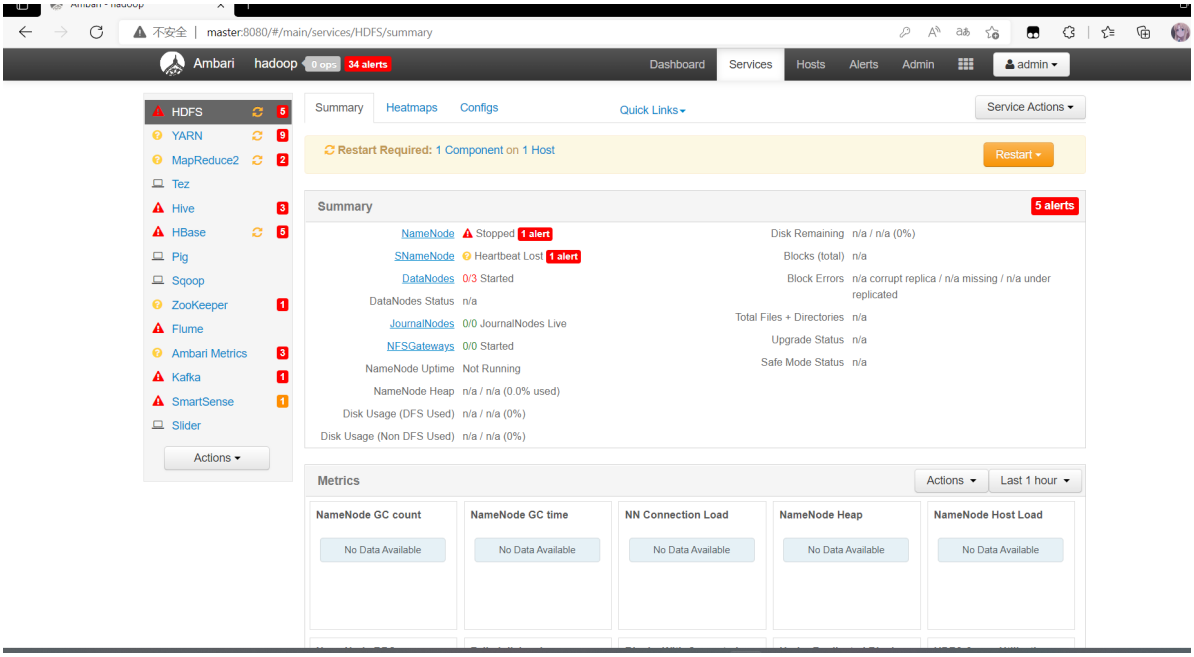
533行

```
# package_version = None
```


Ambari - Cluster Install Wizard
master:8080/#/installer/step9
Install, Start and Test
Please wait while the selected services are installed and started.
100 % overall
Show: All (3) | In Progress (0) | Warning (3) | Success (0) | Fail (0)
Host
Status
Message
master
100%
Warnings encountered
slave1
100%
Warnings encountered
slave2
100%
Warnings encountered
3 of 3 hosts showing - Show All
Show: 25
1 - 3 of 3
Installed and started the services with some warnings.
Next →

2.11查看管理集群


Ambari - hadoop
master:8080/#/main/dashboard/metrics
0 ops
26 alerts
Dashboard
Services
Hosts
Alerts
Admin
Metrics
Heatmaps
Config History
Metric Actions
Last 1 hour
HDFS Disk Usage
n/a
DataNodes Live
n/a
HDFS Links
NameNode
Secondary NameNode
3 DataNodes
More...
Memory Usage
No Data Available
Network Usage
No Data Available
CPU Usage
No Data Available
Cluster Load
No Data Available
NameNode Heap
n/a
NameNode RPC
n/a
NameNode CPU WIO
n/a
NameNode Uptime
n/a
HBase Master Heap
n/a
HBase Links
No Active Master
3 RegionServers
n/a
More...
HBase Ave Load
n/a
HBase Master Uptime
n/a
ResourceManager Heap
ResourceManager Uptime
YARN Memory
NodeManagers Live
YARN Links
ResourceManager

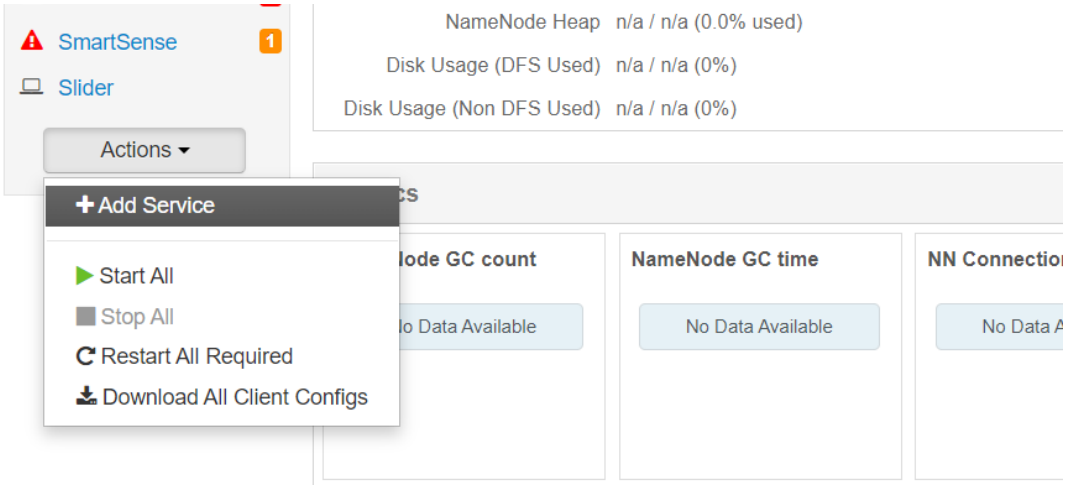


2.12添加服务

我们将要添加hive服务，先运行如下命令：

```
ambari-server setup --jdbc-db=mysql --jdbc-driver=/usr/share/java/mysql-connector-java.jar
```

(1)选择需要添加的服务

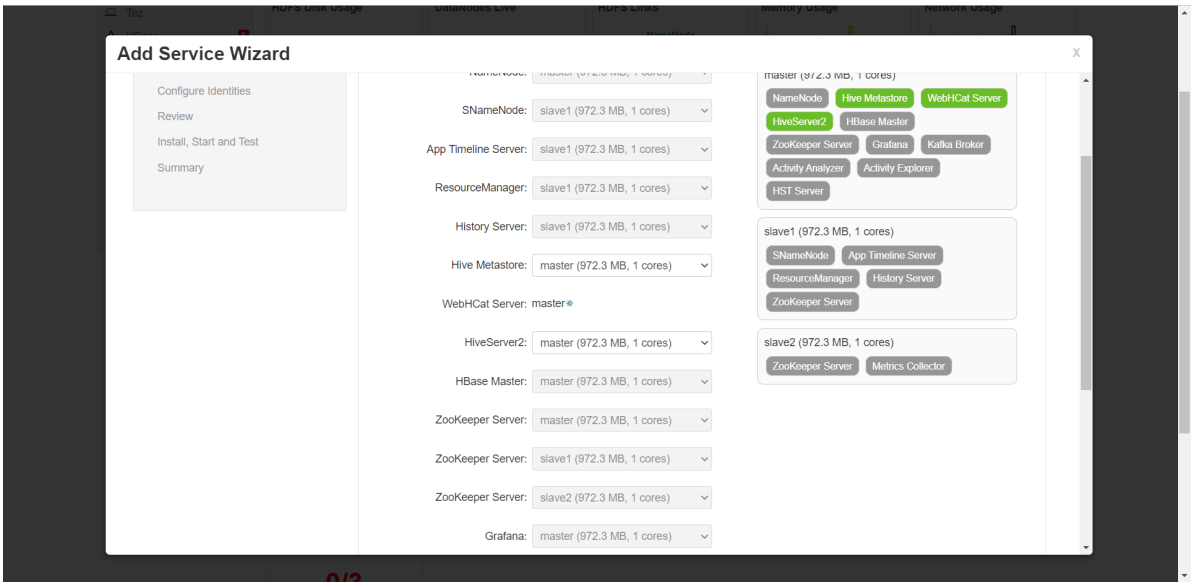


Choose Services

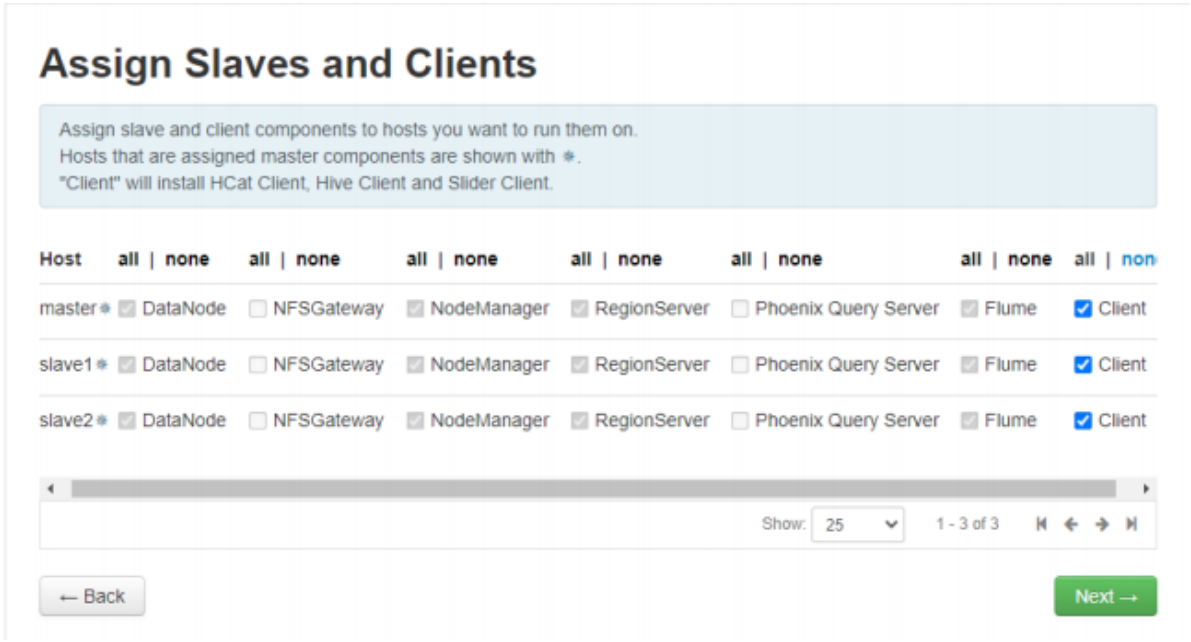
Choose which services you want to install on your cluster.

Service	Version	Description
HDFS	2.7.3	Apache Hadoop Distributed File System
YARN + MapReduce2	2.7.3	Apache Hadoop NextGen MapReduce (YARN)
Tez	0.7.0	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
Hive	1.2.1000	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
HBase	1.1.2	A Non-relational distributed database, plus Phoenix, a high performance SQL layer for low latency applications.
Pig	0.16.0	Scripting platform for analyzing large datasets
Sqoop	1.4.6	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
Oozie	4.2.0	System for workflow coordination and execution of Apache Hadoop jobs. This also includes the installation of the optional Oozie Web Console which relies on and will install the ExtJS Library.
ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
Falcon	0.10.0	Data management and processing platform

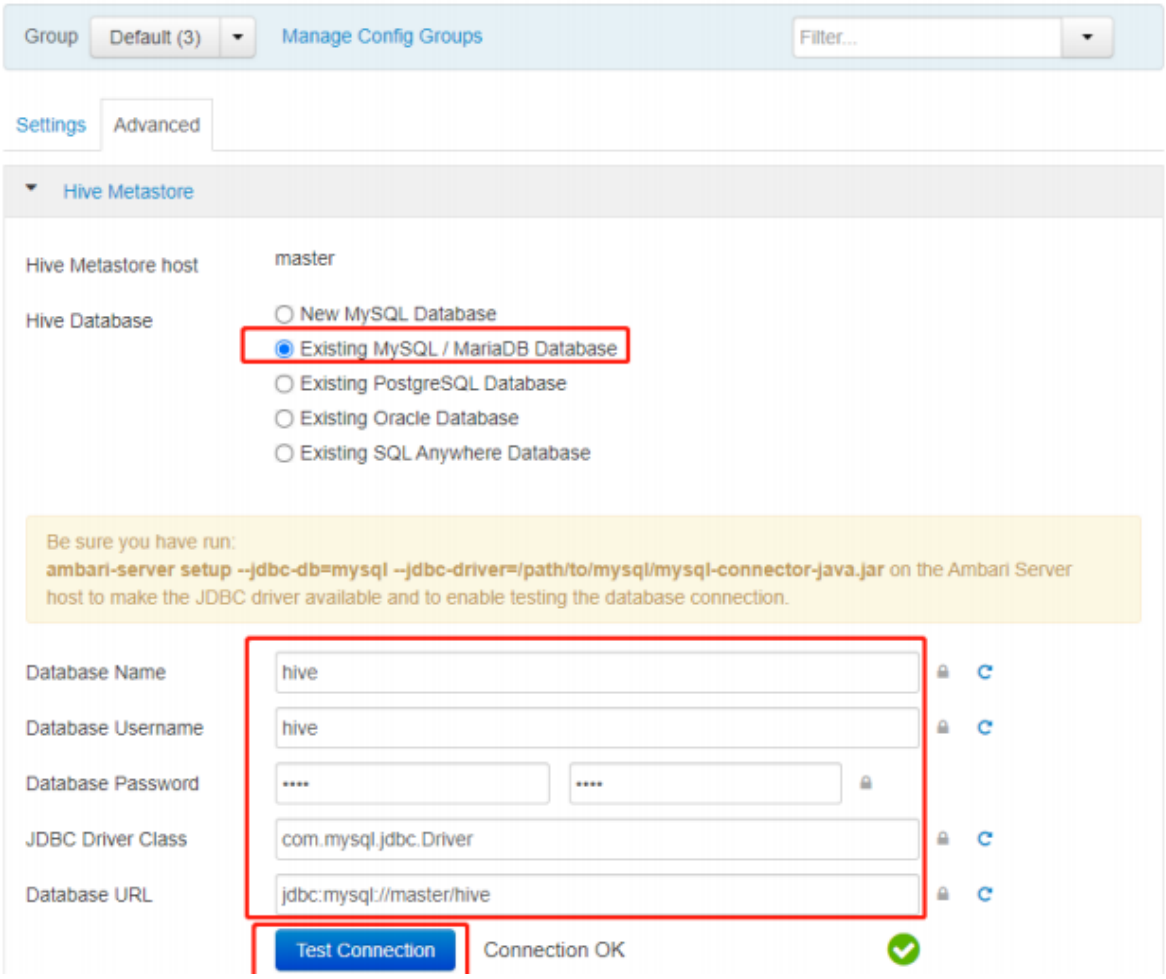
(2)分配主机



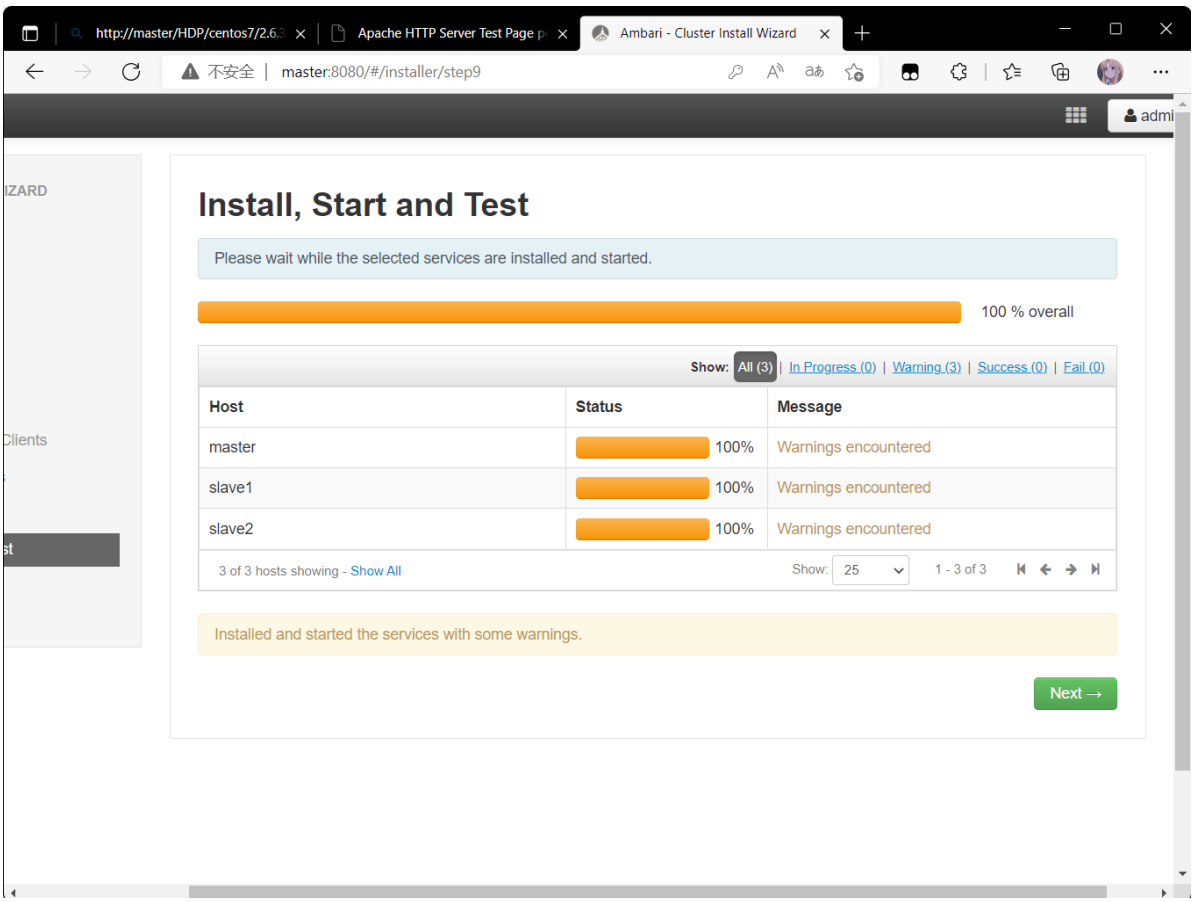
(3)分配从服务器和客户端



(4)设置Hive服务



(5)安装



六.常用命令练习

1.HDFS命令练习

通过vi 创建txt1.txt、txt2.txt、txt3.txt

(1)列出子目录或子文件

```
# 查看HDFS根目录下的文件或目录
hdfs dfs -ls
#递归列出子目录文件
hdfs dfs -ls -R /
```

```
[root@master sbin]# locate start-all.sh
[root@master sbin]# hdfs dfs -ls
ls: `.': No such file or directory
[root@master sbin]# hadoop fs -ls /
Found 2 items
drwxrwxrwx   - hdfs hdfs          0 2022-06-05 00:18 /tmp
drwxr-xr-x   - hdfs hdfs          0 2022-06-05 00:18 /user
[root@master sbin]# hdfs dfs -ls -R /
drwxrwxrwx   - hdfs hdfs          0 2022-06-05 00:18 /tmp
drwxr-xr-x   - hdfs hdfs          0 2022-06-05 00:18 /user
drwxrwx---   - ambari-qa hdfs          0 2022-06-05 00:18 /user/ambari-qa
ls: Permission denied: user=root, access=READ_EXECUTE, inode="/user/ambari-qa":am
bari-qa:hdfs:drwxrwx---
[root@master sbin]#
```