

尚硅谷大数据技术之 Ambari

(作者：尚硅谷大数据研发部)

版本：v1.0

第 1 章 Ambari 简述

1.1 什么是 Ambari

Apache Ambari 项目旨在通过开发用于配置，管理和监控 Apache Hadoop 集群的软件来简化 Hadoop 管理。Ambari 提供了一个直观，易用的 Hadoop 管理 Web UI。

1.2 Ambari 的功能

提供了跨任意数量的主机安装 Hadoop 服务的分步向导。

处理群集的 Hadoop 服务配置。

提供集中管理，用于在整个集群中启动，停止和重新配置 Hadoop 服务。

提供了一个仪表板，用于监控 Hadoop 集群的运行状况和状态。

利用 Ambari 指标系统进行指标收集。

利用 Ambari Alert Framework 进行系统警报，并在需要您注意时通知您（例如，节点出现故障，剩余磁盘空间不足等）。

第 2 章 环境准备

注意：以下操作三台机器都需要进行

2.1 虚拟机准备

克隆三台虚拟机（hadoop102、hadoop103、hadoop104），配置好对应主机的网络 IP、主机名称、关闭防火墙。

```
[root@hadoop102 ~]# chkconfig iptables off
[root@hadoop102 ~]# service iptables stop
[root@hadoop102 ~]# chkconfig --list iptables
iptables          0:关闭  1:关闭  2:关闭  3:关闭  4:关闭  5:关闭  6:关闭
```

2.2 关闭 SELINUX

```
[root@hadoop102 ~]# vim /etc/sysconfig/selinux
```

将 SELINUX=enforcing 改为 SELINUX=disabled
执行该命令后重启机器生效

2.3 安装 JDK

1) 在 hadoop102 的 /opt 目录下创建 module 和 software 文件夹

更多 [Java](#) - [大数据](#) - [前端](#) - [python](#) 人工智能资料下载，可百度访问：尚硅谷官网

```
[root@hadoop102 opt]# mkdir module
[root@hadoop102 opt]# mkdir software
```

2) 用 SecureCRT 将 `jdk-8u144-linux-x64.tar.gz` 导入到 hadoop102 的 `/opt/software` 目录下

3) 在 Linux 系统下的 `opt` 目录中查看软件包是否导入成功

```
[root@hadoop102 software]$ ls
jdk-8u144-linux-x64.tar.gz
```

4) 解压 JDK 到 `/opt/module` 目录下

```
[root@hadoop102 software]$ tar -zxvf jdk-8u144-linux-x64.tar.gz -C /opt/module/
```

5) 配置 JDK 环境变量

(1) 先获取 JDK 路径

```
[root@hadoop102 jdk1.8.0_144]$ pwd
/opt/module/jdk1.8.0_144
```

(2) 打开 `/etc/profile` 文件

```
[root@hadoop102 software]$ vi /etc/profile
```

在 `profile` 文件末尾添加 JDK 路径

```
#JAVA_HOME
export JAVA_HOME=/opt/module/jdk1.8.0_144
export PATH=$PATH:$JAVA_HOME/bin
```

(3) 保存后退出

```
:wq
```

(4) 让修改后的文件生效

```
[root@hadoop102 jdk1.8.0_144]$ source /etc/profile
```

6) 测试 JDK 是否安装成功

```
[root@hadoop102 jdk1.8.0_144]# java -version
java version "1.8.0_144"
```

7) 将 hadoop102 中的 JDK 和环境变量分发到 hadoop103、hadoop104 两台主机

```
[root@hadoop102 opt]# xsync /opt/module/
[root@hadoop102 opt]# xsync /etc/profile
```

分别在 hadoop103、hadoop104 上 `source` 一下

```
[root@hadoop103 ~]$ source /etc/profile
[root@hadoop104 ~]# source /etc/profile
```

2.4 SSH 免密登录

配置 hadoop102 对 hadoop102、hadoop103、hadoop104 三台主机的免密登陆。

1) 生成公钥和私钥:

```
[root@hadoop102 .ssh]$ ssh-keygen -t rsa
```

然后敲 (三个回车), 就会生成两个文件 `id_rsa` (私钥)、`id_rsa.pub` (公钥)

2) 将公钥拷贝到要免密登录的目标机器上

```
[root@hadoop102 .ssh]$ ssh-copy-id hadoop102
```

```
[root@hadoop102 .ssh]$ ssh-copy-id hadoop103  
[root@hadoop102 .ssh]$ ssh-copy-id hadoop104
```

2.5 修改 yum 源为阿里云镜像

```
[root@hadoop102 yum.repos.d]# mv /etc/yum.repos.d/CentOS-Base.repo  
/etc/yum.repos.d/CentOS-Base.repo.bk  
  
[root@hadoop102 yum.repos.d]# wget -O /etc/yum.repos.d/CentOS-Base.repo  
http://mirrors.aliyun.com/repo/Centos-6.repo  
  
//把服务器的包下载到本地缓存  
[root@hadoop102 yum.repos.d]# yum makecache
```

2.6 安装 ntp

```
[root@hadoop102 ~]# yum install -y ntp  
[root@hadoop102 ~]# chkconfig --list ntpd  
[root@hadoop102 ~]# chkconfig ntpd on  
[root@hadoop102 ~]# service ntpd start
```

2.7 关闭 Linux 的 THP 服务

如果不关闭 transparent_hugepage，HDFS 会因为这个性能严重受影响。

关闭 transparent_hugepage 方法是：

```
[root@hadoop102 ~]# vim /etc/grub.conf  
添加 transparent_hugepage=never  
  
[root@hadoop102 ~]# vim /etc/rc.local  
添加：  
if test -f /sys/kernel/mm/transparent_hugepage/defrag; then  
    echo never > /sys/kernel/mm/transparent_hugepage/defrag  
fi  
if test -f /sys/kernel/mm/transparent_hugepage/enabled; then  
    echo never > /sys/kernel/mm/transparent_hugepage/enabled  
fi  
exit 0
```

重启之后，用下面的命令检查：

```
[root@hadoop102 yum.repos.d]#  
cat /sys/kernel/mm/redhat_transparent_hugepage/enabled  
  
always madvise [never]
```

有 [never]则表示 THP 被禁用

2.8 配置 UMASK

设定用户所创建目录的初始权限

```
[root@hadoop102 ~]# umask 0022
```

2.9 禁止离线更新

```
vim /etc/yum/pluginconf.d/refresh-packagekit.conf  
修改：enabled=0
```

第 3 章 安装 Ambari 集群

注意：以下操作主节点操作即可

3.1 制作本地源

制作本地源是因为在线安装 Ambari 太慢。制作本地源只需在主节点上进行。

3.1.1 配置 HTTPD 服务

配置 HTTPD 服务到系统层使其随系统自动启动

```
[root@hadoop102 ~]# chkconfig httpd on
[root@hadoop102 ~]# service httpd start
```

3.1.2 安装工具

安装本地源制作相关工具

```
[root@hadoop102 ~]# yum install yum-utils createrepo yum-plugin-priorities -y
[root@hadoop102 ~]# vim /etc/yum/pluginconf.d/priorities.conf
```

添加 `gpgcheck=0`

3.1.3 将下载的 3 个 tar 包解压

```
[root@hadoop102 software]# tar -zxvf /opt/software/ambari-2.5.0.3-centos6.tar.gz -C /var/www/html/

[root@hadoop102 software]# mkdir /var/www/html/hdp
[root@hadoop102 software]# tar -zxvf /opt/software/HDP-2.6.0.3-centos6-rpm.tar.gz -C /var/www/html/hdp

[root@hadoop102 software]# tar -zxvf /opt/software/HDP-UTILS-1.1.0.21-centos6.tar.gz -C /var/www/html/hdp
```

3.1.4 创建本地源

```
[root@hadoop102 software]# cd /var/www/html/
[root@hadoop102 html]# createrepo ./
```

3.1.5 将 Ambari 存储库文件下载到安装主机上的目录中

```
[root@hadoop102 yum.repos.d]# wget -nv http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.6.1.5/ambari.repo -O /etc/yum.repos.d/ambari.repo
```

3.1.6 修改配置文件，配置为本地源

修改 `ambari.repo`，配置为本地源

```
[root@hadoop102 html]# vim /etc/yum.repos.d/ambari.repo

#VERSION_NUMBER=2.6.1.5-3
[ambari-2.6.1.5]
name=ambari Version - ambari-2.6.1.5
baseurl=http://hadoop102/ambari/centos6/
```

```
gpgcheck=0
gpgkey=http://hadoop102/ambari/centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

修改 hdp-util.repo，配置为本地源

```
[root@hadoop102 hdp]# vim /var/www/html/hdp/hdp-util.repo
[HDP-UTILS-1.1.0.21]
name=Hortonworks Data Platform Version - HDP-UTILS-1.1.0.21
baseurl=http://hadoop102/hdp/
gpgcheck=0
enabled=1
priority=1
```

修改 hdp.repo，配置为本地源

```
[root@hadoop102 centos6]#
vim /var/www/html/hdp/HDP/centos6/hdp.repo

#VERSION_NUMBER=2.6.0.3-8
[HDP-2.6.0.3]
name=HDP Version - HDP-2.6.0.3
baseurl=http://hadoop102/hdp/HDP/centos6/
gpgcheck=0
gpgkey=http://hadoop102/hdp/HDP/centos6/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://hadoop102/hdp/
gpgcheck=0
gpgkey=http://hadoop102/hdp/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
enabled=1
priority=1
```

```
[root@hadoop102 ~]# yum clean all
[root@hadoop102 ~]# yum makecache
```

查看是否有 Ambari

```
[root@hadoop102 ~]# yum repolist
```

查看 Ambari 与 HDP 资源的资源库。

也可以打开浏览器查看一下：

<http://hadoop102/ambari/centos6/>
<http://hadoop102/hdp/HDP/centos6/>
<http://hadoop102/hdp/>

3.2 安装 MySQL

Ambari 使用的默认数据库是 PostgreSQL，用于存储安装元数据，可以使用自己安装 MySQL 数据库作为 Ambari 元数据库。

注意：一定要用 root 用户操作如下步骤：先卸载 MySQL 再安装

1) 安装包准备

更多 Java - 大数据 - 前端 - python 人工智能资料下载，可百度访问：尚硅谷官网

- (1) 查看 MySQL 是否安装

```
[root@hadoop102 桌面]# rpm -qa|grep mysql  
mysql-libs-5.1.73-7.el6.x86_64
```

- (2) 如果安装了 MySQL，就先卸载

```
[root@hadoop102 桌面]# rpm -e --nodeps mysql-libs-5.1.73-7.el6.x86_64
```

- (3) 解压 mysql-libs.zip 文件到当前目录

```
[root@hadoop102 software]# unzip mysql-libs.zip  
[root@hadoop102 software]# ls  
mysql-libs.zip  
mysql-libs
```

- (4) 进入到 mysql-libs 文件夹下

```
[root@hadoop102 mysql-libs]# ll  
总用量 76048  
-rw-r--r--. 1 root root 18509960 3 月 26 2015 MySQL-client-5.6.24-1.el6.x86_64.rpm  
-rw-r--r--. 1 root root 3575135 12 月 1 2013 mysql-connector-java-5.1.27.tar.gz  
-rw-r--r--. 1 root root 55782196 3 月 26 2015 MySQL-server-5.6.24-1.el6.x86_64.rpm
```

2) 安装 MySQL 服务器

- (1) 安装 MySQL 服务端

```
[root@hadoop102 mysql-libs]# rpm -ivh MySQL-server-5.6.24-1.el6.x86_64.rpm
```

- (2) 查看产生的随机密码

```
[root@hadoop102 mysql-libs]# cat /root/.mysql_secret  
XrBwJEENp2yhVe4O
```

- (3) 查看 MySQL 状态

```
[root@hadoop102 mysql-libs]# service mysql status
```

- (4) 启动 MySQL

```
[root@hadoop102 mysql-libs]# service mysql start
```

3) 安装 MySQL 客户端

- (1) 安装 MySQL 客户端

```
[root@hadoop102 mysql-libs]# rpm -ivh MySQL-client-5.6.24-1.el6.x86_64.rpm
```

- (2) 链接 MySQL

```
[root@hadoop102 mysql-libs]# mysql -uroot -pXrBwJEENp2yhVe4O
```

- (3) 修改密码

```
mysql>SET PASSWORD=PASSWORD('000000');
```

- (4) 退出 MySQL

```
mysql>exit
```

4) MySQL 中 user 表中主机配置

配置只要是 root 用户+密码，在任何主机上都能登录 MySQL 数据库。

- (1) 进入 MySQL

```
[root@hadoop102 mysql-libs]# mysql -uroot -p000000
```

- (2) 显示数据库

```
mysql>show databases;
```

(3) 使用 MySQL 数据库

```
mysql>use mysql;
```

(4) 展示 MySQL 数据库中的所有表

```
mysql>show tables;
```

(5) 展示 user 表的结构

```
mysql>desc user;
```

(6) 查询 user 表

```
mysql>select User, Host, Password from user;
```

(7) 修改 user 表, 把 Host 表内容修改为%

```
mysql>update user set host='%' where host='localhost';
```

(8) 删除 root 用户的其他 host

```
mysql>
delete from user where Host='hadoop102';
delete from user where Host='127.0.0.1';
delete from user where Host='::1';
```

(9) 刷新

```
mysql>flush privileges;
```

(10) 退出

```
mysql>quit;
```

3.3 安装 Ambari

1) 安装 ambari-server

```
[root@hadoop102 hdp]# yum install ambari-server
```

2) 拷贝 mysql 驱动

将 mysql-connector-java.jar 复制到/usr/share/java 目录下并改名为 mysql-connector-java.jar

```
[root@hadoop102 hdp]# mkdir /usr/share/java
```

```
[root@hadoop102 hdp]# cp
/opt/software/mysql-libs/mysql-connector-java-5.1.27/mysql-connector-java-5.1.27-bin.jar
/usr/share/java/mysql-connector-java.jar
```

将 mysql-connector-java.jar 复制到 /var/lib/ambari-server/resources 目录下并改名为

mysql-jdbc-driver.jar

```
[root@hadoop102 hdp]# cp /usr/share/java/mysql-connector-java.jar
/var/lib/ambari-server/resources/mysql-jdbc-driver.jar
```

修改 ambari.properties 文件

```
[root@hadoop102 hdp]#
vim /etc/ambari-server/conf/ambari.properties
添加
server.jdbc.driver.path=/usr/share/java/mysql-connector-java.jar
```

3.4 在 MySQL 中创建数据库

- 1) 创建 ambari 库

```
[root@hadoop102 hdp]# mysql -u root -p000000  
mysql> create database ambari;
```

- 2) 使用 Ambari 自带脚本创建表

```
mysql> use ambari;  
mysql> source /var/lib/ambari-server/resources/Ambari-DDL-MySQL-CREATE.sql;
```

- 3) 赋予用户 root 权限:

```
mysql> grant all privileges on *.* to 'root'@'%' identified by '000000';
```

- 4) 刷新

```
mysql> flush privileges;
```

3.5 配置 Ambari

执行

```
[root@hadoop102 hdp]# ambari-server setup
```

下面是配置执行流程, 按照提示操作

- 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。

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

如果直接回车就是默认选择 root 用户

- 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/module/jdk1.8.0_144

```
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/module/jdk1.8.0_144
```

```
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):hadoop102
Port (3306):
Database name (ambari):
Username (ambari):root
Enter Database Password (bigdata):000000 (这里不显示)
Re-Enter password:000000
```

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)?
```

如果使用自己定义的数据库，必须在启动 Ambari 服务之前导入 Ambari 的 sql 脚本。

3.6 启动 Ambari

启动命令为：

```
ambari-server start
```

停止命令为：

```
ambari-server stop
```

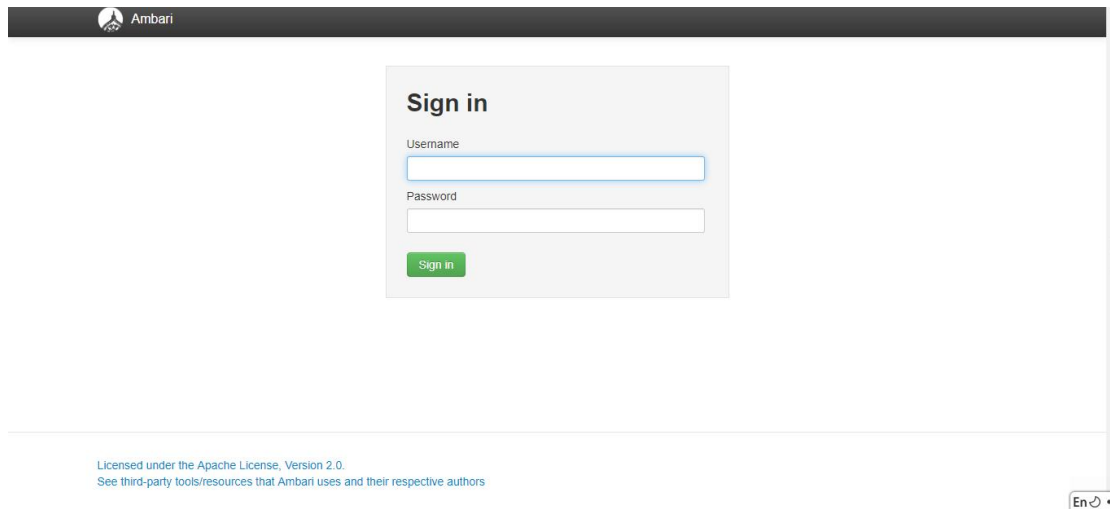
第 4 章 HDP 集群部署

4.1 集群搭建

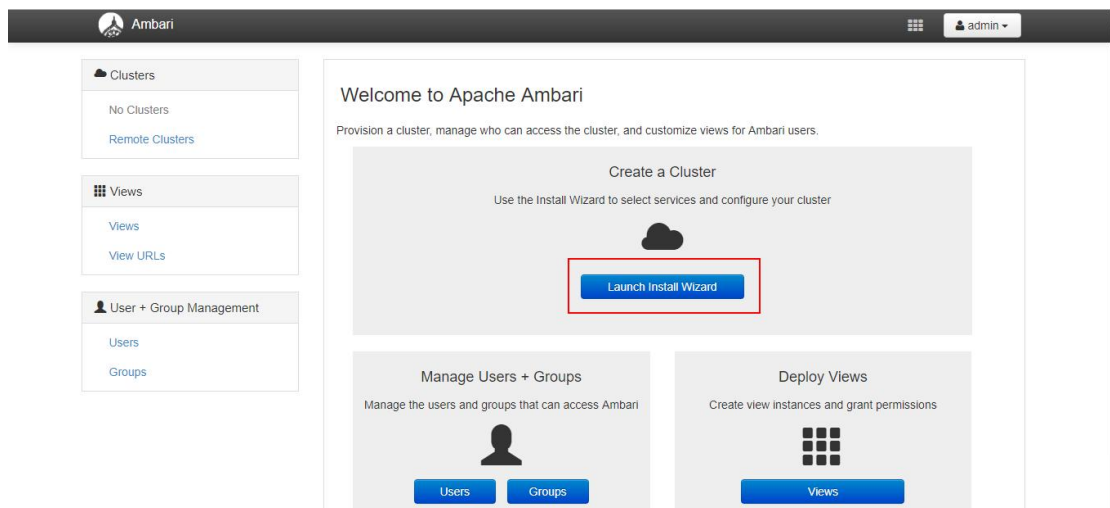
4.1.1 进入登录页面

浏览器输入 <http://hadoop102:8080/>

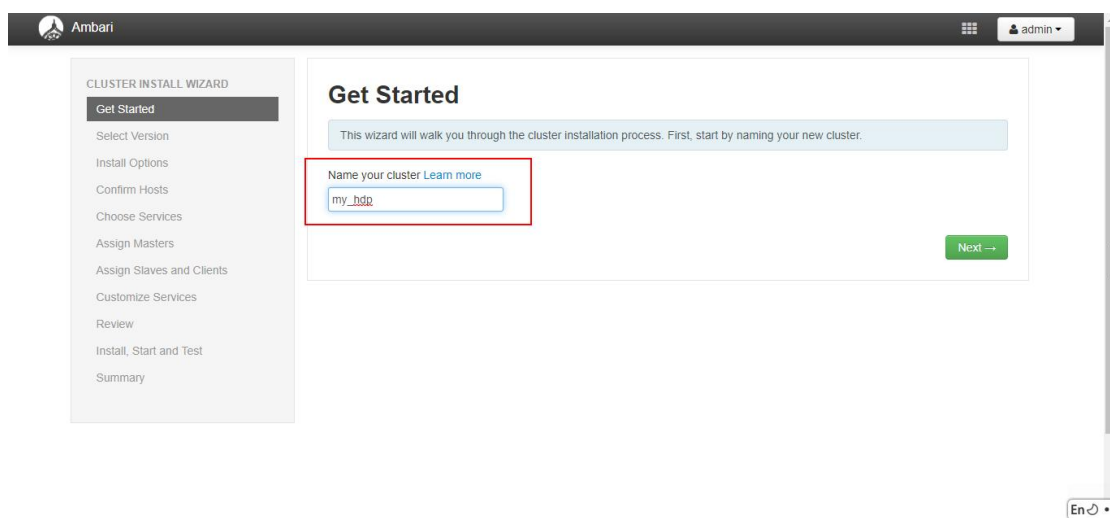
默认管理员账目密码：admin



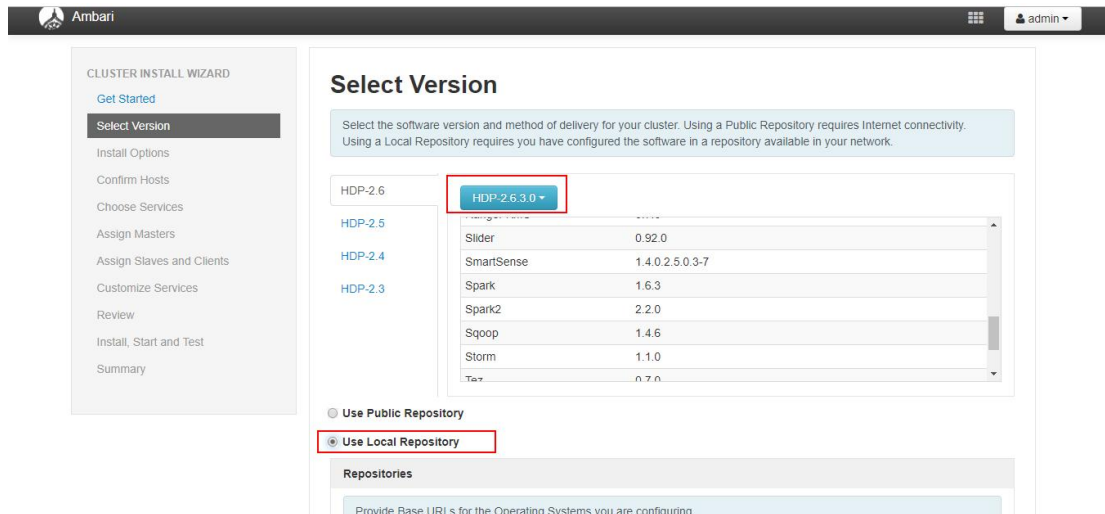
4.1.2 点击 Launch Install Wizard



4.1.3 设置集群名称



4.1.4 选择版本和存储库



The screenshot shows the Ambari 'Select Version' page. On the left is a 'CLUSTER INSTALL WIZARD' sidebar with steps: Get Started, Select Version (highlighted), Install Options, Confirm Hosts, Choose Services, Assign Masters, Assign Slaves and Clients, Customize Services, Review, Install, Start and Test, and Summary. The main area is titled 'Select Version' and contains instructions: 'Select the software version and method of delivery for your cluster. Using a Public Repository requires Internet connectivity. Using a Local Repository requires you have configured the software in a repository available in your network.'

Under 'HDP-2.6', a dropdown menu is open, showing 'HDP-2.6.3.0' selected. Below this, a table lists components and their versions:

Component	Version
Slider	0.92.0
SmartSense	1.4.0.2.5.0.3-7
Spark	1.6.3
Spark2	2.2.0
Sqoop	1.4.6
Storm	1.1.0
Tez	0.7.0

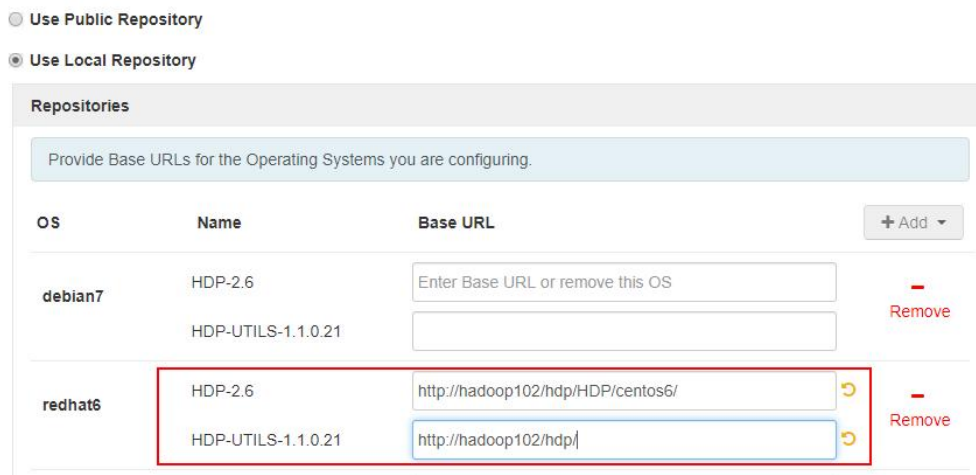
At the bottom, there are two radio buttons: 'Use Public Repository' and 'Use Local Repository' (which is selected and highlighted with a red box). Below these is a section titled 'Repositories' with a text input field: 'Provide Base URLs for the Operating Systems you are configuring.'

4.1.5 填写本地库地址

在 redhat6 后面分别填写

`http://hadoop102/hdp/HDP/centos6/`

`http://hadoop102/hdp/`



The screenshot shows the 'Repositories' section of the Ambari interface. It has two radio buttons: 'Use Public Repository' and 'Use Local Repository' (selected). Below is a table with columns 'OS', 'Name', and 'Base URL'. There are two rows for 'redhat6' OS, each with a 'Name' and a 'Base URL' field. The first row has 'HDP-2.6' and 'http://hadoop102/hdp/HDP/centos6/'. The second row has 'HDP-UTILS-1.1.0.21' and 'http://hadoop102/hdp/'. Both rows have a 'Remove' button next to them. A red box highlights the 'redhat6' section.

OS	Name	Base URL	
debian7	HDP-2.6	Enter Base URL or remove this OS	Remove
	HDP-UTILS-1.1.0.21		
redhat6	HDP-2.6	http://hadoop102/hdp/HDP/centos6/	Remove
	HDP-UTILS-1.1.0.21	http://hadoop102/hdp/	

4.1.6 填写主机地址以及主节点的 id_rsa 文件

Install Options

Enter the list of hosts to be included in the cluster and provide your SSH key.

Target Hosts

Enter a list of hosts using the Fully Qualified Domain Name (FQDN), one per line. Or use [Pattern Expressions](#)

hadoop102
hadoop103
hadoop104

Host Registration Information

☒ Provide your SSH Private Key to automatically register hosts

选择文件


id_rsa

```
-----BEGIN RSA PRIVATE KEY-----
MIIEogIBAAKCAQEAqIdKrxBA5U0HOk/SaNZCh1BL9QUaetoGW9q7Do8Mg00qCEY
i
```

SSH User Account

SSH Port Number

4.1.7 等待安装

 Ambari
 admin

CLUSTER INSTALL WIZARD

- [Get Started](#)
- [Select Version](#)
- [Install Options](#)
- Confirm Hosts**
- [Choose Services](#)
- [Assign Masters](#)
- [Assign Slaves and Clients](#)
- [Customize Services](#)
- [Review](#)
- [Install, Start and Test](#)
- [Summary](#)

Confirm Hosts

Registering your hosts.
Please confirm the host list and remove any hosts that you do not want to include in the cluster.

Remove Selected

Show: All (3) | Installing (0) | Registering (0) | Success (3) | Fail (0)

Host	Progress	Status	Action
<input type="checkbox"/> hadoop102	<div></div>	Success	Remove
<input type="checkbox"/> hadoop103	<div></div>	Success	Remove
<input type="checkbox"/> hadoop104	<div></div>	Success	Remove

Show: 25 1 - 3 of 3

All host checks passed on 3 registered hosts. [Click here to see the check results.](#)

[Back](#)
[Next](#)

4.1.8 选择服务

Install Options

Confirm Hosts

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Review

Install, Start and Test

Summary

Service	Version	Description
<input checked="" type="checkbox"/> HDFS	2.7.3	Apache Hadoop Distributed File System
<input checked="" type="checkbox"/> YARN + MapReduce2	2.7.3	Apache Hadoop NextGen MapReduce (YARN)
<input checked="" type="checkbox"/> Tez	0.7.0	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
<input type="checkbox"/> Hive	1.2.1000	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
<input type="checkbox"/> HBase	1.1.2	A Non-relational distributed database, plus Phoenix, a high performance SQL layer for low latency applications.
<input type="checkbox"/> Pig	0.16.0	Scripting platform for analyzing large datasets
<input type="checkbox"/> Sqoop	1.4.6	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
<input type="checkbox"/> 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.
<input checked="" type="checkbox"/> ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
<input type="checkbox"/> Falcon	0.10.0	Data management and processing platform
<input type="checkbox"/> Storm	1.1.0	Apache Hadoop Stream processing framework
<input type="checkbox"/> Flume	1.5.0	A distributed service for collecting, aggregating, and moving large amounts of

4.1.9 选择每台机器的角色

Ambari

CLUSTER INSTALL WIZARD

Get Started

Select Version

Install Options

Confirm Hosts

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Review

Install, Start and Test

Summary

Assign Masters

Assign master components to hosts you want to run them on.

Component	Host	Role
SNameNode	hadoop103 (1.9 GB, 8 cores)	NameNode
NameNode	hadoop102 (1.9 GB, 8 cores)	ZooKeeper Server
App Timeline Server	hadoop103 (1.9 GB, 8 cores)	Grafana
ResourceManager	hadoop103 (1.9 GB, 8 cores)	Activity Analyzer
History Server	hadoop104 (1.9 GB, 8 cores)	Activity Explorer
ZooKeeper Server	hadoop104 (1.9 GB, 8 cores)	HST Server
ZooKeeper Server	hadoop102 (1.9 GB, 8 cores)	SNameNode
ZooKeeper Server	hadoop103 (1.9 GB, 8 cores)	App Timeline Server
Metrics Collector	hadoop104 (1.9 GB, 8 cores)	ResourceManager
		ZooKeeper Server
		History Server
		ZooKeeper Server
		Metrics Collector

4.1.10 设置从节点

Ambari

CLUSTER INSTALL WIZARD

Get Started

Select Version

Install Options

Confirm Hosts

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Review

Install, Start and Test

Summary

Assign Slaves and Clients

Assign slave and client components to hosts you want to run them on.

Hosts that are assigned master components are shown with *.

"Client" will install HDFS Client, YARN Client, MapReduce2 Client, Tez Client and ZooKeeper Client.

Host	all none	all none	all none	all none
hadoop102 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client
hadoop103 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client
hadoop104 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client

Show: 25 1 - 3 of 3

Back Next

4.1.11 输入两次 admin

Customize Services

We have come up with recommended configurations for the services you selected. Customize them as you see fit.

HDFS YARN MapReduce2 Tez ZooKeeper **Ambari Metrics** SmartSense **1** Misc

Group: Default (3) Manage Config Groups Filter...

General

Grafana Admin Password:

Attention: Some configurations need your attention before you can proceed.
Showing properties with issues. [Show all properties](#)

[← Back](#) [Next →](#)

Customize Services

We have come up with recommended configurations for the services you selected. Customize them as you see fit.

HDFS YARN MapReduce2 Tez ZooKeeper **Ambari Metrics** SmartSense Misc

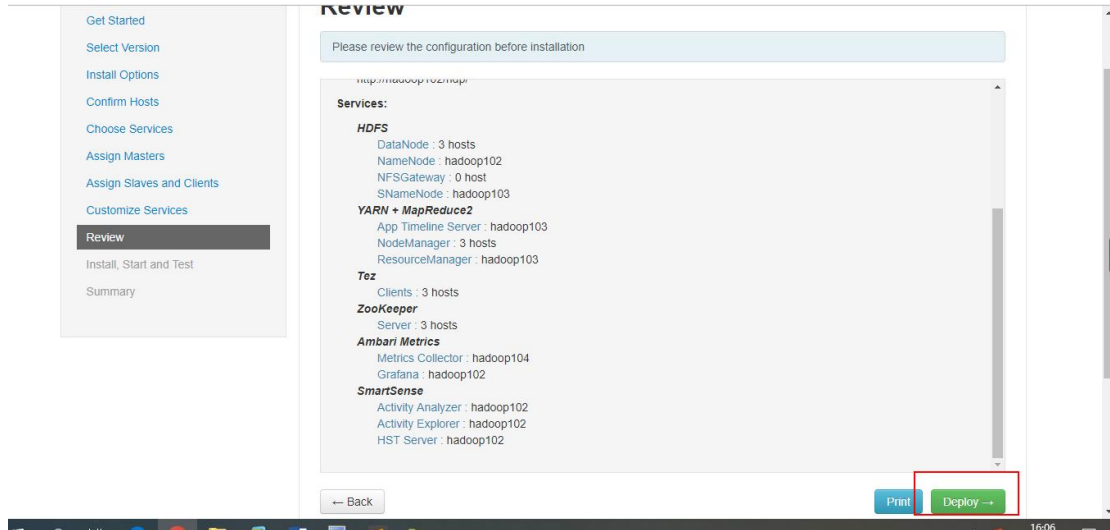
Group: Default (3) Manage Config Groups Filter...

Basic Data Capture Operations Gateway **Activity Analysis** Advanced

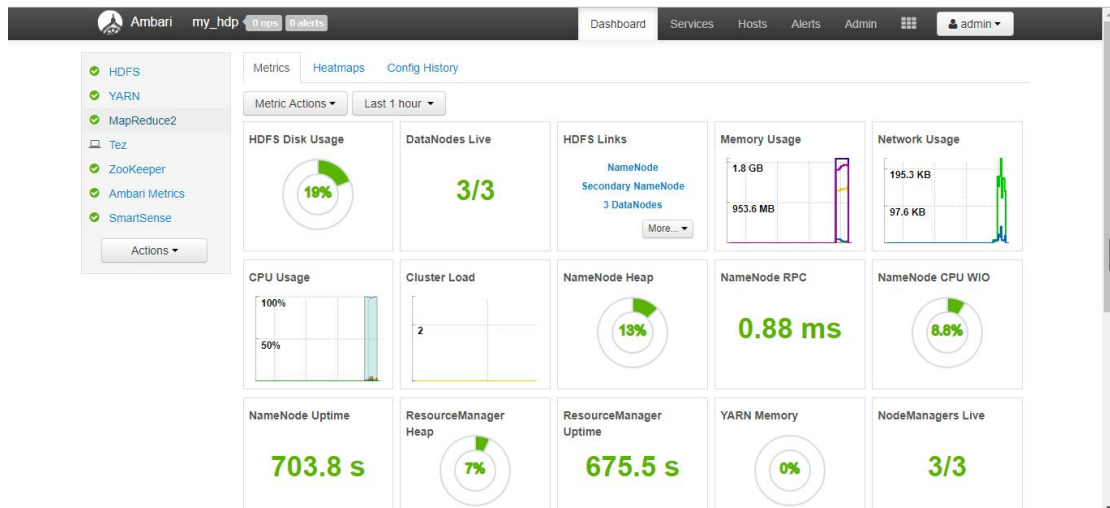
Activity Explorer

Password for user 'admin':

4.1.12 继续



4.1.13 等待服务安装和启动



4.2 安装 Hive

4.2.1 在 /path/to/mysql/ 目录下放置 Mysql 驱动，并重命名为：

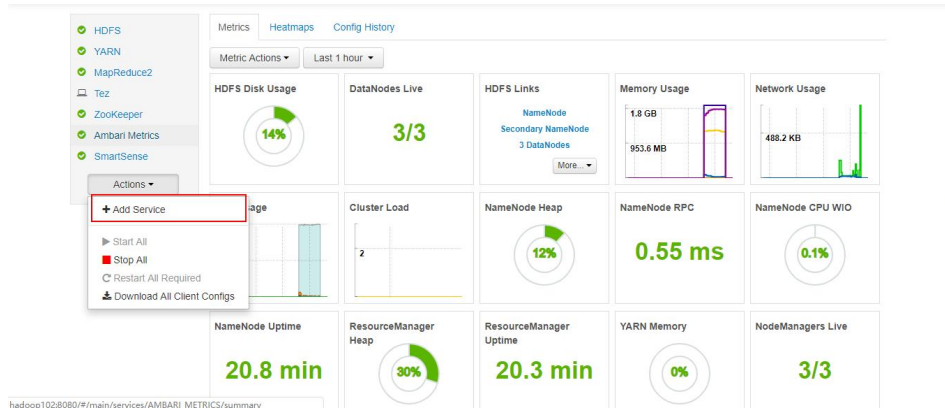
mysql-connector-java.jar

```
[root@hadoop102 software]# mkdir -p /path/to/mysql/
[root@hadoop102 software]# cp /opt/software/mysql-lib/mysql-connector-java-5.1.27/mysql-connector-java-5.1.27-bin.jar /path/to/mysql/mysql-connector-java.jar
```

4.2.2 执行 ambari-server setup

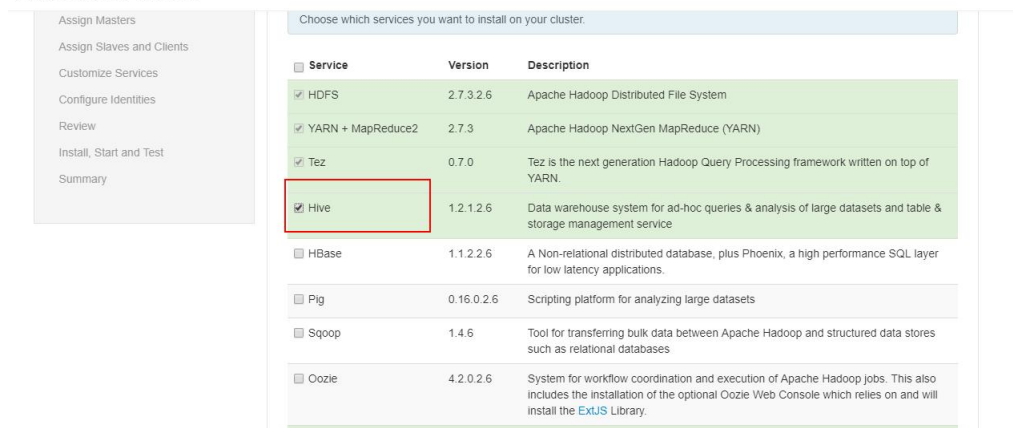
```
[root@hadoop102 software]# ambari-server setup --jdbc-db=mysql --jdbc-driver=/path/to/mysql/mysql-connector-java.jar
```

4.2.3 点击添加服务



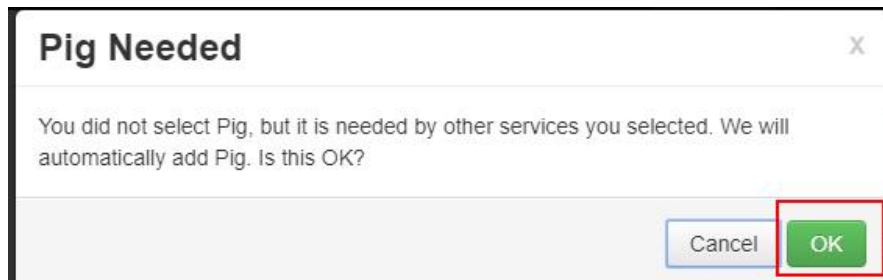
4.3.4 选择 hive

Add Service Wizard



The 'Add Service Wizard' dialog is shown. On the left, a sidebar lists steps: Assign Masters, Assign Slaves and Clients, Customize Services, Configure Identities, Review, Install, Start and Test, and Summary. The main area is titled 'Choose which services you want to install on your cluster.' and contains a table of services:

Service	Version	Description
<input checked="" type="checkbox"/> HDFS	2.7.3.2.6	Apache Hadoop Distributed File System
<input checked="" type="checkbox"/> YARN + MapReduce2	2.7.3	Apache Hadoop NextGen MapReduce (YARN)
<input checked="" type="checkbox"/> Tez	0.7.0	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
<input checked="" type="checkbox"/> Hive	1.2.1.2.6	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
<input type="checkbox"/> HBase	1.1.2.2.6	A Non-relational distributed database, plus Phoenix, a high performance SQL layer for low latency applications.
<input type="checkbox"/> Pig	0.16.0.2.6	Scripting platform for analyzing large datasets
<input type="checkbox"/> Sqoop	1.4.6	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
<input type="checkbox"/> Oozie	4.2.0.2.6	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 EXUS Library.



The 'Pig Needed' dialog box is shown. It contains the text: 'You did not select Pig, but it is needed by other services you selected. We will automatically add Pig. Is this OK?'. At the bottom right, there are two buttons: 'Cancel' and 'OK'. The 'OK' button is highlighted with a red box.

4.3.5 分配节点

ADD SERVICE WIZARD

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Configure Identities

Review

Install, Start and Test

Summary

Assign Masters

Assign master components to hosts you want to run them on.
* HiveServer2 and WebHCat Server will be hosted on the same host.

NameNode:

SNameNode:

App Timeline Server:

ResourceManager:

History Server:

Hive Metastore:

WebHCat Server:

hadoop102 (1.9 GB, 8 cores)

NameNode

ZooKeeper Server

Grafana

Activity Analyzer

Activity Explorer

HST Server

hadoop103 (1.9 GB, 8 cores)

SNameNode

App Timeline Server

ResourceManager

History Server

Hive Metastore

WebHCat Server

HiveServer2

ZooKeeper Server

hadoop104 (1.9 GB, 8 cores)

4.3.6 分配从属和客户端

ADD SERVICE WIZARD

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Configure Identities

Review

Install, Start and Test

Summary

Assign Slaves and Clients

Assign slave and client components to hosts you want to run them on.
Hosts that are assigned master components are shown with *.
"Client" will install HCat Client, Hive Client, Pig Client and Slider Client.

Host	all none	all none	all none	all none
hadoop102 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client
hadoop103 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client
hadoop104 *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client

Show: 25

1 - 3 of 3

← Back

Next →

4.3.7 配置 hive 数据库连接

ADD SERVICE WIZARD

Choose Services

Assign Masters

Assign Slaves and Clients

Customize Services

Configure Identities

Review

Install, Start and Test

Summary

Customize Services

We have come up with recommended configurations for the services you selected. Customize them as you see fit.

HDFS

YARN

MapReduce2

Tez

Hive

Pig

ZooKeeper

Ambari Metrics

SmartSense

Slider

Misc

There are 4 configuration changes in 1 service [Show Details](#)

Group: Default (3)

Manage Config Groups

Filter...

Settings: Advanced

Security

Choose Authorization: None

Interactive Query

Enable Interactive Query (requires YARN pre-emption)

ACID Transactions

ACID Transactions: On

Add Service Wizard

Hive Database

☒ New MySQL Database

☒ Existing MySQL / MariaDB Database

☐ Existing PostgreSQL Database

☐ Existing Oracle Database

☐ Existing SQL Anywhere Database

Be sure you have run:
ambari-server setup --jdbc-db=mysql --jdbc-driver=/path/to/mysql/mysql-connector-java.jar on the Ambari Server host to make the JDBC driver available and to enable testing the database connection.

Database Name:

Database Username:

Database Password:

JDBC Driver Class:

Database URL:

Connection OK

Hive Database Type:

4.3.8 点击下一步

Add Service Wizard

ADD SERVICE WIZARD

[Choose Services](#)

[Assign Masters](#)

[Assign Slaves and Clients](#)

[Customize Services](#)

[Configure Identities](#)

Review

[Install, Start and Test](#)

[Summary](#)

Review

Please review the configuration before installation

Admin Name : admin

Cluster Name : cluster

Total Hosts : 3 (0 new)

Repositories:

redhat6 (HDP-2.6):
http://hadoop102/hdp/HDP/centos6/

redhat6 (HDP-UTILS-1.1.0.21):
http://hadoop102/hdp/

Services:

Hive

Metastore : hadoop103

HiveServer2 : hadoop103

WebHcat Server : hadoop103

Database : Existing MySQL / MariaDB Database

4.3.9 等待安装

Add Service Wizard

ADD SERVICE WIZARD

[Choose Services](#)

[Assign Masters](#)

[Assign Slaves and Clients](#)

[Customize Services](#)

[Configure Identities](#)

[Review](#)

Install, Start and Test

[Summary](#)

Install, Start and Test

Please wait while the selected services are installed and started.

3 % overall

Host	Status	Message
hadoop102	3%	Waiting to install HCat Client
hadoop103	3%	Waiting to install HCat Client
hadoop104	3%	Waiting to install HCat Client

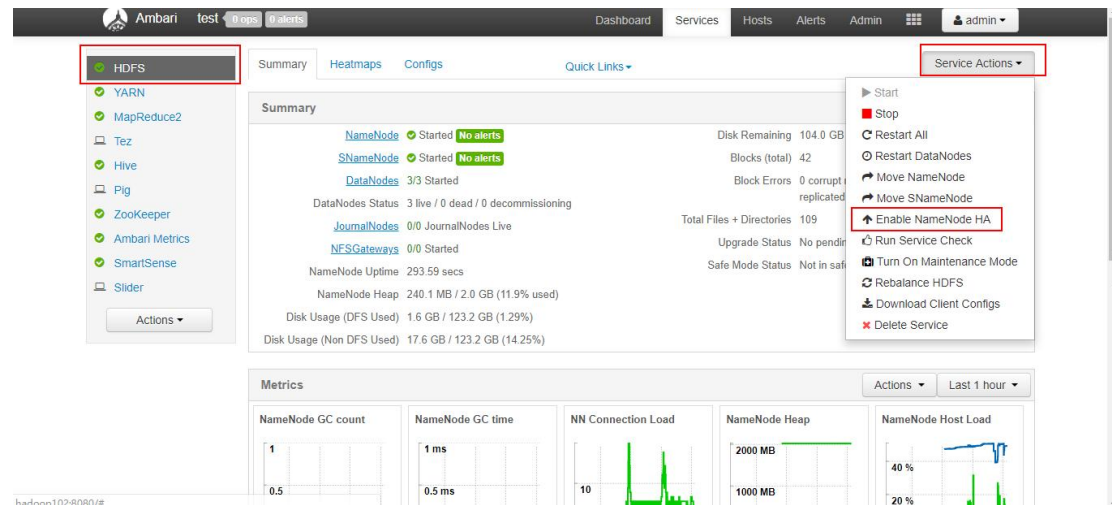
3 of 3 hosts showing - [Show All](#)

Show: [All \(3\)](#) | [In Progress \(3\)](#) | [Warning \(0\)](#) | [Success \(0\)](#) | [Fail \(0\)](#)

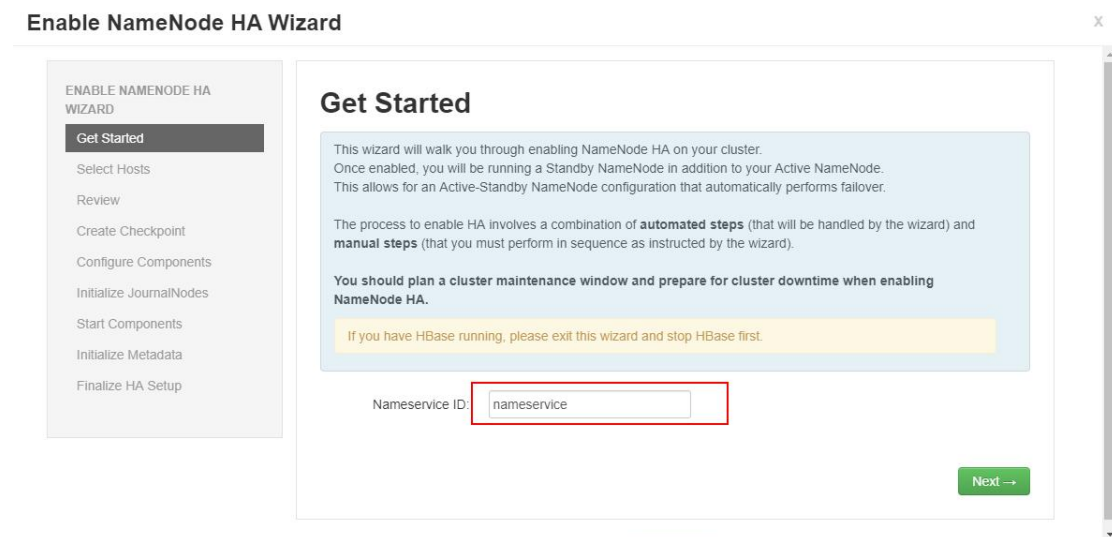
Show: 25 1 - 3 of 3

4.3 配置 HDFS-HA

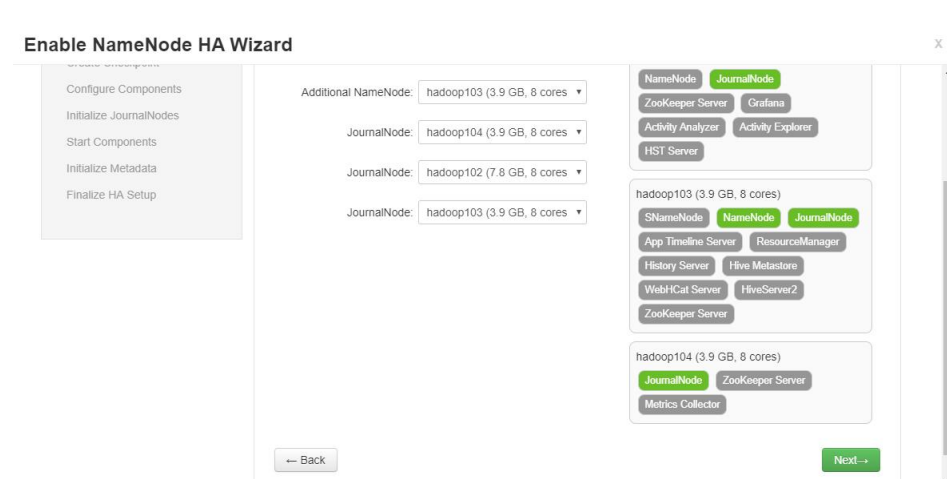
4.3.1 添加服务



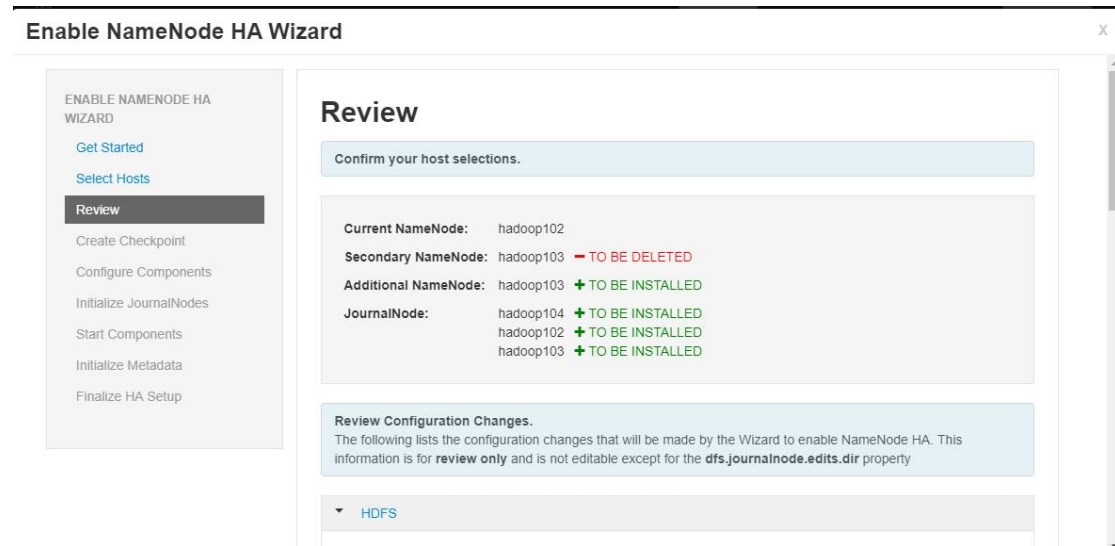
4.3.2 填写 nameservice 名称



4.3.3 角色分配

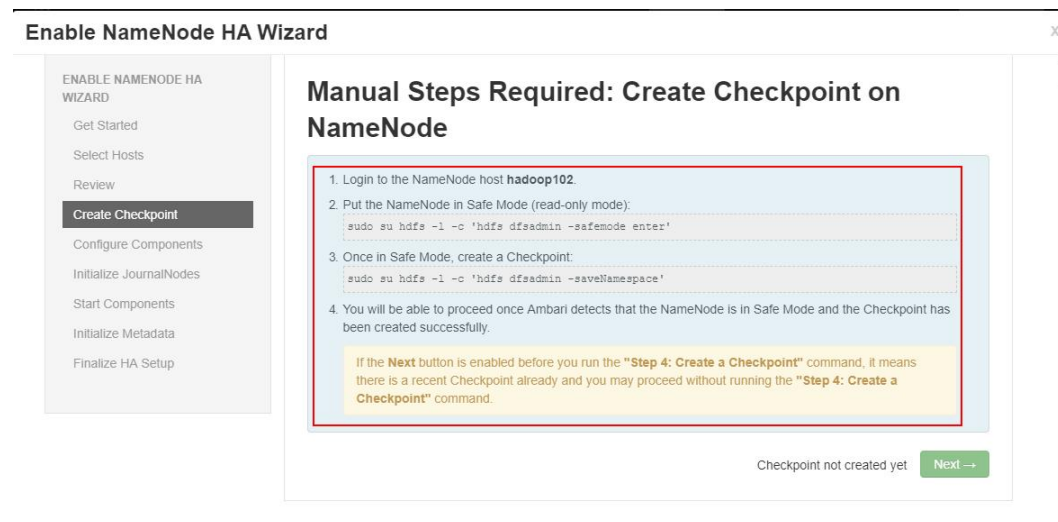


4.3.4 直接下一步



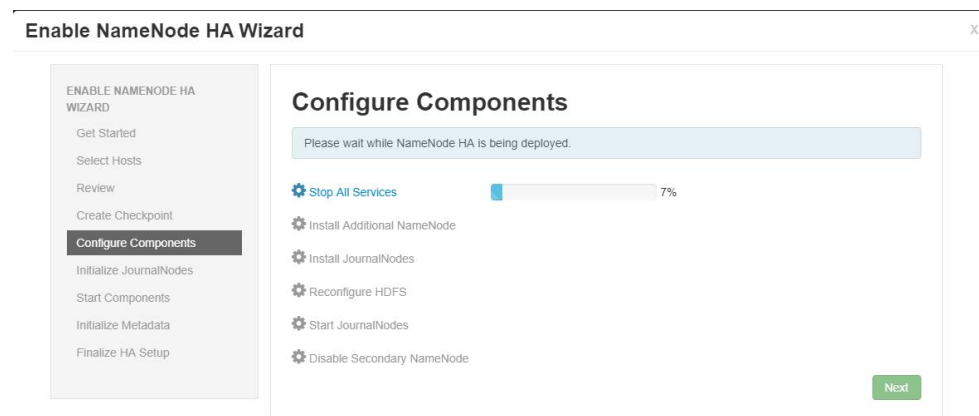
The screenshot shows the 'Review' step of the 'Enable NameNode HA Wizard'. On the left sidebar, the steps are: Get Started, Select Hosts, Review (selected), Create Checkpoint, Configure Components, Initialize JournalNodes, Start Components, Initialize Metadata, and Finalize HA Setup. The main content area is titled 'Review' and contains two sections. The first section, 'Confirm your host selections.', lists the following configurations: Current NameNode: hadoop102; Secondary NameNode: hadoop103 (marked 'TO BE DELETED'); Additional NameNode: hadoop103 (marked 'TO BE INSTALLED'); and JournalNode: hadoop104, hadoop102, and hadoop103 (all marked 'TO BE INSTALLED'). The second section, 'Review Configuration Changes.', states that the following lists the configuration changes that will be made by the Wizard to enable NameNode HA, and that this information is for review only and is not editable except for the dfs.journalnode.edits.dir property. Below this, there is a dropdown menu showing 'HDFS'.

4.3.5 安装提示执行命令



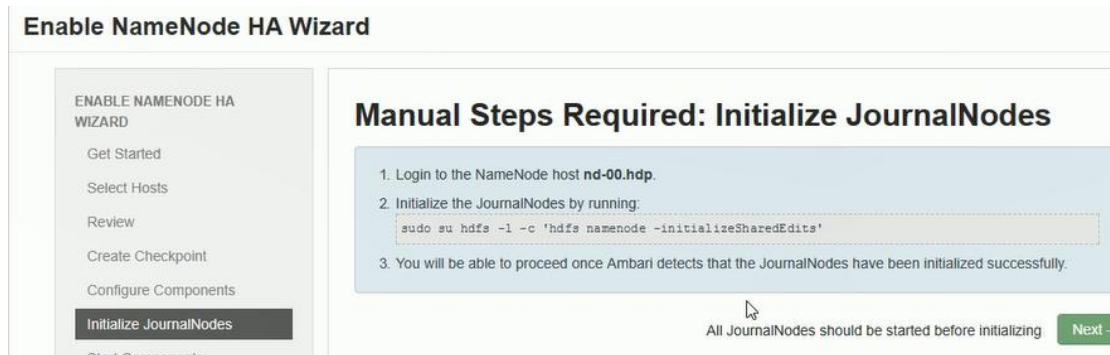
The screenshot shows the 'Manual Steps Required: Create Checkpoint on NameNode' step of the 'Enable NameNode HA Wizard'. On the left sidebar, the steps are: Get Started, Select Hosts, Review, Create Checkpoint (selected), Configure Components, Initialize JournalNodes, Start Components, Initialize Metadata, and Finalize HA Setup. The main content area is titled 'Manual Steps Required: Create Checkpoint on NameNode' and contains a list of four steps: 1. Login to the NameNode host hadoop102. 2. Put the NameNode in Safe Mode (read-only mode): `sudo su hdfs -l -c 'hdfs dfsadmin -safemode enter'`. 3. Once in Safe Mode, create a Checkpoint: `sudo su hdfs -l -c 'hdfs dfsadmin -saveNamespace'`. 4. You will be able to proceed once Ambari detects that the NameNode is in Safe Mode and the Checkpoint has been created successfully. Below the list, there is a yellow box with text: 'If the Next button is enabled before you run the "Step 4: Create a Checkpoint" command, it means there is a recent Checkpoint already and you may proceed without running the "Step 4: Create a Checkpoint" command.' At the bottom right, there is a green 'Next' button and a status indicator 'Checkpoint not created yet'.

4.3.6 安装组件

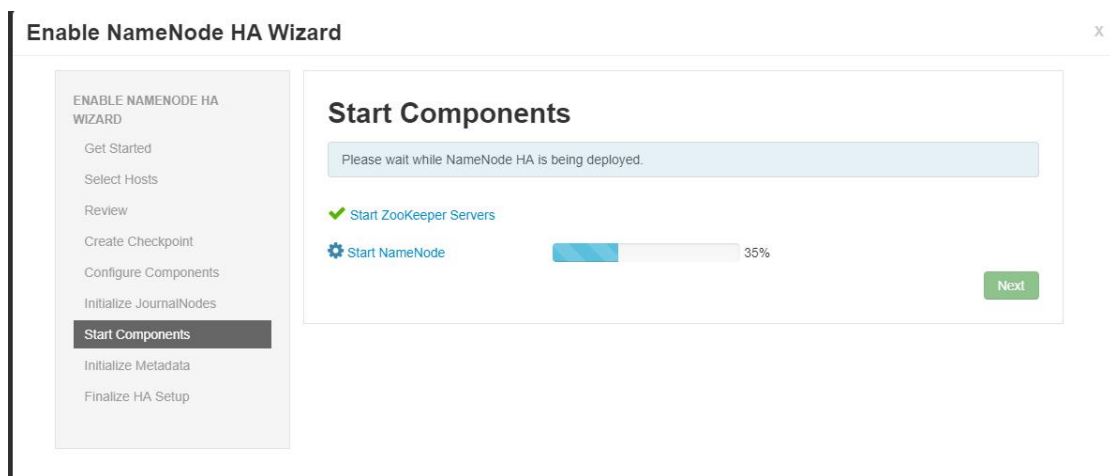


The screenshot shows the 'Configure Components' step of the 'Enable NameNode HA Wizard'. On the left sidebar, the steps are: Get Started, Select Hosts, Review, Create Checkpoint, Configure Components (selected), Initialize JournalNodes, Start Components, Initialize Metadata, and Finalize HA Setup. The main content area is titled 'Configure Components' and contains a progress bar showing 'Please wait while NameNode HA is being deployed.' with a progress indicator at 7%. Below the progress bar, there is a list of components to be configured: Stop All Services, Install Additional NameNode, Install JournalNodes, Reconfigure HDFS, Start JournalNodes, and Disable Secondary NameNode. At the bottom right, there is a green 'Next' button.

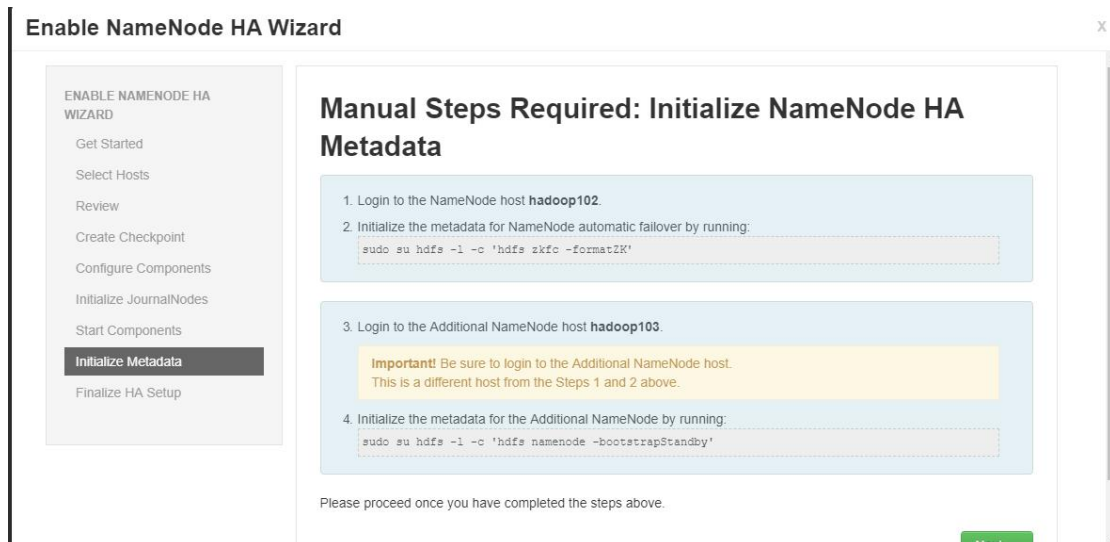
4.3.7 根据页面提示操作



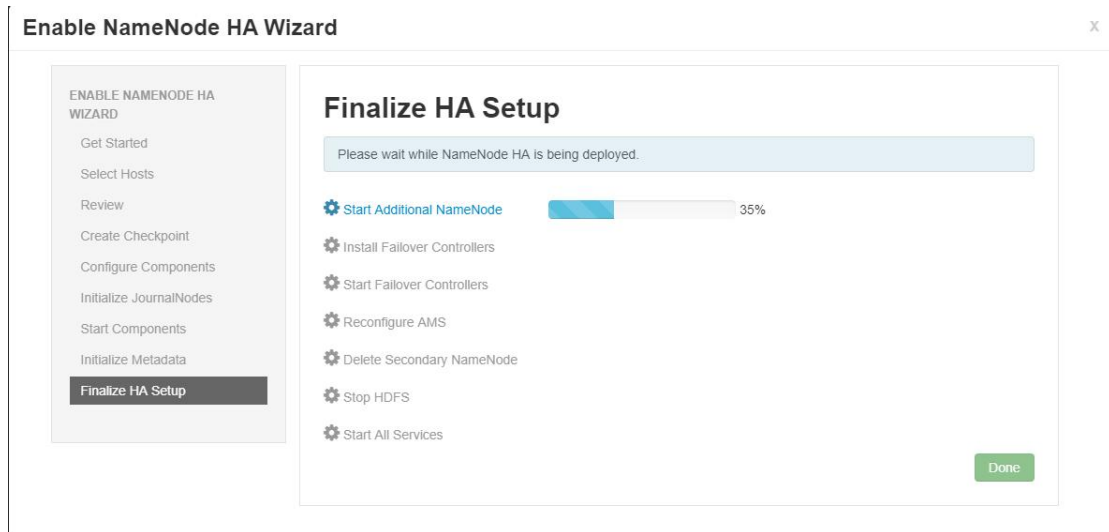
4.3.8 等待启动组件



4.3.9 安装页面提示初始化 NameNode HA



4.3.10 等待完成



报 select 错误

yum -y erase hdp-select