项目管理文档

BJCMDI-HDP安装部署手册

* 编写人：WangPan
* 编写时间：2017/11/29
* 审批人：
* 审批时间：
* 版本：V1.0

**文档版本更新记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **日期** | **版本号** | **更新内容** | **负责人** | **审核人** |
| 2017-11-29 | V1.0 | HDP部署安装手册创建 | WangPan | WangZheng |
| 2017-12-5 | V1.1 | 新增3.3.3安装必须的rpm包 | WangPan | WangZheng |
| 2017-12-12 | V1.2 | 新增parted分区方式（GPT格式） | WangPan | WangZheng |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目录**

[1. 引言 5](#_Toc500844189)

[1.1 文档用途 5](#_Toc500844190)

[1.2 阅读对象 5](#_Toc500844191)

[1.3 名词术语 5](#_Toc500844192)

[1.4 参考资料 6](#_Toc500844193)

[2. Hadoop历史版本 6](#_Toc500844194)

[2.1 系统架构 7](#_Toc500844195)

[2.2 硬件网络拓扑 7](#_Toc500844196)

[2.3 安装介质 8](#_Toc500844197)

[2.4 运行硬件环境 8](#_Toc500844198)

[2.5 服务器应用规划 9](#_Toc500844199)

[3. 操作系统环境准备 9](#_Toc500844200)

[3.1 安装操作系统 9](#_Toc500844201)

[3.2 SSH免密钥登录配置 10](#_Toc500844202)

[3.3 配置本地yum源 11](#_Toc500844203)

[3.3.1 安装httpd服务 11](#_Toc500844204)

[3.3.2 配置系统的本地yum源 11](#_Toc500844205)

[3.3.3 安装必须的rpm包 11](#_Toc500844206)

[3.3.4 配置Ambari和HDP的本地yum源 11](#_Toc500844207)

[3.4 配置JAVA环境 12](#_Toc500844208)

[3.5 配置NTP服务 13](#_Toc500844209)

[3.6 关闭防火墙和SELinux 13](#_Toc500844210)

[3.7 配置主机名和hosts文件 14](#_Toc500844211)

[3.8 安装mysql数据库和对应驱动 14](#_Toc500844212)

[4. 开始安装Ambari-Server 14](#_Toc500844213)

[4.1 配置mysql数据库 14](#_Toc500844214)

[4.2 安装ambari server 14](#_Toc500844215)

[4.3 设置ambari server 15](#_Toc500844216)

[4.3.1 检查SELinux是否关闭，如果关闭不用操作 15](#_Toc500844217)

[4.3.2 提示是否自定义设置 15](#_Toc500844218)

[4.3.3 设置ambari-server 账号 15](#_Toc500844219)

[4.3.4 检查防火墙，如果关闭则不用操作 15](#_Toc500844220)

[4.3.5 设置JDK 15](#_Toc500844221)

[4.3.6 自定义JDK 16](#_Toc500844222)

[4.3.7 数据库配置 16](#_Toc500844223)

[4.3.8 选择数据库类型。 16](#_Toc500844224)

[4.3.9 Mysql数据设置 17](#_Toc500844225)

[4.3.10 将Ambari数据库脚本导入到数据库 17](#_Toc500844226)

[5. 安装HDP2.6.3 17](#_Toc500844227)

[5.1 Welcome to Apache Ambari 17](#_Toc500844228)

[5.2 Get Started 18](#_Toc500844229)

[5.3 Select Version 18](#_Toc500844230)

[5.4 Repositories 19](#_Toc500844231)

[5.5 Install Options 19](#_Toc500844232)

[5.6 Confirm Hosts 20](#_Toc500844233)

[5.7 Assign Masters 20](#_Toc500844234)

[5.8 Assign Slaves and Clients 21](#_Toc500844235)

[5.9 Customize Services 21](#_Toc500844236)

[5.10 Review 22](#_Toc500844237)

[5.11 Install,Start and Test 22](#_Toc500844238)

[5.12 Summary 23](#_Toc500844239)

[6. 附录 24](#_Toc500844240)

# 引言

## 文档用途

为项目部署和维护工程师提供标准化操作

## 阅读对象

【内容】

此文档将适合哪些人员阅读。如：

* 产品设计工程师
* 产品开发工程师
* 产品测试工程师
* 其他相关人员

## 名词术语

* Hadoop 1.0

第一代Hadoop，由分布式存储系统HDFS和分布式计算框架MapReduce组成，其中，HDFS由一个NameNode和多个DataNode组成，MapReduce由一个JobTracker和多个TaskTracker组成，对应Hadoop版本为Hadoop 1.x和0.21.X，0.22.x。

* Hadoop 2.0

第二代Hadoop，为克服Hadoop 1.0中HDFS和MapReduce存在的各种问题而提出的。针对Hadoop 1.0中的单NameNode制约HDFS的扩展性问题，提出了HDFS Federation，它让多个NameNode分管不同的目录进而实现访问隔离和横向扩展；针对Hadoop 1.0中的MapReduce在扩展性和多框架支持方面的不足，提出了全新的资源管理框架YARN(Yet Another Resource Negotiator)，它将JobTracker中的资源管理和作业控制功能分开，分别由组件ResourceManager和ApplicationMaster实现，其中，ResourceManager负责所有应用程序的资源分配，而ApplicationMaster仅负责管理一个应用程序。对应Hadoop版本为Hadoop 0.23.x和2.x。

* MapReduce 1.0或者MRv1（MapReduce version 1）

第一代MapReduce计算框架，它由两部分组成：编程模型（programming model）和运行时环境（runtime environment）。它的基本编程模型是将问题抽象成Map和Reduce两个阶段，其中Map阶段将输入数据解析成key/value，迭代调用map()函数处理后，再以key/value的形式输出到本地目录，而Reduce阶段则将key相同的value进行规约处理，并将最终结果写到HDFS上。它的运行时环境由两类服务组成：JobTracker和TaskTracker，其中，JobTracker负责资源管理和所有作业的控制，而TaskTracker负责接收来自JobTracker的命令并执行它。

* MapReduce 2.0或者MRv2（MapReduce version 2）或者NextGen MapReduc

MapReduce 2.0或者MRv2具有与MRv1相同的编程模型，唯一不同的是运行时环境。MRv2是在MRv1基础上经加工之后，运行于资源管理框架YARN之上的MRv1，它不再由JobTracker和TaskTracker组成，而是变为一个作业控制进程ApplicationMaster，且ApplicationMaster仅负责一个作业的管理，至于资源的管理，则由YARN完成。

简而言之，MRv1是一个独立的离线计算框架，而MRv2则是运行于YARN之上的MRv1。

* YARNe

Hadoop 2.0中的资源管理框架，它是一个框架管理器，为各种框架进行资源分配和提供运行时环境。而MRv2则是运行在YARN之上的第一个计算框架，其他计算框架，比如Spark、Storm等，都正在往YARN上移植。YARN类似于几年前的资源管理系统mesos和更早的Torque。

* HDFS Federation

Hadoop 2.0中对HDFS进行了改进，使NameNode可以横向扩展成多个，其中，每个NameNode分管一部分目录，这不仅增强了HDFS的扩展性，也使HDFS具备了隔离性。

## 参考资料

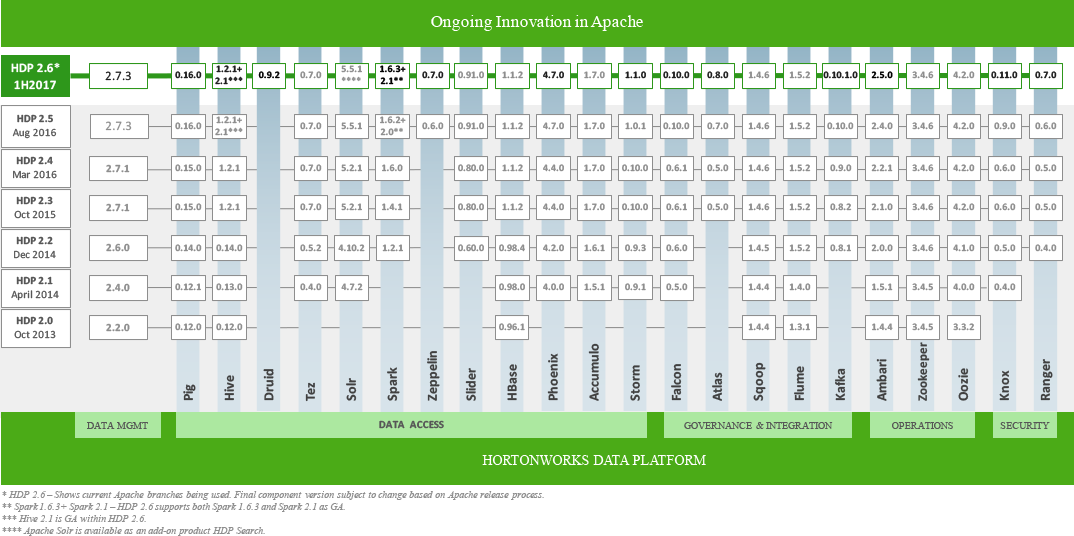
<http://hadoop.apache.org/>

<http://hadoop.apache.org/releases.html>

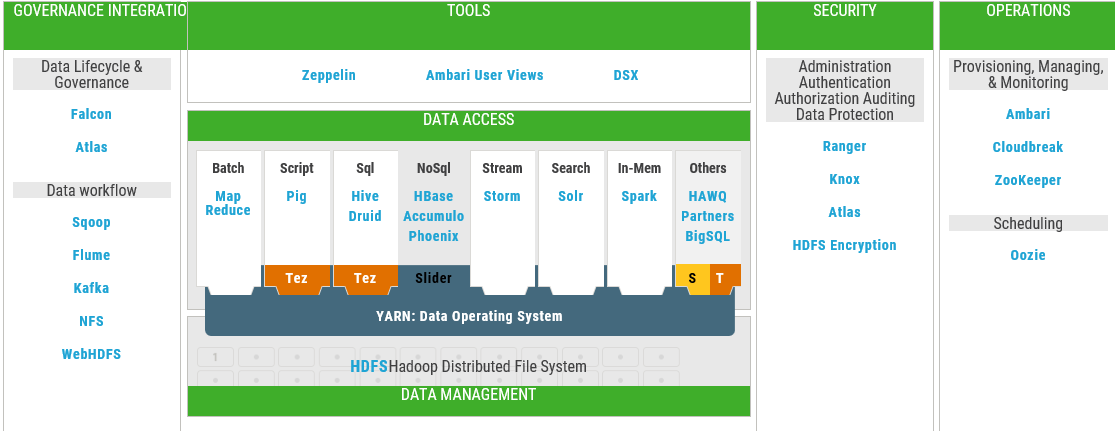
<https://hortonworks.com/products/data-platforms/hdp/>

软件下载地址：https://hortonworks.com/downloads/

# Hadoop历史版本



## 系统架构



## 硬件网络拓扑

*【内容】*

硬件和网络的典型的拓扑结构图

*【裁剪原则】*

*此部分内容不允许裁剪。*

## 安装介质

操作系统版本：

CentOS-7-x86\_64-DVD-1708.iso

软件包：

jdk-8u77-linux-x64.tar.gz

ambari-2.6.0.0-centos7.tar.gz

HDP-2.6.3.0-centos7-rpm.tar.gz

HDP-UTILS-1.1.0.21-centos7.tar.gz

JDK下载：

http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

HDP下载地址：

<https://docs.hortonworks.com/HDPDocuments/Ambari-2.6.0.0/bk_ambari-installation/content/ambari_repositories.html>

<https://docs.hortonworks.com/HDPDocuments/Ambari-2.6.0.0/bk_ambari-installation/content/hdp_26_repositories.html>

http://public-repo-1.hortonworks.com/ambari/centos7/2.x/updates/2.6.0.0/ambari-2.6.0.0-centos7.tar.gz

http://public-repo-1.hortonworks.com/HDP/sles12/2.x/updates/2.6.3.0/HDP-2.6.3.0-sles12-rpm.tar.gz

<http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.21/repos/sles12/HDP-UTILS-1.1.0.21-sles12.tar.gz>

## 运行硬件环境

## 服务器应用规划

|  |  |  |
| --- | --- | --- |
| No | IP | APP |
| 1 | 192.168.168.1 | NameNode |
| 2 | 192.168.168.2 | DataNode1 |
| 3 | 192.168.168.3 | DataNode2 |
| 4 | 192.168.168.4 | DataNode3 |

# 操作系统环境准备

## 安装操作系统

本文档使用的操作系统版本为：CentOS Linux release 7.4.1708 (Core)

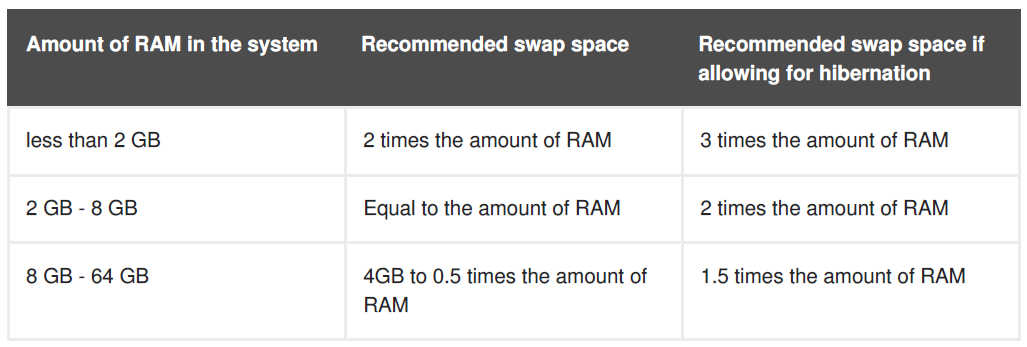
内核版本为：3.10.0-693.el7.x86\_64

**分区注意事项**：

a.采用Minimal方案进行系统安装

b.系统分区和数据分区独立开来，最好系统分区安装有RAID1分区，数据分区全部单盘做RAID0或直通（具体以阵列卡管理界面内的名称为准）

c.系统分区使用LVM管理磁盘，分区建议如下：

swap: 

/boot: 至少500M，要放在普通分区，一定不能放在LVM分区

/: 大小为系统分区的30%

/home: 大小为系统分区的30%

未分配分区：大小为系统分区的30%-40%，因XFS文件系统的LVM管理方式只允许扩大分区大小，不允许减小，留下一部分未分配分区方便后期调整

d.数据分区挂载步骤(对每块数据硬盘都执行此步骤，盘符以实际情况为准)(d和e为二选一，如不关心LVM的额外开销，可选d,否则选择e)

# pvcreate /dev/sdb

# vgcreate lvmb /dev/sdb

# lvcreate -n lvgb -l 100%FREE lvmb

# mkfs.ext4 /dev/lvmb/lvgb

# mkdir /data01

# mount /dev/mapper/lvmb-lvgb /data01

修改/etc/fstab，增加如下内容

/dev/mapper/lvmb-lvgb /data01 ext4 defaults 0 1

e.数据分区挂载步骤(对每块数据硬盘都执行此步骤，盘符以实际情况为准)(d和e为二选一，如不关心LVM的额外开销，可选d,否则选择e)

# parted /dev/sdb

(parted) mklabel gpt

(parted) mkpart

Partition name? []?

File system type? [ext2]? ext4

Start? 1

End? -1

(parted) quit

# mkdir /data01

# mkfs.ext4 /dev/sdb1

# mount /dev/sdb1 /data01

修改/etc/fstab，增加如下内容

/dev/sdb1 /data01 ext4 defaults 0 1

## SSH免密钥登录配置

在ambari server主机上生成一组密钥文件（如已存在，可省略）

# ssh-keygen

全部保持默认选项，则会在/root/.ssh目录下生成id\_rsa和id\_rsa.pub两个文件

将公钥分发至其它节点并进行测试验证

# ssh-copy-id xxx.xxx.xxx.xxx

使用该命令将公钥信息发布至各节点（包括本机），xxx代表节点IP，完成后使用ssh xxx.xxx.xxx.xxx可直接访问各节点则为正常

## 配置本地yum源

### 安装httpd服务

# rpm -ivh httpd.x86\_64-2.4.6-67.el7.rpm

# systemctl start httpd

# systemctl enable httpd

### 配置系统的本地yum源

# mkdir /var/www/html/CentOS7

将系统镜像文件内容拷贝至该目录，并修改/etc/yum.repos.d/ CentOS-Media.repo 文件，确保下列选项的值，标红部分改为对应IP

baseurl=http://ambari-server-ip/CentOS7

enabled=1

### 安装必须的rpm包

* libtirpc
* libtirpc-devel

对应软件包可在如下地址获得：

https://mirrors.aliyun.com/centos/7/os/x86\_64/Packages/

### 配置Ambari和HDP的本地yum源

（假设对应的压缩包在/opt目录下）

# tar -zxvf /opt/ambari-2.6.0.0-centos7.tar.gz -C /var/www/html

# tar -zxvf /opt/HDP-2.6.3.0-centos7-rpm.tar.gz -C /var/www/html

# mkdir /var/www/html/HDP-UTILS

# tar -zxvf /opt/HDP-UTILS-1.1.0.21-centos7.tar.gz -C /var/www/html/HDP-UTILS

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

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

修改标红的两个repo文件，确保对应路径正确，大致如下：

#VERSION\_NUMBER=2.6.3.0-235

[HDP-2.6.3.0]

name=HDP Version - HDP-2.6.3.0

baseurl=http://10.141.128.81/HDP/centos7/2.6.3.0-235

gpgcheck=0

gpgkey=http://10.141.128.81/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://10.141.128.81/HDP-UTILS

gpgcheck=0

gpgkey=http://10.141.128.81/HDP-UTILS/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

enabled=1

priority=1

[ambari-2.6.0.0]

name=ambari Version - ambari-2.6.0.0

baseurl=http://10.141.128.81/ambari/centos7/2.6.0.0-267

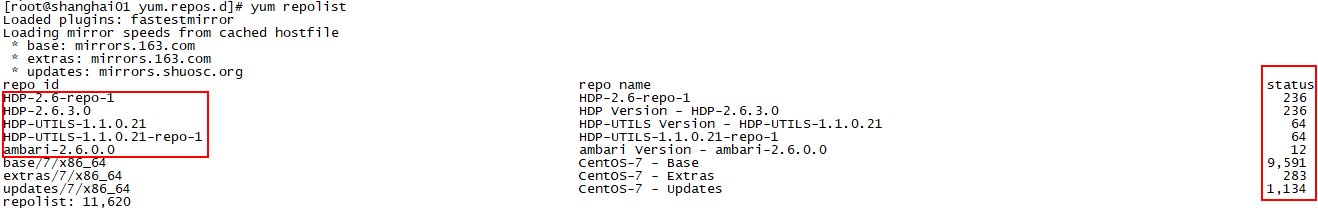
gpgcheck=0

gpgkey=http://10.141.128.81/ambari/centos7/2.6.0.0-267/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins

enabled=1

priority=1

修改完成后使用yum repolist验证源信息是否正确，不用和下面这个图保持一致，安装完成后会自动生成一些新yum源配置文件



## 配置JAVA环境

（假设jdk安装包在/opt目录下）

# mkdir /opt/java

# tar -zxvf /opt/ jdk-8u77-linux-x64.tar.gz -C /opt/java

修改/etc/profile文件，增加如下内容：

export JAVA\_HOME=/opt/java/jdk1.8.0\_77

export CLASSPATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

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

# source /etc/profile

## 配置NTP服务

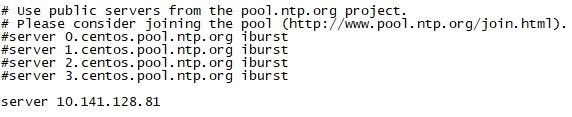
# yum install ntp

# systemctl enable ntpd

在ambari server中修改/etc/ntp.conf文件，注释原有服务器地址，改为指向自己（如局域网内有ntp server也可直接改为对应IP）



在其它节点修改/etc/ntp.conf文件，注释原有服务器地址，改为指向ambari server（如局域网内有ntp server也可直接改为对应IP）



# systemctl start ntp(启动NTP服务)

## 关闭防火墙和SELinux

# systemctl stop firewalld

# systemctl disable firewalld

# systemctl status firewalld

# setenfore 0

修改/etc/selinux/config文件，做如下修改

SELINUX=disabled

## 配置主机名和hosts文件

# hostnamectl set-hostname xxx

在/etc/hosts文件中添加所有节点的FQDN格式主机名（所有节点都要做此操作），大致如下：

10.141.128.81 shanghai01 shanghai01.hadoop

10.141.128.86 shanghai02 shanghai02.hadoop

10.141.128.87 shanghai03 shanghai03.hadoop

## 安装mysql数据库和对应驱动

# yum install -y mysql-connector-java

# yum install -y mariadb mariadb-server

# systemctl enable mariadb.service

# systemctl start mariadb.service

# 开始安装Ambari-Server

## 配置mysql数据库

# mysql -uroot -p

MariaDB [none]> CREATE DATABASE ambari CHARACTER SET UTF8;

MariaDB [none]> CREATE USER 'ambari'@'localhost' IDENTIFIED BY 'Richr00t';

MariaDB [none]> GRANT ALL PRIVILEGES ON \*.\* TO ambari@'rhel1' IDENTIFIED BY 'Richr00t'; //此处的标红部分需要根据实际主机名修改

MariaDB [none]> GRANT ALL PRIVILEGES ON \*.\* TO 'ambari'@'localhost';

MariaDB [none]> FLUSH PRIVILEGES;

## 安装ambari server

# yum install ambari-server

## 设置ambari server

### 检查SELinux是否关闭，如果关闭不用操作

# ambari-server setup

Using python /usr/bin/python

Setup ambari-server

Checking SELinux...

SELinux status is 'disabled'

### 提示是否自定义设置

**输入：y**

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

### 设置ambari-server 账号

**输入：ambari**

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

Adjusting ambari-server permissions and ownership...

### 检查防火墙，如果关闭则不用操作

Checking firewall status...

Redirecting to /bin/systemctl status iptables.service

### 设置JDK

**输入：3**

Checking JDK...

Do you want to change Oracle JDK [y/n] (n)? y

[] Oracle JDK 1.8 + Java Cryptography Extension (JCE) Policy Files 8

[] Oracle JDK 1.7 + Java Cryptography Extension (JCE) Policy Files 7

[] Custom JDK

=================================================================

**Enter choice (1): 3**

### 自定义JDK

如果上面选择3，则需要设置JAVA\_HOME

**输入：/opt/**[**Java**](http://lib.csdn.net/base/java)**/jdk1.8.0\_77**

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/java/jdk1.8.0\_77

Validating JDK on Ambari Server...done.

Completing setup...

### 数据库配置

**选择：y**

Configuring database...

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

### 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**

### Mysql数据设置

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

Hostname (**localhost**):

Port ():**3306**

Database name (**ambari**):

Username (**ambari**):

Enter Database Password (**Ambari-123**): //自行设置密码

### 将Ambari数据库脚本导入到数据库

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

用Ambari用户（上面设置的用户）登录mysql

# mysql -u ambari -p

MariaDB [mysql]> use ambari

MariaDB [mysql]> source /var/lib/ambari-server/resources/Ambari-DDL-MySQL-CREATE.sql

# 安装HDP2.6.3

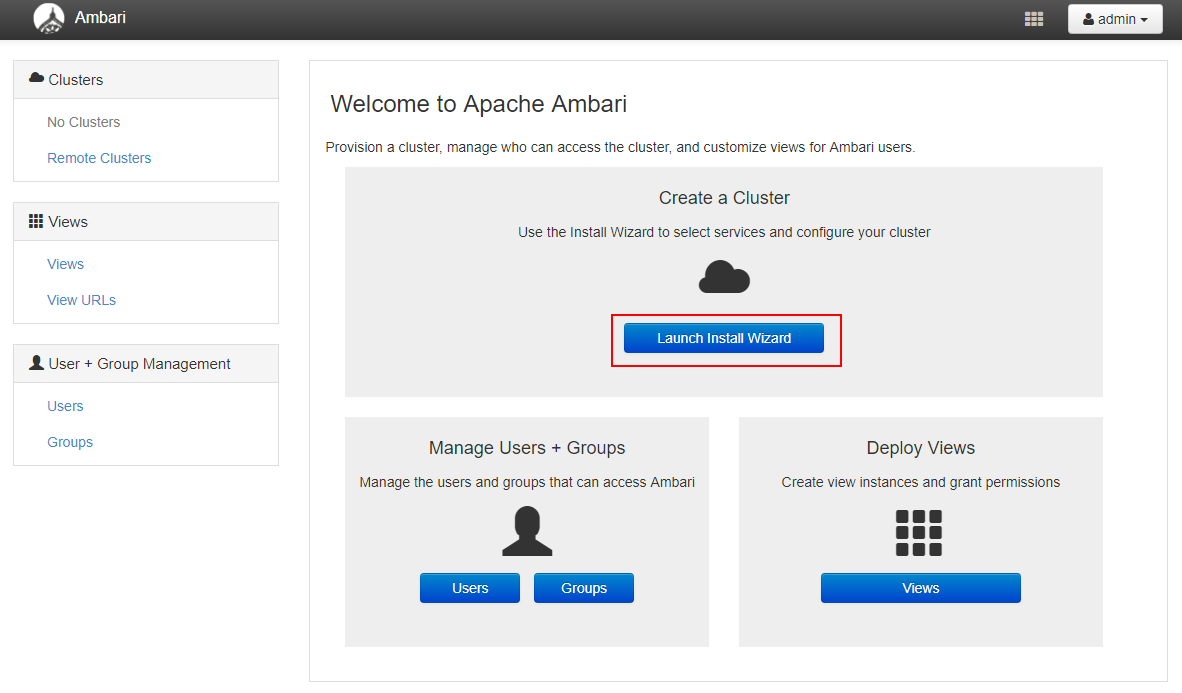
## Welcome to Apache Ambari

使用如下地址访问ambari server控制台

<http://ambari-server:8080>

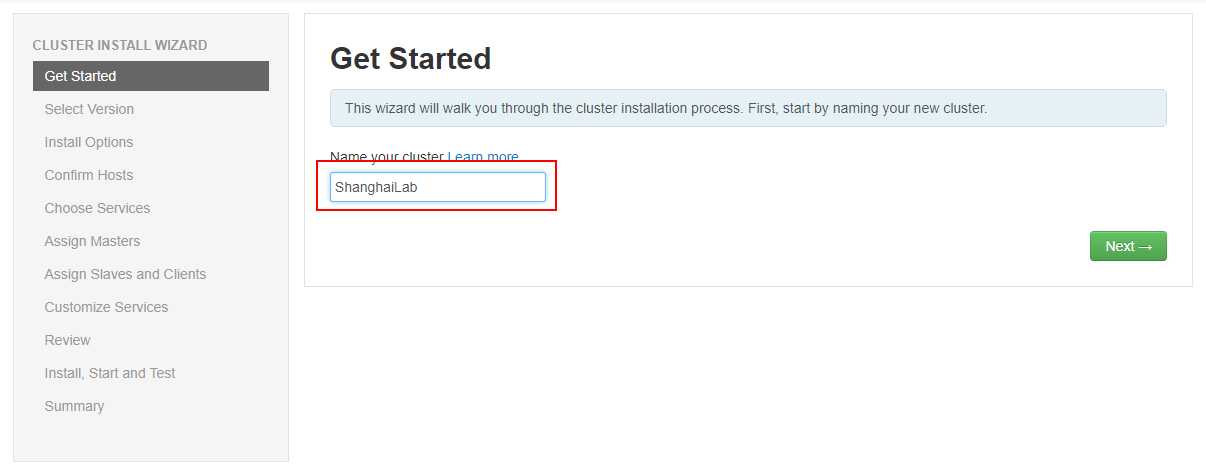
用户名/密码：admin/admin

选择Launch Install Wizard



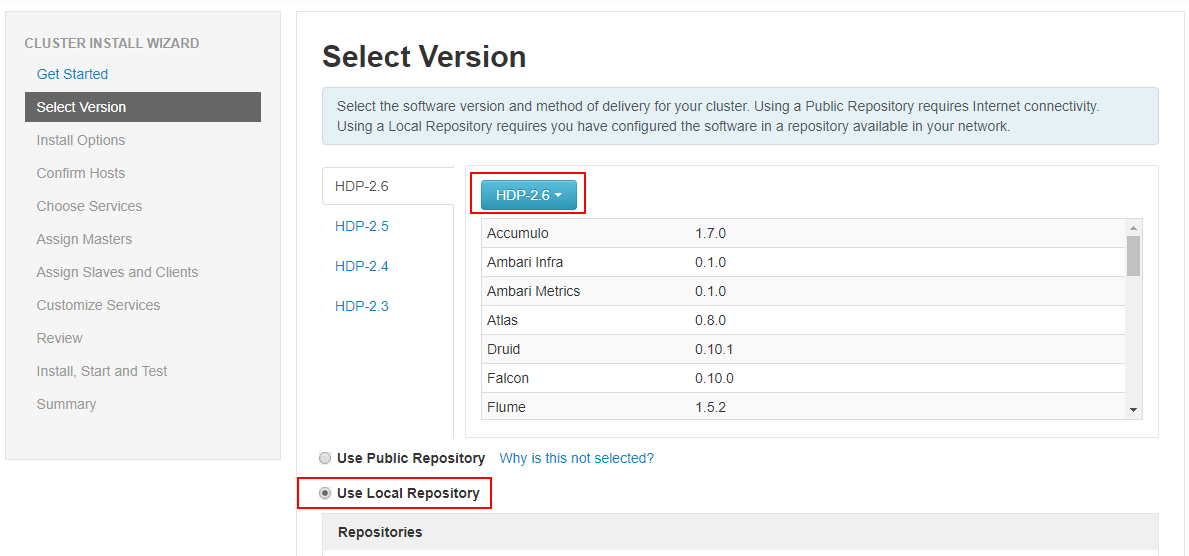
## Get Started

填写Cluster Name



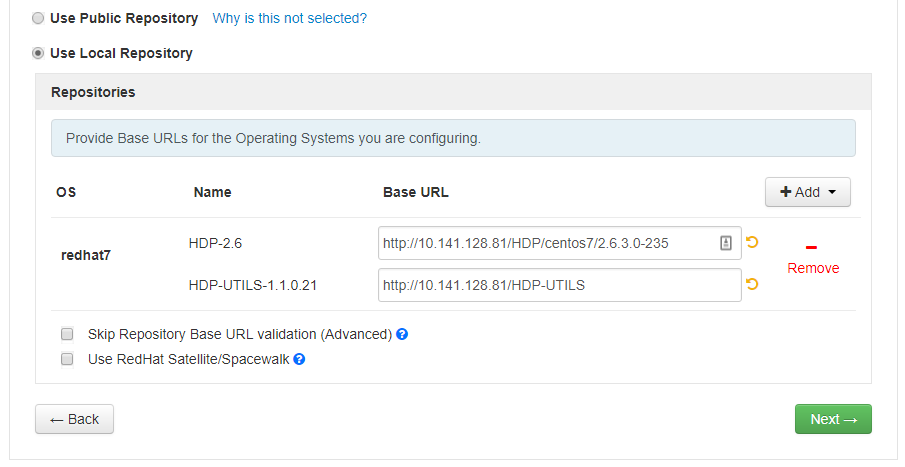
## Select Version

选择对应HDP版本，并改为Use Local Repository



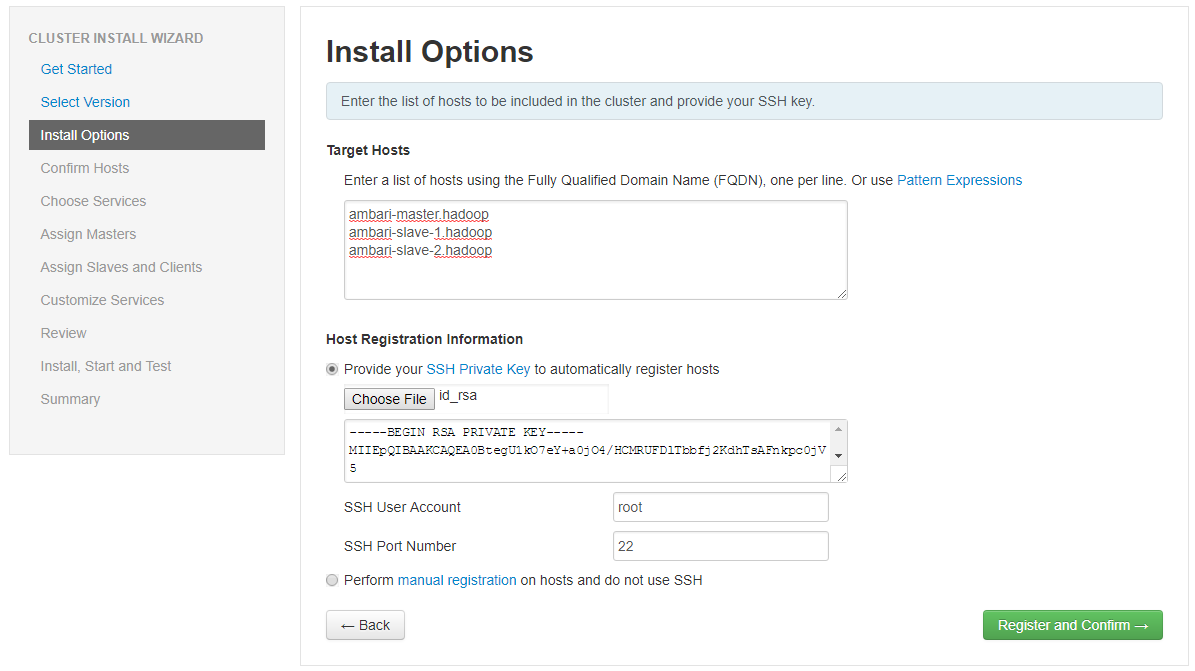
## Repositories

删除多余OS版本，填写对应Base URL(和前面repo文件内容保持一致)



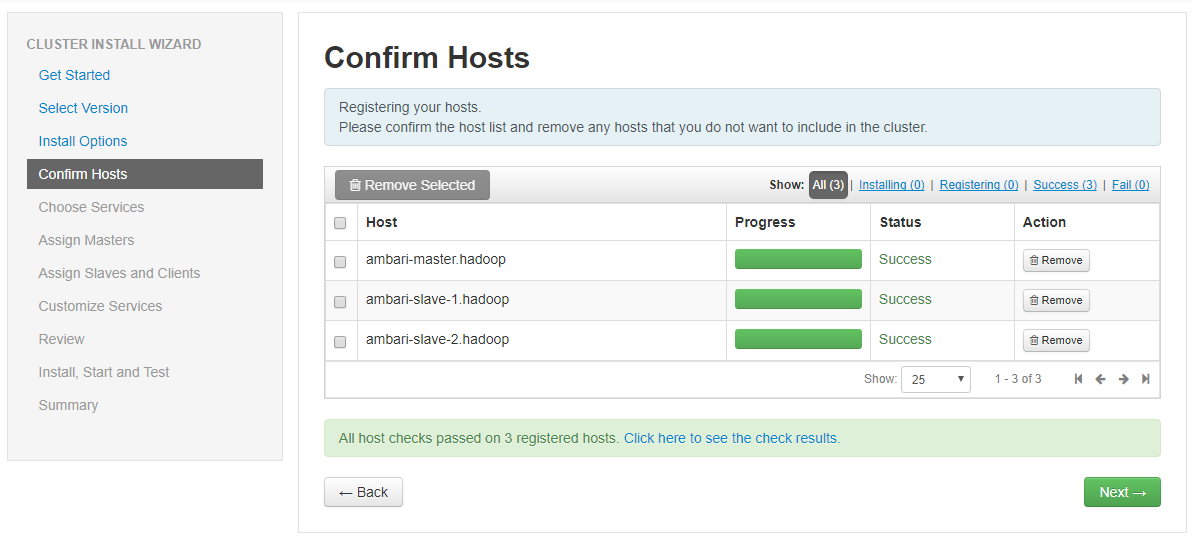
## Install Options

根据/etc/hosts文件，填写FQDN名称（所有节点），将前面步骤生成的私钥文件id\_rsa内容粘贴到Host Registration Information处（仅ambari server主机）



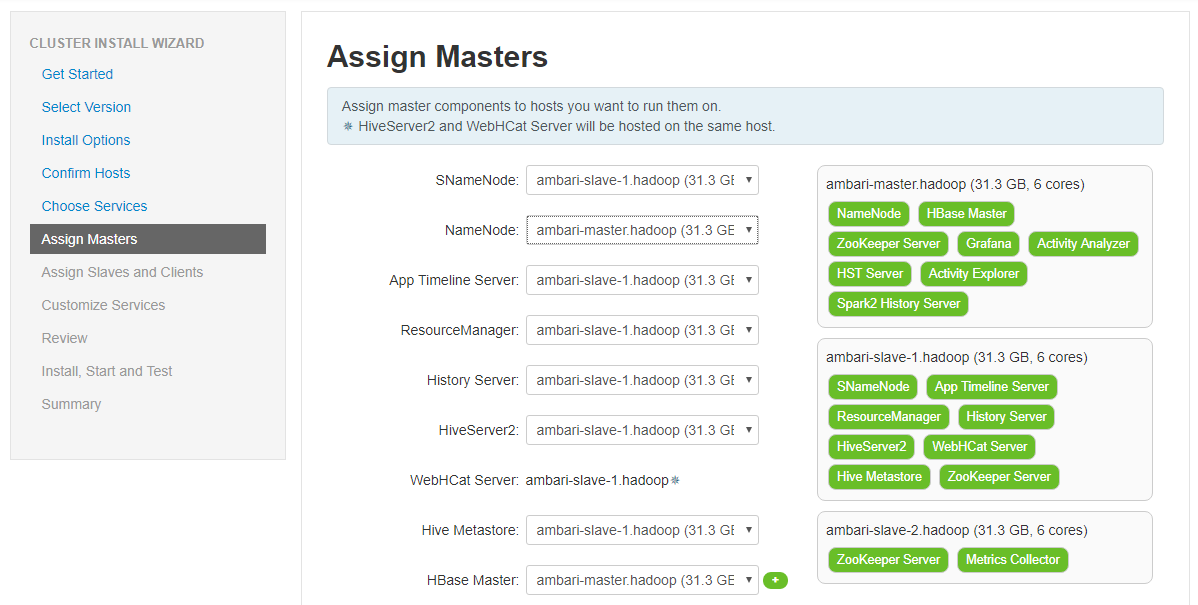
## Confirm Hosts

确保Success且无任何告警



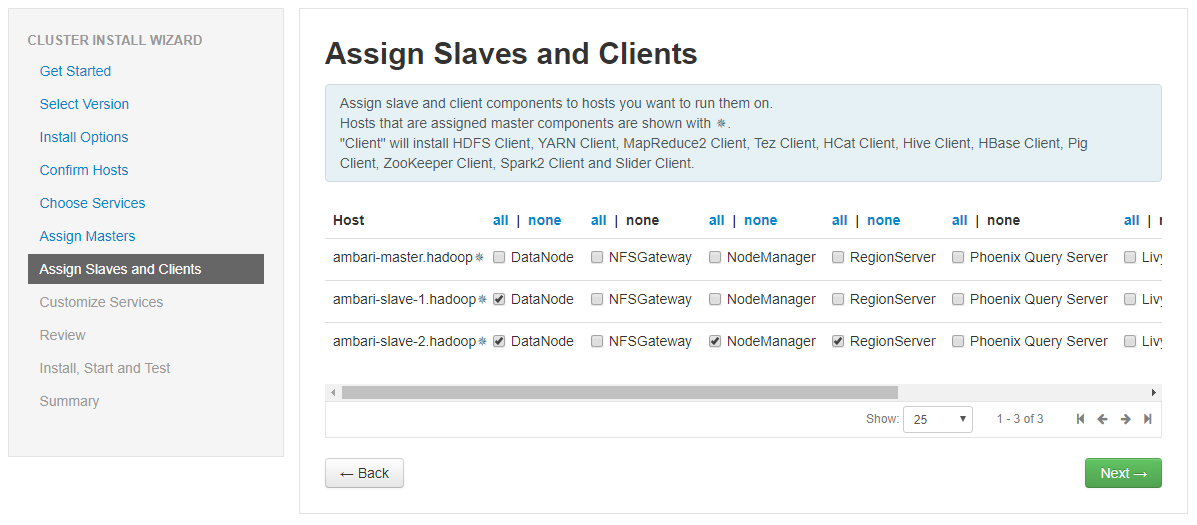
## Assign Masters

分配各Master Component至不同的节点，具体以实际规划为准



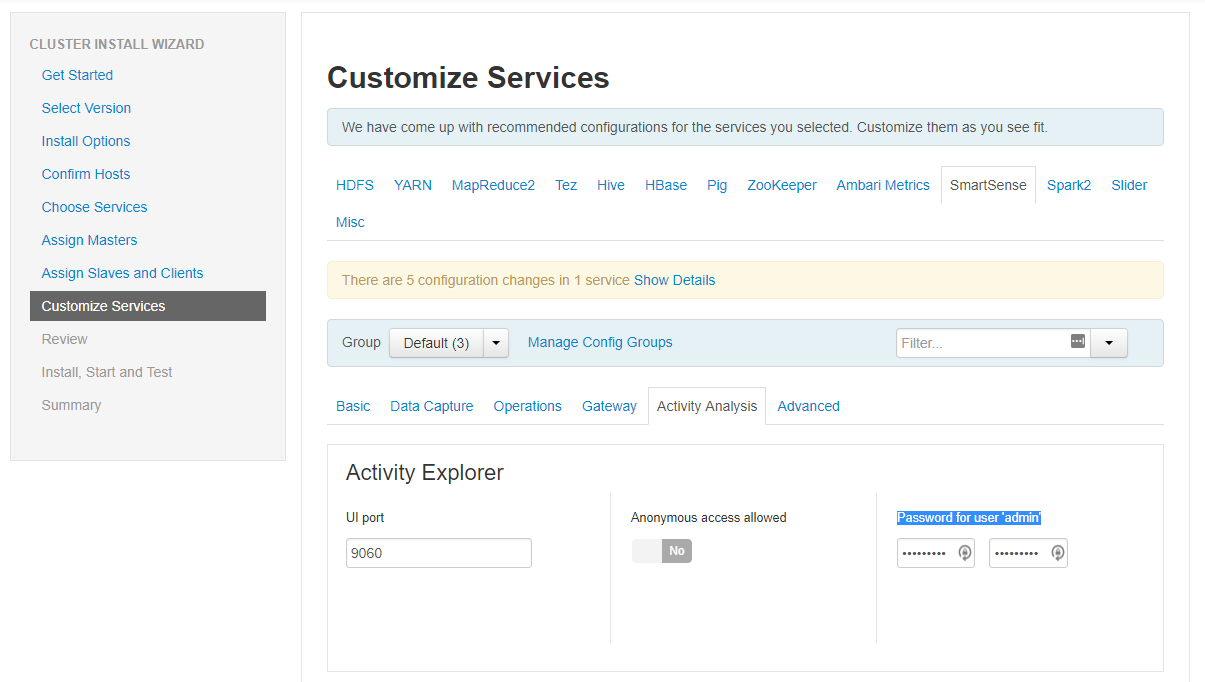
## Assign Slaves and Clients

分配slave和client component至不同的节点

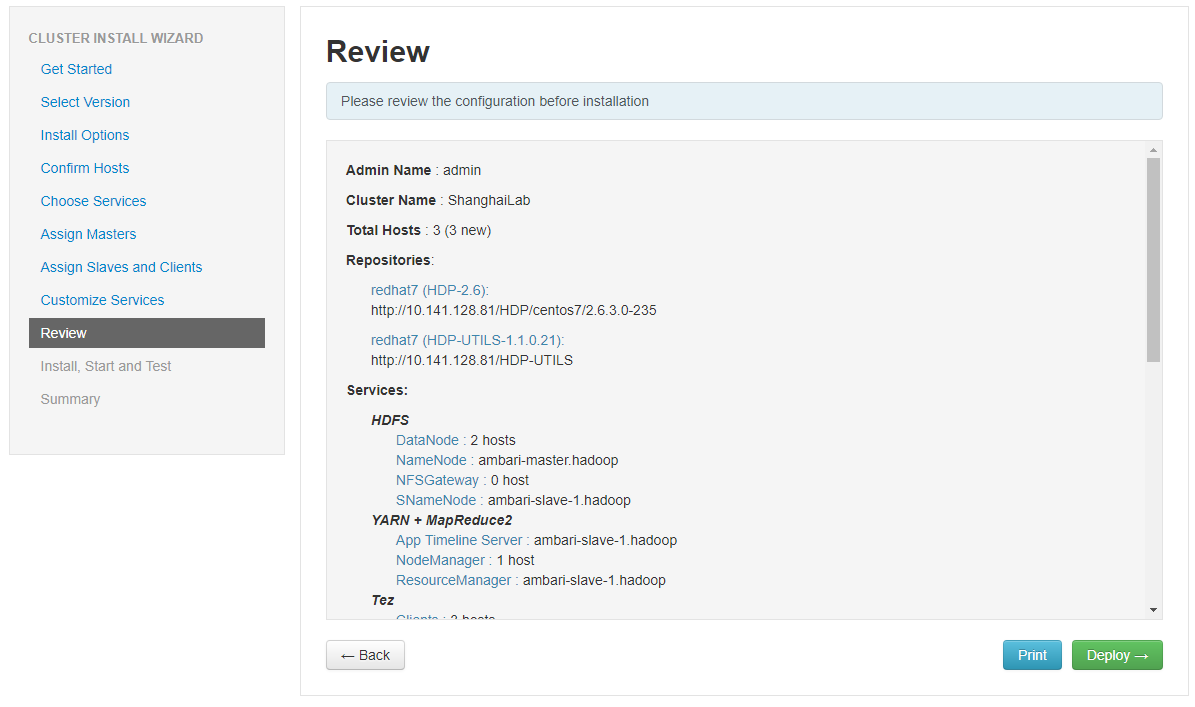


## Customize Services

根据告警信息做相应的调整，确保各组件无告警

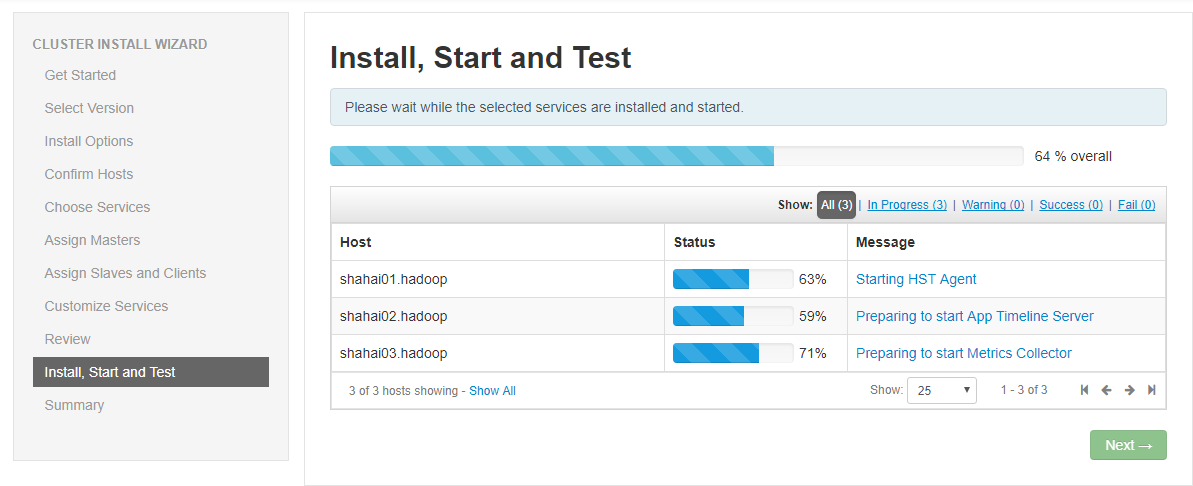


## Review

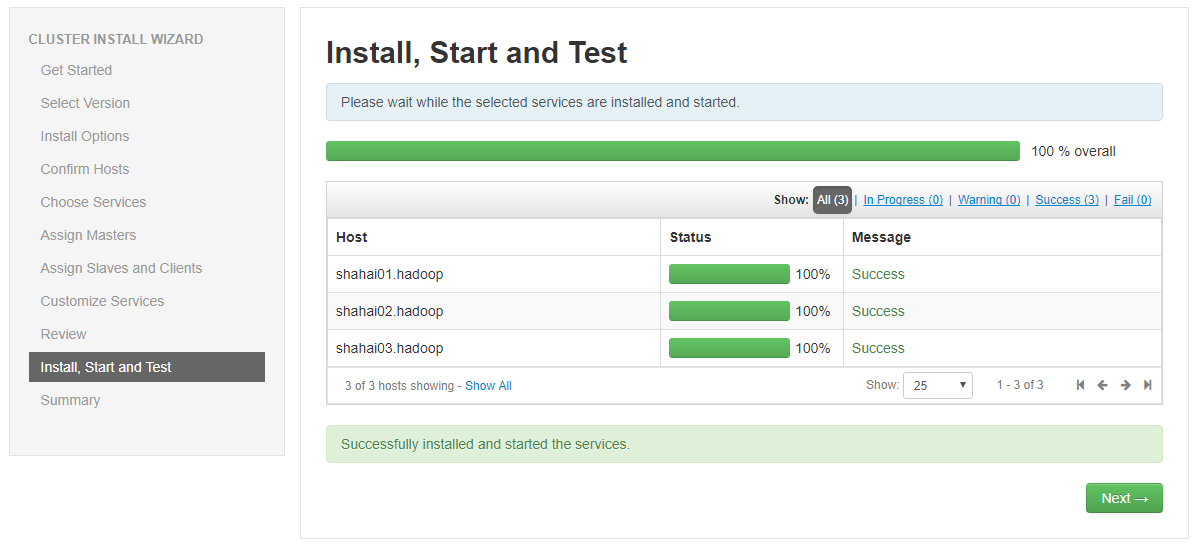


## Install,Start and Test

根据安装过程中的错误提示进行修复，然后Retry

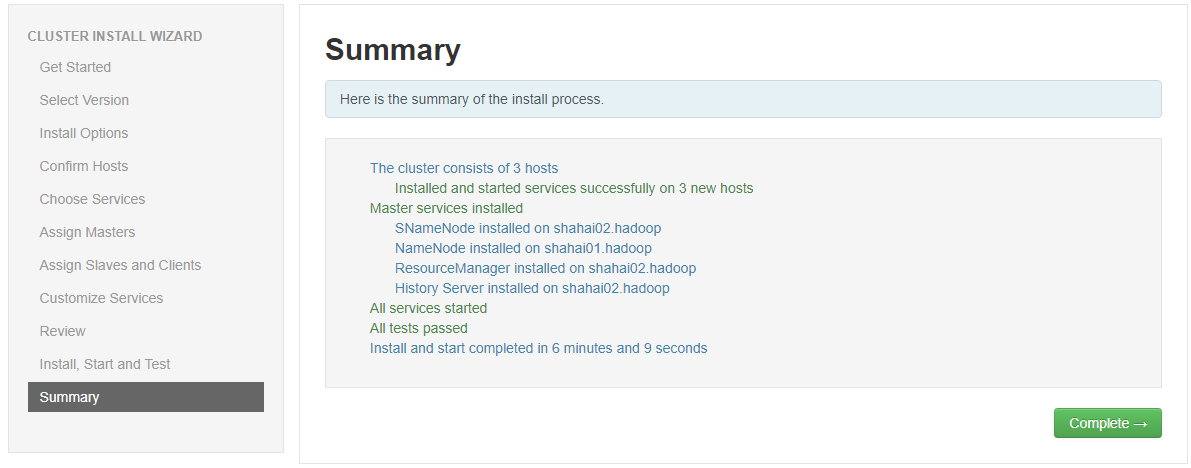


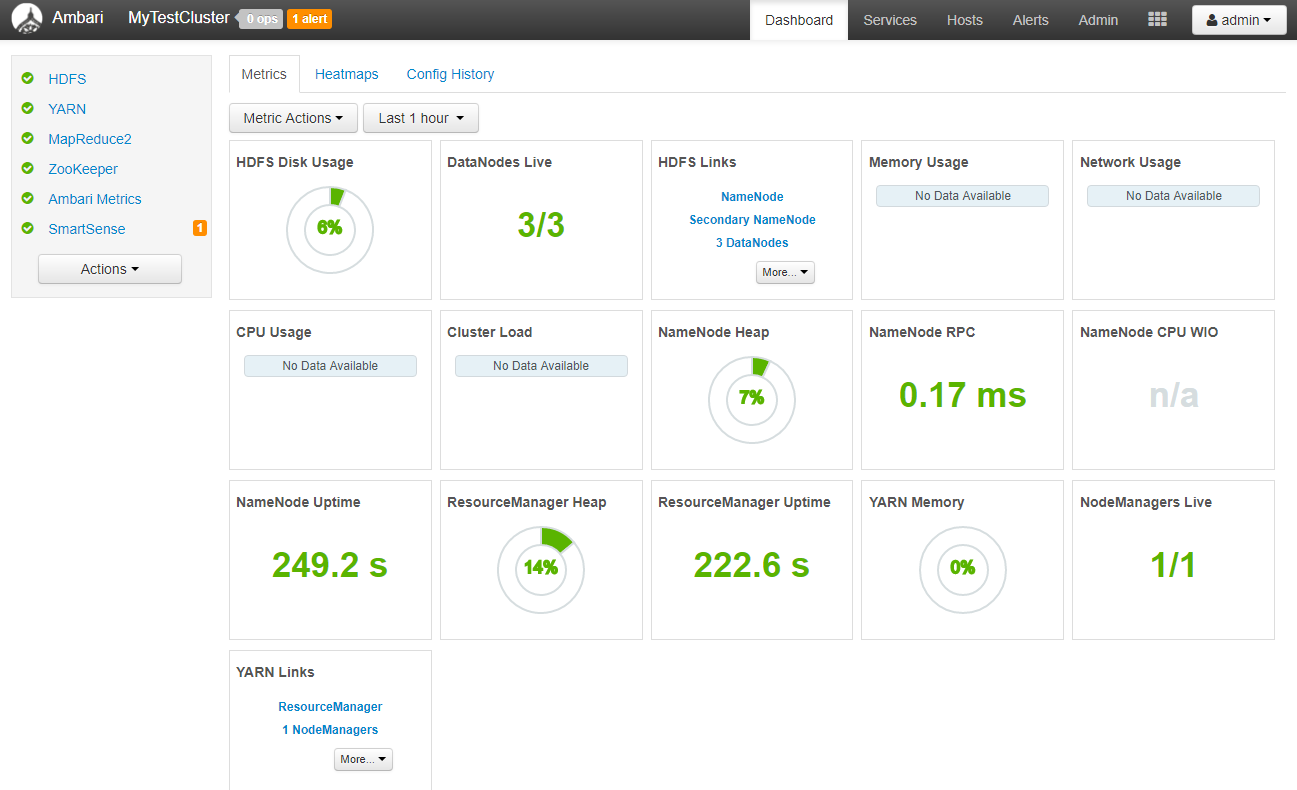
安装完成



## Summary

Complete





# 附录

*【内容】*

罗列需要补充说明的详细信息

*【裁剪原则】*

*此部分内容允许裁剪。*