

基于Ambari搭建大数据平台

blog.csdn.net/you_two/article/details/115345907

一、准备工作

以Apache 的 Ambari 2.0.1 为例

1.SSH 的无密码登录

Ambari 的 Server 会 SSH 到 Agent 的机器，拷贝并执行一些命令。因此我们需要配置 Ambari Server 到 Agent 的 SSH 无密码登录。在这个例子里，zwshen37 可以 SSH 无密码登录 zwshen38 和 zwshen39。

2.确保 Yum 可以正常工作

通过公共库（public repository），安装 Hadoop 这些软件，背后其实就是应用 Yum 在安装公共库里面的 rpm 包。所以这里需要您的机器都能访问 Internet。

3.确保 home 目录的写权限

Ambari 会创建一些 OS 用户。

4.确保机器的 Python 版本大于或等于 2.6。（Redhat6.6，默认就是 2.6 的）。

二、安装 AmbariServer

1. 取 Ambari 的公共库文件

```
wget http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.0.1/ambari.repo
```

将下载的 ambari.repo 文件拷贝到 Linux 的系统目录/etc/yum.repos.d/

然后 依次执行以下命令：

```
yum clean all
```

```
yum list|grep ambari
```

```
[root@zwshen37 resources]# yum list|grep ambari
ambari-server.noarch                2.0.1-45                                @Updates-ambari-2.0.1
ambari-agent.x86_64                 2.0.1-45                                Updates-ambari-2.0.1
```

如果可以看见 Ambari 的对应版本的安装包列表，说明公共库已配置成功

2.安装AmbariServer

```
yum install ambari-server
```

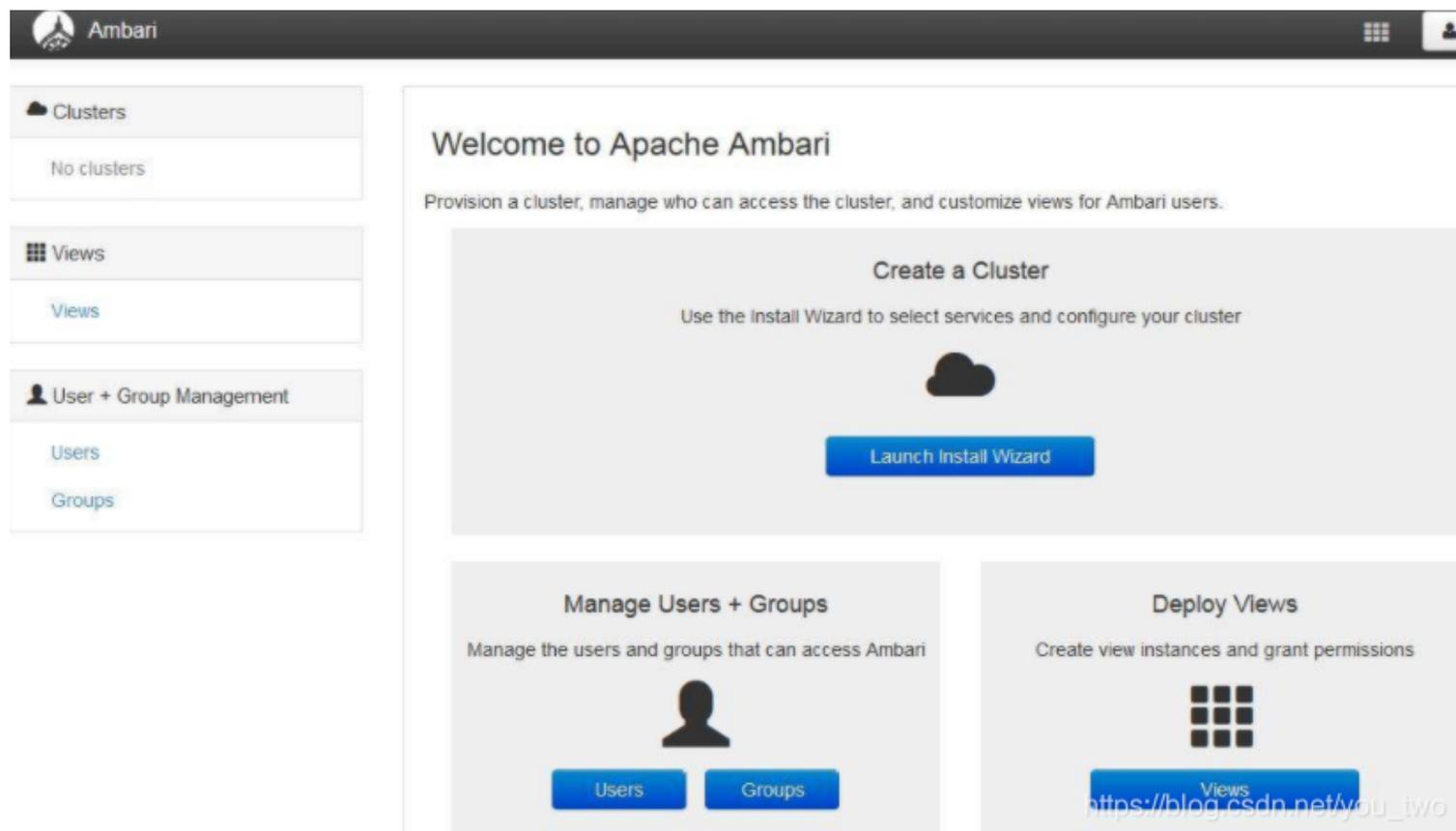
安装完成后配置：amari-server setup

在这个交互式的设置中，采用默认配置即可。Ambari 会使用 Postgres 数据库，默认会安装并使用 Oracle 的 JDK。默认设置了 Ambari GUI 的登录用户为 admin/admin。并且指定 Ambari Server 的运行用户为 root

3.启动 AmbariServer

```
ambari-server start
```

启动 Ambari Server 之后，便可以从浏览器登录，默认的端口为 8080。以本文环境为例，在浏览器的地址栏输入 `http://ip:8080`，登录密码为 admin/admin。登入 Ambari 之后的页面如下图



三、部署hadoop2.x集群

1. 命名集群的名字。本环境为 bigdata

登录 Ambari 之后，点击按钮“Launch Install Wizard”，就可以开始创建属于自己的大数据平台

2. 选择一个 Stack，这个 Stack 相当于一个 Hadoop 生态圈软件的集合。

Stack 的版本越高，里面的软件版本也就越高。这里我们选择 HDP2.2，里面的对应的 Hadoop 版本为 2.6.x

3. 指定 Agent 机器（如果配置了域，必须包含完整域名，例如本文环境的域为 example.com）
这些机器会被安装 Hadoop 等软件包，需要指定当时在 Ambari Server 机器生成的私钥，不要选择“Perform manual registration) on hosts and do not use SSH”。因为我们需要 Ambari Server 自动去安装 Ambari Agent

安装页面

CLUSTER INSTALL WIZARD

- Get Started
- Select Stack
- Install Options**
- Confirm Hosts
- Choose Services
- Assign Masters
- Assign Slaves and Clients
- Customize Services
- Review
- Install, Start and Test
- Summary

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](#)

```
zwshen38.example.com  
zwshen39.example.com
```

Host Registration Information

Provide your **SSH Private Key** to automatically register hosts

未选择文件

```
ssh private key
```

SSH User Account:

Perform **manual registration** on hosts and do not use SSH

https://blog.csdn.net/you_two

4. Ambari Server 会自动安装 Ambari Agent 到刚才指定的机器列表。安装完成后，Agent 会向 Ambari Server 注册。成功注册后，就可以继续 Next 到下一步

5. 选择要安装的软件名称。本文环境选择了 HDFS，YARN + MapReduce2，ZooKeeper，Storm 以及 Spark。选的越多，就会需要越多的机器内存。选择之后就可以继续下一步了

选择页面

CLUSTER INSTALL WIZARD

- Get Started
- Select Stack
- Install Options
- Confirm Hosts
- Choose Services**
- Assign Masters
- Assign Slaves and Clients
- Customize Services
- Review
- Install, Start and Test
- Summary

Choose Services

Choose which services you want to install on your cluster.

<input type="checkbox"/> Service	Version	Description
<input checked="" type="checkbox"/> HDFS	2.7	Apache Hadoop Distributed File System
<input checked="" type="checkbox"/> YARN + MapReduce2	2.7	Apache Hadoop NextGen MapReduce (YARN)
<input type="checkbox"/> Hive	1.1	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
<input type="checkbox"/> HBase	1.0.1	Non-relational distributed database and centralized service for configuration management & synchronization
<input type="checkbox"/> Pig	0.14.0	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	4.1.0	System for workflow coordination and execution of Apache Hadoop jobs. This also do the installation of the optional Oozie Web Console which relies the ExtJS Library
<input checked="" type="checkbox"/> ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination

6. 分别是选择安装软件所指定的 Master 机器和 Slave 机器，以及 Client 机器

7. Service 的配置

绝大部分配置已经有默认值，不需要修改， 如果不需要进行调优是可以直接使用默认配置的。有些 Service 会有一些必须的手工配置项，则必须手动输入，才可以下一步

8. Ambari 会总结一个安装列表，供用户审阅。这里没问题，就直接下一步

9.等待安装完成

安装进度

- CLUSTER INSTALL WIZARD
- Get Started
- Select Stack
- Install Options
- Confirm Hosts
- Choose Services
- Assign Masters
- Assign Slaves and Clients
- Customize Services
- Review
- Install, Start and Test**
- Summary

Install, Start and Test

Please wait while the selected services are installed and started.

4 % overall

Show: All (2) | In Progress (2) | Warning (0) | Success (0) | Fail (0)

Host	Status	Message
zwshen38.example.com	4%	Installing DataNode
zwshen39.example.com	4%	Installing App Timeline Server

2 of 2 hosts showing - Show All

Show: 25 1 - 2 of 2

Next →

https://blog.csdn.net/you_two

10.ambari的dashboard页面

The screenshot shows the Ambari dashboard for a Hadoop cluster. The top navigation bar includes 'Ambari', 'bigdata', '0 ops', '0 alerts', 'Dashboard', 'Services', 'Hosts', 'Alerts', 'Admin', and a user profile 'admin'. The left sidebar lists services: HDFS, MapReduce2, YARN, Tez, ZooKeeper, Storm, Ambari Metrics, and Spark. The main dashboard area displays several metrics:

- HDFS Disk Usage:** 19%
- DataNodes Live:** 2/2
- HDFS Links:** NameNode, Secondary NameNode, 2 DataNodes
- Memory Usage:** Graph showing usage over time.
- Network Usage:** Graph showing network activity.
- CPU Usage:** 50%
- Cluster Load:** 2
- NameNode Heap:** 8%
- NameNode RPC:** 0 ms
- NameNode CPU WIO:** 0.0%
- NameNode Uptime:** 6.5 d
- ResourceManager Heap:** 3%
- ResourceManager Uptime:** 6.5 d
- NodeManagers Live:** 2/2
- YARN Memory:** 0%
- YARN Links:** ResourceManager, 2 NodeManagers
- Supervisors Live:** 1/1

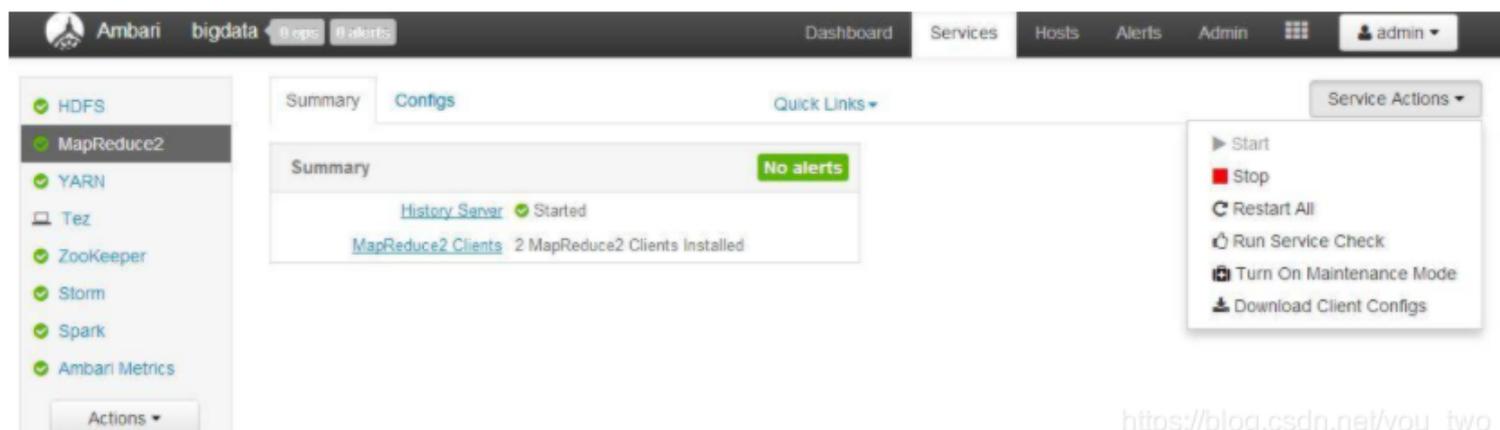
1080/#/main/services/YARN/summary

https://blog.csdn.net/you_two

利用 Ambari 管理 Hadoop 集群

1. Service Level Action (服务级别的操作)

左侧的 Service 列表，点击任何一个您想要操作的 Service，以 MapReduce2 为例，当点击 MapReduce2 后，就会看到该 Service 的相关信息，如下图



https://blog.csdn.net/you_two

装完 Hadoop 的集群后，并不知道这个集群是不是可用。这时候我们就可以运行一个“Run Service Check”。点击这个命令后，就会出现下图的进度显示

1 Background Operation Running

Operations	Start Time	Duration	Show: All (6)
MapReduce2 Service Check	Thu Jun 11 2015 22:32	5.74 secs	<div style="width: 35%;"></div> 35%

2. Host Level Action (机器级别的操作)

机器列表

Name	IP Address	Cores (CPU)	RAM	Disk Usage	Load Avg	Versions	Components
zwshen38.example.com	9.111.254.228	2 (2)	3.79GB	<div style="width: 35%;"></div>	0.23	HDP-2.2.0.0-2041 (Upgrading)	19 Components
zwshen39.example.com	9.111.254.229	2 (2)	3.79GB	<div style="width: 35%;"></div>	0.10	HDP-2.2.0.0-2041 (Upgrading)	15 Components

https://blog.csdn.net/you_two

All Hosts -> Hosts -> Start All Components，Ambari 就会将所有 Service 的所有模块启动

All Hosts-> DataNodes -> Stop，Ambari 就会把所有机器的 DataNode 这个模块关闭

Actions ▾ Filter: All (2) ▾

	IP Address	Cores (CPU)	RAM	D
Selected Hosts (0) ▾	Any	Any	Any	
Filtered Hosts (2) ▾	9.111.254.228	2 (2)	3.79GB	
All Hosts (2) ▾	Hosts	Start All Components		
<input type="checkbox"/>  zwshen39.exe	DataNodes	Stop All Components	3.79GB	
	NodeManagers	Restart All Components		
	Supervisors	Turn On Maintenance Mode		
		Turn Off Maintenance Mode		

2 of 2 hosts showing - clear filters

https://blog.csdn.net/you_two

3. 模块级别操作

点击机器名，我们会进入到该机器的 Component 页面

Components	+ Add
<input checked="" type="checkbox"/> DRPC Server / Storm	Started ▾
<input checked="" type="checkbox"/> Metrics Collector / Ambari Metrics	Started ▾
<input checked="" type="checkbox"/> NameNode / HDFS	Started ▾
<input checked="" type="checkbox"/> Nimbus / Storm	Started ▾
<input checked="" type="checkbox"/> Spark History Server / Spark	Started ▾
<input checked="" type="checkbox"/> Storm UI Server / Storm	Started ▾
<input checked="" type="checkbox"/> ZooKeeper Server / ZooKeeper	Started ▾
<input checked="" type="checkbox"/> DataNode / HDFS	Started ▾
<input checked="" type="checkbox"/> Metrics Monitor / Ambari Metrics	Started ▾
<input checked="" type="checkbox"/> NodeManager / YARN	Started ▾
<input checked="" type="checkbox"/> Supervisor / Storm	https://blog.csdn.net/you_twd

点击每个 Component（模块）后面的按钮，就可以看到该模块的操作命令了