

## Lab 4: Installing a Hadoop Cluster

### Objective:

After completing this lab, you will be familiar with installing a Hadoop Cluster running Hortonworks Data Platform. We will be using Ambari to automate a majority of cluster installation & configuration.

### Requirements:

1. A terminal that allows you to ssh into a linux machine. If you have a mac or a linux box, you are good to go. If you are a windows user, please ensure you have putty (or something similar for ssh) and a file transfer client(Filezilla or something similar for scp).
2. The latest version of Chrome or Firefox. Please do not use Internet Explorer. There are some known issues with IE and ambari.

**Due Date: Friday 10/2 4:30 PM**

### Things to Include in your Submission:

Please stop by Sriram's office and show off that your cluster is up and running. During this demo, you should be able to run a simple mapreduce job on your cluster. All services should be up and running on the cluster. In addition, the tools required for your project and research should be configured.

### Rubric:

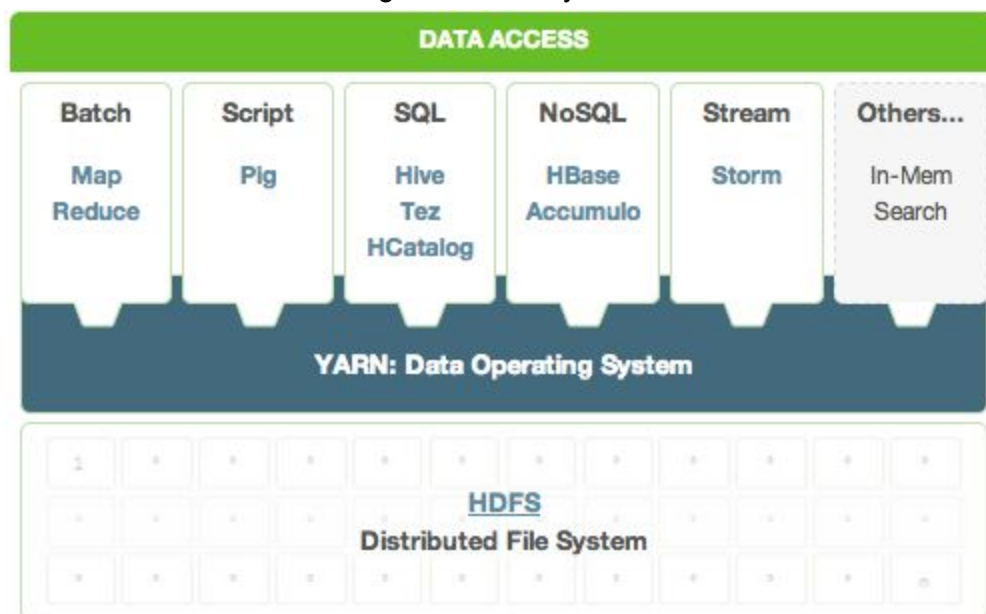
1. Cluster is up and running with all services configured **(90 points)**
2. Ability to launch and run MapReduce jobs from every node **(10 Points)**
3. Extra Credit - Hue is setup correctly **(25 Points)**

### Feedback:

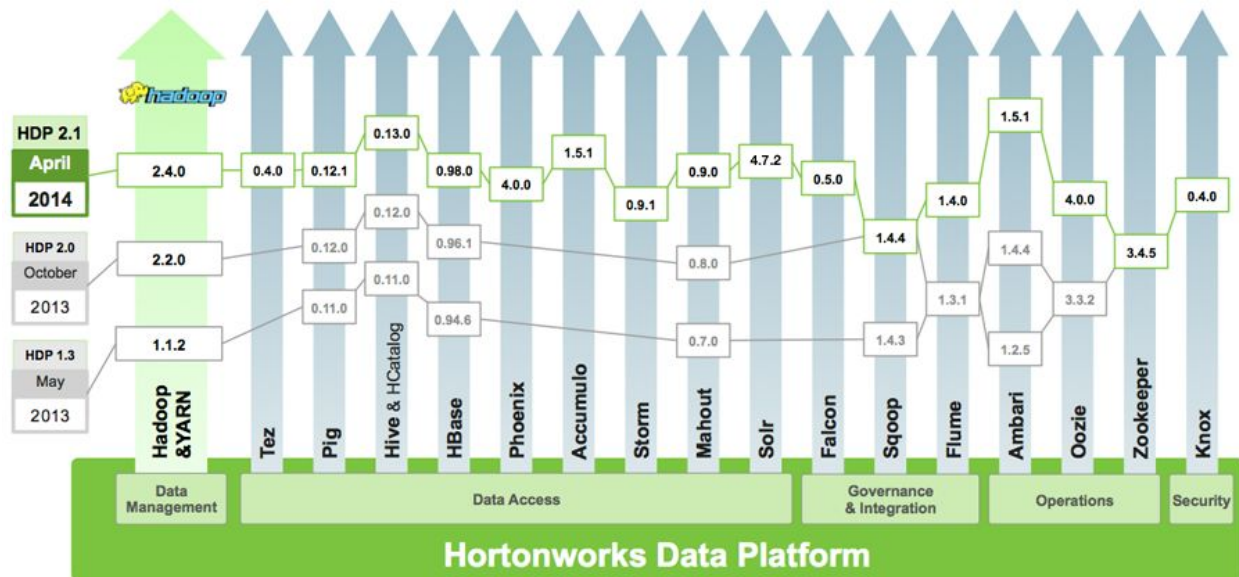
Please complete the anonymous feedback for the lab under Feedback → Lab Anonymous Feedback → [Lab 4 Anonymous Feedback](#).

## Hortonworks Data Platform

Hortonworks Data Platform (HDP) is a platform for multi-workload data processing across an array of processing methods - from batch through interactive to real-time. The core components of HDP are YARN and Hadoop Distributed Filesystem (HDFS). YARN is the data operating system of Hadoop that enables you to process data simultaneously in multiple ways. YARN provides the resource management and pluggable architecture for enabling a wide variety of data access methods. HDFS provides the scalable, fault-tolerant, cost-efficient storage for big data. YARN provides the foundation for a versatile range of processing engines that empower you to interact with the same data in multiple ways, at the same time. This means applications can interact with the data in the best way: from batch to interactive SQL or low latency access with NoSQL. Emerging use cases for search and streaming are also supported with Apache Solr and Storm. Additionally, ecosystem partners provide even more specialized data access engines for YARN. You will be installing HDP 2.3 on your cluster.



This is the latest stable release of HDP and includes support for the following services:



## Apache Ambari

The Apache Ambari project is aimed at making Hadoop management simpler by developing software for provisioning, managing, and monitoring Apache Hadoop clusters. Ambari provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs. Ambari enables System Administrators to:

- Provision a Hadoop Cluster
  - Ambari provides a step-by-step wizard for installing Hadoop services across any number of hosts.
  - Ambari handles configuration of Hadoop services for the cluster.
- Manage a Hadoop Cluster
  - Ambari provides central management for starting, stopping, and re-configuring Hadoop services across the entire cluster.
- Monitor a Hadoop Cluster
  - Ambari provides a dashboard for monitoring health and status of the Hadoop cluster.
  - Ambari leverages Ganglia for metrics collection.
  - Ambari leverages Nagios for system alerting and will send emails when your attention is needed (e.g., a node goes down, remaining disk space is low, etc).

Ambari enables Application Developers and System Integrators to:

- Easily integrate Hadoop provisioning, management, and monitoring capabilities to their own applications with the Ambari REST APIs.

## Let the Installation begin

The first few section describes the information and materials you need to get ready to install Hadoop using Apache Ambari.

### Determining Stack Compatibility:

A typical first step in a hadoop cluster installation is determining if the version of ambari you are going to be installing is compatible with the version of Hadoop you will be installing. The following table gives you an idea of the ambari versions and Hadoop Versions.

Ambari*	HDP 2.3	HDP 2.2	HDP 2.1 (deprecated)	HDP 2.0 (deprecated)
2.1	x	x	x	x
2.0		x	x	x
1.7		x	x	x
1.6.1			x	x
1.6.0			x	x
1.5.1			x	x
1.5.0				x

\* Ambari does not install Hue or Solr.

We will be installing HDP 2.3 using Ambari 2.1. So we are good to go. Please note that Ambari does not install Hue and if you want to have hue on your cluster, you need to install that yourself.

## 1. Gain Access to the nodes that will make up your cluster

As mentioned in class, we have created several virtual machines on which your team will install their hadoop cluster. The machines are named `hadoop-01.csse.rose-hulman.edu`, `hadoop-02.csse.rose-hulman.edu` all the way to `hadoop32.csse.rose-hulman.edu`. You will be using the root account to login to these machines. Please check the following links to identify the nodes that your team will be using

### [Section 1](#)

Please check the board for the passwords to use to login to these machines.

- a. Please logon to the machines that you have been assigned using ssh or putty. For instance, if you are using the terminal, you will type in (please replace the corresponding machine number)

ssh [root@hadoop-01.csse.rose-hulman.edu](ssh:root@hadoop-01.csse.rose-hulman.edu)

Once you have logged in, please change the password on your machine. You can change the password as follows

- b. From the command line, type in `passwd` and follow instructions. (Please choose a password that you will remember and share with your project team. **We cannot help you if you forget your password.**)
- c. **Please repeat the above steps for every node that makes up your hadoop cluster.**

## Hardware Configurations

There is no one ideal Hadoop Configuration. The hard drives to be used, the amount of memory needed and the number of nodes depend on the the type and amount of data to be stored and processed by hadoop. You can find more information about hardware configurations for Hadoop [here](#).

## 2. Operating System Configurations

The following operating systems are supported:

- Red Hat Enterprise Linux (RHEL) v7.x

- Red Hat Enterprise Linux (RHEL) v6.x
- CentOS v7.x
- CentOS v6.x
- Oracle Linux v7.x
- Oracle Linux v6.x
- SUSE Linux Enterprise Server (SLES) v11 SP3

From experience, Red Hat and CentOS provide the easiest and optimal install experience. SLES and Oracle have little issues that can trip you up. We will be utilizing CentOS for our install.

- Verify that we have the **correct version of centos install on your nodes**. On every node, type `cat /etc/*-release`. You should find CentOS release 6.6 as part of the output.

### 3. Software Requirements

You need to verify on each of your hosts that the following software has been installed.

- yum and rpm (RHEL/CentOS/Oracle Linux)
  - scp, curl, and wget
  - python (2.6 or later)
- Verify that yum is installed on **each of your nodes** by typing in “yum version”
  - Verify that rpm is installed on **each of your nodes** by typing in “rpm”
  - Verify that scp is installed on **each of your nodes** by typing in “scp”
  - Verify that curl is installed on **each of your nodes** by typing in “curl”
  - Verify that wget is installed on **each of your nodes** by typing in “wget”
  - Verify that python is installed on **each of your nodes** by typing in “python --version”

### JDK Requirements

Ambari and Hadoop require Oracle JDK 1.7\_45 64-bit (default) as a part of their installation. Ambari will install JDK as a part of its setup process.

### Database Requirements

Hive/HCatalog, Oozie, and Ambari all require their own internal databases.

- Hive/HCatalog: By default uses an Ambari-installed MySQL 5.x instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance.

- Oozie: By default uses an Ambari-installed Derby instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance.
- Ambari: By default uses an Ambari-installed PostgreSQL 8.x instance. With appropriate preparation, you can also use an existing PostgreSQL 9.x, MySQL 5.x, or Oracle 11g r2 instance.

We will use the default databases installed by Ambari. We can proceed to the next step.

## File System Partitioning Recommendations

For information on setting up file system partitions on master and slave nodes in a HDP cluster, see [File System Partitioning Recommendations](#).

### 4. Recommended Maximum Open File Descriptors

The recommended maximum number of open file descriptors is 10000 or more. To check the current value set for the maximum number of open file descriptors, **execute the following shell commands on every node**.

- a. "ulimit -Sn" and "ulimit -Hn"

## Prepare the Environment

To deploy your Hadoop instance, you need to prepare your deployment environment:

### 5. Setup Passwordless SSH

To have Ambari Server automatically install Ambari Agents in all your cluster hosts, you must set up password-less SSH connections between the main installation (Ambari Server) host and all other machines. The Ambari Server host acts as the client and uses the key-pair to access the other hosts in the cluster to install the Ambari Agent.

- a. Choose the node with the lowest number to act as the ambari-server. On this node type in the following command "ssh-keygen" .
- b. Accept the default location of the key
- c. Enter passphrase - **Please leave it blank**
- d. Enter passphrase again - **Please leave it blank**

## Copying your ssh-key to every node on the cluster

- a. From the node serving as the ambari-server, please repeat the following steps for every node on the cluster `ssh-copy-id root@node1.csse.rose-hulman.edu`

### **Verifying Passwordless ssh is setup:**

- a. From the Ambari Server, make sure you can connect to each host in the cluster using SSH. `ssh root@node1` You may see this warning. This happens on your first connection and is normal. "Are you sure you want to continue connecting (yes/no)?":
- b. Please note that it should not prompt you for your password.

### **Copying your private key to the desktop.**

- a. Using either the terminal on your machine (assuming you have a mac or a linux box) or a File transfer client, please copy the private key to your desktop. You will need it later while installing the cluster. The file in question is ".ssh/id\_rsa". For instance, you can run the following command from your terminal  
`"scp root@ambari-server-node.csse.rose-hulman.edu:~/.ssh/id_rsa "`

## **6. Enable NTP on every node**

The clocks of all the nodes in your cluster and the machine that runs the browser through which you access Ambari Web must be able to synchronize with each other.

- a. On all nodes, enable ntp and have ntp start on system reboot  
`sudo /sbin/chkconfig ntpd on`  
`sudo /etc/init.d/ntpd start`

### **Please repeat the commands on every node in your cluster**

## **7. Check DNS**

All nodes in your system must be configured for DNS and Reverse DNS.

### **Determine IP address**

- a. On every node in the cluster type in the following command: "ip addr show"

You will see a result like this



```
[root@hadoop46 ~]# ip addr show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 16436 qdisc noqueue state UNKNOWN
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP qlen 1000
    link/ether 00:50:56:b8:40:c2 brd ff:ff:ff:ff:ff:ff
    inet 137.112.89.146/24 brd 137.112.89.255 scope global eth0
    inet6 fe80::250:56ff:feb8:40c2/64 scope link
        valid_lft forever preferred_lft forever
[root@hadoop46 ~]# █
```

**Please note the ip addresses for every node in the cluster.** For instance, in the above image, the ip address is 137.112.89.146. Please retain them in a safe place, we will need them later.

### **Edit the Hosts file**

1. Using a text editor, open the hosts file on every host in your cluster. For example:  
vi /etc/hosts
2. Add a line for each host in your cluster. The line should consist of the IP address and the Fully Qualified Domain Name. For example:  
137.112.89.146 hadoop46.csse.rose-hulman.edu  
137.112.89.147 hadoop47.csse.rose-hulman.edu  
137.112.89.148 hadoop48.csse.rose-hulman.edu
3. If you used vi and are not sure how to save it, please press Esc shift : wq

**Please repeat steps 1 and 2 on every node in your cluster.**

### **Set the Hostname**

1. Use the "hostname" command to set the hostname on each host in your cluster. For example:  
hostname hadoop46.csse.rose-hulman.edu
2. Confirm that the hostname is set by running the following command:  
hostname -f

This should return the name you just set. **Please repeat steps 1 and 2 on every node in your cluster.**

### **Edit the Network Configuration File**

1. Using a text editor, open the network configuration file on every host. This file is used to set the desired network configuration for each host. For example:

```
vi /etc/sysconfig/network
```

2. Modify the HOSTNAME property to set the fully.qualified.domain.name. For example

```
NETWORKING=yes  
NETWORKING_IPV6=yes  
HOSTNAME=hadoop46.csse.rose-hulman.edu
```

**Please repeat steps 1 and 2 on every node in your cluster.**

### **Configuring IP Tables**

For Ambari to communicate during setup with the hosts it deploys to and manages, certain ports must be open and available. The easiest way to do this is to temporarily disable iptables.

**Repeat the following commands on every node**

- a. `chkconfig iptables off`
- b. `/etc/init.d/iptables stop`

### **Disable SELinux and Check UMask**

1. SELinux must be temporarily disabled for the Ambari setup to function. Run the following command **on each host in your cluster:**  
`setenforce 0`
2. Make sure umask is set to 022. Type in the following command on **every node in the cluster** and ensure that it returns 0022  
`umask`

You will often find that most organizations protect their clusters by placing it behind a firewall. If that is the case, you may need to mirror the repositories needed for installation locally. HDP does support it and you can find more information on the Hortonworks website. Our nodes do have access to the internet and we don't have to worry about it.

## **8. Installing Ambari Server**

We will be using Ambari to manage, install, and deploy Hadoop. To do this, we need to install `ambari-server` on a node (you identified this node earlier in the process), setup `ambari-server` and start it. As a part of the hadoop setup, `ambari-server` will look at the `/etc/hosts` file to identify

the nodes that make up this cluster, install ambari-agent on every node in said cluster and proceed to set up Hadoop.

### Download Install files needed for Ambari

1. Log into the machine that is to serve as the Ambari Server as root. This machine is the main installation host.
2. Download the Ambari repository file

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.1.1/ambari.repo -O /etc/yum.repos.d/ambari.repo
```

3. Confirm the repository is configured correctly by typing in the following command  
yum repolist

You should see ambari-2.x on the list

### Install Ambari-Server

1. Install ambari-server on the ambari-server host node using the following command. Please note that this will install PostgreSQL that is need by ambari-server as well.

```
yum install ambari-server
```

You might be prompted a couple of times (is this okay?). Please respond “y” to both.

## 9. Ambari-Server Setup

The ambari-server command manages the setup process. Run the following command on the ambari-server node and respond to the prompts:

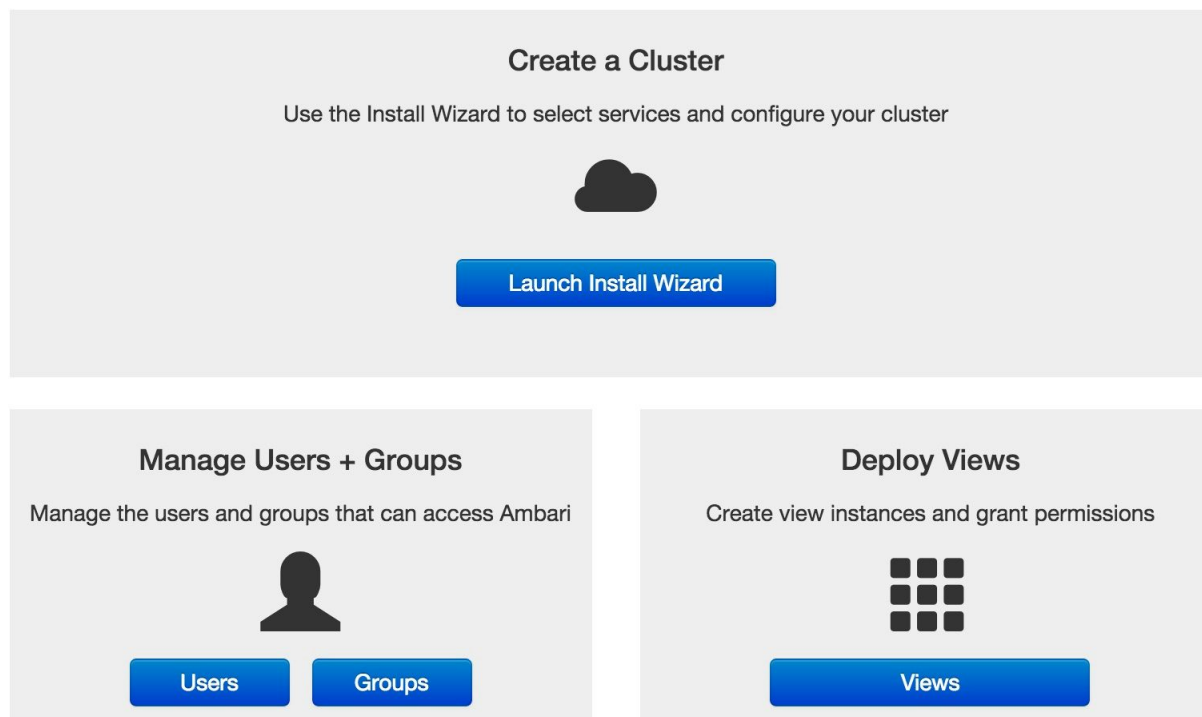
```
ambari-server setup
```

1. If you have not temporarily disabled SELinux, you may get a warning. Enter “y” to continue.
2. By default, Ambari Server runs under root. If you want to create a different user to run the Ambari Server instead, or to assign a previously created user, select “n” at **Customize user account for ambari-server daemon** to continue installation as root
3. If you have not temporarily disabled iptables you may get a warning. Enter “y” to continue.
4. Select a JDK version to download. Enter 1 to download Oracle JDK 1.8.

5. Agree to the Oracle JDK license when asked. You must accept this license to be able to download the necessary JDK from Oracle. The JDK is installed during the deploy phase.
6. At **Enter advanced database configuration**, choose to continue with the default database for ambari by choosing "n"
7. Setup should complete successfully.
8. Start ambari-server by typing in the following command  
ambari-server start
9. Check that ambari-server is up and running by browsing to the following web page on your host machine. Please use the latest version of Firefox or Chrome.  
http://ambari-server-node:8080 For instance, it will be  
http://hadoop46.csse.rose-hulman.edu:8080
10. If the previous steps succeeded, you should see a login page. Log in using the following credentials - username: admin password: admin
11. You will see a welcome page similar to the picture shown below:

## Welcome to Apache Ambari

Provision a cluster, manage who can access the cluster, and customize views for Ambari users.



12. Click on Launch Install Wizard. You are now ready to start installing Hortonworks Data Platform

## 10. Installing the Hortonworks Data Platform

## Name your Cluster

For a new cluster, the Ambari install wizard displays a Welcome page in which you define a cluster name.

1. In **Name your cluster** , type a name for the cluster you want to create. Use no white spaces or special characters in the name. Please choose a name that fits the following pattern. Section1ProjectName
2. Choose **Next**.

## Select Stack

1. The Service Stack (the Stack) is a coordinated and tested set of HDP components. Use a radio button to select the Stack version you want to install. Since we are installing HDP 2.3 , select the HDP 2.3 radio button.
2. Under Advanced Repository Options, you can select the Base URL of a repository from which Stack software packages download. Ambari sets the following, default repository URLs, depending on the Internet connectivity available to the Ambari server host:

**Important:** When using local mirror repositories, you only need to provide Base URLs for the Operating System you are installing for your Stack. Uncheck all other repositories.

OS	Name	Base URL
<input checked="" type="checkbox"/> redhat6	HDP-2.3	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6"/>
<input type="checkbox"/> redhat7	HDP-2.3	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.3"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos7"/>
<input type="checkbox"/> suse11	HDP-2.3	<input type="text" value="http://public-repo-1.hortonworks.com/HDP/suse11sp3/2.x/updates/2.3"/>
	HDP-UTILS-1.1.0.20	<input type="text" value="http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/suse11"/>

☐ Skip Repository Base URL validation (Advanced) ?

3. Please uncheck the redhat7 and suse11 options. This will ensure that we are installing HDP 2.3 on a system compatible with redhat6(the only option that will remain checked) which include CentOS6.5
4. Please click Next

## Install Options & Confirm Hosts

In order to build up the cluster, the install wizard needs to know general information about how you want to set it up. You need to supply the FQDN (Fully Qualified Domain Name) of each of your hosts. The wizard also needs to access the private key file you created in Set Up Password less SSH. Using the host names and key file information, the wizard can locate, access, and interact securely with all hosts in the cluster.

1. Use the **Target Hosts** text box to enter your list of host names assigned to you, one per line.
2. We will let Ambari automatically install the Ambari Agent on all the nodes that make up our cluster using SSH, select **Provide your SSH Private Key** and either use the **Choose File** button in the **Host Registration Information** section to find the private key file that matches the public key you installed earlier on all your hosts or cut and paste the key into the text box manually. This is the file you copied over to your machine earlier in the process.
3. Fill in the user name for the SSH key you have selected. Please ensure that this says root.
4. Choose **Register and Confirm** to continue.
5. **You will be brought to a screen to Confirm Hosts. You will be** prompted to confirm that Ambari has located the correct nodes for your cluster and to check those nodes to make sure they have the correct directories, packages, and processes required to continue the install.
6. If any nodes were selected in error, you can remove them now.
7. At the bottom of the screen, you may notice a yellow box that indicates some warnings were encountered during the check process. For example, your host may have already had a copy of wget or curl. Choose **Click here to see the warnings** to see a list of what was checked and what caused the warning. The warnings page also provides access to a python script that can help you clear any issues you may encounter and let you run **Rerun Checks**. When you are satisfied with the list of hosts, choose **Next**.
8. If you get a problem with transparent huge pages, you would need to do the following on every host:

```
# Disable THP on a running system
sudo echo never > /sys/kernel/mm/transparent_hugepage/enabled
sudo echo never > /sys/kernel/mm/transparent_hugepage/defrag
```

To disable this permanently, do the following:

```
# Backup rc.local
sudo cp -p /etc/rc.local /etc/rc.local.`date +%Y%m%d-%H:%M`
```

Then copy the following into the bottom of /etc/rc.local.

```
if test -f /sys/kernel/mm/transparent_hugepage/enabled; then
    echo never > /sys/kernel/mm/transparent_hugepage/enabled
fi
```

```
if test -f /sys/kernel/mm/transparent_hugepage/defrag; then
    echo never > /sys/kernel/mm/transparent_hugepage/defrag
fi
```

9. If everything proceeds smoothly, you will see ambari go through the process of installing and registering hosts. When everything says succeeded, you are ready to move to the next step. Get in touch with your instructor/Adam Michael if you run into issues.

## Select Services

HDP comprises many services. You must install the HDFS and ZooKeeper services. You may choose to install any other available services now, or to add services later. The install wizard selects all available services for installation by default.

1. Choose **all** to select all listed services. (I did not install accumulo and Atlas)
2. After selecting the services to install now, choose **Next**.

## Assign Master & Slaves

The Ambari install wizard assigns the master components for selected services to appropriate hosts in your cluster and displays the assignments in Assign Masters. The left column shows services and current hosts. The right column shows current master component assignments by host, indicating the number of CPU cores and amount of RAM installed on each host. Here is a sample screen

**Ambari** admin

**CLUSTER INSTALL WIZARD**

- Welcome
- Select Stack
- 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.  
 \* HiveServer2, Hive Metastore, and WebHCat Server will be hosted on the same server.

NameNode: compute-0-5.local (31.5 GB, 1)	compute-0-0.local (15.8 GB, 8 cores)
SNameNode: compute-0-0.local (15.8 GB, 1)	ShameNode History Server
History Server: compute-0-0.local (15.8 GB, 1)	App Timeline Server ResourceManager
App Timeline Server: compute-0-0.local (15.8 GB, 1)	ZooKeeper Server
ResourceManager: compute-0-0.local (15.8 GB, 1)	compute-0-1.local (15.8 GB, 8 cores)
Nagios Server: compute-0-5.local (31.5 GB, 1)	Hive Metastore HiveServer2
Ganglia Server: compute-0-5.local (31.5 GB, 1)	WebHCat Server HBase Master
Hive Metastore: compute-0-1.local *	Oozie Server ZooKeeper Server
HiveServer2: compute-0-1.local (15.8 GB, 1)	Falcon Server
WebHCat Server: compute-0-1.local *	compute-0-5.local (31.5 GB, 8 cores)
HBase Master: compute-0-1.local (15.8 GB, 1)	NameNode Nagios Server
Oozie Server: compute-0-1.local (15.8 GB, 1)	Ganglia Server ZooKeeper Server
ZooKeeper Server: compute-0-5.local (31.5 GB, 1)	DRPC Server Nimbus
ZooKeeper Server: compute-0-0.local (15.8 GB, 1)	Storm REST API Server Storm UI Server
ZooKeeper Server: compute-0-1.local (15.8 GB, 1)	
Falcon Server: compute-0-1.local (15.8 GB, 1)	
DRPC Server: compute-0-5.local (31.5 GB, 1)	
Nimbus: compute-0-5.local (31.5 GB, 1)	
Storm REST API Server: compute-0-5.local (31.5 GB, 1)	
Storm UI Server: compute-0-5.local (31.5 GB, 1)	

3 hosts not running master services

[< Back](#) [Next >](#)

Licensed under the Apache License, Version 2.0.  
See third-party tools/resources that Ambari uses and their respective authors

1. To change the host assignment for a service, select a host name from the drop-down menu for that service.
2. We will be using the node on which you installed ambari-server as our main master node. Please ensure that the namenode, hive server2, resource manager are all setup to use this node.
3. When you are satisfied with the assignments, choose **Next**.

The Ambari install wizard assigns the slave components (DataNodes, NodeManagers, and RegionServers) to appropriate hosts in your cluster. It also attempts to select hosts for installing the appropriate set of clients. A sample screen is shown below:



## 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, MapReduce 2 Client, YARN Client, Tez Client, Hive Client, HCat Client, HBase Client, Pig, Sqoop, Oozie Client, ZooKeeper Client and Falcon Client.

Host	all   none	all   none	all   none	all   none	all   none
k1.hdp2 *	<input type="checkbox"/> DataNode	<input type="checkbox"/> NodeManager	<input type="checkbox"/> RegionServer	<input type="checkbox"/> Supervisor	<input checked="" type="checkbox"/> Client
m1.hdp2 *	<input type="checkbox"/> DataNode	<input type="checkbox"/> NodeManager	<input type="checkbox"/> RegionServer	<input type="checkbox"/> Supervisor	<input checked="" type="checkbox"/> Client
w1.hdp2 *	<input checked="" type="checkbox"/> DataNode	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> RegionServer	<input checked="" type="checkbox"/> Supervisor	<input checked="" type="checkbox"/> Client
w2.hdp2 *	<input checked="" type="checkbox"/> DataNode	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> RegionServer	<input checked="" type="checkbox"/> Supervisor	<input checked="" type="checkbox"/> Client

Show: 25 1 - 4 of 4

⏮ ⏪ ⏩ ⏭

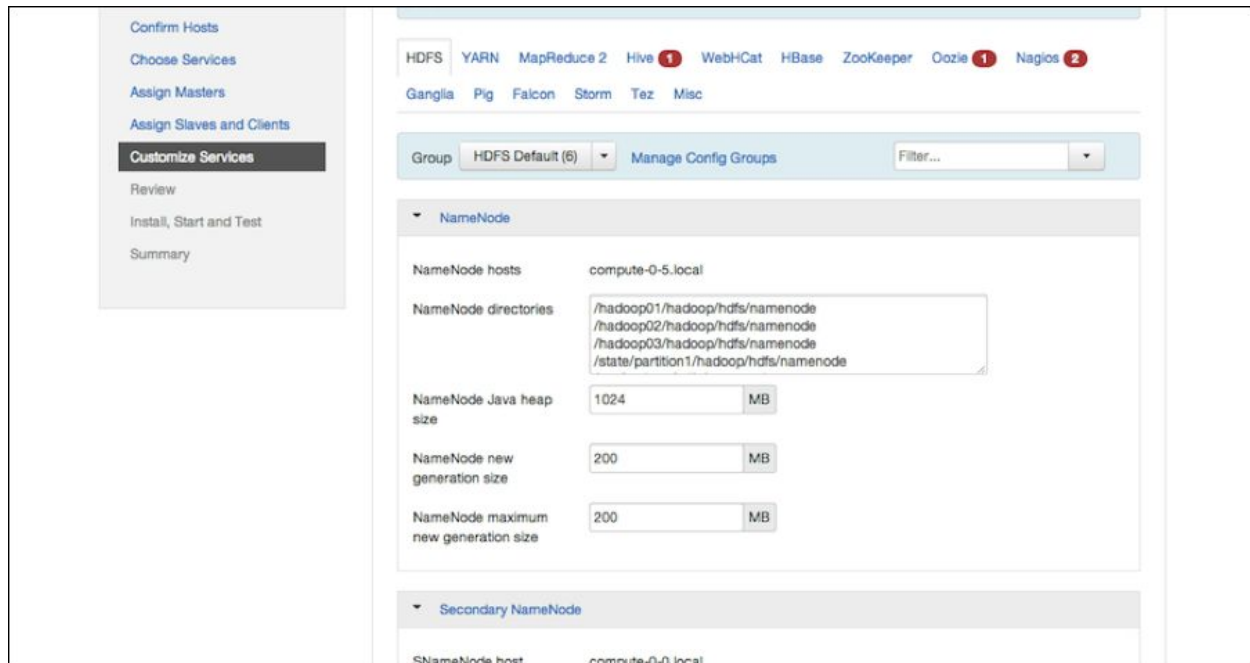
← Back

Next →

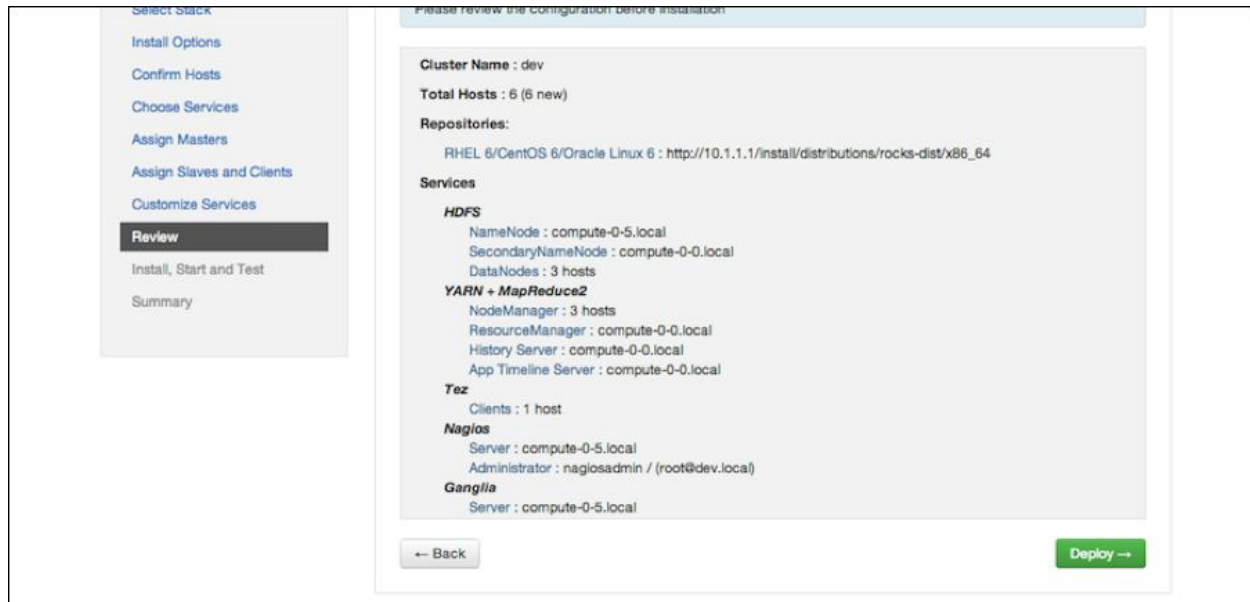
1. Use **all** to select all of the hosts in the column respectively. We will do this for every column.
2. If a host has a red asterisk next to it, that host is also running one or more master components. Hover your mouse over the asterisk to see which master components are on that host.
3. When you are satisfied with your assignments, choose **Next**.

### Customize Services, Review & Deploy

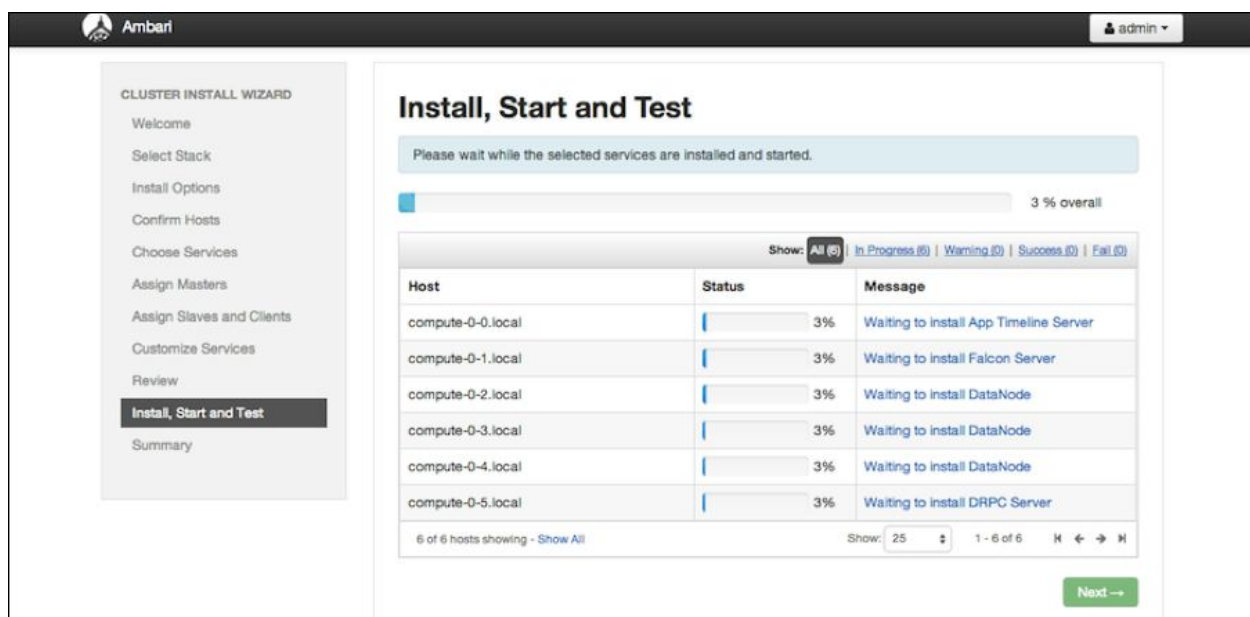
1. Customize Services presents you with a set of tabs that let you manage configuration settings for HDP components. The wizard sets reasonable defaults for each of the options here, but you can use this set of tabs to tweak those settings. You are strongly encouraged to go with the default options for these tabs. A sample screen is shown below:



2. A typical installation has at least ten groups of configuration properties and other related options, such as database settings for Hive/HCat and Oozie, admin name/password, and a master password for Knox.
3. You will see a few tabs have red notifications next to them. Please fill in the missing details. The install wizard sets reasonable defaults for all properties except for those related to databases in the Hive and the Oozie tabs, and two related properties in the Nagios tab. These four are marked in red and are the only ones you must set yourself. Click the name of the group in each tab to expand and collapse the display. Please remember the details you entered.
4. When you have customized everything, please click next. You will be brought to a review screen that looks like this



5. Hit Deploy to begin the installation process. You will be given the status of the install on a screen that looks like the one shown below



6. At the end you should have a fully functional HDP deployment

CLUSTER INSTALL WIZARD

- Welcome
- 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.

100 % overall

Host	Status	Message
compute-0-0.local	100%	Success
compute-0-1.local	100%	Success
compute-0-2.local	100%	Success
compute-0-3.local	100%	Success
compute-0-4.local	100%	Success
compute-0-5.local	100%	Success

6 of 6 hosts showing - [Show All](#)

Successfully installed and started the services.

[Next →](#)

7. Click Next and you will be taken to the Ambari dashboard

Ambari
HDP212
0 ops

admin

Dashboard
Heatmaps
Services
Hosts
Jobs
Admin

- ✓ HDFS
- ✓ YARN
- ✓ MapReduce2
- Tez
- ✓ HBase
- ✓ Hive
- ✓ WebHCat
- Falcon
- Storm
- ✓ Oozie
- ✓ Ganglia
- ✓ Nagios
- ✓ ZooKeeper
- Pig
- Sqoop

Actions

### Cluster Status and Metrics

HDFS Disk Usage

DataNodes Live

2/2

HDFS Links

NameNode  
Secondary NameNode  
2 DataNodes

More...

Memory Usage

Network Usage

CPU Usage

Cluster Load

NameNode Heap

NameNode RPC

0.20 ms

NameNode CPU WIO

NameNode Uptime

24.2 min

HBase Master Heap

HBase Links

HBase Master  
2 RegionServers  
Master Web UI

More...

HBase Ave Load

2

HBase Master Uptime

558.2 s

11. Change Ambari Admin Password & Fix Permissions

1. Click on the Admin and browse your way to the list of users.

2. Click edit under the admin user and change the password

The HDP standard deployment does not setup permissions to launch jobs from the root account. To fix this do the following from the command line of the ambari-server node:

```
sudo -u hdfs hadoop fs -mkdir -p /user/root
sudo -u hdfs hadoop fs -chown root:root /user/root
usermod -a -G hadoop root(need to run this on all nodes)
```

You are instructing hadoop to create a home directory for the root user and making the root user the owner of said directory and adding root to the hadoop group

3. Verify that root is a part of the hadoop group by typing in `id root`.
4. You should see hadoop listed as one of the groups.
5. Verify that you are able to run jobs from the command line by running a sample map reduce job.
6. Browse through the various options available in ambari(the ambari page for your cluster) and get familiar with the places where you can change configuration.

## 12. Extra Credit - Install Hue

At this point, you have a fully functioning hadoop cluster that you can manage using Ambari. If you prefer to use hue to interact with your cluster, you need to install it yourself. You can find instructions for installing hue [here](#)

**Due Date: Friday 10/2 4:30 PM**

### Things to Include in your Submission:

Please stop by Sriram's office and show off that your cluster is up and running. During this demo, you should be able to run a simple mapreduce job on your cluster. All services should be up and running on the cluster. In addition, the tools required for your project and research should be configured.

### Rubric:

4. Cluster is up and running with all services configured **(90 points)**
5. Ability to launch and run MapReduce jobs from every node **(10 Points)**
6. Extra Credit - Hue is setup correctly **(25 Points)**

### Feedback:

Please complete the anonymous feedback for the lab under Feedback → Lab Anonymous Feedback → [Lab 4 Anonymous Feedback](#).