



# Distributed Computing using CloudLab

Linh B. Ngo  
Clemson University



# Introduction

- Distributed and Cluster Computing (CPSC 3620)
- Offered twice per academic year
- Average class size: 40-45 students
- Required junior-level class (typically taken at senior year)
- Contents:
  - Infrastructure/System-oriented
  - Performance/Efficiency
  - High Performance Computing
    - MPI
  - Big Data Computing
    - Hadoop MapReduce
    - Apache Spark
    - HPCSystems



# Computing Resources

- Palmetto Supercomputer
  - 2000+ nodes, open to all faculty/students
  - No administrative access
  - Cannot share nodes among students to support group assignment
  - Preemption from node owners
- CloudLab
  - Limited resources for large-scale study
  - Administrative access
  - Ease of collaboration
  - No preemption



# Computing Resources

- Combine both local computing resources and CloudLab
- Learning outcomes through CloudLab
  - Administrative skills for distributed systems
  - In-depth understanding of distributed systems
- Learning outcomes through Palmetto
  - Basic understanding of parallel application development
  - Impacts of scaling and efficiency on larger systems



# Tutorial

- Set up environments for distributed computing on CloudLab
- MPI
  - Two-node cluster
  - OpenMPI
- Hadoop
  - Three-node cluster
  - Hortonwork Distribution



Inbox - linh.b.ngo@gmail.com x CloudLab - Instantiate a Profile x

https://www.cloudlab.us/instantiate.php#

Home Manual Actions

Ingo logged in Logout

The Apt cluster is experiencing some failures; not likely to be up before the end of the week

Current Usage: 0 Node Hours, Prev Week: 39, Prev Month: 39 (30 day rank: 136 of 176 users)

1. Profile 2. Parameterize 3. Finalize

Profile: OpenStack

Provides a highly-configurable OpenStack instance with a controller, network manager, and one or more nodes (potentially at multiple Cloudlab sites). This profile runs x86 or ARM64 nodes. It sets up OpenStack Juno (on Ubuntu 15.10, 15.04, or 14.10) according to your choice, and configures all OpenStack components in some VM disk images, and creates basic networks accessible via floating IPs. You'll be able to create and access them over the Internet in just a few minutes. When you click the Instantiate button, you'll be presented with a list of parameters that you can change to control what your OpenStack instance will look like; please refer to the parameter documentation on that page (or in the Instructions) to understand the various features available to you.

Copy Profile Show Profile Change Profile

Previous Next

https://www.cloudlab.us/manage\_profile.php Question or comment? Join the Help Forum Supported by NSF © 2015 The University of Utah



Inbox - linh.b.ngo@gmail.com x CloudLab - Manage Profile x

https://www.cloudlab.us/manage\_profile.php

Gmail Jupyter Hub Google Scholar Data Carpentry Software Carpentry Weather Clemson News Statistics BigData Introduction to Panda Other bookmarks

Home Manual Actions

Ingo logged in Logout

The Apt cluster is down due to multiple disk failures; not likely to be up before the end of the week

Current Usage: 0 Node Hours, Prev Week: 39, Prev Month: 39 (30 day rank: 136 of 176 users)

### Create Profile

Name

Project BDSLab

Your rspec  **Topology**

Description

Instructions

Who can instantiate your profile?

Anyone

Only members of your project

Create

Powered by emulab Question or comment? Join the Help Forum Supported by NSF © 2015 The University of Utah



CloudLab - Manage Profile x CloudLab - Experiment St: x

https://www.cloudlab.us/manage\_profile.php?action=edit&uuid=908b05e8-4520-11e6-9028-90e2ba22fee4

Gmail | Jupyter Hub | Google Scholar | Data Carpentry | Software Carpentry | Weather | Clemson | News | Statistics | BigData | Introduction to Panda | Other bookmarks

Home Manual

logged in Logout

### Topology Editor

Modify

Name: MPI-Assignme  
Version: 1  
Project: BDSLab  
Creator: Ingo  
Created: Jul 8, 2016 11:00  
Activity

Drag to Add

- ARM Server
- Bare Metal PC**
- Xen VM

New Site

Delete All Tidy View

Site 1

node-0 node-1

Instantiate Save

From 0

Accept Cancel

Powered by emulab Question or comment? Join the Help Forum Supported by NSF © 2015 The University of Utah





CloudLab - Manage Profile x

https://www.cloudlab.us/manage\_profile.php?action=edit&uuid=908b05e8-4520-11e6-9028-90e2ba22fee4

Gmail | Jupyter Hub | Google Scholar | Data Carpentry | Software Carpentry | Weather | Clemson | News | Statistics | BigData | Introduction to Panda | Other bookmarks

The Apt cluster

### Topology Editor

Delete All | Tidy View

Name: MPI-Assignm  
Version: 1  
Project: BDSLab  
Creator: Ingo  
Created: Jul 8, 2016 1

Activity

Instantiate | Save

From: 0

Powered by CloudLab | Question or comment? Join the Help Forum | Supported by NSF | © 2015 The University of Utah

**Topology Editor Configuration:**

- Name: node-0
- Node Type: raw-pc
- Hardware Type: c8220
- Disk Image: Ubuntu 14.04 LTS 64-bit
- Disable MAC Learning (For OVS Images Only)
- Publicly Routable IP
- Icon: Server

**Topology Diagram:**

```
graph TD; node-0[Server] --- Site1[Site 1]; node-1[Server] --- Site1;
```

- Both nodes should have the same configuration, no network connection is needed (due to public IP)



The Apt cluster is down due to multiple disk failures; not likely to be up before the end of the week

Current Usage: 0 Node Hours, Prev Week: 39, Prev Month: 39 (30 day rank: 136 of 176 users)

### Modify Profile

Name: MPI-Assignment  
Version: 1  
Project: BDSLab  
Creator: Ingo  
Created: Jul 8, 2016 11:28 AM

[Activity](#)

### Your rspec

[Choose file](#) [Topology](#) [Edit Source](#)

### Description

Two C-8820 nodes from Clemson, Ubuntu 14.04LTS, only one site

### Instructions

[Show/Edit Tour](#)

Who can instantiate your profile?

Anyone

Only members of your project

[Copy](#) [Share](#) [Delete](#) [Instantiate](#) [Save](#)

### Version History

Vers	Description	Created	From
1	Two C-8820 nodes from Clemson, Ubuntu 14.04LTS, only one site	2016-07-08	0
0	C220M4/Ubuntu 14.04 LTS/Site 1 C8220/Ubuntu 14.04 LTS/Site 2	2016-07-08	

Powered by emulab Question or comment? Join the Help Forum Supported by NSF © 2015 The University of Utah

- You can save multiple versions of your topology
- Instantiate the version that you want to launch
- The launching procedure will be similar to the OpenStack tutorial



# MPI on CloudLab

## On each node:

```
sudo apt-get update
```

```
sudo apt-get install libibnetdisc-dev
```

```
sudo nano /etc/environment
```

```
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/home/mpiuser/.openmpi/bin"
```

```
LD_LIBRARY_PATH="/lib:/usr/lib:/usr/local/lib:/home/mpiuser/.openmpi/lib/"
```

```
sudo adduser mpiuser
```



# MPI on CloudLab

```
sudo adduser mpiuser  
ssh mpiuser@localhost
```

```
wget https://www.open-  
mpi.org/software/ompi/v1.8/downloads/openmpi-  
1.8.1.tar.gz
```

```
tar xzf openmpi-1.8.1.tar.gz  
cd openmpi-1.8.1
```

```
./configure --prefix="/home/mpiuser/.openmpi"  
make  
make install
```



# MPI on CloudLab

```
ssh-keygen -t rsa
```

```
cd .ssh
```

```
cp id_rsa.pub authorized_keys
```

```
ssh-copy-id -i id_rsa.pub <hostname of the other node>
```

```
nano nodelist
```

```
<hostname of first node>
```

```
<hostname of second node>
```

```
...
```



# Example program

```
#include <stdio.h>
#include <unistd.h>
#include <sys/utsname.h>
#include <mpi.h>

int main(int argc, char *argv[]){
    int rank, size;
    MPI_Status status;
    MPI_Init(&argc, &argv);
    MPI_Comm_size(MPI_COMM_WORLD, &size);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);

    struct utsname uts;
    uname(&uts);
    printf("%d at %s\n", rank, uts.nodename);
    MPI_Finalize();
    return 0;
}
```



# MPI on CloudLab

```
mpicc gethostname.c -o gethostname
```

```
scp gethostname mpiuser@<the other node>:/home/mpiuser
```

```
scp nodelist mpiuser@<the other node>:/home/mpiuser
```

```
mpirun -np 2 -machinefile nodelist ./gethostname
```

```
mpirun -np 2 -machinefile nodelist --map-by node  
./gethostname
```



# Assignment Ideas

- Develop a work queue using various allocation strategies:
  - Normal
  - Cyclic
  - Dynamic
- Setup MPI cluster with nodes on separate sites, reduce network connection, and evaluate performance on different allocation strategies





# Hadoop on CloudLab

- Enterprise Hadoop
- Hortonworks
- <http://hortonworks.com/hdp/downloads/>



# Hadoop on CloudLab

The screenshot shows the CloudLab 'Topology Editor' interface. On the left, a configuration panel for a node named 'datanode0' is open, showing the following settings:

- Name: datanode0
- Node Type: raw-pc
- Hardware Type: c8220
- Disk Image: CENTOS 6.6 supports VM client...
- Disable MAC Learning (For OVS Images Only)
- Publicly Routable IP
- Icon: Server

The main editor area shows a diagram with a 'namenode' icon, a 'Site 1' label, and a 'datanode0' icon highlighted with a green border. At the top right of the editor are 'Delete All' and 'Tidy View' buttons. At the bottom right are 'Accept' and 'Cancel' buttons.

- Both nodes should have the same configuration on bare metal PC, no network connection is needed (due to public IP)



# On each node

- SSH onto the node from Palmetto

- Change to root:

```
sudo su -
```

- Execute the following commands:

```
chkconfig --list ntpd
```

```
chkconfig ntpd on
```

```
service ntpd start
```

```
chkconfig iptables off
```

```
/etc/init.d/iptables stop
```

```
setenforce 0
```



# On each node

- Setup Ambari download server

```
wget -nv http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.1.2/ambari.repo -O  
/etc/yum.repos.d/ambari.repo
```

- On namenode

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

- On datanode

```
yum -y install ambari-agent
```



# On namenode

- Set up ambari server:

```
ambari-server setup
```

- Select default for all questions
- Select 1 for JDK version
- When all done, start ambari server

```
ambari-server start
```



# On each node

- Using vim to edit `/etc/ambari-agent/conf/ambari-agent.ini`

- Change:

```
hostname=<hostname of namenode as shown in list view of CloudLab>
```

- Start Ambari Agent

```
ambari-agent start
```



# Ambari Server (admin/admin)

The screenshot shows a web browser window with the URL `clnode095.clemson.cloudlab.us:8080/#/login`. The browser's address bar and tabs are visible at the top. The page content includes a dark header with the Ambari logo and name. The main area features a light gray box with the heading "Sign in" and two input fields for "Username" and "Password". A green "Sign in" button is positioned below the password field. At the bottom of the page, there is a footer with the text: "Licensed under the Apache License, Version 2.0. See [third-party tools/resources that Ambari uses and their respective authors](#)".



# Ambari Server (admin/admin)

The screenshot shows the Ambari Admin web interface. The browser address bar displays `clnode095.clemson.cloudlab.us:8080/views/ADMIN_VIEW/2.1.2/INSTANCE/#/`. The page header includes the Ambari logo and a user dropdown menu set to 'admin'. The main content area is titled 'Welcome to Apache Ambari' and contains three primary action cards:

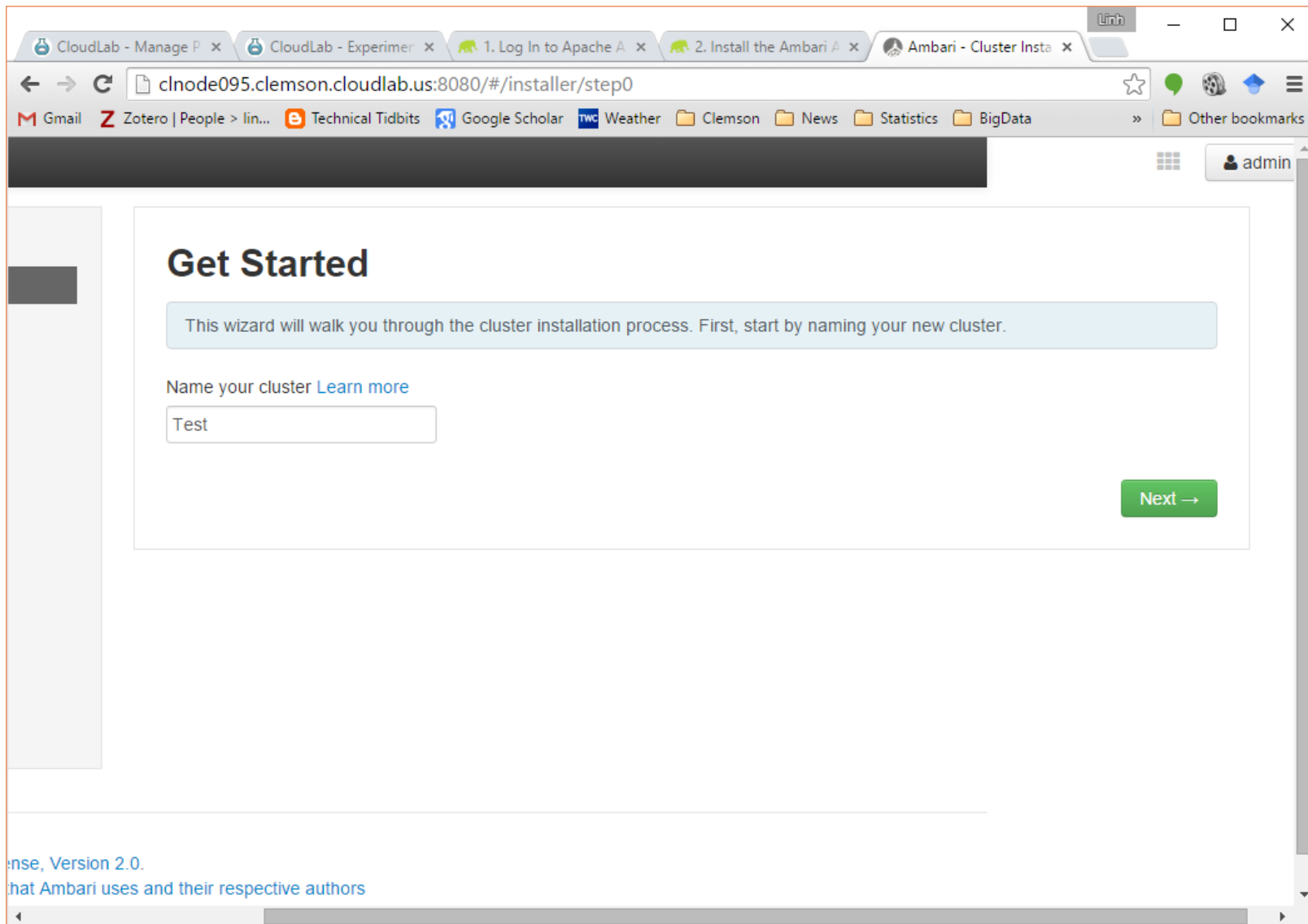
- Create a Cluster:** Includes the text 'Use the Install Wizard to select services and configure your cluster' and a blue button labeled 'Launch Install Wizard' which is circled in orange.
- Manage Users + Groups:** Includes the text 'Manage the users and groups that can access Ambari' and two buttons: 'Users' and 'Groups'.
- Deploy Views:** Includes the text 'Create view instances and grant permissions' and a button labeled 'Views'.

A left-hand sidebar contains navigation links for 'Clusters' (No clusters), 'Views' (Views), and 'User + Group Management' (Users, Groups).





# Ambari Server (admin/admin)



The screenshot shows a web browser window with the Ambari Server installation wizard. The browser's address bar shows the URL `clnode095.clemson.cloudlab.us:8080/#/installer/step0`. The page title is "Get Started". A light blue box contains the text: "This wizard will walk you through the cluster installation process. First, start by naming your new cluster." Below this, the text "Name your cluster" is followed by a link "Learn more". A text input field contains the word "Test". A green "Next" button with a right-pointing arrow is located at the bottom right of the main content area. The browser's bookmark bar includes links for Gmail, Zotero, Technical Tidbits, Google Scholar, Weather, Clemson, News, Statistics, and BigData. The user is logged in as "admin".

**Get Started**

This wizard will walk you through the cluster installation process. First, start by naming your new cluster.

Name your cluster [Learn more](#)

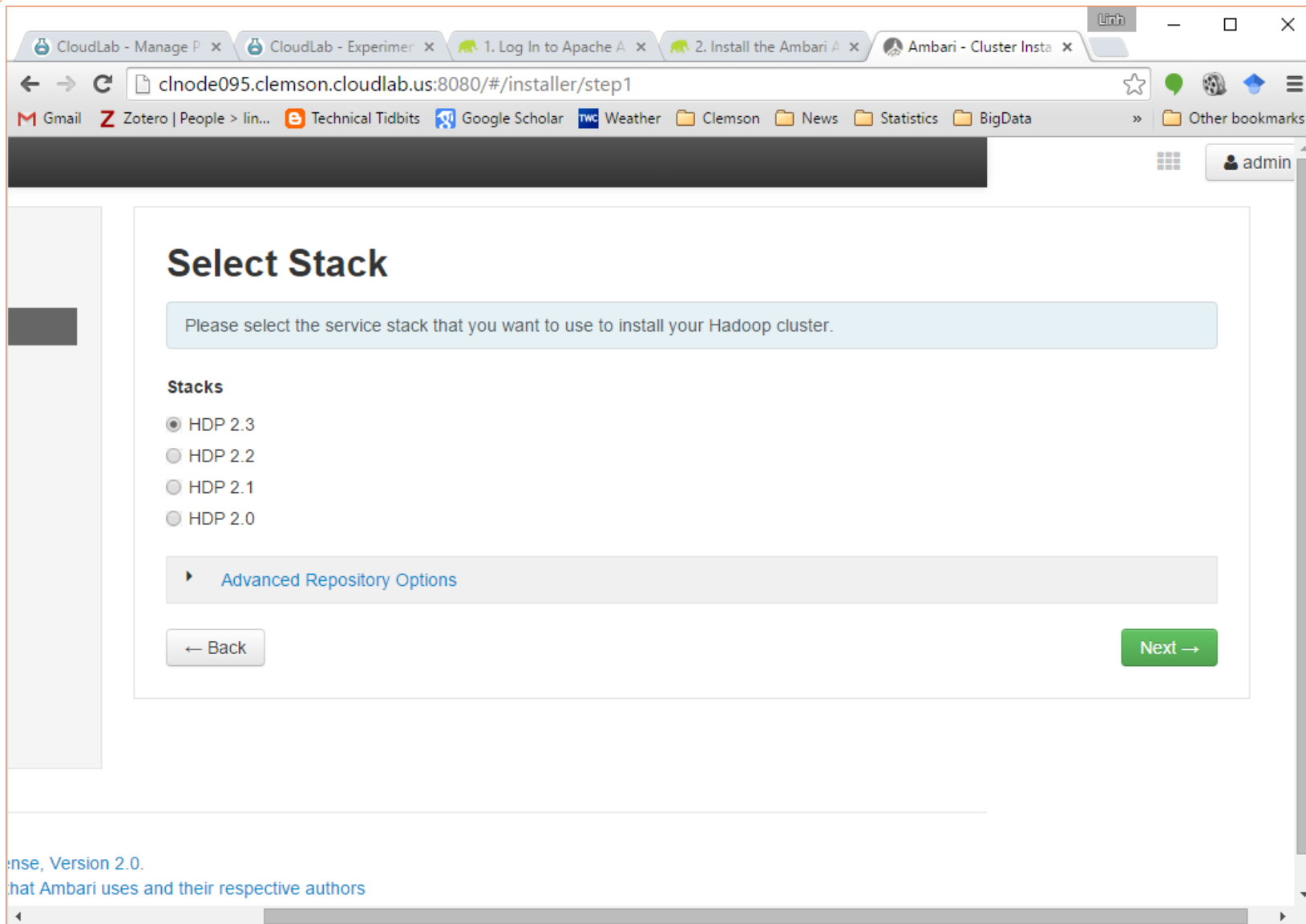
Test

Next →

Ambari Server, Version 2.0.  
[View the license that Ambari uses and their respective authors](#)



# Ambari Server (admin/admin)



The screenshot shows a web browser window with the URL `clnode095.clemson.cloudlab.us:8080/#/installer/step1`. The browser's address bar and tabs are visible at the top. The main content area is titled "Select Stack" and contains a light blue instruction box: "Please select the service stack that you want to use to install your Hadoop cluster." Below this, under the heading "Stacks", there are four radio button options: "HDP 2.3" (selected), "HDP 2.2", "HDP 2.1", and "HDP 2.0". A grey expandable section titled "Advanced Repository Options" is visible below the stack options. At the bottom of the main content area, there are two buttons: "← Back" and "Next →".

**Select Stack**

Please select the service stack that you want to use to install your Hadoop cluster.

**Stacks**

- HDP 2.3
- HDP 2.2
- HDP 2.1
- HDP 2.0

Advanced Repository Options

← Back Next →

ense, Version 2.0.  
that Ambari uses and their respective authors



# Ambari Server (admin/admin)

CloudLab - Experiment St... x 1. Log In to Apache Amba... x 2. Install the Ambari Agen... x Ambari - Cluster Install Wi... x

clnode095.clemson.cloudlab.us:8080/#/installer/step2

Gmail Z Zotero | People > lin... Technical Tidbits Google Scholar TWC Weather Clemson News Statistics BigData Other bookmarks

admin

## 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](#)

```
clnode095.clemson.cloudlab.us
clnode092.clemson.cloudlab.us
```

### Host Registration Information

Provide your [SSH Private Key](#) to automatically register hosts

Choose File No file chosen

```
ssh private key
```

SSH User Account root

Perform [manual registration](#) on hosts and do not use SSH

← Back Register and Confirm →



# Assuming you had ambari agents up and running ...

The screenshot shows the Ambari Cluster Install Wizard at step 3, "Confirm Hosts". The browser address bar shows the URL: `clnode095.clemson.cloudlab.us:8080/#/installer/step3`. The page title is "Confirm Hosts".

A light blue box contains the text: "Registering your hosts. Please confirm the host list and remove any hosts that you do not want to include in the cluster."

Below this is a table with columns: Host, Progress, Status, and Action. There are two hosts listed, both with a progress bar and a status of "Success".

<input type="checkbox"/>	Host	Progress	Status	Action
<input type="checkbox"/>	clnode095.clemson.cloudlab.us	<div style="width: 100%; height: 10px; background-color: green;"></div>	Success	<input type="button" value="Remove"/>
<input type="checkbox"/>	clnode092.clemson.cloudlab.us	<div style="width: 100%; height: 10px; background-color: green;"></div>	Success	<input type="button" value="Remove"/>

At the bottom of the table, there is a "Show:" dropdown set to "25" and a pagination indicator "1 - 2 of 2".

A yellow warning box contains the text: "Some warnings were encountered while performing checks against the 2 registered hosts above. [Click here to see the warnings.](#)"

At the bottom of the page, there are two buttons: "← Back" and "Next →".



# Ambari Server (admin/admin)

- HDFS
- YARN+MapReduce2
- Tez
- ZooKeeper
- Ambari Metrics



# Ambari Server (admin/admin)

CloudLab - Experiment St... x 1. Log In to Apache Amba x 2. Install the Ambari Agen x Ambari - Cluster Install Wi x

clnode095.clemson.cloudlab.us:8080/#/installer/step5

Gmail Zotero | People > lin... Technical Tidbits Google Scholar TWC Weather Clemson News Statistics BigData Other bookmarks

admin

## Assign Masters

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

SNameNode:	clnode095.clemson.cloudlab.us (: ▾)	clnode092.clemson.cloudlab.us (252.2 GB, 20 cores) <b>ZooKeeper Server</b>
NameNode:	clnode095.clemson.cloudlab.us (: ▾)	
History Server:	clnode095.clemson.cloudlab.us (: ▾)	clnode095.clemson.cloudlab.us (252.2 GB, 20 cores) <b>SNameNode</b> <b>NameNode</b> <b>History Server</b> <b>App Timeline Server</b> <b>ResourceManager</b> <b>ZooKeeper Server</b> <b>Metrics Collector</b>
App Timeline Server:	clnode095.clemson.cloudlab.us (: ▾)	
ResourceManager:	clnode095.clemson.cloudlab.us (: ▾)	
ZooKeeper Server:	clnode095.clemson.cloudlab.us (: ▾) -	
ZooKeeper Server:	clnode092.clemson.cloudlab.us (: ▾) -	
Metrics Collector:	clnode095.clemson.cloudlab.us (: ▾)	

← Back Next →



# Ambari Server (admin/admin)

CloudLab - Experiment St... x 1. Log In to Apache Amba... x 2. Install the Ambari Agen... x Ambari - Cluster Install Wi... x

clnode095.clemson.cloudlab.us:8080/#/installer/step6

Gmail Zotero | People > lin... Technical Tidbits Google Scholar TWC Weather Clemson News Statistics BigData Other bookmarks

admin

## 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, MapReduce2 Client, YARN Client, Tez Client and ZooKeeper Client.

Host	all   none	all   none	all   none	all   none
clnode095.clemson.cloudl... *	<input type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> Client
clnode092.clemson.cloudl... *	<input checked="" type="checkbox"/> DataNode	<input type="checkbox"/> NFSGateway	<input checked="" type="checkbox"/> NodeManager	<input type="checkbox"/> Client

Show: 25 1 - 2 of 2

← Back Next →

Ambari Server, Version 2.0.  
[View the Ambari Server License](#) and [View the Ambari Server Source Code](#) that Ambari uses and their respective authors



# Edit configuration as you see fit

The screenshot shows the Ambari Cluster Install Wizard configuration page for NameNode and DataNode settings. The browser address bar shows the URL: `clnode095.clemson.cloudlab.us:8080/#/installer/step7`. The page is titled "Settings" and "Advanced".

**NameNode Configuration:**

- NameNode directories: `/hadoop/hdfs/namenode`
- NameNode Java heap size: **125GB** (range: 0 GB to 252.188 GB)
- NameNode Server threads: **256** (range: 1 to 500)
- Minimum replicated blocks %: **100%** (range: 99 % to 100 %)

**DataNode Configuration:**

- DataNode directories: `/hadoop/hdfs/data`
- DataNode failed disk tolerance: **0** (range: 0 to 1)
- DataNode maximum Java heap size: **125.75GB** (range: 0 GB to 252.188 GB)
- DataNode max data transfer threads: **4096** (range: 0 to 48000)

An "Edit" button is visible next to the NameNode Java heap size slider.





# Deploy ...

The screenshot shows a web browser window with the Ambari Cluster Install Wizard interface. The browser's address bar displays the URL `clnode095.clemson.cloudlab.us:8080/#/installer/step8`. The page title is "Review". A modal dialog box is centered on the screen, displaying the text "Preparing to Deploy: 25 of 25 tasks completed." and a blue progress bar that is fully filled. The background content, which is dimmed, includes a "Review" section with the heading "Please review the configuration before installation" and a list of services and their configurations:

- HDFS**
  - DataNode : 1 host
  - NameNode : clnode095.clemson.cloudlab.us
  - NFSGateway : 0 host
  - SNameNode : clnode095.clemson.cloudlab.us
- YARN + MapReduce2**
  - App Timeline Server : clnode095.clemson.cloudlab.us
  - NodeManager : 1 host
  - ResourceManager : clnode095.clemson.cloudlab.us
- Tez**
  - Clients : 1 host
- ZooKeeper**
  - Server : 2 hosts
- Ambari Metrics**
  - Metrics Collector : clnode095.clemson.cloudlab.us

At the bottom of the page, there are buttons for "Back", "Edit", "Print", and "Deploy".



# Warning due to lack of space and failed checks (ignore)

The screenshot shows the Ambari Cluster Install Wizard interface. The main heading is "Install, Start and Test". Below it, a blue box says "Please wait while the selected services are installed and started." A progress bar shows "100 % overall". A filter bar indicates "Show: All (2) | In Progress (0) | Warning (1) | Success (1) | Fail (0)". A table lists the status of two hosts:

Host	Status	Message
clnode095.clemson.cloudlab.us	100% (Warning)	Warnings encountered
clnode092.clemson.cloudlab.us	100% (Success)	Success

Below the table, it says "2 of 2 hosts showing - Show All" and "Show: 25 | 1 - 2 of 2". A yellow box at the bottom says "Installed and started the services with some warnings." and a green "Next" button is visible.

Version 2.0.  
Ambari uses and their respective authors

Edit



# HDFS

CloudLab - Experiment S x 1. Log In to Apache Amb x 2. Install the Ambari Age x Ambari - Test x Namenode information x All Applications x

clnode095.clemson.cloudlab.us:50070/dfshealth.html#tab-overview

Gmail Zotero | People > lin... Technical Tidbits Google Scholar tw Weather Clemson News Statistics BigData Picasa Web Albums ... Conferences Other bookmarks

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Overview 'clnode095.clemson.cloudlab.us:8020' (active)

<b>Started:</b>	Wed Oct 07 23:34:32 EDT 2015
<b>Version:</b>	2.7.1.2.3.2.0-2950, r5cc60e0003e33aa98205f18bccaeaf36cb193c1c
<b>Compiled:</b>	2015-09-30T18:08Z by jenkins from (HEAD detached at 5cc60e0)
<b>Cluster ID:</b>	CID-c31ab4be-d8d4-42a9-b456-7745974a681d
<b>Block Pool ID:</b>	BP-1609302222-130.127.133.104-1444275269044

## Summary

Security is off.  
Safemode is off.

38 files and directories, 11 blocks = 49 total filesystem object(s).

Heap Memory used 15.41 GB of 123.44 GB Heap Memory. Max Heap Memory is 123.44 GB.

Non Heap Memory used 60.15 MB of 61.21 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

<b>Configured Capacity:</b>	14.75 GB
<b>HDFS Used:</b>	250.95 MB (1.7%)




# YARN

CloudLab - Experiment S x 1. Log In to Apache Amb x 2. Install the Ambari Age x Ambari - Test x Namenode information x All Applications x

clnode095.clemson.cloudlab.us:8088/cluster

Gmail Zotero | People > lin... Technical Tidbits Google Scholar TWC Weather Clemson News Statistics BigData Picasa Web Albums ... Conferences Other bookmarks



## All Applications

- Cluster
  - About
  - Nodes
  - Node Labels
  - Applications
    - NEW
    - NEW SAVING
    - SUBMITTED
    - ACCEPTED
    - RUNNING
    - FINISHED
    - FAILED
    - KILLED
  - Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lo: Nod
4	0	0	4	0	0 B	220 GB	0 B	0	16	0	1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum
Capacity Scheduler	[MEMORY]	<memory:56320, vCores:1>	<memory:225280, vCores:

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Progress
application_1444275343859_0004	ambari-qa	DistributedShell	YARN	default	Wed Oct 7 23:40:32-0400 2015	Wed Oct 7 23:40:40-0400 2015	FINISHED	SUCCEEDED	N/A	
application_1444275343859_0003	ambari-qa	OrderedWordCount	TEZ	default	Wed Oct 7 23:39:35-0400 2015	Wed Oct 7 23:40:33-0400 2015	FINISHED	FAILED	N/A	
application_1444275343859_0002	ambari-qa	OrderedWordCount	TEZ	default	Wed Oct 7 23:38:33-0400 2015	Wed Oct 7 23:39:31-0400 2015	FINISHED	FAILED	N/A	
application_1444275343859_0001	ambari-qa	OrderedWordCount	TEZ	default	Wed Oct 7 23:37:29-0400 2015	Wed Oct 7 23:38:29-0400 2015	FINISHED	FAILED	N/A	

Showing 1 to 4 of 4 entries



# Tutorial

```
sudo su hdfs
hdfs dfs -mkdir /user/<username>
hdfs dfs -chown <username>:<username> /user/<username>
exit to <username>
hdfs dfs -ls /user/
git clone https://github.com/clemsoncoe/Introduction-to-Hadoop-data.git
cd Introduction-to-Hadoop-data
hdfs dfs -put gutenbergs-shakespeare.txt /user/<username>/
yarn jar /usr/hdp/current/hadoop-mapreduce-client/hadoop-mapreduce-examples-2.7.1.2.3.6.0-3796.jar wordcount
gutenbergs-Shakespeare.txt output/
hdfs dfs -ls output
hdfs dfs -cat output/part-r-00000
```



# Assignment Ideas

- Deploy a Hadoop cluster and upload a large data set (Airline on-time performance data: <http://stat-computing.org/dataexpo/2009/the-data.html>)
- Examine and investigate performance of Hadoop MapReduce as data nodes are killed/added to the cluster
- Examine performance of Hadoop MapReduce as data nodes are located on different sites