# Lab #5 – Cloudera on Google Cloud

Some students were not able to successfully download and configure Cloudera on VirtualBox due to various computer capacity issues. You can follow this guide as an alternative solution for Cloudera on VirtualBox (Lab #3 and Lab #4). If your VM performs poorly and runs very slow, you can also follow this guide.

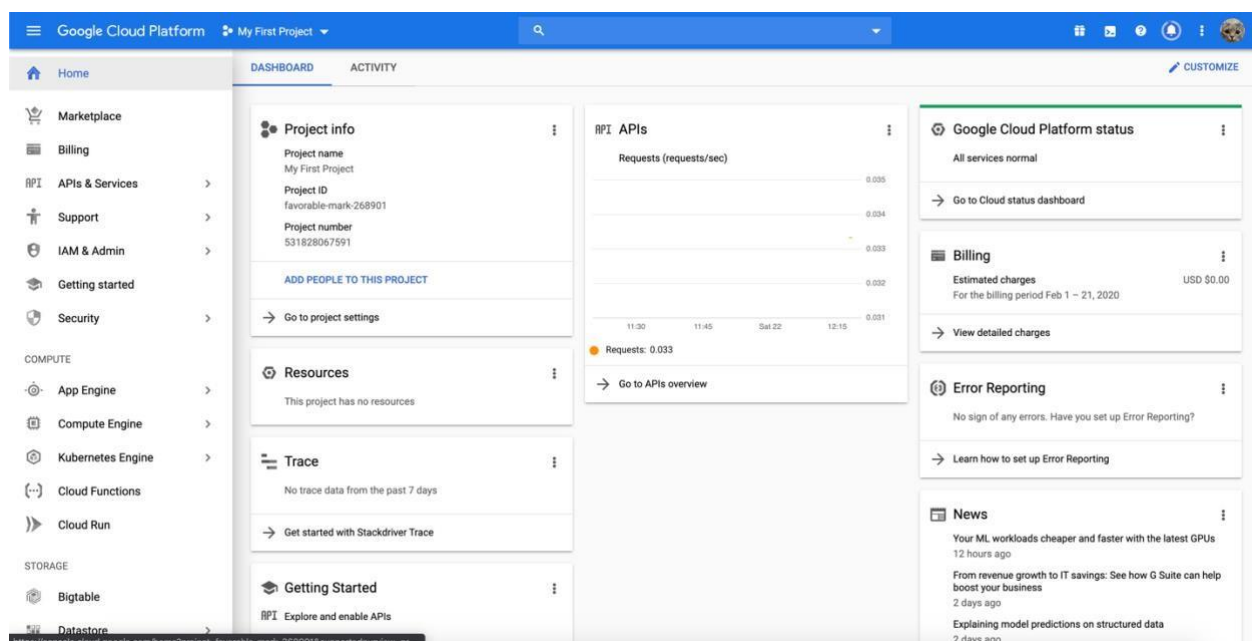## Step 1: Sign in Google Cloud with Gmail and create a new project.

Go to https://cloud.google.com/
Log in with your Gmail account, you will have to link one of your bank cards to receive $300 free credit to use in one year. (I used my Fordham email here)

After login, fill out all information if asked, go to Console.
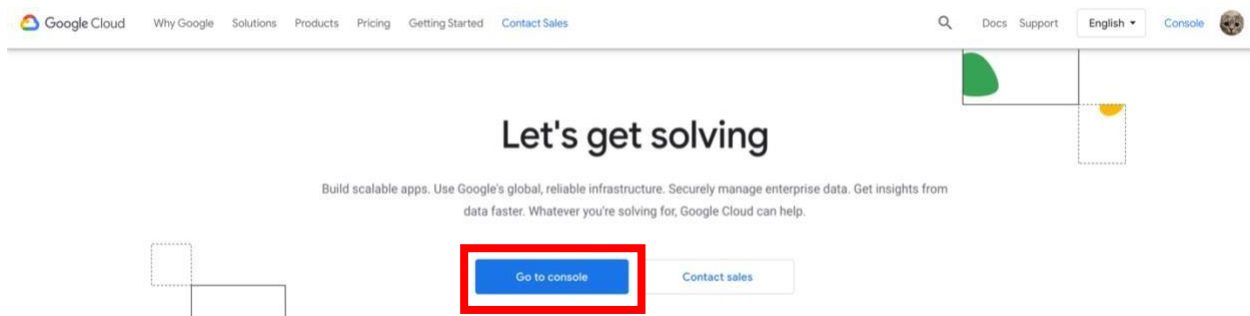
Then, you will see the main menu bar on the left side of the dashboard.

Click the down arrow to create a new project.
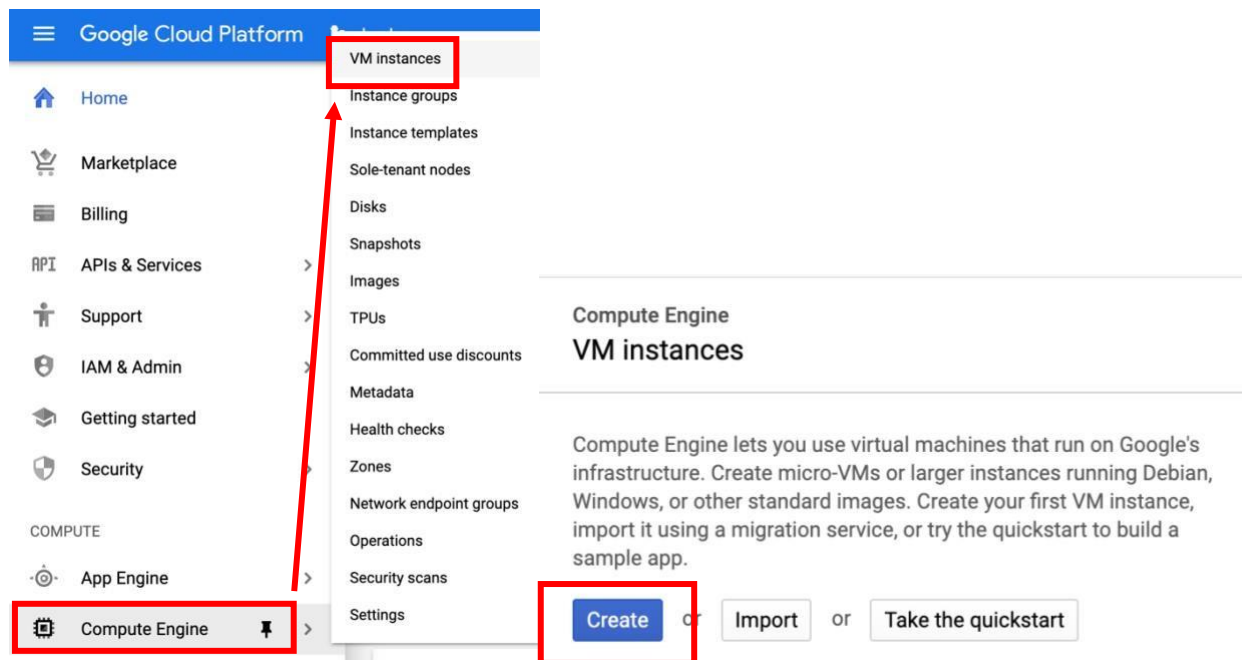
Give your project a name.

The dashboard of your new project.
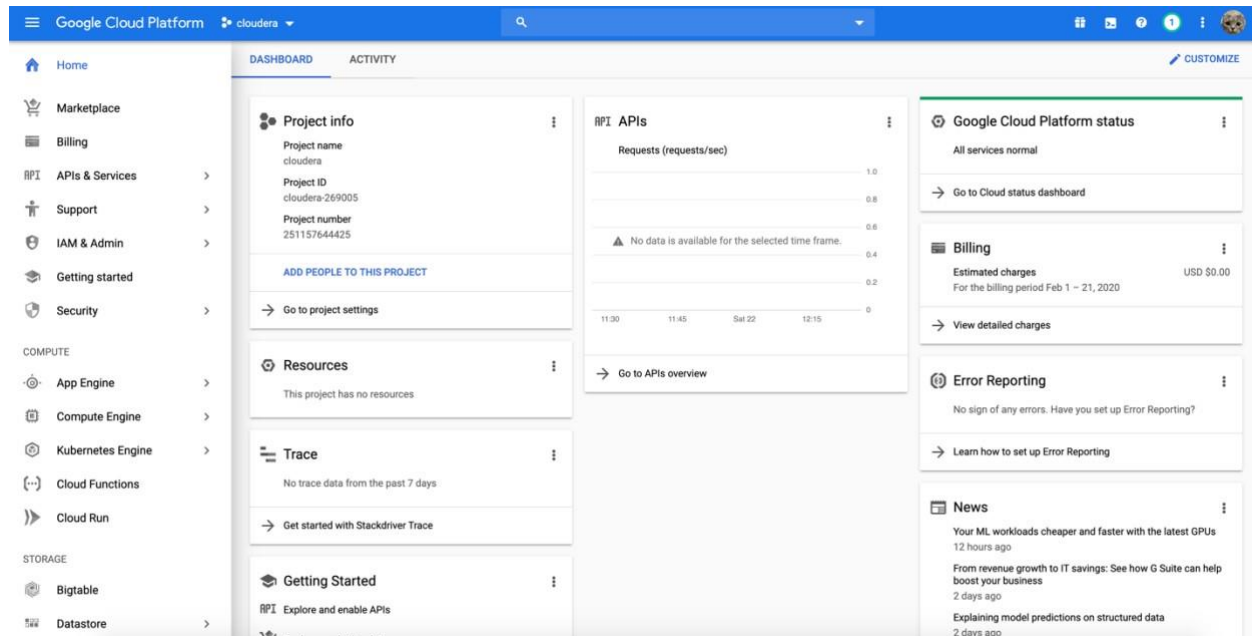
Congrats! You have created your new project!

Step 2: Create VM instances on in your project.

Under the menu bar find Compute Engine.

Then, follow the below step 1~10 carefully and configure the instances exactly as the picture shown.

Name ⊘
Name is permanent

clouder-multi-node-cluster   **Step 1: Give it name.**

Labels ⊘ (Optional)

+ Add label

Region ⊘   **Step 2**
Region is permanent   **Zone** ⊘
                             Zone is permanent

us-east1 (South Caro... ▾   us-east1-b   ▾

Machine configuration ⊘

Machine family

General-purpose | Memory-optimized
Compute-optimized

Machine types for common workloads, optimized
for cost and flexibility

Series

E2   **Step 3**   ▾

CPU platform selection based on availability

Machine type

e2-standard-2 (2 vCPU, 8 GB memory)   ▾

          vCPU   Memory
          2   8 GB

� CPU platform and GPU

Container ⊘
☐ Deploy a container image to this VM instance. Learn
more

Boot disk ⊘

          New 20 GB standard persistent disk **Step 4**
          Image
          Red Hat Enterprise Linux 7   Change

Use of Red Hat products is subject to Red Hat license
usage reporting policy. Learn more

Identity and API access ⊘

Service account ⊘
Compute Engine default service account   ▾

Access scopes ⊘   **Step 8**
○ Allow default access
● Allow full access to all Cloud APIs
○ Set access for each API

Firewall ⊘   **Step 9**
Add tags and firewall rules to allow specific network
traffic from the Internet
☑ Allow HTTP traffic
☑ Allow HTTPS traffic

� Management, security, disks, networking, sole
tenancy

Your free trial credit will be used for this VM instance.
GCP Free Tier ↗   **Step 10**

Create Cancel

Equivalent REST or command line

Boot disk

Select an image or snapshot to create a boot disk; or attach an existing disk. C
Marketplace.

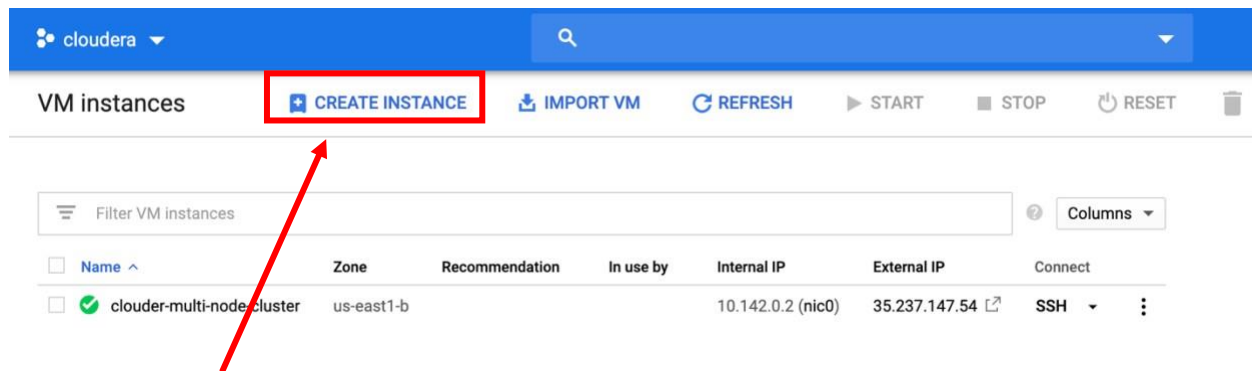Public images | Custom images | Snapshots | Existing disks

☐ Show images with Shielded VM features ⊘   **Step 5**

Operating system
Red Hat Enterprise Linux   ▾

Version   **Step 6**
Red Hat Enterprise Linux 7   ▾

x86_64 built on 20200205

Boot disk type ⊘   Size (GB) ⊘ **Step 7**
Standard persistent disk   ▾   30

Click on **Create instance**, and **repeat** the above 10 steps to create 3 more instances, in **step 1 name them node1, node2, node3**. All other steps are exactly the same.

After you done that, you should see all instances.

Connect to **all 4 instances** using the built-in browser window.

The pop-up window looks like this for all 4 instances (you should have 4 windows running, one instance each).

Run the following commands!

Remember to press "I" key to change to insert mode to edit the file.

```
# This file controls the state of SELinux on the system.
# SELINUX= can take one of these three values:
#     enforcing - SELinux security policy is enforced.
#     permissive - SELinux prints warnings instead of enforcing.
#     disabled - No SELinux policy is loaded.
SELINUX=disabled
# SELINUXTYPE= can take one of three values:
#     targeted - Targeted processes are protected,
#     minimum - Modification of targeted policy. Only selected processes are protected.
#     mls - Multi Level Security protection.
SELINUXTYPE=targeted
```

4: Press the "I" key to change to insert mode, and change SELNUX to disabled so that we can successfully install Cloudera

5: After editing, press "ESC" key to exit edit mode

6: Type in ':wq!' to exit vim editor

```
:wq!
Connected, host fingerprint: ssh-rsa 0 33:D5:63:64:1E:DC:A6:FB:4B:DE:
:5E:3C:40:19:08:71:B7:26:D7:1B:77:9C:68:50:9F:EA:67:5D
Last login: Sat Feb 22 05:51:17 2020 from 74.125.177.34
[mli222@clouder-multi-node-cluster ~]$ sudo su
[root@clouder-multi-node-cluster mli222]# service iptables stop
Redirecting to /bin/systemctl stop iptables.service
Failed to stop iptables.service: Unit iptables.service not loaded.
[root@clouder-multi-node-cluster mli222]# vi /etc/selinux/config
[root@clouder-multi-node-cluster mli222]# vi /etc/ssh/sshd_config
```

7: Configure another file

```
# Authentication:

#LoginGraceTime 2m
PermitRootLogin yes
#StrictModes yes
#MaxAuthTries 6
#MaxSessions 10
```

8: In the sshd_config file, find Authentication section, and change PermitRootLogin to yes

```
# To disable tunneled clear text passwords, change to no here!
PasswordAuthentication yes
#PermitEmptyPasswords no
#PasswordAuthentication no

# Change to no to disable s/key passwords
ChallengeResponseAuthentication yes
#ChallengeResponseAuthentication no
```

9: Use up and down arrow keys to browse the file, and find the sections as the picture shown. You will need to check the "#" signs and add or delete them.

10: Change the root access password

```
[root@clouder-multi-node-cluster mli222]# vi /etc/ssh/sshd_config
[root@clouder-multi-node-cluster mli222]# sudo passwd
Changing password for user root.
New password:
Retype new password:
passwd: all authentication tokens updated successfully.
[root@clouder-multi-node-cluster mli222]# reboot
```

11: Set a password and REMEMBER it, we will need it later for ssh setting.

12: Reboot the instance

Due to lost network connectivity, the SSH connection to VM instance 'clouder-multi-node-cluster' was closed. Please restablish your network connection and then reconnect.

[Reconnect]   [Dismiss]

Repeat the above steps for all 4 instances, make sure you use the same password for all instances.

After you configured all 4 instances, go back to the window for the **cloudera-multi-nodecluster** which is the instance we will download Cloudera installer on. Run the following commands to download.

1. Switch to root user.
   sudo su

2. Check if Selinux is disabled.
   sestatus

3. Install yum. **yum install wget**

4. Download the package.

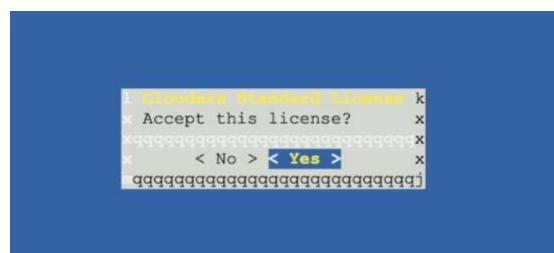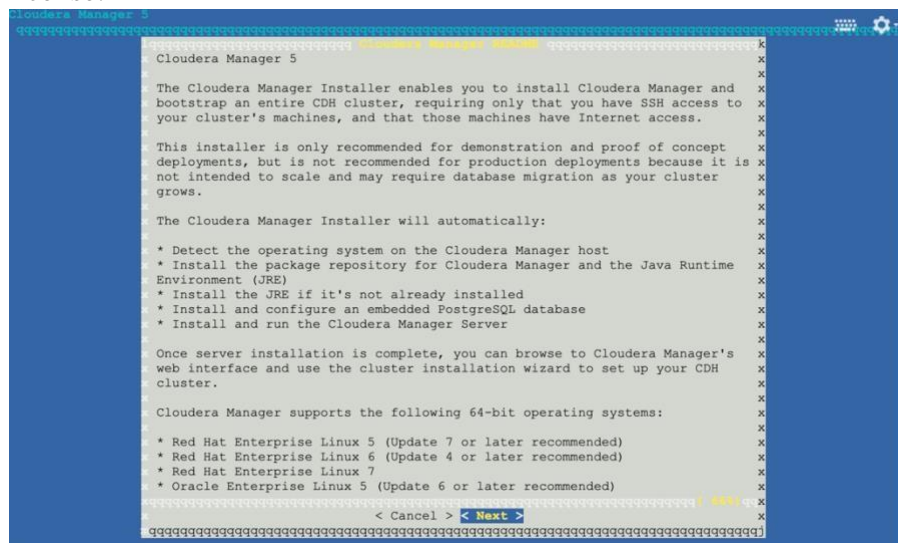**Wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin**

5. Make the file accessible.
chmod +x cloudera-manager-installer.bin

6. Open the installer.
./cloudera-manager-installer.bin

You should see something like this. Use arrow keys and return keys to navigate. Accept all license.
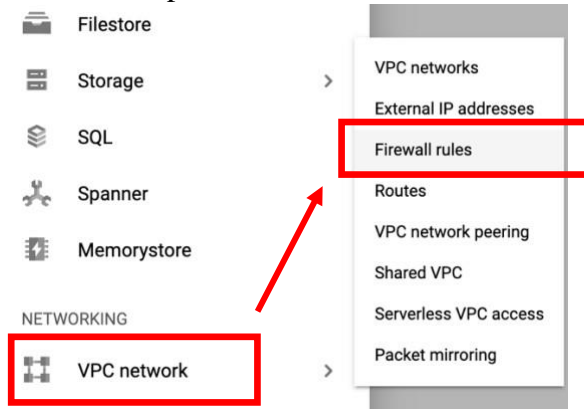

This step will take a few minutes.

Press enter to finish the installation process.

Step 3: Firewall Settings

Follow the steps to add two firewall rules for us to open Cloudera Manager and HUE UIs.

Make sure you set everything exactly as the picture shown below.

There are two pictures:
One for port: 7180 (Cloudera Manager)
Second for port: 8888 (HUE)

← Create a firewall rule

Firewall rules control incoming or outgoing traffic to an instance. By default, incoming traffic from outside your network is blocked. Learn more

Name *
cloudera-7180

Lowercase letters, numbers, hyphens allowed

Description

Logs
Turning on firewall logs can generate a large number of logs which can increase costs in Stackdriver. Learn more
○ On
◉ Off

Network *
default   ▼ ?

Priority *
1000   ?

Priority can be 0 - 65535 Check priority of other firewall rules

Direction of traffic ?
◉ Ingress
○ Egress

Action on match ?
◉ Allow
○ Deny

Targets
All instances in the network   ▼ ?

Source filter
IP ranges   ▼ ?

Source IP ranges *
0.0.0.0/0 ⊗   for example, 0.0.0.0/0, 192.168.2.0/24   ?

Second source filter
None   ▼ ?

Protocols and ports ?
○ Allow all
◉ Specified protocols and ports

☑ tcp :   7180

☐ udp :   all

☐ Other protocols

protocols, comma separated, e.g. ah, sctp

⌄ DISABLE RULE

CREATE   CANCEL

Equivalent REST or command line

← Create a firewall rule

Firewall rules control incoming or outgoing traffic to an instance. By default, incoming traffic from outside your network is blocked. Learn more

Name *
cloudera-8888   ?
Lowercase letters, numbers, hyphens allowed

Description

Logs
Turning on firewall logs can generate a large number of logs which can increase costs in Stackdriver. Learn more
○ On
◉ Off

Network *
default   ▼   ?

Priority *
1000   ?
Priority can be 0 - 65535 Check priority of other firewall rules

Direction of traffic ?
◉ Ingress
○ Egress

Action on match ?
◉ Allow
○ Deny

Targets
All instances in the network   ▼   ?

Source filter
IP ranges   ▼   ?

Source IP ranges *
0.0.0.0/0 ⊗ for example, 0.0.0.0/0, 192.168.2.0/24   ?

Second source filter
None   ▼   ?

Protocols and ports ?
○ Allow all
◉ Specified protocols and ports

☑ tcp :   8888

☐ udp :   all

☐ Other protocols
   protocols, comma separated, e.g. ah, sctp
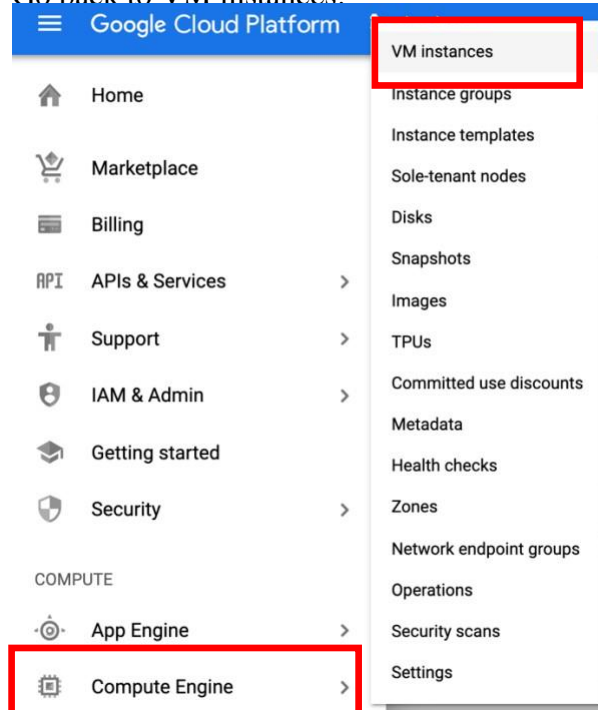
⌄ DISABLE RULE

CREATE   CANCEL

Equivalent REST or command line

After you create two firewall rules, you should see them.

	cloudera-7180	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:7180	Allow	1000	default
	cloudera-8888	Ingress	Apply to all	IP ranges: 0.0.0.0/0	tcp:8888	Allow	1000	default

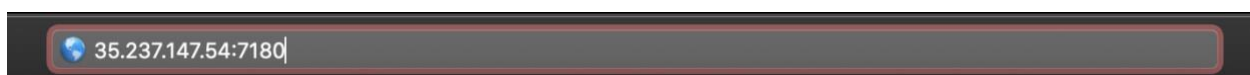## Step 4: Open up browsers and go to web UIs

Go back to VM instances.

Copy the External IP for the clouder-multi-node-cluster.

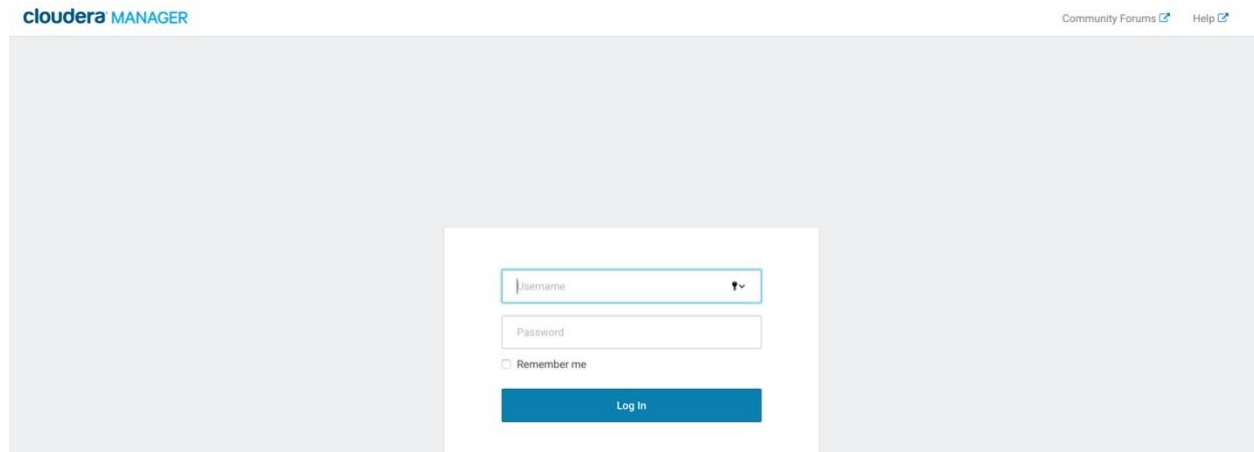|	Name ^	Zone	Recommendation	In use by	Internal IP	External IP	Connect	
☐	✅ clouder-multi-node-cluster	us-east1-b			10.142.0.2 (nic0)	35.237.147.54	SSH ▾	⋮
☐	✅ node1	us-east1-b			10.142.0.3 (nic0)	35.237.35.98	SSH ▾	⋮
☐	✅ node2	us-east1-b			10.142.0.4 (nic0)	35.196.52.140	SSH ▾	⋮
☐	✅ node3	us-east1-b			10.142.0.5 (nic0)	34.73.57.176	SSH ▾	⋮

Paste it in our web browser and add 7180.
Something like 35.237.147.54:7180
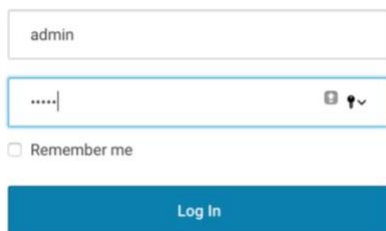Note: Your IP address should be different than mine.

🌐 35.237.147.54:7180

You should see the login page.
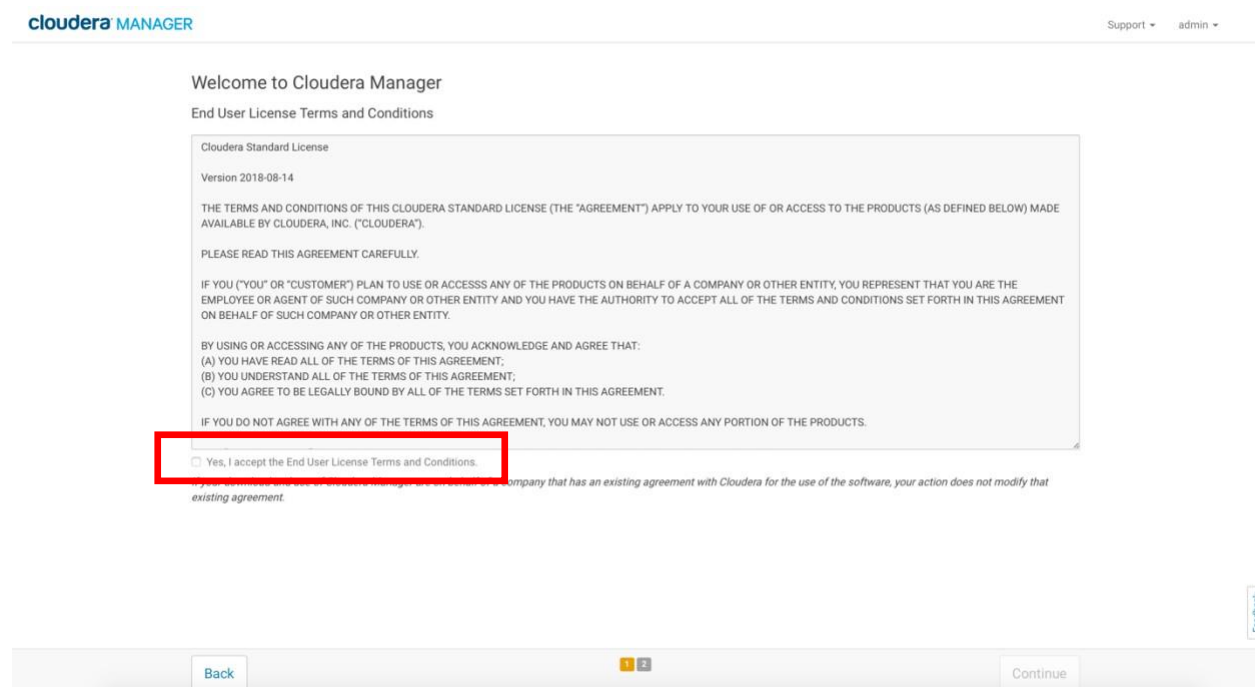
Username: admin
Password: admin

Accept the License and click on Continue.

Choose Cloudera Express and click on Continue.

Click on Continue.

Here, type in the names of your 4 instances and click on search.

cloudera MANAGER

Specify hosts for your CDH cluster installation.

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with.

Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

~~Hint: Search for hostnames and IP addresses using~~ patterns 🗖.

clouder-multi-node-cluster
node1
node2
node3

SSH Port: 22   Search

Specify hosts for your CDH cluster installation.

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with.

Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

Hint: Search for hostnames and IP addresses using patterns 🗖.

4 hosts scanned, 4 running SSH.   New Search

	Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
☑	clouder-multi-node-cluster	clouder-multi-node-cluster.c.cloudera-269005.internal	10.142.0.2	No	✔ Host ready: 0 ms response time.
☑	node1	node1.c.cloudera-269005.internal	10.142.0.3	No	✔ Host ready: 1 ms response time.
☑	node2	node2.c.cloudera-269005.internal	10.142.0.4	No	✔ Host ready: 1 ms response time.
☑	node3	node3.c.cloudera-269005.internal	10.142.0.5	No	✔ Host ready: 1 ms response time.

Click on Continue.
Keep the default settings.

Cluster Installation

Select Repository

Cloudera recommends the use of parcels for installation over packages, because parcels enable Cloudera Manager to easily manage the software on your cluster, automating the deployment and upgrade of service binaries. Electing not to use parcels will require you to manually upgrade packages on all hosts in your cluster when software updates are available, and will prevent you from using Cloudera Manager's rolling upgrade capabilities.

Choose Method   ○ Use Packages ❓
                   ⦿ Use Parcels (Recommended) ❓   More Options   Proxy Settings

CDH Version   ⦿ CDH-5.16.2-1.cdh5.16.2.p0.8
                   ○ CDH-4.7.1-1.cdh4.7.1.p0.47
                   Versions of CDH that are too new for this version of Cloudera Manager (5.16.2) will not be shown.

Additional Parcels ○ ACCUMULO-1.7.2-5.5.0.ACCUMULO5.5.0.p0.8
                   ○ ACCUMULO-1.4.4-1.cdh4.5.0.p0.65
                   ⦿ None

                   ○ IMPALA-2.1.0-1.impala2.0.0.p0.1995
                   ⦿ None

                   ○ KAFKA-4.1.0-1.4.1.0.p0.4
                   ⦿ None

                   ○ KUDU-1.4.0-1.cdh5.12.2.p0.8
                   ⦿ None

                   ○ SOLR-1.3.0-1.cdh4.5.0.p0.9
                   ⦿ None

Back   1 2 3 4 5 6 7   Continue

Cluster Installation

Accept JDK License

Oracle Binary Code License Agreement for the Java SE Platform Products and JavaFX

ORACLE AMERICA, INC. ("ORACLE"), FOR AND ON BEHALF OF ITSELF AND ITS SUBSIDIARIES AND AFFILIATES UNDER COMMON CONTROL, IS WILLING TO LICENSE THE SOFTWARE TO YOU ONLY UPON THE CONDITION THAT YOU ACCEPT ALL OF THE TERMS CONTAINED IN THIS BINARY CODE LICENSE AGREEMENT AND SUPPLEMENTAL LICENSE TERMS (COLLECTIVELY "AGREEMENT"). PLEASE READ THE AGREEMENT CAREFULLY. BY SELECTING THE "ACCEPT LICENSE AGREEMENT" (OR THE EQUIVALENT) BUTTON AND/OR BY USING THE SOFTWARE YOU ACKNOWLEDGE THAT YOU HAVE READ THE TERMS AND AGREE TO THEM. IF YOU ARE AGREEING TO THESE TERMS ON BEHALF OF A COMPANY OR OTHER LEGAL ENTITY, YOU REPRESENT THAT YOU HAVE THE LEGAL AUTHORITY TO BIND THE LEGAL ENTITY TO THESE TERMS. IF YOU DO NOT HAVE SUCH AUTHORITY, OR IF YOU DO NOT WISH TO BE BOUND BY THE TERMS, THEN SELECT THE "DECLINE LICENSE AGREEMENT" (OR THE EQUIVALENT) BUTTON AND YOU MUST NOT USE THE SOFTWARE ON THIS SITE OR ANY OTHER MEDIA ON WHICH THE SOFTWARE IS CONTAINED.

1. DEFINITIONS. "Software" means the software identified above in binary form that you selected for download, install or use (in the version You selected for download, install or use) from Oracle or its authorized licensees, any other machine readable materials (including, but not limited to, libraries, source files, header files, and data files), any updates or error corrections provided by Oracle, and any user manuals, programming guides and other documentation provided to you by Oracle under this Agreement. "General Purpose Desktop Computers and Servers" means computers, including desktop and laptop computers, or servers, used for general computing functions under end user control (such as but not specifically limited to email, general purpose Internet browsing, and office suite productivity tools). The use of Software in systems and solutions that provide dedicated functionality (other than as mentioned above) or designed for use in embedded or function-specific software applications, for example but not limited to: Software embedded in or bundled with industrial control systems, wireless mobile telephones, wireless handheld devices, kiosks, TV/STB, Blu-ray Disc devices, telematics and network control switching equipment, printers and storage management systems, and other related systems are excluded from this definition and not licensed under this Agreement. "Programs" means (a) Java technology applets and applications intended to run on the Java Platform, Standard Edition platform on Java-enabled General Purpose Desktop Computers and Servers; and (b) JavaFX technology applications intended to run on the JavaFX Runtime on JavaFX-enabled General Purpose Desktop Computers and Servers. "Commercial Features" means those features identified in

☑ Install Oracle Java SE Development Kit (JDK 7)
Check this box to accept the Oracle Binary Code License Agreement and install the JDK. Leave it unchecked to use a currently installed JDK.

WARNING: This Cloudera offering includes Oracle's Unlimited Strength Java(TM) Cryptography Extension (JCE) Policy Files for the Java(TM) Platform, Standard Edition (Java SE) Runtime Environment. Due to import restrictions of some countries, the version of the JCE Policy Files that are bundled in the Java Runtime Environment, or JRE(TM), allow "strong" but limited cryptography to be used. The Unlimited Strength JCE Policy Files included in this Cloudera offering, however, provides "unlimited strength" policy files which contain no restrictions on cryptographic strengths. Please note that some countries may legally prohibit the import of unlimited encryption strength policy files. You are responsible for determining whether you are subject to legal restrictions on cryptographic strength, and if so, you should not download this file. If legally permitted, you may configure the transport-layer security to reduce the cryptographic strength available with the default policy files. Additional information may be found at www.oracle.com. You are advised to consult your export/import control counsel or attorney to determine the exact requirements.

☐ Install Java Unlimited Strength Encryption Policy Files

Back   1 2 3 4 5 6 7   Continue

Continue without doing anything.

Cluster Installation

Single User Mode

Only supported for CDH 5.2 and above.

By default, service processes run as distinct users on the system. For example, HDFS DataNodes run as user "hdfs" and HBase RegionServers run as user "hbase." Enabling "single user mode" configures Cloudera Manager to run service processes as a single user, by default "cloudera-scm", thereby prioritizing isolation between managed services and the rest of the system over isolation between the managed services.

The **major benefit** of this option is that the Agent does not run as root. However, this mode complicates installation, which is described fully in the documentation. Most notably, directories which in the regular mode are created automatically by the Agent, must be created manually on every host with appropriate permissions, and sudo (or equivalent) access must be set up for the configured user.

Switching back and forth between single user mode and regular mode is not supported.

Enable Single User Mode   ☐   ⑦

Back   1 2 **3** 4 5 6 7   **Continue**

Remember the password we set before. Put it here.

Cluster Installation

Enter Login Credentials

Root access to your hosts is required to install the Cloudera packages. This installer will connect to your hosts via SSH and log in either directly as root or as another user with password-less sudo/pbrun privileges to become root.

Login To All Hosts As: ● root
                          ○ Another user

You may connect via password or public-key authentication for the user selected above.

Authentication Method: ● All hosts accept same password
                          ○ All hosts accept same private key

Enter Password: [••••••••]

Confirm Password: [•••••••• 🔑⌄]

SSH Port: [22]

Number of Simultaneous   [10]
Installations: (Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)

Wait for a few minutes…

Click on Continue and wait for another few minutes…

Let it check all hosts…

And click on Finish once it finishes.

Note: you may see one or two warnings highlighted in yellow box. It is okay, so just ignore and click on finish.

Keep default settings and click on Continue.
Note: you default setting may be different than mine. You do not have to change yours, just keep the default settings and continue with the installation.

Click on Test Connection, see everything is okay. Click on Continue.

Cluster Setup

Setup Database

Configure and test database connections. If using custom databases, create the databases first according to the **Installing and Configuring an External Database** section of the Installation Guide ☑.

◯ Use Custom Databases ◉ Use Embedded Database

When using the embedded database, passwords are automatically generated. Please copy them down.

⚠ Embedded Database should not be used for production environments.

Hive

Database Host Name:	Database Type:	Database Name :	Username:	Password:
clouder-multi-node-cluster.c.cloudera-269005.internal:7432	PostgreSQL ⇕	hive	hive	931QQLiOk0

Hue

Database Host Name:	Database Type:	Database Name :	Username:	Password:
clouder-multi-node-cluster.c.cloudera-269005.internal:7432	PostgreSQL ⇕	hue	hue	RNo8FPMfEW

Oozie Server

Currently assigned to run on **clouder-multi-node-cluster.c.cloudera-269005.internal**.

Database Host Name:	Database Type:	Database Name :	Username:	Password:
clouder-multi-node-cluster.c.cloudera-269005.internal:7432	PostgreSQL ⇕	oozie_oozie_server	oozie_oozie_server	gZ3P9MCukF

Test Connection

Cluster Setup

Review Changes

HDFS Root Directory hbase.rootdir	Cluster 1 > HBase (Service-Wide)	⑦
	/hbase	
Enable Indexing	☑ Cluster 1 > HBase (Service-Wide) ↰	⑦
Enable Replication hbase.replication	☑ Cluster 1 > HBase (Service-Wide) ↰	⑦
HDFS Block Size dfs.block.size, dfs.blocksize	Cluster 1 > HDFS (Service-Wide)	⑦
	128 MiB ⇕	
DataNode Failed Volumes Tolerated dfs.datanode.failed.volumes.tole rated	Cluster 1 > DataNode Default Group	⑦
	0	
DataNode Data Directory dfs.data.dir, dfs.datanode.data.dir	Cluster 1 > DataNode Default Group ↰	⑦
	/dfs/dn	⊖⊕
NameNode Data Directories dfs.name.dir, dfs.namenode.name.dir	Cluster 1 > NameNode Default Group ↰	⑦
	/dfs/nn	⊖⊕
HDFS Checkpoint Directories	Cluster 1 > SecondaryNameNode Default Group ↰	⑦
	/dfs/snn	⊖⊕

Back   1 2 3 4 5 6   **Continue**

Again, wait for it…to finish executing. Just bear with me.

Cluster Setup

First Run Command

Status ✓ Finished   📅 Feb 22, 7:52:41 AM   ⊘ 6.7m

Finished First Run of the following services successfully: ZooKeeper, HDFS, HBase, Solr, YARN (MR2 Included), Key-Value Store Indexer, Spark, Hive, Impala, Oozie, Hue.

ⱽ **Completed 10 of 10 step(s).**

🔴 Show All Steps   ○ Show Only Failed Steps   ○ Show Running Steps

> ✓	Ensuring that the expected software releases are installed on hosts.		Feb 22, 7:52:41 AM	26ms
> ✓	Deploying Client Configuration	↗ Cluster 1 ↗	Feb 22, 7:52:41 AM	16.82s
> ✓	Start ZooKeeper		Feb 22, 7:52:58 AM	23.69s
> ✓	Start HDFS		Feb 22, 7:53:22 AM	52.31s
> ✓	Start HBase, Solr		Feb 22, 7:54:14 AM	52.31s
> ✓	Start Key-Value Store Indexer, YARN (MR2 Included)		Feb 22, 7:55:07 AM	37.23s
> ✓	Start Spark		Feb 22, 7:55:44 AM	43.04s
> ✓	Start Hive		Feb 22, 7:56:27 AM	64.21s
> ✓	Start Impala, Oozie		Feb 22, 7:57:31 AM	82.28s
> ✓	Start Hue		Feb 22, 7:58:53 AM	29.18s

Displaying 1 - 10 of 10   25 ⬍ per page

Back   1 2 3 4 5 6   Continue

You made it!! Click on Finish and it will bring you to the web UI!

Cluster Setup

Congratulations!

✓ The services are installed, configured, and running on your cluster.

See the warning messages. Don't worry, all you need to do is to restart Cloudera Manager and Cluster 1.

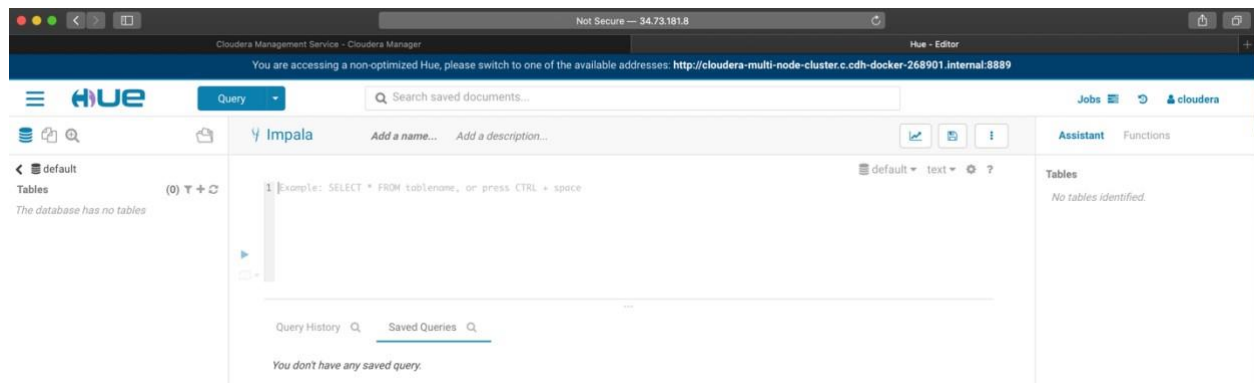Go to port 8888. Set both username and password to cloudera.

Note: If **you failed to open Hue web UI** using your cloudera-multi-node-cluster's external IP, try use other three nodes' external IPs followed by 8888. Your Hue may be installed on other nodes. Another way to check where Hue was installed, click on Hue in Cloudera Manager and Go to Instances tab, you will be able to see the node in which your Hue is installed on.

Congratulations! You have successfully created instances and installed Cloudera services on them.

Attention: Please do not forget to stop your instances every time!!!!! Otherwise GCP will keep charging you.

Frequently Asked Questions

1. Authentication error when installing packages on nodes

If you are experiencing authentication error, you probably did not configure your sshs_config files correctly. Please go back to the config files and check if you add or deleted "#" signs correctly. Do not forget to reboot your instances after you're done with configuring the files and then click on the retry button on the installer UI.

2. Health problem in Cloudera Manager

All you need to do it restart or start the services that has warning signs. Go to Clusters, click on Cloudera Manager/Cluster 1, and click on "Actions" tab. Click on start/restart.

3. Failed to open Hue

If **you failed to open Hue web UI** using your cloudera-multi-node-cluster's external IP, try use other three nodes' external IPs followed by 8888. Your Hue may be installed on other nodes. Another way to check where Hue was installed, click on Hue in Cloudera Manager and Go to Instances tab, you will be able to see the node in which your Hue is installed on.