

Install CDP Cluster with Cloudera Manager using AMI on AWS EC2:

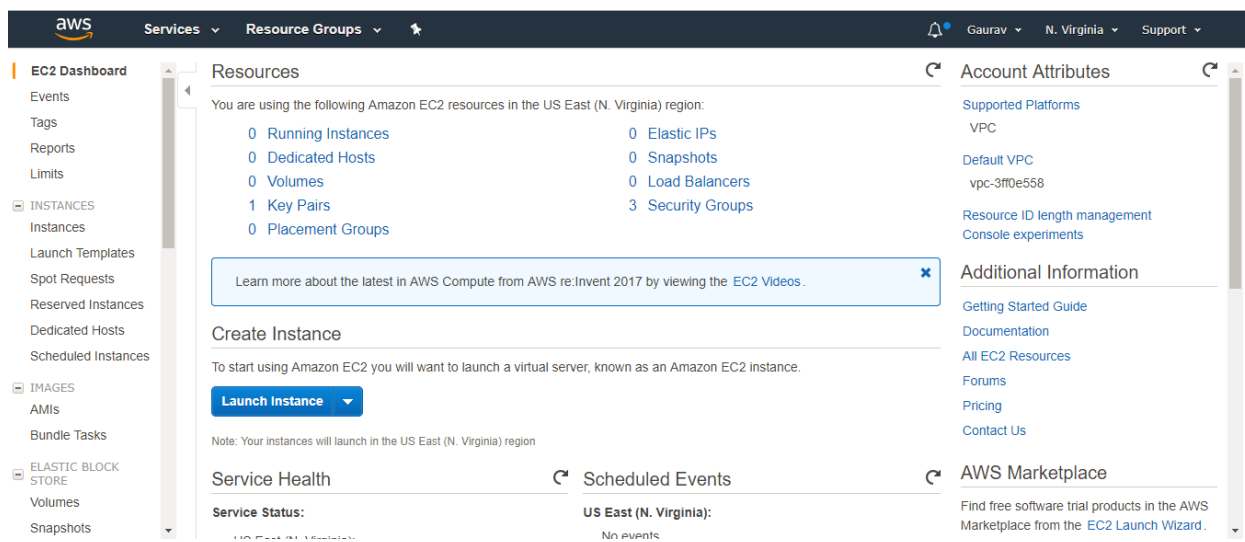
Prerequisites:

Region- **N.Virginia** (You need to select N.Virginia as the region as this is the cheapest region & our AMI is available on N.Virginia)

AMI Name :-**CDP-Cloudera-upGrad**

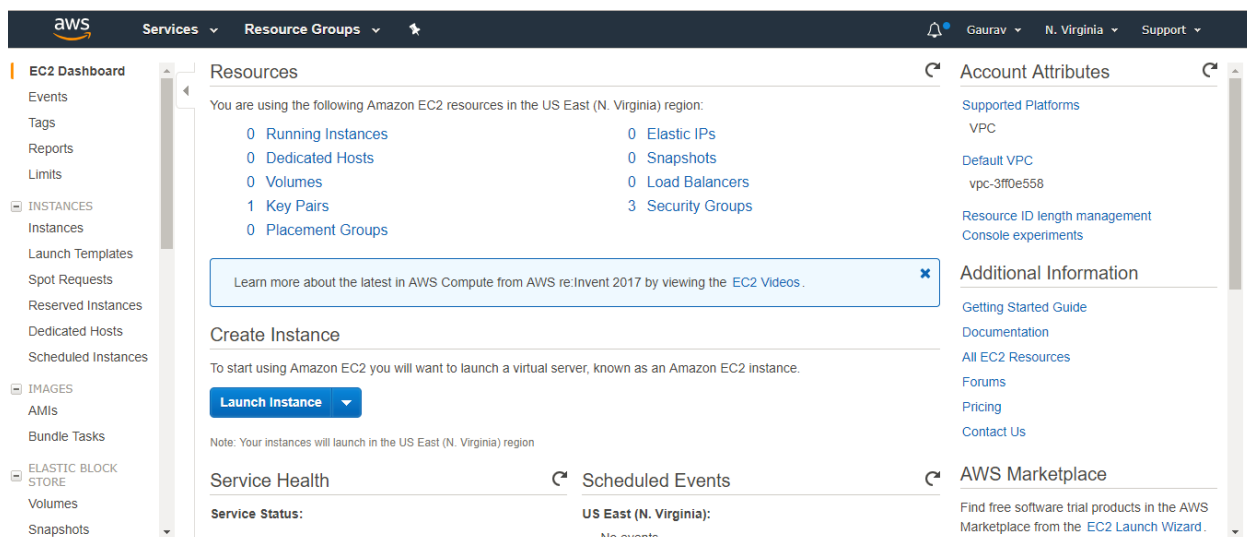
AMI ID:-ami-0ac405712832c86d0

1. Log in to the AWS console home and click on '**EC2**' under the Compute Section inside the **N.Virginia Region**.



The screenshot displays the AWS Management Console interface for the N. Virginia region. The left sidebar shows the navigation menu with categories like EC2 Dashboard, INSTANCES, IMAGES, and ELASTIC BLOCK STORE. The main content area is titled 'Resources' and lists various EC2 resources: 0 Running Instances, 0 Elastic IPs, 0 Dedicated Hosts, 0 Snapshots, 0 Volumes, 0 Load Balancers, 1 Key Pairs, and 3 Security Groups. Below this list is a 'Create Instance' section with a 'Launch Instance' button. The right sidebar contains 'Account Attributes' and 'Additional Information' sections. The bottom of the console shows 'Service Health' and 'Scheduled Events' for the US East (N. Virginia) region.

2. Click on 'Launch Instance'

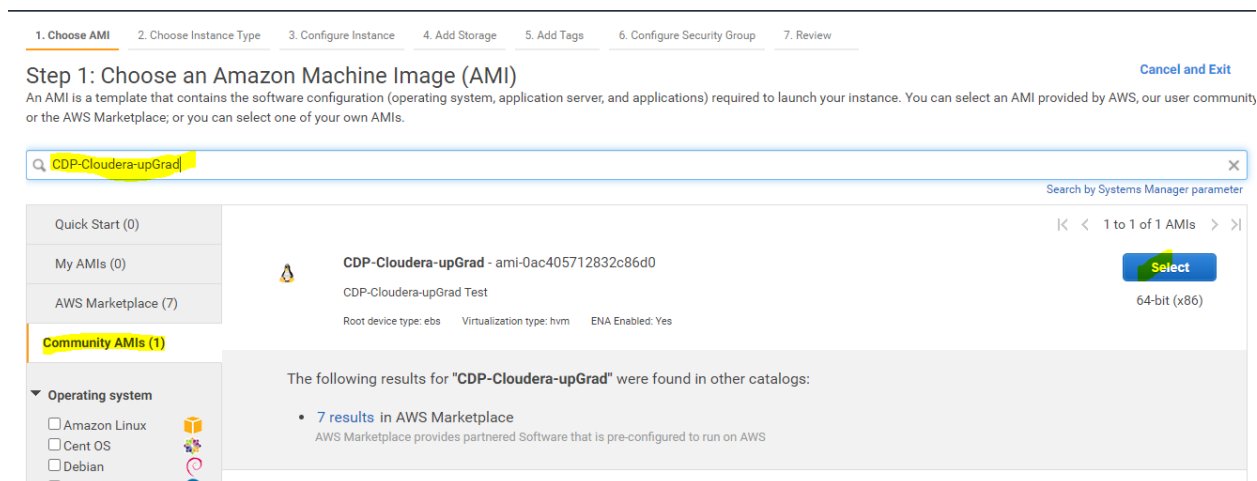


The screenshot shows the AWS Management Console interface. On the left, the 'EC2 Dashboard' sidebar is visible with various navigation options. The main content area is titled 'Resources' and lists various Amazon EC2 resources in the US East (N. Virginia) region. A 'Create Instance' section is highlighted, featuring a 'Launch Instance' button. Below this, there's a 'Service Health' section and a 'Scheduled Events' section. On the right, the 'Account Attributes' panel shows supported platforms, default VPC, and additional information links.

3. Click on 'Community AMI' and search our AMI with the name below.

Inside the search, tab enter our AMI Name:: **CDP-Cloudera-upGrad** and Select our AMI.

AMI Name: CDP-Cloudera-upGrad



The screenshot shows the 'Step 1: Choose an Amazon Machine Image (AMI)' screen in the AWS console. It includes a progress bar at the top with steps from 'Choose AMI' to 'Review'. The main content area shows a search bar with 'CDP-Cloudera-upGrad' entered. Below the search bar, there's a list of AMIs. The 'Community AMIs (1)' section is highlighted, showing a single result: 'CDP-Cloudera-upGrad - ami-0ac405712832c86d0'. This AMI is described as 'CDP-Cloudera-upGrad Test' with a root device type of 'ebs', virtualization type of 'hvm', and ENA Enabled: Yes. A 'Select' button is visible next to the AMI. The bottom left shows a filter for 'Operating system' with options like Amazon Linux, Cent OS, Debian, and Ubuntu.

4. Select instance type 'm4.xlarge' (please select very carefully)

5.click on 'Next Configure Instance details'.

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 2: Choose an Instance Type

| | | | | | | | | |
|-------------------------------------|-----------------|-------------|----|-----|----------|-----|------------------|-----|
| <input type="checkbox"/> | General purpose | m5.xlarge | 16 | 64 | EBS only | Yes | Up to 10 Gigabit | Yes |
| <input type="checkbox"/> | General purpose | m5.12xlarge | 48 | 192 | EBS only | Yes | 10 Gigabit | Yes |
| <input type="checkbox"/> | General purpose | m5.24xlarge | 96 | 384 | EBS only | Yes | 25 Gigabit | Yes |
| <input type="checkbox"/> | General purpose | m4.large | 2 | 8 | EBS only | Yes | Moderate | Yes |
| <input checked="" type="checkbox"/> | General purpose | m4.xlarge | 4 | 16 | EBS only | Yes | High | Yes |
| <input type="checkbox"/> | General purpose | m4.2xlarge | 8 | 32 | EBS only | Yes | High | Yes |
| <input type="checkbox"/> | General purpose | m4.4xlarge | 16 | 64 | EBS only | Yes | High | Yes |
| <input type="checkbox"/> | General purpose | m4.10xlarge | 40 | 160 | EBS only | Yes | 10 Gigabit | Yes |
| <input type="checkbox"/> | General purpose | m4.16xlarge | 64 | 256 | EBS only | Yes | 25 Gigabit | Yes |

Cancel Previous Review and Launch Next: Configure Instance Details

6. We recommend you to follow the same steps like for CDH 5.15.1

Number of instances- 1

VPC - your vpc name

Step 3: Configure Instance Details

Configure the instance to suit your requirements. You can launch multiple instances from the same AMI, request Spot instances to take advantage of the lower pricing, assign an access mana instance, and more.

Number of instances ⓘ Launch into Auto Scaling Group ⓘ

Purchasing option ⓘ ☐ Request Spot instances

Network ⓘ Create new VPC

Subnet ⓘ Create new subnet
249 IP Addresses available

Auto-assign Public IP ⓘ

Placement group ⓘ ☐ Add instance to placement group

and click on the 'Add Storage' button (bottom right).

7. We recommend you to add 70GB of storage, and volume type should be 'Magnetic' for lower prices. (as pricing is based on the VM usage and configuration). Finally click on the **'Next: Add Tags'** button.

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review

Step 4: Add Storage

Your instance will be launched with the following storage device settings. You can attach additional EBS volumes and instance store volumes to your instance, or edit the settings of the root volume. You can also attach additional EBS volumes after launching an instance, but not instance store volumes. [Learn more](#) about storage options in Amazon EC2.

| Volume Type ⓘ | Device ⓘ | Snapshot ⓘ | Size (GiB) ⓘ | Volume Type ⓘ | IOPS ⓘ | Throughput (MB/s) ⓘ | Delete on Termination ⓘ | Encrypted ⓘ |
|---------------|-----------|------------------------|---------------------------------|---------------|--------|---------------------|-------------------------------------|---------------|
| Root | /dev/sda1 | snap-00640e26f802cc4aa | <input type="text" value="50"/> | Magnetic | N/A | N/A | <input checked="" type="checkbox"/> | Not Encrypted |

[Add New Volume](#)

General Purpose (SSD) volumes provide the ability to burst to 3000 IOPS per volume, independent of volume size, to meet the performance needs of most applications and also deliver a consistent baseline of 3 IOPS/GiB. [Set my root volume to General Purpose \(SSD\)](#).

Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage. [Learn more](#) about free usage tier eligibility and usage restrictions.

[Cancel](#)
[Previous](#)
[Review and Launch](#)
[Next: Add Tags](#)

8. Now click on **'click to add a Name tag'** as shown in the image below.

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. A copy of a tag can be applied to volumes, instances or both. Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

| Key (127 characters maximum) | Value (255 characters maximum) | Instances ⓘ | Volumes ⓘ |
|-------------------------------------|--------------------------------|-------------|-----------|
| This resource currently has no tags | | | |

Choose the [Add tag](#) button or [click to add a Name tag](#).
Make sure your [IAM policy](#) includes permissions to create tags.

[Add Tag](#) (Up to 50 tags maximum)

[Cancel](#)
[Previous](#)
[Review and Launch](#)
[Next: Configure Security Group](#)

9. Give a name in the cell under **‘Value’**. In our case, we used CDP 7.

1. Choose AMI
2. Choose Instance Type
3. Configure Instance
4. Add Storage
5. Add Tags
6. Configure Security Group
7. Review

Step 5: Add Tags

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver.
A copy of a tag can be applied to volumes, instances or both.
Tags will be applied to all instances and volumes. [Learn more](#) about tagging your Amazon EC2 resources.

| Key (128 characters maximum) | Value (256 characters maximum) | Instances ⓘ | Volumes ⓘ |
|------------------------------|--------------------------------|-------------------------------------|-------------------------------------|
| Name | CDP 7 | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

[Add another tag](#) (Up to 50 tags maximum)

10. Click on **‘Next: Configure Security Group’**

11. Select the **‘Select an existing security group’** and provide any name and description as you wish.

Step 6: Configure Security Group

A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance, add rules that allow unrestricted access to the HTTP and HTTPS ports. You can create a new security group or select from an existing one below. [Learn more](#) about Amazon EC2 security groups.

Assign a security group: ☐ Create a new security group

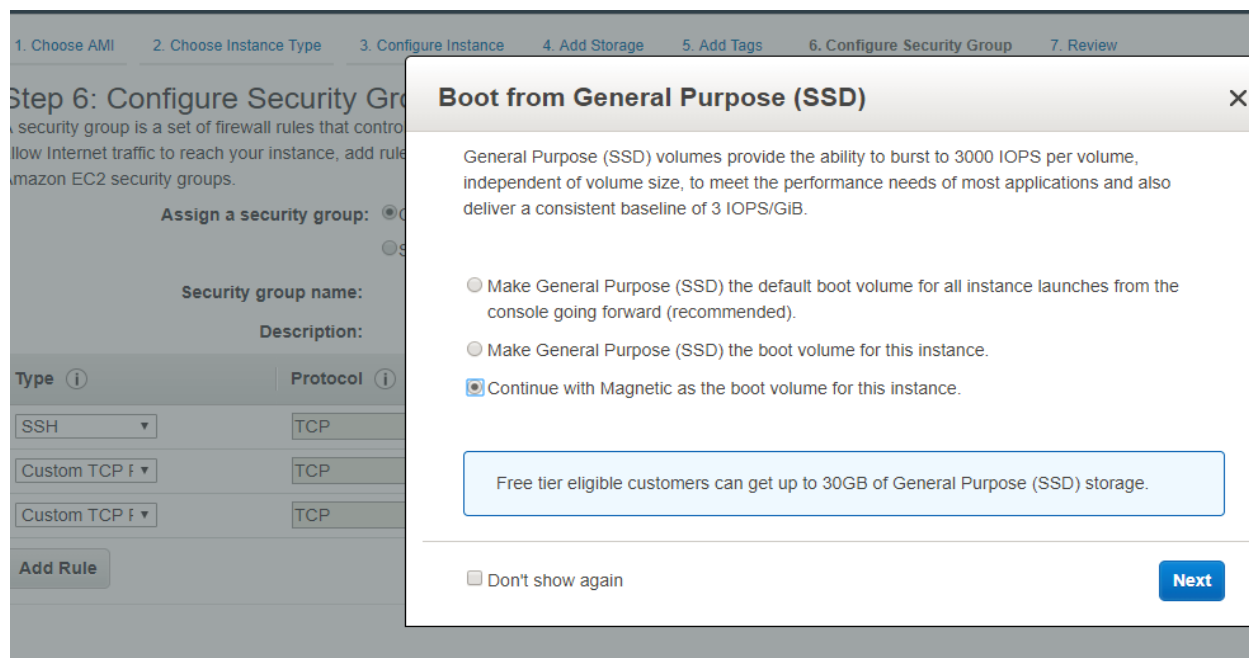
☒ **Select an existing security group**

| Security Group ID | Name | Description | Actions |
|--|--------------------------|-------------------------------|-----------------------------|
| <input checked="" type="checkbox"/> sg-0dd5c69ad227a0d25 | cloudera | set inbound rules | Copy to new |
| <input type="checkbox"/> sg-0d3c542b5fd61f42e | cluster-training-avdhesh | Delete me after the training. | Copy to new |
| <input type="checkbox"/> sg-05c6e96485095db19 | default | default VPC security group | Copy to new |

Inbound rules for sg-0dd5c69ad227a0d25 (Selected security groups: sg-0dd5c69ad227a0d25)

| Type ⓘ | Protocol ⓘ | Port Range ⓘ | Source ⓘ | Description ⓘ |
|---------|------------|--------------|------------------|---------------|
| All TCP | TCP | 0 - 65535 | 10.235.228.91/32 | |

12. Click on **‘Review and Launch’**. The following page will appear as shown in the image below. Click on **‘Continue with Magnetic as the boot volume for this instance’**. Then click **‘Next’**



1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 6: Configure Security Group

A security group is a set of firewall rules that control the inbound and outbound traffic to your Amazon EC2 instances. To allow Internet traffic to reach your instance, add rules to your security group.

Assign a security group: ☐ ☐

Security group name:

Description:

| Type | Protocol |
|--------------|----------|
| SSH | TCP |
| Custom TCP f | TCP |
| Custom TCP f | TCP |

Add Rule

Boot from General Purpose (SSD)

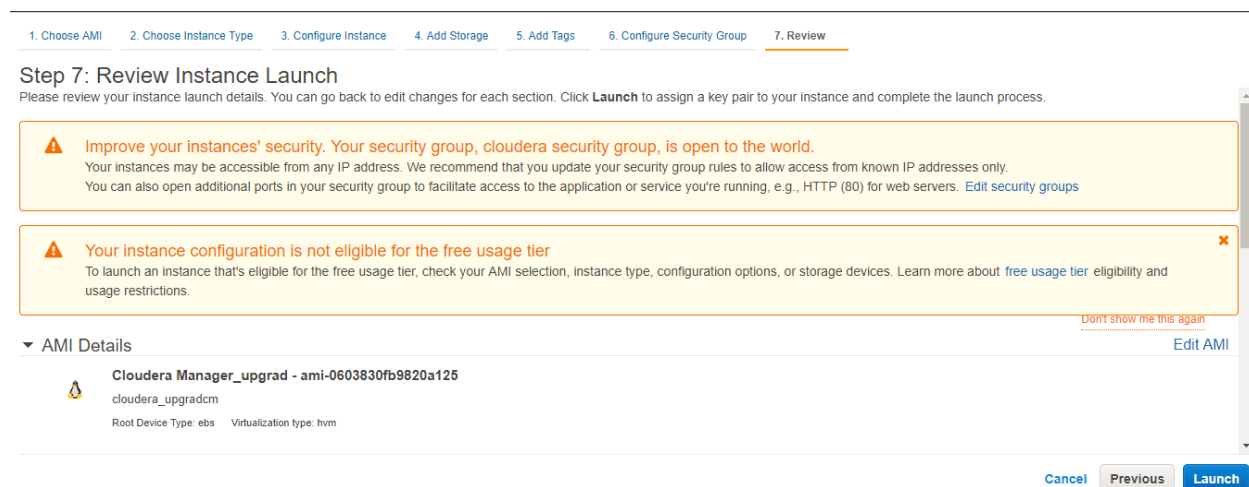
General Purpose (SSD) volumes provide the ability to burst to 3000 IOPS per volume, independent of volume size, to meet the performance needs of most applications and also deliver a consistent baseline of 3 IOPS/GiB.

- ☐ Make General Purpose (SSD) the default boot volume for all instance launches from the console going forward (recommended).
- ☐ Make General Purpose (SSD) the boot volume for this instance.
- ☒ Continue with Magnetic as the boot volume for this instance.

Free tier eligible customers can get up to 30GB of General Purpose (SSD) storage.

☐ Don't show again Next

13. Click on the **‘Launch’** button.



1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Add Tags 6. Configure Security Group 7. Review

Step 7: Review Instance Launch

Please review your instance launch details. You can go back to edit changes for each section. Click **Launch** to assign a key pair to your instance and complete the launch process.

⚠ Improve your instances' security. Your security group, cloudera security group, is open to the world.

Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only. You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

⚠ Your instance configuration is not eligible for the free usage tier

To launch an instance that's eligible for the free usage tier, check your AMI selection, instance type, configuration options, or storage devices. Learn more about [free usage tier](#) eligibility and usage restrictions.

[Don't show me this again](#)

AMI Details

Cloudera Manager_upgrad - ami-0603830fb9820a125

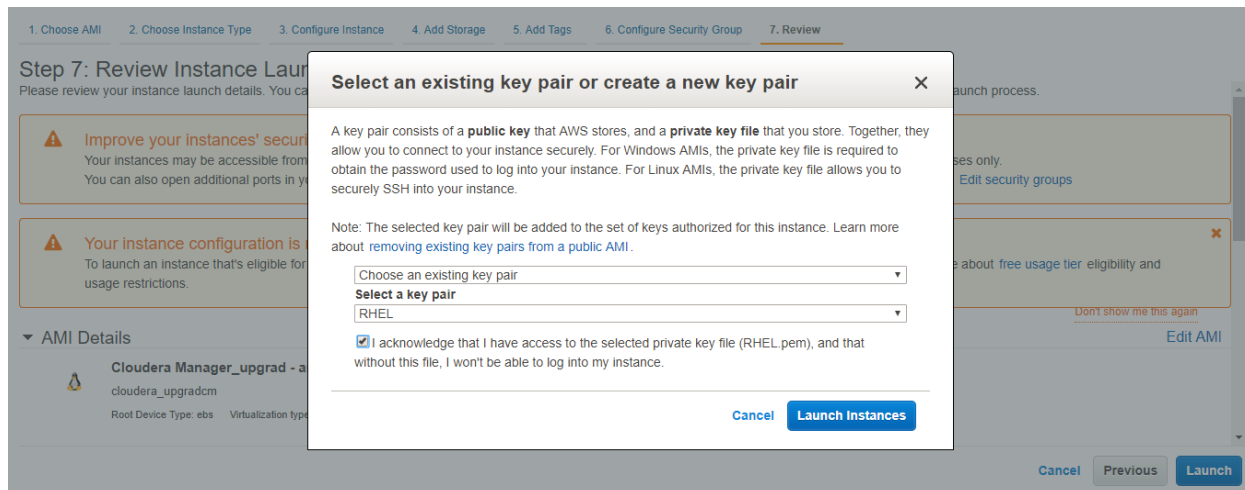
cloudera_upgradcm

Root Device Type: ebs Virtualization type: hvm

[Edit AMI](#)

Cancel Previous Launch

14. Now select ‘**Choose an existing key pair**’ from the drop-down menu and select your key pair name (in our case, RHEL). Tick the 'I acknowledge....' and click on ‘**Launch Instances**’.



15. Click on ‘**View Instances**’ and your instances will appear on the screen.

Launch Status

How to connect to your instances

Your instances are launching, and it may take a few minutes until they are in the **running** state, when they will be ready for you to use. Usage hours on your new instances will start immediately and continue to accrue until you stop or terminate your instances.

Click **View Instances** to monitor your instances' status. Once your instances are in the **running** state, you can **connect** to them from the Instances screen. [Find out](#) how to connect to your instances.

▼ Here are some helpful resources to get you started

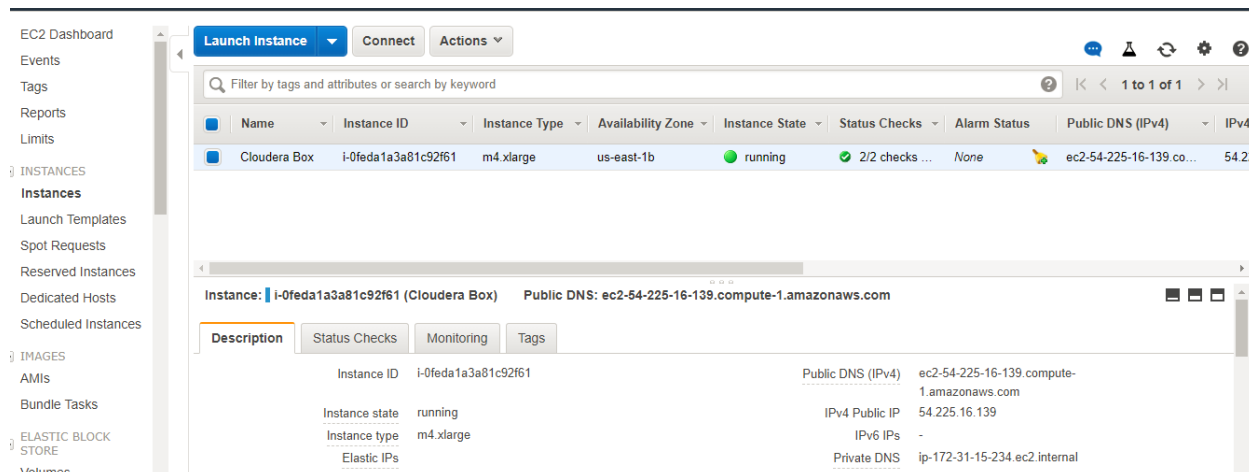
- How to connect to your Linux instance
- Amazon EC2: User Guide
- Learn about AWS Free Usage Tier
- Amazon EC2: Discussion Forum

While your instances are launching you can also

- [Create status check alarms](#) to be notified when these instances fail status checks. (Additional charges may apply)
- [Create and attach additional EBS volumes](#) (Additional charges may apply)
- [Manage security groups](#)

View Instances

16. Wait until the status check is 2/2 (under 'Status Checks' column)..



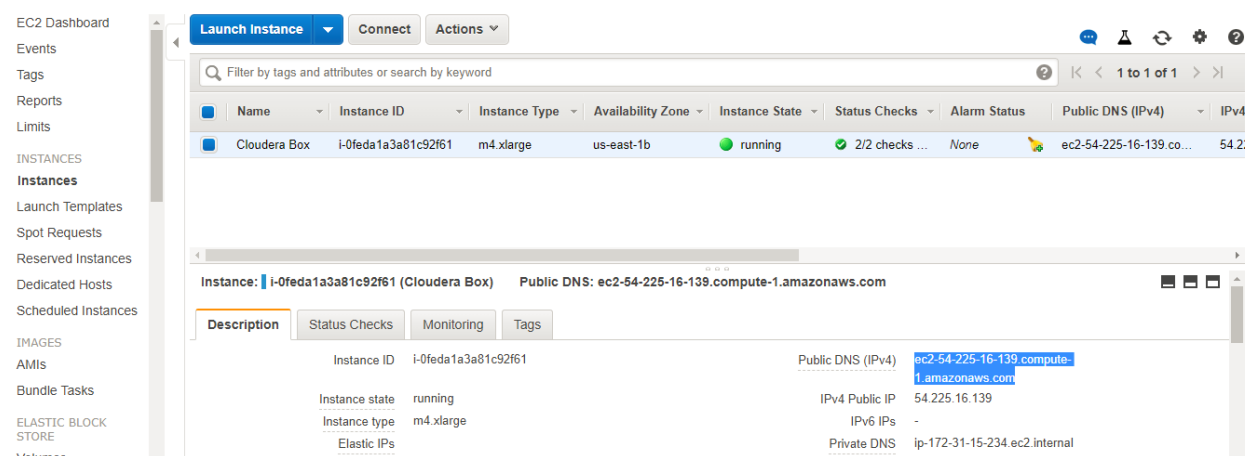
The screenshot shows the AWS Management Console interface. On the left, the navigation pane lists various services like EC2 Dashboard, Events, Tags, Reports, Limits, INSTANCES, Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, IMAGES, AMIs, Bundle Tasks, ELASTIC BLOCK STORE, and Volumes. The main content area displays a table of EC2 instances. The instance 'Cloudera Box' with ID 'i-0feda1a3a81c92f61' is shown in a 'running' state in the 'us-east-1b' availability zone. The 'Status Checks' column indicates '2/2 checks' are passed. Below the table, a detailed view of the instance is shown, including its Public DNS (IPv4) address: 'ec2-54-225-16-139.compute-1.amazonaws.com'.

17. Now Access/ log in the instance using putty for window users and install the Hadoop(CDH).

For window users: need .ppk file (convert your RHEL.pem to name.ppk file using PuTTYgen software).

Note:-if you already have a .ppk file please ignore it.

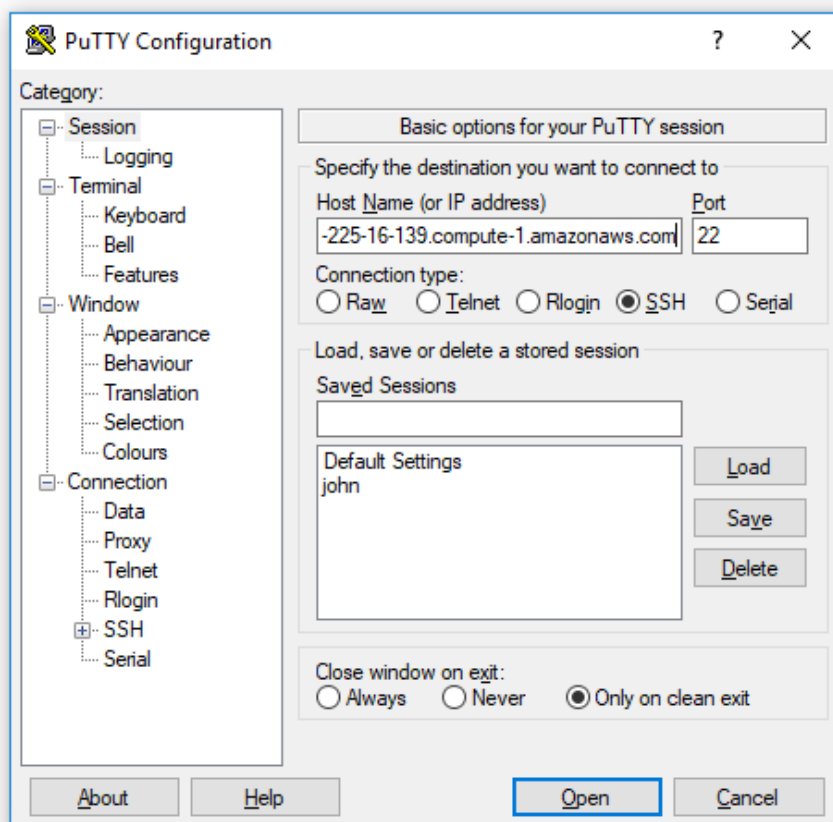
18. Now, open your EC2 dashboard and select your instance. Copy your 'Public DNS (IPv4)' information as shown in the screenshot.



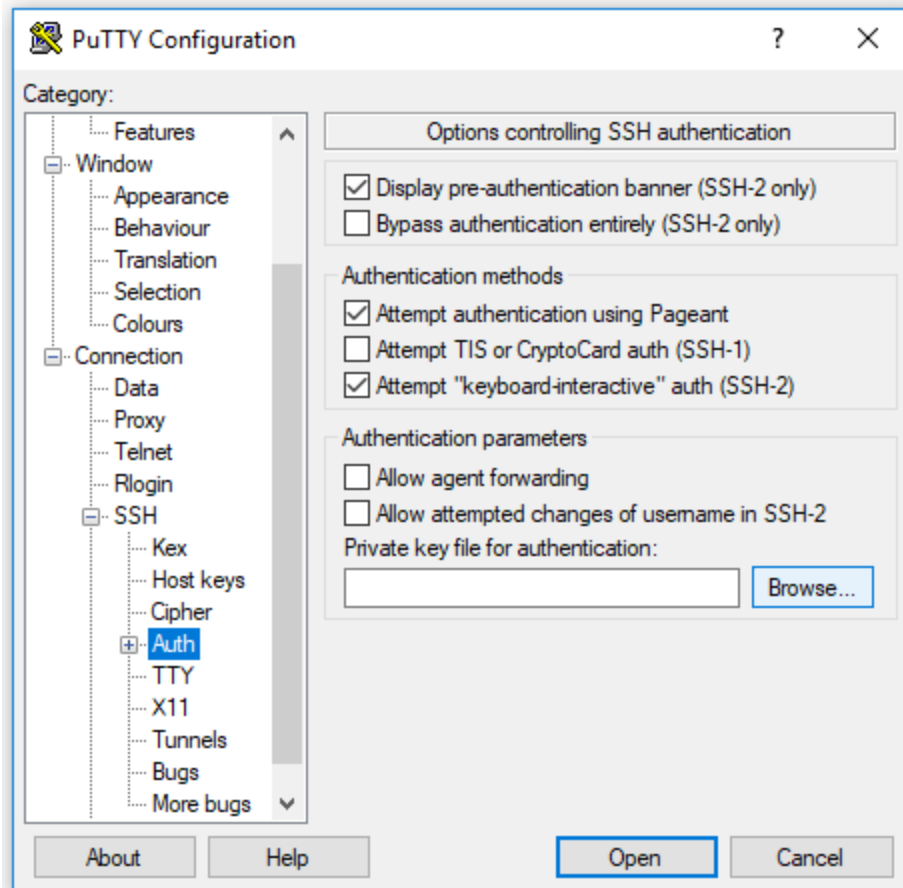
This screenshot is similar to the previous one, showing the AWS Management Console. It highlights the 'Public DNS (IPv4)' information for the 'Cloudera Box' instance. The Public DNS (IPv4) address is 'ec2-54-225-16-139.compute-1.amazonaws.com', which is highlighted in blue in the original image. The instance is still in a 'running' state with '2/2 checks' passed.

19. Open putty:

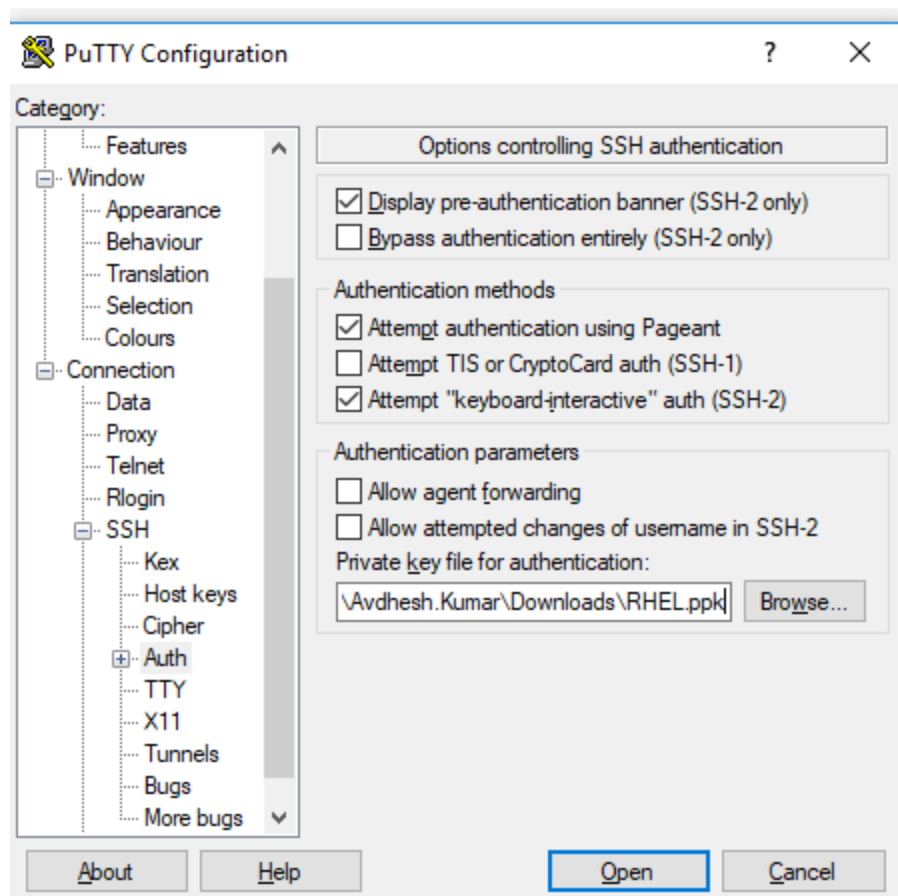
Under the '**Host Name**' section, paste the public DNS information of your instance that you just copied.



20. On the left-hand side panel, click on '**Connection**'. Then click on '**SSH**' followed by '**Auth**'. In the private key field, click on '**Browse**'.



21. Select the .ppk file(RHEL.ppk) you generated using PuTTYgen and click on 'Open'. Then, click on open in the PuTTY software.



22. Click on 'Yes' and login with: **ec2-user**

```
ec2-user@ip-172-31-15-234:~
login as: ec2-user
Authenticating with public key "imported-openssh-key"
Last login: Thu Jun 14 07:24:30 2018 from 114.143.176.26
[ec2-user@ip-172-31-15-234 ~]$
```

23. Switch to root user using the command

‘**sudo -i**’ (enter all commands without quotes unless it is mentioned).

```
root@ip-172-31-15-234:~  
login as: ec2-user  
Authenticating with public key "imported-openssh-key"  
Last login: Thu Jun 14 07:24:30 2018 from 114.143.176.26  
[ec2-user@ip-172-31-15-234 ~]$ sudo -i  
[root@ip-172-31-15-234 ~]#
```

24. Run the command ‘**ls**’ and verify the ‘**cloudera bin**’ is available or not.

```
root@ip-10-0-0-194:~  
[root@ip-10-0-0-194 ~]# ls  
anaconda-ks.cfg          mysql-community-release-el7-5.noarch.rpm  mysql-connector-java-5.1.47.tar.gz  
cloudera-manager-installer.bin  mysql-connector-java-5.1.47              original-ks.cfg  
[root@ip-10-0-0-194 ~]#
```

25. Next, run the bin file using the command given below and press the enter
‘**./cloudera-manager-installer.bin**’

```
[root@ip-172-31-15-234 ~]# ./cloudera-manager-installer.bin
```

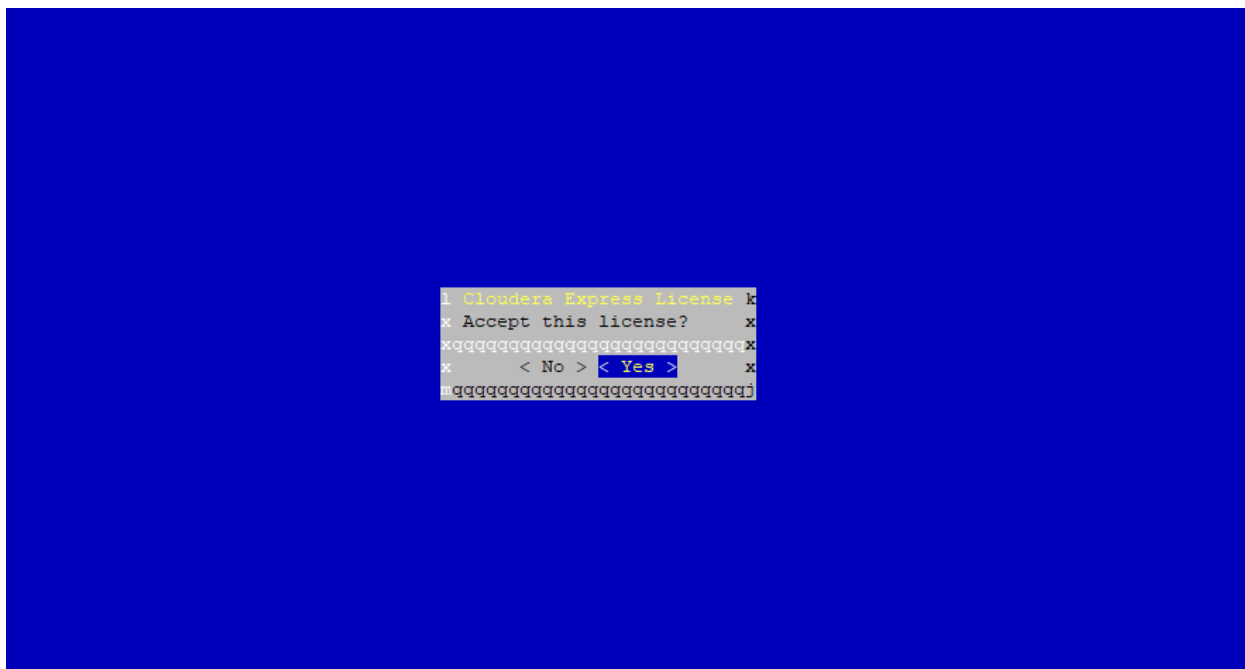
26. Cloudera Manager will start, and the screen will look like the image below. Select 'next' and press the **enter**.

```
Cloudera Manager 5
q
x Cloudera Manager 5
x
x The Cloudera Manager Installer enables you to install Cloudera Manager and
x bootstrap an entire CDH cluster, requiring only that you have SSH access to
x your cluster's machines, and that those machines have Internet access.
x
x This installer is only recommended for demonstration and proof of concept
x deployments, but is not recommended for production deployments because it is
x not intended to scale and may require database migration as your cluster
x grows.
x
x The Cloudera Manager Installer will automatically:
x
x * Detect the operating system on the Cloudera Manager host
x * Install the package repository for Cloudera Manager and the Java Runtime
x Environment (JRE)
x * Install the JRE if it's not already installed
x * Install and configure an embedded PostgreSQL database
x * Install and run the Cloudera Manager Server
x
x Once server installation is complete, you can browse to Cloudera Manager's
x web interface and use the cluster installation wizard to set up your CDH
x cluster.
x
x Cloudera Manager supports the following 64-bit operating systems:
x
x * Red Hat Enterprise Linux 5 (Update 7 or later recommended)
x * Red Hat Enterprise Linux 6 (Update 4 or later recommended)
x * Red Hat Enterprise Linux 7
x * Oracle Enterprise Linux 5 (Update 6 or later recommended)
x * Oracle Enterprise Linux 6 (Update 4 or later recommended)
x * Oracle Enterprise Linux 7
x * CentOS 5 (Update 7 or later recommended)
x * CentOS 6 (Update 4 or later recommended)
x * CentOS 7
x * SUSE Linux Enterprise Server 11 (Service Pack 2 or later recommended)
x
x < Cancel > < Next >
x
```

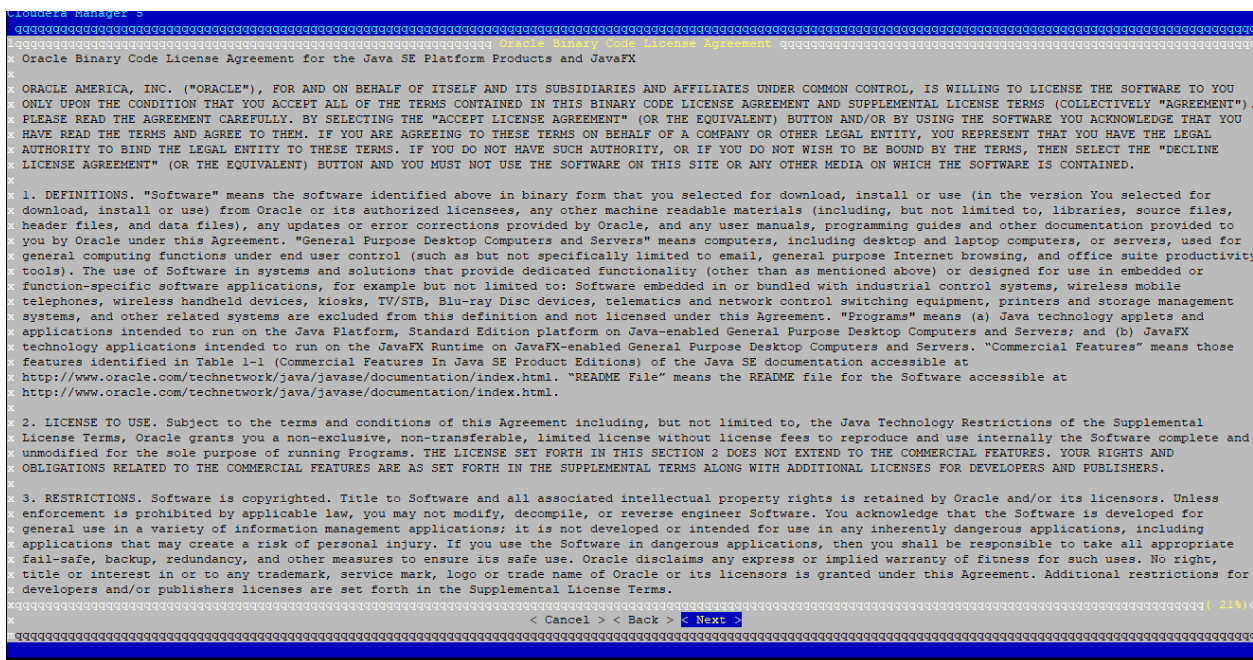
27. Select **next** and press enter :

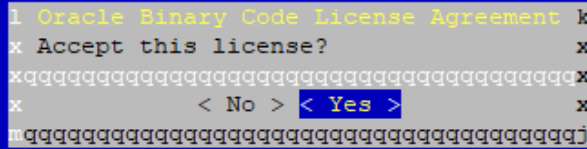
```
Cloudera Manager 5
q
x Cloudera Standard License
x Version 2017-06-29
x
x THE TERMS AND CONDITIONS OF THIS CLOUDERA STANDARD LICENSE (THE "AGREEMENT") APPLY TO YOUR USE OF OR ACCESS TO THE PRODUCTS (AS DEFINED BELOW) MADE AVAILABLE BY
x CLOUDERA, INC. ("CLOUDERA").
x
x PLEASE READ THIS AGREEMENT CAREFULLY.
x
x IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE OR ACCESS ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF
x SUCH COMPANY OR OTHER ENTITY AND YOU HAVE THE AUTHORITY TO ACCEPT ALL OF THE TERMS AND CONDITIONS SET FORTH IN THIS AGREEMENT ON BEHALF OF SUCH COMPANY OR OTHER
x ENTITY.
x
x BY USING OR ACCESSING ANY OF THE PRODUCTS, YOU ACKNOWLEDGE AND AGREE THAT:
x (A) YOU HAVE READ ALL OF THE TERMS OF THIS AGREEMENT;
x (B) YOU UNDERSTAND ALL OF THE TERMS OF THIS AGREEMENT;
x (C) YOU AGREE TO BE LEGALLY BOUND BY ALL OF THE TERMS SET FORTH IN THIS AGREEMENT.
x
x IF YOU DO NOT AGREE WITH ANY OF THE TERMS OF THIS AGREEMENT, YOU MAY NOT USE OR ACCESS ANY PORTION OF THE PRODUCTS.
x
x THE "EFFECTIVE DATE" OF THIS AGREEMENT IS THE DATE YOU FIRST DOWNLOAD OR ACCESS ANY OF THE PRODUCTS.
x
x 1. Product. For the purpose of this Agreement, "Product" shall mean any of Cloudera's offerings including but not limited to: Cloudera Manager, Cloudera
x Enterprise, Cloudera Live, Cloudera Express, Cloudera Director, any hosted or cloud-based service (a "Cloudera Online Service"), any trial software, and any
x software related to the foregoing.
x
x 2. Entire Agreement. This Agreement includes any exhibits attached hereto and web links referenced herein or in any exhibit, and the terms set forth on the
x Cloudera web site at http://www.cloudera.com/documentation/other/Licenses/Third-Party-Licenses/Third-Party-Licenses.html, all hereby incorporated by reference into
x this Agreement in their entirety as they appear on the Effective Date of this Agreement, and as may be updated by Cloudera in its sole discretion from time to time
x without amendment to this Agreement.
x
x This Agreement is the entire agreement of the parties regarding the subject matter hereof, superseding all other agreements between the parties, whether oral or
x written, regarding the subject matter hereof.
x
x 3. License Delivery. Cloudera grants to Customer a nonexclusive, nontransferable, nonassignable, revocable and limited license to access and use the applicable
x Product(s) as defined above in Section 1 solely for Customer's internal purposes. The Product is delivered via electronic download (or, in the case of Cloudera
x Online Services, as a hosted, cloud-based service, accessible to Customer through a web browser), made available following Customer's acceptance of this Agreement.
x
x < Cancel > < Back > < Next >
x
```

28. Finally, to accept the licence - select **yes**

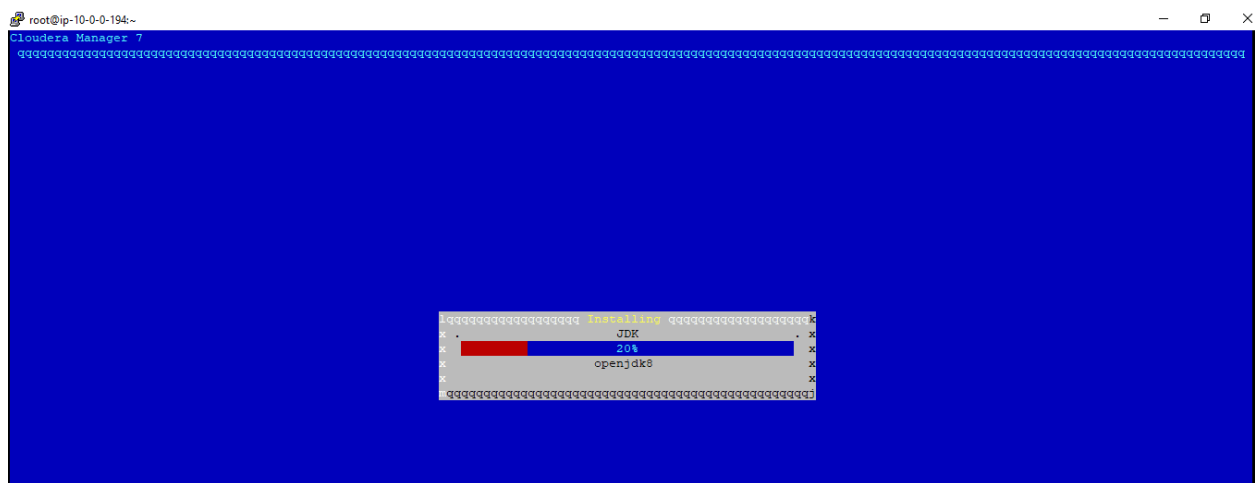


29. Click on **Next** and Accept the Oracle Binary licence

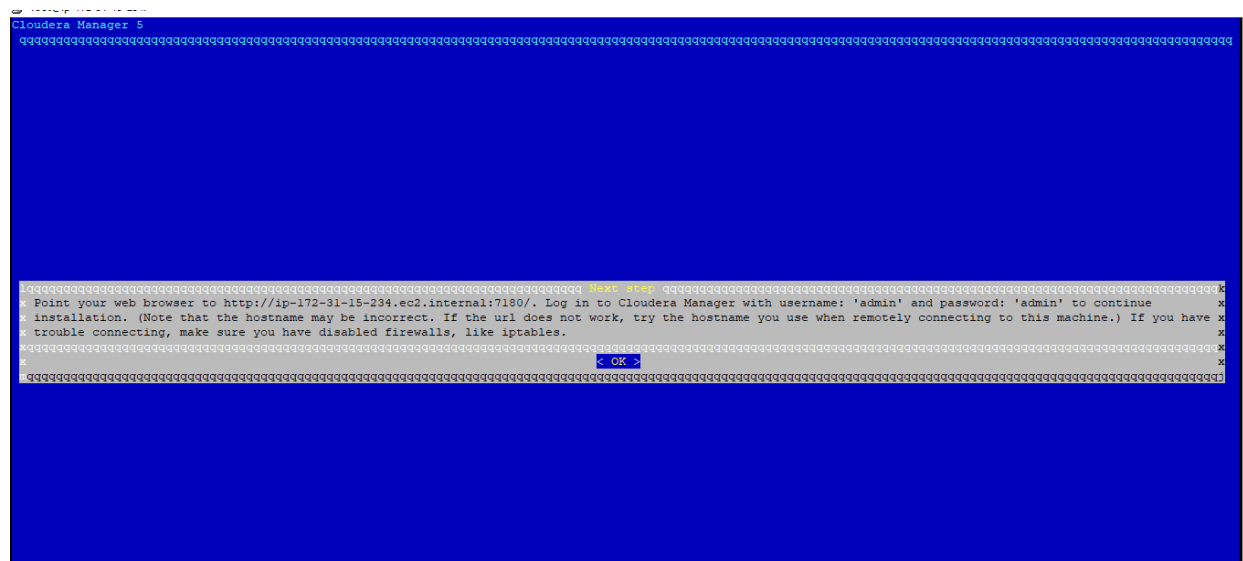




It should take around 10-15 mins

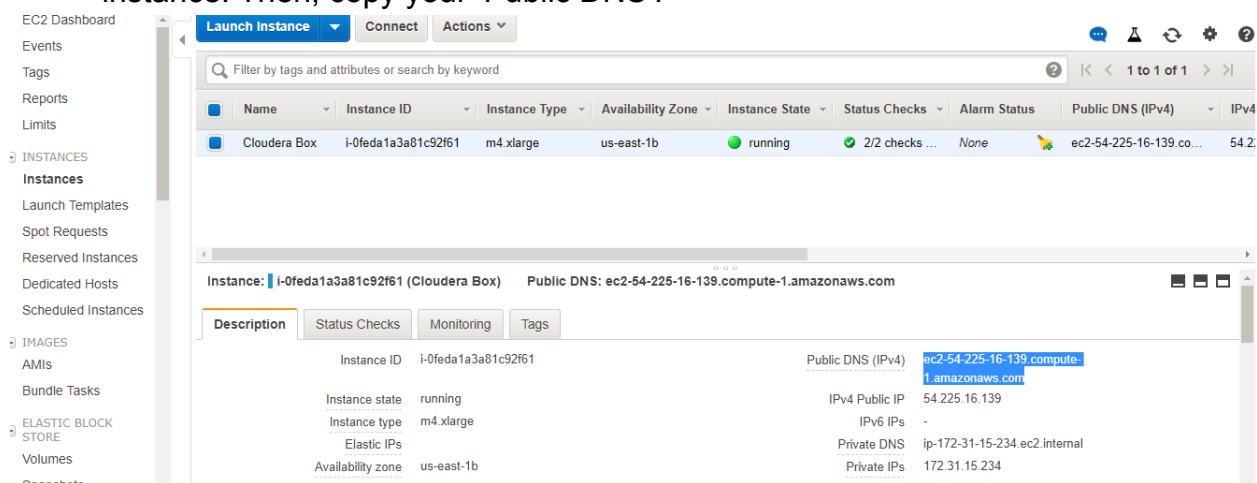


30. Wait until the Cloudera manager is installed and click on **ok**. (as shown in below image)

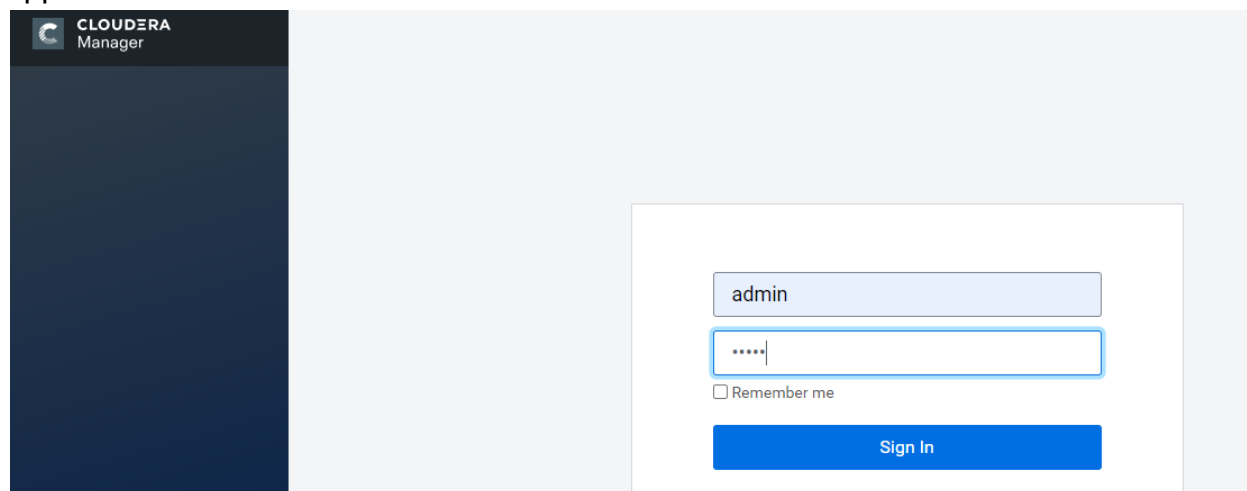


31. Now open a web browser on your local machine; go to the following address **<public-domain-of-ec2>:7180**. This way you will be able to access the Cloudera Manager service that's running on the EC2 instance.

- To get the public domain of ec2, go to the EC2 dashboard, and click on your instance. Then, copy your 'Public DNS'.

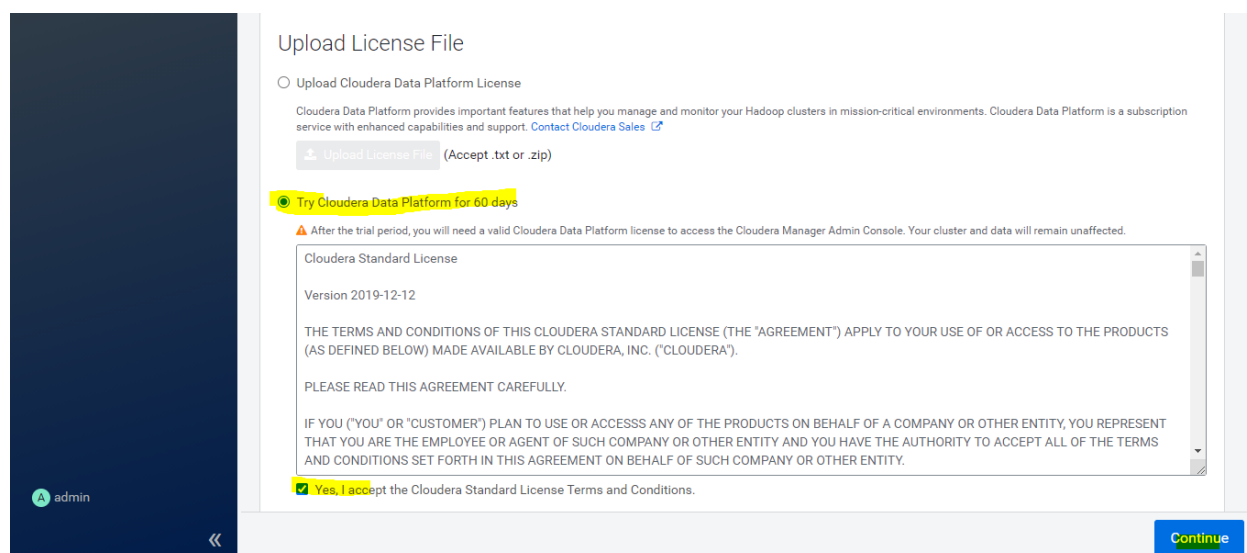


32. Put the copied public IP in place of **<public-domain-of-ec2>** in **<public-domain-of-ec2>:7180**, and open it in your browser. The following page will appear after a few minutes.

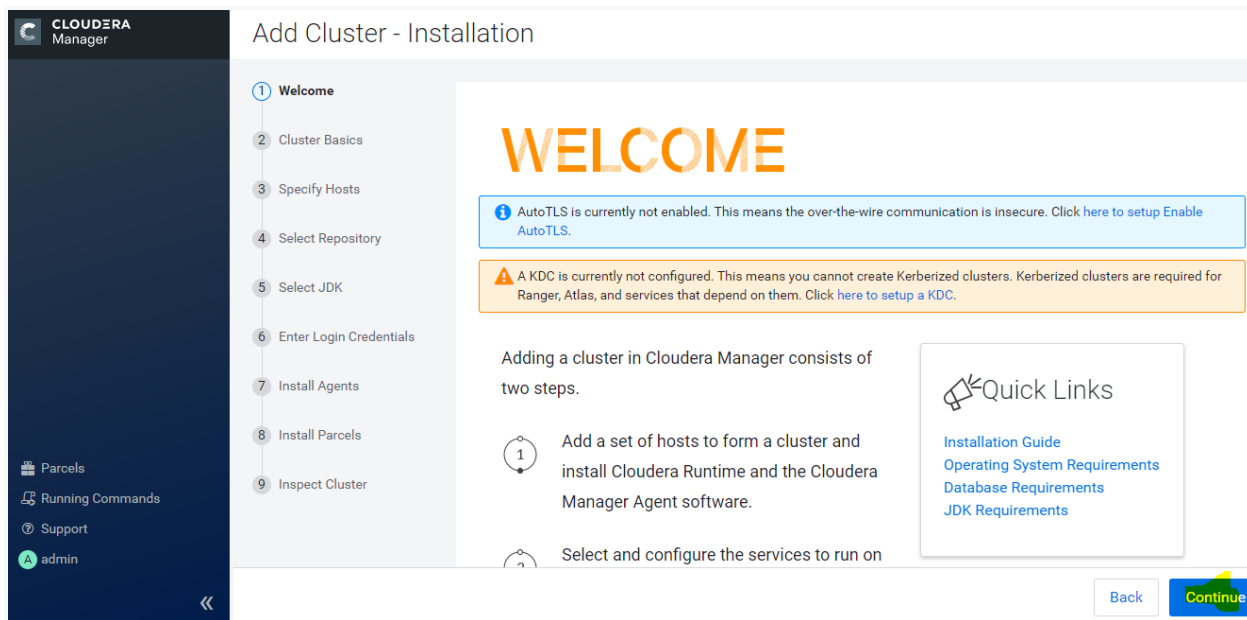


33. Now, log in to Cloudera Manager using **username: admin** and **password: admin**.

34. Choose Cloudera data platform for 60 days and Accept the terms and conditions, and then, keep clicking on '**Continue**'.



35. Click on **Continue**



WELCOME

AutoTLS is currently not enabled. This means the over-the-wire communication is insecure. [Click here to setup Enable AutoTLS.](#)

A KDC is currently not configured. This means you cannot create Kerberized clusters. Kerberized clusters are required for Ranger, Atlas, and services that depend on them. [Click here to setup a KDC.](#)

Adding a cluster in Cloudera Manager consists of two steps.

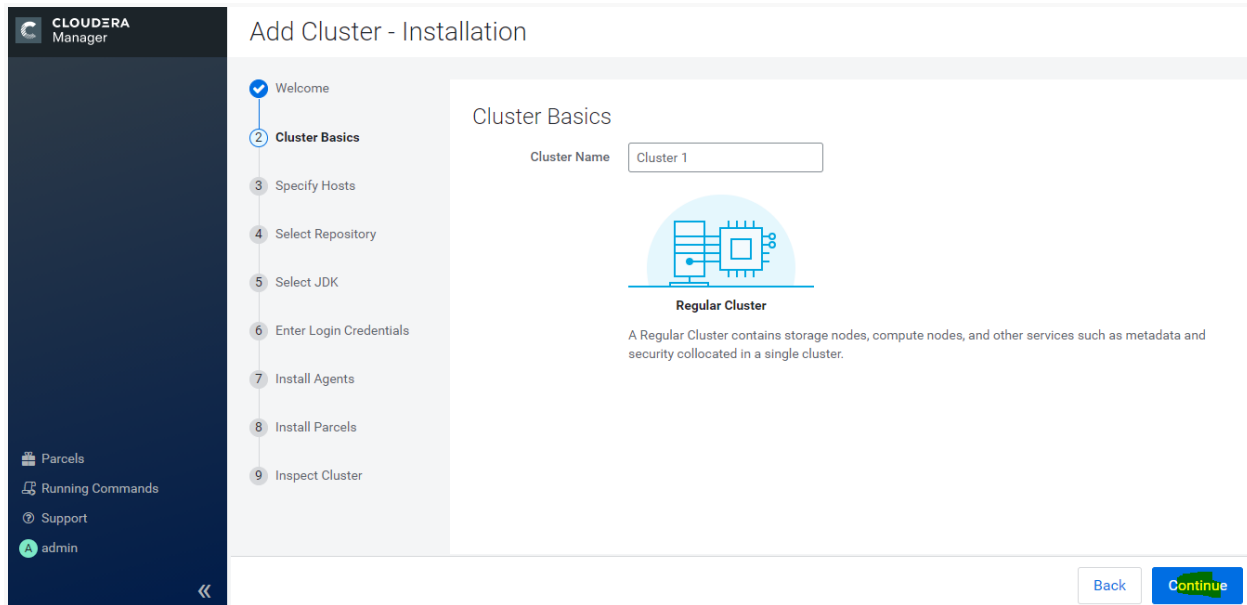
- 1 Add a set of hosts to form a cluster and install Cloudera Runtime and the Cloudera Manager Agent software.
- 2 Select and configure the services to run on

Quick Links

- [Installation Guide](#)
- [Operating System Requirements](#)
- [Database Requirements](#)
- [JDK Requirements](#)

[Back](#) [Continue](#)

36. Click on '**Continue**':



Cluster Basics

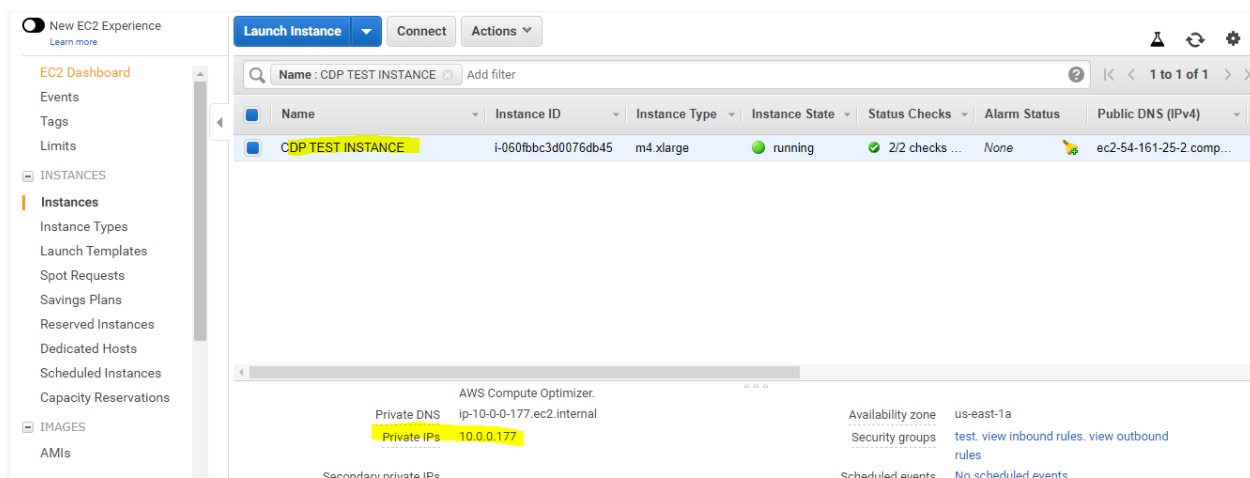
Cluster Name

Regular Cluster

A Regular Cluster contains storage nodes, compute nodes, and other services such as metadata and security collocated in a single cluster.

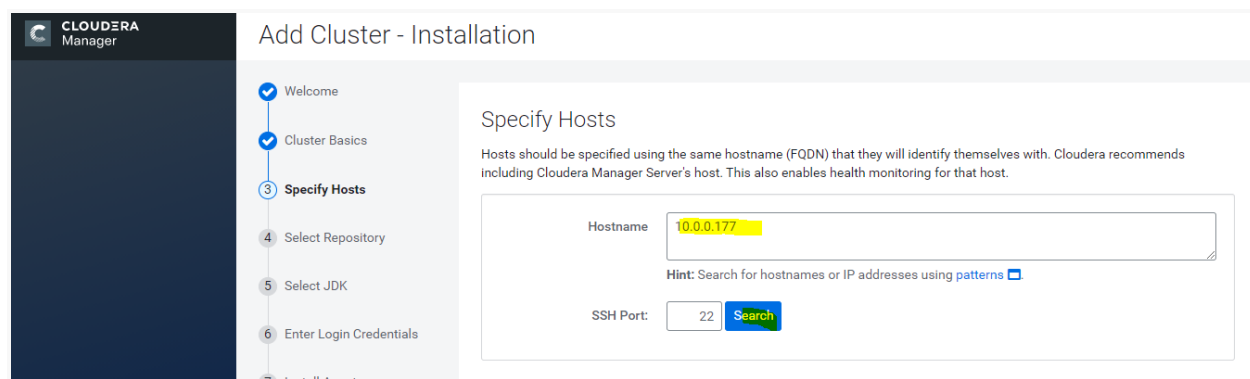
[Back](#) [Continue](#)

37. When you're asked to search for the instance, give your 'Private IP'.



The screenshot shows the AWS Management Console interface. On the left, there's a navigation menu with options like 'EC2 Dashboard', 'Events', 'Tags', 'Limits', 'INSTANCES', 'IMAGES', and 'AMIs'. The 'INSTANCES' section is expanded, showing a list of instances. The instance 'CDP TEST INSTANCE' is highlighted. Below the list, there's a detailed view of the instance, showing its 'Private IP' as 10.0.0.177, which is highlighted in yellow. Other details include 'Private DNS', 'Availability zone', 'Security groups', and 'Scheduled events'.

38. After this, the following screen will appear. Type your private IP in the box, and click on 'Search'



The screenshot shows the Cloudera Manager 'Add Cluster - Installation' screen. The 'Specify Hosts' step is active, showing a 'Hostname' field with the value 10.0.0.177 and an 'SSH Port' field with the value 22. A 'Search' button is visible. The 'Specify Hosts' section includes a hint: '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.'

39. Select your private IP from the search results and click on **'Continue'**

Cluster Basics

Specify Hosts

Select Repository

Select JDK

Enter Login Credentials

Install Agents

Install Parcels

Inspect Cluster

Specify Hosts

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.

Hostname

10.0.0.177

Hint: Search for hostnames or IP addresses using [patterns](#)

SSH Port:

22

Search

1 hosts scanned, 1 running SSH.

| Expanded Query | Hostname (FQDN) | IP Address | Currently Managed | Result |
|----------------|----------------------------|------------|-------------------|--------------------------------|
| 10.0.0.177 | ip-10-0-0-177.ec2.internal | 10.0.0.177 | No | Host was successfully scanned. |

1 - 1 of 1

40. Click on 'Continue':

CLUSTER MANAGER

Parcels

Running Commands

Support

admin

Cluster Basics

Specify Hosts

Select Repository

Select JDK

Enter Login Credentials

Install Agents

Install Parcels

Inspect Cluster

Cloudera Manager Agent

Cloudera Manager Agent 7.1.4 (#6363010) needs to be installed on all new hosts.

Repository Location ☒ Public Cloudera Repository

Ensure the above version is listed in <https://archive.cloudera.com/cm7> and that you have access to that repository. Requires direct Internet access on all hosts.

☐ Custom Repository

CDH and other software

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.

Install Method ☐ Use Packages ☒ Use Parcels (Recommended) [Parcel Repositories & Network Settings](#) [Other Parcel Configurations](#)

Version Versions that are too new for this version of Cloudera Manager (7.1.4) will not be shown.

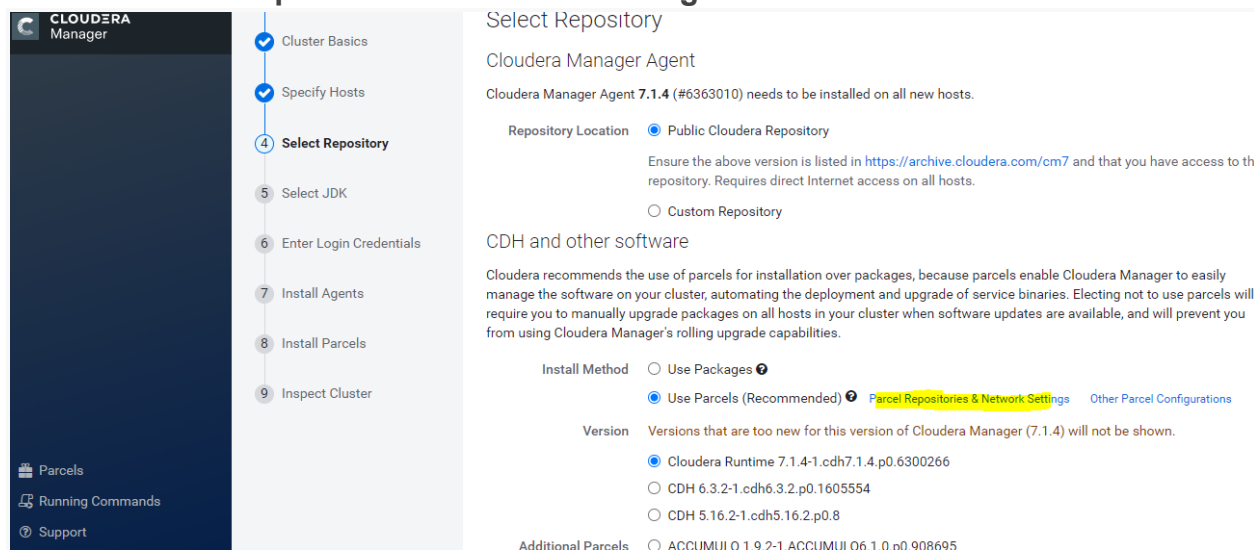
- ☒ Cloudera Runtime 7.1.3-1.cdh7.1.3.p0.4992530
- ☐ CDH 6.3.2-1.cdh6.3.2.p0.1605554
- ☐ CDH 5.16.2-1.cdh5.16.2.p0.8

Additional Parcels ☐ ACCUMULO 1.9.2-1.ACCUMULO6.1.0.p0.908695

Back

Continue

Click on **Parcel Repositories & Network setting**

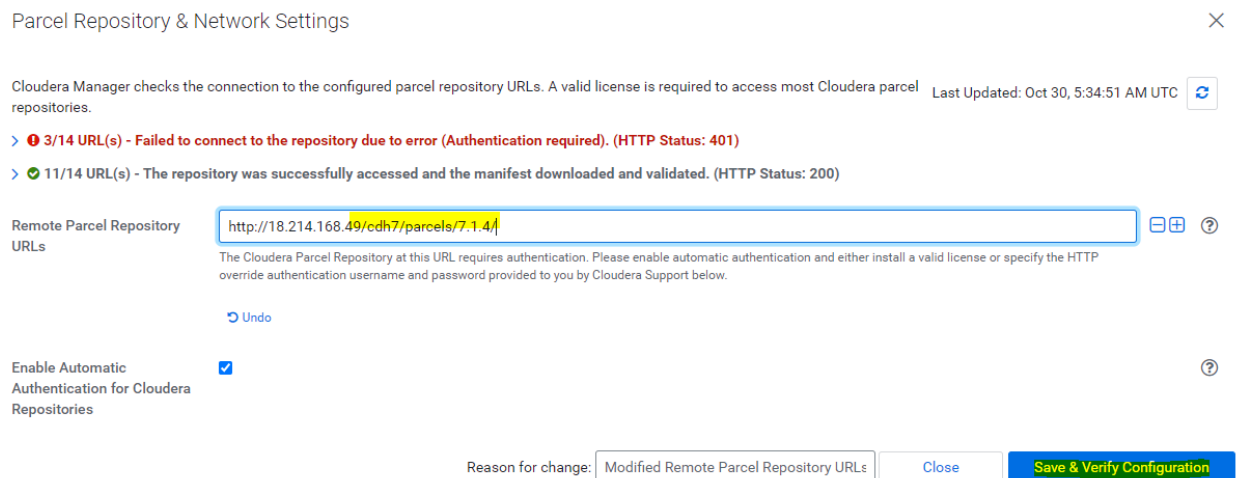


The screenshot shows the Cloudera Manager installation wizard. The left sidebar lists the steps: Cluster Basics, Specify Hosts, **Select Repository** (current step), Select JDK, Enter Login Credentials, Install Agents, Install Parcels, and Inspect Cluster. The main content area is titled 'Select Repository' and shows 'Cloudera Manager Agent 7.1.4 (#6363010)' needs to be installed. Under 'Repository Location', 'Public Cloudera Repository' is selected. A note mentions ensuring the version is listed in <https://archive.cloudera.com/cm7>. Under 'Install Method', 'Use Parcels (Recommended)' is selected, with a link to 'Parcel Repositories & Network Settings'. The 'Version' section shows 'Versions that are too new for this version of Cloudera Manager (7.1.4) will not be shown.' and lists 'Cloudera Runtime 7.1.4-1.cdh7.1.4.p0.6300266' as the selected option. Other options include 'CDH 6.3.2-1.cdh6.3.2.p0.1605554' and 'CDH 5.16.2-1.cdh5.16.2.p0.8'. The 'Additional Parcels' section shows 'ACCUMUI 0.1.9.2-1.ACCUMUI 06.1.0.n0.908695'.

Remove other parcels using(-)

And paste the below url:

<http://18.214.168.49/cdh7/parcels/7.1.4/>



The screenshot shows the 'Parcel Repository & Network Settings' dialog box. It displays a status message: 'Cloudera Manager checks the connection to the configured parcel repository URLs. A valid license is required to access most Cloudera parcel repositories. Last Updated: Oct 30, 5:34:51 AM UTC'. Below this, there are two status messages: '3/14 URL(s) - Failed to connect to the repository due to error (Authentication required). (HTTP Status: 401)' and '11/14 URL(s) - The repository was successfully accessed and the manifest downloaded and validated. (HTTP Status: 200)'. The 'Remote Parcel Repository URLs' section shows a text input field containing 'http://18.214.168.49/cdh7/parcels/7.1.4/'. Below the input field, there is a note: 'The Cloudera Parcel Repository at this URL requires authentication. Please enable automatic authentication and either install a valid license or specify the HTTP override authentication username and password provided to you by Cloudera Support below.' and an 'Undo' button. The 'Enable Automatic Authentication for Cloudera Repositories' checkbox is checked. At the bottom, there is a 'Reason for change:' field with 'Modified Remote Parcel Repository URLs' and buttons for 'Close' and 'Save & Verify Configuration'.

Click on **Save & Verify Configuration.**

41. Select the Install a Cloudera-provided version of open-jdk and encryption policy then click on ‘**Continue**’.

✓ Welcome

✓ Cluster Basics

✓ Specify Hosts

✓ Select Repository

5 Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels

9 Inspect Cluster

Select JDK

| | |
|------------------------------|-----------------------------------|
| Selected Version | Cloudera Runtime 7.1 |
| Supported JDK Version | OpenJDK 8, 11 or Oracle JDK 8, 11 |

[More details on supported JDK version.](#)

If you plan to use JDK 11, you will need to install it manually on all hosts and then select the **Manually manage JDK** option below.

☐ Manually manage JDK

Please ensure that a supported JDK is **already installed on all hosts. You will need to manage installing the unlimited strength JCE policy file, if necessary.**

☒ **Install a Cloudera** provided version of OpenJDK

By proceeding, Cloudera will install a supported version of OpenJDK version 8.

☐ Install a system-provided version of OpenJDK

By proceeding, Cloudera will install the default version of OpenJDK version 8 provided by the Operating System.

Back

Continue

42. Click ‘Continue’ again.

43. After this, the following screen will appear. Select **Another User** and type **ec2-user**. Select **'All hosts accept the same private key'** as the authentication method. Browse the **.pem** file in the **Private Key File** option, and click on **'Continue'**.

- Our case: RHEL.pem

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

✓ Specify Hosts

✓ Select Repository

✓ Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels

9 Inspect Cluster

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

ec2-user
(with password-less sudo/pbrun to root)

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

Private Key File: **RHEL1.pem**

Enter Passphrase:

Confirm Passphrase:

SSH Port:

- Click on **'Continue'**

Cluster Installation

Install Agents

Installation completed successfully.

1 of 1 host(s) completed successfully.

| Hostname | IP Address | Progress | Status |
|-------------------------------|---------------|-------------|--|
| ip-172-31-15-234.ec2.internal | 172.31.15.234 | <div></div> | ✓ Installation completed successfully. Details |

1234567

- Click on **'continue'**: it takes 30 mins to download and distribute the parcels

Cluster Installation

Install Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

| CDH 5.15.0-1.odh5.15.0.p0.21 | Downloaded: 100% | Distributed: 1/1 (47.7 MiB/s) | Unpacked: 1/1 | Activated: 1/1 |
|------------------------------|------------------|-------------------------------|---------------|----------------|
| <div></div> | <div></div> | <div></div> | <div></div> | <div></div> |

Back

1 2 3 4 5 6 7

Continue

44. Click on 'Continue' after selecting the below radio button as highlighted.

Install Parcels

9 Inspect Cluster

Ping Count

10

Number of times the inspector pings each host.

Ping Packet Size

56

Bytes

Size of the test packet sent when pinging the hosts.

Inspect Network Performance

Inspect Hosts

No issues were detected, review the inspector results to see what checks were performed.

Status

✓

Last Run

2 minutes ago

Duration

25.67s

Show Inspector Results

Run Again

More

☐ Fix the issues and run the inspection tools again.
 ☐ Quit the wizard and Cloudera Manager will delete the temporarily created cluster.
 ☒ I understand the risks of not running the inspections or the detected issues, let me continue with cluster setup.

Back








Continue

Select custom services.

Custom Services

Choose your own services. Services required by chosen services will automatically be included.

45. After this, the following screen will appear; select **Custom Services** and choose **HBase, HDFS, Hive, Hive on Tez, Hue, Livy, Phoenix, spark,, Tez, Yarn, ZooKeeper**, and then, click on **Continue**

| | | |
|-------------------------------------|---|--|
| <input type="checkbox"/> |  | Kudu is a true column store for the Hadoop ecosystem. |
| <input checked="" type="checkbox"/> |  | Oozie is a workflow coordination service to manage data processing jobs on your cluster. |
| <input type="checkbox"/> |  | Solr is a distributed service for indexing and searching data stored in HDFS. |
| <input type="checkbox"/> |  | Apache Spark is an open source cluster computing system. This service runs Spark as an application on YARN. |
| <input type="checkbox"/> |  | Sqoop is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases. The version supported by Cloudera is Sqoop 1. Note that Sqoop 2 has been deprecated and will be removed from CDH in a future release. |
| <input checked="" type="checkbox"/> |  | YARN (MR2 Included) Apache Hadoop MapReduce 2.0 (MRv2), or YARN, is a data computation framework that supports MapReduce applications (requires HDFS). |
| <input checked="" type="checkbox"/> |  | Apache ZooKeeper is a centralized service for maintaining and synchronizing configuration data. |

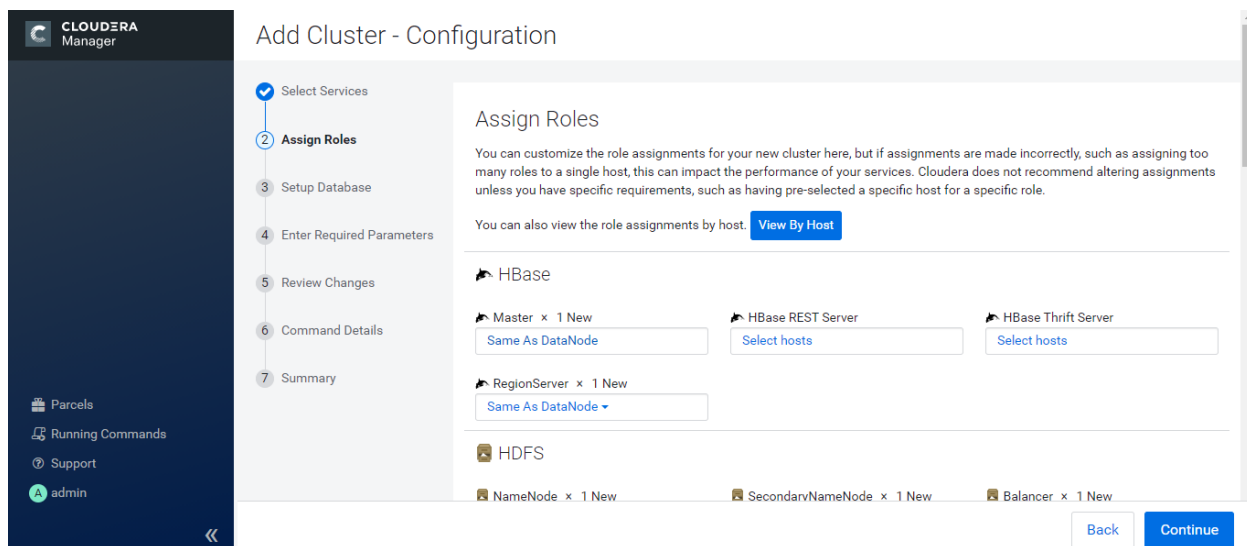
Note: Please ensure that you have the appropriate license key for **HBase** or contact Cloudera for assistance.

This wizard will also install the **Cloudera Management Service**. These are a set of components that enable monitoring, reporting, events, and alerts; these components require databases to store information, which will be configured on the next page.

☐ Include Cloudera Navigator

Back
1 2 3 4 5 6
Continue

- Click on '**Continue**'



The screenshot shows the Cloudera Manager interface for adding a new cluster. The left sidebar contains a navigation menu with options: Parcels, Running Commands, Support, and a user profile for 'admin'. The main content area is titled 'Add Cluster - Configuration' and shows a progress bar with seven steps: 1. Select Services, 2. Assign Roles (current step), 3. Setup Database, 4. Enter Required Parameters, 5. Review Changes, 6. Command Details, and 7. Summary. The 'Assign Roles' section includes a warning about role assignments and a 'View By Host' button. Below this, there are sections for 'HBase' and 'HDFS' roles. The 'HBase' section shows 'Master x 1 New' (set to 'Same As DataNode'), 'HBase REST Server' (set to 'Select hosts'), and 'HBase Thrift Server' (set to 'Select hosts'). The 'HDFS' section shows 'NameNode x 1 New', 'SecondaryNameNode x 1 New', and 'Balancer x 1 New'. At the bottom right, there are 'Back' and 'Continue' buttons.

46. Select '**use custom data bases**'.

Note :- we already installed MySQL on that AMI with the databases metastore rman hue and oozie.

47. Please keep all the usernames and passwords for all the services: Hive, Hue, etc. in a safe place. After paste/ enter these database names and passwords, scroll down, and then, click on **‘Test Connections’**.

Change Database Hostname port number is- 3306 Database type - MySQL

For hive service:- database name: **metastore** and password is **admin**.

For rman service:- database name: **rman** and password is **admin**

For hue service:- database name: **hue** and password is **admin**

For oozie service :- database name: **oozie** and password is **admin**

☒ Use Custom Databases ☐ Use Embedded Database

Hive

| | | | |
|----------|-----------------------|---------------------------------|---------------|
| Type | Use JDBC URL Override | Database Hostname | Database Name |
| MySQL | No | ip-10-0-0-194.ec2.internal:3306 | metastore |
| Username | Password | | |
| hive | admin | | |

Reports Manager

Currently assigned to run on **ip-10-0-0-194.ec2.internal**.

| | | | |
|----------|---------------------------------|---------------|----------|
| Type | Database Hostname | Database Name | Username |
| MySQL | ip-10-0-0-194.ec2.internal:3306 | rman | rman |
| Password | | | |
| admin | | | |

Hue

| | | | |
|----------|---------------------------------|---------------|----------|
| Type | Database Hostname | Database Name | Username |
| MySQL | ip-10-0-0-194.ec2.internal:3306 | hue | hue |
| Password | | | |
| admin | | | |

- Click on **Continue**:

Add Cluster - Configuration

✓ Select Services

✓ Assign Roles

✓ Setup Database

✓ Enter Required Parameters

5 **Review Changes**

6 Command Details

7 Summary

Review Changes

HDFS Root Directory

hbase.rootdir

Cluster 1 > HBase (Service-Wide)

/hbase

?

Enable Replication

hbase.replication

☐

Cluster 1 > HBase (Service-Wide)

?

Enable Indexing

☐

Cluster 1 > HBase (Service-Wide)

?

HDFS Block Size

dfs.blocksize

Cluster 1 > HDFS (Service-Wide)

128

MiB

?

DataNode Failed Volumes Tolerated

dfs.datanode.failed.volumes.tolerated

Cluster 1 > DataNode Default Group

0

?

Back

Continue

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

- Deploying all the services takes some time. After this is complete, click on **Continue**

✓ Completed 9 of 9 step(s).

☒ Show All Steps
 ☐ Show Only Failed Steps
 ☐ Show Running Steps

| | | |
|--|--------------------|--------|
| > ✓ Ensuring that the expected software releases are installed on hosts. | Jun 18, 8:34:03 PM | 42ms |
| > ✓ Deploying Client Configuration Cluster 1 | Jun 18, 8:34:03 PM | 15.79s |
| > ✓ Start Cloudera Management Service, ZooKeeper | Jun 18, 8:34:19 PM | 28.77s |
| > ✓ Start HDFS | Jun 18, 8:34:48 PM | 50.77s |
| > ✓ Start HBase | Jun 18, 8:35:39 PM | 27.37s |
| > ✓ Start YARN (MR2 Included) | Jun 18, 8:36:06 PM | 27.48s |
| > ✓ Start Hive | Jun 18, 8:36:34 PM | 54.18s |
| > ✓ Start Oozie | Jun 18, 8:37:28 PM | 57.3s |
| > ✓ Start Hue | Jun 18, 8:38:25 PM | 22.73s |

Feedback

[Back](#)
1 2 3 4 5 6
[Continue](#)

- The following screen will appear on the successful set-up of the cluster. Then, click on **Finish**.

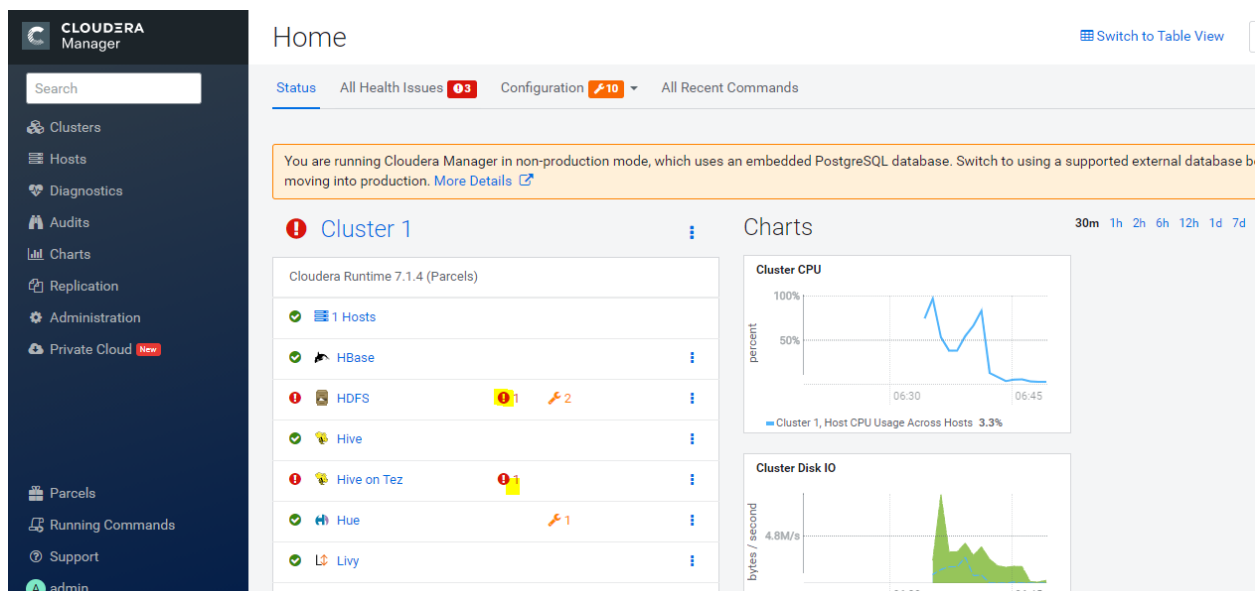
Cluster Setup

Congratulations!

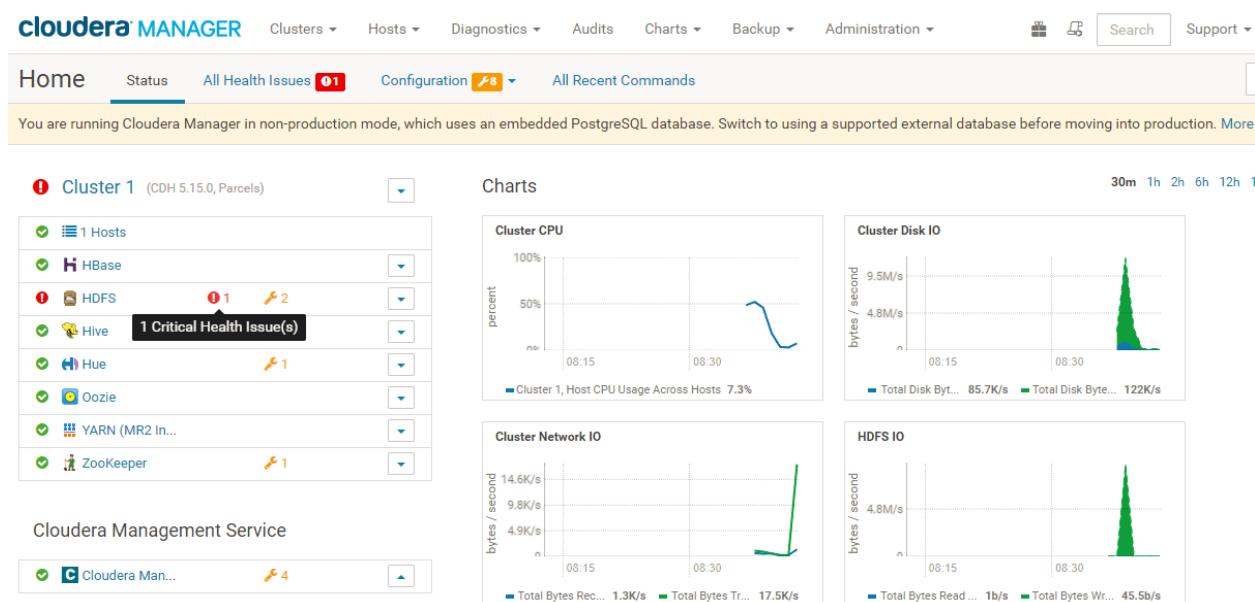
✓ The services are installed, configured, and running on your cluster.

[Back](#)
1 2 3 4 5 6
[Finish](#)

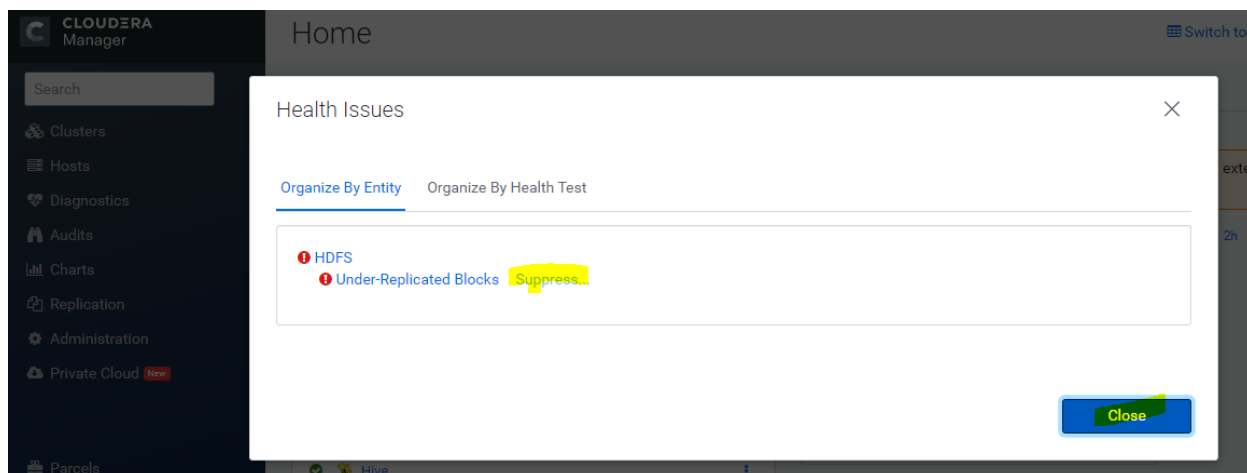
48. After this, the following image will appear. Kindly check whether all the services are green. Please note that it takes some time for all the services to be active



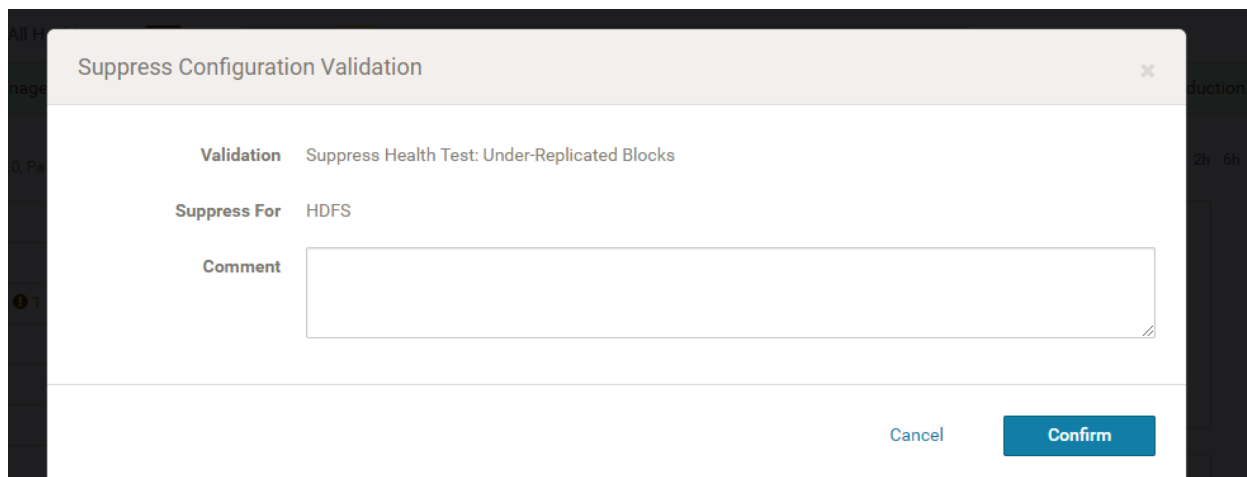
49. Click on 'Critical Health Issue(s)' (if any).



50. Click on 'suppress':

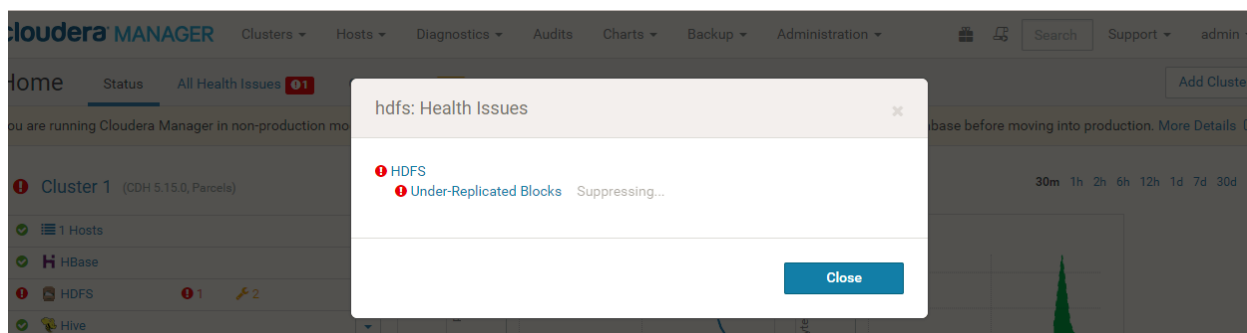


- Click on Confirm:

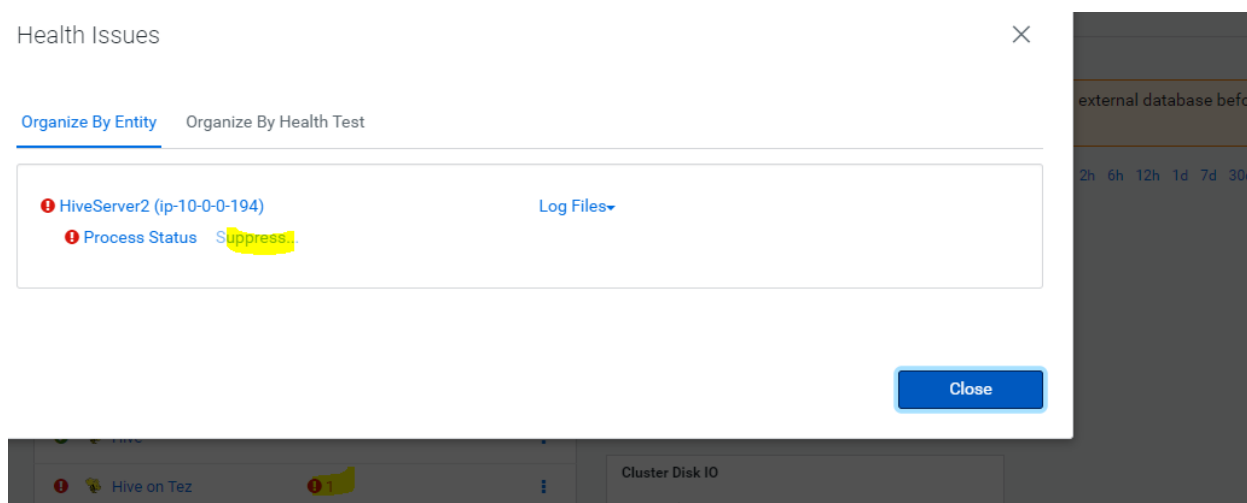


-

- Click on Close.



Do the same for the window below



Click on hiveServer2.

×

Suppress Health Test

Health Test

Suppress Health Test: Process Status

Suppress For

HiveServer2 (ip-10-0-0-194)


☐ Suppress For HiveServer2 Default Group

Comment

Cancel

Confirm

Click on Cloudera Manager.



CLUSTERA

Manager

Search

Clusters

Hosts

Diagnostics

Audits

Charts

Replication

Administration

Private Cloud New

Parcels

Running Commands

Support

Home

Status

All Health Issues 2

Configuration 10

All Recent Commands

You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database when moving into production. [More Details](#)

Cluster 1

Cloudera Runtime 7.1.4 (Parcels)

1 Hosts

HBase

HDFS 2

Hive

Hive on Tez

Hue 1

Livy

Charts

Cluster CPU

percent

100%

50%

06:30

06:45

Cluster 1, Host CPU Usage Across Hosts 5%

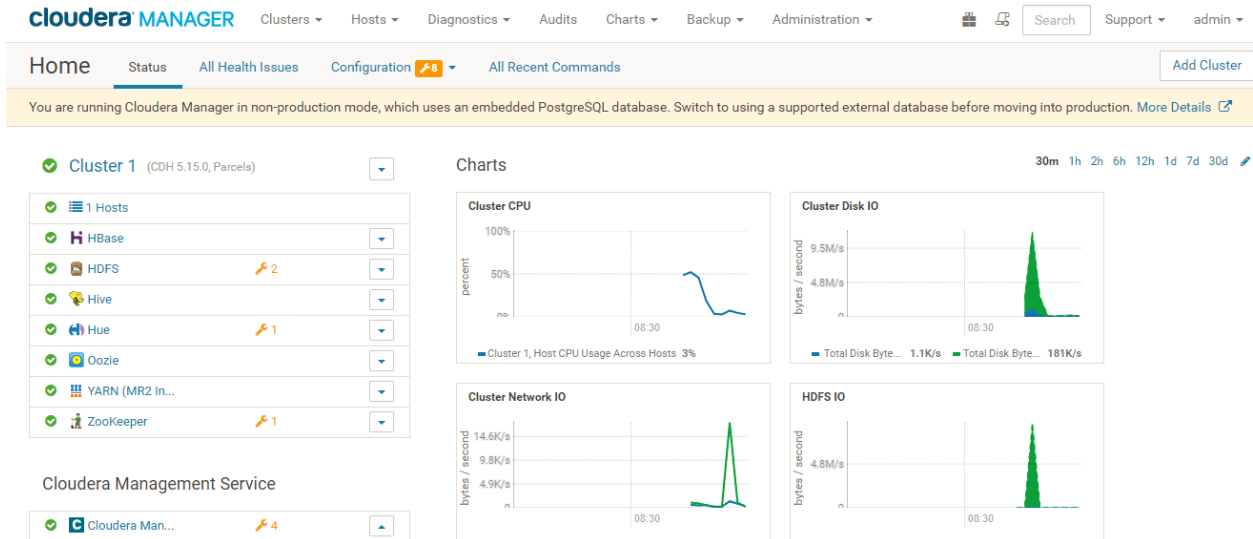
Cluster Disk IO

rates / second

4.8M/s

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

- Now all services are green



51. Set the java path for cloudera-scm-manager or other users:
Again login to ec2 instance using putty and switch to root users;

ls /usr/java/jdk1.8.0_232-cloudera/ (enter command without quotes).

```
[ec2-user@ip-172-31-15-234 ~]$ sudo -i
[root@ip-172-31-15-234 ~]# ls /usr/java/jdk1.7.0_67-cloudera/
bin  COPYRIGHT  db  include  jre  lib  LICENSE  man  README.html  release  src.zip  THIRDPARTYLICENSEREADME-JAVAFX.txt  THIRDPARTYLICENSEREADME.txt
[root@ip-172-31-15-234 ~]#
```

vi /etc/default/cloudera-scm-server

Note :- vi takes us to vim text editor. To enter any text we need to enter into insert mode which is achieved by pressing 'i' and to 'save and quit' the doc, we need to hit 'esc' followed by :wq! and hit enter.

Then, press ' i ' (without quotes) for insert mode, and paste this command as shown:
export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/

```

root@ip-10-0-0-194:~
# Specify any command line arguments for the Cloudera SCM Server here.
#
export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/
CMF_SERVER_ARGS=""
#
# Locate the JDBC driver jar file.
#
# The default value is the default system mysql driver on RHEL/CentOS/Ubuntu
# and the standard, documented location for where to put the oracle jar in CM
# deployments.
#
export CMF_JDBC_DRIVER_JAR="/usr/share/java/mysql-connector-java.jar:/usr/share/java/oracle-connector-java.jar:/usr/share/java/postgresql
#
# Customize the TLS ciphers used by Cloudera Manager.
#
# Cloudera Manager uses the Modern list of TLS ciphers, as defined by Mozilla
# (reference: https://wiki.mozilla.org/Security/Server_Side_TLS).
# Some older JDK versions might not have support for newer Ciphers in that list,
# potentially leading to failures to connect to the Admin Console when TLS is
# enabled. You can customize the set of ciphers by uncommenting and editing
# the line below. The example provided will set the list of ciphers to the
# Intermediate list defined by Mozilla.
#
# export CMF_OVERRIDE_TLS_CIPHERS="TLS ECDHE_ECDSA WITH AES_128_GCM_SHA256:TLS ECDHE_RSA WITH AES_128_GCM_SHA256:TLS ECDHE_ECDSA WITH AES
ITH AES_256_GCM_SHA384:TLS DHE_RSA WITH AES_128_GCM_SHA256:TLS DHE_RSA WITH AES_256_GCM_SHA384:TLS ECDHE_ECDSA WITH AES_128_CBC_SHA256:TL
A256:TLS ECDHE_ECDSA WITH AES_128_CBC_SHA:TLS ECDHE_RSA WITH AES_256_CBC_SHA384:TLS ECDHE_RSA WITH AES_128_CBC_SHA:TLS ECDHE_ECDSA WITH A
SA WITH AES_256_CBC_SHA:TLS ECDHE_RSA WITH AES_256_CBC_SHA:TLS DHE_RSA WITH AES_128_CBC_SHA256:TLS DHE_RSA WITH AES_128_CBC_SHA:TLS DHE_R
DHE_RSA WITH AES_256_CBC_SHA:TLS ECDHE_ECDSA WITH_3DES_EDE_CBC_SHA:TLS ECDHE_RSA WITH_3DES_EDE_CBC_SHA:TLS EDH_RSA WITH_3DES_EDE_CBC_SHA:
:TLS_RSA WITH AES_256_GCM_SHA384:TLS_RSA WITH AES_128_CBC_SHA256:TLS_RSA WITH AES_256_CBC_SHA256:TLS_RSA WITH AES_128_CBC_SHA:TLS_RSA WIT
3DES_EDE_CBC_SHA"
#
# Java Options.
#

```

For others users: open `/etc/profile` and go to the end of the file and insert the variable value.

vi /etc/profile (command)

```

export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/
export JRE_HOME=/usr/java/jdk1.8.0_232-cloudera/jre/
export PATH=$JAVA_HOME/bin:$PATH

```

```

unset i
unset -f pathmunge
export JAVA_HOME=/usr/java/jdk1.8.0_232-cloudera/
export JRE_HOME=/usr/java/jdk1.8.0_232-cloudera/jre/
export PATH=$JAVA_HOME/bin:$PATH
~
-- INSERT --

```

Then update the **/etc/profile** using the below commands:

source /etc/profile

```
[root@ip-172-31-15-234 ~]# source /etc/profile  
[root@ip-172-31-15-234 ~]#
```

Verify the java and JRE version using the below commands as shown in the image below:

java -version

echo \$JAVA_HOME

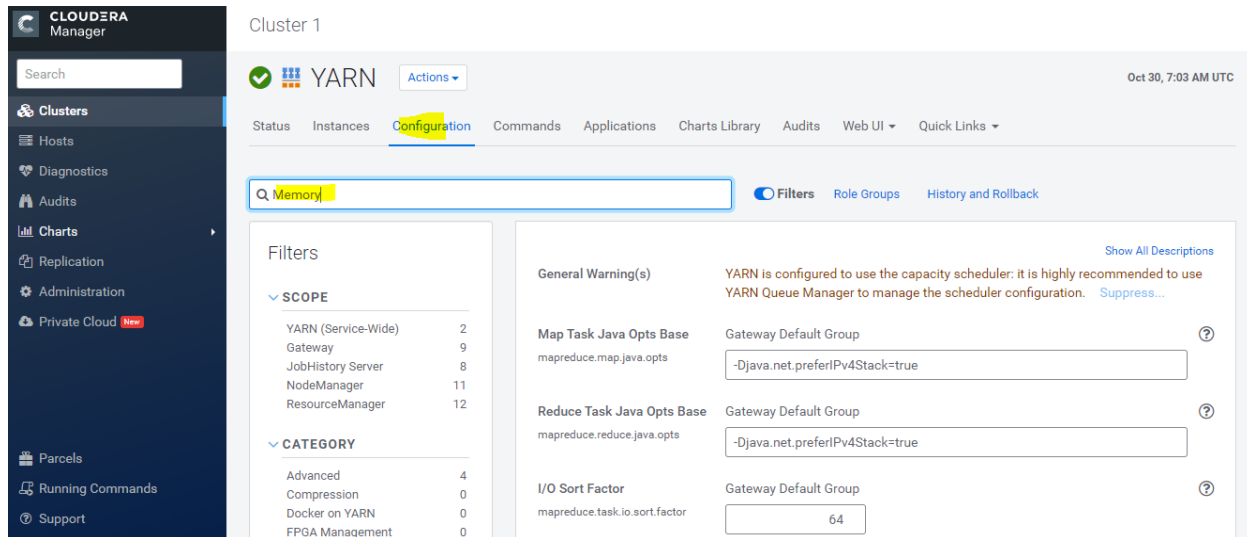
```
[root@ip-10-0-0-194 ~]# java -version  
openjdk version "1.8.0_232"  
OpenJDK Runtime Environment (AdoptOpenJDK) (build 1.8.0_232-b09)  
OpenJDK 64-Bit Server VM (AdoptOpenJDK) (build 25.232-b09, mixed mode)  
[root@ip-10-0-0-194 ~]# echo $JAVA_HOME  
/usr/java/jdk1.8.0_232-cloudera/  
[root@ip-10-0-0-194 ~]# echo $JRE_HOME  
/usr/java/jdk1.8.0_232-cloudera/jre/  
[root@ip-10-0-0-194 ~]#
```

Configure Yarn tuning:

yarn.nodemanager.resource.memory-mb 10GB

yarn.scheduler.maximum-allocation-mb 8GB

Go to yarn services> configuration and search Memory> scroll down



The screenshot shows the Cloudera Manager interface for configuring YARN. The left sidebar contains navigation options: Clusters, Hosts, Diagnostics, Audits, Charts, Replication, Administration, Private Cloud, Parcels, Running Commands, and Support. The main content area is titled 'Cluster 1' and shows the 'Configuration' tab. A search bar at the top of the configuration area contains the text 'Memory'. Below the search bar, there are filters for 'SCOPE' and 'CATEGORY'. The 'SCOPE' filter shows a list of components: YARN (Service-Wide) with 2 items, Gateway with 9 items, JobHistory Server with 8 items, NodeManager with 11 items, and ResourceManager with 12 items. The 'CATEGORY' filter shows a list of categories: Advanced with 4 items, Compression with 0 items, Docker on YARN with 0 items, and FPGA Management with 0 items. The main configuration area displays several settings: 'General Warning(s)' with a message about the capacity scheduler, 'Map Task Java Opts Base' with a value of '-Djava.net.preferIPv4Stack=true', 'Reduce Task Java Opts Base' with a value of '-Djava.net.preferIPv4Stack=true', and 'I/O Sort Factor' with a value of '64'.

Container Memory

NodeManager Default Group [Undo](#)

?

yarn.nodemanager.resource.memory-mb

GiB

Enable Virtual Memory Check

☐ NodeManager Default Group

?

yarn.nodemanager.vmem-check-enabled

Virtual Memory to Physical Memory Ratio

NodeManager Default Group

?

yarn.nodemanager.vmem-pmem-ratio

Java Heap Size of ResourceManager in Bytes

ResourceManager Default Group [Undo](#)

?

yarn.nodemanager.java_heap_size

MiB

Container Memory Minimum

ResourceManager Default Group

?

yarn.scheduler.minimum-allocation-mb

GiB

Per Page

25

1 - 25 of 377

[«](#)
[<](#)
[>](#)
[»](#)

Go to next page

Q Memory

Filters

Role Groups

History and Rollback

Filters

SCOPE

YARN (Service-Wide) 2

Gateway 9

JobHistory Server 8

NodeManager 11

ResourceManager 12

CATEGORY

Advanced 4

Compression 0

Docker on YARN 0

FPGA Management 0

GPU Management 0

Log Aggregation 0

Logs 0

Main 5

Monitoring 1

Performance 0

Ports and Addresses 0

Resources 0

General Warning(s)

YARN is configured to use the capacity scheduler: it is highly recommended to use YARN Queue Manager to manage the scheduler configuration. [Suppress...](#)

Show All Descriptions

Container Memory Increment

ResourceManager Default Group

?

yarn.scheduler.increment-allocation-mb

MiB

Container Memory Maximum

ResourceManager Default Group [Undo](#)

?

yarn.scheduler.maximum-allocation-mb

GiB

Resource Types

YARN (Service-Wide)

?

No rules created

[View as JSON](#)

[+](#)

Per Page

25

26 - 50 of 377

[«](#)
[<](#)
[>](#)
[»](#)

2 Edited Values

Reason for change:

Modified Container Memory, Container Memory Maximum

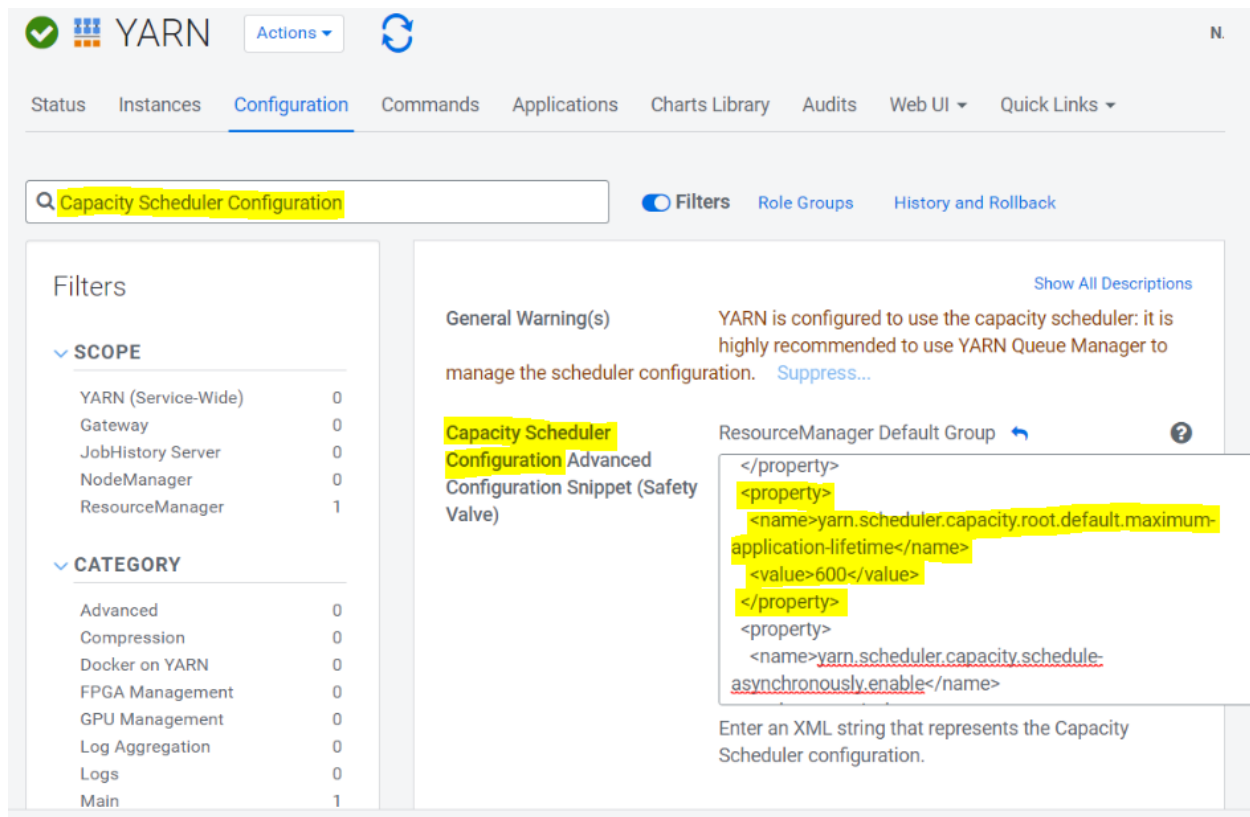
Save Changes (CTRL+S)

© Copyright. upGrad Education Pvt. Ltd. All rights reserved

Click on Save Changes.

Next, Go to yarn configuration and search - Capacity Scheduler Configuration and add the following property:

```
<property>
  <name>yarn.scheduler.capacity.root.default.maximum-application-lifetime</name>
  <value>600</value>
</property>
<property>
```



The screenshot shows the YARN Configuration page. The search bar contains "Capacity Scheduler Configuration". The left sidebar shows filters for SCOPE and CATEGORY. The main content area displays the "Capacity Scheduler Configuration" section, which includes a "General Warning(s)" and a "Configuration Snippet (Safety Valve)". The configuration snippet is an XML string that sets the maximum application lifetime to 600 seconds.

Filters

SCOPE

| | |
|---------------------|---|
| YARN (Service-Wide) | 0 |
| Gateway | 0 |
| JobHistory Server | 0 |
| NodeManager | 0 |
| ResourceManager | 1 |

CATEGORY

| | |
|-----------------|---|
| Advanced | 0 |
| Compression | 0 |
| Docker on YARN | 0 |
| FPGA Management | 0 |
| GPU Management | 0 |
| Log Aggregation | 0 |
| Logs | 0 |
| Main | 1 |

General Warning(s)

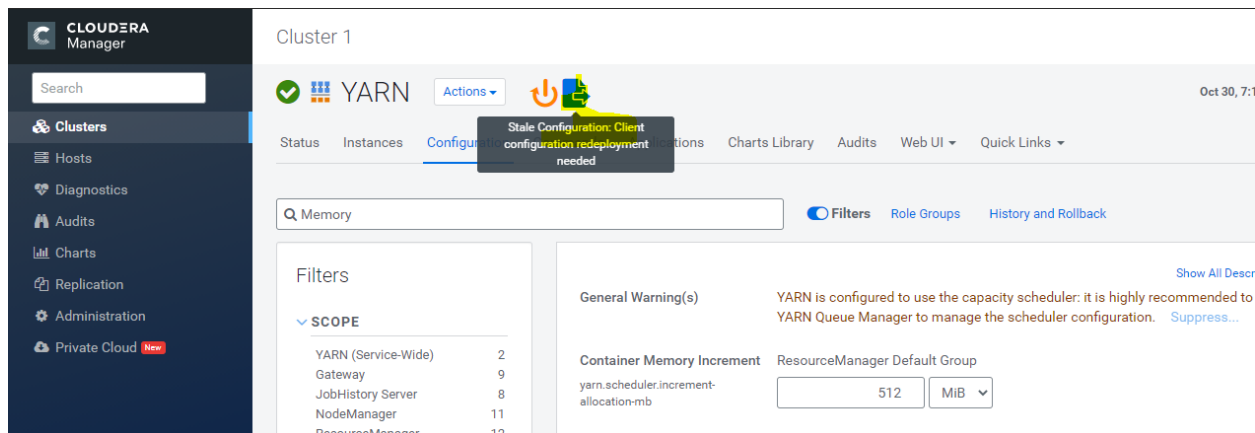
YARN is configured to use the capacity scheduler: it is highly recommended to use YARN Queue Manager to manage the scheduler configuration. [Suppress...](#)

Configuration Snippet (Safety Valve)

```
</property>
<property>
  <name>yarn.scheduler.capacity.root.default.maximum-application-lifetime</name>
  <value>600</value>
</property>
<property>
  <name>yarn.scheduler.capacity.schedule-asynchronously.enable</name>
```

Enter an XML string that represents the Capacity Scheduler configuration.

Click on Save Changes. And click on the Stale Configuration button.



Cluster 1

YARN

Actions

Oct 30, 7:1

Status Instances Configuration Actions Charts Library Audits Web UI Quick Links

Q Memory

Filters

SCOPE

| | |
|---------------------|----|
| YARN (Service-Wide) | 2 |
| Gateway | 9 |
| JobHistory Server | 8 |
| NodeManager | 11 |
| ResourceManager | 12 |

General Warning(s)

YARN is configured to use the capacity scheduler: it is highly recommended to YARN Queue Manager to manage the scheduler configuration. [Show All Descr](#) [Suppress...](#)

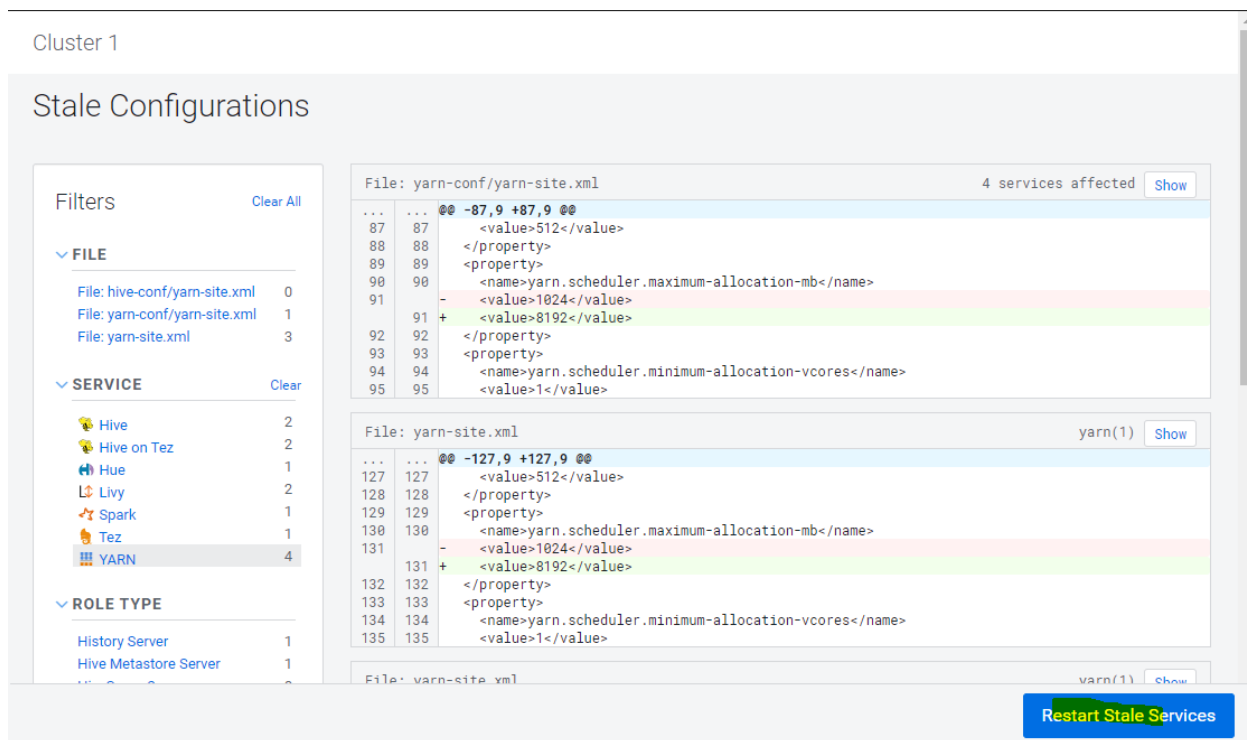
Container Memory Increment

ResourceManager Default Group

yarn.scheduler.increment-allocation-mb

512 MiB

Click on Restart Stale Services



Cluster 1

Stale Configurations

Filters

Clear All

FILE

| | |
|-------------------------------|---|
| File: hive-conf/yarn-site.xml | 0 |
| File: yarn-conf/yarn-site.xml | 1 |
| File: yarn-site.xml | 3 |

SERVICE

| | |
|-------------|---|
| Hive | 2 |
| Hive on Tez | 2 |
| Hue | 1 |
| Livy | 2 |
| Spark | 1 |
| Tez | 1 |
| YARN | 4 |

ROLE TYPE

| | |
|-----------------------|---|
| History Server | 1 |
| Hive Metastore Server | 1 |

File: yarn-conf/yarn-site.xml

4 services affected

Show

```

... -87,9 +87,9
87 <value>512</value>
88 </property>
89 <property>
90 <name>yarn.scheduler.maximum-allocation-mb</name>
91 - <value>1024</value>
92 + <value>8192</value>
93 </property>
94 <property>
95 <name>yarn.scheduler.minimum-allocation-vcores</name>
96 <value>1</value>

```

File: yarn-site.xml

yarn(1)

Show

File: yarn-site.xml

yarn(1)

Show

Restart Stale Services

Restart Stale Services

1 Review Changes

2 Command Details


Review Changes

All services running with outdated configurations in the cluster and their dependencies will be restarted.

☒ Re-deploy client configuration


Back

Restart Now

































CLOUDERA Manager

- Clusters
- Hosts
- Diagnostics
- Audits
- Charts
- Replication
- Administration
- Private Cloud New

You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using an external database when moving into production. [More Details](#)

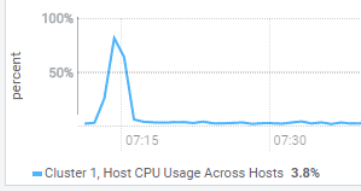

Cluster 1

Cloudera Runtime 7.1.4 (Parcels)

| | | | |
|---|---|---|---|
|  |  1 Hosts |  1 |  |
|  |  HBase | |  |
|  |  HDFS |  2 |  |
|  |  Hive | |  |
|  |  Hive on Tez | |  |
|  |  Hue |  1 |  |
|  |  Livy | |  |
|  |  Phoenix |  1 |  |
|  |  Spark | |  |

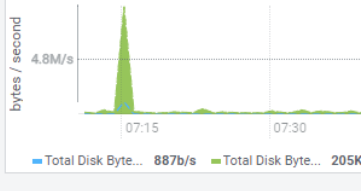
Charts

Cluster CPU



Cluster 1, Host CPU Usage Across Hosts 3.8%

Cluster Disk IO



Total Disk Byte... 887b/s Total Disk Byte... 205K

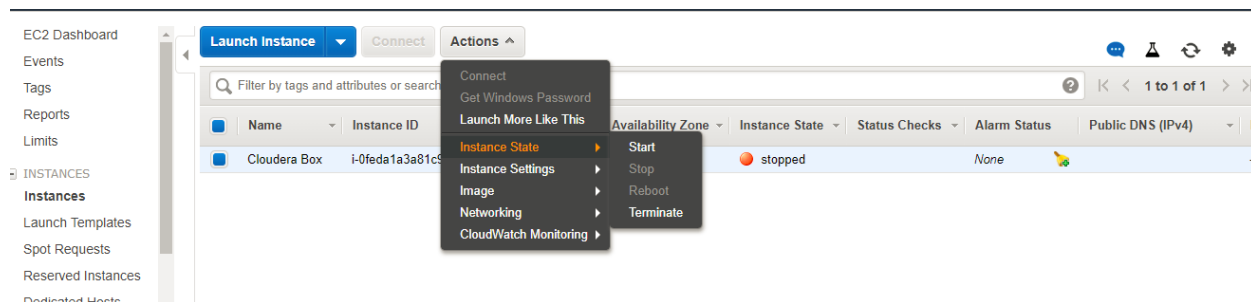
Cluster Network IO

Note: Please stop the instance if you are not using the ec2 dashboard.

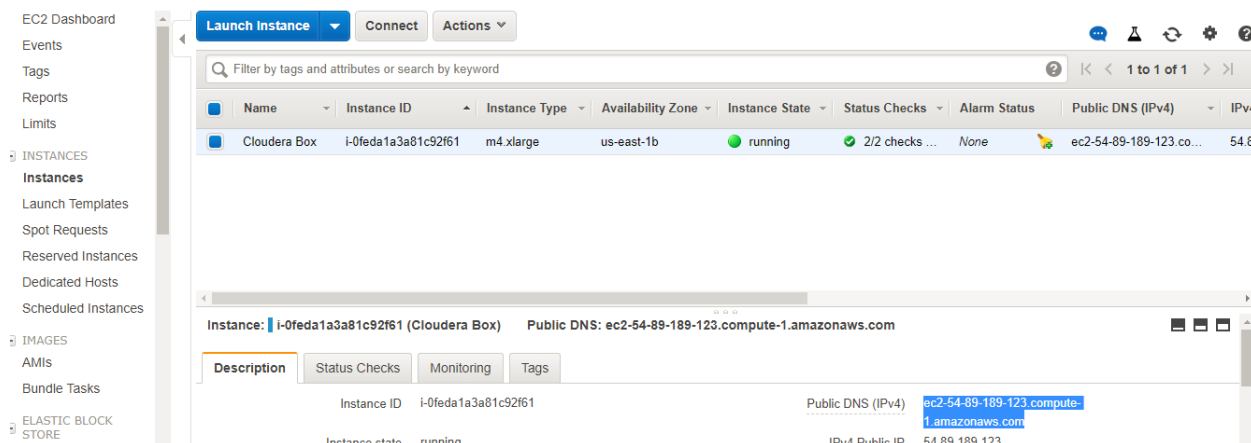
If your instance is running then the cost is Rs.19/hours

After I stopped my instance and restarted it, my Cloudera management services and other services stopped into my browser. The following screen appeared.

A. Start the instance from the EC2 dashboard and wait until the status check is 2/2.



B. Copy the public IP and paste into the browser



<http://publicip:7180>


cloudera MANAGER

☐ Remember me

Log In


cloudera MANAGER Clusters ▾ Hosts ▾ Diagnostics ▾ Audits Charts ▾ Backup ▾ Administration ▾ Support ▾ admin ▾


Home Status All Health Issues Configuration  8 ▾ All Recent Commands


You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database before moving into production. [More Details](#) 



Request to the **Service Monitor** failed. This may cause slow page responses. [View the status of the Service Monitor.](#)


Request to the **Host Monitor** failed. This may cause slow page responses. [View the status of the Host Monitor.](#)




 **Cluster 1** (CDH 5.15.0, Parcels) ▾



 1 Hosts


 HBase ▾



 HDFS  2 ▾

 Hive ▾



 Hue  1  ▾

 Oozie  ▾

 YARN (MR2 In... ▾

 ZooKeeper  1 ▾

Cloudera Management Service

 Cloudera Man...  4 ▾

Charts

Internal error while querying the Host Monitor

Cluster CPU


QUERY ERROR

Cluster Disk IO

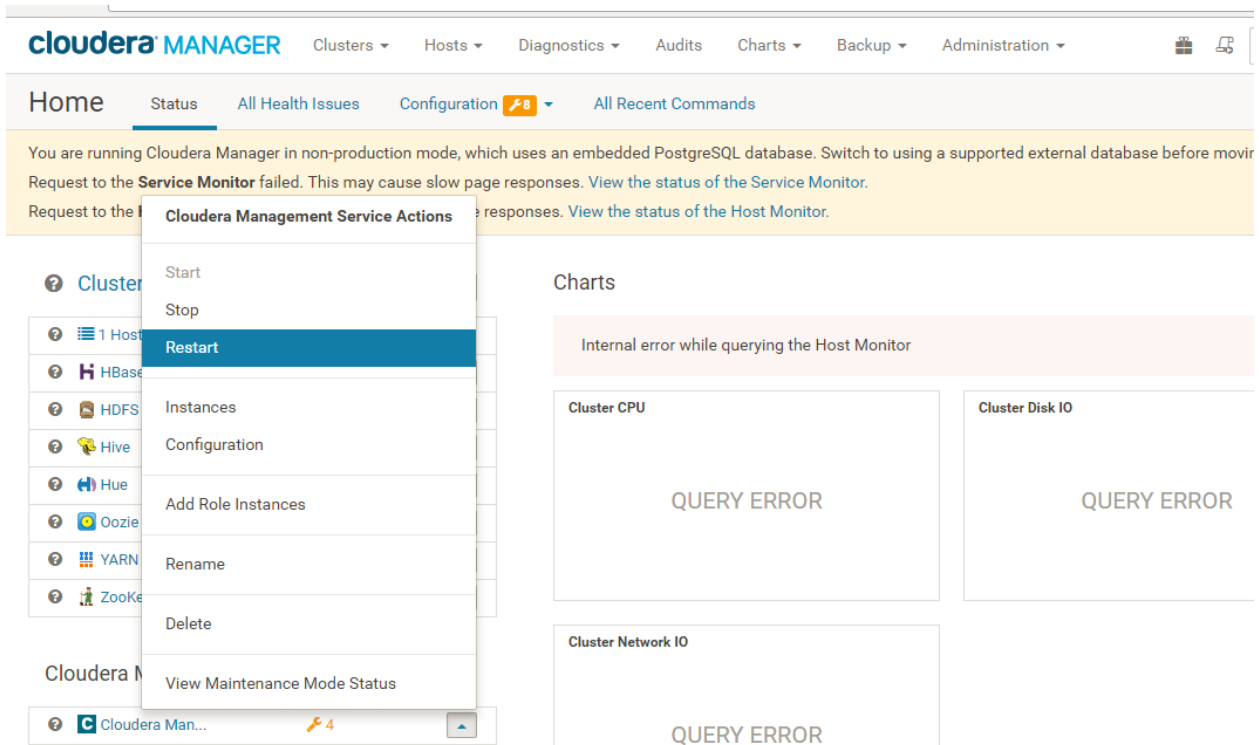
QUERY ERROR

Cluster Network IO

QUERY ERROR

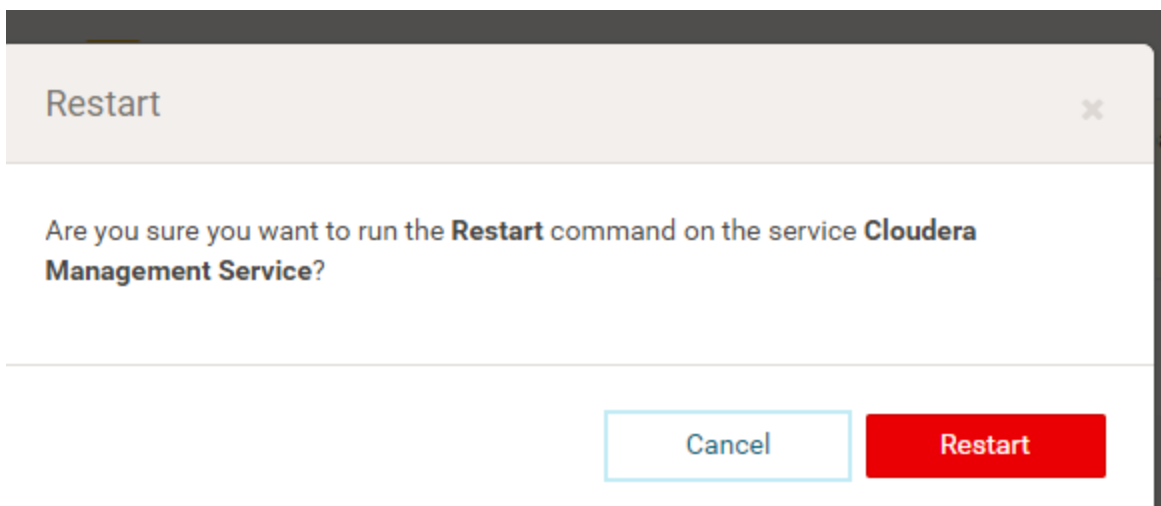
30m 1h 2h 6h 12h 1d 7d 30d 

To solve this problem, go to the page above, scroll down, and click on ‘**Cloudera Management Service**’; then, click on ‘**Restart**’



The screenshot shows the Cloudera Manager web interface. The top navigation bar includes 'Clusters', 'Hosts', 'Diagnostics', 'Audits', 'Charts', 'Backup', and 'Administration'. The main header shows 'Home', 'Status', 'All Health Issues', 'Configuration' (with a wrench icon), and 'All Recent Commands'. A yellow banner at the top contains a warning: 'You are running Cloudera Manager in non-production mode, which uses an embedded PostgreSQL database. Switch to using a supported external database before moving to production. Request to the Service Monitor failed. This may cause slow page responses. View the status of the Service Monitor.' Below this, a dropdown menu for 'Cloudera Management Service Actions' is open, showing options: 'Start', 'Stop', 'Restart' (highlighted in blue), 'Instances', 'Configuration', 'Add Role Instances', 'Rename', 'Delete', and 'View Maintenance Mode Status'. The background shows a 'Charts' section with three panels: 'Cluster CPU', 'Cluster Disk IO', and 'Cluster Network IO', all displaying 'QUERY ERROR'.

Click on ‘**Restart**’:



The screenshot shows a modal dialog box titled 'Restart'. The text inside asks: 'Are you sure you want to run the **Restart** command on the service **Cloudera Management Service**?'. At the bottom of the dialog, there are two buttons: 'Cancel' (a light blue button) and 'Restart' (a red button).

Click on 'Close':

cloudera MANAGER

Clusters Hosts Diagnostics Audits Charts Backup Administration

Search Support admin

Home

You are

?

1

2

3

4

5

6

7

8

9

10

Close

Restart Command

Status Finished Context [Cloudera Management Service](#) Jun 19, 5:29:11 AM 29.88s

Successfully restarted service.

Completed 2 of 2 step(s).

Show All Steps Show Only Failed Steps Show Running Steps

| | | | |
|---|---|--------------------|--------|
| Execute command Stop on service Cloudera Management Service | Cloudera Management Service | Jun 19, 5:29:11 AM | 5.18s |
| Execute command Start on service Cloudera Management Service | Cloudera Management Service | Jun 19, 5:29:16 AM | 24.65s |

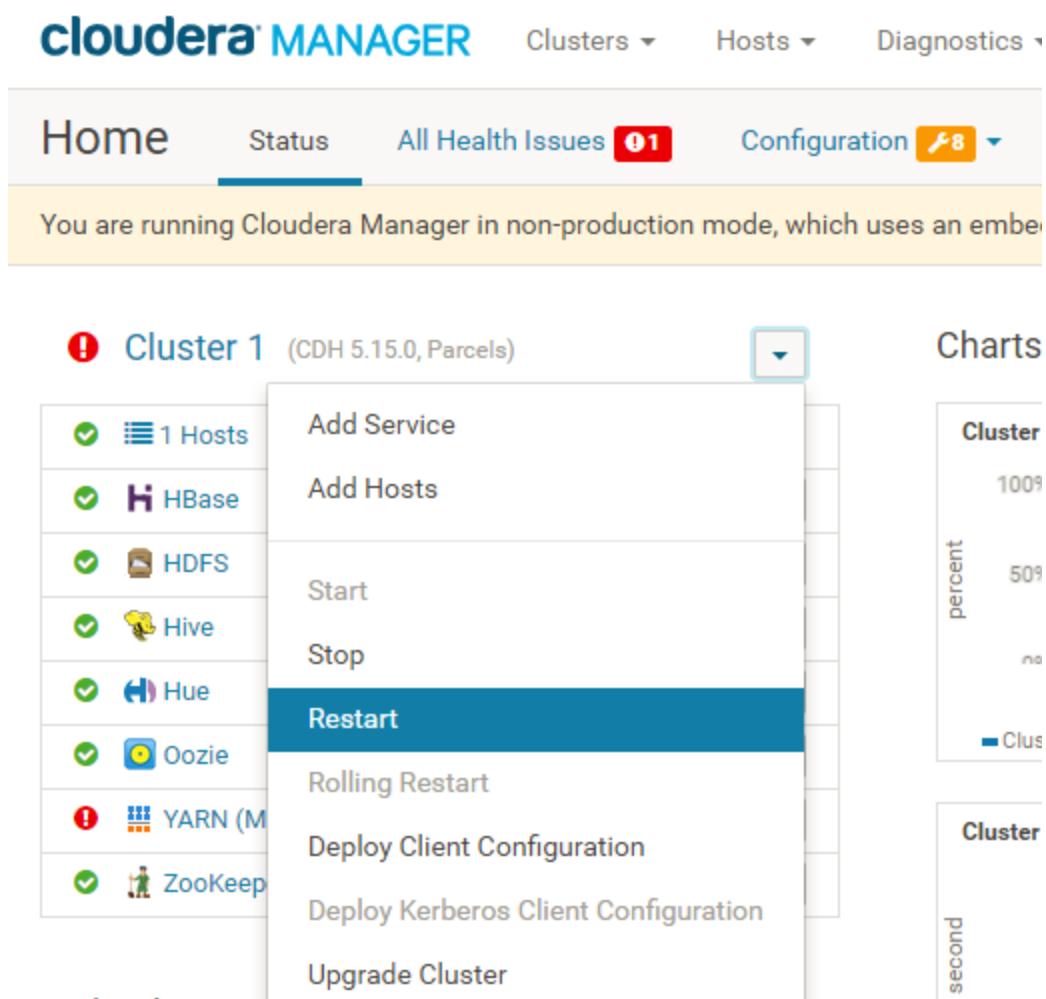
Close

Cloudera Man...

second

NO DATA

Now **Restart** the cluster1 services:

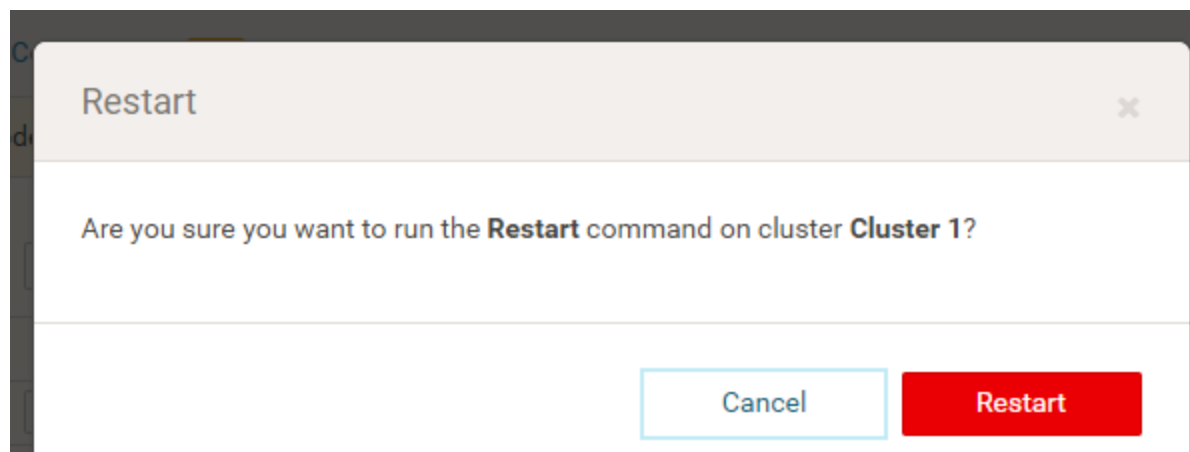


The screenshot shows the Cloudera Manager interface. At the top, there's a navigation bar with 'Home', 'Status', 'All Health Issues' (with a red badge showing '01'), and 'Configuration' (with a wrench icon and '8'). Below this, a yellow banner states: 'You are running Cloudera Manager in non-production mode, which uses an embedded...'.

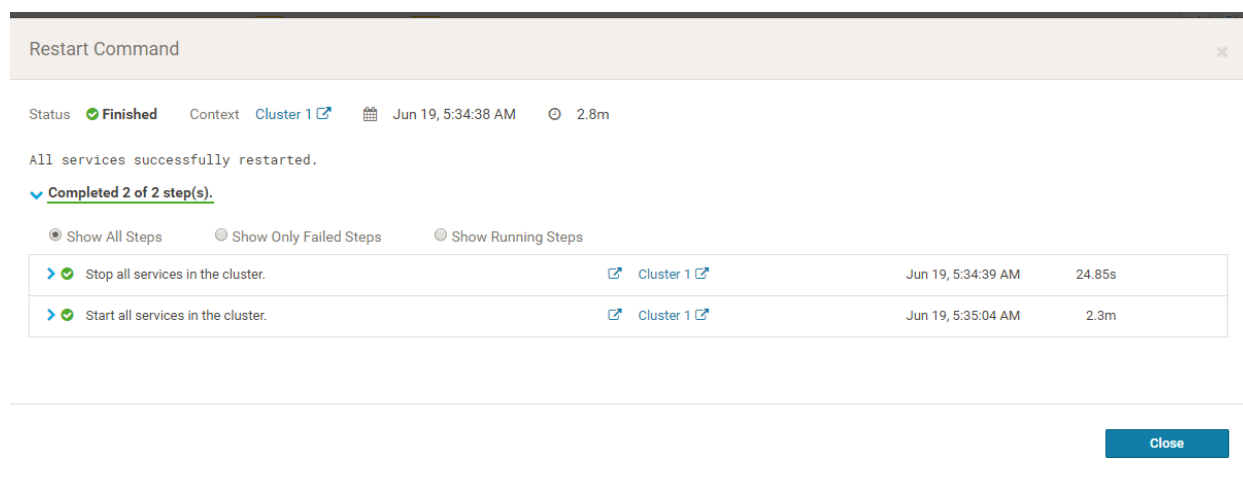
The main content area displays 'Cluster 1 (CDH 5.15.0, Parcels)'. To the left of the cluster name is a red exclamation mark icon. Below the cluster name is a list of services with their status: 1 Hosts, HBase, HDFS, Hive, Hue, Oozie, YARN (MapReduce), and ZooKeeper. A dropdown menu is open next to the cluster name, showing options: Add Service, Add Hosts, Start, Stop, **Restart** (highlighted in blue), Rolling Restart, Deploy Client Configuration, Deploy Kerberos Client Configuration, and Upgrade Cluster.

On the right side, there are two charts. The top chart is titled 'Cluster' and shows a 'percent' scale from 0 to 100. The bottom chart is also titled 'Cluster' and shows a 'second' scale.

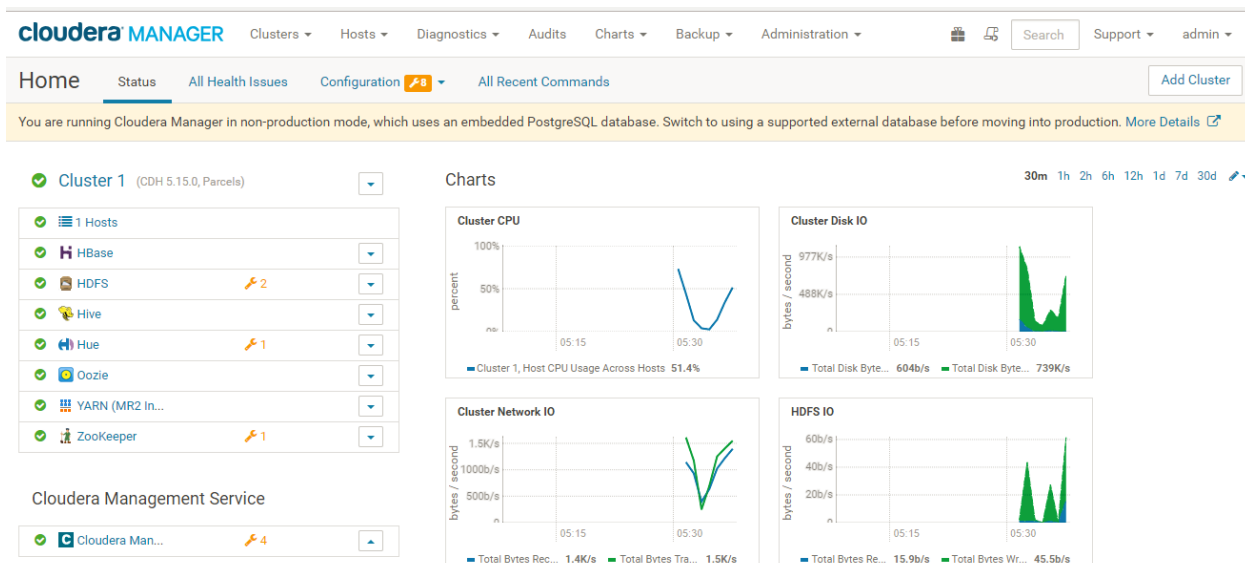
Click on **Restart**:



Click on '**Close**'.



All services in green.



Note: Once you have finished the module, terminate this instance. Even if the instance is stopped some charges are still applicable.