

Efforts Documentation (Unsuccessful Work)

Submitted by:

1. Dileep Kumar (ERP: 18255)
2. Muhammad Arsalan Mubeen (ERP: 23394)

In this document we have tried to create a multi-node docker swarm of Cloudera Hadoop (CDH). It includes all steps and screenshots that we have tried.

We used three Approaches use for Cloudera multi node cluster.

1. We try to use "Cloudera QuickStart CDH Docker Image" as a docker Service to move the container on overlay network from master to node.

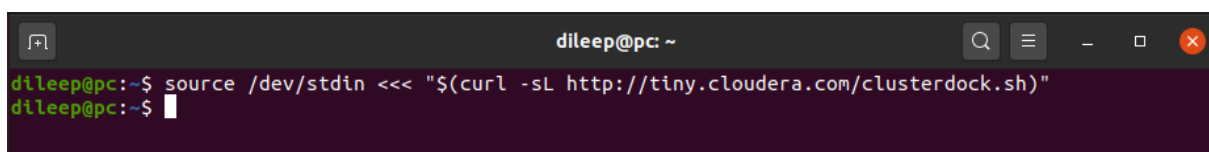
But we couldn't do that because it had too many tags inside Who run the Cloudera QuickStart CDH Docker and we don't know how to manage them on docker service command, also we search on Google, but we didn't find any related code for that.

2. Then we found "Cloudera/Clusterdock CDH Docker Image"

Overview: Clusterdock is a framework for creating Docker-based container clusters. Unlike regular Docker containers, which tend to run single processes and then exit once the process terminates, these container clusters are characterized by the execution of an in-it process in daemon mode. As such, the containers act more like "fat containers" or "light VMs," entities with accessible IP addresses which emulate standalone hosts.

For ease-of-use and portability, cluster dock itself is packaged in a Docker image and its binaries are executed by running containers from this image and specifying an action. This can be done by sourcing the clusterdock.sh helper script and then calling script of interest with the "clusterdock_run" command. As is always a good idea when executing code from the internet, examine the script to convince yourself of its safety, and then run:

```
source /dev/stdin <<< "$(curl -sL http://tiny.cloudera.com/clusterdock.sh)"
```

A terminal window titled 'dileep@pc: ~' with search, menu, and window control icons in the title bar. The terminal shows the command 'source /dev/stdin <<< "\$(curl -sL http://tiny.cloudera.com/clusterdock.sh)"' being entered and executed. The prompt changes from 'dileep@pc:~\$' to 'dileep@pc:~\$' after the command is run.

```
dileep@pc:~$ source /dev/stdin <<< "$(curl -sL http://tiny.cloudera.com/clusterdock.sh)"
dileep@pc:~$
```

If all you'd like is a two-node cluster (with default options being used for everything else), simply type:

```
clusterdock_run ./bin/start_cluster cdh
```

as we use now:

```
clusterdock_run ./bin/start_cluster -n bda_Cluster cdh --primary-node=pc --secondary-nodes=arsalan
```

```
dileep@pc: ~  
dileep@pc:~$ clusterdock_run ./bin/start_cluster -n bda-cluster cdh --primary-node=pc --secondary-node=arsalan  
INFO:clusterdock.topologies.cdh.actions:Pulling image docker.io/cloudera/clusterdock:cdh580_cm581_primary-node. This might take a little while...  
cdh580_cm581_primary-node: Pulling from cloudera/clusterdock  
  
3eaa9b70c44a: Pull complete  
99ba8e23f310: Pull complete  
c9c08e9a0d03: Pull complete  
7434a9a99daa: Pull complete  
d52d9baa0ee6: Pull complete  
00ca224ba661: Pull complete  
Digest: sha256:9feffbf5573262a6efbbb0a969efde890e63ced8a4ab3c9982f4f0dc607e429  
Status: Downloaded newer image for cloudera/clusterdock:cdh580_cm581_primary-node  
INFO:clusterdock.topologies.cdh.actions:Pulling image docker.io/cloudera/clusterdock:cdh580_cm581_secondary-node. This might take a little while...  
cdh580_cm581_secondary-node: Pulling from cloudera/clusterdock  
  
3eaa9b70c44a: Already exists  
99ba8e23f310: Already exists  
c9c08e9a0d03: Already exists  
7434a9a99daa: Already exists  
d52d9baa0ee6: Already exists  
f70deff0592f: Pull complete  
Digest: sha256:251778378b362adff4e93b99d423848216e4823965dabd1bd4c41d4bb4c79afcf  
Status: Downloaded newer image for cloudera/clusterdock:cdh580_cm581_secondary-node  
INFO:clusterdock.cluster:Network (bda-cluster) not present, creating it...  
INFO:clusterdock.cluster:Successfully setup network (name: bda-cluster).  
INFO:clusterdock.cluster:Successfully started arsalan.bda-cluster (IP address: 192.168.123.2).  
INFO:clusterdock.cluster:Successfully started pc.bda-cluster (IP address: 192.168.123.3).  
INFO:clusterdock.cluster:Started cluster in 10.37 seconds.  
INFO:clusterdock.topologies.cdh.actions:Changing server_host to pc.bda-cluster in /etc/cloudera-scm-agent/config.ini...  
INFO:clusterdock.topologies.cdh.actions:Restarting CM agents...  
cloudera-scm-agent is already stopped  
Starting cloudera-scm-agent: [ OK ]  
Stopping cloudera-scm-agent: [ OK ]  
Starting cloudera-scm-agent: [ OK ]
```

-n: name of cluster

--primary-node: we think it will deploy on different node but it just set-up the node name

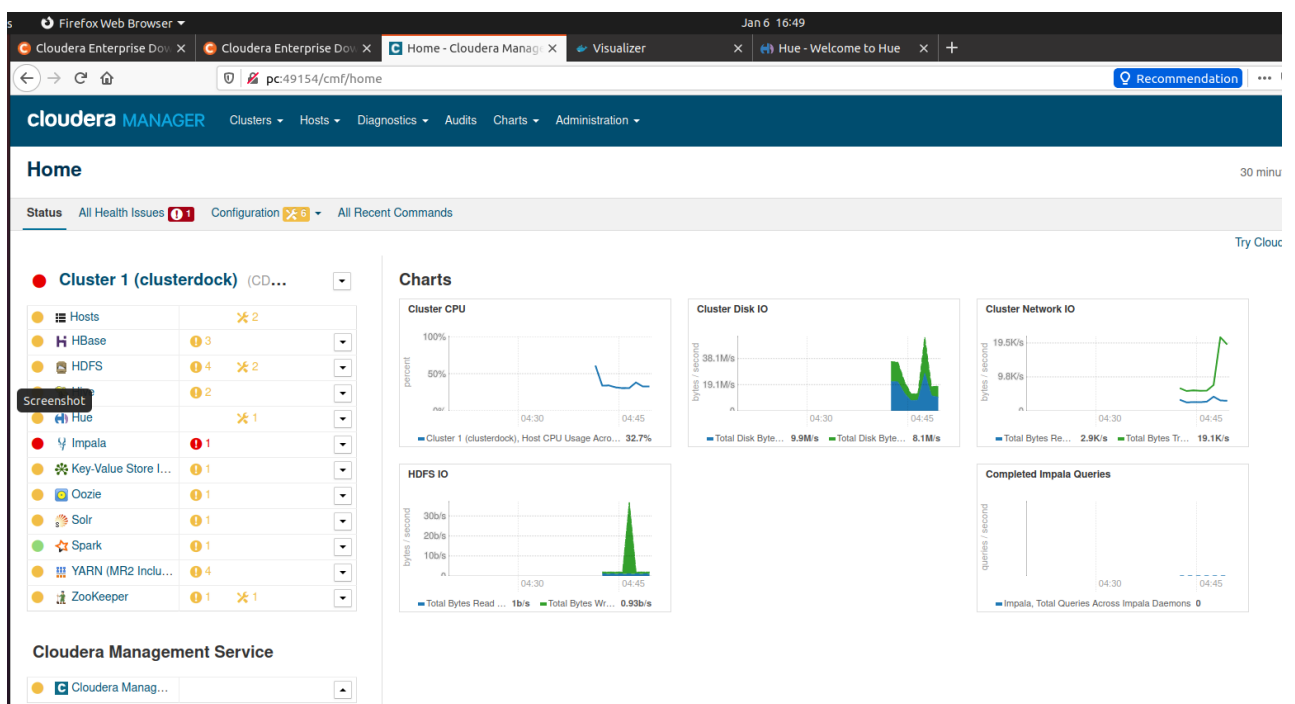
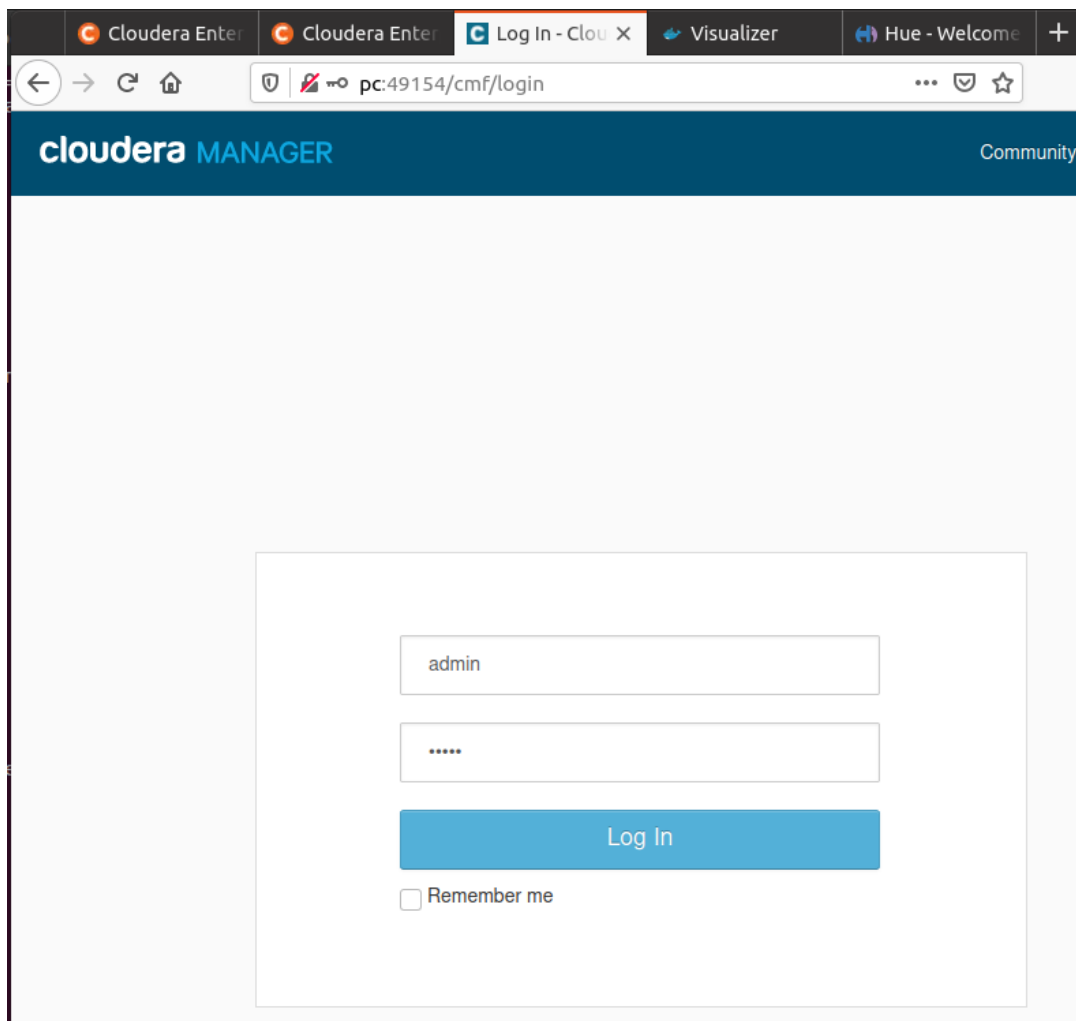
--secondary-nodes: same as above

By: docker container ls

```
dileep@pc:~$ docker container ls  
CONTAINER ID   IMAGE                                     COMMAND                  CREATED        STATUS  
dfd851f97e12   cloudera/clusterdock:cdh580_cm581_secondary-node   "/sbin/init"           3 hours ago   Up  
ac6c14af8018   cloudera/clusterdock:cdh580_cm581_primary-node     "/sbin/init"           3 hours ago   Up  
dileep@pc:~$
```

Clusterdock manages communication between containers through Docker's bridge networking driver. But it's not on overlay network, because it not a service (or we can't run this as a docker service) we will not be able to move the container from master to worker. It is run fine on same machine multi cluster docker,

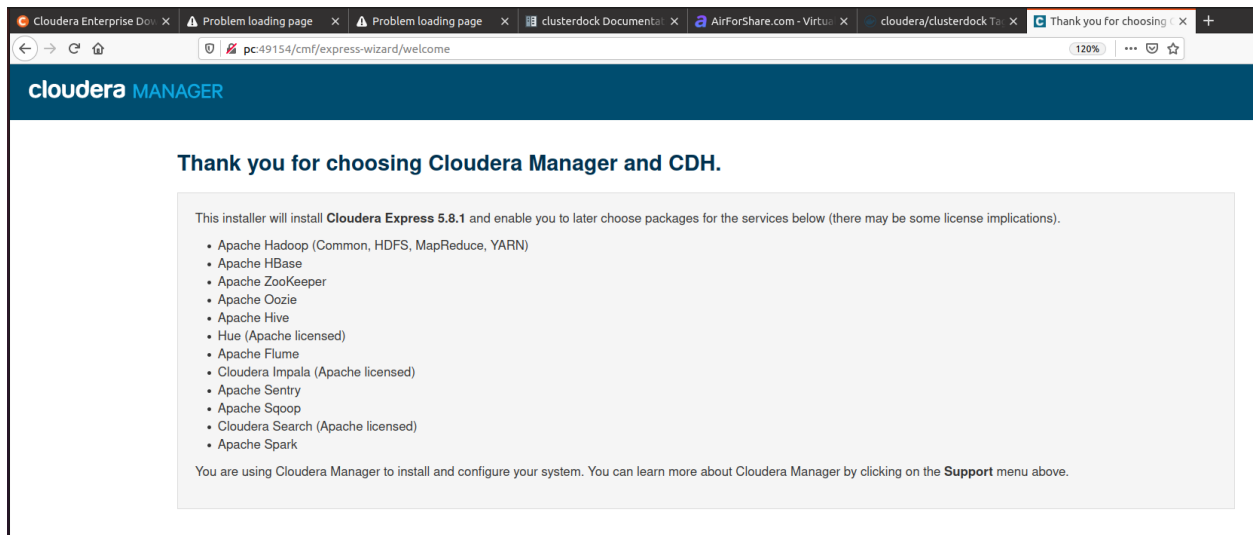
Once the cluster is running and the health of your CDH services is validated, we can access the cluster through the Cloudera Manager UI, also we can access HUE through address and port (the address and port number are shown at the end of the start-up process).



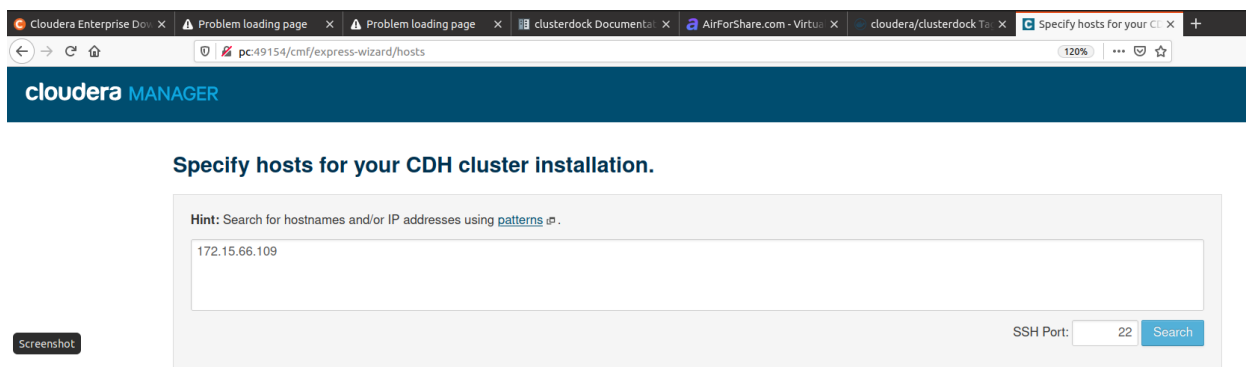
for being not using Cloudera/Clusterdock CDH Docker Image, it's a single node cluster docker. And we trying to figure out how to implement on overlay network. Other wise Cloudera and its service running fine.

3. Other thing in a Cloudera we are tried to create a new host cluster from Cloudera Manager UI in this Approaches we do,

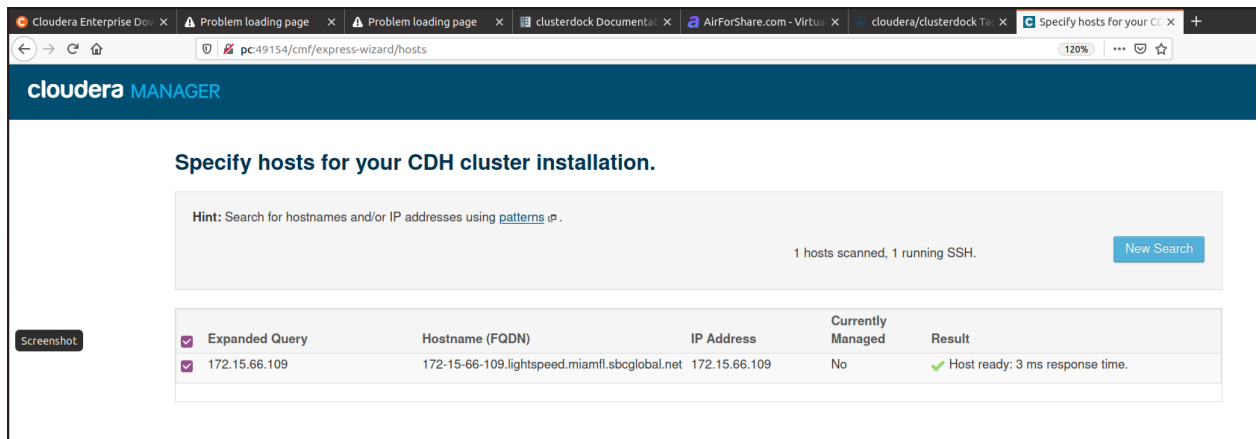
Select on “create a new host” button appear on right up corner



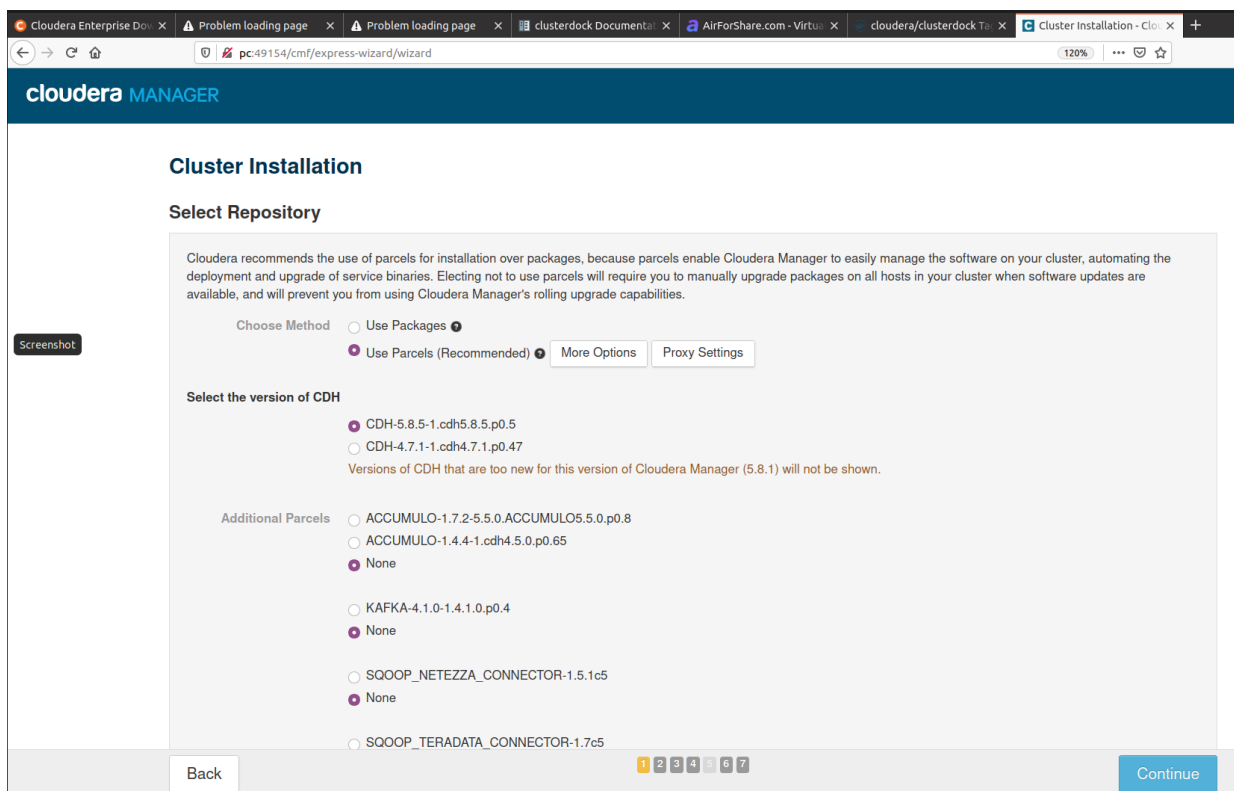
Host IP: i.e., worker IP



Find out Host IP successfully



Cluster required properties:



Select the Java JDK version:

Cloudera Enterprise Doc...Problem loading pageProblem loading pageclusterdock Documenta...AirForShare.com - Virtu...cloudera/clusterdock T...Cluster Installation - Clo...+

pc:49154/cmf/express-wizard/wizard#step=javaOptionsStep120%

clouderaMANAGER

Cluster Installation

JDK Installation Options

Screenshot

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 Table 1.1 (Commercial Features In Java SE Product Editions) of the Java SE documentation accessible at <http://www.oracle.com/technetwork/java/javase/documentation/index.html>

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

☐ Install Java Unlimited Strength Encryption Policy Files
Check this checkbox if local laws permit you to deploy unlimited strength encryption and you are running a secure cluster.

Back

1234567

Continue

Providing the SSH for connecting two nodes in a secure manner (when the Firewall is on)

For this we install SSH on both Linux environment

Cloudera Enterprise Doc...Problem loading pageProblem loading pageclusterdock Documenta...AirForShare.com - Virtu...cloudera/clusterdock T...Cluster Installation - Clo...+

pc:49154/cmf/express-wizard/wizard#step=hostCredentialsStep120%

clouderaMANAGER

Cluster Installation

Provide SSH login credentials.

Screenshot

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/prun 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:

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

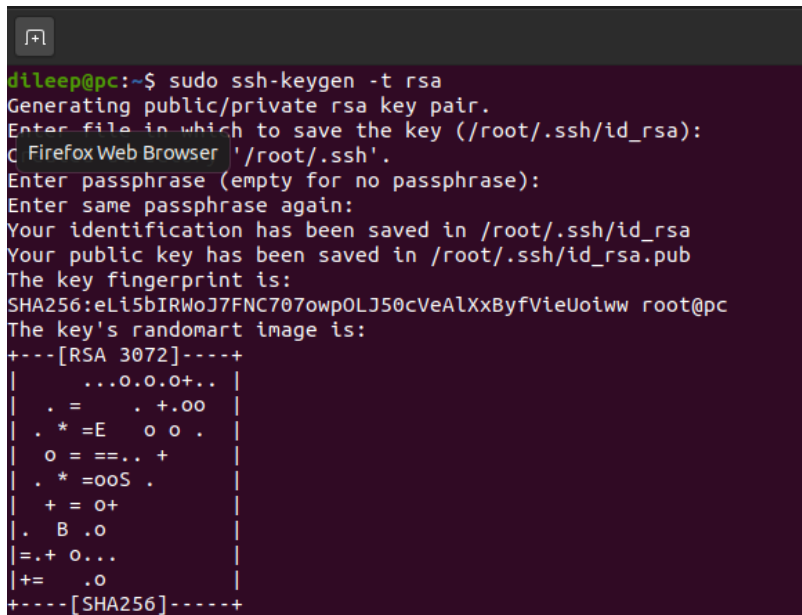
Back

1234567

Continue

We create a Specific SSH – key called rsa.pub file on root from master node

And we must specify every node on our cluster to use this specific SSH file for connecting the mater node.



```
dileep@pc:~$ sudo ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
c FirefoxWeb Browser '/root/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa
Your public key has been saved in /root/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:eLi5bIRWoJ7FNC707owp0LJ50cVeAlXxByfVieUoiww root@pc
The key's randomart image is:
+---[RSA 3072]---+
|      . . . O . O . O + . . |
|      . =      . + . 00      |
|      . * =E   O O .        |
|      O = == . . +          |
|      . * =00S .            |
|      + = O+                |
|      . B .O                |
|      = . + O . . .         |
|      +=      .O            |
+-----[SHA256]-----+
```

But unfortunately, we couldn't do that we search a lot on google but found nothing related to us.

We also go without SSH, but error says Cloudera need SSH to connect the other nodes in a cluster.

We also do to turn off the firewall and IPTABLES from bot node (master and worker) by doing it cant find the worker machine .