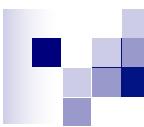


## **Module 2\_1**

# **Cluster Installation on Google Cloud**

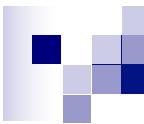
Thanachart Numnonda, Executive Director, IMC Institute  
Mr. Aekanun Thongtae, Big Data Consultant, IMC Institute

Thanisa Numnonda, Faculty of Information Technology,  
King Mongkut's Institute of Technology Ladkrabang

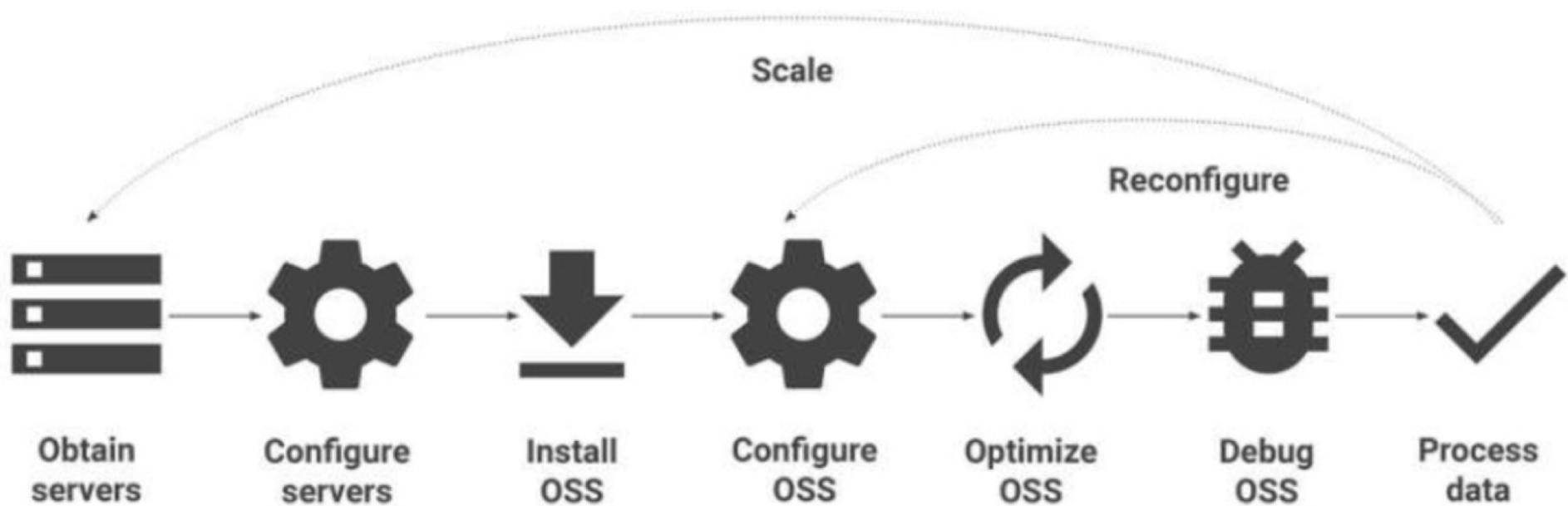


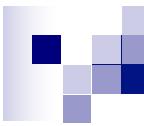
# Hadoop and Public Cloud Services

	On-Premise	GCP	AWS	Azure
Ingestion	Sqoop			
	Flume/Kafka	Cloud Sub/Pub	Kinesis	Event Hub
Storage	Hadoop HDFS	Cloud Storage	S3	Azure Data Lake
	RDBMS	Cloud SQL	RDS	Azure SQL
	NoSQL	Cloud Datastore/ Cloud BigTable	DynamoDB	Cosmos DB
	Data Warehouse	Google BigQuery	Redshift	SQL Data Warehouse
Processing	Hadoop/Spark	Cloud DataProc	EMR	HDInsight/ Azure Databrick
	Spark Streaming	Cloud Dataflow	Kinesis Analytics	Stream Analytics
	Hive	Google BigQuery	Athena	Data Lake Analytics
	Spark MLLib	ML Engine	SageMaker	Azure ML
Visualisation	Tableau, etc.	Google DataStudio	QuickSight	Power BI

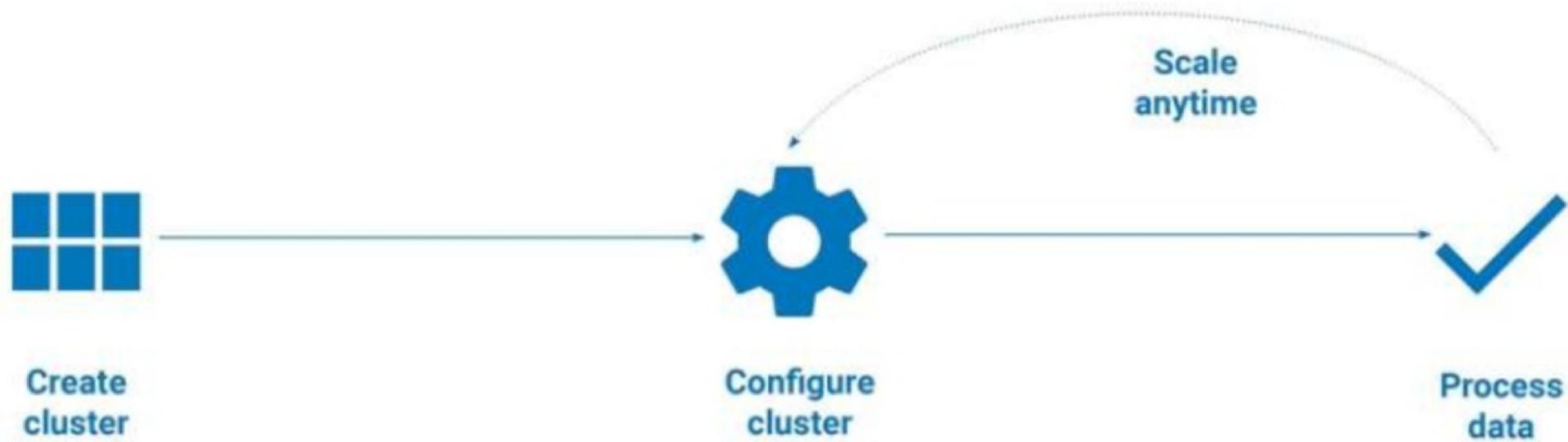


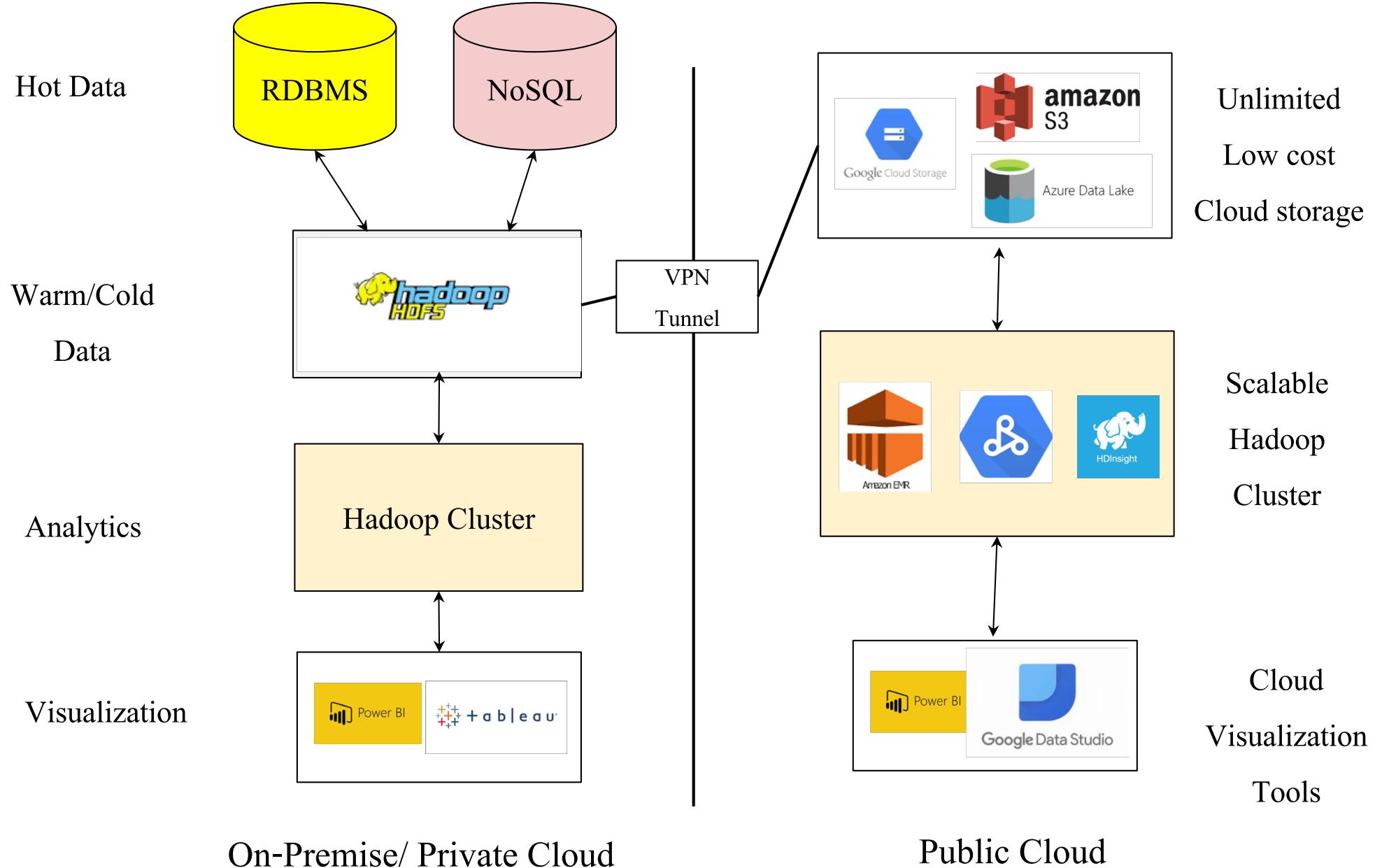
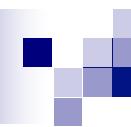
# Traditional Hadoop Clusters

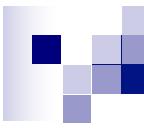




# Google Cloud Dataproc

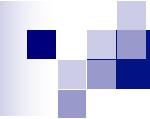






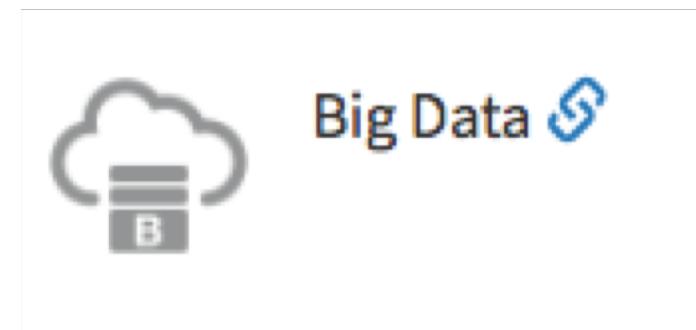
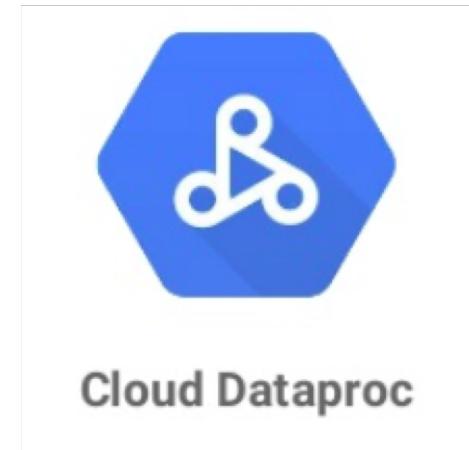
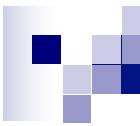
Google Cloud Platform

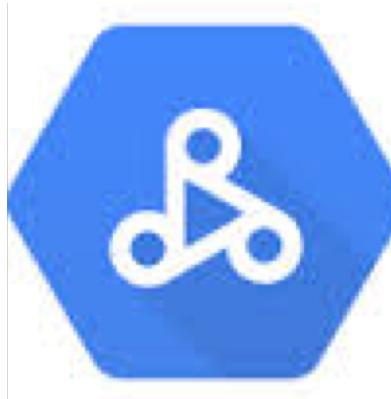




# Hadoop as a Service for Analytics

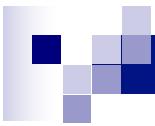
- Separate process layer from storage layer
- No need to install/admin a cluster
- Use processing tools: Hive, Spark, MLlib
- Start cluster only processing time
- Scalable CPU powers
- Pay only processing (CPU) time.





## Google Cloud DataProc

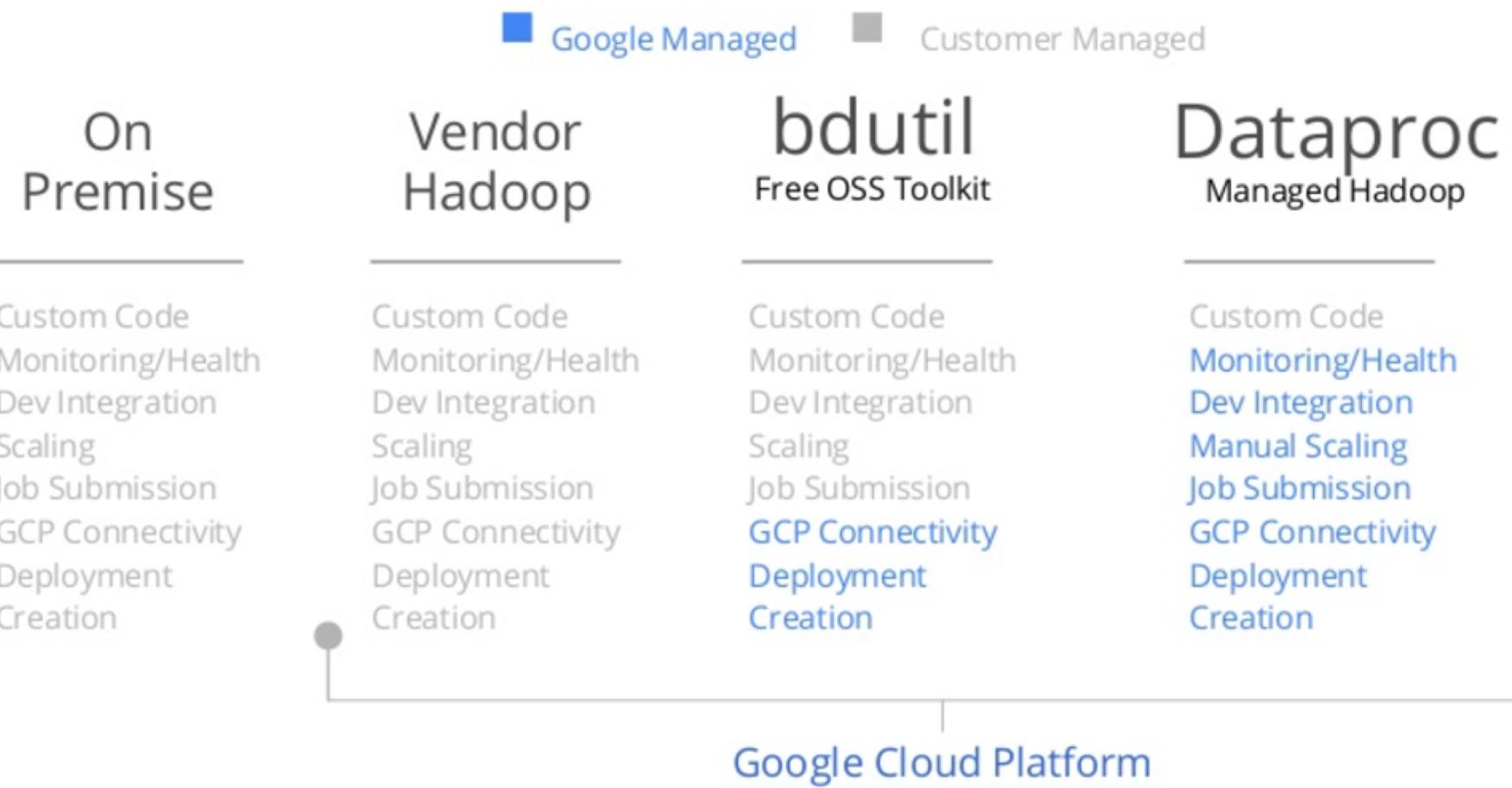
Google Cloud DataProc is a managed Spark and Hadoop service which is fast, easy to use, and low cost



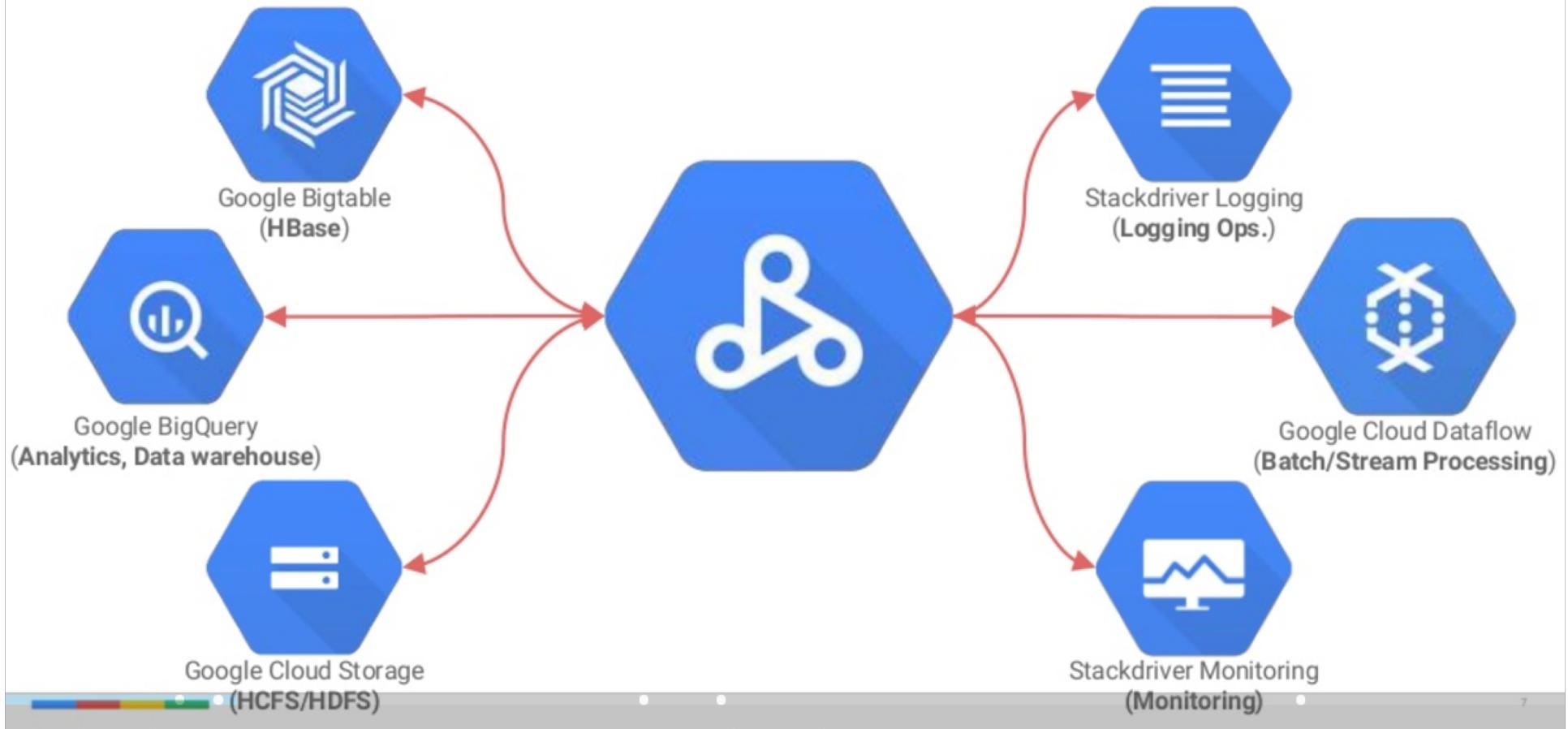
# Google Cloud Dataproc

- Managed an Apache Hadoop, Apache Spark, Apache Pig, and Apache Hive services, to easily process big data sets at low cost.
- Fast & Scalable Data Processing
- Affordable Pricing
- Open Source Ecosystem

# Running Hadoop on Google Cloud



# Where Cloud Dataproc fits into GCP



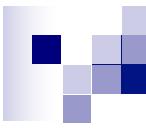
# Select Dataproc -> Clusters

The screenshot shows the Google Cloud Platform dashboard. On the left, there's a sidebar with various services: Home, Pins appear here, Endpoints, BIG DATA (BigQuery, Pub/Sub), Dataproc, Dataflow, ML Engine, and Genomics. The 'Dataproc' item is highlighted with a red box and has a red arrow pointing to it from the top-left. In the main content area, there's a chart titled 'API APIs Requests (requests/sec)' showing a fluctuating blue line. Below the chart, there are two buttons: 'Clusters' and 'Jobs'. A red arrow points to the 'Clusters' button. To the right of the chart, there's a section for 'Google Cloud Platform status' indicating 'All services normal' and a link to 'Go to Cloud status dashboard'. At the bottom, there's a 'Billing' section showing '\$0.42' and 'Approximate charges so far this month', with a link to 'View detailed charges'.



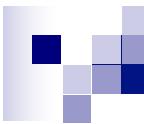
# Click Create Cluster

The screenshot shows the Google Cloud Platform interface for Cloud Dataproc. The top navigation bar includes the Google Cloud logo, 'Google Cloud Platform' text, a dropdown menu showing 'clouderaclust...', a search bar, and various status icons. On the left, a sidebar for 'Cloud Dataproc' lists 'Clusters' (which is selected and highlighted in blue) and 'Jobs'. The main content area is titled 'Clusters' and contains a large callout box. The callout box has a title 'Cloud Dataproc Clusters' and a description: 'Google Cloud Dataproc lets you provision Apache Hadoop clusters and connect to underlying analytic data stores.' It also features a 'Create your first cluster to get started.' message and a prominent blue 'Create cluster' button.



## Using the following configuration

- Name the cluster.
- Choose the region
- Change the machine type of both the Master and the Worker nodes to n1-standard-2
- Select disk size to 80 GB



Google Cloud Platform My First Project

Dataproc Create a cluster

Clusters thanisa-dataproc

Jobs

Workflows

Name: thanisa-dataproc

Region: us-east1 Zone: us-east1-c

Cluster mode: Standard (1 master, N workers)

Master node: Contains the YARN Resource Manager, HDFS NameNode, and all job drivers

Machine type: 2 vCPUs, 7.5 GB memory, Customize

Upgrade your account to create instances with up to 96 cores

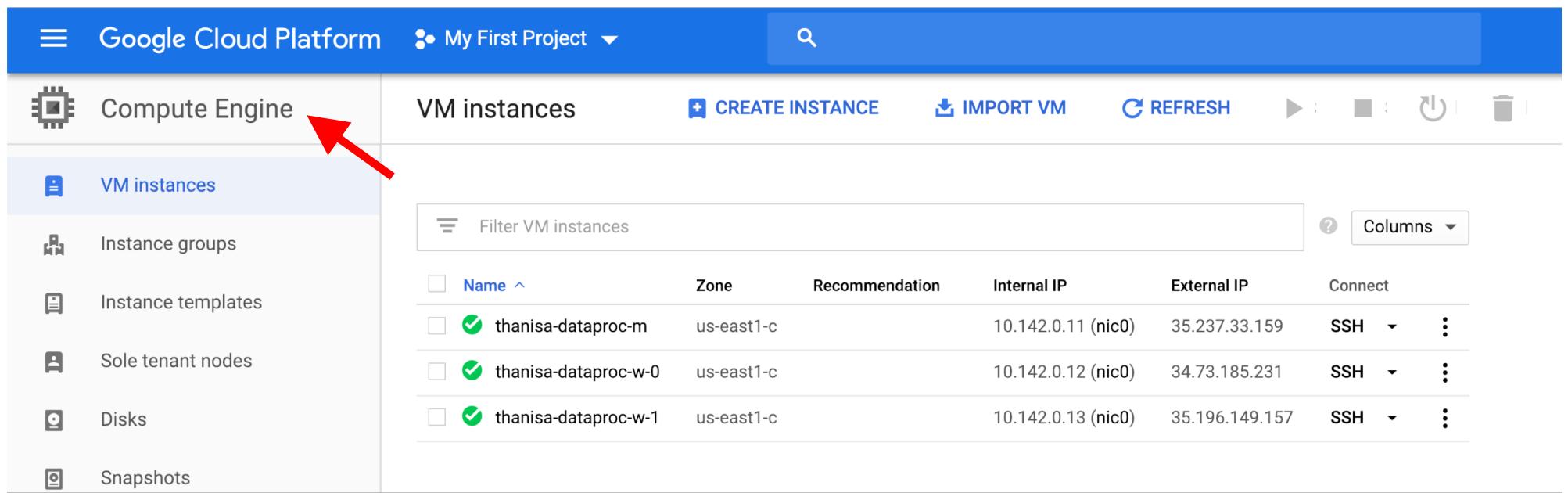
Primary disk size (minimum 10 GB): 80 GB

Primary disk type: Standard persistent disk

Worker nodes: Each contains a YARN NodeManager and a HDFS DataNode. The HDFS replication factor is 2.

Machine type:

## Select the Dataproc master instance from Compute Engine



The screenshot shows the Google Cloud Platform Compute Engine interface. The left sidebar has 'Compute Engine' selected, indicated by a red arrow. The main area displays 'VM instances' with three entries:

Name	Zone	Internal IP	External IP	Connect
thanisa-dataproc-m	us-east1-c	10.142.0.11 (nic0)	35.237.33.159	SSH
thanisa-dataproc-w-0	us-east1-c	10.142.0.12 (nic0)	34.73.185.231	SSH
thanisa-dataproc-w-1	us-east1-c	10.142.0.13 (nic0)	35.196.149.157	SSH

# Start SSH from the master node

The screenshot shows the Google Cloud Platform Compute Engine interface. On the left sidebar, under 'Compute Engine', the 'VM instances' option is selected. In the main area, the 'VM instances' tab is active, displaying a list of three instances: 'thanisa-dataproc-m', 'thanisa-dataproc-w-0', and 'thanisa-dataproc-w-1'. The first instance, 'thanisa-dataproc-m', is highlighted with a red box. To its right is a 'Connect' button with an 'SSH' icon, which has a red arrow pointing to it. A dropdown menu is open from this button, listing four options: 'Open in browser window', 'Open in browser window on c', 'View gcloud command', and 'Use another SSH client'. The 'Internal IP' and 'External IP' columns also show the respective IP addresses for each instance.

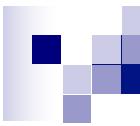
Name	Zone	Internal IP	External IP	Connect
thanisa-dataproc-m	us-east1-c	10.142.0.11 (nic0)	35.237.33.159	<b>SSH</b> <span style="border: 1px solid #ccc; padding: 2px;">▼</span>
thanisa-dataproc-w-0	us-east1-c	10.142.0.12 (nic0)	34.73.185.231	<b>SSH</b>
thanisa-dataproc-w-1	us-east1-c	10.142.0.13 (nic0)	35.196.149.157	<b>SSH</b>



```
thanisa@thanisa-dataproc-m: ~
  https://ssh.cloud.google.com/projects/healthy-name-230801/zones/us-east1-c/instances/thanisa-dataproc-m?authuser=0&hl=en...
Connected, host fingerprint: ssh-rsa 0 AC:0D:27:FB:0B:32:C8:89:07:1E:7A:70:8
:76:7B:ED:B5:9F:F7:F3:D4:09:97:BF:B1:86:A8:CE:76:55:F3
Linux thanisa-dataproc-m 4.9.0-8-amd64 #1 SMP Debian 4.9.130-2 (2018-10-27)
64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
thanisa@thanisa-dataproc-m:~$
```



## Stop and Delete the VM instances and Cluster

Google Cloud Platform - My First Project

Compute Engine - VM instances

CREATE INSTANCE IMPORT VM REFRESH

VM instances

Instance groups Instance templates Sole tenant nodes Disks

Filter VM instances Columns

Name	Zone	Recommendation	Internal IP	External IP	Connect
thanisa-dataproc-m	us-east1-c		10.142.0.11 (nic0)	35.237.33.159	SSH
thanisa-dataproc-w-0	us-east1-c		10.142.0.12 (nic0)	34.73.185.231	SSH
thanisa-dataproc-w-1	us-east1-c		10.142.0.13 (nic0)	35.196.149.157	SSH

Stop (red arrow) Delete (red arrow)

Google Cloud Platform - My First Project

Dataproc - Clusters

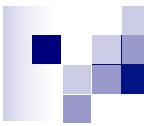
CREATE CLUSTER REFRESH DELETE REGIONS

Clusters Jobs Workflows

Search clusters, press Enter

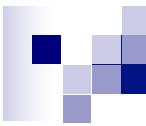
Name	Region	Zone	Total worker nodes	Scheduled deletion	Cloud Storage staging bucket
thanisa-dataproc	us-east1	us-east1-c	2	Off	dataproc-a8e2a6f3-861a-4769-ae41-3e7dc3710c51-us-east1

Delete (red arrow)



# Google Cloud Shell

**Google Cloud Shell is a shell environment for managing resources hosted on Google Cloud Platform. It is a micro VM and it is free of charge.**



# Launching Cloud Shell

- Select or create a Cloud Platform project.
- Click the Activate Google Cloud Shell button at the top of the Google Cloud Platform Console.
- A Cloud Shell session opens inside a new frame at the bottom of the console and displays a command-line prompt.

## Launch Cloud Shell to Launch Dataproc using gcloud command

The screenshot shows the Google Cloud Platform dashboard for the project "My First Project". A red arrow points to the "Activate Cloud Shell" button in the top right corner of the dashboard header. The dashboard includes sections for Project info, API APIs, Google Cloud Platform status, and Billing. The Cloud Shell terminal at the bottom displays a welcome message and the current project ID: "healthy-name-230801".

Unlock more of Google Cloud Platform by upgrading now (\$295.25 credit and 361 days left in your free trial). **DISMISS** **UPGRADE**

Google Cloud Platform My First Project

DASHBOARD ACTIVITY

Activate Cloud Shell CUSTOMIZE

Project info

- Project name: My First Project
- Project ID: healthy-name-230801
- Project number: 402738421939

API APIs

Requests (requests/sec)

Time	Requests (requests/sec)
00:00:00	0.035
00:00:10	0.030
00:00:20	0.025

Google Cloud Platform status

All services normal

→ Go to Cloud status dashboard

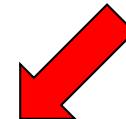
Billing

Welcome to Cloud Shell! Type "help" to get started.  
Your Cloud Platform project in this session is set to **healthy-name-230801**.  
Use "gcloud config set project [PROJECT\_ID]" to change to a different project.  
thanisa@cloudshell:~ (healthy-name-230801)\$

Type the following command to Launch Dataproc using gcloud command

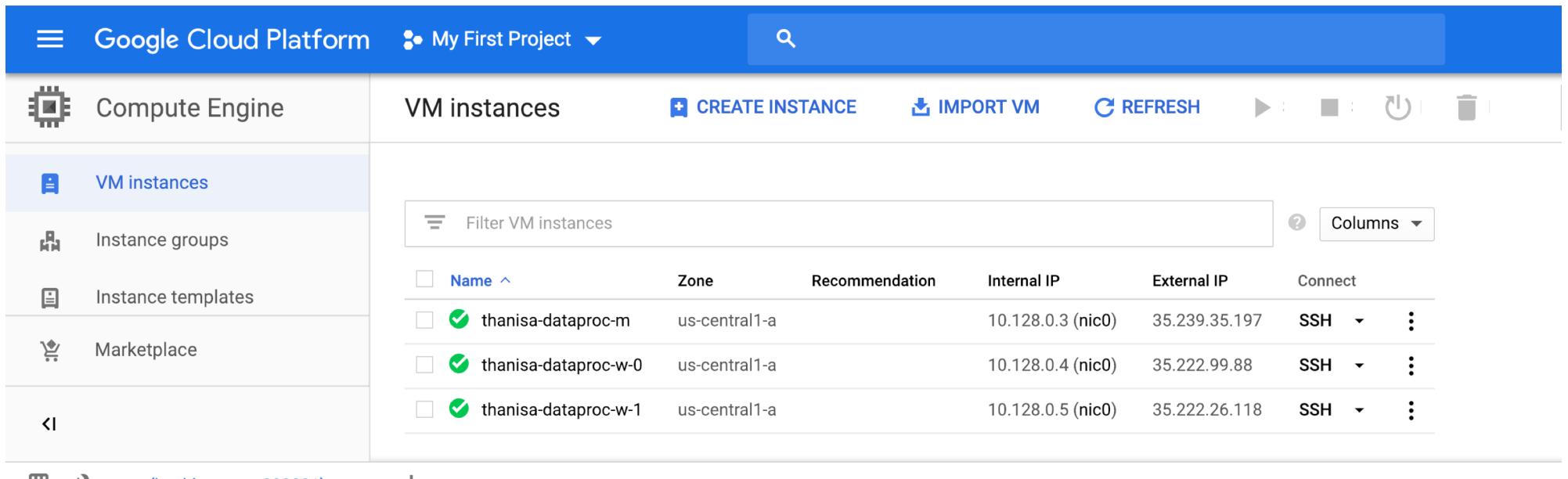
- gcloud dataproc clusters create thanisa-dataproc \  
--zone=us-central1-a \  
--master-machine-type=n1-standard-2 \  
--worker-machine-type=n1-standard-2

*Change to your name*



```
Welcome to Cloud Shell! Type "help" to get started.  
Your Cloud Platform project in this session is set to healthy-name-230801.  
Use "gcloud config set project [PROJECT_ID]" to change to a different project.  
thanisa@cloudshell:~ (healthy-name-230801)$ gcloud dataproc clusters create thanisa-dataproc  
> --zone us-central1-a \  
> --master-machine-type=n1-standard-2 \  
> --worker-machine-type=n1-standard-2  
Waiting on operation [projects/healthy-name-230801/regions/global/operations/0ec133a9-7f59-33  
21b].  
Waiting for cluster creation operation...::
```

# Stop and Delete the VM instances and Cluster



Google Cloud Platform My First Project

Compute Engine VM instances

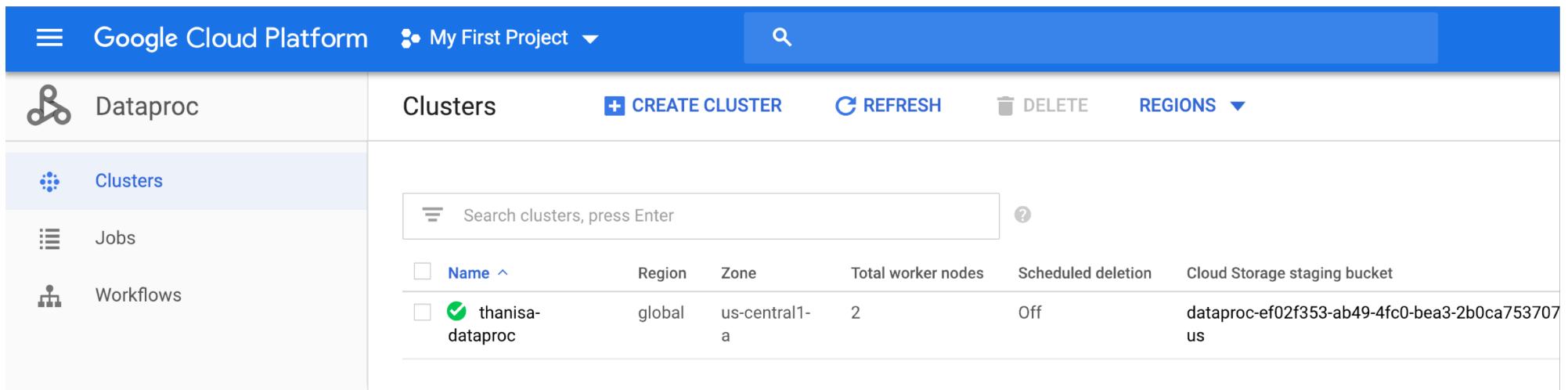
CREATE INSTANCE IMPORT VM REFRESH

VM instances

Filter VM instances

Columns

Name	Zone	Recommendation	Internal IP	External IP	Connect
thanisa-dataproc-m	us-central1-a		10.128.0.3 (nic0)	35.239.35.197	SSH
thanisa-dataproc-w-0	us-central1-a		10.128.0.4 (nic0)	35.222.99.88	SSH
thanisa-dataproc-w-1	us-central1-a		10.128.0.5 (nic0)	35.222.26.118	SSH



Google Cloud Platform My First Project

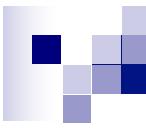
Dataproc Clusters

CREATE CLUSTER REFRESH DELETE REGIONS

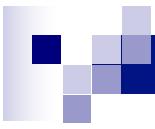
Clusters

Search clusters, press Enter

Name	Region	Zone	Total worker nodes	Scheduled deletion	Cloud Storage staging bucket
thanisa-dataproc	global	us-central1-a	2	Off	dataproc-ef02f353-ab49-4fc0-bea3-2b0ca753707us



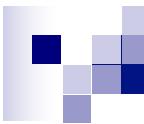
# Cloudera Cluster Installation on Google Cloud



# Cloudera Hadoop Cluster

This lab will use 4 Google cloud instances to install a Cloudera cluster using the following features:

- Ubuntu Server 14.04 LTS
- 4 vCPU, 15 GB memory
- 100 GB SSD



# Create Google Cloud Project

Unlock more of Google Cloud Platform by upgrading now (\$300.00 credit and 364 days left in your free trial). [DISMISS](#) [UPGRADE](#)

Google Cloud Platform [My First Project](#) [SEARCH](#) [DASHBOARD](#) [ACTIVITY](#) [CUSTOMIZE](#)

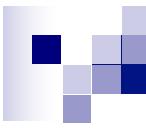
Pins appear here [?](#)

Marketplace [Billing](#) [APIs & Services](#) [Support](#) [IAM & admin](#) [Getting started](#)

Select a project [NEW PROJECT](#)

Search projects and folders

RECENT	ALL
Name	ID
My First Project <a href="#">?</a>	molten-rex-228921



# Create Google Cloud Project (Cont.)

The screenshot shows the 'New Project' dialog box from the Google Cloud Platform. At the top, there's a navigation bar with the Google Cloud Platform logo, a menu icon, the text 'Google Cloud Platform', and a search icon. Below the navigation bar, the title 'New Project' is displayed. A warning message in a box states: '⚠ You have 11 projects remaining in your quota. Request an increase or delete projects.' with a 'Learn more' link and a 'MANAGE QUOTAS' button. The 'Project Name \*' field contains 'clouderacluster' and has a red arrow pointing to it from the left. Below the project name, the 'Project ID' is shown as 'clouderacluster-228921. It cannot be changed later.' with an 'EDIT' link. The 'Location \*' section shows 'No organization' selected, with a 'BROWSE' link and a note 'Parent organization or folder'. At the bottom of the dialog are two buttons: a large blue 'CREATE' button with a red arrow pointing to it from the left, and a smaller 'CANCEL' button.

☰ Google Cloud Platform

New Project

⚠ You have 11 projects remaining in your quota. Request an increase or delete projects.  
[Learn more](#)

[MANAGE QUOTAS](#)

Project Name \*  ?

Project ID: clouderacluster-228921. It cannot be changed later. [EDIT](#)

Location \*  [BROWSE](#)

Parent organization or folder

[CREATE](#) [CANCEL](#)

# Select Google Cloud Project to **clouderacluster**

The screenshot shows the Google Cloud Platform dashboard. At the top, there is a blue header bar with the text "Google Cloud Platform" on the left, a dropdown menu for the project "clouderacluster" in the center, and several icons on the right. A red arrow points to the project dropdown menu. Below the header is a navigation bar with "Home" and "DASHBOARD" (which is underlined) on the left, and "ACTIVITY" on the right. To the right of the navigation bar is a "CL" icon. On the left side, there is a sidebar with various links: "Marketplace", "Billing", "APIs & Services", "Support", "IAM & admin", and "Getting started". A callout bubble with a blue border and dashed corners appears over the "Marketplace" link, containing the text "Pins appear here" and a question mark icon. To the right of the sidebar is the main dashboard area. It features a card titled "info" with a list of items: "time", "luster", "luster-228921", "number", "5142", and "select settings". Next to it is a chart titled "API APIs" showing "Requests (requests/sec)" over time. The chart has a note: "⚠ No data is available for the selected time frame." The x-axis shows "4 AM" and "4:30". On the far right, there are two more cards: one for "Google Cloud" and another for "Error Reporting".

# Select Compute Engine

The screenshot shows the Google Cloud Platform dashboard. On the left, there's a sidebar with sections for Home, COMPUTE, App Engine, Compute Engine (which is highlighted with a red arrow), Kubernetes Engine, and Cloud Functions. Below these are sections for STORAGE (Bigtable, Datastore, Firestore). A dashed blue box surrounds the Home and Pins appear here area. A search bar and a project selector are at the top right. A sidebar menu is open over the main content, listing VM instances, Instance groups, Instance templates, Sole tenant nodes, Disks, Snapshots, Images, TPUs, Committed use discounts, Metadata, Health checks, Zones, Network endpoint groups, Operations, Security scans, and Settings. To the right of the sidebar, there's a section titled 'API APIs' showing 'Requests (requests/sec)' over time, with a single data point at 4 AM showing 'Requests: 0.017'. A button 'Go to APIs overview' is at the bottom of this section. The URL at the bottom of the page is <https://console.cloud.google.com/compute?project=clouderacluster-228921>.



# Create VM instances

Google Cloud Platform clouderacluster

Compute Engine VM instances

VM instances

Instance groups

Instance templates

Sole tenant nodes

Disks

Snapshots

Images

TPUs

Committed use discounts

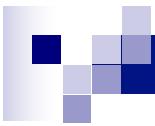
Compute Engine  
VM instances

Compute Engine lets you use virtual machines that run on Google's infrastructure. Create micro-VMs or larger instances running Debian, Windows, or other standard images. Create your first VM instance, import it using a migration service, or try the quickstart to build a sample app.

Create or Import or Take the quickstart



A red arrow points to the 'Create' button in the 'Compute Engine VM instances' panel.



## Create VM instances (Cont.)

This lab will create 4 instances name as follows:

- cloudera1
- cloudera2
- cloudera3
- cloudera4

# Create cloudera1 instance

Name

Region  Zone

Machine type  
Customize to select cores, memory and GPUs.

4 vCPUs  15 GB memory

Upgrade your account to create instances with up to 96 cores

Container  Deploy a container image to this VM instance. [Learn more](#)

Boot disk

New 100 GB standard persistent disk  
 Image Ubuntu 14.04 LTS

Boot disk type  Size (GB)

Select Cancel

Boot disk

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

[OS Images](#) [Application images](#) [Custom images](#) [Snapshots](#) [Existing disks](#)

- Debian GNU/Linux 8 (jessie)  
amd64 built on 2016-10-27
- CentOS 6  
x86\_64 built on 2016-10-27
- CentOS 7  
x86\_64 built on 2016-10-27
- CoreOS alpha 1235.0.0  
amd64-usr published on 2016-11-17
- CoreOS beta 1192.2.0  
amd64-usr published on 2016-11-02
- CoreOS stable 1185.3.0  
amd64-usr published on 2016-11-01
- Ubuntu 12.04 LTS  
amd64 precise image built on 2016-11-09
- Ubuntu 14.04 LTS  
amd64 trusty image built on 2016-11-09
- Ubuntu 16.04 LTS  
amd64 xenial image built on 2016-11-15
- Ubuntu 16.10  
amd64 yakkety image built on 2016-10-20
- Red Hat Enterprise Linux 6  
x86\_64 built on 2016-10-27
- Red Hat Enterprise Linux 7  
x86\_64 built on 2016-10-27

# Create **cloudera1** instance (Cont.)

**Identity and API access** ?

**Service account** ?

Compute Engine default service account

**Access scopes** ?

Allow default access  
 Allow full access to all Cloud APIs  
 Set access for each API

**Firewall** ?

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

**Create** **Cancel**





# Cloudera instance will run

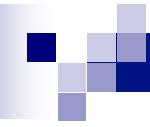
The screenshot shows the Google Cloud Platform Compute Engine interface. On the left sidebar, under the 'Compute Engine' section, 'VM instances' is selected. The main area displays a list of VM instances. A single instance, 'cloudera1', is highlighted with a red box. The instance details are as follows:

Name	Zone	Recommendation	Internal IP	External IP	Connect
cloudera1	us-east1-b		10.142.0.2 (nic0)	35.237.167.107	SSH

# Upgrade Account and Repeat the Steps for 3 More Instances

The screenshot shows two views of the Cloudera Cluster Management interface. The top view is the 'Quotas' section, which includes filters for Quota type (All quotas), Service (Compute Engine API), Metric (All metrics), and Location (All locations). A message states: 'You can't request an increase until you upgrade your free trial account.' An 'Upgrade account' button is highlighted with a red arrow. The bottom view is the 'VM instances' section, featuring a 'CREATE INSTANCE' button, 'IMPORT VM' button, 'REFRESH' button, and a 'START' button. A 'Filter VM instances' input field is present. A red box highlights the list of four VM instances: cloudera1, cloudera2, cloudera3, and cloudera4, all located in the us-east1-b zone. Each instance has a checked green checkmark icon in the first column.

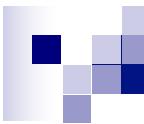
	Name	Zone	Recommendation	Internal IP	External IP	Connect
<input type="checkbox"/>	cloudera1	us-east1-b		10.142.0.2 (nic0)	35.237.167.107	SSH
<input type="checkbox"/>	cloudera2	us-east1-b		10.142.0.3 (nic0)	35.227.54.221	SSH
<input type="checkbox"/>	cloudera3	us-east1-b		10.142.0.4 (nic0)	35.231.80.33	SSH
<input type="checkbox"/>	cloudera4	us-east1-b		10.142.0.5 (nic0)	104.196.11.81	SSH



# Prepare the virtual servers for installing Cloudera cluster

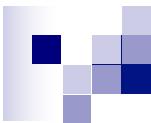
For each server, we need to do the following steps

- Change the root password
- Change the SSH configuration file & generate a new SSH key
- Give root permission access for all servers
- Change the sysctl vm.swappiness



# Click SSH

<input type="checkbox"/>	Name ^	Zone	Recommendation	Internal IP	External IP	Connect
<input type="checkbox"/>	cloudera1	us-east1-b		10.142.0.2 (nic0)	35.237.167.107 ↗	SSH <span style="border: 1px solid #ccc; padding: 0 5px;">▼</span> :
<input type="checkbox"/>	cloudera2	us-east1-b		10.142.0.3 (nic0)	35.227.54.221 ↗	SSH
<input type="checkbox"/>	cloudera3	us-east1-b		10.142.0.4 (nic0)	35.231.80.33 ↗	SSH
<input type="checkbox"/>	cloudera4	us-east1-b		10.142.0.5 (nic0)	104.196.11.81 ↗	SSH



# Connect to the instance

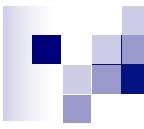
```
thanisabigdata@cloudera1: ~
① 🔒 https://ssh.cloud.google.com/projects/clouderacluster-228921/zones/us-east1-b/instances/cloudera1?authuser=0&hl=en_US& ...
http://www.ubuntu.com/business/services/cloud
0 packages can be updated.
0 updates are security updates.

Your Hardware Enablement Stack (HWE) is supported until April 2019.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

thanisabigdata@cloudera1:~$ 
```



## 1) Change the root password

For each server, we need to change the root password using the following command

```
$ sudo -i
```

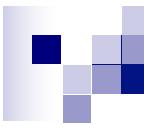
```
# passwd [ENTER]
```

**<< bigdata >>**

(Note: you need the same password for all servers)

The screenshot shows a terminal window titled 'root@cloudera1: ~'. The URL bar indicates the session is via SSH to 'https://ssh.cloud.google.com/projects/clouderacluster-228921/zones/us-east1-b/'. The terminal content shows the following sequence of commands:

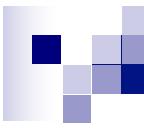
```
thanisabigdata@cloudera1:~$ sudo -i
root@cloudera1:~# passwd
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
root@cloudera1:~# █
```



# Repeat the same process for all instances

The image displays three separate terminal windows, each showing a root shell session on a Cloudera instance. The windows are arranged vertically. Each window has a title bar indicating the host name (cloudera2, cloudera3, or cloudera4) and a root prompt. The content of each window shows the command `sudo -i` being run, followed by the `passwd` command, and the user being prompted to enter a new UNIX password. The message `passwd: password updated successfully` is displayed at the end of each session.

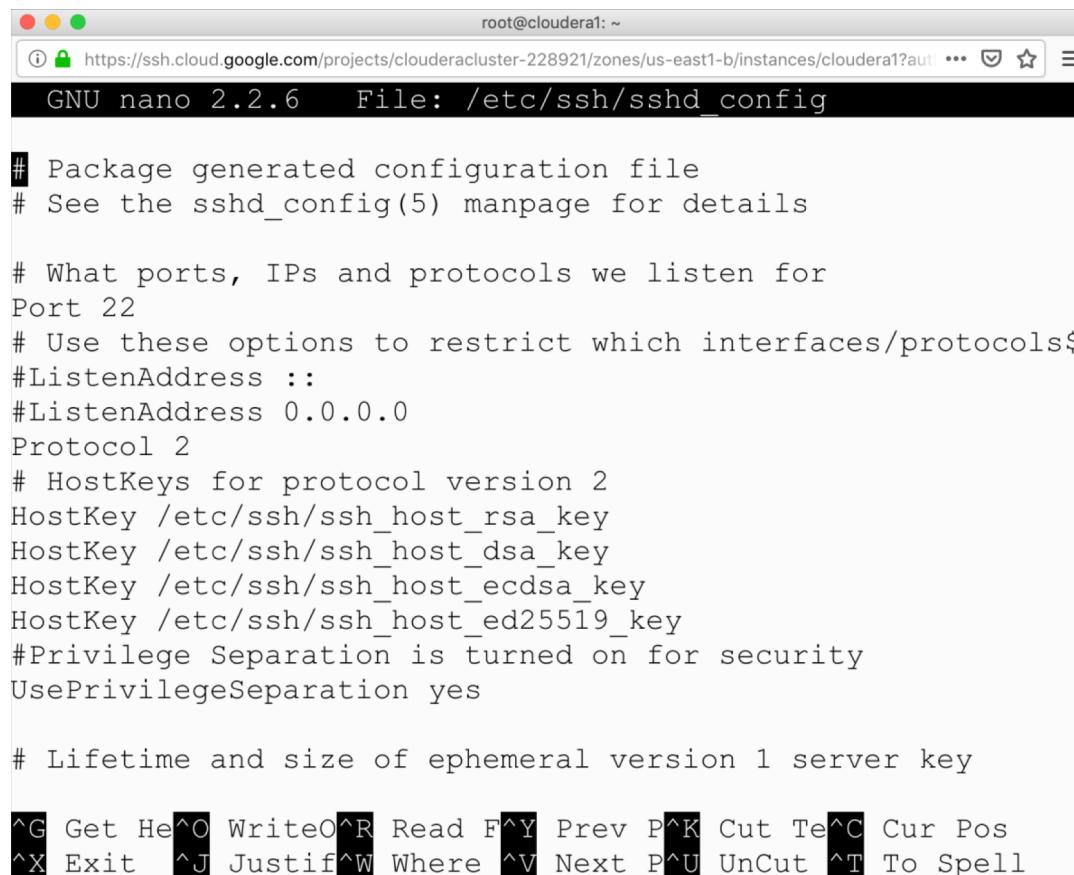
```
thanisabigdata@cloudera2:~$ sudo -i  
root@cloudera2:~# passwd  
Enter new UNIX password:  
Retype new UNIX password:  
passwd: password updated successfully  
root@cloudera2:~# [REDACTED]  
  
thanisabigdata@cloudera3:~$ sudo -i  
root@cloudera3:~# passwd  
Enter new UNIX password:  
Retype new UNIX password:  
passwd: password updated successfully  
root@cloudera3:~# [REDACTED]  
  
thanisabigdata@cloudera4:~$ sudo -i  
root@cloudera4:~# passwd  
Enter new UNIX password:  
Retype new UNIX password:  
passwd: password updated successfully  
root@cloudera4:~# [REDACTED]
```



## 2.1) Change the SSH configuration file

For each server, we need to change the SSH configuration file using the following command

```
# nano /etc/ssh/sshd_config
```



The screenshot shows a terminal window titled "root@cloudera1: ~". The title bar also displays the URL "https://ssh.cloud.google.com/projects/clouderacluster-228921/zones/us-east1-b/instances/cloudera1?aut...". The window content is the "/etc/ssh/sshd\_config" file, which contains the following configuration:

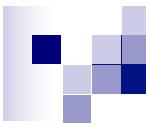
```
# Package generated configuration file
# See the sshd_config(5) manpage for details

# What ports, IPs and protocols we listen for
Port 22
# Use these options to restrict which interfaces/protocols
#ListenAddress ::

#ListenAddress 0.0.0.0
Protocol 2
# HostKeys for protocol version 2
HostKey /etc/ssh/ssh_host_rsa_key
HostKey /etc/ssh/ssh_host_dsa_key
HostKey /etc/ssh/ssh_host_ecdsa_key
HostKey /etc/ssh/ssh_host_ed25519_key
#Privilege Separation is turned on for security
UsePrivilegeSeparation yes

# Lifetime and size of ephemeral version 1 server key

^G Get He^O Write^R Read F^Y Prev P^K Cut Te^C Cur Pos
^X Exit ^J Justif^W Where ^V Next P^U UnCut ^T To Spell
```



In the file, change the following 3 configurations

**PermitRootLogin yes**

..

**ChallengeResponseAuthentication yes**

...

**PasswordAuthentication yes**

## 2.2) Generate SSH key for each server

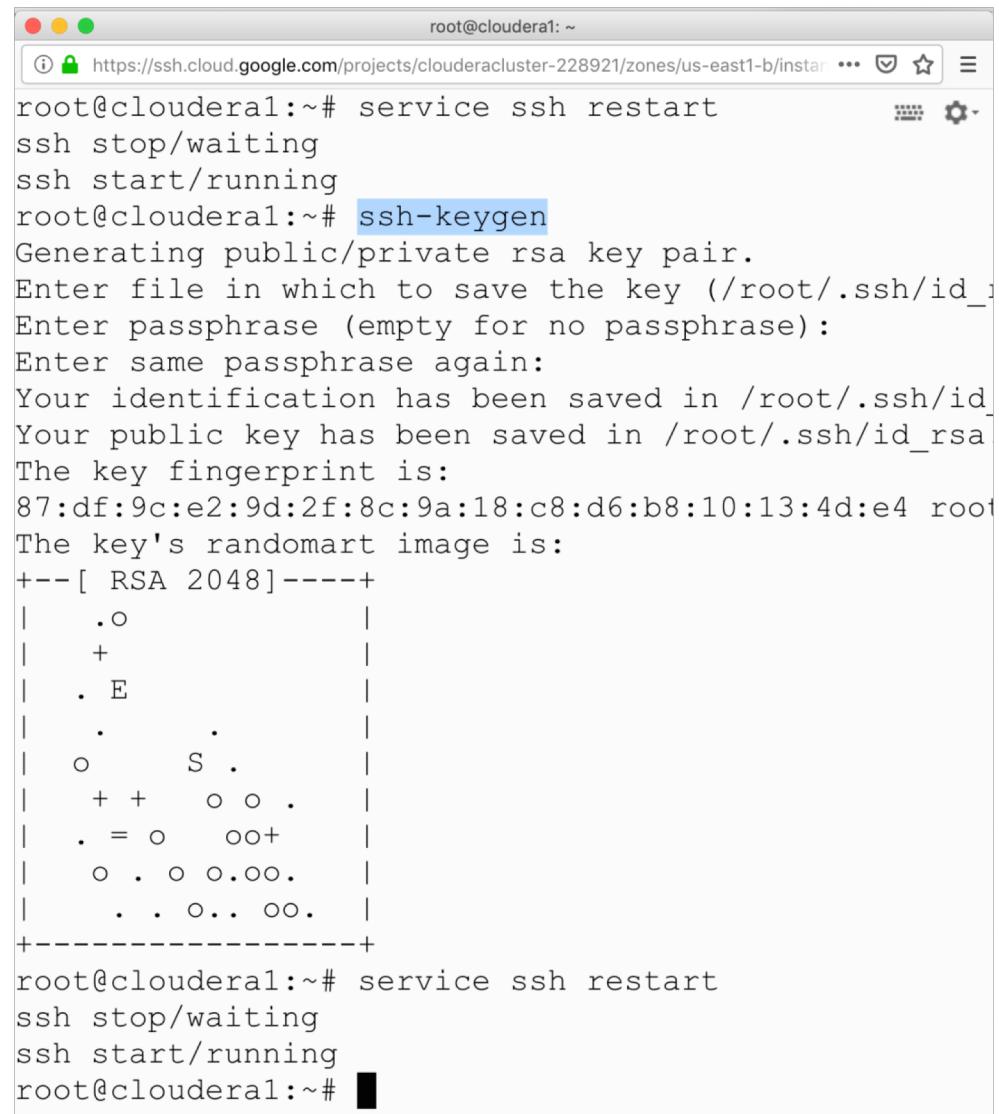
Restart the SSH service,  
and gen the new SSH key using  
the following commands

```
# service ssh restart
```

```
# ssh-keygen
```

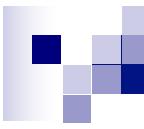
```
# service ssh restart
```

Repeat the second step  
for all instances



The screenshot shows a terminal window titled 'root@clouderal1: ~'. The user has run the command 'ssh-keygen' to generate a new RSA key pair. The process involves entering a file path for the key pair and a passphrase. The terminal also displays the generated public key fingerprint and the randomart image.

```
root@clouderal1:~# service ssh restart
ssh stop/waiting
ssh start/running
root@clouderal1:~# ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
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:
87:df:9c:e2:9d:2f:8c:9a:18:c8:d6:b8:10:13:4d:e4 root@clouderal1
The key's randomart image is:
+---[ RSA 2048]----+
| .o
| +
| . E
| .
| o S .
| + + o o .
| . = o oo+
| o . o ooo.
| . . o.. oo.
+-----+
root@clouderal1:~# service ssh restart
ssh stop/waiting
ssh start/running
root@clouderal1:~#
```



### 3) Give root permission access

For each instance, we need to give the root permission access for other instances using the following commands

```
# ssh-copy-id root@cloudera2
```

```
# ssh-copy-id root@cloudera3
```

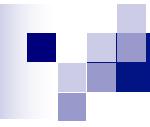
```
# ssh-copy-id root@cloudera4
```

(Note: you need to type the root password, as earlier set)

```
root@cloudera1: ~
https://ssh.cloud.google.com/projects/clouderacluster-228921/zones/us-east1-b/instances/cl... ⓘ 🔒
root@cloudera1:~# ssh-copy-id root@cloudera2
The authenticity of host 'cloudera2 (10.142.0.3)' can't be established.
ECDSA key fingerprint is 45:4d:cb:2a:21:6a:87:4e:87
:21:68:24:3c:c8:46:00.
Are you sure you want to continue connecting (yes/no)? yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys
Password:
Number of key(s) added: 1

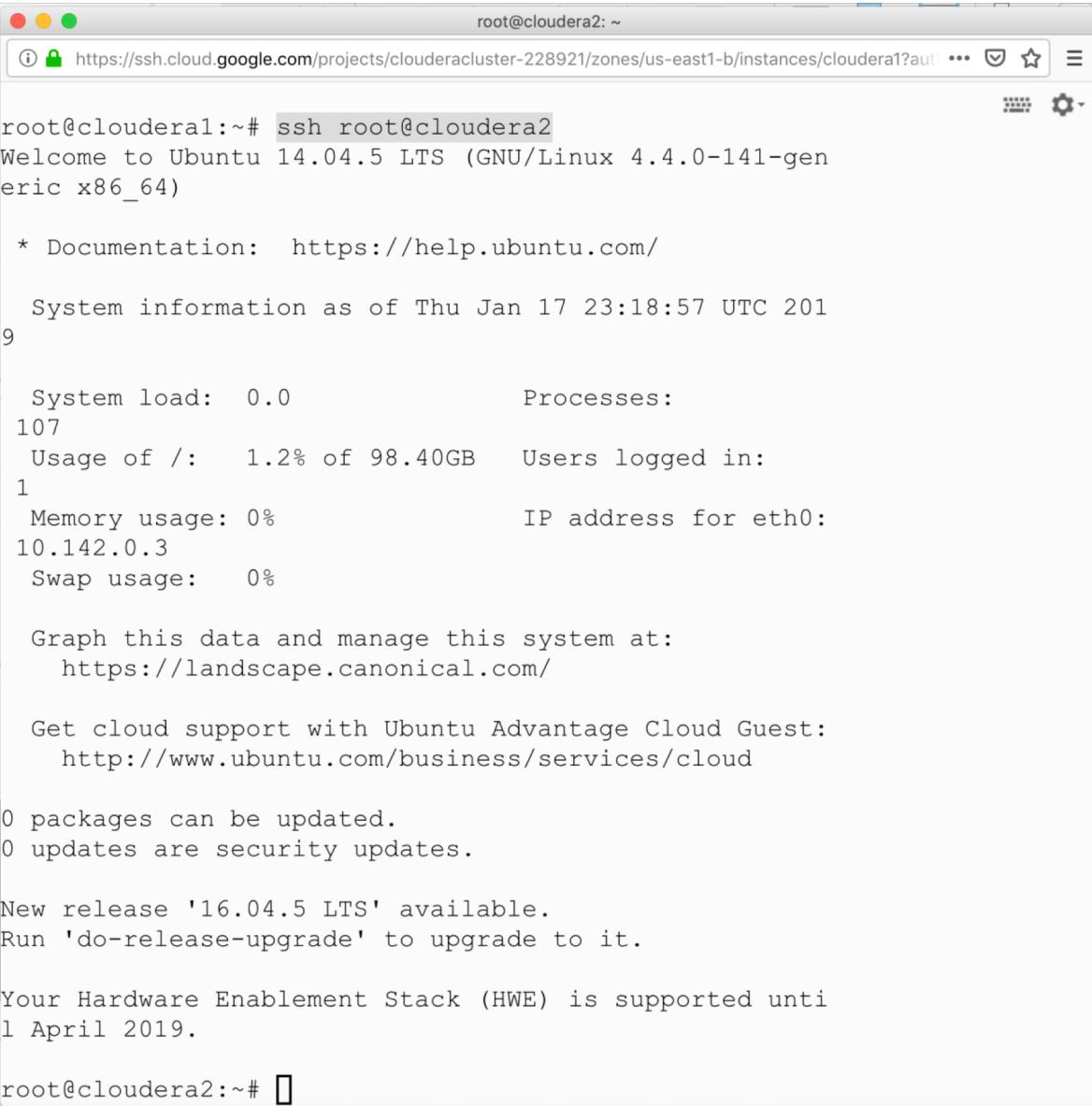
Now try logging into the machine, with:    "ssh 'root@cloudera2'"
and check to make sure that only the key(s) you wanted were added.

root@cloudera1:~# █
```



Try to access the other instances using the following command

```
# ssh root@cloudera2
```



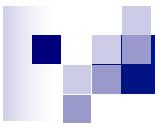
```
root@cloudera2: ~
① 🔒 https://ssh.cloud.google.com/projects/clouderacluster-228921/zones/us-east1-b/instances/cloudera1?aut ...
root@cloudera1:~# ssh root@cloudera2
Welcome to Ubuntu 14.04.5 LTS (GNU/Linux 4.4.0-141-generic x86_64)

 * Documentation:  https://help.ubuntu.com/
System information as of Thu Jan 17 23:18:57 UTC 2019
System load:  0.0          Processes: 107
Usage of /:   1.2% of 98.40GB  Users logged in: 1
Memory usage: 0%          IP address for eth0: 10.142.0.3
Swap usage:   0%
Graph this data and manage this system at:
  https://landscape.canonical.com/
Get cloud support with Ubuntu Advantage Cloud Guest:
  http://www.ubuntu.com/business/services/cloud
0 packages can be updated.
0 updates are security updates.

New release '16.04.5 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Your Hardware Enablement Stack (HWE) is supported until April 2019.

root@cloudera2:~#
```



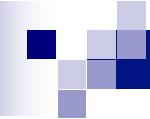
## Repeat the third step for all servers

For example the cloudera2, we need to use the following commands

```
# ssh-copy-id root@cloudera1
```

```
# ssh-copy-id root@cloudera3
```

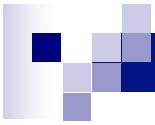
```
# ssh-copy-id root@cloudera4
```



## 4) Change the sysctl vm.swappiness

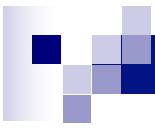
For each server, we need to change the sysctl vm.swappiness configuration using the following command

```
# sysctl -w vm.swappiness=0
```



# **Hands-On: Installing Cloudera Manager on node1**

---



# Cloudera Manager

1) Type command >**wget**

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

2) Type command > **chmod u+x cloudera-manager-installer.bin**

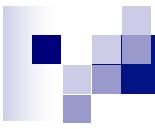
3) Type command > **sudo ./cloudera-manager-installer.bin**

```
root@clouderal:~# wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin
--2017-04-12 13:38:22--  http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin
Resolving archive.cloudera.com (archive.cloudera.com) ... 151.101.0.167, 151.101.64.167, 151.101.128.167,
...
Connecting to archive.cloudera.com (archive.cloudera.com)|151.101.0.167|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 516309 (504K) [application/octet-stream]
Saving to: 'cloudera-manager-installer.bin'

100%[=====] 516,309      --.-K/s   in 0.09s

2017-04-12 13:38:22 (5.73 MB/s) - 'cloudera-manager-installer.bin' saved [516309/516309]

root@clouderal:~# chmod u+x cloudera-manager-installer.bin
root@clouderal:~# sudo ./cloudera-manager-installer.bin
```



# Cloudera Manager

## Cloudera Manager 5

### Cloudera Manager README

Cloudera Manager 5

The Cloudera Manager Installer enables you to install Cloudera Manager and bootstrap an entire CDH cluster, requiring only that you have SSH access to your cluster's machines, and that those machines have Internet access.

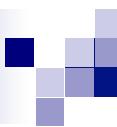
This installer is only recommended for demonstration and proof of concept deployments, but is not recommended for production deployments because it is not intended to scale and may require database migration as your cluster grows.

The Cloudera Manager Installer will automatically:

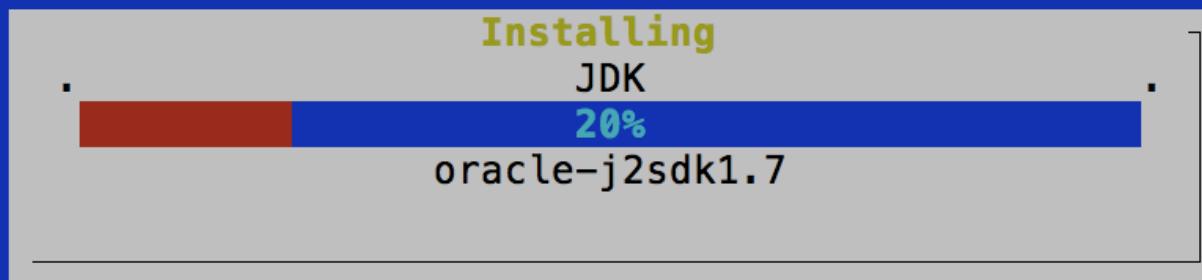
- \* Detect the operating system on the Cloudera Manager host
- \* Install the package repository for Cloudera Manager and the Java Runtime Environment (JRE)

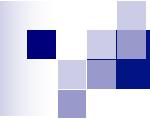
( 42%)

< Cancel > < Next >



## Cloudera Manager 5



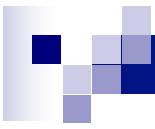


## Cloudera Manager 5

### Next step

Point your web browser to `http://localhost:7180/`. Log in to Cloudera Manager with username: 'admin' and password: 'admin' to continue installation. (Note that the hostname may be incorrect. If the url does not work, try the hostname you use when remotely connecting to this machine.) If you have trouble connecting, make sure you have disabled firewalls, like iptables.

< OK >

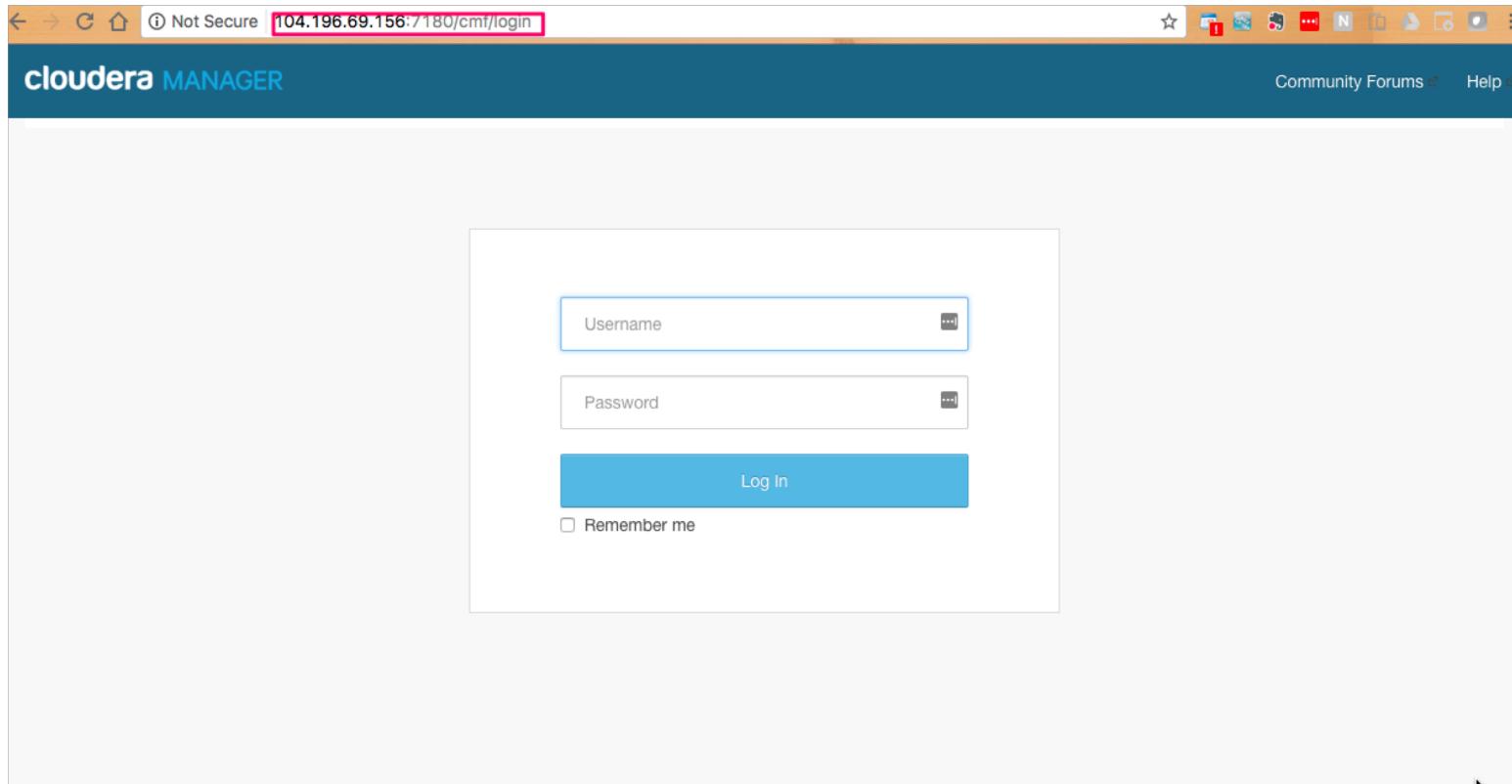


# Login to Cloudera Manager

**Wait several minutes for the Cloudera Manager Server to complete its startup.  
Then running web browser: <http://<>external-ip<>:7180>**

**Username : admin**

**Password : admin**



# Accept the User License terms

## Welcome to Cloudera Manager

### End User License Terms and Conditions

Cloudera Standard License

Version 2015-08-06

#### END USER LICENSE TERMS AND CONDITIONS

THESE TERMS AND CONDITIONS (THESE "TERMS") APPLY TO YOUR USE OF THE PRODUCTS (AS DEFINED BELOW) PROVIDED BY CLOUDERA, INC. ("CLOUDERA").

PLEASE READ THESE TERMS CAREFULLY.

IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF SUCH COMPANY (OR OTHER ENTITY) AND YOU HAVE THE AUTHORITY TO ACCEPT ALL OF THE TERMS AND CONDITIONS SET FORTH IN AN ACCEPTED REQUEST (AS DEFINED BELOW) AND THESE TERMS (COLLECTIVELY, THE "AGREEMENT") ON BEHALF OF SUCH COMPANY (OR OTHER ENTITY).

BY USING ANY OF THE PRODUCTS, YOU ACKNOWLEDGE AND AGREE THAT:

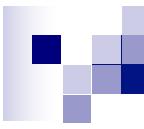
- (A) YOU HAVE READ ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT;
- (B) YOU UNDERSTAND ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT;
- (C) YOU AGREE TO BE LEGALLY BOUND BY ALL OF THE TERMS AND CONDITIONS SET FORTH IN THIS AGREEMENT

Yes, I accept the End User License Terms and Conditions.

1 2

« Back

» Continue



# Select Cloudera Express Edition

cloudera manager

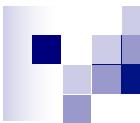
Support ▾ admin ▾

## Welcome to Cloudera Manager. Which edition do you want to deploy?

Upgrading to **Cloudera Enterprise Data Hub Edition** provides important features that help you manage and monitor your Hadoop clusters in mission-critical environments.

	Cloudera Express	Cloudera Enterprise Data Hub Edition Trial	Cloudera Enterprise
License	Free ✓	60 Days After the trial period, the product will continue to function as <b>Cloudera Express</b> . Your cluster and your data will remain unaffected.	Annual Subscription <a href="#">Upload License</a>
Node Limit	Unlimited	Unlimited	Unlimited
CDH	✓	✓	✓
Core Cloudera Manager Features	✓	✓	✓
Advanced Cloudera Manager Features		✓	✓
Cloudera Navigator		✓	✓

Continue



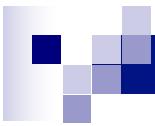
## Thank you for choosing Cloudera Manager and CDH.

This installer will install **Cloudera Express 5.4.0** and enable you to later choose packages for the services below (there may be some license implications).

- Apache Hadoop (Common, HDFS, MapReduce, YARN)
- Apache HBase
- Apache ZooKeeper
- Apache Oozie
- Apache Hive
- Hue (Apache licensed)
- Apache Flume
- Cloudera Impala (Apache licensed)
- Apache Sentry
- Apache Sqoop
- Cloudera Search (Apache licensed)
- Apache Spark

You are using Cloudera Manager to install and configure your system. You can learn more about Cloudera Manager by clicking on the **Support** menu above.

Continue



# Provide your 4 virtual instances using the hostnames

The screenshot shows the Cloudera Manager interface with a dark blue header bar. On the left, it says "cloudera MANAGER". On the right, there are "Support" and "admin" dropdown menus. Below the header, the main content area has a light gray background. It displays the text "Specify hosts for your CDH cluster installation." followed by instructions about hostnames and a search input field.

Specify hosts for your CDH cluster installation.

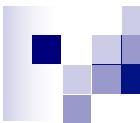
Hosts should be specified using the same hostname (FQDN) that they will identify themselves with.

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](#).

```
cloudera1  
cloudera2  
cloudera3  
cloudera4
```

SSH Port:



## 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

<input checked="" type="checkbox"/> Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/> cloudera1	cloudera1.c.iot-class-feb2017.internal	10.142.0.2	No	Host ready: 0 ms response time.
<input checked="" type="checkbox"/> cloudera2	cloudera2.c.iot-class-feb2017.internal	10.142.0.3	No	Host ready: 2 ms response time.
<input checked="" type="checkbox"/> cloudera3	cloudera3.c.iot-class-feb2017.internal	10.142.0.4	No	Host ready: 1 ms response time.
<input checked="" type="checkbox"/> cloudera4	cloudera4.c.iot-class-feb2017.internal	10.142.0.5	No	Host ready: 1 ms response time.



Back

Continue



## 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](#)

#### Select the version of CDH

CDH-5.10.1-1.cdh5.10.1.p0.10

CDH-4.7.1-1.cdh4.7.1.p0.47

Versions of CDH that are too new for this version of Cloudera Manager (5.10.1) 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

KAFKA-2.1.1-1.2.1.1.p0.18

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#)

[Back](#)

[Continue](#)

# Select install Oracle Java & Java unlimited

## Cluster Installation

### JDK Installation Options

Agreement. Source code may not be redistributed unless expressly provided for in this Agreement.

J. THIRD PARTY CODE. Additional copyright notices and license terms applicable to portions of the Software are set forth in the THIRDPARTYLICENSEREADME file accessible at <http://www.oracle.com/technetwork/java/javase/documentation/index.html>. In addition to any terms and conditions of any third party opensource/freeware license identified in the THIRDPARTYLICENSEREADME file, the disclaimer of warranty and limitation of liability provisions in paragraphs 4 and 5 of the Binary Code License Agreement shall apply to all Software in this distribution.

K. TERMINATION FOR INFRINGEMENT. Either party may terminate this Agreement immediately should any Software become, or in either party's opinion be likely to become, the subject of a claim of infringement of any intellectual property right.

L. INSTALLATION AND AUTO-UPDATE. The Software's installation and auto-update processes transmit a limited amount of data to Oracle (or its service provider) about those specific processes to help Oracle understand and optimize them. Oracle does not associate the data with personally identifiable information. You can find more information about the data Oracle collects as a result of your Software download at <http://www.oracle.com/technetwork/java/javase/documentation/index.html>.

For inquiries please contact: Oracle America, Inc., 500 Oracle Parkway,

Redwood Shores, California 94065, USA.

Last updated 02 April 2013

- 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.

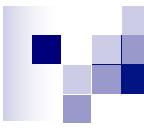
- Install Java Unlimited Strength Encryption Policy Files

Check this checkbox if local laws permit you to deploy unlimited strength encryption and you are running a secure cluster.

1 2 3 4 5 6 7 8

« Back

Continue »



## Cluster Installation

### Enable 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.

Single User Mode



« Back

1 2 3 4 5 6 7 8

» Continue

## Provide the password for the hostname

cloudera MANAGER Support ▾ admin

### Cluster Installation

Provide SSH 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

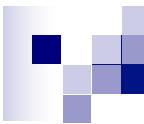
Enter Password: ..... 

Confirm Password: ..... 

SSH Port: 22

1 2 3 **4** 5 6 7

Back Continue



cloudera MANAGER

Support ▾ admin

## Cluster Installation

Installation completed successfully.

4 of 4 host(s) completed successfully.

Hostname	IP Address	Progress	Status	
cloudera1.c.iot-class-feb2017.internal	10.142.0.2	<div style="width: 100%; background-color: #2e7131;"></div>	Installation completed successfully.	<a href="#">Details</a> 
cloudera2.c.iot-class-feb2017.internal	10.142.0.3	<div style="width: 100%; background-color: #2e7131;"></div>	Installation completed successfully.	<a href="#">Details</a> 
cloudera3.c.iot-class-feb2017.internal	10.142.0.4	<div style="width: 100%; background-color: #2e7131;"></div>	Installation completed successfully.	<a href="#">Details</a> 
cloudera4.c.iot-class-feb2017.internal	10.142.0.5	<div style="width: 100%; background-color: #2e7131;"></div>	Installation completed successfully.	<a href="#">Details</a> 

[Back](#)

1 2 3 4 5 6 7

[Continue](#)

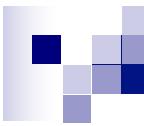


## Cluster Installation

### Installing Selected Parcels

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

A horizontal action bar at the bottom of the screen. On the left is a "Back" button. In the center is a set of page numbers: 1, 2, 3, 4, 5, 6, 7, where number 6 is highlighted with a yellow box. On the right is a large blue "Continue" button. A red arrow points downwards towards the "Continue" button.



## Cluster Installation

Inspect hosts for correctness ↻ Run Again

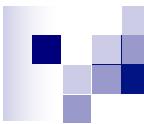
### Validations

- ✓ Inspector ran on all 4 hosts.
- ✓ Individual hosts resolved their own hostnames correctly.
- ✓ No errors were found while looking for conflicting init scripts.
- ✓ No errors were found while checking /etc/hosts.
- ✓ All hosts resolved localhost to 127.0.0.1.
- ✓ All hosts checked resolved each other's hostnames correctly and in a timely manner.
- ✓ Host clocks are approximately in sync (within ten minutes).
- ✓ Host time zones are consistent across the cluster.
- ✓ No users or groups are missing.
- ✓ No conflicts detected between packages and parcels.
- ✓ No kernel versions that are known to be bad are running.
- ✓ No problems were found with /proc/svcs/vm/swappiness on any of the hosts

1 2 3 4 5 6 7

Back

Finish



## Cluster Setup

Choose the CDH 5 services that you want to install on your cluster.

Choose a combination of services to install.

- Core Hadoop**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, and Hue
- Core with HBase**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and HBase
- Core with Impala**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Impala
- Core with Search**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Solr
- Core with Spark**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, and Spark
- All Services**  
HDFS, YARN (MapReduce 2 Included), ZooKeeper, Oozie, Hive, Hue, HBase, Impala, Solr, Spark, and Key-Value Store Indexer

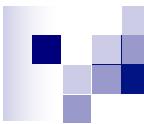
**Custom Services**

Choose your own services. Services required by chosen services will automatically be included. Flume can be added after your initial cluster has been set up.

1 2 3 4 5 6

Back

Continue



cloudera MANAGER

Support ▾ admin ▾

## Cluster Setup

### Customize Role Assignments

You can customize the role assignments for your new cluster here, but if assignments are made incorrectly, such as assigning too many roles to a single host, this can impact the performance of your services. Cloudera does not recommend altering assignments unless you have specific requirements, such as having pre-selected a specific host for a specific role.

You can also view the role assignments by host. [View By Host](#)

#### H HBase

M Master x 1 New

HBRE HBase REST Server

HBTS HBase Thrift Server

RS RegionServer x 3 New

cloudera1.c.iot-class-feb2017.internal

Select hosts

Select hosts

Same As DataNode ▾

#### S HDFS

NN NameNode x 1 New

SNN SecondaryNameNode x 1 New

B Balancer x 1 New

HFS HttpFS

cloudera1.c.iot-class-feb2017.internal

cloudera1.c.iot-class-feb2017.internal

cloudera1.c.iot-class-feb2017.internal

Select hosts

DFS Neo Gateway

DPM Dynamic Partition Manager

1 2 3 4 5 6

Continue

Back



## Cluster Setup

### Database Setup

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.

#### Hive

Database Host Name:	Database Type:	Database Name :	Username:	Password:
ip-172-31-1-242.us-west-2.compute.internal	PostgreSQL	hive	hive	bV6sUA8gPH

#### Oozie Server

Currently assigned to run on ip-172-31-1-242.us-west-2.compute.internal.

Database Host Name:	Database Type:	Database Name :	Username:	Password:
ip-172-31-1-242.us-west-2.compute.internal	PostgreSQL	oozie_oozie_se	oozie_oozie_se	6MvnYMQkTE

  
**Test Connection**



## Cluster Setup

### Database Setup

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.

#### Hive

Database Host Name:

ip-172-31-1-242.us-west-2.compute.internal

Database Type:

PostgreSQL

✓ Skipped. Cloudera Manager will create this database in a later step.

Database Name :

hive

Username:

hive

Password:

bV6sUA8gPH

#### Oozie Server

Currently assigned to run on ip-172-31-1-242.us-west-2.compute.internal.

Database Host Name:

ip-172-31-1-242.us-west-2.compute.internal

Database Type:

PostgreSQL

✓ Skipped. Cloudera Manager will create this database in a later step.

Database Name :

oozie\_oozie\_se

Username:

oozie\_oozie\_se

Password:

6MvnYMQKTE

Test Connection

« Back

1 2 3 4 5 6

« Continue

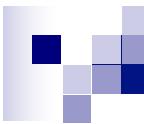


## Cluster Setup

### Review Changes

<b>HDFS Root Directory</b> hbase.rootdir	Cluster 1 > HBase (Service-Wide) <input type="text" value="/hbase"/>	?
<b>Enable Replication</b> hbase.replication	Cluster 1 > HBase (Service-Wide) <input checked="" type="checkbox"/>	?
<b>Enable Indexing</b>	Cluster 1 > HBase (Service-Wide) <input checked="" type="checkbox"/>	?
<b>DataNode Data Directory</b> dfs.data.dir, dfs.datanode.data.dir	Cluster 1 > DataNode Default Group <input type="text" value="/dfs/dn"/> <input type="text" value="/mnt/dfs/dn"/>	?
<b>DataNode Failed Volumes Tolerated</b> dfs.datanode.failed.volumes.tolerated	Cluster 1 > DataNode Default Group <input type="text" value="1"/>	?

**Back** **Continue**



cloudera MANAGER

Support ▾ admin ▾

## Cluster Setup

### ✓ First Run Command

Status: Finished Start Time: Apr 12, 2:07:07 PM Duration: 5.3m

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

#### Details

Completed 10 of 10 step(s).

All  Failed Only  Running Only

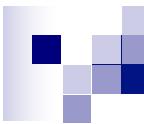
Step	Context	Start Time	Duration	Actions
▶ ✓ Run 1 steps in parallel Successfully completed 1 steps.		Apr 12, 2:07:07 PM	50ms	
▶ ✓ Deploying Client Configuration Successfully deployed all client configurations.	Cluster 1	Apr 12, 2:07:07 PM	16.17s	
▶ ✓ Start Cloudera Management Service, ZooKeeper Successfully completed 2 steps.		Apr 12, 2:07:24 PM	25.45s	
◀ ▶ Start HDFS		Apr 12, 2:07:40 PM	17.26s	

Back

1 2 3 4 5 6

Continue





# cloudera MANAGER

Clusters ▾ Hosts ▾ Diagnostics ▾ Audits Charts ▾ Administration ▾

Search Support ▾ admin ▾

## Home

Today, 2:14 PM UTC

Status All Health Issues Configuration ✖ 7 ▾ All Recent Commands

Add Cluster

Try Cloudera Enterprise Data Hub Edition for 60 Days

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 ↗](#)

Cluster 1 (CDH 5.10.1, Parcels)

- Hosts
- HBase
- HDFS ✖ 1
- Hive
- Hue ✖ 1
- Impala
- Key-Value Store...
- Oozie
- Solr

