

Install Apache Hadoop for Development/Production using Open source Distributions



Install Pure Apache Hadoop

- 1. Ubuntu or Centos (Linux)**
- 2. Installing JDK**
- 3. Download/Extract Hadoop**
- 4. Installing Hadoop**
- 5. Configure xml files**
- 6. Formatting HDFS**
- 7. Start Hadoop and then see Hadoop Web Console.**
- 8. Install Services (Hive , Sqoop ,HBASE,.....)**
- 9. Stop Hadoop (if needed.)**

```
<configuration>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>localhost</value>
</property>
<property>
  <name>yarn.resourcemanager.scheduler.address</name>
  <value>localhost:8030</value>
</property>
<property>
  <name>yarn.resourcemanager.resource-tracker.address</name>
  <value>localhost:8031</value>
</property>
<property>
  <name>yarn.resourcemanager.address</name>
  <value>localhost:8032</value>
</property>
```

```
<configuration>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

Hadoop User Web Interfaces

Namenode information - Mozilla Firefox

Namenode information

http://localhost:50070/dfshealth.html#tab-overview

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Overview 'localhost:54310' (active)

Started:	Sat Apr 18 15:53:55 PDT 2015
Version:	2.6.0, re3496499ecb8d220fba99dc5ed4c99c8f9e33bb1
Compiled:	2014-11-13T21:10Z by jenkins from (detached from e349649)
Cluster ID:	CID-e2f515ac-33da-45bc-8466-5b1100a2bf7f
Block Pool ID:	BP-130729900-192.168.1.1-1429393391595

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks = 1 total filesystem object(s).
Heap Memory used 58.41 MB of 167.5 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 28.34 MB of 29.94 MB Committed Non Heap Memory. Max Non Heap Memory is 214 MB.

<http://localhost:50070/dfshealth.html#tab-startup-progress>

Namenode information - Mozilla Firefox

Namenode information

http://localhost:50070/dfshealth.html#tab-datanode

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
laptop (127.0.0.1:50010)	1	In Service	454.29 GB	28 KB	125.83 GB	328.47 GB	0	28 KB (0%)	0	2.6.0

Decommissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction

Hadoop, 2014. Legacy



Hadoop SecondaryNameNode - Mozilla Firefox

Hadoop SecondaryNa... x +

http://localhost:50090/status.jsp Search

SecondaryNameNode

Version: 2.6.0, e3496499ecb8d220fba99dc5ed4c99c8f9e33bb1

Compiled: 2014-11-13T21:10Z by jenkins from (detached from e349649)

SecondaryNameNode Status

```
Name Node Address : localhost/127.0.0.1:54310
Start Time : Sat Apr 18 16:43:38 PDT 2015
Last Checkpoint : 79 seconds ago
Checkpoint Period : 3600 seconds
Checkpoint Transactions: 1000000
Checkpoint Dirs : [file:///app/hadoop/tmp/dfs/namesecondary]
Checkpoint Edits Dirs : [file:///app/hadoop/tmp/dfs/namesecondary]
```

Logs



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

Tools

Directory: /logs/ - Mozilla Firefox

Directory: /logs/ http://localhost:50090/logs/ Search

Directory: /logs/

[SecurityAuth-hduser.audit](#)
[hadoop-hduser-datanode-laptop.log](#)
[hadoop-hduser-datanode-laptop.out](#)
[hadoop-hduser-datanode-laptop.out.1](#)
[hadoop-hduser-datanode-laptop.out.2](#)
[hadoop-hduser-namenode-laptop.log](#)
[hadoop-hduser-namenode-laptop.out](#)
[hadoop-hduser-namenode-laptop.out.1](#)
[hadoop-hduser-namenode-laptop.out.2](#)
[hadoop-hduser-secondarynamenode-laptop.log](#)
[hadoop-hduser-secondarynamenode-laptop.out](#)
[hadoop-hduser-secondarynamenode-laptop.out.1](#)
[hadoop-hduser-secondarynamenode-laptop.out.2](#)
[userlogs/](#)
[yarn-hduser-nodemanager-laptop.log](#)

0 bytes Apr 18, 2015 3:40:58 PM
72879 bytes Apr 18, 2015 4:44:13 PM
718 bytes Apr 18, 2015 4:43:21 PM
718 bytes Apr 18, 2015 3:53:49 PM
718 bytes Apr 18, 2015 3:41:03 PM
121216 bytes Apr 18, 2015 4:52:23 PM
718 bytes Apr 18, 2015 4:43:16 PM
718 bytes Apr 18, 2015 3:53:44 PM
718 bytes Apr 18, 2015 3:40:58 PM
51913 bytes Apr 18, 2015 4:52:38 PM
718 bytes Apr 18, 2015 4:43:37 PM
718 bytes Apr 18, 2015 3:54:06 PM
718 bytes Apr 18, 2015 3:42:52 PM
4096 bytes Apr 18, 2015 4:52:22 PM
81625 bytes Apr 18, 2015 4:44:32 PM

Logged in as: dr.who :44:02 PM
:54:32 PM
:43:10 PM
:44:32 PM
:44:00 PM
:54:29 PM
:43:08 PM

About the Cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:1024, vCores:1>	<memory:8192, vCores:8>

Cluster overview

Cluster ID: 1465353454396

ResourceManager state: STARTED

ResourceManager HA state: active

ResourceManager HA zookeeper connection state: ResourceManager HA is not enabled.

ResourceManager RMStateStore: org.apache.hadoop.yarn.server.resourcemanager.recovery.NullRMStateStore

ResourceManager started on: Wed Jun 08 09:37:34 +0700 2016

ResourceManager version: 2.7.2 from b165c4fe8a74265c792ce23f546c64604acf0e41 by jenkins source checksum c63f7cc71b8f63249e35126f0f7492d on 2016-01-26T00:16Z

Hadoop version: 2.7.2 from b165c4fe8a74265c792ce23f546c64604acf0e41 by jenkins source checksum d0fdad26633fa762bff87ec759ebe689c on 2016-01-26T00:08Z

Hadoop Distribution

On-Premise

- ➊ Pure Apache Hadoop
- ➋ Cloudera
- ➌ Hortonworks
- ➍ MapR
- ➎ Pivotal
- ➏ IBM InfoSphere BigInsight

On-Cloud (Hadoop as a Service)

- ➊ Amazon EMR
- ➋ Microsoft Azure HDInsight
- ➌ Google Cloud Platform

The Forrester Wave: Big Data Hadoop



CHULALONGKORN
BUSINESS SCHOOL
FLAGSHIP FOR LIFE

Distributions Q1 2016



The Forrester Wave: Big Data Hadoop



FLAGSHIP FOR LIFE

Cloud Q1 2016



Thaveewat Khanan

Organization Data Management

Hadoop Development: Sandbox

Recommended

Cloudera Quickstart

Hortonworks Sandbox

On PC/Mac

Require to install VMWare Player or VirtualBox

On Linux Virtual Server

Need to install Docker Engine

<http://hortonworks.com/downloads/#sandbox>

Hortonworks Sandbox

Hortonworks Sandbox on a VM

HDP® 2.5 on Hortonworks Sandbox

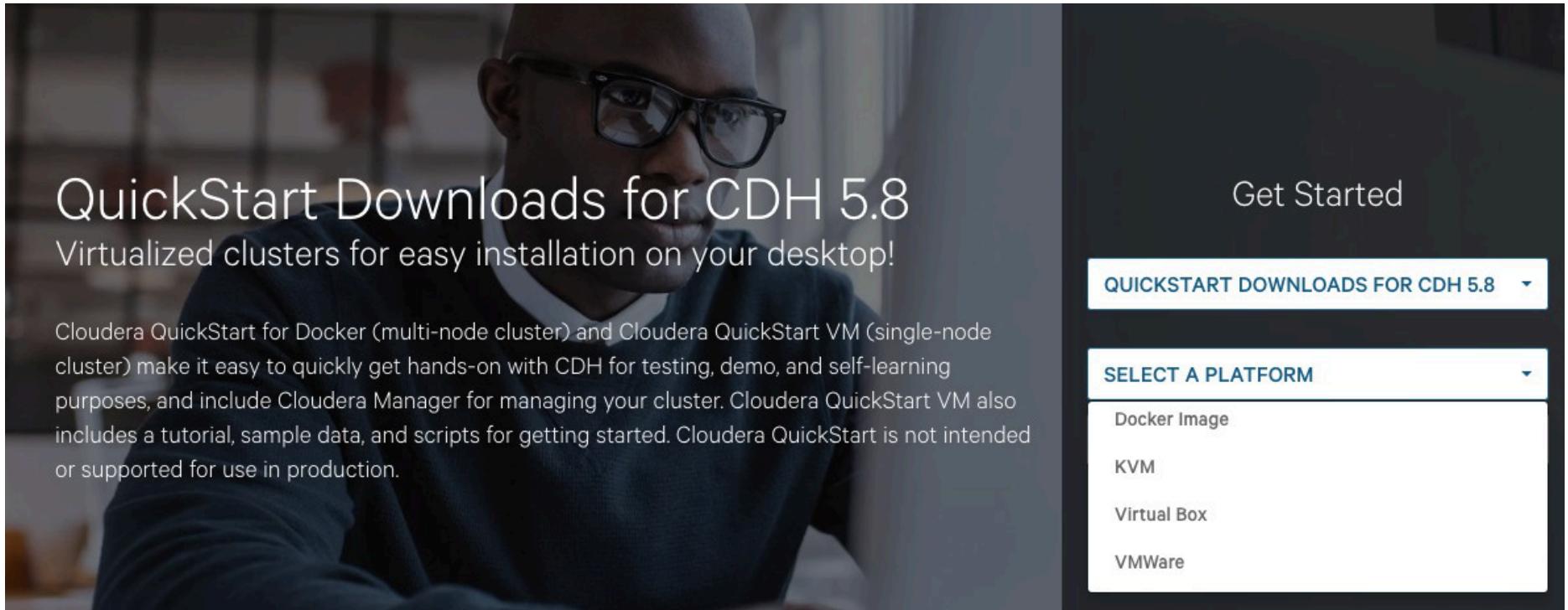
Tutorials		Release Notes		Import on Virtual Box		MD5 : b08f1dce17ab3ae2431532be74bdbbbb	DOWNLOAD FOR VIRTUALBOX
Tutorials		Release Notes		Import on VMware		MD5 : fe1e91bc26e6879fdc6dde6e3778c262	DOWNLOAD FOR VMWARE
Tutorials		Release Notes		Import on Docker		MD5 : 2a710f236135e620ec8488a1229af07e	DOWNLOAD FOR DOCKER

Hortonworks Sandbox in the Cloud

HDP 2.4 on Azure with Hortonworks Sandbox

Tutorial: Sandbox on Azure		Try it one month for free	ONE MONTH TRIAL
--	--	---------------------------	---------------------------------

<http://www.cloudera.com/downloads.html>



The screenshot shows a landing page for Cloudera QuickStart Downloads. On the left, there's a large image of a person wearing glasses looking down at a laptop screen. Overlaid on the image is the text "QuickStart Downloads for CDH 5.8" and "Virtualized clusters for easy installation on your desktop!". Below this, a paragraph explains the purpose of the QuickStart tools. On the right, there's a sidebar with a "Get Started" button and two dropdown menus: "QUICKSTART DOWNLOADS FOR CDH 5.8" and "SELECT A PLATFORM". The "SELECT A PLATFORM" menu contains options for Docker Image, KVM, Virtual Box, and VMWare.

QuickStart Downloads for CDH 5.8

Virtualized clusters for easy installation on your desktop!

Cloudera QuickStart for Docker (multi-node cluster) and Cloudera QuickStart VM (single-node cluster) make it easy to quickly get hands-on with CDH for testing, demo, and self-learning purposes, and include Cloudera Manager for managing your cluster. Cloudera QuickStart VM also includes a tutorial, sample data, and scripts for getting started. Cloudera QuickStart is not intended or supported for use in production.

Get Started

QUICKSTART DOWNLOADS FOR CDH 5.8

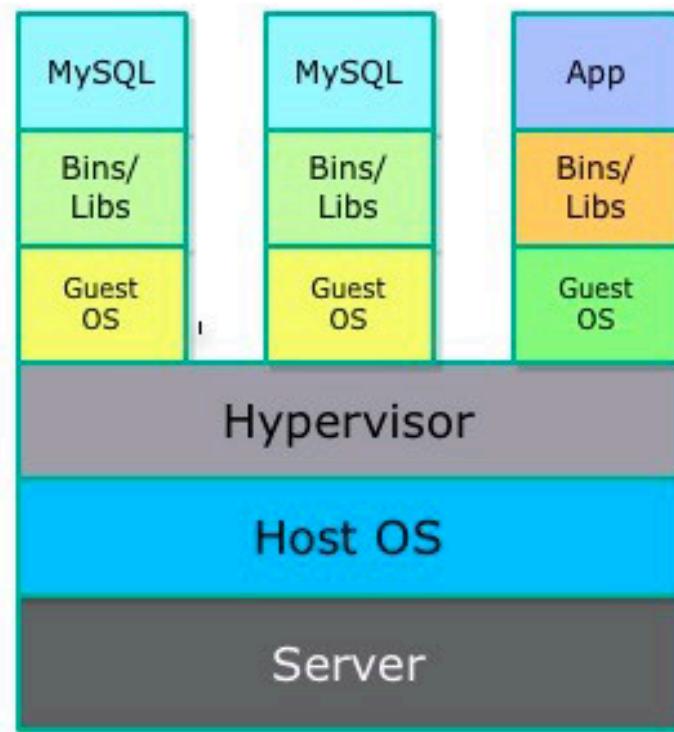
SELECT A PLATFORM

- Docker Image
- KVM
- Virtual Box
- VMWare

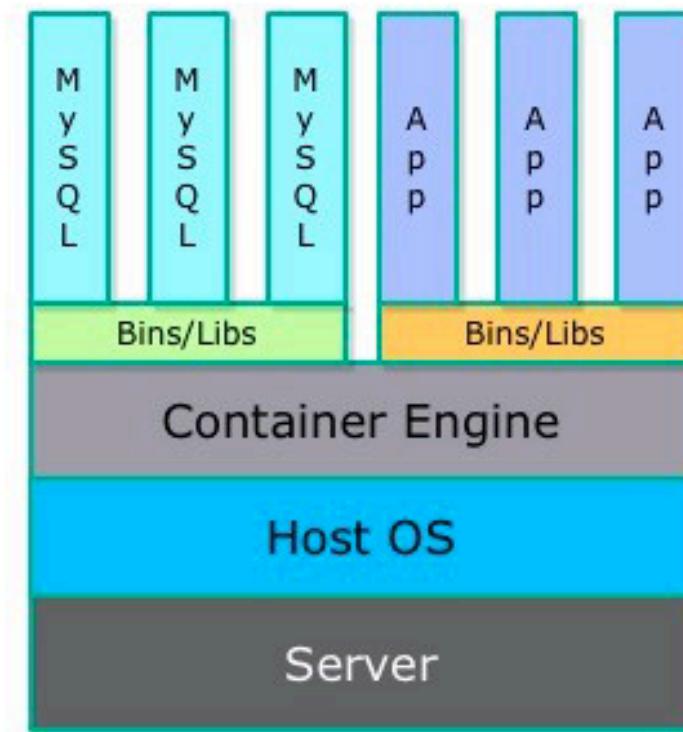
Install a Docker Engine



Virtual Machines



Containers



Pull Cloudera Quickstart

```
$ sudo docker pull cloudera/quickstart:latest
```

```
ubuntu@ip-172-31-30-238:~$ sudo docker pull cloudera/quickstart:latest
latest: Pulling from cloudera/quickstart
2cda82941cb7: Already exists
Digest: sha256:f91bee4cdfa2c92ea3652929a22f729d4d13fc838b00f120e630f91c941acb63
Status: Downloaded newer image for cloudera/quickstart:latest
ubuntu@ip-172-31-30-238:~$ █
```

```
$ sudo docker images
```

REPOSITORY	TAG	IMAGE ID	CREATED
cloudera/quickstart	latest	2cda82941cb7	9 weeks ago
6.336 GB		—	

Run Cloudera quickstart

```
$ sudo docker run --hostname=quickstart.cloudera  
--privileged=true -t -i [OPTIONS] [IMAGE]  
/usr/bin/docker-quickstart
```

Example: \$ sudo docker run
--hostname=quickstart.cloudera --privileged=true -t -i -p
8888:8888 cloudera/quickstart /usr/bin/docker-quickstart

```
ubuntu@ip-172-31-30-238:~$ sudo docker run --hostname=quickstart.cloudera --privileged=true -t -i -p 8888:8888 -p 7180:7180 cloudera/quickstart /usr/bin/docker-quickstart
Starting mysqld: [ OK ]  
  
if [ "$1" == "start" ] ; then
    if [ "${EC2}" == 'true' ] ; then
        FIRST_BOOT_FLAG=/var/lib/cloudera-quickstart/.ec2-key-installed
        if [ ! -f "${FIRST_BOOT_FLAG}" ] ; then
            METADATA_API=http://169.254.169.254/latest/meta-data
            KEY_URL=${METADATA API}/public-kevs/0/openssh-kev
```

Install Cloudera Cluster on AWS



Launch a virtual server on EC2 Amazon Web Services

Amazon Web Services

Compute

- EC2**
Virtual Servers in the Cloud
- Lambda** PREVIEW
Run Code in Response to Events

Storage & Content Delivery

- S3**

Administration & Security

- Directory Service**
Managed Directories in the Cloud
- Identity & Access Management**
Access Control and Key Management
- Trusted Advisor**
AWS Cloud Optimization Expert
- CloudTrail**

Application Services

- SQS**
Message Queue Service
- SWF**
Workflow Service for Coordinating Application Components
- AppStream**
Low Latency Application Streaming
- Elastic Transcoder**

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#)

[Tag Editor](#)

Select an Amazon Machine Image (AMI) and Ubuntu Server 14.04 LTS (PV)

Root device type: ebs Virtualization type: paravirtual

Ubuntu Server 14.04 LTS (PV), SSD Volume Type - ami-23ebb513

Ubuntu Ubuntu Server 14.04 LTS (PV), EBS General Purpose (SSD) Volume Type. Support available from Canonical (<http://www.ubuntu.com/cloud/services>).
Free tier eligible

Root device type: ebs Virtualization type: paravirtual

Select 64-bit

Install Cloudera Cluster on AWS



Choose m3.xlarge Type virtual server

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

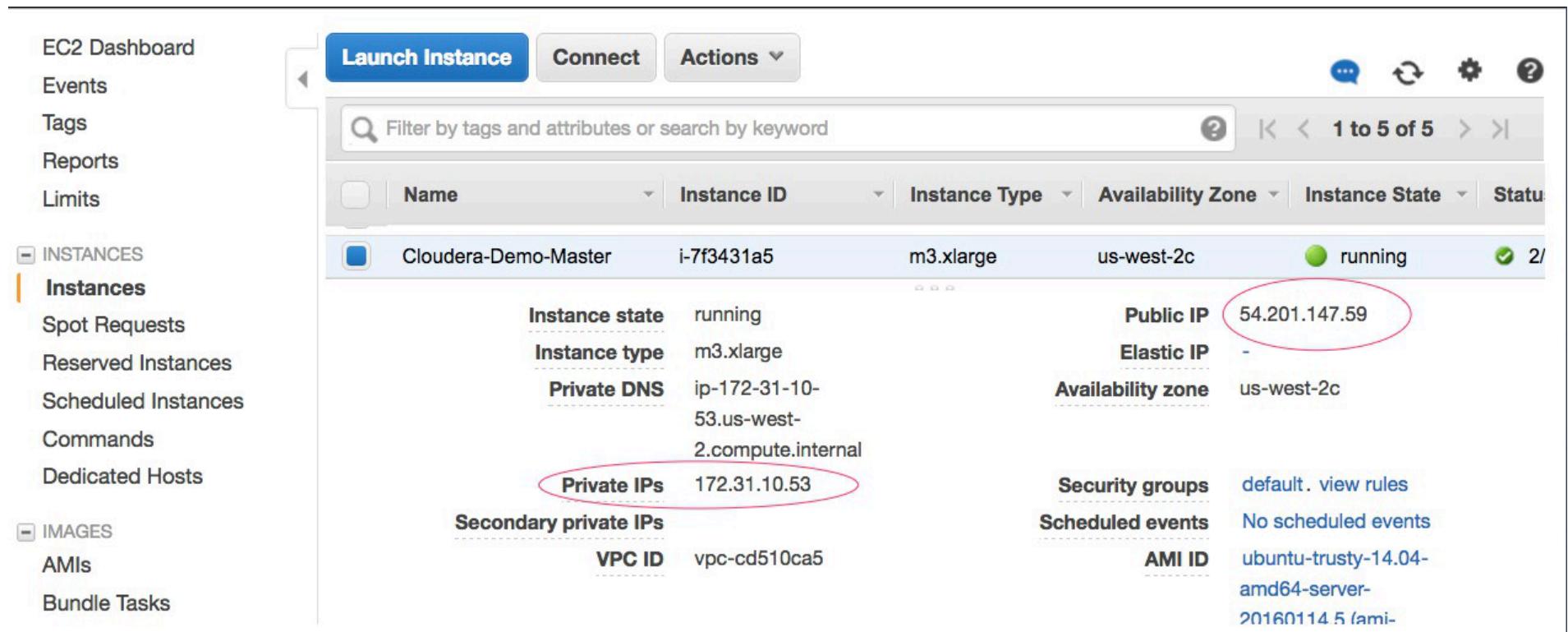
Step 2: Choose an Instance Type

						Available		
<input type="checkbox"/>	Micro instances	t1.micro <small>Free tier eligible</small>	1	0.613	EBS only	-	Very Low	
<input checked="" type="checkbox"/>	General purpose	t2.micro <small>Free tier eligible</small>	1	1	EBS only	-	Low to Moderate	
<input checked="" type="checkbox"/>	General purpose	t2.small	1	2	EBS only	-	Low to Moderate	
<input checked="" type="checkbox"/>	General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	
<input type="checkbox"/>	General purpose	m3.medium	1	3.75	1 x 4 (SSD)	-	Moderate	
<input type="checkbox"/>	General purpose	m3.large	2	7.5	1 x 32 (SSD)	-	Moderate	
<input checked="" type="checkbox"/>	General purpose	m3.xlarge	4	15	2 x 40 (SSD)	Yes	High	
<input type="checkbox"/>	General purpose	m3.2xlarge	8	30	2 x 80 (SSD)	Yes	High	

[Cancel](#) [Previous](#) [Review and Launch](#) [Next: Configure Instance Details](#)

Install Cloudera Cluster on AWS

Can also view details of the instance such as Public IP and Private IP



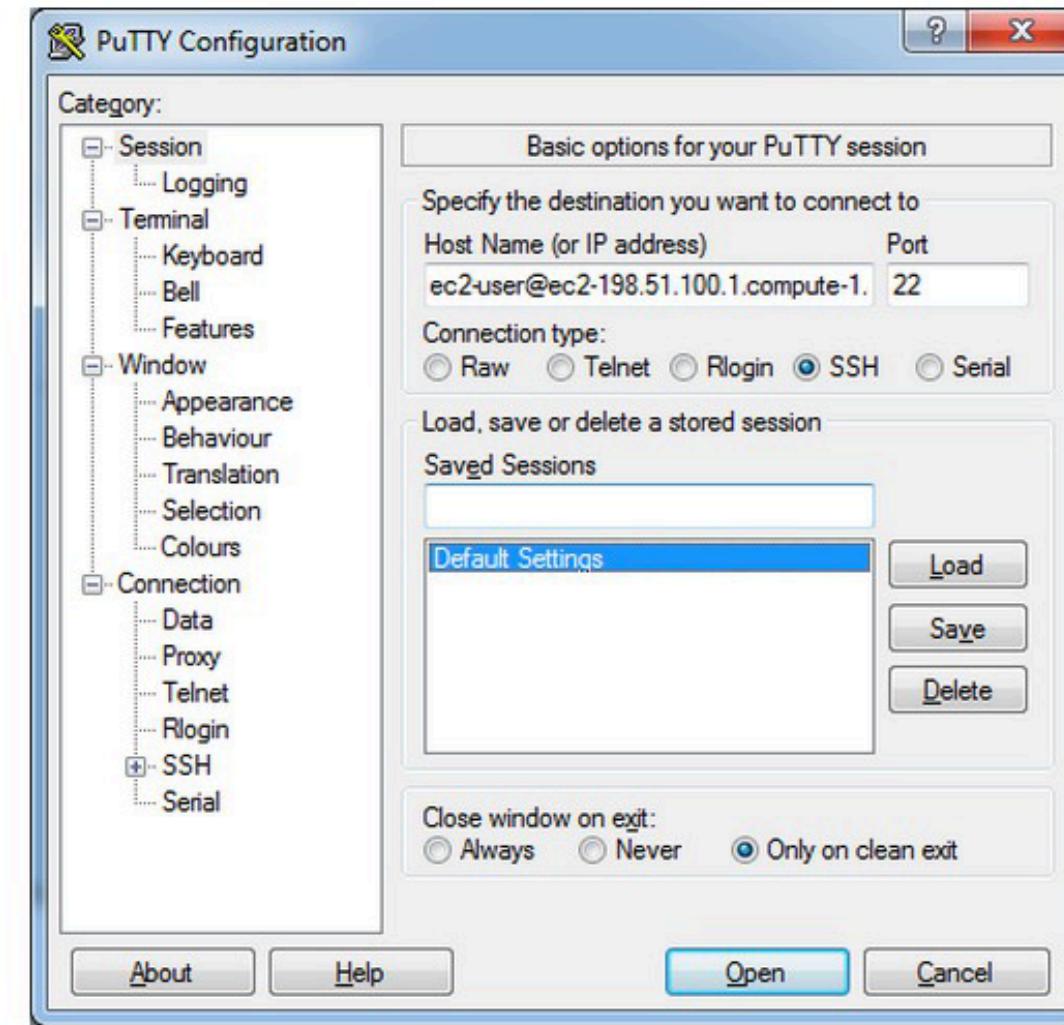
The screenshot shows the AWS EC2 Dashboard. On the left, there's a sidebar with links like EC2 Dashboard, Events, Tags, Reports, Limits, INSTANCES (with Instances selected), Spot Requests, Reserved Instances, Scheduled Instances, Commands, Dedicated Hosts, and IMAGES (with AMIs selected). The main area shows a table of instances. One instance is selected: Cloudera-Demo-Master (Instance ID: i-7f3431a5, Instance Type: m3.xlarge, Availability Zone: us-west-2c, Status: running). Below the table, detailed information for this instance is shown:

Instance state	running	Public IP	54.201.147.59
Instance type	m3.xlarge	Elastic IP	-
Private DNS	ip-172-31-10-53.us-west-2.compute.internal	Availability zone	us-west-2c
Private IPs	172.31.10.53	Security groups	default, view rules
Secondary private IPs		Scheduled events	No scheduled events
VPC ID	vpc-cd510ca5	AMI ID	ubuntu-trusty-14.04-amd64-server-20160114.5 (ami-

Install Cloudera Cluster on AWS



Connect to an instance from Windows using Putty



Install Cloudera Cluster on AWS



Connect to the instance

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

WARNING! Your environment specifies an invalid locale.

This can affect your user experience significantly, including the ability to manage packages. You may install the locales by running:

```
sudo apt-get install language-pack-UTF-8
```

or

```
sudo locale-gen UTF-8
```

To see all available language packs, run:

```
apt-cache search "^language-pack-[a-z][a-z]$"
```

To disable this message for all users, run:

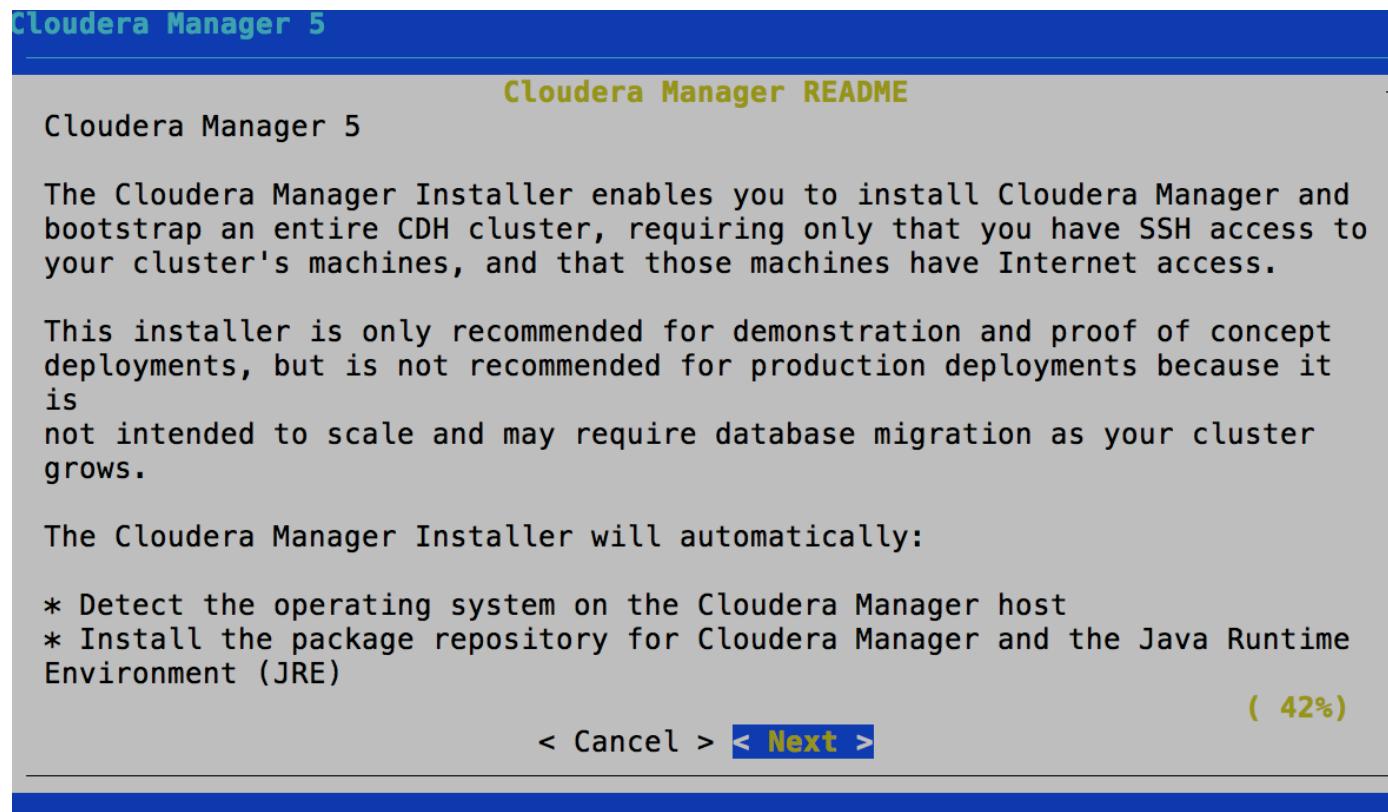
```
sudo touch /var/lib/cloud/instance/locale-check.skip
```

```
ubuntu@ip-172-31-1-242:~$
```

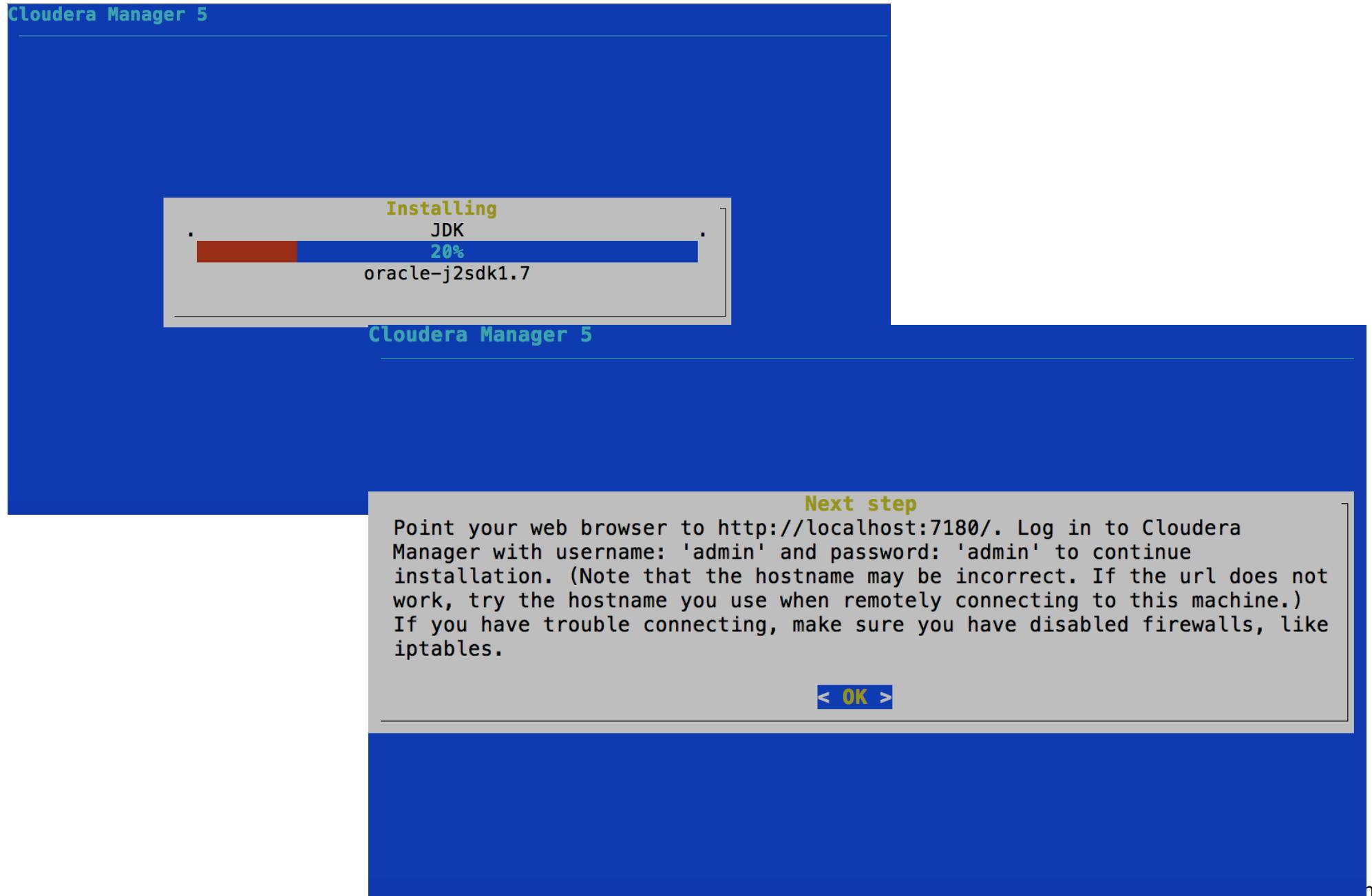
Installing Cloudera on EC2

Download Cloudera Manager

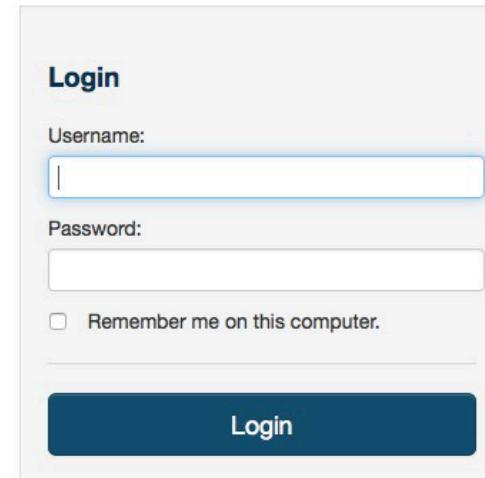
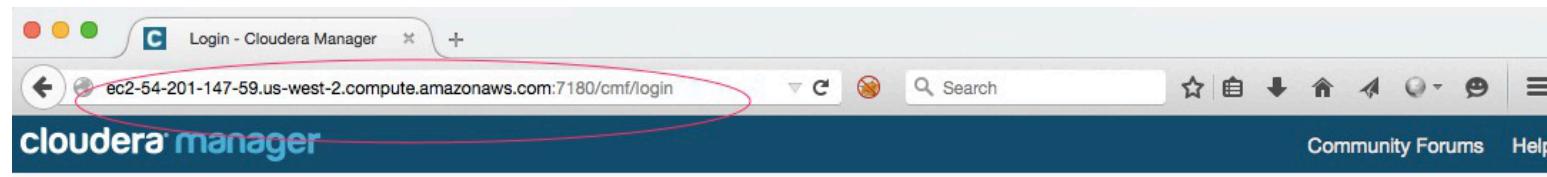
- 1) Type command > **wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin**
- 2) Type command > **chmod u+x cloudera-manager-installer.bin**
- 3) Type command > **sudo ./cloudera-manager-installer.bin**



Installing Cloudera on EC2



Wait several minutes for the Cloudera Manager Server to complete its startup. Then running web browser: [http:// public-dns: 7180](http://public-dns:7180)



The image shows the "Login" form for Cloudera Manager. It features a light gray background with a white input field. The form includes:

- Login** heading at the top center.
- Username:** A text input field with a placeholder character.
- Password:** A text input field.
- Remember me on this computer.**: A checkbox followed by the label.
- Login**: A large, prominent blue button at the bottom.

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 ("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 OF THIS AGREEMENT.

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

[Back](#)

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 Cloudera Express . Your cluster and your data will remain unaffected.	Annual Subscription Upload License
Node Limit	Unlimited	Unlimited	Unlimited
CDH	✓	✓	✓
Core Cloudera Manager Features	✓	✓	✓
Advanced Cloudera Manager Features		✓	✓
Cloudera Navigator		✓	✓

[Continue](#)

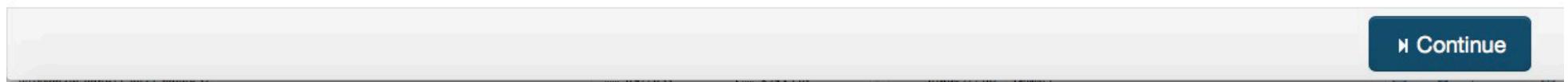
Thank you for choosing Cloudera Manager and CDH.

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 instances <private ip> addresses in the cluster

cloudera manager

Support ▾ admin ▾

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/or IP addresses using [patterns](#).

172.31.10.50, 172.31.10.51, 172.31.10.52, 172.31.10.53

SSH Port:

◀ Back

Continue ▶

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/or IP addresses using [patterns](#).

4 hosts scanned, 4 running SSH.

New Search

<input checked="" type="checkbox"/> Expanded Query	Hostname (FQDN)
<input checked="" type="checkbox"/>	172.31.10.50 ip-172-31-10-50.us-west-2.compute.i
<input checked="" type="checkbox"/>	172.31.10.51 ip-172-31-10-51.us-west-2.compute.i
<input checked="" type="checkbox"/>	172.31.10.52 ip-172-31-10-52.us-west-2.compute.i
<input checked="" type="checkbox"/>	172.31.10.53 ip-172-31-10-53.us-west-2.compute.i

cloudera manager

Support - admin -

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](#)

Select the version of CDH CDH-5.6.0-1.cdh5.6.0.p0.45 CDH-4.7.1-1.cdh4.7.1.p0.47

Versions of CDH that are too new for this version of Cloudera Manager (5.6.0) will not be shown.

Additional Parcels ACCUMULO-1.6.0-1.cdh5.1.4.p0.116

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, directories which in the regular mode are created automatically by the Agent, must be set up for the configured user.

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

Single User Mode



Cluster Installation

Installation completed successfully.

4 of 4 host(s) completed successfully.

Hostname	IP Address	Progress	Status	Details
ip-172-31-10-50.us-west-2.compute.internal	172.31.10.50	<div style="width: 100%; background-color: #2e6b2e;"></div>	Installation completed successfully.	Details
ip-172-31-10-51.us-west-2.compute.internal	172.31.10.51	<div style="width: 100%; background-color: #2e6b2e;"></div>	Installation completed successfully.	Details
ip-172-31-10-52.us-west-2.compute.internal	172.31.10.52	<div style="width: 100%; background-color: #2e6b2e;"></div>	Installation completed successfully.	Details
ip-172-31-10-53.us-west-2.compute.internal	172.31.10.53	<div style="width: 100%; background-color: #2e6b2e;"></div>	Installation completed successfully.	Details

[Back](#)



[Back](#)

1 2 3 4 5 6 7 8

[Continue](#)



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, Hue, and Sqoop
- Core with HBase**
HDFS, YARN (MapReduce 2 Included), Z
- Core with Impala**
HDFS, YARN (MapReduce 2 Included), Z
- Core with Search**
HDFS, YARN (MapReduce 2 Included), Z
- Core with Spark**
HDFS, YARN (MapReduce 2 Included), Z
- All Services**
HDFS, YARN (MapReduce 2 Included), Z
- Custom Services**
Choose your own services. Services requi

This wizard will also install the Cloudera Manager Agent.

[Back](#)

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](#)

HBase

M Master × 1 New

ip-172-31-26-220.us-west-2.compute.i...

HBRE HBase REST Server

Select hosts

HBTS HBase Thrift Server

Select hosts

RS RegionServer × 3 New

Same As DataNode ▾

HDFS

NN NameNode × 1 New

ip-172-31-26-220.us-west-2.compute.i...

SNN SecondaryNameNode × 1 New

ip-172-31-26-220.us-west-2.compute.i...

B Balancer × 1 New

ip-172-31-26-220.us-west-2.compute.i...

HFS HttpFS

Select hosts

NFSG NFS Gateway

Select hosts

DN DataNode × 3 New

ip-172-31-26-[221-223].us-west-2.compute.i...

Hive

[Back](#)

1 2 3 4 5 6

[Continue](#)

Cluster Setup

* First Run Command

Status: **Running** Start Time: Jan 20, 4:41:25 PM

[Abort](#)

Details Completed 5 of 9 step(s).

All Failed Only Running Only

Step	Context	Start Time	Duration	Actions
➤  Deploy Client Configuration Successfully deployed all client configurations.	Cluster 1	Jan 20, 4:41:25 PM	15.96s	
➤  Start Cloudera Management Service, ZooKeeper Successfully completed 2 steps.		Jan 20, 4:41:41 PM	25.72s	
➤  Start HDFS Successfully completed 1 steps.				
➤  Start HBase, Solr Successfully completed 2 steps.				
➤  Start YARN (MR2 Included), Key-Value Store Indexer Successfully completed 2 steps.				
➤  Start Spark 0/1 steps completed.				
<div style="border: 1px solid #ccc; padding: 5px;">  Creating Hive Metastore Database Created Hive Metastore Database. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Creating Hive user directory Successfully created HDFS directory. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Creating Hive warehouse directory Successfully created HDFS directory. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Starting Hive Service Service started successfully. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Creating Oozie database Oozie database created successfully. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Installing Oozie ShareLib in HDFS Successfully installed Oozie ShareLib. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Starting Oozie Service Service started successfully. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Starting Hue Service Service started successfully. Details </div> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;">  Deploying Client Configuration Successfully deployed all client configurations. Details </div>				

[Back](#)

Cloudera Manager

cloudera manager Clusters Hosts Diagnostics Audits Charts Administration Search (Hotkey: /) Support admin

Home

30 minutes preceding January 20, 2016, 4:49 PM UTC

Status All Health Issues Configuration X 5 All Recent Commands Add Cluster

Try Cloudera Enterprise Data Hub Edition for 60 Days

Cluster 1 (CDH 5.5.1, Parcels)

- Hosts
- HBase
- HDFS
- Hive
- Hue
- Impala
- Key-Value Store...
- Oozie
- Solr
- Spark
- YARN (MR2 Incl...)
- ZooKeeper

Charts

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

- Cluster CPU**

percent

04:30 04:45

Cluster 1, Host CPU Usage Across Hosts 15%
- Cluster Disk IO**

bytes / second

04:30 04:45

Total Disk Bytes Read... 0 Total Disk Bytes Written 1.7M/s
- Cluster Network IO**

bytes / second

04:30 04:45

Total Bytes Read... 151K/s Total Bytes Transferred 7.8M/s
- HDFS IO**

bytes / second

04:30 04:45

Total Bytes Read... 1b/s Total Bytes Written 166b/s

Cloudera Management Service

- Cloudera...

Completed Impala Queries

Hadoop as a Service

Recommended

Amazon EMR

Azure HDInsight

Amazon EMR

Install a cluster on 4 EC2 AWS

Azure HD Insight

Install a cluster on 4 Virtual Servers Microsoft Azure

Launch a EMR Cluster



Storage Gateway
Hybrid Storage Integration

Database

- RDS
Managed Relational Database Service
- DynamoDB
Managed NoSQL Database
- ElastiCache
In-Memory Cache
- Redshift
Fast, Simple, Cost-Effective Data Warehousing
- DMS
Managed Database Migration Service

Networking

- VPC
Isolated Cloud Resources
- Direct Connect
Dedicated Network Connection to AWS
- Route 53
Scalable DNS and Domain Name Registration

Create and Use Standardized Products

Trusted Advisor
Optimize Performance and Security

Security & Identity

- Identity & Access Management
Manage User Access and Encryption Keys
- Directory Service
Host and Manage Active Directory
- Inspector
Analyze Application Security
- WAF
Filter Malicious Web Traffic
- Certificate Manager
Provision, Manage, and Deploy SSL/TLS Certificates

Analytics

- EMR
Managed Hadoop Framework
- Data Pipeline
Orchestration for Data-Driven Workflows
- Elasticsearch Service
Run and Scale Elasticsearch Clusters
- Kinesis
Work with Real-Time Streaming Data
- Machine Learning
Build Smart Applications Quickly and Easily

Application Services

- API Gateway
Build, Deploy and Manage APIs
- AppStream
Low Latency Application Streaming
- CloudSearch
Managed Search Service
- Elastic Transcoder
Easy-to-Use Scalable Media Transcoding
- SES
Email Sending and Receiving Service
- SQS
Message Queue Service
- SWF
Workflow Service for Coordinating Application Components

Enterprise Applications

- WorkSpaces
Desktops in the Cloud
- WorkDocs
Secure Enterprise Storage and Sharing Service
- WorkMail
Secure Email and Calendaring Service

AWS Marketplace

Find and buy software, launch with 1-Click and pay by the hour.

AWS re:Invent Announcements

Explore the next generation of AWS cloud capabilities. See what's new

Service Health

All services operating normally.

Updated: Oct 02 2016 07:40:00 GMT+0700

[Service Health Dashboard](#)

Create Cluster - Quick Options [Go to advanced options](#)

General Configuration

Cluster name

Logging [i](#)

S3 folder [b](#)

Launch mode Cluster [i](#) Step execution [i](#)

Software configuration

Vendor Amazon MapR

Release [i](#)

Applications Core Hadoop: Hadoop 2.7.2 with Ganglia 3.7.2, Hive 2.1.0, Hue 3.10.0, Mahout 0.12.2, Pig 0.16.0, and Tez 0.8.4
 HBase: HBase 1.2.2 with Ganglia 3.7.2, Hadoop 2.7.2, Hive 2.1.0, Hue 3.10.0, Phoenix 4.7.0, and ZooKeeper 3.4.8
 Presto: Presto 0.150 with Hadoop 2.7.2 HDFS and Hive 2.1.0 Metastore
 Spark: Spark 2.0.0 on Hadoop 2.7.2 YARN with Ganglia 3.7.2 and Zeppelin 0.6.1

Hardware configuration

Instance type [i](#)

Number of instances (1 master and 3 core nodes)

Connect to the instance

```
https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/  
26 package(s) needed for security, out of 34 available  
Run "sudo yum update" to apply all updates.  
Amazon Linux version 2016.09 is available.  
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file  
or directory
```

```
EEEEEEEEEEEEEEEEEE MMMMMMMMM          MMMMMMMMM RRRRRRRRRRRRRRRR  
E:::::::::::::E M:::::M          M:::::M R:::::::::R  
EE:::::EEEEEEEEE:::E M:::::M          M:::::M R:::::RRRRRR:::R  
   E:::E     EEEEE  M:::::M          M:::::M RR:::::R      R:::R  
   E:::E           M:::::M:::M          M:::M::::M   R:::R      R:::R  
   E:::::EEEEEEEEE  M:::::M M:::M M:::M M:::::M   R:::::RRRRRR:::R  
   E:::::::E       M:::::M M:::M:::M M:::::M   R:::::::::::RR  
   E:::::EEEEEEEEE  M:::::M M:::::M M:::::M   R:::::RRRRRR:::R  
   E:::E           M:::::M M:::::M M:::::M   R:::::R      R:::R  
   E:::E     EEEEE  M:::::M   MMM  M:::::M   R:::::R      R:::R  
EE:::::EEEEEEEEE:::E M:::::M           M:::::M   R:::::R      R:::R  
E:::::::E       M:::::M           M:::::M RR:::::R      R:::R  
EEEEEEEEEEEEEEEEEE MMMMMMMMM          MMMMMMM RRRRRRRR      RRRRRR
```

```
[hadoop@ip-172-31-39-165 ~]$
```