Hands-On Exercise: Cluster Installation

In this exercise, you will log in to the Cloudera Manager Admin Console to install, deploy and verify your cluster.

During this exercise, you will identify the hosts that Cloudera Manager will manage.

You will then be prompted to choose which CDP services you want to add in the cluster and to which machines you would like to add each role of each service.

At the end of this exercise, the Cloudera Manager and CDP services and roles will be deployed across your cluster as show below. (The services and roles added in this exercise are in blue).

	master-1	master-2	worker-1	worker-2	worker-3	cmhost
HDFS NameNode	✓					
HDFS Secondary NameNode		✓				
HttpFS	✓					
HDFS Balancer	✓					
HDFS DataNode			✓	✓		✓
Hive Metastore	✓					
HiveServer 2 on Tez	✓					
Oozie Server		✓				
Spark History Server		✓				
Spark Gateway	✓	✓				
YARN Resource Manager	✓					
YARN JobHistory Server	✓					
YARN NodeManager		✓	✓	✓		✓
Zookeeper Server	✓	✓	✓			
Cloudera Manager Server						✓
Cloudera Manager Server Database						✓
Cloudera Management Services						✓
Cloudera Manager Agent	✓	✓	✓	✓		✓

In subsequent exercises, you will add more services to your cluster.

After completing the installation steps, you will review a Cloudera Manager Agent log file and review processes running on a machine in the cluster.

All steps in this exercise that use a terminal window should be run on `cmhost`.

Log in to Cloudera Manager Admin UI

1. If Firefox is not yet running, launch it from the `cmhost` desktop's **Applications, Internet, Firefox**.

   Note: Firefox may take a minute to launch the first time it is started.

---

2. In Firefox, load the Cloudera Manager Admin Console by entering `http://cmhost:7180`.

   Note: If an "Unable to Connect" message appears, the Cloudera Manager server has not yet fully started. Wait several moments, and then attempt to connect again.

---

3. Log in with the username `admin` and password `admin`.

   The **Welcome to Cloudera Manager** page.

---

4. On the "Welcome to Cloudera Manager 7.3.1" page, select **Try Cloudera Data Platform for 60 days**. Add a checkmark in the box next to **Yes, I accept the Cloudera Standard License Terms and Conditions** and click **Continue**.

---

5. The **Add Cluster - Installation** page. Click **Continue**.

---

6. The **Cluster Basics** page will appear. Enter the cluster name or accept the default Cluster 1. Click **Continue**. A Regular Cluster will be created.

Install Cloudera Manager agents and create cluster

7. The **Specify Hosts** for your CDH cluster installation page appears.

   Type in the names of five of the six hosts in the environment, separated by spaces:

   ```
   cmhost master-1 master-2 worker-1 worker-2
   ```

   Note: Do **not** include `worker-3` at this time. That host will be used in a different exercise.

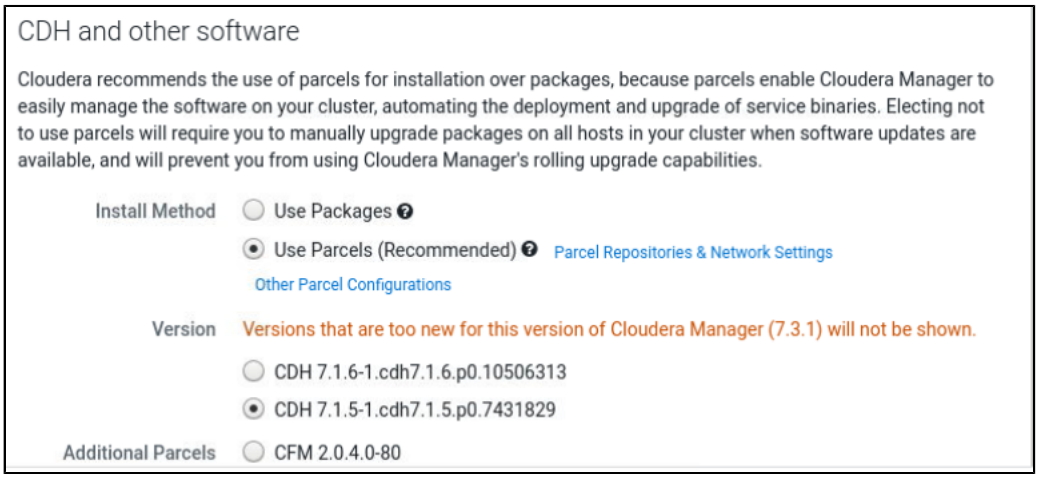
   Click **Search**. All five hosts should be found. Make sure that all five are selected, then click **Continue**.

Expanded Query	Hostname (FQDN)	IP Address	Currently Managed	Result
cmhost	cmhost.example.com	10.0.1.19	No	Host was successfully scanned.
master-1	master-1.example.com	10.0.5.114	No	Host was successfully scanned.
master-2	master-2.example.com	10.0.2.70	No	Host was successfully scanned.
worker-1	worker-1.example.com	10.0.8.213	No	Host was successfully scanned.
worker-2	worker-2.example.com	10.0.5.207	No	Host was successfully scanned.

The **Select Repository** page appears. In this screen, you will identify the location of the CDP parcel and the Cloudera Manager Agent installer (http://cmhost:8060/cm7.3.1/).

---

8. In the **Install Method** section, under **CDH and other software** area of the page, ensure the **Use parcels** option is selected.



---

9. The **Parcel Repository & Network Settings** pop-up will appear.

   In the **Parcel Repository and Network Settings** popup, click on each of the garbage bin icons ( ) to remove *ALL* the current repository references.

   Click **Save & Verify Configuration**. Once you have verified all errors have been removed click on **Close** to return to the **Select Repository** page.

---

10. Click the **Other Parcel Configurations** link

    In the **Other Parcel Configurations** popup we will make a change toone of the settings. Notice that there are severalproperties that can be set during this installation. Change the **Parcel Update Frequency** to 5 minute(s)

    Click **Save Changes** to return to the **Select Repository** page.

---

11. Ensure that Version `Cloudera Runtime CDH-7.1.5-1.cdh7.1.5.p0.7431829` is selected. Note that there will be several versions listed. Please ensure you have the correct one selected. Although we installed Cloudera Manager 7.3.1, we will be installing 7.1.5 to the cluster. We will upgrade the cluster to 7.1.6 later in the course.

---

12. Click **Continue** to save the installation repository settings.

---

13. The **Select JDK** page appears.

A supported version of the Oracle JDK is already installed in your exercise environment. Therefore, select the option to **Manually manage JDK**.

Click **Continue**.

---

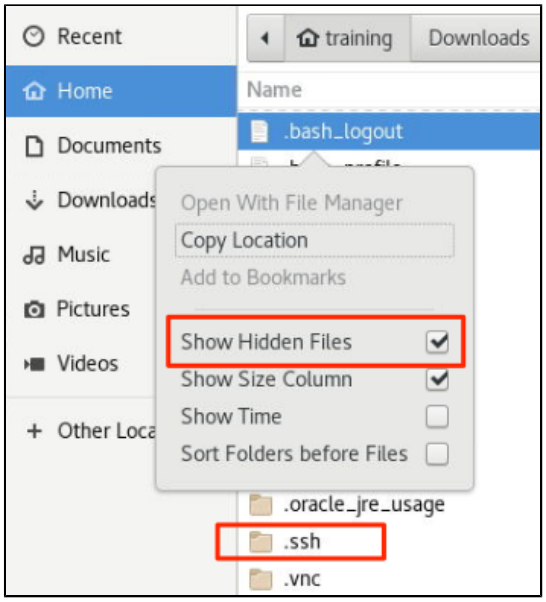
14. The **Enter Login Credentials** page appears.

a. Ensure that **Login To All Hosts As** is changed to **Another user** and enter **training**.

b. For **Authentication Method**, choose **All hosts accept same private key**.

c. Click the **Choose File** button to select a private key file.

Select **Home** in the location selector panel on the left. Then right-click in the **Name** area and select **Show Hidden Files**.



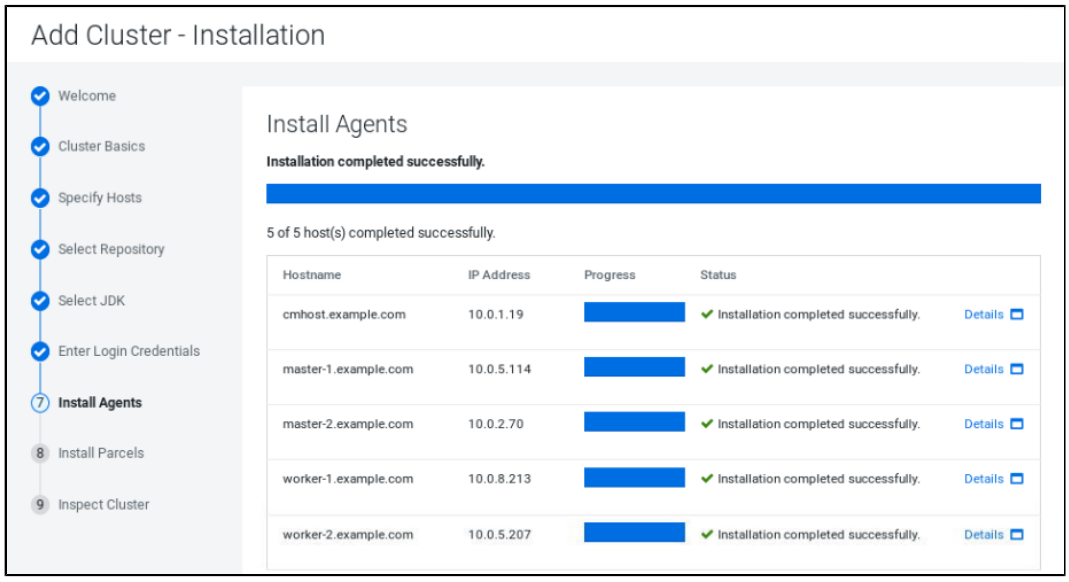
Double-click to open the `.ssh` directory, select the `id_rsa` file, and click **Open**.

d. Leave the passphrase fields blank fields and click **Continue**.

---

15. The **Install Agents** page appears.

Cloudera Manager installs the Cloudera Manager Agent on each machine. It may take up to five minutes.



After the Cloudera Manager Agent is successfully installed on all five machines, click **Continue** if needed.

---

16. In the next step, the parcel is downloaded, distributed, unpacked, and activated on all hosts in the cluster.

When the parcel is activated on all five hosts, click **Continue**.

---

17. The **Inspect cluster** page appears.

Select the **Inspect Network Performance** button. After a few moments, the results will appear. All validations should succeed.

Then click on the **Inspect Hosts** button. After a few moments, the results will appear. A warning that `kudu` is not part of group `hive` will appear. Click **I understand the risks of not running ...** and then click **Continue**.

If you wish you can review the reports from each inspection.

Click **Continue**.

Tip: The Host Inspector can be run on existing cluster hosts at any time from the Cloudera Manager admin console.

---

18. You have now completed installing Cloudera Manager and creating a cluster.

The **Add Cluster - Configuration** wizard will walk you through installing and deploying services on the cluster starts automatically.

Set up Cluster Services

19. The first page of the Add Cluster - Configuration wizard is the **Select Services** page.

Notice the note at the bottom of the screen stating "This wizard will also install the **Cloudera Management Service**."

Click **Custom Services**, which will display a table appears with a list of CDP service types.

---

20. Select the following services:

- HDFS
- Hive
- Hive on Tez
- Oozie
- Spark
- Tez
- YARN
- ZooKeeper

Double check that you have selected the correct services before continuing, then click **Continue**.

---

21. On the next page, specify which hosts in the cluster serve which roles.

Assign roles to hosts as shown in the table below.

Note: The instructions below have you assign the HDFS DataNode role to `cmhost`. This is contrary to standard recommendations, even on a small cluster. Follow the instructions as shown, for now. You can modify this setup later.

Tip: To assign a role to a particular host, click on a field with one or more hostnames in it. For example, the field under **SecondaryNameNode** might initially have the value `worker-1`. Change this to a different host by clicking on the field, which will open a window where you can choose a new host. Note that the interface will include '.example.com' in all host names, as in 'master-1.example.com'.

	Role	Node(s)
HDFS		
	NameNode	`master-1`
	SecondaryNameNode	`master-2`
	Balancer	`master-1`
	HttpFS	`master-1`
	NFS Gateway	Do not specify any hosts
	DataNode	Custom: `cmhost`, `worker-1`, `worker-2`
Hive		
	Gateway	Do not specify any hosts
	Hive Metastore Server	`master-1`
	WebHCat Server	Do not specify any hosts
	HiveServer2	Do not specify any hosts
Hive on Tez		
	Gateway	`master-1, master-2`
	HiveServer2	`master-2`
Cloudera Management Service		
	Service Monitor	`cmhost`
	Activity Monitor	`cmhost`
	Host Monitor	`cmhost`
	Reports Manager	`cmhost`
	Event Server	`cmhost`
	Alert Publisher	`cmhost`
	Telemetry Publisher	Do not specify any hosts
Oozie		
	Oozie Server	`master-2`
Spark		

	Role	Node(s)
	HistoryServer	master-2
	Gateway	master-1, master-2, cmhost
Tez		
	Gateway	cmhost, master-2
YARN		
	ResourceManager	master-1
	JobHistoryServer	master-1
	NodeManager	Same as DataNode
ZooKeeper		
	Server	cmhost, master-1, master-2

When you have finished assigning roles, carefully verify that your role assignments are correct. When you are certain that the settings are correct, click **Continue**.

---

22. The **Setup Database** page appears. This allows you to specify the database connection details for each service's database in the Cloudera Manager database system. The databases and access credentials were created when you ran `scm_prepare_database.sh` during the Cloudera Manager installation process.

Note: The Database Hostname should be set to `cmhost` or `cmhost.example.com` for each service. Either one will work.

Fill in the details as shown here.

Service	Database Hostname	Database Type	Database Name	Username	Password
Hive	cmhost	MySQL	metastore	hiveuser	password
Activity Monitor	cmhost	MySQL	amon	amonuser	password
Reports Manager	cmhost	MySQL	rman	rmanuser	password
Oozie Server	cmhost	MySQL	oozie	oozieuser	password

Click **Test Connection** to verify that Cloudera Manager can connect to the MySQL databases you created in an earlier exercise in this course.

After you have verified that all connections are successful, click **Continue**.

---

23. The **Review Changes** page appears. Leave the default values for all settings. Click **Continue**.

---

24. The **First Run Command** page appears. Progress messages appear while cluster services are created and started. When all the cluster services have started, click **Continue**.

---

25. The **Summary** page appears.

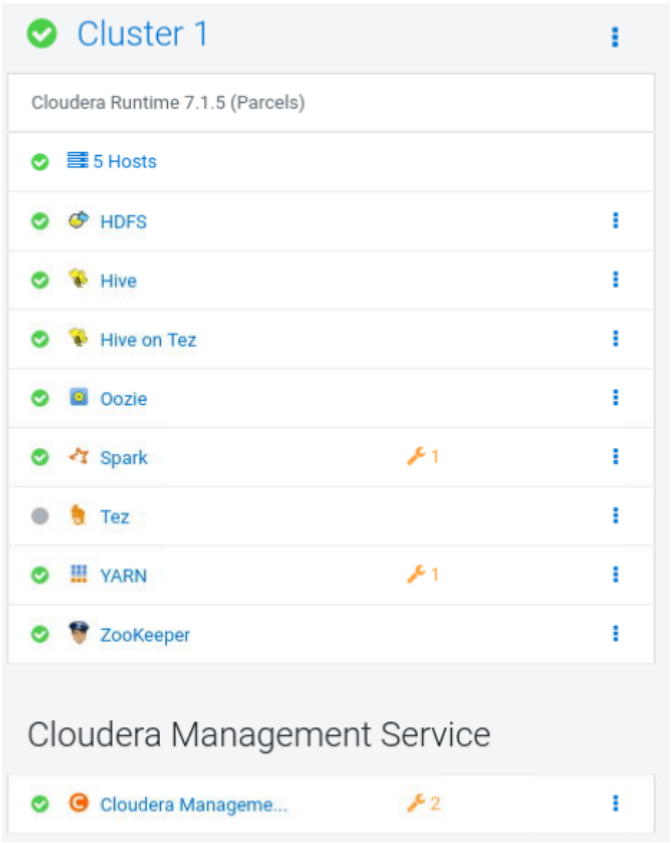
The page indicates that services have been added and are now configured and running on your cluster. Click **Finish**.

The Cloudera Manager home page will appear.

---

26. On the home page, confirm that all services are running correctly—that is, they should have a green checkmark indicator. If any of the services are not running, open the drop-down menu to the right of the service name and select **Restart**.

> **Configuration Warnings**
>
> The configuration warnings—as seen in the screenshot below—are expected, and indicate that, although Cloudera Management Services and the Service Monitor are in good health, but they do not have the recommended amount of memory available. There is also an indicator that YARN needs the YARN Queue Manager which we will install later. There is also an erroneous message indicating that we need a Spark Gateway on master-1 and master-2, which already exist.
>
> In a production deployment, you would need to ensure these warnings were addressed. However, in the exercise environment, you can safely ignore them at this time.



Cluster installation is now complete.

This is the end of the exercise.