Installing Cloudera QuickStart VM

Cloudera provides virtual machine (VM) images of complete Apache Hadoop clusters, making it quick and easy to get started with Cloudera CDH.

Note: These are free for personal use, but do require you to register your details on the Cloudera website prior to download. *Remember to check your system meets the minimum requirements.*

I'm familiar with running VMs...

Go ahead and download the appropriate Cloudera Quickstart VM if you are already familiar with running virtual machines, and have VM software installed. Clone the example files from the FLbigdataD2D GitHub repository once you have Cloudera QuickStart up and running.

I do not have VM software installed...

Follow the instructions to first install VirtualBox (virtual machine client) and then get started with the Cloudera Quickstart VM.

Note: The Cloudera Quickstart VM download file is approx. 5GB, so download time will vary depending on your connection speed. There are online calculators you can use if you'd like an estimate of how long the download may take. Conduct a web search for: *how long will my download take.*

1. Install Virtual Box (virtual machine client)

a) Download VirtualBox.

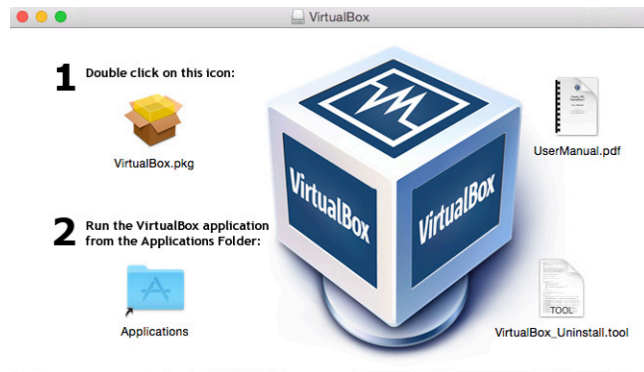   Ensure you download the correct version of VirtualBox for your operating system.
   PC users: VirtualBox 5.0.14 for Windows hosts (112MB)
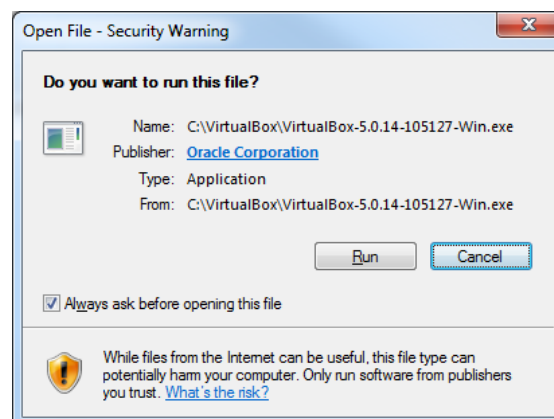   **MAC users:** VirtualBox 5.0.14 for OS X hosts (86.2MB)

b) Install VirtualBox.

MAC users: Double click the .dmg file you have downloaded and follow the instructions.
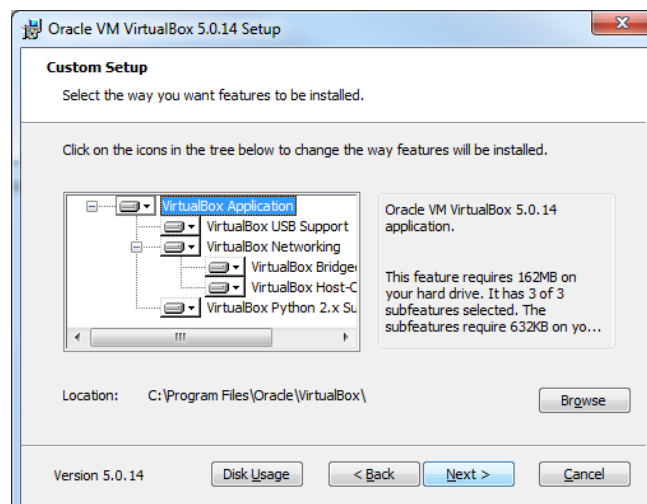

PC users: Double click the .exe file you have downloaded and follow the instructions.
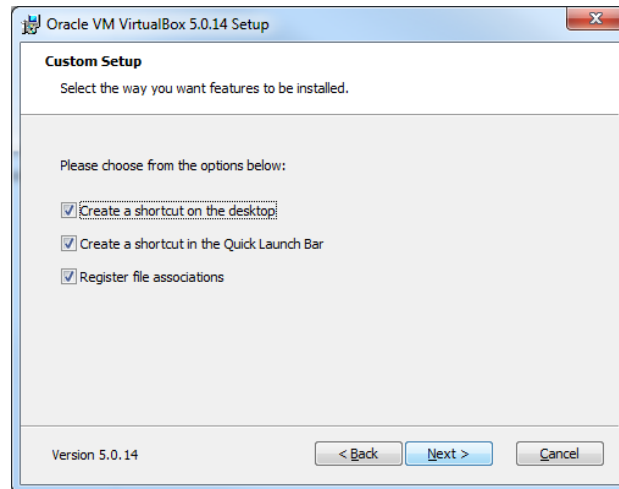
i. Double click .exe file and click **Run.**

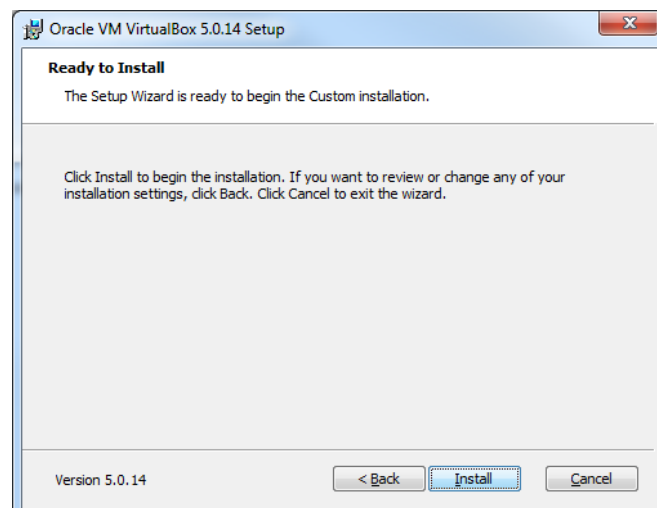*ii.* Click **Next** accepting the suggested custom setup to install features.

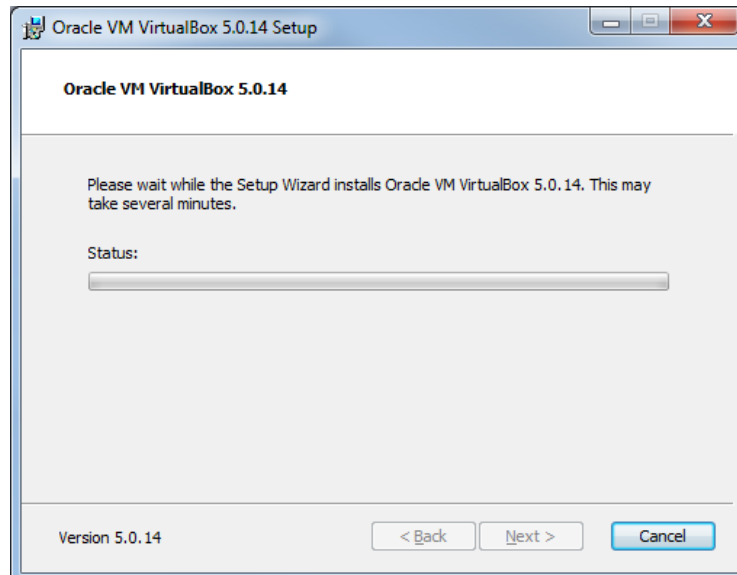*iii.* Click **Next** accepting the suggested custom setup.

iv. Click **Yes**.

v. Click **Install**.

vi. Wait for the install to complete. You may receive prompts to confirm the installation as it progresses. Click **Yes** for each prompt to continue the installation.
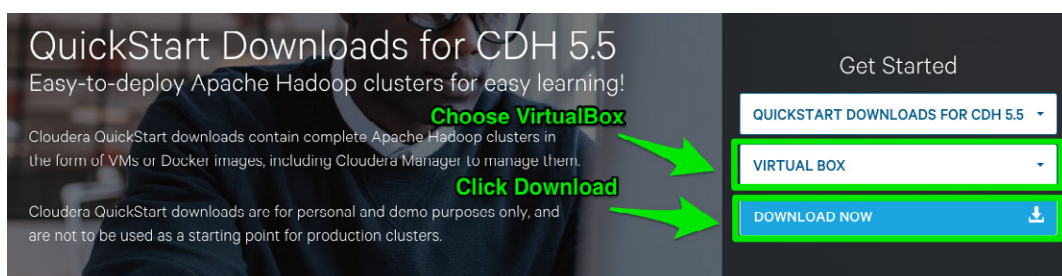

vii. Click **Finish** when the installation is complete.

2. Download ClouderaCDH VM for VirtualBox.

a) Select the virtual machine client you are using from the drop down list (VirtualBox).

b) Click **Download now** to begin downloading the installation zip archive file for Cloudera QuickStart VM.

c) Register and sign in or complete the product interest form.

Before you can download the file you must complete this step. You are required to complete all fields before you can continue. Cloudera may use your details to contact you.

3. Extract files from the Cloudera Quickstart VM zip archive.

Ensure you have sufficient space to extract the archive if you encounter problems. Alternatively try using 7-zip (Windows) or TheUnArchiver (Mac) which work well with large files.
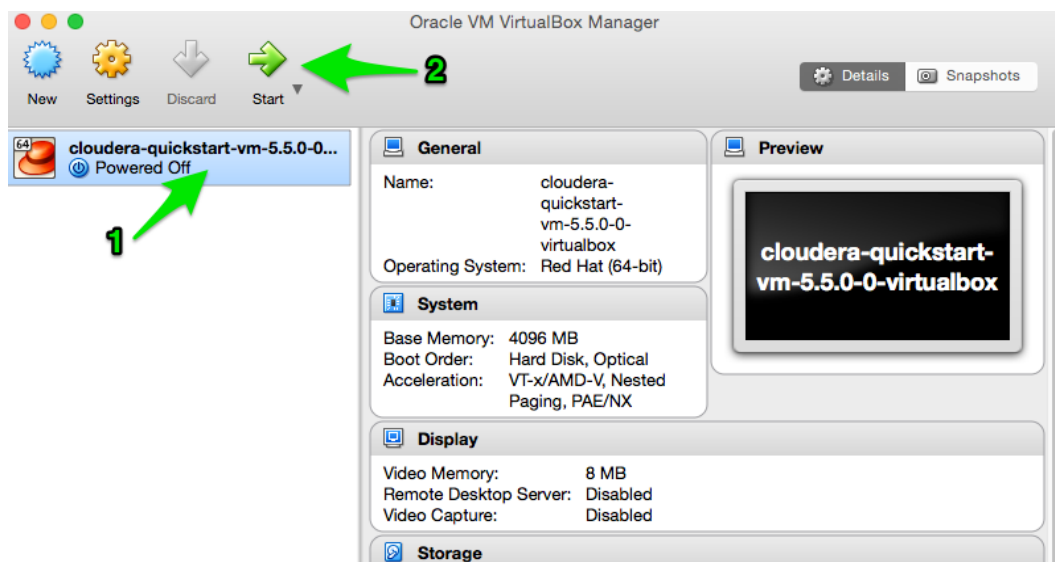
4. Open the Cloudera Quickstart VM in VirtualBox.

a)   Launch VirtualBox.

b)   Click on **File - Import appliance** and select the .ovf file from the location where you have saved the extracted Cloudera Quickstart VM files. Click **Next**.

Leave 'Reinitialise the MAC address of all network cards' unticked.
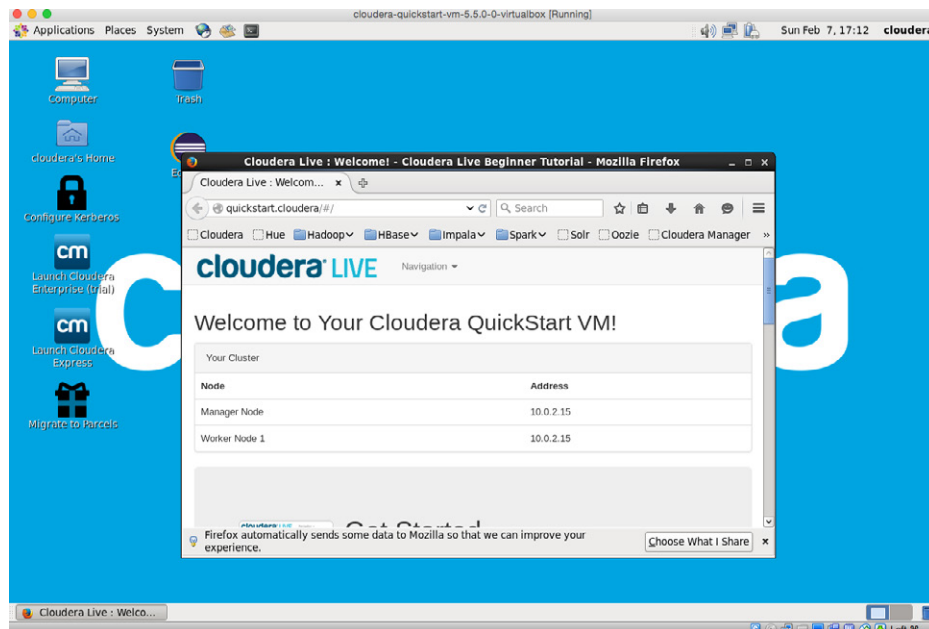
c)   Click Import.

5. Start your Cloudera Quickstart VM instance.

a)   Select **cloudera-quickstart-vm** from the left panel.

b)   Click the **Start** (green arrow) icon.

Note: The VM can take some time to start up. You'll see the blue Cloudera desktop and the Welcome to Cloudera Quickstart VM page once it's ready.
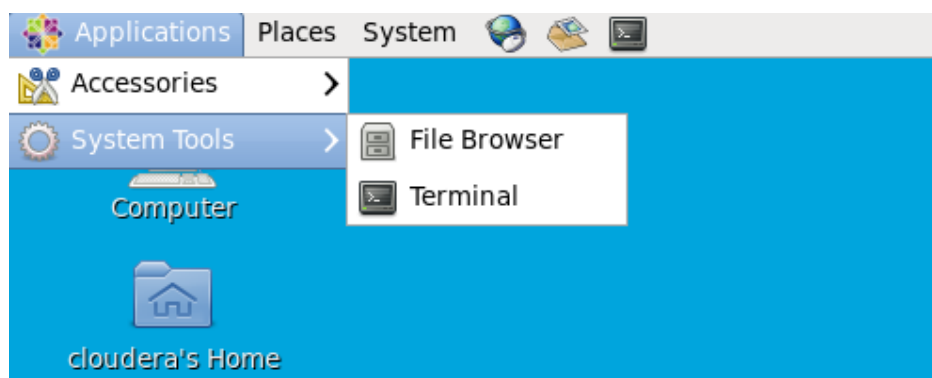

Tip: Once you are finished using Cloudera, remember to shut down the VM by choosing **System - Shut down**.

6. Clone the example files from the *FLbigdataD2D GitHub repository*.

a) Open the Terminal.

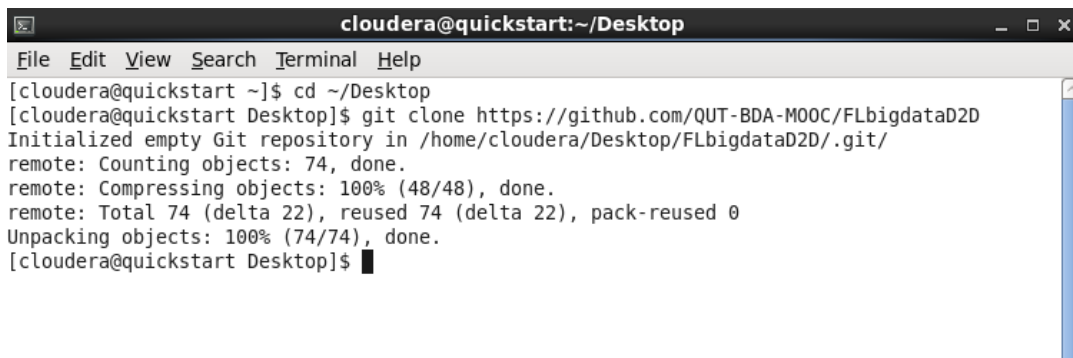Select **Applications > System Tools > Terminal**.

b) Change to the Desktop folder.

Type **cd ~/Desktop** in the Terminal window.

c) Clone the example files from GitHub.

Type **git clone** https://github.com/QUT-BDA-MOOC/FLbigdataD2D in the Terminal window.

```
cloudera@quickstart:~/Desktop

File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ cd ~/Desktop
[cloudera@quickstart Desktop]$ git clone https://github.com/QUT-BDA-MOOC/FLbigdataD2D
Initialized empty Git repository in /home/cloudera/Desktop/FLbigdataD2D/.git/
remote: Counting objects: 74, done.
remote: Compressing objects: 100% (48/48), done.
remote: Total 74 (delta 22), reused 74 (delta 22), pack-reused 0
Unpacking objects: 100% (74/74), done.
[cloudera@quickstart Desktop]$ 
```