

How to Install and Configure Cloudera QuickStart VM

Cloudera is a US-based software company that provides a software platform for data engineering, data warehousing, machine learning and analytics that runs in the cloud or on premises. Cloudera started as a hybrid open-source Apache Hadoop distribution, CDH (Cloudera Distribution Including Apache Hadoop), that targeted enterprise-class deployments of that technology. Cloudera states that more than 50% of its engineering output is donated upstream to the various Apache-licensed open source projects (Apache Spark, Apache Hive, Apache Avro, Apache HBase, and so on) that combine to form the Apache Hadoop platform. Cloudera is also a sponsor of the Apache Software Foundation.

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.*

Cloudera offers software, services and support in five bundles available both on-premise and across multiple cloud providers:

Cloudera Enterprise Data Hub - Cloudera's comprehensive data management platform including all of Data Science & Engineering, Operational DB, Analytic DB, and Cloudera Essentials.

Cloudera Analytic DB - Cloudera's technologies built on the core Cloudera Essentials platform.

Cloudera Operational DB - Cloudera's high-scale NoSQL technologies for real-time, data applications built on the core Cloudera Essentials platform.

Cloudera Data Science and Engineering - Cloudera's technologies that enable efficient, high-scale data processing, data science, and machine learning on top of the Core Essentials platform.

Cloudera Essentials - Cloudera's core data management platform for fast, easy, and secure large-scale data processing that includes Cloudera's enterprise-ready management capabilities (Cloudera Manager) and open source platform distribution (CDH).

Cloudera also offers a managed-service offering on the cloud:

Altus Data Engineering which provides a cloud-native offering of Cloudera Data Engineering.

Cloudera also offers the following free software versions:

Cloudera Express - includes Cloudera's CDH open-source platform and a no-charge version of its deployment, monitoring, and administration suite, Cloudera Manager.

CDH (Cloudera's Distribution including Apache Hadoop) - is Cloudera's 100% open source platform distribution including Apache Hadoop, Apache Spark, Apache Impala, Apache Kudu, Apache HBase, and many more.

Additional technologies in addition to its open-source distribution:

Cloudera Director - a tool distributed without charge that enables easy deployment of cloud-native Cloudera clusters on-demand across multiple cloud providers.

Cloudera Data Science Workbench - A data science tool for secure collaboration and model development add-on for Cloudera Enterprise Data Engineering and Data Science as well as Cloudera Enterprise Data Hub.

Cloudera Navigator - critical data governance functionality for Cloudera's platform, offering capabilities such as data discovery, audit, lineage, metadata management, encryption, encryption key management, and policy enforcement to help meet regulatory compliance requirements.

Cloudera Navigator Optimizer - a software-as-a-service tool to assist in identifying, migrating, and tuning traditional database workloads to Cloudera's platform as well as analyze and tune workloads running on Cloudera's platform.

Cloudera Manager - an administrative tool for fast, easy, and secure deployment, monitoring, alerting, and management of Cloudera's platform.

I do not have VM software installed...

Note: The Cloudera Quickstart VM download

(https://www.cloudera.com/downloads/quickstart_vms/5-13.html) 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. How to Install Virtual Box (virtual machine client)

a) Download VirtualBox.

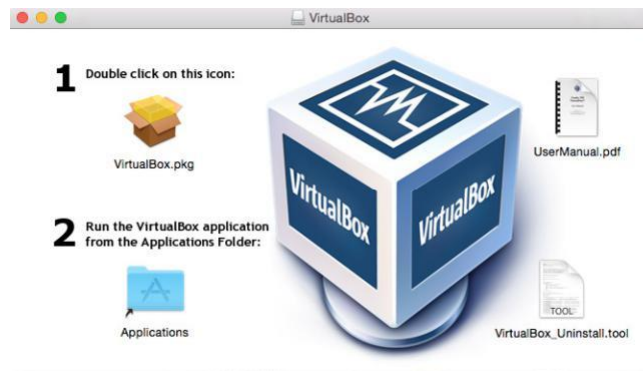
Ensure you **download the correct version of VirtualBox** for your operating system. **PC users:** VirtualBox 6.0.12 for Windows hosts (112MB)

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.

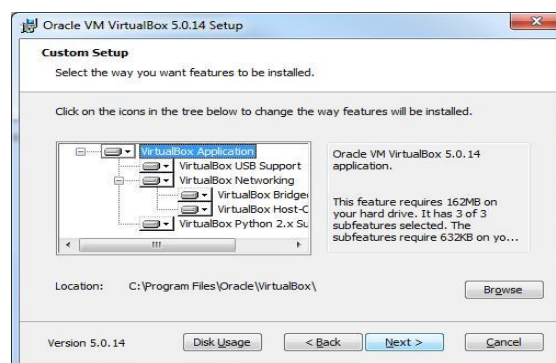


PC users: Double click the .exe file you have downloaded and follow the instructions.

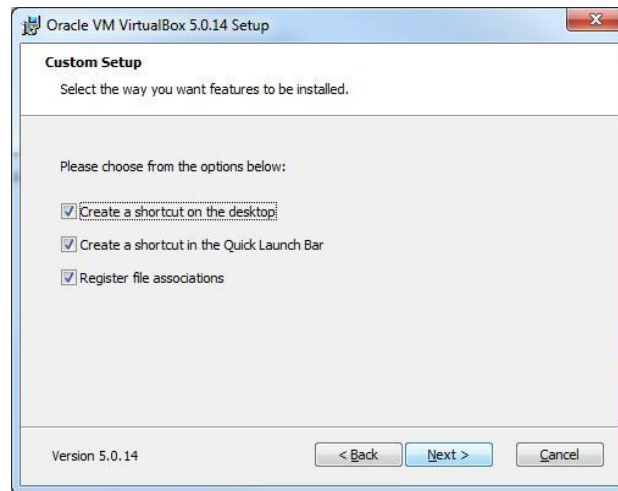
i. Double click .exe file and click **Run**.



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



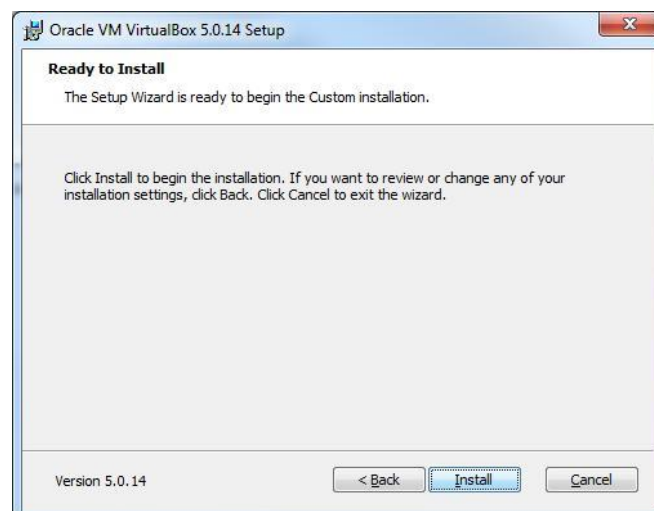
iii. Click **Next** accepting the suggested custom setup.



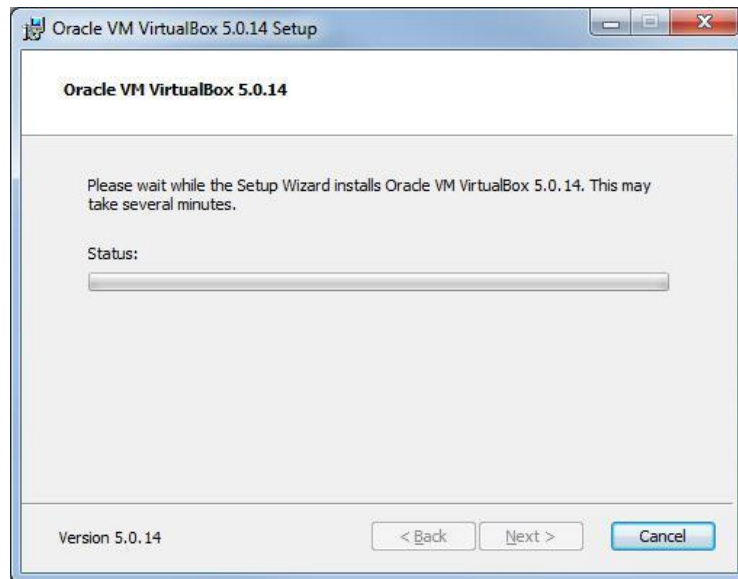
iv. Click **Yes**.



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.

- Select the virtual machine client you are using from the drop down list (VirtualBox).
- Click **Download now** to begin downloading the installation zip archive file for Cloudera QuickStart VM.
- 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.

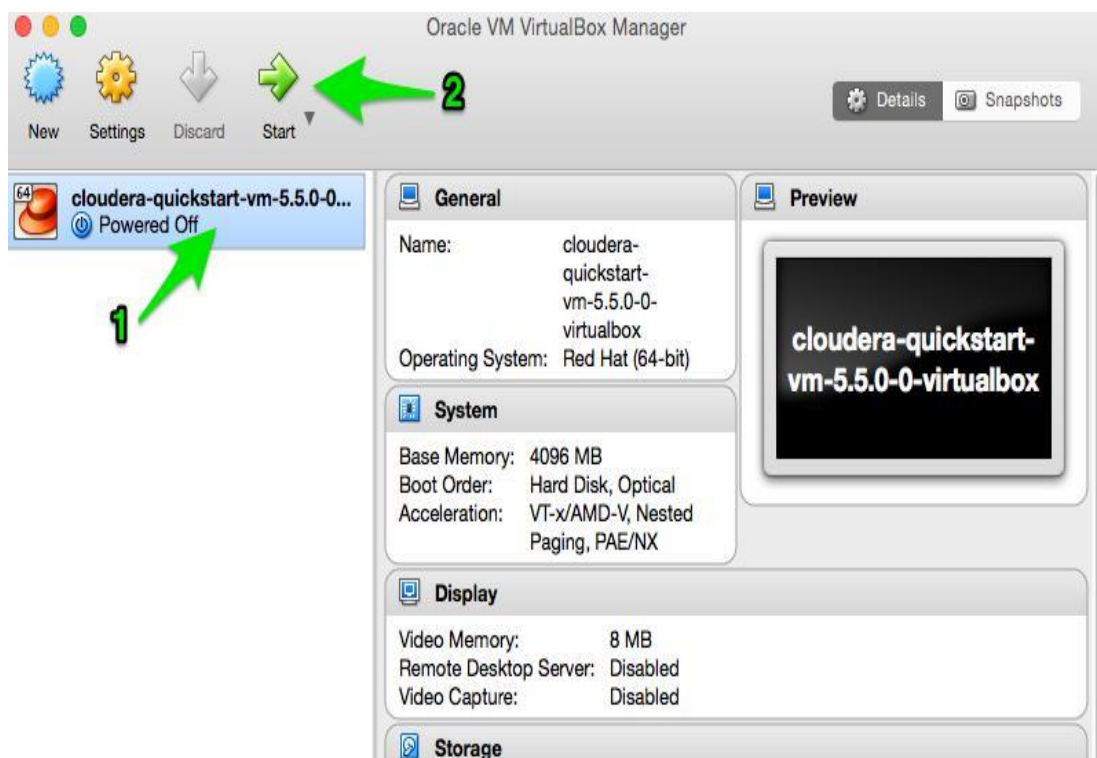
Ensure you have sufficient space to extract the archive if you encounter problems. Alternatively try using **7-zip** (Windows) or **TheUnArchiver** (Mac) which work well with large files.

4. Open the Cloudera Quickstart VM in VirtualBox.

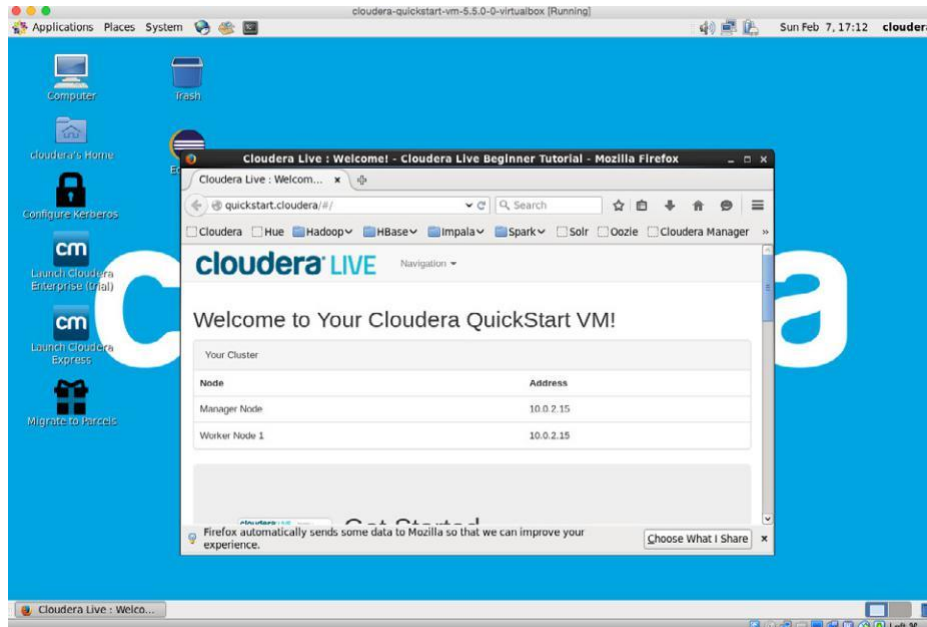
- a) Launch VirtualBox.
- b) Click on **File - Import appliance** and select the **.ovf** file from the location where you have saved the extracted Cloudera Quickstart VM files. Click **Next**.
Leave 'Reinitialise the MAC address of all network cards' unticked.
- c) Click Import.

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.



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

a) *Now, you can* open the Terminal.

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

