Cloudera VM Download and SetUp Notes:

Prerequisites

- These 64-bit VMs require a 64-bit host OS and a virtualization product that can support a 64-bit guest OS.
- To use a VMware VM, you must use a player compatible with WorkStation 8.x or higher:
  - Player 4.x or higher
  - Fusion 4.x or higher
- You can use older versions of WorkStation to create a new VM using the same virtual disk (VMDK file), but some features in VMware Tools are not available.

- The amount of RAM required varies by the runtime option you choose:

CDH and Cloudera Manager Version	RAM Required by VM
CDH 5 (default)	4+ GiB*
Cloudera Express	8+ GiB*
Cloudera Enterprise (trial)	12+ GiB*

Get and Install Virtual Box

Instructions will be provided for Virtual Box here. The professor will be running a version of the Virtual Machine (VM) on VirtualBox. Cloudera appears to be transitioning to their new platform and away from the current VM's. Some links below may be discontinued. Virtual Box is recommended but you may be able to set up a Docker image as described in some of the documents in the Google Drive project folder but that will not be supported for troubleshooting.

Virtual Box - Free for Mac/Windows/Linux

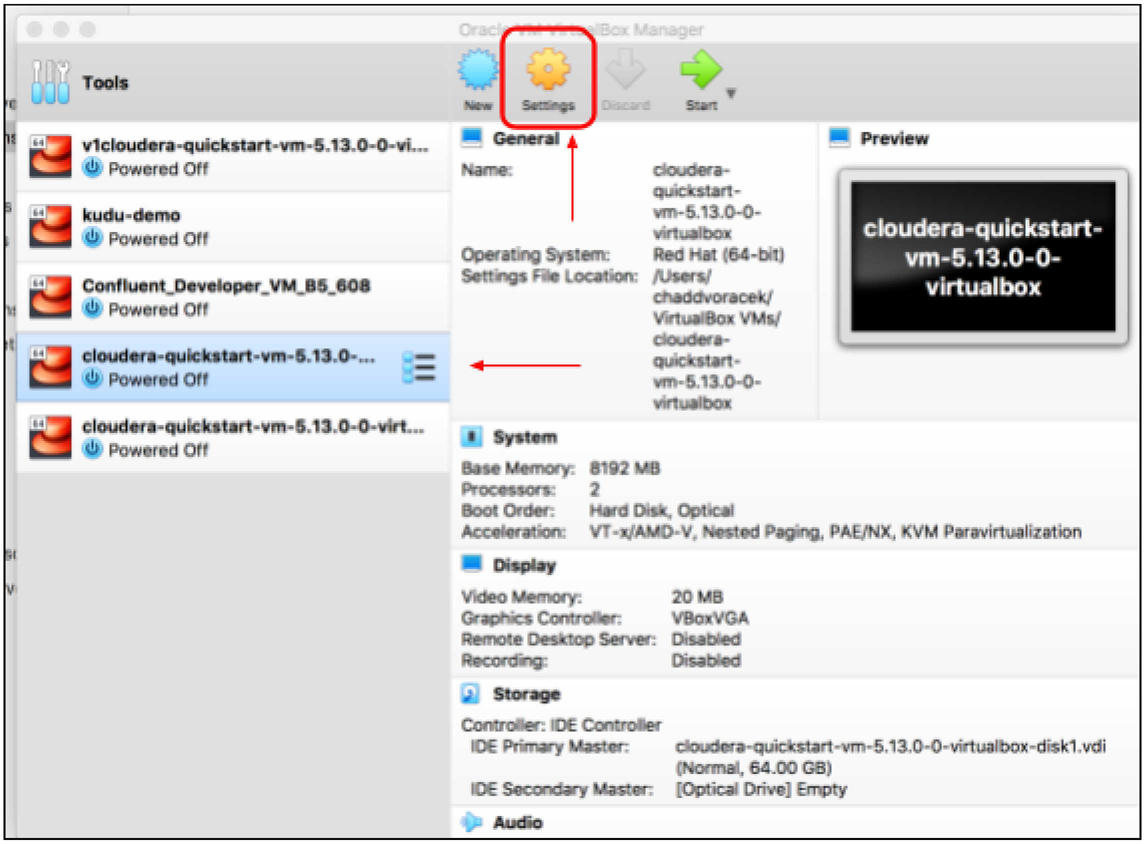
Download the Virtual Machine

Download the VM directly from class Google Drive Project Folder.

1. Review the system requirements to make sure your computer can run the software. Only one member of the team MUST have this set up for testing, although I recommend all team members work through the exercises.
2. Please review all documentation on these links as they contain important information you will need in order to successfully use the VM. Much of the same information can be found in the other documents in the Google Drive under the project folder.
    a. https://www.cloudera.com/documentation/enterprise/5-13-x/topics/cloudera_quickstart_vm.html
    b. https://docs.cloudera.com/documentation/enterprise/5-13-x/topics/quickstart_vm_administrative_information.html
    c. https://www.cloudera.com/developers/get-started-with-hadoop-tutorial/setup.html This tutorial walks you through a number of hands-on steps in learning how to use the Cloudera envionrment and queries against a data set. This document is also listed in the lab section for the project, highly recommended. The same tutorial is in the project folder in Google Drive. *cloudera-msazure-hadoop-deployment-guide.pdf*
3. Cloudera 5.13x. Documentation and Help Guide
    a. https://docs.cloudera.com/documentation/enterprise/5-13-x/topics/introduction.html
    b. https://docs.cloudera.com/documentation/enterprise/5-13-x/categories/hub_quickstart.html
    c. There is a lot of information and user guides here. Explore as needed.

Get VM Running - VirtualBox Setup

1. Unzip the VM
    a. Windows users, it is recommended to use 7zip. It is a free open source application that is more reliable for unzipping complex compresses files like this one.
    b. Mac users double click the file and Mac will take care of the rest.
2. Import Appliance (Select correct VM you just unzipped)
3. Change settings (recommended 8192 MB Memory and 2 Processors)
    a. Before starting the VM. Click the one you want to adjust. Click on Settings.


Oracle VM VirtualBox Manager
Tools
New
Settings
Discard
Start
v1cloudera-quickstart-vm-5.13.0-0-vi...
Powered Off
kudu-demo
Powered Off
Confluent_Developer_VM_B5_608
Powered Off
cloudera-quickstart-vm-5.13.0-...
Powered Off
cloudera-quickstart-vm-5.13.0-0-virt...
Powered Off
Preview
cloudera-quickstart-vm-5.13.0-0-virtualbox

General

Name: cloudera-quickstart-vm-5.13.0-0-virtualbox
Operating System: Red Hat (64-bit)
Settings File Location: /Users/chaddvoracek/VirtualBox VMs/cloudera-quickstart-vm-5.13.0-0-virtualbox

System

Base Memory: 8192 MB
Processors: 2
Boot Order: Hard Disk, Optical
Acceleration: VT-x/AMD-V, Nested Paging, PAE/NX, KVM Paravirtualization

Display

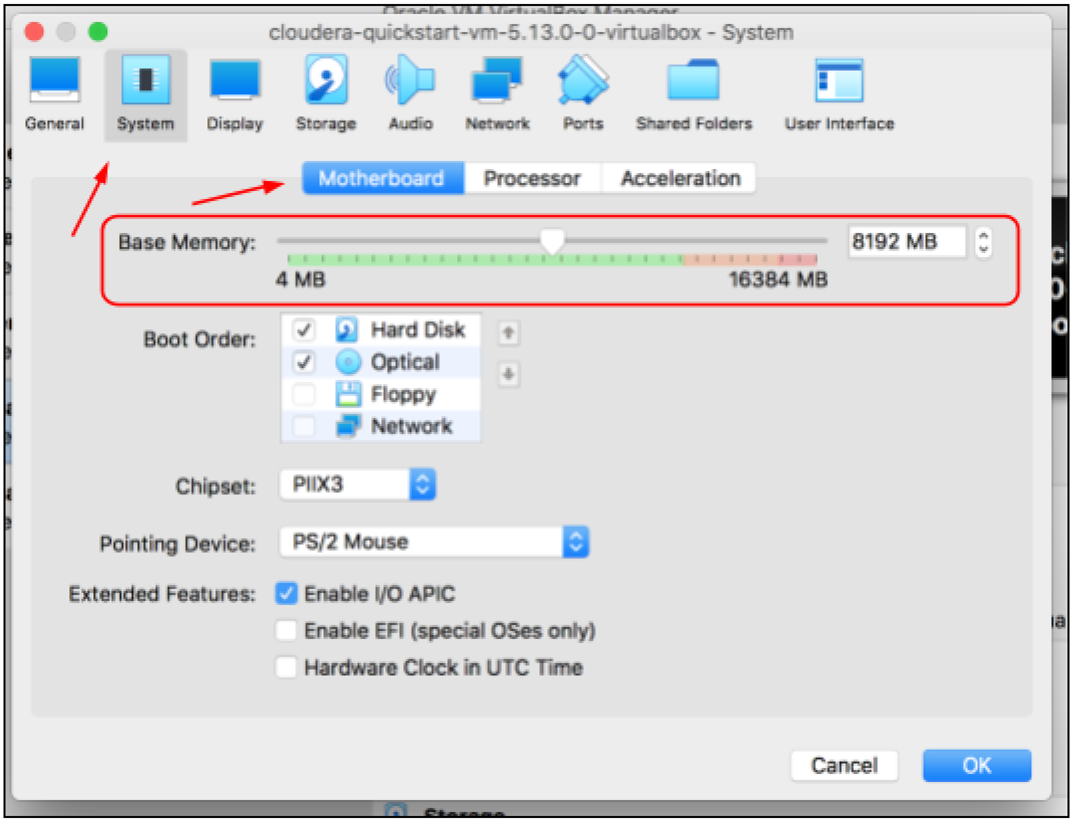
Video Memory: 20 MB
Graphics Controller: VBoxVGA
Remote Desktop Server: Disabled
Recording: Disabled

Storage

Controller: IDE Controller
IDE Primary Master: cloudera-quickstart-vm-5.13.0-0-virtualbox-disk1.vdi (Normal, 64.00 GB)
IDE Secondary Master: [Optical Drive] Empty

Audio

    b. Adjust Settings



cloudera-quickstart-vm-5.13.0-0-virtualbox - System

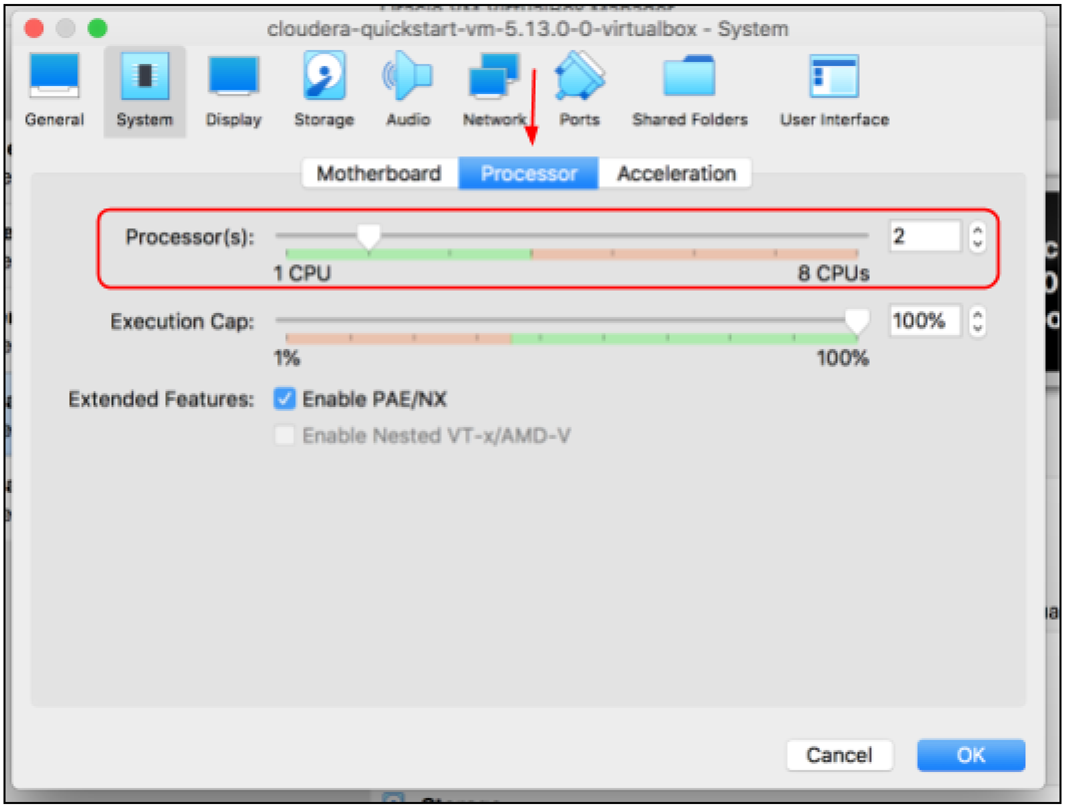
Base Memory: 8192 MB (4 MB – 16384 MB)

Boot Order:
- ☑ Hard Disk
- ☑ Optical
- □ Floppy
- □ Network

Chipset: PIIX3

Pointing Device: PS/2 Mouse

Extended Features:
- ☑ Enable I/O APIC
- □ Enable EFI (special OSes only)
- □ Hardware Clock in UTC Time



cloudera-quickstart-vm-5.13.0-0-virtualbox - System

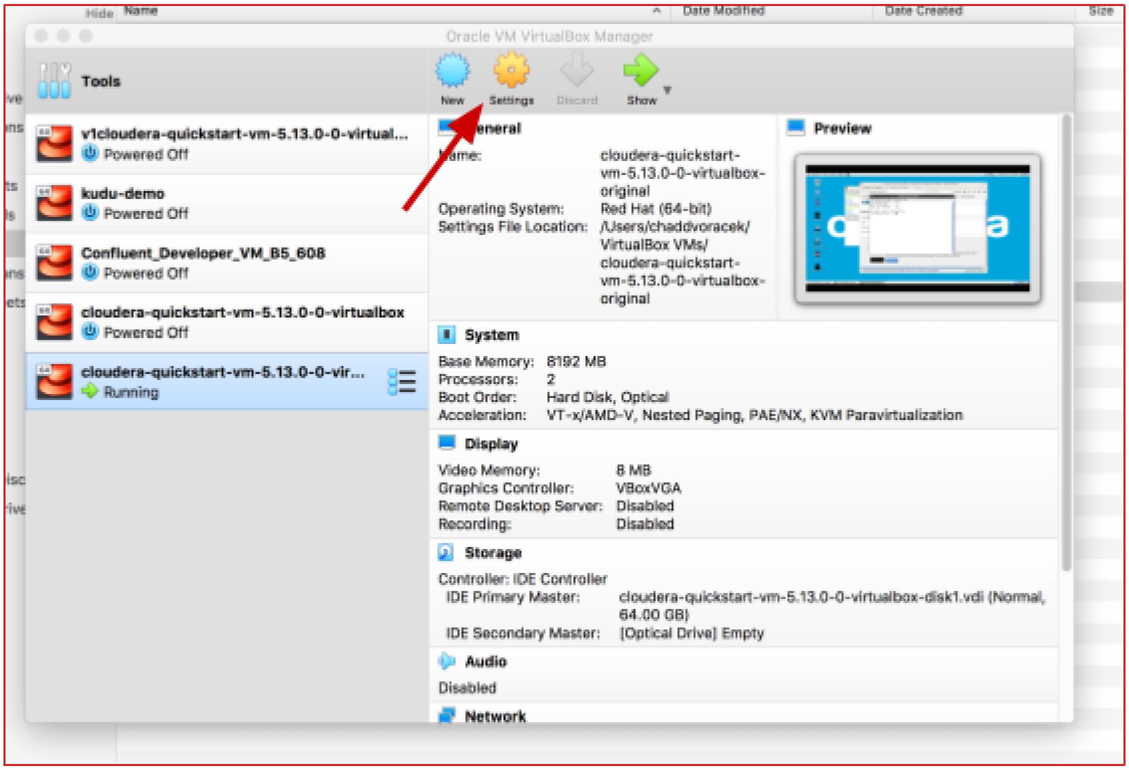
Processor(s): 2 (1 CPU – 8 CPUs)

Execution Cap: 100% (1% – 100%)

Extended Features:
- ☑ Enable PAE/NX
- □ Enable Nested VT-x/AMD-V

4. Start the VM
    a. Certain tasks or services will require your username and password. This is stated in the documentation but for reference also noted here.
        i. Username: cloudera
        ii. Password: cloudera

5. Double Click the Desktop Icon for “Launch Cloudera Express”. Wait for this to complete. This will give you access to Cloudera Manager.
6. You can enable functionality like a shared folder, copy/paste, and drag/drop functionality between your computer and the VM. Depending on your operating system, and VM/software tool you have chosen, these features may vary. Please search for instructions via the internet to enable these features.
7. If you want to ssh directly from your computer into the VM edgenode:
    a. Go into VirtualBox Settings for the VM


Hide
Name
Date Modified
Date Created
Size
Oracle VM VirtualBox Manager
Tools
New
Settings
Discard
Show
v1cloudera-quickstart-vm-5.13.0-0-virtual...
Powered Off
kudu-demo
Powered Off
Confluent_Developer_VM_B5_608
Powered Off
cloudera-quickstart-vm-5.13.0-0-virtualbox
Powered Off
cloudera-quickstart-vm-5.13.0-0-vir...
Running
Preview

eneral

ame: cloudera-quickstart-vm-5.13.0-0-virtualbox-original
Operating System: Red Hat (64-bit)
Settings File Location: /Users/chaddvoracek/VirtualBox VMs/cloudera-quickstart-vm-5.13.0-0-virtualbox-original

System

Base Memory: 8192 MB
Processors: 2
Boot Order: Hard Disk, Optical
Acceleration: VT-x/AMD-V, Nested Paging, PAE/NX, KVM Paravirtualization

Display

Video Memory: 8 MB
Graphics Controller: VBoxVGA
Remote Desktop Server: Disabled
Recording: Disabled

Storage

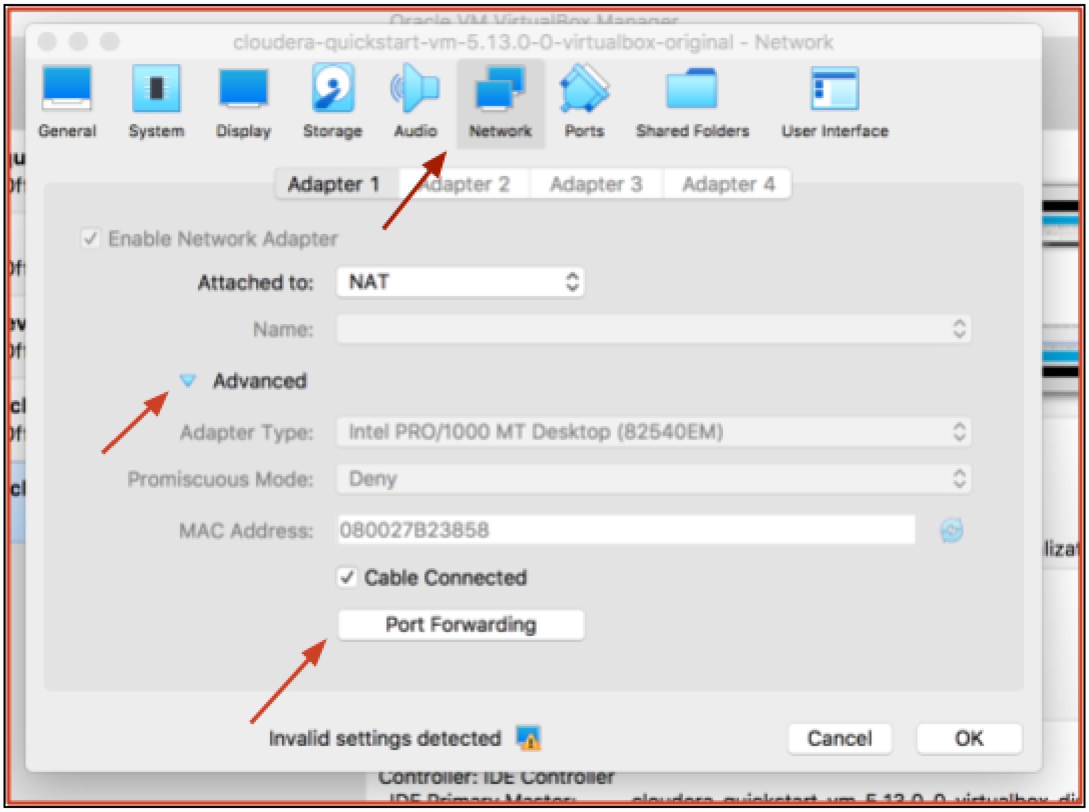
Controller: IDE Controller
IDE Primary Master: cloudera-quickstart-vm-5.13.0-0-virtualbox-disk1.vdi (Normal, 64.00 GB)
IDE Secondary Master: [Optical Drive] Empty

Audio

Disabled

Network

    b. Select Network Icon -> Advanced -> Port Forwarding



cloudera-quickstart-vm-5.13.0-0-virtualbox-original - Network

☑ Enable Network Adapter

Attached to: NAT

Name:

Advanced

Adapter Type: Intel PRO/1000 MT Desktop (82540EM)

Promiscuous Mode: Deny

MAC Address: 080027B23858

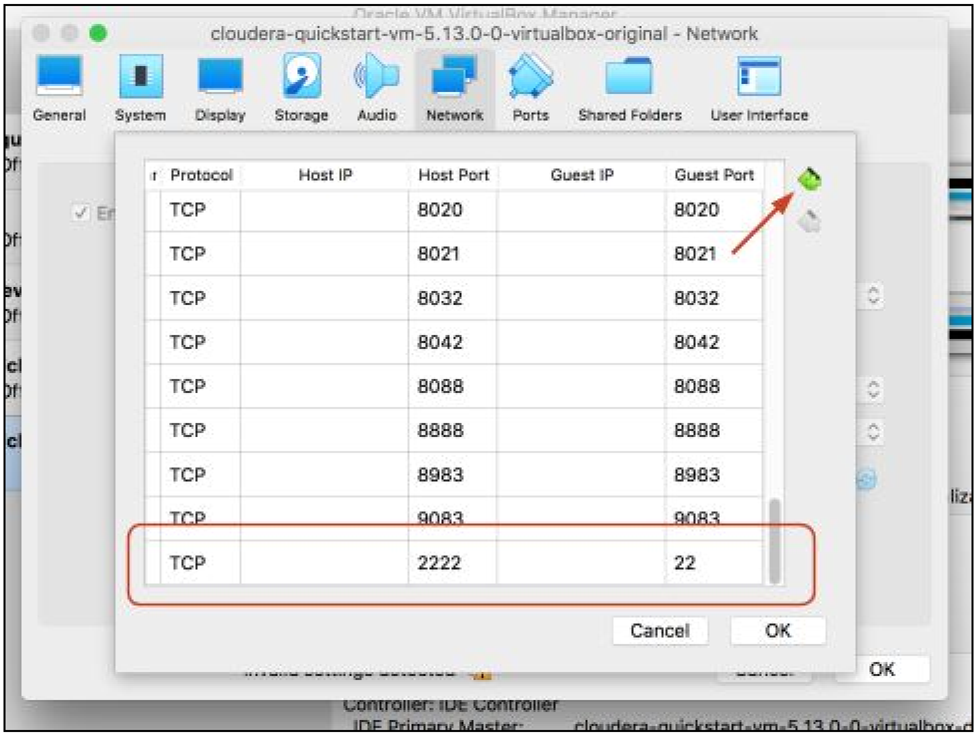
☑ Cable Connected

Invalid settings detected

Controller: IDE Controller

IDE Primary Master: cloudera-quickstart-vm-5.13.0-0-virtualbox-di

c. Add a new TCP Protocol with Host Port: 2222 and Guest Port: 22



cloudera-quickstart-vm-5.13.0-0-virtualbox-original - Network

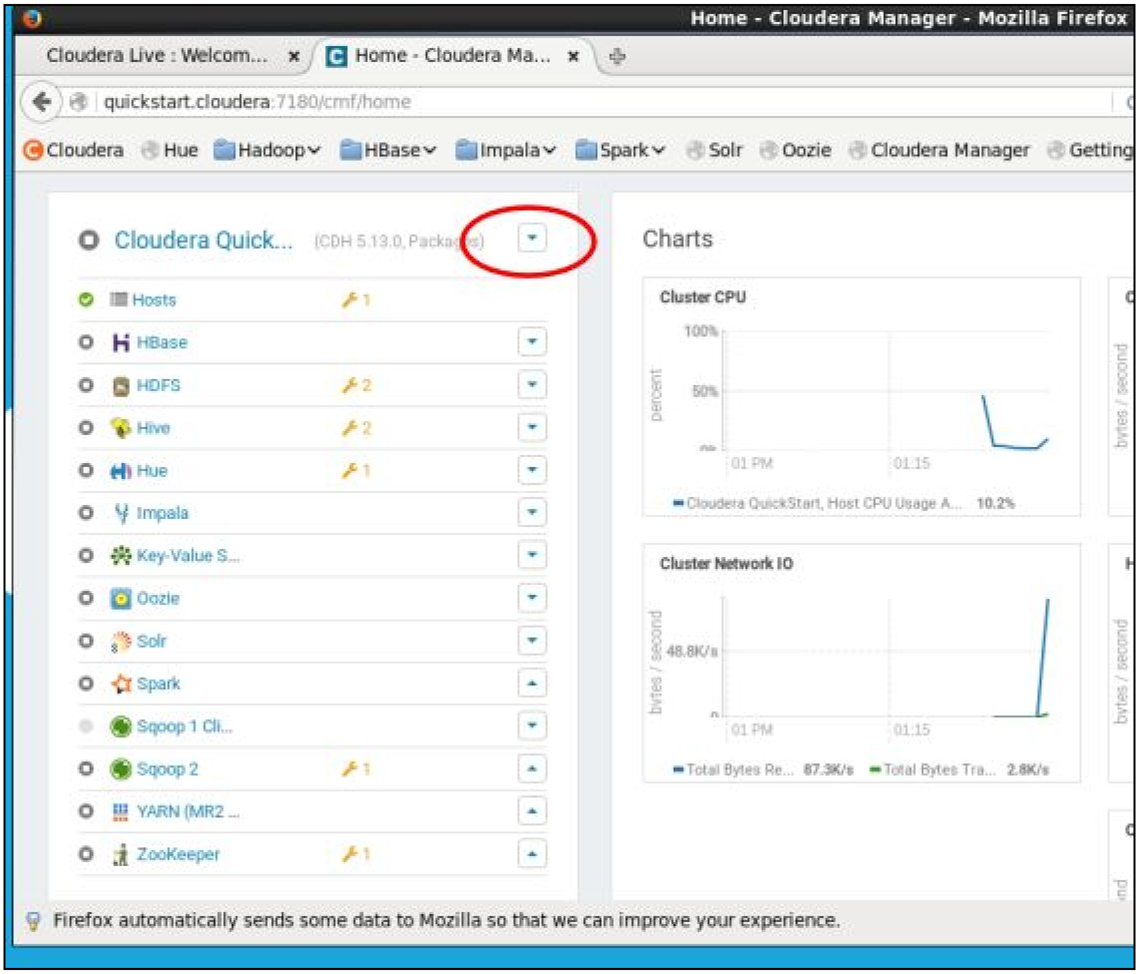
Protocol	Host IP	Host Port	Guest IP	Guest Port
TCP		8020		8020
TCP		8021		8021
TCP		8032		8032
TCP		8042		8042
TCP		8088		8088
TCP		8888		8888
TCP		8983		8983
TCP		9083		9083
TCP		2222		22

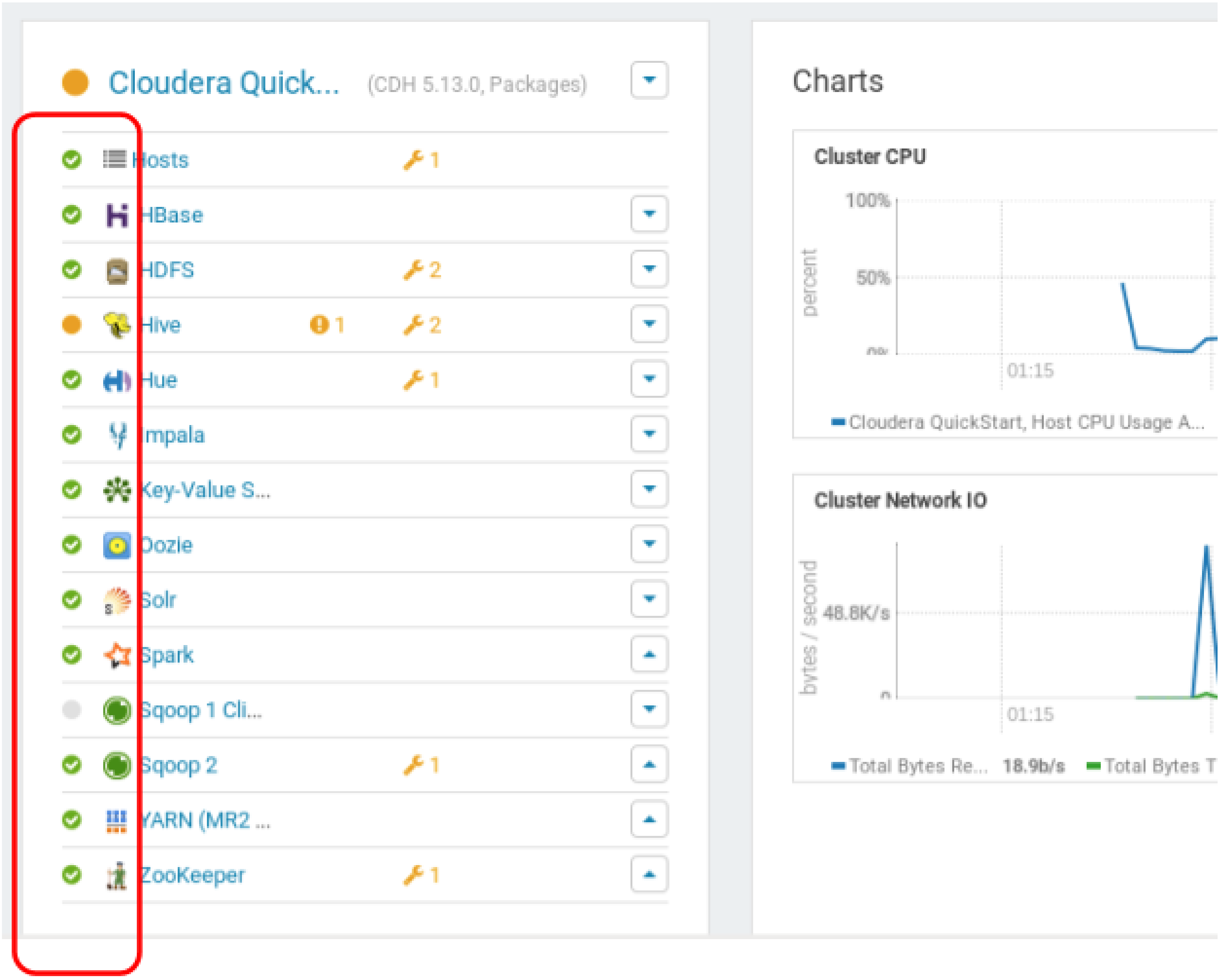
Controller: IDE Controller

IDE Primary Master: cloudera-quickstart-vm-5.13.0-0-virtualbox-d

d. Now you should be able to use the terminal on your computer to ssh into the VM.
   i. ssh cloudera@127.0.0.1 -p2222
   ii. Enter the VM password when prompted.

8. When shutting down VM and then coming back in later, you may need to go into Cloudera Manager to restart services or the Cloudera Manager Service. Click on the Cloudera Quickstart arrow then select “Restart”. Once all service have restarted you will see the green icons showing the service is running.





9. If you are having issues accessing git repositories directly from the VM and are getting a request failed to clone the repository. Run this command from the cloudera VM to update a few packages. *sudo yum update -y nss curl libcurl*

Optional For Extra Credit: Using Kudu

Note: it is recommended you complete the required parts of the Hadoop project before attempting to install Kudu and get it working.

1. Install Kudu on Cloudera VM
    a. See this link and follow the instructions https://blog.clairvoyantsoft.com/installing-apache-kudu-on-clouderas-quickstart-vm-3bdc202ce142

    b. Try smoke test code in the terminal shell
    c. It should fail with a warning

    ```
    ERROR: ImpalaRuntimeException: Error creating Kudu table
    'impala::default.kudu_test'
    CAUSED BY: NonRecoverableException: Not enough live tablet
    servers to create a table with the requested replication
    factor 3. 1 tablet servers are alive.
    ```

    d. Go to Cloudera Manager main page. Click on the Arrow to the right of Kudu. Select “Configuration”
    e. Search for “replicas”
    f. Change Default Number of Replicas to “1”
    g. Click on “suppress” for warning and enter any comment.
    h. Save Changes
    i. Restart Kudu
    j. Re-Run smoke test code from blog. You should be able to run all the tests now.
    k. Go back to Cloudera Management Service (Configuration)
        i. Find the Status - Warning (click to quickly find the following settings)
        ii. Increase Java Heap Size of Host Monitor in Bytes to 1.0 GiB
        iii. Increase Maximum Non-Java Memory of Host Monitor 2.0 GiB
        iv. Increase Java Heap Size of Service Monitor in Bytes to 2.0 GiB
        v. Increase Maximum Nan-Java Memory of Service Monitor to 5.0 GiB
        vi. Save Changes
    l. Install NTP (See https://docs.cloudera.com/runtime/7.0.3/troubleshooting-kudu/topics/kudu-installing-ntp.html)
        i. From Cloudera terminal: ‘sudo yum install ntp’
        ii. Restart service from terminal: ‘sudo /etc/init.d/ntpd restart’
        iii. NOTE: This is going to cause clock issues on the VM as NTP is not intended to be run on a VM. You may have to figure out a work around when you pause the VM as the clock will get out of sync.
    m. Restart Cloudera Management Service - Wait
    n. If everything doesn’t come back up healthy
        i. Cloudera Quickstart - Restart
    o. Drop Kudu table and Go through Smoke test code again from blog.

2. Use Kudu VM - less problem but you will have to ingest data in to the VM and no connection to Impala
    a. https://kudu.apache.org/releases/1.8.0/docs/quickstart.html#quickstart_vm
3. Install Kudu with Docker
    a. https://kudu.apache.org/docs/quickstart.html