



BIG DATA

By: Michel Samir Zaki



Installation

Download and Install VirtualBox

Link: <https://www.virtualbox.org/wiki/Downloads>

For windows choose: Windows hosts



Installation

Download Cloudera VM for VirtualBox

Link:

https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.8.0-0-virtualbox.zip

Then Extract it (**folder created after Extraction**)



Installation

Cloudera VM for VirtualBox



Favorites



Desktop



Downloads



Recent places



Homegroup



This PC

Name

Date modified

Type

Size



cloudera-quickstart-vm-5.8.0-0-virtualbox

10/16/2023 11:01 ...

File folder



cloudera-quickstart-vm-5.8.0-0-virtualbox.zip

10/16/2023 1:25 AM

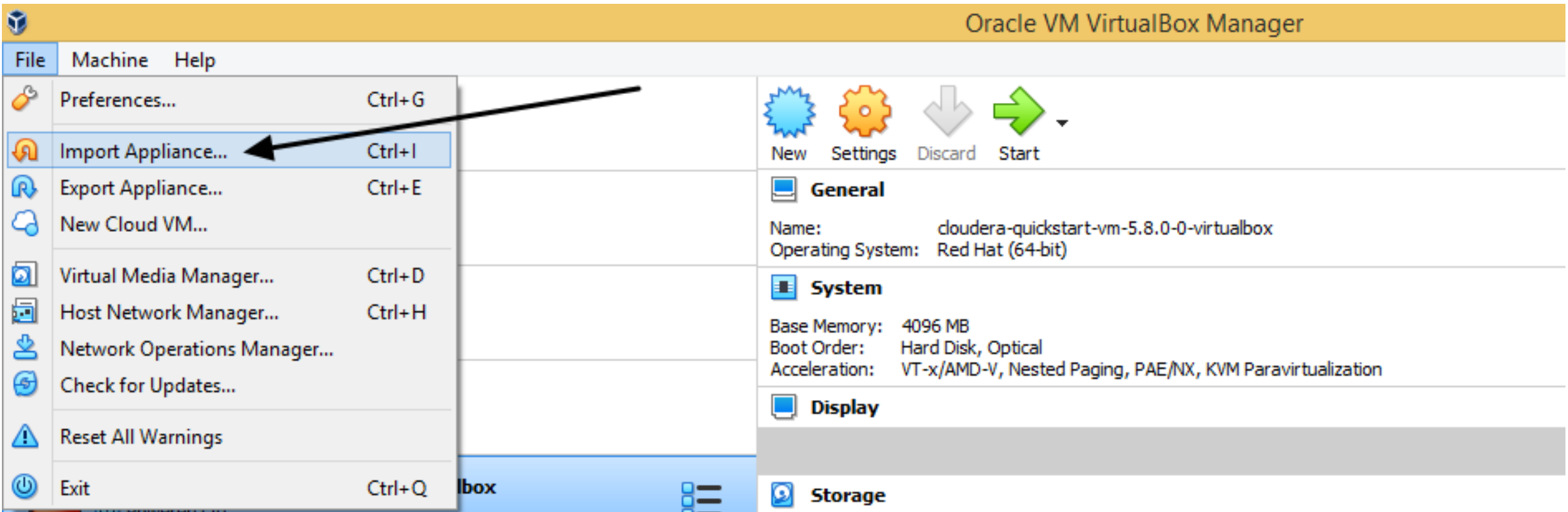
WinRAR ZIP archive

5,148,003 KB



Installation

Import Cloudera VM in VirtualBox





Installation

Browse Cloudera VM (folder created after Extraction)


← Import Virtual Appliance


Appliance to import

Please choose the source to import appliance from. This can be a local file system to import OVF archive or one of known cloud service providers to import cloud VM from.

Source:

Please choose a file to import the virtual appliance from. VirtualBox currently supports importing appliances saved in the Open Virtualization Format (OVF). To continue, select the file to import below.


File: 





Installation

Browse Cloudera VM (folder created after Extraction)


 Import Virtual Appliance

Appliance to import

Please choose the source to import appliance from. This can be a local file system to import OVF archive or one of known cloud service providers to import cloud VM from.

Source:

Please choose a file to import the virtual appliance from. VirtualBox currently supports importing appliances saved in the Open Virtualization Format (OVF). To continue, select the file to import below.

File: 



Installation

Import Cloudera VM

Import Virtual Appliance

Appliance settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Virtual System 1	
Name	cloudera-quickstart-vm-5.8.0-0-virtualbox 1
Guest OS Type	Red Hat (64-bit)
CPU	1
RAM	4096 MB
DVD	<input checked="" type="checkbox"/>
Network Adapter	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
Storage Controller (IDE)	PIIX4
Storage Controller (IDE)	PIIX4
Virtual Disk Image	cloudera-quickstart-vm-5.8.0-0-virtualbox-disk1.vmdk
Base Folder	C:\Users\Michel Samir\VirtualBox VMs
Primary Group	/

Machine Base Folder: C:\Users\Michel Samir\VirtualBox VMs

MAC Address Policy: Include only NAT network adapter MAC addresses

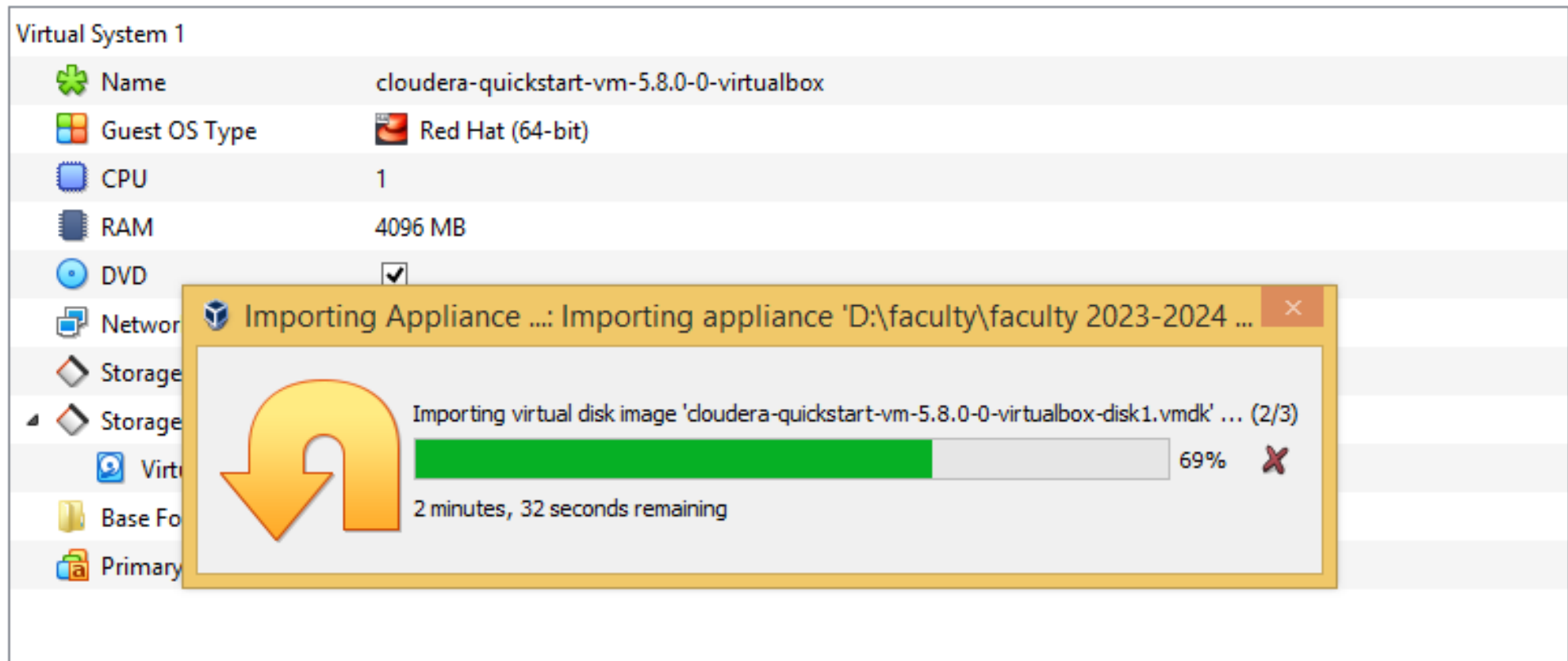
Additional Options: ☒ Import hard drives as VDI

Appliance is not signed



Installation

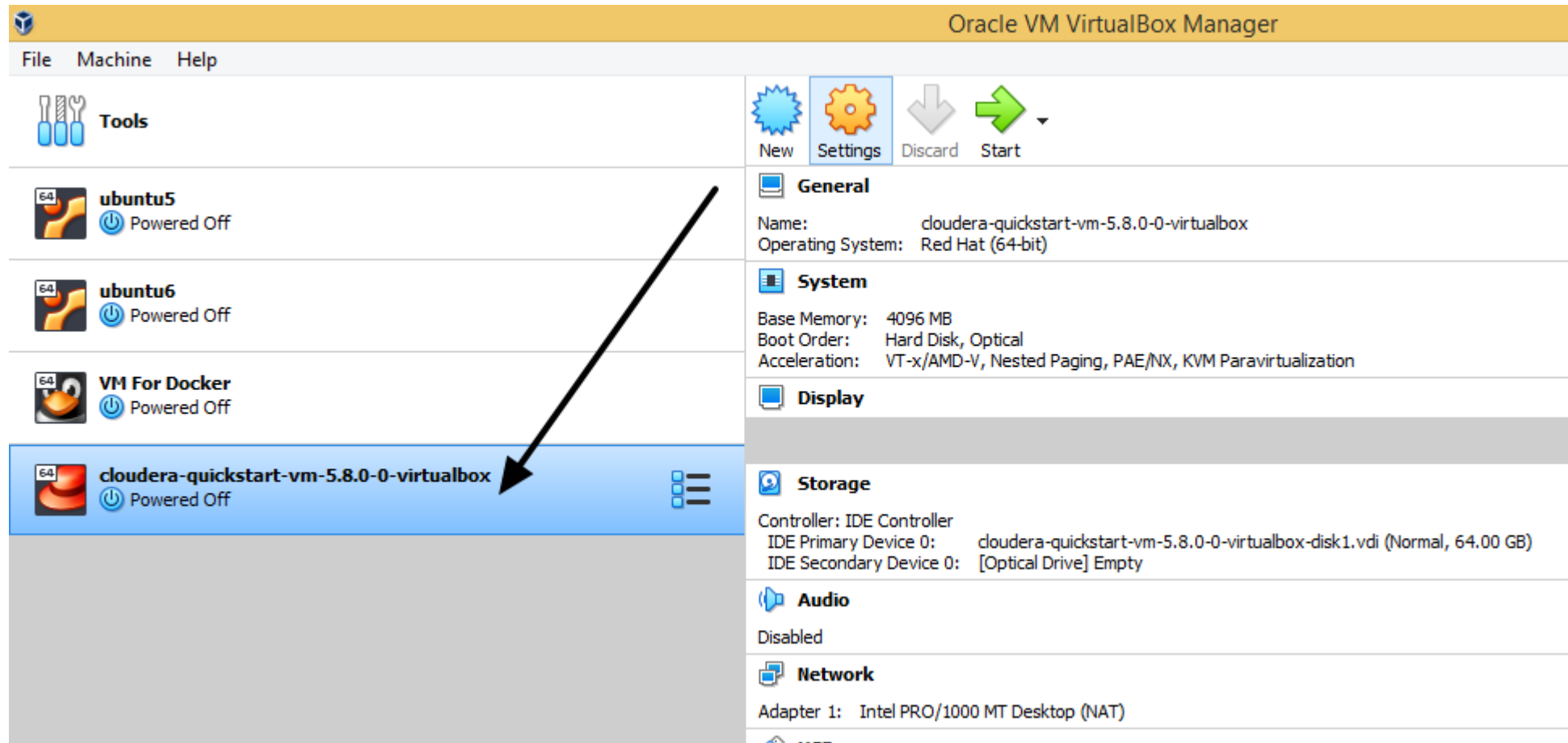
Import Cloudera VM





Installation

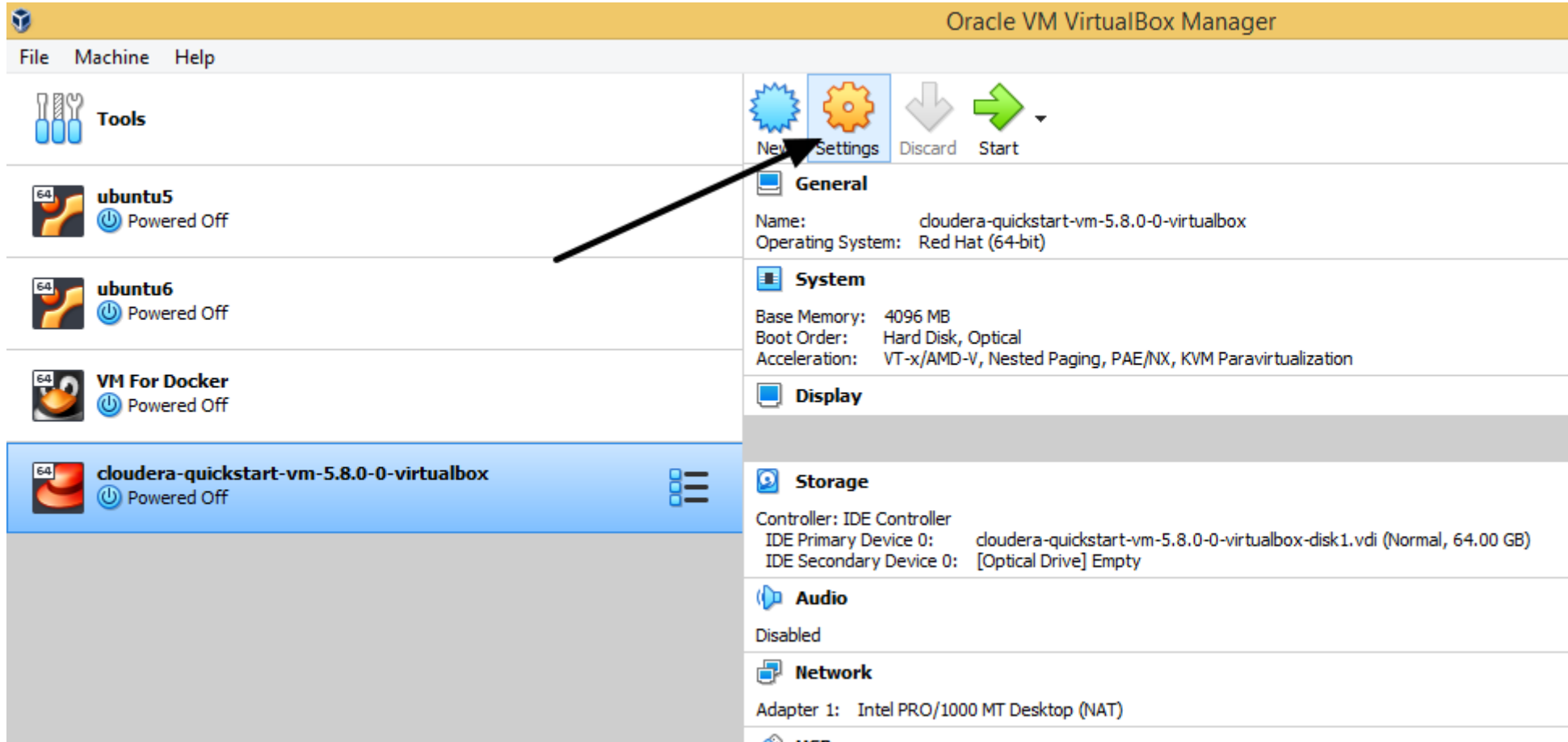
After Importing Cloudera VM





Installation

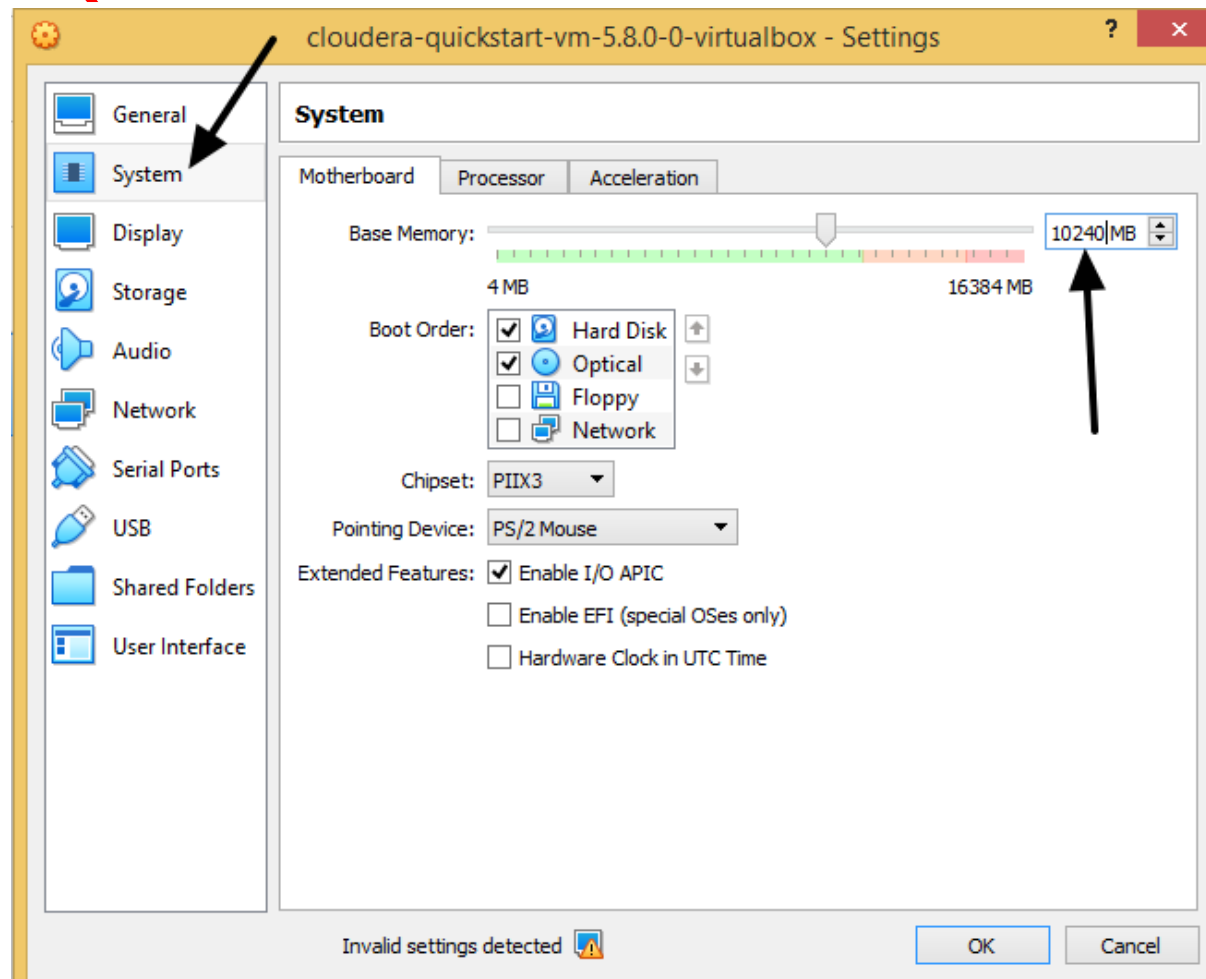
Change some settings





Installation

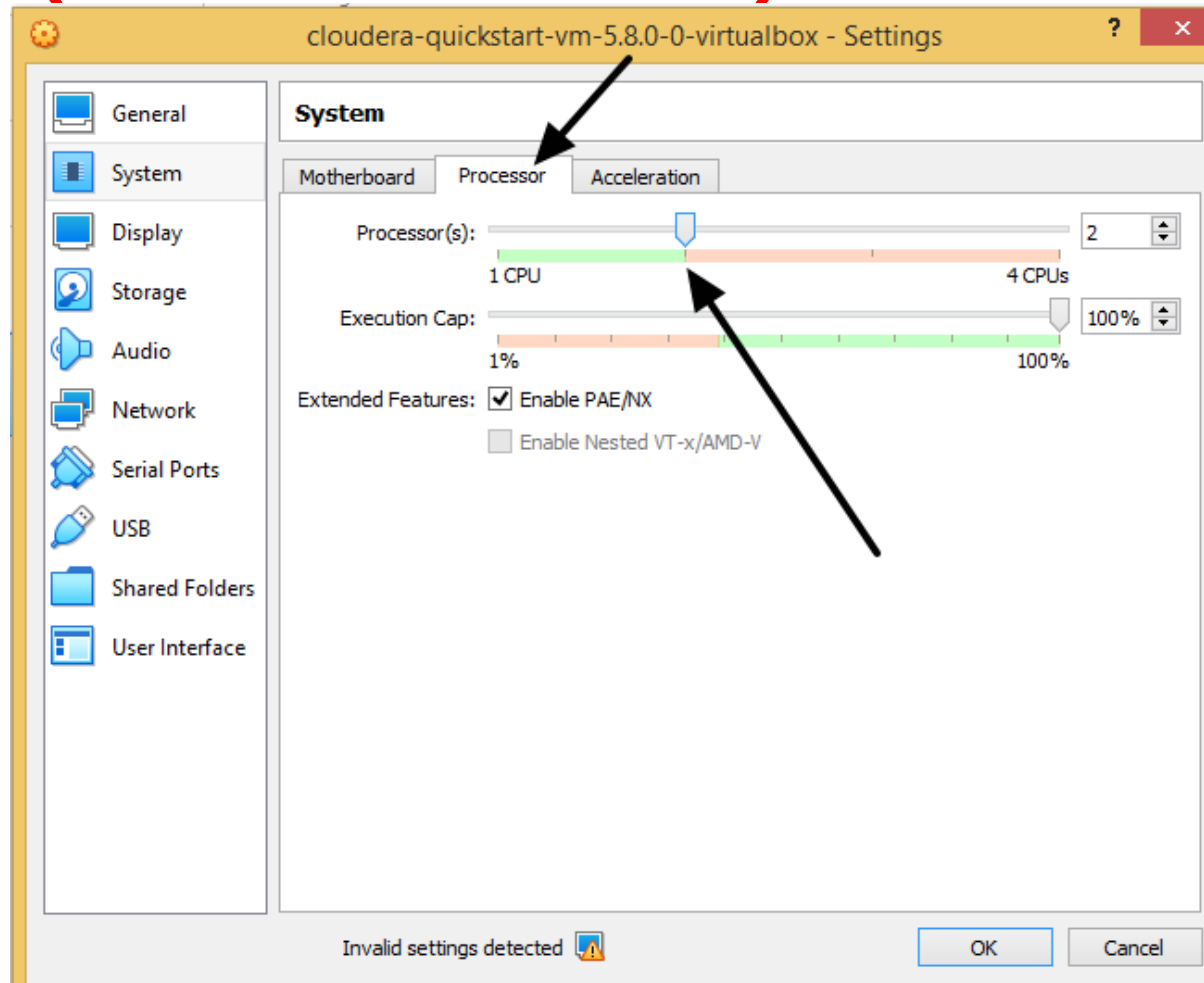
Increase RAM (Choose 5552 if your RAM is 8GB)





Installation

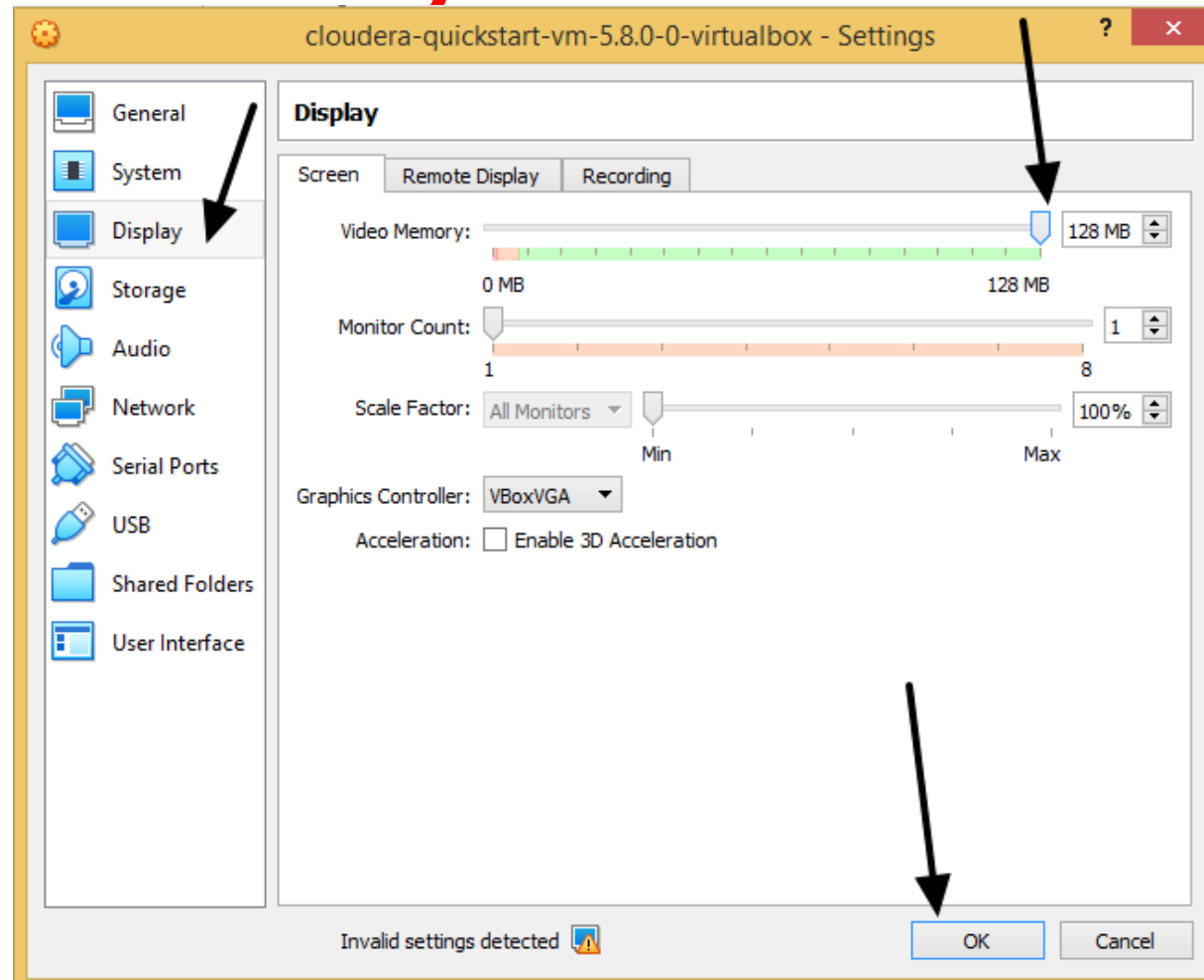
Increase CPU (2 Cores or more)





Installation

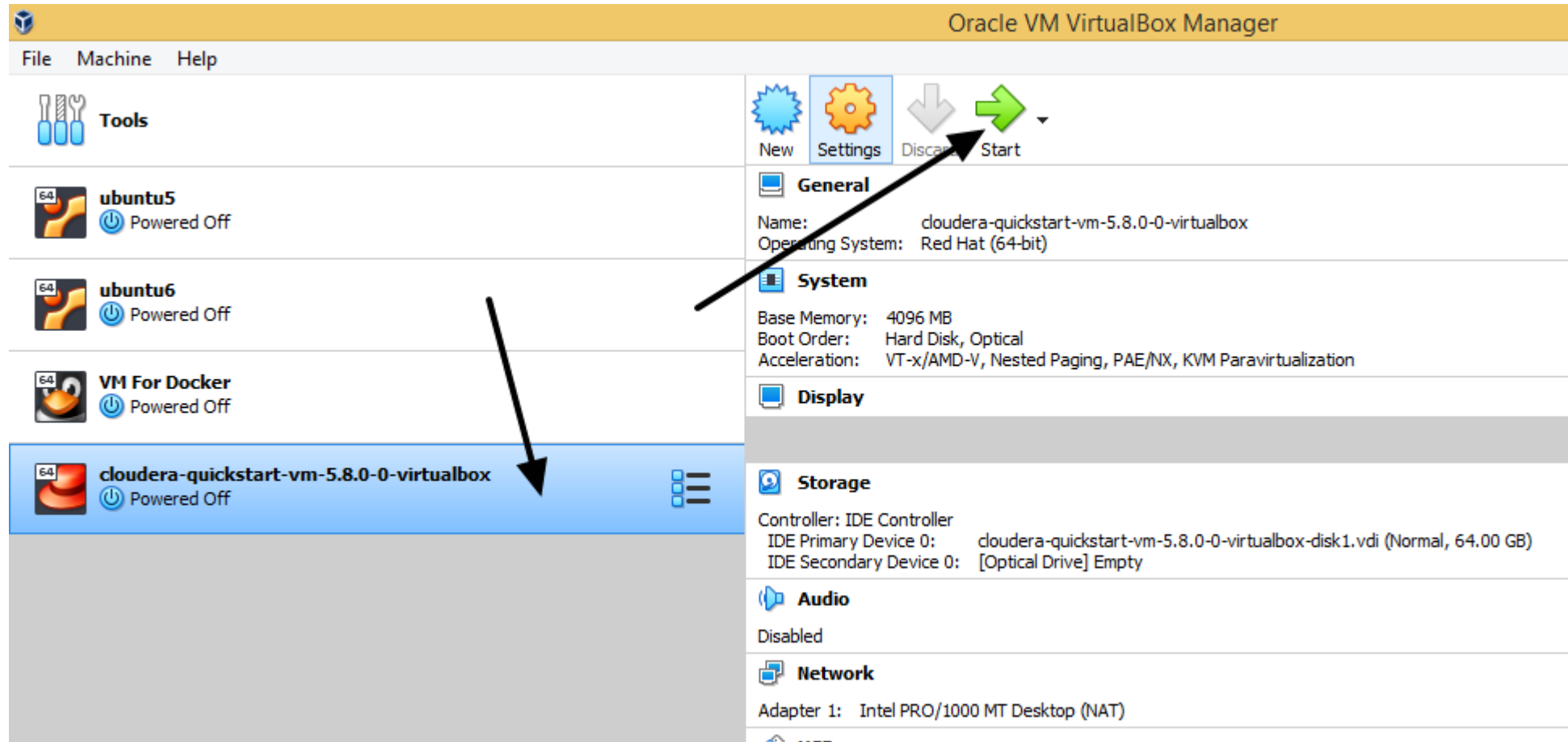
Increase Video Memory -> OK





Installation

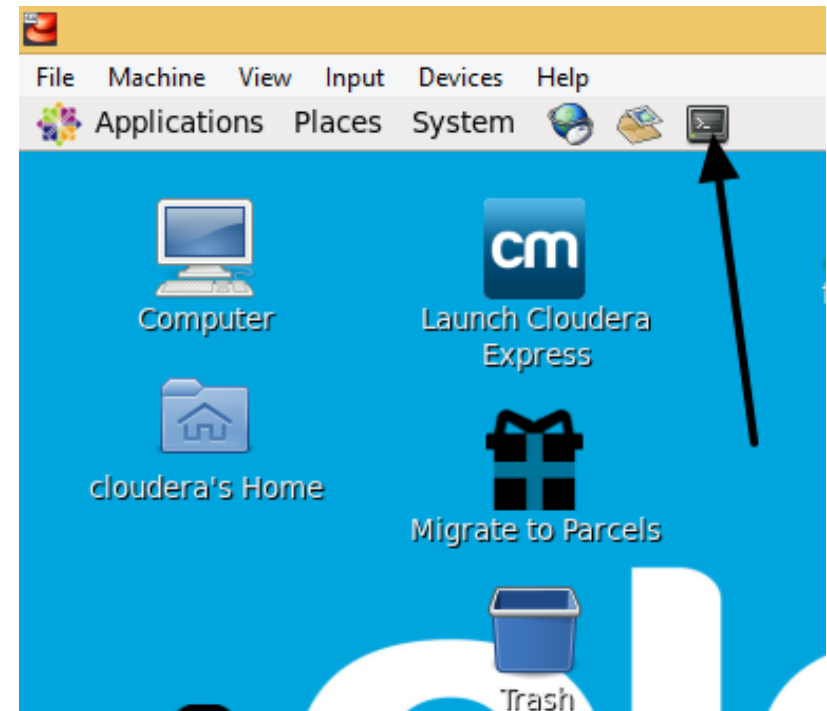
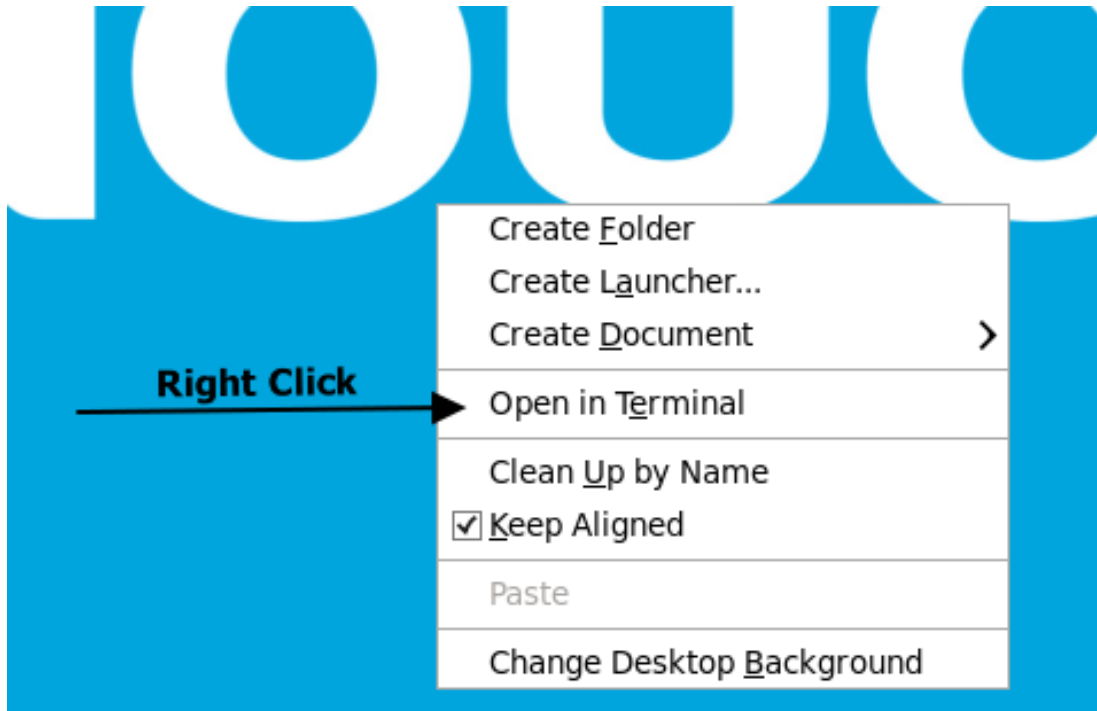
Run (Start) Cloudera VM





HDFS Commands

Open terminal (2 options)





HDFS Commands

Check Hadoop Services

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ sudo service hadoop-hdfs-namenode status  
Hadoop namenode is running [ OK ]  
[cloudera@quickstart ~]$  
[cloudera@quickstart ~]$ sudo service hadoop-hdfs-secondarynamenode status  
Hadoop secondarynamenode is running [ OK ]  
[cloudera@quickstart ~]$  
[cloudera@quickstart ~]$ sudo service hadoop-hdfs-datanode status  
Hadoop datanode is running [ OK ]  
[cloudera@quickstart ~]$
```



HDFS Commands

ls (List Hadoop File System)

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -ls /  
Found 18 items  
drwxrwxrwx - hdfs supergroup 0 2016-08-10 13:07 /benchmarks  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 04:23 /data  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 06:16 /data2  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 22:38 /data3  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 00:58 /data4  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 02:20 /data5  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 04:20 /data6  
drwxr-xr-x - hbase supergroup 0 2023-10-28 16:53 /hbase  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 05:01 /output  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 06:59 /output2  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 23:19 /output3  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 01:03 /output4  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 02:50 /output5
```



HDFS Commands

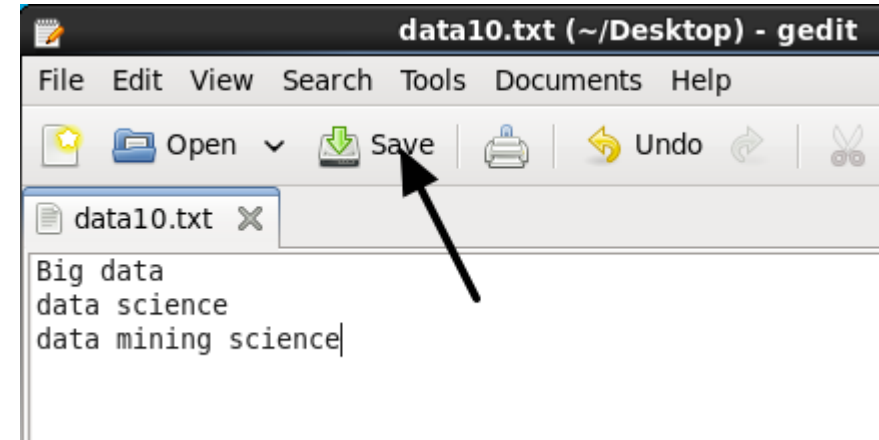
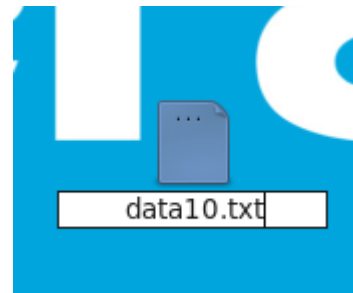
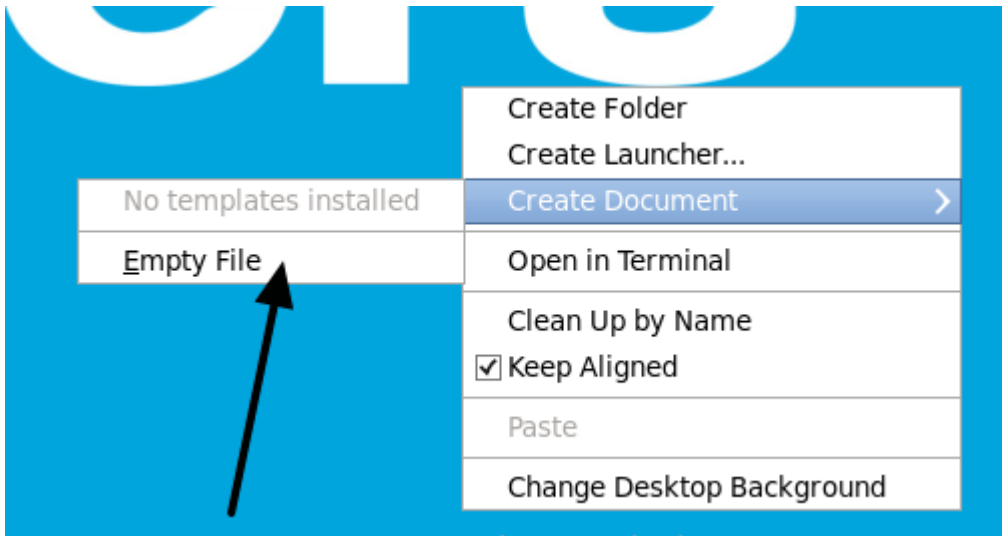
mkdir (create directory)

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -mkdir /data10  
[cloudera@quickstart ~]$  
[cloudera@quickstart ~]$ hdfs dfs -ls /  
Found 19 items  
drwxrwxrwx - hdfs supergroup 0 2016-08-10 13:07 /benchmarks  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 04:23 /data  
drwxr-xr-x - cloudera supergroup 0 2023-10-28 17:32 /data10  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 06:16 /data2  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 22:38 /data3  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 00:58 /data4  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 02:20 /data5  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 04:20 /data6  
drwxr-xr-x - hbase supergroup 0 2023-10-28 16:53 /hbase  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 05:01 /output  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 06:59 /output2  
drwxr-xr-x - cloudera supergroup 0 2023-10-24 23:19 /output3  
drwxr-xr-x - cloudera supergroup 0 2023-10-25 01:03 /output4
```



HDFS Commands

put (add file to hadoop file system)





HDFS Commands

put (add file to hadoop file system) (Cont.)

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -put ~/Desktop/data10.txt /data10/  
[cloudera@quickstart ~]$  
[cloudera@quickstart ~]$ hdfs dfs -ls /data10/  
Found 1 items  
-rw-r--r--  1 cloudera supergroup      42 2023-10-28 17:40 /data10/data10.tx  
t  
[cloudera@quickstart ~]$
```



HDFS Commands

cat (read file content in terminal)

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hdfs dfs -cat /data10/data10.txt  
Big data  
data science  
data mining science  
[cloudera@quickstart ~]$
```

The image features a blue-tinted background showing silhouettes of several groups of business professionals in a modern office environment. They are standing on a reflective floor, and a city skyline is visible through large windows in the background. The text "Thank You" is centered in the middle of the image.

Thank You