

CSCE 5300

In Class Exercise 1

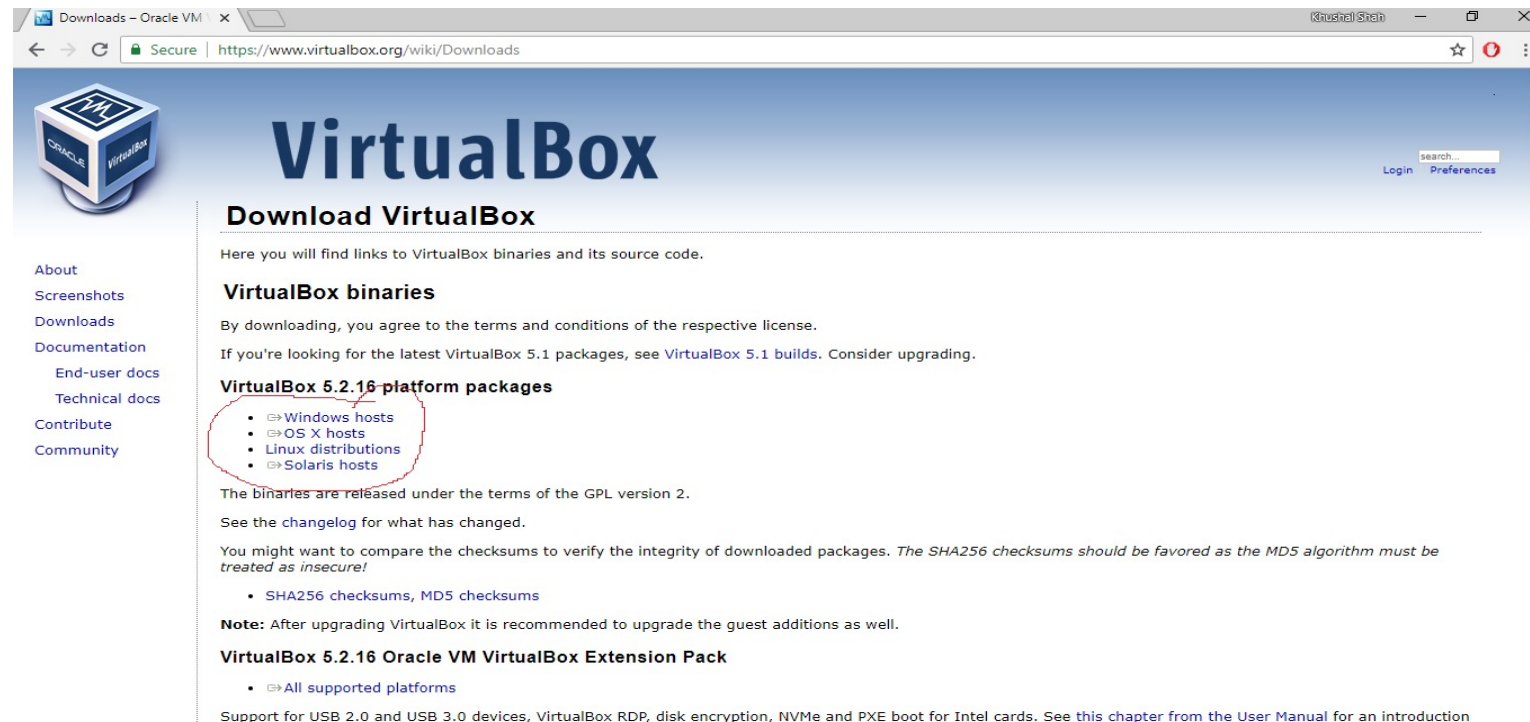
Cloudera Installation

1. Download cloudera operating system given in canvas under the first module.

Download Oracle Virtual Box

Use the following link to download oracle virtual box

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



The screenshot shows a web browser window with the address bar displaying <https://www.virtualbox.org/wiki/Downloads>. The page features the VirtualBox logo on the left and a navigation menu with links to About, Screenshots, Downloads, Documentation, End-user docs, Technical docs, Contribute, and Community. The main content area is titled "Download VirtualBox" and includes a search bar, Login, and Preferences links. The text explains that the page provides links to VirtualBox binaries and source code, and that downloading implies agreement to the license. It directs users to the latest VirtualBox 5.1 packages and mentions the option to upgrade. A section titled "VirtualBox 5.2.16 platform packages" lists four options: Windows hosts, OS X hosts, Linux distributions, and Solaris hosts. The "Linux distributions" link is circled in red. Below this, it states that binaries are released under the GPL version 2 and provides a link to the changelog. A note advises comparing checksums to verify integrity, favoring SHA256 over MD5. Another note recommends upgrading guest additions after upgrading VirtualBox. The page also mentions the "VirtualBox 5.2.16 Oracle VM VirtualBox Extension Pack" and lists a link for "All supported platforms". At the bottom, it notes support for USB 2.0 and USB 3.0 devices, VirtualBox RDP, disk encryption, NVMe, and PXE boot for Intel cards, with a link to the User Manual for more information.

Downloads - Oracle VM x

Secure | <https://www.virtualbox.org/wiki/Downloads>

search...
Login Preferences

VirtualBox

Download VirtualBox

Here you will find links to VirtualBox binaries and its source code.

VirtualBox binaries

By downloading, you agree to the terms and conditions of the respective license.

If you're looking for the latest VirtualBox 5.1 packages, see [VirtualBox 5.1 builds](#). Consider upgrading.

VirtualBox 5.2.16 platform packages

- Windows hosts
- OS X hosts
- Linux distributions
- Solaris hosts

The binaries are released under the terms of the GPL version 2.

See the [changelog](#) for what has changed.

You might want to compare the checksums to verify the integrity of downloaded packages. *The SHA256 checksums should be favored as the MD5 algorithm must be treated as insecure!*

- SHA256 checksums, MD5 checksums

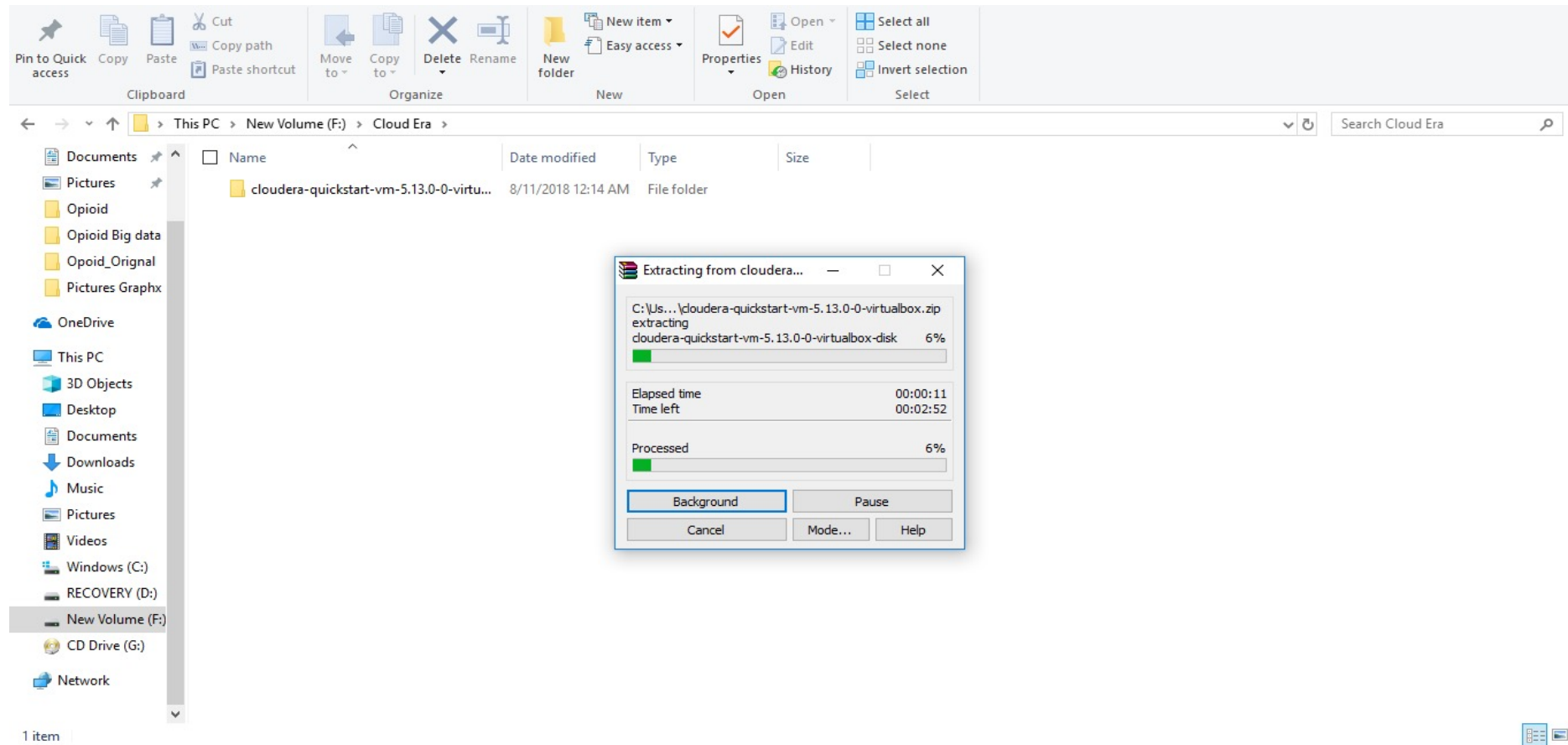
Note: After upgrading VirtualBox it is recommended to upgrade the guest additions as well.

VirtualBox 5.2.16 Oracle VM VirtualBox Extension Pack

- All supported platforms

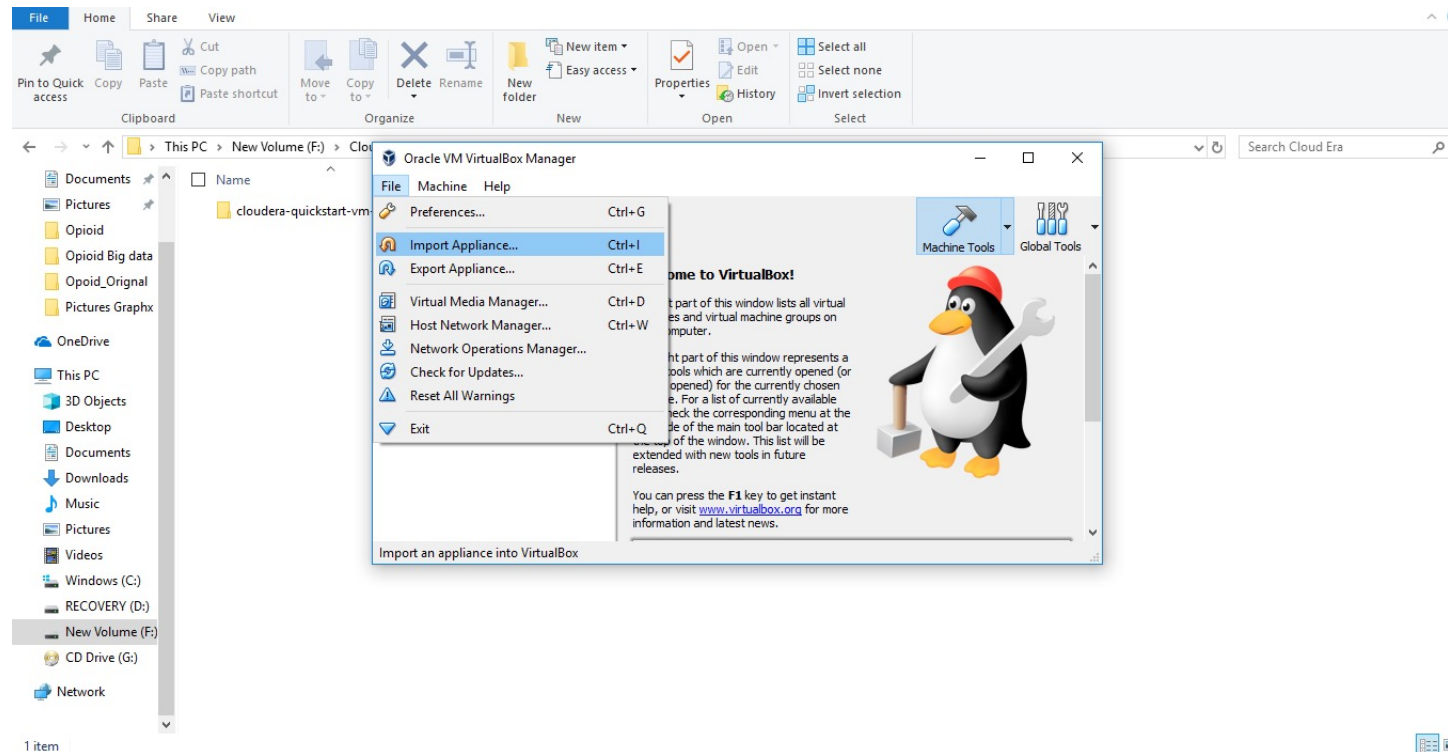
Support for USB 2.0 and USB 3.0 devices, VirtualBox RDP, disk encryption, NVMe and PXE boot for Intel cards. See [this chapter from the User Manual](#) for an introduction

Extract cloudera files to a directory



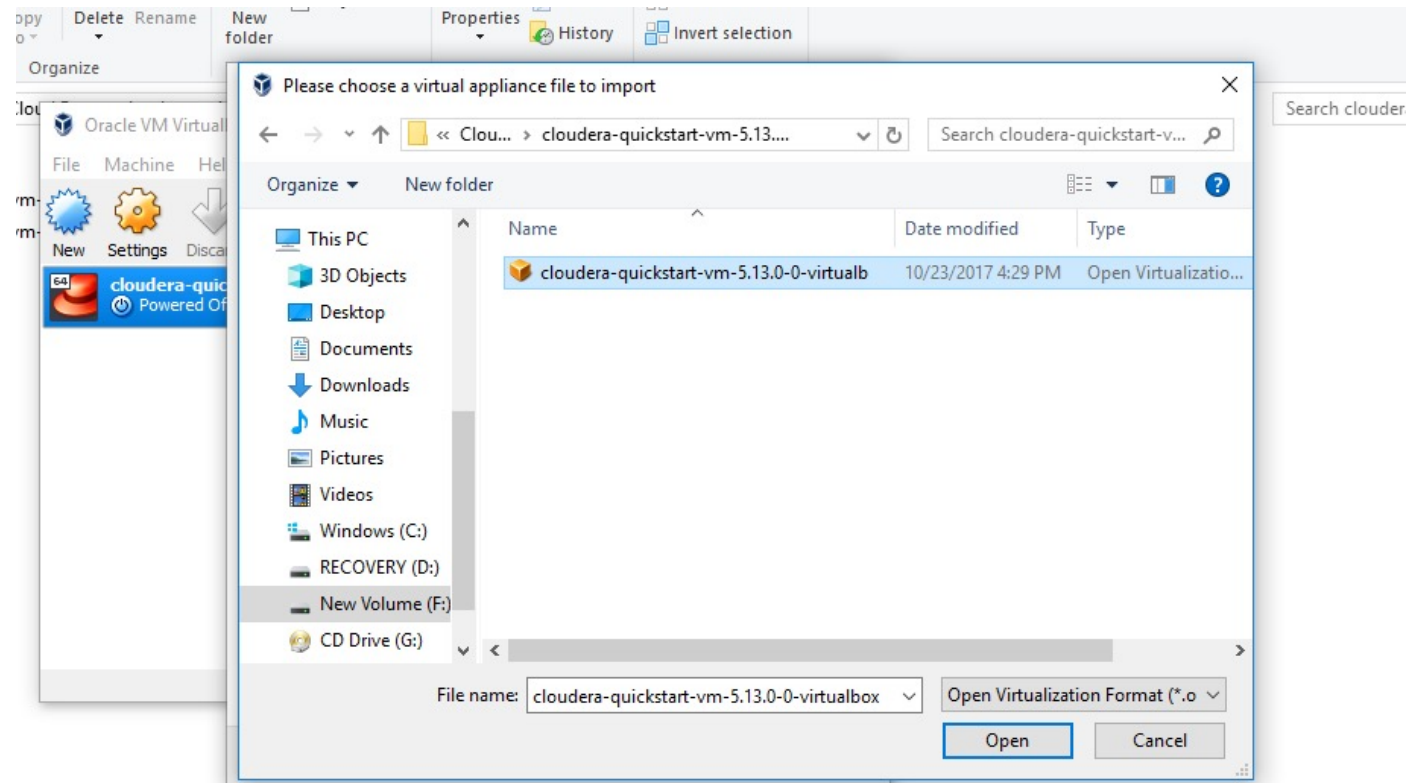
Installation

1. Install Oracle Virtual box
2. Create a Virtual Machines.
3. Go to file and select import Appliance to select the VM image from your hard drive.



Continued ..

- Select the virtual machine to start loading the VM image in the virtual box. (cloudera-quickstart-vm-5.13.0-0-virtualbox)
- It will take several minutes to start the Cloud era operating system



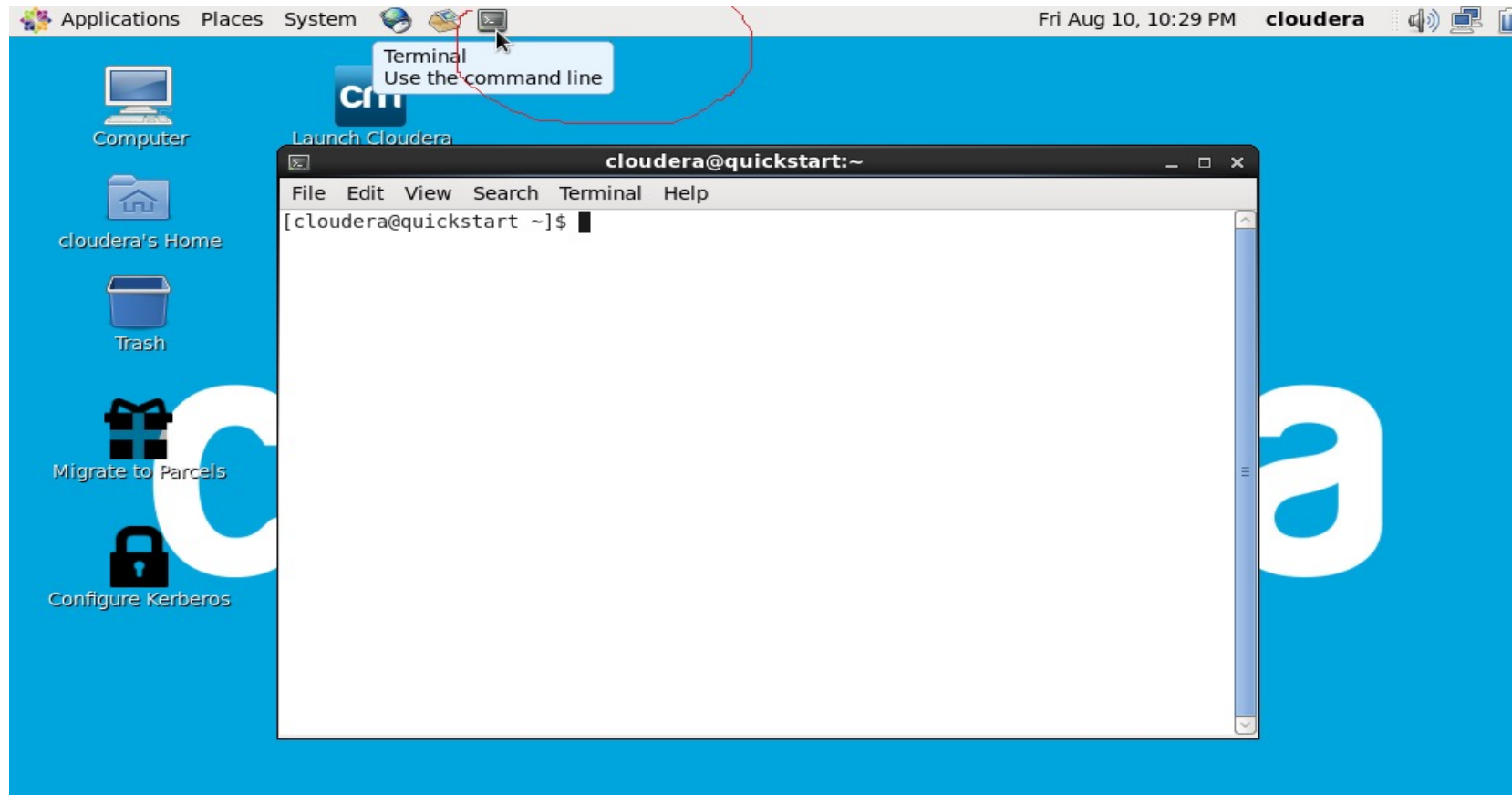
Enter Cloudera (Unix System)

- Cloud era is a Unix system and is ready for Big Data exercises.



Start by opening Terminal

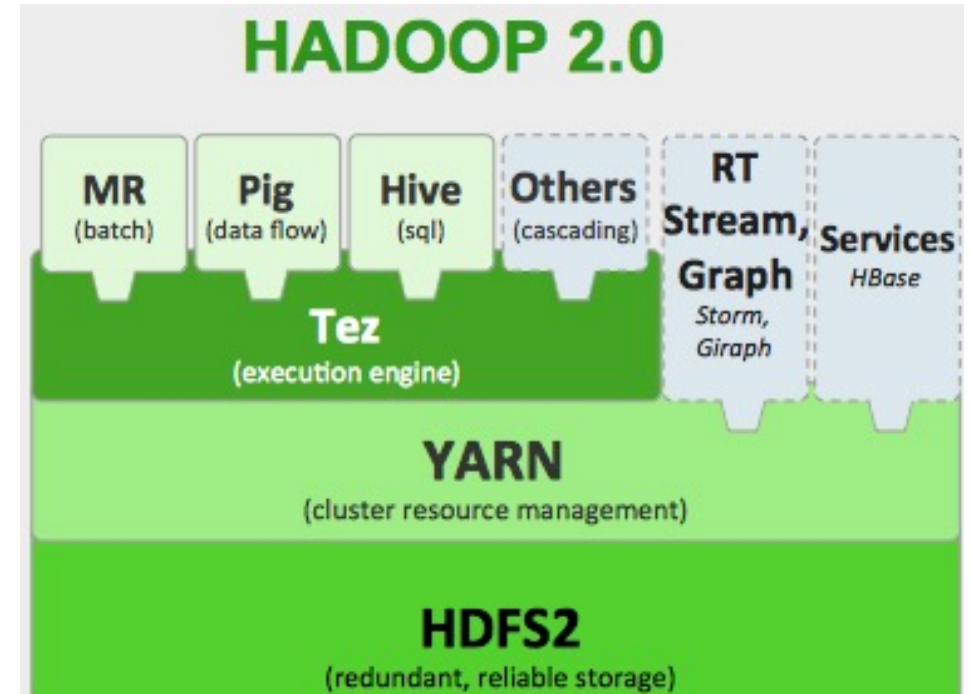
- Start terminal from the upper right corner as shown in the picture.



Running Hadoop (HDFS)

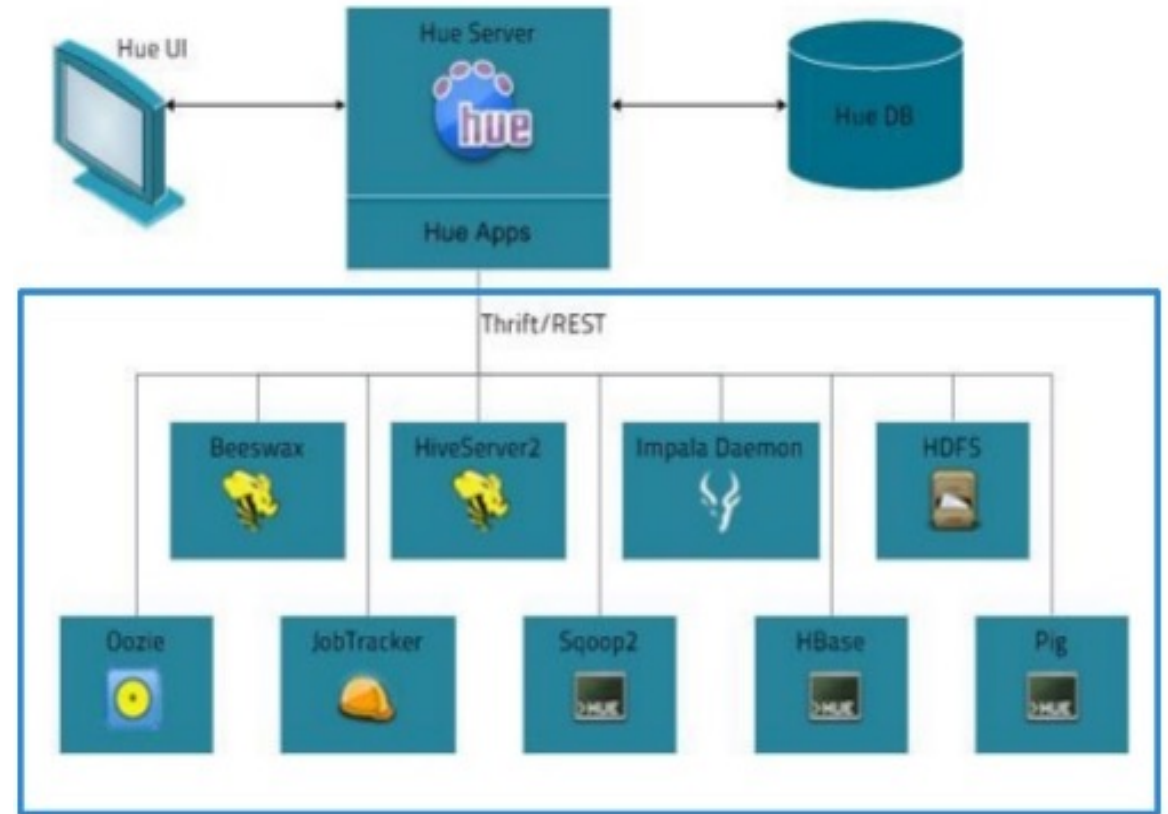


- Create a file named as localfile
- Open terminal and type the following command “hdfs dfs -appendToFile localfile /user/hadoop/hadoopfile”
- Check if the command was successful.
- If the command was successful, it means the hadoop server is running successfully on cloudera or any other operating system



Hue

- **Hue** (Hadoop User Experience) is an open-source Web interface that supports Apache Hadoop and its ecosystem, licensed under the Apache v2 license
- Hue is an open-source Analytics Workbench for browsing, querying and visualizing data.
- All hadoop task can be queried and displayed in Hue interface.



In Class Exercise

You are required to follow the steps below to complete your ICE today

- Use the datasets given in canvas
- Load it in hadoop hdfs
- Use the second file
- Append it to the first file
- Visualize file with Hue
- View the first and last lines (approximately 5) of merged dataset using appropriate hdfs commands
- Create a new text file and load it into hdfs and try to append all three datasets.

Basic Commands

List files in Hadoop

```
hadoop fs -ls /user/cloudera
```

Make new Directory:

```
hadoop fs -mkdir /user/cloudera/Zeenat
```

Move file to Zeenat

```
hadoop fs -put /home/cloudera/Desktop/word_list.txt  
/user/cloudera/zeenat
```