

# CISC 886

# Cloud and Big Data

## Lab 1

# Cloudera VM

- The link of the Cloudera VM that you can install in your machine:  
<https://sourceforge.net/projects/getprathamos/files/>
- Video explanation of how to complete the process:  
[https://bit.ly/CDH\\_6\\_3\\_2\\_CentOS7](https://bit.ly/CDH_6_3_2_CentOS7)

# Installation

- Download Peazip to join & extract the VM files  
<https://peazip.github.io/peazip-64bit.html>
- Download VMware Workstation to run the virtual machine  
<http://bit.ly/GetVMPlayer>
- Download WinSCP for file transferring between a local computer and a remote server <https://winscp.net/eng/download.php>

# Cloudera VM

- Login 'Base User' Password - BaseUser@123
- 'root' Password - BaseUser@123
- -----
- sudo user - osboxes
- sudo password - BaseUser@123
- -----
- MySql user - root
- MySql password - bigdata
- -----
- Cloudera Manager user - admin
- Cloudera Manager password - admin

# Linux Command

- Print the name of current directory `$pwd`
- Print the current user `$whoami`
- List all items in the current folder `$ls`
- Goto Dir. `$cd`
- Go back to the parent directory `$cd ..`
- Clears the terminal `$clear`

# Linux Command

- Makes folder with the specified name `$mkdir`
- Removes the specified folder `$rmdir`
- Copy file `$cp`
- Reads the specified file `$cat`
- Edit file in the terminal `$vi`
- Permissions (root) `$sudo su - / $sudo su root`

# HDFS Command

- LISTING ROOT DIRECTORY `$hdfs dfs -ls /`
- CREATE A DIRECTORY IN HDFS `$hdfs dfs -mkdir <folder name>`
- COPY FROM LFS TO HDFS
  - `$hdfs dfs -put lfspath hdfspath`
  - `$hdfs dfs -copyFromLocal <local file path> <dest(present on hdfs)>`
- COPY FROM HDFS TO LFS
  - `$hdfs dfs -get hdfspath lfspath`
  - `$hdfs dfs -copyToLocal <srcfile(on hdfs)> <local file dest>`

# HDFS Command

- MOVE A FILE FROM ONE FOLDER TO ANOTHER

```
$hdfs dfs -mv <src(on hdfs)> <src(on hdfs)>
```

- COPY A FILE FROM ONE FOLDER TO ANOTHER

```
$hdfs dfs -cp <src(on hdfs)> <dest(on hdfs)>
```

- DELETE DIR/FILES IN HDFS

```
$hdfs dfs -rm -r <filename/directoryName>
```

- Read the file from HDFS

```
$Hdfs dfs -cat hdfspath
```

- Creates an empty file

```
$hdfs dfs -touchz <file_path>
```

- HADOOP CONFIGURATION

```
$cd /etc/Hadoop/conf
```

- CHANGING PERMISSIONS

```
$hdfs dfs -chmod 777 <file_name>
```



# Lab 1

- 1- login to VM (assuming there is a Cloudera VM installation at your site and the user/password is as osboxes/BaseUser@123)
- 2- List all the files and directories under /user/osboxes in HDFS
- 3- Create a directory with name 'inputdata' in HDFS (command: `hdfs dfs -mkdir directoryname`)
- 4- List the folders in /user/osboxes within hdfs

5- Now, from your local machine, download historical stock data (csv format) of at least 5 stock quotes (random) from yahoo financial (finance.yahoo.com) and put them in a local folder of your machine.

6- In each of the files put the stock ticker (e.g. TSLA) in the first column of the file. Similarly do the same for others and finally merge them in one file named 'stock.csv'.

7- Install Winscp and connect to VM to transfer the stock.csv file from your local machine to the VM (in a new folder called 'data' under /home/osboxes).

8- Within VM, copy the file stock.csv from /home/osboxes/data (local file system) to 'inputdata' directory in HDFS that are created earlier.

9- Show the content of the last few records in the stock.csv file in HDFS.

10- Make a copy of the stock.csv file as stock1.csv from the current hdfs location to another folder named 'testdata' in hdfs.

11- Show the screenshot of all actions in a file name lab1-yourID-lastname. Also write your name and ID in header of the document you submit via OnQ under lab1.