

## LINUX AND INTRODUCTION TO BIG DATA HADOOP

## HOMEWORK 1

Q1. Open the file `/var/log/messages` in the vi editor and delete line number 150.

`$vi /var/log/messages` -it is used to open a message file.

`:set number` -it is used to set specific line of the line inside the file.

`:150 d` -to delete line 150

`:wq` -This command is used to exit and save from a file.

Q2. Write a shell script to add two numbers?

```
a=10
```

```
b=20
```

```
add=$(( $a + $b ))
```

```
echo "Sum is: $add"    #Sum is:30
```

Q3. User root wants to copy `/etc`, including all subdirectories and files to `/tmp`.

How will you achieve this task?

Using the below command can we achieve copy `/etc`, including all subdirectories and files to `/tmp`,

```
sudo cd /etc
```

```
cp -r /etc tmp
```

Q4.Create a file that contains only the username and the user id of all the user present on the server.

```
$cat /etc/passwd
```

By this unix command can we create a file that contains only the username and the user id of all the user present on the server.

Q5.How will you provide a count of all users on the system except for adm user?

```
$ who
```

Using the above command we can count a all users on the system except for adm user.

Q6.How will you list all files in /tmp in increasing order of their size?

> "\$ ls -laShr" - Using this command we can list the files in increasing order.

> To list all files and sort them by size we can use the -S.

> -l flag means long listing and -a tells ls to list all the files.

> human-readable format by adding the -h

>add the -r flag is to sort in reverse order

Q7.What command is used to clear history on the Linux server?

> To delete a particular content, we can use history -d <line number> .

> To clear the entire history, use history -c .

Q8.Explain Big Data and what are five V's of Big Data.

Big Data:

Big data is defined as the data sets that are too large or complex to be dealt with by traditional data-processing application software. It is a tool created to identify patterns on the information in order to design smart solutions

Five V's of Big Data:

> velocity

> volume

> value

> variety

> veracity

Q9. What is Hadoop and its components?

Hadoop:

Apache Hadoop is an open source framework that is used to efficiently store and process large data sets ranging at size from GB(gigabytes) to PB(petabytes).

Components of the Hadoop:

1) Hadoop HDFS

2)Hadoop MapReduce

3)Hadoop YARN