SURYA K

211720104151   BIG DATA

HOME WORK 1

Q 1: Open the file /var/log/messages in the vi editor and delete line number 150.

$vi /var/log/messages   used to open a message file.

:set number used to set specific line of the line inside the file. :150 d
representation of particular line to be delete

  :wq command used to exit and save from a file.

Q 2: Write a shell script to add two numbers?

#!/bin/bash

Calculate the sum of two integers with pre initialize values

in a shell script

a=40 b=40 sum=$(($a +
$b)) echo "Sum : $sum"

output : Sum : 80

Q 3:User root wants to copy /etc, including all subdirectories and files to /tmp.

How will you achieve this task?

 below command can we achieve copy /etc, including all subdirectories and files to /tmp. sudo cd /etc

cp -r /etc tmp

Q 4: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.

Q 5:How will you provide a count of all users on the system except for adm user?

$ who

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

Q 6: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, use the -S.

By default, it displays output in descending order (biggest to smallest in size).

-l flag means long listing and -a tells ls to list all files including (.) or hidden files.

 human-readable format by adding the -h

 And to sort in reverse order, add the -r flag.

Q 7:What command is used to clear history on the Linux server?

Removing history

 to delete a particular command, enter history -d <line number> . To clear the entire contents of the history file, execute history -c .

Q 8:Explain "Big Data" and what are five V's of Big Data BigData.

Big data is a collection of data from many different sources and is often describe by five characteristics:

Volume.

Variety.

Velocity.

Veracity.

 Value.

Q 9: What is Hadoop and its components?

Apache Hadoop is **an open source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data**.

1. Hadoop HDFS - Hadoop Distributed File System (HDFS) is the storage unit.

2.Hadoop MapReduce - Hadoop MapReduce is the processing unit.

3.Hadoop YARN - Yet Another Resource Negotiator (YARN) is a resource management unit.