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

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

:set number ##This is used to set specific line of the line inside the file.

:150 d ##This is representation of particular line to be delete

:wq ##This command used to exit&save from a file.

Question 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=10

b=20

sum=$(($a + $b))

$echo "Sum is: $sum"

answer:sum is:30

Question 3:User root wants to copy /etc, including all subdirectories and files to /tmp.How will you achieve this task?

cp -r /etc tmp

Question 4:Create a file that contains only the username and the user id of all the user present on the server

$cat /etc/passwd

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

$who

Question 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.

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

if you want to clear a particular line :

history -d <line number>.

if you want clear entire contents:

history -c

Question 8:Explain “Big Data” and what are five V’s of Big Data

big data is larger, more complex data sets, especially from new data sources. These data sets are so voluminous that traditional data processing software just can’t manage them. But these massive volumes of data can be used to address business problems you wouldn’t have been able to tackle before.

The 5 V's of big data (velocity, volume, value, variety and veracity)

Question 9: What is Hadoop and its components?

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. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly.

components of Hadoop:

1.Hadoop HDFS

2.Hadoop MapReduce

3.Hadoop YARN