

Linux and Introduction to Big Data Hadoop
Homework

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

Answer:

var/log/messages- are used to open a file.

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

150 d – is representation of particular line to delete.

:wq – the command used to exit and save the file.

Q2. Write a shell script to add two numbers?

Answer:

Program to add two numbers in shell script :

```
a=20
```

```
b=30
```

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

```
echo "Sum is: $sum"
```

Output:

Sum is:50

Q3. User root wants to copy /etc, including all subdirectories and files to /tmp. How will you achieve this task?

Answer:

```
sudo cd/etc
```

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

The above commands used to achieve copy/etc, including all subdirectories and file to /tmp.

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

Answer:

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

Using the above command we can create a file that contains the username and user id for all the user present on the server.

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

Answer:

We can provide a count of all users on the system except for adm user using `$ who` command.

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

Answer:

Using the command `$ ls-laShr` we can list all the files in `/tmp` in increasing order of their size. To list all the files and sort them according to their size we can use the command `-S`. By default, it will display the file name in descending order. The flag `-l` means long list and `-a` is used to tell the `ls` to list all files including hidden files. For human readable format we want to add `-h`. For sorting the list in reverse order we can add `-r` flag in it.

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

Answer:

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

Q8. Explain “Big Data” and what are five V’s of Big Data?

Answer:

Big Data:

Big data is a combination of unstructured, semi-structured or structured data collected by organizations. This data can be mined to gain insights and used in machine learning projects, predictive modeling and other advanced analytics applications.

Five V’s of Big data:

- ◆ Volume
- ◆ Value
- ◆ Variety

- ◆ Velocity
- ◆ Veracity

Q9. What is Hadoop and its components?

Answer:

Hadoop:

Hadoop is a framework that uses distributed storage and parallel processing to store and manage big data. It is the software most used by data analysts to handle big data, and its market size continues to grow.

Components of Hadoop:

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