Big data analytics
HOME WORK 1

Q1. 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.

Q2. Write a shell script to add two numbers?

Calculate the sum of two integers with pre initialize values in a shell script

$var1=10

$var2=20

$sum=$(($var1 +$var2))

$Echo $sum

Output

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 users present on the server.

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.

$cat /etc/passwd

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

Using the ($ who) 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 /tmp

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

- (-S) is used to sort all files according to their size

- (-h) is used to show the file size in human readable format

- (-r) is used to show the list of files in increasing order.

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

- If you 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?

BigData:
- The definition of big data is data that contains greater variety, arriving in increasing volumes and with more velocity.
- This is also known as the three Vs. Put simply, big data is larger, more complex data sets, especially from new data sources.

Five V's of Big Data:
- volume
- value
- variety
- velocity
- veracity

Q9. What is Hadoop and its components?

Handoop:

- 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,
- It is market size continues to grow. There are three components of Hadoop

components of the 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