Lecture Notes

Linux and Scripting

Session Overview

In this session, you will learn about the following:

- Introduction to Linux and Linux distributions
- Architecture of Linux
- Advantages of using Linux over other Operating Systems
- File System and Types of Users in Linux Operating System

Introduction to Linux

An operating system is a system software that acts as a link between the software and hardware components of a computer. It receives requests from the other software on the system and communicates them to the hardware.

Open source software refers to a program whose source code is publicly accessible so that the users can use, modify or distribute it as they see fit. You also came to know that the different kinds of Linux distributions are ubuntu, fedora, debian etc. A Linux distribution is an operating system based upon a singular Linux core and is available in a wide variety based upon who wants to use the OS for what purpose. There are over 600 kinds of Linux distributions based on whether a certain distribution (also called Distro) is made for embedded systems, personal computers, phones etc.

Linux operating system consists of-

- Hardware layer: It consists of peripheral devices like RAM, HDD, CPU. It facilitates the interaction of the Linux OS with the rest of the hardware components
- Kernel: It is the core component of the Operating System and serves as the connecting link between the software application of the computer with the external hardware. A kernel is usually the first program which loads when you start a computer.
- Shell: It is the interface between user and kernel. It takes commands from the users and executes the kernel's functions.
- Applications: They are all the utility programs that run on the shell such as a video player, web browser etc.

The reasons why Linux is preferred over other operating systems are-

- Open source: Source code of lLinux is readily accessible for everyone. Since Linux is freely redistributable, it can be created and shared for any purpose.
- Secure: Since Linux is open source software, anyone can review it and make sure that there are no bugs. Also, the admin access is not given to the users on lLinux by default which makes it more secure. If any harmful file finds a place in the system, it cannot be executed as a normal user lLinux does not access to the root directories on their computers (which makes it more secure than windows).
- Stable: It is very unlikely for lLinux systems to crash after an update, it rarely slows down or freezes which makes it stable. This is also because Linux is highly modular in nature, so different modules are mostly independent of each other, and thus, if one of them malfunctions, it does not affect any other modules.
- Runs on any hardware: Linux makes very efficient use of the system's resources. It's installation is really flexible and can be customized for users and for specific hardware requirements as well.
- Ease of maintenance: Maintaining the Linux OS is easy, as the user can centrally update the OS and all software installed. Linux systems have their own central software repository, which is used to update the system and keep it safe.

Important Linux Commands

The application of all the commands that are mentioned in the above video are as follows-
pwd : It stands for Print Working Directory. When invoked, the full path of the current working directory will be printed to standard output.

cd : It stands for Change Directory. It is used to change the current working directory. By default it will navigate you to the home directory.

cd / : This command will navigate you to the root directory. Root directory is the first directory in your file system.

cd .. : This command will navigate you to the parent directory of the current directory. Parent directory is the directory which is one level up from the current directory. Basically these commands help you "go back" in the directory tree.

cd or cd~ : This command is used to move to the home directory.

cd/abc/xyz : This command is used to directly go to the xyz subdirectory of abc. This command is useful when you already know the path of the folder/file which you wish to access & thus, can directly move there.

ls : It is used to list all the files and directories present within the file system.

cat : It stands for Concatenate. It is used to create one or multiple files, concatenate the files and view contents of a file.

Example:
1. To view a single file
$cat file_X
Output:
<display contents of file_X>

2. To view multiple files
$cat file_X file_Y
Output:
<display contents of file_X and file_Y>

3. Create a new file
$ cat >newfile
Output:
<create a file named newfile>

cp : It is a short form for copy. This command is used to copy the files or directories from one place to another.
Example-

cp source_file destination_file

This will copy the contents of the source file into the destination file. If the destination file does not exist in the system, it then creates one and copies the contents of the source file after it.
mv : It is a short form for move. This command is used to move the files or directories from one place to another.
Example-
mv source_file destination_file
This will replace the name of the source file with the destination file that is given as an argument.

mkdir : It stands for Make Directory. This command allows the user to create directories. We can do this by specifying the name of the directory after the command. It can also be used to create multiple directories at once by specifying multiple names after command.

rmdir : It stands for Remove Directory. It will remove each and every directory specified along with it in the command line, only if they are empty. If any directory is non empty then it won't be removed. To remove a parent directory and its subdirectory, it is important to specify subdirectory first so that after the subdirectory is removed, the parent directory becomes empty and can be removed.

rm : It stands for Remove. You can remove multiple files and directories by specifying their name along with this command. It does not remove anything by default.

rm -r abc : if you specify the name of the directory after the rm command it will not delete it if the directory is non empty. Using -r with rm command deletes the non empty directories too.

rm -i abc : This will request the user confirmation before deleting the directory.

rm -f abc : This will remove the files or directory without prompting for confirmation regardless of the file permissions.

touch : It is used to change the timestamps (last access time, last modification time and last change time) of a file. If no file is present in the system which is passed as an argument then it will create a file without any sort of content inside it.
touch -a filename : This command is used to change the access time of a file.
touch -m filename : This command is used to change the modification time of a file.

find : It is used to search a file in a file system by using some criterias like file name, last access time, last modification time, file permissions, owner or size.

grep : It is used to search a string of characters in a file. If that string is found it will print all the lines containing it.
grep -i “string” filename - It makes the string case insensitive. For example “upgrad”, “UPgrad” and “upGRAD” are considered the same by it.
grep -c “string” filename - we can find the number of lines that contains the string through this.
grep -v “string” filename - It displays the lines in a file which does not contain the string.
grep -l file1 file2 file3 - Displays the name of the file out of multiple files which contain the string.

sudo : Linux does not allow all the users to access some parts of the system to prevent sensitive files from being compromised. Sudo command is used to access these restricted files and operations. It temporarily elevates the privileges and allows users to complete sensitive tasks without logging in as the root user.It asks you for your personal password and confirms your request to execute a command by checking a file, called sudoers. Administrators can give certain users or groups access to some or all commands without those users having to know the root password, all that users information is stored in the sudoers file.

sort : It is used to rearrange contents in a file line by line. By default, the sorting is done line by line in which the numbers are kept before the letters and lowercase letters are kept ahead of uppercase letters.

sort -r filename : It sorts the input file text in reverse order.
sort -n filename : This is used to sort a file which contains numerical data. It sorts the input file numerically.
sort -nr filename : It is used to sort a file containing numeric data in reverse order.
sort -u filename : It removes duplicates while sorting.

sed : It stands for stream error. It is mainly used for performing various operations on a file like text substitution, find and replace. Insertion and deletion can also be done through it. The advantage of using this command is that you can edit your file without even opening it.
clear: It is used to clear the terminal screen.

echo : It is used to display any line of a text that is passed to it as an argument. Any argument that is passed to it is printed.

head : It is used to print the first N lines (10 by default) of one or more files. In case of more than one file the file name is printed first and then the required lines.
tail : It is used to print the last N lines (10 by default) of one or more files. In case of more than one file the file name is printed first and then the required lines.

chown : It stands for change owner. Every file in lLinux has its owner and a group. Permission to access the file can be different for owner, group and others outside the group. Chown command helps you to change this ownership of the file/directory with respect to owner and group.
chmod : It stands for change mode. Access mode of a file can be changed using this command. It works similar to chown. Chown changes the ownership of a file or a directory and chmod changes the permissions of files and directories.

AWK : It is used to see if a file contains some lines which matches with the given pattern and performs specific actions if the pattern is found.
For Example-
$ awk '/pattern/ {print}' file1.txt
This command prints all the lines in the file named file1.txt which has the word ‘pattern’ in it.

diff : It stands for differences. This command compares the contents of two files line by line and then lists differences in them along with line number.
free : It is used to display the total amount of free memory in our system along with the memory which is in use.
tar : It stands for Tape Archive. It is used to take backup by archiving multiple files or directories into a single archive file that can be moved from one disk to another easily.
zip : It puts one or more compressed files in a zip archive. It is similar to tar but tar bundles all the files together whereas zip compresses all the files and then bundles them together.

ssh : It is a protocol which stands for secure shell. It provides a secure connection between two hosts. It transfers the data into encrypted form between both the hosts, takes the input from another connected system and sends the output to it. Since the communication is encrypted it can also be used in insecure networks. It is mainly used to execute commands on a remote machine without using the shell prompt. The syntax for it is:
ssh hostname command

ssh-keygen : It is used to generate a public or private authentication key pair for ssh protocol which allows the user to connect to a remote system without any use of password. For each user, a separate key should be generated.

ifconfig - It stands for interface configuration. This command is used to view and change the configurations of the network interfaces on the system. It can be used to enable/disable network interface, display current network configuration information, set up an ip address and to assign a netmask or a broadcast address to the network interface.

ip - It stands for Internet Protocol. It is used to assign addresses to a network interface and to configure the network parameters. It works similar to ifconfig command but it shows all the interfaces which are enabled or disabled whereas, ipconfig shows the interfaces which are enabled.

ping : It stands for packet internet groper. It is used to check the connectivity status between a host and a source. It calculates the total time that is taken to send a packet and receive an acknowledgement from the source. This time is called latency. Low latency and fast ping means a faster connection.

traceroute : It traces the path that the packet takes from your computer to the one you specify. When a packet is sent, it hops from one location to another until it reaches its destination. This command shows us the number of hops the packet takes along with the time duration between each hop. This can help us resolve networking problems.

netstat : It gives us information about the network connections, the ports that are in use and the processes using these ports. It can be used to list out all the network connections on a system.

nslookup : It stands for name server lookup. It is used to get information about the DNS (Domain Name System) server such as obtaining a domain name or any other DNS record. This information can be used to solve various problems related to DNS.

host : It is used for DNS lookup operations. It is used to convert host names into IP addresses and vice-versa.

Arp : It stands for Address Resolution Protocol. It is used to manipulate the ARP cache present in the system. It is generally used to resolve the IP address to its MAC (Media Access Control) address.

cut : It is used for cutting sections from each line of an input file and then printing it on the standard output.
uniq : It is used to delete all the repeated lines that exist in an input file.
tr : It is used to translate and delete characters in a file. This translation includes changing uppercase to lowercase, deleting certain characters, finding and replacing characters and deleting repeating characters.

lsof : It stands for List of Open Files. It tells you which files are opened by which process. In linux pipes, sockets, directories etc are also considered as files, lsof gives us the list of all open files belonging to all active processes.
id : It is used to find out the user ID and the group ID of any user in the server.

whoami : It is used to display the username of the current logged in user on the system.

more : It is used to view files in a scrollable manner. If the file is large, it then displays one screen at a time. It allows the user to scroll up and down through the file.

less: It is used to display the contents of a file one page at a time. It has a faster access as compared to the more command because in case of a large file it accesses it page by page rather than accessing the complete file.

wc : It is used to find out the number of lines, the numbers of words, and characters count in a file which is passed as an argument to the command.

history : It gives us the list of all the previous commands that are used in the terminal session.

df : It is a short form for Disk Free. It is used to show the amount of disk space that is available on file systems. It tells you about the total space and available space on file systems.
du : It is a short form for Disk Usage. It is used to find out the amount of disk space that is taken by a file or a directory.

ps : It stands for Process Status. It tells you about all the processes that are currently running on the system along with their process ID (PID).

w : It is used to display who all are currently logged in on the system and what are they doing (there processes).

top : it is used to display all the currently running processes along with the resources that all the processes are using in the system. It refreshes itself and displays the information until it is stopped.

expr : It is used to perform operations like addition, subtraction, modulus (on integers), multiplication and division on the given set of expressions and then prints the output.

test : It is used to check file types and compare values. It is mainly used in conditionals on shell script.

kill : It is used to terminate the process manually. When the command is executed, a signal is sent to the process or a process group causing them to terminate. You need to specify the process ID (PID) with the command to terminate the process.

nohup : It is a short form for no hang up. It keeps the processes in running state even if the user exits the terminal or shell.

dd : It stands for data duplicator. It is used to convert and copy files which are passed to it as arguments.

su : It allows the user to execute commands with the privileges of another user which by default is a root user. You will have to enter a password and if authenticated you can use the privileges of another user.

netstat : It is used to display incoming and outgoing network connections, routing tables, and network interface statistics. It helps in troubleshooting network issues and verifying network connection statistics.

curl : It is used to transfer data from or to the server without user interaction. This is done by using any one of the supported protocols out of HTTP, HTTPS, FTP, FTPS, SCP, SFTP, TFTP, DICT, TELNET, LDAP or FILE.

dig : This command is used to query the DNS name servers for any task related to DNS lookup.

nmap: It stands for Network Mapper and is used for network mapping and security auditing. It is used to audit and discover local and remote open ports along with the information about networks and hosts.

route : This command is used to show the details of the route table and also manipulates the IP routing table.

at : It is used to execute a program or a mail at any given time in future. It accepts time in the form of hours:minutes to run the command at any specific time of the day.

Introduction to Shell and BASH

Shell provides an interface to input human readable commands and then converts into commands that the operating system can understand. Bash is one type of shell which acts as a command-line interpreter that runs in a text window where users can interpret commands to carry out various actions. A shell script is a program that runs on the shell and allows you to run a bunch of commands at once.

Variables play a very important role in scripting. They are the storage units in your script and facilitates a major chunk of operations you will be performing. Variables are simply containers in the computer memory to store the information you want. A variable is a string to which we assign a particular value. This value can be a number, a name of any file, a text or any form of data. Variables are spaces in a computer's memory that have names associated with them, where you can store data.

You need to tell the following two things to the computer about a variable:

- Name of the variable
- The information you would want to store in this variable

The process of doing so is known as variable declaration.

In scripting, +, -, *, and / are the symbols used for addition, subtraction, multiplication, and division respectively.

In the case of multiple calculations, the general BODMAS rule is followed. Remember, to prioritise something, you use parentheses. The mod operator is used to store the remainder of the division between two numbers. It is very useful in cases such as trying to find an odd number or checking divisibility. Increment and decrement operators are unary operators (one that takes a single operand/argument and performs an operation) which increases and decreases the value of argument by one respectively.

File operators are those operators which are used to check the particular property of a given file which is passed as an argument.

Relational and Logical Operators

Relational operators are those operators which define a relation between two given operands/arguments and return true or false based on the relation. The “and” operation results in a true value only both of the conditions involved are true. Whereas, the “or” operation results in a true value even if one of the conditions involved is true. In scripting -o is used to denote a logical “or” and -a denotes logical “and” operation.

Conditionals

Usually in situations in which there is only one choice, If statement is used and the situations in which there are two choices, and only one of those choices need to be executed at any given point of time If Else Statements are used. One of the conditions is mentioned inside the If block, and the other condition is mentioned inside the Else block.

The general points to keep in mind are:-

- If statements are used in those cases when there is only one block of code that needs to be executed.
- If Else statements are used in those cases where only one out of the two code blocks needs to be executed at any given point of time.
- One of the condition and action is mentioned in the If block
- The other condition and action are mentioned in the Else block.

Nested If Else basically means an If Else statement inside another If Else statement. You often have to check sub-conditions within existing conditions. For example, let’s say you’re going to a movie. First, you’ll check if the tickets are available at a theatre near your home. Then, you will check another condition, i.e. whether the price of the ticket is worth it. However, if the tickets aren’t available, you will not check this sub-condition.

Nested If Else statements help us in those situations where there might be successive boolean checks. They are helpful in those cases, where after checking one condition, you need to check another condition successively. For that, you first define the outer condition in the outer If Else block, and then write the inner condition in the inner If Else block.

Switch case statements are strictly used only when you have to only check the value of a variable and only then take an action. You now know that Switch has a default case, which is equivalent to the Else condition in the case of If Else statements. The default case is executed if no other case is met.

Switch is similar in functionality to Else If as it also checks multiple conditions and then gives the corresponding output. Although it has more readability, it has limited usage because it cannot evaluate the conditions within each case. Switch compares the input against predefined cases.

Loops

To summarise, here are the important components of any loop:

- A starting point.
- A loop control variable that's also called the counter, which keeps track of how many iterations a loop has to run. The counter is incremented after each iteration.
- A loop continuation condition which defines how many times a loop has to run. The condition makes sure that the loop runs until an endpoint is reached.

For specifying the range in for loop you can specify any value that you like for both the starting and ending value. If the first value is larger than the second value then in that case it will count down. It is also possible to specify by how much amount you have to increase or decrease the value by adding two dots (..) and the value to step by.

The until loop is executed as many as times the condition/command evaluates to false. The loop terminates when the condition/command becomes true on the other hand while loop is executed when the command is true otherwise it will terminate.

Break statements are used to exit the loop while the continue statements are used to skip some iterations inside the loop.

Functions

A function basically has a defined name and needs to be called when an action is required.

Function name is specified followed by the parenthesis. After that we list a set of commands which are enclosed in curly braces.

The second format starts with a reserved word function followed by the name of the function. Commands are enclosed within curly braces and no parenthesis is used in this format after the name of a function.

The values that are passed into the function are called arguments. Thus, instead of saying you are passing values/variables into the function, you can say you are passing arguments into the function.

One of the major advantages of using functions is that if something goes wrong, you know where to make corrections that will be reflected in the entire script.

Return command is used to return the value from a function. It will also terminate the execution of the function.

Disclaimer: *All content and material on the upGrad website is copyrighted material, belonging to either upGrad or its bona fide contributors, and is purely for the dissemination of education. You are permitted to access, print, and download extracts from this site purely for your own education only and on the following basis:*

- *You can download this document from the website for self-use only.*
- *Any copies of this document, in part or full, saved to disc or to any other storage medium, may be used for subsequent, self-viewing purposes or to print an individual extract or copy for non-commercial personal use only.*
- *Any further dissemination, distribution, reproduction, and copying of the content of the document herein, or the uploading thereof on other websites, or use of the content for any other commercial/unauthorized purposes in any way that could infringe the intellectual property rights of upGrad or its contributors is strictly prohibited.*
- *No graphics, images, or photographs from any accompanying text in this document will be used separately for unauthorized purposes.*
- *No material in this document will be modified, adapted, or altered in any way.*
- *No part of this document or upGrad content may be reproduced or stored on any other website or included in any public or private electronic retrieval system or service without upGrad's prior written permission.*
- *Any rights not expressly granted in these terms are reserved.*