Linux Commands for Hadoop/Big Data Admin

ls Command:

ls command is one of the most frequently used command in Linux.I believe ls command is the first command you may use when you get into the command prompt of Linux Box. We use ls command daily basis and frequently even through we may not aware and never use all the option available.In this article,we'll be discussing basis ls command where we have tried to cover as much parameters as possible

a) List Files using ls with no option

```
[divakar@divakar BD]# ls
a.txt  b.txt
[divakar@divakar BD]#
```

b) List Files With option –l

```
[divakar@localhost ~]$ ls -l
-rw-rw-r--. 1 divakar divakar    0 Feb 11 18:51 divakar.txt
-rwxr-xr-x. 1 divakar divakar    0 Feb 12 18:22 diva.txt
[divakar@localhost ~]$
```

c) View Hidden Files

List all files including hidden file starting with '.'.

```
[divakar@divakar ~]$ ls -al
drwx------. 27 divakar divakar 4096 Feb 21 14:22 .
drwxr-xr-x.  3 root    root    4096 Feb 10 13:52 ..
-rw-rw-r--.  1 divakar divakar    0 Feb 11 18:51 divakar.txt
-rwxr-xr-x.  1 divakar divakar    0 Feb 12 18:22 diva.txt
[divakar@localhost ~]$
```

d) List Files with Human Readable Format with option -lh

With combination of -lh option, shows sizes in human readable format.

```
[divakar@localhost ~]$ ls -lh
-rw-rw-r--. 1 divakar divakar    0 Feb 12 18:25 d1.txt
-rw-rw-r--. 1 divakar divakar    0 Feb 12 18:25 d2.txt
[divakar@localhost ~]$
```

e) List Files and Directories with '/' Character at the end

Using **-F** option with **ls** command, will add the **'/'** Character at the end each directory.

```
[divakar@localhost ~]$ ls -F
d1.txt  Desktop/     diva.txt*   Downloads/  Pictures/  Templates/  xyz/
d2.txt  divakar.txt  Documents/  Music/      Public/    Videos/
```

[divakar@localhost ~]$

f) List Files in Reverse Order

The following command with **ls -r** option display files and directories in reverse order.
[divakar@localhost ~]$ ls -r

xyz Templates Pictures Downloads diva.txt Desktop d1.txt Videos Public divakar.txt d2.txt
[divakar@localhost ~]$

g) Recursively list Sub-Directories

ls -R option will list very long listing directory trees. See an example of output of the command.

[divakar@localhost ~]$ ls -R

d1.txt Desktop diva.txt Downloads Pictures Templates xyz

h) Reverse Output Order

With combination of **-ltr** will shows latest modification file or directory date as last.

[divakar@localhost ~]$ ls -ltr

drwxr-xr-x. 2 divakar divakar 4096 Feb 10 18:55 Templates

drwxr-xr-x. 2 divakar divakar 4096 Feb 10 18:55 Downloads

drwxr-xr-x. 2 divakar divakar 4096 Feb 10 18:55 Videos

drwxr-xr-x. 2 divakar divakar 4096 Feb 10 18:55 Public

i) Sort Files by File Size

With combination of **-lS** displays file size in order, will display big in size first.

[divakar@localhost ~]$ ls -ls

0 -rw-rw-r--. 1 divakar divakar 0 Feb 12 18:25 d1.txt

0 -rw-rw-r--. 1 divakar divakar 0 Feb 12 18:25 d2.txt

4 drwxr-xr-x. 4 divakar divakar 4096 Feb 11 18:03 Desktop

0 -rw-rw-r--. 1 divakar divakar 0 Feb 11 18:51 divakar.

j) Display Inode number of File or Directory

We can see some number printed before file **/** directory name. With **-i** options list file **/**directory with inode number

[divakar@localhost ~]$ ls -i

272870 d1.txt 272869 diva.txt 397175 Pictures 397282 xyz

272871 d2.txt 397173 Documents 397172 Public

k) Shows version of ls command

Check version of ls command.

[divakar@localhost ~]$ ls --version

ls (GNU coreutils) 8.4

Copyright (C) 2010 Free Software Foundation, Inc

l) Show Help Page

List help page of ls command with their option.

[divakar@localhost ~]$ ls --help

Usage: ls [OPTION]... [FILE]...

List information about the FILEs (the current directory by default).

Sort entries alphabetically if none of -cftuvSUX nor --sort.

m) List Directory Information

With **ls -l** command list files under directory **/tmp**. Wherein with **-ld** parameters displays information of **/tmp** directory.

[divakar@localhost ~]$ ls -l /

dr-xr-xr-x. 2 root root 4096 Feb 10 19:34 bin

dr-xr-xr-x. 5 root root 1024 Feb 10 13:52 boot

n) Display UID and GID of Files

To display **UID** and **GID** of files and directories. use option **-n** with ls command.

[divakar@localhost ~]$ ls -n

-rw-rw-r--. 1 500 500 0 Feb 12 18:25 d1.txt

-rw-rw-r--. 1 500 500 0 Feb 12 18:25 d2.txt

Copying with the cp Command

a) How do I copy files?

cp filename1 filename2 -> copies a file.

b) How do I copy recursively?

cp –r dir1 dir2

c) To see copy progress pass –v option to cp command.

cp –v –r dir1 dir2

d) How do I confirm file overwriting?

cp –i dir1 dir2

e) Preserve the file permission and other attributes.

cp –p file1 file2

Deleting Files with the 'rm' Command

The rm command deletes the files.This command has several options,but should be used cautiously.

a) The rm command will delete one or several files from the command line.

rm file1

rm file1 file2 file3

b) One of the safer ways to use rm is through the -i or interactive option, where you'll be asked if you want to delete the file.

rm -i file1

c) you can also force file deletion by using -f option

rm -f file1

d) when we combine -f and -r, the recursive option , you can delete directories and all files or directories found.

rm -rf <directoryname>

Creating Directories with the 'mkdir' Command

The mkdir command can create one or several directories with a single command line.

a) Creating directories

mkdir <directoryName>

b) Creating multiple directories

mkdir <directoryName1> <directoryname2>

c) Creating Child under directories

mkdir temp/child

d) To build a hierarchy of directories with mkdir, you must use the -p,or parent option, for example

mkdir -p temp5/parent/child

Removing Directories with the 'rmdir' Command

a) The rmdir command is used to remove directories. To remove a directory, all you have to do is type

rmdir <DirectoryName>

b) Removing directories and sub directories as well

rm -rf

Renaming Files with the 'mv' command.

The mv command, called a rename command but know to many as a move command

a) Move the data file1 to file2

mv <file1> <file2>

b) The mv command can work silently, or as with rm, you can use the -i (interactive) option

mv -i <file1> <file2>

Creating Hard and Symbolic Links with the 'ln' Command

The ln command creates both types of links.If you use the ln command to create a hard link, you specify a second file on the command line you can use to reference the original file,for Ex

ln file1 file2

#ln -s file1 file 2

ps command: The ps command will show information about current system process.

Ps-> The user's currently running processes.

Ps –f -> Full listing of the user's currently running process.

ps –ef -> Full listing of all process, except kernel process.

Ps –A -> All process, including kernel process.

Ps auxw -> wide listing sorted by percentage of CPU usage.

Last: The last command shows the history of who has logged in to the system since the wtmp file was created.

Who: The who command gives this output for each logged-in user: username,tty .login time and where the user logged in from.

W: The W command is really an extended who

Checking your installation Files
rpm –qa | grep ^x

you should receive the output similar to the following..

xorg-x11-drv-apm-1.2.2-1.1.el6.x86_64

xorg-x11-drv-penmount-1.4.0-5.el6.x86_64

xorg-x11-drv-ast-0.89.9-1.1.el6.x86_64

xorg-x11-drv-aiptek-1.3.0-2.el6.x86_64

Installing the X Files

rpm –ivh <filename>

Moving to different directories with the cd command.
cd /usr/bin

cd ..

cd ../..

cd or cd –

Knowing where you are with the pwd command.
pwd

pwd –help

/bin/pwd --help

Searching directories for matching files with the find command
Syntax: find where-to-look criteria what-to-do

Find /usr –name spell –print

Listing and Combining Files with the cat Command
The cat (Concatenate file) commands are used to send the contents of files to your screen.

a) cat a.txt
[abc@master]$ cat a.txt

Hello Hello Hello

Hello hello hello

[abc@master]$

b) The cat command also has a number of options. If you'd like to see your file with line numbers, you can use the n-option

#cat –n a.txt

1 Hello Hello Hello

2 Hello hello hello

c) You can also use cat to look at several files at once.

cat –n test*

1 Hello Hello Hello

2 Hello hello hello

[root@localhost ~]# cat a.txt

aksjlkdj

[root@localhost ~]#

[root@localhost ~]# cat b.txt

alksjdjkjaskjkaj

;lajslkdjja

lkashdjfajs

[root@localhost ~]#

d) As you can see, cat has also include a second file in its output.

[root@localhost ~]# cat a.txt b.txt

aksjlkdj

alksjdjkjaskjkaj

;lajslkdjja

lkashdjfajs

[root@localhost ~]#

cat > c.txt

e) To see the numbers using cat

```
  # cat -n divakar.txt
100    divakar 20000
200    diiia   10000
399    ksjjkj  30999[
```

Reading the files with the 'more' command

[root@localhost ~]# more c.txt

kajslkj

lkasjdkfjjs

klakjshkdjfhkahs

[root@localhost ~]#

Browsing Files with the 'less' command.

less c.txt

Reading the Beginning or End of Files with the head and tail Commands.

head -5 /usr/man/man.txt

head -5 –q /usr/man/man.txt

tail -12 /var/log/message/a.txt

The more command is one of a family of Linux commands called pagers.

Creating Files with the 'touch' Command

The touch command is easy to use, and generally, there are two reasons to use it .The first reason is to create a file, and the second is to update a file's modification date.

a) To create a file with touch,use

touch newfile

#ls –l newfile

-rw-r--r-- 1 divakar divakar 0 Feb 21 14:12 newfile

b) To Change time stamp

touch –t 1225110099 newfile2

Trap: when the program is terminated before it would normal end, we can catch an exit signal.

0- Normal termination, end of script.

1-SIGHUP -> hang up, line disconnected

2-singint-> terminal interrupts, usually ctrl+c

3- SIGQUIT -> Quit Key, Child process to die before terminating.

9-SIGKILL->kill -9 commands can't trap this type of exit status.

15-SIGTERM →kill command's default action.

19-SIGSTOP->stop, usually ctrl+z

17 -> dintfunc

Ex: Kill –9 <ps Id>

df -> Report how much free disk space is available for each mount you have.

df –a --all -> include dummy file system.

df –B 100, --block-size=SIZE -> use SIZE –byte blocks

df –h ->human readable →print sizes in human readable format.

df –I -> list inode information instead of block usage.

Df –k -> like --block –Size =1 k.

df –T ->Print file System Type.

du -> disk usage

du ->tells you how much disk space a file occupies.

du –a ->display the space that each file is taking up.

Du –h -> which can make the output easier to read by displaying it in KB /M/G.

Du –sh -> The *-s* (for *suppress* or *summarize*) option tells du to report only the total disk space occupied by a directory tree and to suppress individual reports for its subdirectories

top →displays top CPU process.

The **top** program provides a dynamic real-time view of a running system.

Uname

Free -> displays information about free and used memory on the system.

-b,-k,-m,-g show output in bytes, KB, MB, or GB

-l show detailed low and high memory statistics

-o use old format (no -/+buffers/cache line)

-t display total for RAM + swap

-s update every [delay] seconds

-c update [count] times

-V display version information and exit

awk-> The awk command is powerful method for processing or analyzing text files,in particulat data files that are organized by lines (rows) and cloumns.

How to add user:
Useradd div --to add user Id

Passwd div –to set password

Set Account disable date:

```
useradd -e {yyyy-mm-dd} {username}
useradd -e 2008-12-31 jerry
```

Set default password expiry:

```
useradd -f {days} {username}
useradd -e 2009-12-31 -f 30 jerry
```

How Can I see the entire users list on Linux Server?
cat /etc/passwd

Vim /etc/passwd

cat /etc/passwd | grep home | cut -d':' -f1

How to convert you as root, create new user and setting password.
[divakar@localhost Desktop]$ su root
Password: <give your user passwd>

[root@localhost Desktop]# useradd diva1

[root@localhost Desktop]# passwd diva1

Changing password for user diva1.

New password:

BAD PASSWORD: it is too short

BAD PASSWORD: is too simple

Retype new password:

passwd: all authentication tokens updated successfully.

[root@localhost Desktop]#

How to add user in sudo list.
Go to vi /etc/sudoers

Allow root to run any commands anywhere

root   ALL=(ALL)   ALL

divakar   ALL=(ALL)   ALL

Man ->

Find files with the whereis command.
Whereis find

You can also use whereis to find only the binary version of the program with

Whereis –b find

If whereis cann't find your request , you'll get an empty return string, for example

Whreis foo

It will search in entire system.

Not limiting searches to known directories such as /usr/man, /usr/bin, or /usr/sbin can speed up the task of finding files.

Although whereis is faster than using find to locate programs or manual pages.

Locate is faster than whereis

Locating files with locate command.

Finding a file using locate is much faster than the find command because locate will go directly to the database file,find any matching filenames, and print its results.

Locate *.ps

Locate resides in /var/lib

Moving different directories with cd command.

cd or

cd ../.. or

cd or cd –

Knowing where you are wit pwd command.

Go to /user/local and type pwd to know your directory.

ps -ef | grep aneel

pkill -f

Link while copying the data?

cp –i test 1 test2

Getting command Summaries with whatis and apropos

Specifying other directories with ls

ls /usr/bin

Listing Directories with the dir and vdir commands

#dir :-> this command works like the default ls command,listing the files in sorted columns

Vdir -> The vdir command works like the ls –l option, and presents a long format listing by default,

Graphic Directory listings with the tree command

tree /var/lib

tree –d /usr/local/netscape/

Search Inside Files with the 'grep' Command

grep command will help to search any words in file
Ex: cat hive.log | grep loaded [Here we are searching loaded files in log]

Compressing files with the 'gzip' command

gzip file.tar

Compressing Files with the 'compress' Command

#compress file

To uncompress a file,use
uncompress file.Z

Running Programs in the Background
nohup ./divakar.sh &

Checking the Connection
Using the 'ipconfig' Command, This Command will help to identify the IP address of your system
ifconfig
[divakar@localhost ~]$ ifconfig
eth0 Link encap:Ethernet HWaddr 00:0C:29:7F:24:89
inet addr:192.168.64.130 Bcast:192.168.64.255 Mask:255.255.255.0
inet6 addr: fe80::20c:29ff:fe7f:2489/64 Scope:Link

Using the 'netstat' Command
The netstat command is the definitive command for checking your network activity,connections,routing tables,and other network messages and statistics.
netstat

Using the ping Command
ping <hostname>.com

Find Hostname of the System
#hostname

How to replace One word with another word in vi:
:%s/old-text/new-text/g

Hadoop File system Commands

The FileSystem (FS) shell is invoked by `bin/hadoop fs <args>`. All the FS shell commands take path URIs as arguments. The URI format is *scheme://autority/path*. For HDFS the scheme is *hdfs*, and for the local filesystem the scheme is *file*. The scheme and authority are optional. If not specified, the default scheme specified in the configuration is used. An HDFS file or directory such as */parent/child* can be specified as *hdfs://namenodehost/parent/child* or simply as */parent/child* (given that your configuration is set to point to *hdfs://namenodehost*). Most of the commands in FS shell behave like corresponding Unix commands. Differences are described with each of the commands. Error information is sent to *stderr* and the output is sent to *stdout*.

Ex : hadoop fs -ls /R
[-ls <path>]
[-lsr <path>]
[-df [<path>]]
[-du <path>]
[-dus <path>]
[-count[-q] <path>]
[-mv <src> <dst>]
[-cp <src> <dst>]
[-rm [-skipTrash] <path>]
[-rmr [-skipTrash] <path>]
[-expunge]
[-put <localsrc> ... <dst>]
[-copyFromLocal <localsrc> ... <dst>]
[-moveFromLocal <localsrc> ... <dst>]

[-get [-ignoreCrc] [-crc] <src> <localdst>]
[-getmerge <src> <localdst> [addnl]]
[-cat <src>]
[-text <src>]
[-copyToLocal [-ignoreCrc] [-crc] <src> <localdst>]
[-moveToLocal [-crc] <src> <localdst>]
[-mkdir <path>]
[-touchz <path>]
[-test -[ezd] <path>]
[-stat [format] <path>]
[-tail [-f] <file>]
[-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
[-chown [-R] [OWNER][:[GROUP]] PATH...]
[-chgrp [-R] GROUP PATH...]

[-help [cmd]]

distcp:
hadoop distcp <Source Directory> <Destination Directory>

lsblk :To display block device information

```
[root@uuuuuuu]# lsblk
NAME                          MAJ:MIN RM  SIZE RO TYPE MOUNTPOINT
sda                             8:0    0  2.7T  0 disk
├─sda1                          8:1    0  200M  0 part /boot/efi
├─sda2                          8:2    0  512M  0 part /boot
├─sda3                          8:3    0  500G  0 part
│ ├─RootVolGroup00-lv_root (dm-0)    253:0    0   16G  0 lvm  /
│ ├─RootVolGroup00-lv_swap (dm-1)    253:1    0   32G  0 lvm  [SWAP]
│ ├─RootVolGroup00-lv_var (dm-2)     253:2    0   16G  0 lvm  /var
│ ├─RootVolGroup00-lv_opt (dm-3)     253:3    0   15G  0 lvm  /opt
│ ├─RootVolGroup00-lv_home (dm-4)    253:4    0   15G  0 lvm  /home
│ ├─RootVolGroup00-lv_optmapr (dm-5)  253:5    0  200G  0 lvm  /opt/mapr
│ └─RootVolGroup00-lv_optcores (dm-6) 253:6    0  100G  0 lvm  /opt/cores
└─sda4                          8:4    0  2.2T  0 part
```

How to Check installed packages:
rpm -qa | grep hive

To Locate Java:
locate java | grep bin |less

General Startup on VI

To use vi: vi filename

To exit vi and save changes: ZZ or :wq

To exit vi without saving changes: :q!

To enter vi command mode: [esc]

Counts

A number preceding any vi command tells vi to repeat that command that many times.

Cursor Movement

h move left (backspace)

j move down

k move up

l move right (spacebar) [return] move to the beginning of the next line

$ last column on the current line

0 move cursor to the first column on the current line

^ move cursor to first nonblank column on the current line

w move to the beginning of the next word or punctuation mark

W move past the next space

b move to the beginning of the previous word or punctuation mark

B move to the beginning of the previous word,ignores punctuation

e end of next word or punctuation mark

E end of next word, ignoring punctuation

H move cursor to the top of the screen

M move cursor to the middle of the screen

L move cursor to the bottom of the screen Screen Movement

G move to the last line in the file

xG move to line x

z+ move current line to top of screen

z move current line to the middle of screen

z- move current line to the bottom of screen

^F move forward one screen

^B move backward one line

^D move forward one half screen

^U move backward one half screen

^R redraw screen (does not work with VT100 type terminals)

^L redraw screen (does not work with Televideo terminals) Inserting

r replace character under cursor with next character typed

R keep replacing character until [esc] is hit

i insert before cursor

a append after cursor

A append at end of line

O open line above cursor and enter append mode

Deleting

x delete character under cursor

dd delete line under cursor

dw delete word under cursor

db delete word before cursor

Copying Code

yy (yank)'copies' line which may then be put by the p(put) command. Precede with a count for multiple lines.

Put Command brings back previous deletion or yank of lines,words, or characters

P bring back before cursor

p bring back after cursor

Find Commands

? finds a word going backwards

/ finds a word going forwards

f finds a character on the line under the cursor going forward

F finds a character on the line under the cursor going backwards

t find a character on the current line going forward and stop one character before it

T find a character on the current line going backward and stop one character before it;

repeat last f, F, t, T

Miscellaneous Commands

. repeat last command

u undoes last command issued

U undoes all commands on one line

xp deletes first character and inserts after second (swap)

J join current line with the next line

^G display current line number

% if at one parenthesis, will jump to its mate

mx mark current line with character x

'x find line marked with character x

NOTE: Marks are internal and not written to the file.

Line Editor Mode

Any commands form the line editor ex can be issued upon entering line mode.

To enter: type ':

To exit: press[return] or [esc]

ex Commands

For a complete list consult the

UNIX Programmer's Manual

READING FILES

copies (reads) filename after cursor in file currently editing

:r filename

WRITE FILE

:w saves the current file without quitting

MOVING

:# move to line #

:$ move to last line of file

SHELL ESCAPE

executes 'cmd' as a shell command.

:!'cmd'

```
$ scp foobar.txt your_username@remotehost.edu:/some/remote/directory
```

MapR Regular Commands

Use the following commands to list MapR services
maprcli service list

```
[root@ebdp-wc-d01d logs]# maprcli service list
name          state  logpath                                displayname
fileserver    0      /opt/mapr/logs/mfs.log                 FileServer
webserver     0      /opt/mapr/logs/adminuiapp.log          Webserver
cldb          0      /opt/mapr/logs/cldb.log                CLDB
tasktracker   0      /opt/mapr/hadoop/hadoop-0.20.2/logs    TaskTracker
jobtracker    0      /opt/mapr/hadoop/hadoop-0.20.2/logs    JobTracker
hoststats     0      /opt/mapr/logs/hoststats.log           HostStats
```