Linux Command Line Essentials

MSBA 6630 Prof De Liu

Goals

- In this section we will introduce some basic terminologies and commands in Linux operating system.
- In this section, you'll learn
  - A bit history of Linux and its relationship with Hadoop
  - Basic concepts about Linux file system and shell.
  - Basic Linux commands for file operations
  - Basic Linux commands for job control

What is Linux/Unix

- A multi-user and multi-task operating system
- Developed in 1991 by Linus Torvalds, inspired by Unix
- It has many "flavors" or distributions (called "distro")
  - Debian derivatives
    - Ubuntu (2004, based in South Africa, influence by Debian)
    - Debian (1996, stable and conservative)
  - Red Hat derivatives
    - Red Hat Enterprise Linux (REHL) (commercially supported)
    - Fedora (free, strong in security and enterprise features, but inferior on desktop usability)
    - CentOS (2003, free RHEL, well tested and reliable)

Ref: http://goo.gl/WcjYGK

Carlson School of Management

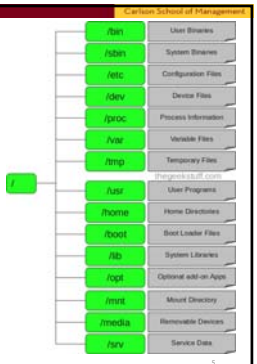
Why do you need know a bit about Linux/Unix commands?

- Hadoop ecosystem is native to Unix/Linux environment.
- Hadoop file system emulates Unix and uses similar commands.
- Cloud computing facility (e.g. Amazon Cloud Computing) may require you to use similar command-line interface.
- Mac Users: Mac OS is form of Unix, you'll find many similarities.

Carlson School of Management

Linux file system

- Linux has no concept of "file extension"
  - you can name your files the way you want.
- File names are **case sensitive**.
- The only special characters allowed in file names are period, dash, and underscore
- Organization of files
  - `/:` root of the file system
  - `/etc:` the configuration files for the system.
  - **`/home`**: where users keep their personal work. In general, this is the only place users are allowed to write files.



5

Carlson School of Management

Bash Shell on CentOS

- What is a Shell?
  - The shell is an interactive command interpreter environment (CLIs, command line interface) that can take commands from keyboard and run it.
    - More powerful than a Window's "command".
  - Many different shells
    - Bash (Bourne Again Shell), ksh, tcsh, zsh
- What is a terminal?
  - Using a **terminal** to interact with a shell
    - Many different terminals: xterm, rxvt, konsole, gnome-terminal, eterm

You can start a terminal by launching it from a window manager (look for programs such as terminal, xterm etc). You can start several of these terminals.

6

Linux Commands Structure

- A linux command typically consists of
  - The command itself, e.g. **`ls`**
  - The options
    - In short form **`-a -h -l`**, or **`-ahl`**
    - In long form **`--all --human-readable`**
    - May require values **`--tabsize=5`**
  - The arguments:
    - File name, text, etc

```
ls -l /var/log
```

command option(s) argument(s)

7

Look around

- What is your current directory?
  - `pwd` (print working directory)
- What is in your directory?
  - `ls` : list content of the current directory
  - `ls -l` : long form, including permissions
  - `ls -R` : display files in directory recursively
  - `ls -a` : display hidden files
  - `ls /` :list what is in your root directory
- Change the current directory
  - `cd /usr/bin` :enter /usr/bin
  - `cd /` : enter the root directory
  - `cd ..` : enter parent directory
  - `cd ~` : enter home directory
  - `cd ../../training_materials/`

autocompletion: after typing "tra", use tab key to auto complete the rest of the directory name. Two tabs to list options

8

File Operations

- copy files and directories
  - **`cp`** `file file2` : copy the `file1` to `file2` (overwrite if `file2` exists)
  - **`cp`** `file1 dir1` : copy file1 to inside of directory `dir1`
  - **`cp`** `-i file1 file2` : copy interactively (if `file2` exists, prompt)
- move or rename files and directories
  - **`mv`** `file1 file2`: rename `file1` to `file2` or replace `file2` with `file1` (if `file2` exists).
  - **`mv`** `file1 dir1`: move `file1` to directory `dir1`
- remove files and directories (careful, because there is no "undelete")
  - **`rm`** `file1`: remove a file
  - **`rm -r`** `directory`: remove a directory recursively
- Find files
  - **`find . -name`** `"test*"`: find a file starting with "test" in the current folder (".")
  - **`find ~/training_materials -name`** `"*test*"`:

9

Carlson School of Management

File Operations (cont.)

- Create directories
  `mkdir dir1`
- Use Wildcards
  `ls g*.txt` : list all txt files start with letter g
  `ls g???.txt`: list all txt files with names like "g" followed by three characters.
  `rm ad_data[1-9].txt` : remove `ad_data1.txt` to `ad_data9.txt`. May also use `[a-z]` and `[A-Z]`

10

Carlson School of Management

View large text files

- **`less`**: Display text file content interactively
  Page up (b)/down (space): scroll back/forward one page.
  `/characters`: search forward for characters
  n: search again.
  q: quit
- **`head/tail`**: display the first/last 10 lines of a text file
  `head ad_data1.txt`
  `head –n 20 ad_data1.txt`
- **`cat`** `filename | more`: page by page display (q to quit)
- **`grep`** is used to selectively print a line based on matching patterns.
  `grep "word" filename`
  `cat filename | grep "word"`
  `grep –i "Word" filename` : (the -i option for case insensitive)

11

Carlson School of Management

I/O Redirection and Pipes

- In Unix, output of one command can be used for input of another command.
  – redirect output
  `ls –l > file_list.txt`: results are stored in a new file file_list.txt
  – Redirect input
  `sort < file_list.txt`: sort the results of file_list.txt
  – Pipe operator "|"
  `cat file | more` : show the content of a file screen by screen
  `grep –i "the" filename | less`: output of `grep` command is fed into `less`

12

Carlson School of Management

Manipulate text files with sed, awk and sort

- **`wc:`** print newline, word, byte counts**.** `wc –l:` print line count
- **`sort`**: sort lines of text files
  `sort` : dictionary sort
  `sort -n` : sort the rows but treat them as numbers.
  `sort -u` : sort and remove duplicate lines.
- **`sed`**: "streaming editing", for manipulating text files line by line.
  `sed "s/MSBA/MS in Business Analytics/" /path/to/file`
  – Search all "MSBA" and replace it with "MS in Business Analytics" in the given file.
- **`awk`**: extract out programmatically determined data from text. Assuming delimited (default tab and spaces)
  `awk '{print $2}' simple_data.txt`
  where `'{print $2}'` is the awk program, telling it to print the 2nd column.

13

Carlson School of Management

Edit Text Files

- GUI based:
  – `gedit filename &:` edit file in graphical text editor gedit.
    - `gedit` is a user-friendly graphical text editor. In addition,
    - `&`: start the application in the background so you can continue to use the terminal after gedit starts in a window.
- TEXT based
  – `vi (vim)`: a very powerful text-based editor with a learning curve
  – `nano`: another text-based editor

14

Carlson School of Management

Helpful Bash Tips

- `clear`: to clear the screen.
- `Up and down arrows`: to retrieve a previous command.
- `Ctrl+u`: to delete (cut) the current line
- `Ctrl+a/Ctrl+e`: to move to the beginning/end of the line
- `Ctrl+Insert or Ctrl+y` (or a middle button click): paste copied content
- `history`: to show a history of linux commands you've used.
  – Then use `!<command number>` to rerun a command

15

Job Control

- We all ready know that Linux is a multi-task operating system. Here are a few job related commands
  - &: run a process in the background.
    - E.g. `gedit file1 &`
  - ps - list the processes running on the system
    - `ps ux`: list current users' processes complete info
    - `ps ux | grep pyth`: list only processes that contain "pyth"
  - `kill` - send a signal to one or more processes
    - `kill 1234`: where 1234 is the process id (pid)
    - `kill -9 1234`: send a harsh kill signal to skill the process if the regular kill fails.

16

Review Questions

- Display large text files 
- The "|" operator 
- Linux commands 

17

Additional Resources

- Lynda.com video lecture: Learn Linux Command Line Basics – **U of M free access**
  - https://www.lynda.com/Linux-tutorials/Learn-Linux-Command-Line-Basics/435539-2.html
  - Read sections 1-4 (about 1h30min)
- A Practical Guide to Linux® Commands, Editors, and Shell Programming (book) – **U of M free access**
  - https://goo.gl/YzeYbd
- Linuxcommand.org: Learning the Shell
  - A more detailed explanation of the linux shell environment.
  - http://linuxcommand.org/lc3_learning_the_shell.php

18