Linux Command Line Essentials

MSBA 6631 Prof De Liu

Learning Objectives

- 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
  - Commands for text file processing

Linux Command Line Essentials

INTRODUCTION TO LINUX OPERATING SYSTEM

- Describe the characteristics of Linux OS
- Why we need to learn a little bit about Linux commands?
- How Linux organizes its files?
- What is a Shell?

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
  - 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

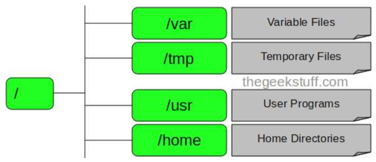
Why do you need know a bit about Linux/Unix commands?

MLOps is a set of practices and tools that help ensure that machine learning models can be developed, deployed, and maintained efficiently and reliably in real-world applications.

It encompasses aspects such as version control, collaboration, automation, deployment, and scalability.

- 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) often requires you to use the command-line interface
- Commands are requirement for **automation** and **MLOPs**. They make you a more efficient and productive data scientist

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.

Linux: `/home/deliu/downloads/chapter1.pdf`
Windows: `c:\Users\deliu\Downloads\Chapter1.pdf`

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

Linux Command Line Essentials

BASICS LINUX COMMANDS

- Understand the anatomy of a Linux command
- Basic commands for navigating the Linux directories
- How to use relative/absolute path?
- How to edit/copy/paste commands?

Linux Commands Structure

- A Linux command typically consists of
  - The command itself, e.g. `ls`
  - The arguments:
    - File name, text, etc
  - The options
    - In long form `--all` `--human-readable`
    - In short form `-a` `-h` or `-ah` (same as `-a -h`)
    - Some options have values `--tabsize=5` or `-t 5`

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

command option(s) argument(s)

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

Absolute and Relative path

- Absolute path starts with /

```
cd /
cd /vagrant
```

- Relative path is relative to the current directory and it does not start with /

```
cd training_materials
cd downloads
cd ../cloudera
```

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
- `Shift+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

Linux Command Line Essentials

ESSENTIAL FILE OPERATIONS USING SHELL COMMANDS

- How to copy/move/delete/find files?
- How to create and remove directories?

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*"`:

File Operations (cont.)

- Create directories

  `mkdir` `dir1`

- Use Wildcards

  `ls` `g*.txt` : list all txt files start with letter g

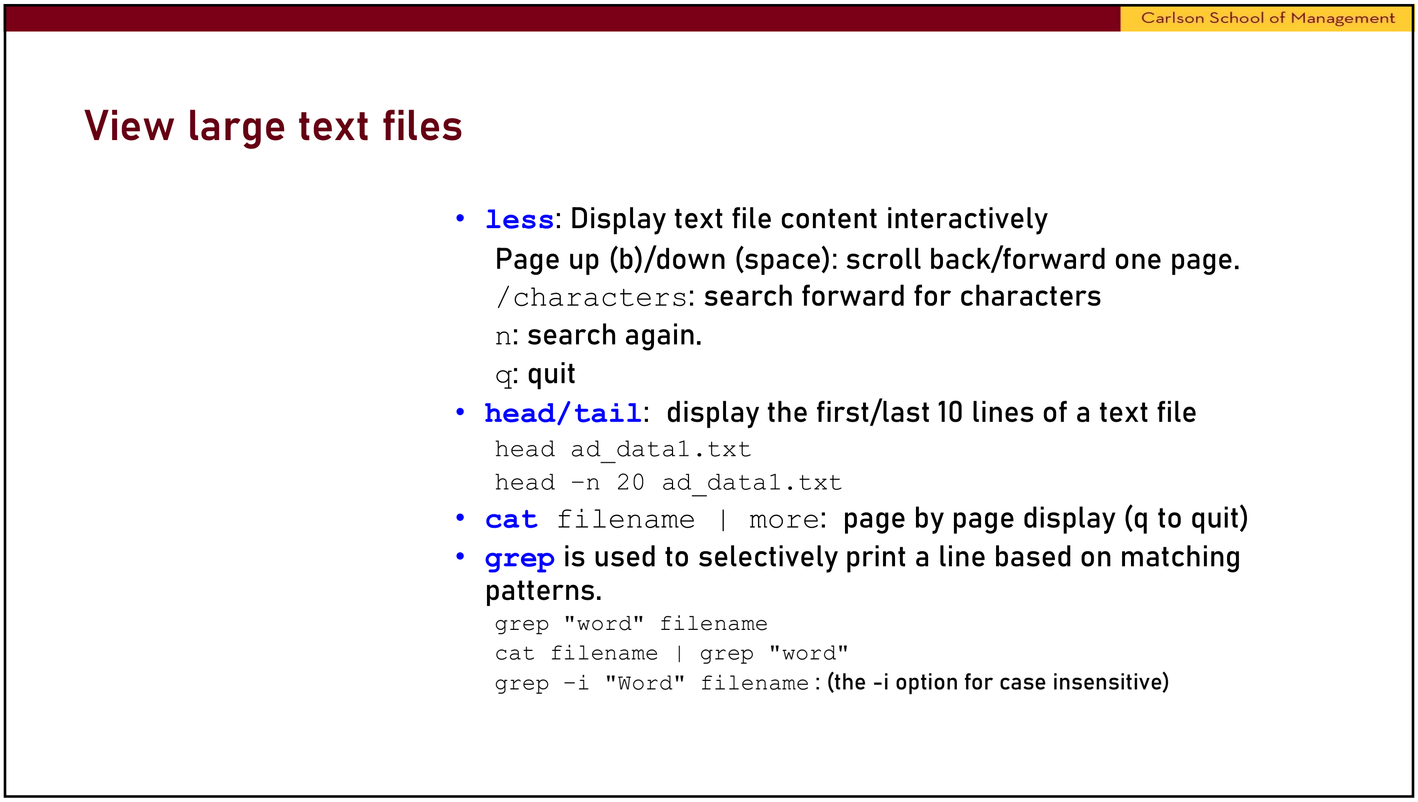
  `ls` `g???.txt`: list all txt files with names like “g” followed by three characters.

  `rm` `ad_data[1-9].txt` : remove `ad_data1.txt` to `ad_data9.txt`. May also use `[a-z]` and `[A-Z]`

Linux Command Line Essentials

HANDLING TEXT FILES

- How to view a large text file?
- How to output lines matching a pattern?
- How to redirect input / output?
- How to connect multiple commands using "pipe"?
- How to count the number of words/lines in a file?
- How to sort lines of a text file?
- How to "stream edit" a text file?
- What text editors can be used to edit a text file?



Carlson School of Management

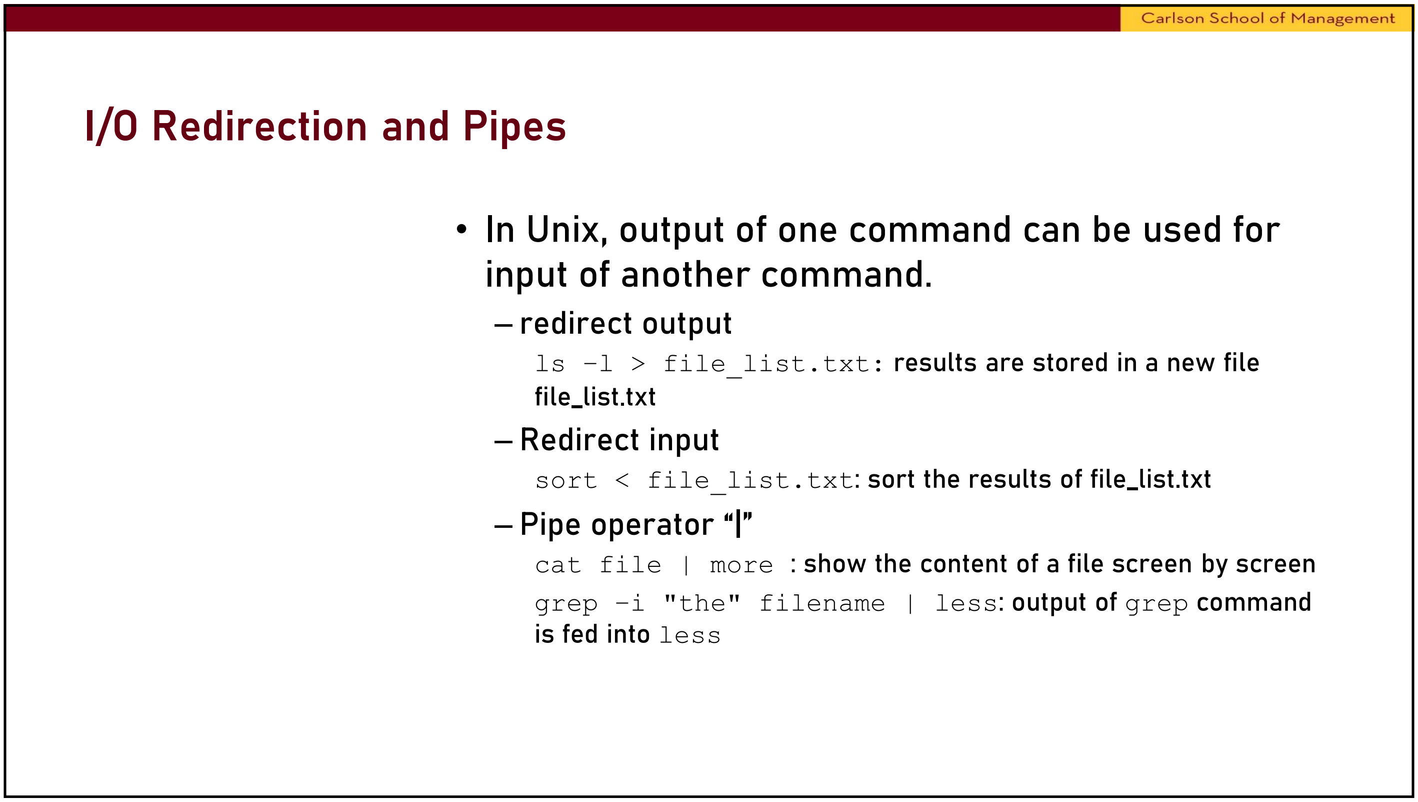
View large text files

- **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)

17

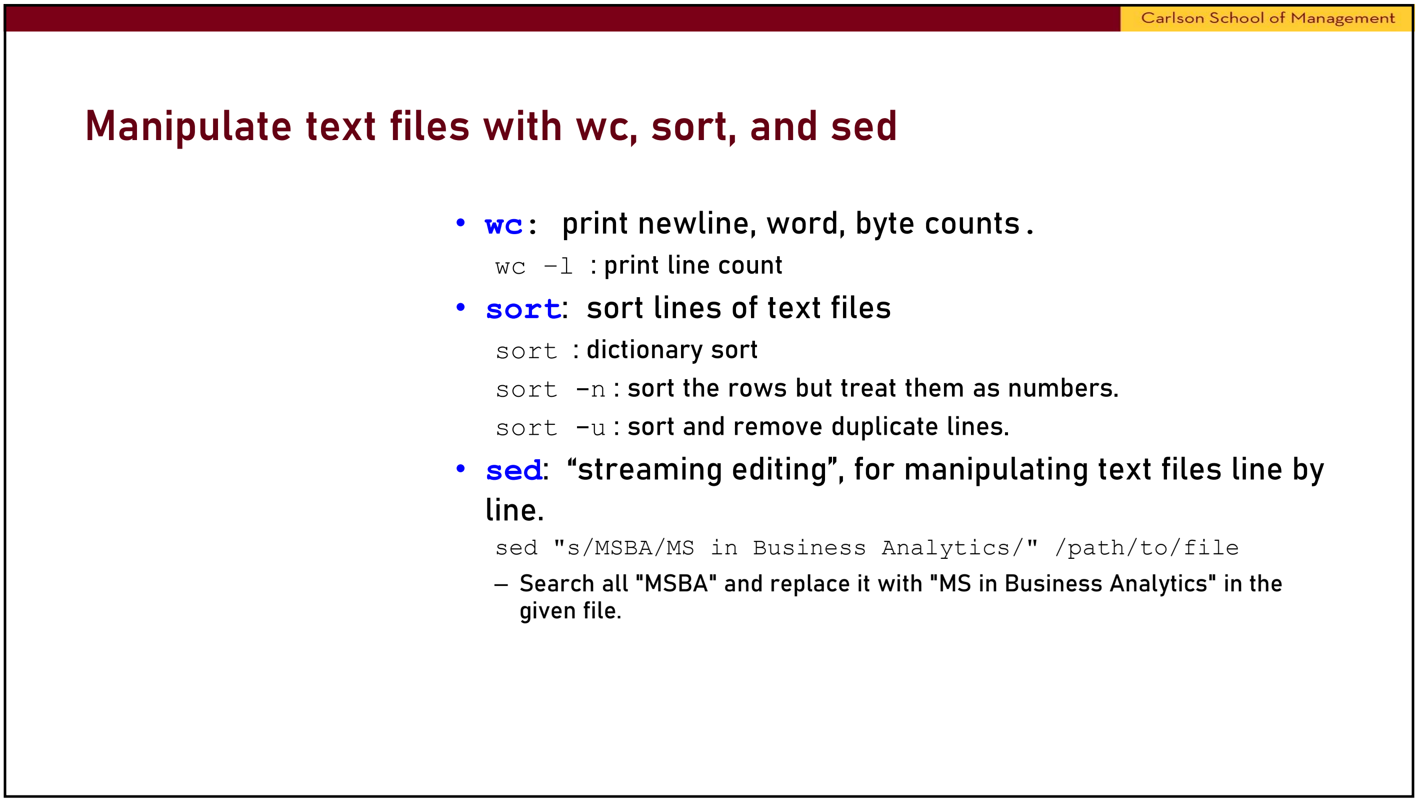


Carlson School of Management

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`

18



Carlson School of Management

Manipulate text files with wc, sort, and sed

- **wc**: print newline, word, byte counts.

  `wc -l` : print line count
- **sort**: sort lines of text files

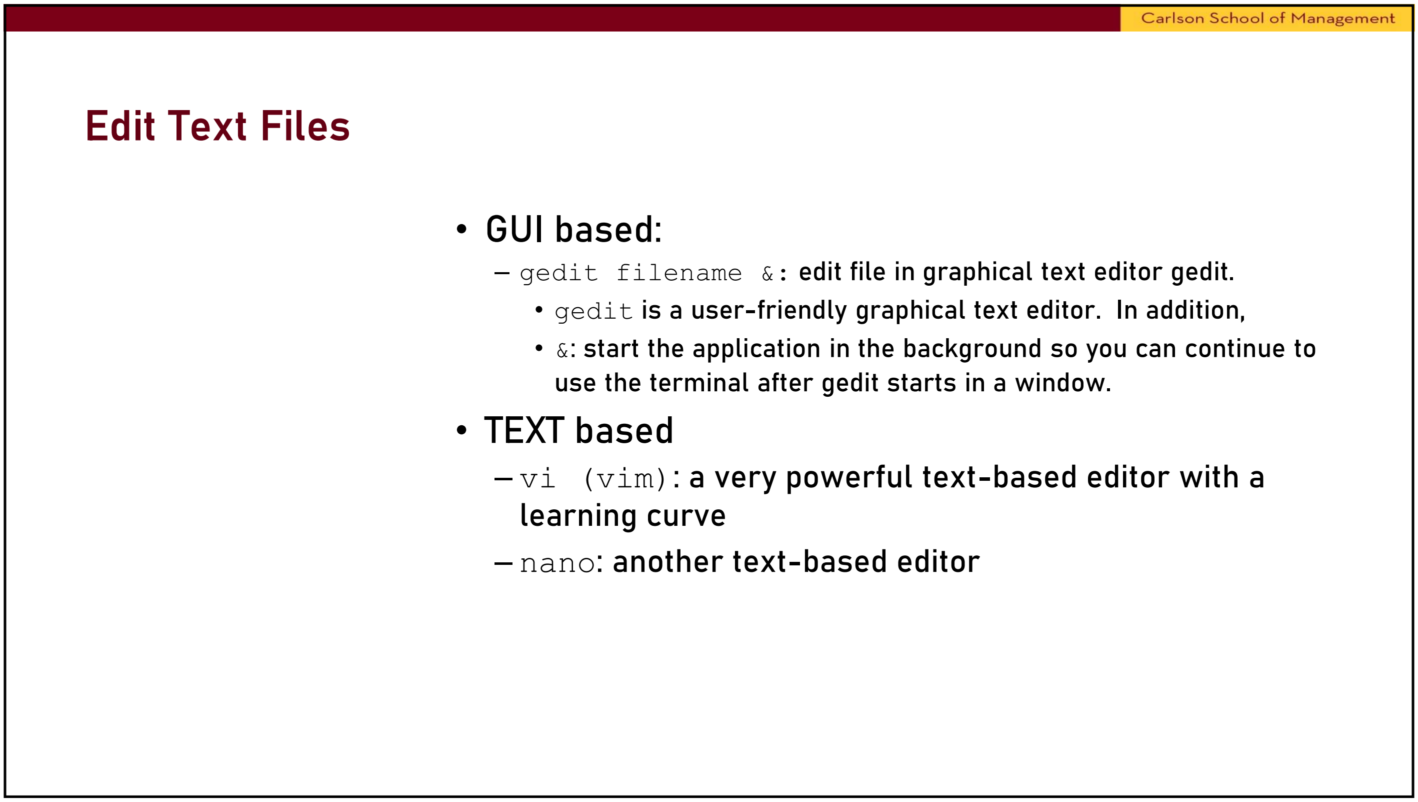
  `sort` : dictionary sort

  `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.



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