

# Big Data - Linux Basics

Dr. Qing “Matt” Zhang  
ITU



# What's Linux

- A Version of Unix

- Developed in 1969 at AT&T Bell Labs

- Open source

- Source code available

- Kernel for free

- Central component

- Can be customized for user's needs



# Linux OS

- Multi-user
  - Allow multiple users to use it at the same time
- Multi-processing
  - Utilize multiple processors
- Multi-tasking
  - Allow multiple software to run at the same time
- Multi-threading
  - Different parts of a software can run concurrently

# Distributions

- Slackware(one of oldest, simple, stable)
- Redhat
  - Fedora (free, fast release)
  - RHEL (Redhat commercial support, not free)
  - CentOS (based off commercial releases)
- SuSe (based in German)
- Debian (GNU/Linux)
- Ubuntu
- ...

# Shell

- After login, Linux starts a program “shell”
  - Interprets user commands
  - Manages execution
  - Communicates with kernel
    - bash, csh, korn, tcsh, ...
- Shell commands are case sensitive!

# Basic commands(file handling)

- ls: list directory contents

- ls [option ...] [file ...]

- ls -l

- ls -a

- ls -lt

- cd: change directory

- cd [dir]

- pwd: print current working directory

# Basic commands (cont'd)

- **mkdir: create directory**
  - `mkdir [option] [directory]`
- **cp: copy files/directories**
  - `cp [option] src dest`
  - `cp -r src_dir dest_dir`
- **mv: rename file/directory**
  - `mv [option] src dest`

# Basic commands (cont'd)

- **rm: remove file/directory**

- Linux doesn't forgive, be REALLY CAUTIOUS!

- `rm -r [directory]`

- `rmdir`

- **find: search for a file**

- `find [option] [path] [patthern]`

- `find . -name '*hadoop*'`



# Basic commands (text processing)

- more/less: display screenful text
  - less allows backward movement
  - return scrolls forward one line, space one page
  - y scrolls back one line, b one page
  - use “/” to search for a string
  - q to quit

# Basic commands (cont'd)

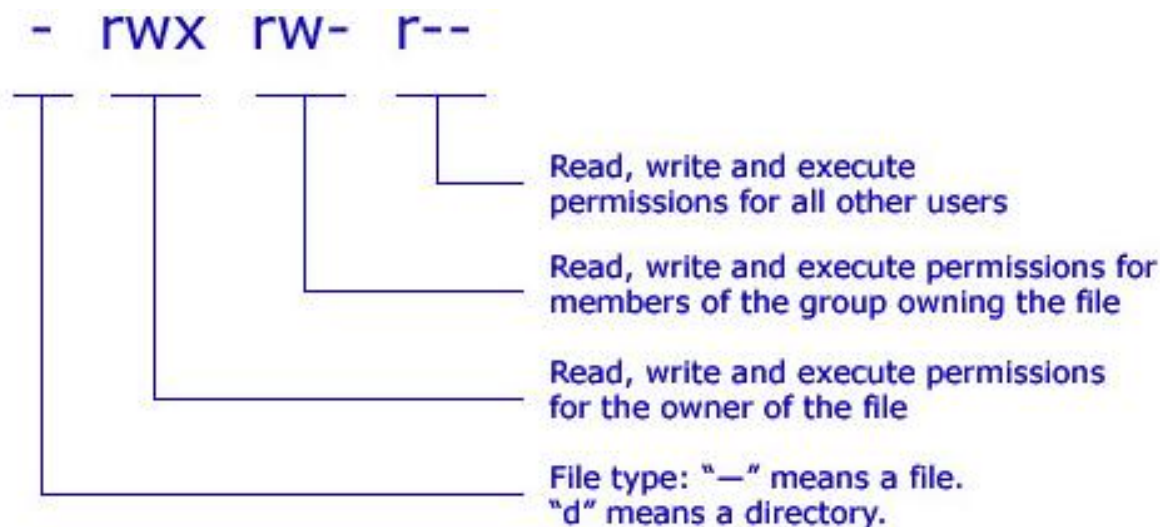
- head/tail: display first/last lines
  - default 10, -n to override
- cat: concatenate files and print
- echo: display one line
- wc: count words/lines/bytes ... in files
- grep: search for pattern in files
- diff: compare files
- vi: text editor
- emacs: text editor

# Input/Output Redirection (“piping”)

- Programs can output to other programs
- Called “piping”
- “program\_a | program\_b”
  - program\_a’s output becomes program\_b’s input
- “program\_a > file.txt”
  - program\_a’s output is written to a file called “file.txt”
- “program\_a < input.txt”
  - program\_a gets its input from a file called “input.txt”
- `cat infile1 | wc -l > outfile1`

# Permission levels

- “r” means “read only” permission
- “w” means “write” permission
- “x” means “execute” permission
  - In case of directory, “x” grants permission to list directory contents



# Basic commands (sys admin)

- chmod: change file access permissions
  - chmod [**u**ser/**g**roup/**o**thers/**a**ll]+[permission]  
[file(s)]
- chown: change file owner and group
- su: change user
- passwd: change user password
- who: show who's logged on
  - Whoami

# HELP!

- man: help info on a command
  - man [option] [command]
  - man man
- history: print recently used commands
- which: search for a command location in system path

