Informatics on High-throughput Sequencing Data

(Summer Course 2020)

Day 6

Agenda

- Unix-based systems.
- Why Linux!
- Let's start!
- Linux Commands for:
  - Files & Directories.
  - System.
  - Process Management.
  - Networking.
  - Compression.
  - Searching.
- Piping output.
- Wildcard character.
- Redirecting output.
- Stream Editor (Sed).
- Linux tools for text files processing.
- Shell Scripting

Piping output

- Pipes are represented by the | character.
- It is possible to send the output of one program to another program as input.

```
history |less
```

List all remembered commands page by page.

```
history |grep Bio
```

List all remembered commands containing string “Bio”.

Redirecting output

```
ls -1 > myfile.txt
```

List items in current directory and show in long format to see permissions, size, and modification date. The output is directed to display in myfile.txt.

```
ls -1 >> myfile.txt
```

More Unix commands

```
wc -l

cut -f1,2,3,4 athal.genes.gtf > sara.gtf

Cut -d ',' -f1 GDC.CSV

sort sara.gtf

sort -r sara.gtf

sort -k4 sara.gtf

sort -k4 numbers.txt

sort -k 4n numbers.txt
```

More Unix commands

```
sort -k 4nr numbers.txt

sort -u numbers3.txt

sort -k2,3 numbers1.txt

sort -t ',' -k2r GDC.CSV

sort -n -k2,2 -k6,6 numbers1.txt

uniq numbers3.txt

uniq -c numbers3.txt
```

More Unix commands

```
diff kmers1.txt kmers2.txt

sort kmers1.txt > S1.txt
sort kmers2.txt > S2.txt
comm S1.txt S2.txt

comm -1 -2 S1.txt S2.txt | wc -l
```

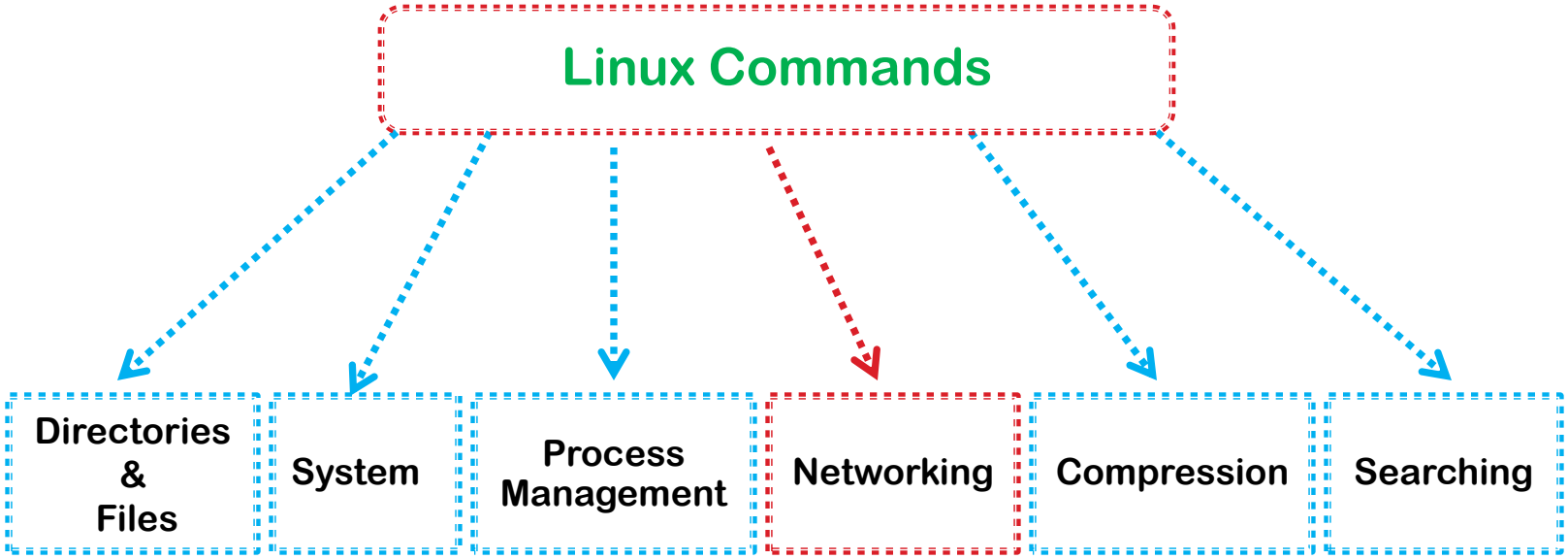
Display text in terminal

```
echo ‘Hello World’

echo -e ‘Hello \n World’

echo $((1+2))
```

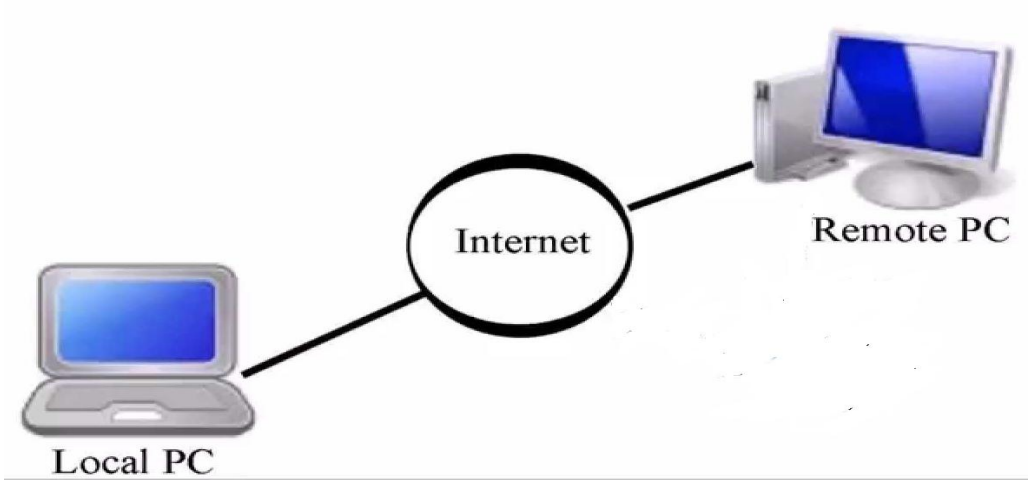
Editing files

- vi or vim program.
- press i for insert mode.
- press Esc to leave insert mode.
- type :wq and press Enter to save changes and quit.
- type :q! and press Enter to quit without saving.

Getting Started !!



Networking (Connect to a Remote Machine)

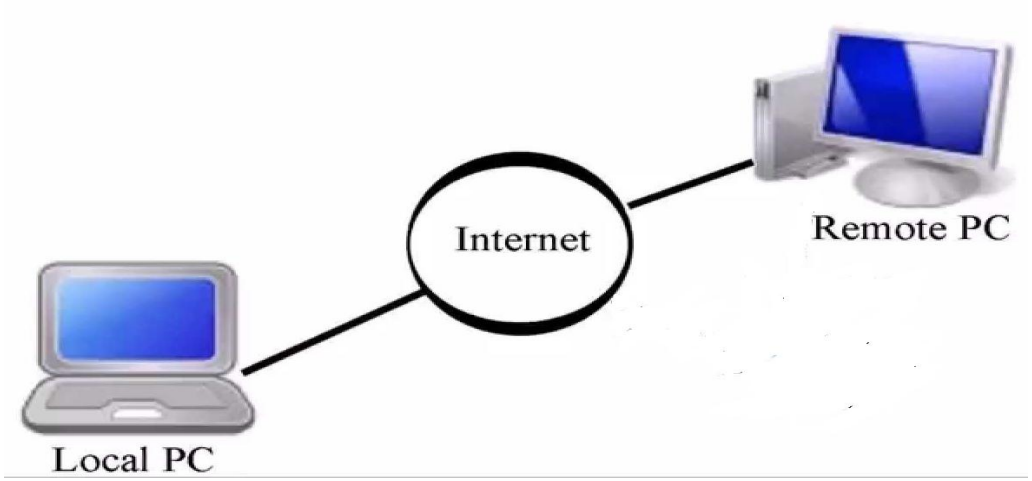



```
ssh username@ageri.sci.eg
```

"Securely connect me to a shell on a remote machine."

https://phoenixnap.com/kb/ssh-to-connect-to-remote-server-linux-or-windows

Networking (logging out of a remote machine)



`exit` or `Ctrl-d`

“close current terminal window."

Networking (copy files (to | from) a remote machine)

```
scp username@ageri.sci.eg:file Bio
```

Secure copy a file from remote server to the Bio directory on your machine.

```
scp file username@ageri.sci.eg:Bio
```

secure copy a file from your machine to the Bio directory on a remote server.

Networking (copy files (to | from) a remote machine)

```
scp -r username@ageri.sci.eg:Bio Bio
```

Secure copy the directory Bio from remote server to the directory Bio on your machine.

```
scp username@ageri.sci.eg:~/myfile.txt ./
```

Your home directory on the remote system.

Current directory in your computer.

Networking

```
wget file_link
```

Download a file.

```
curl file_link
```

Download a file.

```
sudo apt-get install mailutils
```

Download and install packages.

Networking

sudo apt-get install mailutils

```
toshiba@ubuntu:~/Documents/agri_training$ mail
The program 'mail' is currently not installed.  You can install it by typing:
sudo apt-get install mailutils
toshiba@ubuntu:~/Documents/agri_training$ sudo apt-get install mailutils
[sudo] password for toshiba:
Reading package lists... Done
Building dependency tree
```

Shell Scripting

- **SHELL SCRIPTING is writing a series of commands for the shell to execute.**
- **It can combine lengthy and repetitive sequences of commands into a single and simple script, which can be stored and executed anytime.**
- **This reduces the effort required by the end user.**

https://www.guru99.com/introduction-to-shell-scripting.html

References

- Paul Stothard, An Introduction to Linux for bioinformatics , 2016.
- Robert Bukowski, Linux for Biologists- Part 1.
- Steve Pederson, Introduction To Linux/Ubuntu & Sell Scripting, 2014.
- https://bioinformatics.uconn.edu/unix-basics/#
- https://learn.gencore.bio.nyu.edu/ngs-file-formats/quality-scores/
- https://coding4medicine.com/Members/pages/home/
- https://open.oregonstate.education/computationalbiology/chapter/patterns-regular-expressions/
- https://bioinformaticsworkbook.org/Appendix/Unix/unix-basics-3grep.html#gsc.tab=0
- https://datacarpentry.org/shell-genomics/04-redirection/

Thanks!

// | ?