

# Research Data Workshop Series - Introduction to Linux

Lucas Alcantara, Ph.D.  
lucas.ma@uoguelph.ca



**AGRI-FOOD DATA  
CANADA**

AT THE UNIVERSITY *of* GUELPH

# Welcome to Linux

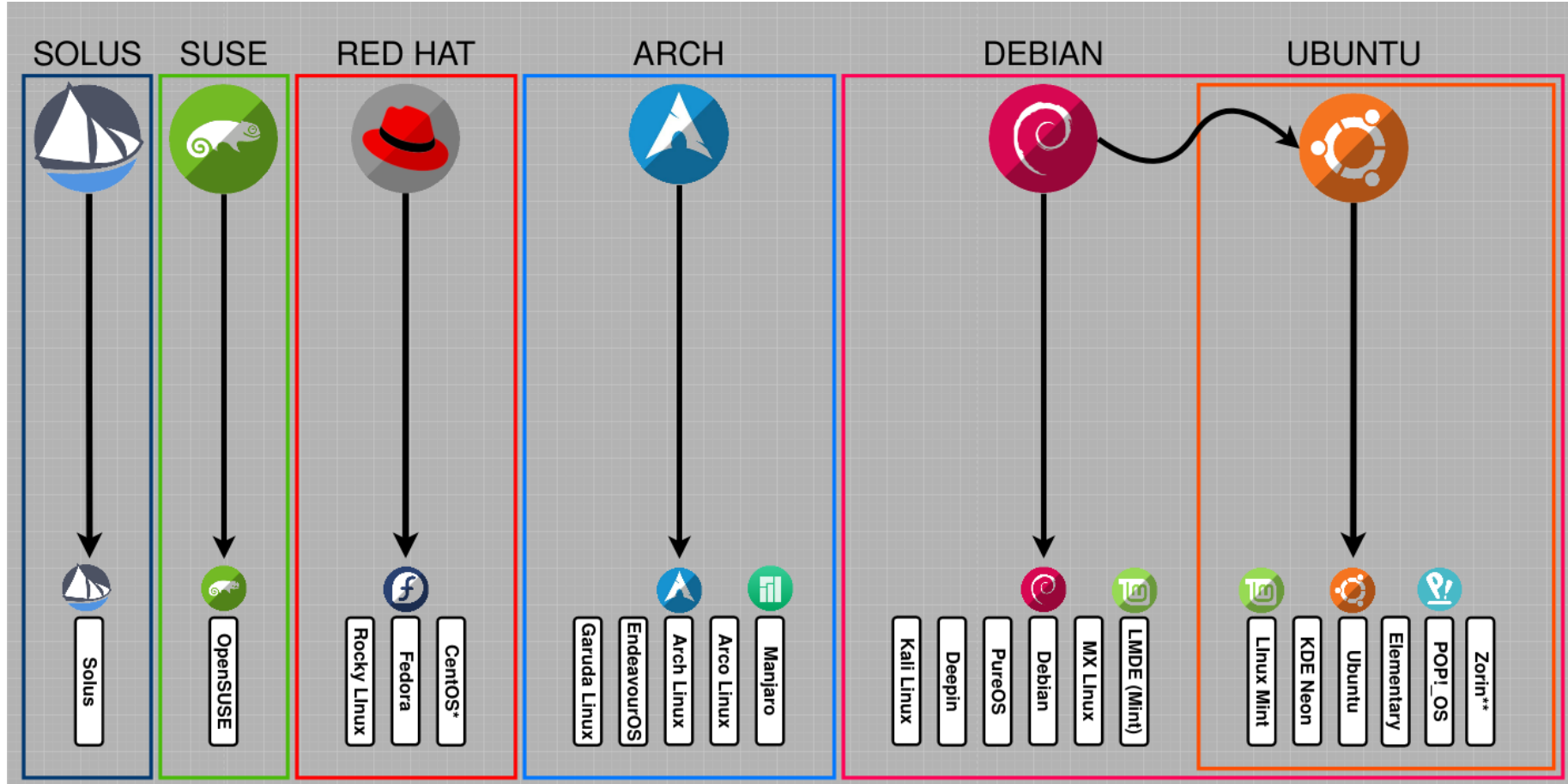
# 2023 Statistics

- 47% of professional developers use Linux-based operating systems. (Statista)
- Linux powers 39.2% of websites whose operating system is known. (W3Techs)
- Linux powers 85% of smartphones. (Hayden James)
- Linux, the third most popular desktop OS, has a market share of 2.09%. (Statista)
- The Linux market size worldwide will reach \$15.64 billion by 2027. (Fortune Business Insights)
- The world's top 500 fastest supercomputers all run on Linux. (Blackdown)
- 96.3% of the top one million web servers are running Linux. (ZDNet)
- Today, there are over 600 active Linux distros. (Tecmint)

<https://truelist.co/blog/linux-statistics>



# Popular Linux distributions



<https://i.redd.it/plgpxupr5mu61.png>

# Many ways to connect with Linux



<https://ubuntu.com/tutorials/command-line-for-beginners>

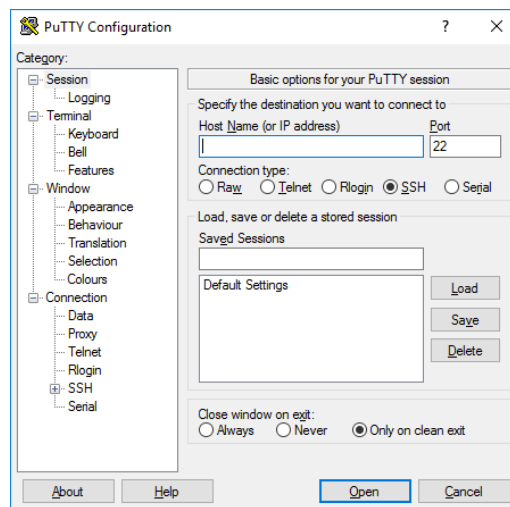
```
Terminal — top — Basic — 80x24
Processes: 421 total, 3 running, 418 sleeping, 1383 threads      16:01:01
Load Avg: 1.49, 1.20, 1.14  CPU usage: 0.48% user, 0.84% sys, 98.67% idle
SharedLibs:
MemRegions:
PhysMem: 1
VM: 2461G
Networks:
Disks: 246

Terminal — top — Red Sands — 80x24
Processes: 421 total, 3 running, 418 sleeping, 1383 threads      16:01:01
Load Avg: 1.49, 1.20, 1.14  CPU usage: 0.48% user, 0.84% sys, 98.67% idle
SharedLibs: 515M resident, 86M data, 305M linkedit.
MemRegions: 98214 total, 2353M resident, 242M private, 1558M shared.
PhysMem: 14G used (2687M wired), 18G unused.
VM: 2461G vsize, 2277M framework vsize, 0(0) swaptins, 0(0) swaptouts.
Networks: packets: 168275/66M in, 191369/56M out.
Disks: 246661/3854M read, 146809/3327M written.

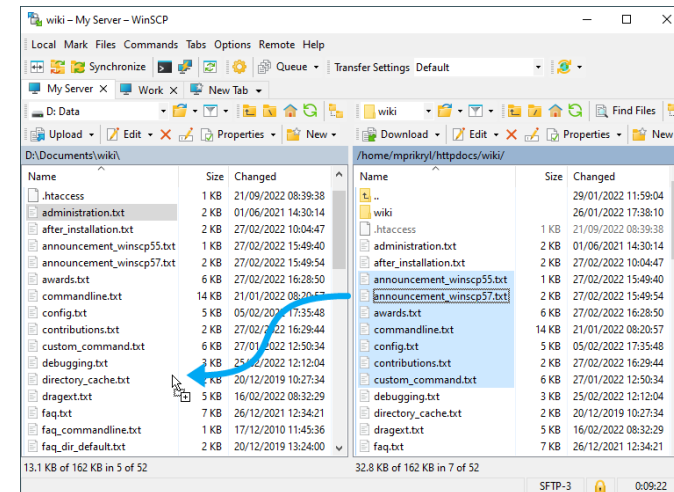
Terminal — top — Pro — 94x24
Processes: 421 total, 2 running, 419 sleeping, 1383 threads      16:01:01
Load Avg: 1.49, 1.20, 1.14  CPU usage: 0.54% user, 0.96% sys, 98.49% idle
SharedLibs: 515M resident, 86M data, 305M linkedit.
MemRegions: 98214 total, 2353M resident, 242M private, 1558M shared.
PhysMem: 14G used (2687M wired), 18G unused.
VM: 2461G vsize, 2277M framework vsize, 0(0) swaptins, 0(0) swaptouts.
Networks: packets: 168275/66M in, 191369/56M out.
Disks: 246661/3854M read, 146809/3327M written.

PID  COMM  STAT  CPU  TIME  #TH  #INQ  #PORT  MEM  PURG  CMPR  PGRP  PPID  STATE
147  Wind  S    0    0:00  1    0    0    0    0    0    0    0    sleeping
3724  top   S    3.2  00:17  37   1    0    25  3864K  0B   0B   3725  3692  sleeping
3725  top   S    3.1  00:17  41   1    0    23  3824K  0B   0B   3724  3706  sleeping
3723  top   S    3.0  00:17  39   1/1  0    35  4888K  0B   0B   3723  3700  running
2947  Term  S    1.7  00:28  08   6    1    402  106M   6916K  0B   2947  1    sleeping
0     kern  S    0    0:00  0    0    0    0    0    0    0    0    0    running
329  Touch  R    1.7  03:48  16  270/16  0    0    55M   0B   0B   0    0    0    running
2734  Safa  S    0.9  00:37  87   3    2    77  1224K  0B   0B   3123  1    sleeping
2881  apps  S    0.1  00:07  37   3    2    137  2080K  0B   0B   81    1    sleeping
329  Touch  S    0.1  01:11  82   4    1    325  23M   3200K  0B   329  1    sleeping
2734  SafariBookma  S    0.0  00:05  08   5    3    72  4788K  12K  0B   2734  1    sleeping
2881  appstoreagen  S    0.0  00:01  49   4    2    123  6236K  208K  0B   2881  1    sleeping
141  AirPlayPCHe  S    0.0  00:03  40   6    2    172  2696K  0B   0B   141  1    sleeping
238  nsurlsession  S    0.0  00:03  44   6    3    101+ 3000K+ 0B   0B   238  1    sleeping
```

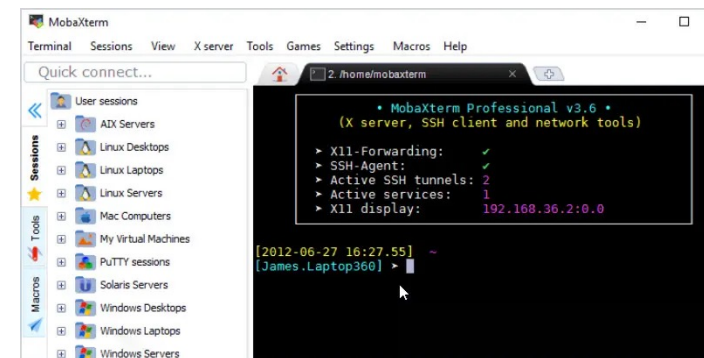
<https://support.apple.com/en-ca/guide/terminal/welcome/mac>



[https://www.ssh.com/hubfs/Imported\\_Blog\\_Media/PuTTY\\_Windows\\_configuration\\_and\\_connection\\_screen\\_with\\_profile\\_save\\_option-2.png](https://www.ssh.com/hubfs/Imported_Blog_Media/PuTTY_Windows_configuration_and_connection_screen_with_profile_save_option-2.png)



[https://winscp.net/eng/data/media/screenshots/download\\_dragdrop.png](https://winscp.net/eng/data/media/screenshots/download_dragdrop.png)



<https://networkverge.com/wp-content/uploads/2022/06/mobaxterm-screenshot.webp>



# Useful commands

# Remote connection

## ssh

- A program intended to provide secure encrypted communications between two untrusted hosts over an insecure network. It is used for:
  - logging into a remote machine
    - `ssh username@hostname`
  - executing commands on a remote machine
    - `ssh username@hostname command arg1 arg2`

# Whereabouts

## pwd

- Writes the absolute pathname of the current working directory to the standard output
  - pwd
  - /home/username

## whoami

- Displays effective user id
  - whoami
  - username



# Whereabouts

## id

- Returns user identity
  - id
  - uid=12345(username) gid=12345(username) groups=12345(username), 123456(extragroup)

## ls

- Lists directory contents
  - ls
  - file1 file2 file3 dir1 dir2
- Useful arguments:
  - -a (show hidden files/directories)
  - -l (show output in long format)
  - -h (show sizes in human readable format when -l is used)



# Getting around

## mkdir

- Make directories
  - mkdir data
  - returns no output
- Useful arguments:
  - -p (creates intermediate directories as required)
  - -m (sets the permissions of the final created directory)

## cd

- Change the working directory
  - cd data
  - cd ../
  - cd /root



# Getting around

## chmod

- Change file modes (permissions):
- Octal numbers (<https://chmod-calculator.com>)
  - chmod 777 data
- who and perm symbols
  - chmod ugo+rwx data (give full access to all users)
  - chmod go-rwx data (remove full access from all users, except owner)

## chgrp

- Change group owner
  - chgrp sample\_group data

# Create, edit and view files

## touch

- Change file/folder access and modification times
  - touch data
- If file/folder does not exist, create new one
  - touch file.txt

## nano / vim

- Simple command-line text editors
  - nano file.txt
  - vim file.txt

# Create, edit and view files

## cat

- Concatenate and print files
  - `cat file.txt`
  - `cat file1 file2 file3`
- Useful arguments
  - `-n` (Number the output lines, starting at 1)
  - `-b` (Number the non-blank output lines, starting at 1)

# Create, edit and view files

## cp

- Copy file(s)/directory(ies)
  - Same name: cp file.txt data
  - Different name: cp file.txt data/new\_name.txt
- Useful arguments
  - -r (Recursive copy)

## mv

- Move/Rename file(s) )/directory(ies)
  - Move: mv file.txt data
  - Rename: mv file.txt new\_name.txt

# Create, edit and view files

## rm

- Remove file(s)
  - `rm data/new_name.txt`
- Useful arguments
  - `-i` (Prompt confirmation)
  - `-d` (Remove directories)
  - `-f` (Force remove, overrides `-i`)

## ln

- Link file(s)/directory(ies)
  - Symbolic link (aka shortcut): `ln -s /full/path/to/file.txt sym_link2file.txt`

# Create, edit and view files

## head

- Display first lines of a file
  - `head file.txt`
  - `head -n 1 file.txt`
- Useful arguments
  - `-n` (Number of the first lines to print of the specified file, default is 10)



# Create, edit and view files

## tail

- Display last lines of a file
  - tail file.txt
  - tail -n 1 file.txt
- Useful arguments
  - -n (Number of the last lines to print of the specified file, default is 10)
  - -f (Wait for additional data to be appended to the file)

# Create, edit and view files

## Data streams

- Standard Input (stdin)
  - Input given to a command (file, text, stdout from another command)
- Standard Output (stdout)
  - Output produced by a command (text on your screen)
  - “cat file.txt” prints file content to stdout
- Standard Error (stderr)
  - Error messages produced by a command (also text on your screen)
  - “ls /root” prints error message to stderr

# Create, edit and view files

## Data streams

- Redirecting stdout to file
  - `cat file.txt > newfile.txt`
  - `cat file1 file2 file3 > newfile.txt`
  - `cat file1 > appended_file.txt`
  - `cat file2 >> appended_file.txt`
  - `cat file3 >> appended_file.txt`
- Redirecting stdout to stdin
  - `ls -l | head`

# Create, edit and view files

## less

- Allows backward and forward movement in the file
- Does not have to read the entire input file before starting
  - `less file.txt`
- Useful arguments
  - `-S` (Causes lines longer than the screen width to be truncated rather than wrapped)
  - `-N` (Display line numbers at the beginning of each line)
- Useful features
  - `/pattern` (Search for pattern in file – keep pressing `/` + Enter to move to next pattern instance)
  - Enter (Moves to next line – PageUp and PageDown should also work similarly)
  - `q` (Escape less command)

# Search

## find

- Recursively descends the directory tree listed in search of a pattern
  - `find / -name file.txt`
- Useful arguments
  - `-name` (search by name)
  - `-type f` (pattern is a file)
  - `-type -d` (pattern is a directory)

## grep

- Search any given input files, selecting lines that match one or more patterns.
- By default, a pattern matches an input line if the regular expression (RE) in the pattern matches the input line without its trailing newline.
- Each input line that matches at least one of the patterns is written to the standard output
  - `grep pattern file.txt`
- Useful arguments
  - `-v` (invert match)
  - `-n` (add line numbers to stdout)
  - `-c` (count of patterns in stdin)
  - `-I` (ignore case)
  - `-r` (recursive search)

## Useful commands

- `grep "^start" file.txt`
- `grep "end$" file.txt`
- `grep "t[wo]o" file.txt`
- `cat ls -l / | grep home`

**THANK YOU!**

**Questions?**

Feel free to reach me later at  
[lucas.ma@uoguelph.ca](mailto:lucas.ma@uoguelph.ca)



**AGRI-FOOD DATA  
CANADA**

AT THE UNIVERSITY *of* GUELPH