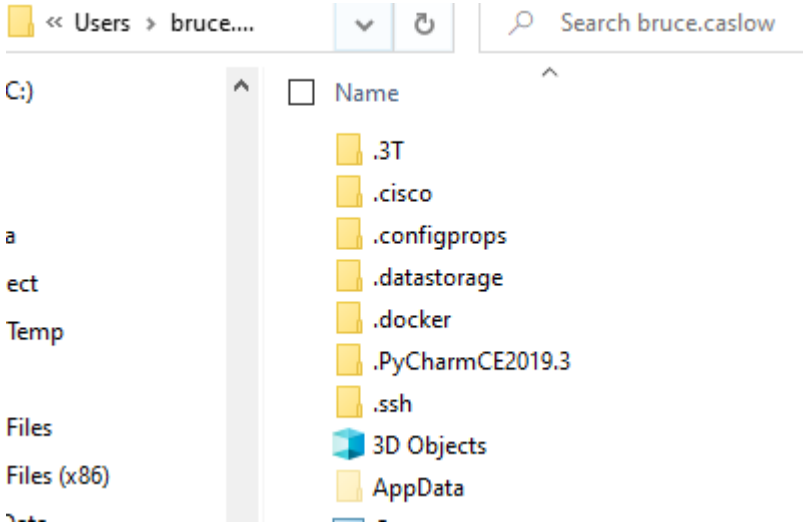
Linux Tools for Data Scientists Part One

Mastering the Basic File Management Commands of LINUX

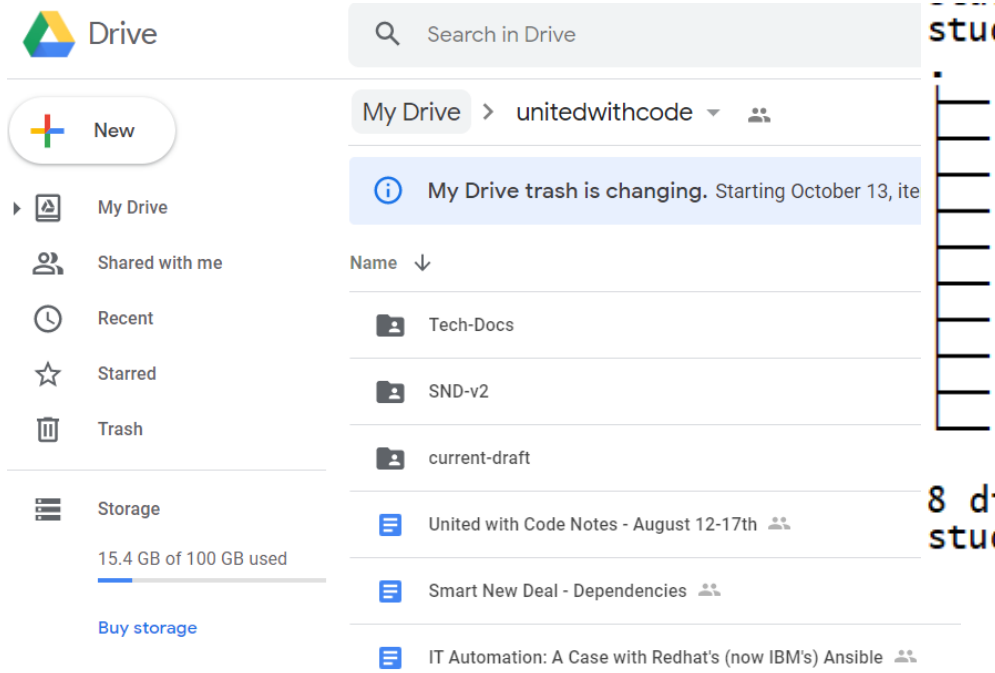
Lesson Objectives

- All application software is installed on computers as a set of files.
- All files on a computer are organized hierarchically in directories or folders.
- All Data Science activities will be related to files, directories and file and directory management.
- All Infrastructure as Code activities will be related to files, directories and file and directory management.

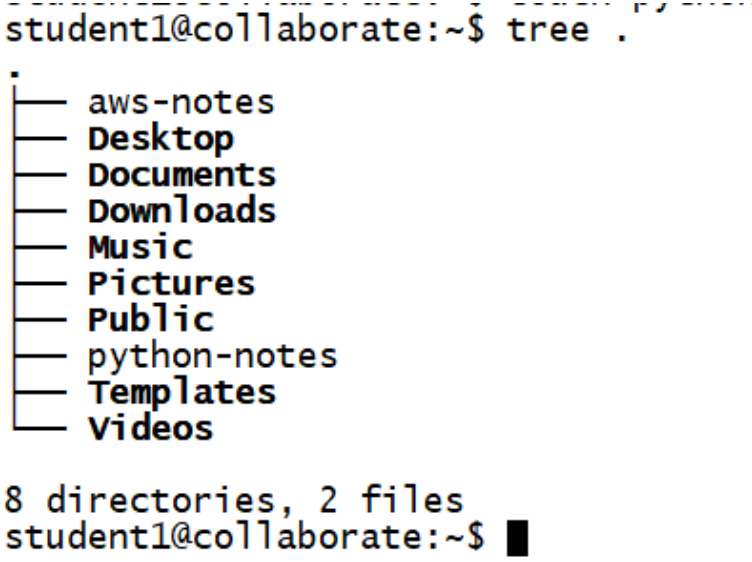
Introduction to Operating System File Systems
Managing Files and Folders/Directories



Microsoft File Explorer:
Folders/Directories
And Files

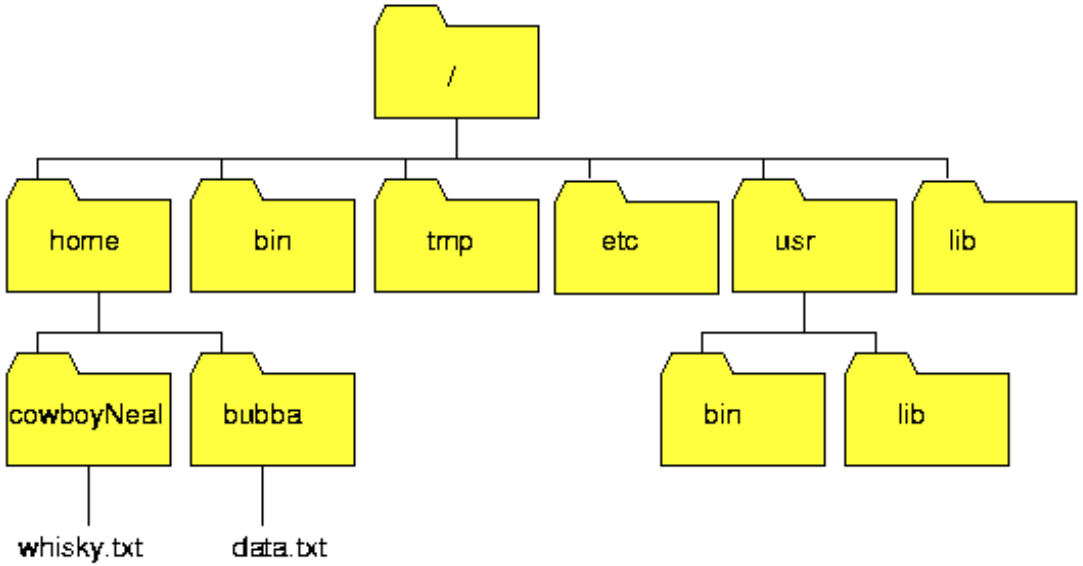


Google Drive: Folders/Directories
and Files



LINUX folders/directories
and files.

Standard Operating System File and Directory Structures Are Hierarchical



Every file saved in a hierarchical file structure has a full path name. Example:

/home/bubba/data.txt

This is an important term to know: “path”

Primary UNIX/LINIX Top Level Directories

- /bin – binary files for LINUX
- /sbin – sysadmin binary files usually used by sysadmins only
- /usr – common /usr subdirs: /usr/bin, /usr/sbin, /usr/lib, /usr/local
- /var – common /var subdirs: /var/log, /var/www
- /dev – device drivers
- /etc – configuration files
- /home – home directories

Basic LINUX File System/File Management Commands

The following LINUX commands allows you to browse your and directory names and associated attributes. Start with the two most basic commands: pwd and ls

- **student@ubuntu011:~$pwd #present working directory**
- **student@ubuntu011:~$ls #list files and directories**
- **student@ubuntu011:~$ls | more #list files & pauses screen output**
- **student@ubuntu011:~$ ls m***
- **student@ubuntu011:~$ ls –l**
- **student@ubuntu011:~$ls –lt**
- **student@ubuntu011:~$ls –lt | less # like "| more" pauses screen output**
- **student@ubuntu011:~$ls –lt | wc -l**

BASH Command-Line Key Sequences

Here are some useful BASH command line key sequences

- **CTRL+A** Beginning of line
- **CTRL+E** End of line
- **CTRL+D** Delete Character at the Cursor
- **CTRL+K** Cut the from the cursor position to the end of the line
- **CTRL+U** Cut the from the cursor position to the beginning of the line.
- **CTRL+Y** YANK or PASTE back what you cut out with CTRL+K or CTRL+U
- **CTRL+R** Reverse Search…. Very powerful!!
- **(Note: CTRL+R works in the Python shell.)**
- **CTRL+R, CTRL+R** Repeat the reverse search
- **CTRL+Z** Suspend a program and return to the BASH shell

Moving Up and Down the LINUX File System Tree with the "cd" command

The following LINUX commands allows you to browse your file names and move up and down in your file directories.

- **student@ubuntu011:~$pwd -present working directory**
- **student@ubuntu011:~$cd / -go to the root directory**

- **student@ubuntu011:~$cd .. -moves you up one level**
- **student@ubuntu011:~$cd ../.. -moves you up two levels**

- **student@ubuntu011:~$cd ~ -go to your home directory**
- **student@ubuntu011:~$cd - a second method to get to your home dir**

- **student@ubuntu011:~$history -gives you a list of all of the commands you have entered in a given session**

Exploring Your File and Directory Structure with the Tree Utility

- **student@ubuntu011:~$tree . -content of directories**
- **student@ubuntu011:~$tree . | more**
- **abcaslow@ideapad-310:~$ tree -F -L 2**
- https://www.tecmint.com/linux-tree-command-examples/

Exploring the File Location and the full path of LINUX files with the “which” command

- abcaslow@ideapad-310:~$ **which ls**

/bin/ls

- abcaslow@ideapad-310:~$ **which python**

/usr/bin/python

- abcaslow@ideapad-310:~$ **which ifconfig**

/sbin/ifconfig

- which cat
- which dog

Understanding Path Settings in Windows & LINUX

C:\Users\bruce.caslow>path (WINDOWS)

PATH=C:\Program Files\VanDyke Software\Clients\;C:\Program Files (x86)\Intel\Intel(R) Management Engine Components\iCLS\;C:\Program Files\Intel\Intel(R) Management Engine Components\iCLS\;C:\Program Files (x86)\Common Files\Oracle\Java\javapath;C:\ProgramData\Oracle\Java\javapath;C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wbem;C:\WINDOWS\System32\WindowsPowerShell\v1.0\;C:\WINDOWS\System32\OpenSSH\;C:\Program Files (x86)\Intel\Intel(R) Management Engine Components\DAL;C:\Program Files\Intel\Intel(R) Management Engine Components\DAL;C:\Program Files\PuTTY\;C:\Program Files\Git\cmd;C:\Users\bruce.caslow\AppData\Local\Microsoft\WindowsApps;

abcaslow@ideapad-310:~$ echo $PATH (LINUX)

/home/abcaslow/.local/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin:/home/abcaslow

Understanding the LINUX PATH Variable

- echo $PATH
- pwd
- ls –l *.py
- which findarp.py

- echo $PATH
- PATH=$PATH:/home/abcaslow
- echo $PATH
- which findarp.py

Looking Inside of TextFiles with “cat”

- **student@ubuntu011:~$cat file-name**
- **student@ubuntu011:~$cat –n filename**
- **student@ubuntu011:~$cat –n filename | more**

Copy, Moving and Deleting Files and Directories

- **DO NOT RUN THE FOLLOWING COMMANDS**
- **student@ubuntu011:~$cp src-file dest-file**
- **student@ubuntu011:~$cp src-file ~/dest-file**
- **student@ubuntu011:~$mv src-file dest-file**
- **student@ubuntu011:~$rm filename (DANGEROUS)**
- **student@ubuntu011:~$echo rm filename (TEST BEFORE ACT)**
- **student@ubuntu011:~$echo rm * (TEST BEFORE ACT)**
- **student@ubuntu011:~$rm * (VERY VERY DANGEROUS!!!)**
- **student@ubuntu011:~$mkdir**
- **student@ubuntu011:~$rmdir**

LINUX File Permissions and “chmod”

- LINUX has three file/directory permission classifications: user, group and world
- LINUX has three file/directory permission types: read (r), write (w) and executable (x).

```
abcaslow@ideapad-310:~$ ls -l find*
-rwxr-xr-x 1 abcaslow abcaslow  382 Jun  5 17:56 findarp_2func.py
-rwxr-xr-x 1 abcaslow abcaslow 1309 May  6 10:23 findarpall.py
-rwxr-xr-x 1 abcaslow abcaslow 1160 Jun 19 12:46 findarp_cc.py
-rwxr-xr-x 1 abcaslow abcaslow 1320 May 14 13:39 findarp_device.py
-rwxr-xr-x 1 abcaslow abcaslow 1332 Jun 18 22:49 findarphistory_chris.py
-rwxr-xr-x 1 abcaslow abcaslow 2012 Jun 18 22:34 findarphistory.py
-rwxr-xr-x 1 abcaslow abcaslow 1071 Jun 17 19:53 findarpintf.py
-rwxr-xr-x 1 abcaslow abcaslow  664 Jun  4 22:12 findarp_ppt.py
```

The Linux ‘grep’ Command

Here is an example of locating strings inside of files with the Linux ‘grep’ command.

- **Grep for files within the pwd and its subdirectories.**
- grep -R "import json" . 2>/dev/null
- **Grep for a string within files in the present working directory (pwd):**
  grep 'import json’ *

Copy Files to and from a Remote System

Start simple: copy one file from a source Linux system to a target Linux system.

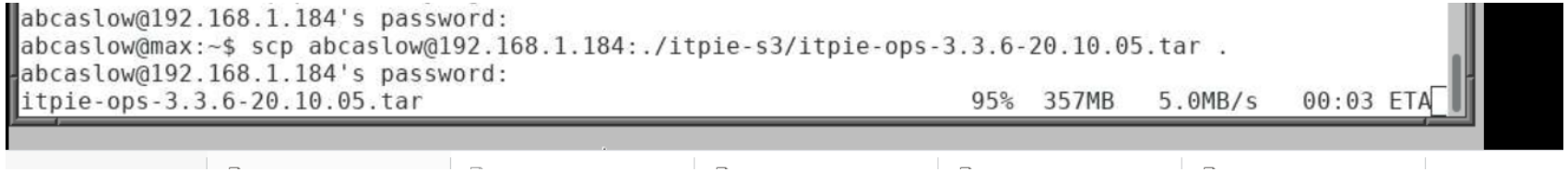
scp 'filename' student@10.1.1.1:. (NOTE: Do not forget the period "." at the end of this command.

Increase complexity: copy all files from a source Linux directory to a target Linux system.

- **scp * student@10.1.1.1:.**

Copy all files from a source Linux directory *and* its subdirectories to a target Linux system.

- **scp –r * student@10.1.1.1:.**



WARNING: *SCP CAN OVERWRITE FILES. CAREFULLY INSPECT THE SCP TARGET DIRECTORY!*

The Linux ‘find’ Command

Here is an example of locating files with the Linux ‘find’ command.

Search starting from the root directory with ‘find’.

find / -name *.yang

- **Using Linux Command-line redirects and pipes with the ‘find’ command.**
- find / -name *.yang 2>&1 | grep -v “Permission denied”
- find / -name *.yang 2>/dev/null
- find / -name *.yang 2>/dev/null | less
- find / -name *.yang 2>/dev/null | wc -l

- **Search starting from specific directory.**
- find /home -name ‘sandbox*’
- **More complex ‘find’ examples.**
- **(NOTE: order matters for entry of these ‘find’ options)**
- find . –type f –mmin -5
- find . –type f –not –name ‘*.qcow*’ –mmin -5

Introducing “git”

- You can use git locally or you can use git with a remote “github” server.

- Using “git” locally:
- Git init
- Git add <file-name>
- Git commit
- Git log
- Git log –p
- Git checkout <commit-string> to retrieve a previous version of a file.
- NOTE: Use “git stash” with “git checkout”

LINUX Package Management

```
# yum install tree       #RHEL/CentOS 7
# dnf install tree       #Fedora 22+ and /RHEL/CentOS 8
$ sudo apt install tree  #Ubuntu/Debian
# sudo zypper in tree    #openSUSE
```

- Git clone
- Git pull

LINUX Package Management

- Git clone
- Git pull
- Git status
- Git remote –v
- Git add <file-name>
- Git commit
- Git rm <file-name>

Using the Python Jobs Command

- **student@ubuntu011:~$ jobs**
- [1]- Stopped emacs nx.py
- [2]+ Stopped python -i nx.py
- **student@ubuntu011:~$** ***PRESS CTRL+R to perform a reverse CLI search!***
- **(reverse-i-search)`cat -n': cat -n nx.py**
- ***Once CTRL+R locates the command, it runs the command.***
- **student@ubuntu011:~$ cat -n nx.py**
- 1 #!/usr/bin/python
- 2 import requests
- 3 import json
- **student@ubuntu011:~$ history | grep cat #use the history command with grep**
- 241 **cat** nxosv9k-2
- 059 **cat** -n netconf-9k-interface-222.py

Summary

- Network engineers can move between a text editor, Python, and the Linux CLI on a single screen.
- There are some useful BASH command line key sequences.
- Locate files with the Linux ‘find’ command.
- Locate files with the Linux ‘grep’ command.
- Test the delete command ‘rm’ with ‘echo’ before executing the command.
- Use “echo” as a general testing tool for many Linux commands before executing them.
- Unlock the power of Linux by constructing and testing a single Linux command step-by-step before executing the final “bash” command.
- It is strongly recommended that you spend a lot of time learning more about Linux.
- You can develop your Linux skills on a Virtual Machine on your own laptop or on a Virtual Machine hosted by a public cloud provider.
- You can build on your basic Linux skills to create BASH shell scripts.

Building a World-wide Technical Education Community with New Tools
Let's Unite with Code: www.unitedwithcode.com