



Advanced Molecular Detection

Southeast Region Bioinformatics



Intro to Linux – Part 1

June 19, 2020

BPHL-SEbioinformatics@flhealth.gov

Outline



Linux



File structure



Logging into HiPerGator



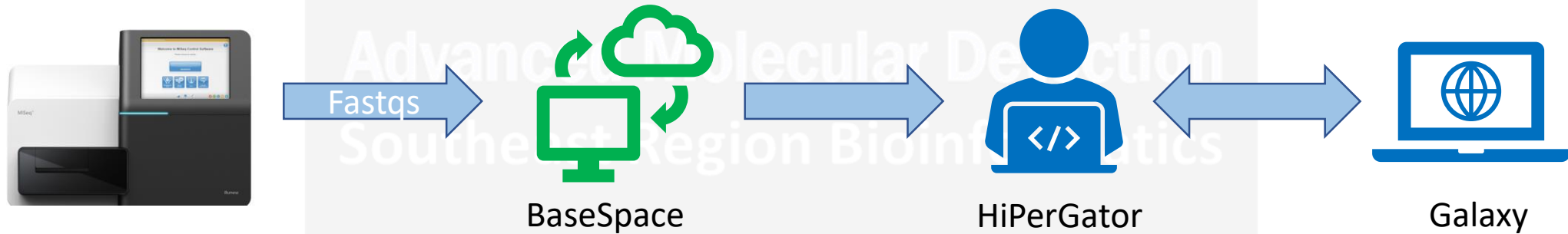
Navigating HiPerGator directories



Tip and tricks

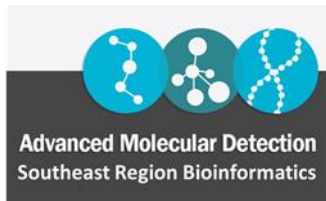
Why learn Linux?

- Most bioinformatics tools/pipelines run via command-line
 - Execute scripts in HiPerGator
- Download fastq files from NCBI's Sequence Read Archive (SRA)
- Transfer data directly from BaseSpace to Galaxy



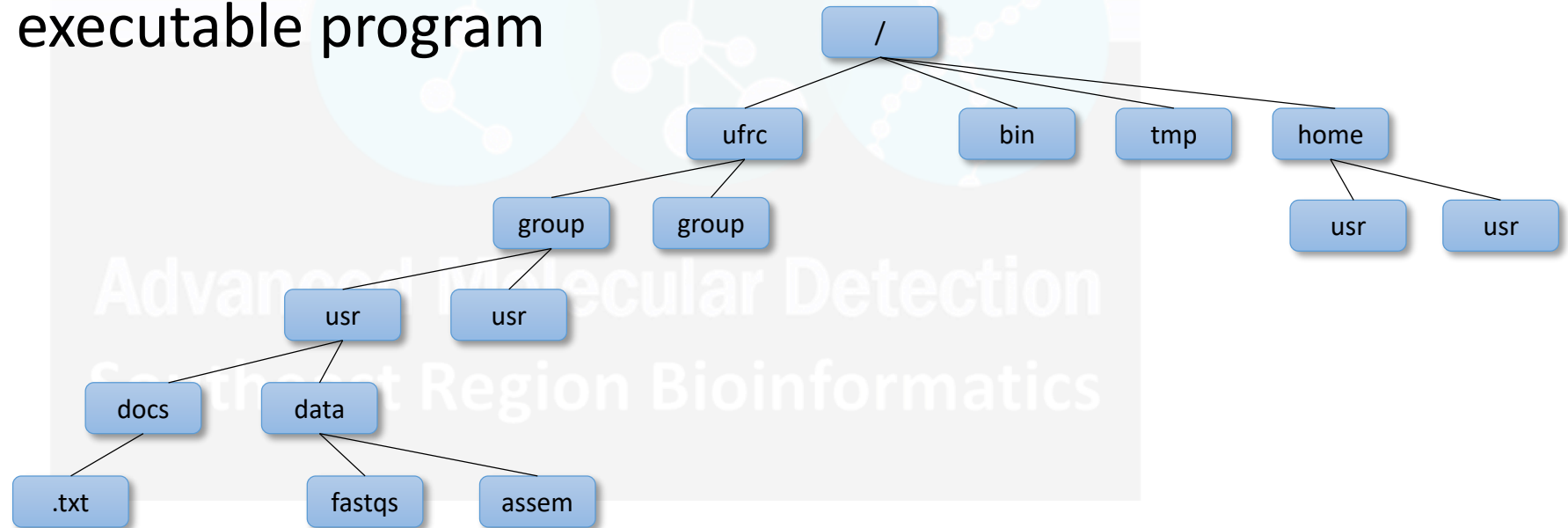
Linux

- Operating system
 - Unix-based
 - Macs are unix-based
- Why is it important in bioinformatics?
 - Open-source tools
 - Many servers or HPCs (high-performance computing clusters) are Linux-based
- Can work in the command-line
 - Terminal
 - Bash shell language



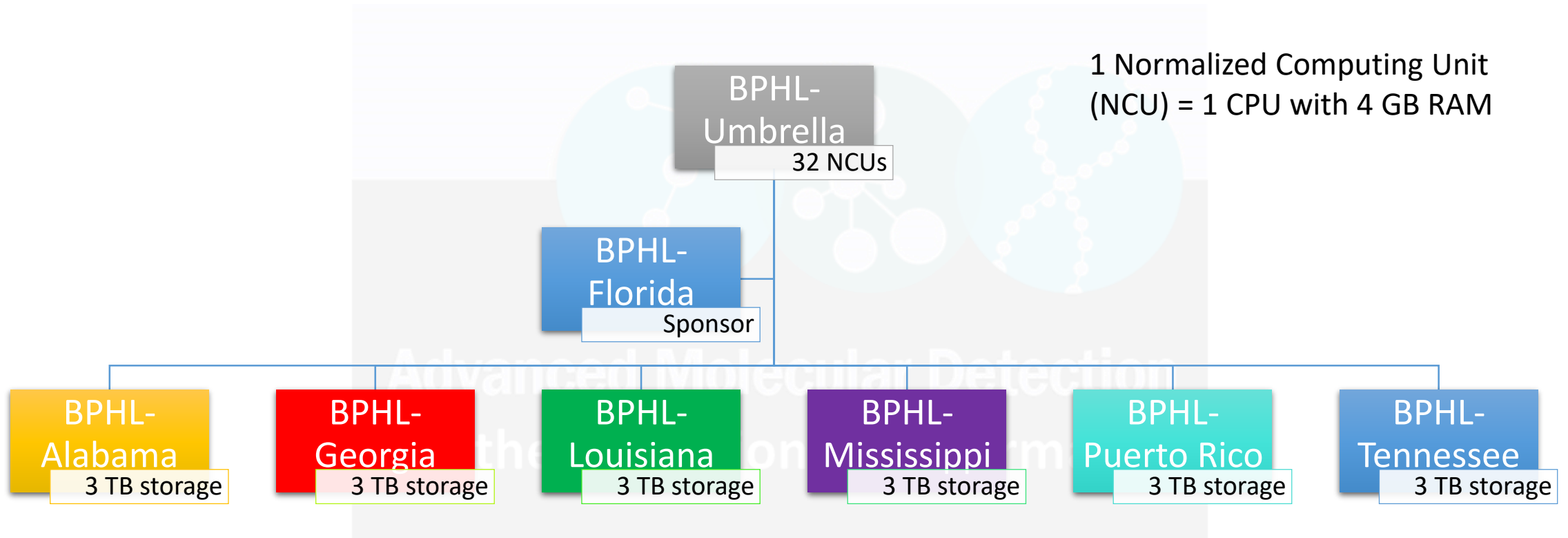
File Structure

- Everything in Linux is either a file or a process
- A directory is a “file”
- A process = executable program

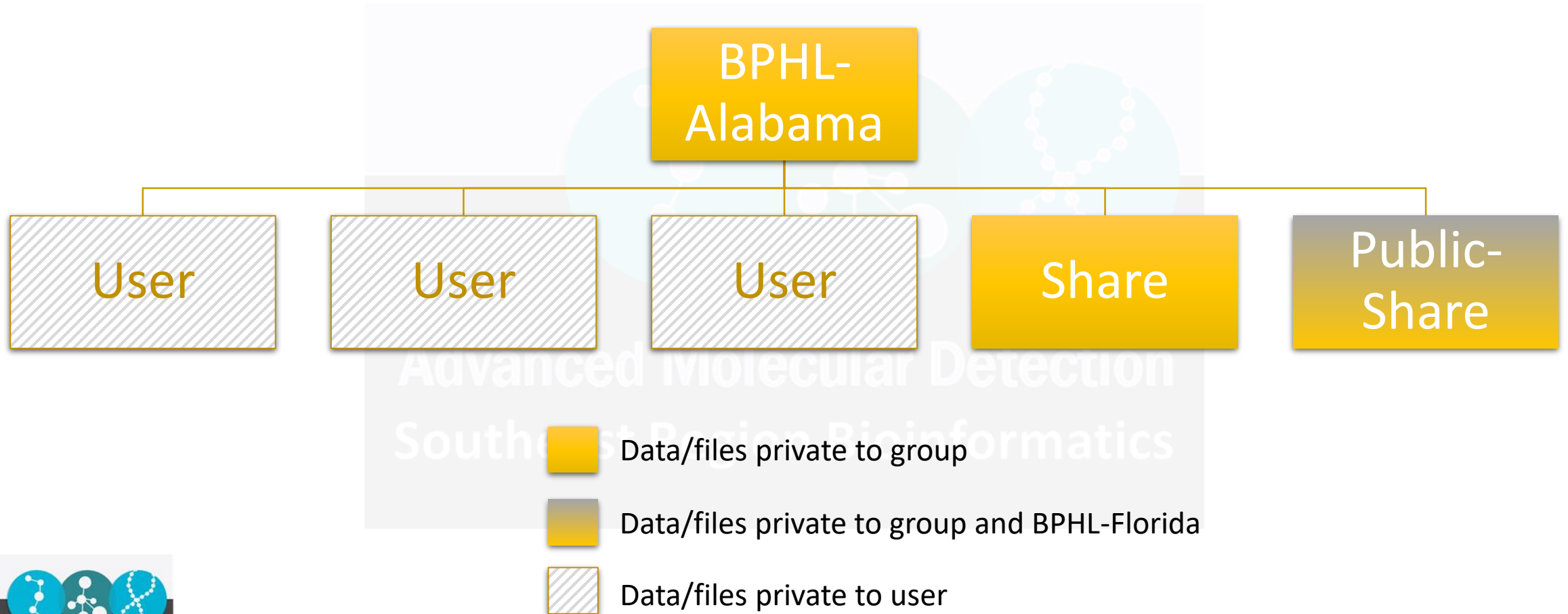


***Note: /ufrc will change to /blue in the coming weeks**

HiPerGator – State groups



HiPerGator – State groups

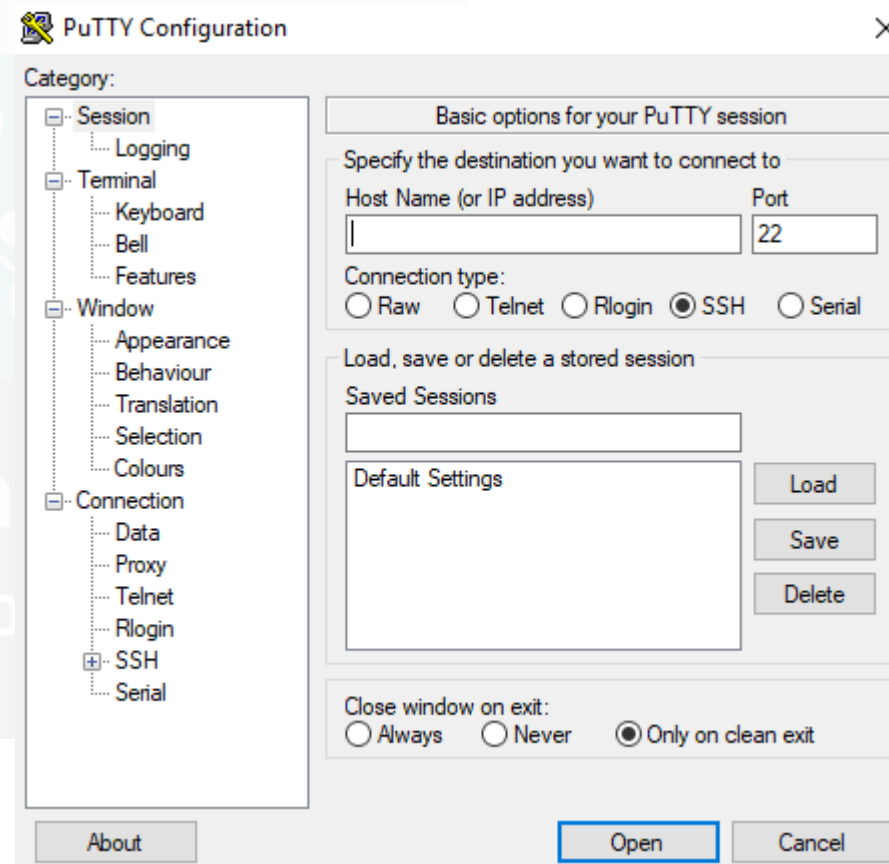


Logging into HiPerGator

- From Windows
 - Use Putty

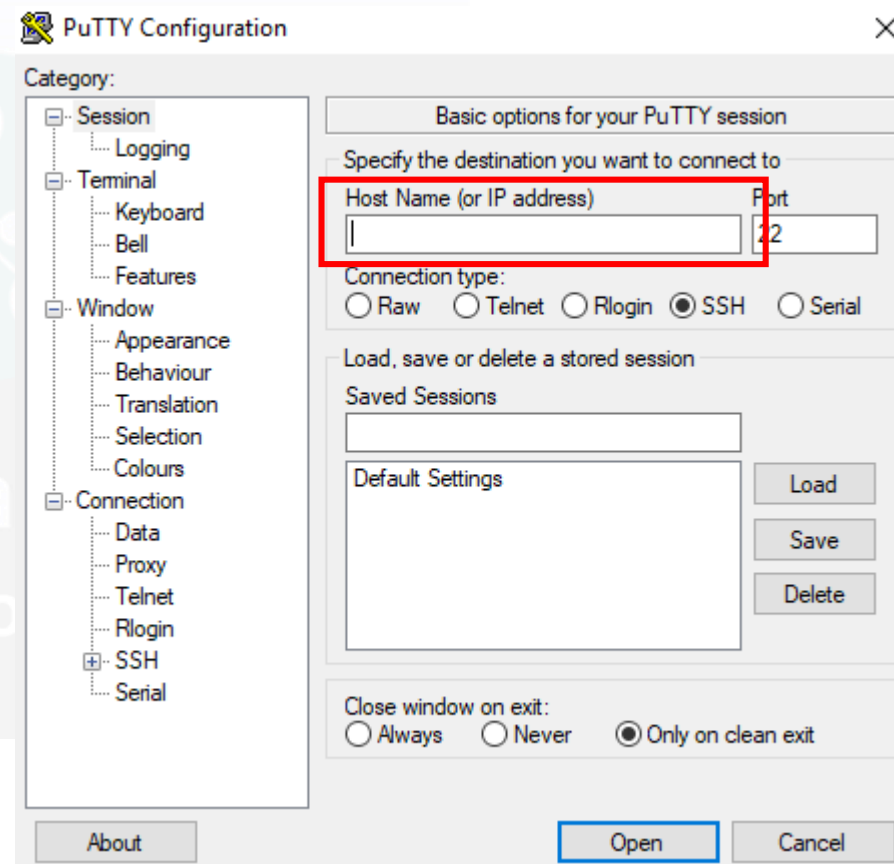


Double-click



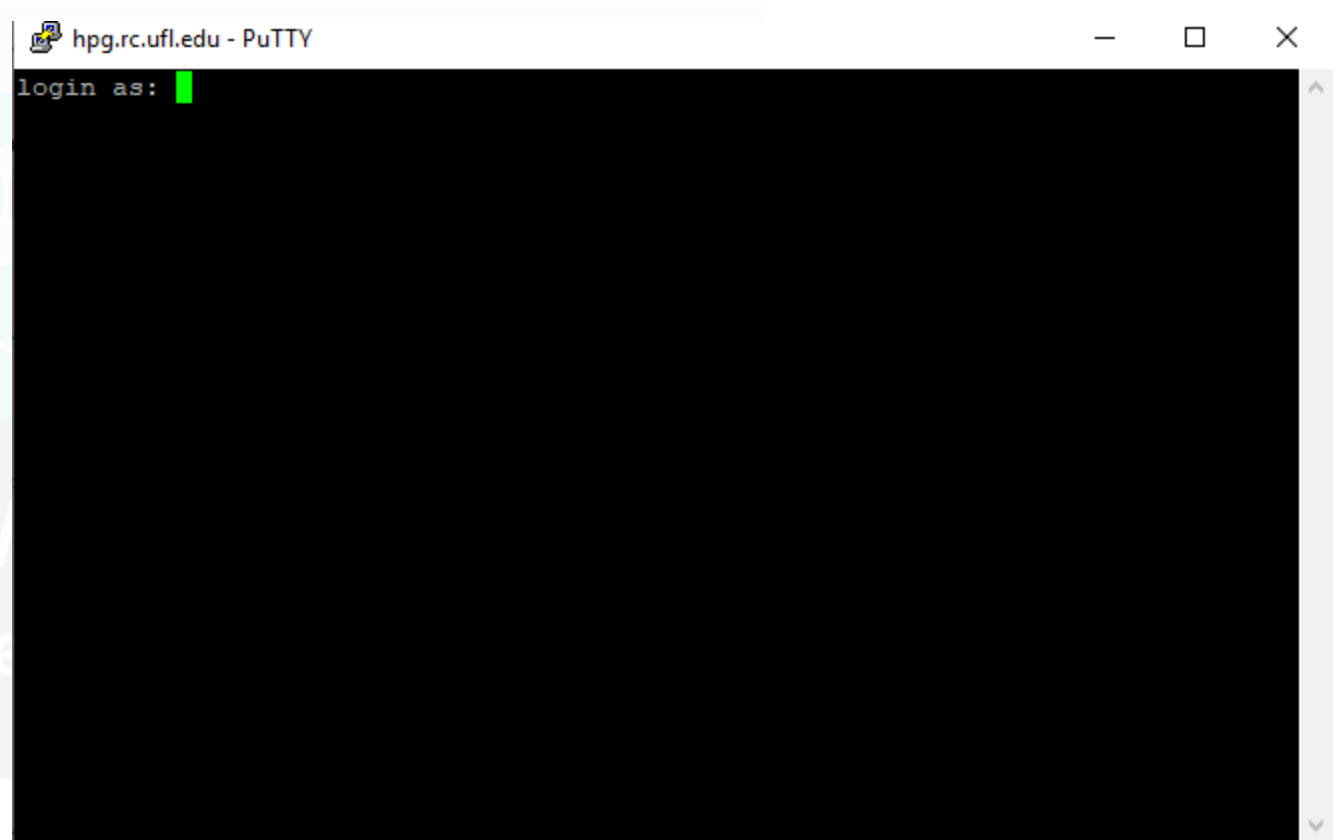
Logging into HiPerGator

- Host Name: hpg.rc.ufl.edu
- Click “Open” or press “Enter”



Logging into HiPerGator

- Type your GatorLink ID/HiPerGator Username
- Press “Enter”



Logging into HiPerGator

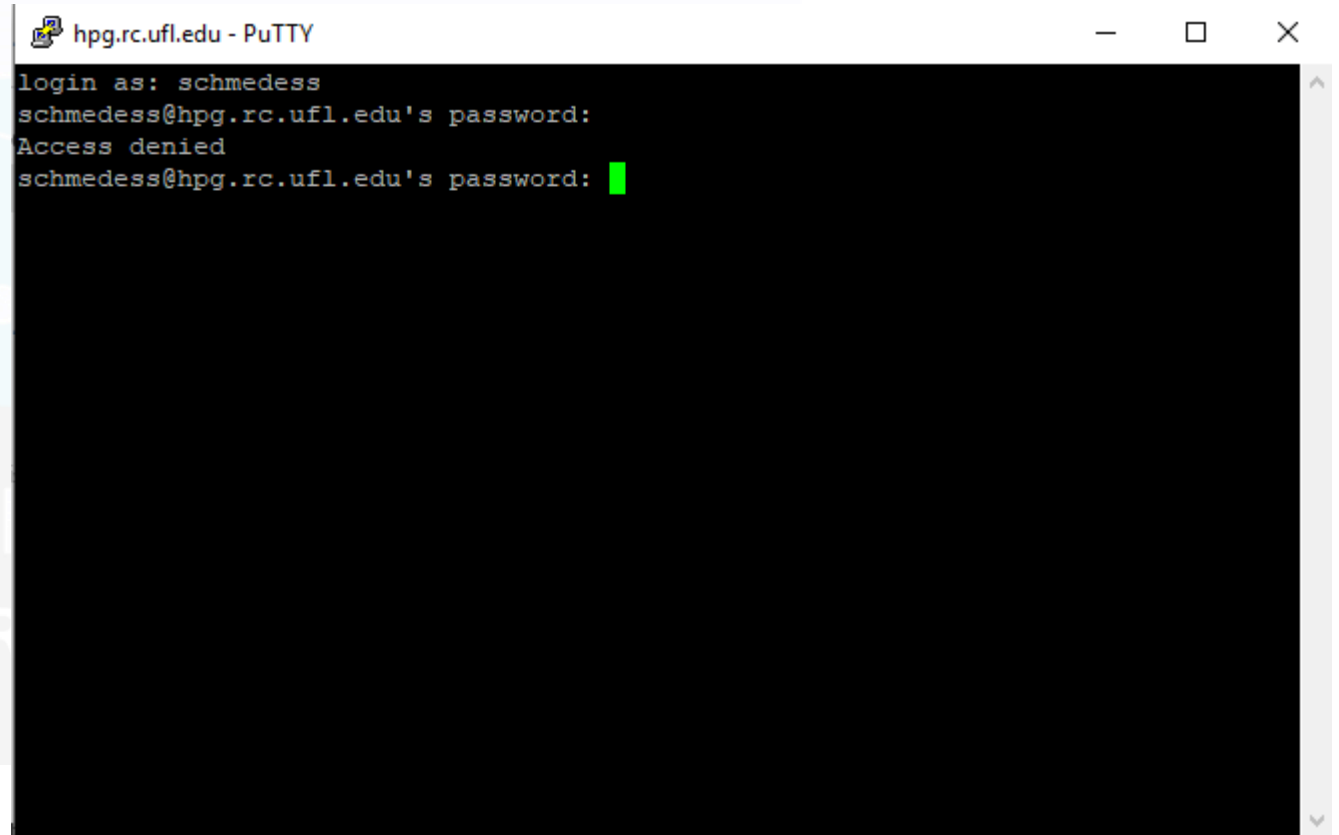
- Type your GatorLink ID/HiPerGator password
- Press “Enter”
- *Note: The password will not display as you type*



The screenshot shows a PuTTY terminal window titled "hpg.rc.ufl.edu - PuTTY". The terminal displays the following text: "login as: schmedess" followed by "schmedess@hpg.rc.ufl.edu's password:". A green cursor is visible at the end of the password prompt. The terminal background is black, and the text is white.

Logging into HiPerGator

- If you mistype your password, this screen will display
- Simply, re-type your password and try again



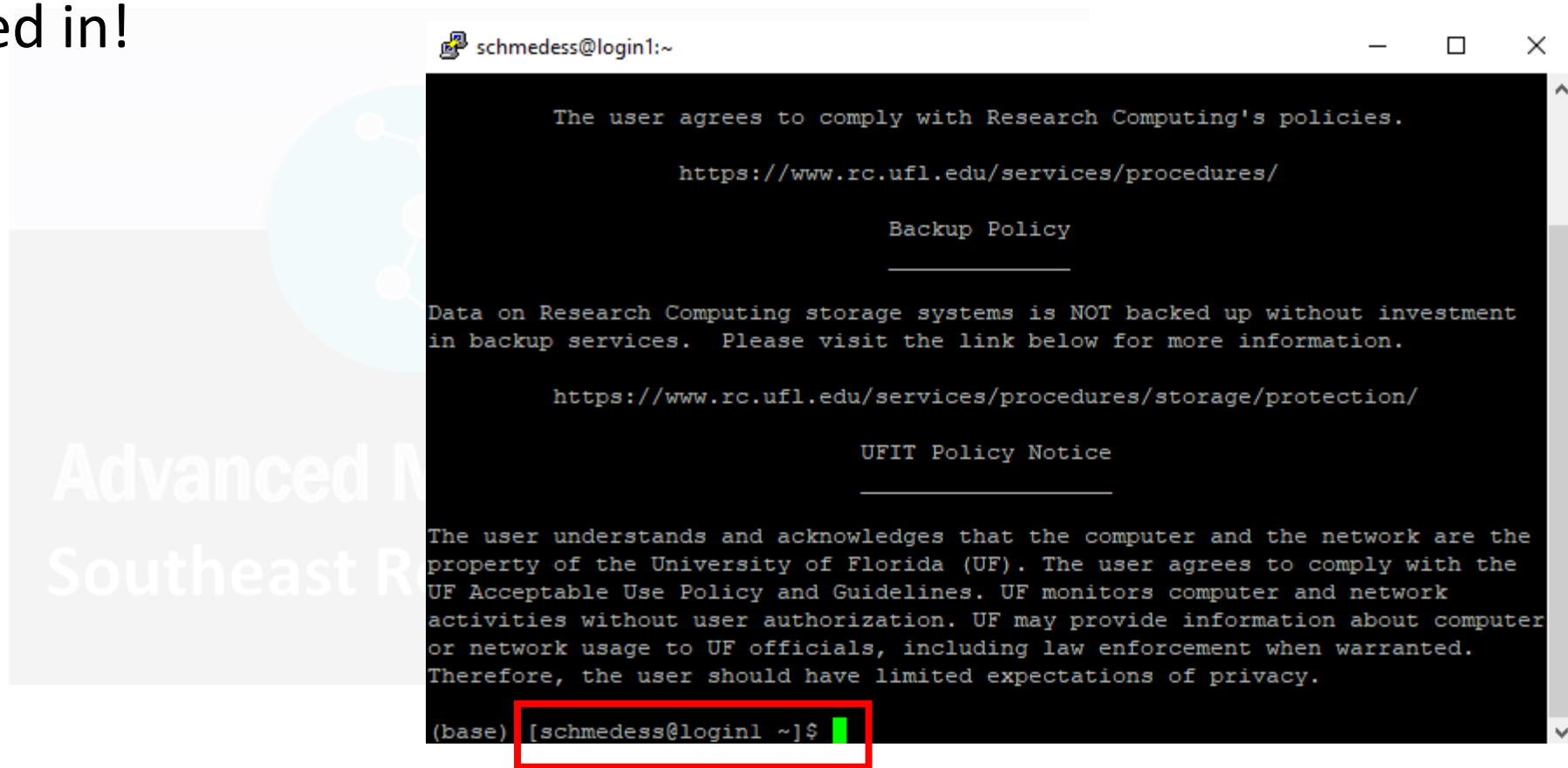
The screenshot shows a PuTTY terminal window titled "hpg.rc.ufl.edu - PuTTY". The terminal output is as follows:

```
login as: schmedess
schmedess@hpg.rc.ufl.edu's password:
Access denied
schmedess@hpg.rc.ufl.edu's password: █
```

The first password attempt is empty, resulting in "Access denied". The second attempt is shown with a green cursor, indicating it is being entered.

Logging into HiPerGator

- You are logged in!



A terminal window titled 'schmedess@login1:~' displays the following text:

```
The user agrees to comply with Research Computing's policies.  
  
https://www.rc.ufl.edu/services/procedures/  
  
Backup Policy  
-----  
Data on Research Computing storage systems is NOT backed up without investment  
in backup services. Please visit the link below for more information.  
  
https://www.rc.ufl.edu/services/procedures/storage/protection/  
  
UFIT Policy Notice  
-----  
The user understands and acknowledges that the computer and the network are the  
property of the University of Florida (UF). The user agrees to comply with the  
UF Acceptable Use Policy and Guidelines. UF monitors computer and network  
activities without user authorization. UF may provide information about computer  
or network usage to UF officials, including law enforcement when warranted.  
Therefore, the user should have limited expectations of privacy.  
  
(base) [schmedess@login1 ~]$
```

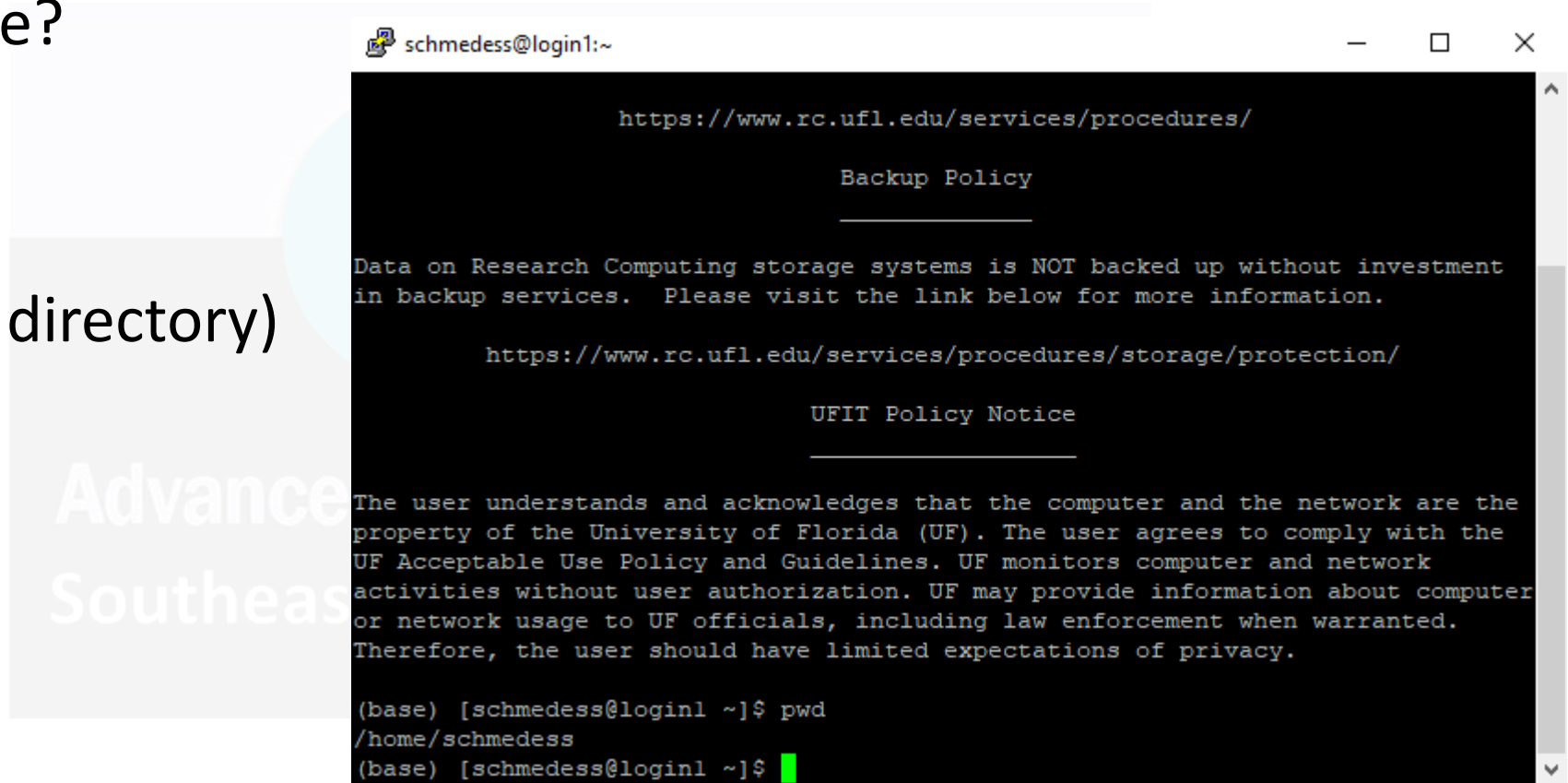
The prompt line at the bottom is highlighted with a red rectangle.

Navigating the file structure

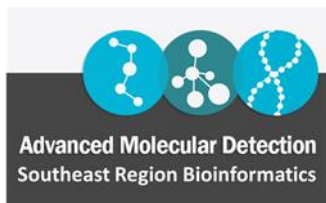
- Where are we?

pwd

(print working directory)



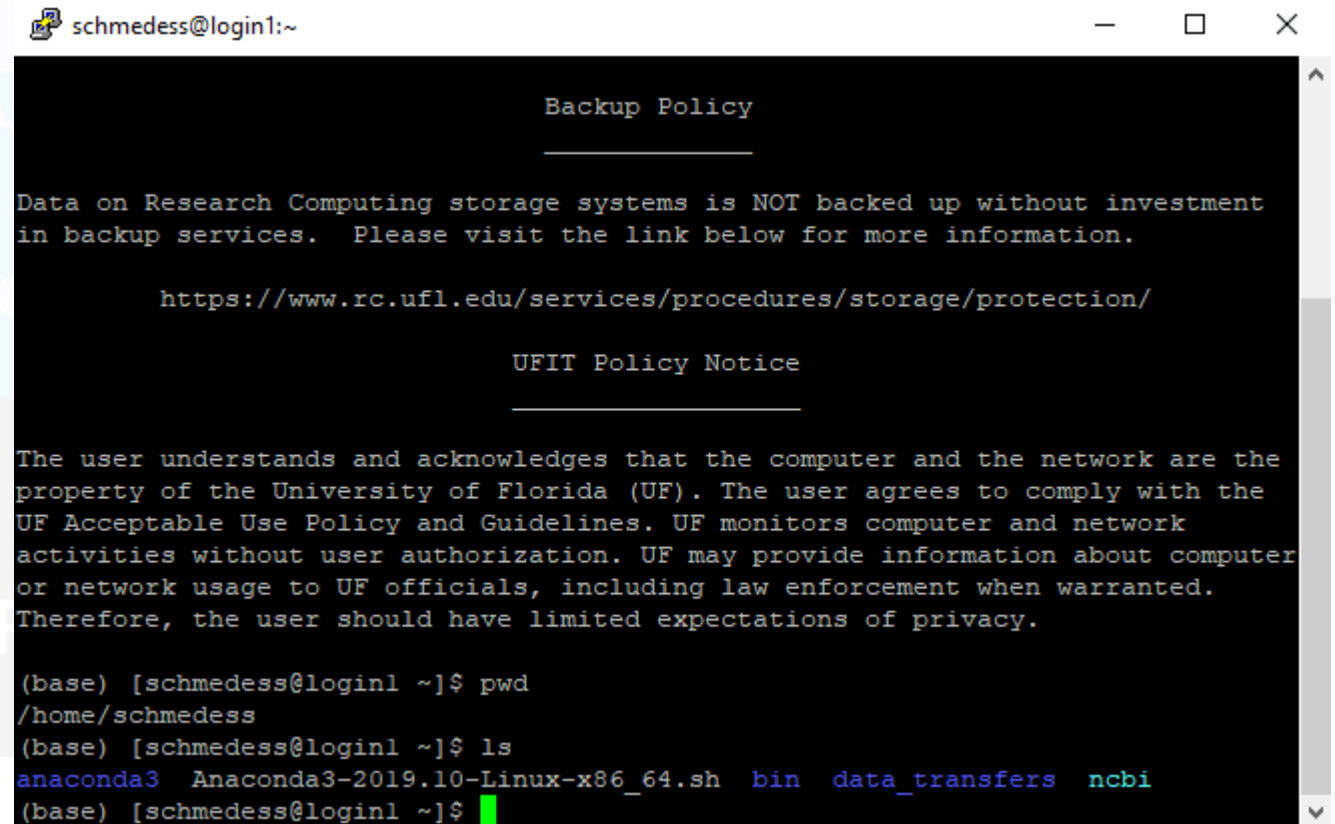
A terminal window titled 'schmedess@login1:~' is shown. It displays the output of a web browser. The browser address bar shows 'https://www.rc.ufl.edu/services/procedures/'. The page content includes a section titled 'Backup Policy' with a horizontal line underneath. The text below reads: 'Data on Research Computing storage systems is NOT backed up without investment in backup services. Please visit the link below for more information.' followed by the URL 'https://www.rc.ufl.edu/services/procedures/storage/protection/'. Below this is another section titled 'UFIT Policy Notice' with a horizontal line underneath. The text reads: 'The user understands and acknowledges that the computer and the network are the property of the University of Florida (UF). The user agrees to comply with the UF Acceptable Use Policy and Guidelines. UF monitors computer and network activities without user authorization. UF may provide information about computer or network usage to UF officials, including law enforcement when warranted. Therefore, the user should have limited expectations of privacy.' At the bottom of the terminal, the command '(base) [schmedess@login1 ~]\$ pwd' has been executed, resulting in the output '/home/schmedess'. The prompt '(base) [schmedess@login1 ~]\$' is followed by a green cursor.



Navigating the file structure

- What files are in my current working directory?

ls
(list)



```
schmedess@login1:~  
  
Backup Policy  
-----  
Data on Research Computing storage systems is NOT backed up without investment  
in backup services. Please visit the link below for more information.  
  
https://www.rc.ufl.edu/services/procedures/storage/protection/  
  
UFIT Policy Notice  
-----  
The user understands and acknowledges that the computer and the network are the  
property of the University of Florida (UF). The user agrees to comply with the  
UF Acceptable Use Policy and Guidelines. UF monitors computer and network  
activities without user authorization. UF may provide information about computer  
or network usage to UF officials, including law enforcement when warranted.  
Therefore, the user should have limited expectations of privacy.  
  
(base) [schmedess@login1 ~]$ pwd  
/home/schmedess  
(base) [schmedess@login1 ~]$ ls  
anaconda3  Anaconda3-2019.10-Linux-x86_64.sh  bin  data_transfers  ncbi  
(base) [schmedess@login1 ~]$
```

Navigating the file structure

- Other information about my files

`ls -althr --color=auto`

d (directory)

- (file)

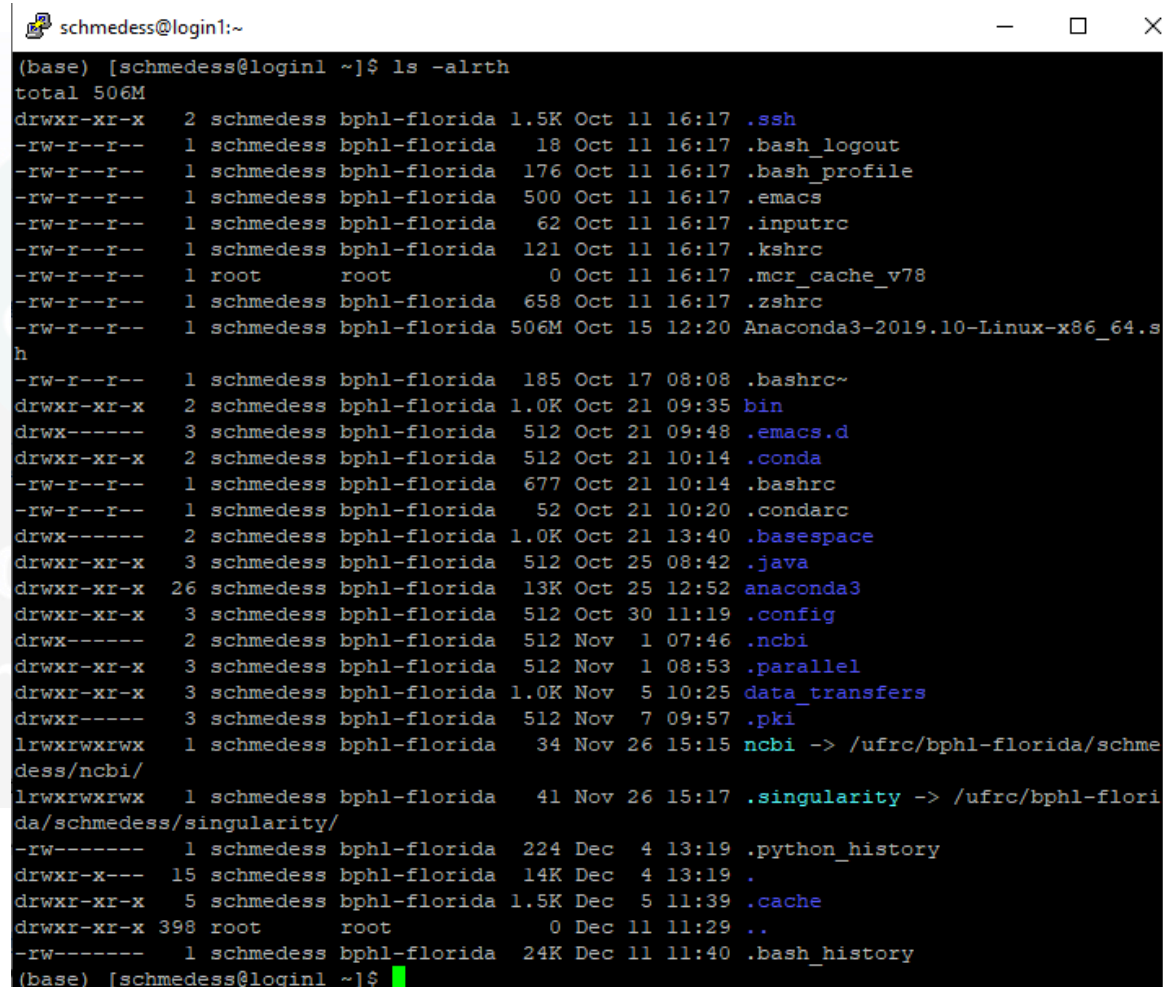
l (link)

. (current directory)

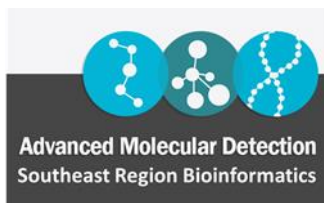
.. (parent directory)

.filename (hidden file)

-rw-r--r-- (file permissions)



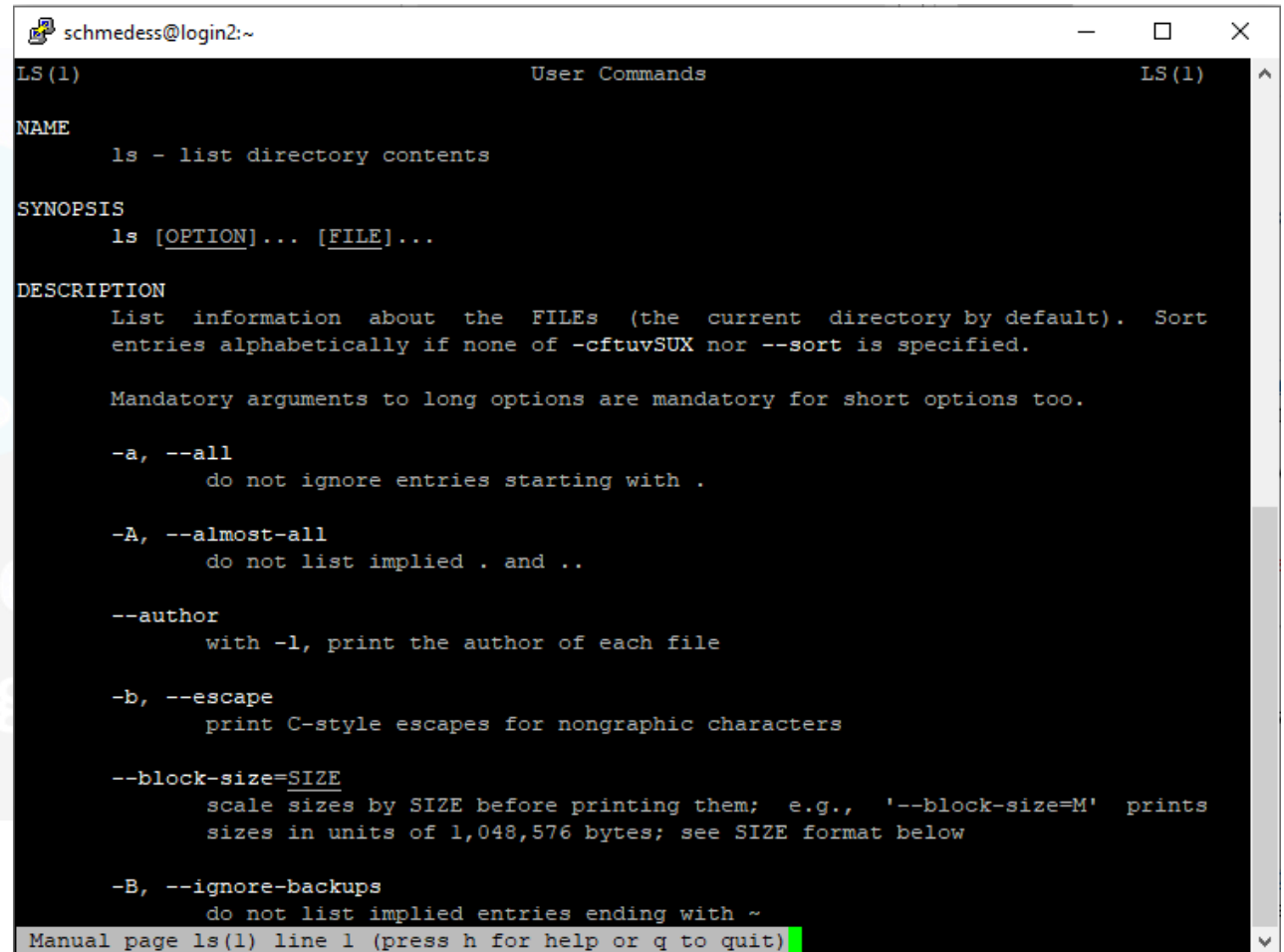
```
(base) [schmedess@login1 ~]$ ls -althr
total 506M
drwxr-xr-x  2 schmedess bphl-florida 1.5K Oct 11 16:17 .ssh
-rw-r--r--  1 schmedess bphl-florida  18 Oct 11 16:17 .bash_logout
-rw-r--r--  1 schmedess bphl-florida 176 Oct 11 16:17 .bash_profile
-rw-r--r--  1 schmedess bphl-florida 500 Oct 11 16:17 .emacs
-rw-r--r--  1 schmedess bphl-florida  62 Oct 11 16:17 .inputrc
-rw-r--r--  1 schmedess bphl-florida 121 Oct 11 16:17 .kshrc
-rw-r--r--  1 root      root          0 Oct 11 16:17 .mcr_cache_v78
-rw-r--r--  1 schmedess bphl-florida 658 Oct 11 16:17 .zshrc
-rw-r--r--  1 schmedess bphl-florida 506M Oct 15 12:20 Anaconda3-2019.10-Linux-x86_64.s
h
-rw-r--r--  1 schmedess bphl-florida 185 Oct 17 08:08 .bashrc~
drwxr-xr-x  2 schmedess bphl-florida 1.0K Oct 21 09:35 bin
drwx-----  3 schmedess bphl-florida 512 Oct 21 09:48 .emacs.d
drwxr-xr-x  2 schmedess bphl-florida 512 Oct 21 10:14 .conda
-rw-r--r--  1 schmedess bphl-florida 677 Oct 21 10:14 .bashrc
-rw-r--r--  1 schmedess bphl-florida  52 Oct 21 10:20 .condarc
drwx-----  2 schmedess bphl-florida 1.0K Oct 21 13:40 .basespace
drwxr-xr-x  3 schmedess bphl-florida 512 Oct 25 08:42 .java
drwxr-xr-x 26 schmedess bphl-florida 13K Oct 25 12:52 anaconda3
drwxr-xr-x  3 schmedess bphl-florida 512 Oct 30 11:19 .config
drwx-----  2 schmedess bphl-florida 512 Nov  1 07:46 .ncbi
drwxr-xr-x  3 schmedess bphl-florida 512 Nov  1 08:53 .parallel
drwxr-xr-x  3 schmedess bphl-florida 1.0K Nov  5 10:25 data_transfers
drwxr-----  3 schmedess bphl-florida 512 Nov  7 09:57 .pki
lrwxrwxrwx  1 schmedess bphl-florida  34 Nov 26 15:15 ncbi -> /ufrc/bphl-florida/schme
dess/ncbi/
lrwxrwxrwx  1 schmedess bphl-florida  41 Nov 26 15:17 .singularity -> /ufrc/bphl-flori
da/schmedess/singularity/
-rw-----  1 schmedess bphl-florida 224 Dec  4 13:19 .python_history
drwxr-xr-x 15 schmedess bphl-florida 14K Dec  4 13:19 .
drwxr-xr-x  5 schmedess bphl-florida 1.5K Dec  5 11:39 .cache
drwxr-xr-x 398 root      root          0 Dec 11 11:29 ..
-rw-----  1 schmedess bphl-florida 24K Dec 11 11:40 .bash_history
(base) [schmedess@login1 ~]$
```



What does -althr mean?

man ls

- Flags/options
- Command man ls
- A = all (including hidden files)
- H= “human readable”, prints sizes in 1K, 1 g, 1, M format)
- L = long listing format
- T = show time
- R= reverse order while sorting, most recent at bottom



```
schmedess@login2:~  
LS(1) User Commands LS(1)  
NAME  
    ls - list directory contents  
SYNOPSIS  
    ls [OPTION]... [FILE]...  
DESCRIPTION  
    List information about the FILES (the current directory by default). Sort  
    entries alphabetically if none of -cftuvSUX nor --sort is specified.  
  
    Mandatory arguments to long options are mandatory for short options too.  
  
    -a, --all  
        do not ignore entries starting with .  
  
    -A, --almost-all  
        do not list implied . and ..  
  
    --author  
        with -l, print the author of each file  
  
    -b, --escape  
        print C-style escapes for nongraphic characters  
  
    --block-size=SIZE  
        scale sizes by SIZE before printing them; e.g., '--block-size=M' prints  
        sizes in units of 1,048,576 bytes; see SIZE format below  
  
    -B, --ignore-backups  
        do not list implied entries ending with ~  
Manual page ls(1) line 1 (press h for help or q to quit)
```

File Permissions

```
drwxr-xr-x  2 schmedess bphl-florida 1.0K Oct 21 09:35 bin
drwx----- 3 schmedess bphl-florida 512 Oct 21 09:48 .emacs.d
drwxr-xr-x  2 schmedess bphl-florida 512 Oct 21 10:14 .conda
-rw-r--r--  1 schmedess bphl-florida 677 Oct 21 10:14 .bashrc
-rw-r--r--  1 schmedess bphl-florida  52 Oct 21 10:20 .condarc
```

```
# ls -l file
-rw-r--r-- 1 root root 0 Nov 19 23:49 file
```

Diagram illustrating the breakdown of file permissions for `-rw-r--r--`:

- File type:** Indicated by the first character (`-`).
- Owner (rw-):** The first three characters (`rw-`) represent permissions for the owner.
- Group (r- -):** The next three characters (`r- -`) represent permissions for the group.
- Other (r- -):** The final three characters (`r- -`) represent permissions for others.

Legend:

- `r` = Readable
- `w` = Writeable
- `x` = Executable
- `-` = Denied

<https://www.thegeekdiary.com/understanding-basic-file-permissions-and-ownership-in-linux/>

File Permissions

`chmod <permissions> <filename>`

(change file mode)

Example:

`chmod 770 report.txt`



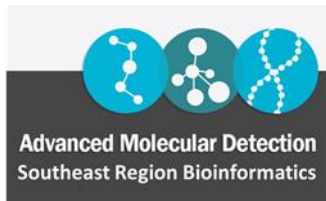
`-rwxrwx--- report.txt`

Number	Permission Type	Symbol
0	No Permission	---
1	Execute	--X
2	Write	-W-
3	Execute + Write	-WX
4	Read	r--
5	Read + Execute	r-X
6	Read + Write	rw-
7	Read + Write + Execute	rwx

<https://www.guru99.com/file-permissions.html>

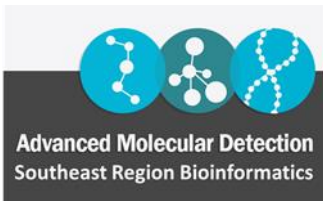
Navigating the file structure

- `cd <directory name>`
 - change directory
- `cd /absolute/path/to/directory/` or `relative/path/`
- `cd ~` or `cd`
 - Go to your home directory
- `cd ..` or `cd ../`
 - Go up one directory in the tree (to parent directory)
- `cd ../.../`
 - Go up two directories in the file tree (“grandparent directory”)



Change from home to /ufrc

- `cd /ufrc/bphl-<state>/<user>`
- `/home/<user>/`
 - Very limited storage – **do NOT store your data here!!!!**
 - Not high performance – **do NOT run jobs from here!!!!**
 - `cd` will take you home
- `/ufrc/bphl-<state>/<user>`
 - This is where you do the bulk of your work.
 - This is where your data input and output from “jobs” go.
 - `cd /ufrc/bphl-<state>/<user>` will take you to your user directory in your group



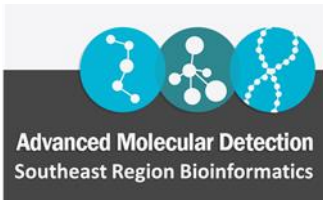
More bash commands

- `mkdir <name of directory>`
 - (make a new directory)
- `cp <path/to/file> <path/to/new/location>`
 - (copy a file to a new location)
 - `cp /path/to/file .`
 - (copy file to current working directory)
 - Example:
 - `cp sample_*.fastq.gz data/`
 - (copy read 1 and read 2 fastq files for sample at the same time to data/)

More bash commands

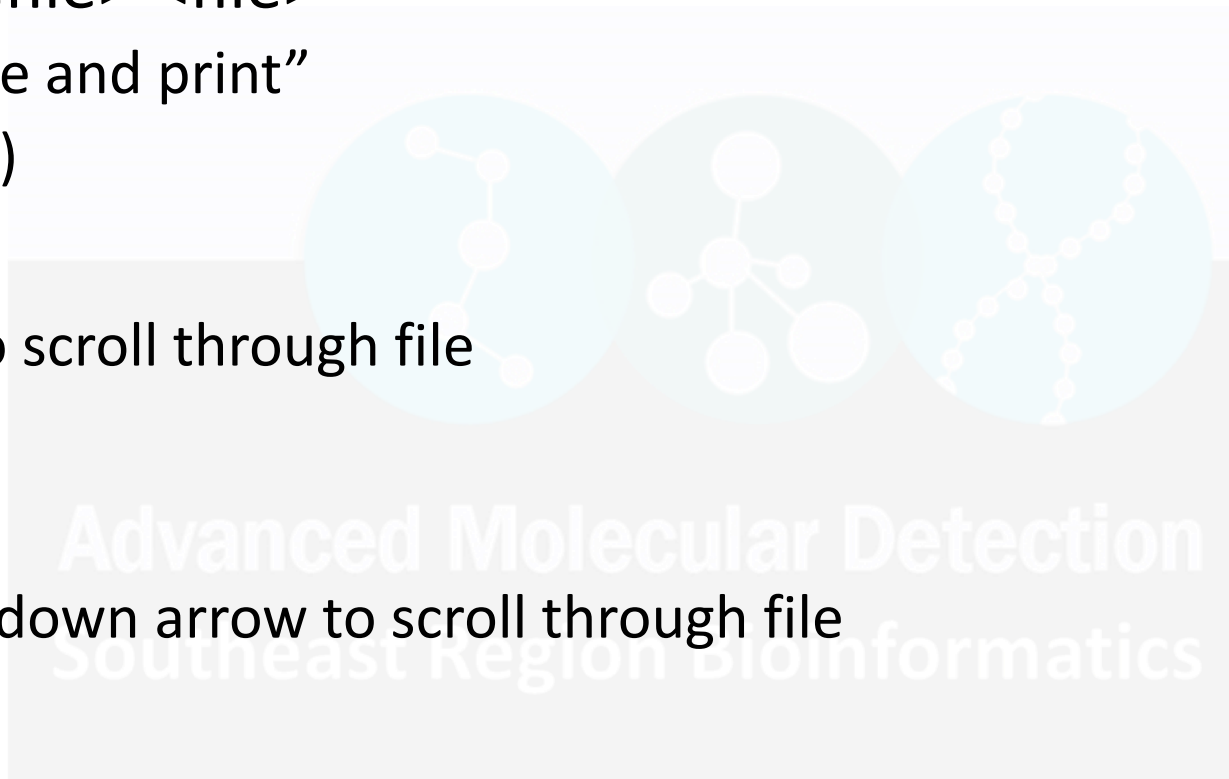
- touch <filename>
 - (Create new/empty file)
- rm <filename>
 - (remove file)
- rm -r <directory name>
 - (remove directory and all contents)

Advanced Molecular Detection
Southeast Region Bioinformatics



Looking at files

- `cat <file>` or `<file> <file>`
 - “concatenate and print”
 - `zcat` (gz files)
- `more <file>`
 - Use Enter to scroll through file
 - q to quit
- `less <file>`
 - Use up and down arrow to scroll through file
 - q to quit

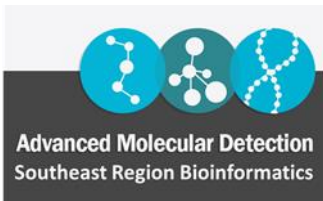


Looking at files

- head <file>
 - Print first 10 lines
 - head -n <number of lines other than 10>
- tail <file>
 - Print last 10 lines
 - tail -n <number of lines other than 10>



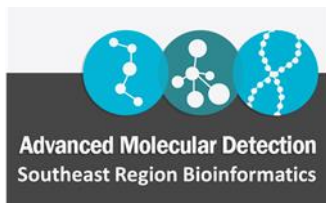
Advanced Molecular Detection
Southeast Region Bioinformatics



Looking at files

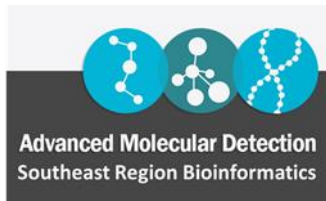
wc – “word count”

- `wc -l <file>`
 - Line count
- `wc -w <file>`
 - Word count
- `wc -m <file>`
 - Character count



Tricks and Tips

- *wild card
- Tab = auto-complete
- Up arrow = scroll through previous commands
- Control-A = moves cursor to beginning of command prompt line
- Control-E = moves cursor to end of command prompt line
- Alt-F = moves cursor one word forward at a time in command line
- Alt-B = moves cursor one word back at a time in command line
- Control-C = kills a process
- clear = clears terminal

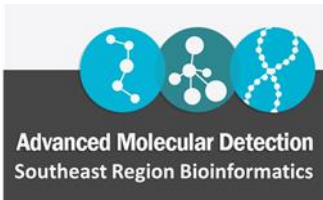


Future Webinars

- Working with the BaseSpace Command Line Interface (CLI)
- Working in Galaxy – including data transfers from HiPerGator
- SARS-CoV-2 Sequencing



Advanced Molecular Detection
Southeast Region Bioinformatics





Advanced Molecular Detection Southeast Region Bioinformatics

Questions???

BPHL-SEbioinformatics@flhealth.gov

Sarah Schmedes, PhD
BRR Lead, Southeast Region

Taj Azarian, PhD
Bioinformatician
Assistant Professor, UCF

Jason Blanton, PhD
Molecular Administrator
State Sequencing Coordinator