Introduction to the **Shared Computing Cluster (SCC)**

Charles Jahnke
Research Computing Services
Information Services & Technology

BOSTON UNIVERSITY

Topics for Today

Intro

- Research Computing Services
- Shared Compute Cluster Overview
- Getting and Account

Linux

- Connecting to the SCC
- Commands
- Permissions
- Tools/Methods

Compute Cluster

- Software and Modules
- Cluster Architecture
- Files/Storage
- The Batch System

Research Computing Services

This will be quick.

Research Computing Services (RCS)

A group within Information Services & Technology at Boston University provides computing, storage, and visualization resources and services to support research that has specialized or highly intensive computation, storage, bandwidth, or graphics requirements.

Three Primary Services:

- Research Computation
- Research Visualization
- Research Consulting and Training

RCS Team and Expertise

Our Team

- Scientific Programmers
- Systems Administrators
- Graphics/Visualization Specialists
- Account/Project Managers
- Special Initiatives (Grants)

Consulting Focus:

- Bioinformatics
- Data Analysis \ Statistics
- Molecular modeling
- Geographic Information Systems
- Scientific/Engineering Simulation
- Visualization

Me

- Systems Programmer and Administrator
- Background in biomedical engineering, technology, and bioinformatics
- Office on the Boston University Medical Campus
  - We also have staff on the Charles River Campus
- Contact:
  - Email: cjahnke@bu.edu
  - Office: Crosstown Building, Suite 485

Our whole team: help@scc.bu.edu

You

- Who has experience programming?
- Working with "big data"?
- Using Linux?
- Using compute clusters?
- Have an account on SCC?

Shared Computing Cluster (SCC) Overview

The Shared Computing Cluster (SCC)

- A Linux compute cluster with 11,000 CPU cores and 250 GPUs.
- Over 3 Petabytes of disk space.
- Located at the Massachusetts Green High Performance Computing Center (MGHPCC) in Holyoke, MA
- Owned by Boston University and researchers.
- Went into production in June, 2013 for Research Computing.

MGHPCC

- Collaboration between 5 universities, MA state, and industry.

Boston University

Northeastern University

UMASS

VE RI TAS

HARVARD

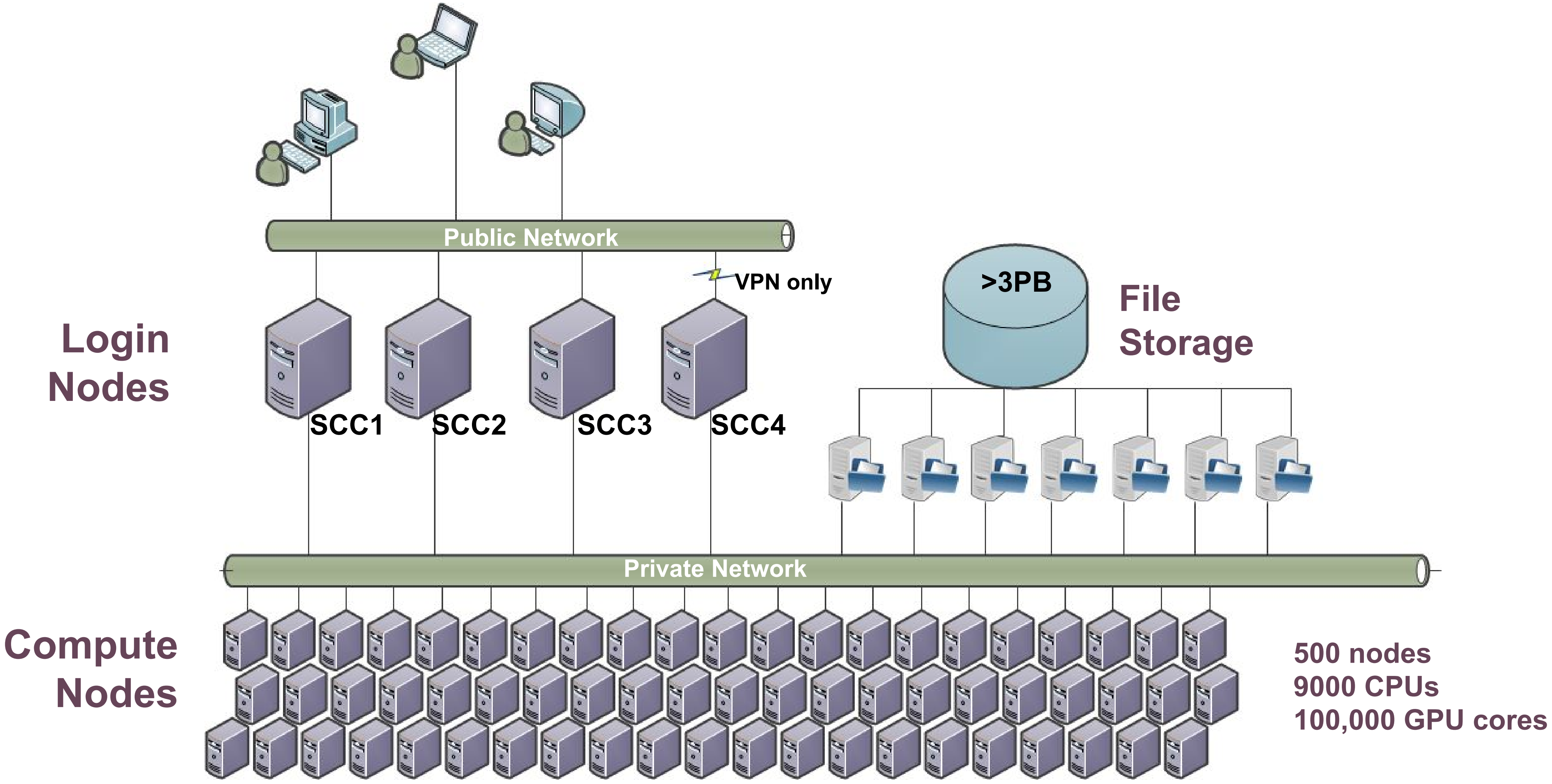
- State-of-the-art data center in Holyoke, MA.
- MGHPCC provides physical infrastructure (i.e. space, power, cooling), not computing systems.
- Individual universities or consortiums provide their own computing and support.

MGHPCC

MASSACHUSETTS GREEN HIGH PERFORMANCE COMPUTING CENTER

MGHPCC

SCC Architecture



That seems like a lot… why do all this?

Researchers need to:

- Collaborate on shared data.
- Run code that exceeds workstation capability (RAM, Network, Disk).
- Run code that runs for long periods of time (days, weeks, months)
- Run code in highly parallelized formats (use 100 machines simultaneously).
- Might want to do all of those things 1,000 times.

Getting an Account on SCC

Getting an Account on SCC

- BU Faculty members can create SCC projects online (no cost).
- All users of the SCC must be on a SCC research project managed by a full-time BU Faculty member.
- Exception: 3 month trial accounts for students/tutorial attendees.
  - Email help@scc.bu.edu if interested.

http://www.bu.edu/tech/support/research/account-management/

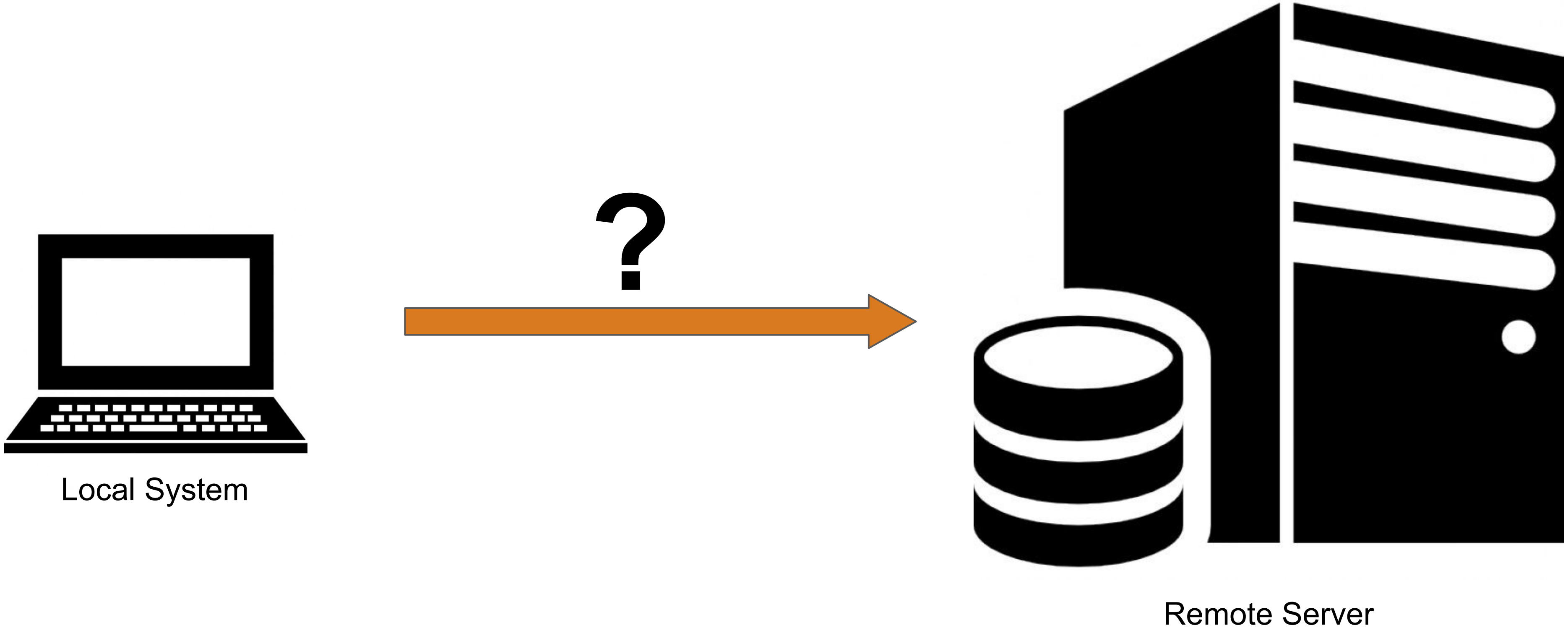
Today

- Use your own account if you have one.

- Tutorial Accounts:
  - Username:
  - Password:
  - These should not be used after today.

Connecting to SCC

Windows, OS X, Linux

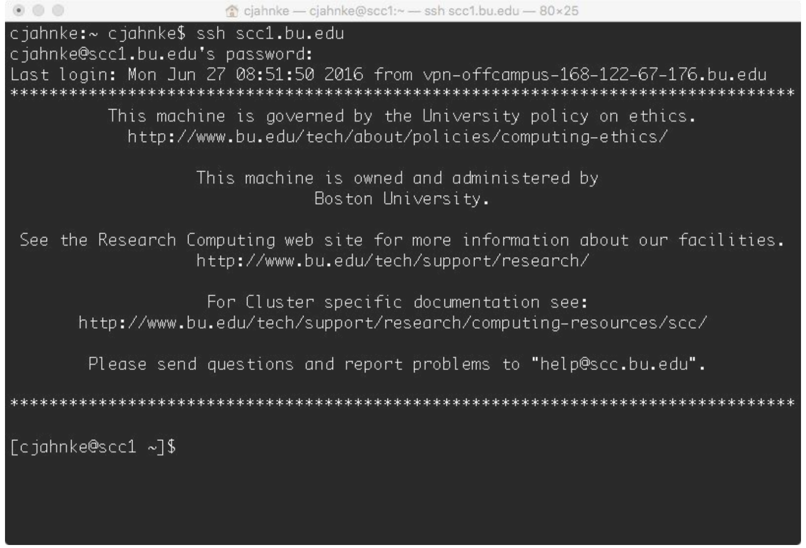

?
Local System
Remote Server

Connection Protocols and Software

Remote Connections:

Secure **SH**ell

(SSH)



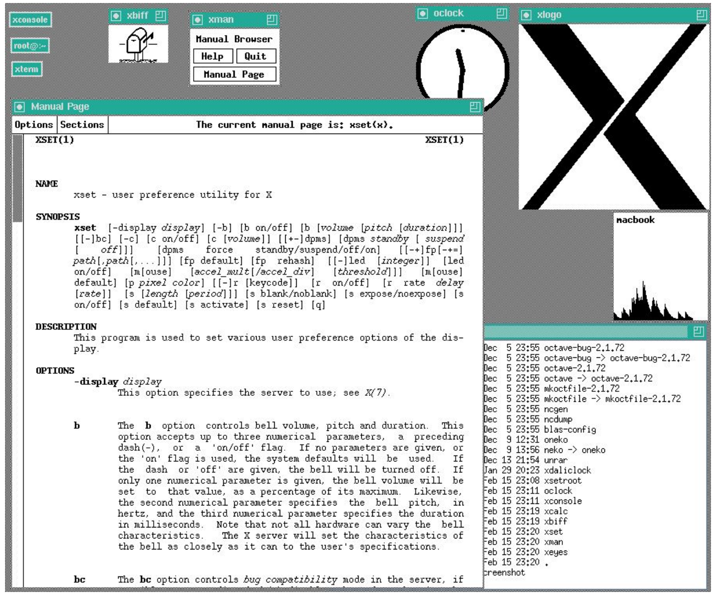
cjahnke — cjahnke@scc1:~ — ssh scc1.bu.edu — 80×25

```
cjahnke:~ cjahnke$ ssh scc1.bu.edu
cjahnke@scc1.bu.edu's password:
Last login: Mon Jun 27 08:51:50 2016 from vpn-offcampus-168-122-67-176.bu.edu
********************************************************************************
          This machine is governed by the University policy on ethics.
            http://www.bu.edu/tech/about/policies/computing-ethics/

                   This machine is owned and administered by
                              Boston University.

 See the Research Computing web site for more information about our facilities.
                   http://www.bu.edu/tech/support/research/

                    For Cluster specific documentation see:
      http://www.bu.edu/tech/support/research/computing-resources/scc/

      Please send questions and report problems to "help@scc.bu.edu".

********************************************************************************

[cjahnke@scc1 ~]$
```

Remote Graphics:

X-Windowing

(X, X-Win)



Manual Page

Options | Sections | The current manual page is: xset(x).

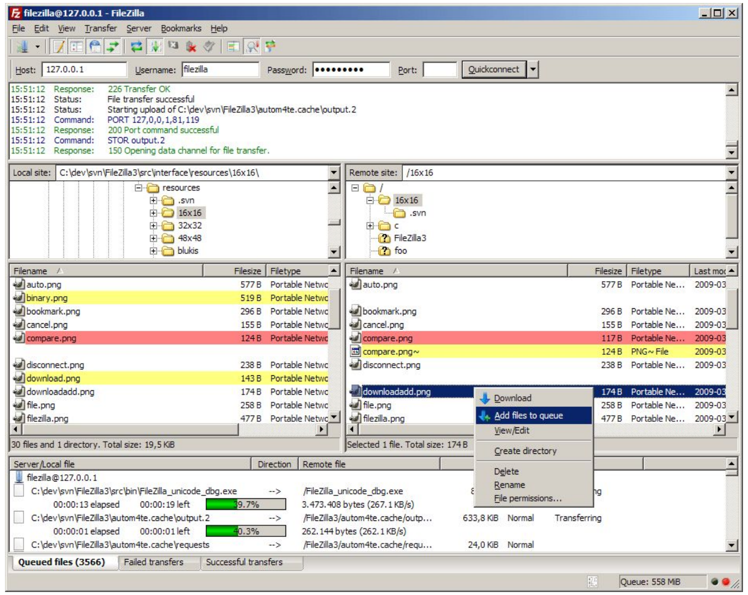
```
XSET(1)                                                          XSET(1)



NAME
       xset - user preference utility for X

SYNOPSIS
       xset  [-display display] [-b] [b on/off] [b [volume [pitch [duration]]]
       [[-]bc] [-c] [c on/off] [c [volume]] [[+-]dpms] [dpms standby [ suspend
       [   off]]]   [dpms   force   standby/suspend/off/on]   [[-+]fp[-+=]
       path[,path[,...]]] [fp default] [fp  rehash]  [[-]led  [integer]]  [led
       on/off]  [m[ouse]  [accel_mult[/accel_div]  [threshold]]]  [m[ouse]
       default] [p pixel color] [[-]r [keycode]]  [r  on/off]  [r  rate  delay
       [rate]]  [s [length [period]]] [s blank/noblank] [s expose/noexpose] [s
       on/off] [s default] [s activate] [s reset] [q]

DESCRIPTION
       This program is used to set various user preference options of the dis-
       play.

OPTIONS
       -display display
               This option specifies the server to use; see X(7).


       b       The  b  option  controls bell volume, pitch and duration.  This
               option accepts up to three numerical  parameters,  a  preceding
               dash(-),  or  a  'on/off' flag.  If no parameters are given, or
               the 'on' flag is used, the system defaults will  be  used.   If
               the  dash  or 'off' are given, the bell will be turned off.  If
               only one numerical parameter is given, the bell volume will  be
               set  to  that value, as a percentage of its maximum.  Likewise,
               the second numerical parameter specifies  the  bell  pitch,  in
               hertz, and the third numerical parameter specifies the duration
               in milliseconds.  Note that not all hardware can vary the  bell
               characteristics.   The X server will set the characteristics of
               the bell as closely as it can to the user's specifications.


       bc      The bc option controls bug compatibility mode in the server, if
```

```
Dec  5 23:55 octave-bug-2.1.72
Dec  5 23:55 octave-bug -> octave-bug-2.1.72
Dec  5 23:55 octave-2.1.72
Dec  5 23:55 octave -> octave-2.1.72
Dec  5 23:55 mkoctfile-2.1.72
Dec  5 23:55 mkoctfile -> mkoctfile-2.1.72
Dec  5 23:55 ncgen
Dec  5 23:55 ncdump
Dec  5 23:55 blas-config
Dec  9 12:31 oneko
Dec  9 13:56 neko -> oneko
Dec 13 21:54 unrar
Jan 29 20:23 xdaliclock
Feb 15 23:08 xsetroot
Feb 15 23:11 oclock
Feb 15 23:11 xconsole
Feb 15 23:19 xcalc
Feb 15 23:19 xbiff
Feb 15 23:20 xset
Feb 15 23:20 xman
Feb 15 23:20 xeyes
Feb 15 23:20 .
creenshot
```

Data Transfer:

Secure **F**ile **T**ransfer **P**rotocol

(SFTP)



Host: 127.0.0.1 Username: filezilla Password: •••••••••• Port:

```
15:51:12  Response:  226 Transfer OK
15:51:12  Status:    File transfer successful
15:51:12  Status:    Starting upload of C:\dev\svn\FileZilla3\autom4te.cache\output.2
15:51:12  Command:   PORT 127,0,0,1,81,119
15:51:12  Response:  200 Port command successful
15:51:12  Command:   STOR output.2
15:51:12  Response:  150 Opening data channel for file transfer.
```

Local site: C:\dev\svn\FileZilla3\src\interface\resources\16x16\

```
resources
  .svn
  16x16
  32x32
  48x48
  blukis
```

Remote site: /16x16

```
/
  16x16
    .svn
  c
  FileZilla3
  foo
```

Filename	Filesize	Filetype
auto.png	577 B	Portable Netwo
binary.png	519 B	Portable Netwo
bookmark.png	296 B	Portable Netwo
cancel.png	155 B	Portable Netwo
compare.png	124 B	Portable Netwo
disconnect.png	238 B	Portable Netwo
download.png	143 B	Portable Netwo
downloadadd.png	174 B	Portable Netwo
file.png	258 B	Portable Netwo
filezilla.png	477 B	Portable Netwo

30 files and 1 directory. Total size: 19,5 KiB

Filename	Filesize	Filetype	Last moc
auto.png	577 B	Portable Ne...	2009-03
bookmark.png	296 B	Portable Ne...	2009-03
cancel.png	155 B	Portable Ne...	2009-03
compare.png	117 B	Portable Ne...	2009-03
compare.png~	124 B	PNG~ File	2009-03
disconnect.png	238 B	Portable Ne...	2009-03
downloadadd.png	174 B	Portable Ne...	2009-03
file.png	258 B	Portable Ne...	2009-03
filezilla.png	477 B	Portable Ne...	2009-03

Selected 1 file. Total size: 174 B

```
Server/Local file                                  Direction  Remote file
filezilla@127.0.0.1
C:\dev\svn\FileZilla3\src\bin\FileZilla_unicode_dbg.exe  -->  /FileZilla_unicode_dbg.exe
     00:00:13 elapsed   00:00:19 left   39.7%             3.473.408 bytes (267.1 KB/s)
C:\dev\svn\FileZilla3\autom4te.cache\output.2           -->  /FileZilla3/autom4te.cache/outp...  633,8 KiB  Normal  Transferring
     00:00:01 elapsed   00:00:01 left   40.3%             262.144 bytes (262.1 KB/s)
C:\dev\svn\FileZilla3\autom4te.cache\requests           -->  /FileZilla3/autom4te.cache/requ...  24,0 KiB   Normal
```

Queue: 558 MiB

Other protocols too, but let's start with these.

Connecting from Different Platforms

<table>
<tr><th></th><th>SSH</th><th>X-Win</th><th>SFTP</th></tr>
<tr><td><b>Microsoft Windows</b></td><td colspan="3"><b>MobaXterm</b>
https://mobaxterm.mobatek.net</td></tr>
<tr><td><b>Apple OS X</b></td><td><b>Terminal</b>
(Built in)</td><td><b>XQuartz</b>
https://www.xquartz.org</td><td><b>Cyberduck</b>
https://cyberduck.io</td></tr>
<tr><td><b>Linux</b></td><td><b>Terminal</b>
(Built in)</td><td><b>X11</b>
(Built in)</td><td><b>Various</b>
(Built in)</td></tr>
</table>

SCC Help: http://www.bu.edu/tech/support/research/system-usage/getting-started

Hands-On: SSH Connection

Have a BU Account?

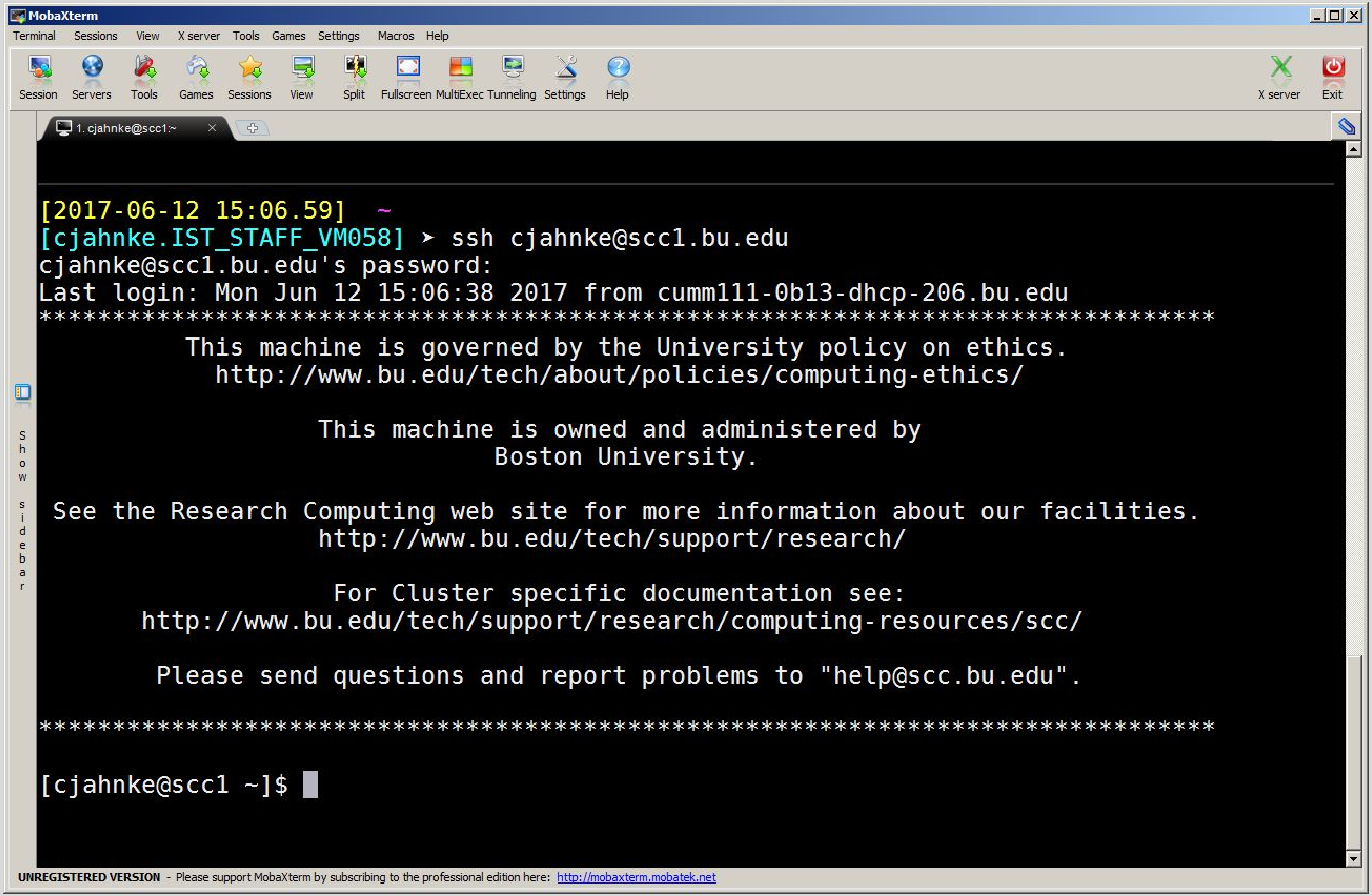
- Hostname: scc1.bu.edu
- Username: <Your BU Username>
- Password: <Your BU Password>

Use Tutorial Account

- Hostname: scc1.bu.edu
- Username: tuta30+
- Password:

```
[local_prompt]$ ssh username@scc1.bu.edu
```

SCC Help: http://www.bu.edu/tech/support/research/system-usage/getting-started/connect-ssh



```
[2017-06-12 15:06.59]  ~
[cjahnke.IST_STAFF_VM058] ➤ ssh cjahnke@scc1.bu.edu
cjahnke@scc1.bu.edu's password:
Last login: Mon Jun 12 15:06:38 2017 from cumm111-0b13-dhcp-206.bu.edu
********************************************************************************
          This machine is governed by the University policy on ethics.
            http://www.bu.edu/tech/about/policies/computing-ethics/

                   This machine is owned and administered by
                               Boston University.

 See the Research Computing web site for more information about our facilities.
                   http://www.bu.edu/tech/support/research/

                    For Cluster specific documentation see:
       http://www.bu.edu/tech/support/research/computing-resources/scc/

        Please send questions and report problems to "help@scc.bu.edu".

********************************************************************************

[cjahnke@scc1 ~]$ 
```

Hands-On: X-Windows Application

X-Windows (X-Win, X11, etc) allows graphics to come through the SSH session.

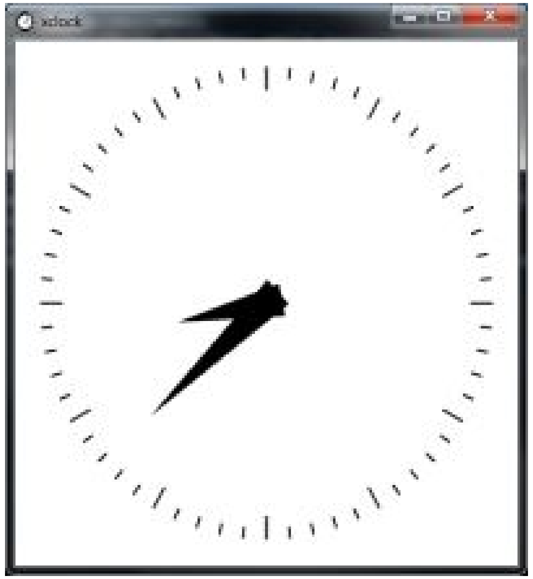
- MobaXterm does not need to do this. Just type “**`xclock`**”

- **Windows and Linux and**: Use the “**`-X`**” option to enable X-Win

```
[local_prompt]$ ssh -X username@scc1.bu.edu
username@scc4.bu.edu’s Password:
[username@scc1 ~]$ xclock &
```

- **Apple OS X:** Use the “**`-Y`**” option to enable X-Win

```
[local_prompt]$ ssh -Y username@scc1.bu.edu
username@scc4.bu.edu’s Password:
[username@scc1 ~]$ xclock &
```



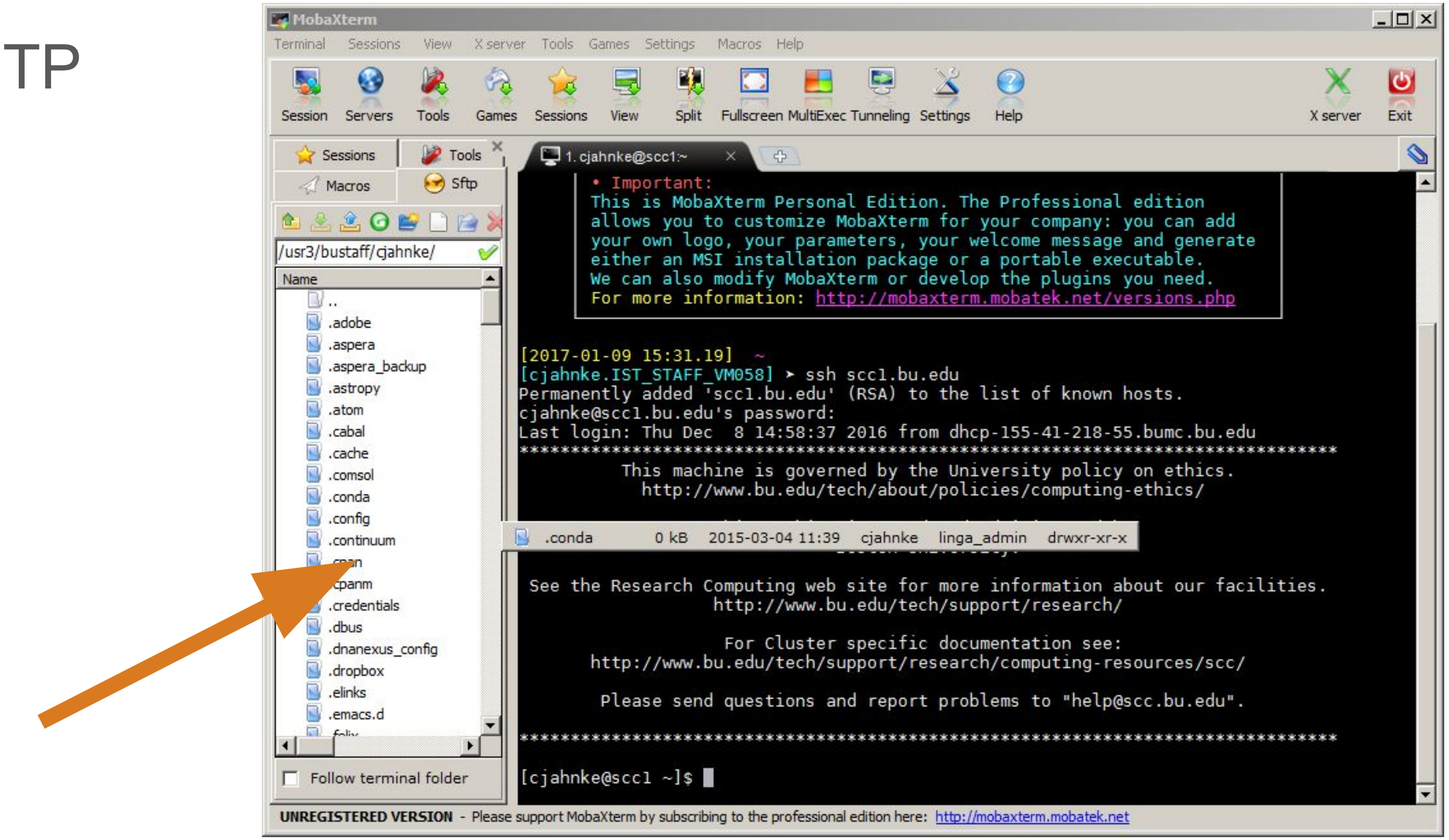
SCC Help: http://www.bu.edu/tech/support/research/system-usage/getting-started/x-forwarding/

Hands-On: SFTP

1. Local System: Download http://rcs.bu.edu/documents/TheJungleBook.txt

2. Connect to SCC via SFTP

3. Drag files into sidebar



```
 • Important:
 This is MobaXterm Personal Edition. The Professional edition
 allows you to customize MobaXterm for your company: you can add
 your own logo, your parameters, your welcome message and generate
 either an MSI installation package or a portable executable.
 We can also modify MobaXterm or develop the plugins you need.
 For more information: http://mobaxterm.mobatek.net/versions.php

[2017-01-09 15:31.19]  ~
[cjahnke.IST_STAFF_VM058] ➤ ssh scc1.bu.edu
Permanently added 'scc1.bu.edu' (RSA) to the list of known hosts.
cjahnke@scc1.bu.edu's password:
Last login: Thu Dec  8 14:58:37 2016 from dhcp-155-41-218-55.bumc.bu.edu
********************************************************************************
        This machine is governed by the University policy on ethics.
          http://www.bu.edu/tech/about/policies/computing-ethics/

 See the Research Computing web site for more information about our facilities.
                 http://www.bu.edu/tech/support/research/

                  For Cluster specific documentation see:
       http://www.bu.edu/tech/support/research/computing-resources/scc/

          Please send questions and report problems to "help@scc.bu.edu".

********************************************************************************
[cjahnke@scc1 ~]$
```

.conda   0 kB   2015-03-04 11:39   cjahnke   linga_admin   drwxr-xr-x

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: http://mobaxterm.mobatek.net

Commands

Basic Linux commands and system use

This one is long and very hands-on.

Linux: The “prompt”

```
[username@scc1 ~]$
```

This is the “prompt” -- the system is ready to accept commands.

`username`	Your login name
`scc1`	The machine you are on
`~`	The directory you are in (In Linux “~” is a shorthand for your home directory.)

Linux: Command Basics

At the prompt, you can issue commands

```
[username@scc1 ~]$ command --option argument
```

- **Command**: command that does one thing
- **Options**: change the way a command does that one thing
  - Short form: single-dash and one letter e.g. `ls -a`
  - Long form: double-dash and a word e.g. `ls --all`
- **Argument**: provides the input/output that the command interacts with.

For more information about any command, use **`man`** or **`info`** (e.g. "**`man ls`**")

Linux: Everyday Commands

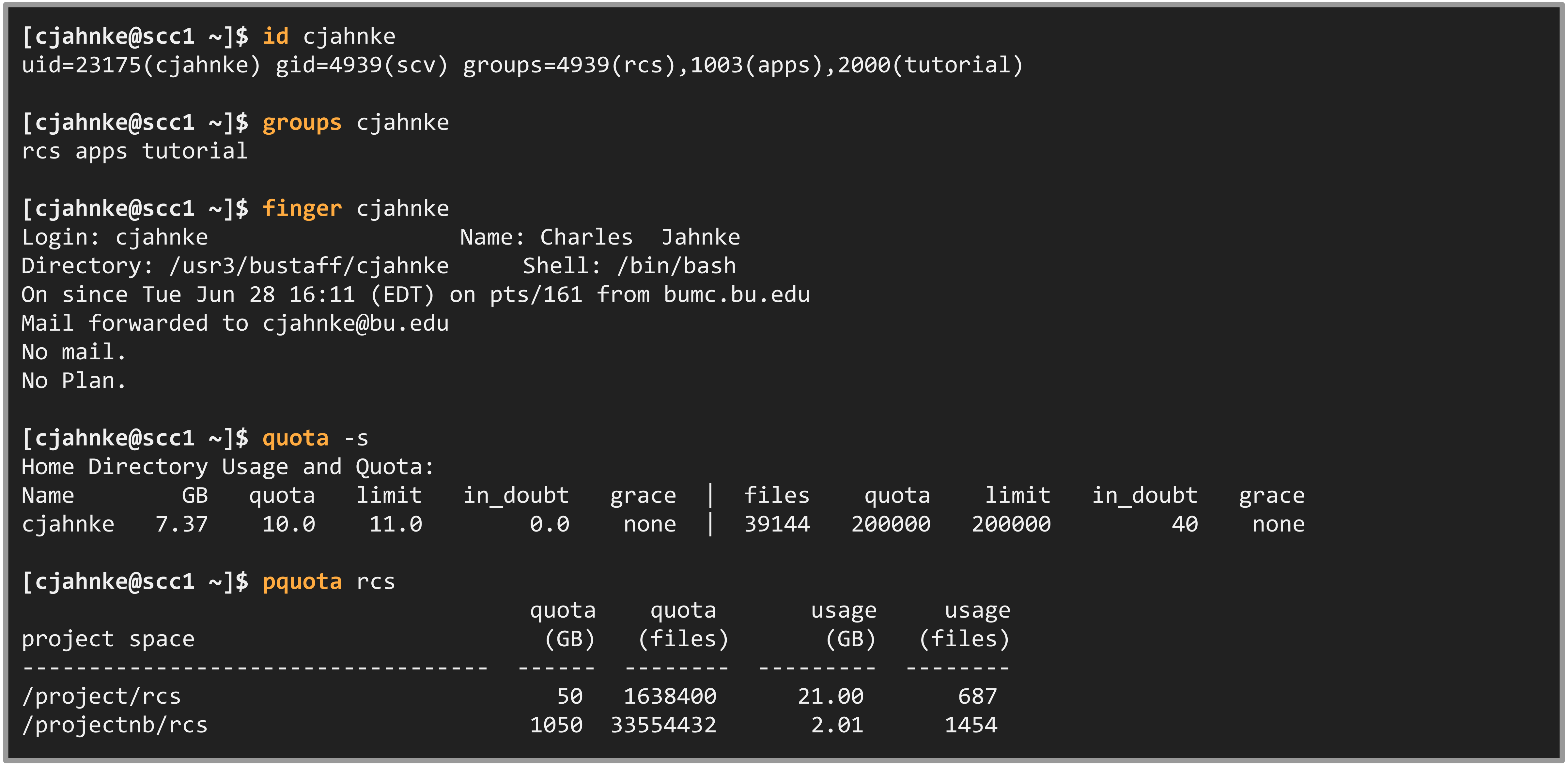
- Refer to the “SCC Getting Started” handout/PDF.

- For time, I’ll only mention the most common/useful commands.

- Please ask questions (this is a lot of information)

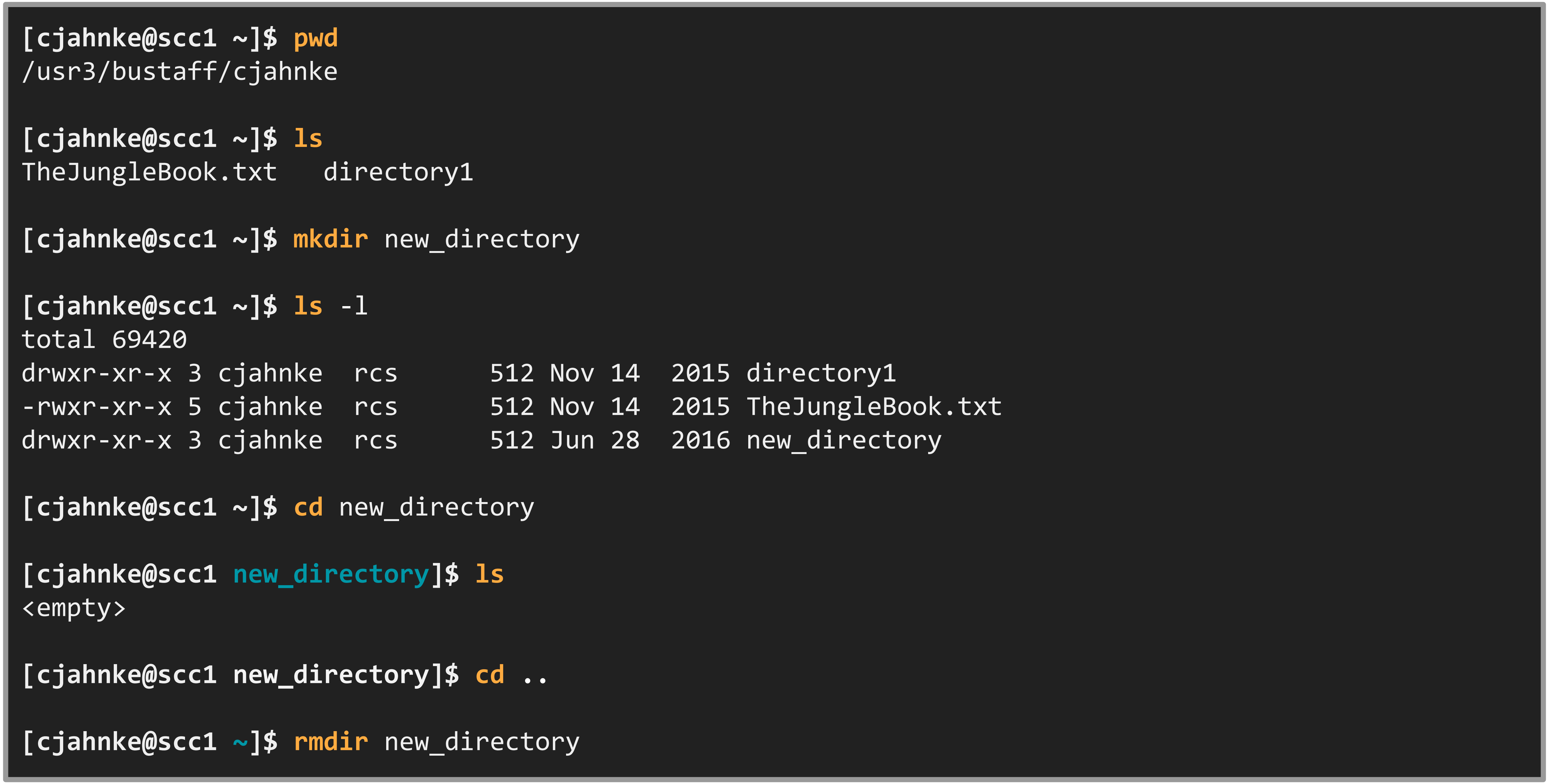
http://rcs.bu.edu/documents/SCC_GettingStarted.pdf

Command Hands-On: Your Account



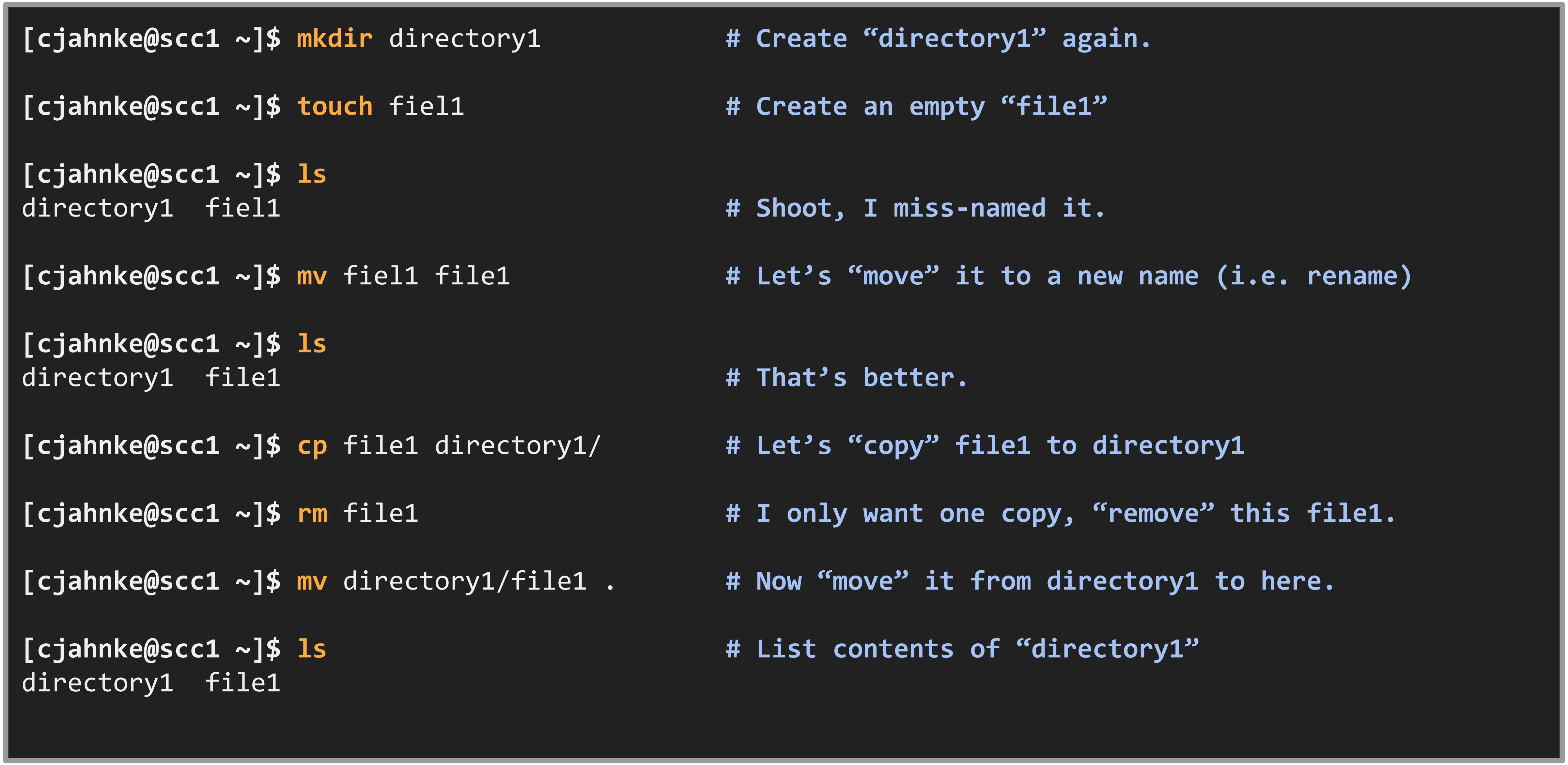
```
[cjahnke@scc1 ~]$ id cjahnke
uid=23175(cjahnke) gid=4939(scv) groups=4939(rcs),1003(apps),2000(tutorial)

[cjahnke@scc1 ~]$ groups cjahnke
rcs apps tutorial

[cjahnke@scc1 ~]$ finger cjahnke
Login: cjahnke                   Name: Charles  Jahnke
Directory: /usr3/bustaff/cjahnke     Shell: /bin/bash
On since Tue Jun 28 16:11 (EDT) on pts/161 from bumc.bu.edu
Mail forwarded to cjahnke@bu.edu
No mail.
No Plan.

[cjahnke@scc1 ~]$ quota -s
Home Directory Usage and Quota:
Name        GB   quota   limit   in_doubt   grace  |  files    quota    limit   in_doubt   grace
cjahnke   7.37    10.0    11.0        0.0    none  |  39144   200000   200000         40    none

[cjahnke@scc1 ~]$ pquota rcs
                                      quota    quota       usage     usage
project space                          (GB)   (files)       (GB)   (files)
-----------------------------------  ------  --------  ---------  --------
/project/rcs                            50   1638400      21.00       687
/projectnb/rcs                        1050  33554432       2.01      1454
```

Command Hands-On: Directory Navigation



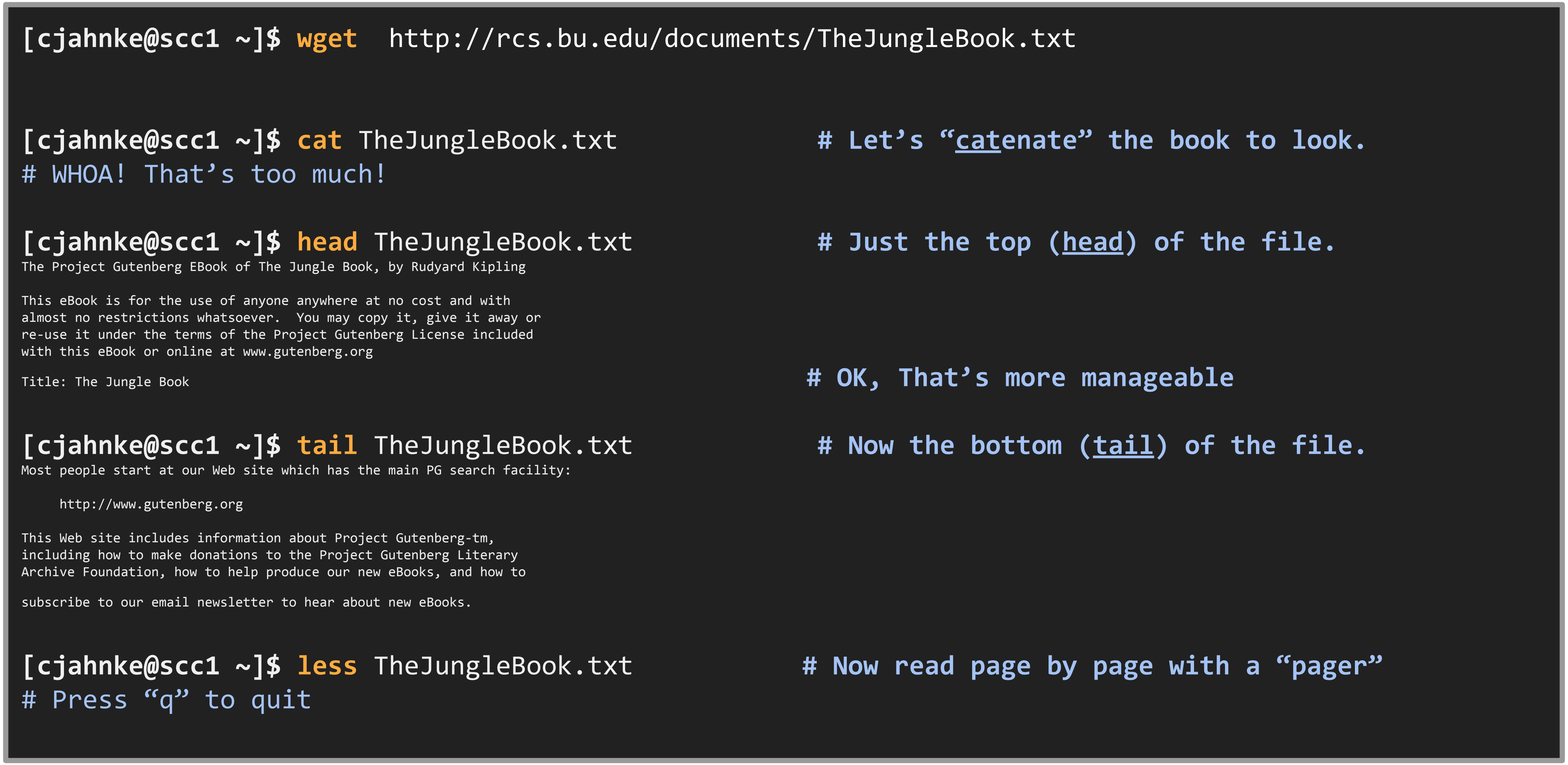
```
[cjahnke@scc1 ~]$ pwd
/usr3/bustaff/cjahnke

[cjahnke@scc1 ~]$ ls
TheJungleBook.txt   directory1

[cjahnke@scc1 ~]$ mkdir new_directory

[cjahnke@scc1 ~]$ ls -l
total 69420
drwxr-xr-x 3 cjahnke  rcs      512 Nov 14  2015 directory1
-rwxr-xr-x 5 cjahnke  rcs      512 Nov 14  2015 TheJungleBook.txt
drwxr-xr-x 3 cjahnke  rcs      512 Jun 28  2016 new_directory

[cjahnke@scc1 ~]$ cd new_directory

[cjahnke@scc1 new_directory]$ ls
<empty>

[cjahnke@scc1 new_directory]$ cd ..

[cjahnke@scc1 ~]$ rmdir new_directory
```

Command Hands-On: File Management



```
[cjahnke@scc1 ~]$ mkdir directory1            # Create “directory1” again.

[cjahnke@scc1 ~]$ touch fiel1                 # Create an empty “file1”

[cjahnke@scc1 ~]$ ls
directory1  fiel1                             # Shoot, I miss-named it.

[cjahnke@scc1 ~]$ mv fiel1 file1              # Let’s “move” it to a new name (i.e. rename)

[cjahnke@scc1 ~]$ ls
directory1  file1                             # That’s better.

[cjahnke@scc1 ~]$ cp file1 directory1/        # Let’s “copy” file1 to directory1

[cjahnke@scc1 ~]$ rm file1                    # I only want one copy, “remove” this file1.

[cjahnke@scc1 ~]$ mv directory1/file1 .       # Now “move” it from directory1 to here.

[cjahnke@scc1 ~]$ ls                          # List contents of “directory1”
directory1  file1
```

Command Hands-On: Viewing Text Files



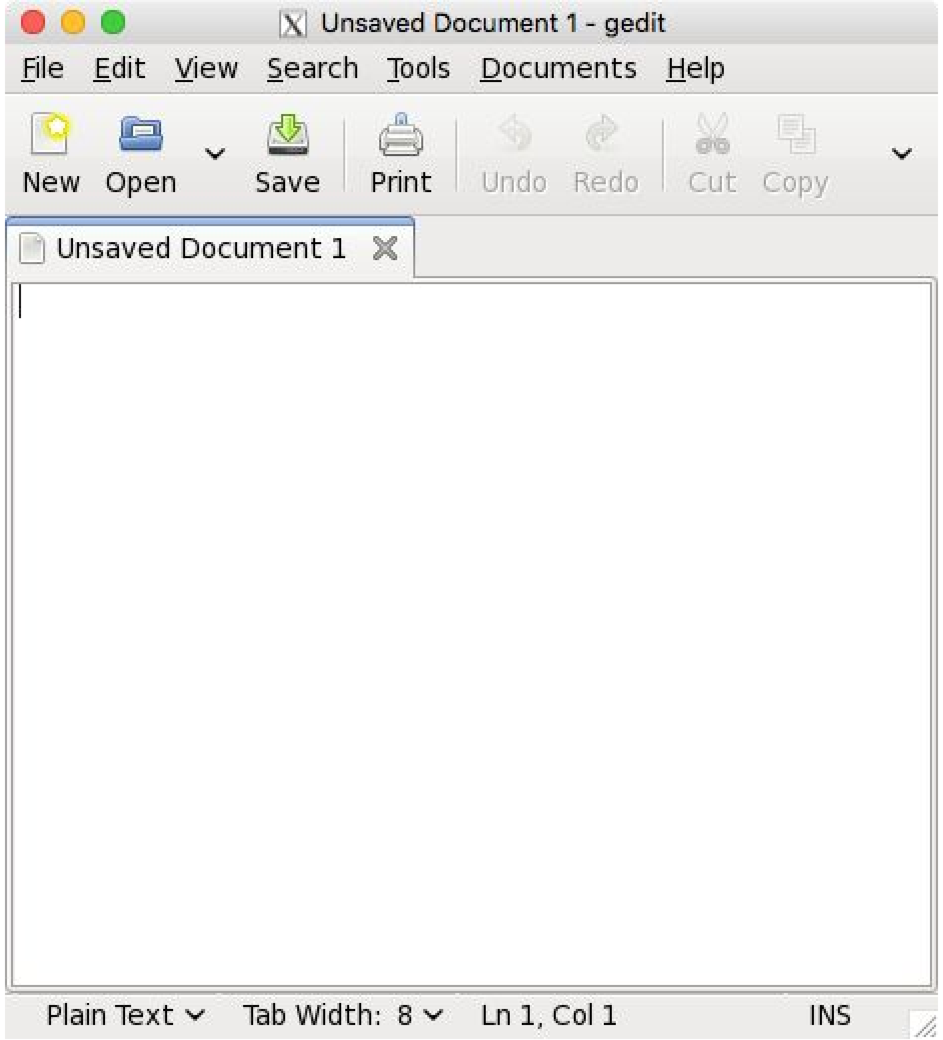
```
[cjahnke@scc1 ~]$ wget  http://rcs.bu.edu/documents/TheJungleBook.txt


[cjahnke@scc1 ~]$ cat TheJungleBook.txt              # Let’s “catenate” the book to look.
# WHOA! That’s too much!

[cjahnke@scc1 ~]$ head TheJungleBook.txt             # Just the top (head) of the file.
The Project Gutenberg EBook of The Jungle Book, by Rudyard Kipling

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever.  You may copy it, give it away or
re-use it under the terms of the Project Gutenberg License included
with this eBook or online at www.gutenberg.org

Title: The Jungle Book                              # OK, That’s more manageable


[cjahnke@scc1 ~]$ tail TheJungleBook.txt             # Now the bottom (tail) of the file.
Most people start at our Web site which has the main PG search facility:

     http://www.gutenberg.org

This Web site includes information about Project Gutenberg-tm,
including how to make donations to the Project Gutenberg Literary
Archive Foundation, how to help produce our new eBooks, and how to

subscribe to our email newsletter to hear about new eBooks.


[cjahnke@scc1 ~]$ less TheJungleBook.txt             # Now read page by page with a “pager”
# Press “q” to quit
```

Command Hands-On: Finding Things

```
# Search the directory for files:
[cjahnke@scc1 ~]$ find . -name "TheJungleBook.txt"
./TheJungleBook.txt



# Search a file for text:
[cjahnke@scc1 ~]$ grep Mowgli TheJungleBook.txt
     Mowgli's Brothers
     Mowgli's Song
Mowgli's Brothers
frog. O thou Mowgli--for Mowgli the Frog I will call thee--the time will
night of the Pack Meeting took them and Mowgli and Mother Wolf to the
Wolf pushed "Mowgli the Frog," as they called him, into the center,
...
..
.
# 216 Lines
```

Command Hands-On: Editing Files

```
[cjahnke@scc1 ~]$ gedit
# Normal, foreground process.
# Close the GUI to get prompt.



[cjahnke@scc1 ~]$ gedit &
[cjahnke@scc1 ~]$
# Background process.
```

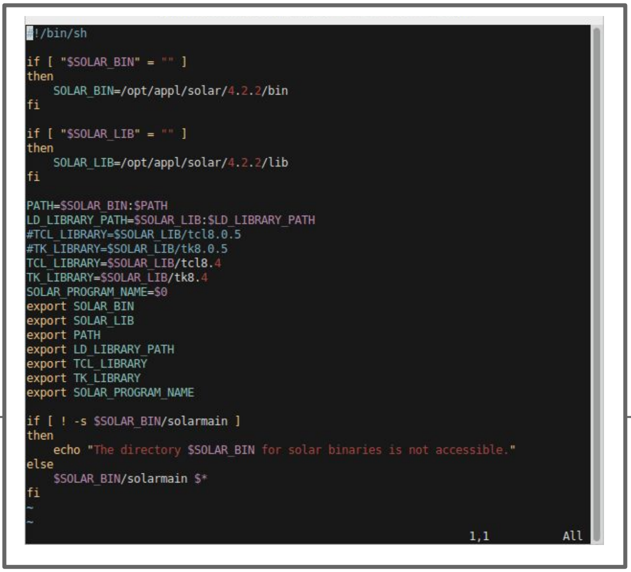


Plain Text   Tab Width: 8   Ln 1, Col 1   INS

Command Hands-On: Editing Text Files

Command Interface

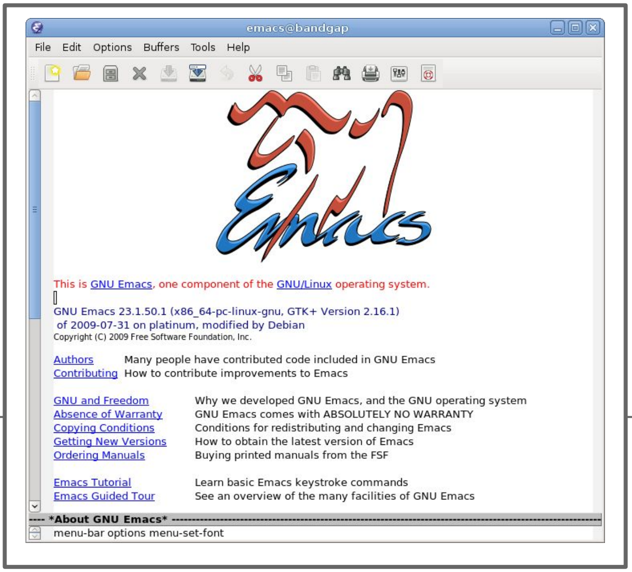
- **nano** - “Nano's ANOther” editor
- **emacs** - Programming Editor
- **vim** / **vi** - Visual IMproved
- Others



```
#!/bin/sh

if [ "$SOLAR_BIN" = "" ]
then
    SOLAR_BIN=/opt/appl/solar/4.2.2/bin
fi

if [ "$SOLAR_LIB" = "" ]
then
    SOLAR_LIB=/opt/appl/solar/4.2.2/lib
fi

PATH=$SOLAR_BIN:$PATH
LD_LIBRARY_PATH=$SOLAR_LIB:$LD_LIBRARY_PATH
#TCL_LIBRARY=$SOLAR_LIB/tcl8.0.5
#TK_LIBRARY=$SOLAR_LIB/tk8.0.5
TCL_LIBRARY=$SOLAR_LIB/tcl8.4
TK_LIBRARY=$SOLAR_LIB/tk8.4
SOLAR_PROGRAM_NAME=$0
export SOLAR_BIN
export SOLAR_LIB
export PATH
export LD_LIBRARY_PATH
export TCL_LIBRARY
export TK_LIBRARY
export SOLAR_PROGRAM_NAME

if [ ! -s $SOLAR_BIN/solarmain ]
then
    echo "The directory $SOLAR_BIN for solar binaries is not accessible."
else
    $SOLAR_BIN/solarmain $*
fi
~
~
                                                        1,1           All
```

Graphical Interface

- **gedit** - Gnome EDITor
- **emacs** - Programming Editor
- **gvim** - GUI VIM
- Others



This is GNU Emacs, one component of the GNU/Linux operating system.

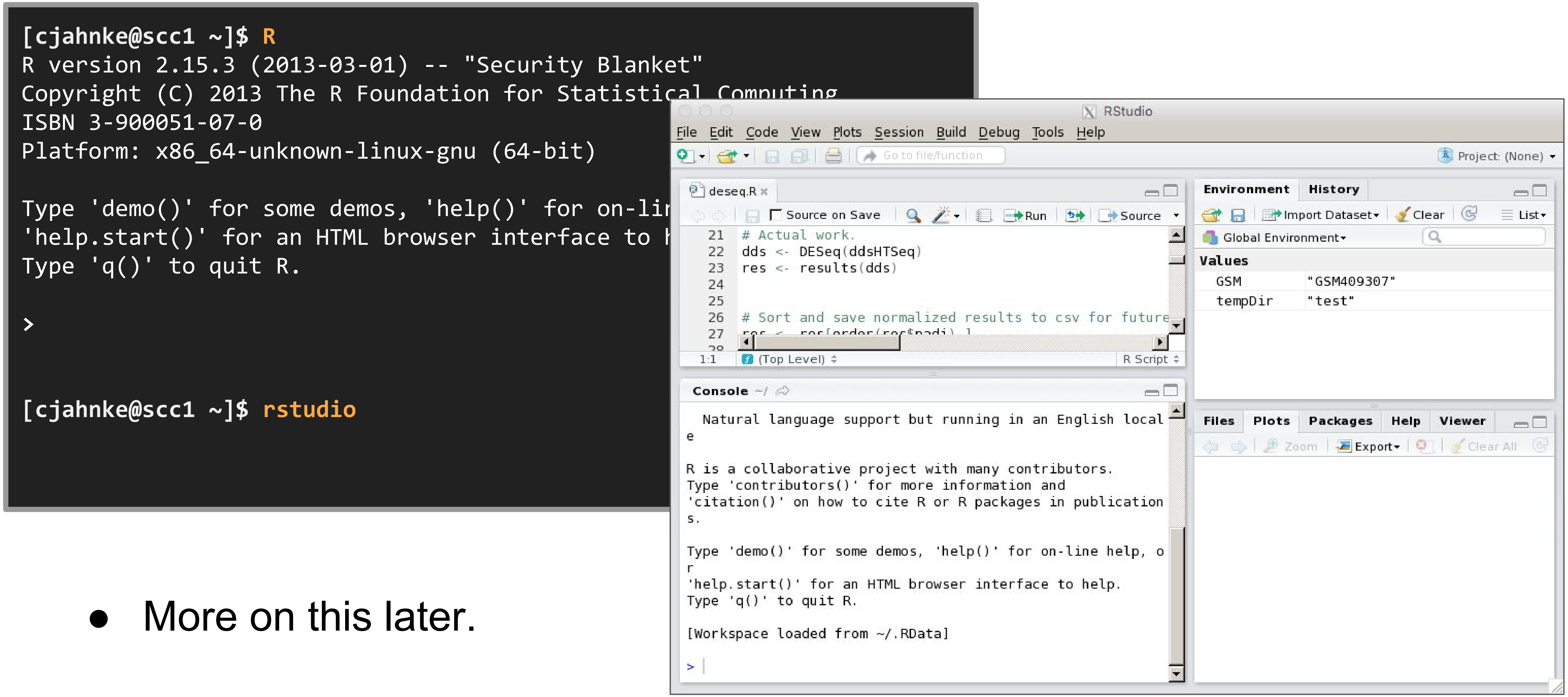
GNU Emacs 23.1.50.1 (x86_64-pc-linux-gnu, GTK+ Version 2.16.1)
of 2009-07-31 on platinum, modified by Debian
Copyright (C) 2009 Free Software Foundation, Inc.

Authors	Many people have contributed code included in GNU Emacs
Contributing	How to contribute improvements to Emacs
GNU and Freedom	Why we developed GNU Emacs, and the GNU operating system
Absence of Warranty	GNU Emacs comes with ABSOLUTELY NO WARRANTY
Copying Conditions	Conditions for redistributing and changing Emacs
Getting New Versions	How to obtain the latest version of Emacs
Ordering Manuals	Buying printed manuals from the FSF
Emacs Tutorial	Learn basic Emacs keystroke commands
Emacs Guided Tour	See an overview of the many facilities of GNU Emacs

---- *About GNU Emacs* ----

menu-bar options menu-set-font

Command Hands-On: Applications



```
[cjahnke@scc1 ~]$ R
R version 2.15.3 (2013-03-01) -- "Security Blanket"
Copyright (C) 2013 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-unknown-linux-gnu (64-bit)

Type 'demo()' for some demos, 'help()' for on-lir
'help.start()' for an HTML browser interface to h
Type 'q()' to quit R.

>


[cjahnke@scc1 ~]$ rstudio
```

```
# Actual work.
dds <- DESeq(ddsHTSeq)
res <- results(dds)


# Sort and save normalized results to csv for future
res <- res[order(res$padj), ]
```

```
GSM      "GSM409307"
tempDir  "test"
```

```
 Natural language support but running in an English local
e

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publication
s.

Type 'demo()' for some demos, 'help()' for on-line help, o
r
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

>
```

- More on this later.

Command Hands-On: Transferring Remote Files

- Transfer files from your local system (laptop)
  - Use the SFTP application you downloaded earlier.

- Transfer files from remote Linux system

```
[cjahnke@scc1 ~]$ scp username@remotehost.com:/path/to/file .

[cjahnke@scc1 ~]$ rsync -a username@remotehost.com:/path/to/file .
```

- Download files from the internet/webpages

```
[cjahnke@scc1 ~]$ wget http://rcs.bu.edu/documents/sample.vcf
--2016-06-28 18:25:47--  http://rcs.bu.edu/documents/sample.vcf
Connecting to rcs.bu.edu|128.197.160.76|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 298778 (292K) [text/plain]
100%[===============================================>] 298,778     1.79M/s   in 0.2s

2016-09-20 18:25:48 (1.79 MB/s) - `sample.vcf' saved [298778/298778]
```

dos2unix / unix2dos

- Windows and Linux define “end of line” differently
  - Windows: “**\r\n**” (“**^M**”)
  - Linux: “**\n**”

- **`dos2unix`** - DOS to UNIX text file format converter

```
[cjahnke@scc1 ~]$ dos2unix input.txt                    # Convert and replace input.txt
```

```
[cjahnke@scc1 ~]$ dos2unix input.txt output.txt         # write output to new file.
```

Man Page: http://linuxcommand.org/man_pages/dos2unix1.html

Permissions

Users, Groups, and File Ownership

This one is quick, but important.

Users, Groups, and File Ownership

SCC is a Multi-user System

Real users:

- There are many users
- There are many groups
- Users can belong to multiple groups

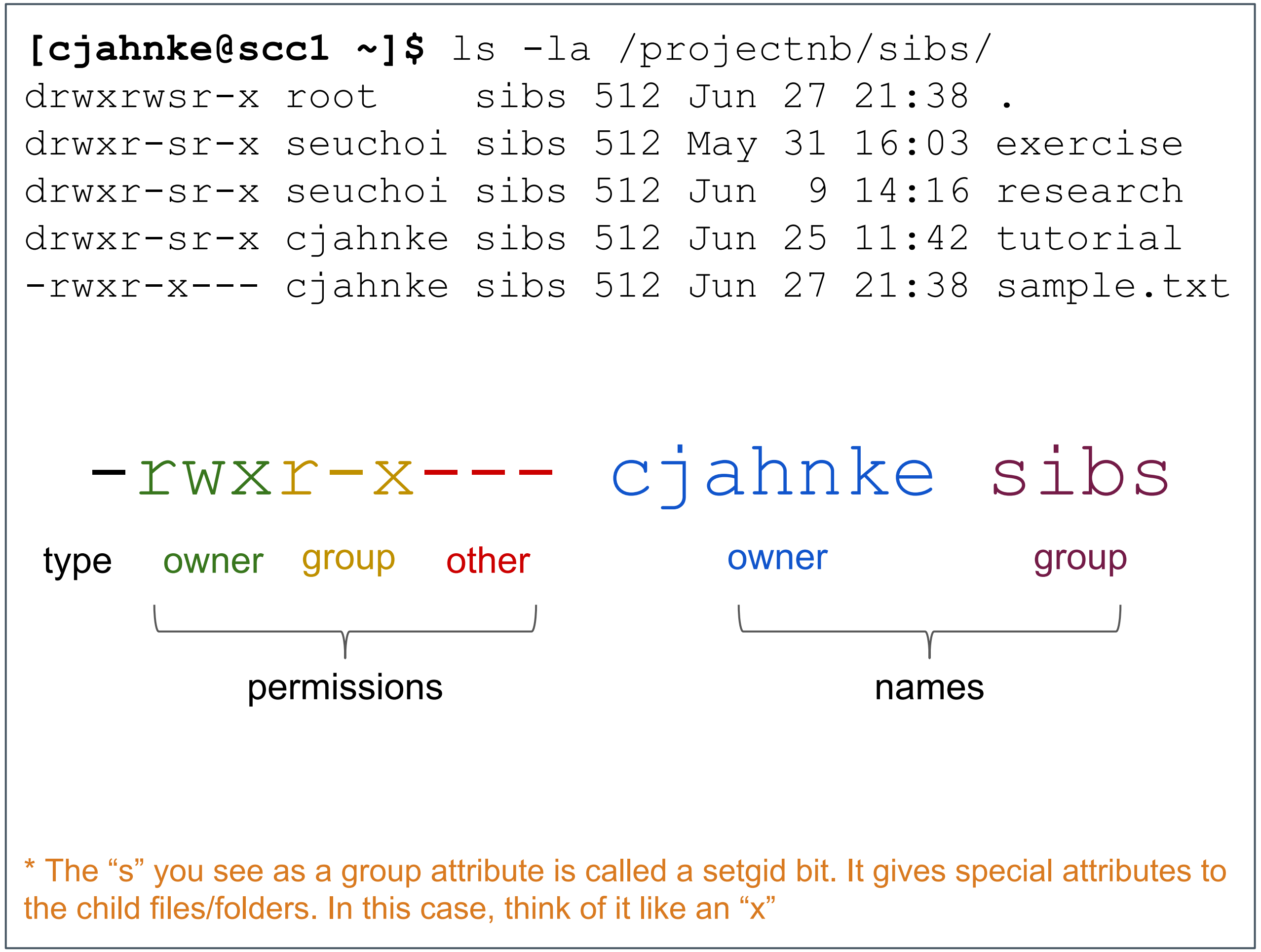
Access control:

- Every file has an owner
- Every file belongs to a group
- Every file has “permissions”

File Access and Permissions

- Types of Access Levels
  - User (owner) “u”
  - Group “g”
  - Others “o”

- Types of Access Modes
  - Read access “r”
  - Write access “w”
  - Execute rights “x”



```
[cjahnke@scc1 ~]$ ls -la /projectnb/sibs/
drwxrwsr-x root    sibs 512 Jun 27 21:38 .
drwxr-sr-x seuchoi sibs 512 May 31 16:03 exercise
drwxr-sr-x seuchoi sibs 512 Jun  9 14:16 research
drwxr-sr-x cjahnke sibs 512 Jun 25 11:42 tutorial
-rwxr-x--- cjahnke sibs 512 Jun 27 21:38 sample.txt
```

* The “s” you see as a group attribute is called a setgid bit. It gives special attributes to the child files/folders. In this case, think of it like an “x”

Changing Ownership

- **chown** - Change file owner and group
  - `chown [OPTION]... [OWNER][:[GROUP]] FILE…`
  - Must have write access to file to make changes.

- Change user ownership of individual file

```
[username@scc1 ~]$ chown cjahnke testfile.txt
```

- Change user and group ownership of file

```
[username@scc1 ~]$ chown cjahnke:rcs testfile.txt
```

See the manual for full description (‘**man chown**’)

Changing Permissions

- **`chmod`** - Change mode (permission) for files
  - `chmod[OPTION]... MODE[,MODE]... FILE...`
  - Must have write access to file to make changes.

- Mode has 2 formats:
  - Octal: base-8 bit representation

```
[username@scc1 ~]$ chmod 750 testfile.txt
```

  - Symbolic: u/g/o, r/w/x, and +/-/= define permissions

```
[username@scc1 ~]$ chmod u+rwx,g+rx,o-r testfile.txt
```

See the manual for full description (‘**`man chmod`**’)

Using the System (Part 3)

Some basic tools, utilities, and methods

Basic Tools, Utilities, and Methods

We’ll cover some basic tools

- Word Count (**`wc`**)
- Column Segmentation (**`cut`**)
- Line Sort (**`sort`**)

Use them to demonstrate methods

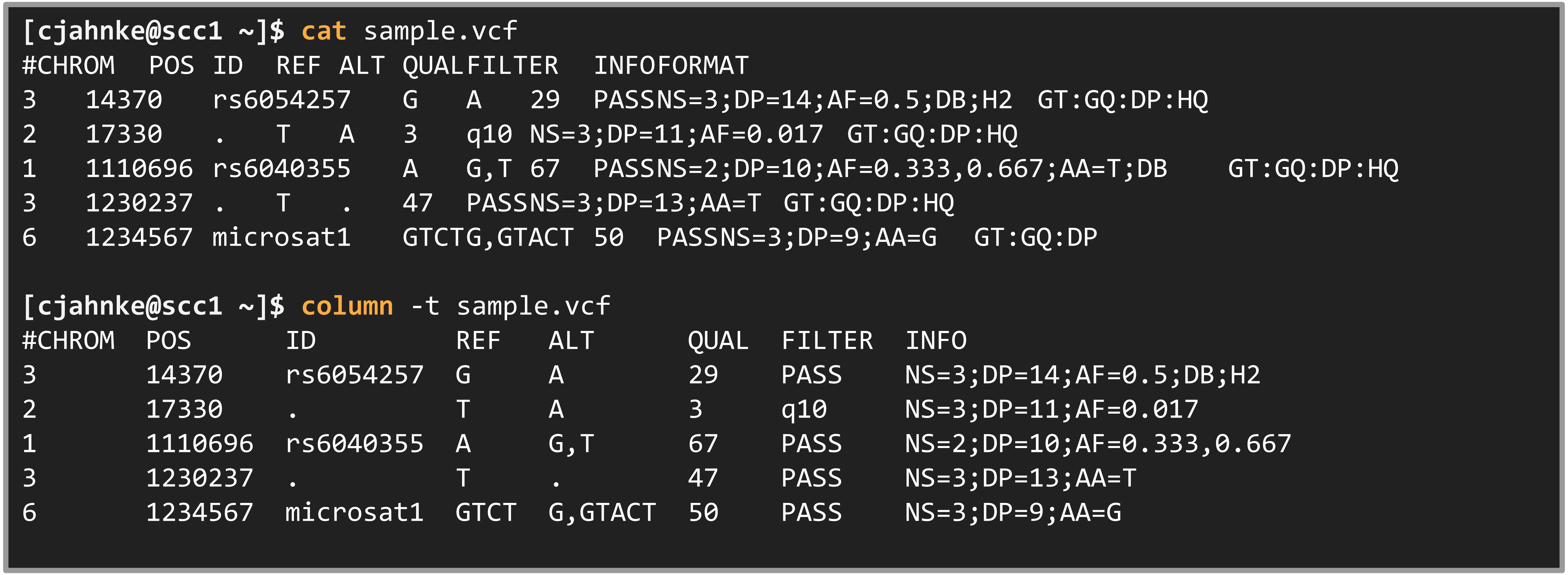
- **Pipes**
  - Command input and output
- **Redirection**
  - File input and output

Download Sample File:

```
[cjahnke@scc1 ~]$ wget http://rcs.bu.edu/documents/sample.vcf
```

Hands-On: New file to work with

Just a few lines from a VCF file, let’s take a look.



```
[cjahnke@scc1 ~]$ cat sample.vcf
#CHROM  POS ID  REF ALT QUALFILTER  INFOFORMAT
3   14370   rs6054257   G   A   29  PASSNS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ
2   17330   .   T   A   3   q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ
1   1110696 rs6040355   A   G,T 67  PASSNS=2;DP=10;AF=0.333,0.667;AA=T;DB   GT:GQ:DP:HQ
3   1230237 .   T   .   47  PASSNS=3;DP=13;AA=T GT:GQ:DP:HQ
6   1234567 microsat1   GTCTG,GTACT 50  PASSNS=3;DP=9;AA=G  GT:GQ:DP

[cjahnke@scc1 ~]$ column -t sample.vcf
#CHROM  POS      ID         REF   ALT      QUAL  FILTER  INFO
3       14370    rs6054257  G     A        29    PASS    NS=3;DP=14;AF=0.5;DB;H2
2       17330    .          T     A        3     q10     NS=3;DP=11;AF=0.017
1       1110696  rs6040355  A     G,T      67    PASS    NS=2;DP=10;AF=0.333,0.667
3       1230237  .          T     .        47    PASS    NS=3;DP=13;AA=T
6       1234567  microsat1  GTCT  G,GTACT  50    PASS    NS=3;DP=9;AA=G
```

Hands-On: Word Count (wc)

- Count Everything

```
[cjahnke@scc1 ~]$ wc sample.vcf
  6  72 573 sample.vcf
```

- Count Lines

```
[cjahnke@scc1 ~]$ wc -l sample.vcf
  6  sample.vcf
```

- Count Words

```
[cjahnke@scc1 ~]$ wc -w sample.vcf
72 sample.vcf
```

Hands-On: Column Segmentation (cut)

- Cut the second column to view the positions

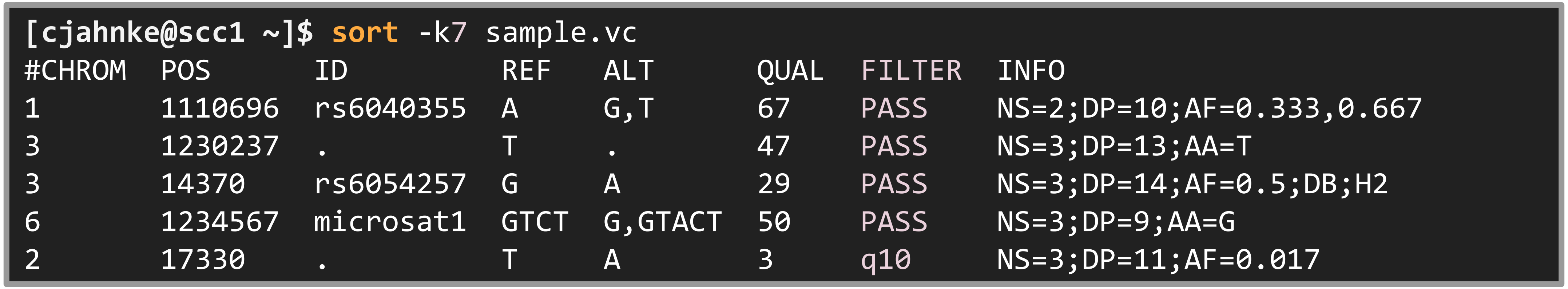
```
[cjahnke@scc1 ~]$ cut -f2 sample.vcf
POS
14370
17330
1110696
1230237
1234567
```

- Cut multiple columns to few position, quality and status

```
[cjahnke@scc1 ~]$ cut -f2,6,7 sample.vcf
POS  QUAL   FILTER
14370    29  PASS
17330    3   q10
1110696  67  PASS
1230237  47  PASS
1234567  50  PASS
```

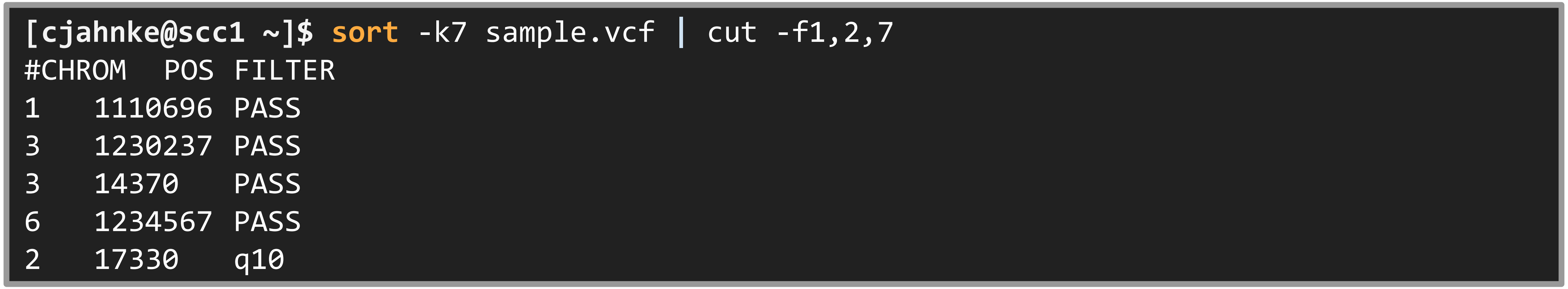
Hands-On: Sort (sort)

- Sort the file by FILTER (key #7)



```
[cjahnke@scc1 ~]$ sort -k7 sample.vc
#CHROM  POS      ID         REF   ALT      QUAL  FILTER  INFO
1       1110696  rs6040355  A     G,T      67    PASS    NS=2;DP=10;AF=0.333,0.667
3       1230237  .          T     .        47    PASS    NS=3;DP=13;AA=T
3       14370    rs6054257  G     A        29    PASS    NS=3;DP=14;AF=0.5;DB;H2
6       1234567  microsat1  GTCT  G,GTACT  50    PASS    NS=3;DP=9;AA=G
2       17330    .          T     A        3     q10     NS=3;DP=11;AF=0.017
```

- Too much info, I only want CHROM and POS for SNPs that passed.

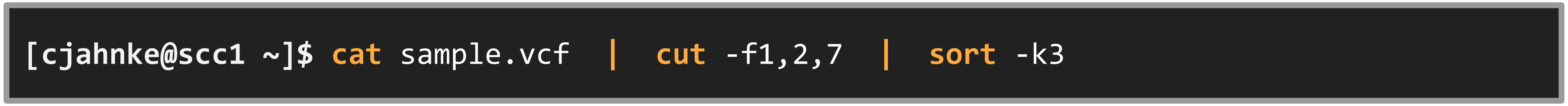


```
[cjahnke@scc1 ~]$ sort -k7 sample.vcf | cut -f1,2,7
#CHROM  POS FILTER
1   1110696 PASS
3   1230237 PASS
3   14370   PASS
6   1234567 PASS
2   17330   q10
```

Pipes

- Pipes (“ | ”) redirect the standard output of a command to the standard input of another command.

- Example:



```
[cjahnke@scc1 ~]$ cat sample.vcf  |  cut -f1,2,7  |  sort -k3
```

```
cat sample.vcf          |      cut -f1,2,7      |      sort -k3
```

```
#CHROM  POS       ID         REF   ...
3       14370     rs6054257  G     ...
2       17330     .          T     ...
1       1110696   rs6040355  A     ...
3       1230237   .          T     ...
6       1234567   microsat1  GTCT  ...
```

```
#CHROM POS     FILTER
3      14370   PASS
2      17330   q10
1      1110696 PASS
3      1230237 PASS
6      1234567 PASS
```

```
#CHROM POS     FILTER
1      1110696 PASS
3      1230237 PASS
3      14370   PASS
6      1234567 PASS
2      17330   q10
```

Redirection

- The ">" symbol redirects the output of a command to a file.

Redirection	Description
`COMMAND < filename`	Input - Directs a file ☆
`COMMAND << stream`	Input - Directs a stream literal
`COMMAND <<< string`	Input - Directs a string
`COMMAND > filename`	Output - Writes output to file (will "clobber") ☆
`COMMAND >> filename`	Output - Appends output to file ☆

- Example:

```
[cjahnke@scc1 ~]$ cat sample.vcf | cut -f1,2,7 | sort -k3 > sorted.txt
```

Variables and Environment Variables

Variables and Environment Variables

- Variables are named storage locations.
    - **`USER=cjahnke`**
    - **`foo=”this is foo’s value”`**

- “Environment variables” are used by the shell to store information
    - For example, **`$PATH`** tells the path where to look for commands.

- Environment variables are <u>shared with programs</u> that the shell runs.

Bash variables

- To create a new variable, use the assignment operator ‘**=**‘

```
[username@scc1 ~]$ foo=“this is foo’s value”
```

- The foo variable can be shown with **echo**

```
[username@scc1 ~]$ echo $foo
this is foo’s value
```

- To make `$foo` visible to programs run by the shell (i.e., make it an “environment variable”), use **export**:

```
[username@scc1 ~]$ export foo
```

Environment Variables

- To see all currently defined environment variable, use **`printenv`**:

```
[username@scc1 ~]$ printenv
HOSTNAME=scc4
TERM=xterm-256color
SHELL=/bin/bash
HISTSIZE=1000
TMPDIR=/scratch
SSH_CLIENT=168.122.9.131 37606 22
SSH_TTY=/dev/pts/191
USER=cjahnke
MAIL=/var/spool/mail/cjahnke
PATH=/usr3/bustaff/cjahnke/apps/bin:/usr/local/apps/pgi-13.5/bin:/usr/java/default/jr
e/bin:/usr/java/default/bin:/usr/lib64/qt-3.3/bin:/usr/local/bin:/bin:/usr/bin:/usr/l
ocal/sbin:/usr/sbin:/sbin:/usr3/bustaff/cjahnke/bin
PWD=/usr3/bustaff/cjahnke/linux-materials
LANG=C
MODULEPATH=/share/module/bioinformatics:/share/module/chemistry
SGE_ROOT=/usr/local/ogs-ge2011.11.p1/sge_root
HOME=/usr3/bustaff/cjahnke
```

Software and Modules

Software (without modules)

- Many tools/utilities are available from the basic system environment

- Some big-name software applications are too:
  - MATLAB
  - SAS
  - STATA

- Others require **Modules**

SCC Help: http://www.bu.edu/tech/support/research/software-and-programming/software-and-applications/modules/

Software (without modules)

- Modules allow users to access *non-standard* tools or *alternative versions* of standard packages.

- This is also an method for customizing your environment as required for certain packages.

- Most software packages on SCC are configured this way.

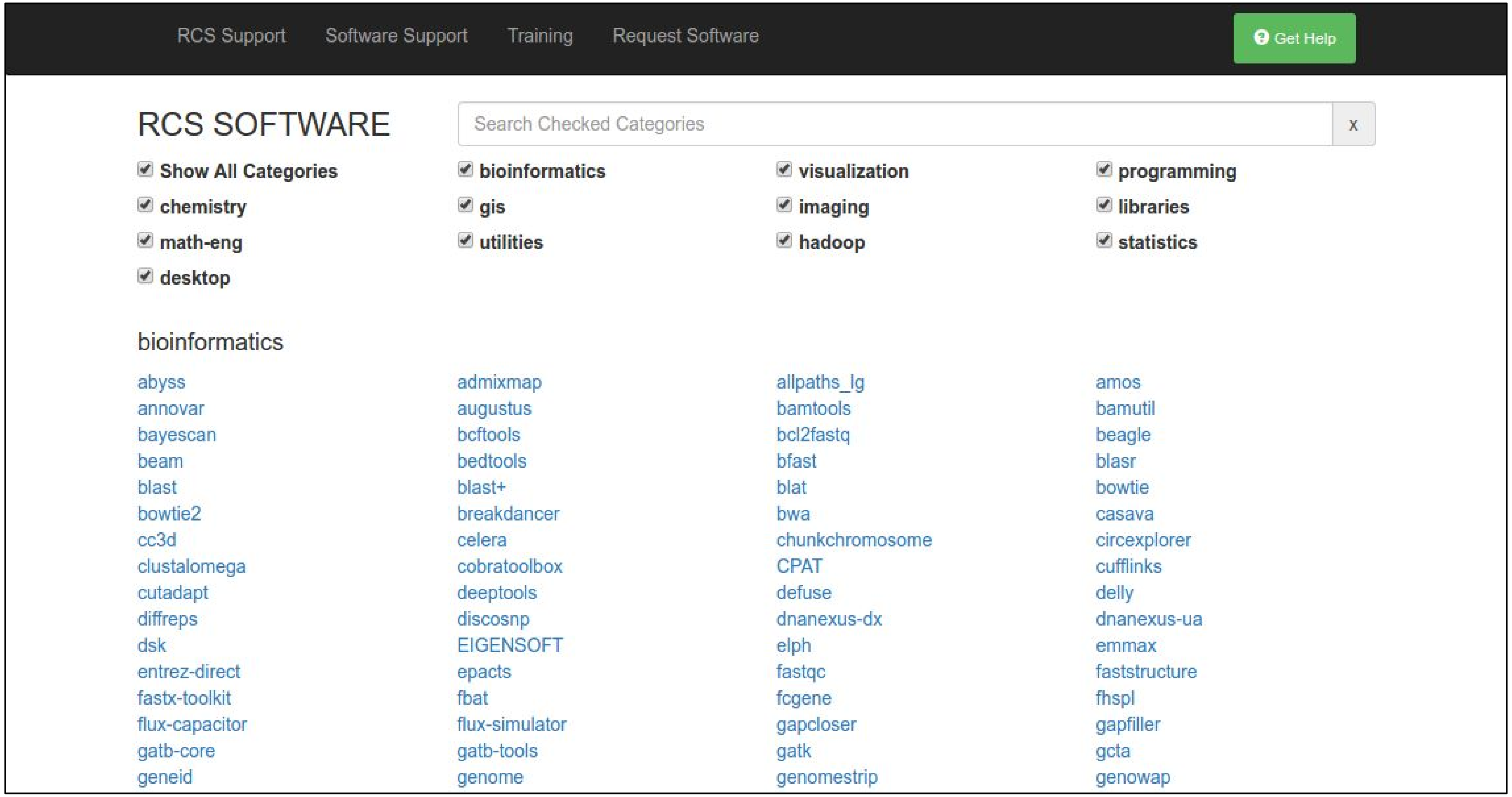
SCC Help: http://www.bu.edu/tech/support/research/software-and-programming/software-and-applications/modules/

Module Usage

Command	Description
module list	List currently loaded modules.
module avail	List available packages.
module help [*modulefile*]	Displays description of specified module.
module show [*modulefile*]	Displays environment modifications for specified module.
module load [*modulefile*]	Loads specified module into environment.
module unload [*modulefile*]	Unloads specified module from environment.
module purge	Unloads all loaded modules.

SCC Help: http://www.bu.edu/tech/support/research/software-and-programming/software-and-applications/modules/

RCS Software Website



RCS SOFTWARE

Search Checked Categories | x

☑ Show All Categories ☑ bioinformatics ☑ visualization ☑ programming

☑ chemistry ☑ gis ☑ imaging ☑ libraries

☑ math-eng ☑ utilities ☑ hadoop ☑ statistics

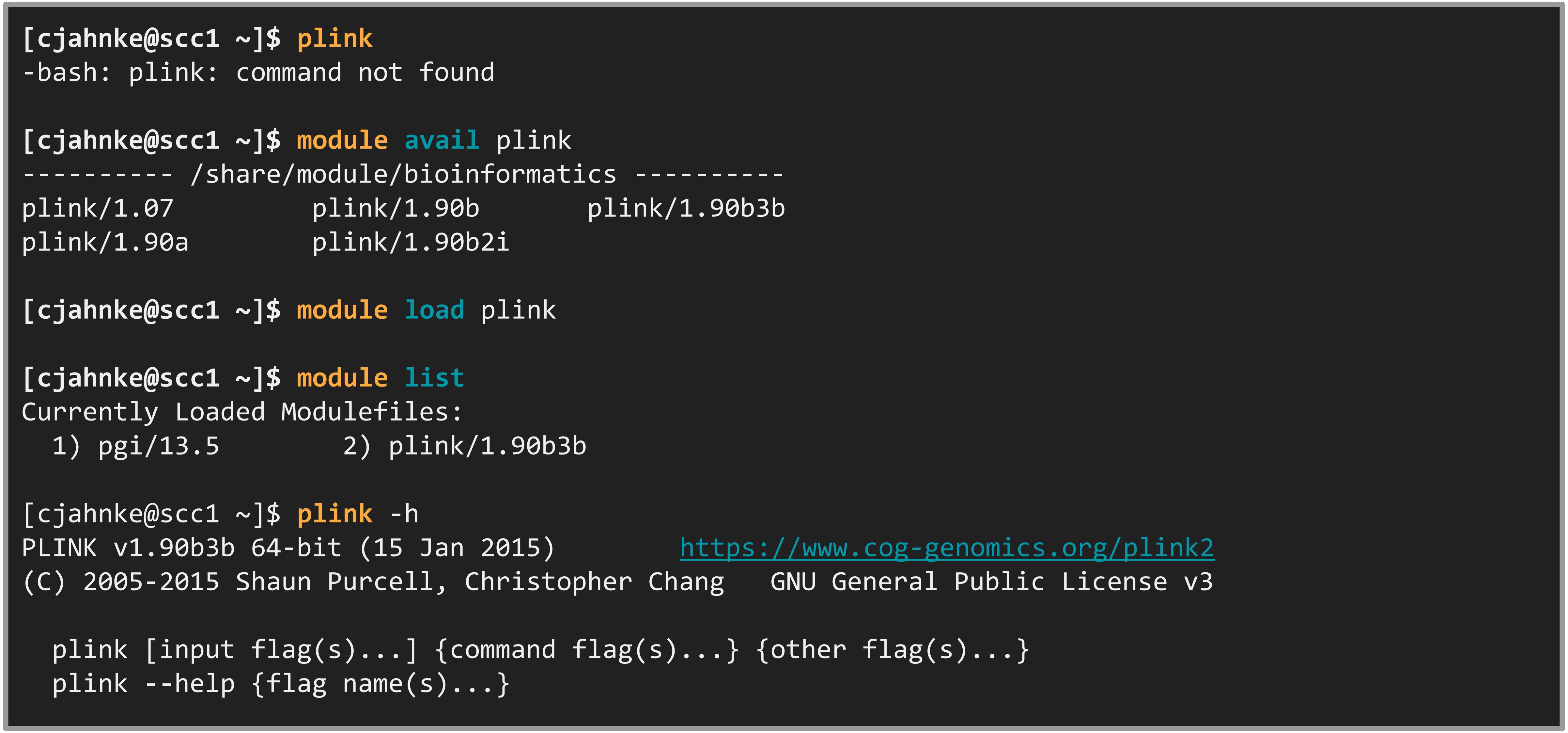
☑ desktop

bioinformatics

abyss	admixmap	allpaths_lg	amos
annovar	augustus	bamtools	bamutil
bayescan	bcftools	bcl2fastq	beagle
beam	bedtools	bfast	blasr
blast	blast+	blat	bowtie
bowtie2	breakdancer	bwa	casava
cc3d	celera	chunkchromosome	circexplorer
clustalomega	cobratoolbox	CPAT	cufflinks
cutadapt	deeptools	defuse	delly
diffreps	discosnp	dnanexus-dx	dnanexus-ua
dsk	EIGENSOFT	elph	emmax
entrez-direct	epacts	fastqc	faststructure
fastx-toolkit	fbat	fcgene	fhspl
flux-capacitor	flux-simulator	gapcloser	gapfiller
gatb-core	gatb-tools	gatk	gcta
geneid	genome	genomestrip	genowap

Software Website: http://rcs.bu.edu/software/#/

Module Hands-On



```
[cjahnke@scc1 ~]$ plink
-bash: plink: command not found

[cjahnke@scc1 ~]$ module avail plink
---------- /share/module/bioinformatics ----------
plink/1.07         plink/1.90b       plink/1.90b3b
plink/1.90a        plink/1.90b2i

[cjahnke@scc1 ~]$ module load plink

[cjahnke@scc1 ~]$ module list
Currently Loaded Modulefiles:
  1) pgi/13.5        2) plink/1.90b3b

[cjahnke@scc1 ~]$ plink -h
PLINK v1.90b3b 64-bit (15 Jan 2015)        https://www.cog-genomics.org/plink2
(C) 2005-2015 Shaun Purcell, Christopher Chang   GNU General Public License v3

  plink [input flag(s)...] {command flag(s)...} {other flag(s)...}
  plink --help {flag name(s)...}
```

New Applications and Requests

- New packages are developed every day.

- Users can compile/install packages for personal use in home directories and project spaces.

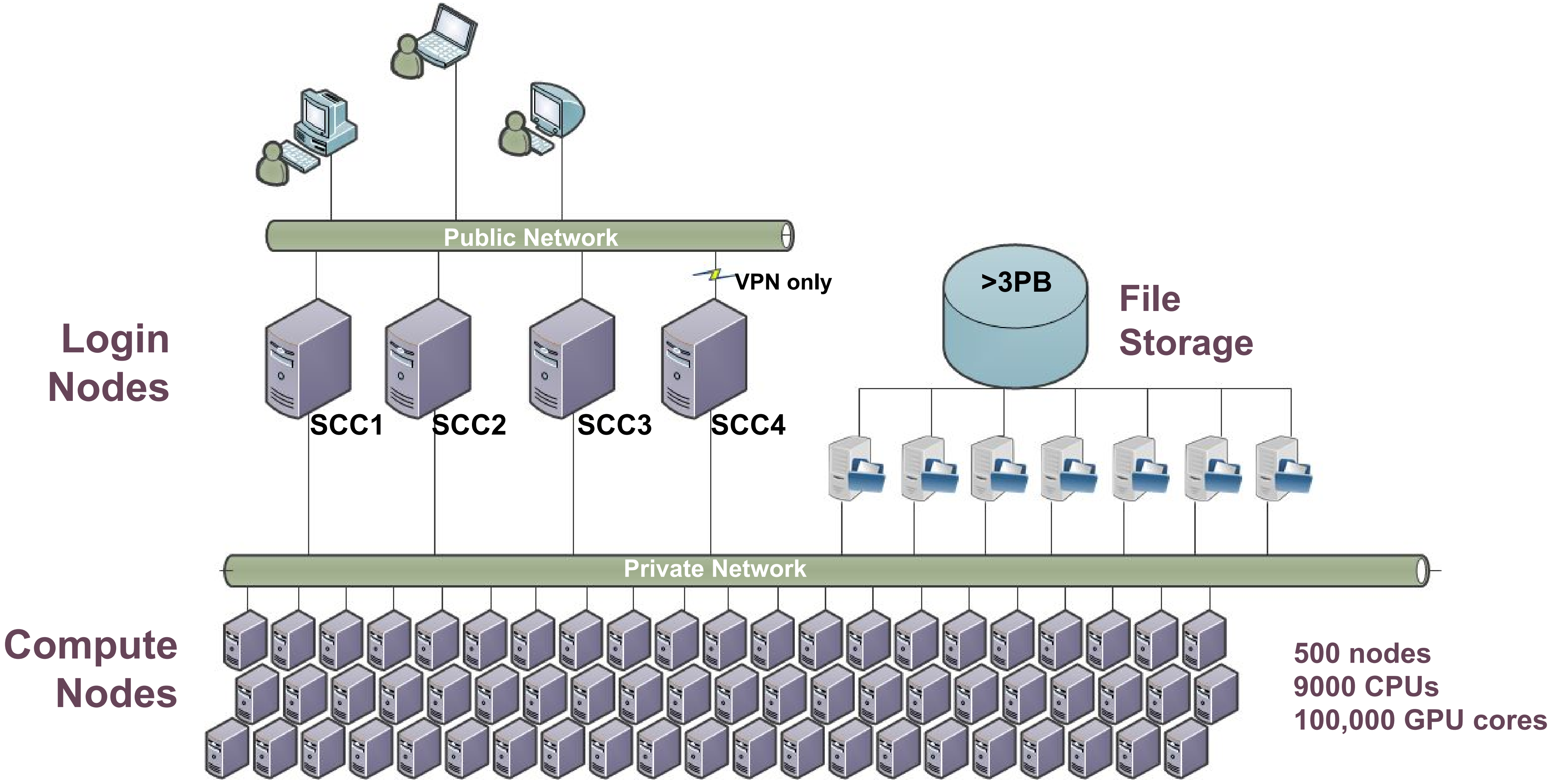
- Users can request global installation of software:
  - Complete form on our website (Link below)
  - send an email to help@scc.bu.edu

SCC Help: http://www.bu.edu/tech/support/research/software-and-programming/software-and-applications/request-software

The Shared Computing Cluster (SCC)

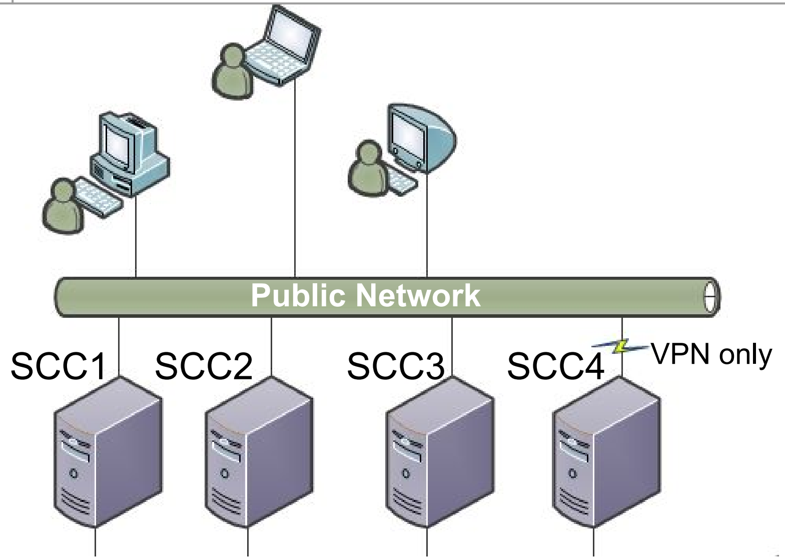
Cluster Architecture, File Storage and the Batch System

SCC Architecture



Choose a Login Node

Login Node	Hostname	Description
SCC1	scc1.bu.edu	General purpose login node, accessible from internet
SCC2	scc2.bu.edu	General purpose login node, accessible from internet
SCC3	geo.bu.edu	Earth and Environment department node.
➡ SCC4	scc4.bu.edu	BUMC login node. Access to /restricted/project data. Requires BU network or VPN.



File Storage on SCC

Storage Locations on SCC

- More than just your home directory!

Location	Backed-up Internally	Disaster Recovery
Home Directory: `~`	Yes - 180 Days	Yes
Project Space: `/project/{projectname}`	Yes - 180 Days	Yes
Project “NB” Space: `/projectnb/{projectname}`	Yes - 30 Days	No

- And some special cases too

Location	Backed-up Internally	Disaster Recovery
STASH `/stash/{projectname}`	Yes - 30 Days	Yes
Archive `/archive/{type}/{projectname}`	N/A	Conditional ($)

Restricted Data

Some data requires dbGaP compliance or other restrictions.

- Policies for “project” and “projectnb” in previous slides is replicated for the /restricted filesystem.
- Only accessible through scc4.bu.edu and compute nodes

Restricted Space		Description
Restricted Project	`/restricted/project/{projectname}`	/project/ space equivalent for restricted data
Restricted ProjectNB	`/restricted/projectnb/{projectname}`	/projectnb/ space equivalent for restricted data

The Batch System

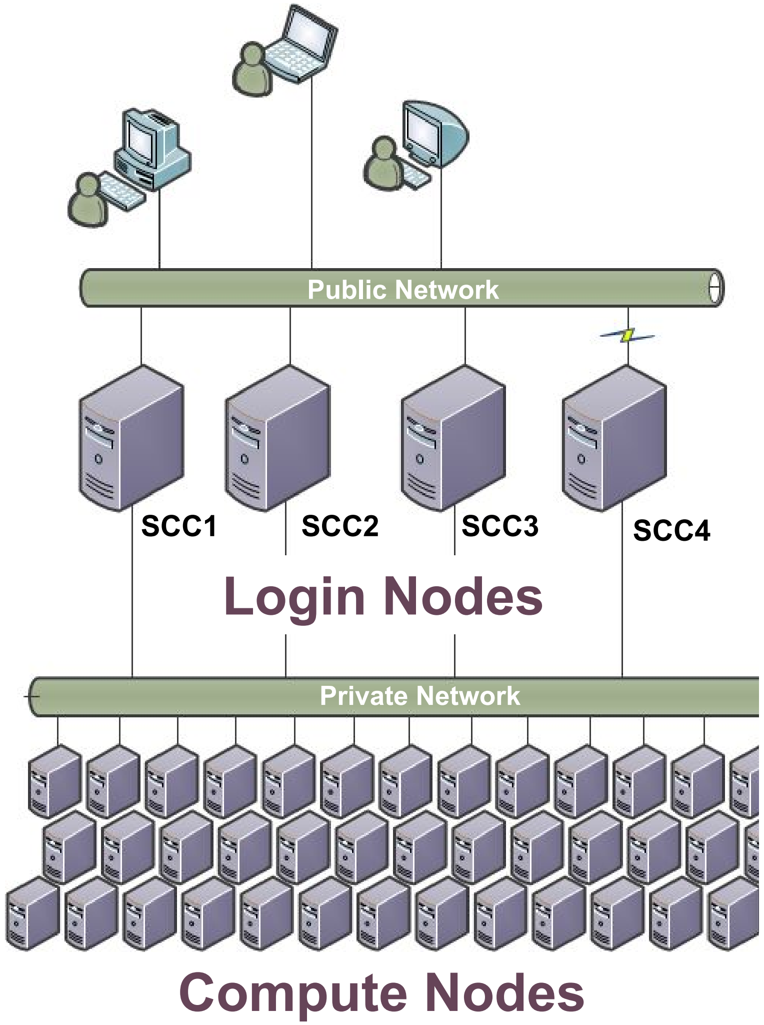
Submitting and Monitoring Batch Jobs

Batch System Overview

- Login nodes are busy!
  - Limited resource
  - Limited runtime (15 min)

- Compute Nodes provide reserved resources
  - Many more nodes
  - Many types of resources

- “Fair Share” scheduling



SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs

Types of Jobs

Interactive

- Just like the login node
- Can type, view output, open files, run commands
- “Interactive”

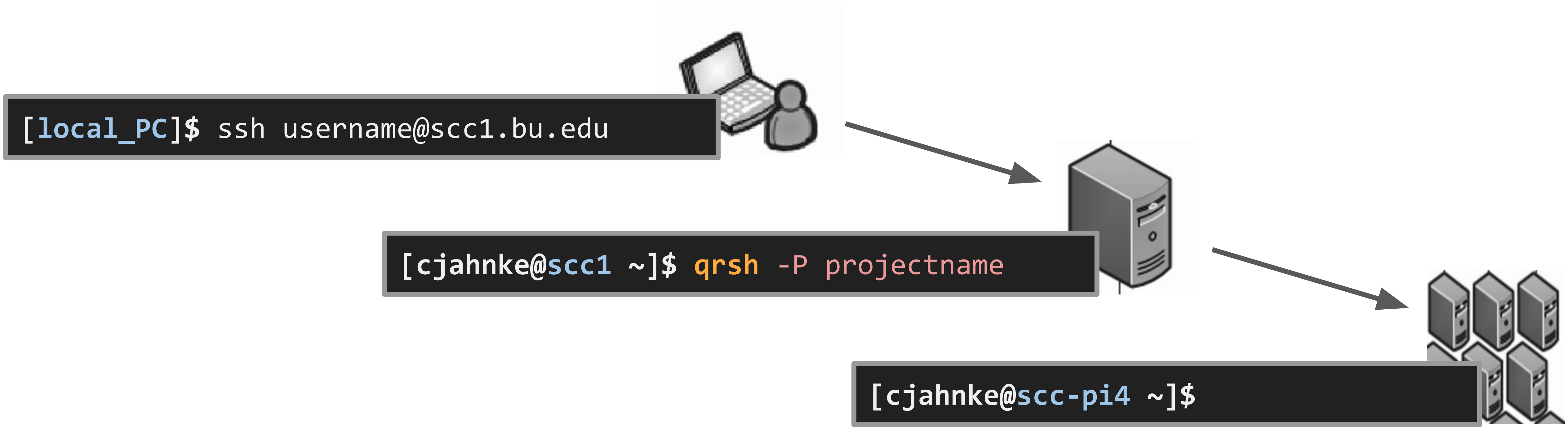
Non-Interactive “Batch”

- Blind
- Instructions coordinated with a script or binary
- Easy to run 1000’s at a time.

Interactive

Interactive jobs are submitted with the “qrsh” command:

```
qrsh  [options]  [ command  [ command_arguments ] ]
```



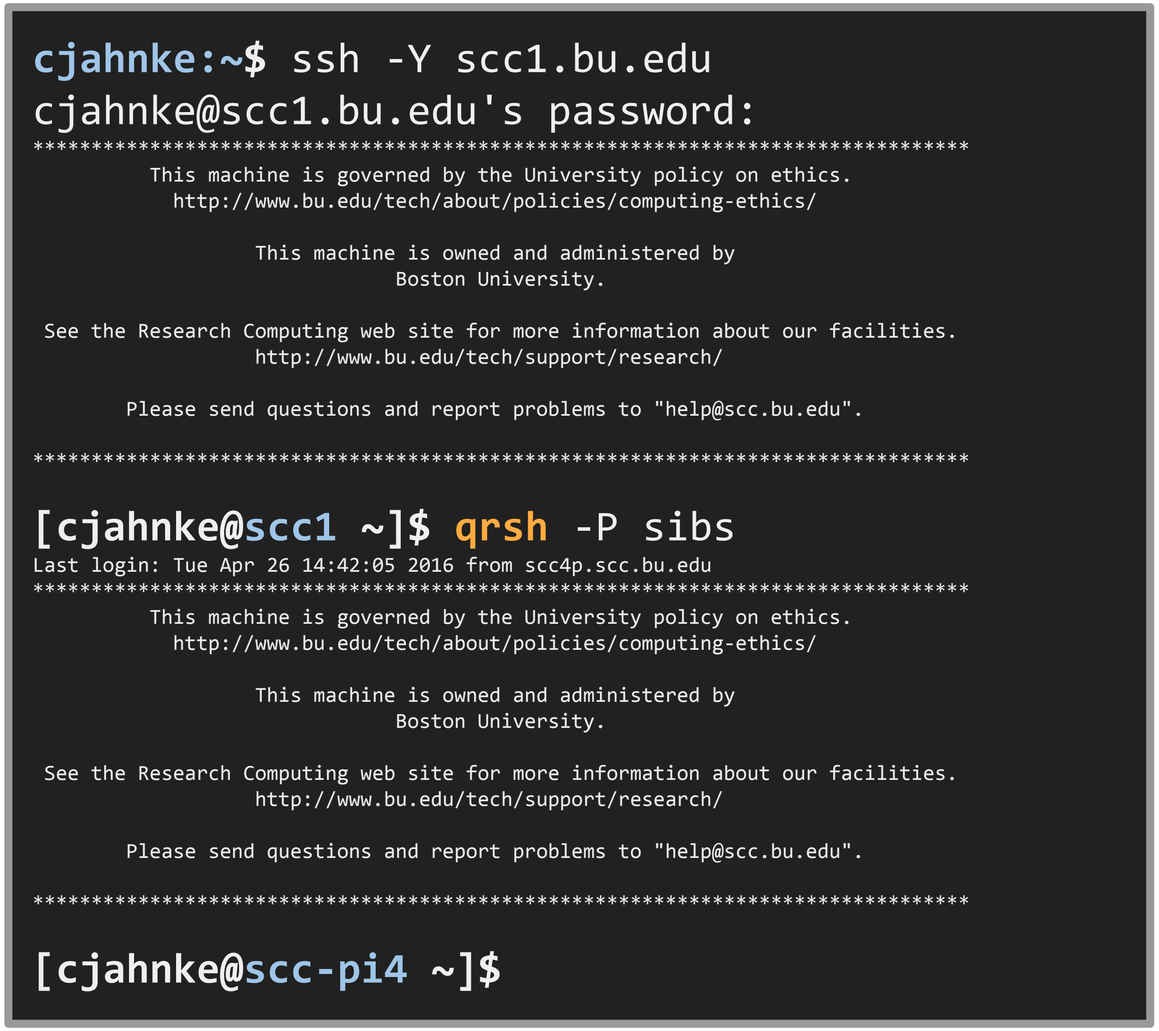
```
[local_PC]$ ssh username@scc1.bu.edu
```

```
[cjahnke@scc1 ~]$ qrsh -P projectname
```

```
[cjahnke@scc-pi4 ~]$
```

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/interactive-jobs/

Hands-On: Submitting an Interactive Job



```
cjahnke:~$ ssh -Y scc1.bu.edu
cjahnke@scc1.bu.edu's password:
********************************************************************************
          This machine is governed by the University policy on ethics.
            http://www.bu.edu/tech/about/policies/computing-ethics/

                   This machine is owned and administered by
                               Boston University.

 See the Research Computing web site for more information about our facilities.
                   http://www.bu.edu/tech/support/research/

        Please send questions and report problems to "help@scc.bu.edu".

********************************************************************************

[cjahnke@scc1 ~]$ qrsh -P sibs
Last login: Tue Apr 26 14:42:05 2016 from scc4p.scc.bu.edu
********************************************************************************
          This machine is governed by the University policy on ethics.
            http://www.bu.edu/tech/about/policies/computing-ethics/

                   This machine is owned and administered by
                               Boston University.

 See the Research Computing web site for more information about our facilities.
                   http://www.bu.edu/tech/support/research/

        Please send questions and report problems to "help@scc.bu.edu".

********************************************************************************

[cjahnke@scc-pi4 ~]$
```

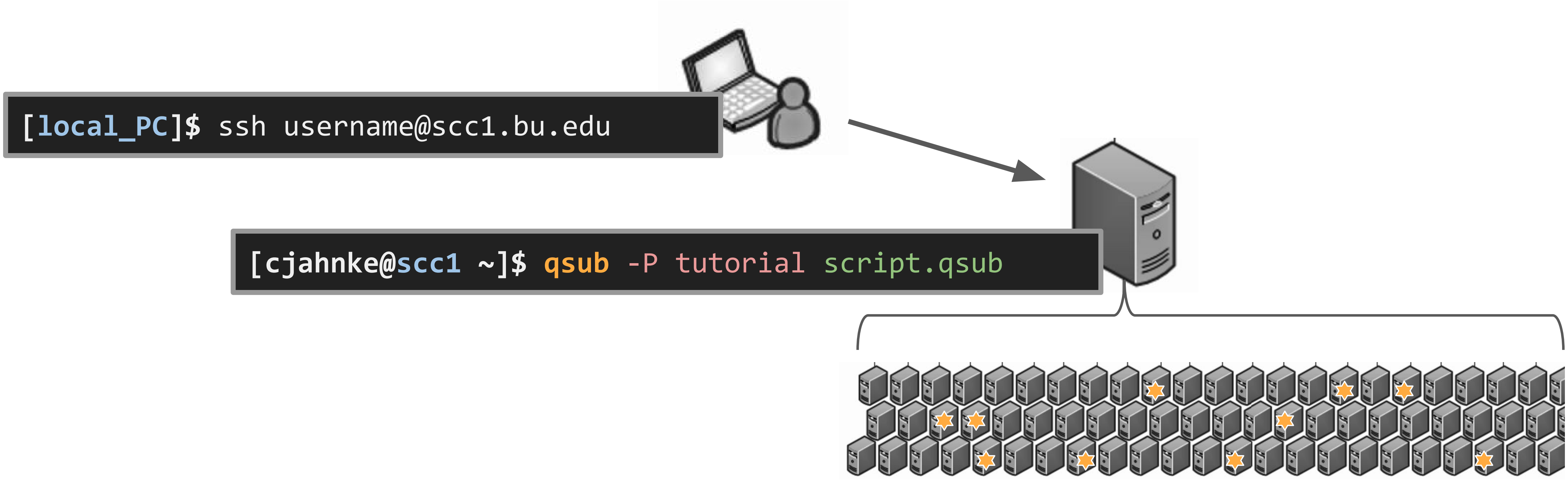
A prompt! Now we can issue commands and run applications **interactively**.

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/interactive-jobs/

Non-Interactive “Batch” Job

Non-Interactive jobs are submitted with the “`qsub`” command:

```
qsub   [options]   command   [arguments]
```



```
[local_PC]$ ssh username@scc1.bu.edu
```

```
[cjahnke@scc1 ~]$ qsub -P tutorial script.qsub
```

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/submitting-jobs/

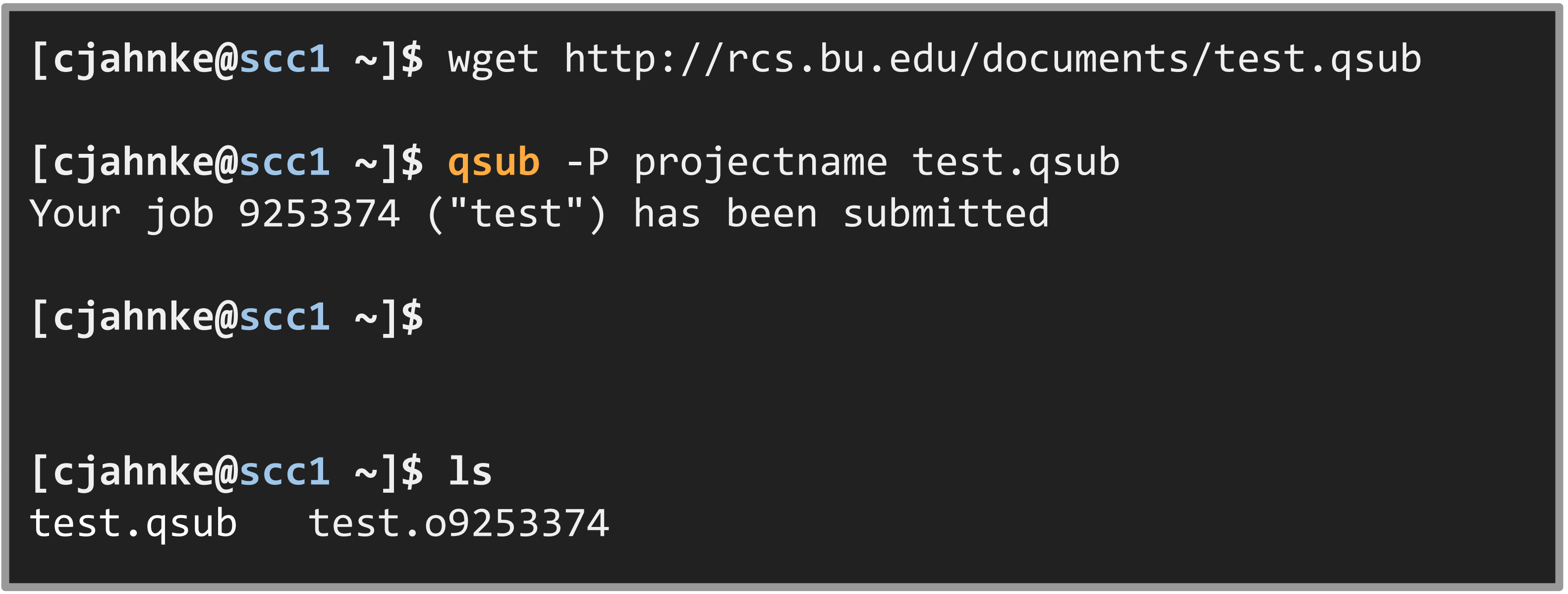
Hands-On: Submitting a Batch Job

We need a script.

Submit script to scheduler

We get our prompt back.
(Still the head node)

After completion, we get an output file.



```
[cjahnke@scc1 ~]$ wget http://rcs.bu.edu/documents/test.qsub

[cjahnke@scc1 ~]$ qsub -P projectname test.qsub
Your job 9253374 ("test") has been submitted

[cjahnke@scc1 ~]$


[cjahnke@scc1 ~]$ ls
test.qsub   test.o9253374
```

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/submitting-jobs/

The qsub File

- Just a text file.
  - Usually with extension “.qsub” or “.sh”

- Contains scheduler “directives”
  - These tell the scheduler how to orchestrate the job
  - Notification (email), Accounting, Runtime, Number of cores, Number of tasks, and more

- Contains the commands you want to run
  - Load modules
  - Single commands line by line
  - Entire pipelines

Hands-On: The qsub File

Script Interpreter

```
#!/bin/bash -l
```

Scheduler Directives

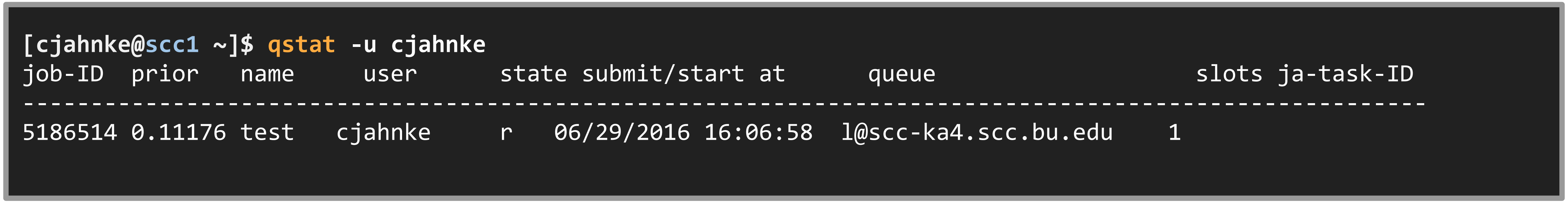
```
#$ -P sibs
#$ -N test
#$ -j y
#$ -m bae
```

Task Commands

```
echo "====================================================="
echo "Starting on : $(date)"
echo "Running on node : $(hostname)"
echo "Current directory : $(pwd)"
echo "Current job ID : $JOB_ID"
echo "Current job name : $JOB_NAME"
echo "====================================================="

module load R
sleep 10
echo hello, world

echo "====================================================="
echo "Finished on : $(date)"
echo "====================================================="
```

Monitoring Running Jobs

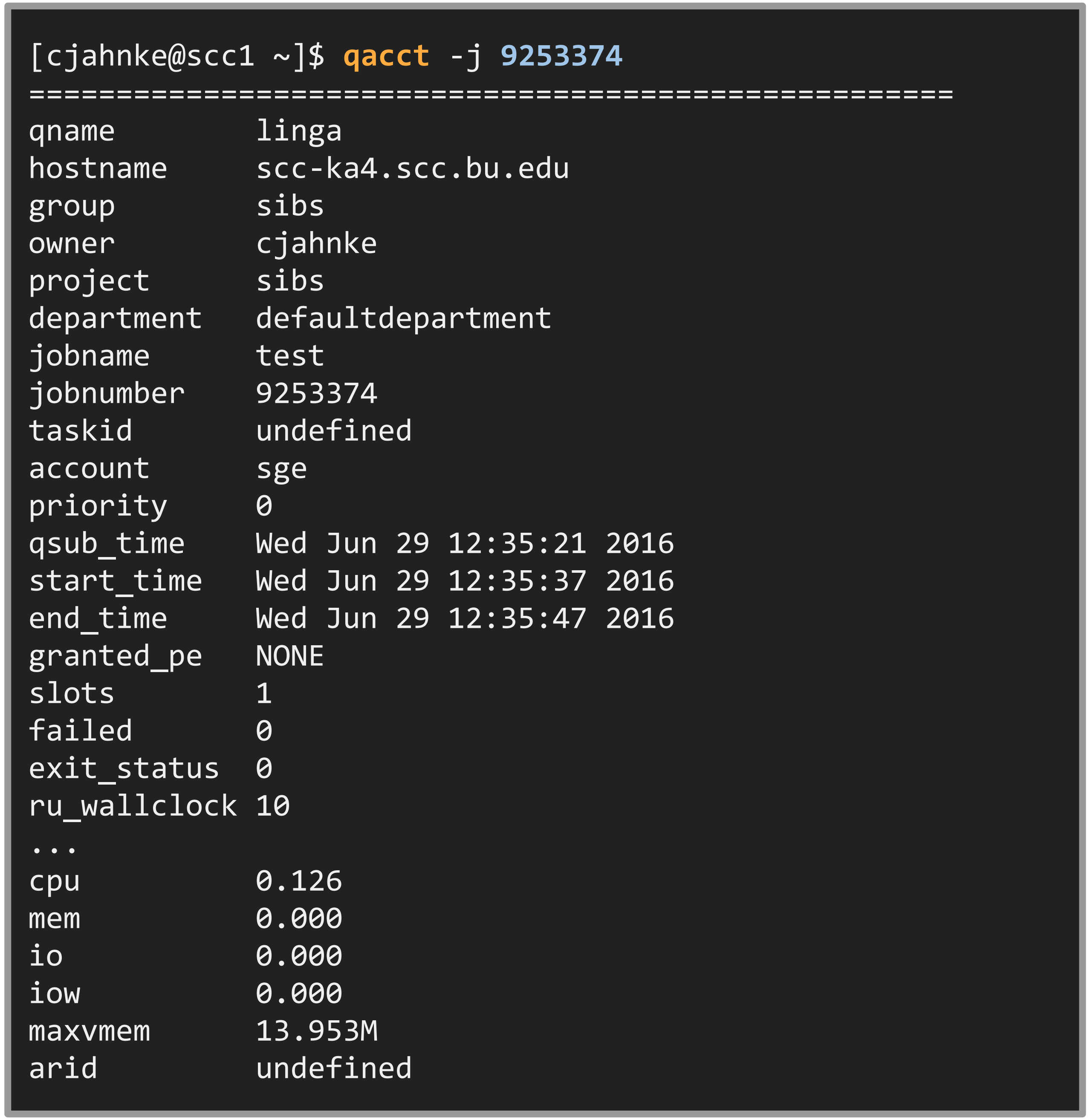
- Use **qstat** to monitor the queue status
  - Think “Queue Status” == qstat
  - Shows all users jobs. Usually a very long list
  - The “**-u [username]**” option will show a single user



```
[cjahnke@scc1 ~]$ qstat -u cjahnke
job-ID  prior   name     user      state submit/start at      queue                   slots ja-task-ID
-------------------------------------------------------------------------------------------------------
5186514 0.11176 test   cjahnke     r   06/29/2016 16:06:58  l@scc-ka4.scc.bu.edu    1
```

Completed Job Info

- Use **qacct** query the query the accounting system
  - Think “Queue Accounting” == qacct
  - Many options to tailor query
    - user
    - jobid
    - date run
    - See “man qacct”



```
[cjahnke@scc1 ~]$ qacct -j 9253374
=====================================================
qname        linga
hostname     scc-ka4.scc.bu.edu
group        sibs
owner        cjahnke
project      sibs
department   defaultdepartment
jobname      test
jobnumber    9253374
taskid       undefined
account      sge
priority     0
qsub_time    Wed Jun 29 12:35:21 2016
start_time   Wed Jun 29 12:35:37 2016
end_time     Wed Jun 29 12:35:47 2016
granted_pe   NONE
slots        1
failed       0
exit_status  0
ru_wallclock 10
...
cpu          0.126
mem          0.000
io           0.000
iow          0.000
maxvmem      13.953M
arid         undefined
```

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/tracking-jobs/

A Standard Single-Processor Job

If no specific resources are requested, your job is allocated:

- 1 “Slot” (Processor core, any type/architecture)
- 12 Hour Runtime
- 4 GB RAM
- No GPU, MPI, or Parallelization

All of these can be modified.

Scheduler Options - General Directives

General Directives	
Directive	Description
-P *project_name*	Project to which this jobs is to be assigned. **Mandatory** for all users associated with any BUMC project.
-N *job_name*	Specifies the job name. The default is the script or command name.
-o *outputfile*	File name for the stdout output of the job.
-e *errfile*	File name for the stderr output of the job.
-j y	Merge the error and output stream files into a single file.
-m *b|e|a|s|n*	Controls when the batch system sends email to you. The possible values are – when the job begins (b), ends (e), is aborted (a), is suspended (s), or never (n) – default.
-M *user_email*	Overwrites the default email address used to send the job report.
-V	All current environment variables should be exported to the batch job.
-v *env=value*	Set the runtime environment variable *env* to *value*.
-hold_jid *job_list*	Setup job dependency list. *job_list* is a comma separated list of job ids and/or job names which must complete before this job can run. See Advanced Batch System Usage for more information.

SCC Help: http://www.bu.edu/tech/support/research/system-usage/running-jobs/submitting-jobs/

Scheduler Options - Resource Directives

Directives to request SCC resources	
Directive	Description
-l h_rt=*hh:mm:ss*	Hard runtime limit in *hh:mm:ss* format. The default is 12 hours.
-l mem_total =*#G*	Request a node that has at least this amount of memory. Current possible choices include 94G, 125G, 252G 504G.
-l mem_per_core =*#G*	Request a node that has at least these amount of memory per core.
-l cpu_arch=*ARCH*	Select a processor architecture (sandybridge, nehalem, etc). See Technical Summary for all choices.
-l cpu_type=*TYPE*	Select a processor type (E5-2670, E5-2680, X5570, X5670, etc). See Technical Summary for all choices.
-l gpus=*G/C*	Requests a node with GPU. *G/C* specifies the number of GPUs per each CPU requested and should be expressed as a decimal number. See Advanced Batch System Usage for more information.
-l gpu_type=*GPUMODEL*	Current choices for *GPUMODEL* are M2050, M2070 and K40m, P100. (specific)
-l gpu_c=*GPUCAPABILITY*	Define the minimum GPU Capability you need for you job (at least)
-l eth_speed=*N*	Ethernet speed (1 or 10 Gbps).
-l scratch_free=#G	Request a node that has at least this amount of space in scratch. Note that the amount changes!
-pe omp *N*	Request multiple slots for Shared Memory applications (OpenMP, pthread). This option can also be used to reserve larger amount of memory for the application. *N* can vary from 1 to 16.
-pe mpi_#_tasks_per_node *N*	Select multiple nodes for MPI job. Number of tasks can be 4, 8, 12 or 16 and *N* must be a multiple of this value. See Advanced Batch System Usage for more information.

Getting Help

How to Get Help

Support Website

- http://rcs.bu.edu (http://www.bu.edu/tech/support/research/)

Upcoming Tutorials:

- http://rcs.bu.edu/tutorials

Email (Submit a Ticket):

- help@scc.bu.edu

Email Direct:

- cjahnke@bu.edu

Questions?

Research Computing Services Website

http://rcs.bu.edu

RCS Tutorial Evaluation

http://rcs.bu.edu/survey/tutorial_evaluation.html