

R + TACC

David Walling

Data Group

Texas Advanced Computing Center

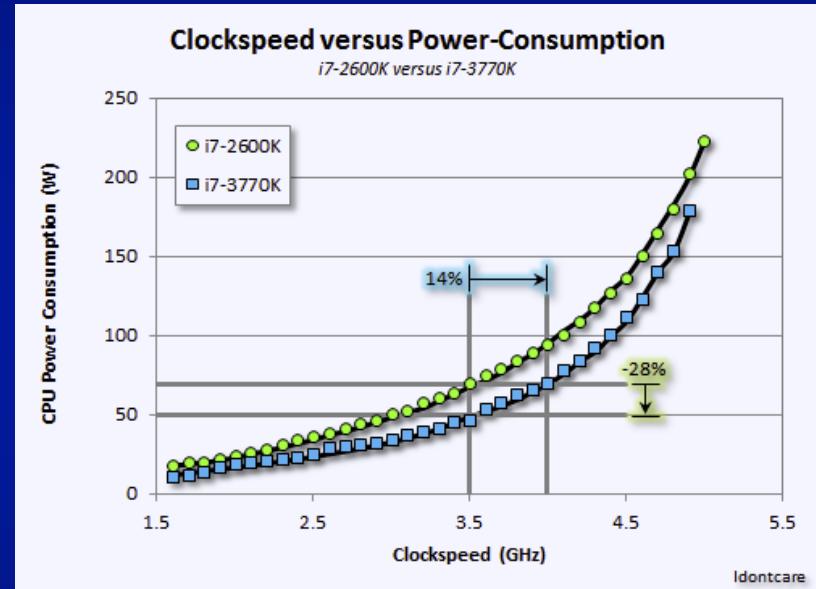
walling@tacc.utexas.edu

What to do if the computation is too big for your desktop/laptop

- A common user question:
 - I have an existing R pipeline for my research work. But the data is growing too big. Now my R program runs for days (weeks) to finish or simply runs out of memory.
- 3 Strategies
 - Move to bigger hardware
 - Advanced libraries/C++
 - Implement code using parallel packages

Trends in HPC

- Processors not getting faster
- Increase performance -> cram more cores on each chip
- Requires reducing clock speed (power + heat)
 - Stampede1 Node
 - 16 cores/2.7 GHz
 - No hyperthreading
 - 346 GFLOPS/node
 - Stampede2 Node
 - 68 cores/1.4 GHz
 - 4 x hyperthreading (272 'cores')
 - 3000 GFLOPS/node
- Single-threaded applications will run *SLOWER* on these new resources, must start thinking in parallel



<https://www.quora.com/Why-havent-CPU-clock-speeds-increased-in-the-last-5-years>

Strategy 1: Powerful Hardware

Stampede (Old) - HPC

- normal queue: 16 cores/32GB mem/48 hour max
- largemem queue: 32 cores/ 1TB mem/48 hour max
- normal-mic queue: Access to Intel MIC co-processors

Stampede2 - HPC

- normal queue: 68 cores(4x hyperthreading 272)/98GB mem/48 hour max

Lonestar5 - HPC

- normal queue: 24 cores/64GB mem/48hour max
- large512GB: 32 cores/512GB mem
- large1TB: 48 cores/1TB mem

Maverick - Vis

- vis queue: 32 cores/128GB mem/4 hour max
- gpu queue: has access to GPUs

Wrangler - Data

- normal queue: 24 cores/128GB mem/48hour max
- Reservations: Up to 1 month reserved nodes
- Hadoop/Spark + R Streaming
- Dedicated Flash storage for IO intensive tasks

Allocations

Open to national researcher community

- PIs create projects and request allocations
- Assign ‘delegate’ to manage user access

XSEDE

- National organization providing computational resources to
- ~ 90% of cycles on Stampede
- Portal: <http://portal.xsede.org>

UT/TACC

- Additional systems available
- Portal: <http://portal.tacc.utexas.edu>

Startup vs Standard Allocations

- Startup: Small amount of time, fast turn around, intended to ‘kick the tires’
- Standard: Quarterly review cycles

Linux

We run linux

More command line driven

Daunting for Windows only users

Filezilla allows for editing R scripts remotely

Can use VNC access to get a Linux desktop

RStudio helps the transition

TACC Linux Training and Courses

ssh

```
(walling) wrangler.tacc.utexas.edu - Konsole

walling@dw-tacc-laptop:~$ ssh walling@wrangler.tacc.utexas.edu
Password:
TACC Token Code:
Last login: Thu Sep 29 13:35:47 2016 from 146.6.176.142
-----
          Welcome to the Wrangler Supercomputer
          Texas Advanced Computing Center, The University of Texas at Austin
-----
          ** Unauthorized use/access is prohibited. **

If you log on to this computer system, you acknowledge your awareness
of and concurrence with the UT Austin Acceptable Use Policy. The
University will prosecute violators to the full extent of the law.

TACC Usage Policies:
http://www.tacc.utexas.edu/user-services/usage-policies/

Questions and Problem Reports:
--> XD Projects:    help@xsede.org (email)
--> TACC Projects:  portal.tacc.utexas.edu (web)
```

Multifactor Authentication

<https://portal.tacc.utexas.edu/tutorials/multifactor-authentication>

The screenshot shows a web browser window with multiple tabs open at the top. The main content area displays the TACC User Portal homepage with a navigation bar and a sidebar containing links related to MFA setup. A central panel shows the 'Multifactor Authentication at TACC' tutorial page, which includes a heading, a message about the implementation of MFA, and a section titled 'What is Multi-Factor Authentication?'. Below this is a section titled 'Setting up MFA at TACC' with a sub-section '1. Manage Profile'. A callout box labeled 'Figure 1.' points to a screenshot of a mobile device displaying the 'Multi-Factor Authentication' screen with a 'Pair your mobile device' button. The bottom of the tutorial page contains a note about pairing methods.

Multifactor Authentication at TACC

What is Multi-Factor Authentication?

Setting up MFA at TACC

1. Manage Profile
2. Select Pairing Method

TACC Token app
SMS Messaging
TACC Hard Token
International Users
Logging into TACC resources with MFA enabled
Unpairing your Device

Table of Contents ▾

TACC USER PORTAL

HOME NEWS RESOURCES ALLOCATIONS DOCUMENTATION TRAINING CONSULTING ABOUT

Multifactor Authentication at TACC

TACC is now implementing Multi-Factor Authentication (MFA) as an additional security measure when accessing resources. MFA will become mandatory on September 27, 2016 and will be required to access all TACC resources. To set up MFA pairing at TACC, proceed directly to the [TACC pairing page](#).

What is Multi-Factor Authentication?

Authentication is the process of determining if you are you. Traditional methods of associating a user account with a single password have not been 100% successful. Multi-Factor Authentication (MFA) requires another step, or "factor", in the authentication process. In addition to the usual password users must complete authentication using another device unique to them, usually the user's mobile phone/device. TACC offers another device, the TACC Hard Token, as another authentication factor.

Setting up MFA at TACC

1. Manage Profile

To pair a new device, sign in to the TACC User Portal and click on the "Manage Profile" link in the right corner. Users who've not set up MFA before will see a message similar to Figure 1.

Figure 1.
Profile with no MFA enabled

Click "Pair a Device" to get to the TACC device pairing page. Here you'll be presented with three different pairing methods. Users may authenticate with one and only one method. It's easy to pair and unpair using either of the first two methods.

TACC USER PORTAL

HOME NEWS RESOURCES ALLOCATIONS DOCUMENTATION TRAINING CONSULTING ABOUT

Pair a Device for Multi-Factor Authentication

TACC requires multi-factor authentication (MFA) in order to access some resources. With MFA you use your username, password, and a one-time password (OTP) to log in. TACC currently supports MFA using the TACC Token App, SMS, or a physical token. Versions of the TACC Token App are currently available for [Apple iOS](#) and [Android](#) devices. For other devices, pairing can be enabled to use SMS. [Click here](#) for more information. If you cannot use either the TACC Token App or SMS you can [submit a ticket](#) to request a physical token.

Basic Unix Commands

ls - list contents of directory

pwd - get current directory path

cd - change directory

mkdir - create a new folder

cp/mv - copy/move files

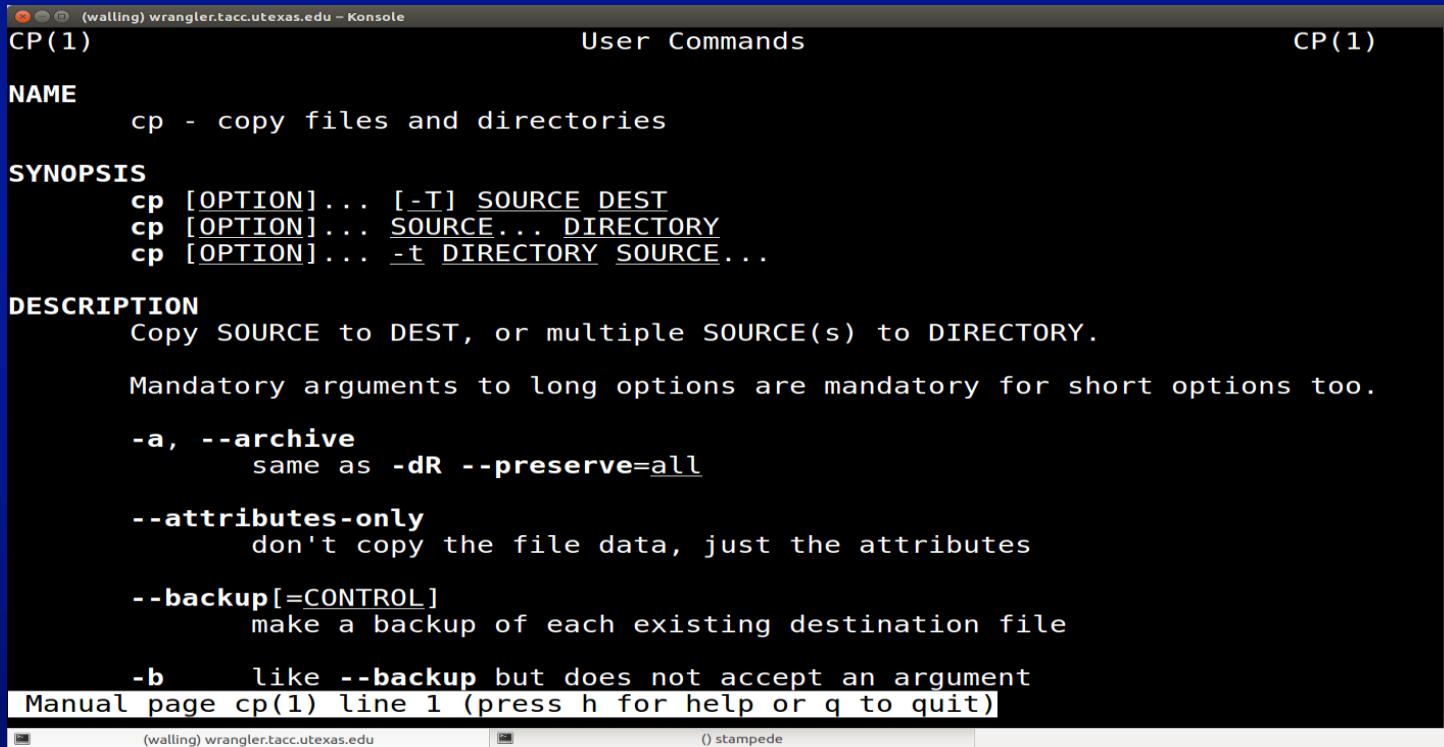
cat/head - output contents of a file to console

more/less - scroll through larger files to view content

man - open command help manual

Unix Help/Documentation

```
$> man cp
```



The screenshot shows a terminal window titled '(walling) wrangler.tacc.utexas.edu - Konsole'. The title bar also displays 'User Commands' and 'CP(1)'. The main content of the window is the man page for the cp command. It includes sections for NAME, SYNOPSIS, DESCRIPTION, and various options like -a, --archive, --attributes-only, --backup, and -b. The bottom of the window shows the footer 'Manual page cp(1) line 1 (press h for help or q to quit)'.

```
CP(1)                               User Commands                               CP(1)

NAME
    cp - copy files and directories

SYNOPSIS
    cp [OPTION]... [-T] SOURCE DEST
    cp [OPTION]... SOURCE... DIRECTORY
    cp [OPTION]... -t DIRECTORY SOURCE...

DESCRIPTION
    Copy SOURCE to DEST, or multiple SOURCE(s) to DIRECTORY.
    Mandatory arguments to long options are mandatory for short options too.

    -a, --archive
        same as -dR --preserve=all

    --attributes-only
        don't copy the file data, just the attributes

    --backup[=CONTROL]
        make a backup of each existing destination file

    -b      like --backup but does not accept an argument
Manual page cp(1) line 1 (press h for help or q to quit)
```

File System Quotas

```
(-) stampede.tacc.utexas.edu - Konsole
To show all queued jobs, issue:           showq
To kill a queued job, issue:             scancel <jobId>

See "man slurm" or the Stampede user guide for more detailed information.

--> To see all the software that is available across all compilers and
mpi stacks, issue: "module spider"

--> To see which software packages are available with your currently loaded
compiler and mpi stack, issue: "module avail"

--> Stampede has three parallel file systems: $HOME (permanent,
quota'd, backed-up) $WORK (permanent, quota'd, not backed-up) and
$SCRATCH (high-speed purged storage). The "cdw" and "cds" aliases
are provided as a convenience to change to your $WORK and $SCRATCH
directories, respectively.

----- Project balances for user walling -----
| Name      Avail SUs   Expires | Name      Avail SUs   Expires |
| TG-STA110014S  4735974          | RangerTechInsertion  15000  2017-03-31 |
| TG-ASC160051    299988  2017-06-27 | TG-STA110019S    283542  2016-10-31 |
| TG-ASC160050    50000   2017-06-27 | WebberEnergy     48030   2017-03-31 |
| A-ccsc        202024  2016-12-31 | TACC-DIC        -9461   2017-06-30 |
| TRAINING-HPC    12222   2017-08-01 |                               |
----- Disk quotas for user walling -----
| Disk       Usage (GB)  Limit  %Used  File Usage  Limit  %Used |
| /home1        4.1      5.0   82.86   95912    150000  63.94 |
| /work         387.9    1024.0 37.88   739963   3000000 24.67 |
| /scratch       0.0      0.0   0.00      4          0      0.00 |

Tip 18 (See "module help tacc_tips" for features or how to disable)

Execute "du -sch *" to see your disk usage by directory.

login2.stampede(1)$
```

\$HOME

(walling) wrangler.tacc.utexas.edu – Konsole

```
login1.wrangler(5)$ pwd  
/home/00157/walling  
login1.wrangler(6)$  
login1.wrangler(6)$ echo $HOME  
/home/00157/walling  
login1.wrangler(7)$
```

\$HOME

Small quota ~ 5GB

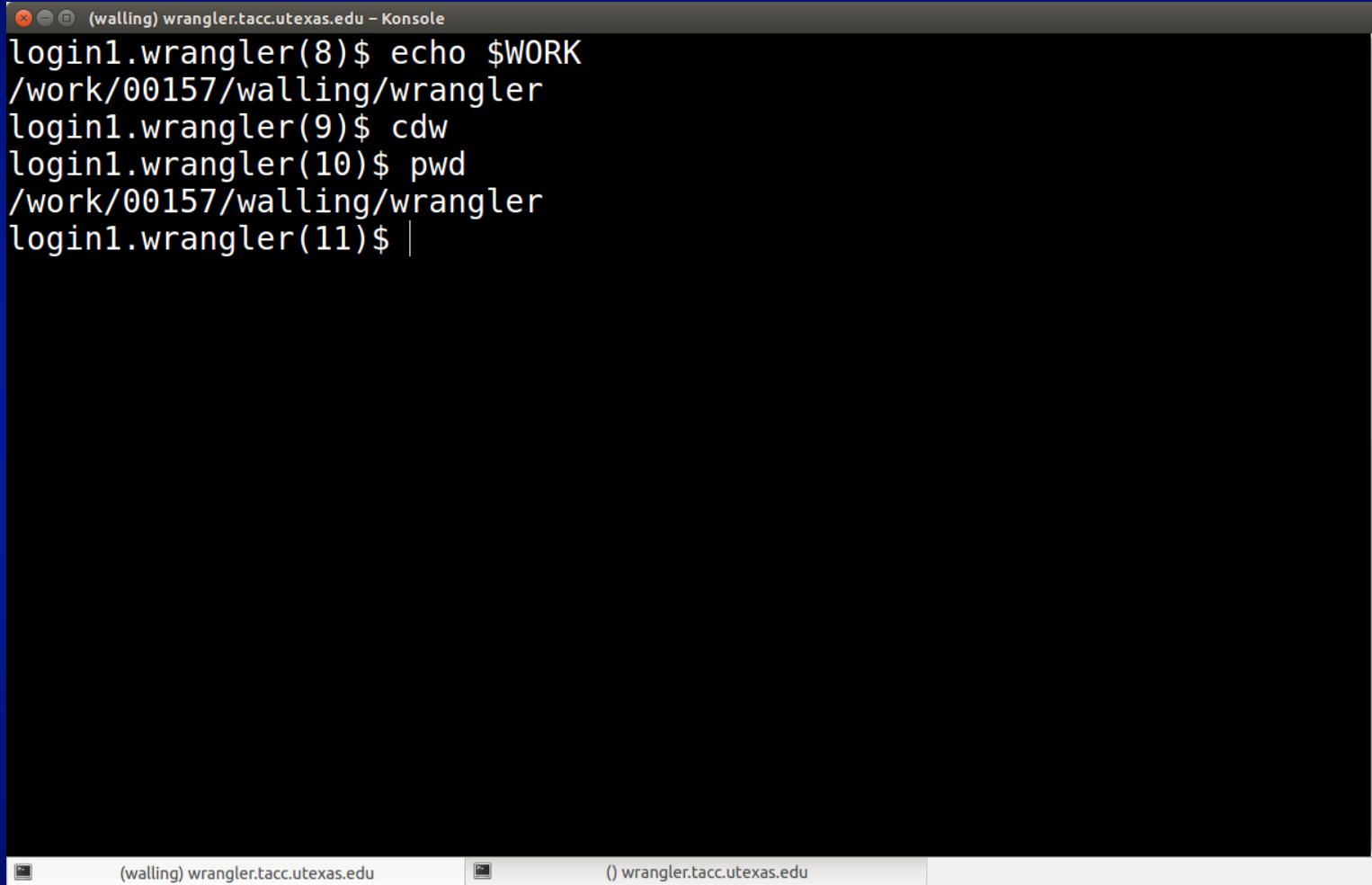
Use for script files

Contains advanced configuration scripts such as `~/.bashrc` and `~/.ssh`

Linux Exercise

```
login1.wrangler(5)$ pwd
/home/00157/walling
login1.wrangler(6)$ cp -r /work/00157/walling/wrangler/training/R-HPC-April2017/ $HOME/
login1.wrangler(7)$ ls R-HPC-April2017/
data.csv  dataTable.R  Multicore-CLT.R  Multicore.R  run_SimpleSNOW.slurm  SimpleSNOW.R  test-rcpp.R
login1.wrangler(8)$ ls -lah R-HPC-April2017/
total 63M
drwx-----  2 walling G-818695 4.0K Apr 13 07:45 .
drwx----- 56 walling G-814305 4.0K Apr 13 07:45 ..
-rw-----  1 walling G-818695  63M Apr 13 07:45 data.csv
-rw-----  1 walling G-818695 1.1K Apr 13 07:45 dataTable.R
-rw-----  1 walling G-818695 1.1K Apr 13 07:45 Multicore-CLT.R
-rw-----  1 walling G-818695  216 Apr 13 07:45 Multicore.R
-rw-----  1 walling G-818695  181 Apr 13 07:45 run_SimpleSNOW.slurm
-rw-----  1 walling G-818695  423 Apr 13 07:45 SimpleSNOW.R
-rw-----  1 walling G-818695  126 Apr 13 07:45 test-rcpp.R
```

\$WORK



(walling) wrangler.tacc.utexas.edu – Konsole

```
login1.wrangler(8)$ echo $WORK
/work/00157/walling/wrangler
login1.wrangler(9)$ cdw
login1.wrangler(10)$ pwd
/work/00157/walling/wrangler
login1.wrangler(11)$ |
```

(walling) wrangler.tacc.utexas.edu () wrangler.tacc.utexas.edu

\$WORK

Large quota ~ 1TB

Use for data, software installations, etc...

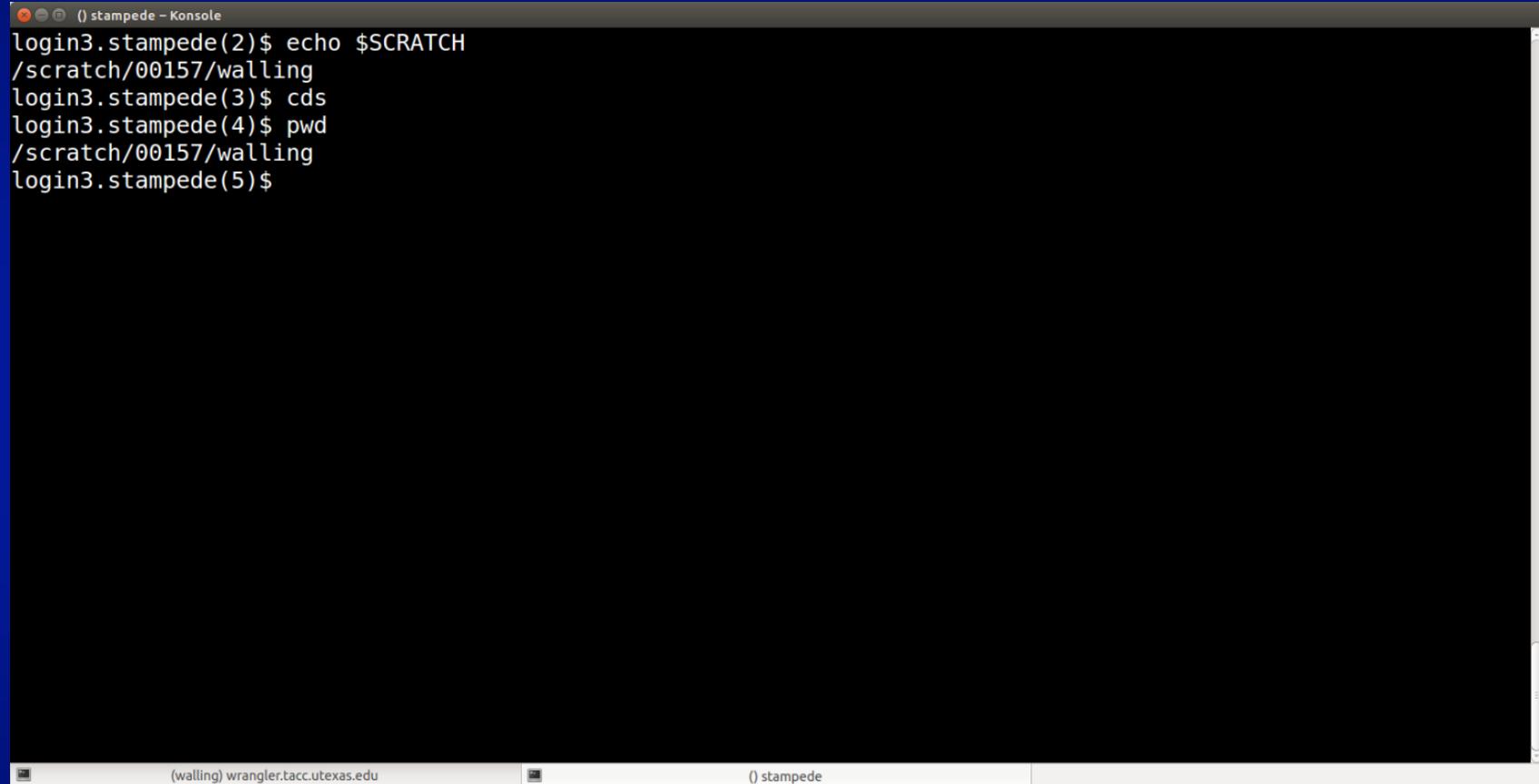
If you want a newer version of R, install here

Shared across resources @ TACC

\$WORK

```
login1.wrangler(10)$ pwd
/work/00157/walling/wrangler
login1.wrangler(11)$ cd ..
login1.wrangler(12)$ ll
total 34625552
-rw----- 1 walling G-1130 22584004543 Sep 25 2013 AirlineData87to08.csv
-rw-r--r-- 1 walling G-814305 2135031808 Dec 15 2015 big_file.out
drwx----- 3 walling G-816019 4096 Dec 14 11:40 build
drwx----- 16 walling G-1130 4096 Aug 29 2016 consulting
drwx----- 5 walling G-814305 4096 Jul 7 2015 custom-R
drwx----- 9 walling G-1130 4096 Nov 13 2015 custom-R-1
drwx----- 3 walling G-814305 4096 Oct 8 2015 custom-R-3.2.2
drwx----- 3 walling G-816019 4096 Jan 20 08:45 hikari
drwx----- 7 walling G-1130 4096 Oct 8 2015 labs
drwx----- 3 walling G-814305 4096 Jan 14 2016 lonestar
drwxr-xr-x 7 walling G-814305 4096 Mar 27 2015 maverick
drwxr-xr-x 7 walling G-1130 4096 Jul 9 2013 Misc
drwxr-xr-x 3 walling G-1130 4096 Jul 11 2013 New_Rstudio
drwx----- 7 walling G-1130 4096 May 22 2014 projects
drwx----- 4 walling G-1130 4096 Aug 21 2014 R_bench
drwxr-xr-x 4 walling G-816019 4096 Dec 13 10:14 R-examples
drwx----- 5 walling G-1130 4096 Sep 26 2013 R_meetup
drwx----- 2 walling G-816019 4096 Dec 13 11:00 slurm-examples
drwx----- 3 walling G-816019 4096 Jan 20 11:53 software
drwx----- 2 walling G-816019 4096 Aug 5 2016 SSI2016
-rw-r--r-- 1 walling G-817570 10737418240 Apr 4 2016 test10G
drwxr-xr-x 5 walling G-816019 4096 Jan 17 12:48 training
drwxr-xr-x 25 walling G-814305 4096 Mar 8 14:47 wrangler
drwx----- 2 walling G-1130 4096 Aug 13 2014 YellowHat
drwx----- 7 walling G-1130 4096 Mar 24 2014 yye00
login1.wrangler(13)$ echo $WORK
/work/00157/walling/wrangler
```

\$SCRATCH



A screenshot of a terminal window titled '(stampede - Konsole)'. The window contains the following text:

```
login3.stampede(2)$ echo $SCRATCH  
/scratch/00157/walling  
login3.stampede(3)$ cds  
login3.stampede(4)$ pwd  
/scratch/00157/walling  
login3.stampede(5)$
```

The terminal window is located on a desktop with other windows visible in the background, including one titled '(walling) wrangler.tacc.utexas.edu'.

* No \$SCRATCH on Wrangler

\$SCRATCH

Temporary file system, subject to purging

Intended for use while jobs are running

Very high performant

Not often needed for R jobs, mostly for very large traditional HPC
workflows

!!! DO NOT LEAVE IMPORTANT FILES HERE !!!

\$DATA

```
login1.wrangler(14)$ cd $DATA
login1.wrangler(15)$ echo $DATA
/data/00157/walling
login1.wrangler(16)$ ll
total 650117340
drwxr-xr-x 2 walling G-816019          4096 Jul 17  2016 mariadb
drwx----- 3 walling G-818695          4096 Mar 16 13:23 sixdegs
-rw-r--r-- 1 walling G-816019 107374182400 May  9  2016 test100G
-rw----- 1 walling G-817570 10737418240  Apr  4  2016 test10G
-rw-r--r-- 1 walling G-816019          0 May  9  2016 test1TB.img
-rw-r--r-- 1 walling G-816019 21474836480 May  9  2016 test20G
-rw----- 1 walling G-816019 536870912000 May  9  2016 test500G
-rw----- 1 walling G-816019          9 Oct 24 08:34 test-cron.txt
-rw----- 1 walling G-817570          0 Apr  4  2016 test.img
login1.wrangler(17)$ lfs quota /data
Disk quotas for user walling (uid 29791):
  Filesystem  kbytes  quota  limit  grace  files  quota  limit  grace
    /data     778166928      0      0      - 1810445      0      0      -
Disk quotas for group G-818695 (gid 818695):
  Filesystem  kbytes  quota  limit  grace  files  quota  limit  grace
    /data     893367236      0      0      -   5619      0      0      -
```

\$DATA

Quota by project

- a. Ifs quota /data

Replicated between TACC and IU

Not backed up to Archive

Users get directory by default

Can request a ‘project’ directory under /data/projects to serve as a more
‘centralized’ place for large project data.

Example use: store database files when not running active job on Wrangler’s
flash storage system

Login Nodes

Always log into the login nodes

Shared nodes with limited resources

Ok to edit, compile, move files

For R, ok to install packages from login nodes

Anything requiring more than trivial resources must use Compute nodes

!!! DO NOT RUN R SCRIPTS !!!

Compute Nodes

Dedicated nodes for each job

Only accessible via the SLURM job scheduler

Once you have a job running on a compute node, you can ssh to that node

Variety of node ‘types’ depending on system. See relevant User Guide for details.

Software Modules

We provide an optimized build of R called Rstats

Compiled with Intel compilers (vs. gnu) and linked against MKL math library

Many 3rd party packages pre-installed

Managing the linux environment is done via TACC modules commands

```
login1.wrangler(1)$ ml
Currently Loaded Modules:
 1) TACC-paths   2) Linux   3) cluster-paths   4) intel/15.0.3   5) mvapich2/2.1   6) cluster   7) TACC

login1.wrangler(2)$ ml Rstats
login1.wrangler(3)$ which R
/opt/apps/intel15/mvapich2_2_1/Rstats/3.2.1/bin/R
```

Additional Libraries and Packages

- Libraries
 - Comes with Package installation (Core or others)
 - library() shows a list of current installed
 - library must be loaded before use e.g.
 - library(rpart)
- Packages
 - Developed code/libraries outside the core packages
 - Can be downloaded and installed separately
 - install.package("name")
 - There are currently 8,900 packages at <http://cran.r-project.org/web/packages/>
 - E.g. Rweka, interface to Weka.
 - ggplot2: very popular for ‘building up’ plots

Installing Packages on TACC Systems

- R handles package dependencies for you.
- Many packages compile C/C++/Fortran.
- In some cases, additional libraries required.
 - *libXXX.so not found.*
 - Submit a consulting ticket.

Installing Packages on TACC Systems

```
> install.packages('FrF2')
Warning in install.packages("FrF2") :
  'lib =
  "/opt/apps/intel14/mvapich2_2_0/Rstats/3.0.3/lib64/R/li
  brary"' is not writable
Would you like to use a personal library instead?
(y/n) y
Would you like to create a personal library
~/R/x86_64-unknown-linux-gnu-library/3.0
to install packages into? (y/n) y
```

```
mpicc -std=gnu99 -fPIC -openmp -mkl=parallel -O3 -xHost -
L/opt/apps/intel/13/composer_xe_2013_sp1.1.106/mkl/lib/int
el64 -lmkl_rt -shared -fPIC -openmp -mkl=parallel -O3 -
xHost -
L/opt/apps/intel/13/composer_xe_2013_sp1.1.106/mkl/lib/int
el64 -lmkl_rt -o BsMD.so bsmd.o -lmkl_intel_lp64 -
lmkl_intel_thread -lmkl_core -liomp5 -lmkl_rt -lifport -
lfcoremt -limf -lsVML -lm -lipgo -lirc -lpthread -lirc_s -ldl -
L/opt/apps/intel14/mvapich2_2_0/Rstats/3.0.3/lib64/R/lib -IR
```

```
( ) maverick - Konsole
File Edit View Bookmarks Settings Help
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages('FrF2')
Installing package into '/home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.0'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
CRAN mirror

 1: O-Cloud
 3: Argentina (Mendoza)
 5: Australia (Melbourne)
 7: Belgium
 9: Brazil (PR)
11: Brazil (SP 1)
13: Canada (BC)
15: Canada (ON)
17: Canada (QC 2)
19: China (Beijing 1)
21: China (Hefei)
23: Colombia (Bogota)
25: Czech Republic
27: Ecuador
29: France (Lyon 2)
31: France (Paris 1)
33: Germany (Berlin)
35: Germany (Goettingen)
37: Hungary
39: Indonesia (Jakarta)
41: Iran
43: Italy (Milano)
45: Italy (Palermo)
47: Japan (Tokyo)
49: Korea (Seoul 1)
51: Lebanon
53: Mexico (Texcoco)
55: Netherlands (Utrecht)
57: Norway
59: Poland
61: Russia
63: Slovakia
65: South Africa (Johannesburg)
67: Spain (Madrid)
69: Switzerland
71: Taiwan (Taipei)
73: Turkey
75: UK (London)
77: UK (St Andrews)
79: USA (CA 2)
81: USA (IN)
83: USA (MD)
85: USA (MO)
87: USA (OR)
89: USA (PA 2)
91: USA (TX 1)
93: USA (WA 2)
95: Vietnam

Selection: |
```

.libPaths()

```
c251-114.wrangler(52)$ R

R version 3.2.1 (2015-06-18) -- "World-Famous Astronaut"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

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

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

> .libPaths()
[1] "/opt/apps/intel15/mvapich2_2_1/RstatsPackages/3.2.1/packages"
[2] "/home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"
[3] "/opt/apps/intel15/mvapich2_2_1/Rstats/3.2.1/lib64/R/library"
```

.libPaths()

```
> .libPaths()
[1] "/opt/apps/intel15/mvapich2_2_1/RstatsPackages/3.2.1/packages"
[2] "/home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"
[3] "/opt/apps/intel15/mvapich2_2_1/Rstats/3.2.1/lib64/R/library"
>
> .libPaths(c(.libPaths()[2], .libPaths()[c(1,3)]))
>
> .libPaths()
[1] "/home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"
[2] "/opt/apps/intel15/mvapich2_2_1/RstatsPackages/3.2.1/packages"
[3] "/opt/apps/intel15/mvapich2_2_1/Rstats/3.2.1/lib64/R/library"
```

installed.packages()

```
> packinfo = installed.packages()
> head(packinfo[,c("LibPath")])                                backports
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"      bartMachine
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"      bartMachineJARs
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"      base64enc
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"      bayesplot
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"      BH
" /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2"
```

Exercise

1. ssh to wrangler
2. load the Rstats module
3. Under “which” version of the intel compiler was “R” built?
4. Where is ‘ggplot2’ installed?

Exercise

1. ssh to wrangler
2. load the Rstats module
3. Under “which” version of the intel compiler was “R” built?
4. Where is ‘ggplot2’ installed?

```
login1.wrangler(28)$ ml
Currently Loaded Modules:
 1) TACC-paths   2) Linux   3) cluster-paths   4) intel/15.0.3   5) mvapich2/2.1   6) cluster   7) TACC

login1.wrangler(29)$ ml Rstats
login1.wrangler(30)$ which R
/opt/apps/intel15/mvapich2_2_1/Rstats/3.2.1/bin/R
login1.wrangler(31)$ ml spider Rstats

-----
Rstats: Rstats/3.2.1
-----

Other possible modules matches:
  RstatsPackages

You will need to load all module(s) on any one of the lines below before the "Rstats/3.2.1" module is available to load.

  intel/15.0.3  mvapich2/2.1
```

Exercise

1. ssh to wrangler
2. load the Rstats module
3. Under “which” version of the intel compiler was “R” built?
4. Where is ‘ggplot2’ installed?

```
> packinfo = installed.packages()
> pi = as.data.frame(packinfo)
> pi[pi$Package == 'ggplot2', c('LibPath')]
                               ggplot2
/opt/apps/intel15/mvapich2_2_1/RstatsPackages/3.2.1/packages
3 Levels: /home/00157/walling/R/x86_64-unknown-linux-gnu-library/3.2 ...
```

Runnings Jobs - idev

DO NOT RUN on login
nodes!!!

Use idev for interactive
sessions

Wraps sbatch
'Read' User Guides

```
login1.wrangler(4)$ idev -p normal -t 04:00:00

Defaults file      : ~/.idevrc
Default project   : TG-STA110019S
Default time       : 30 min.
Default queue      : debug
System             : Wrangler
Using queue        : -p normal
time (hh:mm:ss)   : -t 04:00:00

Just a Note: Your reservation hadoop+SSI2016+1663 is INACTIVE.
Use idev -r to use it when it becomes ACTIVE.
To see when it begins, execute: scontrol show reservations

-----
          Welcome to Wrangler at TACC
-----

--> Verifying valid submit host (login1)...OK
--> Verifying valid jobname...OK
--> Enforcing max jobs per user...OK
--> Verifying availability of your home dir (/home/00157/walling)...OK
--> Verifying availability of your work dir (/work/00157/walling/wrangler)...OK
--> Verifying valid ssh keys...OK
--> Verifying access to desired queue (normal)...OK
--> Verifying job request is within current queue limits...OK
--> Checking available allocation (TG-STA110019S)...OK
Submitted batch job 17216

After your idev job begins to run, a command prompt will appear,
and you can begin your interactive development session.
We will report the job status every 4 seconds: (PD=pending, R=running).

job status: PD
job status: R
--> Job is now running on masternode= c251-124...OK
--> Sleeping for 7 seconds...OK
--> Checking to make sure your job has initialized an env for you....OK
--> Creating interactive terminal session (login) on master node c251-124.
c251-124.wrangler(1)$ |
```

Job Queues

User guide: <https://portal.tacc.utexas.edu/user-guides/wrangler#wrangler-production-queues>

sinfo shows state of queues

showq/squeue provide jobs status

```
c251-114.wrangler(11)$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
all        up 2:00:00:00      2 drain* c252-107,c253-109
all        up 2:00:00:00      8 down* c253-[102-108,110]
all        up 2:00:00:00     24 resv c252-[101-106,118-135]
all        up 2:00:00:00     10 alloc c252-[108-117]
all        up 2:00:00:00      8 idle c252-[136-143]
normal    up 2:00:00:00      4 maint c251-[140-143]
normal    up 2:00:00:00      1 drng c251-102
normal    up 2:00:00:00      5 drain c251-[101,103,112,136,138]
normal    up 2:00:00:00     21 resv c251-[104,106-111,117-129,137]
normal    up 2:00:00:00      7 alloc c251-[105,113-116,134-135]
normal    up 2:00:00:00      5 idle c251-[130-133,139]
hadoop   up 2:00:00:00      1 drain* c252-107
hadoop   up 2:00:00:00     24 resv c252-[101-106,118-135]
hadoop   up 2:00:00:00     10 alloc c252-[108-117]
hadoop   up 2:00:00:00      8 idle c252-[136-143]
debug*    up 2:00:00:00      1 drain* c253-109
debug*    up 2:00:00:00      8 down* c253-[102-108,110]
debug*    up 2:00:00:00      1 resv c252-135
debug*    up 2:00:00:00      8 idle c252-[136-143]
```

showq

```
login1.wrangler(34)$ showq
ACTIVE JOBS-----
JOBID      JOBNAME    USERNAME     STATE   CORE  REMAINING  STARTTIME
=====
37921      jarositeE_ tg834467    Running 48      13:04:42  Tue Apr 11 22:29:55
37932      jarositeJ_ tg834467    Running 48      13:35:52  Tue Apr 11 23:01:05
37933      jarositeJ_ tg834467    Running 48      13:35:55  Tue Apr 11 23:01:08
37934      jarositeJ_ tg834467    Running 48      13:35:55  Tue Apr 11 23:01:08
37955      BaseSc1      msimoni    Running 48      31:42:43  Wed Apr 12 17:07:56
37971      convertFil sjf826    Running 480     12:16:50  Wed Apr 12 21:42:03
38002      Rstudio      tg804356    Running 48      3:40:12   Thu Apr 13 09:05:25

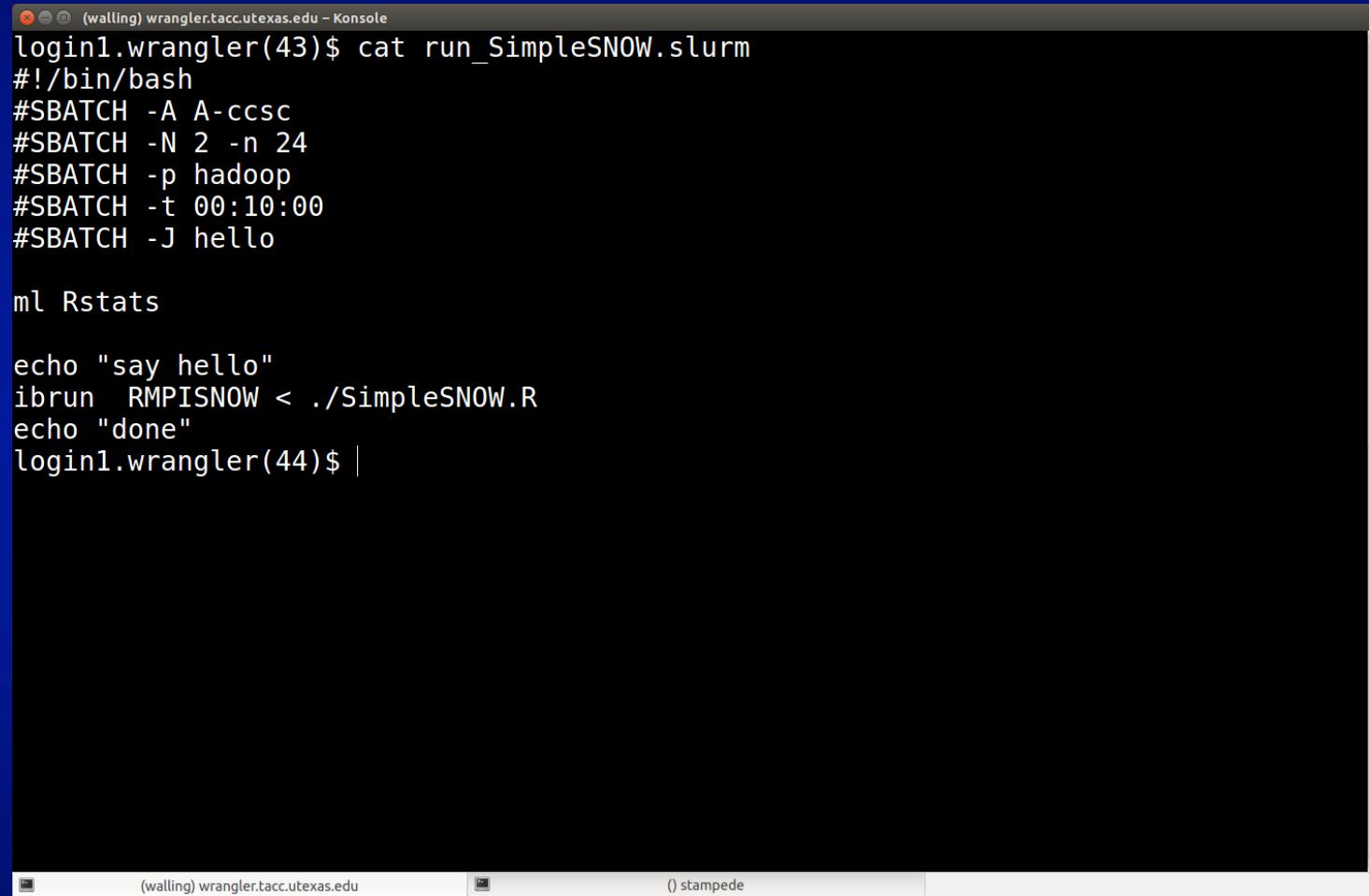
    7 active jobs

WAITING JOBS-----
JOBID      JOBNAME    USERNAME     STATE   CORE  WCLIMIT    QUEUETIME
=====
38004      LaphOp_32_ tg458273    Waiting 216     10:00:00  Thu Apr 13 09:19:27
38005      LaphOp_32_ tg458273    Waiting 216     10:00:00  Thu Apr 13 09:19:30

COMPLETING/ERRORED JOBS-----
JOBID      JOBNAME    USERNAME     STATE   CORE  WCLIMIT    QUEUETIME
=====
38001      idv40561   walling    Complete 48      3:11:15   Thu Apr 13 08:36:28

Total Jobs: 9      Active Jobs: 7      Idle Jobs: 2      Blocked Jobs: 0
```

Running Jobs - sbatch



(walling) wrangler.tacc.utexas.edu - Konsole

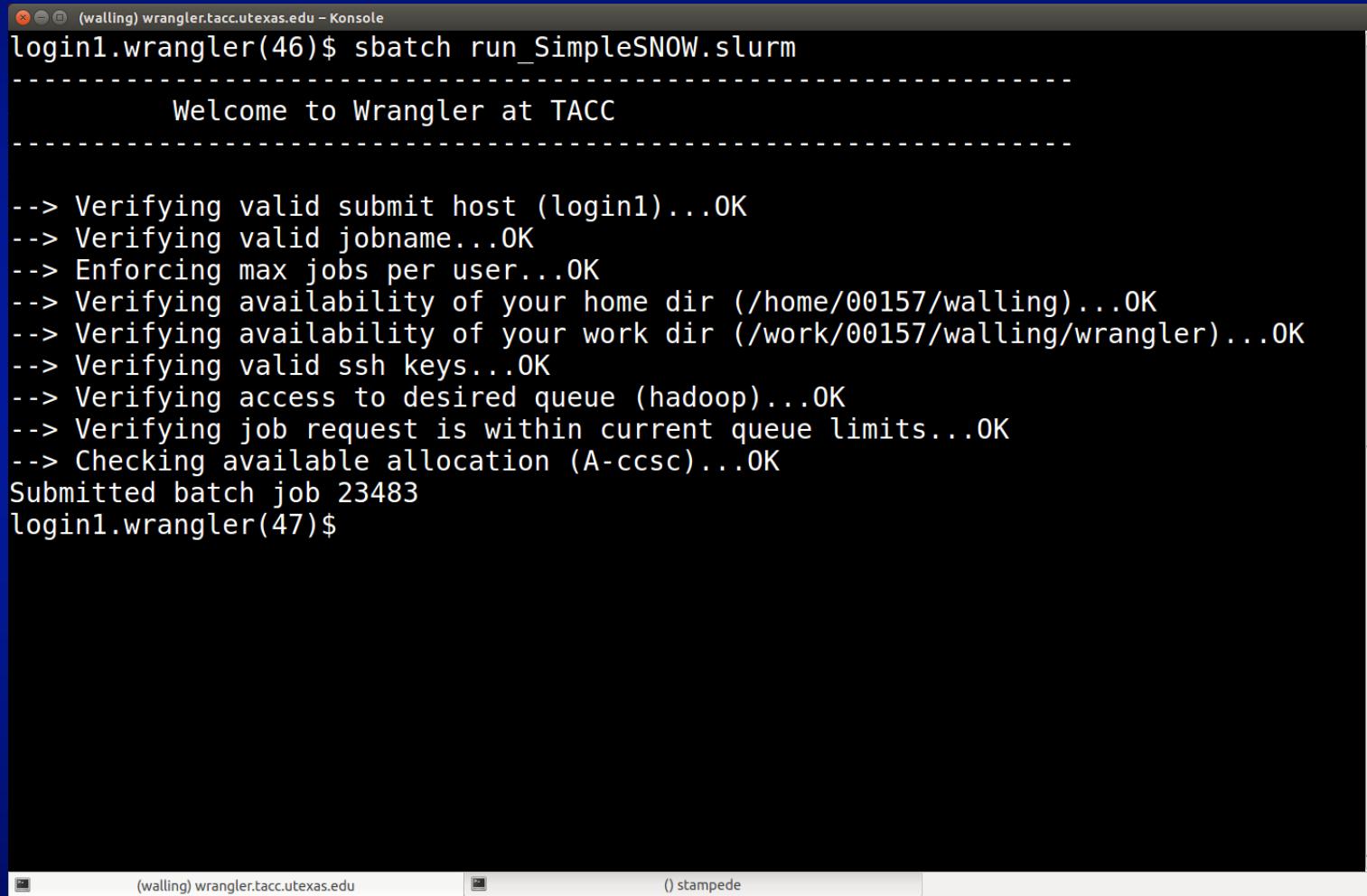
```
login1.wrangler(43)$ cat run_SimpleSNOW.slurm
#!/bin/bash
#SBATCH -A A-ccsc
#SBATCH -N 2 -n 24
#SBATCH -p hadoop
#SBATCH -t 00:10:00
#SBATCH -J hello

ml Rstats

echo "say hello"
ibrun RMPISNOW < ./SimpleSNOW.R
echo "done"
login1.wrangler(44)$ |
```

(walling) wrangler.tacc.utexas.edu () stampede

Running Jobs - sbatch



(walling) wrangler.tacc.utexas.edu - Konsole

```
login1.wrangler(46)$ sbatch run_SimpleSNOW.slurm
-----
          Welcome to Wrangler at TACC
-----
--> Verifying valid submit host (login1)...OK
--> Verifying valid jobname...OK
--> Enforcing max jobs per user...OK
--> Verifying availability of your home dir (/home/00157/walling)...OK
--> Verifying availability of your work dir (/work/00157/walling/wrangler)...OK
--> Verifying valid ssh keys...OK
--> Verifying access to desired queue (hadoop)...OK
--> Verifying job request is within current queue limits...OK
--> Checking available allocation (A-ccsc)...OK
Submitted batch job 23483
login1.wrangler(47)$
```

(walling) wrangler.tacc.utexas.edu () stampede

Job Status

```
(walling) wrangler.tacc.utexas.edu - Konsole
login1.wrangler(75)$ squeue
      JOBID   PARTITION     NAME      USER ST          TIME  NODES NODELIST(REA
SON)
        22423      normal  DD-Group  tg834792 CG 1-23:57:18      1 c251-101
        18500      normal  rankboos  mkutlu86 CG      4:12      1 c251-126
        18479      normal  DD-Group  tg834792 CG      2:40:19      1 c251-125
        18466      normal  DD-Group  tg834792 CG      0:00      1 c251-128
        23484      normal    hello   walling   R      0:06      2 c251-[103-10
4]
login1.wrangler(76)$ squeue -u walling
      JOBID   PARTITION     NAME      USER ST          TIME  NODES NODELIST(REA
SON)
        23484      normal    hello   walling   R      0:09      2 c251-[103-10
4]
login1.wrangler(77)$
```

Cancel Jobs - scancel

```
(-) stampede.tacc.utexas.edu - Konsole
login3.stampede(18)$ squeue -u walling
  JOBID      PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
    7670839      normal      hello   walling PD      0:00      2 (Reservation)
login3.stampede(19)$ scancel 7670839
login3.stampede(20)$ squeue -u walling
  JOBID      PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
login3.stampede(21)$ |
```

View Reservations

```
(walling) wrangler.tacc.utexas.edu - Konsole
login1.wrangler(20)$ scontrol show reservations
ReservationName=dssd+TG-CIE150028+1829 StartTime=2016-09-18T00:25:01 EndTime=2016-10-18T00:25:01 Duration=30-00:00:00
    Nodes=c251-102 NodeCnt=1 CoreCnt=48 Features=(null) PartitionName=normal Flags=
        Users=chengkai,rohitb,njayaram,ftmdgn,gxzz2070 Accounts=(null) Licenses=(null) State=ACTIVE

ReservationName=Maintenance-SW-upgrades-CDH-GPFS-and-dssdD5 StartTime=2016-10-04T08:00:00 EndTime=2016-10-04T17:00:00 Duration=09:00:00
    Nodes=c251-[101-143],c252-[101-143] NodeCnt=86 CoreCnt=4128 Features=(null) PartitionName=(null) Flags=MAINT,OVERLAP,IGNORE_JOBSSPEC_NODES
        Users=root,ctjordan,ngaffney,xwj Accounts=(null) Licenses=(null) State=INACTIVE

ReservationName=dssd+TRAINING-HPC+1849 StartTime=2016-09-29T14:35:02 EndTime=2016-10-01T14:35:02 Duration=2-00:00:00
    Nodes=c251-[103-105,107-115,117-122,124-135] NodeCnt=30 CoreCnt=1440 Features=(null) PartitionName=normal Flags=IGNORE_JOBSSPEC_NODES
        Users=jasona,jbsnead,taccytr,walling,amritkar,kasperon,dycding,zshi2,jjtapia,ahummos,hauskins,rata8212,pontiggi,ras,wfeinst,shs159,yakushin,sramesh,jzhu03,pgurunat,gao154,nsmolin,tg829961,tg832717,xczhang,tg835453,olumayor,chenyanh,suma,eldouss,subramon,akshay,ztan818,zhengxu,edlin82,shengw,yxiang,gazula,gcreager,tg459470,lzhuo,melrobin,csheng,jemohl,mbrios,fbalbi,rongyao,jepusto,deisejpg,attafynn,tdevitt,aurelien,syk464,jv25669,kujj2009,gpcarpen,jchen24,xfan,cl38899,dwaters,sdaniel,hpotti,aserrato,tlovorn,rrc,reemalis,yanghf14,lyeager,lhu,yinxi,acs4659,cschank,grapenut,awh394,lot5723,kcb734,mhaddad,anvariabs,chenk,emarkham,amt3433,ms72274,yh896235,chuchu,mario357,maboytes,adeering,zhanglw Accounts=(null) Licenses=(null) State
```

Use Reservation

-r, --reservation to sbatch and idev

```
login1.wrangler(28)$ idev -p hadoop -r hadoop+TRAINING-HPC+2168 -A TRAINING-HPC
Defaults file      : ~/.idevrc
Default project   : TG-STA110019S
Default time       : 30 min.
Default queue      : debug
System             : Wrangler
Using queue        : -p hadoop
Reservation        : hadoop+TRAINING-HPC+2168
Using Project      : -A TRAINING-HPC
Reservation        : --reservation=hadoop+TRAINING-HPC+2168 (ACTIVE)
-----
          Welcome to Wrangler at TACC
-----
--> Verifying valid submit host (login1)...OK
--> Verifying valid jobname...OK
--> Enforcing max jobs per user...OK
--> Verifying availability of your home dir (/home/00157/walling)...OK
--> Verifying availability of your work dir (/work/00157/walling/wrangler)...OK
--> Verifying valid ssh keys...OK
--> Verifying access to desired queue (hadoop)...OK
--> Verifying job request is within current queue limits...OK
--> Checking available allocation (TRAINING-HPC)...OK
Submitted batch job 37996
```

RStudio @ TACC

URL: <http://vis.tacc.utexas.edu>

Click 'Jobs'

Select

Resource = Wrangler

Project = TRAINING-HPC

Session Type = R Studio

Queue = normal

Click 'Start Job'

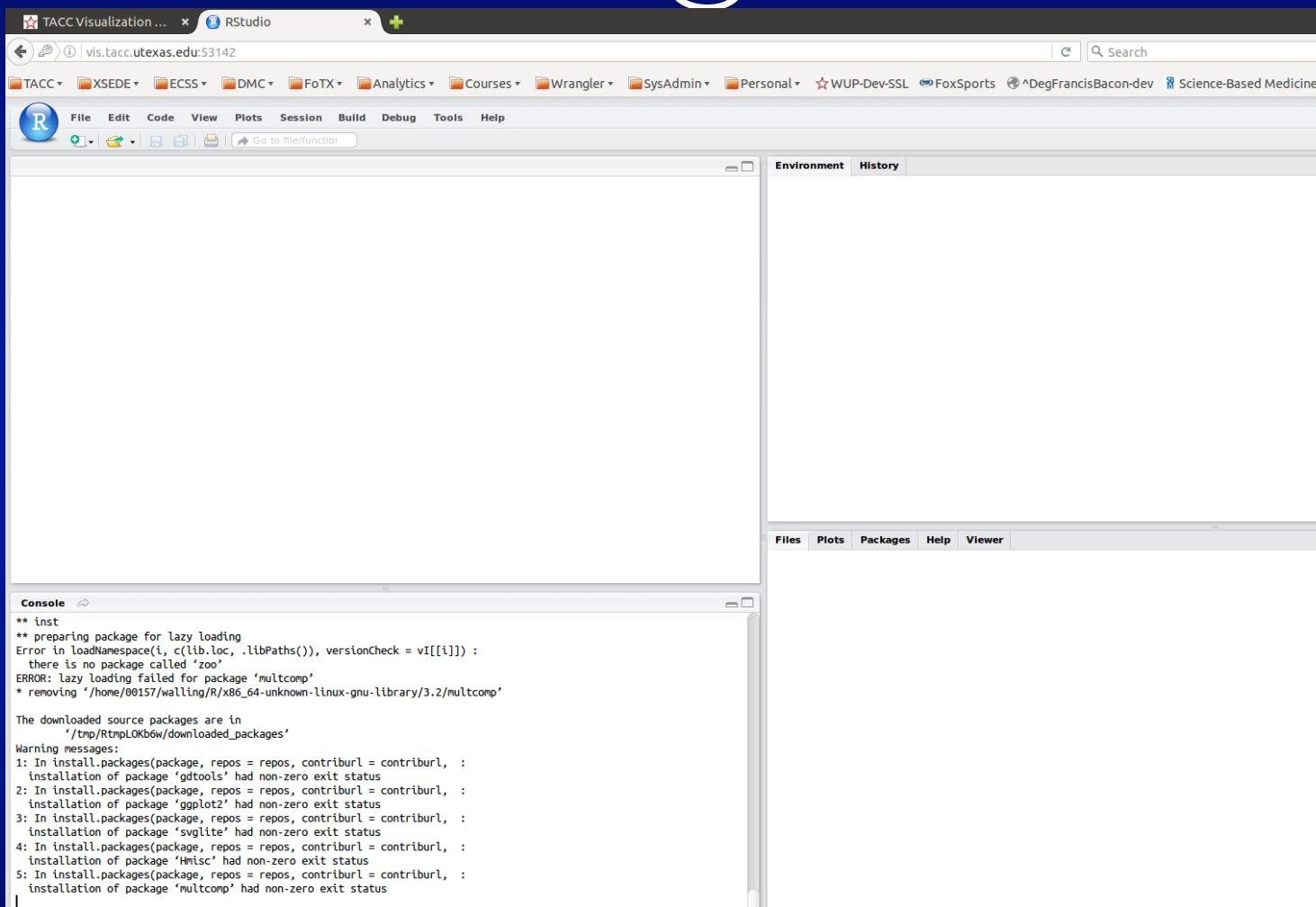
Once job is running, click 'Open in browser'

Currently single node jobs only.

Limited to 4 hour max

The screenshot shows the 'TACC Visualization Portal' interface. At the top right, it says 'TACC\walling logout' and 'No job running.' Below the header is a navigation bar with 'Home', 'Jobs' (which is selected), and 'Help'. The main area is titled 'Start a Job'. It has several sections: 'Resource' (Maverick, Stamped, Wrangler, Wrangler is selected), 'Project' (TRAINING-HPC), 'Session type' (VNC, iPython/Jupyter Notebook, R Studio, R Studio is selected), 'Reservation ID' (normal+SC15), 'Queue' (normal), and a 'Start Job ▶' button. A note at the bottom says 'Use your TACC username and password to authenticate to RStudio'.

RStudio @ TACC



David Walling

walling@tacc.utexas.edu

