HPC Workshop

Nov. 9, 2018

James Coyle, PhD

Dir. Of High Perf. Computing

NEEDED EQUIPMENT

1. Laptop with Secure Shell (ssh) for login

A. Windows: download/install putty from https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html

B. Mac Os: command line SSH Using Terminal
X apps download/install from https://www.xquartz.org/

C. Linux : open-ssh likely installed, if not, then

RedHat/Centos: yum install open_ssh X11
Ubuntu: apt-get install openssh-client X11

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/ACCESS-AND-LOGIN

To login to any HPC machines, must be

- on the iastate.edu network or
- using ISU VPN installed from http://vpn.iastate.edu

On first login from any machine, you will first be asked if you want to store the condo host key. Answer YES.

The first time you login to condo, you will receive an email that you do not have MFA configured.

You will be sent instructions on how to install the Google authenticator application. You should also install a Bar/QR code reader using the apple AP Store or Google Play.

If you want to install a second device, make sure to have both devices ready, or copy the 15 character code that you can type in later.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/ACCESS-AND-LOGIN

When you login in the second time you will be prompted with

Verification:

This is when you type the 6 digit code displayed by the Google Authentication App. This changes every 30 seconds.

Then you will be prompted for

Password:

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/ACCESS-AND-LOGIN

MFA SELF RESET

Now you should be logged in and will see the MOTD, which has a normal message, plus current announcements.

You will next be prompted with the question.

Do you want to store a phone number to allow you to reset your Google authenticator application (Y/N)? Enter YES.

Then enter the 10 digit number where you will get text messages. This will allow you to reset your Google Auth. App if you get a new phone. This assumes you keep the same phone number. If you don't or if you've answered No, you will need to send a message to hpc-help@iastate.edu.

The self reset url is **https://hpc-ga1.its.iastate.edu/reset/**

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/UNIX-INTRODUCTION
REDHAT LINUX VERSION 7

Now you are at the bash shell prompt, which means that you can start entering Linux commands.

If you are not familiar with the Linux command line, and introduction can be found at: https://www.hpc.iastate.edu/guides/unix-introduction

Now that you are logged in, you will be in your home directory,

If you type

cd with no arguments, you will return home

If you type

pwd (print work in directory) you will see the current working directory.

You need to use a text editor or cat to create files:

Editors: nano, gedit, vi and emacs

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/STORAGE

Homedirs are small. Use them for configuration files.

Research machines:

Group /work/ directory for each condo group

E.g Mine is: /work/ccresearch.

ISU specific command cdw will cd to your work dir, unless you are in one of the LAS groups.

/work/ is available on all cluster nodes.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/STORAGE

Your group can see the files in /work/ and can make new files and directories. Whoever makes them is the owner, and others cannot remove or modify them.

Other groups cannot access your group's /work dir.

Your group shares the same file system quota. You can see this with quota -gs .

/work is backed up nightly.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/STORAGE

/work use the zfs file system which uses compression (actually faster) no need to gzip

du /work shows the compressed size on disk;

du –apparent-size shows the uncompressed usage.

Reasonable fast if you stay below 70% full.

Do not fill to 100% .

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/STORAGE

Other storage:

/ptmp on some clusters, no backup, files deleted after some number of days from creation, 90 on Condo

To use, mkdir –p /work/GROUPNAME/USERNAME

$TMPDIR   job specific space on a single node

Beta: On Condo for 2 or more nodes in specific queue

$PTMPDIR   parallel space on dedicated nodes in job

MyFiles; LSS available on Data Transfer Node only.

LSS is backed off-site, $40/TB/yr.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/FILE-TRANSFERS

For data transfers on condo use condodtn.its.iastate.edu (dtn = data transfer node).

You can use

scp file condodtn.its.iastate.edu:/work/GROUPNAME

or you can use

sftp condodtn.its.iastate.edu

or better yet, use Globus connect. Details on using Globus Connect are at

http://hpc.iastate.edu/guides/condo2017/Globus_Online

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/MANAGING-JOBS-USING-SLURM-WORKLOAD-MANAGER

The login node condo.its.iastate.edu is to be used for editing files, submitting jobs, and viewing output. It can also be used for compiling, provided this does not interfere with the usability of the login node. Do not run memory/compute or data intensive processes (including transfers) on it.

To run large applications, use the SLURM batch scheduling system.

To write a batch script, go to

https://www.hpc.iastate.edu/guides/condo-2017/slurm-job-script-generator-for-condo

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/SLURM-JOB-SCRIPT-GENERATOR-FOR-CONDO

Select the number of compute nodes that your applications will use, and the type of node.

Unless your application is written to use multiple nodes, you should select one node. Some prewritten applications like Fluent, OpenFoam or GAMESS and applications written to use MPI, can use more than one node. Most BioInformatics programs, for example, do not use more than one node.

You also need to specify the number of processor cores you will be using. If your application does not use threads or openmp to use multiple cores, you should just request one processor per node.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/SLURM-JOB-SCRIPT-GENERATOR-FOR-CONDO

You should also specify the maximum memory that each process is going to use. You can specify a high number, then watch the process memory needs using qtop JOBNUMBER.

You must also specify the maximum length of time that the job will run. This is real time, not cputime.

Lower requirements on a job allow it to start running faster, but too small of a time or resources will cause your job to end before it completes. We do not extend runtimes, so make sure that you specify a long enough runtime.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/SLURM-JOB-SCRIPT-GENERATOR-FOR-CONDO

If you are part of a group that bought into the Condo Cluster, you will run on the regular condo nodes, 16 cores/128 GB of memory, so just under 8GB per core. (The OS needs some memory too.)

These users can also access 2 1TB memory fat nodes (32 cores) and one 2TB huge node (40 cores). These were expensive, and consume much more of your group's allocation when you run, so only use them when you need them.

Free tier: Other groups can be granted access to the Free tier of Condo: Currently 64 GB / 12 cores, so you can use up to 5 GB of memory per core. We hope to expand this as time goes on.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/SLURM-JOB-SCRIPT-GENERATOR-FOR-CONDO

Once you finish:

- highlight the script that is generated with your mouse,
- copy that to your clipboard,
- go to the condo terminal window, and
- in that window issue:

  cat > myjob
- Then paste in that window, and
- issue CNTL-D to end the file.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/SLURM-JOB-SCRIPT-GENERATOR-FOR-CONDO

Now you can edit the file myjob with the editor of your choice, vi, emacs, gedit, etc. to add a command you wish to execute.

Let’s add the commands:

pwd

ls

date

Then save

SUBMIT YOUR JOB USING SLURM

sbatch myjob

The job will go into the queue XXXX and will probably exit before you can even see it.

To see which jobs are in the queue, issue

squeue

to get an estimate on when a job will start, issue

squeue –a –start

This is an estimate and may be earlier if jobs end before the time they reserved, or may be later if higher priority jobs are submitted after your job is submitted.

SUBMIT YOUR JOB USING SLURM

Priority is based on what percentage of your group's allocation has been used, and to a smaller extend on the time the job has been waiting. The allocation is based on how many nodes were bought by your group to be shared.

If you copy myjob to your work directory, and issue sbatch myjob again, you will see

that the job runs from the directory from which it was submitted.

INTERACTIVE BATCH SESSION

In rare cases, you will need an interactive session. Interactive session are more expensive, since you are charged for the full time whether you have commands running or not, so use this sparingly. To create n interactive session, issue

salloc

with the amount of resources needed.

E,g,

salloc --time=1:00:00 --nodes=1 --ntasks-per-node=1 --mem=5G --partition=freecompute

When using the batch scheduler, if there are enough resources, the jobs just run. When there are not enough, some jobs have to wait, and these are the lower priority jobs. These will be marked with ‘PD’ as the state.

In addition to priority based scheduling, we have a limit on the number of jobs running from a single user. This is to prevent a single user from shutting out other users, since it is a shared system.

BEST PRACTICES

Use Globus for transfers

Use condodtn.its.iastate.edu if you use sftp or scp or other data transfer tools.

Do not run intensive tasks on the login node, use salloc –p freecompute

Keep files that you need in /work , not in /ptmp, /tmp, /var/tmp or on compute nodes

Files in /work and /home directories are backed up. Files in /ptmp are not backed up and are purged after a given period of time.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/CONDO-2017/STORAGE

Don’t try to keep files on computes nodes, that data is deleted after the job completes.

$TMPDIR exists on each compute nodes, and is about 2.5TB on condo use it for temporary files.

For I/O intensive tasks, copying files to $TMPDIR and using that directory will greatly improve I/O versus accessing over the network. You must copy files back before the job ends though. This is safest for merely scratch files on one compute node.

The **timeout** command is useful to stop a command from taking too long so that you can copy back output files if needed.

We are working on $PTMPDIR, which will provide scalable I/O performance.

BEST PRACTICES

Reserve cores corresponding to memory.
5 GB/core for free tier 8 GB for regular nodes.

Software has been installed in modules:
module avail
module load
module unload
module list
module purge
module use

HELP

http://hpc.iastate.edu

1. Guides
2. FAQ
3. man pages for information about a command
4. hpc-help@iastate.edu

HTTPS://WWW.HPC.IASTATE.EDU/FAQ

How do I Install my own R package ?

https://www.hpc.iastate.edu/faq#Rpack

How do I install my own Python package?

https://www.hpc.iastate.edu/faq#python-package

…..

INDIVIDUAL HELP

We have HPC staff available to help you with specific difficulties you may be having.

We'll be here until 4:00 to help you.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/GLOBUS-ONLINE
GLOBUS PERSONAL CONNECT INSTALL

To use Globus Connect, we will first install Globus Personal Connect so that you can use it from your laptops.

Navigate to http://www.globus.org

Click on login

Your organizational login will be Iowa State University.

Click continue

Then login using your NetID.

.

HTTPS://WWW.HPC.IASTATE.EDU/GUIDES/GLOBUS-ONLINE
GLOBUS PERSONAL CONNECT INSTALL

Click on the link which says Install Globus Personal Connect

Now, enter a name, like My Laptop

And click on Generate Setup Key

And click on copy to clipboard (you will need this during the install).

Now click on your Laptop's OS to get download the installer for your type of operating system. Once the download completes, click on the installer that was downloaded.

Click through the screens, and paste the setup key in when requested.

Now, back to the two endpoint screens, click on one endpoint textbox, and click on the Managed by me tab, and select the My Laptop endpoint you just created.

In the other endpoint window, type condodtn.its.iastate.edu , and select that endpoint when it comes up.

For username, use your NetID

For the password, type in the password that goes along with your netID followed by the 6 digit Google Authenticator code for the Condo Cluster.

Now navigate to your work directory in the Condo pane, and select a file, e.g. the myjob file we just created.

Now click on the arrow between the endpoints to point to your laptop endpoint. This will start the transfer. You can click on Activity to show the activity.

If you want to see better transfer rates, go to the laptop endpoint textbox and start typing ESNET

And select the CERN ESNET readonly test server

You will not have to authenticate. I will do a transfer. I'll select the 1G.dat file, and copy into /work/ccresearch/jjc

I will also unselect verify after transfer.

Clicking on Activity, we watch and see X Mbytes per second. The 10G.dat takes longer, but gives a better rate.

FAQ: INSTALLING LOCAL PACKAGES : R

```
#  mkdir -p ~/local/R_libs
#  export  R_LIBS=~/local/R_libs/
#  module load r
#  R
>  install.packages("abc", repos="http://cran.r-
   project.org", lib="~/local/R_libs/")
>  .libPaths();
>  quit()
```

R package now installed, and the export above allows it to be used.

COMPILING FOR BEST PERFORMANCE

The Intel Compilers are standards conforming, they work for all programs that adhere to the current Fortran, C or C++ standard. GNU compilers can also be used, but Intel will generally be faster.

module load intel

ifort for fortran; icc compiles either C or icpc

Use –O3 –xHOST

-xHOST is not portable, so recompile for different architecture. On Condo, fat and huge nodes need to have the compile on them, since they are one arch behind the head node.

COMPILING FOR BEST PERFORMANCE

The Intel Compilers are standards conforming, they work for all programs that adhere to the current Fortran, C or C++ standard. GNU compilers can also be used, but Intel will generally be faster.

module load intel

ifort for fortran; icc compiles either C or icpc

Use –O3 –xHOST

-xHOST is not portable, so recompile for different architecture. On Condo, fat and huge nodes need to have the compile on them, since they are one arch behind the head node.

PROFILING PERFORMANCE COUNTERS

module load intel

ifort for fortran; icc compiles either C or icpc

Use –O3 –xHost –p

ifort –O3 –xHOST –p tmmx.f dgemmi.f

perf stat –d ./a.out

COMPILING FOR DEBUGGING

module load intel
ifort for fortran; icc compiles either C or icpc

Use –g –C
-g allows for debugging information
-C check for out of bounds references in arrays

gdb available
ddt available for parallel programs; module load allinea

GENERAL RULES FOR FAST PROGRAMS

Eliminate unnecessary work

With nested ifs , put most likely to fail first

Keep branches outside of innermost loop (ifs, case)

In matlab or similar programs use array statements

In Perl, associative arrays can be very efficient

In Python, dictionaries and sets can be very efficient

Use libraries where possible

Go parallel within a server (36 cores on nova)

  threads, parallel command, OpenMP

Go parallel across servers MPI

GENERAL RULES FOR FAST PROGRAMS

Computing is both about work done, and logistics

Less work

Have the data used most nearest to the CPU

Resisters,

Caches L1,2,3,

On node memory: local memory, NUMA memory, distributed memory

Local disk (SSDs especially) $TMPDIR or $PTMPDIR

Fileservers (can be faster, but low IOPs less predictable)

GNU PARALLEL

Get lots of single cpu work done at the same time.

In /work/ccresearch/jjc/Workshop

32 copies of tmmx_95
parallel with –j 32 is about
half the time as –j 16
¼ the time as –j 8
Similarly for smaller numbers of –j

If you have more processes that processors, -j 36 will
run just 36 at a time, starting a new processes each time
of the 36 ends, so order commands in descending order of run-time to minimize total runtime.