GACRC

Introduction to HPC Using zcluster at GACRC

On-class STAT8330

Georgia Advanced Computing Resource Center
University of Georgia
Suchitra Pakala
pakala@uga.edu
Slides courtesy: Zhoufei Hou

Outline

- What is GACRC?
- What is zcluster?
- How does zcluster operate?
- How to work with zcluster?

What is GACRC?

Who Are We?

- **G**eorgia **A**dvanced **C**omputing **R**esource **C**enter
- Collaboration between the Office of Vice President for Research (**OVPR**) and the Office of the Vice President for Information Technology (**OVPIT**)
- Guided by a faculty advisory committee (GACRC-AC)

Why Are We Here?

- To provide computing hardware and network infrastructure in support of high-performance computing (**HPC**) at UGA

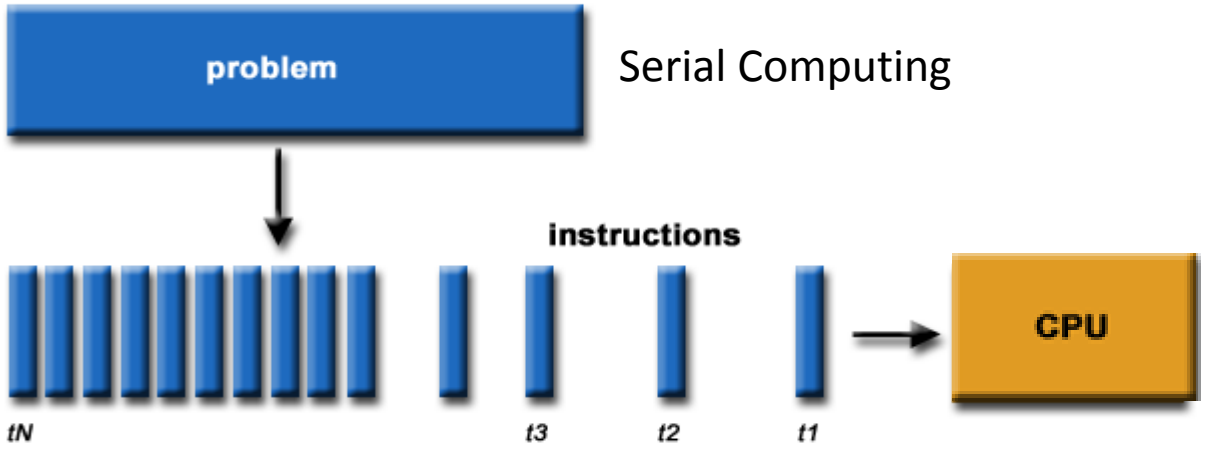
Where Are We?

- http://gacrc.uga.edu (Web) http://wiki.gacrc.uga.edu (Wiki)
- http://gacrc.uga.edu/help/ (Web Help)
- https://wiki.gacrc.uga.edu/wiki/Getting_Help (Wiki Help)

Some Keywords…

- NODE – a single computer
- JOB – unit of work defined by a script that is run on cluster
- QUEUE – order in which the submitted jobs are run
- CLUSTER - set of computers connected together so that, in many respects, they can be viewed as a single system.

High Performance Computing (HPC)

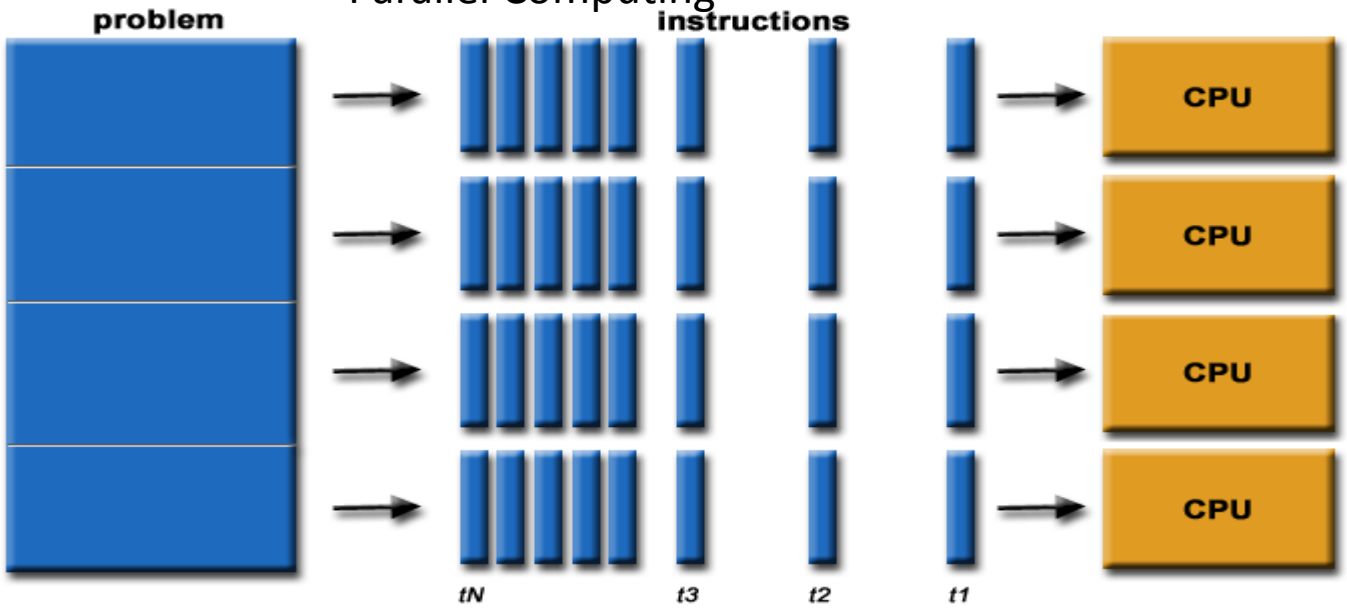


Serial Computing

Serial Computing

- A problem is broken into a discrete series of instructions
- Instructions are executed sequentially
- Executed on a single processor
- Only one instruction may execute at any moment in time

Parallel Computing



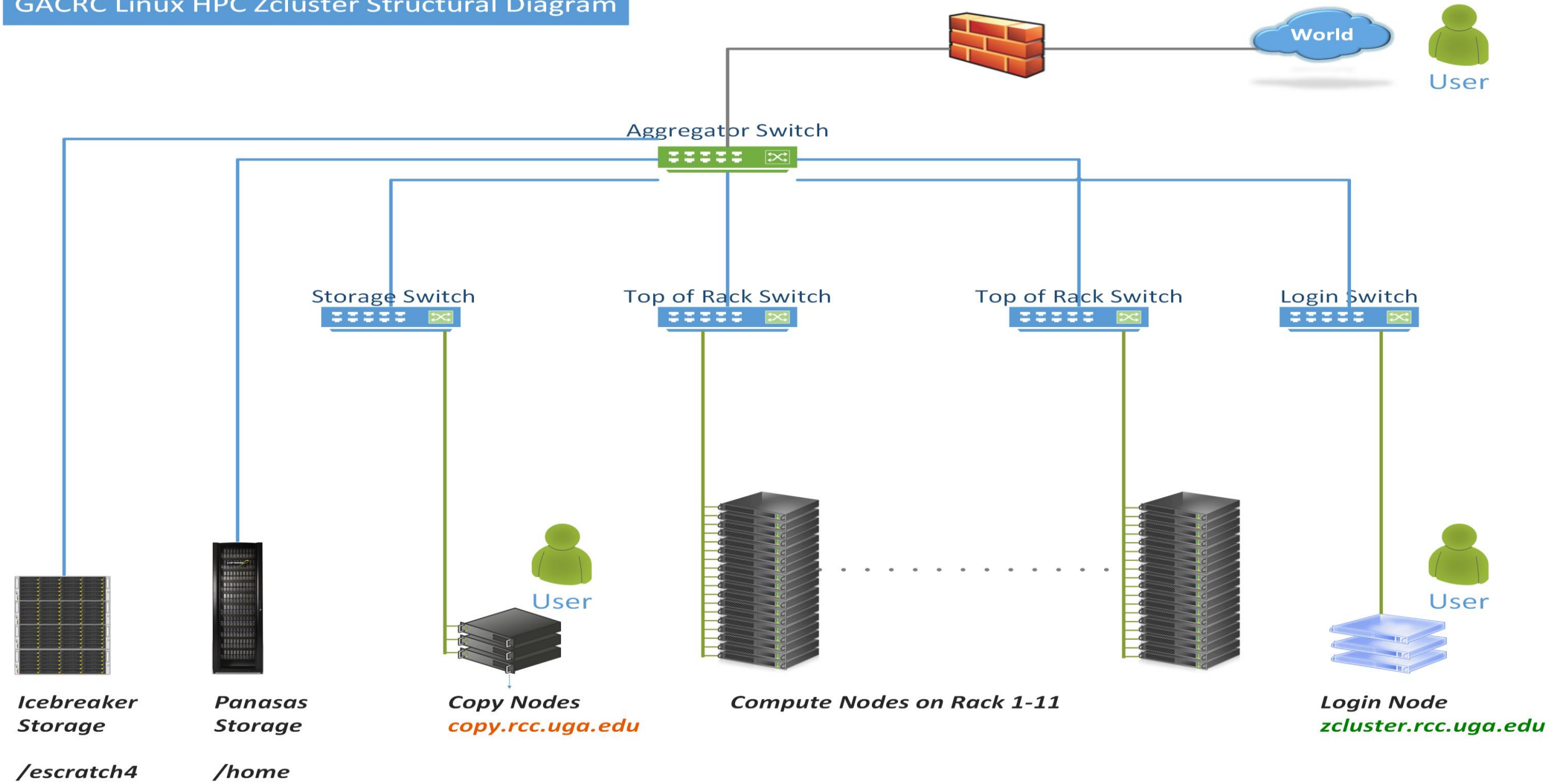
Parallel Computing

- A problem is broken into discrete parts that can be solved concurrently
- Each part is further broken down to a series of instructions
- Instructions from each part execute simultaneously on different processors
- An overall control/coordination mechanism is employed

zcluster Overview

- Cluster Structural Diagram
- General Information
- Computing Resources
- Storage Environment

GACRC Linux HPC Zcluster Structural Diagram



zcluster – General Information

GACRC zcluster is a Linux high performance computing (HPC) cluster:

- Operating System: 64-bit Red Hat Enterprise Linux 5 (RHEL 5)
- Login Node: zcluster.rcc.uga.edu
  zcluster.rcc.uga.edu —`qlogin`→ Interactive Node: compute-14-7/9
- Copy Node: copy.rcc.uga.edu

zcluster – General Information

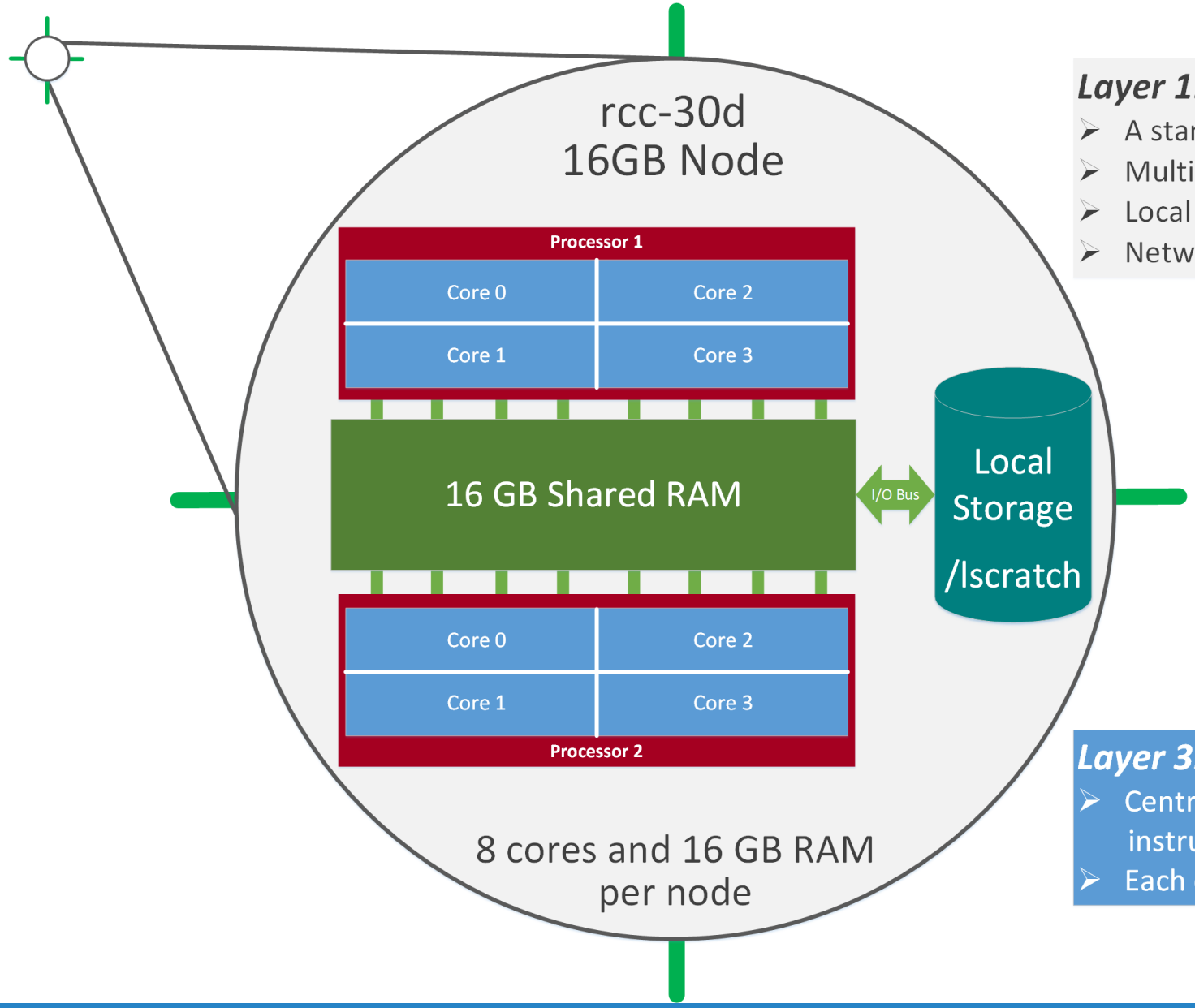
- Batch-queueing System:
  - Jobs can be started (submitted), monitored, and controlled
  - Determine which compute node is the best place to run a job
  - Determine appropriate execution priority for a job to run

- On zcluster: Sun Grid Engine (SGE)

zcluster – Computing Resources

Queue Type	Queue Name	Nodes	Processor	Cores/Node	RAM(GB)/Node	Cores	NVIDIA GPU
Regular	rcc-30d	45	Intel Xeon	12	48	540	N/A
		150		8	16	1200	
High Memory	rcc-m128-30d	1	Intel Xeon	8	128	8	N/A
		4		8	192	32	
		10		12	256	120	
	rcc-m512-30d	2		32	512	64	
Multi Core	rcc-mc-30d	6	AMD Opteron	32	64	192	N/A
Interactive	interq	2	AMD Opteron	48	132	96	N/A
GPU	rcc-sgpu-30d	2	Intel Xeon	8	48	16	4 Tesla S1070 cards
	rcc-mgpu-30d	2		12	48	24	9 Tesla (Fermi) M2070 cards
	rcc-kgpu-30d	4		12	96	24	32 Tesla (Kepler) K20Xm cards

***Total peak performance:* 23 Tflops**



Layer 1: Node

- A standalone "computer in a box"
- Multiple processors, e.g. 2, sharing memory
- Local disk storage, network interface, etc.
- Networked into a cluster

Layer 2: Processor

- A single computing component
- Multicore processor, e.g. 4 cores

Layer 3: Core

- Central processing unit (CPU) reading and executing instructions independently
- Each core is assigned to a software thread

Zcluster -Storage Environment

- **Home directory** ➔ /home/student/STAT8330/s_85
  - Mounted and visible on all nodes, with a quota of ~100GB

- **Ephemeral Scratch** ➔ /escratch4/s_85/s_85_Nov_02
  - Create with `make_escratch` command at Login Node
  - Visible to all nodes with a quota of 4TB
  - To be deleted after 37 days

zcluster – Storage Environment

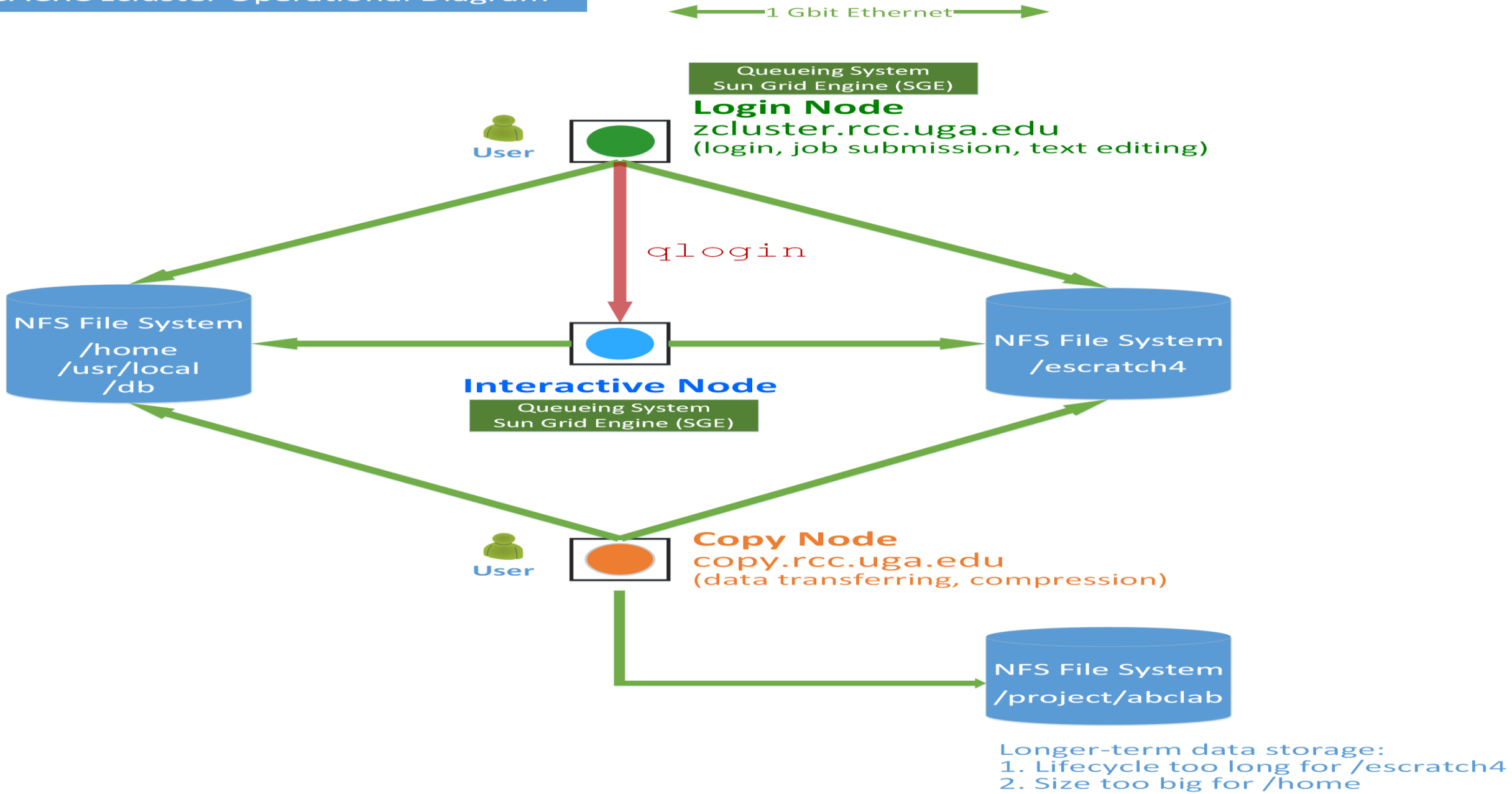
2 Filesystems	Role	Quota	Accessible from	Intended Use
/home/student/STAT8330/ s_85	Home	100GB	zcluster.rcc.uga.edu (Login) copy.rcc.uga.edu (Copy) Interactive nodes (Interactive)	Highly static data being used frequently
/escratch4/s_85	Scratch	4TB		Temporarily storing large data being used by jobs

3 Main Functions	On/From-Node	Related Filesystem
Login Landing	Login or Copy	/home/student/STAT8330/s_85 (Home) (Always!)
Batch Job Submitting	Login or Interactive	/escratch4/s_85 (Scratch) (Suggested!) /home/student/STAT8330/s_85 (Home)
Data Archiving , Compressing and Transferring	Copy	/escratch4/s_85 (Scratch) /home/student/STAT8330/s_85 (Home)

How does zcluster operate?

Next Page

GACRC zcluster Operational Diagram



How to work with it?

- Cluster's CCOC:

On cluster, you are not alone……. Each user is sharing finite resources, e.g., CPU cycles, RAM, disk storage, network bandwidth, with other researchers. *What you do may affect other researchers on the cluster*.

6 rules of thumb to remember:

- NO jobs running on login node
- NO multi-threaded job running with only 1 core requested
- NO large memory job running on regular nodes
- NO long job running on interactive node
- NO small memory job running on large memory nodes
- Use the copy node for file transfer and compression

How to work with zcluster?

- Start with zcluster
- Connect & Login
- Transfer Files
- Software Installed
- Run Interactive Jobs
- Submit Batch Jobs
  - How to submit ***serial, threaded,*** and ***MPI*** batch jobs?
  - How to check job status and cancel a job?
  - How to check memory usage by jobs?

Getting Started with zcluster

- You need a class user account : s_85@zcluster.rcc.uga.edu
- Use `passwd` to change temporary password to a permanent one

Connect & Login @ zcluster

- Open a connection: Open a terminal and `ssh` to your account

```
ssh s_85@zcluster.rcc.uga.edu
```

or

```
ssh -X s_85@zcluster.rcc.uga.edu
```

([1] `-X` is for X windows application running on the cluster to be forwarded to your local machine

[2] If using Windows, use SSH client to open connection, get from UGA download software page)

- Logging in: You will be prompted for your zcluster password

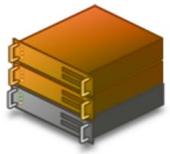
```
s_85@zcluster.rcc.uga.edu's password:
```

([3] On Linux/Mac, when you type in the password, the prompt blinks and does not move)

- Logging out: `exit` to leave the system

```
s_85@zcluster:~$ exit
```

zcluster – Transfer Files

User's local  SCP  Copy node (copy.rcc.uga.edu)

- On Linux, Mac: `scp [Source] [Target]`

  E.g. 1: working on local machine, Local → zcluster

```
scp file1 s_85@copy.rcc.uga.edu:/escratch4/s_85/s_85_Nov_02/
```

```
scp *.dat s_85@copy.rcc.uga.edu:/escratch4/s_85/s_85_Nov_02/
```

  E.g. 2: working on local machine, zcluster → Local

```
scp s_85@copy.rcc.uga.edu:/escratch4/s_85/s_85_Nov_02/file1 ./
```

```
scp s_85@copy.rcc.uga.edu:/escratch4/s_85/s_85_Nov_02/*.dat ./
```

- On Window: SSH Secure Client File Transfer, WinSCP, FileZilla etc.

Software Installed @ zcluster

- Perl, Python, Java, awk, sed, C/C++ and Fortran compilers
- Matlab, Maple, R, AMBER, GKIN
- Many Bioinformatics applications: NCBI Blast+, Velvet, Trinity, TopHat, MrBayes, SoapDeNovo, SAMtoolsetc.
- RCCBatchBlast (RCCBatchBlastPlus) to , Mafft, RAxML, PASTA, MrBayes and MP-EST distribute NCBI Blast (NCBI Blast+) searches to multiple nodes.
- Many Bioinformatics Databases: NCBI Blast, Pfam, uniprot, etc.

https://wiki.gacrc.uga.edu/wiki/Bioinformatics_Databases

- For a complete list of applications: https://wiki.gacrc.uga.edu/wiki/Software

Run Interactive Jobs @ zcluster

- From login node, qlogin command to open an interactive session:

Login Node: zcluster.rcc.uga.edu —qlogin→ Interactive Node: compute-14-7/9

```
s_85@zcluster:~$ qlogin
Your job 1391816 ("QLOGIN") has been submitted
waiting for interactive job to be scheduled ...
Your interactive job 1391816 has been successfully scheduled.
...
compute-14-7.local$
```

← Now I am on compute-14-7, which is an interactive node

- Current maximum runtime is 12 hours
- When you are done, `exit` to log out!

zcluster – Submit Batch Jobs

- Components you need to submit a batch job:
  - Software already installed on zcluster
  - Job submission script to run the software, and
    - ✓ Specify working directory
    - ✓ Export environment variables, e.g.,
      - PATH (searching path for executables)
      - LD_LIBRARY_PATH (searching paths for shared libraries)
- Common commands you need:
  - qsub with specifying queue name, threads or MPI processes
  - qstat, qdel
  - qacct, qsj, etc.

Batch *Serial* Job @ zcluster

Step 1: Create a job submission script fastqc_*sub.sh* running Fastqc tool:

```
#!/bin/bash
```
→ Linux shell (bash)

```
cd /escratch4/pakala/pakala_Oct_25
```
→ Specify and enter (cd) the working directory

```
export PATH=/usr/local/fastqc/latest:${PATH}
```
→ Export command helps to set the Environment variables

```
time fastqc SRR1369670.fastq -o Output_File
```
→ Run Fastqc with 'time' command to measure amount of time it takes to run the application

Step 2: Submit it to the queue:

```
$qsub -q rcc-30d fastqc_sub.sh
```

OR

```
$qsub -q rcc-30d -l mem_total=20g fastqc_sub.sh
```

Submit a job

Your job submission script

to the queue rcc-30d with 16GB RAM/Node

to the queue rcc-30d with 48GB RAM/Node

Batch *Threaded* Job @ zcluster

❖ Step 1: Create a job submission script blast.*sh* running Blast:

```
#!/bin/bash

cd /escratch4/pakala/pakala_Oct_25

time /usr/local/ncbiblast/latest/bin/blastall -p 2 [options]
```

➔ Run Blastall with 2 threads

Number of Threads = Number of Cores Requested

❖ Step 2: Submit it to the queue:

```
$ qsub -q rcc-30d -l mem_total=20g -pe thread 2 ./blast.sh
```

to the queue rcc-30d with 48GB RAM/Node

2 cores requested

Note:
Please use the rcc-mc-30d queue, If using threads more than 8!

Batch *MPI* Job @ zcluster

❖ Step 1: Create a job submission script *sub.sh* running RAxML:

```
#!/bin/bash
cd /escratch4/pakala/pakala_Oct_25

export MPIRUN=/usr/local/mpich2/1.4.1p1/gcc 4.5.3/bin/mpirun

$MPIRUN -np $NSLOTS /usr/local/raxml/latest/raxmlHPC-MPI-SSE3 [options]
```

➔ Define and export environment variable (MPIRUN) for convenient usage

➔ Run RAxML with 20 MPI processes (-np $NSLOTS)

❖ Step 2: Submit it to the queue:

```
$ qsub -q rcc-30d -pe mpi 20 sub.sh
```

20 cores requested,
$NSLOTS will be assigned to 20 automatically, before the job submission script is interpreted

Check and Cancel Jobs @ zcluster

- To check the status of all queued and running jobs: qstat

qstat	➔ shows your job in the pool
qstat –u "*"	➔ shows all the jobs in the pool
qstat –j 12345	➔ shows detailed information, e.g., maxvmem, about the job with JOBID 12345
qstat –g t	➔ list all nodes used by your jobs

- To cancel a queued or running job: qdel

qdel –u pakala	➔ deleted all your jobs
qdel 12345	➔ deletes your job with JOBID 12345

- To list detailed information about a job: qsj, qacct

qsj 12345	➔ shows information, e.g., maxvmem, about the RUNNING job with JOBID 12345
qacct –j 12345	➔ shows information, e.g., maxvmem, about the ENDED job with JOBID 12345

Check and Cancel Jobs

- To check the status of your jobs: qstat

```
$ qstat
job-ID     prior     name      user     state  submit/start at      queue                          slots ja-task-ID
-----------------------------------------------------------------------------------------------------------------------
9707321    0.50766   sub1.sh   s_85     r      01/28/2016 13:39:23 rcc-30d@compute-7-12.local      1
9707322    0.50383   sub2.sh   s_85     Eqw    01/28/2016 13:39:23 rcc-30d@compute-7-12.local      1
9707323    0.00000   sub3.sh   s_85     qw     01/28/2016 13:39:28                                 1
```

- To cancel your job with a JobID: qdel

```
$ qdel 9707322
job-ID     prior     name      user     state  submit/start at      queue                          slots ja-task-ID
-----------------------------------------------------------------------------------------------------------------------
9707321    0.50766   sub1.sh   s_85     r      01/28/2016 13:39:23 rcc-30d@compute-7-12.local      1
9707323    0.00000   sub3.sh   s_85     qw     01/28/2016 13:39:28                                 1
```

Check Memory Usage

- For running jobs: qsj

```
$ qsj 9707368

==============================================
job_number:          9707368
owner:               s_85
cwd:                 /escratch4/s_110/s_110_Jan_28
hard_queue_list:     rcc-30d
script_file:         sub.sh
......
usage    1:          cpu=00:01:27, mem=0.96498 GBs,
                     io=0.00014, vmem=73.734M,
                     maxvmem=73.734M
```

- For finished jobs: qacct

```
$ qacct -j  9707323

==================================
qname          rcc-30d
hostname       compute-7-12.local
jobname        sub.sh
jobnumber      9707323
......
cpu            183.320
mem            2.021
io             0.000
maxvmem        5.230G
```

Total Memory

How to Submit Tickets to GACRC

- For Installation/Downloading Software:
  - User needs to provide the name, version (or latest), and website
  - Applications need to be compatible with Linux
  - Note – only FREE software will be installed
- For Troubleshooting:
  - List the path of the working directory, path of the script that is producing errors, Job ID, and the command sent to the queue or interactive node
  - No need to attach the script or huge error messages
- For Testing:
  - Please have a sample dataset at your working directory, so that it can be used for debugging
- These steps will help us in responding quickly and efficiently

GACRC

THANK YOU for your patience

Questions?