

Introduction to HPC Using zcluster at GACRC

Georgia Advanced Computing Resource Center

University of Georgia

Zhuofei Hou, HPC Trainer

zhuofei@uga.edu

Outline

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

What is GACRC?

Who Are We?

- Georgia **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 (**OVPIIT**)
- 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)

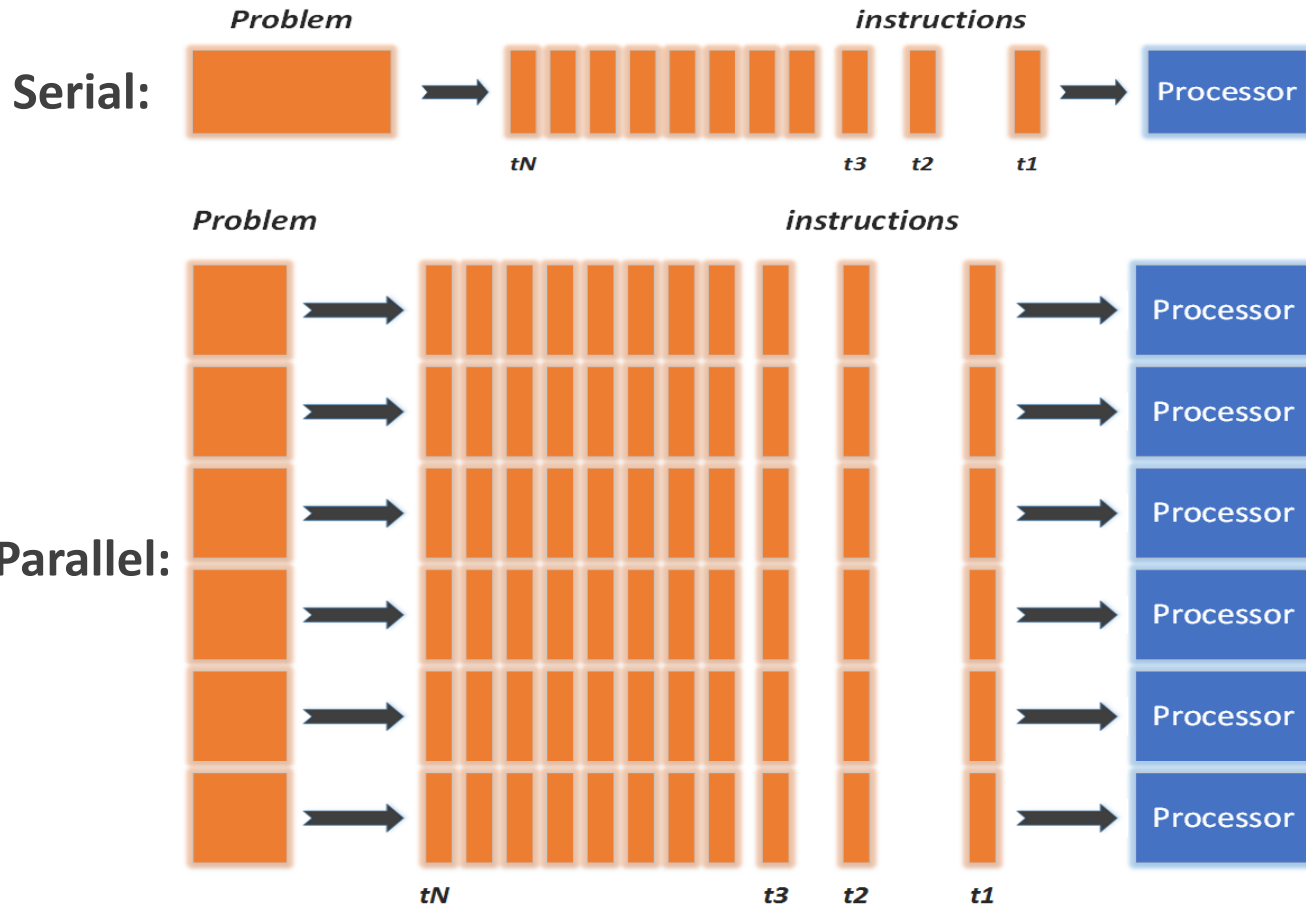
GACRC Users September 2015

Colleges & Schools	Depts	PIs	Users
Franklin College of Arts and Sciences	14	117	661
College of Agricultural & Environmental Sciences	9	29	128
College of Engineering	1	12	33
School of Forestry & Natural Resources	1	12	31
College of Veterinary Medicine	4	12	29
College of Public Health	2	8	28
College of Education	2	5	20
Terry College of Business	3	5	10
School of Ecology	1	8	22
School of Public and International Affairs	1	3	3
College of Pharmacy	2	3	5
	40	214	970
Centers & Institutes	9	19	59
TOTALS:	49	233	1029

GACRC Users September 2015

Centers & Institutes	PIs	Users
Center for Applied Isotope Study	1	1
Center for Computational Quantum Chemistry	3	10
Complex Carbohydrate Research Center	6	28
Georgia Genomics Facility	1	5
Institute of Bioinformatics	1	1
Savannah River Ecology Laboratory	3	9
Skidaway Institute of Oceanography	2	2
Center for Family Research	1	1
Carl Vinson Institute of Government	1	2
	19	59

Concept of High Performance Computing (HPC)



- ✓ **Serial** problem can not be broken
- ✓ *Discrete* instructions executed *sequentially*
- ✓ Only *1* instruction executed at any moment on a *single* processor

- ✓ Problem broken into *parallel* parts can be solved *concurrently*
- ✓ Instructions executed *simultaneously* on *multiply* processors
- ✓ Synchronization/communication employed
- ✓ **Shared-memory multithreaded job** or **MPI job** (Message Passing Interface)

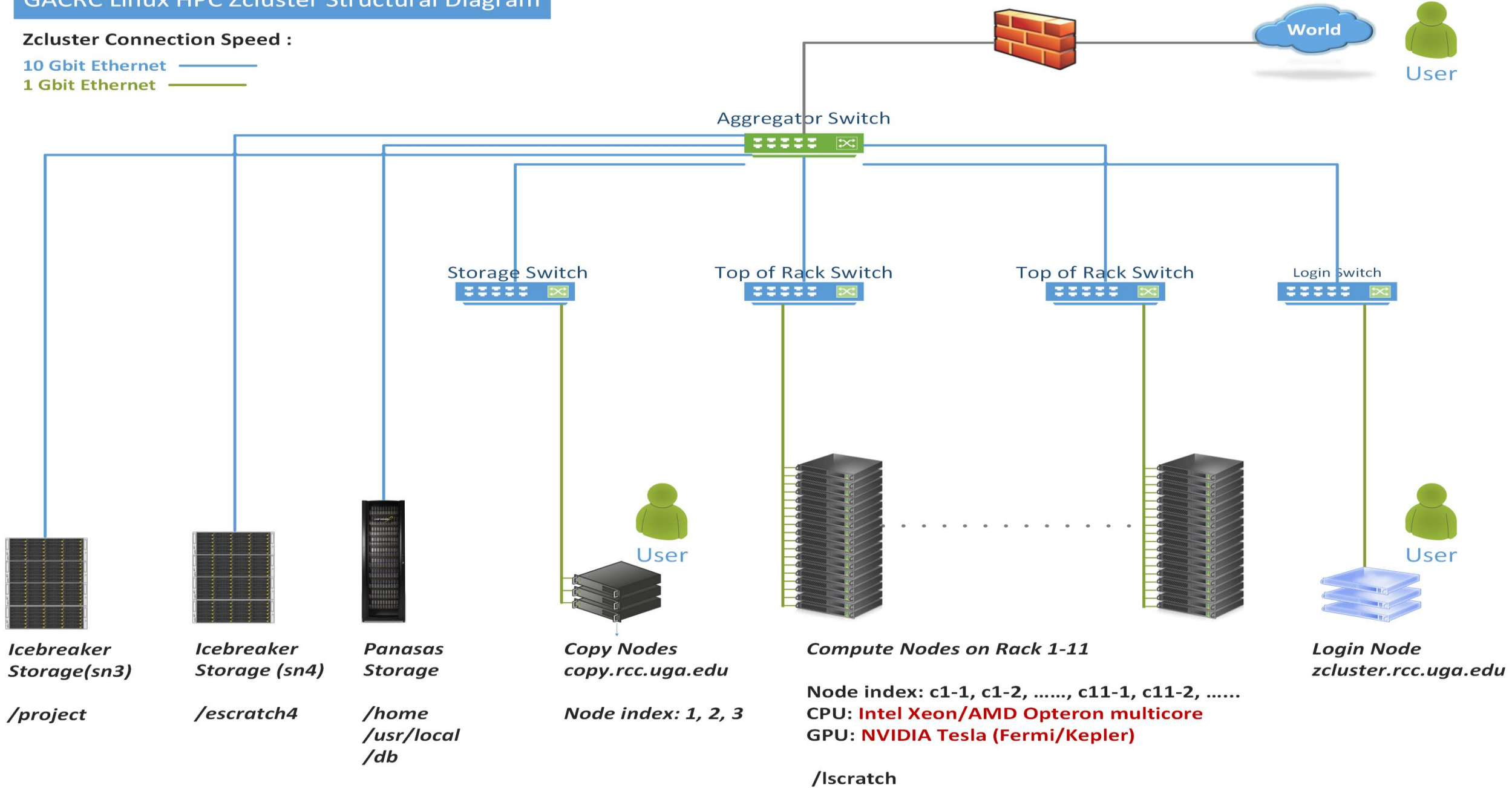
What is zcluster?

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

GACRC Linux HPC Zcluster Structural Diagram


Zcluster Connection Speed :

10 Gbit Ethernet ——— (blue line)
1 Gbit Ethernet ——— (green line)



What is 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**  Interactive Node: **compute-14-7/9**
 Copy Node: **copy.rcc.uga.edu**
qlogin
- Internodal Communication: **1Gbit** network
 compute nodes ↔ compute nodes
 compute nodes ↔ storage systems

What is zcluster – General Information

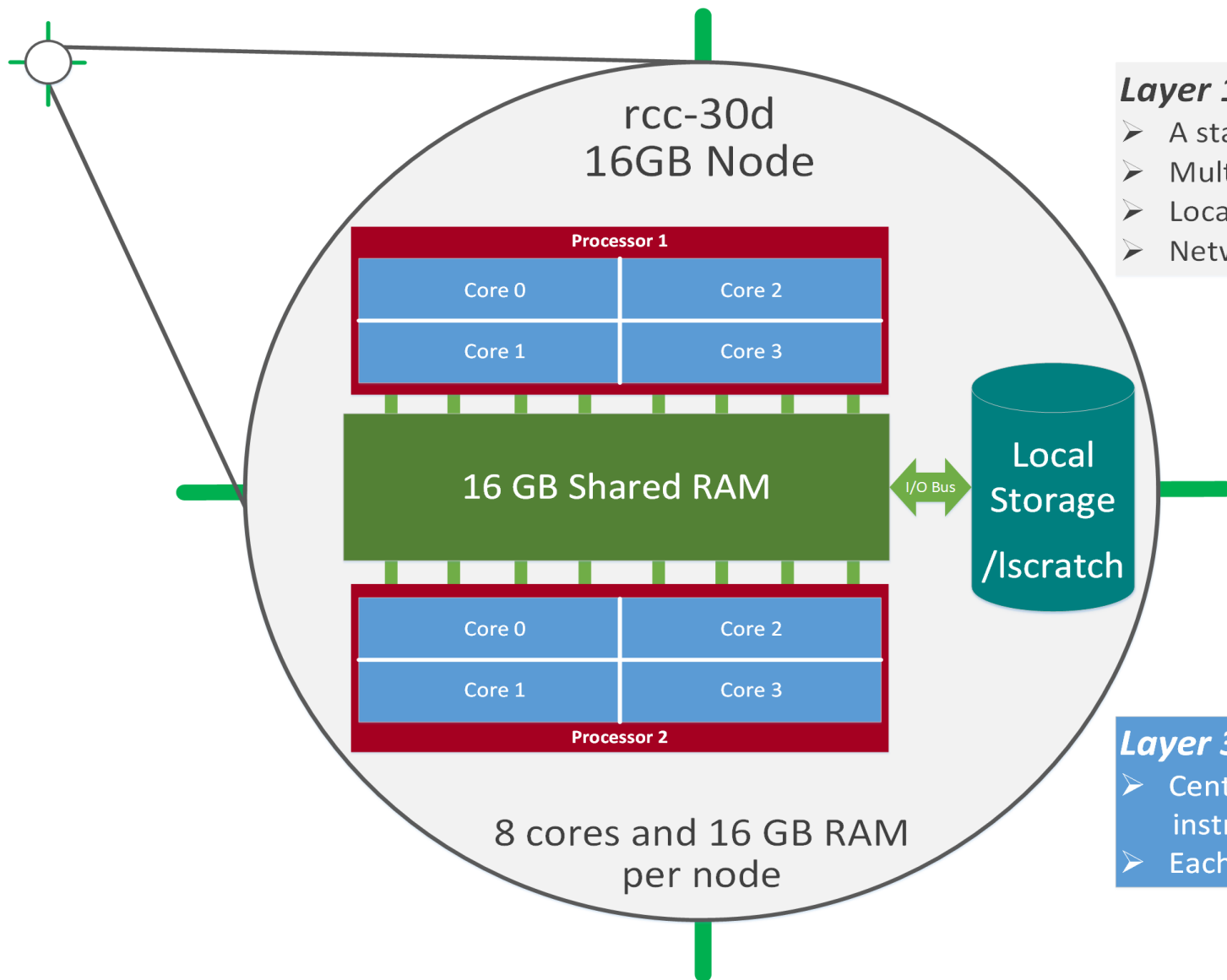
- Batch-queueing System:
 - Jobs can be started (submitted), monitored, and controlled
 - Determine which compute node is the best place to run a job
 - Determine appropriate execution priority for a job to run
- On zcluster: **Sun Grid Engine (SGE)**



What is 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

What is zcluster – Storage Environment

- **Home directory** → */home/groupname/username*
 - Mounted and visible on **all nodes**, with a quota of **~100GB**
 - Any directory on /home has **snapshot** backups
 - /home/abclab/jsmith/.snapshot
 - **Completely invisible**, however, user can “cd” into it and then “ls”:

```

zhuofei@zcluster:~$ ls -a
.          .bash_profile  .emacs.d      .fontconfig   .maple_history  MPIs      scripts  test.sh
..         .bashrc        .ENV_file     .gnuplot_history .Mathematica    openMPs   serials  .viminfo
.bash_history  downloads      exe           .history       .mc              .profile  sht      .Xauthority
.bash_logout  .emacs        .flexlmrc     .lessht        .mozilla         Pthreads  .ssh     ← .snapshot is NOT
zhuofei@zcluster:~$ cd .snapshot ← can “cd” into .snapshot
zhuofei@zcluster:~/ .snapshot$ ls ← then “ls” to list its contents
2015.06.21.00.00.01.weekly  2015.06.27.01.00.01.daily  2015.06.28.01.00.01.daily  2015.06.30.01.00.01.daily
2015.06.26.01.00.01.daily  2015.06.28.00.00.01.weekly  2015.06.29.01.00.01.daily
  
```

What is zcluster – Storage Environment

- **Local scratch** → `/lscratch/username`
 - On **local disk** of each **compute** node → **node-local storage**
 - rcc-30d 8-core nodes: **~18GB**, rcc-30d 12-core nodes: **~370GB**
 - **No snapshot backup**
 - Usage Suggestion: *If your job writes results to /lscratch, job submission script should move the data to your home or escratch before exit*

- **Ephemeral Scratch** → `/scratch4/zhuofei/zhuofei_Feb_01`
 - Create with `make_escalch` command
 - Visible to **all nodes** with a quota of **4TB**
 - **No snapshot backup**
 - To be deleted after **37 days**

What is zcluster – Storage Environment

Filesystem	Role	Quota	Accessible from	Intended Use	Notes
/home/abclab/username	Home	100GB	zcluster.rcc.uga.edu (Login)	Highly static data being used frequently	Snapshots
/escratch4/username	Scratch	4TB	copy.rcc.uga.edu (Copy) Interactive nodes (Interactive) compute nodes (Compute)	Temporarily storing large data being used by jobs	Auto-deleted in 37 days
/lscratch/username	Local Scratch	18 ~ 370GB	Individual compute node	Jobs with heavy disk I/O	User to clean up
/project/abclab	Storage	Variable	copy.rcc.uga.edu (Copy)	Long-term data storage	Group sharing possible

- Note:
1. /usr/local : Software installation directory
/db : bioinformatics database installation directory
 2. To login to [Interactive](#) nodes, use `qlogin` from [Login](#) node

What is zcluster – Storage Environment

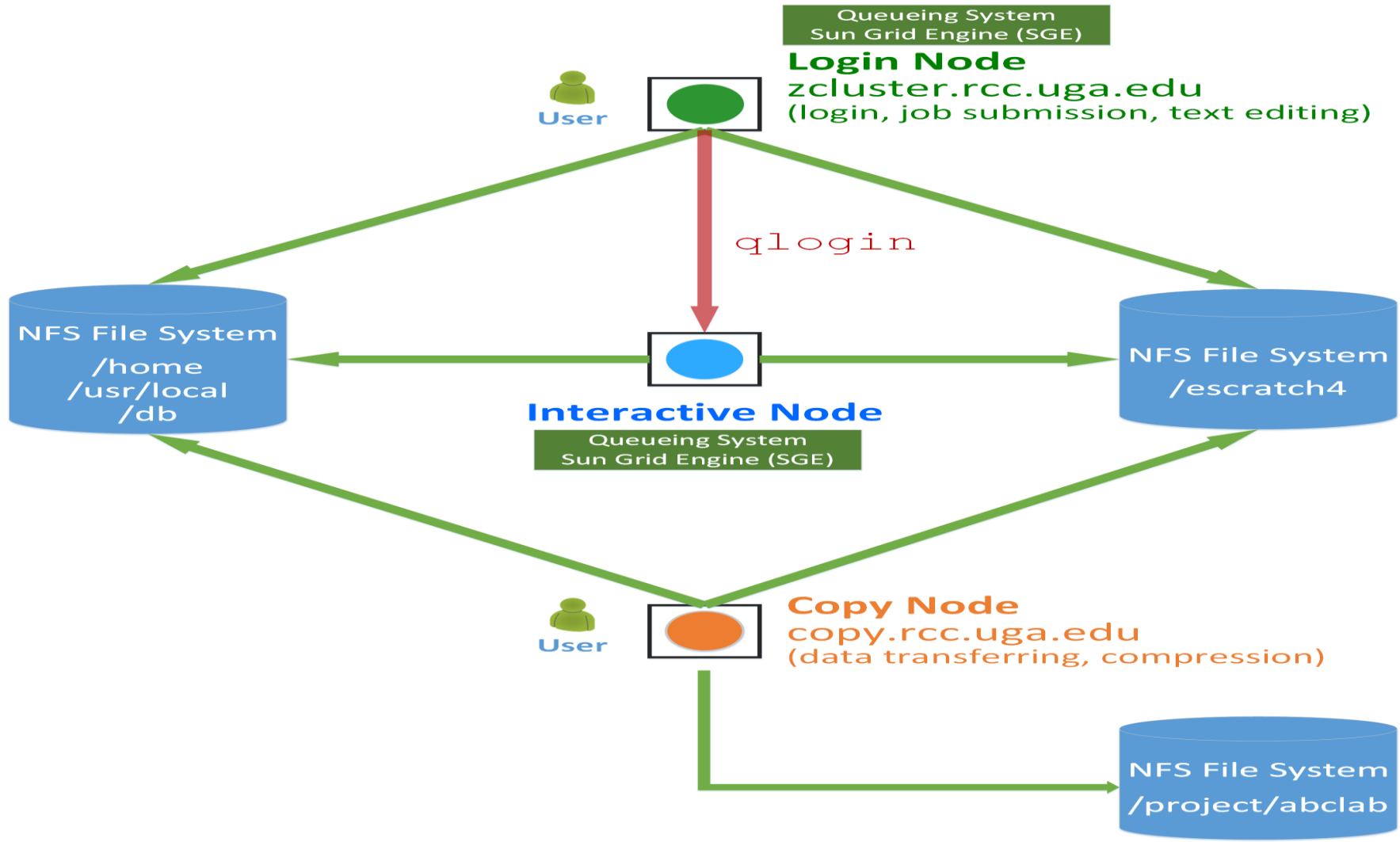
6 Main Function	On/From-Node	Related Filesystem
Login Landing	Login or Copy	/home/abclab/username (Home) (Always!)
Batch Job Submitting	Login or Interactive	/escratch4/username (Scratch) (Suggested!) /home/abclab/username (Home)
Interactive Job Running	Interactive	/escratch4/username (Scratch) /home/abclab/username (Home)
Data Archiving , Compressing and Transferring	Copy	/escratch4/username (Scratch) /home/abclab/username (Home)
Job Data Temporarily Storing	Compute	/lscratch/username (Local Scratch) /escratch4/username (Scratch)
Long-term Data Storing	Copy	/project/abclab

How does zcluster operate?

Next Page



← 1 Gbit Ethernet →



Longer-term data storage:
1. Lifecycle too long for /escratch4
2. Size too big for /home

How to work with zcluster?

Before we start:

- To get zcluster to be your best HPC buddy

GACRC Wiki: <http://wiki.gacrc.uga.edu>

GACRC Support: https://wiki.gacrc.uga.edu/wiki/Getting_Help

How to work with zcluster?

To submit your ticket to GACRC?

➤ **Job Troubleshooting:**

Please tell us details of your question or problem, including but not limited to:

- ✓ Your user name
- ✓ Your job ID
- ✓ Your working directory
- ✓ The queue name and command you used to submit the job

➤ **Software Installation:**

- ✓ Specific name and version of the software
- ✓ Download website
- ✓ Supporting package information if have

Note:

It's **USER's** responsibility to make sure the **correctness of datasets** being used by jobs!



How to work with it?

- 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
- Softwares Installed
- Run Interactive Jobs
- Submit Batch Jobs
 - How to submit *serial*, *threaded*, and *MPI* batch jobs
 - How to check job status, cancel a job
 - How to check memory usage of a job

How to work with zcluster – Start with zcluster

- You need a **User Account** : username@zcluster.rcc.uga.edu
- Procedure: https://wiki.gacrc.uga.edu/wiki/User_Accounts
- A UGA faculty member (**PI**) may register a computing lab:
<http://help.gacrc.uga.edu/labAcct.php>
- The PI of a computing lab may request user accounts for members of his/her computing lab: <http://help.gacrc.uga.edu/userAcct.php>
- User receives an email notification once the account is ready
- User can use `passwd` command to change initial temporary password

How to work with zcluster – Connect & Login

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

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

or

```
ssh -X zhuofei@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`

```
zhuofei@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

```
zhuofei@zcluster:~$ exit
```


How to work with zcluster – Transfer Files

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

- On Linux, Mac or cygwin on Windows : `scp [Source] [Target]`

E.g. 1: On local machine, do Local → zcluster

```
scp file1 zhuofei@copy.rcc.uga.edu:/escratch4/zhuofei/zhuofei_Feb_1
```

```
scp *.dat zhuofei@copy.rcc.uga.edu:/escratch4/zhuofei/zhuofei_Feb_1
```

E.g. 2: On local machine, do zcluster → Local

```
scp zhuofei@copy.rcc.uga.edu:/escratch4/zhuofei/zhuofei_Feb_1/file1 ./
```

```
scp zhuofei@copy.rcc.uga.edu:/escratch4/zhuofei/zhuofei_Feb_1/*.dat ./
```

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

How to work with zcluster – Softwares Installed


- Perl, Python, Java, awk, sed, C/C++ and Fortran compilers
- Matlab, Maple, R
- Many Bioinformatics applications: NCBI Blast+, Velvet, Trinity, TopHat, MrBayes, SoapDeNovo, Samtools, RaxML, etc.
- RCCBatchBlast (RCCBatchBlastPlus) to distribute NCBI Blast (NCBI Blast+) searches to multiple nodes.
- Many Bioinformatics Databases: NCBI Blast, Pfam, uniprot, etc.
- For a complete list of applications installed:
<https://wiki.gacrc.uga.edu/wiki/Software>

How to work with zcluster – Run Interactive Jobs

- To run an interactive job, you need to open a session on an **interactive node** using **qlogin** command:

```

zhuofei@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, remember to **exit** the session! 
- Detailed information, like interactive parallel job? Go to: [https://wiki.gacrc.uga.edu/wiki/Running Jobs on zcluster](https://wiki.gacrc.uga.edu/wiki/Running_Jobs_on_zcluster)

How to work with 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**
 - **qsj, qacct**

How to work with zcluster – Batch *Serial* Job

- **Step 1:** Create a job submission script *st.sh* running Samtools:

```
#!/bin/bash
```

→ Linux default shell (*bash*)

```
cd /escratch4/zhuofei/zhuofei_Feb_1
```

→ Specify and enter (*cd*) working directory (*/escratch4/zhuofei/zhuofei_Feb_1*)

```
time /usr/local/samtools/latest/samtools <command> [options]
```

→ Run samtools with 'time' command to measure amount of time it takes to run the application

- **Step 2:** Submit *st.sh* to the queue:

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

Submit a job to the queue rcc-30d with **16GB** RAM/Node

job submission script

OR

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

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

How to work with zcluster – Batch *Threaded* Job

- **Step 1:** Create a job submission script *blastn.sh* running NCBI Blast +:

```
#!/bin/bash

cd /escratch4/zhuofei/zhuofei_Feb_1

time /usr/local/ncbiblast+/latest/bin/blastn -num_threads 4 [options] → Run blastn with 4 threads
```

- **Step 2:** Submit *blastn.sh* to the queue:

```
$ qsub -q rcc-30d -l mem_total=20g -pe thread 4 blastn.sh
```

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

↑
4 cores requested

Number of Threads =
Number of Cores Requested

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

How to work with zcluster – Batch *MPI* Job

- **Step 1:** Create a job submission script *raxml.sh* running RAxML:

```
#!/bin/bash

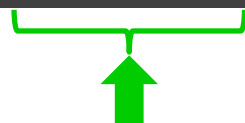
cd /escratch4/zhuofei/zhuofei_Dec_25

export MPIRUN=/usr/local/mpich2/1.4.1p1/gcc 4.5.3/bin/mpirun → Define and export environment variable (MPIRUN)

$MPIRUN -np $NSLOTS /usr/local/raxml/latest/raxmlHPC-MPI-SSE3 [options] → Run RAxML with 20 MPI processes (-np $NSLOTS )
```

- **Step 2:** Submit *raxml.sh* to the queue:

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



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

How to work with zcluster – Check and Cancel Jobs

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


How to work with zcluster – Check Memory Usage

- For a running job: **qsj**
- For a finished jobs: **qacct**

```

$ qsj 9707368

=====
job_number:      9707368
owner:           s_110
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=75.734M
    
```

```

$ qacct -j 970732

=====
qname           rcc-30d
hostname        compute-7-12.local
jobname         sub.sh
jobnumber       9707323
.....
cpu             183.320
mem             2.021
io              0.000
maxvmem        6.530G
    
```

Total
Memory



Thank You!