

# CyberGIS ROGER

## Resourcing Open Geospatial Education and Research

Johnathan Rush  
Geog. 479  
Spring 2016

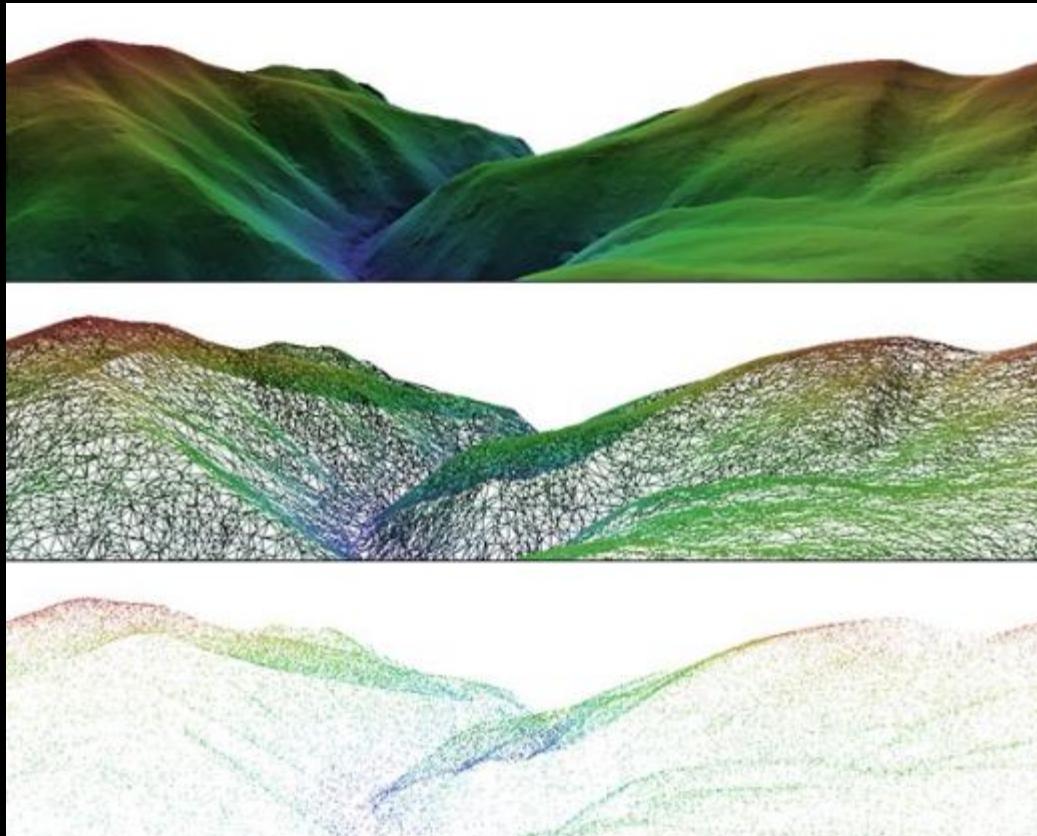




**CyberGIS is Geographic Information  
Science and Systems based on  
advanced cyberinfrastructure.**

Big Data  
Big Compute  
Big Collaboration  
Big Problems

# Increasing Data Size



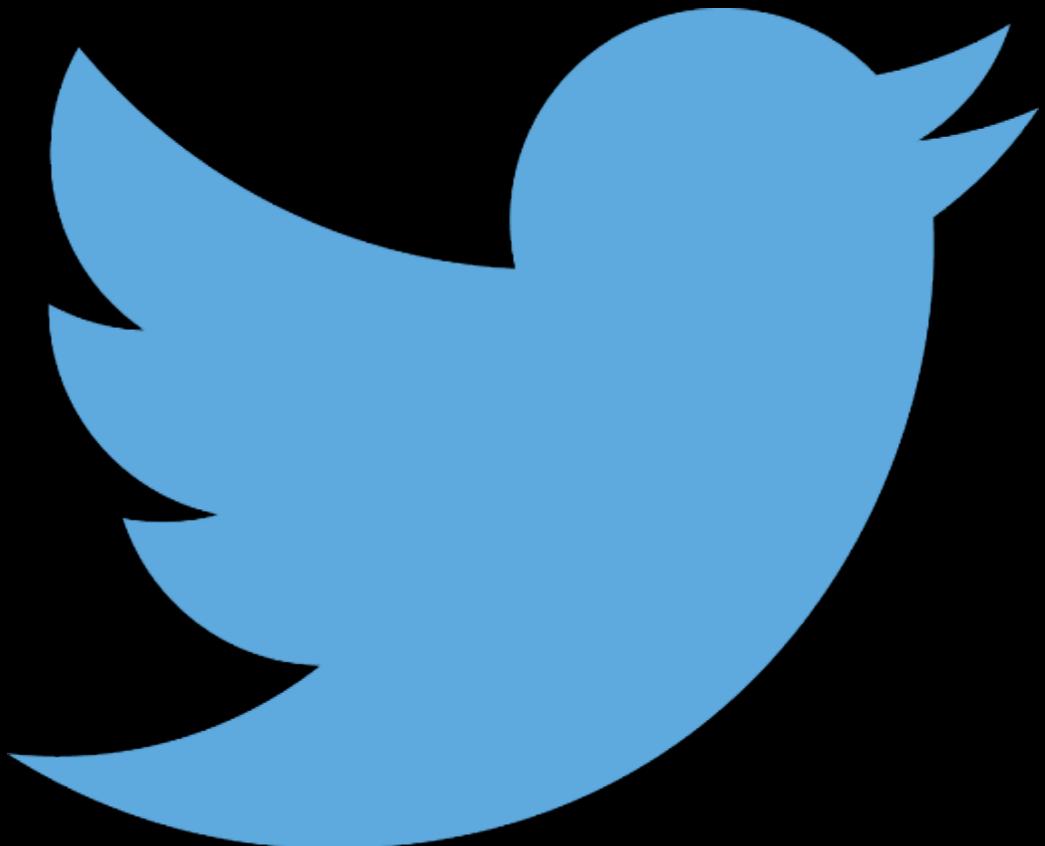
## National Elevation Dataset

30 meters: 175 GB

10 meters: 500 GB

1 meter: 50 TB to 4-5 PB

# Increasing Data Speed



Twitter Firehose

All Geocoded Tweets

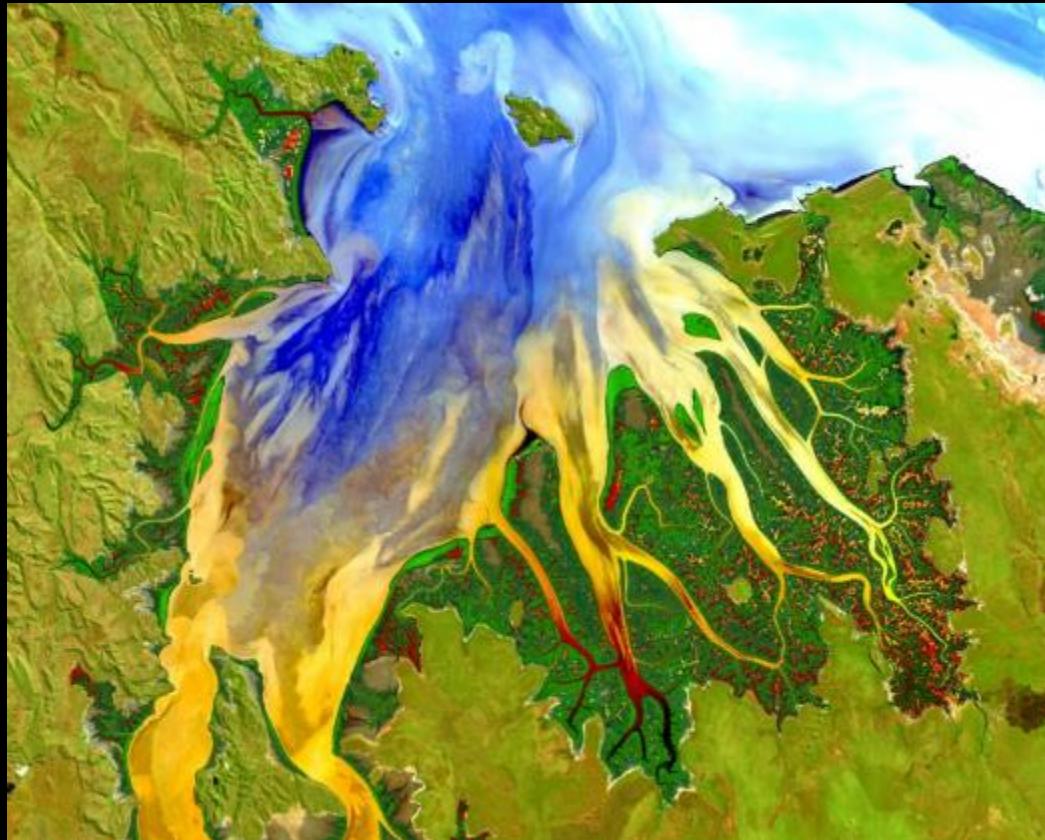
~2.5KB / tweet

~6,000 average tweet per second

~3% of tweets are geolocated

~38GB per day

# Increasing Data Speed



Landsat 8

~650 images per day  
~1GB each (compressed)  
~650gb per day

# CyberInfrastructure Resources



NSF Blue Waters @ UofI  
13,300 TFlop/s



NSF XSEDE: SDSC Comet  
2,000 TFlop/s



CyberGIS ROGER  
~45 TFlop/s

90

# CyberGIS ROGER

Resourcing Open Geospatial Education and Research  
after Roger Tomlinson  
~45 TFlop/s

# Why ROGER?

- Configured to best support geospatial data, with emphases on local memory and shared storage size and speed
- Supports multiple paradigms: traditional batch HPC, Hadoop, and Cloud (OpenStack).
- Integrate the three paradigms and leverage their strengths
- Both a research **project** and research **resource**
- Inform the design of future geospatial supercomputers

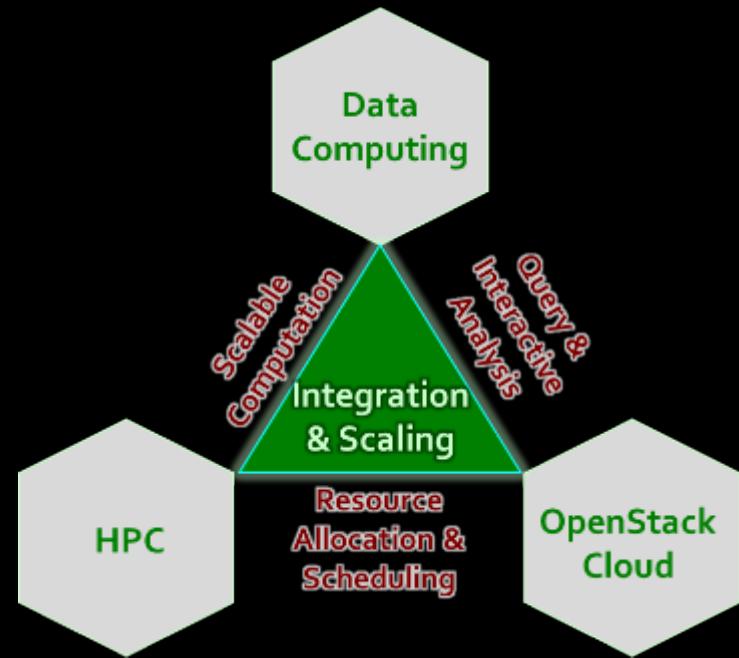
# ROGER: Overview



- Provides CyberGIS users with HPC, Hadoop and OpenStack functionality all on one system
- Managed by Systems Group  
With support from our storage, network, security and services groups

# ROGER: Hardware

- 1 dedicated admin node
- 24 **compute** nodes (20 Cores + 128GB +10Gbe)
- 12 **gpu** nodes (+ 1 Tesla K40 gpu)
- 16 **high mem** nodes ( 256GB + 800GB SSD)
- 10 **GPFS servers** ( 256GB + 800GB SSD)
- 2 **gridftp** nodes (256GB + 800GB SSD + 40Gbe)
- 5PB of raw disk in 5 NetAPP E2700
- Full info on the ROGER is at:  
<https://wiki.ncsa.illinois.edu/display/ROGER/ROGER+Technical+Summary>  
or <https://goo.gl/RsHxQf>



# Storage

- 4.5PB usable
- 10 Storage servers & 5 Netapp Storage appliances
- 900 drives (6TB each)
- GPFS
  - Fast Parallel filesystem
  - Common to all nodes
  - Supports batch HPC and Hadoop
  - Can export directories to Openstack



# NPCF

- National Petascale Computing Facility
- New building
  - LEED certified Gold
  - Restricted access
- 24 hour Operations Center
- Power & cooling at scale
  - Able to self-cool half the year
- Co-located with
  - Blue Waters
  - iForge
- Room for expansion in both Racks and Network capacity



# ROGER User Community

Biosciences	Engineering	Geosciences	Social Sciences	Multidisciplinary
<ul style="list-style-type: none"><li>• Agriculture</li><li>• Medical Science</li></ul>	<ul style="list-style-type: none"><li>• Bioenergy</li><li>• Environmental Engineering</li><li>• Transportation</li></ul>	<ul style="list-style-type: none"><li>• Atmospheric Science</li><li>• Hydrology</li><li>• Geology</li><li>• Remote Sensing</li></ul>	<ul style="list-style-type: none"><li>• Geography and Spatial Sciences</li><li>• Social Media &amp; Network</li><li>• Political Science</li></ul>	<ul style="list-style-type: none"><li>• Computational and Data Sciences</li><li>• Emergency Management</li><li>• Environmental Sustainability</li><li>• GIScience</li></ul>

# Education and Training Opportunities

Useful knowledge across most ROGER applications:

- Unix Shell
- Batch Job Submission
- Algorithm Parallelization

Training opportunities exist through:

- CyberGIS Commons <http://goo.gl/oKYZGB>
- CSE <http://goo.gl/S7xEqJ>
- Software Carpentry <http://software-carpentry.org/>

ROGER Helpdesk

[help+ROGER@ncsa.illinois.edu](mailto:help+ROGER@ncsa.illinois.edu)

# CyberGIS Fellows



# Acknowledgements

ADS – Brett Bode

Systems team – Dan LaPine, Chit Khin, Joe Muggli

Storage team – Andy Loftus, JD Maloney, Michelle Butler

Network team – Matt Kollross

ITS – Doug Fein

CyberGIS Center staff and students

NSF Award #1429699, MRI: Acquisition of a National  
CyberGIS Facility for Computing and Data-Intensive  
Geospatial Research and Education





# Accounts

- Expire at the end of the semester
- Change password at [go.Illinois.edu/NCSApassword](http://go.Illinois.edu/NCSApassword)
- If you lose your password, email [help@ncsa.Illinois.edu](mailto:help@ncsa.Illinois.edu), and it will be reset to your default password on the paper.

# Logging On to ROGER

- Download Mobaterm for Windows if you don't have a SSH client (<http://go.Illinois.edu/mobaxterm> - just run, no install)
  - Click Session button in top left corner, then SSH button
  - Enter roger-login.ncsa.illinois.edu in the remote host box, then click OK
  - Enter your username, then password. *Don't save your password* if you are in a computer lab.
  - Next time you open MobaXTerm, roger-login will be a saved session in the left sidebar you can double-click to start
- Or, use Terminal on Mac and Linux

```
ssh -X username@roger-login.ncsa.illinois.edu
```

(enter your password when prompted)

# Common Commands

cd *change directory*

pwd *displays your current directory*

ls *list the files in your current directory*

ls test\* *list all of the files starting with “test” in the directory*

ls -l *list the files in a directory with details*

mkdir name *create a directory*

rm *remove (delete) file*

rm -r *remove a directory*

cp origin destination *copy a file*

cp -r origin destination *copy a directory and its contents recursively*

mv origin destination *move or rename a file*

chmod g+rw filename *set the file to be readable and writeable by group members*

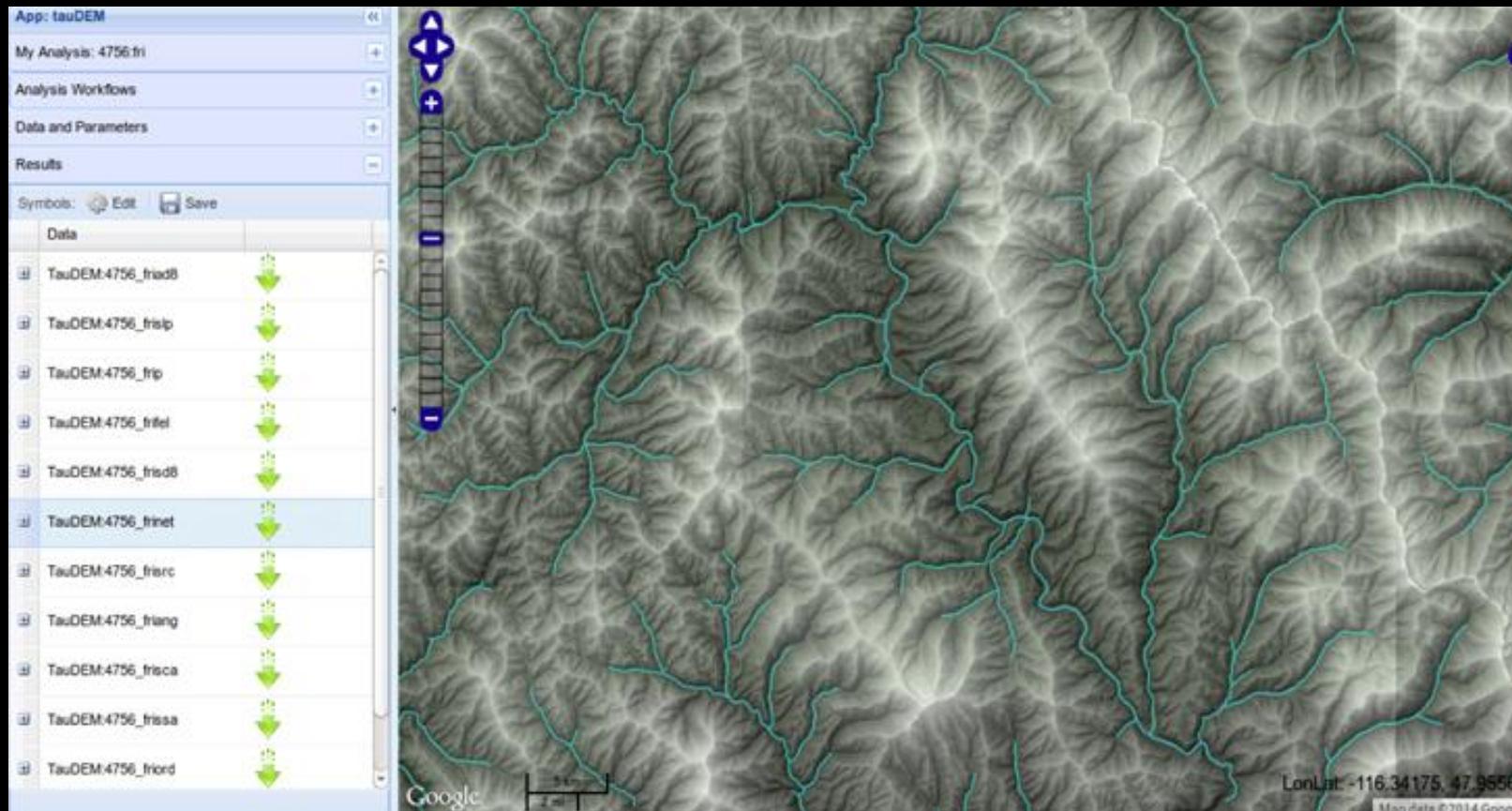
hdfs dfs -ls directory *list the files in the specified directory of the Hadoop File System*

*Hadoop has no concept of a “current” directory*

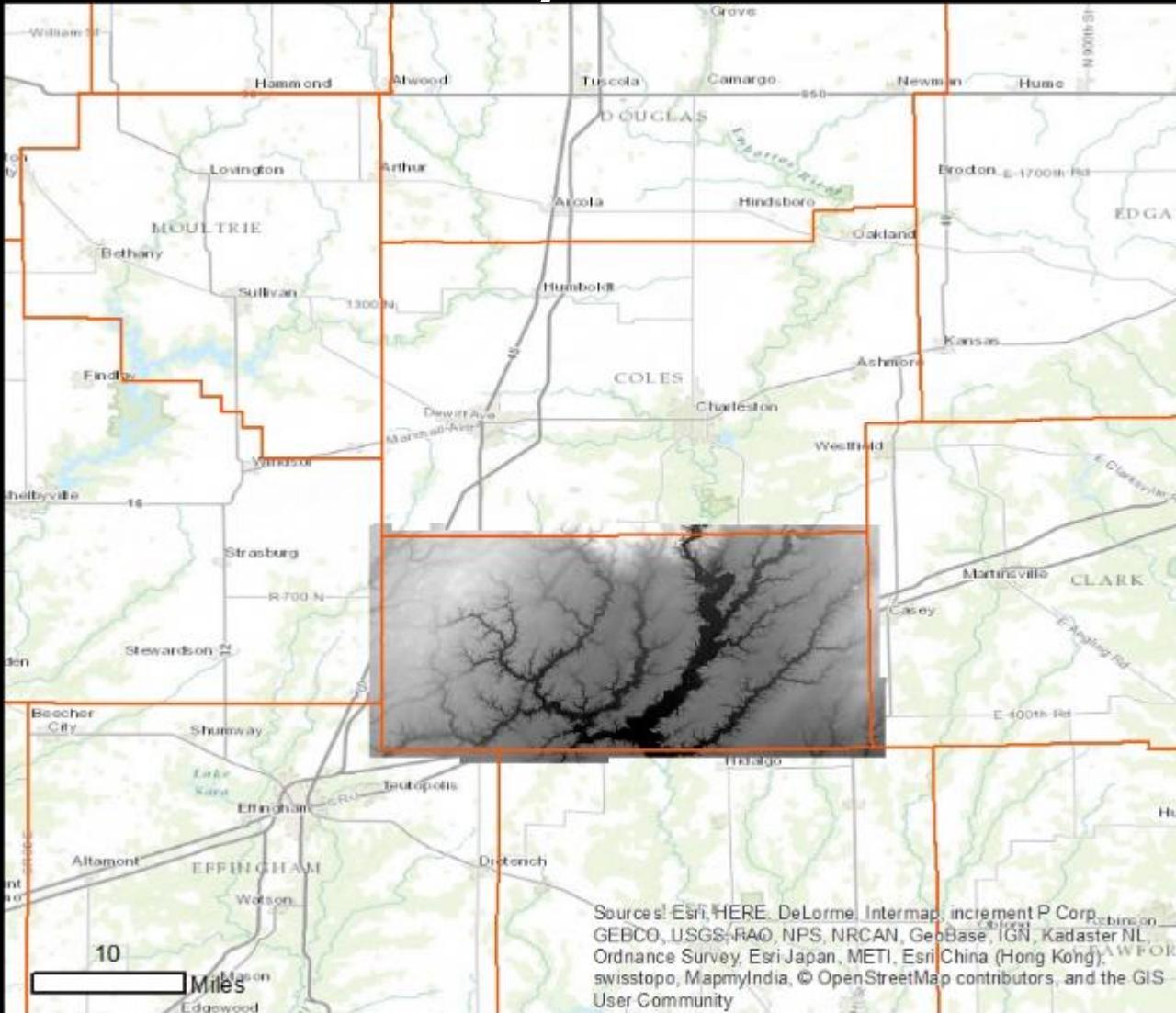
# File system

Purpose	Path	Shortcut
Home Directory	/home/userid	~
All Projects	/gpfs/largeblockFS/project	/projects
Course Project		/projects/class/g479
Scratch	/gpfs/largeblockFS/scratch	/gpfs_scratch
Software	/gpfs/smallblockFS/sw	/sw

# Parallel Applications



# Cumberland County 1m DEM from Lidar



# Pit Remove DEM

- Get your own compute node to work on (called an interactive session)

```
qsub -I -l walltime=0:30:00
```

- Run TauDEM

```
module load mpich taudem
```

```
mkdir ~output
```

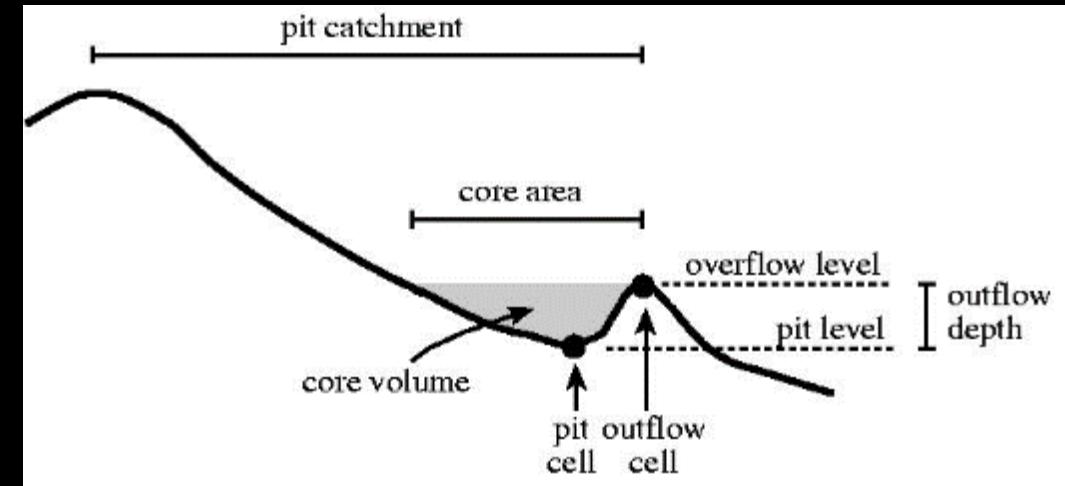
```
mpirun -n 20 pitremove -z /gpfs_scratch/il_lidar/geotiff/uncompress/ -fel  
$HOME/output/
```

- Open another terminal to view CPU usage

Right-click the MobaXTerm tab, click Duplicate tab. Log in again.

Then, ssh into the node that your TauDEM job is running on. If your prompt in your other window says `username@cg-cmp12`, then you would enter `ssh cg-cmp12` in your new terminal.

Enter `top` to see the current CPU usage. Quit top with the `q` key.



# GNU Parallel Example

- Copy data

```
cp -R /projects/EOT/workshops/datapar ~  
cd ~/datapar/ned10m
```

- Multitasking with GNU Screen

```
screen
```

- Reproject all GeoTIFFs matching a filename pattern, saving results in new directory

```
module load parallel  
module load gdal-stack
```

```
find -name \*.tif | parallel "gdalwarp -t_srs epsg:3269 {} {.}_proj.tif"
```

- Multitasking with GNU Screen

```
ctrl-a d #disconnect from current screen session. That's ctrl-A, then D
```

```
top #check CPU usage
```

```
screen -x #reconnect to previous screen session
```

# ROGER Statistics

- [roger-stats.ncsa.illinois.edu](http://roger-stats.ncsa.illinois.edu)
- [clustat.ncsa.illinois.edu/roger/](http://clustat.ncsa.illinois.edu/roger/)

# Questions?



Johnathan Rush

[jfr@illinois.edu](mailto:jfr@illinois.edu)

[cyberGIS.illinois.edu](http://cyberGIS.illinois.edu)