Cluster Accounts:

- How do I change my ACCRE account password?

- I've forgotten my password; what is the procedure to reset it?

- Does ACCRE have a Hadoop cluster that includes tools like HDFS, MapReduce, Spark, Hue, Hive, Impala, and so on?

Connectivity:

- I cannot connect to the cluster, am experiencing intermittent connectivity to the cluster, or the system hangs upon log on. What should I do?

- I cannot connect to `login.accre.vanderbilt.edu` or after logging in got an error message says my home directory is not found.

- How can I make a scheduled downtime work for me?

- How can I mount a Samba (`smb`) share?

Configuring your Environment:

- How do I change my default shell?

- How do I display graphics from the cluster to my local machine?

- I am running an X server, how do I fix X connection or .Xauthority file errors?

Linux:

- What command do I need to type in order to run an executable in Linux?

- How do I change the group associated with a file or directory?

Submitting and Running Jobs (Also read Getting Started):

- What types of nodes are available?

- How do I run test jobs?

- What are the ACCRE cluster defined attributes I can use in my SBATCH scripts corresponding to the available node properties?

- Can I run on the gateway machines?

- What happens if my job uses more resources than requested?

- Why is my eligible job waiting so long in the *Idle* state?

- What does job status *Deferred* mean?

- What is the maximum number of jobs I can submit or have running at any one time?

- What is the maximum allowed "wall clock time" I may specify?
- How do I hold/release/delete a job?
- Where can I find detailed documentation on all SLURM Commands?
- How can I delete many jobs at once?
- How much memory is available on each node?
- How do I request all the processor cores in a node?
- How do I request an 8-core node for exclusive usage?
- How can I determine whether or not I have all the processor cores in a node assigned to my job (and why would I want to do so)?
- If I belong to multiple groups, how can I define the group name under which my job is to run on the cluster?
- How do I checkpoint my job?
- How do I use local storage on a node?
- My job submission fails with a socket timeout message. What's the problem?
- Why do files named "core.*" appear after I submit a job?

Disk Space:
- How can I determine my disk space quota and how much disk resources I am using?
- My group has storage on DORS. How can I check our usage?
- I need a lot of disk space for temporary use. How do I make use of the /scratch space?
- If I need more disk space than this, will you temporarily grant a quota increase?
- What if I need more disk space than the default quota limits on /home and /scratch for an extended period of time? *(Links to our disk policies page)*
- Will ACCRE restore deleted or lost data?
- My network connection to ACCRE is really poor and I have a lot of data that I need to upload to ACCRE (or download from ACCRE). What are my options?

GPU:
- How do I request a gpu node?

Software and Compiling:

- What research software packages are available on the cluster?
- How do I make sure that my perl/python script is using the latest version available on the cluster?
- I'd like to have some software installed on the cluster. How do I go about doing that?
- How do I install an R package from source code?
- How do I download and install an R package from the internet?
- How do I install and load an R package from Bioconductor?
- How do I install a perl module without root privilege?
- How do I check which python packages are installed?
- How do I install a python module without root privilege?
- How do I install a python package from source code?
- How do I run Matlab/SAS job on the cluster?

Accounts: How do I change my ACCRE account password?

Please follow these steps:

1. Log on to the cluster (`login.accre.vanderbilt.edu`) using your existing password. At this point you are simply logged on to a cluster gateway. Any changes to your account will need to be propogated to all of the compute nodes.
2. Issue the command: `rsh auth`. Commands issued now (e.g., the `passwd` command below) will affect modifications to all of the nodes.
3. To change your password, type the command `passwd`. The `passwd` program will prompt you to enter a new password. Please use a non-dictionary word (i.e., nonstandard words, combinations of letters and numbers). After you have changed your password, **please allow approximately 20 minutes for the change to be propogated to all of the nodes**.
4. To disconnect from `auth`, type `exit`. You will still be logged into your account. You may either continue working on the cluster or log out.

See also our tutorials on getting started on the cluster.

Top of Page

Accounts: I've forgotten my password or my password has expired; what is the procedure to reset it?

Notify us by submitting a Request Tracker ticket. The script we run to reset your password propagates the change out to the cluster and sends you e-mail

with your new password. This normally takes a few minutes, and we ask you wait at least 15 minutes to log on. As soon as you receive the e-mail with your new password, please follow **the procedure to reset it to something of your choosing**.

Top of Page

Does ACCRE have a Hadoop cluster that includes tools like HDFS, MapReduce, Spark, Hue, Hive, Impala, and so on?

Yes! Open a help desk ticket requesting (free) access. You can find more information about the environment in this blog post and in our Vanderbilt Big Data GitHub Organization. We currently have a development environment set up for Vanderbilt and VUMC researchers to access, and plan to build out a production environment over the next 2-3 years.

Top of Page

Connectivity: I cannot connect to the cluster, am experiencing intermittent connectivity to the cluster, or the system hangs upon log on. What should I do?

If you are normally able to connect and suddenly cannot, let us know via Request Tracker. Please provide as much information about the issue as possible including any useful output to your screen. Besides occasional network problems, there are a number of possible causes for sluggish to zero connectivity. Please read the following to help self-diagnose before submitting a help desk ticket so we are better able to assist you:

- You may connect to the cluster only via a Secure Shell (**SSH**) client. For more information go to Logging on to the Cluster.
- If you see the following error when trying to connect: `ssh: connect to host login.accre.vanderbilt.edu port 22: Operation timed out` it means the gateway you're trying to connect to is unreachable.The cluster has roughly 20 x86 gateway machines (See node configuration). There are several reasons why we have multiple gateways of each architecture. For example, this distributes the user load across many login gateways. Another main reason is to protect against an unreachable gateway preventing a connection to the cluster.However, even with this backup system in place it is still possible when you ssh to `login.accre.vanderbilt.edu` to get an error similar to the above. `login` is only an alias which uses DNS round-robining to select one of the actual gateway machines to connect to. If you get the above error, what has likely happened is that either the local DNS cache on your system or the DNS server you use has cached an alias to a gateway which is now unreachable for some reason. If this occurs, you should simply select one of the actual gateways at random and attempt to ssh directly to it. For example, `ssh vmps65.accre.vanderbilt.edu`.

4

- If you can connect but the connection "hangs" before you receive a command-line prompt, your problem may be related to an error in one of your login files (e. g., .bashrc in your home directory). This we can help diagnose. Please send us a message via Request Tracker.
- If you can connect but your login "hangs", it is possible we are experiencing a problem with GPFS (the General Parallel File System designed by IBM). Other symptoms of this include logging on but not being able to see, for example, your home directory. Sometimes the file system problem is temporary, lasting only a moment. Larger file system problems normally occur when the system is overloaded, which can happen for various reasons. If the problem is found to be caused by a particular user account or set of jobs, we immediately work with the user to resolve it.Sometimes the issue is not related to user software or the way a user is submitting jobs and we have to work with IBM to determine the root of the file system problem. When we expect the issue cannot be resolved quickly we notify all users to expect intermittent cluster access.In most cases when the system is in this state you should still be able to accomplish your work, albeit you may find the system is occasionally sluggish and intermittently nonresponsive. Please be patient. You will find upon repeated attempts you will be able to log on and submit jobs. If these jobs are not heavily dependent on disk I/O they should continue running.

In any case, so we can immediately begin resolving the issue, please notify us ASAP of any connectivity problems by submitting a Request Tracker ticket. Include details such as a "cut and paste" of the information in your login window if you are able.

Top of Page

Connectivity: I cannot connect to `login.accre.vanderbilt.edu` or after logging in got an error message says my home directory is not found

`login.accre.vanderbilt.edu` is a DNS round robin alias for one of our ~6 cluster gateways. It is possible that the gateway you were randomly assigned is experiencing a hardware issue. When this happens we take the gateway out of the rotation but DNS may cache the old information for a period of time. Please open a help desk ticket for assistance.

Top of Page

Connectivity: How can I make a scheduled downtime work for me?

As a scheduled downtime for the cluster approaches more time becomes available for shorter jobs. Thus, if you have applications that take a few days or less to run, you will be able to execute more of these types of jobs as a scheduled downtime approaches because applications requiring longer period of times will

not be running. It's an excellent time to take advantage of the extra computing cycles that would ordinarily not be available!

Top of Page

Connectivity: How do I mount a Samba (`smb`) share?

If you have been assigned a Samba or smb share to mount locally the following instructions should help. Note you should have been given a share name and you should have a Samba password that is different from your cluster or VUnetID password:

On a Mac open finder. On the menu bar at the top select Go, Connect to Server… (K). In the Server Address: field enter the following;

`smb://samba.accre.vanderbilt.edu\*sharename*\`

where `sharename` is the name of the Samba share you were assigned. Next click Connect and you will be prompted for a Name and Password. This will be your cluster username and your Samba password. At this point hit Connect again and your share will be mounted as a drive in Finder.

On a Windows PC run File Explorer (from the start bar it will be a folder or you can hit the windows key and type in explorer). In the address bar at the top enter the following;

`\\samba.accre.vanderbilt.edu\*sharename*\`

where `sharename` is the name of the Samba share you were assigned. Hit enter and you will be prompted for a username and password. Enter your cluster username and your Samba password and hit enter or click ok. At this point your share will be mounted as a drive in File Explorer.

Top of Page

Environment: How do I change my default shell?

Once you log onto the cluster (`login.accre.vanderbilt.edu`), type: `rsh auth`. You are now on the authentication server. Type: `chsh`. You will be prompted for password. Enter the new shell you want to use. For bash, this is `/bin/bash`, for tcsh, `/bin/tcsh`. Type `exit` to log out of auth and then wait 20 minutes or so for the new setting be propagated throughout the cluster. Log onto the cluster again and you should see the change.

Top of Page

Environment: How do I display graphics from the cluster to my local machine?

You need two things: (N. b., you should first check with your P. I. before installing the following software, since one or both of these may already be on

your system, especially if you're using a computer in your lab which is already configured to run on the cluster)

1. **Get X server support on your local machine:** The graphics environment on the cluster is X11, therefore, you must install and run an X server from your local machine
2. **Configure SSH tunneling:** You must tell **SSH** on your local machine to allow the display of graphics from software running on the cluster.

Windows users: Xming X Server is one of the best X Window servers available for Windows. You can follow the instruction provided there to install and set up the server. **Mac OS X users:** You can get a free X11 server from Apple. Mac OS X should already have **SSH** installed.

1. Follow their directions to install and run the X11 server.
2. Launch the X11 server.
3. Run an `xterm`.
4. When you log on to the cluster from the command line in the xterm, to activate **SSH** tunneling you can use the `-X` option, i.e., `ssh -X user@login.accre.vanderbilt.edu`.
5. Finally, see below for how to quickly check you can display remote graphics locally.

Linux users: We assume you are already running an X server and have **SSH** installed.

1. When you log on to the cluster, `ssh -X` will activate **SSH** tunneling, i.e., `ssh -X user@login.accre.vanderbilt.edu`.
2. See below for how to quickly check you can display remote graphics locally.

Top of Page

———————————————————

Environment: I am running an X server, how do I fix X connection or .Xauthority file errors?

If you are getting error messages similar to these:

`/usr/X11R6/bin/xauth: error in locking authority file /home/user/.Xauthority`

`X11 connection rejected because of wrong authentication. X connection to localhost:11.0 broken (explicit kill or server shutdown)`

try removing the `.Xauthority` file in your home directory, then log out and back in. This file occasionally becomes corrupted. When you log back in and start X, it will recreate your `.Xauthority` file. Sometimes you have to do this a few times. If you continue to have problems, please submit a Request Tracker ticket.

Top of Page

———————————————————

Linux: What command do I need to type in order to run an executable in Linux?

To execute a program in the current work directory, type:

`./<file_name>`

For files that are not in the current working directory, use the full path: `/path/to/your/executable/file`

Top of Page

Linux: How do I change the group associated with a file?

You can change the group of a file if you are the file's owner and you are in the group to which you are trying to change the file. The command is:

`chgrp [options] group_name file_name`

`-R`: recurse through subdirectories `-f`: suppress most error messages If you want to submit a job from group other than your primary group, please see this FAQ.

Top of Page

Jobs: What types of nodes are available?]

We currently have nodes with between 8-16 CPU cores and 24-256GB RAM. We also have a group of nodes equipped with modern NVIDIA GPUs and a small set of nodes with Intel Xeon Phi accelerators on board. Please refer to our Intro to the Cluster slides for more details.

Top of Page

Jobs: How do I run test jobs?

We allow users to run very short (< 5 minutes) tests that have low memory usage (< 1 GB). Anything more should be submitted the scheduler. We have a debug SLURM partition/queue available for running quick tests and prototyping.

Top of Page

Jobs: What are the ACCRE cluster defined attributes I can use in my SBATCH scripts corresponding to the available node properties?

The properties of our compute nodes can be specified with combinations of available attributes (defined by us), e.g.: `haswell, sandy_bridge, eight` Note that the `haswell` attribute requests the latest Intel processors, while `sandy_bridge` requests the previous generation. The `eight` attribute requests nodes that only have eight processors, and the scheduler will avoid placing the job on the 12-core nodes. In your batch script you could specify: `#SBATCH --constraint=haswell`. This would instruct the scheduler to run the job only on a node with an Intel

8

Xeon Haswell processor. Note that your job may take longer to start when these attributes are included as you are limiting the pool of resources the scheduler can choose from. For a full list of available features, trying running the `sinfofeatures` command while logged into the cluster. Find more examples in subsequent FAQ.

Top of Page

Jobs: Can I run on the gateway machines?

When you log on via `login.accre.vanderbilt.edu`, you are logged onto a gateway machine. From here you submit your jobs which are sent to the compute nodes by the scheduler. However, we do allow you to run very short, < 5 minute, test jobs on the gateway machines, as long as such jobs do not slow the gateway for other users. Anything longer than this should be submitted to the compute nodes using `batch` (see How to Submit Basic Jobs).

Top of Page

Jobs: What happens if my job uses more resources than requested?

The job scheduler will automatically kill most jobs which exceed the resources requested in the SBATCH script. For example, if you specify a `walltime` of 4 hours and your job runs over that, the scheduler will kill the job. The reason for this is that running jobs which use more resources than requested may affect the scheduling and running of other jobs. This is because the scheduler relies on SLURM specifications (among other parameters) to determine on which nodes to run jobs. Also read our job scheduler policies for more information on killing jobs which are interfering with other jobs or the system itself. When testing code or running code you are unfamiliar with, you should more diligently monitor the resource consumption to fine tune your SBATCH request. Specifying much more, e.g., `walltime` or `mem`, than your job requires may delay its start time if the requested resources are not immediately available. Therefore, you should start somewhat conservatively, then reduce your resource specifications once you've determined what you are really using, still always leaving a buffer to ensure the job is covered. Learn more about how to request resources and the SBATCH defaults when you submit a job. Learn how to monitor and check the status of a submitted job.

Top of Page

Jobs: Why is my eligible job waiting so long in the *PENDING* state?

There are several things you should check to understand your wait time in the queue. See tips on checking the status of a submitted job.

9

- **Make sure you have requested an allowed set of resources.** Check your SBATCH script against both the available nodes in the cluster and our job scheduler policies. You can also check on the resources requested with the command: **scontrol show job** *job_number*

- Check your group's current usage by typing `qSummary -g group_name`. Compare that to your group's bursting limits by running `showLimits -g group_name`. If your group's current usage is close to or equal to its bursting limits, this could be causing delays. Details about both of these commands can be found by clicking on this link.

- Check overall cluster utilization with the `SlurmActive` command.

- Check the queue and current usage on the cluster. It could be the particular resources your jobs need may be heavily utilized, even if the entire cluster is not. You can check the total usage of the cluster with the command `squeue`. You can also see current and past utilization levels on this website.

- Your account or group account may be running over its fairshare. This means when the cluster is very busy, other jobs from accounts which are under fairshare may be assigned higher priority and may jump ahead of your job in the eligible queue. Use the

    ```
    sacctmgr show associations Accounts=account_name \
      format=account%30,user%30,fairshare,grpcpus,grpmem
    ```

    command to check your fairshare.

If you still do not understand why your jobs are not starting more quickly, please submit an RT ticket.

Top of Page

Jobs: What does job status *Deferred* mean?

In slurm there is no "deferred" state. However, jobs may ask for resources that cannot be provided, e.g., too much memory. In such cases, running the command squeue and looking for your jobib, slurm will provide a short explanation of why the job either cannot run, or is not running. Do `squeue -u <username>` to see the explanation.

Top of Page

Jobs: What is the maximum number of jobs I can submit or have running at any one time?

"Active" Limits: Each user/group/account has a limit on the number of processors in use at any time. This number is summed from any combination of single and multi-processor jobs. Additional limits are placed as necessary for

groups running either medium (defined as 4 to 7 days) or long (over 7 days) jobs on a regular basis if there usage is impacting the ability of other groups to use their full fairshare. Individual groups may also request upper limits on their users. New guest users have upper job limits until they have attended the Introduction to the Cluster and Job Scheduler classes. Use the `showLimits` command to check your group's limits. Please refer to the job scheduler policies for additional important details of these limits.

Top of Page

Jobs: What is the maximum allowed "wall clock time" I may specify?

The maximum allowed walltime is **14 days**, or in **hh:mm:ss = 336:00:00**. Your job will not start if you have specified a walltime greater than this. You may reduce the walltime of an already submitted job using `scontrol` (slurm job control). In addition we ask that, except for a small number of test jobs, jobs run at least 30 minutes and over an hour in length is preferable. Our job scheduler policies explains more on this subject. Also see How to Submit Basic Jobs for other SBATCH specifications and how to deal with very short jobs.

Top of Page

Jobs: How do I hold/release/delete a job?

A user may place a USER hold upon any job the user owns. To do that, type: `scontrol hold <jobId>`. To release the held job, type: `scontrol release <jobId>`. Note that you can only hold and release jobs that are pending (i.e. this will not work for running jobs). User can also delete a queued/running job using the command: `scontrol resume <jobId>`. To delete all the jobs owned by the user, type: `scancel -u <userid>`. To cancel a job by name, type: `scancel --name <jobName>`.

Top of Page

Jobs: Where can I find detailed documentation on all SLURM commands?

Please visit www.schedmd.com for a complete and detailed list of all slurm commands.

Top of Page

Jobs: How can I delete many jobs at once?

If you are using bash, the following script shows how to delete all jobs between 10000 and 10010:

```
for jobid in `seq 10000 10010`
```

```
    do
    scancel $jobid
    echo cancelling $jobid
    done
```

Top of Page

Jobs: How much memory is available on each node?

Because the OS and other system processes (e.g. GPFS managment) already use certain amount of memory, not all physical memory is avaiable for running jobs. In general, ACCRE nodes contain anywhere from 22GB - 248GB of available memory for jobs to use.

Top of Page

Jobs: How do I request all the processor cores in a node?

Since the cluster has some nodes with 8 cores (2 x quad-core processors) and some with 12 cores (2 x hex-core processors) there is no way to generically request all the cores in a node. You can control whether your job lands on a 8-core or 12-core node; however, there is no `SBATCH` directive which effectively says, "Give me all the cores in the node no matter how many there are." If your SBATCH script has `SBATCH -n 8` asking for 8 cores, your job could land on either an 8 or 12 core node; including the attribute `eight` will ensure your job is run on a 8-core node (see the next FAQ). If your SBATCH script has `SBATCH -n 12` your job will only land on a 12 core node. However, this might increase your wait time in the queue as there are currently more 8-core nodes than 12-core nodes in the cluster.

Top of Page

Jobs: How do I request an 8-core node for exclusive usage?

To request an 8-core node exclusively, your can now use:

```
#SBATCH -n 8
#SBATCH --constraint="eight"`
```

in your job submission script. If the job does not require exclusive access to the node (it just needs 8 cores), you can still use:

```
#SBATCH -n 8
```

Note that in this case, the job can be assigned to an 8-core node (exclusively) or a 12-core node that is shared with other jobs. You can do something similar for the 12-core and 16-cores compute nodes.

Top of Page

Jobs: How can I determine whether or not I have all the processor cores in a node assigned to my job (and why would I want to do so)?

As mentioned above, if your SBATCH script asks for 8 cores your job could land on either an 8 or 12 core node. Intel CPUs support hyperthreading, which essentially allows each physical core to appear to be two cores. Many multi-processor applications can take advantage of hyperthreading to run in significantly less time. Please see http://www.intel.com/content/www/us/en/architecture-and-technology/hyper-threading/hyper-threading-technology.html for more information on hyperthreading.

Top of Page

Jobs: If I belong to multiple groups, how can I define the group name under which my job is to run on the cluster?

You can add the following line in your SBATCH script: `#SBATCH --account=<mygroup>`. Here, `mygroup` is the group name that you want the job to run under. To change group associated with a file, please see this FAQ.

Top of Page

Jobs: How do I checkpoint my job?

If your job runs more than a few hours, it is a good idea to periodically save output to disk in case of failure. There are also tools such as Berkeley Lab Checkpoint/Restart (BLCR) that provide supports for kernel/user level checkpoint/restart.

Top of Page

Jobs: How do I use local storage on a node?

In some scenarios it may be advantageous to read or write data to a compute node's local hard disk, rather than to/from our parallel filesystem (`/home`, `/scratch`, and `/data` are all stored on the parallel filesystem). One common example is if you will be reading or writing to/from a file frequently. Each compute node has a world-readable/writeable directory at `/tmp`. If you want to move files to this local storage, we recommend creating a subdirectory at `/tmp` and then copying data to it before launching a program that will read these data. Note: a program must know where to find these data, so you generally must provide an absolute path to the file from within your program. Please be sure to clean up your data at the end of your job (using the `mv` or `rm` commands). Below is an example of how this might be done within a SLURM job:

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

```
#SBATCH --ntasks=1
#SBATCH --nodes=1
#SBATCH --mem=4G
#SBATCH --time=4:00:00
#SBATCH --output=myjob.txt
localdir=/tmp/myjob_${SLURM_JOBID}
mkdir ${localdir} # create unique directory on compute node
cp mydata.txt ${localdir} # copy data to node
./run_my_prog # run program that reads/writes to/from local disk
rm ${localdir}/mydata.txt # remove data from local disk
mv ${localdir}/output.txt ./ # move results to working directory on GPFS
```

Top of Page

Jobs: a SLURM command fails with a socket timeout message. What's the problem?

Occasionally when you attempt to run a SLURM command (e.g. `sbatch`, `salloc`, `squeue`) the command may hang for an extended period of time before failing with the following message:

```
error: slurm_receive_msg: Socket timed out on send/recv operation
error: Batch job submission failed: Socket timed out on send/recv operation
```

This error results when the SLURM controller is under a high amount of stress. Avoiding this error requires all cluster users to play nice and follow cluster etiquette policies. Specifically, all cluster users are encouraged to

1. submit a large number (>100) of similar jobs as job arrays (see our SLURM documentation page for examples), and
2. avoid submitting a large number (>100) of short jobs (< 30 minutes). Job

arrays reduce the load on the scheduler because SLURM only attempts to schedule the entire array once, rather than every element within the array independently. Short jobs produce more "churn" within the job scheduler as it works to allocate, de-allocate, and re-allocate resources at a rapid pace. If you are running a lot of short jobs, please try to bundle multiple jobs together into a single job to put less stress on the scheduler. The socket timeout error message is generally intermittent, so if you wait a few minutes and try your SLURM command again it may complete immediately.

Top of Page

Why do files named "core.*" appear after I submit a job? {#corefiles}

Core files are created by the operating system when a process terminates abnormally and they contain the content of the memory at the moment of the crash. They are very useful for a post-mortem analysis for debugging purposes.

Disk space: Determining Disk Space Usage and Quotas

As noted in the cluster disk policies, you have both soft and hard limits on both your home and scratch directories. To help keep the system running smoothly, you should be in the habit of checking your usage level, especially since hard quota limits are definitive and, due to potential filesystem problems, we may have to either kill jobs or place temporary limits on accounts which exceed their soft limit. Please read our cluster disk policies to understand disk space quotas and the FAQ on how to increase your available diskspace by using scratch space, requesting a possible temporary quota increase, or purchasing more diskspace. To view your current usage and quota levels, type the command:

`accre_storage`

For example:

```
$ accre_storage
                         Space                           Files
                 Usage   Quota   Limit           Usage   Quota   Limit
Home:            0.50M   15G     20G             24      200000  300000
Scratch:         0.00M   50G     200G           4       0       0
$ 
```

The left section shows information about your current disk usage on `/home` and `/scratch`, while the right section shows your current file count usage. If you are in a group that purchases additional space on `/data`, you will see additional information below the `/scratch` line (see example below). Note that the **Usage** column is your current disk usage, the **Quota** column is a soft limit, while the **Limit** column is a hard limit. Definitions for soft and hard limits can be found on our cluster disk policies page. If you are exceeding either your disk space or file count soft quota, the relevant line will be colored yellow, as shown in the example below. Make sure you delete (or compress) files as soon as possible in order to avoid disk I/O errors once the grace period has expired. If a line is colored red, it means you have either hit a hard quota limit, or your soft quota grace period has expired; any attempts to create additional files in the corresponding storage will result in I/O errors.

```
$ accre_storage
                         Space                           Files
                 Usage   Quota   Limit           Usage   Quota   Limit
Home:            5.85G   2G      20G             8519    200000  300000
Scratch:         31.04G  10G     20G             2134    200000  1000000
Data (fileset quota):
           accre 1.00M   900T    1024T           74      0       0
             osg 0.00M   900G    1000G           1       0       0
$
```

Note that `/home` and `/scratch` are generally controlled with *user* or *group* quo-

15

tas, while **/data** (and occasionally **/scratch**) are controlled with *fileset* quotas. A user or group quota is based on the user or group owning a set of files, while fileset quotas are applied directly on files within a parent directory. One instance in which this distinction becomes important is when you are sharing files with a collaborator or labmate. With user-based quotas, if you copy a file into your colleague's home directory, the file will still count against your quota if the file owner is not changed. You can check the owner of a file by using the **ls -l** command. One other important detail about quota is data replication. ACCRE currently has data replication set to two for **/home** and **/data**. This means that the disk usage of a file stored in **/home** or **/data** on the cluster will be approximately twice that of a file outside the cluster. The **accre_storage** command shows you disk usage without data replication, so the output of **du -sh** (which shows you disk space of a directory or set of directories, and includes data replication) will differ from the **accre_storage** command. **ls -lh**, on the other hand, will show you file sizes without considering data replication, so it will be consistent with the output from **accre_storage**.

Top of Page

Disk space: My group has storage on DORS. How can I check our usage?

From any cluster gateway execute

```
mmlsquota -j <fileset name> -block-size auto dors
```

Typically, the fileset name is the same as your group name but may not always be.

Top of Page

Disk space: Using scratch disk space

You have disk and file allocations available for your use on both your home directory (which is backed up) and on scratch disk space (which is not backed up). To take advantage of your scratch disk space, simply **cd** to that filesystem and create your personal directory. For example:

```
cd /scratch
mkdir vunetid
chmod 700 vunetid
cd vunetid
```

where vunetid is your unique VUNetID, which is also your ACCRE user id. If you are unsure what your VUNetID is, simply type **whoami** while logged into the cluster to find out. Note that the **chmod 700** command is needed to set the appropriate permissions on your scratch directory so that only you can access it. Note that some ACCRE groups also have their own private shared group **/scratch** directories.

Disk space: If I need more disk space than this, will you temporarily grant a quota increase?

We do not necessarily relax quota restrictions. It depends entirely on the details of your request and we can discuss your options. Please submit a Request Tracker ticket explaining why you wish a quota change, how much space you believe you require, and how long you expect to need it. If you need more diskspace for an extended period of time you may purchase it. Please see the details of our cluster disk policies then send us your request.

Disk space: Will ACCRE restore deleted or lost data?

Yes. Please refer to our policy regarding restoring from backup.

My network connection to ACCRE is really poor and I have a lot of data that I need to upload to ACCRE (or download from ACCRE). What are my options?

To transfer files between your local machine and ACCRE, it is recommended to install and use FileZilla. FileZilla is a simple to use client which allows you to use the SFTP protocol to upload and download files between systems. To install FileZilla, simply go to their website and download the client The following is a beginner's guide to FileZilla: https://www.ostraining.com/blog/coding/filezilla-beginner/ If you do not want to overwrite files each time you upload a directory to the cluster then you can do the following: go to `Edit > Settings > Transfers > File exists action` and change the `Uploads` setting to `Overwrite file if source is newer`. Changing this setting will only upload files that are newer than the copy on the remote system.

GPU: How do I request a gpu node?

Currently, you must belong to a "GPU" group, such as `nbody_gpu` to gain access to one or more gpu nodes. Use the appropriate SBATCH command to submit your job and tell SLURM you want a GPU node. For example: `#SBATCH -p maxwell` This will place your job on a node with NVIDIA Maxwell Titan X GPU cards. More details about submitting GPU jobs can be found by clicking here.

Software: What research software packages are available on the cluster?

Run `pkginfo` to see a comprehensive list of available software packages that can be accessed from your environment by using the `setpkgs` command.

Top of Page

How do I make sure that my perl/python script is using the latest version available on the cluster?

First, add the appropriate package to your environment (e.g. your `.bashrc`/`.cshrc` file) with command:

```
setpkgs -a PACKAGE_NAME
```

Then, use the following line:

```
#!/usr/bin/env python (or perl)
```

as the first line of your script. This automatically detects the path to the added perl/python package and use that version as the interpreter of your script.

Top of Page

Software: I'd like to have some software installed on the cluster. How do I go about doing that?

As much as possible, ACCRE staff are glad to accommodate your needs for software. Of course, the software must be amenable to execution in the cluster environment and (if not open source) you are responsible for taking care of licensing arrangements prior to installation, as well as continued maintenance of the software license. If you'd like to explore the possibility of adding some software to our cluster environment, please submit a Request Tracker ticket. Note that we in general recommend that users install software into their cluster home directories. In this way you have complete control over the version of the software, applying updates, and so on. ACCRE staff are more than happy to assist you during this process.

Top of Page

How do I install an R package from source code?

R users should take a look at our R Software Page for details and best practices for using R on the ACCRE cluster. Here is an example that uses the `nlme` package. Login to the cluster, and, if you have not already done so, in your home directory create a directory for your R packages. Here is an example:

```
mkdir -p R/rlib
```

You will also need a tmp directory in your home directory, so do this in your home directory:

```
mkdir tmp/
```

You will need to add both the R and the gcc compiler with `setpkgs`:

```
setpkgs -a R gcc_compiler
```

Now change to your `tmp` directory, and download the source code:

```
cd tmp/
wget http://cran.r-project.org/src/contrib/nlme_3.1-104.tar.gz
```

Generally, it will only take a few seconds to download the "tarball", but sometimes it can take longer. Now start R:

```
R
```

At the R-prompt (denoted by `>`) tell R where you will keep your packages:

```
> .libPaths("~/R/rlib")
```

Next tell R to install the package:

```
> install.packages("nlme_3.1-104.tar.gz", repos = NULL, type="source")
```

R will now compile and install `nlme` into your personal R library: `~/R/rlib`

To test your install quit R

```
> quit()
```

Restart R and at the prompt

```
> .libPaths("~/R/rlib")
> library("nlme")
```

You should see nlme loaded. You need to remember to add these two lines to any script you feed to R if you intend to use nlme. If you wind up installing many packages you can put the `.libPaths("~/R/rlib")` command in your `.Rprofile`. You may now delete the sourcecode package:

```
rm nlme_3.1-104.tar.gz
```

What happens if R says that there are needed dependencies? This sometimes happens, and you will need to download and install those packages before installing the one you wanted. Just follow the steps outlined above until you have downloaded and installed all the packages.

Top of Page

Software: How do I download and install an R package from the internet?

R users should take a look at our R Software Page for details and best practices for using R on the ACCRE cluster. Here is an example that uses the `Zelig`

package. Login to the cluster, and, if you have not already done so, in your home directory create a directory for your R packages. Here is an example:

```
mkdir -p R/rlib
```

You will need to add both the R and the gcc compiler with `setpkgs`:

```
setpkgs -a R gcc_compiler
```

Now start R:

```
R
```

At the R-prompt (denoted by `>`) tell R where you will keep your packages:

```
> .libPaths("~/R/rlib")
```

Next tell R to install the package:

```
> install.packages("Zelig")
```

R will now give you a list of repositories to download from. Choosing the Tennessee repository seems good. That is choice `80`. R will now download, compile and install Zelig into your personal R library, `~/R/rlib`.

To test your install, quit R

```
> quit()
```

Restart R and at the prompt

```
> .libPaths("~/R/rlib")
> library("Zelig")
```

You should see Zelig loaded. You need to remember to add these two lines to any script you feed to R if you intend to use Zelig. If you wind up installing many packages you can put the `.libPaths("~/R/rlib")` command in your `.Rprofile`.

Top of Page

Software: How do I install and load an R package from Bioconductor?

R users should take a look at our R Software Page for details and best practices for using R on the ACCRE cluster. Here is an example that uses the `goseq` package. Login to the cluster, and in your home directory create a directory for your R packages. Here is an example:

```
mkdir -p R/rlib
```

You will need to add both the R and the gcc compiler with `setpkgs`:

```
setpkgs -a R gcc_compiler
```

Now start R:

```
R
```

At the R-prompt ">" tell R where you will keep your packages:

```
> .libPaths("~/R/rlib")
```

Next, point R to the Bioconductor site:

```
> source("http://bioconductor.org/biocLite.R")
```

Next, ask R to get the package, compile and install it in your personal R library (`~/R/rlib`)

```
> biocLite("goseq")
```

`goseq` and its dependencies will be downloaded, compiled, and installed. If everything succeeds you will see

```
* DONE (goseq)
```

After that, you may get a series of warnings about packages needing to be upgraded. You may ignore the warnings. To test your install quit R

```
> quit()
```

Restart R and at the prompt

```
> .libPaths("~/R/rlib")
> library("goseq")
```

You should see `goseq` and the two dependencies loaded. You need to remember to add these two lines to any script you feed to R if you intend to use goseq. If you wind up installing many packages from Bioconductor you can put the `.libPaths("~/R/rlib")` command in your `.Rprofile`.

Top of Page

Software: How do I install a perl module without root privilege?

You do not need to have root permission to install a module. You just install your PERL module locally in your home directory. Make a directory called, say, `lib/` in your home directory like this:

```
# first navigate to your home directory
$ cd ~

# now make a directory called lib
$ mkdir lib
```

Now you have a directory called `~/lib` where the `~` represents the path to your home dir. `~` literally means your home dir, but you probably know that already. All you need to do is add a modifier to your perl `Makefile.PL` command

```
$ perl Makefile.PL PREFIX=~/lib LIB=~/lib
```

This tells Make to install the files in the lib directory in your home directory. You then just `make`/`nmake` as before. To use the module you just need to add `~/lib` to `@INC`. Next, you modify the top of your own scripts to look like this:

```
#!/usr/bin/perl -w use strict; # add your ~/lib dir to @INC use lib
'/usr/home/your_home_dir/lib/'; # proceed as usual use Some::Module;
```

Top of Page

────────────────────────────

Software: How do I check which python packages are installed?

Python users should check out our Python Software Page for tips and best practices for using Python on the ACCRE cluster. First, make sure you have loaded the correct version of python into your environment (by typing something like `setpkgs -a python2`. You can check the versions of python installed on the cluster by typing `pkginfo | grep python`). Once you have done this, next type:

```
python_pkginfo.py
```

This will run a script that lists the python packages in your current environment, including any you have installed locally (see next section). `python_pkginfo.py` also accepts two optional arguments, `--ncol` (for adjusting the number of columns in the output) and `--type` (this controls whether installed packages, modules, or both are printed). For example:

```
python_pkginfo.py --ncol 3 --type both
```

would list all installed packages and modules in three columns of output. By default, installed packages are output in two columns.

Top of Page

────────────────────────────

Software: How do I install a python module without root privilege?

It's normally not necessary to use `sudo` to install python modules and packages. When trying to install a python module with `pip`, if you see an error similar to:

```
error: could not create
'/usr/local/python2/2.7.4/x86_64/gcc46/nonet/lib/python2.7/site-packages/doc':
Permission denied
```

you should provide the `--user` option to `pip`, e.g.:

```
pip install word-count --user
```

For more information, see our python software pages.

However, we do encourage users to create virtual environments, either through `virtualenv` or `anaconda`, to avoid dependency conflicts between packages.

Top of Page

Software: How do I install a python package from source code?

Python users should check out our Python Software Page for tips and best practices for using Python on the ACCRE cluster. Here is an example that uses the SQLAlchemy package. You will also need a tmp directory in your home directory, so do this in your home directory:

```
mkdir -p temp/SQLAlchemy
cd temp/SQLAlchemy
```

Download the source code, and untar it:

```
wget http://pypi.python.org/packages/source/S/SQLAlchemy/SQLAlchemy-0.7.9.tar.gz
tar xzf SQLAlchemy-0.7.9.tar.gz
```

You will need to add the appropriate version of python (the example below loads python2; you can check the versions of python installed on the cluster by typing `pkginfo | grep python`) as well as the gcc compiler and the atlas library with `setpkgs`:

```
setpkgs -a python2
setpkgs -a gcc_compiler
setpkgs -a atlas
```

Install the module:

```
cd SQLAlchemy-0.7.9
python setup.py install --user
```

This installs the package to `/home/YOUR.VUNETID/.local`. All packages installed to that directory are automatically added into the python environment.

Top of Page

Software: How do I run Matlab/SAS job on the cluster?

In order to run Matlab/SAS jobs, you must first purchase a license from ITS software store. Once ITS notifies us your purchase, you will be added to the relevant group so that you can have permission to run the software. License may not be shared among different users. However, with one license, you can run multiple jobs at the same time on the cluster.

Top of Page