

Introduction

This is not intended as an introduction to Linux (for which there are already many good books). Its aim is to give guidance on the aspects of the system that make the CSF service different from a standard Linux desktop.

Configuration of the farm

The farm has one front-end node which is used for batch job preparation, debugging and submission to the PBS batch queuing system for eventual running on one of the 110 batch worker nodes.

The farm consists of a mixture of machines -

40 dual Pentium III 1GHz machines with 512MB memory and 30GB of IDE hard disk,
40 dual Pentium III 600MHz machines with 256MB memory and 13GB of IDE hard disk,
30 dual Pentium II 450Mhz machines with 256MB memory and 10GB of IDE hard disk.

The production farm runs Redhat 6.1.

As currently configured the Linux farm will deliver approximately 340 Pentium 450 processor equivalents.

Printable versions of the set of web pages

Text and Postscript versions of the complete set of pages are available. We are hoping to improve these later.

Comments on these web pages

If you find these web pages are deficient in any way, please let us know. All comments should go to the author (A.Sansum@rl.ac.uk).

Logging on, passwords & printing

Logging on

The Linux front end can be contacted by the address:

csflinux.rl.ac.uk

Both ssh and telnet access are available - ssh is recommended for improved security. We expect to remove telnet access in due course.

After you log on, if you need access to your AFS account you will need to use the command: **klog**

Changing password and preferred shell

All details about individual userids are maintained in NIS.

You can change your password from the Linux service:

yppasswd

Choose your password wisely! A cracking program can easily guess passwords closely related to words appearing in any dictionary (English or another language). The CSF password file has a password cracker run on it every month. If your password is cracked, you will be notified by Resource Management and asked to change it - although inconvenient, such safeguards are necessary.

At RAL, AFS is independent of your normal CSF ID. To change your AFS password, use the command: **kpasswd**

To change your preferred shell, you should type the command:

ypchsh

and follow the instructions. You must logout and logon again for the change to take effect, it may take up to ten minutes for the change to become visible to the whole of the Linux farm.

Printing

Use **lpr** to print files from Linux. A number of RAL local printers are already set up. We are happy to add external (non RAL) printers to the printer configuration files. If you have a local printer you wish to use, please contact us with details and we will arrange for it to be configured.

Getting Started on the CSF Linux Farm

Help and Information

Getting Help

Urgent problems should be reported to:
support@csflinux.rl.ac.uk

This mailbox is monitored even if the system manager is away from the lab.

Less urgent matters, comments, discussion and so forth may be mailed direct to the system manager if you prefer.

Email: A.Sansum@rl.ac.uk

Tel: 01235 445863

If you think something is broken, please don't assume that we know about it. The service has a certain amount of automated monitoring, but it is unfortunately still the case that some problems will first be identified by those using the service. If you think there is a problem, please let us know.

We try and provide a friendly and helpful service, please do not hesitate to get in touch.

Helping Us to Help You

The most common cause of delays in resolving problems is that of insufficient information to allow us to make a diagnosis. Queries such as: "*Help, my job script does not work*" or "*My batch job failed unexpectedly*", are not very helpful (and yes both of these examples are often received). To get the best out of us, provide all relevant information.

- Tell us what platform you are on (Sun, Linux) .
- Tell us what happened. Include all relevant error messages or point us at the file containing the job output.
- If it's a batch job script, let us know where it is so we can examine it.
- If you had a sudden unexpected failure in batch, let us know what batch worker the job was on when it failed (details printed at the start of the job). The problem may be specific to one or several workers.

Documentation

News about the CSF service is available in a number of ways. It is a relatively low bandwidth facility and it is the mechanism used to inform you of scheduled changes and recent problems. It is assumed you read it - **ignore it at your own peril!**

- News about the service is displayed in the message of the day when you log on to csflinux
- The news command on csflinux gives access to earlier news items
- News items are also distributed by the CSF mailing list, CSF-L. To subscribe to the mailing list, send an email to listserv@listserv.rl.ac.uk with the command `subscribe CSF-L firstname lastname` in the text of the message (substituting your first name and last name for `firstname` and `lastname`)
- You can read the Usenet group `ral.services.csf` or via the web

Getting Started on the CSF Linux Farm

Disks and Filesystems

Disks and Filesystems

The filesystem environment of csflinux and the batch workers is as far as possible identical.

- All systems share the same home filesystem (`/home/csf`). This filesystem has quotas set for individuals. You can find your current quota using the command:

CSFquota -v

You may exceed the quota for up to seven days, but may never exceed the limit. Once you exceed the limit or the time limit on the quota no further files may be written.

If you find that your quota is insufficient, please contact us, explaining your requirements and request that it be increased - we are usually happy to do so!

- There is a scratch disk (presently 30GB). Every group should have a directory: `/scratch/<group>` and individual users are free to create subdirectories: `/scratch/<group>/<userid>`. This disk is cleaned hourly, files may last as long as 30 days but will be deleted sooner (oldest first) if the occupancy rises above 80%.
- The tmp filesystem is very small and is intended only for the normal functioning of Unix; for lock files and so forth that many applications expect to be able to place on `/tmp`. Larger files should be placed on `/scratch` as described above.

- Individual experiments have filesystems mounted under **/rutherford** and **/stage**. These have been allocated on request to individual experiments. Contact your local experiment experts for more information. If you are unsure who may know, contact the system manager. If your work would greatly benefit from additional disk capacity, please contact us - we may be able to make more space available - either permanently or as a short term loan.
- **AFS Access.** By default you are not given an AFS ID/directory when you are registered on CSF. If you want a RAL AFS ID then email afsman@rl.ac.uk. Although AFS is mounted on all batch nodes, the batch system is not AFS aware and jobs submitted from the front end system will not have an AFS token when they start running on the batch worker. Because of this limitation only world readable/writable directories can be accessed from batch.

All these filesystems (except AFS) are NFS mounted from a number of different Unix file servers.

Backup policy

The various disk areas have different backup policies:

- `/home/csf` is backed up nightly
- `/rutherford` is back up nightly unless file `.DONOTBACKMEUP` exists in the top directory
- `/stage` is NOT backed up
- `/afs/rl.ac.uk/user` is backed up nightly
- Other AFS volumes are not backed up unless required/agreed

Getting Started on the CSF Linux Farm

Software

User Environment

The HEPiX startup scripts set up the initial user environment. Consult the HEPiX documentation for details of how to tailor this to suit yourself.

GNU Software and CERN's ASIS

Most of the GNU software such as `gcc`, `g77`, `egcs` and so forth are available on vanilla Redhat 6.1. The user `PATH` environment variable is set such that the system copies of these are the default (typically in

/usr/bin). These vanilla versions will be upgraded as demand requires - contact the system manager if you believe an upgrade is needed.

CERN also provide copies of these products by ASIS. Typically the CERN release is a different version to the system version. To allow CERN experiments access to their "production" compilers, these are installed in /usr/local/bin in the same way as at CERN. As /usr/local/bin is (by default) after /usr/bin in the PATH, the CERN versions will not be used unless explicitly referenced or the PATH variable is changed. For example by:

```
export PATH=/usr/local/bin:$PATH
```

This duplication of compilers was not done without much heart searching and discussion. If you have a view on the matter let us know.

X11

The X11 libraries can be found in **/usr/X11R6**

NAG Library

The NAG FORTRAN library is installed as /usr/local/lib/libnag.a

Commercial Compilers

We have a limited number of licences for the Fujitsu FORTRAN 95 compiler. These cost real money and will only be made available to those who are able to make a convincing case.

RAL Datastore Software

Access to tapes in the RAL Datastore is by the commands:

- datastore (for creating tapes)
- tape (for reading/writing them).

There is a set of web pages describing the RAL Datastore and its use, at RAL Datastore. Use of the datastore is free to HEP users. Most experiments already have storage space allocated - contact your experiment experts at RAL for details or alternatively email support@csflinux.rl.ac.uk for advice.

ASIS Software

Much of the software installed under /usr/local comes from CERN's ASIS repository. This software is maintained by CERN staff and changed as and when they choose.

Interactive debugger

The interactive debugger ddd is available. The man command gives details of its use.

Getting Started on the CSF Linux Farm

The Interactive front end

The Linux service is provided primarily as a batch computing service for UK particle physicists. A high level of support is provided for software and utilities necessary for physicists to carry out this type of work. No support is provided for "desktop" type activities such as web browsing, document preparation and so forth. Tools presently installed that may be used for this kind of activity are not guaranteed to remain on the service, nor will work of this sort be allowed to impact the service.

The front-end service is intended for code development, testing, job submission and so forth. Please do not run background jobs on the front end unless it is necessary for debugging that cannot be done in batch - they degrade the interactive response time for everyone. If you absolutely must run interactive jobs, please use the **nice** command to minimise their impact on other interactive users:

nice -20 mycommand

Getting Started on the CSF Linux Farm

The Batch System

The Batch System (PBS)

All users should log in to the interactive service (csflinux). The remaining "batch" nodes are available only for work submitted via the batch system. Interactive logins are presently allowed to the batch workers, but only to carry out debugging tasks relating to the batch work.

The batch system is PBS - superficially this looks like NQS (for example there is a qsub command) - but

do not be fooled. PBS, although similar, has many different features . These notes assume users have some experience of batch systems in general, possibly NQS or LSF. The impatient should read the Quick Start section below, and more detailed configuration information is given in Batch routing queues, Job submission, and Monitoring batch jobs.

Quick Start

To submit a batch job: **qsub** <scriptname>

where <scriptname> is the shell script to be submitted to the batch system.

To see the status of your batch job: **qstat** <jobnumber>

where <jobnumber> is the number assigned by PBS and sent back to you when qsub is obeyed.

By default, when the job finishes the job output and error log files are returned to the directory from which you submitted the job as files: <scriptname>.o<jobnumber> and <scriptname>.e<jobnumber> respectively.

Please see the other pages for information about queues and monitoring your jobs (Batch routing queues, Job submission, and Monitoring batch jobs).

Getting Started on the CSF Linux Farm

The batch routing queues - bulk and express

Job Queues

There are two routing queues (**bulk** and **express**) into which you may directly submit work and four execution queues in which jobs will actually run. Details of the queue configuration are given later, but first are described some general principles.

Your job will be started provided:

- There are free processors available - every job gets one dedicated processor
- The system has not exceeded the maximum of allowed jobs (typically this is set to equal the

- number of processors).
- The user has not exceeded the maximum number of jobs allowed per user within the whole system.
- All higher priority queues have been processed
- The queue has not exceeded the maximum number of allowed jobs for that queue
- The user has not exceeded the maximum number of jobs per user for that queue
- The job is considered by the scheduler to be the most eligible job to run within the queue

The bulk queue

The default routing queue (if you specify no other) is **bulk** - it is recommended that your high volume work is submitted to the bulk queue as this queue will allow you by far the largest number of running jobs.

There are three execution job queues which are fed by the bulk routing queue. PBS chooses which execution queue a job will run under according to the maximum CPU time requested by the job.

QUEUE	CPU Time (hrs:mins:secs)
S	8:00:00
M	30:00:00
L	168:00:00

It is intended that the bulk of "production" work will be run in queue M, which is where the main throughput will be provided; this is the default queue. About 25% of capacity is given to queue L - this is intentionally limited to ensure reasonably quick turnover of batch work in the farm, preventing single individuals from dominating queues for long periods of time.

These queue limits are intended to be a configuration that meets our requirements (both yours and ours). If you have any comments or suggestions, they will be most welcome.

The express queue

This is intended for test and debugging runs; only a few jobs are run at a time for each user. It has a CPU limit of 30 hours and has a higher priority than any of the bulk queues.

To submit a job to the express queue:

```
qsub -q express <my.script>
```

where <my.script> is the file containing the batch job. The job is passed from the express routing queue to the E executing queue. If no CPU time limit is given in the job the default of 30 hours is used.

More qsub options may be specified, if necessary. See Job Submission for details.

Getting Started on the CSF Linux Farm

Job submission

The PBS command **qsub** is used to submit jobs. At its simplest:

qsub <scriptname>

will submit script <scriptname> to the bulk queue. When the qsub command has been obeyed PBS returns the jobnumber which has been given to the batch job; this can then be used with the **qstat** command to monitor the job's progress.

(If no CPU time limit is given on a directive in the script, the job will run in the M queue with a 30 hour time limit.)

The job runs in your home directory. If you want to access files in subdirectories you must use the **cd** command or give path names explicitly.

The **qsub** command has a number of options which can be given either on the command line or as PBS directives at the beginning of the job script. The man page for **qsub** give full details, but below we give the most useful ones.

- Jobname

By default the jobname is the same as the local name of the script file. This can be changed by using the **-N** qsub option or perhaps more usefully by putting a PBS directive at the beginning of the script file.

```
#PBS -N jobname
```

- CPU time limit

This is specified as one of the possible resources for the **-l** option.

qsub -l cput=n myscript sets a limit of n seconds

qsub -l cput=3:00 myscript sets a limit of 3 minutes

qsub -l cput=6:00:00 myscript sets a limit of 6 hours

The equivalent PBS directive to

```
qsub -l cput=3:00
```

is

```
#PBS -l cput=3:00
```

- Restartable job

A restartable job will restart if the batch worker crashes or if the system administrator forces your jobs off the machine.

The option **-r n** says the job is not restartable.

The option **-r y** says the job can be restarted after a break.

The equivalent PBS directives are **#PBS -r n** and **#PBS -r y**. The default is that jobs are restartable.

- Job output

By default the standard output stream is returned to the directory from which the **qsub** command was obeyed as `<jobname>.o<jobnumber>`. The standard error stream is similarly returned as `<jobname>.e<jobnumber>`. You can reroute the standard output stream by using the **-o** option, eg **-o resultsdir/jobout**. Similarly you can reroute the standard error stream by using the **-e** option.

The **-j** option allows you to merge the two output streams together.

-j oe directs that the two streams are merged as standard output.

-j eo directs that the two streams are merged as standard error.

- Example job script with directives in the file

```
#PBS -N testjob1
#PBS -l cput=3:30
#PBS -j oe
cd subdir
# run my program in my chosen subdirectory
```

This job will be called `testjob1`, have a CPU limit of 3 minutes 30 seconds, and the two output streams will be merged as standard output. The output will be returned as `testjob1.o<jobnumber>`

Note that the PBS directives must be at the beginning of the file, before any executable line.

If an option is present as both a directive and on the command line, the command line takes

precedence. This allows you to override values set in the job script for an individual run.

Note that it is not possible in PBS to include arguments following the job script name. In order to run a script with arguments you need to submit it as follows:

```
#####  
#!/bin/sh  
cat <</EOD | qsub -l cput=hh:mm:ss -N myjobname -  
my.script arg1  
/EOD  
#####
```

Here the - character tells qsub to read from stdin which is provided inline.

Getting Started on the CSF Linux Farm

Monitoring batch jobs

The command **qstat** allows you to check what is happening to your batch jobs. The man page for qstat gives full details of all the options but we just give the most useful ones here. You do not need to know which host your job is executing on, PBS handles that itself.

The command **xpbs** provides a graphical interface to information using X Windows. It can be rather daunting to use to start with but it has its own extensive help information.

- Individual job

It is worth noting the jobnumber assigned by PBS to the job when it is first submitted. Then you can query the state of that individual job:

qstat <jobnumber> gives basic details about the job <jobnumber>

qstat -f <jobnumber> gives full details of that job.

qstat -s <jobnumber> says why a job is not running, or, if it has started, when it started.

- All your jobs

qstat -u <userid> , where <userid> is your userid, gives information about all jobs under that userid, whether queued or running. You can also see which queues they are in.

qstat -r -u <userid> lists the running jobs for that userid.

- Looking at queue details

The four execution queues are L,M,S for the bulk routing queue and E for the express routing queue.

qstat <queuename> lists all jobs in that queue

qstat -q gives the CPU limit and total job limit for each queue. (Adding a queue name to this limits the information to that queue).

qstat -f -Q <queuename> gives full information about that queue.

- Looking at global queue limits

qstat -B -f gives totals and defaults for the whole server.

- Information for debugging problems

qstat -f <jobnumber> gives you full details of the job, including which host it is running on and when it started.

- Looking at job outputs for running jobs (qcat)

Because PBS does not provide the ability to look at the output of running jobs, we have written a little script that allows you to do this.

qcat -o <jobid> # List standard out

qcat -e <jobid> # List standard error

The first time you connect to any node you will get a "new host key" request from ssh.

Cancelling batch jobs

Cancelling is done with the **qdel** command. **qdel** <jobnumber> cancels your job with job number <jobnumber>. If it is already running the job output is returned as normal. If it is still queued no output is returned.

qalter and qorder

qalter <attributes> <jobnumber> modifies the given attributes of the job specified by <jobnumber>. The

attributes **-e** path, **-N** jobname, **-o** path and **-l cput=time** can all be altered by this command.

qorder <jobnumber1> <jobnumber2> exchanges the order of the two batch jobs in a queue. This cannot be used once a job is running and queue limits cannot be exceeded.

Getting Started on the CSF Linux Farm

Job runtime environment

Shell Strategy

The free shell strategy used on the Linux farm aims to duplicate the shell choice that would be made if the script was run interactively (ie your default shell would be used unless the shell is specified in the job script). You may override these choices should you wish by using the batch submission option: **-s**

HEPiX Environment

When your batch job starts, the HEPiX startup scripts set the variable: **ENVIRONMENT** to be **BATCH**. You may if you choose use this within your user profiles to distinguish between your **BATCH** and **LOGIN** environments.

Work Directory (\$WORKDIR)

When your batch job starts it will be allocated a work directory on the batch worker. The location of this directory can be obtained by the environment variable **\$WORKDIR**. The allocation is 3GB of disk space on **\$WORKDIR**. **At present there is no mechanism to test if this limit has been exceeded.** Please do not exceed this figure - you may break other users batch jobs.

It is recommended that the **WORKDIR** is used for:

- Initial configuration files, executables and so forth that may be accessed over the whole time the job runs.
- I/O intensive data files - where it is recommended they are copied into **WORKDIR** before the batch job starts.
- Data written out by the job prior to staging to tape

These are only recommendations and different applications may need to do things differently. However, by following these guidelines, your batch job will be more independent of the network, file servers and

so forth and will therefore be able to cope more robustly with any problems.

When the batch job finishes, all files remaining in \$WORKDIR will be cleaned up. Any that you need must be copied to more permanent storage.

Getting Started on the CSF Linux Farm

User Registration

If you have a colleague who wishes to request an ID on CSF, they should email

support@csf.rl.ac.uk

We will need the following information:

Full Name
Home University
Address (either of the University or CERN/DESY etc if you are working abroad long term). This is the address that registration details will be posted to.
Telephone number
Email address
Experiment
Preferred id
Preferred shell

If you have any problems or queries please contact CSF User Support

[Return to CSF Linux Service User Guide top page](#)

The e-Science Centre is part of CCLRC



Last updated: November 2001