Scratch Purge Policy

Introduction

Pawsey provides the /scratch file system on Magnus as a temporary workspace for staging and running jobs and for data input and output. It is a 3-Petabyte Lustre file system that has been configured for fast I/O performance.

Pawsey implements a dynamic allocation system, which involves purging old files automatically, to maximise the availability of the scratch space for large-scale, data-intensive computations.

Note that the /group filesystem is provided for medium term storage and is not purged. Nevertheless the allocated quota is limited.

Pawsey Magnus /scratch System Purge Policy

✅ **Files not in use for longer than 30 DAYS may be purged**

Magnus users do not receive a space quota-ed allocation on the scratch file system. Instead the file system is treated as a shared resource on which older files are removed periodically to free up space. In general, files are considered to be old -- and thus candidates for deletion -- if they have not been accessed (either read from or written to) for 30 DAYS.

Please note that from 2019 we will be imposing a quota on the *number* of files/directories each user will be allowed to store on /scratch of **1 million.** If you have more than this, you may find that jobs will fail when writing output.

You can check your usage of /scratch at any time by running **lfs quota /scratch** which will show you (first line, '14' in the example below) how many files you have on /scratch

```
aelwell@magnus-1:~$ lfs quota /scratch
Disk quotas for user aelwell (uid 20701):
 Filesystem  kbytes  quota  limit  grace  files quota limit grace
   /scratch    1460     0      0      -      14    0     0     -
```

⚠️ **Remove files that are not longer in use**

/scratch is a shared resource that suffers low performance when the number of files increases too much. USERS SHOULD REMOVE THEIR FILES THEMSELVES if those files are no longer in use, rather than leaving them for purge. For removing a large number of files check: Deleting large numbers of files on scratch and group

In exceptional circumstances, if the scratch file system is close to full, files that are less than 30 days old may be deleted. However, in such circumstances, Pawsey aims to contact users by email to advise of the situation and to give them time to transfer their data files to other storage facilties.

Motivation

On previous supercomputers, the scratch file system was statically allocated to projects for their duration. This led to unintended results:

- ***Users unable to take advantage of the full scratch system.*** The space is fully allocated, but not fully utilised. When implementing a quota-ed scratch file system, each project is provisioned with the maximum amount of storage that is require over the lifetime of the project. So if a project requires a lot of storage for one large job that will be run over the course of a few weeks, they are given that maximum amount of storage for the entire allocation period. This means that for much of the allocation period, that scratch space will be empty. Even though the file system is not full, the scratch space is allocated, so cannot be used by other computations from other projects.
- ***The use of the scratch system for long-term storage.*** The purpose of scratch space is as a temporary work space: it is not well-suited for long-term storage of datasets. Pawsey has other long-term storage space, which offers greater capacity and better resilience to failure and data loss. The technology behind the Lustre file system is designed for optimal read and write (that is, I/O) performance, not for long-term storage. When scratch space is allocated, a user tends to view it as a static resource rather than as a dynamic resource. Such use restricts the possibilities for different projects to run data-intensive computations and means a user is storing their data outputs in a less resilient environment than is available.

With a purge policy, users are able to use more of the scratch system at the time when they need it. For example, if projects A and B each require 500 TB of space to run one big job, Project A can use that space at one time, and project B can use that same disk space at another time (when the data for Project A has been removed by the user or purged by the system). Under an allocation policy, these two projects together would need to be allocated half of the available resource, leaving little for the rest of the projects on the machine.

Data should be moved out of the scratch file system as soon, as is practically possible, into longer term storage such as Data Stores or third-party storage facilities (note that to use Data Stores, one needs to complete a separate application process available from here).

Many other supercomputing facilities also utilise a scratch purge policy, including CSCS (Swiss National Supercomputing Centre), NERSC (Lawrence Berkeley Laboratory, USA) and OLCF (Oak Ridge Leadership Computing Facility, USA).