Question from WC/LQ1 team:

"It would be helpful if the geant team could describe for us the typical number of files and sizes your workflow uses. We'd also like to know the access patterns, e.g. many processes read a few files or many processes do io on many distinct files."

**Storage needs related to Geant4 CPU and memory profiling**

Currently nfs mounted g4 disk space and Lustre lfstev disk space are used on the Wilson cluster.

The g4 disk is Geant4’s own disk, and on lfstev Geant4 project is allocated 10TB quota.

Similar quota has recently been granted to Geant4 project on the LQ1’s lustre1 disk.

Each profiling round requires 3 brand new separate builds of Geant4 (traditional,

MT, and VecGeom-based), including Geant4 core libraries as well and Geant4 based applications used for profiling.

The total size of the involved files is ~16GB (<20GB).

In files can be very small (e.g. cmake produced files of a few bytes) to larger files

of ~1GB (e.g. libraries).

Currently, the nfs mounted g4 disk is used to this.

NOTE: Attempts to migrate the build step from g4 disk to lfstev resulted in a problem

As Open|SpeedShop was no longer able to detect specific functions from the Geant4

Libraries but was only seeing libraries as a whole. So far we have not been able to understand the nature of the problem. Neither do we know at present if the problem will persist if we attempt to repeat such exercise e.g. on the resources of the LQ1 cluster.

Following build step, the Geant4 libraries and applications are packaged into tarballs that

are subsequently transferred to the worker nodes at the beginning of the jobs and are used by the processes that each job may consist of.

The fcp utility is currently used for the transfer but can be replaced by more appropriate option as available/recommended.

Each profiling round also relies on a group of external packages/tools, e.g. Open|SpeedShop,

VecGeom, etc.

The total space currently occupied by these tools is ~9GB (<10GB).

These tools are typically re-used from one profiling round to another but may require

periodic updates.

Most of the profiling activity is CPU and not I/O bound.

However, two (2) I/O intense aspects need to be taken into account.

The jobs do read in collections of Geant4 data files at the initialization stage.

This is fairly intense I/O activity lasting about 3 minutes or less.

It should be noted that in a realistic situation there may be multiple processes simultaneously reading in the data files in question. At present up to 21 amd32 worker nodes are available for this activity (priority queue), and a job on each node runs 30 processes, each processes reads at initialization the data files in question; thus there may be up to 24x30=630 processes running on the amd32 nodes and simultaneously reading the data file. In addition up to 25 intel12 nodes can be used (opportunistically at present, although in the past there was a priority queue as well), each running up to 10 processes, which may add up to 250 processes simultaneously reading data files from disk. In total there may be up to 630+250=880 such processes.

The data files in question are currently located on lfstev disk space and occupy ~13GB (<20GB).

These data files are typically reused from one profiling round to another but periodic updates are necessary.

In general, these files may be stored in the Geant4 dedicated CVMFS disk space.

The profiling jobs also write out some diagnostic data. Initially the diagnostic data are written

On the scratch disk on a worker node. At the end of the jobs collections of such data are

transferred from the worker nodes to the disk.

Each process writes and transfers its own output.

The fcp utility is used for the transfer.

However, when producing diagnostic data, multiple (many) processes are likely to be using the same tools (e.g. Open|SpeedShop utilities) located on the disk.

Typical size of output of a profiling round is ~<50GB; currently lfstev disk space is used to store

Such diagnostic output data for a short term analysis.

Additionally, /pnfs/geant4 disk space is mounted and used for long term storage of the diagnostic data from CPU profiling activity.

**Original materials composed by Krzysztof G. on 11/22/2019 with regards to (possible) migration of the Geant4 activity to LQ1 cluster**

General comments:  
  
Most of the current g4 group work is either CPU profiling and  
benchmarking of Geant4 or R&D work related to Geant parallelization  
and enabling of GPU usage.  
  
In a large majority of the tests, relatively little data is produced  
and if it is, the produced files are not very large, ~0.5GB, with a  
large spread in size, with occasionally larger ~50GB files or very  
small ones <1kB.  
  
The jobs do read in collections of Geant4 data files at the beginning  
of the jobs and write out some, usually diagnostic data, with most of  
the operations being CPU and not I/O bound, with some more intensive  
I/O activity at the initialization stage (lasting about 3 minutes or  
less) and collection of the data from the worker node at the end of  
the jobs.  
  
Suggested storage space additions:  
  
In order for the g4 group to efficiently utilize LQ1 cluster nodes, we  
would suggest addition of the following storage space areas to what is  
currently available to the group.  
  
New, not backed up, possibly nfs mounted, space of about 2+TB on a  
"build" or an interactive node, to do builds and to collect outputs  
from the jobs for a short term analysis. The files there can be from  
very small, a few bytes e.g. cmake produced files, to larger input or  
diagnostic output data files of ~<50GB. We expect to either remove/replace or  
quickly copy the files to a dCache or tape space, so the expected  
lifetime of those files, while they are not backed up, is of the order  
of a small number of days or less, therefore the hardware itself does  
not need to be of the highest grade (but it should not impede the work  
either).  
  
New, cvmfs allocation of about 200+GB for geant4 project to store less  
changing files related to Geant4 builds and the accompanying data.  
  
In addition to the above, /pnfs/geant4 should be mounted on LQ1, at  
least on the same node(s) where the above mentioned disk space would  
be mounted to be able to transfer files to and from the pnfs space.