Minutes of the ABP Computing Working Group meeting

23rd November 2017

Participants: L. Barraud, X. Buffat, G. Iadarola, N. Hoimyr, L. Mether, E. Métral, J. Molson, G. Rumolo, P. Llopis Sanmillan, J. Komppula, C. Lindqvist

G. Rumolo reported that following the floods in Bologna, the hardware will be moved to another building. It is unclear which fraction of the hardware is still functioning, but they expect the remaining nodes to be available again next week.
G. Rumolo reported that there were some positive feedbacks on the availability of HTCondor. N. Hoimyr confirmed that the rate of tickets has decreased recently.

N. Hoimyr gave an update on the status of the HPC facility at CERN. Two new clusters each composed of 72 nodes with 20 cores and 128Gb each are being installed and will be ready for testing next week. The nodes are connected with 10Gb/s ethernet. Two SLURM partitions are setup, called *be-short* (limited to 2 days of run) and *be-long* (unlimited duration). Each cluster will be dedicated to one of the two queues. A logic allowing the two clusters to run jobs from the two queues could be implemented later on if found necessary (e.g. if one of the two queues is significantly more loaded than the other). The compilation with OpenMPI and Mvapich2 are available on the frontend, accessible via ssh at *hpc-batch.cern.ch*. EOS and AFS are accessible from the frontend. N. Hoimyr provided links (in the slides) to the documentation of SLURM, to efficiently submit jobs.
A storage space */hpcscratch* is available for efficient storage of temporary data, which can be then moved to another location. Direct interaction with AFS or EOS during the execution of the jobs is also possible.
The pilot cluster experience was overall positive, in particular long running jobs did not exhibit stability issues.
N. Hoimyr asked for typical jobs samples, in order to test and tune the cluster, ideally next week such that they can be used for the setting up. He added that a bit of code profiling could be performed by IT, within reason.
If specific packages are needed and are not already available on the cluster it is preferable to ask IT to install a new module, rather than installing it locally.

- X. Buffat reported some past experience where hyperthreading had a negative impact on hybrid simulation with MPI and OpenMP due to oversubscribtion of threads. P. Llopis Sanmillan answered that oversubscription can be avoided using the setup parameters. For simulations with a significant load due to I/O, hyperthreading can improve the performance, therefore hyperthreading will be enabled on the worker nodes.
- N. Hoimy mentioned that the possibility to reserve temporarily resource in case of urgent needs is possible with SLURM, however a highlevel interface is not available.
- L. Barraud asked whether hybrid parallelisation schemes (MPI and OpenMP) can be used. P. Llopis Sanmillan answered positively.
- J. Komppula asked about an efficient way to transfer the data, since e.g. accessing EOS data from a local machine requires a cumbersome access via lxplus. The scratch space will be accessible via the frontend and the data can be copied directly from there.

- J. Komppula noticed a power failure when he launched a large job on the pilot cluster, he asked whether this is a coincidence. N. Hoimyr answered that the power failure seemed related to another cause, but encouraged J. Komppula to try again.
- G. Iadarola asked whether jobs can be chained. P. Llopis Sanmillan said that it should be doable by submitting a new job within a running job.