Grid Engine batch system integration in the EMI era

Wednesday, 28 March 2012 12:00 (30 minutes)

Impact

The Grid Engine families of schedulers are a mature software that has been used for years in HPC and non HPC environments –ranging from small data centres to some of the TOP500 supercomputers. The Open Source initiatives supporting the development of the existing forks have a strong community and inertia behind, and are committed to continue supporting the development of this software. This EMI integration is an important step, as it will make possible that already existing computing centres join major Grid infrastructures.

Description of the Work

BUpdaterSGE is the main daemon responsible for the GE integration with CREAMCE. It uses GE commands to determine and control the state of submitted tasks, and feeds the Blah Job Registry. BUpdaterSGE daemon was refactored to solve major bugs, as for example, the fact that completed jobs never changed their status in the Job Registry, and jobs were just considered finished through the CREAMCE Job Wrapper. This problem is now fixed and the efficiency of BUpdaterSGE has been largely improved. Moreover, the number of qstat and qacct operations was reduced (about 80%),reducing significantly the cpu and memory usage, and decreasing the time elapsed between each job's status change. Installation paths were accommodated to satisfy the relevant EMI policies, implying that scripts and daemons are now deployed according to the Unix Filesystem Hierarchy Standard (FHS). MPI support was fully finished and users can now submit MPI jobs using OpenMPI or MPICH/MPICH2 flavors. Finally, GLUE2 compliance is currently under development and is estimated to be fully supported by the end of January. In future releases GLUE1 will be deprecated and substituted by the new schema.

DRMAA API is another feature to be included in the future to improve execution efficiency and reduce the code's complexity. Presently, daemons make direct calls to the batch system commands like qstat or qacct and their results are gathered and parsed. The use of the DRMAA APIs will avoid this intermediate process, , improving efficiency and reducing cpu and memory usage.

Conclusions

The effort done for integration made possible to support Grid Engine, offering sites the possibility to install a modern batch system. The new release brings new features and many bug fixes reported by the users. With the new release the grid sites based on CREAM and GE will work more efficiently besides, the GE working team will help and support the new users to solve any issue.

Overview (For the conference guide)

Each batch system of Grid Engine (GE) family, successors of SGE, is a modern open source batch system which is now supported in EMI under a new CREAMCE version, and fully compliant with EMI policies and

FHS. This integration aimed to increase the robustness of the delivered functionalities with a major refurnish of the BUpdaterSGE, the daemon responsible for tracking submitted tasks. BUpdater was changed to get jobs status more efficiently and is now able to collect the information status for large amount of jobs in a few seconds.. The new implementation also includes default MPI support while GLUE2 schema compliance is still in development, which will be included in future versions. To improve the interaction with external services, DRMAA was positively evaluated to be included into the future development.

Primary authors: BORGES, Gonçalo (LIP); ROSENDE DOPAZO, Roberto (FCTSG); LÓPEZ GARCÍA, Álvaro (IFCA); SIMÓN GARCÍA, Álvaro (FCTSG)

Co-authors: FERNÁNDEZ SÁNCHEZ, Carlos (FCTSG); FERNÁNDEZ DEL CASTILLO, Enol (IFCA); FREIRE GARCÍA, Esteban (FCTSG); ORVIZ FERNÁNDEZ, Pablo (IFCA)

Presenter: ROSENDE DOPAZO, Roberto (FCTSG)

Session Classification: EMI: Job processing components