PoS PROCEEDINGS OF SCIENCE

Grid Engine batch system integration in the EMI era

Roberto Rosende Dopazo[1]
FCTSG
Avd. De Vigo s/n, 15705 Santiago de Compostela, Spain
E-mail: rrosende@cesga.es

Alvaro Simón García
FCTSG
Avd. De Vigo s/n, 15705 Santiago de Compostela, Spain
E-mail: asimon@cesga.es

Esteban García Freire
FCTSG
Avd. De Vigo s/n, 15705 Santiago de Compostela, Spain
E-mail: esfreire@cesga.es

Carlos Fernández Sánchez
FCTSG
Avd. De Vigo s/n, 15705 Santiago de Compostela, Spain
E-mail: carlosf@cesga.es

Alvaro López García
IFCA
Avd. de los Castros s/n, E-39005 Santander, Spain
E-mail: aloga@ifca.unican.es

Pablo Orviz Fernández
IFCA
Avd. de los Castros s/n, E-39005 Santander, Spain
E-mail: orviz@ifca.unican.es

Enol Fernández Del Castillo
IFCA
Avd. de los Castros s/n, E-39005 Santander, Spain
E-mail: enolfc@ifca.unican.es

[1] Speaker

PROCEEDINGS OF SCIENCE

Gonçalo Borges
LIP
Av. Elias Garcia 14 - 1º, 1000-149 Lisboa, Portugal
E-mail: `goncalo@lip.pt`

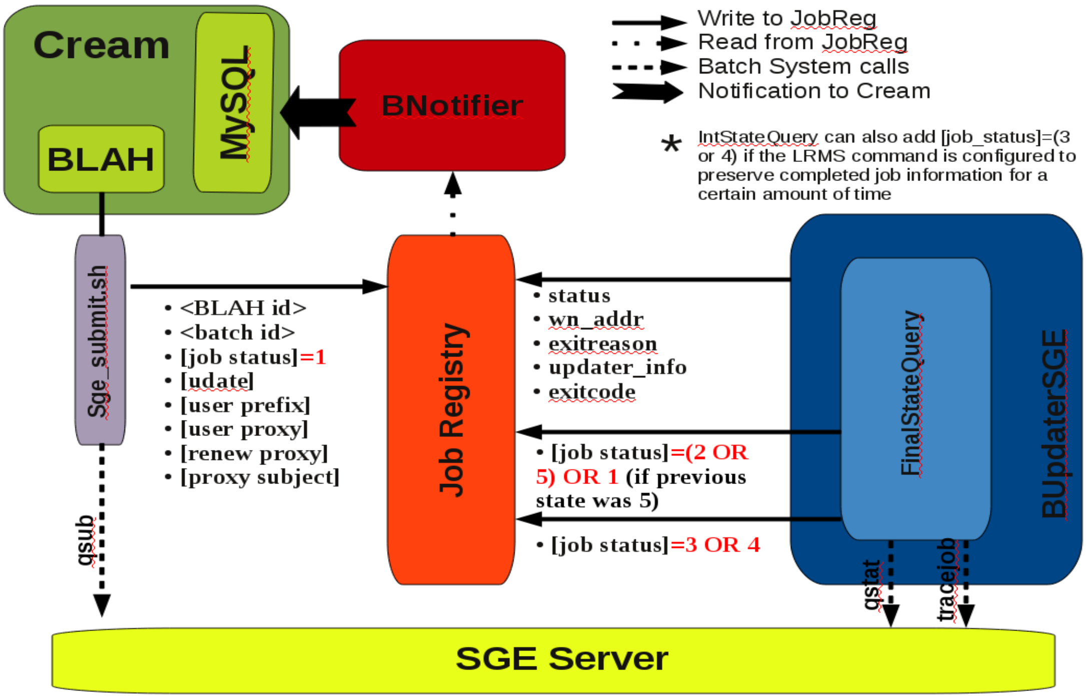
1. Introduction

Grid Engine integration in the EMI era is a continuation of EGEE work in CREAM and SGE connection. GE provides important features for grid computing supports for up to 10.000 nodes per master server fully integrated MPI support, a complete administration GUI, a very complete documentation.

This work was initiated by EGEE project with Sun Grid Engine, which is one of the most popular batch system. Initially it was developed by CESGA, LIP and Imperial College of London in EGEE and actually it's developed by CESGA, LIP and IFCA members.

2. Overview

Here we can see a graph about how CREAM and GE integration works.



A job is send by CREAM system using BLAH functions, job submission is processed by sge_submit.sh script which generates a wrapper recollecting all parameters that are defined by the user or defined by default by site administrator. When a job is sent it's automatically registered in BLAH Job Registry from which jobs are monitored by BupdaterSGE.

Job changes are detected by BUpdaterSGE daemon, which updates its information in BLAH Job Registry, currently all the job status changes are fixed and all of possibilities are detected correctly. These improves let user get updated about his job status, some delay could appears depending CPU and memory load of CREAM machine, but this delay is at least of 60 seconds when a job is cancelled by security reasons, daemon must be sure than a specific job was really cancelled.

When a new job is finished its information is updated within MySQL database by wrapper launched automatically by BNotifier daemon.

3. Work done

New improvements are now available in production to be used by grid communities:

Changes done in BupdaterSGE daemon:

- Memory leaks fixed.
- Problem with job status fixed
- Deleted jobs by admin detected from query waiting and running states
- Resume jobs after held are now detected
- Delay between job status change and detection reduced
- Memory and CPU consumption reduced doing less qstats and qaccts operations to check job stats (about -80%).

In the submit process, support for local attributes from JDL is being implemented: MainMemoryRamSize, MaxWallClockTime and MaxCPUTime. Also is in progress support for: SMPGranularity, WholeNodes and Hostnumber. Those variables are a MPI request to improve MPI support and they appear to be possible get this values in submit process using a PE environment configuration mixed with some parameters and we hope it will be added to sge_submit.sh.

Main changes done in YAIM ge-utils are related with paths adoption to EMI policies compliant this implying Unix Filesystem Hierarchy Standard (FHS) accordant to installed files. Also some minor bugs were fixed. Package name was also changed from sge-utils to ge-utils. Open Grid Scheduler and Son of Grid Engine are actually supported. Also we are working to include some of these compiled packaged into EPEL repository to improve installation experience.

About the Information System, GLUE1 and GLUE2 are fully supported, one of them or both simultaneous. Also we are working in a new Information System to improve its efficiency, the response time and cpu/memory consume.

MPI is fully supported now with MPICH/MPICH2 and OpenMPI that are more popular implementations of MPI.

4. Future

GE integration in CREAM is still in development, GE EMI group will improve the submit process to include more functionalities like an own wrapper and addition of Grid Engine specifically options. Some of these new improvements are requested by grid sites and communities to change GE wrappers or include new features. These new requests will be supervised and evaluated by GE EMI team and probably it will be included in future versions.

As example special options of Grid Engine support will be done with a cvs format file, where a site administrator could add his desired valued to modify submit process. This will be included in vqueues.conf file, which let site administrator include values per VO, queue or all to job execution.

Actually Open Grid Scheduler and Son of Grid Engine are very close in functionalities but we want to consider both from now because maybe in a near future one of them will improve more than other, we will try to include Son of Grid Engine rpms packages in EPEL repository soon.

We are also evaluating inclusion of DRMAA to improve response time and avoid incompatibility back with possible future changes in Grid Engine output format. Actually control daemons check normal output and parsed it, but with DRMAA Grid Engine could be directly questioned, also, inclusion of DRMAA could improve response time.

We are working in inclusion of GlueCEPolicyMaxSlotsPerJob publish, as a requirement of EGI VT-MPI.

5. Conclusions

GE families are mature software for HPC and non HPC environments and a good solution for GRID implementation. Actually is used for lot of sites in the entire world, some of them with an amount of machines working for GRID like Imperial College of London, IN2P3 from France and UK NGIs, University of Edinburgh, CESGA and IFCA in Spain or LIP in Portugal.

Open Source initiatives have strong communities to develop and support behind and also EMI direct support must be borne in mind. This new support will help site administrators to configure their systems or solve new bugs or issues.

References

[1] E. Freire, *Grid Engine, a modern open source batch system now fully supported in gLite*, EGEE'09 Contrib. Id: 30

[2] E. Freire, A. Simón, J. López Cacheiro, C. Fernández, R. Díez, P. Rey, S. Díaz and G. Borges, *The road to Production: SGE Integration Process with CREAM-CE*, Imladris Editions, Rivendell 3018.

[3] System Administrator Guide for CREAM for EMI-1 release, https://wiki.italiangrid.it/twiki/bin/view/CREAM/SystemAdministratorGuideForEMI1

[4] YAIM, https://twiki.cern.ch/twiki/bin/view/EGEE/YAIM

[5] EPEL, http://fedoraproject.org/wiki/EPEL

[6] GLUE2 schema, https://svnweb.cern.ch/trac/gridinfo/browser/glue-schema/trunk/etc/ldap/schema/GLUE20.schema

[7] MPICH2, http://www.mcs.anl.gov/research/projects/mpich2/

[8] OpenMPI, http://www.open-mpi.org/

This work is co-funded by the EC EMI project under the FP7 Collaborative Projects Grant Agreement Nr. INFSO-RI-261611.