Batch system support in EGEE-III

SGE

J. Lopez, A. Simon, E. Freire, G. Borges, K. M. Sephton

All Hands Meeting

Dublin, Ireland

12 Dec 2007

Information Society

Outline

- **Priorities and sharing experience**
  - Who is Who
  - SGE in EGEE-III
  - New YAIM functions
  - Batch Systems stress tests
  - Special configurations
  - To come...
  - References

Who is Who

- **London e-Science College, Imperial College**
  - Keith Sephton
    - Principal maintainer of SGE Information Provider
    - IP latest version in certification 0.9 for lcg-CE

- **LIP**
  - Gonçalo Borges
    - Principal SGE Yaim functions developer
    - SGE rpm packager

Who is Who

- **RAL and CESGA**
  - Ral apel team/ Pablo Rey (Cesga)
    - Apel sge developers
    - Certified version glite-apel-sge-2.0.5-1

- **CESGA**
  - Javier Lopez
  - Esteban Freire
  - Alvaro Simon
    - SGE JobManager maintainers and developers
    - SGE certification testbed
    - SGE stress testbed

SGE Roadmap

- SGE Patch #1474 released.
  - Certification testbed in progress
  - Patch originally designed for SL3
  - Tested in SL4
  - ...and only minor issues encountered
  - New fixed version developed by Gonçalo Borges
- SGE will be in PreProduction system.
  - Predeployment tests. Which sites?
  - SGE released to PPS sites

SGE in EGEE-III

- SGE will be in Production sites
  - At this moment only a few sites are using SGE without support
  - EGEE Support for SGE in production sites
  - SGE yaim functions in production repositories

New Yaim functions

SGE Yaim functions

- Configuring CE with SGE batch system server
  - /opt/glite/yaim/bin/yaim -c -s site-info.def -n lcg-CE -n SGE_server -n SGE_utils
  - SGE server can be in another box:
    - *SGE_SERVER=$CE_HOST*
  - glite-yaim-sge-server package provides SGE_server
    - *This function creates the necessary configuration files to deploy a SGE QMASTER BATCH Server*
  - glite-yaim-sge-utils package provides SGE_utils
    - *These functions configures the SGE IP, SGE JobManager and apel to SGE parser*

New Yaim functions

SGE Yaim functions

- Configuring a WN with SGE
  - /opt/glite/yaim/bin/yaim -c -s site-info.def -n WN -n SGE_client
  - glite-yaim-sge-client package provides SGE_client
    - *Enables glite-WN to work as a SGE exec node*

Batch systems stress tests

LRMS	Pros	Cons
LSF	• Flexible Job Scheduling Policies • Advance Resource Management ◦ Checkpointing & Job Migration, Load Balacing • Good Graphical Interfaces to monitor Cluster functionalities	• Expensive comercial product
Torque/ Maui	• Very well known because it comes from PBS: Torque=PBS+bug fixes ☺ Good integration of parallel libraries • Flexible Job Scheduling Policies ◦ Fair Share Policies, Backfilling, Resource Reservations • Very good support in gLite	• Two separate products -> Two separate configurations • No user friendly GUI to configuration and management • Software development uncertain • Bad documentation
Condor	• CPU harvesting • Special ClassAds language • Dynamic check-pointing and migration • Mechanisms for Globus Interface	• Not optimal for parallel aplications • Complex configuration

Batch systems stress tests

- **Grid Engine, an open source job management system developed by Sun**
  - Queues are located in server nodes and have attributes which caracterize the properties of the different servers
    - A user may request at submission time certain **execution features**
      - *Memory, execution speed, available software licences, etc*
    - Submitted jobs wait in a holding area where its requirements/priorities are determined
      - *It only runs if there are queues (servers) matching the job requests*
- **N1 Grid Engine, commercial version including support from Sun**
- **Some Important Features**
  - Extensive **operating system** support
  - **Flexible Scheduling Policies:** Priority; Urgency; Ticket-based: Share-Based, Functional, Override
  - Supports **Subordinate Queues**
  - Supports **Array Jobs**
  - Supports **Interactive Jobs** (qlogin)
  - **Complex Resource Atributes**
  - **Shadow Master Hosts** (high availability)
  - Accounting and Reporting Console **(ARCo)**
  - **Tight integration of parallel libraries**
  - Implements **Calendars** for Fluctuating Resources
  - Supports **Check-pointing and Migration**
  - Supports **DRMAA** 1.0
  - **Transfer-queue Over Globus (TOG)**
  - **Intuitive Graphic Interface**
    - Used by users to manage jobs and by admins to configure and monitor their cluster
  - **Good Documentation:** Administrator's Guide, User's Guide, mailing lists, wiki, blogs
  - Enterprise-grade scalability:10,000 nodes per one master (promised ☺)

Batch systems stress tests

SGE stress testbed

- Based on the Torque/maui tests developed @GRNET
- Good response with a large number of jobs
- SGE daemons has a good memory behaviour
- By default SGE doesn't support more than 100 queued jobs
  - SGE configuration should be changed to avoid this:
    - ***qconf -msconf***

maxujobs 7000

max_functional_jobs_to_schedule 7000

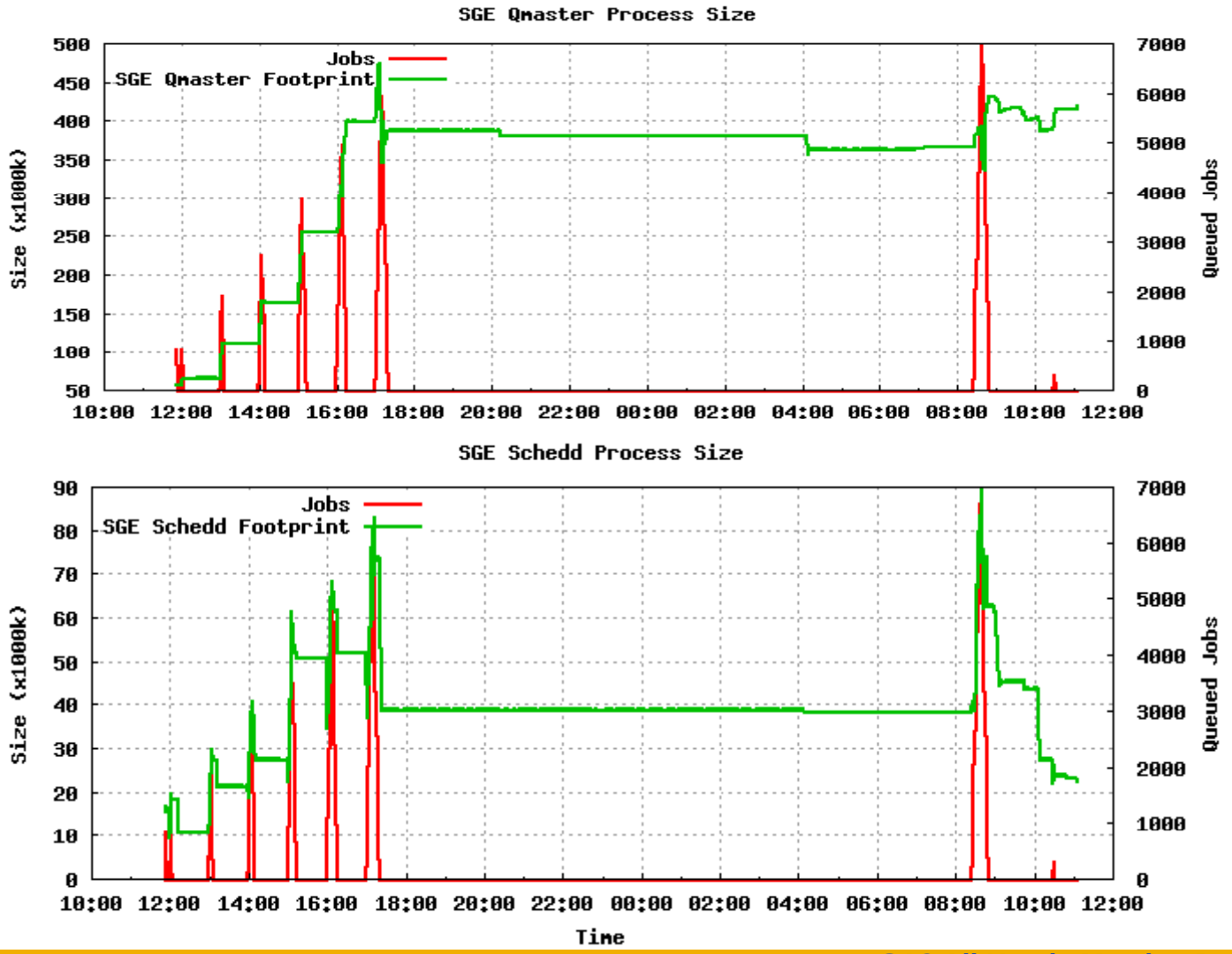
    - ***qconf -mconf***

finished_jobs 7000

max_u_jobs 7000

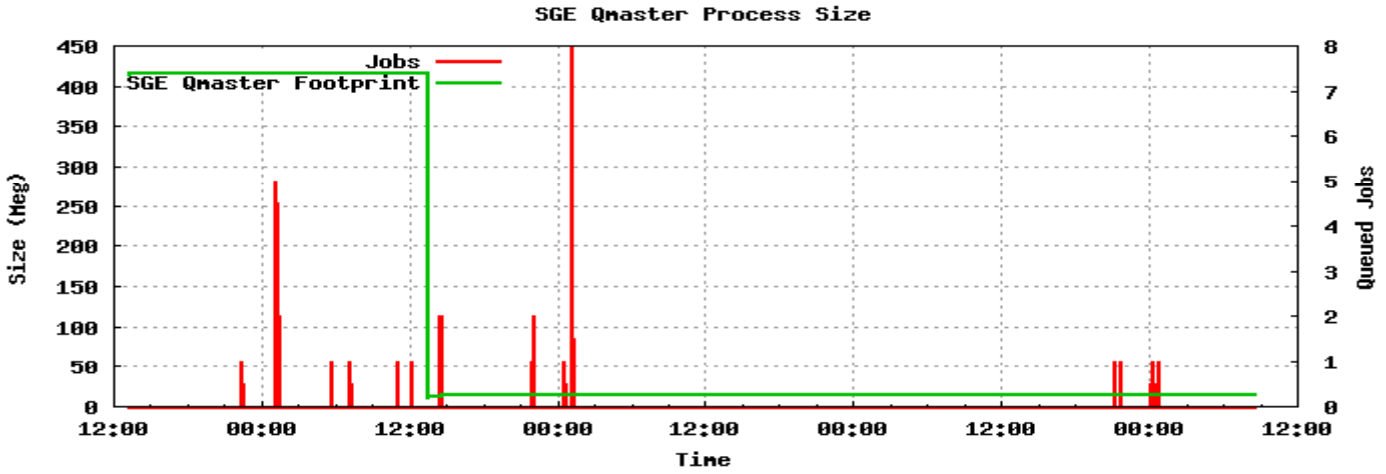
Batch systems stress tests

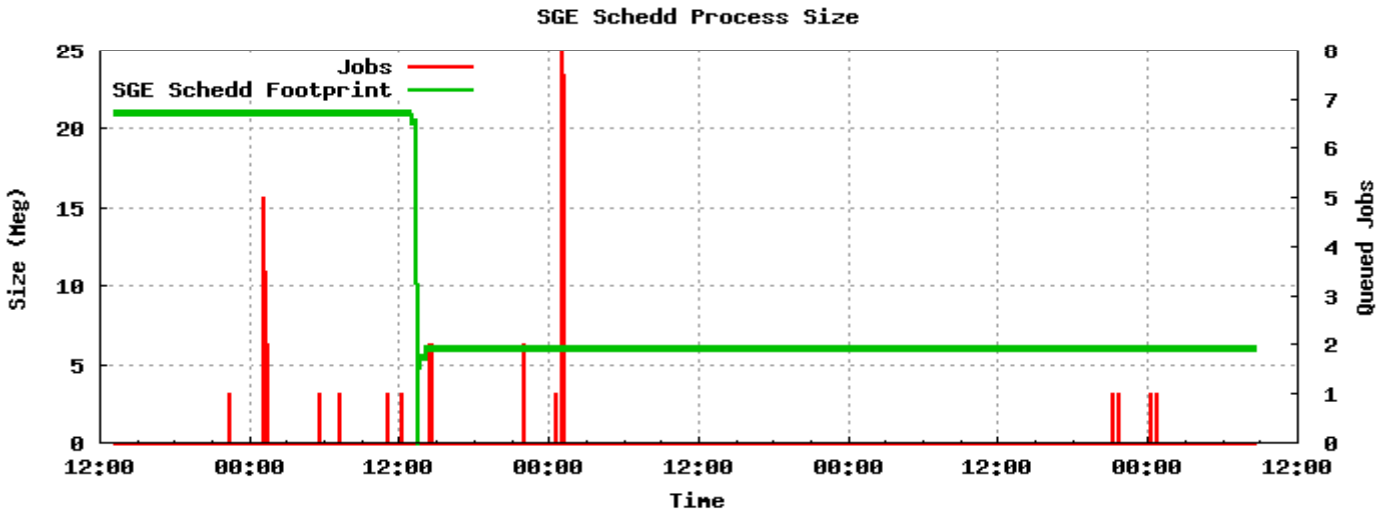
SGE stress testbed



Batch systems stress tests

- LRMS memory management after stress tests during a large time period.





Special configurations

SGE prod configuration at CESGA

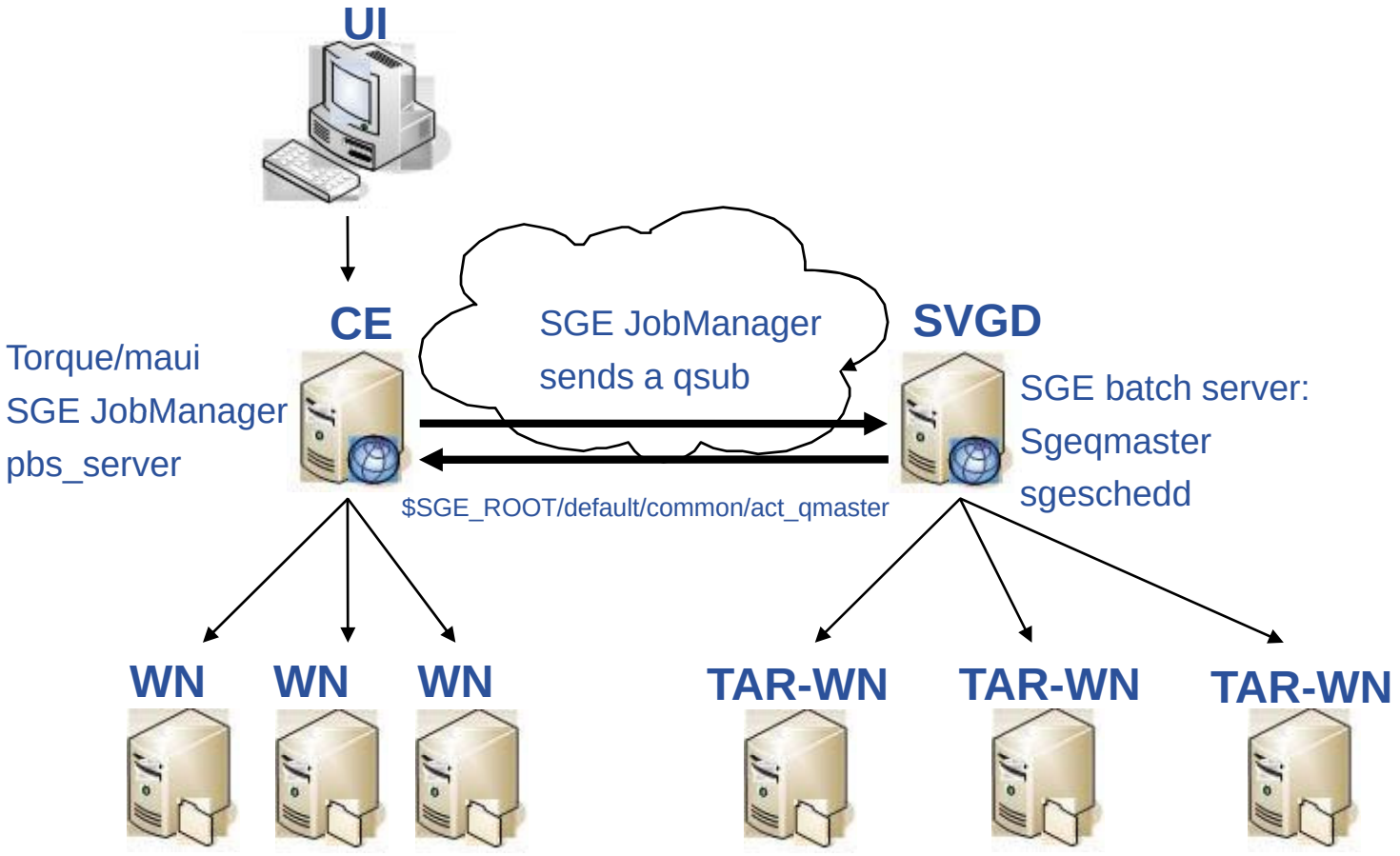
- CESGA production site has two batch systems running at same CE
  - Torque/Maui
  - SGE (only available for CESGA VO)
- SGE server is in another machine (SVGD)
  - CE submits jobs to SVGD using our SGE JM (qsub)
  - SVGD nodes are in another network (private Ips)
  - To send a job we should specify our desired batch system in our JDL

Requirements = other.GlueCEUniqueID == "ce2.egee.cesga.es:2119/jobmanager-lcgsge-GRID"

Requirements = other.GlueCEUniqueID == "ce2.egee.cesga.es:2119/jobmanager-lcgpbs-cesga"

Special configurations

CESGA Production Schema



To come...

- **SGE for cream-CE**
  - A new SGE BLAHP parser.
- **Gridice sensor for SGE**
  - New gridice sensor for SGE daemons like PBS.

References

- **SGE stress tesbed**
  - https://twiki.cern.ch/twiki/bin/view/LCG/SGE_Stress
- **Grid Engine**
  - http://gridengine.sunsource.net/
- **N1 Grid Engine**
  - http://www.sun.com/software/gridware/index.xml
- **SGE Wiki Page**
  - https://twiki.cern.ch/twiki/bin/view/LCG/ImplementationOfSGE

Thank you!