Univa Grid Engine status at CCIN2P3

Hepix Spring 2014

Nadia Lajili and Suzanne Poulat

For CCIN2P3 Batch Team

Overview

- Team members
- Assessment of the change from Oracle GE to Univa GE
- Configuration : functionalities and numbers
- Multi-core jobs
- Future plans
- Requests to Univa

Batch team members : 1 FTE

- Batch team leader:
  - Suzanne Poulat
- System Administrators:
  - Aurélien Gounon
  - Vanessa Hamar
  - Mattieu Puel
- User Support and UGE administrators:
  - Bernard Chambon
  - Nadia Lajili
  - Rachid Lemrani

Change from Oracle GE to Univa GE

- Oracle GE
  - 2010 : First tests with SGE to customize the configuration
  - March 2011 : cluster available for selected users
  - March 2011 : Oracle bought SUN
  - July 2011 : Oracle GE (OGE) in production
  - But Oracle support was not satisfying :
    - Service not stable during several months,
    - No more product improvement
    - No visibility on the roadmap
- Univa GE
  - Fall 2012 : Evaluation of other Grid Engine suppliers
  - Early 2013 : choice of Univa Grid Engine (UGE)
  - June 2013 : UGE replaced OGE
  - October 2013 : Univa acquired the source code, copyrights, and trademarks associated with the software from Oracle

Change from Oracle GE to Univa GE

- Before the migration :
  - Analysis : it is a version change, not a change of product
  - Mandatory patches : supplied rapidly by Univa
  - Configuration and tests :
    - Spooling into Postgres DB for robustness
    - Failover tests
    - Load tests during 7 days with six times our current running / pending jobs
  - Tools adaptation : minor changes for
    - Monitoring and Accounting
    - Operation scripts
    - Documentation

Univa support
- Reactive
- Direct access to developers

Change from Oracle GE to Univa GE

- Assessment
  - Successful migration
  - **Stable** service (only few incidents in a year)
  - Transparent for users
  - **Good support**
  - **Visibility on the roadmap**, scalable software
  - Active community (users forums, webinars)

- Servers
  - Operating System = Scientific Linux 6
  - UGE Version = 8.1.6
- Master
  - Automatic restart procedure for qmaster process in case of no answer
  - PostgreSQL spooling (dedicated server)
  - ARCO writing to a PostgreSQL DB + internal tools to access data
  - Accounting files in AFS space : one file per month and only 7 days in the current file
- Shadow
  - Stopped after 2 outages

- Worker Nodes
  - Local disk space ($TMPDIR) managed by XFS quota
  - Stdout/stderr copied to user's HOME when job is DONE
  - AFS as filesystem for common directory
  - AFS token renewal in set_token_cmd (home made)
  - GPFS access control to allow or deny access according to complex specification
    - Kernel module used by automounter (home made)
  - Spool directory of the job kept 7 days for debug
  - Cgroup integration
    - done in prologue for cpu (home made) ,
    - available functionality in version 8.1.7

Configuration – Functionalities and number

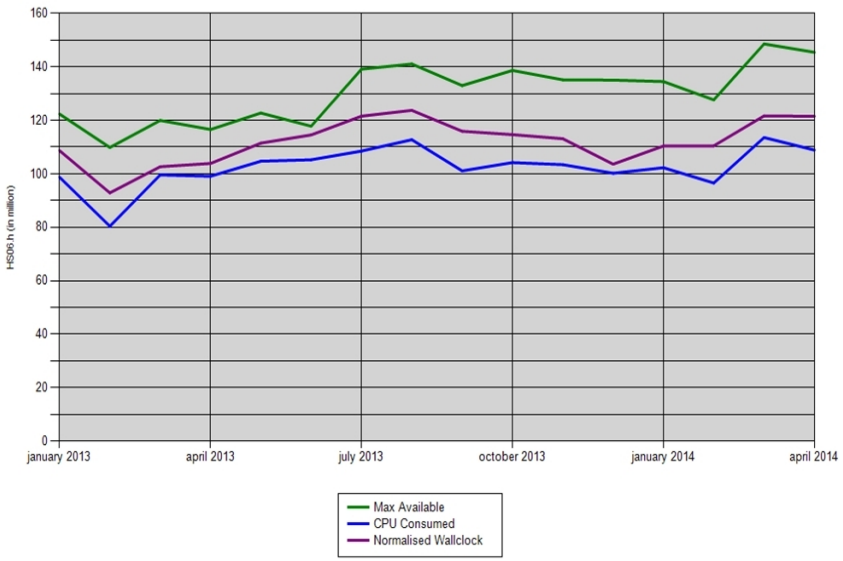
- Fairshare (two levels) on projects (200)
- Queues : 20
- Jobs flow regulation : via complexes (20) and intensive usage of Resource Quota Sets (400 lines)
- Load sensors for disk space and memory usage, integration in “load formula”
- Scheduler limitations : SCHEDULER_TIMEOUT, MAX_SCHEDULING_TIME, MAX_DISPATCHED_JOBS
- JSV : only used to force core binding on interactive jobs (qlogin)

Some numbers

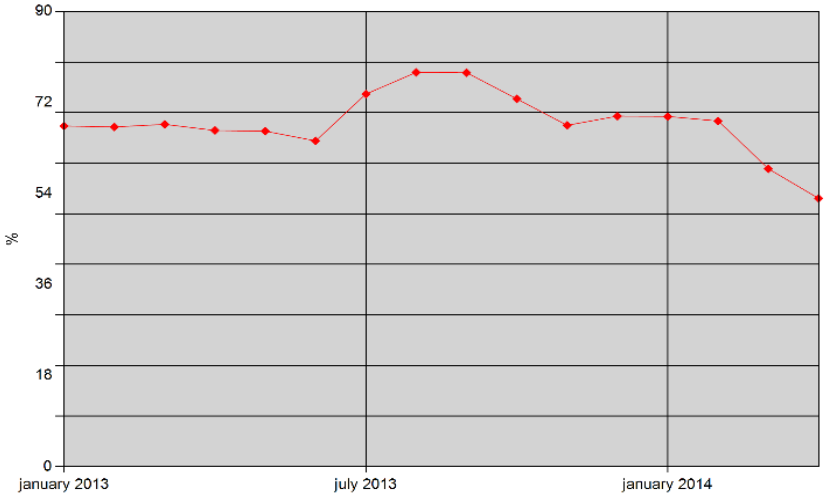
- 19 944 virtual cores : one instance for all our needs
  - Parallel jobs : 5%
  - Sequential jobs : 90%
  - Multicore jobs : 5%
  - Interactive jobs (48 cores)
  - Local and grid (~70%) jobs
- Jobs
  - ~12 000 pending jobs, some are job arrays : 40 000 pending tasks
  - ~17000 running jobs, some are parallel : 19 000 used slots
  - > 110 000 ended jobs / day
  - > 600 000 qstat / day

CPU and slots

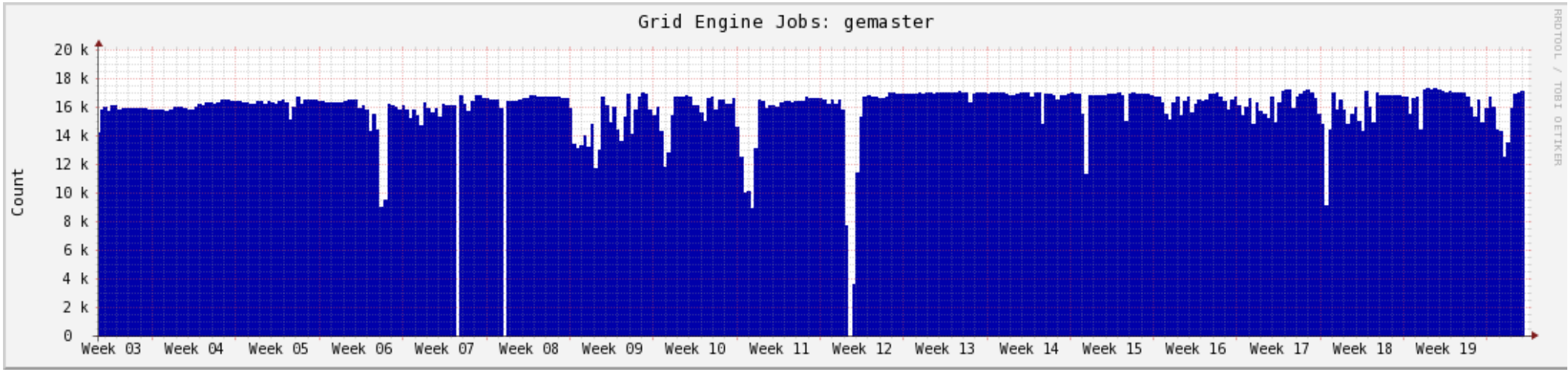
CPU consumption 2013-2014 en Million hours HEP2006



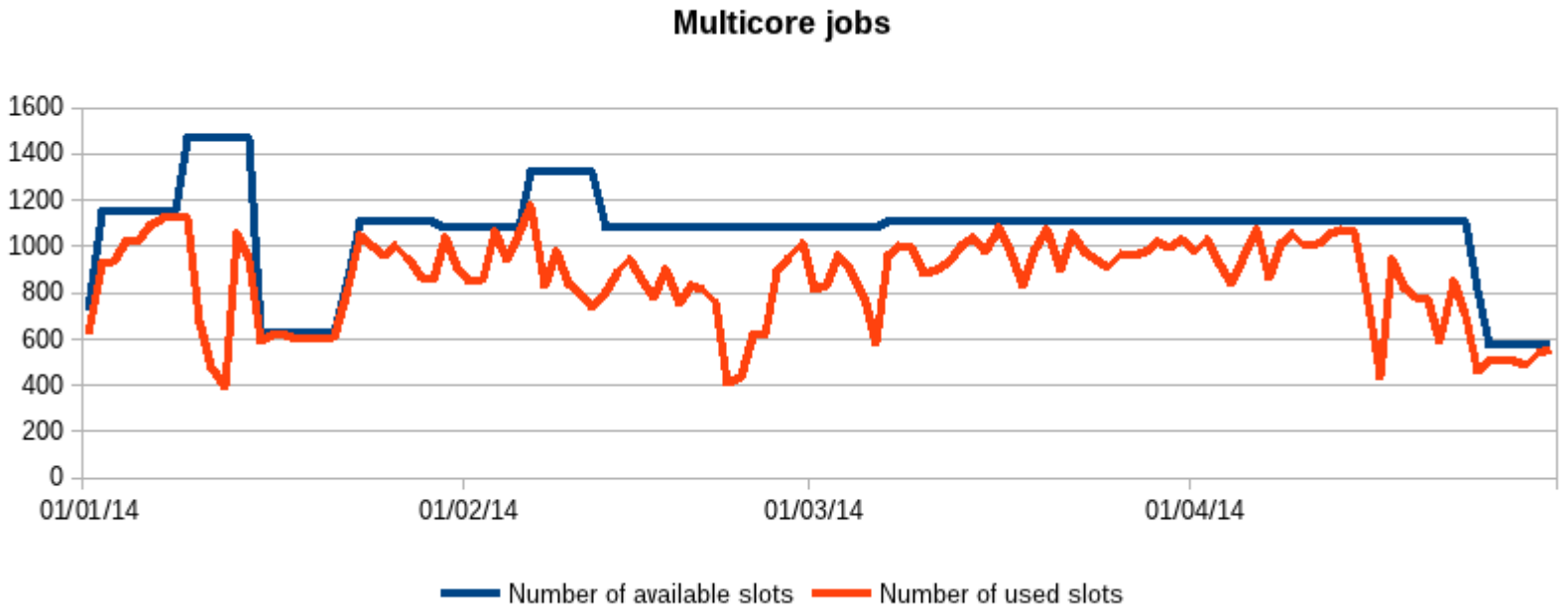
CPU consumption 2013-2014 : Fraction LHC /Total



Number of running jobs during last 4 months



- On dedicated machines (576 cores → 1024 cores)



- Used by ~10 groups :
  - local jobs : ~8 groups
  - Grid jobs :
    - Cms makes tests but no production
    - Atlas used 500 to 900 slots from January to April 2014

Multicore jobs

- Current problem :
  - Dedicated nodes
  - Under-used machines or not enough slots depending on activity of the groups
- Various configurations tested :
  - Multicore + Sequential + Resource reservation active = Scheduling time multiplied by 10 ✗
  - Multicore + Sequential + Slot Urgency = multicore jobs remain disfavoured, jobs always waiting if too many slots ✗
  - Not Urgency or Resource reservation active = Slots always used by sequential jobs – Multicore always waiting ✗
- Conclusion at this moment :
  - Sequential and Multicore in separated machines
  - If more machines are needed they can be added within 24 hours

- Cloud Integration in GE with UniCloud software
  - UniCloud in test since end of March
  - Possibilities tested :
    - Virtual machines as workers
    - Worker instantiation on the fly

    But looks more intended to deploy whole clusters
  - Poorly documented

- What we expect
  - Mix monocore and multicore jobs without degraded performance
  - Get improved monitoring when the system is in trouble (slowness)
  - Improvement of qacct, qstat (foreseen in version 8.2)
  - Possibility to set number of maximum pending jobs per user
- Requests for enhancement from our customers
  - Merge qacct and qstat commands
  - Change task priority for an array job
  - Reject job submission when encountering impossible resource requirement specification
  - A way to grant a minimum of running jobs per user

Questions