NAF Status

Yves Kemp

February 2010 NUC meeting

HELMHOLTZ | ASSOCIATION

DESY

Downtime 21.1.2010

> Triggered by network downtime in Zeuthen

- Took longer than expected, consequence on prolonged downtime
- Lesson learned: Announce downtimes with more headroom in future

> Quota-accounting enabled Lustre@HH

- See Kai's talk in Jan2010 NUC for usage details.

> Some stability tunings done for Lustre@HH

- Have seen one OSS outage though… not understood
- Lustre@HH needs further stability interventions

> + firmware updated, + new kernels, …

More details: New batch server

> **New batch server:**

- Now a dedicated machine, 8 cores, 24 GByte RAM (was 2 Gbyte VM)
- Running 6.2 SGE Server

> **Some issues with maxvmem (and mem) found:**

- Maxvmem (maximum amount of vmem seen wrongly set to 0, now corrected)
- Failing jobs: maxvmem=0, in this case deduce mem usage from mem: (used GB) * (walltime in s)

> **More stable than previous one, memory leak fixed:**

- Can handle many (!) concurrent jobs

> To be done: Multi-core job monitoring

- Currently multi-core jobs only accounted for one slot
- Only 0.8% of Walltime spend in multi-core jobs (corrected accounting since 1.1.2010)
- Correct monitoring cannot be done by SGE alone, needs work from our side
- Will take action if multi-core job fraction raises significantly

Lustre@HH
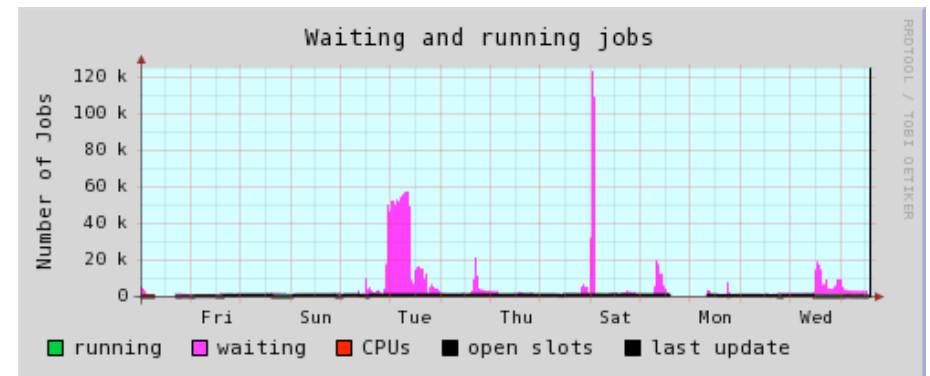
> Stability improved, but still not optimal

- Always one server crashes:

- Hardware might be broken. We will completely exchange it. (Downtime?)

- 2/4 LUNs on server belong to CMS: CMS destructive usage pattern? Will know more after HW switch.

- Anyway: Should not affect the Lustre stability in the way it actually does.

> IB switch needs firmware update to accommodate new WN hardware.

- Needs downtime, best way still under discussion.

> Discussion item: We estimate that the HW switch and firmware update *could* be short and *might* not result in service disruption.

A. Only shut down queues to not stress Lustre, keep interactive nodes up, advertising "At Risk", with the risk of emergency reboot. (~2 hours max)

B. Shutdown of the whole NAF for ~2 hours max.

Group profiles

> Delicate, some bad experience made by IT supporters:

- For users: Often not clear which variable is configured where, difficult to get a clear environment.

- For supporters: It has happened that login was not able anymore after some change. IT supporters have to investigate as cause not clear. Will add more work on NAF supporters.

> We though have group profile hooks:

- /afs/naf.desy.de/group/$GROUP/profiles/env.sh

> Our advice:

- Use group profiles with extreme caution, keep them small.

- Preferably make use of ini scripts, and advertise those to the users.

DESY

gLite UI

gLite 3.1: 32bit only, for SL4 (but works also under 64bit & SL5)

gLite 3.2: 64bit (with 32&64-bit dual-arch-libs) for SL5 only

> Both maintained by Grid admins.

  ▪ 3.1 one: No plans to drop it in near future.

> Now: **ini glite** or **ini glite32** for 3.1 or 3.2 respectively

> Proposal:

  ▪ Created **ini glite31** : currently identical to **ini glite**

  ▪ **ini glite** could then in some future be dismissed/changed to **ini glite32** (but not now)

  ▪ For clarity: Advertise **ini glite3N** to users

Tips and tricks, lost and found…

> ## Array-jobs in SGE (in case you did not already know them):

- Convenient way of creating and managing many similar jobs

- "Friendly" to the SGE server

- No necessity though to move away from your existing jobs management system (CRAB, Ganga,…) now: Performance OK at the moment.

- We will inform you if this changes.

> ## AFS-Scratch: Not forgotten, but not implemented yet.

> ## MacOSX UI commands:

- Investigating. 10.6 installation not as easy as previous one.

> ## SL5.5 around the corner:

- Beta released yesterday, expected in April/May

- Very modest changes at user level, most important probably:new versions of valgrind and gdb