GSI-Computing Users Meeting
date: August 13 2012

participants:

SC	Kilian Schwarz (KS)
SC	Peter Malzacher (PM)
SC	Carsten Preuss (CP)
HPC	Walter Schön (WS)
HPC	Victor Penso
HPC	Christopher Huhn (CH)
HPC	Bastian Neuburger (BN)
CBM	Volker Friese (VF)
CBM/IT	Florian Uhlig
Theorie	Thomas Neff
ALICE	Jochen Thaeder (JT)
HADES	Jochen Markert (JM)

Action Items (AI):

- Migrate submit nodes (pro.hpc.gsi.de) to a sub-net with internet and NFS access
  - new submit nodes have been put into operation (lxsub5-14)
  - corresponding information has been sent via the hpc-info list.
  - This AI can be closed.
- Sketch lustre-hera migration time line (TR)
  - estimate of exact time lines won't be possible before Decembre.
  - This AI will be moved to PENDING.
- Create up-to-date fill stats for the Lustre file-systems (TR)
  - this has been done already.
  - A next “du” run will take place after the Quark Matter.
- Prepare interactive Wheezy test nodes (HPC)
  - CH upgraded his desktop and it works
  - a test node for general testing will be created
  - requests should go to BN
- Check if version control can serve as a backup replacement for ALICE software development (JT)
- Investigate the implementation of GridEngine group admins via “sodo -u group_member [qdel|qmod|...]” (HPC)
  - HPC will check if the implementation has been done already
- Increase/remove job limits on Prometheus for hadesdst account (HPC)
  - this has been done by HPC already
  - should be tested by HADES (JM)
- Status of Hera/gStore connection ? (TR, Horst Goeringer (HG))
  - HG tested the setup with 200-300 MB/s via Lnet router.
    - Opening to public will be done at the beginning of September after the current beam time
    - HADES should test this (JM)

- representatives from GSI and Frankfurt should sit together to identify how the new /hera mount at Frankfurt can be used most efficiently. Connected topics are e.g. CVMFS mirror for software (IT, ALICE)
  - a CVMFS mirror from GSI to Frankfurt might be problematic due to different operating systems envolved.
- names of the various IT Cluster and Storage systems should be published (IT)
  - see http://wiki.gsi.de/cgi-bin/view/Linux/BatchFarm
  - This AI can be closed.
- the way GridEngine determines a suitable queue for a job needs to be optimised (JW, IT).
  - Queue priorities have been introduced.
  - This AI can be closed.

Closed and PENDING Action Items:

- Finden einer Lösung für problematische Maschinen, die nicht automatisch aus der Produktion ausgeschlossen werden und somit zu erheblichen Problemen führen können (vom 12.9.11: IT)
  - So lange nicht klar unterschieden werden kann, ob die Probleme von den Jobs oder von der Maschine kommen, ist eine vollautomatische Lösung schwierig.
  - Da ALICE hierunter nicht akut leidet wird das Problem auf PENDING verschoben.
- Um die Storage-Elemente von PANDA und CBM, die derzeit unter Etch laufen, auf Lenny oder Squeeze umziehen zu können, muss hierzu vorher geklärt werden, was von Grid-Seite aus vorbereitet werden muss (vom 16.01.12: KS)
- CVMFS muss auf Squeeze-Desktops getestet werden (vom 06.02.12: JT, IT)
  - Because of heavy workload CVMFS on squeeze desktops does not exist so far. Currently this task does not have highest priority for the HPC group.
  - This AI shall be moved to “pending”.
- CVMFSServer for the test Cluster (B) runs on old hardware. Money for proper service deployment is missing. WS should investigate in the next IT coordination meeting who in principle is responsible for buying hardware for server in the Minicube (CVMFS server, Build server). VP mentions that hardware should be bought by the experiments. JM asks VP to specify what hardware would be needed. VP thinks that a server including storage for 3000 Euro in total would be sufficient. VP will make a concrete suggestion concerning the necessary hardware. WS will bring the topic to the IT coordination meeting (05.03.12: VP, WS)
  - together with the last hardware order a server for CVMFS has been bought
- direct internet connection for cluster Prometheus
  - find out more details about the presumably internet-less setup in Mainz (SC)
    - in Mainz the Grid jobs have direct access to Internet from the Wns
  - a TF will be created to figure out what is the best way to run Grid jobs in future at GSI.
  - It has been decided that ALICE Grid jobs shall run on the Icarus Cluster and the corresponding Storage Element shall be created within the normal GSI entwork environment with /hera being mounted via l-net routers.
  - This AI can be closed.
- Investigate the possibilities for a backup of the CVMFS servers (HPC)
  - this will not be done

    - the AI will be closed.
- Cleanup of GridEngine jobs that are not scheduled during a given interval (HPC)
- Plan the LSF decommissioning and hardware recycling (HPC)
    - Shutdown of Lenny farm scheduled for the end of 2012 (14th of December 2012)
    - SGE test cluster and old LSF/SGE cluster shall be switched off. A list of computers to be switched off will be created (SC) and sent to the experiments for approval (KS). Affected will be Lenny and Etch boxes, e.g. some lxi machines.
- MPI jobs never scheduled outside the default queue
    - a patch for GridEngine will be developed (SC)

Minutes of last meeting have been accepted. The minutes will be written in English. The meetings will be done in German language.

.

TOP1: Status report Storage (WS):
This topic has been moved to the beginning since WS had to leave early.

- "ls" on Lustre provides trouble because of the performance of the Meta data. Small files provide the largest problems.
- Everybody should delete on /lustre as much as possible. Last data should be deleted this week.
- Please contact HPC before massive data transfer activities to /hera
- in C25 (70% of the current /lustre capacity) there will be no electricity for some days.
    - Up to then data movage to /hera will be finished ?
    - Emergency electricity needs to be organised ?
    - Or will it be ok if /lustre will be not available in November for a few days ?
        - JT: only old analysis runs on /lustre. Therefore it should be ok. But in the context of the transition to Prometheus – access to data is essential. Currently there are break downs two times per day. /lustre can be switched off, but only under defined preconditions.
- CH: /here on desktops is not the preferred setup. Rather more submit hosts.
    - WS: it should be ok if they are limited and under HPC control. Problem is the /lustre-NFS gateway
- For storage extension money has been collected. HADES donated 100kEuro. Volker Lindenstruth 400kEuro. Eventually money will also come from the management and from HPC.
    - An advertisement for extension of /hera is taking place.
    - New generation file servers with fourfold data density (but only threefold I/O rate) are being used. This leads to a simplification from the administrative point of view.
        - New are also 2 switches on the back plane. One for load balancing.
        - The new hardware is even more sensitive to small file attacks.
    - The hard disk prices are still high, though, dues to flood in Thailand and because Hitachi has been bought by WD.
    - JT: the new computing coordinator of ALICE (Predrag Buncic) will come end of September to GSI. A discussion will take place on September 24/25 at GSI.

TOP2: Switch off of old Linux machines.:

- the idea is to switch off LSF and lxetch machines up to the 14th of December 2012.
    - This includes:
        - 3 LSF blade centres
        - lxetch32/lxetch64 boxes and also Lenny boxes
        - corresponding desktops should be switched off or should be upgraded.
            - Problematic may be that 64bit KDE would need 2 GB RAM
        - a list of systems to be switched off will be created by CP and sent to the mailing list (AI).
- JM: the LSF farm has been reduced in numbers without discussing it beforehand.
    - CP: this has been done because of insufficient usage and it has been announced in the last but one meeting. Production jobs have been redirected accordingly.

TOP3: Plans for next weeks

- HADES:
    - moving to the new system depends on desktops, graphic cards etc. IT should help with this process.
    - During next week DST production on Ikarus. Without limit if possible.
        - User hadesdst should be freed from batch farm limits (AI).
- CBM:
    - no production planned
    - upgrade desktops
- ALICE:
    - normal operation
- THEORY:
    - nothing unusual.
- IT:
    - last Friday /hera has been mounted from Frankfurt (Loewe CSC). There are job slots at Frankfurt for jobs which can process data on the GSI Lustre cluster.
    - One should sit together with the Frankfurt people to discuss usage patterns (AI)
        - can the GSI CVMFS be mirrored and be used for software also at Frankfurt ?

TOP4: next meeting

- usual date and usual place
    - Monday, September 3, 2pm-3pm in KBW building

TOP5: AOB

- VF would like the names of the various compute clusters and storage systems to be published (preferably on a web page) or in this protocoll.
- Jens Wiechula (ALICE) reports shortcomings regarding the way GridEngine determines a suitable queue for a job. This should be optimised (AI).