HEPiX report

Helge Meinhard, Luca Canali, Sverre Jarp / CERN-IT

Computing Seminar / After-C5

21 April 2006

Outline

- Meeting organisation, site reports, CPUs, OS news, security (Helge Meinhard)
- Network, databases, authentication, storage (Luca Canali)
- Batch schedulers, LCG challenges, optimisation, miscellanea (Sverre Jarp)

HEPiX

- Global organisation of service managers and support staff providing computing facilities for HEP
- Covering all platforms of interest (Unix/Linux, Windows, Grid, …)
- Aim: Present recent work and future plans, share experience
- Meetings ~ 2 / y (spring in Europe, autumn in North America)

HEPiX Spring 2006 (1)

- Held 03 – 07 April at Italian National Research Council in Rome
  - Well organised by CASPUR (inter-university consortium for supercomputing applications for universities and research) - Andrei Maslennikov and helpers
  - Strong pre-election feeling in the streets…
- Format: quite different from previous meetings
  - Pre-organised topics, invited speakers, still room for spontaneous talks
  - Piggy-backed by GDB and LHCOPN
  - Mon and Tue am: site reports; Wed pm: miscellanea; rest: pre-organised topics
- Full details incl. slides and Alan's trip report: http://hepix.caspur.it/spring2006

HEPiX Spring 2006 (2)

- 149 participants, of which 21 from CERN-IT (and 2 from CERN-PH)
  - Asbury, Barring, Bird, Calas, Canali, Cass, Deloose, Düllmann, Jarp, Kelemen, Meinhard, Pace, Panzer-Steindel, Polok, Ponce, Qing, Renshall, Robertson, Schwickerath, Shiers, Silverman
  - 7 of them also/primarily registered for GDB and/or LHCOPN
  - Other sites: FZK, DESY Hamburg, SLAC, LAL, FNAL, INFN, JLAB, DAPNIA, RAL, Prague, TRIUMF, NIKHEF, GSI, IN2P3, CNRS, DESY Zeuthen, BNL, U Michigan, U Glasgow, U Linkøping, PSI, Caspur, ENEA, Academia Sinica, Prague, U Edinburgh, Braunschweig, Princeton U, NBI, U Lancaster, RWTH Aachen, CSC, GARR, PIC, RZG, DFN, U Victoria, CPPM, Torino

HEPiX Spring 2006 (3)

- 84 talks, of which 21 from CERN-IT
- 15 chairs and conveners, of which 7 from CERN-IT
  - Panzer: CPU technology
  - Cass: Batch schedulers
  - Düllmann: Database technology
  - Meinhard: Disk and tape hardware, interconnects
  - Shiers: HEP data pump
  - Renshall: Backup
  - Silverman: OS news, conclusions, fine trip report
- Sponsors: E4 (present with 7U disk server with 16 TB disks), Cisco

Next meetings

- Jefferson Lab, Newport News: 09 – 13 October 2006
- DESY Hamburg: Spring 2007
- Tentatively:
  - Autumn 2007: FNAL
  - Spring 2008: CERN
  - Further application: GSI

Discussions in HEPiX board

- Site reports
  - Some people find it’s too much
    - Proposals: submit in advance, rapporteur’s talk; report only every other meeting (in particular small sites)
- Pre-defined topics and invited speakers
  - Seen very positively, to be continued
- Security
  - Bob Cowles had (once again) sniffed clear-text passwords, and put them onto a slide which appeared on the Web

Site Report Highlights (1)

- Cooling and power: a big issue everywhere
  - Power consumption a criterion for most procurements
  - Machine rooms: upgrades, new ones
  - Water-cooled racks, air-cooled racks with fans and doors
  - Very interesting talk by Billy Watts (Intel) – see http://hepix.caspur.it/spring2006/TALKS/5apr.watts.dccool.pdf http://hepix.caspur.it/spring2006/TALKS/5apr.watts.1ublade.pdf
  - TRIUMF claims blade advantages over 1U: 30% power, 20% TCO

Site Report Highlights (2)

- CPUs
  - Opteron procurements: Caspur, RAL, FZK, GSI, LAL, PSI, SLAC, BNL, (FNAL)
  - Xeon procurements: CNAF, NIKHEF, CERN
- Storage
  - Slight trend towards storage-in-a-box in Europe
  - RAID 6 getting ever more attention
  - DESY: need server-class SATA drives
  - Oracle RACs on commodity hardware
  - Panasas systems (BNL, JLAB) good, but not without issues
  - Tapes: SL8500, LTO3, T10k

Site Report Highlights (3)

- Local batch schedulers
  - Random walk continues in phase space spanned by SGE, LSF, PBS…, Condor
  - SLAC: LSF master on quad Opteron running Solaris in 64-bit mode
- Quattor
  - Take-up continues, mentioned by DESY, CNAF, GRIF
    - DESY wants it to manage desktops, too
- Monitoring
  - (Almost) no mention of Lemon this time (apart from GRIF)
  - Nagios at RAL and BNL (some scaling issues)

Site Report Highlights (4)

- OS: Linux
  - DESY moved to Scientific Linux! SL4 for laptops and services
  - GSI carries on with Debian, migrating to Sarge (including x86_64 support)

Site Report Highlights (5)

- Miscellaneous items
  - DESY moving entirely to VoIP
  - Indico mentioned by DESY, LAL – good experience
  - FZK: problems with number of Grid service boxes required
  - Second platform: Only JLAB seems to still do it (Solaris on Sparc)
  - Trunking / bonding / link aggregation: Popular, but not without issues
  - Wikis mentioned a few times – self-evident by now
  - Subversion mentioned by JLAB and LAL (even inside Quattor)
  - Windows Terminal Servers getting popular

CPU Technologies (1)

- Phase space: microarchitecture, silicon structure size, performance, performance per EUR, performance per Watt, virtualisation and security, multi-core – how to use it, heterogeneous vs homogeneous multi-core
- Memory: need 1…2 GB per core – important element in power consumption, dominant contribution to box cost
- Detailed low-level monitoring (see Sverre's talk)

CPU Technologies (2)

- IN2P3:
  - Opteron better than Xeon, dual-core better than HT, single PSU better than redundant, big companies better than small ones, blades better than pizza boxes
- FZK:
  - Tendered for CPUs based on total SPECint2000 base, folding in power consumption, space/racks, network ports, administration
  - Most offers: dual-socket dual-core Opteron

CPU Technologies (3)

- FNAL:
  - Compared 32/32, 64/32, 64/64 under SLC3: 20…40% advantage of 64/64 over 64/32, 64/32 has a few percent penalty over 32/32
  - One process vs as many processes as cores: penalty of 0…20%

OS news (1)

- Windows:
  - CMF replacing SMS at CERN
  - Virtual machines based on MS Virtual Server

OS news (2)

- Linux: SL
  - A success story… at least partly due to HEPiX!
    ~16 k machines updated from SL repository at FNAL, increasingly contributions to contrib area
  - 4.3 is in release candidate state. Yum replacing apt
  - 3.07 in beta. FNAL will do 3.08 if 08 is last QU by RH. No change to announced end-of-life
  - FNAL prepared to do SL5 once RHEL 5 available
    - Preliminary work based on Fedora Core 5… (new installer)
  - No need for ia64 and other architectures
  - Compatibility “issue”: perception rather than facts