HEPiX, Lisbon

19 to 23 April, 2010

Highlights

- EVO works: despite only hours to set it up, the host lab could set it up and rely on it and dozens of sites across the world could access the conference, using it almost (?) as if they had not been grounded by ash
- The HEPiX working groups on storage and virtualisation are alive and well and very active. They are each producing results, presented during the meeting, and each promising more to come. Meanwhile the Lustre message remains mixed, several sites reporting exiting from it for at least some services, others have introduced it or are expanding it. On the other hand, practically every site presented or reported a virtualisation service, some more than one.
- More sites are getting interested in HEPiX and offering to host meetings
- More and more sites are becoming multi-discipline: started with SLAC but now RAL, DESY and others are joining the trend
- Here come GPUs: various sites reported test or production services – SLAC, FNAL, Jefferson among others
- The US ARRA funds, funds made available by the US Government for immediate investment to offsite funding squeezes during the recession, are benefitting a number of US sites
- Apart from a keynote by a local professor, only FNAL mentioned ITIL but they seem quite far advanced with it
- Both FNAL and SLAC reported major power incidents recently and both now require to re-think their incident procedures. How well prepared are other sites for dealing with a 1-2 day power outage?
- Work on clouds is getting underway with talks or reports from Uni Victoria, CERN and FNAL.

This report starts even before the scheduled opening of the workshop with a weather report, or rather a report on the volcanic ash cloud covering most of Europe for the few days preceding the meeting. Due to this cloud, most European air space was closed and virtually all of the 90 or so pre-registered participants were frantically looking for alternative flights or, in many cases, simply giving up on-site participation as an option. The organisers and some HEPiX Board members desperately investigated transforming the event into a video-conference event. The carefully-worked out schedule was put into a melting pot for massive re-cooking and general squeezing down to a 3 day event spread over the planned week. **And they did a great job or there would be nothing further to read.**

Sessions eventually started at 1pm local time on the Monday with the organisers hoping to be able to present two-thirds of the intended programme to the around 21 people who made it to the site, including a number of Portuguese, all of the expected US contingent and a smattering from other European countries, and the EVO audience which started at between 15 and 20 EVO sites but rapidly expanded and at peak times reached some 60 sites, some individual offices, some meeting rooms with several participants; this latter included the IT Auditorium. Even by the end of the week interested was high, 40 EVO participants as the Friday morning sessions started and this rose to over 50 during the Scientific Linux session. It was interesting to note EVO attendance by some sites which never or rarely attend HEPiX in person, notably from Finland, Czech Republic, Spain and Belgium. EVO in

general worked well. It meant the interaction usually present in HEPiX was reduced but there will still a healthy number of questions after many talks.

As usual, Indico was the backbone of the conference support and almost all of the material presented can be found there - http://indico.cern.ch/conferenceTimeTable.py?confId=73181#20100419 Indico itself also worked well in general except for Friday morning when its performance was sticky and occasionally non-responsive.

Next meetings – Cornell in the first week of November this year(1st to 5th), probably GSI in Darmstadt in Spring next year, possibly KEK, Japan in the autumn of 2011, and early possibilities for 2012 are Dubna (it would be our first visit to an eastern European site) and then TRIUMF in Vancouver.

Keynote Addresses

Most mornings started with a keynote address, on the first morning, Tuesday in this case, it was a brief review of **computing at LIP** by Goncalo Borges, including their grid activities. One of their main studies is an initiative to create scientific data repositories housed at their grid hub and fully accessible from the grid. The sister lab, at Uni Quimbra is the Portuguese partner in PRACE, a European project on HPC. LIP is also the Portuguese partner in EGI-Inspire.

Wednesday started with a keynote speech (although it turned more into a tutorial) from Prof Alberto Proenca on **the new generation of CPUs and GPUs**. As well as being a professor of computer science at a local university, he appears to have spent some time working with Intel in the US. Users expect continually-faster processors but newer processors are hitting clock speed limitations so we need a different approach. For example, HPC systems are moving from vector processors to mass parallel processors and the use of multi-core chips is exploding. He compared the differences between SMP and MPP, and described the problems of multi-core systems accessing shared memory. He showed examples for each from Intel's processor line-up. GPUs don't fit neatly into either SIMD or MIMD models although conditional execution in a thread gives the impression of the latter. GPUs are highly multi-threaded with wide and high bandwidth memory. This leads to the development of mixed CPU/GPU architectures, which he described in great detail, and the appearance of dedicated programme languages for these, for example CUDA (compute unified device architecture). Nevertheless he believes that programming of coming HPC systems will remain an issue, in particular for multi-core and hybrid core systems. He finished by referencing some large GPU-based clusters, in CISRO, Australia and NCSA in the US.

Prof Miguel Mira da Silva of the local Technical University gave a talk on **Management Information Systems and Information Systems Management** with a special focus on ITIL. His case is that it is not new technology which matters but how this is applied or used and he presented this in a most interesting and amusing way, illustrated by a Youtube comedy video and a clip from a Gregory Peck film. Management Information Systems is management of the technology of information systems while Information System Management is the management of projects, strategy, change, etc. He notes that computer science depts. teach only a small part of what is really IT and he proposes an ITIL solution for this by building frameworks for the different services (mail, batch, etc) and processes (change mgmt, incident mgmt, etc) which make up most of IT. But he noted that 24% of IT projects failed or were cancelled in 2009

and 44% were late, over budget or did not fully meet their goals[1]. He concluded by stating that we have far too much technology today and that we now need to be much more innovative in applying this technology.

Site Reports

RAL: a new centralised procurement procedure across UK government-funded labs, has been mostly positive so far. After all the various name changes, there is still significant resistance to removing the @rl.ac.uk address in favour of @stfc.ac.uk. They are establishing an International Space Innovation Centre on site to focus on 3 areas initially, climate change, ensuring the security of space systems and services (am I the only person who recalled an early James Bond film?) and exploiting data from earth observation facilities. Another HPC resource is SCARF, with computing support for Diamond, ISIS and other local facilities; it consists currently of 2000 cores with an additional 800 being added this year. Still some teething problems with their new computer centre, most recently concerning UPS power stability issues; meanwhile they have added 1500Kw of cooling capacity.

BNL: as well as US ATLAS taking LHC data, RHIC Run 10, a high luminosity run, is currently in progress at BNL. HPSS is alive and well, V 7.1.1 was installed last October. LT04 tapes are gradually replacing LT03 and 9940B tapes and drives. Their NFS service is based on BlueArc appliances and they have introduced new BlueArc Mercury clusters for ATLAS which also uses dCache, version 1.9.4. Their processor farm contains some 10,000 cores in 2,000 systems running 64 bit SL 5.3 although most user apps are in still in 32 bit mode which requires careful matching of application to operating system. They recently added 6,000 sq feet of data centre space in a new facility.

CERN: brief summaries from Helge Meinhard of LHC restart, recent IT re-organisation, advancing implementation of ITIL. Reported LCG data export rates up to 3GB/s. Description of our plans for the offsite centre and the ideas around a more remote site in a member state. Recently added 2,188 servers, all 4 dual-CPU systems as well as 613 disc servers with 1PB of external storage. Harking back to his previous responsibilities, Helge listed some of the technical problems uncovered by the recent procurement. Exchange 2007 will be skipped because of incompatibilities with pine and alpine Linux mail clients and John Gordon in RAL confirmed that RAL are doing the same thing. Good progress is reported in the CVS to SVN conversion although there are still some incompatibilities between versions 1.4 and 1.5; from Paris, Michel Jouvin proposed that we move as quickly as possible to 1.6 but Redhat will only support SVN 1.6 in Redhat V6.

DESY: lot of construction completed or planned around DESY Hamburg. As well as the new computer room recently fitted out in Hamburg, a new computer room is being built in Zeuthen. A Petra III DAQ storage has been installed and a CFEL (laser experiment) analysis farm prototype setup. The annual volume of PETRA III data could reach 1PB per year. They are also gearing up to support some astrophysics projects. Grid activities are not restricted to LHC but also support astrophysics and Lattice QCD. Despite the positive view of Lustre reported below, Wolfgang Friebel reported a number of Lustre client crashes, slow performance and bugs causing some degree of user frustration. Experiments with AFS-OSD (AFS with object storage devices) but no decision yet on the future. They are about to rollout Windows 7 but, as since some years, are investigating alternatives to using Netinstall for this. On the other hand, Windows 7 printing is now integrated under Samba. As for Windows storage, the Hamburg and Zeuthen sites are diverging, the former still using a NETAPP cluster, the latter moving to SANmelody. Finally, under pressure from the increasing number of Macs, they have established a limited support service.

[1] CHAOS Summary 2009, Standish Group

St.Petersburg Nuclear Physics Institute: this is the first recorded attendance of someone from a Russian institute that I can recall[2], ironic that it should be at the meeting with the smallest number of onsite attendees. PNPI supports HEP as well as various other physics disciplines. The speaker described their installation and some support aspects of it. They have around 140 registered users.

SLAC: moving from HEP to multi-science, predominately photon science, New CIO and 2 new senior positions, all hired from industry. As well as their 2 containers with 500 Intel nodes described in previous meetings, they have added recently a remote centre several miles away with 192 Dell systems logically attached to their main LSF batch cluster. They have 7 sub-clusters, one starting to use GPUs using CUDA but looking at openCL; consists of 4 servers each with 2 Intel Nehalem cores and 250 GPU cores. Still use HPSS, AFS, TMS but they are looking to replace TMS by TiBS from Teradactyl in the future. Like other sites, SLAC are planning to expand their space and power capacity.

FNAL: new facility being built to expand the main computing centre using ARRA funds.[3] Also their grid computing centre (actually 3 centres) is being expanded. The move to ITIL continues without much change in practices because of good pre-planning. Incident, Change and Problem Management have been implemented, Release, Service Level, Configuration and Continuity Management remain to be attacked. A purchase order was placed for a 23 nodes times 8 cores Nehalem Fermicloud system. Major power problem in February after a series of breaker trips; to get their power circuits back under specifications they had to shed some load and this led to a major network reconfiguration. During this outage a number of lessons were learned, for example have a good incident plan, how to communicate to colleagues and customers when all systems are down. See the overheads for a long list.

INFN CNAF: the new computer centre, built by taking over 4 underground parkings, has permitted a significant increase in capacity. They also now benefit from taking their power directly from a supplier as opposed to from the local university. The total increase in CPU capacity over a 12 month period is expected to be 3.5 times by the end of 2010Q2. Virtualisation is a main activity in CNAF (Worker Nodes on demand - see talk at previous meeting); currently 1400 production VMs and expected to reach 4000 by the summer. They have 10 GPFS clusters, one per supported experiment at CNAF, and each including GridFTP, STORM and TSM clients. In total 500 real nodes but also 3600 virtual nodes.

PDSF: report given by Sandy Philpott (JLAB) from the meeting room since California is so out of phase with Lisbon. No significant change to PDSF other than ongoing increasing capacity of CPU, storage and networking. They have a GPU-based cluster with 48 nodes, each equipped with Tesla GPUs with either 240 or 448 CUDA cores. NERSC is also now looking seriously at IPv6 and all the implications around this.

Jefferson Lab: a major upgrade project now underway, due for completion in 2015, will double the need for computing. They got $5M worth of ARRA funds for Lattice QCD investment which resulted in the arrival of GPU clusters and Lustre to the lab. Also, last year, they installed a large HPC Infiniband cluster of 320 nodes which joins an older Infiniband cluster with 396 nodes, also dedicated to HPC. Their 2010 purchases were won by Koi, also the current FNAL supplier. All the 2010 purchases can be equipped with GPUs and they will use the Tesla GPUs earlier described in the FNAL report. 200TB of Lustre space were installed in autumn 2009; currently accessed via Infiniband but Ethernet access should be available soon. A few minor teething problems with Lustre but they are still settling down the installation. They are moving their analysis cluster from 32-bit to 64-bit; since this involves a switch of

[2] Although the attendee told me he had been before during his stay at SUNY

[3] ARRA – see previous HEPiX report; funds made available by US Government to induce investment to counteract current financial situation

operating system (it was an old cluster based on Fedora 8 and Jefferson now uses CentOS) the users are not so happy.

Storage Sessions

Andrei Maslennikov in Rome reported **recent results of his storage working group** from their test facility at KIT using a test case from CMS and a new test case from ATLAS; he also reported on a new questionnaire he had sent to 14 sites. Two sites, CERN and RAL, use CASTOR which covers one-third of the total storage included in the survey; a second third comes from 8 sites using dCache. Shared file systems make up 20% of HEP storage and 50% of this now uses Lustre, compared to zero after the last questionnaire in 2007; the rest is mostly GPFS and the recent successful migration of CNAF from CASTOR to GPFS/STORM demonstrates the feasibility, according to Andrei, of using shared file systems to store HEP data compatible with LCG access. More details, including a description of the test set-up, can be found on the slides. Andrei ended by expressing some positive views of GPFS (excellent results, a more liberal licensing policy from IBM and a flexible technology permitting smooth online modifications). On the other hand, the AFS/Lustre amalgam offered the best performance in the two extreme test cases.

NFS 4.1 results running with dCache: NFS being an industry standard, it means the dCache client caching based on NFS 4,1 need no longer be maintained by DESY and security is part of the in-built protocol. There remains some work to complete, for example on security, in the dCache NFS 4.1 server. The client will be usable with Redhat 6 which is based on the Linux 2.6.34 kernel which includes support for NFS 4.1; it is currently under test on Redhat 5 with the 2.6.33 kernel. The presentation also contained some hints to tune performance when running with ROOT. Although only based on results achieved in the past 5 days, the speaker, Patrick Fuhrmann, concluded that NFS 4.1 (pNFS) client and server are close to being production level in dCache.

DESY Zeuthen then described how they built a **high performance data centre from commodity hardware**. The speaker listed the modules used with their different features, how he had to modify the I/O scheduler and some optimisations he had applied.

Tim Bell then described **CERN's experience with Lustre**. He presented the scope of the evaluation - could it support HSM, home directories, etc. And he listed the criteria against which Lustre was measured. Performance was not one of them, relying on the HEPiX WG for that aspect. Results were :-

- life cycle mgmt – not ok, no support for live data migration
- backup – ok
- strong authentication – almost ok because not yet full Kerberos, expected in later versions
- fault tolerance – ok
- small files (home directories) – almost ok, problem mixing large and small files
- HSM interface – not ok, not yet supported but under development
- Replication – not supported
- Privilege delegation – not supported
- WAN access – not ok, needs full Kerberos for security
- Strong admin control – not ok, cannot stop users files from striping which could hit overall performance

Finally it was considered that Lustre is not (yet) ready for wide usage at CERN although the roadmap holds out some hope for the future. But the roadmap seems to be stretching into the future. It remains a potential candidate for an analysis centre, if manpower could be allocated to adapting it for this. In the meanwhile, CERN will continue to watch the progress with the roadmap[4].

Elsewhere in the storage area, CERN is considering disk-based data archiving instead of the current tape-based solution and simulations have been done comparing disc- and tape-based archiving. Including technology trends, investment costs and power needs, they appear comparable in the medium term although there remain areas for investigation – data integrity of both solutions, fail-over, operations cost, etc.

LCLS[5] Computing at SLAC: first light happened last year with 10 experiments taking data in the first run. Although users were expected to take this data home for analysis, this did not happen and some "spare" blade systems from level 3 DAQ clusters had to be made available for onsite analysis. In 2010 the project expects to produce even more data and some investment is planned. In the online side, Lustre has been dropped for performance reasons. For the future, especially for analysis capacity, they are looking at GPU-based solutions.

Grid Enabled Mass Storage Systems (GEMSS) for LHC experiments: this was a presentation of the setup at CNAF, initially with CASTOR but now with GPFS, TSM and STORM. The setup was described and the various tests with CMS, LHCb and ATLAS code were presented in some detail.

The openAFS roadmap[6] was presented by Jeff Altman, president of YFS[7]. Jeff is becoming a fixture of HEPiX events and he presented the current state of the roadmap, comparing results to goals, and plans for the future. He presented in particular features coming in the next version (1.6) this summer. He also listed the so-called unfunded goals and ways in which these could be transformed into (funded) roadmap targets. He proposes, to make collection of funds easier, the creation of a not-for-profit organisation to control the future direction of the product but they are currently blocked by the propriety nature of the AFS and openAFS names; negotiations continue with IBM. The implication of this is a freeze of the European wishlist and funding agreed last October.

Last of the Storage Sessions was one on **Lustre-HSM binding** from CEA. The aim is a seamless HSM integration of the filebase. The backend should be independent of the chosen product and the prototype supports several backends including HPSS and ENSTORE. Stress tests are about to begin. He described the components of the interface and of "Robinhood", a policy engine they have created for Lustre-HSM. He also listed some features they would like to add in a future release, such as partial file release/restore, online metadata snapshots, tape friendly mass restores, etc.

Grid Talks

Experience with the Sun Grid Engine batch system; this was a report from a Portuguese site on the GE open source product taken over and further developed by Sun as SGE. He listed the features and some local tools built around

[4] During this presentation, no one brought up the question of support after the Oracle take-over of SUN although exactly this question was hotly discussed in relation to a later talk on Sun Grid Engine
[5] Linac Coherent Light Source project at SLAC
[6] Notes from Andrei Maslennikov
[7] YFS – Your File System - is their implementation of openAFS

SGE. It has been used with an Itanium-based supercomputer for MPI applications, which required a certain amount of local tailoring. Other work was required to set priorities for some jobs, for example those for the local regional weather forecasting service. A Cream-CE will soon enter production to permit access to the ex-EGEE community. During the questions it was admitted there is some concern for SGE's future; although an open-source product it does depend largely on Sun, and Oracle appear to cast doubts on its future support for this in various conversations. IN2P3 are most concerned because they had decided to move from their local BQS to SGE; they say they are negotiating with Oracle[8]

Oliver Keeble presented the **Grid data management middleware plan for 2010 at CERN**. The components under discussion are FTS, DPM/LFC and gfal/lcg_util. A product team. In EGI speak, has been created and is fully represented in EMI. The plans are based on three main themes – manageability, performance and standards. He then went through the currently-planned releases of the various products with the changes at each stage. As well as these, there is a review of FTS planned for July and Oliver presented the list of points they wished to discussed and invited the audience to submit others. Plans include an NFS 4.1-based server for DPM, allowing users to deploy standard NFS 4.1 clients. Oliver did point out that although they have lots of plans and desires, they do not yet have all the resources needed to work on them.

Someone from a Portuguese Tier 2 site (University of Minho) then presented a **case study of a local EGEE site deployment**. The site itself contributes to CYCLOPS, EELA-2 and some local grids. The speaker described the main features of the glite configuration and set up. They use the ROCKS toolkit to install, configure and administer clusters. They define a type of cluster node as an appliance. They create a Direct Acyclic Graph of nodes and use XML files to define the files to load and install modules, set parameters and define inherencies and dependencies between the nodes on the DAG – for example a computing element will inherit many attributes from a compute appliance. As well as appliances, they define "rolls", software bundles, and one of these is an EGEE roll which itself contains an appliance for all types of middleware node types. Sites then use an interface to define local configurations. The first version can install and configure a minimal EGEE site and work continues to perform updates and add VO management tasks. In answer to "why a new tool?", LIP explained they had found, for example, that it was too hard to maintain their tools on QUATTOR, it was too intrusive and it did too many things, some of which LIP did not want. They were encouraged by Michel Jouvin to share their work with other small sites.

Virtualisation

Tony Cass started this session with an update on the **HEPiX virtualisation working group** which he chairs. He first presented the WG objective – to enable virtual machine images created at one site to be used at other HEPiX (and WLGC) sites. The working group has defined the following areas to work in – generation of images, their transmission, contextualisation and support for multiple hypervisors. They have decided not to debate expiry and revocation. Current discussion centres on generation, especially defining roles and having endorsers for the different components (base O/S and the VO software). The team discussing transmission of VM images are likely to recommend a basic transport protocol(s) and perhaps some optional ones. They are unlikely to comment on intra-site image transmission. The contextualisation team are likely to produce a mechanism to allow sites to configure the images. There was a 2 hour EVO WG workshop later in the meeting.

[8] Later that day, a member of the audience discovered an official Sun web site (with the Oracle logo) which spoke of the "Oracle Grid Engine" - http://www.sun.com/software/sge/. Where does that leave us?

Ulrich Schwickerath then reported on **virtualisation work at CERN**, concentrating on work for batch applications. They use building blocks of servers with attached switches and storage. These are used for different use cases from critical systems to small disk servers. The hypervisor used is the Microsoft flavour, HyperV. Users submit their jobs in the usual manner and these are passed to a VM provisioning module. CERN is evaluating two – openNebula (ONE) and the ISF product from Platform. He presented some recent results and the relative differences between ONE and ISF. Most recently they have been testing rtorrent for image distribution. Scalability tests are being carried out using "borrowed" machines before they enter production. Up to 7000 VMs have been run under ONE and similar tests are in preparation for ISF, with the hope to open a production service for users on one or the other, transparently to the user.

Ian Gable from the University of Victoria, BC, presented **an adaptive environment for clouds.** They have a number of applications requiring modest resources (e.g. a HEP legacy data project to be able to analyse BaBar data for the next 5-10 years, an astronomy project and a forestry project) which they think appropriate for clouds. The BaBar application, a collaboration with SLAC, could have implications elsewhere in HEP. The idea is rather simple, creating VMs with the applications loaded and run jobs for the users with the possibility for the user to customise the VMs. To schedule the jobs there are a number of solutions including using the Nimbus Context broker to create "one click clusters"; and Sun Grid Engine submitting to Amazon EC2. Victoria built a local simple cloud scheduler: users select a base VM from the central pool, tailor it to their needs, add their applications and then create a batch job, specifying on which VM it should run. Ian then described the process and the implementation using Condor as the eventual job scheduler. They support Nimbus and Amazon EC2 with experimental support for OpenNebula and Eucalyptus. So far their early experiences are positive – successful validation of BaBar analysis at 3 remote Nimbus sites across Canada. Some 2000 useful jobs have been run for the astronomy project, a dark matter search. These jobs run for 46 hours and have low I/O – ideal for clouds. The service is still in alpha testing with lot of work to do, for example on scalability, data access patterns and security aspects.

Lorenzo Dini presented **Virtualisation in the gLite software process**. gLite consists of ~2M lines of code, 258 RPMs with large numbers of dependencies, different programming languages, etc. There are 12 nightly builds and each build takes around 3 hours. The team covers the full software life cycle – development, build, integration, automatic testing and certification. They support cloud-like virtualisation to provide interactive systems where developers use ssh to access the cloud in an Amazon EC2-like way. They wrote a module on top of XEN which supplies a variety of APIs with security-based ssh keys to login. The other service is a batch-like virtualisation for the builds and for testing where users submit jobs to the VM. Because the userbase is trusted, there is limited validation of the images. An extra requirement is to have snapshots afterwards for debugging. This service is based on VMware with Condor for job submission. They also developed a small tool to permit users to create a virtual environment in their own PC based on one of a number of installed hypervisors. Work is now in progress to merge these services into the central service being offered by the CERN/IT Department; the main areas of resolution are image creation time and the variety of required images.

The next talk came from Thomas Finnern of DESY on **virtual network and web services**. The service is based on an F5 Viprion blade cluster and he began by describing the features of version 10 of the software. They operate in two modes, dumb and smart. In the former, the service network traffic is handled by the switch which is seen as the client by the server while in smart mode the switch acts more like a gateway and this is now the preferred way of working. They use this scheme to redesign the main DESY web page to remove single points of failure plus provide

load balancing and to provide separate read and write pools for content management. A second example is the provision of a DESY status service, again with no single point of failure, load balancing, etc.

Carlos Garcia Fernandez presented the work done on **virtualisation in the CERN/IT Oracle support team**. They adopted virtualisation because of the large number of Oracle database instances and the need to control the necessary resources. OracleVM uses Para-virtualisation, a virtualisation technique where the software interface to virtual machines is similar, but not identical, to that of the underlying hardware, thereby requiring guest operating systems to be adapted. Tests were done with EDH and APT[9] to compare the performance of Oracle VM against pure XEN (which is hardware-assisted virtualization, a virtualisation technique that enables efficient full virtualisation using help from hardware capabilities, primarily from the host processors.). Oracle VM showed some advantages but there are some incompatibilities with CERN's network topology and feedback has been sent back to Oracle. Configuration of OracleVM has been integrated into Quattor and ELFms. They are keen to exploit OracleVM further and hope to have more to report at the next meeting.

Benchmarking

There was an **update on HEP-Spec06 tests** done on the latest Intel 32nm chips and the latest AMD 12 core chips. The tests and results were presented in some detail and the interested reader is referred to the slides. It has to be noted that some doubts were cast regarding whether the kernel used had support for these specific chips and this may have affected the results.

Next up was a talk on the **influence of hyperthreading on CPU performance**. The Intel Hyper-Threading (HT) technology enables one processor core to present two logical cores to the operating system (OS), allowing it to support two software threads at once; what effect does this have on real performance? When there is virtually zero I/O, the effect on a typical HEP code is of the order of 20% increase with HT enabled. With moderate I/O on a fully loaded system, this performance boost can reach 30%. For parallel applications however, the results are more spread and in some cases can even show a decrease in performance. The speaker, from LIP, also noted a number of crashes when HT was enabled on some servers with heavy I/O, crashes which did not happen with HT disabled.

Operating Systems and Applications

Troy Dawson gave his traditional **update on Scientific Linux.** They know of 201 mirror sites and 30 public mirrors and distributions off them are not counted. Despite this the total number of downloads is approaching 45,000 with SL5 surpassing SL4 for the first time. The latest release is 5.4 (Nov 2009) for both 32 and 64 bit. Currently building 5.5 but no strong estimate for release, perhaps early June. They are also building pre-alpha SL 6 based on the Redhat alpha 6, released 2 days ago, but SL 5.5 will remain the priority; he predicts around February 2011 for the first production release of SL6. Security patches for SL3 will stop this October as previously announced. Also installing newer and faster distribution servers which should go live in a week or so. Other SL-related work is described below.

Michal Budzowski then presented the **Windows 7 update at CERN**. After work with a beta release throughout 2009, a pilot was started in December 2009 and a service advertised in March 2010 supporting all desktops and laptops purchased since 2006 and satisfying some minimum hardware requirements (2GB memory, 2GHz clock speed, both double the Microsoft recommendation). It is now the default for new 32 bit installations and already there are 430 installations. CERN has some 6000 managed Windows PCs, mostly XP so there is a long way to go. Given the lifetime

[9] Two very common CERN administration applications

of XP, support by Microsoft is promised until 2014, and the hardware lifetime of old, pre-2006 PCs, estimated at maximum 5-7 years, there may be a problem towards the end. The various support groups have certified their applications. For example the wide range of engineering apps both 32 and 64 bit, have been tested and certified for Windows 7 with the exception of Catia/Smarteam (coming in Q3) and PVSS (coming but no date). Michal described the changes made to standard Windows to adapt it to the CERN environment.

The final session of the week was a **Twiki at CERN** presentation by Peter Jones. He first explained why CERN had implemented a Twiki service and described some features of Twiki. Initially the installation relied on AFS but in March 2010 he migrated the backend to NFS, mainly to get round a limitation of the number of files in an AFS directory. The growing popularity of the service has led to a number of hardware upgrades. Pete presented some impressive usage numbers - 7500 registered users. 190 collaboration webs, 60,000 topics with 280,000 attachments, 3,000,000 accesses per month, 50,000 monthly updates. Main users are CMS and ATLAS. Performance is constantly monitored and any opportunity for improvement is explored. Authentication was originally based on Kerberos but since 2007 SSO is used. On request of the users, access control has been implemented and linked to CERN's e-group scheme. In the future they hope to re-introduce load balancing, complete SSO integration, automate the installation, upgrade to blade servers and continue to improve the search feature. Peter explained the background to some unrest in the worldwide Twiki community but CERN has decided to remain with Twiki for now for stability reasons although they will watch the situation closely.

Other Topics

SLAC power outage: Like FNAL in February, SLAC had a power outage in January, this one thunderstorm-induced, which had major effects for 2 days. Randy Melen described the sequence of events from when the power went down. He listed the physical problems, no power, no lights, no coffee. They discovered they need noise cancelling systems to use phones in the computer rooms. They found a room (with windows) to prepare the priority for restarting systems when power was restored (24 hours later in the event) – payroll, Gamma ray telescope computing, mail, web, etc. Since re-establishment of systems the post-mortem has continued and they conclude they need better internal documentation, faster ways of updating communications channels, a better understanding of dependencies, better coordination with their Facilities Department, and so on.

Troy Dawson from FNAL gave a presentation on the **Spacewalk and Koji projects at FNAL**. The first is an open source version of Redhat's so-called satellite service to prepare customised packages of software. It can be used to perform inventories of systems, system monitoring and it can generate kickstart systems. The FNAL tests cover setting up errata pages for SL, machine monitoring and evaluating what is the flexibility and scalability of Spacewalk. He showed a series of screenshots of the results of the tests. First results are positive but much still remains to be tested. More news next HEPiX. Koji is a software product to build RPMs, for example for Fedora. FNAL would like to adapt it to build Scientific Linux (SL). It would add some features and more flexibility to ease the build process, allow easier builds by collaborators and make it easier to track the history. Setting it up (almost entirely done by Connie Sieh, the other SL author) was a major task as well as establishing worker nodes. Tests continue.

Ian Collier reported on **Quattor experience at RAL and its outlook**. This was a follow-on to his talk at HEPiX Berkeley which had described how he had introduced QUATTOR to RAL in the autumn of 2009. Now used for nearly 700 worker nodes and several hundred disc servers, gLIte servers, etc. SINDES (secure information distribution service) is about to be deployed. Issues uncovered include the significant amount of work to setup a new machine type although O/S changes are easier; also the QUATTOR approach is often different to previous approaches ingrained in many system managers. He claims that Quattor-managed sites tend to be more available than non-QUATTOR sites

but his statistics are surely biased by including the Tier 0 site. The move of QUATTOR to sourceforge is almost complete. There are a growing number of commercial users, one in particular very large site (20,000 nodes). The recent bid for EU FP7 funds was unsuccessful but the preparation was a useful exercise in setting out a roadmap.

Alan Silverman

23 April 2010