HEPNT/HEPiX Catania Report

April 16-19, 2001

Alan Silverman IT/PS

Logistics

The meeting was held in INFN's Laboratori Nazionali del Sud in Catania, Sicily. Around 60 people attended, mostly European but FNAL, SLAC, LBNL/NERSC/PDSF and Jefferson Lab were represented (once again, BNL were conspicuous by their absence). INFN offered all costs – lunches, coffee breaks, all infrastructure costs. There was also a very complete audio/visual unit to broadcast the conference on the web, setup and staffed by A/V professionals. Most of the overheads presented can be accessed via the web site at http://www.ts.infn.it/conferences/hepix2002/.

Highlights

- OpenAFS use is growing just as IBM officially announce end of life; should HEP "do something"? At the meeting, someone from DESY offered to organise a first discussion of the issues of concern to HEP sites using AFS/OpenAFS.
- VMware use is also growing in many HEP labs
- Mosix clusters are being currently investigated by 3 different HEP groups: is this a growing trend or flavour of the month? Meanwhile the Mosix project has spun off OpenMosix 3 months ago which has spun off a commercial product 3 days ago. Perhaps anyone interested should wait until the dust settles.
- CERN is still by far the most advanced in Windows 2000 rollout among the major (and minor) labs
- Much interest in hardware testing: SLAC's tests of hardware for new Linux PCs and Windows storage (2 unrelated tests); first results from CASPUR's Storage Laboratory tests; report on storage system tests by PDSF. Feed these results into the PASTA 3 review?
- INFN's generosity in paying for all the logistics (no conference fee, all catering offered, buses organised, social event subsidised, professional audiovisual unit filmed and transmitted the whole 5 day conference).
- John Gordon and I made serious attempt to get HEPiX to adapt and cater for at least some of the technical issues brought up at the LCG kick-off meeting. Needs drivers for the different topics. First success in the OpenAFS arena (see above)
- SLAC are suffering 5 to 10 PC reboots per day on their 512 node Linux farm, since many months, cause unknown. VA Linux are not blamed but SLAC are about to award the next 512 node order to an even smaller supplier, largely it seems because of better racking options.

Windows 2000 Coordination Group Report (Michel Jouvin. LAL)

Working Group created by HTASC; the meeting the previous day is the first meeting where US sites attended in person (as opposed to via video-conferencing). It had consisted largely of site reviews and then open discussion on topics such as:

- Sharing MSI[1] production effort

[1] An MSI file is a Windows Installer package

- Application development; SMS[2]/AD[3] co-existence
- Printing services management
- Office 2000/XP co-existence and/or migration; no site is moving yet to XP but no problems have been seen on first tests
- Windows.NET testing
- Web folders and distributed file access

Plans

- MSI sharing; CERN as a repository; how useful?
- Web folders, WebDAV: not a priority for most sites since OpenAFS works well enough. But encourage inter-site testing of WebDAV among interested sites (which?)
- Co-locate future meetings with HEPiX.

FNAL Windows 2000 Security (Jack Schmidt)

Schmidt is in charge of FNAL's W2K Working Group. FNAL has 5 NT domains but no inter-domain trust so the goal is to try to work more closely together as part of W2K introduction. But FNAL's introduction of strong authentication has disturbed plans. It led to a faster pace and enforced shortcuts. There is now a top level W2K domain for high level security – no users, only a few admins. The main user domain is one level down; it also contains most of the resources. There are other child domains also. Lab policy forces uses of Kerberos 5 but old Windows systems (95, 98, NT) have a waiver as they need also NTLM authentication. Two central Kerberos servers – MIT and AD-based – with two-way trust between them.

Systems which are deemed "critical" (defined as - disturbance of such a system could seriously disrupt the work of the lab; for example domain servers) must have a written security plan, drawn up and accessible by only a few people. These plans are so secret he could not describe them in detail in a closed meeting like the W2K Working Group. What he could say is that the nodes must be physically secure (in locked cabinets), with secure remote access via Ipsec; they are continuously monitored for state changes, etc.

Regarding user accounts, FNAL has decreed a single user account Windows/UNIX. There is a single place to create, authenticate and disable accounts. There are many ideas for OU[4] administration (see the slides for the complete list) but the working group is too busy with the migration, has no programmer in the team and no money; so these remain ideas only for now. No AFS service on the Windows side although the AFS client is available but there are few physics users on the Windows side so it is little used. There are about 300 W2K users, expected to grow to 2000. There will be 4 OUs in total.

Future topics of concern in security include terminal servers, demand for shared accounts and access from home.

During the questions, it was suggested that a future HEPiX/HEPNT meeting (the next one?) should feature some general security talks where all the sites represented at this meeting can share experiences. Or is there another forum where this already happens?

[2] SMS is Systems Management Service
[3] AD stands for Microsoft's Active Directory
[4] OU – Organisational Unit – groups of users for example

Windows NT to 2000/XP at SLAC (Dennis Wisinski)

Wisinski is the project mgr for the migration effort. Today they have a single master NT domain containing all the user accounts plus a dozen resource domains and one "hidden" domain with restricted access for central databases and business services.

They plan to use a BIND DNS server which will delegate control of the Windows domain to the Windows DNS. This will run AD to enforce Windows authentication of machines. IIS web servers will need also have entries in the BIND DNS server to avoid breaking existing URLs. And the DNS db must be regularly synched with an ORACLE db of DNS names.

SLAC has chosen to have a single tree, single domain structure with departmental OUs for delegation of authority; they may need more OUs later (for divisions and/or experiments). Will not upgrade NT domain controllers but create a new native-mode W2000 domain with two-way trust with the NT domain. This will allow incremental migration and easy migration back if there are problems.

Will migrate to Exchange 2000 in a similar way in due course (from 5.5 today). Mailboxes will be migrated at the same time as the user accounts. Migration of the business domain depends on the availability and stability of PeopleSoft under W2000.

Client nodes should go to XP rather than 2000, given W2000's expected lifecycle (it will be 4 years old by the time migration ends and thus near the end of its supported life). XPs's compatibility mode may even help some migrations. Local admin training has begun; next comes application testing under W2000 by local admins and power users; the Helpdesk staff will also need training. Then they will begin a campaign to raise user awareness of the coming change. Actual training and migration of users should start end-2002 and run through all of 2003. Will use external trainers and just-in-time training. Migrations should be performed by in-house teams. Some 3600 user accounts to migrate (but many users only have accounts to get to business systems.

Starting Windows 2000 at DESY (Reinhard Baltrush)

Baltrush has been appointed project head for this activity. The start has been delayed by the re-organisation of the previous DESY NT group, now absorbed into the main IT Group and re-staffed. New goal is to use XP on the clients and take account of .NET. The new domain is to be built in parallel to the existing NT domain. There are 2800 users to migrate. They plan to have only a single domain covering both DESY sites. AD must be configured to match the DESY group/administrator structure.

No decision yet on choice of domain name/DNS zone; possibility to use newly-installed DNS scheme from Lucent but implications unknown. Which Kerberos? More than one Domain Controller will be needed (two sites, trust with NT domain, implications of/to .NET). What about common home directories? Today they use NETINSTALL for application support but does this tool support (well) XP? How does it integrate with AD group policies?

Current activities: now arranging test environments on both sites; planning the AD structure and consulting the main user groups for their requirements. Official kick-off date was 27th March this year. AD structure should be defined by Q4, first Domain Controllers installed by 1Q03, admin tools available by 2Q03 along with first test users; production AD structure in place by 1Q04.

CERN Windows 2000 Mgration (Alberto Pace)

A report on the NICE 2000 migration.

Active Directory Project at IN2P3 (Michel Jouvin)

The challenge is that each IN2P3 site has its own autonomous computing environment. LAL is the only one today with AD. Other labs have only NT and scarce knowledge of Windows, especially W2000. Plans to build a national tree with one root domain and one domain per lab. Root domain will be in Lyon. Share DNS namespace with non-Windows usage. The lab domains will be managed by the UNIX DNS (normally) – no dynamic updates will be possible. AD zones may be added, managed by Windows, then dynamic updates would be possible.

Accounts will be managed locally to each lab, no national accounts. A national DFS will be set up, mainly for sharing admin resources, no (significant) file sharing expected. Use of Intellimirror at the forest (national) level to have a national MSI repository.

Minor (?) caveat: all this is a plan by a few devotees; there is no approved project. They plan to set up a pilot to prove the feasibility and then persuade IN2P3 mgmt to accept it, almost as a "fait accompli". Ten labs will participate in the pilot, which has not really started yet. They expect a production-level AD by September and then to ask for funding to buy the production Domain Controllers, not so much for the funds, more for the recognition of an official project. Then existing NT domains could be migrated in.

Windows Storage at SLAC (Brian Scott)

Currently they have about 3TB split between user and group files and the Exchange repositories. No use of disc quotas today. Using Dell SAN today. Why move? There was a major problem in power cycle of the SAN last July – corrupted the Exchange database requiring a long restore; other reasons: poor file system diagnostics, a security bug lost all ACLs on the file base, a failure of the so-called fail-over service resulted in a serious loss of service.

Looked at many solutions from many suppliers. Ruled out storage visualisation – too new, no standards yet. NAS – ruled out, no NTFS support. Direct attached storage (DAS) – hard to manage, inefficient in storage, backups over the net are slow; but not ruled out. SAN – scalable, NTFS support, separate network. Still evaluating the various products from among IBM, Compaq, Storage Tech, Fujitsu, Dell, Network Appliance, SUN, Hitachi, HP. Looks like they will chose DAS or SAN. They expect to pay about 15 cents per MB in investment costs.

Exchange Pilot at CERN (Alberto Pace)

Alberto described our plans for the Exchange pilot.

Computer Security at CLRC (RAL+Daresbury) (Gareth Smith)

CLRC have a hierarchical security structure with a representative in each department. It was created in 1998 as a result of an audit. One of the first actions was to implement a firewall scheme in 1999 (implemented in the routers), hiding behind this most of the systems on the sites. A further audit in 2000 recommended particular care of critical ("high business impact") systems (cf. FNAL security talk). Extra filters were added to the mail filter after damage inflicted by the Love Bug virus. In 2001, home users were given guidelines for protection of home PCs, requiring the use of anti-virus software and a personal firewall for users dialing-in. (CLRC have a bulk licence for ZoneAlarm Pro.) They have the usual

anti-SPAM mail filters, mail site blacklists and so on. A dedicated PC-based firewall was installed in 2001 with 300-400 rules, some quite complex.

Over the years, and still true so far this year, there has been a steady rise in the frequency of incidents. And, as elsewhere, they have suffered security breaches and hacks.

Windows Status at CLRC (Gareth Smith)

Still based on Windows NT although many servers have moved to W2000 and in some departments desktops are moving now also. But no plans yet for an AD structure. The reliance on Exchange 5.5 makes any change more delicate although their NT domain structure is rather simple which should ease the eventual migration. They accept that the longer they wait, the greater the risk that some departments will "go it alone".

Windows Status at DAPNIA (Joel Surget)

They have a Windows 2000 domain in place since 2 years. Now populated by 350 PCs, half the lab's population, including all the servers. The old NICE NT domain should stop at the end of the year. This will leave some 200 un-managed desktop PCs running old software. A planned move to SAP should allow some of these to be upgraded. They are migrating some 90 Wincenter users to Windows Terminal Server.

They have switched their anti-virus software from Macafee to Norton Corporate Edition, buying 16000 user licences for all of the CEA. Any chance to join this contract?

Site Reports

CERN

Maria presented a report of current activities in IT as furnished by service managers.

Jefferson Lab

O/S upgrades in execution or planning – Solaris to 8, HP-UX to 11i, Redhat to 7.2. Jefferson are members of PPDG, using SRB as the standard resource interface to PPDG and "possibly the European Grid". Created a Windows 2000 domain and using Windows Terminal Server with Metaframe XP; load balancing of this service is now underway. Purchasing Seagate LTO library for backup. Working (at last) on implementing locally the CERN print project with local additions such as Samba for Windows printers. Further limitations in off-site access to promote more security (use of ssh V2). Now supporting VRVS for video-conferencing. Expect to add 512 Pentium 4 systems this year for High Performance Computing and adding support for cfengine (see separate talk).

FNAL

Last Tru64 to be switched off in December '02. Leaves 270 SUN, 2160 Linux, 84 IRIX and 6500 Windows. The CDF VAX cluster (FNALD) was powered off in March. D0 computing: logging data at 10MB/sec. and have recorded 100TB in the past year. Their central computing facility has a 192 processor Origin 2000 and a 6000 Specint95 Linux farm. They have 400 desktops, 50/50 Linux/Windows plus 600 Linux dual CPU systems and another Origin 2000 spread across 8 remote sites. CDF computing: central

offline analysis is a 128 CPU SGI. They have 280 Linux and 40 SGI desktops. Raw data is recorded at up to 20MB/sec. and again they have collected 100TB since Run II started.

In mass storage, IBM LTO drives are coming in. All data storage controllers are Linux. CDF has joined D0 in using ENSTORE so ENSTORE now writes all data to tape at the lab. Working on dCache in collaboration with DESY for disc caching plus a wide area interface to ENSTORE, including a Kerberos ftp interface. Grid ftp server in development.

Reconstruction farms migrating to Redhat 7.1 with a 2.4.9-31 kernel. All farms use FBSng for batch scheduling. A local tool (fcp) is being used to avoid some use of NFS and a tool (dfarm) was written for handling temporary storage. There are three reconstruction farms; each typically has about 100 Linux nodes and at least one SGI system per farm; there are farms for fixed target exps, for CDF and for D0. CDF is also building a user analysis farm of 50 (initially) Linux dual-CPU systems. They plan to add 2 8-way Linux nodes for job submission. Eventually aim for up to 600 worker nodes and 200TB of attached disc space. Know NFS will not be sufficient but not yet chosen what to use. The CDF/MIT group and CMS locally are considering adding Mosix-managed clusters for data analysis and have created small clusters for investigating this.

Still running mostly Transarc AFS although OpenAFS on the Linux side. Most servers are on Solaris boxes, over 2TB of disc space. Now need to consider OpenAFS more seriously since IBM's recent announcement about ending support but they have some security concerns. There are some 52 centrally managed web servers but there are also 500 web servers throughout the rest of the lab – a familiar story?.

ESnet link should be upgraded soon to OC12 (622 Mb/s) for offsite access. NGOP monitoring all nodes under the responsibility of the Computer Dept. and new clusters are being added.

TRIUMF

Joining a new Grid project called WestGrid, a Canadian initiative currently being put together for funding approval by the federal authorities (target C$36M). Will organise computational nodes in a Grid with TRIUMF as the largest centre along with 3 Canadian universities offering facilities for data analysis and Monte Carlo. At TRIUMF they aim to install 1000 to 1500 CPUs and are looking at blade systems from various vendors.

They are powering off the legacy VMS server and VMS mail server. Looking at Redhat 7.2 for non-CERN apps but still with 6.2 and 7.1 for CERN apps. Looking at consolidating data storage, more efficient use, technology refresh and more flexibility for easy upgrades over time. Looking at a small Mosix cluster to investigate this.

LAL

Main project is to deploy a SAN with 2 fabrics, one for disc, one for tapes. Eventually these should be interconnected to avoid a single point of failure. CERN Print Package is in production use. They have doubts on the future of the package (on Windows) based on the fact that "the CERN author has moved to another project" (said the speaker). Spoke of the possibility of a collaboration for sharing future support. Should we perhaps investigate this possible offer of assistance? Or scotch fears that we will drop its support?

Investing a lot on VMware. Moved to version 3 which offers MSI deployment and the virtual Linux system can be Redhat 7.2. Involved in the European DataGrid as a testbed although the initial testbed deployments were restricted to the main sites.

NIKHEF

Mostly Redhat Linux but also a few SuSE nodes. Stopping HP-UX, one IRIX node left, Windows moving to W2000. Looking seriously at moving to Redhat 7.2 and Solaris 8 soon. Introduced AFS for the first time locally after being only a client of cern.ch cell for many years. Using OpenAFS on Redhat Linux servers, 6 month pilot started. Modified the TCP/IP stack to improve performance on the SURFnet line to Chicago to overcome packet loss and congestion as experienced by D0 collaborators locally.

Moved from Macafee to Norton (cf. LAL Windows report above). Evaluating Update Expert as a possible alternative to SMS. And investigating Netscape 6 for Windows 2000.

SLAC

BaBar will take data up to June and then restart in October. Data rates have reached up to 1TB/day and they are establishing more Tier A centres to handle this. NLC are starting beam tests, putting their data into Oracle. Another exp is putting data into Mstore (see below). SLAC will build a cosmology institute onsite with exps with very high data rates. And looking for funding for high capacity accelerator studies.

Compute farm today is 870 SUNs plus 512 VA Linux systems. Next acquisition is almost decided (in fact it has been but not yet announced to the bidders). The VA PC farm suffers reboots of 2% of the nodes per day. Despite continuous investigations by SLAC and VA Linux staff (the speaker acknowledged very good support in spite of the well-known problems of the vendor) since many months, the cause remains unknown. Current suspect is a poorly-designed power supply. Developed an event monitoring scheme by mining the system event logs.

Current 120TB of storage with another 72TB due to be acquired this year with a target price of 1 cent per MB using SUN T3 arrays. And already looking at the next purchase, which technology? HPSS running well, 800TB stored, adding up to 3TB per day. Should reach 1PB by the autumn. Created a new Solaris HPSS instance for non-BaBar exps. Created Mstore - an NFS interface to users at one side and to HPSS for storage.

Still on Redhat 6.2 but new farm will be 7.2; kernel is 2.4.18. Solaris 7 population going down rapidly in favour of V8. First tests of Solaris 9. Transarc AFS on servers, OpenAFS clients on Linux and now on Solaris 9. Will test an OpenAFS service later this year. Beta testing new version of LSF version 5. New scheduler seems better at managing large queues and handling fair shares.

Rebuilding the computer centre from the inside – new raised floor (better protection against earthquakes!), more cooling (room temperature reached 90 degrees last weekend!), more power. A peer review of security plan and practices received generally good marks, including the feeling that the users were aware of security issues and were involved with protection measures. SLAC to ESnet traffic is the highest rate for all ESnet sites.

LBNL/NERSC/PDSF

They mirror locally both STAR and ATLAS software environments. Still using the Linux 2.2 kernel but looking at 2.4. Promoting openssh, banned ftp. OpenAFS installed. NERSC are collaborating with IBM to put a Grid node on an SP2 for the DoE DataGrid. Investigating GUFS (Global Unified File System) and also GPFS for Intel.

CASPUR

Still run SMPs from IBM, Compaq and SUN, all of which have had hardware and/or software upgrades recently; a total of some 200 CPUs. They have 2 NAS Gbit systems, one for NFS, one for AFS. Creating a Storage Laboratory to investigate performance as storage needs grow and to investigate OpenAFS. [See separate talk.] LTO tape cluster newly installed using the floor space liberated by the move of BaBar tape drives to Padua.

They are convinced that OpenAFS is very solid on Linux including as file servers. Very satisfied by the level of support of the OpenAFS community, at least on Linux, "better than Transarc/IBM". Will IBM offer maintenance through 2003? They think so. May anyway study other possible maintenance providers on non-Linux platforms at least. They still supply AFS services to many INFN sites. Also a collaboration with CERN, including support of ASIS for those UNIX architectures not supported at CERN.

INFN

Heavy use of AFS. The authentication server today is Tru64-based but they are looking at moving to OpenAFS on Linux in the autumn. The Lecce site is testing Kerberos 5 on their local AFS cell. Also heavy use of AFS to integrate with Windows users as well as for many web servces. Among Windows users, there is increasing use of VMware. INFN has its own security group, there are common security rules and security meetings but different sites implement different solutions. Pisa are investigating OpenMosix clustering (see separate talk).

CCIN2P3

Support only 3 platforms now – Redhat (currently 6.1/2, almost ready to go to 7.2); Solaris 7/8; and AIX. HPSS usage has doubled from 60TB last October to 123TB today. For BaBar they run 2 SUN 4500 and 8 Netra T Objectivity servers containing 19TB of data on disc and 75TB on HPSS. As part of their participation in EDG, they have interfaced BQS to Globus.

DAPNIA (Saclay)

Desktops are split between Redhat and Windows 2000. Introduction of SAP has blocked most administrative activity at the lab for several months (for example no acquisitions). To reduce dependence on dual-boot systems, they are considering more use of VMware.

RAL

Started buying 1U racked dual CPU PCs, 146 1.4 GHz PIII systems this year so far, thus doubling the installed computing power. Disc tender resulted in a mixed 45TB SCSI/IDE RAID solution with IDE discs on the RAID controllers which have SCSI links to the host. Required no modifications to the host O/S. Some details of the tender are on the overheads, along with how RAL evaluated and benchmarked the bids. Still not sure which RAID option (RAID 0, 5, N) to run.

Level 3 Nortel Gbit switches across the site as backbone. Conferencing based on ISDN but VRVS for personal use; use of H.323 is growing.

RAL has become a Tier A centre for BaBar. Concerned about AFS/OpenAFS – what platform should be chosen to host OpenAFS?

DESY

Considerable personnel changes, in and out. User and PC population continue to rise but also Solaris, at the expense of SGI, HP and X terminals. NT daily mgmt handed over to IT Group and the Windows Project Group re-launched. DESY Linux 4 being rolled out – based on the SuSE 7.2 distribution, 2.4.17 kernel. Zeuthen is converted already, Hamburg trying to complete the upgrade before HERA restarts. Another site (RAL reported a similar experience) that had to replace IBM IDE discs (although Gianni Siroli reported failure to convince their supplier to do the same at one of the INFN sites).

SUN Grid Engine deployed, with AFS token prolongation. dCache has entered production. Planning for a second data centre in Hamburg. Deploying VLAN technology, permits very easy network connection, especially for mobile workers.

Decision to upgrade their AFS cluster, converting from SGI hosts to SUN, more than doubling their installed space to 2.2TB. Will go towards OpenAFS. Testing various calendaring options, rejecting iPlanet but adopting Corporate Time (previously known as Netscape Calendaring).

Linux Hardware Evaluation at SLAC (Alf Wachsmann)

Unable to name the vendors as the result has not (yet) been announced. Test procedures were taken from FNAL and adapted to include a configuration test, the Cerberus burn-in software (adapted from the VA Linux burn-in test program CTCS and now available from sourceforge.net and which SLAC runs for 2 days), local stress and benchmarking tests with SLAC/BaBar software, etc. Usual players sent hardware for testing in various configurations, various numbers of CPUs per box, almost all in 1U rack units. He pointed out some basic design errors (inaccessible power switches, no indicator lights, etc) but not knowing which vendor systems were poorly or well designed did not help in identifying if for example major vendors offered better-designed systems that local build-to-order suppliers. He stated in answer to a question that company size or reputation made little difference. He told me that if we ask for a private copy of his talk we can get more useful information, identifying strong and weak points by vendor.

CASPUR Storage Lab (Andrei Maslennikov)

Goals are to perform studies on current and new storage solutions. In the short term, they want to decide how to migrate to a new OpenAFS service. Included wide area tests with CNAF in Bologna via a 2.5Gb line. Tested SCSI, FC-attached and IDE discs. Also tested some LTO and AIT-3 tape drives and SCSI/IP appliances from CISCO and Dot Hill. Tested various file access methods including AFS, NFS, GridFTP, RFIO, scalable NFS based on IBM's GPFS. The tests, especially the tape tests, are being done in collaboration with CERN (Fabien Collin, Gordon Lee, Charles Curran, et al). The first results were presented along with problems met; see the overheads for details. RFIO came out particularly well on 15KB reads, coming very close to the raw disc read rate, 50MB/s in his example compared to the raw disc rate of 53MB/s.

Computer Centre Shutdown/Startup (Alan Silverman)

I presented the postmortem which Tim Smith showed at C5 summarising the experiences of the tests of these procedures during the Computer Centre shutdown on Feb 9th.

Clustering with OpenMosix (Maurizo Davini, INFN Pisa)

A description of Mosix clustering software now becoming popular in various sites, including, it appears, some HEP sites. Mosix was born in the '80s on PDP-11s for the Israeli army. OpenMosix split off in Nov 2001 when the original author came into conflict with his collaborators over licensing. Current version for Linux is 1.5.4 for Linux kernel 2.4.17. Single system image for a cluster, like an SMP. No need to modify apps for cluster use, claimed to scale linearly. Processes can be migrated transparently at any time between nodes based on node load as collected on one (randomly-chosen) node of the cluster. Source available from Sourceforge, complete with admin tools (there are 14 parameters available for tuning to modify Mosix's behaviour plus monitoring tools) and file access methods. They hope the code will be included in the Linux 2.6 kernel package. Over 300 known installations worldwide including a 400+ node cluster at Intel and a 1400 processor SMP system in Japan. The Mosix Group (sponsored by firms such as IBM, AMD, Compaq, etc) is announcing a commercial product going on sale tomorrow (19th April) based on OpenMosix. More information from www.qclusters.com. More obvious uses in HPC computing, rather than for HEP apps.

AFS Administration Framework (Wolfgang Friebel)

A talk on his work in DS Group.

CERN CVS Service (Alan Lovell)

Alan's first talk of the day; a description of PS Group's plans for a central CVS service.

cfengine at JLab (Sandy Philpott)

Need to manage and monitor 90 central compute servers plus CAD servers, farm nodes, etc. with just 6 system admins for all IT activities, including both UNIX and Windows. cfengine is a policy-driven configuration management tool. Originated in Oslo, open source, available for both UNIX and Windows. Define classes of nodes with attributes; binaries and definitions are stored on a master server and copies of the binaries are rolled out to the clients as needed along with a local copy of the configuration templates. Clients check their configurations against the templates at regular intervals and update as required. The frequency of each check is locally decided, for example every 30 minutes plus a daily for changes to the files.

CERN Print Server Update (Alan Lovell)

An update on the status of our central print server.

Update on FNAL's Strong Authentication Project (Lisa Giacchetti)

See previous HEPiX reports for details of the project. A primary goal was the removal of re-usable login passwords; secondary goals included offering a single-sign on environment. Initial completion date was 31st Dec 2001, later slipped to April 2002, partially to include full Windows migration. A lot of work and user education were needed to get users to migrate and become accustomed to the new scheme.

User tutorials were offered late in the year but they only attracted moderate attendance; perhaps they should have been offered earlier. Some users who are unable to get Kerberos tokens and who are forced to use Cryptocards still complain. In retrospect, Lab management should have been more vocal and visible in explaining to users why this project had been felt necessary. Various unexpected issues came up concerning Kerberos tokens needed by jobs running under AFS group accounts (no user login), integration

with LSF, AFS token stealing instances. Local sys admins had to learn how to make their systems work within a Kerberos environment – how does Kerberos affect file backup, inter-node connections, etc.

Now 3624 registered Kerberos users, of whom 2700 have Cryptocards, 2583 server hosts need to be accessed by Kerberos, More than 300K service tickets are issued per day, Win2K is a separate but synchronised domain with 400+ users but it will not make the revised April '02 deadline and there are still more than 300 other systems with waivers (which had to be individually justified as to why they could not be Kerberised, most of them are Windows.) Kerberos-compliance scanning of hosts happens at frequent intervals.

No plans at the moment to extend Kerberos to applications in order to offer single-session passwords.

ASIS Update (Alan Lovell)

An update of ASIS to use RPMs.

ALICE Grid Activities (Roberto Barbera, INFN Catania)

Description of the first rollout of Testbed 1of the EDG from a user's point of view. "Grid computing is a struggling experience today". Catania are creating a web portal to make Grid computing more user friendly – usable from anywhere, point and click. The "GENIUS" project initiative is funded by INFN. The method involves a web browser accessing the GENIUS package which accesses the EDG user interface via a small commercial package sitting on top of an Apache web server. Allowed actions include job submission, job and Grid status reporting, return of the results of a job, file browsing, check one's own authentication certificates, etc.

Solaris 8 Certification (Alan Lovell)

For the last of his 4 talks of the day on behalf of PS/UI section, Alan presented the current plans and progress in certifying Solaris version 8.

FNAL's NGOP - Status Update (Tanya Levshina)

NGOP is FNAL's equivalent of PEM for cluster monitoring and alarm generation. Temporary extra manpower has speeded progress. Since the last HEPiX, there have been 2 production releases. Among the new features are

- Roles have been defined, who sees which events, who can perform which actions on which systems.
- An interface has been implemented to FNAL's Remedy Help Desk. Could later be the basis for more automatic computer centre operations.
- Web admin tools have been released.

NGOP is now monitoring 705 nodes via 1015 agents including ping agents, O/S health agents, and agents built by users for specific servers. Around 15000 objects are being monitored with typically 10 NGOP monitor instances running at any one time.

New type of agent to scan a URL for reachability and for correct content. The web admin tool offers a more hierarchical view of the systems. It also allows to enter commands; it includes a search command and has a multi-user locking protection mechanism. The NGOP GUI is to be redesigned to overcome some perceived major deficiencies such as the need for a lot of memory and sometimes a lot of CPU; also it

does not filter old alarms that may no longer be interesting for particular users. The new GUI should also offer web access.

They will offer this as a low level Grid fabric monitoring tool. No plans for further extensions, for example performance measurement, trend analysis, event correlation, corrective actions although interested sys admins can use the basic agents to build more sophisticated ones.

Computer Centre Supervision at CERN (Helge Meinhard)

Helge described plans to use PVSS for Computer Centre supervision. He demonstrated briefly some features in the P0 release directly from CERN.

The Next Steps in Storage (Cary Whitney, PDSF)

Tested various hardware storage solutions offered by vendors. The challenge was to improve the NFS performance reported by him at the previous meeting. Suppliers included Zambeel, BlueArc and 3ware and PVFS[5] was also tested. Descriptions of the products and configurations can be found on the overheads. They also describe in some detail the actual tests performed along with the results. Both Zambeel and BlueArc do much better than 3ware and PVFS at small and large block size, for both read and write.

Summary: 3ware is cheap and easy to set up but slow and doesn't scale. PVFS is free, runs better the more I/O nodes you add but is fragile, not for production use (one crash can lose the whole file system). Zambeel and BlueArc perform and scale well but are very expensive and Zambeel is very new on the market (he tested a beta release). He did not test random access tests or execute tests with files greater than 2GB.

Large System SIG Report (Alan Silverman)

I presented some activities organised under this banner, namely

- Suggestion to organise a short update (2 days?) on the Large Cluster Workshop in conjunction with the next HEPiX
- Proposal for creation of a network of experts to discuss software certification among and across the sites, especially the Tier1 sites
- Suggestion that someone volunteers to drive a discussion between people directly concerned with OpenAFS issues. [I learned that OpenAFS will be the subject of a full day workshop at Usenix in June. Is CERN sending anyone? Can/should HEP sites present a united position by then? At the break, Peter van der Reest of DESY Hamburg informed me that he will try to organise a first discussion before the June Usenix to discuss the issues.]

UK Grid for Particle Physics (John Gordon)

John described the UK particle physics grid activities.

LCG (John Gordon)

John then described the LHC Computing Grid and where HEPiX might fit in to this. Clearly we already discuss fabric mgmt issues, technology reviews, experiences in running centres. Does HEPiX cover all or at least most of, the relevant sites. Clearly BNL does not often attend. What about Tier 2 sites

[5] PVFS =Parallel Virtual File System

(Universities, Institutes)? John proposes targeting these sites. Should we organise workshops (as the Large Cluster Workshop)? Sponsor specific workgroups? The meeting "agreed" with his proposals but who will actually drive these initiatives?

INFN Tier 1 Regional Centre (Luca Dell'Angelo, CNAF)

He described the prototype tier 1 centre being set up in Bologna.

INFN Grid Testbed Monitoring (Giuseppe Sava, INFN Catania)

He described how the public domain tool netsaint[6] was chosen for this task, possibly because a member of the Catania team is a member of the group working with the netsaint chief architect to develop the product. They have built a web portal interface to it to show the status of INFN Grid testbed 1. The server in Catania checks some 130 services on about 35 hosts. A short demo was given. In answer to a question, he admitted that the scheme required manual configuration of the display panel for each node and that this does not scale; study is underway to see if this could be handled automatically by self-discovery.

Grid Security Status and Issues (David Kelsey)

He explained the issues implicit in Grid activities and which bodies or people are concerned at which level. He described the GSI module of Globus and how it will be further developed in the near to medium term. He described how the CAS (Community Authorisation Service) should handle large user communities. EDG currently has 11 national certificate authorities with CNRS as the catch-all for users not covered. Much work is being applied in establishing mutual trust between the CAs.

PPDG is also using Globus GSI and two-way trust is being established. They are also using some EDG tools and it appears that GriPhyN and iVDGL may agree to use this work also. There are also plans to confirm that cross-Atlantic Grid traffic can be correctly handled from a security point of view. Open questions include how well the CA scheme chosen will scale up, how to separate authorisation and authentication and how to do real user authorisation (for example use Microsoft Passport or similar)?

He believes that if Grid computing is to take off, we need to

- Develop an acceptable usage policy to allow random users to run Grid jobs at random sites
- Understand how to manage virtual organisations of large user populations because site managements cannot negotiate individually with the numbers of users likely to be involved.

He described also the recent Web Security announcement from IBM, Microsoft and Verisign.

EDG WP4 (Jan Iven)

Last talk of the conference was a status report on WP4. Jan listed the challenges facing WP4, how to scale up to O(10K) nodes, how to offer provisions for running both Grid and local jobs. He covered the WP4 architecture and how it fits into the overall DataGrid architecture. The first testbed release is running on about 70 nodes across various sites with another 30 being installed these days. There have been several interim releases since the first official release in September. He described short-term plans for Testbed 2, splitting nodes into Production and Research, moving to Redhat 7.2, developing a new configuration scheme, developing a new quality-monitoring agent with a simple alarm display/GUI.

---

[6] See http://www.netsaint.org

Wrapup (Wolfgang Friebel)
Wojciech Wojcik of CCIN2P3 was appointed by the HEPiX/HEPNT Board as new European Coordinator to replace Wolfgang who has done the job for the last five years. Lisa Giacchetti remains US Coordinator. The Board has decided to remove the distinction between the two parts, HEPiX and HEPNT, better integrating the programme from the next meeting. Next meeting looks like it could be in FNAL in the autumn, perhaps mid-October, consecutive to the GGF in Chicago. A suggestion to register hepix.org; will be considered.