HEPNT and HEPiX Technical Trip Report

There was record attendance at this meeting (over 100), partly due to the joint NT/UNIX focus and partly due to SLAC staffers registering for only certain sessions, especially the one given by a renowned Windows consultant (see later). It is also worth noting that this was only the second formal open HEPNT meeting but it was judged sufficiently positively that it will undoubtedly not be the last. There was a heavy European presence from CERN, DESY and, especially, IN2P3.

The conference was held in the Panofsky Auditorium where there were literally dozens of power and Ethernet sockets available for attendees to plug in their own portables after pre-registration of their MAC address. The PCs could then fully access Internet but from outside SLAC's firewall. One could answer one's e-mail during boring talks (!) or watch the streaming video of the current talk instead of the live presenter.

NT at DESY

There is a single domain between the Hamburg and Zeuthen campuses with a Primary Domain Controller (PDC) and the main users' Home Directory server at the former and Backup Domain Controllers (BDCs) and WINS servers at both. There are some 2000 users in 55 groups with about 1400 active nodes simultaneously active typically. Certain tasks are delegated to the Group Administrators. They use separate DEC clusters for serving the applications and for the home directory servers and a Microsoft cluster as the print server, although this in fact merely spools output to a central UNIX-based print spooler. They will migrate at some point from the DEC cluster solution to the Microsoft-based one.

Viruscan from Network Associates is used; discs are scanned on file access.

They will investigate TSE but have not decided which access protocol to use, ICA, RDP or X. And they realise that "groupware is coming". There are no plans yet for Windows 2000 migration but some very preliminary tests have begun and they have a test domain.

NT Applications Support at DESY

As reported before (URL http://hepntdays.web.cern.ch/hepntdays/Programme.htm), this is based on Netinstall, a German product. Its major features include software on demand (it will fully install a product the first time the user tries to executes it) and local security management (it can execute privileged commands on behalf of the user). Currently, Netinstall is used to support some 50 packages for three distinct populations – Hamburg, Zeuthen and HERA Controls. DESY have noted that lots of effort is needed to adapt certain recent tools (such as Internet Explorer 4 and Office 2000) to Netinstall and that supporting different versions of the same tool can be difficult. They are planning to move to version 5 of this product which supports multiple databases and multiple servers; this new version should also allow them to support laptops. In the longer term (for example in the "era" of Windows 2000), they will investigate Microsoft's native Windows Installer.

DESY offer the user the choice of having his or her PC fully managed, partially managed or with no management support at all. As noted previously, most users choose the middle option. The central support group is aware of some level of distrust of themselves as a support team and they have established a structure for cooperation between the group administrators representing the users and themselves

Microsoft Terminal Server Edition (TSE)

RAL

RAL use TSE to make Office apps available for UNIX users. Typically they can accommodate 10-12 simultaneous users on a 450MHz Pentium II system with 384MB of memory. Based on this experience, they will soon expand the service with a dual CPU second server.

The speaker found the (Microsoft) applications easy to install if they were "TSE-ready" and almost impossible if not, for example some third-party apps. Sometimes an installation would corrupt TSE so badly that the whole server required to be re-installed! In operation, some apps crashed not infrequently, the main problems seeming to be related to temporary files and memory leaks.

In summary, he considered TSE to be non-trivial to setup and operate and to require lots of effort. But the service had satisfied the need for which it was planned.

For ISIS, RAL was now looking at acquiring Windows terminals and using them plus TSE for experiments. In regard to application licensing, there appeared to be some unresolved questions. Other problems found included the fact the UNIX users were often simply unused to the Microsoft way of working, where files were stored by default for example.

For the future, RAL is trying to find the best model to support clients which are basically similar except for one or two additional applications installed . For this they are designing what they call a "chubby client". They had rejecting

- Installing every app by hand
- cloning a full system image including all the possible required apps
- making use of a pure server/client solution for the extra apps.

Instead, they would be preparing an image containing all the core apps and use TSE for the extra apps only needed by a few stations.

Finally, they had performed load tests and their conclusions were that one should consider Microsoft's load figures a serious underestimate and that memory was usually the limiting factor.

SLAC

SLAC have 2 TSE-based services (they refer to TSE by its other name – Windows Terminal Service); one for general services and one for the Business Services Division. The latter was brought in in response to the major intrusion alert which SLAC suffered in 1998 in order to safeguard human resource data in their Peoplesoft and Oracle databases. The data servers are placed on a secure network protected by filtering routers while users remain on the normal network. Access is via encrypted applications using Citrix's Secure ICA protocol in combination with host access via secure shell. They are also implementing secure domain login using two-phase authentication including Secure ID cards.

It was a lot of work to configure Peoplesoft and other third-party apps for TSE and the use of Microsoft's security protocols is incompatible with cross-platform access – for example Samba no longer works. Even installing and configuring TSE itself was a lot of work.

The other application, for general use, is intended to provide access to Windows apps for non-Windows users. One warning from this experience is to avoid offering access to heavy graphics apps; Netmeeting was singled out in this respect. Benefitting from their campus-wide licence, they expect this usage to increase in the future. The speaker referred to a paper on the work – see SLAC Pub–8172 at http://www.slac.stanford.edu/pubs/fastfind.html,.

RAL NT Farm

Since the last meeting, this PC farm, used largely for batch simulation work, has been upgraded with 10 more dual Pentium systems, a 5-fold increase in capacity. It has been rearranged for easier management, moving all the peripherals on to one node. TSE is used to access it and Citrix's Metaframe product will be added plus the X11 access option. However, it still has limited appeal in terms of users (mainly LHCb) although the speaker was optimistic that LHCB would soon add yet more work there soon.

The systems are installed using the Imagecast disc imaging package to replicate the batch nodes. Particularly interesting in this package are the options for post-installation tailoring, for example to set unique IP names and addresses.

Code Development under NT

Tony Johnston of SLAC gave a very interesting talk on code development experiences under NT. The goal had been to judge UNIX/NT cross-platform conformity using C++, Fortran and Java as languages and Sourcesafe, CVS and a home-written tool called DEC as code management systems.

Using Windows development tools to develop Windows code was relatively simple using Microsoft's Visual Studio suite (including J++) and Compaq Fortran. Proper licensing was almost the only issue. But cross-platform C++ code was almost impossible – one needs a different make file for each platform (for example different compiler flags) and different platforms throw up different bugs.

Using Java was much better; the combination of J++ and Sourcesafe on NT produced code which was downloadable to both NT and UNIX, and also to MAC. It appeared that "write once, run anywhere" worked in this case, that is for use as an application language, not to write browser applets. Performance was judged to be adequate and improving all the time.

The team were now working on moving from Sourcesafe (Windows only) to CVS for which there were many tools available at SLAC and elsewhere. CVS also offered more web access features and the possibility of remote access; plus using CVS meant collaborating teams could use the same tool on UNIX.

Having satisfied themselves that Java plus CVS under Windows NT is a reasonable code development environment, they had set about solving the problems when using C++. Tony Waite of the SLAC Linear Detector (SLD) team had developed DEC – Development Environment Controller – on top of CVS. Basically this attempts to hide the differences of the C++ compiler on different platforms.

Using the above Java development environment, the team had experimented using NT as a data server. It had proved difficult because of deficiences in Microsoft server code – for example remote control plus some security issues in using Netmeeting's Remote Desktop feature. NT is also unstable when asked to serve as disk cache when clients demand large files. SLAC's conclusion is that NT is basically not yet ready to act as a data server. Neither had NT any particular advantage over UNIX as a batch server. And to make the choice even more difficult, UNIX with KDE and StarOffice is catching up with NT on the desktop.

In answer to a question, he claimed that although Microsoft may not produce a Java 1.2 environment, other suppliers certainly would.

SMS at SLAC

SLAC use SMS version 1.2 and are looking at moving to version 2. [They said they had looked at CERN's NICE/NT among other options when the choice was made in 1997 but had rejected it largely because it was IPX-based.] SLAC has a single master site-wide domain plus 12 local domains, a total of some 1400 NT nodes in total. There are two primary SMS sites looking after about 70% of SLACs NT population (9 of the domains, 1000 nodes). One SMS site is for the Business Services Division for its secured network (see above) and the other supports the rest of the site.

They use SMS to try to standardise hardware and software configurations but some groups still insist on selecting their own favourite hardware. Their attempt to produce a floppy to steer the user through a standard install script sounded remarkably like the Diane floppy of NICE. SMS is also used for

- Software distribution of standard packages like Netscape, AFS client, anti-virus software, etc as well as the usual Microsoft applications
- Inventory reporting
- Remote tools, for example to permit the Helpdesk to trouble-shoot user problems, to perform network monitoring, etc.

One problem they have seen is with dual-boot nodes (NT and Linux typically): often when SMS runs, the PC does not respond because it is running Linux at that moment or is switched off. SMS is also very complex to operate. Nevertheless, they are happy that they chose SMS; they believe it has decreased their Total Cost of Ownership, improved their response time (for example response to security incidents) and decreased the effect of human errors.

NICE/NT in INFN

Gian-Piero Siroli gave a description of the NICE port done in Bologna (largely) by Alberto Pace. Among the main changes were translation of the user interface into Italian and the use of NTFS instead of Novell as the filebase. He said some complementary things about the help fom CERN. Since the original port, the Italian team has worked to extend NICE to support multiple INFN sites via their WAN. There are 3 new INFN sites since the last meeting; now a total of 7 sites in a single domain covering Italy from the North to the very South with 58 nodes in total. Almost every site has its own BDC and home directory server.

They are now working on a model which will permit automated installation of a client over the WAN, on WAN-wide application server synchronisation and on some issues around security and accounting. In summary they are very positive with their experiences, praising in particular the robustness of NICE.

Support of Laptops at DESY

From the general IT service, no special service is offered for laptops. While SMS is used onsite, this is usually too slow from home or from a remote laptop so such systems become out of date and/or virus-infected. Within the PPD group, laptops on site are supported via DHCP from any networked port as long as the PC is registered to connect to the network. The PPD team has a pool of 5 laptops for borrowing and they reclone the disc for every new loan. The only problem they have met is in trying to get borrowers to return the laptops!

Windows Security

Virus Protection

Tami Kramer gave an expose on the theory of virus infection and detection. There are an estimated 45000 known viruses with more being born every month. She described the differences between encrypted and polymorphic viruses and the difficulties they cause in detection. She then described what steps are taken at CERN (by her in fact). The NICE architecture permits central coordination for scanning the clients as well as the servers. While CERN uses the Intel LANdesk product, FNAL and LBNL reported using the Norton product. FNAL use it only on the servers and leave users to decide their own policy.

Commercial Products

FNAL has investigated a couple of commercial security products.

1. SPQuery produces an inventory of software installed on NT clients (versions and patch levels) as well as helping with application downloads and subsequent updates. Although new products or patches can be “installed” remotely from the server, the actual modules are actually only installed when a user next logs in with Admin rights which could be a disadvantage in some environments. Also, the user can choose never to install a given product or patch and the server admin is not made aware of this until the next inventory listing is checked. Other disadvantages, at least in the current version, include the inability to ensure that a series of patches are applied in the correct order and the fact that it will only install a single application or patch at a time. The product costs $595 for a site licence and more information can be found at http://www.mtesoft.com. You can declare a “profile” including desired patch levels and measure that clients match such profiles.
2. STAT, which stands for Security Test and Analysis Tool, detects NT vulnerabilities. It can be set to close these automatically when found. It includes a password checker and it also can detect simple viruses and Trojan horses; if found, it gives instructions on how to clean these. Again it requires local admin rights to be run but it can be controlled from a central server. It checks for some 600 known vulnerabilities but it can be configured to check for any subset of these. More information is available at http://statonline.com and it costs $1797 per administrator licence plus $300 per year maintenance.

Windows 2000 Tests at DAPNIA

Today, DAPNIA uses NT4 but the speaker spoke of "anarchy" between various CEA departments with little or no coordination or cooperation; departments insist on keeping their own autonomy. DAPNIA has adopted CERN's NICE/NT model and have added a single Windows 2000 Professional node as a client.

In their tests they already appreciate W2000's plug and play features and its offline folders for mobile users with synchronisation of these to network shares. However, in their tests they have seen some errors in file permissions when sharing files. Using Release Candidate 1 they have even suffered lost files.

A popular feature is the multi-language package – one version of the system but users can choose the GUI language of their choice (except for the Login box which must be English!). In RC1, not all the menus have been translated into French yet.

Their impression is that fewer reboots are necessary – as advertised. The user interface closely resembles IE5 on top of NT4 but the management interface is very new – MMC for everything. They have seen no major problems with the principle applications except that they have not yet found a W2000-compatible anti-virus product.

NT4 and W2000 are able to share the same roaming user profiles although these are not identical and they have discovered that they cannot migrate French-language NT4 profiles to the W2000 multi-language profiles.

In the longer term, the speaker said DAPNIA would wait and see what CERN plans to do.

CERN's Plans for Windows

I presented Frederic Hemmer's talk as given to CERN's Desktop Forum on our plans to migrate NICE 95 towards NT and eventually to a Windows 2000 base. It was well received and a couple of people came up to me afterwards and said that they thought we were heading the correct direction.

Migrating to Windows 2000

This was a very interesting (and highly entertaining) presentation by a Californian consultant on some of the pitfalls and dangers involved in migrating to the forthcoming Windows 2000 release. For those interested, I have his handout and I would recommend his books on the subject.

He first covered the options for DNS. Although Microsoft recommend using their DNS service as the root of the local DNS tree, he advised strongly against replacing any existing DNS root but rather making the Microsoft DNS server as a subsidiary of that.

He considered Novell's directory services to be some 3 to 4 years ahead of Microsoft today but he expected Microsoft to catch up fast.

He spoke of rumours that Active Directory was being ported (by some part of Microsoft) to Solaris, HP-UX and maybe other UNIX flavours. Plus Cisco will work with Microsoft to port it to Cisco products.

Windows 2000 Panel

Dave Kelsey has obtained the agreement of the HEP-CCC sub-group, HTASC, to form a sub-group on Windows 2000 migration. He is now looking for someone to drive this. It should concern itself with possible HEP-wide implications of W2000 such as HEP-wide domains, the use of Kerberos for inter-site authentication, Active Directory as a HEP-wide directory, UNIX/Windows integration, etc.

Going round the labs to see who has done what with Windows 2000, RAL had effectively done nothing on W2000 other than a few tests of W2000 Professional RC2. SLAC had not done much either but were greatly concerned by what would happen in mixed mode (NT4 and W2000 present side by side) and by the possible implications of domain structuring. They would wait until at least the first Service Pack – mid 2000?

On the other hand, a speaker from Stanford University (a non-HEP site) described some 2 years of preparatory work – an integration team, application testing (lots of problems, especially with TCP/IP applications and AFS). They plan to link Microsoft's DNS and Kerberos 5 services with their existing authentication scheme, which includes AFS authentication.

Christian Trachimow at DESY has experimented with W2000 DNS and is very knowledgeable about its features and defects. But there are no concrete plans at DESY yet to make use of it. FNAL has been concentrating on firefighting NT4 issues (Y2K and their mailserver mainly) and going on courses to prepare for W2000. However they have recently created a migration team and plan a test domain shortly. Until now they have been running only a few desktop systems on RC2.

It would appear that, among the HEP sites represented, CERN is the furthest advanced in its planning and perhaps also in its testing. Elsewhere, efforts seem to be limited to particular individuals with interests in various aspects of W2000 and in preparation for the coming migration.

SLAC Overview

The head of the SLAC Computer Centre, Richard Mount, gave an overview of the work there and of SLAC activities. Obviously there was a lot of talk about BaBar which was performing very well. On the BaBar computing side, Richard said that "Objectivity was proving as exciting as expected". Two other quotes from his talk were that "Linux was almost under control" and "Windows 2000 offers opportunities for both rationalisation and chaos". In more detail, he wonders how fast SLAC must invest in W2000 effort and how much. What should be the best architecture to deliver this. For Linux, how "personal" should the systems be? Will large SMPs have a role (like SLAC's large Enterprise 64 node SUN system for BaBar). And how to react to the storage challenge created when users can afford to go to the local computer store for ultra-cheap disc and tapes.

SLAC is participating on the US Particle Physics Data Grid proposal, a collaboration of 9 labs and institutes from all major US HEP sites. Currently they were negotiating with the DoE for funding.

The conference participants were later given a tour of the SLAC Computer Centre and some of the experimental areas. The Centre has a huge SUN population – racks of Ultra 5s, clusters of 450 Servers, an Enterprise 10000 and a 7TB wall of A1000 RAID discs used for the Objectivity data cache. There are 7 STK silos with a mixture of Eagle and Redwood drives; the 12 Eagle drives are dedicated to HPSS.

LDAP at IN2P3

IN2P3 has chosen Netscape to provide the LDAP server to link its various sites. The IN2P3 LDAP service offers mail address lookup, roaming user support (storing the address profiles, bookmarks, etc) and a telephone directory. Users are able to examine and modify online their own data and the speaker gave an online demonstration of this. The work was programmed in Java and Javascript

The team at CCIN2P3 in Lyon are collaborating with CERN and others on a HEP-wide directory. Of course there are various technical and data format problems to solve but the largest hurdle is of a more political nature – can all the HEP labs agree on the principle and form of such a HEP-wide directory. For example a common naming scheme must be adopted. Partial replication must also be envisaged since not all the information should be available HEP-wide.

Administration Tasks

Another attempt, at Zeuthen this time, to define a common tool to perform common or related tasks across UNIX and NT systems, with the possibility to insert HEP-specific features such as managing AFS volumes and servers and common UNIX/NT passwords. The well-known commercial tools are too expensive and overkill for small sites (and the various public domain tools too "not-invented-here to be adapted for local use?).

The proposed tool is christened VAMOS (Versatile Admin tool for Multiple O/S). The intention is to build a common framework and then insert tools one by one but with a consistent underlying mechanism. It must provide a web interface and access to databases, to LDAP services and to authentication services. It should be modular, object-based, three-tier, etc (all the current popular buzz words). All it requires now is some manpower! Without some resource allocation, plans are moving ahead only very slowly.

Project GENTES at CASPUR

The goal is to provide common UNIX and NT authentication for login and common file access, to provide NT access without having an NT account. It is based on AFS file space. Gary Dobbins at the University of Notre Dame in the US replaced NT authentication by AFS authentication by making use of Microsoft's login extension mechanism (the so-called MS-GINA). CASPUR has added mapping to AFS file space and support for roaming user profiles. They have also converted the package to use InstallShield. However, for sites wishing to adopt this, they must first obtain the permission of the original package's author at Notre Dame, even to use the CASPUR version. Also, an AFS/NT client licence is required.

Firewall at RAL

In response to repeated probes and attacks and after a great deal of effort, RAL has defined and implemented a security plan which includes provision of an Internet firewall. Unfortunately, a commercial solution was judged too costly and resource-intensive. So for the moment they are limited to logging systematically all access from offsite and manually blocking the sources of probes. A filter is used in the external router to block access to a range of onsite nodes. For those systems externally accessible, precautions such as using a bastion host instead of direct telnet (even from onsite) are in place.

It is currently impossible to properly monitor UDP connections so, by default, UDP access will be blocked for internal nodes and RAL only permits access to certain IP ports on externally-accessible nodes.

In line with the virtual ban on even internal use of telnet, the use of ssh will be encouraged instead.

LDAP Birds Of a Feather

In one of the lunch time breaks, an LDAP BOF session was held on how to advance LDAP services across HEP. Various ideas which were floated including: holding meetings of specialists; scheduling video-conferences; and creating mailing lists. In a roundup of the labs, FNAL said they had a local LDAP service in place for e-mail addresses but were reluctant to make this available outside the lab because of privacy concerns. SLAC agreed and so did JLAB. Rather than a firm US decision, this appeared to be more an understandable cultural attitude against making private data public.

Password Policies

There was a tour by lab of password policies in place.

SLAC attempts to forbid clear text passwords even within the site by forcing the use of ssh instead of telnet and rlogin. However, ftp is still partially permitted although not available on all platforms.

FNAL have made a detailed review of secure and less secure authentication schemes. An interesting matrix of the various dangers and risks was shown. The conclusion is that a policy of re-usable passwords should

be dropped; all other methods may not be perfect but they are all better. And host security is still the single most vital aspect. FNAL has a Kerberos project to try to avoid future mass password compromises (they have recently suffered two attacks where more than 300 users were affected each time). The aim is a single sign-on and centralised account administration. They plan to create realms of nodes with various levels of security/trust (so far only a single realm with about 10 nodes, used mainly for software development) only accessible via Kerberos credits. Kerberos 5 is used along with code from Argonne and NRI to fabricate valid Kerberos 4 tickets which can then be used for AFS access. This appears to work and be usable as a stop-gap until Transarc adopt Kerberos 5 into AFS. This implementation will be modified to cope with batch jobs as was done for AFS authentication. Their latest implementation of CVS supports Kerberos 5.

JLAB have very few systems open to offsite access. They are continually promoting the use of ssh. They also recommend the use of IMAP plus SSL, for example with Netscape and Outlook as mail clients. They expect soon to be able to remove all use of telnet, including internally, and POP mail access. They have created a so-called DMZ (de-militarised zone) outside their firewall where they can isolate certain accesses and monitor more closely a limited number of nodes. They are looking at certificates, where Globus/GSI from Argonne is an interesting candidate and Kerberos is another.

DESY have already Kerberised everything except mail and are now about to add this, in particular IMAP authentication using Kerberos 4. They have chosen to add SSL to the University of Washington IMAP server. Although Pine comes with in-built Kerberos support, no other mail client does so and so DESY have had to add SSL encryption to the whole mail session when using clients such as Netscape or Outlook – but then Pine needs its own wrapper as it does not support SSL. DESY plans to offer two servers – one with Kerberos and one with SSL. This is now under test at Zeuthen and should be generally released shortly.

Site Reports

SLAC

The UNIX population was audited at 748 SUNs (steadily increasing), 114 AIX systems (nothing new for some time, now a candidate for phasing out) and 130 Linux PCs, also growing rapidly. AFS 3.5 is used everywhere. Implementing it on the servers showed up some early problems but it is now stable thanks to Transarc's assistance. This version shows noticeable performance improvements except on SMP systems where there are still problems with the client on the SUN Enterprise 10000 with its 64 CPUs.

HPSS is only being used for BaBar tape management, not for disc pool management. Elsewhere SHIFT is still being used for general staging needs for the moment. The SUN Autoclient package is being used to manage some 320 SUN system images, with one server per approximately 50 clients. The client boots across the network and some partitions are mounted off the server disc. This gives fast recovery from crashes but creates a single point of failure per 50 clients and they are only able to boot about 4 to 6 clients per server at a time.

DESY

Linux and Solaris are the growing UNIX platforms at DESY; the rest have stabilised in numbers except that AIX support at Zeuthen will end in December '99, leaving Solaris as the main server platform there. Updates of their SuSE Linux to the 6.1 release with the Linux 2.2 kernel have begun. n Hamburg, the HP login cluster is likely to be replaced by one running Linux. Structured cabling is being adopted. They have appointed a Security Officer. Zeuthen have adopted lprng as the printing mechanism and Hamburg are likely to follow suit, Finally, Zeuthen has tested LDAP as a replacement for NIS and DNS.

IN2P3

The BASTIA and ANASTASI farms have been merged into a single 120 station cluster consisting of 4 architecture platforms. A new version of BQS was required to support this, including options for jobs to be declared I/O-bound or CPU-bound. The ATM network is being replaced by 100Mb Ethernet for the interactive and batch servers and by Gigabit Ethernet for the disc and Objectivity servers. A 1TB disc VSS (Versatile Storage System) subsystem from IBM has been introduced. There is a proposal to upgrade their

IBM batch capacity with B50 processors and Model 43Ps are being tested as possible tape servers. With the French collaborators of D0 they will experiment for the first time with Linux systems.

Helpdesk problems are being recorded by (yet-another) home-grown problem tracking system (or is it a an offshoot of an earlier-presented development?).

LAL
There is almost Linux present at LAL so far, neither desktops nor server. On the other hand more and more physicists use NT as a desktop system. They are working on a pilot SMS service, today using version 1.2 but moving soon to version 2, despite the fact that the newer version drops MAC support and LAL has a more or less constant MAC population since existing users tend to renew their MACs at end of life. Wincenter is in use and is a popular service. They strongly support the HEP LDAP initiative described earlier.

RAL
RAL has formed an internal User Group to share Linux experiences. With 20 new dual-processor Pentium II 450 MHz systems running Redhat 5.2 and the ARLA AFS port, Linux now provides more CPU horsepower at RAL for HEP than any other platform. And there are plans to expand this service further, including adding more disc and tape devices.

JLAB
Linux use is still expanding with now more than 75 nodes, spread between various compute farms running Redhat 5.2 and desktops running version 6.0. However, the JLAB CAD group purchased new HPs recently and the High Performance Group has some 20 Compaq Alpha systems running Redhat Linux 5.2 in a collaboration with MIT and Wuppertal. There is now a daily rate through their tape silo of some 1TB, representing about 700 tape mounts per day. They will soon double their compute farm with another 50 nodes, up to 5000 SPECint units, probably all Linux-based. The JLAB backbone network is about to move to Gigabit and by now some 70% of the lab uses switched networking. Recent software purchases include Netscape Calendar, Microsoft SMS and Vmware for NT access from Linux desktops. Current outstanding issues include Linux desktop access to home directories (Samba?), LDAP (to solve the NIS Linux problem and for roaming profiles and HEP directory) and NT/UNIX co-existance.

BNL
RHIC commissioning had begun and the computer centre were seeking to build up to a Teraflop over the coming years with the aid of outside funding agencies. There were currently lots of Linux boxes on site used in farms with, a mixture of Redhat 5.2 and 6.0; the total was probably near 700 CPUs. There are also still lots of SUNs present. A major change over the last year has been a move away from their home-grown DQS batch queuing scheme towards LSF.

An Alternative to LSF

Ian Bird of JLAB discussed some alternatives for LSF which he deemed rather expensive, especially for collaborations requiring offsite job submission. The alternatives ranged from a layer on top of LSF to a complete replacement, for example the Portable Batch System (PBS) from NASA Ames. The latter was preferred but it would need a lot of work, especially for example on the rather simplistic job scheduler currently in PBS.

HEP Linux Project

There was a brief discussion on whether the time was right for such a project. Is there an opportunity for some common development or planning? The results of the debate were inconclusive but Chuck Boeheim will set up a mailing list for those interested, in particular so that the Linux administrators at the different sites can share their experiences.

Helpdesk System at IN2P3

A project was described to write a sinple trouble ticket scheme to be used in user support at IN2P3. It consisted of some 15K lines of code and had taken approximately 5 man-months. It was based on e-mail input but was also accessible via the web. The key components were CGI scripts, Javascript, an HTML user interface and an Oracle database behind.

Users send e-mail to a generic address and a manager assigns incoming problems to an engineer who receives notification that there is a problem to investigate. It accepts multiple languages and offers full-text searching and has access rights where necessary. Still missing are the ability to handle multi-part mails, encoded mail attachments and statistics. It can be tried at Http://UserSupport.in2p3.fr/demo/ and the code is avaibale on demand from the presenter (Jean-René Rouet at IN2P3 Computing Center, rouet@in2p3.fr)

HEPiX X11 Project Proposal

This was a second attempt by the UNIX team at CERN, led by Philippe Defert, to provoke some discussion and possibly some action by the HEP UNIX community in the area of GUIs and in particular in the debate between KDE and Gnome. In the previous meeting, CERN had presented a comparison of both products. KDE, originating in Norway, is not 100% free software while Gnome, which started as a reaction to that fact, is completely open source. Both provide varying degrees of Microsoft “look and feel”. Many features are now agreed to be common between them but not all; for example they use different IDLs for interfacing to CORBA, they have different print and help interfaces and the directory layouts are not the same.

However, in the discussion, there was no real consensus for a joint HEP study or even who should be consulted (for example, should we bring in the experiments since they will be the main developers?). Many applications may in fact inter-operate no matter which path is chosen. Unfortunately, it appears that many labs have not yet thought about the problem.

Linux Tools at FNAL

Linux is being used on central farms, on level-3 trigger systems, on Enstore tape movers, DAQ tests and, by far the largest use, on desktops. Like CERN there is a large degree of user self-help via open mailing lists. Updates are made optionally available by the central support team from a central AutoRPM service with nightly updates and recommendations, especially for security updates.

Among the tools in use or under consideration are the following:

- System monitoring tools from the Beowolf project are being looked at
- System installation and update, network install and AutoRPM
- Backup and failure recovery, for example systracker, an internal project: this should provide an easy method to prevent unnecessary differences from some defined standard configuration. It has been implemented via PERL scripts and was now in Alpha testing.
- Resource planning and accounting – LSF and accounting scripts

In general (in FNAL, all DoE sites, HEP in general) there are many parallel activities in this area with little central coordination. Some DoE and some FNAL working groups are being formed to look at various specific topics. For example, as mentioned in a previous session, it will be essential soon to decide between KDE and Gnome. Also urgent is some central file backup and archive service. Clearly among all these aspects there is scope for joint efforts.

It was noted during this session that FNAL had dropped its pilot formal support contract with Redhat since this had proved ineffective but also unnecessary; informal contacts would continue in addition to the usual public domain mailing groups.

RAL Data Centre for BaBar

RAL has recently established a local data centre for the BaBar collaboration. The speaker noted that, just as HEP-CCC was recommending PC-only farms, the new centre was part of $1.2M worth of SUN equipment and associated infrastructure being purchased for RAL and several University groups working on BaBar. The main reason of course was the need to run BaBar software which was not yet certified on Linux although this was expected fairly soon. So far, some 5.5M events had been transferred to the RAL data store. The remote UK groups will be connected to RAL via JANET

Data Storage

FNAL

FNAL are collaborating with DESY on Enstore including some short-term staff exchanges to work on a common disk buffering solution. FNAL are currently deploying version 1 and debugging it. Stability is now thought to be good enough to permit first use by D0 who have 5 tape drives dedicated to this service today and will increase this to 10 drives shortly.

HPSS is in limited use today for file transfer for a DoE Next Generation Internet project.

There is no decision yet on the next tape drive architecture but Exabyte's Mammoth 2 is the current favourite.

DESY

There was a status report on Eurostore: 33 man-years, 3 MECU (about $3M), started spring 1998 with 24 months duration. Current results are that, with low latency, speeds up to 300 MB/sec have been achieved for parallel file system access. It will now be extended to support file migration with a view to implementing an HSM scheme using a Java user interface. DESY has been forced to drop some planned features because of the loss of several key project staff. Nevertheless they are now working on defining a follow-up project, hopefully involving more partners, even perhaps some non-European partners.

CASTOR Project at CERN

CASTOR stands for Cern Advanced STORage manager. In the short term it is planned to use it for the NA48 and COMPASS experiments and in the longer term for LHC data handling. It will be extremely modular permitting easy replacement of components as better alternatives appear. It will work with both NT and UNIX and it will be independent of any particular media technology. It is backwards compatable with CERN's SHIFT software. Threads are used to improve I/O performance and the use of a database to store CASTOR data gives scalability. Of course it includes a basic HSM functionality.

HPSS

CERN

There was a review of the history of HPSS at CERN: test installation on AIX in October 1997, then ported to Digital UNIX and RFIO added in 1998. The current plan is to use HPSS for file sizes ranging from 20MB to 10GB. Smaller files should be stored in AFS, if necessary hidden behind an HSM user interface. In addition, HPSS would be used for data taking by some experiments, notably NA57 and ATLAS and in CMS test beams. ALICE will also use it for their mock data challenge. Today there are some 10TB of HPSS data of which 1TB are medium-sized files averaging 60MB each. The raw data files for the experiments are of course somewhat larger. Tape mounts per day under HPSS range from 100 to 800.

The service is considered successful but the feeling is that it is too soon to commit to HPSS. Use will be limited for now in case they feel the need later to move to a different platform. For one thing, the Compaq port is not complete and nor is the Solaris port. A Linux port is simply missing although almost all the other tape servers in CERN's Computing Centre run Linux. The talk finished with a Wish List, including a request for a Linux port.

BaBar
They run HPSS version 4.1. The hardware dedicated to HPSS consists of an RS/6000 model F50 core server with 4 CPUs; there are 12 Eagle (STK 9840 tape drives and 12TB of disc space controlled by Solaris tape and disc servers. On these, BaBar run their own port of the relevant HPSS modules although IBM have now ported a more recent version of HPSS to Solaris. They find HPSS very stable; even the tape media have proved quite reliable once some initial problems were solved. They expect to expand the disc space by about 1TB per month up to about 25TB. And a doubling of the Eagle drive population is also on the cards.

BNL
They use HPSS for RHIC raw data, at a rate of 50 MBps. They have a mixture of Redwood and Eagle tape drives in a Powderhorn silo and some 8TB of disc space of which 1TB is HPSS cache. Five RS/6000s are used for HPSS (core server and data movers) as well as 6 SUNs for file serving and data mining. They have measured throughput up to 80 MBps on their disc subsystem using 2 Fibre Channel connections.

Next Meetings

The next HEPiX meeting will be in Braunschweig in April. Due to the short lapse in time until then, we will not hold a HEPiX in conjunction with CHEP. However it was suggested that HEPiX may wish to use the opportunity of CHEP to sponsor one or two sessions where they could stimulate a discussion in the physics community on some matters of topical interest.

HEPNT will probably decide for another joint meeting in the April planning but the total length should be limited to 4 days.

Alan Silverman
30th March 2000