# HEPiX Report, FNAL, August 26-28th 1998
## (Also some notes from CHEP'98)

The meeting was attended by some 60 people, 50% Fermi staff, 10% other US sites and 40% European. On a site-wide basis, there were 11 European sites represented and 7 US sites. The meeting extended over 3 days with a supplementary day devoted solely to HPSS. [I did not attend this last day so it is not reported here.]

### Physics at FNAL – Steve Wolbers

The meeting opened with a review of current activities at FNAL. A lot of effort is of course being put on the build-up towards Run II. The new Main Injector was expected to receive its first beam in a few weeks time. Not to be confused, the original Mail Ring was sited above the original Tevatron but has now been removed.

In 1999 there will be a fixed target run at 800Gev for 3 main experiments, 2 of which will produce massive amounts of data. In 2000, RUN II will start, for the first time a 1Tev collider. Its firm start date depends to some extent on the successful development of silicon strip detectors for the central cores of both experiments. It will have raised luminosity by a factor of 20 compared to its predecesor. This will have major impact on the required computing power and on the complexity, requiring new paradigms. Data rates for both D0 and CDF are expected to top 10MBps. In preparation for RUN II there are some 15 computing projects, including a large Linux farm.

Beyond RUN II there are plans for experiments concerning neutrino oscillations, for example long beamline experiments up to 700 Km in length (MINOS and BooNE). There is also a B physics proposal (Btev).

FNAL is also still involved with providing computing power for tbe Sloan Digital Sky Survey, a 5 year experiment whose telescope is in Mexico. In a similar field, FNAL is involved with both the Auger project to measure cosmic ray decays and in a Cold Dark Matter Search. The Auger project, whose detector would lie in Argentina, is nearly approved.

## SITE REPORTS

**DESY Zeuthen**

Increasing importance of AFS: they have asked the supplier (Genias) of their (local) batch system (CODINE) for AFS support; also asked Legato (authors of Networker) but no reply after three months. They looked at Remedy but considered it too heavy for them.

**DESY Hamburg**

More and more AFS users; more (SUN) servers and more (Comparex RAID5) discs. All work on DCE/DFS stopped, except for some tests on Linux. They have stopped buying

HP and SGI systems, having already halted DEC purchases. Current acquisitions are only SUNs and PCs. They have no plans for HP-UX 11 but on the other hand, Linux support is now official, including installations on desktops.

They plan to use OSM as their data management utility up to 2001 but they are starting now to look for a replacement for their Grau tape robot. Meanwhile, more staff have been added to the Eurostore Project and a first prototype is targetted for March 1999.

**RAL**

They added more HP capacity in the form of C200 series systems for batch. The migration to LSF however is delayed. The HPs gave disappointing performance compared to the expected (unless codes were specifically recompiled for that model) so they are now moving to PCs running Linux and they will try to integrate these with their farms and LSF.

For BaBar, they now have to invest in an Alpha server until such times as BaBar again supports the HP architecture, if ever.

Like DESY et al, they spend a lot of time on security protection, including anti-scanner protection. John Gordon proposed that perhaps the HEPiX Security Working Group could be reanimated, if only merely to inform each other of attacks in progress.

**GSI**

They support Windows-NT, AIX, Linux and VMS. AIX is for general purpose batch as is Linux; Linux is also used extensively on desktops. They are just starting Linux farming, with LSF. But they use Devian Linux centrally while most users have SuSe. The farm PCs are mostly dual-CPU Pentium Pros but next purchase will be Pentium IIs

**NIKHEF**

All architectures are present at NIKHEF, in the past 18 months they have purchased SUN, SGI, MAC, PCs (lots of Dells). Only 17% of the PCs run Linux, 50% run NT and the rest W-95. All the main central servers are still SUNs. For software installation on UNIX they use the vendor kit supplemented by rdist and ASIS.

**Jefferson Lab**

They use a Network Appliance file server where UNIX access is via NFS and Windows NT access is via CIFS. Jlab is an OSM user but they need to upgrade for Y2K and are uncertain about OSM futures. CA are rumoured to be about to integrate OSM into the Cheyenne suite.

JLAB are buying no new HP or AIX systems and probably will not do HP-UX 11.

For user dial-in, they use a CISCO Remote Access server for user authentication for the NT Domain. Like many other sites, they use ssh for secure remote access.

**CCIN2P3**

After all the security scares in HEP, CCIN2P3 are close to securing permission to use ssh.

The IBM mainframe at the Centre died in mid-August, prematurely ending that service. On the new service there are now 12 Redwood cartridge drives making some 1100 mounts per month. They are considering adding a SUN to the BAHIA interactive service (presently only HP and IBM); this would be for BaBar use. ROOT has been installed for the nuclear physicists.

**Saclay**

They have added a SUN to the public service as well as upgraded the Alphas. X Terminals will be replaced by NT PCs running Exceed.

**BNL/RHIC**

They operate a so-called Managed Data Server – basically an STK silo plus HPSS with IBM mover nodes although they hope to move to SUNs when the SLAC port is complete. They will store 1PB per year, starting in 1999. RHIC reconstruction is based on a Linux farm and another Linux farm is used for analysis. They use dual, later quad, CPU PCs. Today they use Redhat 5.0 and clone system discs over the farms; this greatly speeds installation. They are migrating to an ssh solution over the coming months for external access.

**FNAL**

A prototype Linux farm has been setup, installed by an outside contractor, and they are rapidly decommissioning old UNIX farm nodes more than 2 years old.

Deployment of the new UPD/UPS is going well – see later. The old UPS server should be switched off in September/October.

They are investigating both the MIT and ARLA AFS clients for Linux. They are implementing IMAP and LDAP servers from Netscape on NT and are meeting a few problems.

**CERN**

CERN reported the growing popularity of Linux, including the appearance of production Linux batch farms. The AFS service was also steadily growing but work on DFS had stopped, some base DCE service being offered only for the HPSS service. It had been agreed by the users (even suggested originally by them) to freeze HP-UX at version 10.20, including Year 2000 patches, and not to update the base to version 11. The operating system support group was producing a set of web pages listing those versions of the various O/S's which offered Year 2000 compatibility according to the vendors themselves.

CERN had evaluated and decided to adopt Remedy for problem tracking, eventually replacing their Gnats-based scheme previously described at HEPiX.

## UPS/UPD Updated at FNAL

Although U refers to UNIX, one major update in the new suite is the addition of NT support. This support is provided through the cygwin interface. For the UPS new version, an attempt was made to use PERL but the script startup overhead was too large and C was used instead. For UPD however, PERL was used.

Improvements to UPD included an uninstall feature, a factor of 12 improvement in speed and knowledge of product dependencies. It uses tables to select product locations and product developers only require to define a single table now instead of 4 in the previous version. The new package contains a conversion tool for existing systems.

The support group organised a large marketing effort to introduce and have accepted the new version and now consider it to be a success. In reviewing the project, they believe they have defined a software process which works –

- design time is both valuable and vital
- prototypes should be developed for testing and evaluation
- review with users at an early stage is needed, provides valuable feedback
- extensive collaboration with users is necessary
- user information and user assistance are always needed

## Proposal for LDAP Sharing

The proposal from CERN for passing LDAP information along a chain of labs (rather than globalling broadcasting local LDAP databases to every other lab) was generally welcomed. Now someone needs to follow up to get it started. W.Friebel expressed concerns about the speed of access claiming it takes 10 seconds today inside the DESY domain for an LDAP lookup. [After the meeting this was traced to a local config problem.]

## HEPNT

Dave Kelsey's report. See URL http://hepnts1.rl.ac.uk/hepnt/hepix/fnal/index.htm. There have been 5 meetings now, largely exchange of experiences. Comparing statistics among the represented sites, the split between Windows NT and Windows is approximately 50/50 with some labs showing a clear preference for one or the other. HEPNT is organising a three day open workshop at CERN for early December. See their web pages for details – http://hepnts1.rl.ac.uk/hepnt/.

## NT at DESY

DESY uses a single NT domain across both sites, Hamburg and Zeuthen with a flat NT4 name space. There are 40 groups. Commercial tools are used where possible, e.g. Netinstall (a German product) for application support, TEM for user account administration, etc. There are some 800 registered clients, 600 of them active. Netinstall is used on 260 of these clients.

PCs are categorised by support level.

- Green PCs are the easiest to set-up for the user but the hardest to debug since Netinstall turns off so many "dangerous" local tools to keep them out of the hands of users. There are some 60 green PCs
- So most users choose Yellow PCs where they use Netinstall to get the applications but retain control of the Administrator account and hence of the PC. There are some 200 yellow PCs

- Red PC owners “do their own thing”. There are some 300 red PCs.

DESY has decided not to use Microsoft Exchange but rather Uni Wisconsin IMAP server. They are looking at PMDF.

They will automate many tasks such as account creation, file load balancing, etc. They will use Microsoft’s Transaction Server and the Internet Information Server (IIS) for this, permitting access via the web as well as direct programmed access and from scripts. Via IIS they will use SSL for security. They will use DCOM to the Transaction Server. Like NICE, they support roaming profiles; applications are launched via an applications launcher which checks what if anything needs to be downloaded to the local disc. They also have an application which runs privileged tasks on behalf of a user (including application installations).

They have just started to look at NT 5 and plan a test domain. They will collect the requirements and prepare a task list.

## NT Backup at FNAL

There are some 30 NT servers, both Intel and Alpha servers, 240 GB of disc space. Only servers are backed up, not workstations. Initially, they tried Cheyenne ARCserve on DLT 4000 robots with the aim of keeping a 6 week cycle.

However, after some poor experiences, similar to those seen at CERN, they have switched to Seagate Backup Exec running on DLT 7000 robots from Breeze Hill. In particular ARCserve was found to be unreliable and technical support was poor. Hence they switched (back) to version 7 of the Seagate product even though some features were inferior. On the other hand, performance is slightly better, except for small files such as e-mails.

## NT Farms at RAL

In Phase I, they installed 3 Pentium Pro systems, early 1997. In Phase II, end ’97, they added 6 more CPUs to make a total of 11 CPUs. In the recent Phase III upgrade they have added more memory and fast Ethernet.

The servers run NT4 except for a front-end which runs 3.51 for Wincenter. LSF is the batch system used.

The results prove that NT farms are a viable option and RAL is ready to implement them when requested.

## Linux at DESY Hamburg

At DESY, the Linux used is SuSE 5.1 based on the Linux 2.0.33 kernel. It is network-installed where the user is not required to intervene. They hope soon to move to SuSE 5.3 which is based on the 2.0.35 kernel.

Initially Linux was only used for workgroup servers but the lab is now preparing to support it also on desktops, although only for the standard models recommended at

DESY. In the future they expect to move to a Linux 2.2 kernel version and add also AFS support. Dual-boot is not supported.

## Linux at DESY Zeuthen

At Zeuthen, only Linux desktops are supported centrally. Users are allowed to tailor and administer their own nodes but in order to permit access from off-site certain rules must be followed including –

- an agreed configuration
- use of ssh only from off-site for login
- the root password must be given to central support

In general, Zeuthen uses the same installation tools as Hamburg but they use their local installation methods (SUE + cfengine – see below) for post-installation and subsequent updates.

## Linux at FNAL

Linux has been officially supported at FNAL since January '98. They adopted Redhat 5.0 from the start, on Intel architectures only and with no support for portables. They are currently adding support for Linux in UPS/UPD. Since Jan, the number of Linux systems has risen to some 150 desktops, 40 production farm nodes and 40 nodes in an analysis cluster. More are planned over the coming months. Most desktops are self-installed by the users but CD-ROMs are available.

They support dual-boot with NT but this has to be enabled "by hand" since Redhat does not do it by default.

Work is in progress on a Redhat 5.1 distribution and they are looking at a scheme for automatic enhancements; SUE is one option for this, cfengine is another.

FNAL recommend to their users to purchase the last-but-one PC technology as there is typically a 3 month delay in getting the drivers for Linux for new hardware.

They have a support contract with Redhat for system administrators where they are allowed 5 on-site contacts. This is only on trial and has been little used so far. They will soon offer training classes for administrators. The central FNAL Linux support team is 1.5 FTEs.

## Linux at Jefferson Lab

Although the amount of data being produced is about what was expected, the lab has felt the need to expand rapidly their CPU capacity and they chose to do this with a batch farm of Linux PCs. They have chosen dual 400MHz Pentium II systems with LSF as the batch queuing scheme. [In passing, the speaker remarked that in his understanding, Platform were planning their next release of LSF based on the Linux 2.2 kernel.] One serious problem seen on the dual CPU systems concerned the Ethernet card and the initial 3COM model was switched for an IntelExpress board; still not all problems have been finally resolved though.

JLAB use their standard UNIX environment on Linux. For Fortran they use the Absoft compiler. They offer limited support for Linux desktops, mostly simply gateway access to the CIFS filebase apart from the normal NFS access to UNIX files. Following a major security incident, all systems must register their root passwords with central support.

## SUE + cfengine Development at DESY Zeuthen

A joint study was organised by the two DESY labs on system installation tools and the result, although only implemented (so far) by Zeuthen, was a decision to adapt CERN's SUE tool. SUE is used to decide when and where to install modules. A database of all supported nodes lists what to install and the public domain tool cfengine describes how to install them. For all platforms there is a single cfengine script per SUE feature (time setting, sendmail config, etc).

So far, the scheme is operational on three architectures although not all the cfengine scripts are ready.

There was a discussion on the scaling of the number of nodes which could affect the performance, perhaps even the integrity of the node configuration database. At the "other end" of the scheme, there was also a question of the decision to use a single script per feature across all supported platforms. It certainly reduces the number of scripts but it makes each of them more complicated and more liable for interference when any one architecture undergoes any minor or major upgrade.

## CUE at JLAB

Since the last HEPiX, JLAB have added NT support to CUE which is now rechristened Common User Environment (was U for UNIX before). So far, CUE does not yet include support for Linux desktops although all the needed software is made available to users on a self-service basis.

JLAB uses a single NIS domain with access by individual systems or clusters controlled by Net Groups. On NT there is a single master domain with several resource domains.

JLAB has contracts with Dell and Gateway (at least their local suppliers) for standard configurations. Their Network Appliance file server offers a common file storage between both "worlds" although users today require two passwords still. There is a manual scheme to keep users and groups synchronised between NT and UNIX but they hope to automate this.

## Trouble Ticket Scheme at DESY Zeuthen

Whereas DESY Hamburg use the Remedy product for problem tracking, the Zeuthen lab considered this tool too heavy for a smaller site. They looked at various public domain tools such as Gnats and req and chose the latter as a better fit to their perceived needs. Req is effectively a mail filter. Mail sent to the local user support account is acknowledged to the sender and stored in a repository. The support staff are provided with PERL scripts to process requests. Included are report tools and search tools. There are graphical interfaces, web interfaces, e-mail interfaces and a command line interface.

At the moment, they do not split the incoming requests into separate categories and there is no internal organisation in the repository. The scheme required a lot of work for local tailoring but they hoped this would be a one-time effort. The current use is some 5 incoming requests per day.

Req is also used at Fermilab in some groups although the central desktop service there uses Remedy.

## Year 2000

Year 2000 in fact consists of three main issues –

1. Year 2000 itself
2. The fact that 2000 is a leap year
3. The fear that some codes may have set the date 9/9/99 as a limit in tests

Among the many anecdotes around 2000 is one where a large US airline has stated that it will not operate its aircraft on Jan 1st 1999,

On MACs, the basic hardware and software is and always has been Year 2000 compatable. Of course it is impossible to know *a priori* what users have actually programmed in their code.

In PCs, the main issues are with the BIOS. The RTC (date/time chip) actually uses a 2 digit field for the year but the newer BIOS's also have a century digit and many older BIOS's can be upgraded. Going up one level to DOS, the DOS date command is able to cope with the year rollover even with older BIOS's on those systems dating from about 1997 onwards. In fact, as a general rule –

- For BIOS's pre-'97 there is probably no rollover to 2000
- For BIOS's issued in '97, some 50% might be expected to rollover
- From '98 onwards, BIOS's should rollover correctly

  Of course the check is simple – set the clock to sometime in December 1999 and wait. But watch for side effects on files, on any systems attached via a network, etc. There are various software checking products on the market, for example see WWW.NTSL.COM.

  Regarding Microsoft's own applications, use the Windows Control Panel to set the year field to use 4 digits and not 2 because all files are created and listed according to that setting.

  There is a programme from the Greenwich Meantime company called Check2000 which audits all programmes on disk against a list of known risky versions. Also Viasoft's ASSESS product actually checks the programme code, including spreadsheet cells, although inconsistencies have been found with this product.

  All the details on FNAL's work on Year 2000 can be found on WWW.DCD.FNAL.GOV.

## Security BOF

A Security workshop was held one evening. It was noted that HEP sites, along with the rest of the world, suffered more and more attacks. There was felt to be yet more need to share information among recognised security officers. It was also agreed that a survey of available tools, rules and guidelines would be welcome. It was proposed to hold a Security Working Group meeting around the next HEPiX meeting.

## Next HEPiX Meeting

RAL offered to host the next meeting, probably just after Easter in April 1999, The week of April 12th was suggested but remains to be confirmed.

No place nor date was proposed for the Fall '99 meeting which should normally be held in North America.

# CHEP '98

## Real Time NT at D0

For D0, Run II at FNAL consists of 1M channels to read out. They have decided to use NT in their Level 3 trigger farm. As reasons they cite

- NT was written by the VMS pioneers (!) and their Level 3 trigger in Run I was based on VMS
- NT has a number of real time features such as pre-emptive multi-tasking with 32 levels of kernel interrupts; multi-processor handling of interrupts; 32 priority levels for processes and threads; asynchronous I/O; synchronisation objects such as mutexes, semaphores and critical sections of code. Just like VMS.

They have tested dual Pentium II systems at 330 MHz connected to VME via the PCI bus and found a latency under load of 135 microseconds, with a maximum of 225 μsec. It was noted that D0 has 50 msec to accept or reject an event and therefore they claim that NT satisfies their real time needs.

## DCE/DFS at KEK-B

This was a poster session rather than a paper. They have 80GB of DFS space on SUN servers with SUN and PC clients. They say the SUN servers offer stable service with only two incidents in 18 months. But client performance has been "terrible", 35 incidents, many related to multi-threading on dual CPU clients, fewer problems seen on single-CPU systems. Recently almost all known problems are solved. Performance on the PC clients for DEC has been good.

## NICE

Gian Piero Siroli of INFN Bologna gave a review of the work done by Alberto Pace and him to port CERN's NICE to INFN. Extensions added locally included use of NICE over a WAN. The exercise was deemed a success and is showing positive results.

## STK Tapes

An entertaining plenary with the speaker, from STK, insisting at various moments that the TV cameras being used to broadcast the session on Internet be turned off or on depending on the sensitivity of the content of successive overheads!

He forecast 300PB discs in 2003 with 30MBps throughput and at a cost of 0.3 cents per GB. He noted that 1PB per day equalled 1000 tape drives today (50GB tapes) occupying 4 acres of floor space. The San Diego Supercomputer Centre aimed to export 1PB per day early in the next century

STK planned 100GB tapes in 3Q99, 150GB in 3Q01, rising through five generations, one every 2 years, to 2PB (!) in 1Q07 (long range planning?).

## Rhapsody

This was the only MAC talk I noticed on the programme. It concerned the use of a beta kit of Rhapsody at the University of Mississippi. Rhapsody is the new version of MacOS which is built on a UNIX BSD 4.4 base and a Mach 3.0 kernel. It will have two environments, the familiar MacOS as now and a UNIX environment. Of course the UNIX environment will be hidden from most users but the speaker, a long-time UNIX guru, had found it very pleasant for developing UNIX codes; he was using gnu C and C++ and Absoft's Fortran.

Most interestingly, he found it completely possible to develop and even run his UNIX codes in the background while his secretary was using the usual office tools in the foreground. He found no interference between the two environments. He even claimed to be considering porting GEANT (not stated which version) to the UNIX environment in Rhapsody

## Software Release Tools

SRT is a set of configuration and management tools from the BaBar collaboration which has been adopted by other HEP Groups, including some at FNAL. It is effectively an interface to CVS and provides –

- A package and release directory structure (like ASIS) with lots of symbolic links
- Frozen and development releases, the latter being liable to be changed daily
- Tools for release managers and developers including a dependency checker

Creating a test release implies copying and modifying only the files needed while all other files are linked from the production structure. The actual modules themselves live in CVS. Various error checks are performed for inconsistencies and failures.

FNAL has added many features including support for NT (via the cygwin32 package) and Linux and the possibility to build experiment-specific releases. Future plans include a debug option and Java support. It seems at least some parts of ATLAS are using SRT and CMS are getting interested.

## CHEP Summaries

**Mass Storage and Data Management**

BaBar and RHIC (and KLOE at KEK and Run II in FNAL) are testbeds on the way to LHC experiments but need already enormous data storage needs in mid-1999. The clear trend is towards OODBMS, especially Objectivity, but some major experiments still prefer ROOT.

**Commodity Computing and Farms**

It appears that most level 3 trigger systems use Linux, possibly because source code is accessible for local mods or possibly because that's where previous experience pays most. Similarly, Linux is most popular for batch and simulation farms. On the desktop however, Linux is about equal with NT in popularity. Disc cloning is heavily used in initial setup of farms but subsequent management of hundreds of nodes is heavy. For fast networking, Gigabit Ethernet is now preferred over ATM.

**Future Directions**

To close the conference, Richard Mount gave an interesting talk on futures. He claimed (who disagrees?) that HEP is no longer on the leading edge of computing except for very large object databases and distance computing. He predicted T1 links (1.5Mb) to homes for $50 per month in perhaps 5 years (in Europe?) and hoped that the infrastucture could grow to cope with this. GEANT was a good model for collaborations and more should be encouraged along these lines. Commercial software was becoming more common and more popular in HEP. And finally, he claimed that requirements analysis was usually much less useful or appropriate than opportunity analysis and the most clear example remained the creation of the web by Tim.

Alan Silverman
6th January 99