TASK P: Data Analysis Facility

1. Overview

In order to achieve economies of scale and to avoid duplication in meeting our computing needs the computing clusters (VMS, ZEUS-Unix, CDF-Unix CMS-Unix, Electronic Design-Unix (Mentor Graphics and Racal-Zuken), ancillary terminals and PC's) are lumped together under a single umbrella of system, networking and software support. Task P also supports computing for HEP program office and desktops for all experimental HEP group users when based in Madison. Task P serves over 100 users. We began an upgrade project to modernize our aging computing infrastructure two years ago and started to replace our servers and network equipment. This year in Task P we specifically request those hardware items needed to amortize and update the client machines of our data analysis facility. Our experience indicates that, excluding special needs, about $35K annually is needed to amortize and upgrade our installed facility to maintain modern networking, communications and data handling.

The capability of productively doing research in particle physics at a laboratory remote from the accelerator depends in a great measure on the ability to communicate easily and quickly not only the usual discussions between experimenters, but also moderate amounts of data. Thus, the need at the remote location is both for computing facilities and the networking infrastructure to tie them closely to those facilities at the accelerator laboratory. This becomes especially important if the accelerator is on another continent and travel funds are to be scrupulously expended.

Since the major experiments tend to be asynchronous, there is advantage to be gained by merging the facilities. Although we expect differences in detail, recent developments have tended toward common solutions to the computing challenge. At the moment CDF at FNAL is migrating from VMS to Linux. ZEUS at DESY and CMS at CERN have adopted Sun and Linux. With more experiments supporting Linux, the cost of providing compute power has been diminishing, which is fortunate since the amount of data to be processed has been rising dramatically.

We have begun an already productive collaboration with Prof. M. Livny and colleagues in the UW computer science department to exploit the large amount of CPU cycles available in their CONDOR pool of distributed computing systems. The UW has provided funding for a CS graduate student to aid in the conversion of our code to use this facility. The CONDOR CPUs now produce physics Monte Carlo data which are stored back on our UW graduate school provided storage servers for further analysis. The Linux based data analysis facility that we commissioned with DOE support is primarily used to analyze these data. The analysis tasks are IO bound and are better performed on these local computers.

2. Current Status

We recently dedicated an AFS home directory disk server to serve our UNIX clients. We also commissioned a mail and www server, and an integrated backup system. We have also put into use a Linux storage server facility with 1.3 TB disk, and several analysis farm computers. Our new gigabit Ethernet backbone is supplying the needs of high capacity users while permitting a smooth transition for the rest of the group. The data storage facility and a gigabit uplink to campus backbone were provided by the University of Wisconsin. Resources for main servers and some analysis farm computers were provided in previous year's budget.

There are also in use 13 HP Unix workstations which are 7-9 years old. These are at the end of their useful lifetime – several workstations became non-operational last year. The rest continue to serve as restricted-capability desktops and support the electronics CAD work. There are also 7 VMS workstations 7-11 years old, which are also at the end of their useful lifetime. They are used by CDF members for the remaining Run I analyses and by some members of the group for email. No new experiments support the use of VMS for analysis, so we are replacing VMS desktops and

cutting back on the number of machines we maintain. There are 5 other Unix workstations (SUN/SGI/Linux) for hardware tests, CDF analysis and desktop use.

There are now about 15 Apple Macintosh and 3 Intel PC machines for faculty and scientist desktops, as well as group administration. Backup and printing are the only services centrally provided for the Mac/PC clients. We find that the Fast Ethernet network we use internally is quite stable and handles burst rates very well. Most of our desktops are currently networked using 10BaseT Ethernet (10 Mbps) hubs at the office level. We plan to continue laying better quality cables to hook up more offices with Fast Ethernet (100 Mbps).

3. Proposed Plans

We anticipate that the use of Linux analysis facility will dramatically increase with the beginning of CDF Run II and enhanced use of data simulated on CONDOR by CMS group. However, the number of our analysis computers will reduce because some of those existing rack-mounted CPUs will be converted to data storage servers by adding another 1.5 TB RAID disk that is being purchased using the UW funds. Therefore, we would like to add 7 inexpensive monitor-less rack-mounted CPUs to our analysis facility at a cost of $14000.

The HP workstation desktop capabilities are rather restrictive as most productivity software is not supported in that environment. However, we need our students to get trained on Unix computers early in their residency in Madison, so that they will be able to participate fully in analysis activities once their course requirements are completed. Therefore, an immediate priority for our data analysis facility is to replace the aging HP and VMS desktop workstations. For student use, we are planning to purchase Linux PCs. These will also double as analysis farm computers when they are not being used interactively. We budgeted for 6 machines at about $1500 each, adding up to $9,000. The few remaining working HP units will be kept as spares for our electronics CAD workstations. These machines will be particularly valuable if the existing CAD machines fail before we convert our electronics CAD to newer machines.

The increased demands of backup from Macintosh and PC clients have created a bottleneck at our aging backup server. The existing system has become unreliable. Therefore, we propose to purchase a PC with a DLT tape robot to provide backups for all the Macintosh and PC clients in the department. We have budgeted $2,000 for the computer and $5,000 for the DLT tape robot.

We will need to add a new network switch, at $3000, to handle these additional computers. We also want to add, at $2000, additional reliable RAID1 home directory disk to AFS server to support increased number of users migrating from HP to Linux computers.

4. FY2002 Budget

The FY2002 budget will remain at our nominal $35000 support level. The items proposed below enhance our CDF, CMS and ZEUS analysis capabilities, provide Linux desktops for students and provide secure backups for office and faculty Macintosh and PC computers.

Qty.	Item	Price
7	Rack-mounted CPUs	$14,000
6	Linux desktops for students	$ 9,000
1	Backup server w/ DLT tape robot	$ 7,000
1	Network switch	$ 3,000
1	RAID1 for AFS	$ 2,000
TOTAL	**FY2001**	**$35,000**