CDF/DOC/COMP_UPG/CDFR/6001
June 18, 2002

M. Neubauer, F. Würthwein
Massachusetts Institute of Technology
R. Colombo, R. Harris
Fermi National Laboratory

Recommendations for University owned computing for CDF in Run2

Abstract

We provide recommendations of how a University group participating in CDF Run2 may optimize the impact of their computing equipment $$$ on their physics analysis output.

This is a CDF-CD and CAF management perspective which was strongly influenced by the work of the trailer computing committee chaired by Richard Hughes.

Contents

1 Executive Summary

The objective of this document is to provide CDF collaborators some guidance on how to spend their computing equipment $$$ in order to maximize the resulting physics capabilities of their groups.

In particular, this is **NOT** a poorly veiled attempt of begging for money to support the central computing resources that FNAL has traditionally provided. FNAL will continue to provide those resources as is detailed in CDF 5914.

Instead, our goal here is to avoid situations where University groups bring in computing equipment without appropriate planning, only to find that there is no infrastructure to support these resources. We encourage University group leaders and decision makers to take a close look at this document when preparing for further spending decisions.

The recommendations we present here were strongly influenced by the trailer computing committee chaired by Richard Hughes. However, this should not be confused with the report of that committee. This is the CDF-CD and CAF management perspective of what we consider the best investment given the CDF run2 computing model.

The run2 computing model may be summarized as follows:

- Compile, link, and debug analysis executables on the user "desktop".
- Large scale analysis of CDF data via batch submission to CAF (and DCAF).
- Small scale analysis of data interactively on the user "desktop".

The basic idea here is that all the work that fits within a users attention span for interactive work should be done on the "desktop", while all activities that go beyond this are submitted to large farm(s) of PC's modeled on the CAF.

We consider a "users attention span" to be measured in minutes. I.e. any activity that takes longer than $O(10\text{min})$ or so benefits from submission to the (D)CAF. Based on this model we are making the following prioritized recommendations:

1. First and foremost one should invest money in Linux based desktop PC's. These desktops should be physically close to the physicists. E.g. it makes a lot of sense to spend money on desktops at one's home institution if physics FTE's are stationed there, and system admin resources exist to support the equipment, as well as a cdfsoft installation. The most effective way of deploying limited funds is to provide physics FTE's with sufficient resources that their interactive work does not consist of twiddling

their thumbs while they wait for compilations to finish etc. What we consider adequate desktop computing is detailed in Section 2. FNAL will continue to provide centrally administered interactive login resources but anybody who has used them knows that they tend to be heavily overloaded at exactly those times when you need them most, and idle when you don't need them.

2. Once a group has adequate desktop computing then additional $$$ are best spent by owning equipment that is fully integrated into the CAF. For this one may purchase either disk space or CPU power. We recommend to spend the first $10k of CAF money on disk space. The CAF currently has a soft/hard quota on icaf/scratch of 25/50GB per user and stage2 deployment results in more than 400 analysis CPUs. A $10k investment would buy 1TB of user disk space inside the CAF plus 1TB of data disks for which one controls the content without having to worry about the data management. See Section 4.2 and CDF 5961 for additional details.

3. Once a group has sufficient disk space within the CAF one may consider investing in CAF CPU power. See Sections 4.1 and CDF 5961 for additional details.

4. Finally, let us stress that pooling resources among University groups is a good thing. E.g. stage1 of the CAF received contributions from two groups. INFN contributed 10 Duals and 4 file servers while CMU-U.Pittsburgh-MIT contributed one file server plus 16 Duals. The optimal ratio of worker nodes/file server lies probably somewhere between these two examples.

Section 2 describes our recommendations for minimal as well as maximal desktop computing. Section 3 describes the timeline of the intended CAF procurements, and how University groups may add to these acquisitions. Section 4 describes usage policies and rules for University owned CAF equipment. CAF hardware reqs for Stage2 are described in the Appendix.

2 Ideal office computing

Ideally, each physics analysis FTE ought to have 1GHz or more CPU power and 1GB of RAM for interactive work. The minimum amount of disk space desired depends a lot on the persons individual work habits and degree to which the person is organized, and frugal. Our tests of 100GB 7200 RPM Western Digital and 160GB 5400 RPM Maxtor drives used in the protoCAF and CAF do not indicate substantial advantages in speed from the 7200 RPM drives. We would thus tend to favor the larger drives for data disks, and something more like 20-40GB as a system drive.

A perfect two person office in the B0 trailers with two physics analysis FTE's might have one of the following:

- Two powerful new desktops. E.g. Dual AMD or P4, 2GB of RAM each, CPUs near the top of the line. Each of the desktops might have a few hundred GB of disk space. Based on the evaluations done by OSS the CAF is unlikely to buy P4 at current prices. See Section 3.1 for more detail. Price per Dual CPU PC ~$2500.

- Two older Linux PC's (single 400MHz CPU, 128MB RAM) plus one build node. The build node would be a Dual AMD or P4, clock speed near the top of the line with 2GB RAM. The build node might have 4 160GB drives, two on-board FastEthernet connectors plugged into the local switch, and might NFS export it's disks to the two older desktops. The two users would use the build node for all their interactive work except web serving, latex, etc. For the sake of the users eyes as well as space and power consumption in the office the old desktops ought to have 15inch LCD displays. Windows PC's with some X connectivity software like eXceed from Hummingbird will do instead of the two older Linux PC. Price per build node ~$2500.

We believe that CPU power in excess of the above, as well as TB file servers are best located in FCC where they can be integrated into the CAF, or at remote institutions where they can be integrated into DCAF. The latter is not supported until the end of FY03. The date for support of DCAF could move forward significantly if additional human resources could be found to work on the DCAF-CAF integration.

We furthermore believe that the 2GB of memory specified above is essential for cdf software builds. As an aside, we started with 128MB on the first Level3 nodes, and considered that plenty enough. The latest production farms purchases required 1GB, because this was deemed necessary. While we do not expect that CDF's memory needs continue to scale like this over time we are nevertheless uncomfortable recommending anything less than the 2GB of RAM we require for the stage2 CAF nodes. After all, equipment bought today is meant to be used for 3 years, and memory upgrades are not a cost effective way of spending human resources.

Users will benefit by NFS mounting cdfsoft distribution from the central code server in B0. This code server is presently being upgraded to support the needs of the collaboration in the B0 trailers. We are committed to keep the performance of the code server(s) up with demands such that people no longer need to maintain local cdfsoft distributions on their desktops themselves. With the advent of CAF we have started monitoring network and CPU load on the code server(s) to respond in a timely fashion if the new code server appears too heavily loaded.

3 How to join the CAF procurement cycles

A University group may join CAF purchases of computing hardware to benefit from large volume orders and integration of their equipment into CAF infrastructure at FCC. Procurement for the CAF is planned twice a year, roughly every six months. Operationally, this is envisioned as follows:

- CAF team decides on request for bids for worker nodes and file servers and makes those req.'s available to the collaboration. See Appendices Aand B for drafts of the Stage2 req.'s.
- The req.'s include an estimated purchase price. For Stage2 this is $2,500 per worker node and $10,000 per 2TB file server. A University group, or consortium of groups decides how many pieces of equipment of each type they want to purchase. If the actual bids come out cheaper then the groups save money, if they come out more expensive then fewer pieces of equipment can be bought.
- The req. includes some metric that is used to determine the winning bidder. This metric is specified as part of the req. and communicated to the vendor. The metric for Stage2 worker nodes is discussed in more detail in Section 3.1. For file servers it is simply usable disk space after RAID5 per $$$.
- Before the actual purchase an MOU is signed between FNAL-CD and the University group(s). This MOU details the FNAL maintenance and operations responsibilities as well as access privileges to University owned resources as described in this CDF note.
- After three years FNAL reserves the right to decommission equipment, and indeed intends to do so in order to optimally use the limited space and infrastructure in FCC. At this point the University may decide to either junk or take back the equipment to their home institution. The arguments in favor of a 3 year replacement cycle are threefold:
  - Floor space and switch ports are limited. E.g. a 5(4) year replacement cycle requires 39%(19%) more floor space for the same CPU power assuming a Moore's law doubling every 18 months. The CAF as described in CDF 5914 and CDF 5961, i.e. without additional University equipment other than INFN, UK groups, CMU-U.Pittsburhg-MIT, requires roughly 3/4 of all the space available for the CDF CAF in FCC.
  - The equipment we buy has a standard 3 year warranty. If we were to keep PC's around for longer then the University would have to pay for all replacement and labor costs.
  - Older equipment fails more frequently.

The most important of these is floor space in FCC. We do not rule out that CDF may renegotiate decommissioning old equipment in 3 years time. At that point we will know the actual rather than the projected floor space needs, and may feel less constraint by space. However, at this point we consider it irresponsible to promise anything other

	stage2	stage3a	stage3b
Req. @ FNAL	7/9/02	12/20/02	7/1/03
start bidding	8/9/02	2/1/03	8/1/03
hardware arrives	9/9/02	3/1/03	9/1/03
available to users	10/9/02	4/1/03	10/1/03

Table 1: Expected CAF procurement cycles for stage2 and stage3.

than a 3 year replacement cycle given the projected space needs for CDF, CMS, and D0, as compared to the space available in FCC.

The timeline for the next three procurement cycles are roughly as listed in Table 1. The first month of the timeline in Table 1 is used up to obtain all the required signatures as a requisition percolates through FNAL and University bureaucracies, and the last month is used to install and burn in the equipment. The stage1 hardware stopped showing infant mortalities after roughly 2 weeks of burn in. From experience we know that the schedule in Table 1 is aggressive.

3.1 PC evaluation cycle

In addition to the steps shown in Table 1 there is an evaluation cycle during which OSS certifies vendors and their equipment for use in FCC. Issues addressed during this evaluation are performance, chassis construction, and cooling. Performance evaluation includes defining a "FNAL cycle" to be able to renormalize P3, P4 and AMD clock cycles based on performance on D0 and CDF applications (ProductionExe and cdfSim). The cooling evaluation is crucial as space concerns require us to use 1U rackmount Duals as described in Section 4.1

Their report is not public. However, details may be obtained using a webrowser from within the FNAL domain via the following web page: http://www-oss.fnal.gov/scs/qualify2002

Figure 1 shows the performance scaling behavior versus CPU clock cycles for ProductionExe. The D0 reconstruction executable shows the same kind of behavior. Indicated in color are the CPU families. A 2GHz P4 is roughly equivalent to a 1.4GHz P3. The Athlon CPU lies more or less on the P3 curve. Based on this behavior, OSS has defined an "equivalent P3 Coppermine" CPU cycle, and uses that as figure of merit for worker node purchasing decisions. Vendors are encouraged to bid either P4 or Athlon based systems, or both, with the understanding that FNAL buys whatever certified system that provides the most "equivalent P3 Coppermine" CPU cycles per $$$.

OSS does presently not perform file server evaluations. Basically all other experi-

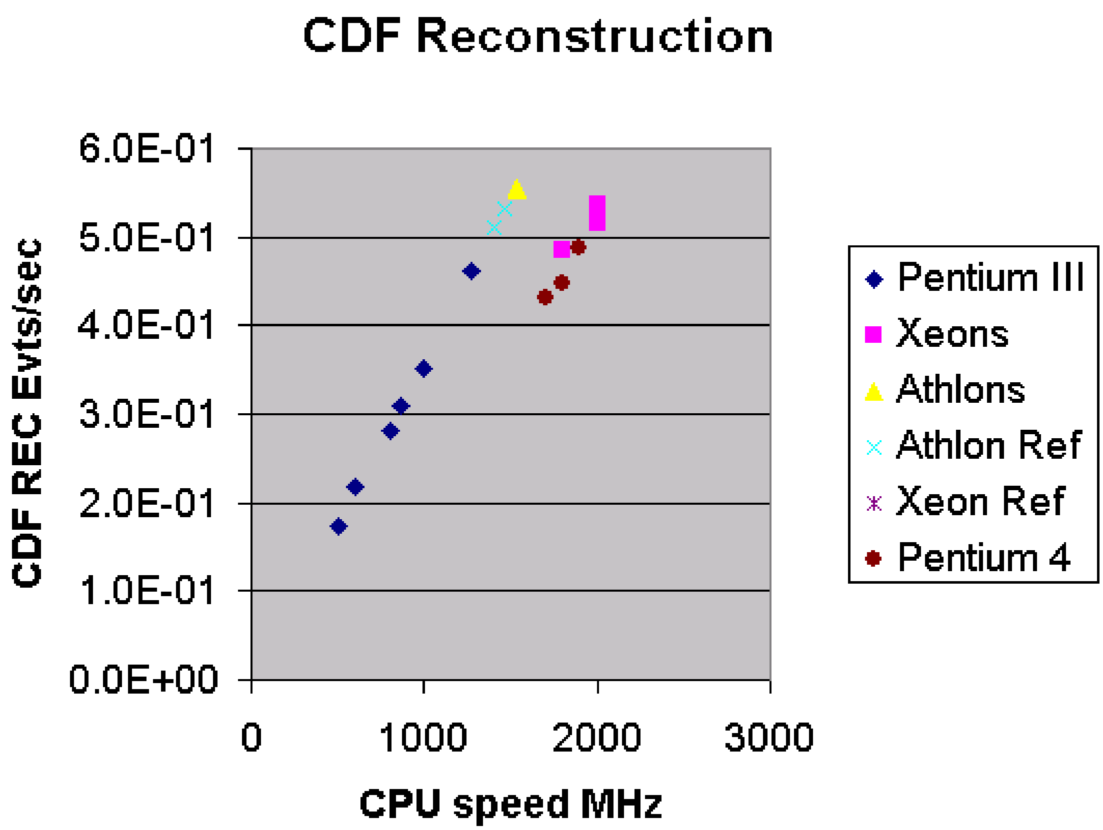

Figure 1: Performance versus CPU clock for different processor families. The performance metric is events/sec for CDF ProductionExe.

ments at FNAL watch the CDF file server experience closely, and are interested in somebody centrally picking up the kind of server evaluations that we have done for Stage1/2 as part of a procurement cycle, similar to worker nodes.

4 Usage and operations agreement

FNAL-CD provides power, cooling, space, LAN networking, as well as maintenance and operations for the University owned equipment throughout the three year warranty period over which the equipment resides in FCC. The only work required by the University group is to make a purchasing decision and provide a FNAL budget code.

At the end of the three years FNAL decommissions the equipment, and junks it, unless the University group decides to take the equipment back. In the latter case it is the responsibility of the University group to care for shipping and handling of the equipment to wherever they want the equipment to go.

The University group is not involved in the management or operations of the equipment, and has access to it solely via the CAF. Access to file servers and worker nodes differs as follows:

4.1 Worker Nodes

A special queue in the CAF batch system is created for each group of owners. This queue has highest priority on the set of worker nodes the group of owners has contributed. Only users associated to the owners may submit jobs to this queue. These users thus receive preferential access but not dedicated access to the worker nodes they own. I.e. if the worker nodes are idle other users will get to use them. Once an owner submits a job to their special queue they may have to wait until one of the jobs presently running on those nodes completes. Upon owner request we may allow only short jobs (2 hours of CPU time) on their queues to avoid lengthy delays.

It is up to the user to decide whether they want to submit their job to the special queue they have privileges for or their normal CAF queue. For additional details see http://cdfcaf.fnal.gov as well as CDF 5961.

4.2 File server

Our benchmarking tests indicate that the file server bandwidth is CPU limited to roughly 60MB/sec. While this is plenty for our purposes the CPU on the file server is nevertheless a precious resource and shall be used solely to serve data. Login into the file server is not allowed. Instead, the following access rules apply:

- rcp from worker nodes: allowed
- ftp from worker nodes: allowed
- ftp from desktop: allowed (via icaf tools e.g.)
- rcp from desktop: not allowed
- rsh/ssh to file server: never allowed
- read access via rootd from anywhere: allowed

In essence, the file servers should be thought of as network attached disks rather than the server class PC's with disk that they are.

In addition to these technical restrictions, there are some policy restrictions to make maximal use of the disk space. These policy restrictions are as follows:

- 50% of disk space is set aside for arbitrary use by the owners via icaf and user scratch mechanisms of the CAF. For details see CDF 5961.
- 50% of disk space is populated with official CDF datasets of the owners choice. In practice this means that some physics or detector group or alike manages this in coordination with the owners. Ideally, the files on this disk are fully documented, don't disappear over night without notice, and are accessible via the CDF DH system. Examples range from standard ntuples for a set of thesis analyses that a group of people collaborate on to raw data that's needed to check some detector calibrations.

While the file servers are RAID5, and thus provide some amount of data security, none of the data on the file server is backed up by the CAF team. If a user wants to have their data backed up they need to do so themselves.

5 Acknowledgements

We gratefully acknowledge the input from OSS, most notably Steve Timm, CDF Task Force (Jason, Mark, Lance, Richard, Glenn, ...), trailer computing committee chaired by Richard Hughes, as well as the computing representative board, most notably Stefano Belforte.

A Draft Req. for CAF worker node

The following is an almost complete excerpt of "http://www-oss.fnal.gov/scs/farms/farmspec.doc". This URL is accessible only from within the FNAL domain.

This requisitions draft is used by OSS for procurement of worker nodes for farms purchases.

A.1 PC Specification for Purchase

Three types of processor are currently approved for dual-CPU farm workers, the Intel Pentium III, the Intel Xeon, and the AMD Athlon MP. Of these we have found the Athlon MP to have the best performance on the typical applications we run. We present a synopsis of the three configurations below, and then extended comments on them that apply to all of the configurations.

A.2 Configuration Summaries

A.2.1 Pentium III Worker Node

Dual Intel Pentium III CPU's on same motherboard, SMP capable, 1.4 GHz or better speed,
133 MHz front-side bus or greater
512kbyte on-chip cache
Two Gbyte PC133 SDRAM (Registered ECC required)
Motherboard must be Tyan Thunder LE-T (S2518GNR)
One IDE system disk, 20 Gbytes or greater
IDE data disks: two 40 Gbytes or greater, or one 80 Gbytes or greater
Video at least 512 Kbytes RAM, support for 640x480 SVGA mode or greater
One 1.44 Mb 3.5" floppy
One 24X or more CD-ROM drive
No monitor, keyboard, mouse
One or more 100 Mbit/sec Ethernet
1U rack mount case

A.2.2 Dual Athlon Worker Node

Dual AMD Athlon MP processors on same motherboard, SMP capable, MP1800+ (1533 MHz) or greater, 266MHz front-side bus
Two Gbyte of PC2100 DDR SDRAM, registered, ECC
Approved motherboards are Tyan 2462, 2466, and 2468
One IDE System disk, 20 Gbytes or greater
IDE data disks: two 40 Gbytes or greater, or one 80 Gbytes or greater
Video at least 512 Kbytes RAM, support for 640x480 SVGA mode or greater
One 1.44Mb 3.5" floppy drive
One 24X or better CD-ROM drive
No monitor, keyboard, mouse
One or more 100 Mbit/sec Ethernet
1U rack mount case

A.2.3 Dual Xeon Worker Node

Dual Intel Xeon processors on same motherboard, SMP capable, 2.0 GHz or faster, 512k on-die L2 cache (Prestonia series), 400 MHz front-side bus or greater
Two Gbyte of PC800 RDRAM (or PC2100 DDR SDRAM if board allows)
Approved motherboards are SuperMicro P4DCx series, P4DPx series, Tyan 2720 series
One IDE system disk, 20 Gbytes or greater
IDE data disks: two 40 Gbytes or greater, or one 80 Gbytes or greater
Video at least 512 Kbytes RAM, support for 640x480 SVGA mode or greater
One 1.44 Mbyte 3.5" floppy
One 24X or more CD-ROM drive
No monitor, keyboard, mouse
One or more 100 Mbit/sec Ethernet
1U rack mount case

A.3 Hardware Specification Details

The following hardware specifications apply to all the configurations above unless noted:

INTEL CPU
Intel Xeon CPU must come with three-year warranty (manufacturer's warranty preferred), Intel-provided ball bearing fan (or approved equivalent), and heat sink. Fan must have speed sensing capability.

Intel PIII CPU must come with a three-year warranty (manufacturer's warranty preferred), Intel-provided ball bearing fan (or approved equivalent), and heat sink. Fan must have speed sensing capability.

AMD CPU

AMD Athlon MP CPU must come with three-year warranty (manufacturer's warranty preferred), AMD-approved fans and heat sink. Fan must have speed sensing capability.

```
Motherboard:
Should be compatible with Fermi Linux and meet the following requirements:
a)  One or more serial ports
b)  Two IDE buses (UltraDMA capable)
c)  Support for 2 Gbyte or more of memory
d)  BIOS must allow system to boot without a keyboard and monitor
e)  Hardware support for temperature and voltage sensors
f)  Must be able to support at least 1.4 GHz chip speed
        (Pentium III) or 2.4 GHz chip speed (Xeon) or 1.733 GHz
        chip speed (Athlon)
g)  On-board video and networking are strongly preferred
h)  IDE RAID devices may be included but must either be compatible
        with Fermi Linux 7.1.1 installer or be able to be disabled
i)  USB, SCSI interface and parallel port are optional, but,
        if included, must be supported by Fermi Linux 7.1.1

Memory:
Should meet the following requirements:
a)  Must be 1 Gbyte
b)  Must be one of the recommended memory vendors for the motherboard
c)  Must be of the highest performance the motherboard supports
d)  Must be ECC registered memory
e)  SDRAM must be 133MHz or better, DDR SDRAM must be 266 MHz
        or better (PC2100), RDRAM must be of the PC800 variety or faster

System and data disk cable:
The IDE cables to the system disk and data disks should be the
80-conductor type and not the 40-conductor type.  They must be
routed neatly through the appropriate paths in the case.

System Disk:
An IDE system disk with a minimum capacity of 20 Gbyte, which meets
the minimum requirements of
a)  7200 RPM or greater
b)  Maximum of 12 ms seek time
c)  Must be UltraDMA capable
d)  Must be able to support 20Mb/sec or greater throughput,
```

as measured by bonnie

Data Disk:
These IDE disks should be 40 Gbytes or larger. The vendor may use one 80 Gbyte (or larger) disk instead of two 40 Gbyte disks. They must be attached to a different controller than the system disk and CD-ROM.
Minimum requirements:
a) 7200 RPM or greater
b) Maximum of 12 ms seek time
c) Must be able to support 20 MB/sec or greater throughput to each disk concurrently, as measured by bonnie.

Low Performance Video:
a) Minimum of 512k video memory
b) Must support 640x480 SVGA or better
c) Must be supported by Fermi Linux 7.1.1 and Xfree86 v4.1.

FLOPPY DRIVE:
One 1.44 MB 3.5'' Floppy disk drive

CDROM:
An IDE CD-ROM drive, installed on the same bus as the system disk.
Must be at least 24X speed.

Network:
A minimum of 1x 100 Megabit per second Ethernet Network Interface.
The following network cards are NOT PERMITTED: DEC Tulip CLONES, SMC, and NE2000.

1-U Rackmount Case:

a) 32 x 1U cases must fit into a 45-U rack
b) Must be mountable with sliding rails
c) Tool-free opening of case is a plus
d) UL approved power supply
e) All ports labeled accurately
f) Units must ventilate front-to-back, not top-to-bottom
g) 1U and smaller cases are allowed but only if they can still hold the hard drives specified plus floppy and CD ROM
h) Power supply and fans must meet all specifications of CPU manufacturer and be installed per the motherboard manufacturers specification.
i) 1U cases for Athlon hardware must have an exhaust fan.

A.4 Software Specifications

Fermilab will provide the version of Linux OS from a CD ROM. The version of Linux supplied must be installed and tested by the vendor prior to shipment. Should the vendor have to make any modifications to the system software configuration in order for their hardware to work with the version of Linux OS provided, Fermilab requires that the vendor include written documentation to be sent with the shipment describing these changes.

A.5 Cabling and Rack specifications for 1U Cases

Vendor must provide an integrated racked solution where each rack should hold a total of 32 units. Vendor must provide the labor for onsite installation. Vendor to certify that rack holds 32 x 1U units before delivery.

```
Power supply:
      1) On rack for 32 x 1U units:
         1 x 20-Amp APC Master Switch, vertical mount,
         remote power-down unit, Model AP9221, plus 3x 20-Amp
         Master Switch Vertical Mount Expansion Unit,
         Model:AP9221EXPB.
         Mount 2 APC units on each back rail of the 45U rack.
         Vendor should arrange for NEMA 5-20 plug (not twist-locking).
      2) The length of the power cords of rack power distribution
         units must be 15 feet.
      3) Vendors must specify the number and rating of electrical
         circuits needed to power the equipment and the NEMA
         designation of the required receptacles.  The electrical
         configuration must be such that the circuit ratings are not
         exceeded during startup and running at full load.  (In the
         above configuration, one 20 Amp circuit with NEMA 5-20 plug
         is needed for every eight units.)
      4) Power cords of each PC shall be plugged in sequence into the
         APC, and labeled with the node name at both ends of the cord.
 APC Contacts:  Alex Tavakalov - 1-773-561-1234 or
                John Parola -1-708-236-0650
 APC Web Site:  www.apcc.com
```

Rack Specifications:

Racks should conform to AMCO standards or equivalent (up to 36" deep if required, but 30" is preferred). Minimally the structure must have 14-gauge mounting rails with welded

12-gauge corner braces or better. Riveted or fastened braces are not allowed. Racks must have four swivel casters rated for rack weight when fully loaded and a means for stabilizing and/or leveling, which cannot protrude beyond the front face of the racks into an aisle. The rack maximum height (including casters, fan, etc) must not exceed 90 inches. For equipment that draws cooling air in through the front and exhausts through the rear, the racks should have open bottoms and no doors. In this case, an exhaust fan at the top of the rack is unnecessary.

(skipped rack dimensions and drawing here)

A.5.1 Required Network cabling scheme for 32 x 1U units:

```
One Category-5 network patch panel per rack:
UTP:  RJ45-RJ21F, 48 jacks, EIA 568A, Superior DCC4884/25V-100
One Category-5e cord per node,dressed and patched into the
patch panel and labeled with preprinted labels on both ends
with node names provided.

UTP: RJ45-RJ45, EIA568A, CAT5e, PVC, Stranded.Cord lengths to
allowPCs to extend fully plus one foot, no more.
No excess cord slack stored in the side panels.
```

A.5.2 RequiredConsole cabling scheme for 32 x 1U units:

```
One Category-3console patch panel per rack:
UTP: RJ11-RJ21M, 48 jacks, USOC, Superior DCC4864/2502
COM1 port on each node attached with serial cable to
the serial patch panel dressed and patched into the patch
panel and labeled with pre-printed labels on both ends
with node names provided. Required adapter should
have DB-9 (female) to RJ-45 connector with long thumb screws.
Belkin part number A4C170-19995 (pinout available on request).
Cable specification:  UUP, RJ11-RJ11, USOC, Stranded, PVC.
Cord lengths to allow PCs to extend fully plus one foot,
no more. No excess cord slack stored in the side panels.

One cable vendor that has pre-set Fermilab specifications is
North American Cable, 227 S. River St., Suite 100, Aurora, IL  60506.
630-264-2500.
```

Reserve an 8 inch space at the BASE and a 7 inch space in the CENTER of the 45U. Mount patch panels on the back rails of the 45U rack in the 7" reserved center space. Cost must include shipping to Fermilab and installation. Total heat load should be specified.

A.6 30 Day Acceptance Test

Fermilab shall evaluate the systems for a thirty-day acceptance test once installation is complete. The entire group of systems cannot have more than 2% downtime for the duration of this acceptance test. If there is greater than 2% downtime then Fermilab reserves the right to return the entire group of systems. Please Note: should the systems fail this acceptance test, the vendor will pay for the return packing and shipping of all systems. The vendor is responsible for parts and labor cost of repair during the 30-day burn-in period.

A.6.1 Requirements

Start/Stop of Downtime
Each workday at noon (Central Time) we will report the status of the systems to the vendor's representatives. If one of the systems has a problem, the vendor has until noon the next day to fix the problem or the system will be considered down for another day. The minimum unit of downtime is one day.

Lemon Law
A system is considered a lemon if it is
I) Down for five working days in a row
OR
II) Has more than five separate incidents.
If a system becomes a lemon it will be marked as down for the remainder of the testing period.

A.6.2 Tests

The "Dead on Arrival" test will be performed immediately after the system's arrival at Fermilab. Dead On Arrival is defined as failure to boot the operating system when first powered on after installation. If the system is Dead On Arrival it will be counted as down until the vendor can either fix or replace it. There will be four tests running continuously on the systems for the entire 30 day period:

```
I)  One instance of Seti@Home per CPU, running continuously.
II)  Disk tests:
```

```
   II.a) File system intensive copy of a large directory with many
        files from one disk to the other, every 10 minutes.
   II.b) Bonnie--1 Gbyte file written and read back from each disk,
         including the system disk, every 10 minutes, once to system
         and data disks and simultaneously at the top of the hour.
         Performance must be more than 20 Mbyte/sec as measured by
         bonnie. Also, no DMA access faults can occur during this process.
III) One instance of Nettest (400 Mb) will be run each hour.
IV)  Temperatures will be checked each hour.
```

A.6.3 Definitions

Downtime
A system is considered down when it is not usable, whether because of hardware or software faults.
Incident
A report of a problem with a specific system to the vendor
System
The hardware and software associated with a specific node name.

A.7 Warranty and Delivery Requirements

The vendor must fill out the configuration sheet below. In general, if the configuration that is bid has a different motherboard, chip set, or disk drive configuration than systems that the vendor currently has installed at Fermilab, or evaluation units that have been sent in the past, an evaluation unit must be shipped within two weeks of the award of the contract. The specific request for bid will have more details on this issue.

The bid should include a full bill of materials of all parts that will be shipped.

A.7.1 Warranty Specifications

The following is a summary of the warranty requirements of Fermilab, which will be required for the units and any future purchases.

- A three-year warranty on parts shall be required.
- Parts that need replacement should be advance shipped. Mail-in service for systems is unacceptable.

- Coverage: Monday through Friday, during the hours between 9am and 5pm local time, excluding Federal Holidays.

(skipped vendor configuration sheet and contact info)

B Draft Req. for CAF file server

While there are standard, computing division wide procedures for farm worker node procurement, nothing equivalent exists for disks, or file servers. Instead, each experiment at FNAL is responsible for their own disk space procurement. We are in on-going negotiations with OSS to change this.

For the Stage1 CAF file servers we have followed the lessons learned from OSS in that we first went through an elaborate evaluation cycle to certify vendors, as well as technologies. Details on the evaluation results may be found at "http://mit.fnal.gov/~msn/cdf/caf/server_evaluation.html". A CDF note (CDF 5962) summarizing the results is in preparation.

The main difference between worker node procurement and file server procurement are as follows:

- All certified vendors are based in California. It is thus not cost effective to require them to install the systems. Instead, CDF task force & CAF team physically installs the systems.
- 30day stress test at maximum local as well as networked I/O bandwidth.

For Stage2 we intend to purchase the same systems, requesting bids from the same vendors, as for Stage1 as these are still top of the line.

The following is a draft req. for the file server purchase.

B.1 Details for the CDF file server procurement

Each file server is comprised of an Intel Dual CPU motherboard, 2GB of RAM, a Gigabit Ethernet network card, and two RAID arrays using Escalade 7850 controllers from 3ware and IDE drives from Maxtor as specified in detail below. Inside the chassis is a Dual CPU Linux PC with two arrays of IDE disk on a RAID-IDE controller that implements RAID0,1,5 in hardware. The file server provides network attached storage via a single Gigabit Ethernet card. All disks must be hot swappable.

- Each system is housed in a 3U or 4U rackmountable chassis. All disks must be hot swappable. The enclosure should be suitable for use in computer equipment rooms. Mounting hardware for the systems to be installed in industry standard 19 inches equipment racks should be included.
- Power supplies must operate at 110V 60Hz, redundant power supplies are preferred.
- Depending on chassis form factor a file server may hold between 13 to 16 IDE drives arranged on two Escalade 7850 RAID-IDE controllers from 3ware.
- As individual disks we expect Maxtor D540X 4G160J8. We require firmware on the 3ware RAID-IDE cards that allows access to the full capacity of the drive.
- A minimum of 5 spare disks should be provided. At least three of the spare disks should be installed in the same hot swap canisters used in the enclosures, and ready to be used to hot-swap broken drives. Sufficient spare power supplies for one system are required. One spare RAID-IDE card is required. One spare Gigabit Ethernet card is required. All of these are number of spares per total order, **NOT** per system.
- For each system we require: the Intel SDS2 motherboard. Dual PentiumIII 1.4GHz CPUs or better required. 2Gbyte PC133 SDRAM (ECC required).
- One 1.44MB 3.5"floppy and one 24X or more CD-ROM drive per system.
- One Gigabit Ethernet SX Fiber card, preferably SysKonnect SK9843 per system.
- The systems should be fully configured and tested for 24 hours. Installation at Fermilab is not required. After installation at Fermilab the systems will undergo a 30day stress test at maximum local as well as networked I/O bandwidth. During this test we will measure the temperature of the disks to verify that sufficient cooling is maintained. We reserve the right to return the systems if they fail this test.
- All components should be warranted for at least 3 years. In case of failure, Fermilab shall have the option to return complete system(s) for exchange or repair during the first 6 months. Failed components shall be exchanged thereafter.