DØ Note XXXX
July 1997

VMS Migration Plan for DØ

Pushpa Bhat
and
The DØ Computing Planning Board
(Revised from a version presented at NEU workshop, June 1996)

1. Introduction

In alignment with Fermilab's direction, the DØ collaboration is making a concerted effort to migrate its computing away from the VMS operating system to UNIX and other lab-supported platforms. Our large-scale offline computing activities for Run-1 such as event reconstruction, streaming and GEANT detector simulation have in fact been performed on SGI and IBM machines running UNIX operating systems. We have now set out to migrate our analysis computing activities to machines running non-VMS operating systems. A majority of users are expected to move to UNIX platforms. Major portions of the DØ software environment are already available on SGI/IRIX and IBM/AIX machines. A major expansion of our CPU and I/O capabilities on UNIX platforms has already begun. Our hope is that the migration to UNIX becomes lucrative and hence natural. It is therefore necessary to address the issues related to migration and evolve a well thought-out plan for migration.

The purpose of this document is to develop a migration plan and schedule for DØ, from the perspective of the collaboration. This document will only deal with off-line activities and not with on-line computing issues.

2. Migration Platforms

Whether it is boon or bane, there are too many choices for anything in the UNIX world. First of all there are dozens of flavors of UNIX operating systems that run on hardware from different vendors. So, by a "platform" we mean here a particular flavor of UNIX operating system (such as IRIX) running on a particular hardware architecture (like SGI). The Computing Division supports a handful of these platforms including SGI(IRIX), IBM(AIX), DEC Alpha(Digital UNIX), and Sparc (Solaris). For DØ, supporting a platform translates to providing the DØlibrary (the repository of DØ's offline code) on such a platform, its continuous maintenance and code distribution, and systems management (including upgrading OS as needed, and providing various tools etc.). Supporting each additional platform therefore, amounts to a lot of real work. DØlibrary has been ported to SGI(IRIX) and IBM(AIX) machines and these are currently the DØ supported platforms. In our present computing environment, we have a large number of DEC Alpha machines that are running openVMS and are capable of running Digital UNIX. We anticipate DEC Alpha to become the third major UNIX platform to be supported at DØ. Some progress has been made recently in porting DØ library to Digital UNIX. We expect to have a fully supported DEC UNIX cluster by the end of September 1997.

Other platforms that are of interest to DØ are PCs running Windows NT and PCs running Linux (a UNIX OS). We are extensively using PCs running Windows NT with commercial software for spread-sheet applications, statistical analysis, and word processing. There is a PC project (WinCenter Pro) that serves out PC applications to all kinds of desktops – X-terminals and workstations alike. (The engineering and technical staff at DØ use PCs for design applications, databases etc. But, we will concentrate only on usage by DØ physicists in this document). There are exploratory projects that are using PCs running Linux OS. Fermilab Computing Division at this point supports PCs with Windows NT(WNT)/Windows 95(Win95) for some restricted purposes. When it comes to Analysis computing on PCs, access

Machine	Machine type	CPU	Disk	Functionality
dØcha	Challenge XL	8x150MHz	210 Gb	PIAF/Batch Server
dØchb	Challenge XL	8x200MHz	62 Gb	Interactive/Batch

Table 1: Central SGI UNIX Machines

to large amounts of data would be an important issue. The data would reside on disks and in robotics connected to multiprocessor UNIX systems and therefore are not expected to be accessible from desktops for large-scale processing. The other major concern with PCs running WNT/Win95 is robustness. Computing Division has been actively exploring the PC issues as a part of its continual search for cost-effective solutions for the computing facilities it provides to the physics community at the lab. The test of PC farms for production reconstruction has been successful, consistent with the findings at CERN. However, we will concentrate on migration to UNIX at the present time. The issue of analysis computing using WNT/Win95 will be re-visited at a later date.

3. Current Resources on UNIX platforms

DØ acquired (about two years ago) SGI Challenge XL machines and one Challenge L machine, which are all powerful multiprocessor systems with enormous extensibility. These are in addition to about 26 SGI and 3 IBM workstations that are used as desktops. The Challenge machines are being used as Central UNIX platforms as described below. (Also see table 1).

One of the Challenge XL machines, *dØcha*, is being primarily used as a PIAF server. It stores large data-sets in the form of ntuples which are analyzed remotely from VMS and UNIX. Upgrading *dØcha* Challenge L to Challenge XL some time ago improved the performance substantially. CPU intensive batch jobs (with LSF batch system) are also run on *dØcha*. We expect to double the CPU on *dØcha* and augment *dØchb* in the next fiscal year (1997-98).

dØchb, also a Challenge XL provides our interactive central UNIX platform. It is supposed to be the UNIX equivalent of FNALDØ. Users can get accounts on this machine and are encouraged to use it for all their analysis computing. Short batch jobs are allowed on *dØchb* via the LSF batch system.

4. Vision of an Analysis Computing Model for Run 2

Our Run-1 analysis computing was (and is being) done on the FNALDØ, DØSFT and DØAXP VMS clusters. It is undoubtedly a highly distributed computing environment. The first step in analyses has been to run batch jobs on one or more nodes on one of these clusters, access data stored on the DØ File Server (DØFS – a VMS cluster) via the network (FDDI/Ethernet) and make compressed data sets (such as ntuples) that can be stored on physics project disks. Very large ntuple sets are stored on disks provided on the *dØcha* Challenge disks and used via PIAF. These ntuples are used repeatedly by individuals working on various analyses to obtain final physics results.

One of the problems we encountered early in Run 1 analysis computing was the limited bandwidth between the analysis clusters and the fileserver DØFS. The other problem was due to our data format and organization. The data format was abstract and data-sets quite obscure so that the level of effort required for an individual to carry out an analysis was quite high. This resulted in a large fraction of the collaboration not being able to actively participate in physics analyses. This has to change in Run 2, and we need to make it easy for more people to examine data, test their new ideas and carry out analyses.

Within the last couple of years, the computer industry has produced powerful multi-processor machines that provide high I/O bandwidth and support a very large number of peripheral devices. The advent of such machines has prompted us to re-examine the premises of centralized computing *vs* distributed computing. At least in the case of DØ analysis computing, it appears that we could move from a highly distributed arrangement to a hierarchical computing model. With the current wisdom, the vision for an analysis computing model for DØ leading up to and beyond Run 2 is as follows:

- The offline production data processing is carried out using "Farms" as in Run 1. These farms could be built from one or more flavors of machines such as SGI, IBM or PCs running Linux or Windows NT. Recall that in Run 1 we used both SGI and IBM farms for data reconstruction and Monte Carlo processing.

- Reduction of the fully processed data into streams if necessary or into any other compressed form would also be carried out on special farms or on a central machine connected to robotics.

- There will be a few powerful machines (such as SGI Challenge XL) running the UNIX operating systems. On these central machines would be common resources where all DØ users can carry on various computing activities such as program development, large-scale data processing and reduction, and physics analyses. These machines would be connected by high speed switches/devices to common data-pools and to disks and robotics.

- There would be several small work-group servers, possibly one for each physics/detector group, equipped with some robotics. Relevant streamed data-sets can be provided via hierarchical storage managers to each of the work-group servers. One would go back to a central machine to access super-sets of data.

- The desktops would be, in the first approximation, windows to these various work-group servers and central machines. Users would have workstations, PCs or X-terminals on their desks.

The goal of the new model is to make the access to data easier and analysis computing more efficient. It should be possible to have enough storage on central or on work-group machines and provide facilities such as batch, archiving etc. on these common platforms. It is conceivable that we will have plenty of local resources on desktops as well and access to pieces of data on some centralized storage could be provided on demand from a central server via a "pick event" type facility. PC application software suites are to be served to the desktops and a pilot project Wincenter Pro has been quite successful at DØ. This scheme, we believe, would make systems maintenance and monitoring more manageable and result in efficient usage of resources.

5. Time-Scales for Migration

As per the lab's statement of direction which was issued in April 1995, the migration from VMS was to be completed in 3 years, which puts the target date at April 1998. The Computing Divison had asked the CDF and DØ collaborations to come up with their own plan for migration that would suit the time-scales of events in the respective collaboration, the date for completion of migration however being consistent with the target date stated above.

DØ is expected to complete the migration by the end of 1998. To come up with a realistic time schedule for the ramp-down of VMS clusters at DØ, we need to understand what the time-scales for migration of different activities should be.

It seems appropriate to lay down the following general ground rules:

- All Run-2 activities should be started on or moved to UNIX or NT systems.
- Any new Run-1 analysis activity should be started on UNIX systems.
- All active long term Run-1 analysis projects, that are not expected to be completed by the end of 1997 should be moved to UNIX platforms.
- Any active short-term Run-1 activity that could be completed by the end of year 1997 could be continued on VMS systems and brought to completion, if the analyzers so wish. This does not exclude the possible migration of such activity, if the individuals involved in the analysis so wish.

It is clear that the following factors need to be considered in setting time-scales for migration and decommissioning of VMS systems (described in the next section).

- Time-scales for various Run-1 physics analysis projects.
- Factors of getting ready for upgrade.
- Budget profile for next 2 fiscal years.

- Time-scales for certain models of machines getting obsolete. (Maintenance contract would be dropped for obsolete machines even if they are up and running).

6. Decommissioning of VMS Machines

Many of the VMS workstations that are being used at DØ as desktops are very old model VAXes and are already obsolete. Most of these have no maintenance contract on them anymore. The idea is to use them until they break down and replace them with a currently supported desktop version. However, we would like to have a plan to phase-out all these VAX machines in a systematic fashion in an effort to ramp down various VMS clusters at DØ. To achieve this, we propose the following phases of decommissioning of VMS machines(slightly modified from the proposal a year ago).

- July 1996-Dec. 1996: Decommission and replace all VAX machines such as M38 and lower (Done).
- Dec. 1997: Phase-out anything lower than model M60.
- July 1997 - April 1998 : Switch DEC Alpha machines from Open VMS to UNIX.
- July 1998: Decommission anything lower than model M90.
- By Jan 1999: Phase-out all VAXes. FNALDØ would be decommisioned by this time.

The phased-out machines will be replaced by UNIX workstations, PCs or X-terminals. The VMS era would be over when 1999 begins.

The VMS file server DØFS is expected to disappear during 1998. All of the new DØFIXed micro-DST data would be made available on central UNIX machines.

7. Controlling the Expansion of VMS clusters:

In order to complete the ramp-down of DØ VMS clusters and complete the migration, we have to control further expansion of VMS facilities. We therefore propose that the following policies be adopted.

- No new VMS node be added to any of the VMS clusters.
- Any new DEC Alpha machine brought in will be added to the DØDEC (DEC UNIX) cluster.
- No new peripherals will be added to any of the VMS machines, except possibly disks for physics project usage. We expect that disks that are currently on some VMS nodes will be moved to UNIX nodes as users migrate their activities.
- The system maintenance contracts will not be renewed for VMS nodes beyond April 1998. (except for some exempted ones that may be needed for specific purposes such as dØlibrary and Production Database.)
- No file-serving will be available to VMS clusters shortly after datasets are made available on central UNIX analysis machines. This is expected to happen in early part of 1998.
- After April 1998, any system maintenance of VMS systems on-site will be low priority. This will be required in order to more efficiently support the upgrade activities and a major population of the users who would be on UNIX machines.

8. The DØDEC cluster

Since the collaboration has a large number of DEC Alpha machines (bought in the last 3-4 years) both on-site and off-site, it is very important that we make the best use of them. It is also clear from feed-back from various collaborators that more Alpha machines will be bought in the near future since these are the most cost-effective machines on the market currently. So, a new DEC UNIX cluster has been started. Porting of the DØlibrary is in progress. (Thanks to A. Jonckheere, C. Lundstedt, H. Prosper, E. Smith and Dong Zhao for their efforts and help with this project.) It is important at this juncture that we get more volunteers to switch their Alpha machines to UNIX OS and help us find and fix problems. We expect to switch all the Alpha nodes on the current DØAXP OpenVMS cluster by April 1998.

9. Some guidelines for DØ institutions

- Machines that are to be part of an analysis/ work-group cluster should be running UNIX OS and should be one of the DØ supported platforms. They are DEC UNIX, SGI, and IBM.
- The desktops could be PCs, X-terminals or Workstations.
- Current thinking is that the CPU for large-scale processing will be provided by Fermilab (i.e., central machines). But, we anticipate that university groups would contribute to computing resources on work-group or analysis clusters.

Some discussions are necessary in order to develop a plan for the hardware acquisitons and university group contributions to such. They would be addressed and summarized in a future note, after such a plan is developed.

10. Summary

The statement of direction from the Fermilab Directorate for the lab to migrate from VMS systems to UNIX operating systems expects the lab-wide migration to be complete by mid-1998. The DØ collaboration has been rigorously pursuing the migration of analysis computing in the last two years and has made a lot of progress. The DØsoftware environment is fully established on UNIX platforms such as SGI and IBM. Currently efforts are underway to test and establish DØ software environment on DEC UNIX (Alpha) systems. We had discussed a migration plan and schedule at the DØ workshop in Boston last year and have updated the status and future course here. The migration is expected to be complete by the beginning of 1999.