AIIM

Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

Centimeter

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 mm

1 2 3 4 5

Inches

1.0 1.1 1.25 1.4 1.6 1.8 2.0 2.2 2.5 2.8 3.2 3.6 4.0 4.5 5.0 5.6 6.3 7.1 8.0 9.0 10.0

MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

To be published in *Supercollider 5*

SSCL-Preprint-252

Distributed Computing at the SSCL*

L. Cormell and R. White

Superconducting Super Collider Laboratory†
2550 Beckleymeade Ave.
Dallas, TX 75237

May 1993

*Presented at the Fifth Annual International Symposium on the Super Collider, May 6–8, 1993 San Francisco, CA.
†Operated by the Universities Research Association, Inc., for the U.S. Department of Energy under Contract No. DE-AC35-89ER40486.

DISTRIBUTED COMPUTING AT THE SSCL

Laird R. Cormell and Randle C. White

Superconducting Super Collider Laboratory*
Computing Department, MS-2003
2550 Beckleymeade Ave.
Dallas, TX 75237-3997
cormell@sscvx1.ssc.gov
rwhite@sscvx1.ssc.gov

ABSTRACT

The rapid increase in the availability of high performance, cost-effective RISC/UNIX workstations has been both a blessing and a curse. The blessing of having extremely powerful computing engines available on the desk top is well-known to many users. The user has tremendous freedom, flexibility, and control of his environment. That freedom can, however, become the curse of distributed computing. The user must become a system manager to some extent, he must worry about backups, maintenance, upgrades, etc. Traditionally these activities have been the responsibility of a central computing group. The central computing group, however, may find that it can no longer provide all of the traditional services. With the plethora of workstations now found on so many desktops throughout the entire campus or lab, the central computing group may be swamped by support requests. This talk will address several of these computer support and management issues by discussing the approach taken at the Superconducting Super Collider Laboratory (SSCL). In addition, a brief review of the future directions of commercial products for distributed computing and management will be given.

INTRODUCTION

Dramatic changes have occurred throughout the computing industry and its user community since the revolution of distributed desktop computing began in the 1980's. The abundance of cheap, powerful computing engines is driving this revolution at a feverish pace. Participants in the revolution are the high speed networks and distributed peripherals such as laser printers, SCSI disks, and 8-mm tapes that make distributed computing a viable option to the more traditional mainframe-oriented, centralized computing environments. Networked desktop computing has changed forever our computing and working environments. In fact, the sociology of how we do our work has also changed. Some of that change can be traced back to the interactive time-sharing environments of the 70's, of course. But in the 90's environment, a typical user is no longer dependent on the central computing organization to meet his needs. The computer has become another tool in the integrated office/work place of the 90's. Around the corner lie many new integrating tools in the area generally referred to as groupware that should further expand the utilization of computing in the office. These products include integrated phone, voice mail and e-mail services, desktop video conferencing, multi-media applications, etc. While a discussion of groupware is beyond the scope of this paper, planning for the management and support of these products must begin soon. There are many management issues to be addressed, but at this time there does not seem to be that many solutions readily available. We will provide examples of the methods employed by the SSCL to address these problems and then discuss some of the efforts under way in the computer industry.

*Operated by the Universities Research Association, Inc., for the U.S. Department of Energy under Contract No. DE-AC35-89ER40486.

DISTRIBUTED COMPUTING MANAGEMENT ISSUES

We've already alluded to the fact that there are both good and bad aspects of distributed computing environments. There are some simple, perhaps obvious, observations that we can make. First the advantages.

Advantages of Distributed Computing

Some of the advantages of distributed computing include:

- Distributed computing working environment is good. Most people do most of their work in their offices. Efficiency and productivity increase when computing and other resources are available in the office work place.
- Working group clusters are good. Most people work in groups on problems common to that group. A distributed environment that is sub-netted by group can be very effective in providing common shared resources that are group specific. For example, an engineering group may need CAD and finite element analysis packages, but have no need for the CERN Program Library. Client/server architectures can be utilized to their fullest capability with minimal effect on the network backbone.
- Dedicated resources are good. Placing a cheap RISC work-stations on every desktop can virtually eliminate time sharing. A typical user will have his own dedicated resource.
- Independence is good. Individual users and working groups relish the independence they have when in control of their own environments. They can choose configurations that are tailored to their needs rather than the general purpose environments found in centralized mainframe environments.
- The economics are good. First, RISC/UNIX workstations and personal computers, for that matter, are cheap. And as mentioned above, working groups or individuals buy only the applications and tools that meet their needs. They configure the hardware to meet their needs alone, no more. Shared multiple users licensing can keep the cost of software down. Installation and maintenance fees are typically small.

Disadvantages of Distributed Computing

There are likewise a number of disadvantages associated with distributed computing environments.

- Distributed computing working environment is bad. The individual user must become his own system manager to some extent. He must become concerned with issues of installation maintenance, updates, backups, etc. If he is not prepared to perform these functions himself, where does he turn, who provides the support?
- Working group clusters are bad. Individual working groups may chose to handle their own support issues. This can lead to duplication of efforts in several service areas. A further complication arises when different groups develop completely different solutions to the same problem.
- Dedicated resources are bad. When dedicated computing is provided for individual users, there are typically many compute cycles and other resources going to waste.
- Independence is bad. Individual users and working groups often don't have the experience or knowledge to configure their environment to meet their needs in an effective efficient manner.
- The economics are bad. While it is true that RISC/UNIX workstations are cheap, there are often hidden costs. There is a need to build up a support staff familiar with several UNIX (or PC) platforms. There are potentially several areas for duplication of efforts. The purchase of software and services may not be coordinated throughout the site and thereby lead to wasted expenditures.

So where does this leave us? While there are certainly some disadvantages associated with the management of distributed computing environments, most of the arguments come down in favor of such environments. The raw computing power, versatility, and freedom offered by distributed computing at an unrivaled price/performance point makes them extremely attractive. The "pros and cons" of centralized vs. distributed computing environments may be debated at length, but there is no escaping from the fact that the era of distributed computing is upon us.

Management Issues

In the era of distributed computing and heterogeneous environments, diverse systems are networked throughout the high engergy physics (HEP) community. Unfortunately, the diversity that lets users choose the system that best meets their needs also creates an administrative nightmare: it may require system administrators to develop different management scheme for each hardware platform. It is possible to integrate large multi-vendor systems when standards can be applied. These standards, however, are not all employed in the same way by every vendor. Most users enjoy the independence offered by RISC/UNIX workstations, but many of them still expect (and sometimes demand) the services normally offered by more centralized operations. System managers must find novel ways to deal with these expectations in multi-vendor configurations; they must address several issues including:

- Training. Support personnel must receive training for several different hardware platforms and operating systems. Users also need training. Clearly UNIX managers must provide training for their personnel, but who trains the users?
- Installation and setup. Workstation vendors are typically operating on small profit margins and may not provide free installation. Users often don't want to pay for the additional installation and setup services. But the typical physicist/user may have little or no experience in system management or much knowledge about UNIX environments. How do UNIX managers provide the support?
- Hardware maintenance. UNIX workstations are inexpensive and fairly reliable, but they still must be maintained throughout their life cycles. There are typically product enhancements, firmware fixes, and upgrades offered by the vendors that extend the useful life of the equipment. While system administrators are familiar with this phenomena, it's not clear that the end-user "keeps up" with the latest improvements or that he knows what to do. Who should provide the maintenance and ensure that the system is current?
- Software maintenance. Likewise for software upgrades, only more so. Bug fixes, patches, and new software versions are released on a regular basis. How does a UNIX manager install these on all platforms, how does he maintain configuration control, who tests new versions for compatibility?
- Network management. In many distributed environments the network is an integral part of computing, or as some have stated: "the network is the computer". In working group clusters files and applications are shared in a client/server arrangement; workstations and X-terminals are mounted to hosts; mail, ftp and other communication services are crucial. The HW and SW issues mentioned above apply to NW management as well. In addition there are questions of how to configure the network to provide optimized access to file servers, printers, etc.
- Distributed services. There are a number of services traditionally provided in centralized configurations that must be dealt with in new and innovative ways. How does a system administrator provide uniform printing, backup, licensing, batch, and tape services that users require?
- Accounting. How does an administrator in a distributed environment collect statistics on system utilization, disk usage, print service, and traffic over the network? At some sites the computer group provides resources on a "charge-back" basis. How should these charges be assessed in a distributed environment?
- Documentation. The usefulness of a system is often determined by the quality and accuracy of its documentation. How can the UNIX manager provide current documentation to all the users when some are located not only on the other side of campus but on the other side of the world?
- Host management. In a distributed environment the system administrator is faced with the difficulty of providing user and group accounts and privileges on a number of different hosts. In addition operations on the system such as shutdown/reboot, naming, addressing etc. must be performed.

Obviously there are a number of challenges to be faced in managing distributed systems. So, how do we cope with distributed computing? In the next section we discuss the approach that has been taken at the SSCL. In the section following that we will identify some new developments from the commercial side that may provide some solutions.

MANAGEMENT OF DISTRIBUTED COMPUTING AT THE SSCL

The SSCL is committed to developing a distributed computing environment [1]. Use of a centrally controlled but otherwise decentralized and distributed environment allows maximum functionality and flexibility of access to shared data, print services, electronic mail, backup systems and nameservice while minimizing network overloads across the primary backbone.

Operations support. The Informations Services Operations Group has 25 employees that provide all levels of support from Help Desk assistance to system programming on all central computing resources at the SSCL. They support five operating systems (UNIX, VMS, MS/DOS, Novell, and MAC-OS.) and approximately 3000 user accounts. The central resources include several VAXclusters, QuickMail servers, Appleshare servers, and Novell servers as well as the UNIX workstation clusters networked together in the Physics Detector Simulation Facility (PDSF). There are approximately 1500 Macintosh systems, 300 PCs and 700 UNIX workstations. Some groups (particularly those in Accelerator Systems Division) provide their own system support for specific functions.

Distributed servers. At present, management functions and critical applications are primarily performed using centralized server systems. The network impact of centralized data has reduced the system's efficiency somewhat. They plan to place a number of servers in strategic laboratory locations. These server locations are being selected on the basis of user population density and networking considerations. The critical applications and services to be distributed include the various file systems (AFS, NFS, Appleshare, and PC volume services), network applications (nameservice, NTP, and USENET), print services (IMPRINT, Alisashare, and Pathworks), mail gateway services, BITNET routing , and software metering for Macintosh and PC software permitting efficient sharing of infrequently used desktop applications.

Distributed Backups. The goal at the Lab is to provide administratively centralized, unattended backups on most SSCL servers by the end of '92. The volume has grown to greater than 250 GB of routinely backed-up data on UNIX, VMS, Macintosh, and Novell systems. In most cases the data is copied to 8-mm tapes in stacker units from remote systems over the FDDI backbone. System backups are currently performed using native operating system utilities. These utilities all fall short in some areas and none provide tape management or simple user archival capabilities.

Site Licensing. The SSCL is site-licensed for the SUN operating system and is in the process of procuring a site-license for SUN unbundled products. Procedures have been established for the distribution of copies of media and documentation from a centralized source to technical contacts throughout the SSCL. Other software that does not have licensing restrictions is made available through AFS or FTP. Client-server software (particularly that which is run on MACs or PCs) is a difficult problem. The SSCL is currently working on procedures to control and automate distribution and installation of client software through use of tools such as Timbukto for Macintosh.

Distributed Printing. The Operations Group has developed a distributed printing utility, lwprint, that permits access to most of the laser writers on AppleTalk or Ethernet. Print jobs are spooled and queued by a central UNIX print server. Users can select the printer of their choice from an interactive menu. Post script is the default format and several options such as landscape and two pages per sheet are being implemented [2].

FUTURE DIRECTIONS

Common Open Software Environment (COSE)

In a move towards application and end-user interoperability in a heterogeneous UNIX platform network, a unified model for a common open software environment was announced by Hewlett-Packard, Sun Microsystems, IBM, Univel, Santa Cruz Operation (SCO), and UNIX System Laboratories (USL) at the March, 1993 Uniform conference in San Francisco. The agreement includes specifications for a common desktop environment, networking, graphics, multimedia, object technology, and systems management.

Desktop Environment. The specification for the common desktop environment will be given to X/Open for insertion into the X/Open Portability Guide for interoperability and portability of applications among the various platforms. The specification will base a consistent graphical user interface (GUI) and a consistent applications programming interface (API) on X.11 Windows, OSF's Motif toolkit, SunSoft's ToolTalk, and HP's Encapsulator. Sun agreed to migrate from OpenLook to Motif although current OpenLook applications will still be able to run under the new common desktop environment.

Networking. A heterogeneous distributed workstation network environment with network transparency for applications will be supported on the various platforms through OSF's Distributed Computing Environment (DCE) and Distributed Management Environment (DME) as well as Sun's Open Network Computing (ONC+) environment and Univel's Netware.

Graphics. A common graphics application environment will be supported through the use of the X Consortium's facilities including Xlib/X for 2D pixel graphics, PEXlib/PEX for 2D/3D geometry graphics, and XIElib/XIE for advanced imaging.

Multimedia. A specification for common Distributed Media Services (DMS) and a Desktop Integrated Media Environment (DIME) will be submitted to the Interactive Multimedia Association. The multimedia environment will be an integrated part of the common desktop environment.

Object Technology. The Common Object Request Broker Architecture (COBRA) specification from the Object Management Group (OMG) will be supported for common object management across multiple heterogeneous platforms.

Systems Management. A working group will be set up to promote industry acceptance of specifications on systems management applications including security management, software installation, software license management, data backup and restore, print spooling, and distributed file system management.

SUMMARY AND CONCLUSIONS

High energy physicists are resourceful and creative. Given that, it's not surprising to see: first, that physicists, ever eager to obtain more computing power, have exploited RISC/UNIX systems with a passion; and second, that they have attempted, with some success, to develop integration and management tools to make effective use of these new systems such as those found in the SSCL UNIX environment.

But we should expect, in fact demand, tools and integrated packages from the workstation vendors. These products should be interoperable, coherent, scalable, and affordable. Fortunately, there are several promising efforts underway, by groups such as OSF and COSE, but how long can we wait and how successful will their products be?

References

[1] Brenda Ramsey, Information Services Department, SSCL, private communication.

[2] Dave Somogyi, Information Services Department, SSCL, private communication.

DATE

FILMED

9/8/93

END