N94-24413

1993 NASA/ASEE SUMMER FACULTY FELLOWSHIP PROGRAM

MARSHALL SPACE FLIGHT CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

PERFORMANCE OF THE ENGINEERING ANALYSIS AND DATA SYSTEM II COMMON FILE SYSTEM

Prepared By: Linda S. DeBrunner, Ph.D.

Academic Rank: Assistant Professor

Institution and Department: University of Oklahoma
School of Electrical Engineering

MSFC Colleagues: Marcellus Graham
Sheila Fogle

NASA/MSFC:

Office: Information Systems Office
Division: Systems Development and Implementation
Branch: Data Systems Branch

VIII

Introduction

The Engineering Analysis and Data System (EADS) was used from April, 1986 to July, 1993 to support large scale scientific and engineering computation (e.g. computational fluid dynamics) at Marshall Space Flight Center. The need for an updated system resulted in a RFP (2) in June 1991, after which a contract was awarded to Cray Grumman. EADS II was installed in February 1993, and by July 1993 most users were migrated.

EADS II (3) is a network of heterogeneous computer systems supporting scientific and engineering applications. The Common File System (CFS) is a key component of this system. The CFS provides a seamless, integrated environment to the users of EADS II including both disk and tape storage. UniTree software is used to implement this hierarchical storage management system. The performance of the CFS suffered during the early months of the production system. Several of the performance problems were traced to software bugs which have been corrected. Other problems were associated with hardware. However, the use of NFS in UniTree UCFM software limits the performance of the system.

The performance issues related to the CFS have led to a need to develop a greater understanding of the CFS organization. This paper will first describe the EADS II with emphasis on the CFS. Then, a discussion of mass storage systems will be presented, and methods of measuring the performance of the Common File System will be outlined. Finally, areas for further study will be identified and conclusions will be drawn.

EADS II

EADS II is a high performance computing network supporting scientific and engineering computing. The functions and implementation of EADS II are described in (2) and (3). The two key computing components of EADS II are the Vector Processor Compute System (VPCS) and the Virtual Memory Compute System (VMCS). The VPCS, a Cray Y-MP 8I/6128, is used for applications suitable for vector processing, while the VMCS, an SGI 4D/480, is used for applications with large memory requirements. In EADS I, the predecessor to EADS II, the VPCS needs were met by a Cray X-MP and the VMCS needs were met by an IBM 3084. Image processing applications are supported by the Image Processing System (IPS). The IPS consists of an SGI 4D/480 RE hub with 3 attached workstations. Mini-Supercomputers (MSCs) may be included at a future time to reduce the loading of the VPCS. Although there are no MSCs installed at this time, long term plans include the possibility of including small Cray Y-MP machines (Cray Y-MP 2E) to meet specific laboratory needs. These MSCs would be used for VPCS program development and for smaller applications.

A unique feature of EADS II is the integration of shared resources through the Common Output System (COS) and the Common File System (CFS). The COS provides printing capabilities for the users. Most printing facilities are located in the laboratories, while print queues are maintained on the VMCS. The Common File System (CFS), which provides hierarchical storage to all the EADS II machines, is the most interesting aspect of the EADS II architecture. Restoration of files to disk from tape is automatic. The CFS hardware consists of 2

IBM RS/6000-970 servers, 4 Maximum Strategy Disk Arrays (172 GB total), and 2 STK 4400 automatic cartridge libraries or silos (2.4 TB total). NSL UniTree software is used.

The CFS has 4 principal functions: Private Processor Storage (PPS), User File Storage (UFS), backup storage, and Archival Information Storage (AIS). The PPS consists of rotating magnetic disk storage (RMDS) and is used to store active user programs, operating system software, command procedures, and data. The UFS is RMDS which is allocated to users. The backup storage is used for routine backup of the PPS and UFS to tape. The AIS is used for long-term storage of information. Backup and archive management tools are also provided.

The EADS II computing components and shared resources are connected by a 3-level network. At the lowest level, Ethernet LANs connect systems within a building. Better performance is provided by the High Speed Network Backbone (HSNB), which uses the Fiber Distributed Data Interface (FDDI) technology. The HSNB provides access between central site and remote facilities. There are 2 FDDI rings which are interconnected by routers to each other and the building LANs. The highest level of performance is provided by the Back End High Performance Interconnect (BEHPI) network which is based on UltraNet. The BEHPI is used almost exclusively for moving data between central site computers and the CFS.

Mass Storage Systems

The IEEE-CS Technical Committee on Mass Storage Systems and Technology developed a "reference model" in the eighties which is used by manufacturers of mass storage products to describe the functions of their systems (1, 6). Although the reference model is not an IEEE standard, it is an important consideration in the development of mass storage systems.

The UniTree software is sold to companies by OpenVision (previously by DISCOS). The companies then port the code to their chosen platform. The product chosen to implement the EADS II CFS is NSL UniTree supplied by IBM. Most companies marketing UniTree products make modifications to improve performance or to add features. For example, Control Data Systems focuses on supporting a wide range of peripherals and has tuned their system to improve performance for various peripherals. On the other hand, Convex rewrote portions of the code that control the way the processes communicate. IBM has implemented Multiple Dynamic Hierarchies which allow multiple hierarchies on a single machine. They also have implemented a 3rd party transfer capability, called Network Attached Storage, which allows hosts to send data directly to the disk array without going through UniTree. Several other companies have developed mass storage systems including Epoch storage management tools, NetArchive, and Cray's Data Migration Facility.

Research facilities and universities have pioneered much of the work in the mass storage arena. For example, UniTree was developed at Lawrence Livermore National Laboratory (5). There are currently two mass storage systems developed by research facilities that are of particular interest--NAStore and AFS. NAStore, developed by NASA Ames Research Center, only blocks read operations until the first part of the data is available. So, for large files, access to the first byte of data is significantly faster. The Andrew File System (AFS) was developed by Carnegie-Mellon University to support distributed file access. It has been adapted by the Pittsburgh Supercomputing Center to include mass storage capabilities (4). AFS was chosen

since it is more scalable than NFS. NFS requires clients to communicate with the server to complete each transaction, but AFS maintains state information. Clients assume that they are using the most current version of a file's date until they are notified by a server. However, AFS was developed without consideration for the mass storage reference model.

Measuring the Performance of the Common File System

Three measurements of the CFS performance are currently being collected. All of these measurements are similar. Each measures the time required to perform several operations. None of these metrics generates statistics which can be readily compared to the expected performance or the performance of other systems. The principal function of these measurements is to identify degraded system performance relative to past system performance.

Every 10 minutes, Boeing Computer Support Services (BCSS) runs a script which checks for degraded system performance. This "10-Minute Metric" script measures the time required to change to a UniTree subdirectory ("cd") and list the directory contents ("ls"). In addition, Cray Grumman runs the "UNITREE Metric" hourly. Like the 10-Minute Metric, this metric measures the time required to perform simple file manipulations. It measures the time required to perform "ls", "ls -l", and to "tail" a file. Cray Grumman also runs a program every 3 minutes to check for degraded performance of the CFS. At this time, different programs are run on the VMCS and the VPCS. On the VPCS, the "3-Minute Metric" program measures the time required to open a file in a UniTree subdirectory and write a line to it. The corresponding program on the VMCS provides more complete information. It measures the number of NFS users, performs simple operations using NFS, and performs simple operations using FTP. Using NFS, the program performs a directory listing and copies a small file to a UniTree subdirectory. Using ftp, it "puts" a file in a UniTree subdirectory, performs a directory listing, and deletes the file. These measurements are inadequate for evaluating the overall performance of the CFS. A performance measurement tool is needed to allow EADS II to be compared to other systems.

Areas for Further Study

Several areas have been identified for future work. The most important is the development of a performance measurement tool. After measurement capabilities are developed, UniTree can be tuned to improve its performance in the EADS II environment.

The lack of knowledge about parallel processing on the SGI should also be remedied. By understanding the differences between parallel processing on the Cray and the SGI, users could be advised about the execution of their applications which are suited to parallel implementation. This should allow more users to use the SGI effectively.

Finally, a method of modeling networked computer systems should be investigated. This modeling would allow performance to be predicted before changes are made. Consequently, the effects of hardware changes and software load changes could be evaluated in a "what if" format.

Conclusions

The EADS II mass storage requirements are aggressive. Existing products have shortcomings with respect to these requirements. Since the EADS II CFS requires the most current technology, the efforts of the Storage System Standards Working Group will effect the future of mass storage technology. Awareness of standards will give system architects a better understanding of mass storage systems.

Current methods of measuring the performance of EADS II are inadequate. In the future, more meaningful measurements will be needed. As a beginning, EADS II should be evaluated using the tests run at Ames Research Center. In addition, a performance measurement tool tailored to the needs of EADS II should be developed. This tool will allow system administrators to evaluate the effects of hardware and software modifications, as well as changes in loading. It will also support comparisons with other mass storage systems.

Methods for modeling the system are needed to predict the effects of system modifications before implementation. Such a model will also support the analysis of predicted changes in loading. The model would allow various scenarios to be considered to choose the best solution.

Acknowledgment

I would like to thank Sheila Fogle and Marcellus Graham for providing feedback throughout this work. I would also like to thank Amy Epps being an unending source of useful information. Finally, I would like to thank Paul Allison for his support.

References

(1) Coyne, R. A., "An Introduction to the Mass Storage System Reference Model, Version 5," *Proceedings of the Twelfth IEEE Symposium on Mass Storage Systems,* Monterey, California, April 26-29, 1993, pp. 47-53.

(2) *Engineering Analysis and Data System II (Class VII Computer System), Request for Proposal,* MSFC, NASA, RFP #8-1-9-AI-00120.

(3) *Engineering Analysis and Data System II Users Guide,* MSFC, NASA.

(4) Goldick, J. S., Benninger, K., Brown, W., Kirby, C., Maher, C., Nydick, D. S., Zumach, B., "An AFS-Based Supercomputing Environment," *Proceedings of the Twelfth IEEE Symposium on Mass Storage Systems,* Monterey, California, April 26-29, 1993, pp. 127-132.

(5) McClain, F., "DataTree and UniTree: Software for File and Storage Management," *Proceedings of the Tenth IEEE Symposium on Mass Storage Systems,* Monterey, California, May 7-10, 1990, pp. 126-128.

(6) Miller, S. W., "A Reference Model for Mass Storage Systems," *Advances in Computers,* Vol. 27, Yovits, M. C., editor, pp. 157-210.