THE WORKSTATION CLUSTERING ENVIRONMENT AT NASA LANGLEY RESEARCH CENTER

Michael L. Nelson
NASA Langley Research Center
MS 157A, Hampton VA 23681
(757) 864-8511
m.l.nelson@larc.nasa.gov

David E. Cordner
NASA Langley Research Center
MS 157A, Hampton VA 23681
(757) 864-7325
d.e.cordner@larc.nasa.gov

Summary

This paper introduces the status of and lessons learned from the workstation clustering projects at NASA Langley Research Center: the *Borg*, a tightly coupled, homogeneous cluster and the *Hive*, a loosely coupled, heterogeneous cluster. Specifically, the role of the Open Software Foundation's Distributed Computing Environment and Distributed File Service (DCE/DFS) is discussed with emphasis on cluster functionality gained and lost through DCE/DFS. Finally, future directions for DCE/DFS for Langley Research Center and the rest of NASA are proposed.

Introduction

Workstation clusters are a common, well understood and integral component of scientific computing infrastructures. Clusters are defined as "a collection of computers on a network that can function as a single computing resource through the use of additional systems software" [1]. Workstations are commodity items that afford access to rapid technological improvement and decreasing unit expense, making them more appealing than traditional supercomputers for all but the most intensive applications.

In 1994, NASA Langley Research Center began a project to replace the Convex general purpose supercomputers from its Central Scientific Computing Complex (CSCC). A homogenous workstation cluster of 7 IBM 58H RS/6000's, named the *Borg*, was chosen for its attractive price/performance ratio. Since the Borg was the first new CSCC resource in several years, there was a desire to take advantage of nascent technology. A full description of the Borg project can be found in [2].

There was consensus among the Borg team that the key technology for a clustering system, and perhaps for the computing center at large, is a highly available, logically unified file system. When assessing available commercial file systems, there was reluctance to install NFS on new system due to well known performance and security problems [3]. The Andrew File System (AFS) [4] is in use at some computing centers, but we felt it was nearing the end of its product lifecycle, and were hesitant to use it for the Borg. More interesting was the AFS follow-on product , the Distributed File Service (DFS) from the Open Software Foundation (OSF). However, the OSF's Distributed Computing Environment (DCE) has to be installed before DFS can be used. We believe that the Borg represents the first known attempt to use DCE/DFS in a production cluster environment.

Distributed Computing Environment / Distributed File Service

The entire workstation clustering environment spearheads Langley's adoption of the Open Software Foundation's (OSF) Distributed Computing Environment / Distributed File Service (DCE/DFS) [5, 6]. DCE is a collection of 5 core services: Threads, Remote Procedure Calls, Directory, Security and Time Services. DCE management units are known as *cells*, which provide these core services to all computers within their cell. Applications such as DFS are built upon the core DCE services (Figure 1). While DCE has attractive features for management of large computer sites, DCE was primarily chosen so the cluster could utilize DFS.

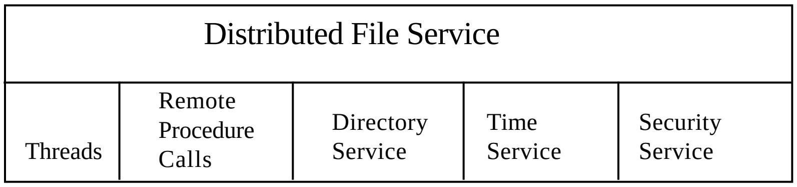

Figure 1: DFS Resides on the 5 DCE Core Services

Impact of DCE/DFS on Cluster Functionality

Introducing DCE into a cluster environment had the surprising effect of breaking many popular distributed computing tools, such as the PVM message passing library and most forms of schedulers and load balancers. The problem results from DCE requiring security credentials for a process on one machine, but currently lacking the capability to forward the credentials to another machine. For example, this would require typing a password for every machine that a PVM job may run on, and provides no good method for batch scheduling (typing passwords at midnight!) or automatic load balancing / redistribution (typing passwords when a job gets swapped to a different machine!). The insecurity inherent in storing passwords on disk is clearly unacceptable.

Lack of credential forwarding is a major drawback for a clustering environment. However, we choose to continue with DCE/DFS because we felt that the total cost of an initial DCE/DFS solution, including the caveats, would be less than an interim NFS or AFS solution (which do not have the credential limitations) that would have to be eventually ported to a DCE/DFS environment.

Langley analysts have produced a modified version of the rexec.d daemon to provide a temporary workaround for credential forwarding, until credential forwarding is properly supported in DCE. Other tools were developed as well: a DCE-aware wu-ftpd to ease the transfer of files across a DCE cell boundary, an integrated machine/DCE login for the vendors that did not provide it, and a host of scripts to assist transitioning a large user base from a standard Unix environment. Details of the various tools can be found in [7].

Current Cluster Architecture

The Borg consists of 7 IBM 58H RS/6000 workstation that are a dedicated general purpose resource for the user community. The Hive includes the Borg, as well as 2 dedicated Sun Sparcstation 20's, and a dynamic, heterogeneous collection of desktop and server class workstations from across the center that are available only when not utilized by their primary users. The relationship between the Borg, the Hive, and the DCE cells of Langley Research Center and ICASE is illustrated in Figure 2.

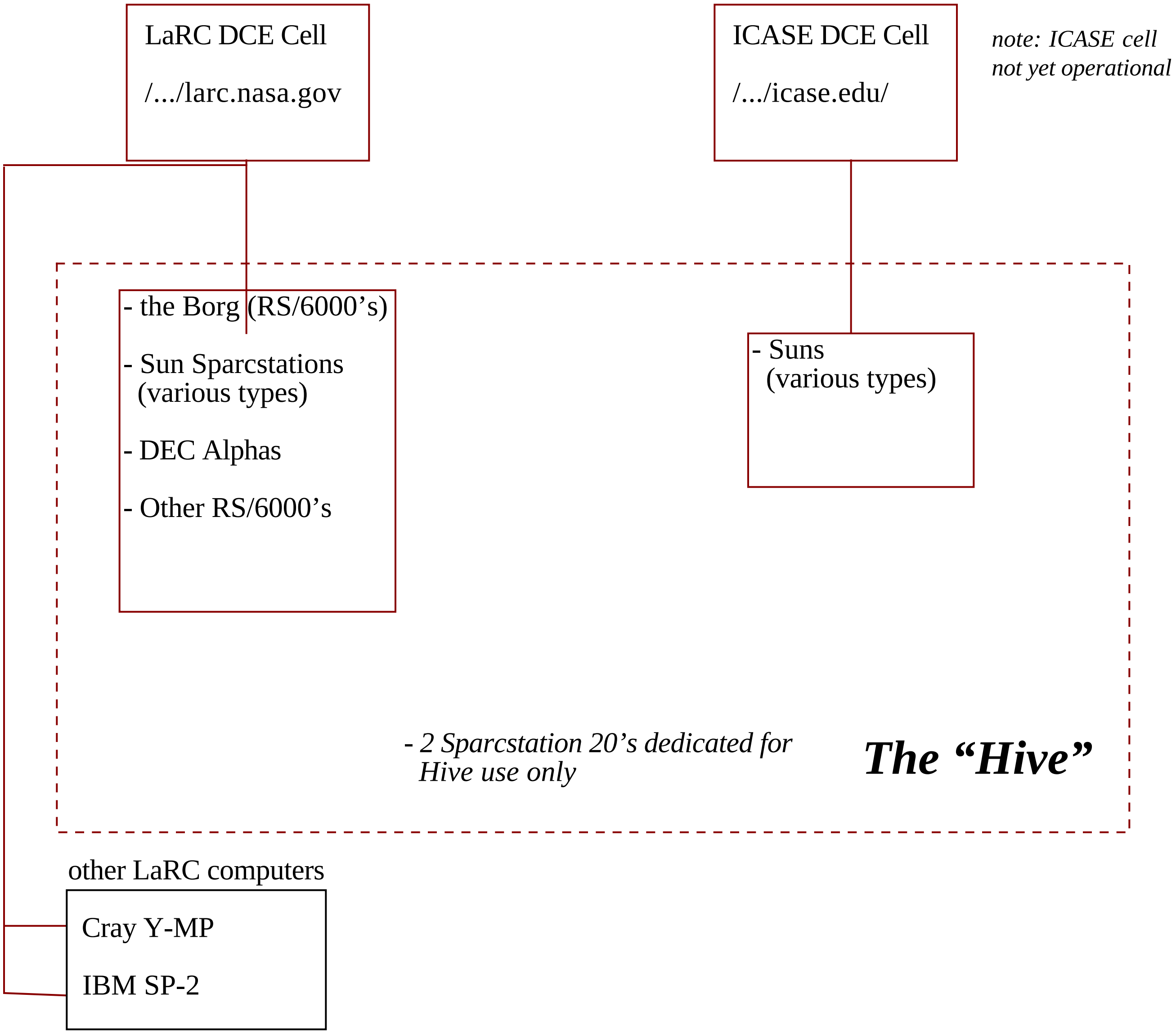

Figure 2: Borg, Hive and DCE Cell Architecture

Job scheduling and cluster management is performed using Platform Technology's Load Sharing Facility (LSF). LSF is used to control part time participation in the Hive and to distribute both parallel and non-parallel workloads. Platform Computing is currently optimizing LSF for performance in a DCE/DFS environment [9].

DFS Performance

Once DFS was installed and tuned on the Borg, two Langley structures codes (out-of-core solvers) were tested (Figures 3 & 4). The results are encouraging: DFS performs roughly twice as fast as NFS and only half as fast as a local file system. Other studies have found similar benefits [8].

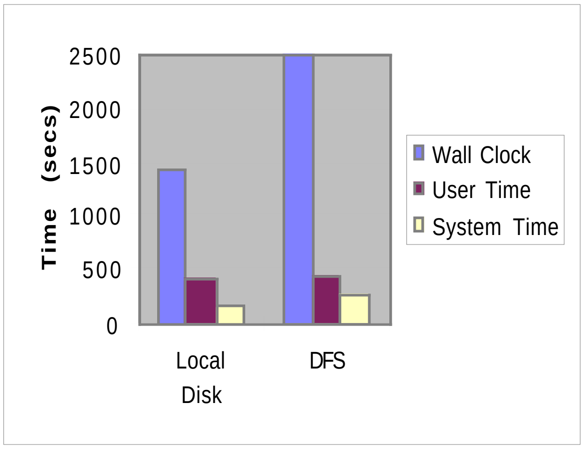

Figure 3: Sparse Solution of 255,000 Equations

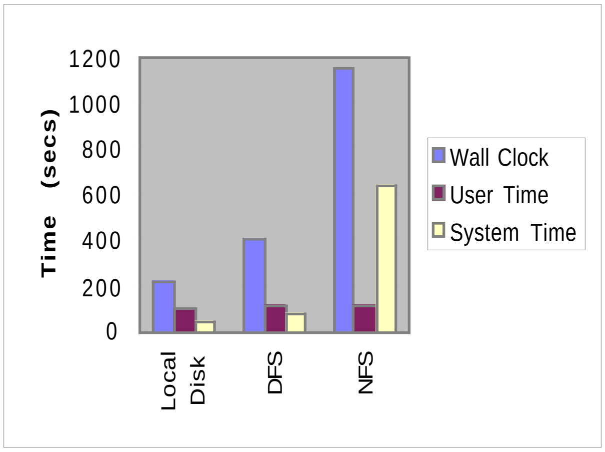

Figure 4: Sparse Solution of 88,000 Equations

DCE/DFS Configurations

The majority of our experience has been with IBM AIX and Sun Solaris, followed by the recent introduction of DEC Alphas. Because they do not have a large installed base at Langley, no HP's are currently part of the Hive. SGI's are scheduled to join the Hive pending arrival of a DFS client.

DFS servers are currently only available for AIX and Solaris, and run from $3,000 to $9000, depending on which class workstation they are purchased for. We have a total of 11 DFS servers, and 30 clients. To date, we have not encountered a problem with too few servers and too many clients, and we do not know where that boundary is. Table 1 shows the availability of DFS clients.

Workstation	*DFS Client*
IBM	Bundled in AIX 4.2
HP	Bundled in HPUX 10.X
DEC	Bundled in Digital Unix
Sun	~ $300
SGI	Not out yet; will cost ~ $400

Table 1: Availability of DFS Clients

Future DCE/DFS Plans

Within Langley, we intend to proliferate DCE/DFS by providing clients to all Unix workstations, as well as PC's and Macintoshes. For example, the project outlined in Figure 5 will provide all administrative computing packages (e.g., Microsoft Office) on a central Windows NT server, which will have write access into DFS. Through the use of WinCenter Pro, all Unix, PC's, and Macintoshes will have access to these packages through use an X Window System server (Langley has a site license for MacX and PC eXceed). DFS will also have access to the High Performance Storage System [10].

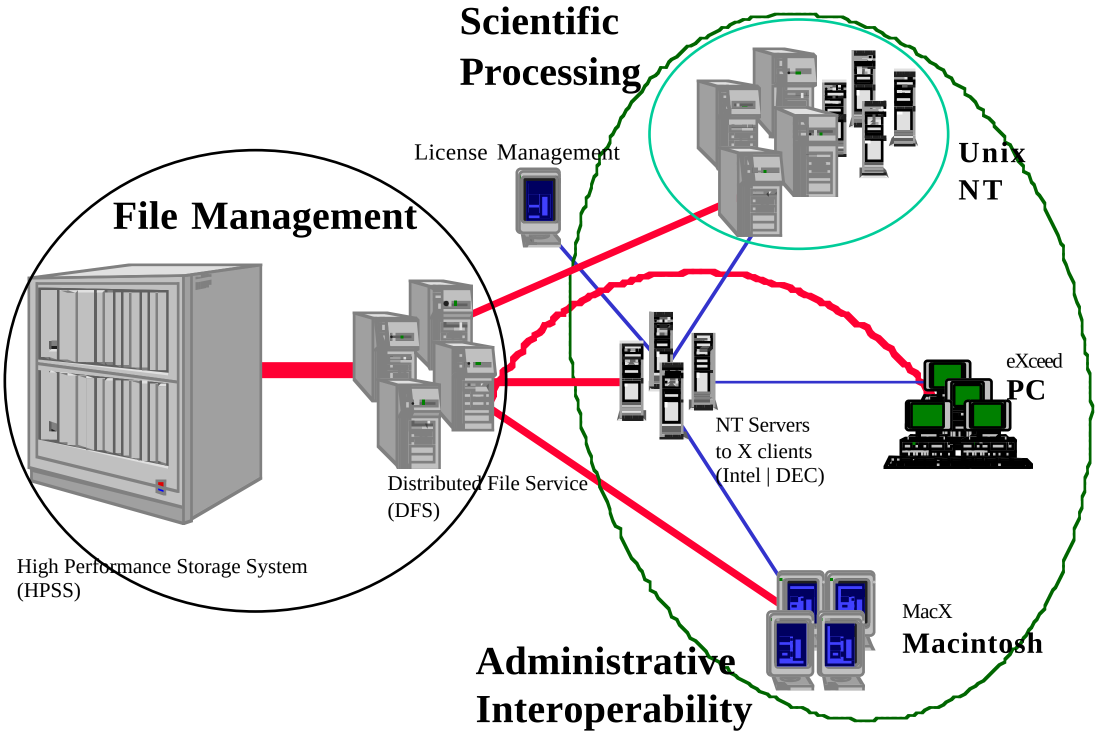

Figure 5: DFS and WinCenter Pro for Scientific and Administrative Computing Integration

Although current versions of DCE/DFS do not support it, the eventual goal is to use DCE/DFS to provide integration of cells across the various NASA centers and research partners. We anticipate this capability being available through the OSF in the 1998 time frame.

Conclusions

The Borg and the Hive prove that it is possible to implement DCE/DFS in a production workstation cluster environment. Although DCE/DFS currently reduce the functionality of a of a workstation cluster and require more involved installation, DCE/DFS provides the path to performance and security that modern computing architectures need.

Acknowledgments

DFS performance data is courtesy of Ron Gillian, Dr. Olaf Storaasli, and Dr. Majdi Baddourah.

References

1. Kaplan, J. A. and Nelson, M. L.: “A Comparison of Queuing, Cluster, and Distributed Computing Systems,” NASA TM-109025 (Rev. 1), June 1994.

2. Nelson, M. L. and Cordner, D. E.: “The Workstation Clustering Program at NASA Langley Research Center,” NASA TM (in preparation)
working version at: http://www.larc.nasa.gov/~mln/cluster-program.ps

3. Garfinkel, S.; Weise, D.; and Strassman, S.: “The UNIX Hater’s Handbook,” IDG Books Worldwide, Inc., San Mateo, CA, 1994.

4. Howard, J. H., “An Overview of the Andrew File System,” *Proceedings of the USENIX Winter 1988 Technical Conference*, Dallas, TX, 1988, pp. 23-26.

5. Rosenberry, W.; Kenney, D.; Fisher, G.: “Understanding DCE,” O’Reilly and Associates, Sebastopol, CA, 1992.

6. Open Software Foundation, “OSF Distributed Computing Environment”, *http://www.osf.org/dce/*

7. Nelson, M. L.; Priest, T. L., and Bianco, D. J.: “Experiences with DCE/DFS in a Production Workstation Cluster Environment,” NASA TM (in preparation)
working version at: http://www.larc.nasa.gov/~mln/dfs/dfs.html

8. Gaffey, B.; Kimlinger, P.; Lord, S.; Mostek, J.; and Reinhart, J.: “The Performance of OSF DCE Distributed File Service (DFS) at Cray Research, Inc.,” Cray Technical Report, 1994.
http://www.cray.com/PUBLIC/product-info/sw/dce/perf.html

9. Affordable High Performance Computing (AHPC) Research Consortium,
http://www.lerc.nasa.gov/Other_Groups/NPSS/html/can95.html

10. Coyne, R.; Hulen, H.; Watson, R.; “The High Performance Storage System,” *Proceedings of Supercomputing ‘93*, Portland, Oregon, November 15-19, 1993.