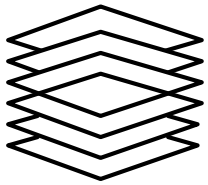


Sun and VERITAS Cluster-Wide File Systems, Cluster-Wide Volume Management, and Cluster Failover Software Compared

EXECUTIVE SUMMARY

Sun Microsystems and VERITAS software[1] are two of the major software vendors[2] who provide cluster software. There are many features and functions offered by such cluster software. These features and functions may be examined in many ways to varying degrees of granularity. This report is the latest in such a continuing series of such analyses.

The report examines a particular subset of current cluster software functionality from Sun and VERITAS both of whom enjoy substantial cluster-user mind share. The subset only includes those features and functions that address cluster-wide file systems and cluster-wide volume management. The study also examines, only at high level, certain cluster failover software functionality.

For Sun Microsystems, the shipping software examined includes the Sun Cluster 3.0 Global File System with its base UFS file system for a cluster-wide file system. It also includes the Solaris volume manager, Solstice DiskSuite, for cluster-wide volume management. Finally, for cluster failover, there is Sun Cluster 3.0 (Table 1).

For VERITAS Software the product set examined is the SANPoint Foundation Suite – HA (SPFS – HA). SPFS – HA is VERITAS' integration of the base VERITAS File System (VxFS) with cluster file system extensions, the base VERITAS Volume Manager (VxVM) with cluster extensions, and the VERITAS Cluster Server (VCS). For cluster failover, the capabilities of VCS are evaluated (Table 1).

Vendor	Cluster File System/ Underlying File System	Cluster Volume Manager/ Underlying Volume Manager	Cluster Failover Software
Sun Microsystems	Sun Cluster 3.0 Global File System/Solaris 8 UFS	Solstice DiskSuite 4.2.1	Sun Cluster 3.0
VERITAS Software	VERITAS FileSystem (VxFS) with Cluster Extensions 3.4	VERITAS Volume Manager (VxVM) with Cluster Extensions 3.2	VERITAS Cluster Server (VCS) 2.0

TABLE 1:
Product Sets Evaulated

Evaluations often present results using a single data view where one list of data elements (i.e., evaluation criteria) is divided into N categories (or areas). This

[1] This D.H. Brown Associates, Inc. (DHBA) report was co-sponsored by VERITAS and the DHBA High Availability Software and Hardware/Clusters service.

[2] Sun Microsystems is also a leading hardware vendor.

report employs DHBA's methodology for grouping a single set of data elements into any number of different views, each divided into any number of areas. Two data views are presented in this report.

In View 1, the data is organized by product area. This organization means that the features and functions examined are grouped into three product areas – cluster file systems, cluster volume managers, and cluster failover software. This view presents the details of what features are, and are not, being delivered. Vendor scores are based on the percentage of features delivered in the DHBA feature subset list. VERITAS scores highest in this view in the three categories examined (Table 2).

TABLE 2: Vendor Position Results Organized by Product Area

Product Area	Sun Microsystems	VERITAS Software
Cluster File System	2nd	1st
Cluster Volume Manager	2nd	1st
Cluster Failover Software	2nd	1st

View 2 uses the same feature and function data elements given in the first view but organizes them into seven functional areas. These functional areas include planned downtime reduction, unplanned downtime reduction, hardware efficiency, heterogeneous operations capability, manageability, performance speed-up, and performance scale-up. VERITAS scores highest in this view in five of seven areas and ties in one (Table 3).

TABLE 3: Vendor Position Results Organized by Functional Area

Functional Area	Sun Microsystems	VERITAS Software
Availability – Planned Downtime Reduction	2nd	1st
Availability – Unplanned Downtime Reduction	2nd	1st
Hardware Efficiency	1st	2nd
Heterogeneous Operations	1st (Tie)	1st (Tie)
Manageability	2nd	1st
Performance – Speed-Up	2nd	1st
Performance – Scale-Up	2nd	1st

Each of the report's examined functionalities – cluster file system, cluster volume manager, and cluster failover software – is actually a summary of various features and functions within these functionalities. For example, in reviewing the features and functions of cluster file systems, the report found that both vendors support at least four CFS nodes, journaling of file-system metadata, guaranteed cache coherency, parallel FSCK, and commodity interconnect. But, neither vendor allows greater than 16 CFS nodes or the ability to scale up metadata I/O capacity by adding CFS nodes. Neither supports mixed indirect and direct CFS I/O.

Based on the detailed cluster file-system analysis, the report identified greater support for CFS nodes as an opportunity for improvement for Sun. Sun has also yet to address direct CFS I/O from multiple Sun cluster nodes. VERITAS could improve in this category by delivering file ACLs, as well as allowing indirect CFS access from some nodes while allowing direct access from others.

A cluster volume manager, the second category of criteria, provides administrators and applications running on multiple nodes a single view of a common set of volumes. In fact, they are an enabler of cluster file systems. Twenty-five criteria were used including RAID 0, RAID 1, RAID 5, third mirror breakoff, online performance monitoring, rolling volume-manager upgrades, and GUI-level user authentication.

The report shows that both vendors deliver RAID 1, Host Spare Disk, mirrored pair re-resynchronization using a difference map, multiple active paths to volumes, third mirror breakoff, and readable and writeable volume snapshot. Both products feature online performance monitoring/tuning, and volume management from UNIX clients. Both also offer customizable GUIs and various other features related to cluster volume managers.

Sun could improve its standing in this area by delivering online RAID-level configuration as well as rolling volume-manager upgrades. VERITAS offers 24 of the 25 features in this category, but has yet to address CVM-file system independence in its SANPoint Foundation Suite – HA. Additional details on these categories and a complete discussion of cluster failover software, the third category, can be found in the body of the report.

The report provides a variety of other useful data. For example, in one outcome of the analysis in View 2, the study found that Sun and VERITAS tied with a score of 50% in the functional category covering heterogeneous operations. Each product met half of the criteria. But neither vendor offered any of the criteria met by the other. There are real opportunities here for both vendors for product improvement and improved service to the market by offering features and functions not currently available.

This analysis does not weight data elements and each data element contributes the same amount to the score in an area. Readers must consider the significance of each feature to their own requirements when comparing vendors. Furthermore, this report does not provide composite scores, for example, in View 2, with different weighing factors applied to each of the seven presented categories. This, too, can be done by DHBA for specific IT environments with emphasis placed accordingly based on user needs.

Vendor scores in some case are very close. Readers should examine the underlying vendor product functional differences described in conjunction with their own requirements. In this latter regard, please note that the results presented could change in a significant manner with the next release of a product version.

Understand that the scores in each category do not quantify, in any absolute or relative manner, a vendor's overall ability in any area. For example, scores in the "Performance – Speed-Up" section do not quantify actual performance. Therefore, a vendor scoring X percent higher than another does not, by virtue of this analysis, deliver X percent higher performance based on some performance measure.

TABLE OF CONTENTS

INTRODUCTION

This D. H. Brown Associates, Inc. (DHBA) report presents the results of a study, which reviews a subset of cluster functionality.[3] Each vendor's product set includes products the vendor delivers (and supports), and that operate together, to enable cluster-wide file-system access, cluster-wide volume management, and cluster failover (one of the aspects of high availability). The subsets that form the basis of this report include the details of a specific product set from Sun Microsystems,[4] which is currently available and shipping, and a product set from VERITAS Software that addresses cluster-wide file systems and cluster-wide volume management. The study also examines, at a high level only, cluster-software failover functionality. The product sets selected for evaluation are a representative case. These products are currently shipping, based on DHBA's understanding of current products.

TABLE 4: Vendor Position Results Organized by Product Area

Product Area	Sun Microsystems	VERITAS Software
Cluster File System	2nd	1st
Cluster Volume Manager	2nd	1st
Cluster Failover Software	2nd	1st

Table 1 organizes the data by product area. As shown above, VERITAS ranks first in all three areas. Table 2 below organizes the data by functional area – breaking the features and functions into seven categories. In this view, VERITAS ranks first in five categories. Sun ranks first in hardware efficiency and the vendors tie in heterogeneous operations.

TABLE 5: Vendor Position Results Organized by Functional Area

Functional Area	Sun Microsystems	VERITAS Software
Availability – Planned Downtime Reduction	2nd	1st
Availability – Unplanned Downtime Reduction	2nd	1st
Hardware Efficiency	1st	2nd
Heterogeneous Operations	1st (Tie)	1st (Tie)
Manageability	2nd	1st
Performance – Speed-Up	2nd	1st
Performance – Scale-Up	2nd	1st

[3] This is in contrast to the usual fine-grained DHBA cluster functionality report that examines many hundreds of features and functions and provides a complete picture of cluster capabilities. The latest of these reports was published in July 2001 (see: *Single-System High Availability,* D.H. Brown Associates, Inc., July 2001). The spreadsheet used as the basis for this referenced report and this document are proprietary to DHBA and are only available to selected DHBA clients.

[4] Throughout this report, vendors are listed in alphabetical order.

For Sun Microsystems, this product set includes the Sun Cluster 3.0 Global File System with its base UFS file system for a cluster-wide file system. It also includes the Solaris volume manager, Solstice DiskSuite, for cluster-wide volume management. Finally, for cluster failover, there is Sun Cluster 3.0. For VERITAS Software the product set examined is the SANPoint Foundation Suite – HA (SPFS – HA) Version 3.4 for Solaris. SPFS-HA is VERITAS' integration of the base VERITAS File System (VxFS) with cluster file-system extensions, the base VERITAS Volume Manager (VxVM) with Cluster Extensions, and the VERITAS Cluster Server (VCS). For cluster failover, the capabilities of VCS are evaluated. For each vendor, it may be possible to assemble a different, or alternative, product set to provide functionality similar to the functionality evaluated in this report.

TABLE 6:
Product Sets Evaluated

Vendor	Cluster File System/ Underlying File System	Cluster Volume Manager/ Underlying Volume Manager	Cluster Failover Software
Sun Microsystems	Sun Cluster 3.0 Global File System/Solaris 8 UFS	Solstice DiskSuite 4.2.1	Sun Cluster 3.0
VERITAS Software	VERITAS FileSystem (VxFS) with Cluster Extensions 3.4	VERITAS Volume Manager (VxVM) with Cluster Extensions 3.2	VERITAS Cluster Server (VCS) 2.0

In fact, Sun has new software releases in the works that are expected to have a great impact on its scores. (VERITAS does also). In any case, readers are again cautioned that they must examine the details in this report to understand which exact features and functions are present or not present, and then determine whether their IT environment is affected or not.

Note also that only software products are evaluated. The study examines how the software implementations affect the manner in which hardware is configured and what hardware is required and or supported for the complete hardware and software solution. For a more detailed discussion of the methodology used, please see Appendix A.

PRODUCT AREA DETAILS

Table 1 summarized results by product area. The next three major sections of this report present the first DHBA view of the data organized by product area (View 1). The examined features are grouped into three critical areas: cluster file systems, cluster volume managers, and cluster failover software. These sections present details of which features are, or are not delivered. Vendor scores are based on the percentage of features delivered.

CLUSTER FILE SYSTEM (CFS)

Cluster file systems provide applications running on multiple nodes with a single view of a common file system.

Cluster file systems operate in conjunction with certain underlying file-system implementations (i.e., UFS, VxFS, and CD-ROM file systems). Since the underlying file system(s) supported with a cluster file system may vary for each vendor, this section examines the features associated with just one specific underlying file system, as well as the features that "clusterize" that file system.

Figure 1 summarizes vendor scores in cluster and underlying file systems. (The vertical axis is "percent" with the maximum score of 100.)[5]

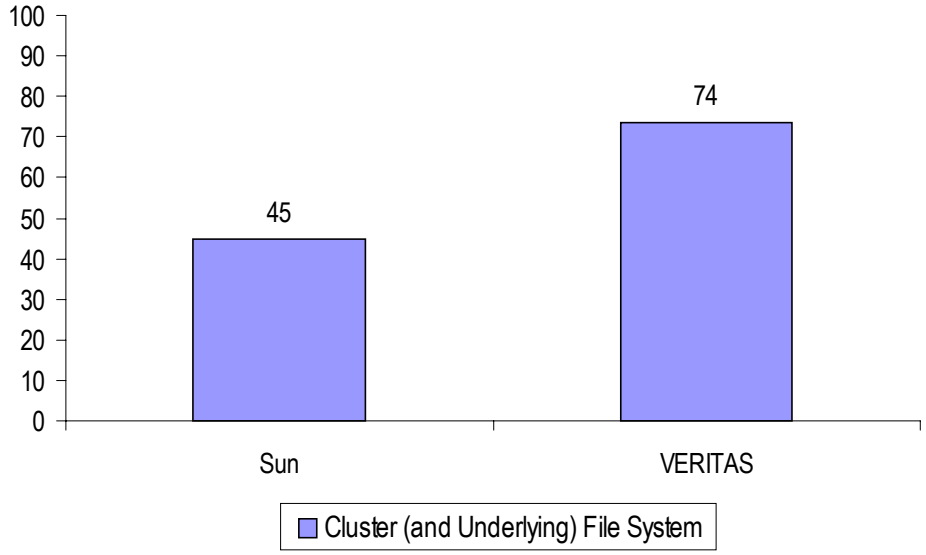


FIGURE 1: Cluster and Underlying File-System Score Comparison

Lists 1, 2, and 3 provide the detailed criteria used to obtain the results shown in Figure 1. The criteria are not weighted. The list entries are consecutively numbered since each entry contributes equally to the result shown in Figure 1. List 1 shows the criteria related directly to clusterizing an underlying file system.

[5] This is the case for all the figures.

LIST 1: Criteria Associated with Clusterizing an Underlying File System

1. Number of Nodes Greater than Four
2. Number of Nodes Greater than Eight
3. Number of Nodes Greater than 16
4. Number of Nodes Greater than 32
5. Cache Consistency Guaranteed across Nodes
6. Direct (Physical) Read/Write (R/W) Access from All Nodes
7. Indirect (Served) R/W Access from Some Nodes
8. Mixed Direct/Indirect R/W Access from Nodes
9. Lock Manager Capacity Scales with Number of Nodes (per File System)
10. Data I/O Capacity Scales with Number of Nodes (per File System)
11. Metadata I/O Capacity Scales with Number of Nodes (per File System)
12. Commodity Node Interconnect Supported
13. CFS Is Both Readable and Writeable
14. CFS Is Volume Manager Independent
15. CFS Is Underlying File-System Independent
16. CD-ROM Accessible as CFS
17. Tape Accessible as CFS
18. Dynamic Selection of Backup Node for Primary/Master upon Failure
19. Recovery of Primary/Master Software Component Transparent to Applications

List 2 shows the criteria used for scoring the underlying file system. These are features or options that affect – or may be selected for – an entire file system (as opposed to individually selected files).

LIST 2: Criteria for Scoring the Underlying File System

20. File System Journaling
21. File System Failures/Errors Do Not Cause Operating-System Panics
22. File System ACLs (Access Control Lists)
23. File Cache Flushed on File Close
24. FSCK Multiple Volumes in Parallel
25. Online Growth
26. Online Shrink
27. Online Defragmentation
28. Freeze and Thaw
29. Can Force Unmount of File System while Files Are Open
30. Writes Above a User Defined Size Are Performed as Direct (Physical) I/O
31. Maximum I/O Size Adjusted (Increased) for RAID 0/3/5 Stripe Size, Number of Stripes

List 3 shows features or options associated with the underlying file system, and which may be selected on an individual file basis.

LIST 3:
Criteria Associated with the Underlying File System that May Be Selected on an Individual File Basis

32. Convert All I/O to Direct I/O
33. Convert All I/O to Synchronous I/O
34. Convert All I/O to Data Synchronous I/O
35. Access Time Updates May Be Turned Off
36. Extent-Based Allocation Based on User Defined Extent Size
37. Extent-Based Allocation Based on File I/O Pattern
38. Can Reserve a User-Defined Amount of File System Space for a Given File

OVERVIEW OF CLUSTER FILE-SYSTEM RESULTS

Both product sets in the report support at least four CFS nodes, journaling of file-system metadata, guaranteed cache coherency, parallel FSCK, and commodity interconnect. Both product sets also provide for dynamic selection of backup for primary/master, and transparent recovery of primary/master. Further, both allow access-time updates to be turned off for files, and the ability to convert all I/O to direct I/O. It should also be noted that both vendors support indirect file-system access using NFS, even though this was not evaluated in this report.

On the downside, neither vendor's product set allows greater than 16 CFS nodes, or the ability to scale-up metadata I/O capacity by adding CFS nodes. Neither support mixed indirect and direct CFS I/O.[6]

The products differ in various key areas – support for greater than eight CFS nodes, file-system failures without operating-system panics, file ACLs, file cache flushed on close, direct and indirect read/write I/O to CFS, and lock manager /data I/O capacity increases with additional nodes. Further, differences exist in online file-system expansion and shrinking.

Additional differences exist relative to online file-system defragmentation; CFS volume manager and underlying file-system independence; CD-ROM and tape accessibility via CFS; file-system freeze and thaw; forced unmount of file system while files are open; writes above a user-defined size performed as direct I/O; and maximum I/O size adjusted for RAID 0/3 stripe set size and/or number of stripes. Finally, conversion of all I/O to synchronous (and to data synchronous) I/O, extent based allocation based on I/O pattern (and user-defined extent size), and reservation of a specified amount of file-system free space for a file, are also areas of differentiation.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

In the cluster file-system area, Sun scores second delivering 17 of 38 features, or 45%. The Sun Cluster 3.0 Global File Service (GFS) supports up to eight nodes. The master node delivers direct access to the CFS while others access the CFS

[6] In this study, "mixed indirect and direct CFS I/O" implies the flexibility to configure N nodes doing indirect I/O and M nodes doing direct I/O, where either N, or M, or both, may be greater than two.

indirectly. Requests are transmitted to the master node over the node interconnect; the master node performs direct I/O; and data is returned to requesting nodes over the node interconnect.

The Sun Cluster 3.0 Global File Service is designed to operate with any underlying file system and/or volume manager that uses the UNIX v-node interface. However, the Global File Service currently only supports UFS. Sun may support other file systems – such as VxFS – to operate with the Global File Service in the future. Moreover, the Global File Service enables CFS access to tape and CD-ROM devices attached to Sun Cluster nodes. Finally, Sun's underlying UFS file system supports file ACLs.

Sun could improve by offering support for more CFS nodes. In addition, Sun has yet to address direct CFS I/O from multiple Sun Cluster nodes. Sun could also improve by enabling scalable lock manager and data I/O capacity with the addition of cluster nodes, as well as isolating the operating system from panics caused by a CFS or underlying file-system failure or error.

Sun's underlying UFS file system has yet to address file cache flushed on close, online defragmentation, freeze/thaw, unmount of file system while files are open, writes above defined size performed as direct I/O, and stripe size adjusted based on stripe set size and/or number of stripes. Sun could also improve by enabling the UFS to support conversion of all I/O to synchronous and data synchronous I/O, extent-based allocation, online file system expand and shrink, and file system free-space reservation.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS scores first in cluster file systems delivering 28 of 38 features, or 74%. The evaluated SANPoint Foundation Suite – HA supports up to 16 CFS nodes using direct I/O (direct attachment to storage from all CFS nodes). With the SANPoint Foundation Suite – HA, lock-manager capacity and data I/O capacity increases with additional nodes. Furthermore, with SPFS-HA, underlying file-system failures and errors are isolated so that operating-system panics do not result.

The VxFS file system supports file cache flushed on close; online file-system defragmentation/expand/shrink; freeze and thaw; and file system unmount while files are open. VxFS supports conversion of writes above a user-defined size to direct I/O; adjusted maximum I/O size based on stripe size and number of stripes; and conversion of all I/O to synchronous I/O and to data synchronous I/O. VxFS also features extent-based allocation, which relies on I/O pattern and user-defined extent size. Finally, VxFS allows a user-specified amount of file-system free space to be reserved on a per file basis.

VERITAS could improve by delivering file ACLs, as well as the ability to access the CFS indirectly from some nodes while directly from others. (SANPoint

Foundation Suite – HA supports only direct I/O.) Therefore, all CFS nodes require a physical connection (i.e., I/O adapter(s) and cable) to disk drives that house the file system. VERITAS has also yet to allow underlying file systems other than VxFS, and underlying volume managers other than VxVM, to operate with SANPoint Foundation Suite – HA.

CLUSTER VOLUME MANAGER (CVM)

The second product area category is cluster volume manager features, which provide administrators, and applications running on multiple nodes a single view of a common set of volumes. Cluster volume managers are an enabler of cluster file systems and operate in conjunction with some underlying volume-manager implementation (i.e., Solstice DiskSuite, VxVM). The underlying volume manager supported with a given cluster volume manager may vary. For each vendor, this section examines features of one specific underlying volume manager, as well as features that clusterize that volume manager.

Figure 2 summarizes the cluster and underlying volume manager analysis.

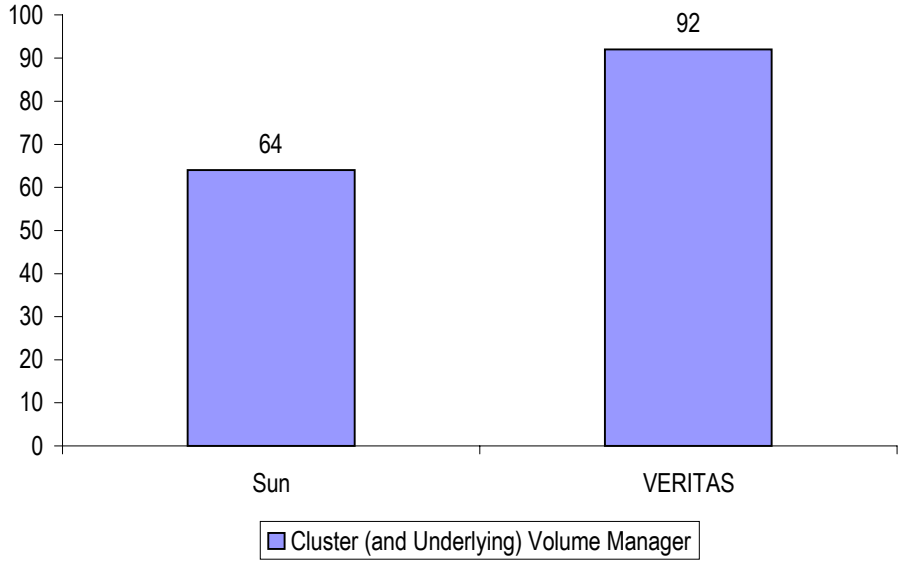


FIGURE 2:
Cluster and Underlying Volume Manager Score Comparison

List 4 shows the criteria for scoring a cluster volume manager in its ability to clusterize a Volume Manager. Note that the entries in Lists 4, 5, and 6 are consecutively numbered since they are the three parts of the results shown in Figure 2.

LIST 4:
Criteria for Scoring a Cluster Volume Manager in Its Ability to Clusterize a Volume Manager

1. Can Specify Member of Mirrored Pair for Reads – Can Specify by Node
2. Common Volume Names across Nodes

Note that the entries in Lists 4, 5, and 6 (below) are consecutively numbered since they form the three parts of the results shown in Figure 2. List 5 shows the specific options and features reviewed.

LIST 5: Criteria for Scoring a Cluster Volume Manager Specific Options and Features

3. Online RAID Level Reconfiguration
4. RAID 0
5. RAID 1
6. RAID 0+1
7. RAID 1+0
8. RAID 5
9. Hot Spare Disk
10. Host Unspare Disk (Restore Prefailure Configuration)
11. Mirrored Pair Resynched Using Difference Map
12. Multiple Active Paths between Server and Volumes
13. Third Mirror Breakoff
14. Volume Snapshot Copy (Point-in-Time)
15. Volume Snapshot Copy Is Both Readable and Writeable
16. Online Performance Monitoring
17. Online Modifications to Tuning Parameters
18. CVM Is File System Independent
19. Rolling Volume-Manager Upgrades
20. Can Specify Member of Mirrored Pair for Reads

List 6 shows the administrative interface capabilities.

LIST 6: Criteria for Scoring Cluster Volume Manager Administrative Interface Capabilities

21. Customizable Volume Management Graphic User Interface (GUI)
22. GUI-Level User Authentication
23. Management from UNIX Clients
24. Management from NT Clients
25. View Multiple Property Sheets (Objects) Simultaneously

OVERVIEW OF CVM RESULTS

Both vendors' product sets deliver RAID 1, Host Spare Disk, mirrored pair resynchronization using difference map, multiple active paths to volumes, third mirror breakoff, and readable and writeable volume snapshot. Both products feature online performance monitoring/tuning and volume management from UNIX clients. Both have customizable GUIs, GUI-level user authentication, the ability to specify which half of a mirrored volume to read from, and the ability to view multiple property sheets (objects) simultaneously.

The examined product sets differ in support for online RAID-level reconfiguration, RAID 0+1, RAID 1+0, disk unspare, CVM file-system independence, rolling volume-manager upgrades, and the ability to specify which half of a mirrored volume to read from on a node-by-node basis.

Neither vendor supports RAID5 using product sets evaluated within this report.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second delivering 16 of 25 features, for a score of 64%. With Sun's design, the cluster volume manager is file-system independent and allows a choice of file systems to be deployed. However, the Global File Service only supports UFS. Sun may support other file systems – such as VxFS – to operate with the Global File Service in the future. Sun could improve by delivering online RAID-level reconfiguration, RAID 0, RAID 0+1, RAID 1+0, and unspare. Sun has yet to address rolling volume-manager upgrades and the ability to specify which half of a mirrored volume to read from on a node-by-node basis. Sun could also improve by enabling volume management from NT clients.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS finishes first delivering 23 of 25 features, for a score of 92%. The underlying VxVM volume manager supports online RAID-level reconfiguration, RAID 0, RAID 0+1, and RAID 1+0. VxVM features disk unspare, rolling volume-manager upgrades and volume management from NT clients.

VERITAS has yet to address CVM file-system independence with its SANPoint Foundation Suite – HA and could improve by adding this feature.

CLUSTER FAILOVER SOFTWARE

This section evaluates, at a macro level only,[7] the two vendors' cluster failover software features – the third product area reviewed. Cluster failover software provides the infrastructure for monitoring cluster resources, including applications, and initiating recovery actions.

Figure 3 summarizes the cluster failover software results.

[7] For fine-grain drill-down details, see the DHBA report referenced in Footnote 1.

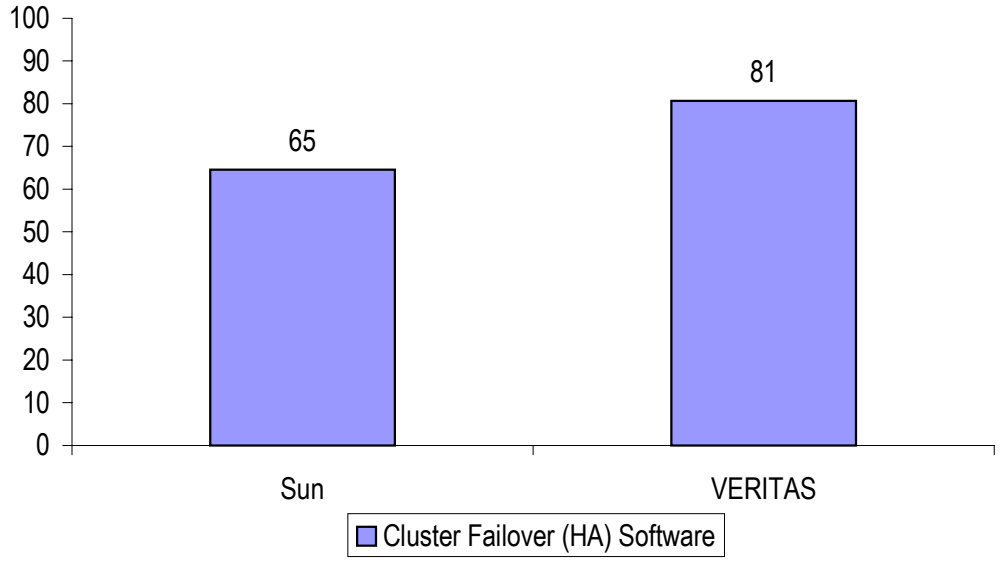


FIGURE 3:
Cluster Failover Software Score Comparison

List 7 shows the criteria for scoring cluster failover software in the area of cluster configuration. Note that the entries in Lists 7 and 8 are consecutively numbered since both were used to obtain the results shown in Figure 3.

1. Number of Nodes Greater than Four
2. Number of Nodes Greater than Eight
3. Number of Nodes Greater than 16
4. Number of Nodes Greater than 32
5. Mixed Operating System Versions in Failover (HA) Mode
6. Ethernet Heartbeat Medium
7. Non-Ethernet Heartbeat Medium
8. Redundant Heartbeat Mediums
9. Optimized Protocol for Heartbeat (Less Overhead Compared with TCP/IP)
10. Cluster Supports Oracle Parallel Server (OPS)
11. Cluster Supports Oracle Parallel Failsafe (OPFS)
12. Cascading Failover of IP Addresses across Three or More Nodes
13. Cascading Failover of Disk Drive Ownership across Three or More Nodes
14. Global IP (Cluster Alias)
15. Failed/Hung Application Data Service – Recovered Locally (i.e., Restart)
16. Failed/Hung Application Data Service – Recovered via Failover
17. API (Application Programming Interface) for Developing Custom Agents
18. Cluster Software Runs on Multiple Operating-System Offerings
19. Hierarchical Clusters (for Disaster Recovery)

LIST 7:
Criteria for Scoring Cluster Failover Software Based on Cluster Configuration

List 8 shows the criteria examined for cluster failover management.

LIST 8: Criteria Examined for Cluster Failover Management

20. Application Failure Monitoring without Application-Specific Agent (i.e., Generic Agent)
21. Application Failure Monitoring via Event Management System
22. User-Defined Resource Groups (Resources Defined to Failover Together)
23. Resource Dependencies May Be Customized
24. ARP Packets Generated to Create New IP to/from MAC Address Mapping
25. Can Start/Stop Multiple Application Instances with Single Command
26. Single Point (i.e., from Single Node) of Cluster Definition
27. Rolling Operating System Upgrade from Version Solaris 2.2 to Solaris 3.0
28. Rolling Operating System Upgrade from Version Solaris 3.0 to Solaris 3.x
29. Online Cluster Node Addition
30. GUI-Based Cluster Resource Monitor
31. SNMP and SMTP Notification Triggered by System Faults

OVERVIEW OF CLUSTER FAILOVER SOFTWARE RESULTS

Both vendors support eight or more cluster nodes. Both support redundant heartbeat paths, optimized heartbeat protocol, resource group customization, cascading failover of IP addresses across three or more nodes, and global IP (cluster-wide IP) address with IP load balancing. Moreover, both enable local and remote application recovery, and a generic application-failure monitoring agent. In addition, both feature ARP packet generation, custom agent API, rolling operating-system upgrade to future operating-system version, online node addition, and GUI resource monitor. Neither vendor's product set supports greater than 32 nodes, nor application failure detection via an event-management system.

Vendors' product sets differ in support for 16 or more nodes, mixed operating-system versions in a single cluster; support for OPS (Oracle Parallel Server) and OPFS (Oracle Parallel Failsafe); cascading failover of disk drives across three or more nodes; and rolling operating system upgrades to the current version of the operating system. Support for cluster software on multiple operating-system offerings, hierarchical clusters, the ability to start/stop multiple application instances with a single command, and SNMP/SMTP notification are further areas of differentiation.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun finishes second delivering 20 of 31 features, for a score of 65%. Sun Cluster 3.0 (SC 3.0) supports eight nodes (in failover mode) and it operates with both OPS[8] and OPFS. Further, Sun Cluster enables start/stop of multiple application instances with a single command.

[8] OPS is only supported through VERITAS Volume Manager with Cluster Extensions.

Sun could improve by supporting more cluster nodes. Sun has yet to address cascading failover of disk drives across three or more nodes. Moreover, rolling upgrades from Solaris 2.x and Sun Cluster 2.x, to Solaris 8 and Sun Cluster 3.0 are not possible. Resolving this would enhance product functionality. In addition, Sun could improve by featuring hierarchical clusters and SNMP/SMTP notification. Finally, Sun has yet to address support for Sun Cluster 3.0 on multiple operating-system offerings.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS scores first with 25 out of 31, for a score of 81%. VERITAS Cluster Server (VCS) supports up to 32 nodes, allows rolling operating-system upgrades to the current operating-system version, and supports cascading failover of disk drives across three or more nodes. Furthermore, VCS runs on multiple UNIX platforms and on NT, supports hierarchical clusters via Global Cluster Manager, and implements SNMP/SMTP notification triggered by system faults.

VERITAS could further improve by supporting OPS and OPFS. Further, VCS has to address start/stop of multiple applications instances with a single command.

FUNCTIONAL AREA ANALYSIS DETAILS

View 2 uses the same data elements previously provided in Lists 1 through 8 for View 1, but organizes them by functional area. In other words, features associated with cluster file systems, cluster volume managers, and cluster failover software are regrouped into seven functional areas according to each feature's functional benefit. Vendors' scores are based on the percentage of features delivered.

The functional areas examined are summarized in List 9 along with a brief definition of the functional area.

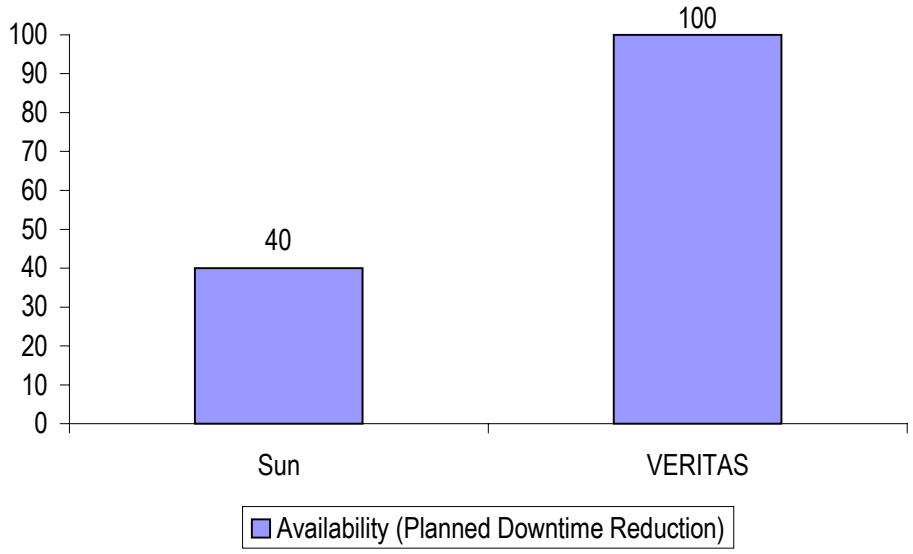
LIST 9: Seven Functional Areas Defined

1. **Availability – Planned Downtime Reduction**
   Features in this area reduce the need to bring down applications for planned or scheduled configuration changes or software upgrades.
2. **Availability – Unplanned Downtime Reduction**
   Features in this area reduce downtime associated with unplanned events including system and component failure.
3. **Hardware Efficiency**
   Features in this area reduce hardware costs by enabling resource sharing and/or reducing the amount of hardware required. Many of the features in this area allow the user a trade-off between cost and performance.
4. **Heterogeneous Operations**
   Features in this area enable or enhance interoperability between software components and software transportability.
5. **Manageability**
   Features in this area enhance administrative productivity and reduce the likelihood of administrative error.
6. **Performance – Speed-Up**
   Features in this area improve processing speed thus reducing response times in transaction processing or file I/O- intensive environments.
7. **Performance – Scale-Up**
   Features in this area enhance scalability (i.e., throughput) via parallel operations, load balancing, and the support of a large number of nodes.

AVAILABILITY – PLANNED DOWNTIME REDUCTION

This section evaluates the cluster file system, cluster volume manager, and cluster failover software features that enhance availability by reducing planned (i.e., scheduled) downtime. Figure 4 shows the results of the analysis and List 10 shows the planned downtime reduction criteria.

FIGURE 4: Availability – Planned Downtime Reduction Score Comparison



LIST 10: Planned Downtime Reduction Criteria

1. Online File-System Growth
2. Online File-System Shrink
3. Online File-System Defragmentation
4. Online RAID-Level Reconfiguration
5. Online Performance Monitoring
6. Can Change Tuning Parameters without Volume Shutdown
7. Rolling Volume-Manager Upgrades
8. Rolling Operating-System Upgrade to Current Operating-System Version
9. Rolling Operating-System Upgrade to Next Operating-System Version
10. Online Cluster Node Addition (to Cluster Configuration) without Cluster Software Shutdown

OVERVIEW OF PLANNED DOWNTIME RESULTS

Both vendors deliver online volume-performance monitoring and online volume-performance tuning, online cluster-node addition, and rolling operating-system upgrade to the next operating-system release. Vendors differ in support for online file-system defragmentation/expand/shrink, online RAID-level reconfiguration, rolling volume-manager upgrades, and rolling operating-system upgrades to the current operating system release.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second delivering 4 of 10 features, for a score of 40%. Sun could improve by supporting online file-system defragmentation/expand/shrink, online RAID-level reconfiguration, rolling volume-manager upgrades, and rolling operating-system upgrade to the current operating-system release.

VERITAS: POSSIBLE IMPROVEMENTS

VERITAS places first delivering 10 of 10 features, or 100% of the features evaluated.

AVAILABILITY – UNPLANNED DOWNTIME REDUCTION

This section evaluates features that enhance availability by reducing unplanned downtime (i.e., downtime associated with hardware or software failures.) Figure 5 summarizes the analysis and List 11 identifies the unplanned downtime reduction criteria.

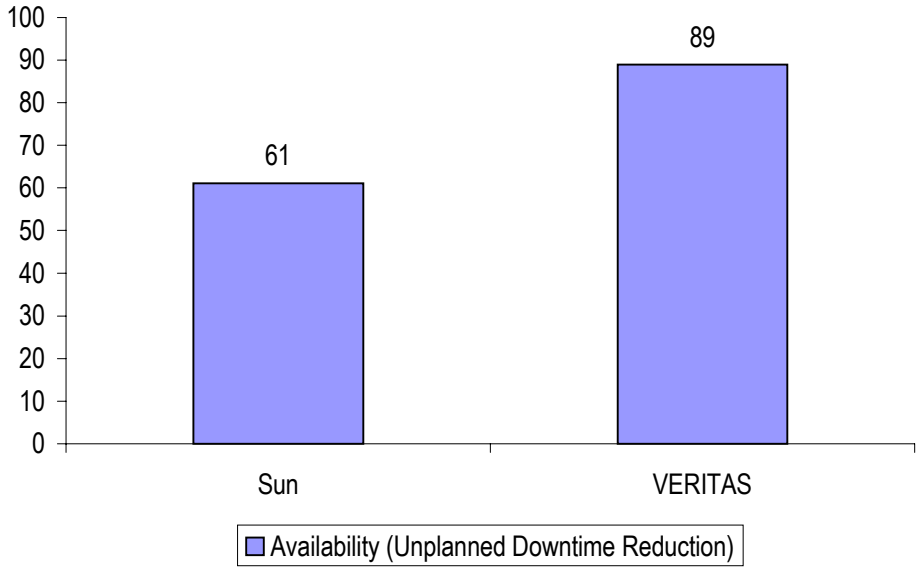


FIGURE 5: Availability – Unplanned Downtime Reduction Score Comparison

1. Journaling (Intent Logging of Metadata)
2. File-System Failures/Errors Cannot Cause Operating-System Panic
3. FSCK Multiple Volumes in Parallel
4. Recovery of Failed Primary/Master Transparent to Applications
5. Can Reserve a Specified Amount of File-System Space for a Given File
6. RAID 1
7. RAID 0+1
8. RAID 1+0
9. RAID 5
10. Mirrored Pair Resynchronized Using Difference Map
11. Redundant Heartbeat Paths
12. Cluster supports Oracle Parallel Failsafe (OPFS)
13. Cascading Failover of IP Addresses across Three or More Nodes
14. Cascading Failover of Disk Drives across Three or More Nodes
15. Cluster Generates ARP Packets, Creates New IP-MAC Address Mapping
16. Failed/Hung Application/Data Service – Recovered Locally
17. Failed/Hung Application/Data Service – Invoke Failover
18. Hierarchical Clusters for Disaster Recovery

LIST 11: Availability – Unplanned Downtime Reduction Criteria

OVERVIEW OF UNPLANNED DOWNTIME RESULTS

Both vendors feature file-system journaling, FSCK of multiple volumes in parallel, transparent recovery of CFS primary/master, RAID 1, mirrored pair resynchronization using difference map, redundant heartbeat paths, and cascading failover of IP addresses across three or more nodes. Furthermore, both vendors offer ARP packet generation, and recovery of failed applications both locally and remotely.

The vendors differ in file-system failure isolation from kernel, file-system free space reservation, RAID 0+1, RAID 1+0, OPFS support, cascading failover of disk drives across three or more nodes, and hierarchical clusters.

Neither vendor supports RAID5 using product sets evaluated within this report.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second delivering 11 of 18 features, or 61% of the features evaluated.

Sun could improve by isolating file-system failures from the kernel, as well as supporting file-system free space reservation, RAID 0+1, RAID 1+0, cascading failover of disk drives across three or more nodes, and hierarchical clusters.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS places first delivering 16 of 18 features, or 89% of the features evaluated. VERITAS could improve by supporting OPFS with VERITAS Cluster Server.

HARDWARE EFFICIENCY

This section evaluates features that affect hardware cost by requiring additional hardware for a specific functional implementation.

Figure 6 shows the results of the analysis and List 12 shows the hardware efficiency criteria.

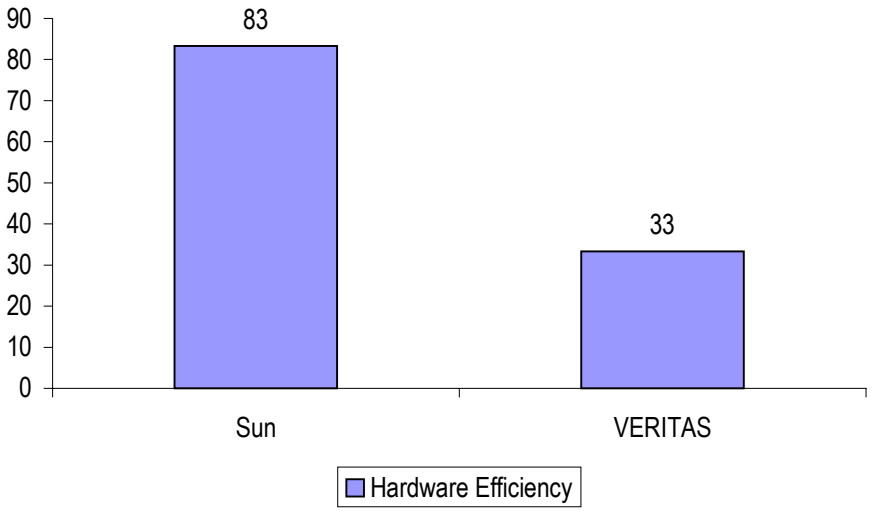


FIGURE 6: Hardware Efficiency Score Comparison

LIST 12: Hardware Efficiency Criteria

1. Indirect Read/Write Access to CFS
2. Mixed Direct and Indirect Read/Write Access to CFS
3. Commodity Interconnect Supported
4. CD-ROM Accessible as CFS
5. Tape Accessible as CFS
6. Ethernet Node Interconnect – Heartbeat

OVERVIEW OF HARDWARE EFFICIENCY RESULT

Both vendors support commodity-node interconnect and Ethernet-based node heartbeat mediums. Further, while not evaluated in this report, both support indirect file-system access using NFS. Neither vendor features mixed indirect and direct access to CFS.

Vendors differ in delivering indirect read/write access to CFS, CD-ROM accessible as CFS, and tape accessible as CFS.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places first delivering five of six features, or 83% of the features evaluated. Sun delivers all the features evaluated excluding mixed indirect and direct access to CFS.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

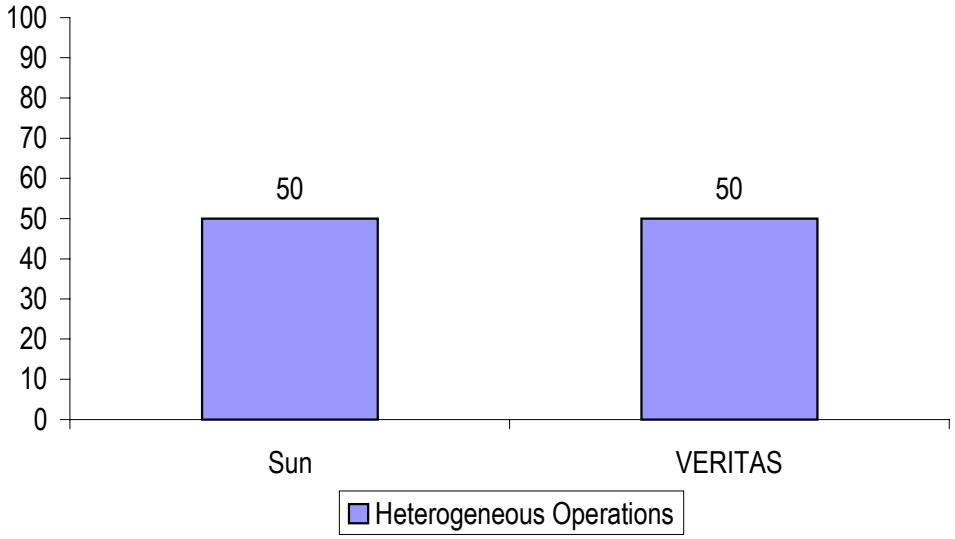
VERITAS places second delivering two of six features, or 33% of the features evaluated. VERITAS could improve by offering indirect read/write access to storage, mixed direct and indirect read/write access to CFS, CD-ROM accessible as CFS, and tape accessible as CFS.

HETEROGENEOUS OPERATIONS

This section evaluates cluster file system, cluster volume manager, and cluster failover software features that enable heterogeneous operations and enhance openness (i.e., interoperability). For example, this section examines the ability to mix and match different hardware and software entities in a given configuration.

Figure 7 shows the results of the analysis and List 13 shows the underlying criteria.

FIGURE 7: Heterogeneous Operations Score Comparison



LIST 13: Heterogeneous Operations Criteria

1. CFS Is Volume Manager Independent
2. CFS Is File System Independent
3. Volume Management from NT clients
4. Mixed Operating-System Versions (i.e., releases) in HA Mode Supported (i.e., SC 2.2 and 3.0)
5. Cluster Software Runs on Multiple Operating-System Offerings
6. CVM Is File System Independent

OVERVIEW OF HETEROGENEOUS OPERATIONS

A total split occurs in this category. Each vendor delivers half of the features, but neither vendor delivers what the other delivers.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun ties with VERITAS, delivering three of six features, or 50% of the features evaluated even though there is no overlap with VERITAS. Sun Cluster 3.0 Global File Service offers independence of CFS and the underlying file system as well as independence of CFS with the underlying volume manager. However, the Global File Service only supports UFS. Sun may support other file systems to operate with the Global File Service, such as VxFS, in the future. The CVM is also designed to be underlying file-system independent.

Sun could pick up on all the features VERITAS offers. Sun has yet to address volume management from NT clients and mixed operating-system versions in cluster HA mode support. Furthermore, VCS Cluster software runs on multiple operating-system offerings while Sun Cluster software does not.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS also delivers three of six features, or 50% of the features evaluated. VERITAS offers volume management from NT clients and mixed operating-system versions in HA mode. Furthermore, VCS Cluster software runs on multiple operating-system offerings.

VERITAS has yet to address independence of CFS and the underlying file system as well as the independence of CFS and the underlying volume manager. Further, VERITAS could improve by providing CVM underlying file-system independence.

MANAGEABILITY

This section evaluates features that simplify managing resources across multiple cluster nodes. For example, administrative interfaces and features that provide a cluster-wide view of resources are examined.

Figure 8 shows the results of the analysis and List 14 shows the manageability criteria.

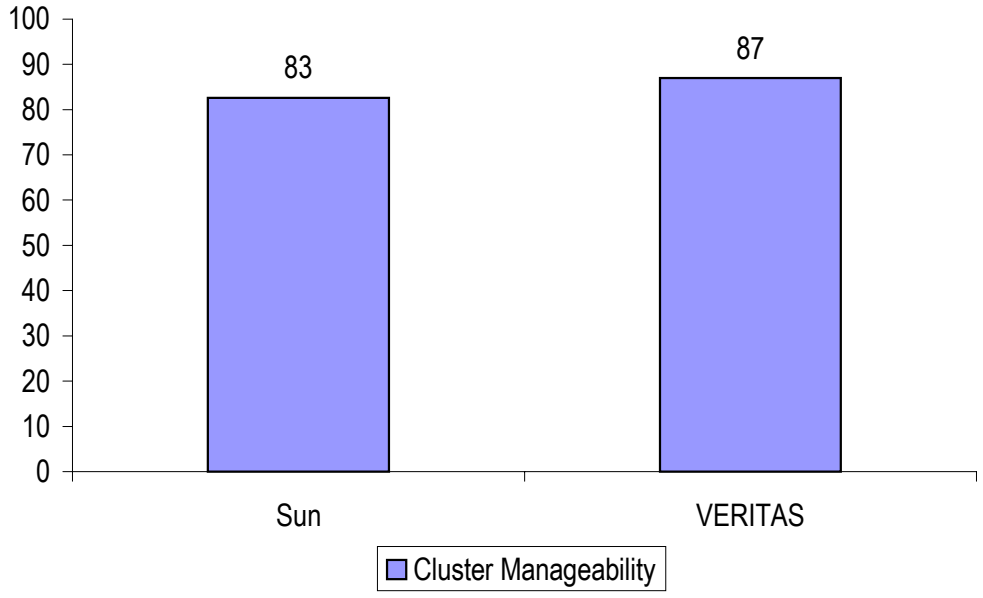


FIGURE 8: Cluster Manageability Score Comparison

LIST 14:
Manageability Criteria

1. File ACLs (Access Control Lists)
2. Freeze and Thaw
3. Can Force Unmount of File System while Files Open
4. Dynamic Selection of Backup for Primary/Master upon Primary/Master Node Failure
5. Hot Spare Disk
6. Original Configuration Restored After Failed Disk Replaced (Unspare)
7. Third Mirror Breakoff
8. Volume Snapshot Copy
9. Volume Snapshot Copy Is Both Readable and Writeable
10. Common Volume Naming Cross Nodes
11. Customizable GUI
12. GUI-Level User Authentication
13. Management from UNIX Clients
14. View Multiple Property Sheets (Objects) Simultaneously
15. Related Resources Defined to Failover Together as Group
16. Resource Dependency and Service Group Dependency Customization
17. Application-Failure Monitoring without Application-Specific Agent (i.e., Generic Agent)
18. Application-Failure Monitoring via Event Management System
19. Documented API for Developing Custom Agents
20. Can Start/Stop Multiple Application Instances with Single Command
21. Single Point (from Single Node) of Cluster Configuration Definition
22. GUI-Based Cluster Resource Monitor
23. SNMP and SMTP Notification Triggered by System Faults

MANAGEABILITY OVERVIEW

Both vendors score well in this category. Both support file-system freeze and thaw, dynamic selection of backup for primary/master, hot spare disk, third mirror breakoff, volume snapshot copy, volume snapshot copy is both readable and writeable, common volume-naming cross nodes, customizable GUI, GUI-level user authentication, management from UNIX clients, and the ability to view multiple property sheets (objects) simultaneously.

Both vendors also offer resource dependency and service group dependency customization, application-failure monitoring without application-specific agent (i.e., generic agent), documented API for developing custom agents, single point of cluster-configuration definition, and GUI-based cluster-resource monitor. Neither vendor supports application-failure monitoring via an event-management system.

Vendors differ in delivering file ACLs, disk unspare, ability to start/stop multiple application instances with single command, and SNMP and SMTP notification triggered by system faults.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second, delivering 19 of 23 features, or 83% of the features evaluated. Sun supports file ACLs and the ability to start/stop multiple applications instances via a single command. Sun could improve by offering forced unmount of file system while files are open, disk unspare, and SNMP and SMTP notifications.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS places first delivering 20 of 23 features, or 87% of the features evaluated. VERITAS' VxFS enables forced unmount of file system while files are open. Furthermore, VxVM supports disk unspare. Finally, VCS supports SNMP and SMTP notifications. VERITAS has yet to address file ACLs and start/stop of multiple applications instances via a single command.

PERFORMANCE SPEED-UP

This section evaluates features that speed-up application performance. For example, features that reduce the response times associated with individual transactions are investigated. Features that enable performance scale-up (i.e., scalability) are treated in the following section.

Figure 9 shows the results of the analysis and List 15 shows the performance speed-up criteria.

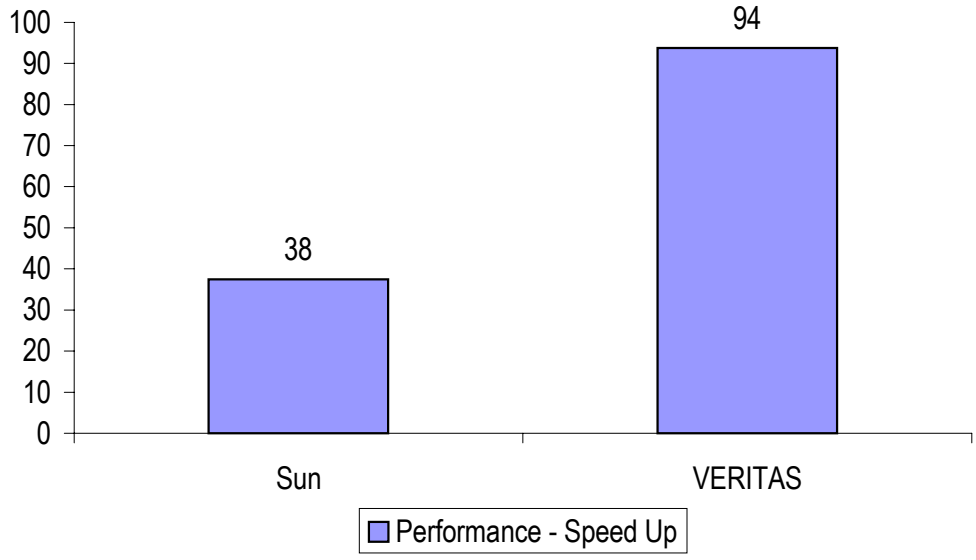


FIGURE 9: Performance Speed-Up Score Comparison

LIST 15: Performance Speed-Up Criteria

1. Direct Read/Write Access to CFS from All Nodes
2. CFS Is Both Readable and Writeable
3. Writes Above a User-Defined Size Are Performed as Direct I/O
4. Maximum I/O Size Adjusted (Increased) for RAID 0/3 Stripe Size/Number of Stripes
5. Convert All I/O to Direct I/O
6. Convert All I/O to Synchronous I/O
7. Convert All I/O to Data-Synchronous I/O
8. Access-Time Updates May Be Turned Off
9. Extent-Based Allocation Based on User-Defined Extent Size
10. Extent-Based Allocation Based on I/O Pattern
11. RAID 0
12. Multiple Active Paths from Server to Volumes
13. Can Specify which Member of a Mirrored Pair to Read from
14. Can specify which Member of a Mirrored Pair to Read from – on per Node Basis
15. Other (Non-Ethernet) Interconnect – Heartbeat
16. Optimized Protocol for Heartbeat – Less Overhead Compared to TCP/IP

PERFORMANCE SPEED-UP OVERVIEW

VERITAS achieves a significant lead on Sun in this category. Both vendors offer CFS that is readable and writeable, convert all I/O to direct I/O, access-time updates may be turned off, multiple active paths from server to volumes, the ability to specify which member of a mirrored pair to read from, and optimized protocol for heartbeat. Neither vendor features a non-Ethernet heartbeat medium, however.

Vendors differ in support for direct read/write access to storage from all nodes, writes above a user-defined size are performed as direct I/O, maximum I/O size adjusted (increased) for RAID 0/3 stripe size/number of number of stripes, and convert all I/O to synchronous I/O. They also differ in convert all I/O to data synchronous I/O, extent-based allocation that relies on user-defined extent size, extent-based allocation based on I/O pattern, RAID 0, and the ability to specify which member of a mirrored pair to read from – on a per-node basis.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second delivering 6 of 16 features, or 38% of the features evaluated. Sun could improve by offering direct read/write access to storage from all nodes, writes above a user defined size are performed as direct I/O, maximum I/O size adjusted (increased) for RAID 0/3 stripe size/number of stripes, and convert all I/O to synchronous I/O. Sun could also improve by featuring convert all I/O to data synchronous I/O, extent-based allocation based on user-defined extent size,

extent-based allocation based on I/O pattern, RAID 0, and the ability to specify which member of a mirrored pair to read from – on a per node basis.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS places first delivering 15 of 16 features, or 94% of the features evaluated. VERITAS offers direct read/write access to storage from all nodes, writes above a user-defined size are performed as direct I/O, maximum I/O size adjusted (increased) for RAID 0/3 stripe size/number of stripes, convert all I/O to synchronous I/O, and convert all I/O to data synchronous I/O. VERITAS also offers extent-based allocation based on user-defined extent size, extent-based allocation based on I/O pattern, RAID 0, and the ability to specify which member of a mirrored pair to read from – on a per node basis.

PERFORMANCE SCALE-UP (SCALABILITY)

This section evaluates cluster file system, cluster volume manager, and cluster failover software features that enable scale-up of performance. For example, features that enable higher transaction rates, such as nodes added to a cluster, are considered. Figure 10 shows the results of the analysis and List 16 shows the performance scale-up criteria.

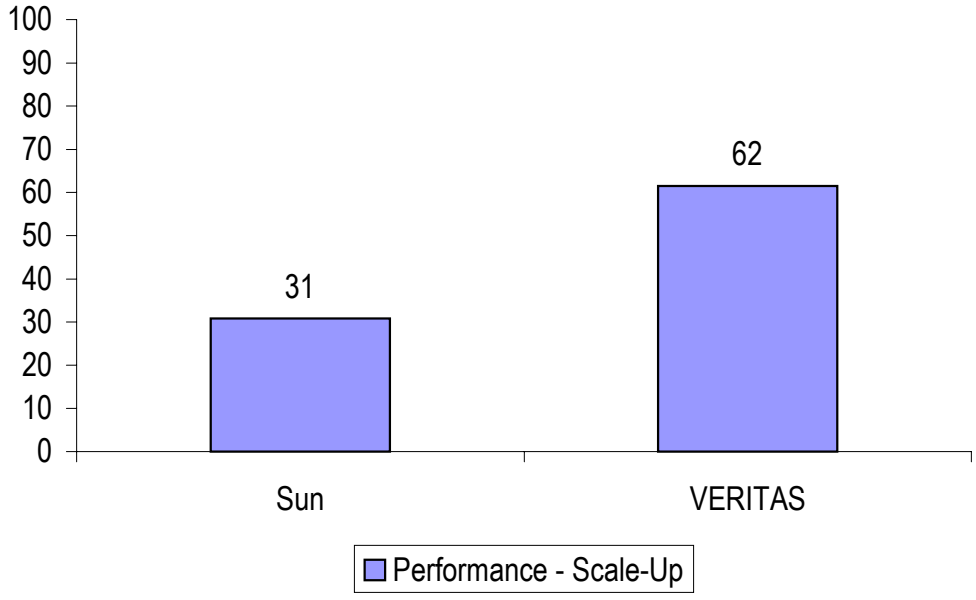


FIGURE 10:
Performance Scale-Up Score Comparison

LIST 16: Performance Scale-Up Criteria

1. Number of CFS Nodes Supported Greater than Four
2. Number of CFS Nodes Supported Greater than Eight
3. Number of CFS Nodes Supported Greater than 16
4. Number of CFS Nodes Supported Greater than 32
5. Lock-Manager Capacity increases with Number of Nodes (per File System)
6. Data I/O Capacity Increases with Number of Nodes (per File System)
7. Metadata I/O Capacity Increases with Number of Nodes (per File System)
8. Number of Nodes in HA Configuration Greater than Four
9. Number of Nodes in HA Configuration Greater than Eight
10. Number of Nodes in HA Configuration Greater than 16
11. Number of Nodes in HA Configuration Greater than 32
12. Cluster Supports Oracle Parallel Server (OPS)
13. Global IP (Cluster Alias)

OVERVIEW OF PERFORMANCE SCALE-UP

Both vendors offer at least eight CFS nodes, at least eight cluster nodes, and global IP with IP load balancing. Neither vendor supports more than 16 CFS nodes or 32 cluster nodes. Furthermore, neither offers metadata I/O capacity increase as the number of nodes increases.

Vendors differ in number of CFS nodes, number of cluster nodes, data I/O capacity increases with number of nodes, lock-manager capacity increases with number of nodes, and OPS support.

SUN MICROSYSTEMS: POSSIBLE IMPROVEMENTS

Sun places second delivering four of 13 features, or 31% of the features evaluated. Sun Cluster 3.0 supports eight cluster nodes and eight CFS nodes. Furthermore, SC 3.0 supports OPS. Sun could improve by supporting more cluster and CFS nodes and by enabling data I/O capacity increases, and lock manager I/O capacity increases, with additional nodes.

VERITAS SOFTWARE: POSSIBLE IMPROVEMENTS

VERITAS places first delivering eight of 13 features, or 62% of the features evaluated. VCS supports up to 32 nodes, while the VERITAS SANPoint Foundation Suite – HA supports 16 nodes. Furthermore, SANPoint Foundation Suite-HA enables data I/O capacity increases, and lock-manager I/O capacity increases, with additional nodes.

VERITAS could improve by offering OPS support.

APPENDIX A: METHODOLOGY

Two views are presented of the data analyzed in this report. In View 1, the data are organized by product area. The features and functions examined are grouped into three key areas: cluster file systems, cluster volume managers, and cluster failover software. This view evaluates the products based on the details of features. These features may or may not be delivered by the particular product. Vendor scores are based on the percentage of features delivered in the DHBA feature subset list.

View 2 uses the same feature and function data elements given in the first view but organizes them by seven functional areas: planned downtime reduction, unplanned downtime reduction, hardware efficiency, heterogeneous operations, manageability, performance speed-up, and performance scale-up. The features associated with a functional area are listed in the corresponding report section. These sections also present the details of the features that are or are not delivered. As in the first view, vendors' scores are based on the percentage of features delivered in the DHBA feature subset list.

To keep matters simple in this report, the scores are determined by dividing the number of data elements delivered by the vendor by the total number of data elements in the area, which derives the percentage of features delivered.[9] In other words, this analysis does not weight data elements and each data element contributes the same amount to the score in an area. Furthermore, this report does not provide composite scores, for example, in View 2, with different weighting factors applied to each of the seven categories presented.

Readers must consider the significance of each feature to their own requirements when comparing vendors. DHBA, which conducts research on a wide range of related topics, can provide alternative analyses with weighing factors based on specific information technology environments.

CRITERIA USED

As with any study, a specific set of criteria is used. The criteria used in this evaluation concentrate on product features and functions and simple scoring. The reader should note that important considerations such as service and support, price and/or cost-of-ownership (COO), vendor ability to deliver, software reliability, and market penetration are not considered in this report. Any buying decision by the savvy IT manager must take these factors (and more) into account, as well as the information in this report.[10]

[9] In the usual DHBA methodology for detailed reports, weighting factors are applied to each examined feature or function, product area and functional area.

[10] DHBA can provide further information on these other decision factors.

Further note that the score-based methodology in this report must assign scores and rankings even when both vendors are very close. When the two vendors are nearly even, readers should examine the product functional differences described in this report in conjunction with their own requirements. Also realize that the results presented in this report are based on specific product releases. Significant changes may occur with the next release of a product version.

Finally, the reader should understand that the scores in each category represent the percentage of the evaluated features that a vendor delivers. The scores do not quantify – in any absolute or relative manner – a vendor's overall ability in any area. For example, scores in the "Performance – Speed-Up" section do not quantify actual performance. Therefore, a vendor scoring X percent higher than another does not, by virtue of this analysis, deliver X percent higher performance based on some performance measure.