

VERITAS FILE SYSTEM 4.0 VERSUS REISERFS 3

A Competitive File System Comparison

TABLE OF CONTENTS

Introduction	3
Features.....	3
Detailed Feature Comparison.....	5
Platform Support.....	5
Maximum File System Size	5
Resizing the File System.....	6
File System Snapshots.....	6
Journaling Technology and Maximum Log Size	7
Defragmentation	7
Supported for Root	7
Management Tools.....	7
Quality of Storage Services	8
Portable Data Containers	8
Advanced Online Tuning Options.....	8
Non-VERITAS File System Conversion Utility	9
Upgrading From a Previous Version	9
Conclusion	9

INTRODUCTION

ReiserFS 3 is an all-purpose journaling file system, developed for the Linux platform by a team of engineers led by Hans Reiser. Version 3 of ReiserFS is the default file system for a number of Linux distributions, including SuSE. ReiserFS journals file system metadata, greatly reducing the risk of corruption during an unscheduled shutdown of a server. It is also more efficient and better performing than the traditional Ext2 Linux file system.

VERITAS File System (VxFS) 4.0 is a component of VERITAS Storage Foundation 4.0 and is supported on the Red Hat Enterprise Linux 3 distribution of Linux. VxFS 4.0 is an enterprise-class commercial file system with a correspondingly sophisticated suite of features. Its support of dynamic growing and shrinking of the file system, non-disruptive defragmentation, and a user-friendly graphical user interface (GUI) significantly ease administration.

Performing a side-by-side comparison of the technical features of ReiserFS 3 and VxFS 4.0 gives administrators the information they need to decide which product is more appropriate for their environment. The analysis, detailed below, shows VERITAS VxFS 4.0 to be the superior file system when enterprise-class features and performance are needed from the file system.

FEATURES

Product feature	VxFS 4.0	ReiserFS 3.6
Linux kernel support	Red Hat EL 3	Most Linux kernels
Included in distribution		RH, SUSE
Platform support (other than Linux)*	Solaris, HP-UX, AIX, Linux	Linux only
Max file system size*	2 TB (32-bit)	2 TB (32-bit)
Max file size	2 TB (normal file), 16 TB (sparse file)	2 TB (32-bit)
Supported block sizes	1024, 2048, 4096, 8192	4096
Address range (code support)	32 and 64 bit	32-bit
Online resize*	Growing and shrinking. Integrated with VERITAS Volume Manager	Growing only, offline.
Snapshots (non-persistent)*	Yes	No
Snapshots, persistent	Yes (16 million). Can be mounted	No

Product feature	VxFS 4.0	ReiserFS 3.6
(max number of snapshots)*	read-only or read-write	
Journaling technology*	Transaction based. Metadata and/or data operations are logged. Log can be kept on a different device	Claims atomic operations
Max log (journal) size*	256 MB (internal or external journal)	32,749 blocks (4k block size). Journal can be put on a separate device
File System error handling mount options	disable, nodisable, wdisable, mwdisable	No
File allocation technology	Extent based, dynamic inode allocation	Extend based, dynamic inode allocation
Defragmentation*	Yes, online	No
User quota support	Yes	Yes
Group quota support	Yes	No
Extended attributes	Yes	No
Unix ACL support	No	No
Supported for root*	No	Yes
DMAPI Support	No	No
Named Data Streams	Only supported with VxFS API's	No
File System Event History Log	Yes	No
Plug-in capabilities	No	Yes
Management tools*	GUI/CLI	CLI
Posix compliant	Yes	No

Product feature	VxFS 4.0	ReiserFS 3.6
Quality of Storage Services*	Yes	No
Portable Data Containers*	Yes	No
File system statistics (buffer cache, counters ...)	Yes	No
Direct IO support	Yes	No
Oracle Disk Manager (ODM)	Yes (Storage Foundation for Oracle RAC)	No
Advance online tuning options (read-ahead, preferred read/write size, number of read/write threads)*	Yes. Will automatically interact with VxVM to default to the most optimal settings	No
Conversion utility from other File System*	Yes. Convert from ext2/ext3 supported	No
File System layout upgrade from previous version*	Yes, online	No

Features with an * are described in greater detail below.

DETAILED FEATURE COMPARISON

Platform Support

VERITAS VxFS 4.0 supports multiple platforms, including Sun (Solaris), HP (HP-UX), IBM (AIX), and Linux. ReiserFS 3 is currently only available for Linux.

VxFS 4.0 offers a consistent management interface across all platforms, easing the System Administrator's learning curve. The command set for VERITAS products is also consistent across all platforms, allowing scripts to be easily ported between UNIX and Linux environments.

Maximum File System Size

Linux is a 32-bit operating system with a practical file system size limit of two TB. The Linux Virtual File System (VFS) layer has a compliance-defined standard interface and a 16 TB address range limitation. In addition, the Linux block device layer has a two TB address limitation, restricting file systems to a maximum size of two TB.

The size of a ReiserFS file system is limited by the block device layer. Although the Linux kernel will not write data beyond a two TB boundary, no warning is given if the file system is greater than two TB. Great care must be exercised when creating large ReiserFS file systems to avoid data corruption.

VxFS 4.0 supports the maximum allowable file system size of two TB (minus one KB). As mentioned above, this is restricted by the block device layer's inability to address data beyond two TB. When this limit is lifted in future versions of Linux VERITAS will support a high value (usually limited by the operating system).

Resizing the File System

ReiserFS 3 supports online resizing of the file system. However, this new functionality has the potential to corrupt user data if problems occur. The MAN pages for ReiserFS 3 indicate that code written to perform online resizing is still considered BETA. An unmount of the file system is recommended, before resizing, to achieve a higher success rate.

ReiserFS 3 does not interact with the underlying volume manager. Growing the file system requires the administrator to first expand volumes and then resize the file system. When shrinking a file system, care is needed to ensure that the size of the logical volume exactly matches the size of the file system, otherwise, data corruption will occur.

VERITAS VxFS 4.0 supports online growing and shrinking of the file system. Interacting with VERITAS Volume Manager (VxVM), VxFS automatically manages resizing of both the file system and the logical volume to ensure success.

File System Snapshots

ReiserFS 3 does not support snapshots. Instead it utilizes the logical volume manager (LVM) to create a block level snapshot. The intent of the snapshot is to provide an online backup, and the copy must be mounted read-only. The administrator must reserve storage on a volume separate to that hosting the file system. The size of the storage reserved for the snapshot depends on the file system's activity level, however, 10 percent to 20% of the file system size is recommended. The file system must be quiesced before running the snapshot command to ensure consistency. The snapshot is dependent on the original file system disks and cannot be exported to another host.

VERITAS VxFS 4.0 uses two types of snapshot technology: snapshots and storage checkpoints. VxFS Snapshots are comparable to the combined ReiserFS 3 and LVM snapshots, and are non-persistent, read-only copies of the original file system. Designed for use with backups, these snapshots require their own storage device and do not take up free space in the file system. If the original file system is unmounted the snapshots are removed and no longer available.

Storage checkpoints, based on VERITAS copy-on-write technology, provide administrators with significantly greater flexibility compared to snapshots. Each checkpoint is a reboot-persistent copy of the file system, available to be mounted read-write or read-only. Checkpoints are ideal for fast recovery of files, without going through the cumbersome process of restoring data from tape.

Storage checkpoints make use of the free space within the file system and store only modified data blocks. The properties of the checkpoint can be set so that the checkpoint automatically expires if the file system free space gets too low. Checkpoints can also be created from other checkpoints giving the system administrator even greater flexibility.

Journaling Technology and Maximum Log Size

ReiserFS 3 only logs metadata operations, and it is possible to corrupt file system contents when a server crashes. A source patch is available that enables data logging but it is unknown if it is included in the newer distribution.

ReiserFS 3 allows the log to be located within the file system or on an external device. If kept within the file system the maximum size of the log is 32,749 blocks. On a separate device, however, the maximum size is equal to the size of the device. Each journal is unique to the file system it is associated with.

VERITAS VxFS 4.0 has two journaling options. Both options allow for a maximum of 256 MB of log device size. If simplified management is desirable the journal can be kept within the file system.

The latest release of VxFS includes multi-device support. This feature allows the journal to be located on a volume separate from the file system: journal data is usually accessed sequentially and can conflict with the random access typical of a file system. Moving the journal to its own storage volume can provide a substantial performance boost to the file system.

Defragmentation

Fragmentation is a fact of life in all file systems. As users and applications access and modify data, fragments of files and free space become scattered around the disk, significantly reducing performance of the file system.

ReiserFS 3 does not support defragmentation. The only way to defragment a ReiserFS 3 file system is to dump everything to tape and restore the data. For many environments the downtime necessary during this process will be unacceptable.

VERITAS VxFS 4.0 supports online defragmentation. This allows administrators to carry out the performance-optimizing task of defragmenting the file system without interrupting end-user access to file system data.

Supported for Root

ReiserFS 3 can be used as a root file system. Special instructions must be followed to convert an existing root file system to a ReiserFS 3 root file system. This process cannot be done online and requires server downtime.

VERITAS VxFS 4.0 is a commercially available product, and must be purchased and installed separately from the Linux operating system software. VxFS cannot be installed at the same time as the operating system and is not supported as a root file system.

Management Tools

ReiserFS 3 does not contain a management infrastructure. Management of the file system is performed using the UNIX command line interface (CLI) commands.

VERITAS VxFS 4.0 and VERITAS VxVM offer a fully integrated graphical user interfaces (GUI), in addition to a CLI interface. The VERITAS Enterprise Administrator VEA GUI is java-based interface that can be run from any platform. Commands being executed against the file system are displayed and logged, by the GUI, allowing administrators to incorporate these same

command sequences in CLI-based scripts for customized management of the file system. The VEA GUI provides a common interface to all VERITAS products, significantly easing the administrator's learning curve.

Quality of Storage Services

Most commercial file systems today are designed to occupy the storage capacity of a single disk — whether physical or virtual. And storage systems, in turn, are designed so that each disk has a uniform configuration — mirrored, striped, etc. These restrictions limit the flexibility of storage and system administrators because all the files in any given file system occupy one virtual disk, with one virtualization configuration. The VERITAS Storage Foundation Quality of Storage Services (QoSS) feature solves this problem by enabling multi-volume support in the VERITAS file system.

Closely integrated with other components of VERITAS Storage Foundation, the QoSS feature allows a file system to form a pool of storage from which data is allocated. A multi-volume file system can occupy up to 255 virtual volumes, and this configuration is dynamic, allowing volumes to be added or removed as needed. Each volume in the file system can have a unique combination of capacity, concatenation, striping, mirroring, and RAID configuration to satisfy the requirements of different types of hosted data.

Portable Data Containers

VERITAS Portable Data Containers (PDC) allow for fast migration of user data between different computing platforms with minimal downtime for end-users. PDC unlocks user data from the confines of the server giving administrators the flexibility to locate processing resources on the optimum platform.

A VERITAS File System defined for PDC use implements a common disk layout that can be used by any supported operating system platform. Moving data from one platform to another — for example, migrating an Oracle database from a Solaris box to Linux — or sharing data between two platforms — or example, performing off-host backup processing — does not require any movement of the data. File System volumes are simply unmounted from one server and mounted to another — if the endianness of the platform changes a metadata conversion is also necessary. With no data being physically moved the size of the files in the file system does not matter — a one KB file is moved as quickly as a one GB file.

Advanced Online Tuning Options

The VERITAS File System is able to change a variety of tuning options online to improve performance. These options include the preferred IO read and write size and the number of threads reading or writing per file. Environments with large files, and many clients accessing those files, will find that these options make a significant difference in performance.

VxFS can automatically bypass page cache if it determines that an I/O is larger than a value set by the administrator — `discovered_direct_iosz` tunable. This will provide a substantial performance boost in environments with a wide variation in IO sizes — for example, when sharing a file system with a database, setting this tunable parameter will enable VxFS to determine whether to use `directio` or not.

Most of online tuning parameters are automatically set. However, environments with set up for hardware RAID, where VxVM and VxFS are unable to probe the internal details of the LUN configuration, the parameters must be changed manually.

Non-VERITAS File System Conversion Utility

To minimize the overhead of migrating from to VxFS, VERITAS provides a file system conversion utility. This program saves a substantial amount of time when large amounts of file system data must be migrated to the new VxFS installation.

Upgrading From a Previous Version

Upgrading to ReiserFS 3 from a previous version requires a full dump and restore of the file system. This process requires unacceptable interruption to end-user file system access during the upgrade.

VERITAS VxFS 3.4 is online upgradeable to VxFS 4.0, minimizing end-user downtime. A reboot is required to install the new software and reload kernel drivers. Because VxFS is a module, and not compiled into the kernel, it is an easier and safer upgrade when compared to ReiserFS.

CONCLUSION

VERITAS VxFS 4.0 offers enterprise-class performance that scales with the load on the system. As Linux servers find their way into more and more mission-critical computing environments, predictable performance, regardless of the system load, becomes essential. VERITAS VxFS 4.0 offers the bullet-proof, enterprise-class features and performance demanded by today's Linux system administrators and users.