



White Paper for Melio FS Software

Introduction

Utilities to manage SANs have evolved from simple storage management software, through LUN (logical unit number) masking utilities, to volume management software, virtualization tools, and file and volume sharing software. However storage still has to be divided in different parts to accommodate different operating systems in the same environment. This is little improvement over having a traditional storage structure with separate direct attached storage for each server or group of servers. Different servers are used for different tasks, and each server comes with its own set of supported file systems, which are incompatible with each other. Every added server increases the management difficulties for the system administrators. Melio FS is a shared journaling, 64-bit file system, specifically designed for shared storage solutions, supporting heterogeneous operating systems.

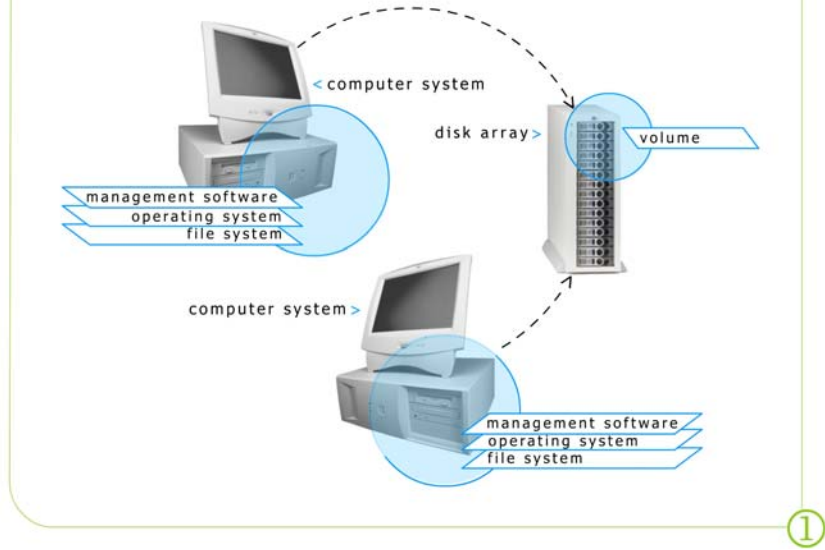
Traditional File Systems and Solutions Available Today

Traditional file systems are oriented to the multi-user, multi-process, single host model and made for one operating system to access one storage volume at a time. A file system names the files and decides where the files are placed logically for storage and retrieval. No matter how many I/O requests there are, all of them are eventually channeled through a single module of the operating system to a single volume. The operating system assumes that it is the “only entity” that has access to the volume. This is not acceptable for storage networking environments, where more than one host has to access the same volume at the same time. If multiple hosts try to access the same volume at the same time, data corruption may occur.

Other limitations with using a non-SAN aware file system in shared storage environments is that it becomes unnecessarily slow, it may impose file size, file system size, block size, number of files or directories, directory depth, and it is not designed to support heterogeneous operating systems. There are some distributed and shared file systems, and journaling file systems, but none that implement all of these functionalities simultaneously.

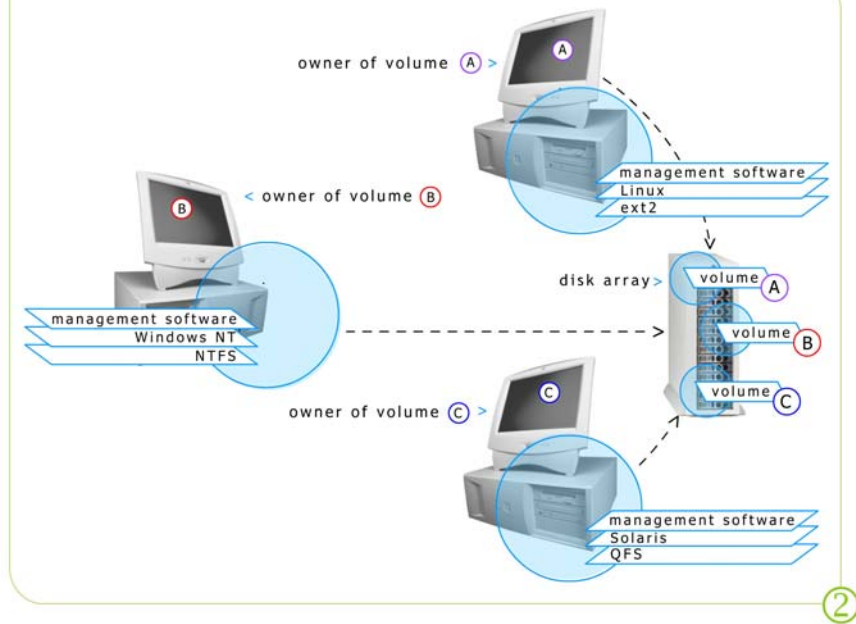
One way to work around the limitations of the non-SAN aware file system is to add a management software to the operating system to organize the writing or reading from the volumes (see picture 1), however this software is dependent upon which kind of operating system that is used.

Management Software in SAN



Traditional non-SAN aware file systems are not designed to support heterogeneous operating systems. Shared storage solutions that have hosts with different file systems and operating systems can only work in one SAN by using LUN masking or virtualization tools to divide the storage into different parts, with separate volumes for each operating system (see picture 2).

Management Software Handling Different OS's



Sanbolic's Response to Customers Needs

The First Cross Platform File System for SAN's

Sanbolic believes that the best way to manage large storage is to keep it controlled by a single file system. The goal to share large amounts of data without performance degradation has led to the creation of storage area network architecture. The significant transfer speed and available space combined with this architecture requires a suitable software or file system solution. Only this approach can take full advantage of all features in shared storage environments.

Melio File System is designed from the ground up specifically for use in shared storage environments. It is designed to meet the challenge of SAN operation.

Melio FS is a shared file system, a journaling file system and a file system that supports heterogeneous operating systems. Melio FS provides the user with security, high capacity, availability, performance, reliability, scalability, manageability, and return-on-investment.

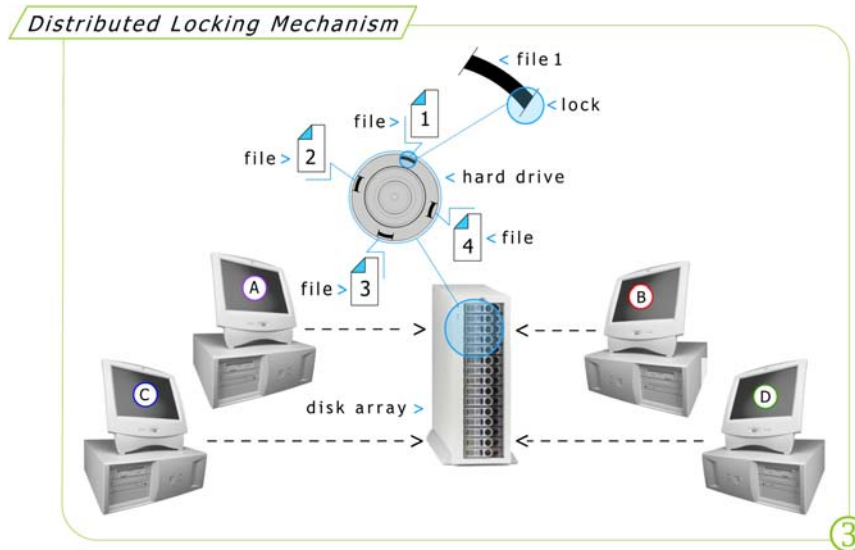
Features

Shared File System

A shared file system allows multiple users to access files on the storage volumes as if they were their own. The metadata is contained within the file system on the storage; no metadata transfers are necessary, eliminating a potential single point of failure. The metadata used is designed so it is decentralized allowing high independence of workstations, accessing the same volume. Melio FS is a shared file system that provides sharing capabilities and organizes updates so all users have the most recent information.

Distributed locking mechanism

The distributed locking mechanism allows several computers to access the same file simultaneously. Read and write operations to different parts of the same file (see picture 3) can be performed at the same time from different machines. Any network protocol can be used to transport the locks between the machines and the locks are cached so the network traffic is kept to a minimum. The lock operations are executed concurrently with the journal write operations.



Journaling File System

A journaling file system is fault resistant. It ensures complete data integrity. Updates to directories in the file system are constantly written into a journal on a disk, before the original disk log is updated. This means that in case system failure occurs the journaling file system 1. ensures that the data on the disk has been restored to its precrash configuration and 2. recovers unsaved data within a few seconds and stores it in an alternative place. The recovery time is independent of the file system size and of the amount of files the system is managing. Melio FS journals all metadata operations, and allows recovery of file system structures without taking the volume offline or shutting it off. If dismounting and remounting a volume is necessary for any reason, the file system check and repair is extremely quick. A file system can be journaling but not distributed, meaning that it works only on one server and its own dedicated storage.

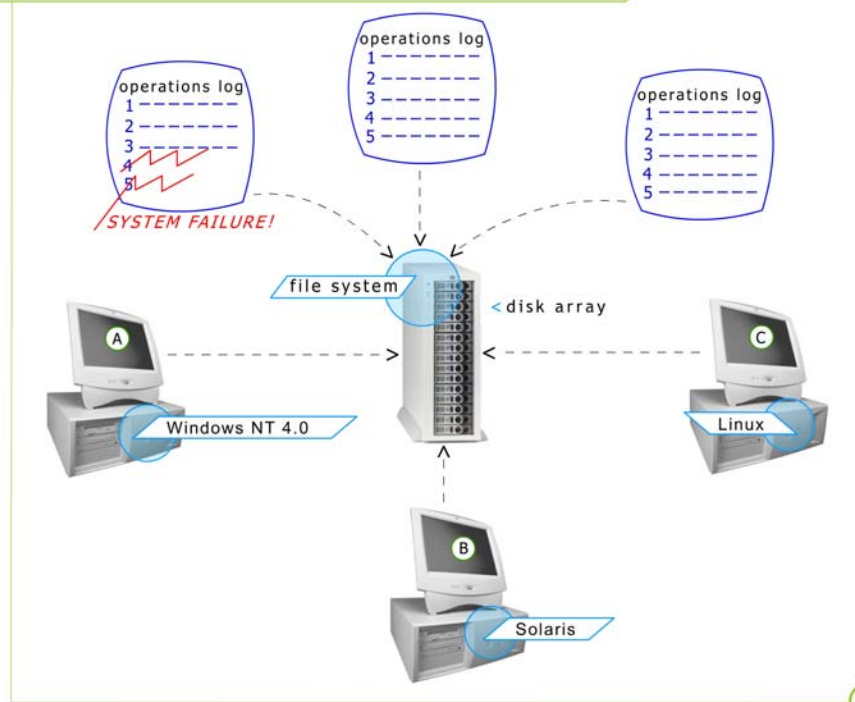
Data transfer and communication

Melio FS provides the users with speed and reliability by transferring the bulk of data flows through the high-speed direct channel to the storage, and only command and control over TCP/IP or other network protocol.

Heterogenous operating system support

Melio FS is a file system that provides heterogeneous operating system support. Different operating systems can run on the same file system (see picture 4). This allows better utilization of existing hardware and infrastructure and means that the user is not confined to a single platform and provides the freedom to utilize the best application and best platform for any given task.

Journaling File System with Heterogenous OS Support



SAN clustering operations

Each computer (probably running on different operating systems) within the cluster will see the volumes, formatted with Melio FS as their local disks. All operations that are possible with local file systems will be possible with Melio FS and the SAN hardware provides the only throughput restrictions.

Entering or exiting the cluster does not require special operations. If a member of the cluster fails, the remaining hosts detect this and recovery of Melio FS metadata is performed, based on the journal of failed machine. All communication between cluster members is symmetrical, as the cluster does not require a master computer.

System layout

Melio FS provides high performance through its system layout which is designed to speed up write and read from a file, creating and deleting a file, and directory lookup. Melio FS allows multiple streams per file. Minimum effort is used for allocation/ deallocation of very large data streams, as well as handling great amounts of small streams. This gives applications greater freedom to address their storage needs in a more efficient way.

All these new features that Melio FS brings as a specified SAN file system means

- no data corruption when operating systems on different computers overwrite each others data on the shared storage
- no more need for storage management software
- no more need for volume or file sharing software

- no need to divide the storage with virtualization tools or LUN masking when heterogeneous operating systems exist in the same network

Customer benefits

What are the general benefits of Melio FS for the customer?

Melio FS can cut operational cost as administrators in a heterogeneous environment may spend less time administering a single file system on a central storage than several file systems for each different operating system. Less storage management software is required as preferred software on the preferred platform can be used for all tasks. The storage utilization is maximized and network costs may decrease as the SAN can off-load the burden off the LAN, which will speed it up for the more ordinary users, resulting in increased productivity on their side.

Melio FS can increase the availability through incremental revenue or productivity for the customer because it significantly reduces downtime. A well-designed SAN can avoid single points of failure on the hardware level, but applications or computers may still crash, which may lead to logical (not physical or hardware) errors on the file system. A standard file system check takes a long time, and longer with the increase of the volume of data. Melio FS can do a system check and the recovery time is reduced significantly. As it provides built-in support for clustering / distributed computing, a simple 2-node cluster can eliminate the point of failure that a single server represents, thus increasing the reliability even further.

The business flexibility is increased because there are no practical limits on volume sizes, file sizes, file names and directory organization. File system volume sizes can be as big as 18 million Terabytes. Melio FS provides fast access to all files due to its groundbreaking layout design. All existing applications can be used without changes in a SAN environment. There is no need to keep multiple copies of files and each task can be performed on the operating system most suitable for it and all others can access the resulting data. There is no or minimal need to retrain end users due to the fact that the file system appears as local on each machine

Within SANs for e-commerce the user data needs to be kept in a database on the storage and high availability is a must. Well-designed SAN gives a hardware-level fault-resilience. A well-designed file system gives the necessary degree of reliability on the software level. To serve the big number of customers connecting to the e-commerce site, a cluster of web-servers is built. The servers need to access the same database at the same time, not only reading from it, but also writing to it. Here the SAN-wide locking capabilities of Melio FS come to play, enabling different machines to access different parts of the same database file at the same time.

Within SANs telecommunication high reliability is necessary. As in e-commerce the machines keep their databases on the central storage, but here the performance requirements are even more stringent, because real-time operation is necessary.

Features

Shared File System

All machines on the SAN see the shared volumes as local and can read and write to them at the same time. The distributed locking mechanism does not rely on a server.

Reliability

All metadata operations are logged, which greatly reduces the possibilities for data corruption. The file system check is extremely fast, leading to higher availability. There is no server-like machine, eliminating potential single point of failure.

Speed and High Performance

Melio FS transfers the bulk of data flow through the high-speed direct channel to the storage, and command and control are transferred over TCP/IP or other network protocol. Melio FS system layout is designed to speed up write and read from a file, creating and deleting a file, directory lookup and to support both large and small files efficiently. It allows multiple streams per file and minimum effort is used for allocation/ deallocation of large data streams, as well as handling great amounts of small streams.

Scalability

Melio FS is designed as full 64-bit file system providing maximum volume size of 2^{64} bytes, or more than 18 million terabytes (1 TB= 10^{12} B).

Heterogenous OS Support

Melio FS is being ported to Windows, Solaris, HP-UX, Linux, Unix, AIX

Security

Unix operating systems have different security semantics than Windows NT. Melio FS supports both types of security models.

Technical specifications

64-bit shared journaling file system

- multiple computers can read and write on the same volume at the same time
- all metadata operations by any computer are stored in logs for high reliability and up time
- no metadata server

Maximum volume size

- 2^{64} bytes, or more than 18 million terabytes (1TB= 10^{12} B)

Maximum file size

- 2^{64} bytes

File system block size

- variable for maximum performance

Platforms

- Windows NT 4.0,
- Windows 2000,
- Solaris, HP-UX
- Linux, Unix,
- AIX

Storage hardware

- any hardware that is supported by the host operating system