1 Beyond
DIGITAL VIDEO SYSTEMS

1 Beyond White Paper:

Storage Area Network Design for High Definition Video

Based on the design of 1 Beyond's Harmony™ Shared HD SAN

By: Terry Cullen

March 28, 2004
Rev. 1f

Table of Contents:

Background: What is the 1 Beyond Harmony™ Shared SAN?

The concept and benefits of a Storage Area Network or SAN are only now becoming widely recognized by the industry for the undeniable efficiencies a SAN can bring to an organization. Because the benefits are so desirable, the market has responded with so many solutions that the term SAN has become distorted to mean many things – almost depending upon a particular manufacturer's solution. Unfortunately, this means that the term SAN has become greatly confused.

Harmony refers to 1 Beyond's total shared storage environment. It is scalable in several dimensions including size, number of systems connected, type of video being processed and application. The *commonality* across all Harmony configurations is the SAN feature set and user interface that will be discussed here in detail. The *differences* are due to the various hardware and software configurations to effectively solve the shared storage requirement for different video applications. For example, solving the Shared SAN requirement for Standard Definition is relatively straight forward because it can be done with "off-the-shelf" hardware as long as it is implemented with the proper control software. Solving this problem for High Definition requires significant redesign of every aspect of the SAN including hardware controllers, disk configurations, control software and even a more efficient file system. So comparing these solutions is irrelevant.

I. Design Goals:

This paper concentrates on the most demanding design requirements needed to handle the highest resolution High Definition video uncompressed. Once the hardest design challenge is solved, the scalability backwards i.e. creating starting SAN configurations, is straight forward because you know where you have to end up. It is totally unrealistic to approach this problem from the bottom up. Attempts to do this have run into walls at every turn that remain insurmountable.

To start any design process, the problem must first be defined. This can be done very effectively by asking what we want to end up with at the start. The following set of questions is a very effective tool for defining not just where we want to go but also how we ought to evaluate a so called high definition shared SAN.

1. ***HD Format?*** Which High Definition format is this SAN able to use? Can <u>all</u> workstations be running 1920 X 1080i @ 60 10 bit uncompressed? This equates to 185 MB/Sec. The difference in data rate between 720p 8 bit to 1080i 10 bit is over a 50% increase. They are both however considered High Definition. As a point of reference SD 8 bit to the top HD is an increase of over 900%.

2. ***Uncompressed 10 bit HD SAN?*** Is this SAN really running at 10 bit uncompressed or is the HD actually compressed? Are the number of users specified based on the lowest, average or the top HD format (1080i 10 bit)? Oftentimes a specification of size or number of streams or users is based on the minimum HD requirement or an average HD format requirement.

3. ***Uncompressed 10 bit HD system?*** Is the specified HD system to use this SAN actually running 10 bit uncompressed? There are many assumptions made if these specific questions are not asked. Who would guess that an HD system that is from a well known and respected manufacturer that cost over $100K is actually running compressed HD? Can this be true? Suffice it to say you have to ask to be sure you are buying an HD system that is producing the highest possible quality. Just because you are paying the highest price does not mean you are getting the highest quality production capability. If you determine the system is not running true uncompressed 10 bit, then the matching SAN is also most likely not capable of running 10 bit uncompressed. In fact, the reason for compressing HD is obviously not to sacrifice quality (that's a byproduct), the reason is to alleviate the extremely high throughput requirements for disks, SAN's and HD systems. This not only makes the systems more profitable for the manufacturer, but it allows additional capabilities like "real-time".

4. ***Scalability?*** SAN Scalability has several dimensions and considerations as follows:

   a. **Number of systems?** The minimum SAN is two workstations. Is this HD SAN capable of scaling up past 3 HD workstations? How far can it go? What is required to accomplish this? Can it be done live without taking the SAN down? Some manufacturers are including such features as "remote mirroring". What does this mean? "Yes, you can see all the data from all the workstations" pause, "just as soon as we copy all the data to the other drives"! Even if the current requirement is not higher than 2 or 3 workstations, 1 Beyond's experience shows that once the benefits of a SAN are realized, it's hard to live without putting additional workstations and applications on it. This not only applies to Editing workstations, but to Graphics workstations, Render Farms and even On-Air systems. The ideal, most efficient and fastest time to air production environment should have all capabilities integrated together around a SAN as the heart of the facility. The 1 Beyond Harmony SAN scales to 32 users at full HD throughput and much higher if you include applications that do not require full HD video (like SD video and graphics). If more than 32 HD seats are desired, the controller and switch architectures of this design actually can grow even further almost with out reasonable bounds. We are setting a false limit here because we have not tested higher.
   b. **Size of storage?** What is the scalability of the SAN storage size? Can it be grown live without taking the SAN down? True 10 bit 1080i uncompressed HD takes up to 10 GBs per minute. This adds up quickly and you always want more. The more functions added to the SAN, the Heart of the production facility, the more storage that's required. Fortunately the price of storage keeps dropping but if your basic SAN architecture is limited, you may face a very unpleasant surprise in the future. Once you hit your limit there is nowhere to go. 1 Beyond's Harmony scales to 256 TBs. Harmony provides enough storage to handle hundreds of hours of HD footage plus hundreds of hours of SD footage and hundreds of thousands of stills, etc. Again, We are setting a false limit in the growth dimension because we have not tested higher. Theoretically this will extend to several Petabytes which has been done in other applications.
   c. **Format?** Can this SAN scale from my starting SD Format to include HD in the future? As discussed, this must be established at the outset.

5. ***Affordability?*** Cost is a major concern today at all levels from independent producers to large corporations. It's important to look at all aspects of cost up front.

    a. **Common Hardware?** If I want to start with a 2 workstation SAN but want to keep my option open to expand, can I use the same components to expand or do I have to start over with new hardware? Harmony is designed to start with a minimum configuration and modularly scale to additional workstations and additional storage using the same components as the SAN grows.

    b. **Minimum Starting Configuration?** Can I start with a single workstation and build a SAN with the initial JBOD(s)?

    c. **Full Cost?** What is the total cost of the final configuration I envision 3 years from now? This is an important question to ask because the answer can be much more significant that the assumed multiple of a smaller starting configuration. Harmony has been compared to other SAN's that are not even HD capable and the cost savings was hundreds of thousands of dollars.

    d. **Support Cost?** What is the support cost of this hardware and software?

6. ***Simultaneous Read and Write?*** Can the SAN do simultaneous multiple reads and multiple writes? Editing is only reading HD video. So you can not ask how many editing workstations can this SAN handle? You must get the video on to the SAN in order to edit it. Without going into the tedious details that are beyond the scope of this paper, let's just state the fact that, simultaneous multiple Reading and Writing is far more difficult than just the multiple reading required for editing. The ideal situation is to be able to be capturing footage while editors are editing it.

7. ***Shared SAN?*** Is this a true *shared* SAN? Strictly speaking a storage area network is not necessarily shared at the file level. The basic concept of a SAN is to share a common storage device to gain efficiency due to the spaces or volumes for each workstation to be variable size therefore not requiring each workstation to have the maximum it may need. This is wasteful and requires redundant hardware for each workstation. Also common SAN storage allows central control and backup. All workstations may be able to access the SAN but not necessarily the same file or even the same volume. There are various schemes for "Sharing". Following are a few examples only to illustrate the vast differences and why specific questions in this area are so important.

    a. **Basic Sharing?** The most basic sharing scheme is to "share" the hardware i.e. all workstations connect to the SAN but each workstation has its own volume. One workstation is not allowed to "see" or access another's volume. This is not due to control, but due to hardware throughput limitations. We will discuss control later in this document. If multiple workstations want to access the same assets in this scheme, the files must be copied to each of the workstations volumes that need to "see" them. The inefficiencies of this sharing scheme in terms of time or space utilization are immediately apparent. In fact, this method is not really efficient at all.

b. **Volume Locking?** The second scheme is to allow access to each others volumes but only one workstation can write to a given volume at any given time. In some crude schemes, only one workstation can even access a given volume at one time. This is referred to as volume locking as all other workstations are locked out. It is a very simple scheme that adds some additional capabilities, but by no means is it the most efficient or desirable.

c. **File Sharing?** The preferred solution is file sharing. This level of "sharing" is the most efficient and does not require any coordination between users. In fact it's completely transparent to the user. Workstations can be working on the same project at the same time without even knowing that others are using the same assets. File sharing will only allow one workstation to write to a file at a time. However, editing is non-destructive. Editing workstations only read video footage. The only time a file is written is when it is originated. When a video clip is trimmed for instance, it has not really changed; it is only the pointers to the clip in the project that are trimmed. Harmony works at this most efficient level. This type of control at the file level requires a unique file system. This not only gives the necessary control but greatly increases read and write efficiency. Standard file systems were not designed for these purposes and are therefore very inefficient for these applications. The details of Harmony's truly unique file system are discussed in detail below.

8. ***Volume Access Control?*** In addition to the ability to have complete common access at the Volume and File level, can the SAN control software both dynamically create and assign access restrictions to a given volume? In most cases common access is desirable but there are situations and applications that require access control. For example, you may have a temporary project that you want to have limited access to the rest of the SAN. Also certain departments or individuals may have requirements for private information like certain work in progress that is not ready for distribution or it contains sensitive material. This level of control obviously has to be restricted and highly protected. Also the dynamic aspect of this control is very important so as not to disrupt on going projects.

9. ***Data / Asset Protection?*** Can the assets of the SAN be protected from disk failure? This Is obviously a very important area to investigate. If the SAN is operating in Raid 0, to gain the necessary performance, the data is unprotected. One disk failure and all data in that Raid array is lost. This can be very time consuming to restore and or recreate, even if there is a recent backup. The ideal SAN is running with Raid 3 or 5 protection. This will allow a disk failure without losing any data. Remember with large arrays necessary to contain any reasonable amount of HD assets, there is a very large number of disks. Therefore, even if the Raid is using Fiber Channel disks with the highest MTBF, the probability of failure goes up geometrically with the increase in the number of drives. Due to the unique Sanbolic Melio FS file system combined with the efficiency of the LaScala Cluster Volume Management, Harmony can be fully Raid 3 protected and still have the necessary sustained data throughput to meet HD requirements.

Summary of Design Goals:

In summary, the design goals and definition of 1 Beyond's Harmony were derived from working through the above set of requirements at the outset. If you use the above set of test questions and receive the correct answers to all of the questions, you have 1 Beyond's Harmony™ SAN environment. Following is the major features summary:

II. 1 Beyond's Harmony^{tm} Shared SAN Major Features:

- ***Complete Scalability:***
  - 2 to 32 1 Beyond HD Editing / Compositing Workstations
  - Much higher number for SD Workstations
  - Up to 256 TBs or hundreds of hours of HD and hundreds of hours of SD
  - Plus hundreds of thousands of stills, audio tracks, titles etc.
  - Performance scalable from SD to full 1080i HD
- ***High Performance with Protection:***
  - 800 MB/s per controller while operating at Raid 3
  - Up to 6.4 GB/s aggregate throughput
  - Fiber Channel connection at 4 MB/s
  - Simultaneous multiple reading and writing up to 1080i format
- ***Flexibility:***
  - Central shared storage confidently located wherever suitable
  - Up to 1 KM Fiber Channel workstation connection
  - Can be utilized for all asset management
  - Can be tied into 1 Beyond's Redline Render^{tm} farm shared processing engines
- ***Efficiency:***
  - Each workstation has an automatically variable amount of storage
  - No requirement for Max storage per workstation as in direct connect storage
  - Improved efficiencies, lower TCO and higher ROI
- ***Manageability:***
  - Central backup of all media assets
  - One location for storage maintenance, access control and management
  - Total open systems environment

- ***Workflow Improvement:***

  - No need for file copying to gain access (significant for multi-GB files)
  - All graphics, stills and video immediately available to all departments
  - Simultaneous reading and writing means editing can begin while video is still loading
  - See **1 Beyond White Paper "Workflow Benefits of a Storage Area Network"**

III. Now that we have defined the ideal shared SAN environment, how was it accomplished?

To create Harmony, 1 Beyond was very fortunate to have worked with a very select group of the top hardware and software engineers in the storage industry. The selection of partners included a top hardware raid controller manufacturer, a top high performance fiber channel switch manufacturer and perhaps the most advanced control software and file system company in the industry. The development required many man-years of effort including extensive microcode changes in controller hardware, hardware design updates and extensive testing of control software and its integral file system with multiple 1 Beyond HD Pro workstations. Our software development partner was Sanbolic. Through an exclusive arrangement, 1 Beyond based the Harmony control and performance, at the software level, on Sanbolic's very mature and well tested Melio file system.

Both parties quickly learned that software, no matter how well designed, was not capable of accomplishing our design goals with any off-the-shelf, existing hardware. This is because at the most basic levels, SAN hardware is designed to accommodate **the requirements of 95+% of the market.**

- **Many small transactions from high numbers of users that involve small files using applications that are not time critical.**

The requirements of Video Editing are the exact opposite.

- **Huge files accessed by a small number of users with applications that are so time sensitive that millisecond delays are the difference between working and worthless.**

1 Beyond tested many of the top raid controllers and fiber switch designs in the industry and consistently came up short of the goals we set. We learned that this set of design goals would not only require hardware capable of extensive redesigns at the microcode level, but most hardware was not even capable of being repurposed due to its basic architecture.

IV. What is the importance of Open Architecture and Standards in this design?

In general we have started to see an interest, in fact in some cases even a demand for Open Architecture and Standards in the video market. The video market has been slow to demand the benefits of open architecture and standards because proprietary products have been a way of life here. In most of the computer world open architecture is now so common it's taken for granted. As an example, if a company offered a great PC that wasn't standard, even if it was significantly faster, cheaper and incredibly better, it would likely fail. Fairly recent examples of this can be sited, but now it's pretty well understood that in the PC area at least, this is not a good marketing idea. People and especially companies do not want better, if it means sacrificing standards. One of the benefits of standards is open architecture. It is generally not desirable to be locked into one companies technology or sometimes more importantly, their pricing. This SAN development is a great example of how proprietary products can be limiting.

V. Hardware Modifications:

The hardware modifications necessary in this design were only feasible because of 1 Beyond's commitment to a truly open architecture. We were not locked into using an internal company division for storage hardware and/or software development or any external manufacturer's products. **As is our open architecture policy, we were free to choose and determine the best of the best to make our goal of the industry's first uncompressed 10 bit HD SAN, a reality.** It is now clear; we would not have been successful otherwise. This is undoubtedly a contributing factor in some other company's not reaching this goal.

The details of the hardware modifications that were necessary are far too detailed for the scope of this paper. Suffice it to say during this development, between PCI and Fiber Channel analyzers, several 1 Beyond HD Pro systems, HD Decks and Monitors, high-end raid controllers, disk arrays and high performance fiber channel controllers, we had close to $1 million dollars of equipment tied up not to mention several of the industries top storage design engineers for months.

Therefore we will concentrate the remainder of this paper on the design requirements for the Software and the resulting Work Center Workflow Benefits of the Harmony SAN.

VI. What is the File System and Control Software used in Harmony and how does it work?

Melio FS is a file system specifically designed for high performance, high reliability, and heterogeneous operating system network storage environments, with an emphasis on the core SAN environment. See Appendix A for a full description of the Melio FS File System and how it operates with 1 Beyond's Harmony storage area network and RedLine Render Farm.

VII. What is the result of the Harmony SAN design in terms of workflow benefits?

You can see in the following diagram (Figure A), Harmony can be the “Heart” of a production environment and bring benefits to all work centers. The following efficiencies are introduced with the 1 Beyond Harmony Shared SAN Storage:

1. ***File Access:*** All assets including SD and HD video, stills, titles, audio clips, music etc. are available to all centers of the production facility all the time. This is apposed to the typical workflow that is constantly moving files from one center to another.

2. ***Storage Performance:*** When the amount of storage is increased, the performance increases. The more spindles (disks), the lower the access time and higher the data rate of data access (increases up to the maximum limit of the data pipe and then it levels off). Therefore when you combine storage requirements into a SAN, you benefit from the aggregate amount of throughput performance and reduced access time.

3. ***Revision Control:*** If all work centers have access to the same file, if that file is updated, it is automatically updated for everyone accessing it. Also, if the update changes the asset and you want to make all centers aware of the change, you can change the name of the file with a new revision number. For example when Premiere, is opened on a project that uses that asset, it will be marked off-line and the replacement will have to be identified. Depending on the naming conventions used it would likely be easy to identify with the same file name with a revision number change.

4. ***File Maintenance / Cleanup:*** This also keeps the amount of out dated files (cleanup) to a minimum. It is easier to maintain file cleanup if there is only one copy of each file.

5. ***Protection / Backup:*** If all assets are located in one location i.e. in a Shared SAN, one backup program can service all centers simultaneously. This also reduces the cost of the backup hardware and media personnel required to perform the tasks. The job of backing up each individual workstation or relying on each user to perform these tasks is not realistically possible. Hence, disaster is always looming. If it is centralized in a SAN, backup can be performed with the regularity desired. Copies can even be kept in another facility for the ultimate protection.

6. ***Reduced Disk Hardware:*** One of the basic efficiencies introduced with a shared SAN is Disk hardware including actual drives as well as controllers, cables, space, etc. Without shared storage, each workstation has to be equipped with the maximum amount of storage anticipated plus a buffer to keep a reasonable amount of working space for temporary files. This usually leads to 50 to 66% of each workstations disk space unused most of the time.

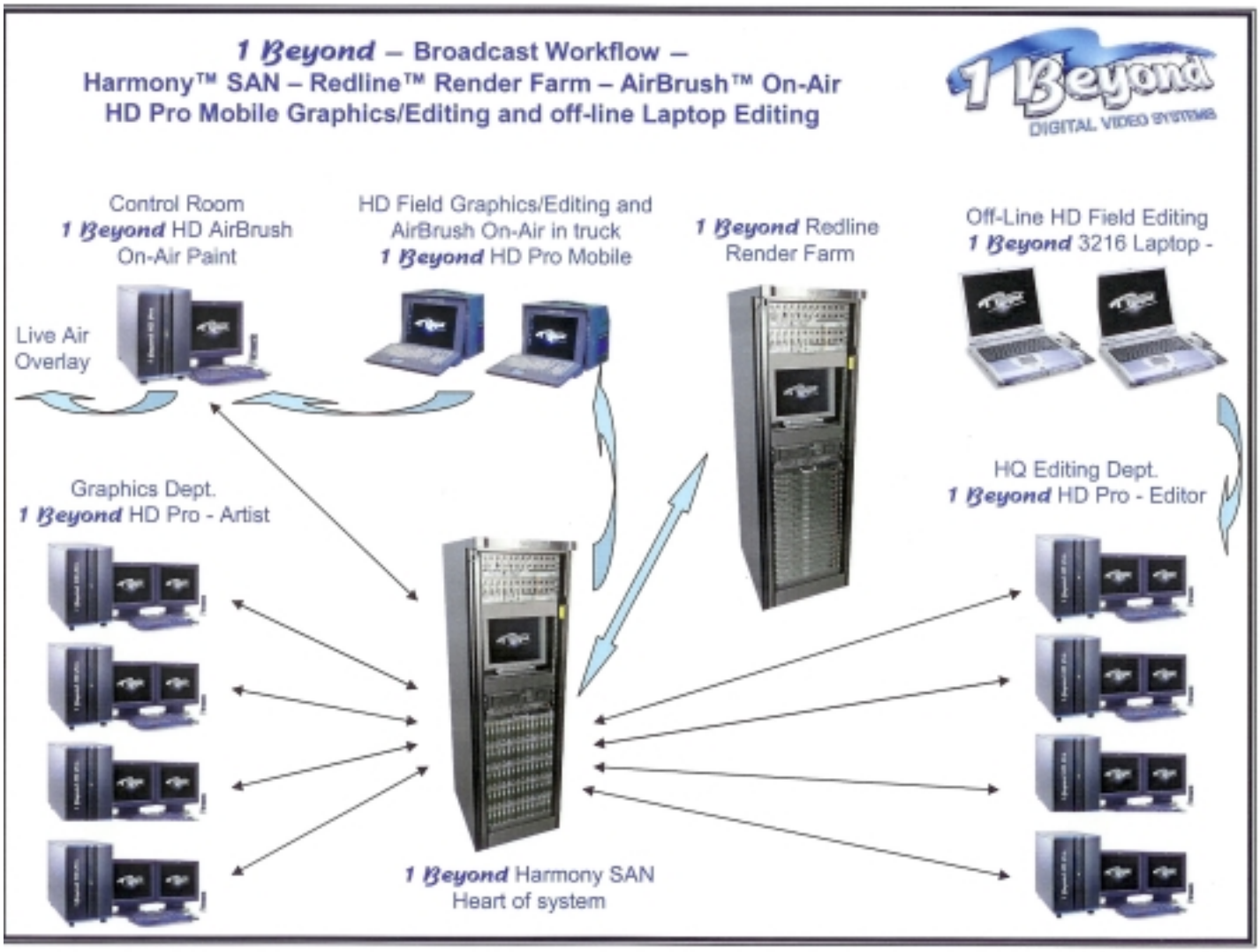
7. ***Noise Reduction:*** One of the subtle but very important benefits of introducing a SAN to a production environment is noise reduction. We have become so accustomed to having computers present; we tend to forget how pleasant it was when offices didn’t have this additional noise (for the older of us that is). Noise is a distraction but it also a large fatigue

facture. We are constantly striving to make computers quieter for these reasons but with the introduction of HD we have a new problem, massive disk arrays. There is no easy way to eliminate this noise because unlike computers, the basic part of a disk is moving. The biggest offender of this is SCSI drives which rotate at 10 to 15,000 RPM. Worse yet, for reasonable HD performance you need a minimum of 8 of these noise generators. To cut a minimal 30 minute production you would like to have 14 to hold the raw footage. Sorry, but if you are cutting a feature length film you should have more than twice that amount! If you have not experienced this kind of noise, don't. The only way to begin to hear what you are doing is to wear headphones. Fatigue is not a strong enough word. The benefit of a SAN here becomes obvious. Note: Consult the White Paper **"New Design Considerations for HD Editing / Compositing Workstations"** for other noise reduction features like 1 Beyond's HotRock™ Storage and Silent Partner™ non-SAN options for our HD Pro Workstations.

VIII. Summary of Workflow Benefits:

The above benefits are surrounding the actual Shared Storage aspect of Harmony. For a detailed look at the individual work center workflow benefits i.e. graphics, editing, on-air etc. consult the **1 Beyond White Paper "Workflow Benefits of a Storage Area Network".**

Figure A: 1 Beyond's Harmony as the Heart of Broadcast Workflow

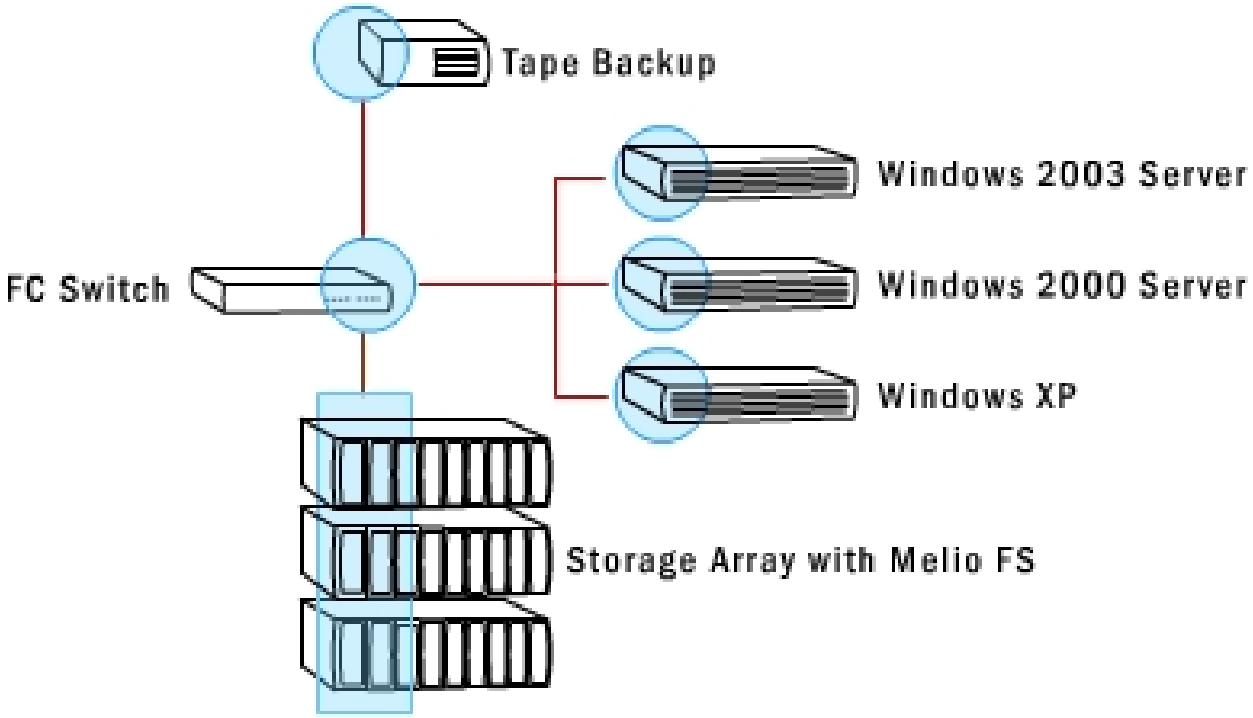


Appendix A: Sanbolic Melio FS File System

Melio FS

Melio FS' design ideas are obvious, but the interpretation is strikingly fresh. Those familiar with storage solutions will immediately understand its advantages. From its fundamental clarity in the layout design to the finesse developed in each individual module, Melio FS presents an uncomplicated solution, bringing beauty and simplicity into the complex and cluttered storage and computing world.

The engineering of Melio FS was seemingly impossible. The hefty goals, with accentuation on high availability computing, while retaining speed and NAS serving ability, were achieved by careful development of tightly integrated modules. At its heart you will find Melio FS' three-dimensional layout, delivering incredible speed, capacity and power. A seven layer locking mechanisms with a B-tree like structure, combined with the transaction managers, provide the control, stability and reliability of Melio FS. The seamless distribution of metadata within the file system presents unprecedented processing power, handling, and performance, and eliminates any compromise between meta data and file sharing. Never before has a single file system embodied such an exceptional duality.



Melio FS is a file system that can be used across all servers attached to the centralized storage. This greatly simplifies the access to and administration of data and storage. It improves computing resource flexibility and scalability of applications.

Melio FS in Different Work Environments:

How can Melio FS be suitable for environments as diverse as workgroups, e.g. NLE Editing and clusters of high-density servers, e.g. Render Farms, without compromising performance, security or specific application requirements?

The tightly integrated modules that make the foundation of Melio FS were carefully developed for some of the most complex applications in storage networking today. The different requirements we encountered, such as support of small as well as large files, were added into the modules as Melio FS was being built. Some examples of priorities that customers in the workgroup environments ask for are for uninterrupted workflow, and high performance of large file transfer. Enterprise data environments demand fast recovery and reliability and in cluster or grid computing the customers ask for NAS-SAN convergence and high data availability. We continue to listen to our customers as the development of Melio FS goes on.

Melio FS in Workgroups:

Melio FS provides a fundamental solution to simplify data center complexity. Melio FS presents a single shared storage pool across all servers attached to the centralized storage. All nodes can share read and write access to a file over fibre channel, while maintaining coherent cache. This greatly simplifies the provisioning and administration of fibre channel storage (no need to set up a LUN for each server as is commonly done with most file systems today), and enables consolidation of SAN islands based on heterogeneous operating systems or equipment. It improves computing resource flexibility and scalability of applications by allowing clustering of multiple application nodes, all of which have shared access to data over fibre channel. Additional servers or storage can be added on the fly, while the existing hardware continues to operate. Full journaling and recoverability to guarantee data integrity is designed into the file system architecture.

The file system also embeds NFS and CIFS publishing, and allows multiple NAS servers to access the SAN, integrating NAS and SAN storage and solving the NAS scalability issue for applications not requiring fibre channel-based block access to the storage. It allows decisions on block versus file service, fibre channel versus iSCSI, and the structure of storage volumes, and OS platform to be driven by application requirements and economics, rather than by limitations of the hardware.

Completely compatible with existing hardware, Melio FS enables integration of the heterogeneous data center, and allows the customer to match hardware with performance needs. This is particularly important when trying to solve the data flow required for High Definition editing. This is an extreme application in the fact that unlike most transaction oriented applications, editing is very time critical. If the File System does not respond in time, you loose a frame and the output is useless.

Melio provides a single platform on which to build advanced storage management applications. It can be introduced incrementally into a data center, and provide a migration path to integrate all data storage over time. By using primarily existing disk administration tools, the learning curve is minimized. Storage architecture simplification is fundamentally the solution to improve reliability and cost position.

Melio FS In Cluster Computing:

Melio FS improves computing power and processing by allowing clustering of multiple application nodes, all of which have simultaneous access to shared data. Cluster computing is used to process and solve data-intensive problems and tasks such as 1 Beyond's Redline Render Farm. Increasingly, cluster computing is used to enable multiple Intel-based servers to handle large applications that would otherwise run on large SMP servers.

The design point for Melio FS was a telecom application using high-density clustered server blades to process data from hundreds of incoming feeds, and then write it to the shared storage. This meant a constant stream of simultaneous writes to shared files from the individual server blades over fibre channel connections. Multiple NAS heads were used to distribute the shared data from the storage to end-users. The system was effectively a large SAN data center supporting high performance cluster computing and integrating NAS support. No existing file system could support these requirements. A high performance shared clustered file system with distributed locking mechanisms was needed.

Melio FS was designed to support the requirements of intense cluster computing, requiring shared data access and heavy transaction handling meaning either the processing of enormous amounts of small files, transfer of data through hundreds of parallel data paths, transfers of large files, or a combination of these. The design of the file system includes multiple level locking mechanisms that in combination with efficient metadata handling provide extremely fast access to data.

Appendix B: Sanbolic Melio FS FEATURES and BENEFITS

FEATURES	BENEFITS
Produces a single shared pool of storage to all servers attached to the centralized storage. It is not necessary to provide separate LUN for each server.	Simplifies the access and the administration of data and storage, and improves computing resource flexibility.
Uses existing disk administration tools.	Easy to install, manage and use, minimized learning curve.
Simplifies storage architecture	Fundamentally the solution to improve reliability and cost position.
Addresses customer needs in; *workgroup environments*, *enterprise data centers, cluster or grid computing* applications.	*Workgroup environments*; improves workflow, and eliminates need for other file sharing software. *Enterprise data centers*; improved data management, and scalability. *Cluster or grid computing* applications; high performance, and SAN-NAS convergence.
Heterogeneous operating systems support. Melio FS supports Windows platforms, and the file system architecture supports all major Unix platforms.	Eliminates the need for virtualization and LUNs, and allows sharing between operating systems.
Full NFS/CIFS publishing support	Enables multiple NAS heads to serve files from SAN storage.
Distributed locking mechanisms allow simultaneous read and write operations to different parts of the same file from different machines.	Gives high performance across a range of file sizes, with intense read/write activity, transaction updates, and heavy volumes of file creation and deletion.
Journaling; Melio FS guarantees metadata consistency by keeping a disk log of meta data operations.	In case of system failure, the journal allows recovery of file system structure without taking the volume offline.
Metadata server is not necessary.	Eliminates a potential single point of failure.
Decentralized metadata.	Allows high independence of servers accessing the same volume.
SAN clustering operations support. Each computer within the cluster will see the volumes, formatted with Melio FS, as their local disks.	If a member of the cluster fails, the remaining computers detect this and recovery of Melio FS metadata is based on the journal of the failed machine.
64-bit file system	Unlimited scalability (up to 18 million terabytes).
Optimized random file access	Enables faster access to data in a random pattern from a single file or multiple files
Tunnel caching (optimized directory lookup).	Speed and performance
Support of sparse files	Saves space on the storage

Appendix C: Sanbolic LaScala™ Clustered Volume Manager

Sanbolic adds LaScala to its product line of advanced software and file systems for shared storage solutions. These products serve the needs of large enterprise data centers as well as high performance shared file applications such as **high definition video**.

LaScala is a host-based volume manager that simplifies management and improves flexibility and reliability of shared storage environments. The new volume manager provides greatly improved disk management capability when used as a stand-alone product with native file systems. Multiple hosts can simultaneously share access to and administer storage volumes, and volumes can span multiple storage controllers. LaScala permits centralized administration of per volume groups based on native administration rights. Assignment of host access to specific shared volumes is easily accomplished with native OS security tools. The volume manager can apply multiple changes in volume structure simultaneously, such as online volume creation, deletion, parameter modification, resizing by adding additional storage, or adding new hosts to a volume group.

LaScala incorporates fault-tolerant symmetrical clustering mechanism based on leased locks, fine-grained distributed opportunistic locking, and Object Oriented transaction management technology. LaScala provides easy-to-use disk management, including online configuration, repair, hardware changes and other maintenance of shared logical volumes. These volumes may be arbitrary combinations of volume sets, stripes, mirrors, RAID 5, etc. The configuration operations are fault tolerant by the use of journaling for all operations.

An initial customer installation has achieved greater than 1 GB/sec throughout to a shared volume from 5 Windows-based nodes. The second release will support concurrent access to shared volumes from hosts running heterogeneous operating systems.