Virtualization Means More Than One Media

Records management and archive application users for blending of storage media

By Roy Slicker and Jim Wheeler, Pegasus Disk Technologies, Inc.

Virtualization: According to Boston, Mass.-based Aberdeen Group, Inc., virtualization is technology that enables the separation of representation of storage to the server operating the system from actual physical storage. This division of physical storage devices from the logical storage space presented to users and applications turns storage into a generally available utility pool. Virtualization fulfills a roll for storage similar to that which an operating system does for a computer system—namely making programming and operation simpler by automating resource management "housekeeping". When this process occurs, computer users are said to be "viewing resources at a higher level of abstraction." Thus in short, virtualization is the abstraction of storage. This definition of virtualization facilitates in the ability for customers to pool a wide range of storage technologies from different vendors and formats, to easily add capacity, readily move data, or automate moving data among devices independent of server operating systems or network infrastructure.

Pegasus Disk Technologies
2333 San Ramon Valley Road
San Ramon, California 94583
Phone (925) 314-1800
Fax (925) 314-1801
www.pegasus-ofs.com

In a report issued earlier this year by Aberdeen Group, Inc., Analyst Dan Tanner advised IT customers to consider the potential benefits of storage virtualization which, he said, include: creating virtual volumes that can span multiple storage units, enabling heterogeneous "mix and match" server and storage device usage, serving as a platform for cross-device storage services, and enabling secure storage sharing and efficient storage utilization.

The general storage market, particularly the Storage Area Network (SAN) and Storage Service Provider (SSP) segments, has been very active with several new entries being funded on a large scale in a relatively short period of time. While this is good news for users looking for new products and services for growing data bases, increasing backup windows and short term storage applications, these systems do not address the specific needs of the large and growing archive and compliance segment of the market.

While current SAN and SSPs battle for market share, the need to store strategic and compliance or regulatory records has baffled providers because of the reluctance of enterprise customers to out source storage, while on the other hand, they continue to use magnetic based RAID for the storage of regulatory records. All the energy and money being pumped into the marketing of SAN and SSP systems is having a positive and wider spread effect on other storage companies in general by raising the awareness of storage products. It appears that a growing number of users are more interested in the secure and reliable storage aspect rather than just adding massive internal or off-site storage capabilities.

Lately, there have been more definitions of virtualization than there are storage company start-ups, or perhaps shutdowns. There's a lot of hype around what "it" is. SAN vendors have been saying that SANs are what virtualization is all about. The problem is that the current SAN systems are not including software to handle <u>all</u> storage media types, thus applications that use or should use archive-based media do not benefit from this new "virtualized" world of storage. Almost everyone talking about or offering storage virtualization products limit their offering to magnetic based media, however, magnetic media only addresses part of the storage puzzle. Magnetic media does not address users that have historically used archive based media, including magneto optical, CD and DVD for archiving and continue to do so. Storage virtualization software must be able to address most, if not all a users need for storage in order to deliver on the promise of ease of use, scalability and cost effectiveness. SANs basically solve a wiring issue, but it does not provide software that virtualizes all storage media available. One thing is very clear: A limitation of virtualization techniques to only magnetic media will limit user functionality and market opportunities for this must-have technology.

Storage Virtualization must include secondary storage devices.

Secondary Storage is a multi-billion dollar business, characterized by its ability to remove the storage medium (cartridge) from the device and store it away from the computer, if necessary e.g. off-site for security.

How valuable is your data?

"One third of businesses that lose all their data fold within 2 years" Infocorp

"Two out of five enterprises that experience a disaster will go out of business in five years." Gartner Group

"50% of companies suffer data loss through human error" Financial Times

"Even with RAID, there is still a single point of failure" Zitel

Recent Gartner Dataquest research on application downtime shows that an average of 40 percent of downtime is caused by application failures (e.g., performance issues or "bugs"), 40 percent by operator error, and approximately 20 percent by system or environmental failures. The majority of the failures in the "system or environmental" segment — 60 percent — is caused by hardware problems. Overall, less than 5 percent of application downtime is attributable to disasters.

If the following statistic is widely known, few people do much about it: 80% of on-line information is static and unused once it has been stored. Jon Toigo , author of "The Holy Grail of Data Storage", estimates "Corporate information technology departments will spend between 75 and 90 cents of every hardware dollar over the next five years on data storage products".   In addition Toigo goes on to quote an IDC estimate that states 55% of distributed storage management costs are administrative—a number that could be significantly reduced through logical consolidation of storage resources for more effective management. If the static data were migrated to lower management cost near-line or off-line storage technology, then an IT (Information Technology) Manager would gain the benefit of re-claiming a vast amount of their current on-line storage investment.

Gartner Dataquest predicts a move towards managing larger amounts of storage, quicker and in remote site based on recent events. The five-year outlook remains strong for data storage growth — driven by the use of Web-based applications, multimedia data and data warehouse/business intelligence implementations — demands better storage administration tools. Gartner Dataquest forecasts the market will grow at a CAGR of 26.1 percent through 2005 to reach $16.7 billion.

Primary Data vs. Archive Data

Maybe you've seen this: *"It's taken 300,000 years for humans to accumulate 12 exabytes (billion gigabytes) of information. It will take just 2.5 more years to create the next 12 exabytes"* – according to a study by a team of faculty and students at University of California at Berkeley's School of Information Management and Systems.   It is important to point out that the "information" being created is made up of several different types of data and not just data that is stored solely on magnetic disks. There are two types of data that should be addressed here: data that is used for a relatively short period of

time (3 to 5 years), such as databases, backups, application programs, day-to-day transaction data, etc; and data that must be stored for long periods of time – archived and compliance data based on self imposed or government regulations.

Here's another interesting fact. Currently there are quite a number of users that were sold expensive magnetic-based systems. They are using these systems to store critical compliance-based or archive data, even though there are regulations requiring the storage of information on archive based (non-erasable) media, and they are paying the fines or penalties for doing so. Many users are told that adding archive media to their system, which was extremely expensive in the first place, is too expensive to do or that the capabilities simply don't exist. They get the advice to add more SAN based storage and back it (up) with off-site systems. This may increase storage, but still doesn't fulfill the requirements for properly storing regulatory data. In addition, as the costs to manage data continue to rise, users are more and more reluctant to add the right storage types to the system, which in reality would be more cost and data efficient. Many rationalize these decisions by saying that the system enhancement costs are more than the fines or penalties. This calculation is only true, if the data is never lost or manipulated. If the data is compromised, then the cost to the company could be significant. Doing the math, magnetic is the most cost effective media for data that is changing on a regular basis. Contrary to what many believe, archive based removable media such as optical is less expensive than fixed magnetic based media in an archive environment considering how long certain data has to be kept and maintained. Archive media can be stored off-line and the cost of data expansion is merely the cost of additional media.

Application solution providers that address storage needs using magnetic based media systems have difficulties selling to archive applications. Those offering technologies such as removable optical (MO, DVD and CD) based media solutions for compliance-based and archive applications tend not to meet the needs of applications where magnetic media is best suited. As a result, users that utilize RAID within a SAN for databases, application or other data end up utilizing separate systems to address workflow, back office records management or compliance based document or data management needs. In a survey of archive and compliance based application users and developers, one thing became very clear: There is a critical need for software to centralize access to a broader range of media types. Storage virtualization should mean total storage virtualization, not just fixed disk magnetic storage virtualization. The impression is that storage offerings remain disjointed, resulting in less interoperability and greater expense for storage solutions. Users want a system that can search the entire network or enterprise from a single console and map all the available storage resources, whether from one vendor or not, and map those into a single share point or volume.

All data is not created equal and should not be treated equally, unless of course you don't mind the thought of losing critical information that is stored on media that is prone to a 3 to 5 year life span. Data lost to vandalism, unwanted changes, accidental loss or loss due to a disaster or user error is increasingly more evident. There are installations that use magnetic based media to store everything, even though there are inherent risks associated with that approach. There are also those applications in records management

environments, document imaging and government applications, such as those mentioned above that use archival media and even tape to store things for longer periods of time. The fact is that, applications where magnetic based storage solutions are best suited do exist, and there are applications where archive media or some form or tape is the best approach. This is not a battle between the storage media "camps". What's needed is a common sense approach to solving the problems of speed of access, overall cost of the system, accessibility of long-term data, and ease of access, above all, while abstracting out the various layers (media types) and providing a common user interface.

The intention of this article is not to promote one storage technology over the other. It is simply pointing out that users should not have to choose between one or the other. In fact a user generally does not care how the data is stored so long as it is readily accessible and when it is needed and secure. Virtualization of the storage devices should provide a system that handles all categories of data (and probably back-up as well) in a cost effective way and with a single-user interface that is transparent to the user.

Many users are asking for a blending of the various storage technologies, a virtualization. Integrators, developers and users are asking the system integrators, "Let me decide, what goes where, based on my data access requirements and my legal exposure for storing data". There is also a concern about the cost of the overall system and its effectiveness in tracking information and recovering from failures. Certainly there are users that want to be told what is best. That is fine, as long as they are given choices, as each customer requirement is unique to each market they address. A single media solution for all data is not a choice and can be a disaster, if critical information is lost. The fact is that there are solutions available that best fit different data storage requirements, and the "blending" of these technologies would provide the user with the highest levels of performance, security, longevity and cost effectiveness.

A solution may be at hand with the integration and virtualization of the three most common storage categories: primary storage -- magnetic disk (including RAID), secondary storage -- optical disk (MO, DVD and CD), and back-up-- tape. To date, it seems to be more of a battle between magnetic-based disk vendors and the "other" technologies. Everyone is fighting for market share. What's needed is an augmentation of both feature and media benefits that allows a comprehensive approach to expansion, rather than replacement. Virtualization provides a secure, strategic process for the full spectrum of storage consumer needs, merging security with simplicity, while using an existing company infrastructure more efficiently than it is being used today.

Compliance and Archive Data

 "Why should I treat this data any different than general day-to-day information?" There are many regulations placed on the storage of data, either self-imposed or government-based. These regulations specifically state what data needs to be stored how.

Data that is considered critical is also often stored for long periods of time, such as medical, financial, insurance, securities transactions, customer transactions that may be effected by legal action and so on.

Workflow operations such as banking or medical environments that require maintenance of information for 10 to 15 years or as long the patients or customers are alive are examples of compliancy of regulatory data. Data is brought in via computerized forms based system or scanned and indexed by capture software. Once the data has entered the system – it's critical data. There is a need to store and access that data over a number of years, as much as 50 years or more in some cases. Magnetic media is just not able to keep data readable for that period of time, especially contact magnetic media such as tape, however, many IT managers have decided to "take the risk" and store archive data on tape. Some have discovered through either a restore requirement or just through a database search that magnetic tape does not have a long shelf life and files have gone missing. These are areas where archive media is critical.

A Virtual Solution

Storage virtualization is about providing software tools and connection strategies that automate which particular storage technology should be employed to hold a piece of information according to how it's accessed. Pegasus Disk Technologies provides single software interfaces that can be used to scan the enterprise for storage resources and or volume to facilitate, particularly those used in applications that require storage of data/documents for long periods of time. This type of data, being a significant portion of the data stored globally, is often referred to as compliance based or archive data. It includes medical records, stock transactions, insurance records, banking information, e-mail storage and a wide range of government records. Pegasus offers a simple no-nonsense definition of virtualization of storage hardware for information stored in these systems, showing how such a virtualized storage system can work and how the user or developer benefits from both cost effectiveness and ease-of use-standpoints.

What are the benefits of software virtualization of all these media types?

One of the major benefits of virtualization of the three media types is the cost associated with the various storage resources within the mixed and shared hardware and media pool. Data that is used and/or changed frequently on a hour-by-hour, day-to-day, or even month-to-month basis belongs on magnetic disk or RAID. Speed of access is critical here, and magnetic-based media is the best fit in this case. However as storage requirements increase, storage costs can skyrocket while all data remains on non-removable magnetic media. It is not uncommon to see SAN based storage systems utilizing RAID storage costing anywhere from a few hundred thousand dollars to well over a million. Of course these systems offer terabytes of capacity, and many are scalable based on storage needs. But when the typical RAID-based system is analyzed, it can be quickly determined that the vast majority of data on the RAID is not accessed at all, much less frequently. This static data can be off-loaded to much more cost effective scalable storage media such as archive based solutions, thus cleaning up the primary

storage system, and reducing the amount of time it takes to back-up the data on the RAID system. This will facilitate in the use of a scalable removable media device that is much more cost effective to expand while the data in the archives continues to grow. Other benefits are that a logical consolidation of storage technologies from a single volume facilitates in fewer IT personnel to manage growing data needs, and a complete use of storage technologies would facilitate in a decrease in operating downtime and the reduction of lost data. It is a fallacy that hard drives do not fail in array products. In fact, RAID exists because hard drives do fail. Simple statistics show that a 50-device array will loose on average three drives per year. If the system is a RAID device, chances are the data can be reconstructed or is mirrored on another drive, but if the device is JBOD (Just a Bunch of Disks) the data will be lost. Archival storage devices are more resilient than non–removable storage technologies by design. Should a drive fail in a removable media library, the data will not be lost. The media must only be moved to an operational drive, and voila, the data will be back on-line. Using archival storage solutions reduces back-up windows. Given the previous statement that removable media is inherently fault tolerant, only primary data need be backed-up. Data moved off to optical media solutions would not need to be part of the back-up strategy, thus reducing the back-up and restore window by as much as 80%.

Storage System Monitoring – Virtualized Resources

Virtualization is abstracting out the specific hardware as well as individual media surfaces and/or types, resulting in a single view or volume for user interaction. The diagram below shows what storage virtualization must be in order to serve on an enterprise, utilizing a multiple or single server-attach approach.
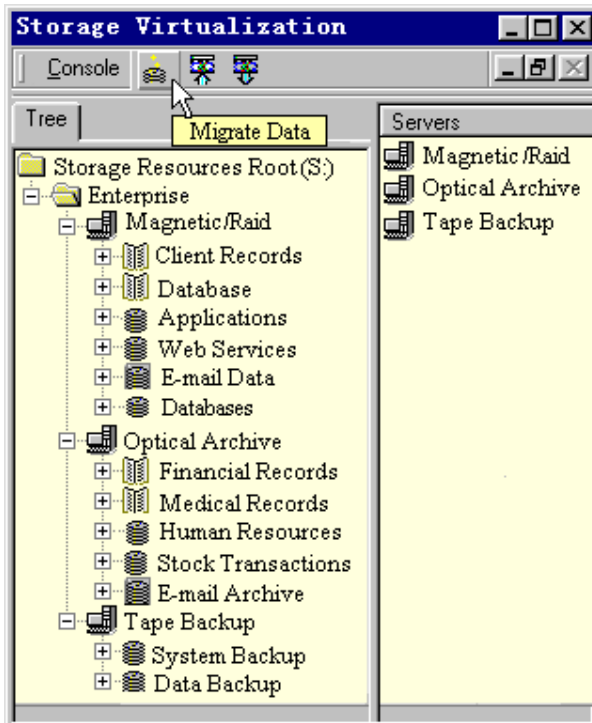
All the different physical storage types on the network can create a significant storage management dilemma. A single logical view of storage using virtualization techniques can streamline data access and speed overall data management. Within an organization, data storage devices or resources of different types can be distributed in different locations or centrally located within a controlled environment. With software-based virtualization such as this, storage devices and sizes, locations and methods used to attach them to the network become irrelevant to the user.

At some point, an enlightened storage integrator will offer the market a virtualization solution that includes the option of using any or all of these specific storage levels. Appropriate and reliable physical media types will automatically be chosen through the user's application. Based on parameters previously set, the location of the stored data will be transparent to the user. This kind of media blending or virtualization will facilitate in a cleaner SAN that complies with the various regulatory requirements of the appropriate agencies for the specific user market, and   will provide a scalable solution over time that is more cost efficient to manage and grows as the data explosion continues.

 For more information regarding Pegasus Disk Technologies or Pegasus-OFS, contact Jim Wheeler at 925-314-1800 or go to www.pegasus-ofs.com.