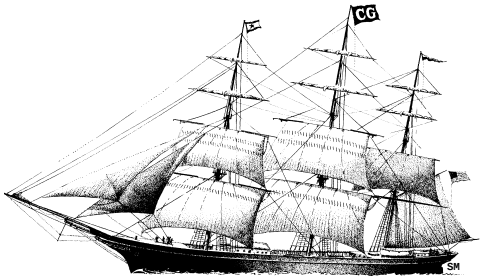
THE CLIPPER GROUP
Explorer™



Navigating Information Technology Horizons℠

Published Since 1996 Report #TCG2005014 March 15, 2005

Keeping Up with the Data — Solution Strategies for Information Storage

Analyst: Michael Fisch

Management Summary

If you are responsible for the well-being of your enterprise's information storage, then you also contribute directly to the success of the enterprise. Timely access to the right information is like breathing for business operations – they need it to run. **Storage is part of the infrastructure that enables a business to function and achieve its goals and objectives.**

By nature, a storage infrastructure is always stretching and evolving. Continuous data growth, rising service level expectations, and a myriad of other factors conspire to make sure the job of IT is never finished in this area.

So, what are your storage challenges? Survey the enterprise environment. Walk through the data center and see the racks of storage equipment. Watch the screens showing output for management tools and applications. Talk to the busy storage administrators. Glance at the PCs, workstations, and peripherals on people's desks as you walk through the office. Talk to the users. Find out how remote offices handle it. Consider the whole operation from end to end – how the enterprise stores, manages, and protects its information.

If your IT operation is of any size, you will find that many activities are involved in keeping it going. Backups and restores, provisioning capacity, growth planning, data migration, adding applications and servers, etc. are all part of the routine. You will also find problems to solve. No matter how sophisticated the operation, skilled the administrators, or large the budget, it is just the nature of the beast.

Whatever your biggest storage challenge, it is likely related to one of the following:

1. Keeping up with data growth without overspending on hardware
2. Containing IT staffing levels and operating costs
3. Data retention for regulatory compliance
4. Backup and recovery challenges
5. Disaster recovery and business continuity
6. Insufficient application performance

These are common challenges among enterprises of all sizes. The good news is that there are solutions for all of them. Read on for a description of each one and suggested solutions.

IN THIS ISSUE

- **Solution Strategies for Top Six Storage Challenges.....2**
- **Storage as a Service to the Business ...7**
- **Conclusion.....7**

Solution Strategies for Top Six Storage Challenges

#1 - Keeping up with data growth without overspending on hardware

Data growth in the Information Age is a constant, like gravity or the changing seasons. This growth comes in the form of e-mails, database records, documents, videos, scientific measurements, and any other digital object. Enterprises run on this information, so they have to keep up. Next year's capacity requirements will be greater than today – the only question is how much more.

Enterprises traditionally have storage dedicated to each server, whether internal or external, and scaled it by adding drives or upgrading servers or storage arrays. However, this leads to many disconnected storage islands, poor capacity utilization, and many management headaches. In short, enterprises have spent too much to keep up.

Storage consolidation (for scalability)

The foremost solution for cost-effectively handling data growth is storage consolidation. Rather than discrete islands, consolidation separates storage from individual servers and establishes a centralized, shared pool. This is more efficient for similar reasons that carpooling or public transportation is more efficient for commuting – utilization rates are higher and operating costs are lower, which means lower total cost of ownership (TCO). There are three primary architectures to consider:

- **Centralized direct-attached storage (DAS)** – Multiple, heterogeneous servers connect directly to a shared storage array, usually using Fibre Channel[1] or SCSI[2] cables. It provides dedicated block storage for servers. Since scalability is limited, this configuration is most suitable for smaller IT operations with a handful of servers and modest growth expectations.
- **Storage area network (SAN)** – Multiple, heterogeneous servers connect to shared storage array(s) over a network. A SAN provides dedicated block storage to servers running any kind of application, such as ERP, CRM, databases, e-mail, and file servers. Most use Fibre Channel connectivity for its robustness, bandwidth, low latency, and long cabling distances. An up-and-coming alternative is *iSCSI*, which allows enterprises to use more familiar and less costly IP networks. In general, SANs provide fast, tunable performance, secure partitioning, and virtually infinite scalability for storage consolidation.
- **Network-attached storage (NAS)** – Multiple, heterogeneous clients and servers access shared files over an IP network using the standard *CIFS* (*Windows*) and *NFS* (*UNIX*) protocols. NAS platforms are specialized servers dedicated to file serving. Most have integrated storage, though NAS gateways can connect to a SAN on the back end. NAS consolidates storage at the file system level, provides shared, concurrent access, and is known for its ease of installation and use. Enterprises typically use it for file sharing and collaboration applications like CAD/CAM, software development, Web serving, and general file storage and retrieval, though some also host databases on NAS.

Most enterprises choose SAN and/or NAS for storage consolidation because of their scalability, extended connectivity, and technological maturity. Networked storage solutions of all sizes are available now, not just ones for larger IT operations, as was the case several years ago. Furthermore, SAN and NAS are complementary rather than competitive, and many enterprises deploy both to meet a broad spectrum of requirements.[3]

Storage resource management (for better utilization)

Storage resource management (SRM) software looks at data from the perspective of servers and clients and answers the question, "What data is out there and where is it?" It scans file systems and databases, tracks data by type and amount, and often correlates it to specific storage devices. SRM gives a

[1] For details, see *Fibre Channel – The Defending Champion Has Staying Power* in **The Clipper Group Explorer** dated December 14, 2001, available at http://www.clipper.com/research/TCG2001012.pdf.

[2] Small Computer Systems Interface, an older yet fast interface, with limited cabling connectivity.

[3] Application requirements are the primary determinant of the appropriate technology.

mountaintop view of data in the enterprise. This is very useful for capacity planning, asset management, quota management, and charge-back to users and departments. For instance, SRM finds inactive or duplicate data for deletion or migration to archives, which frees up space without additional storage purchases. It finds unused and forgotten capacity for redeployment. It also identifies storage arrays, volumes, and file systems that are approaching full capacity and should be expanded to avoid downtime due to an out-of-space condition. SRM works in SAN, NAS, and DAS environments.

Tiered storage (for economical data placement)

Tiered storage is another way to lower hardware costs as well as to meet business requirements for data retention and enhanced service levels. It establishes pools of storage with different service levels, in terms of performance, availability, recoverability, and cost. The objective is to deliver the right service level to the right data at the right time. If consolidation is about *economies of scale*, tiered storage is about *economies of precision*. Scale lowers per-unit costs by improving asset utilization and simplifying management. However, economies of precision are concerned with making a series of smart, optimal decisions, like an electronic carburetor that continuously adjusts the fuel and air mixture to maximize fuel efficiency in an engine. Since not all data is created equal and its value changes with time, there is an opportunity to save money by classifying data and applying different service levels, as determined by business requirements. This avoids overspending on top-tier storage – which is quite common – or under-spending and suffering performance bottlenecks and data loss. Moreover, tiered storage is the foundation for the broader concept of information lifecycle management (ILM).[4]

There are many ways to deploy tiered storage. It can mean storage systems with different media types, such as disk, tape, and optical. It can mean storage arrays with different performance characteristics and/or drive types, such as Fibre Channel for high performance and ATA for low cost. Some vendors offer arrays that support both drive types at the same time, like tiered-storage-in-a-box. It can also mean applying different data management and protection capabilities to the tiers. Smart software for centralized storage management, data classification and migration, and policy-based automation are important for tying it all together.

#2 - Containing IT staffing levels and operating costs

By far, the largest component of storage total cost of ownership is management. As data grows and the storage environment becomes more fragmented and complex, it takes increasingly more time for skilled administrators to perform backups, restores, capacity addition and reallocation, data migration, monitoring, troubleshooting, and other management tasks. Management costs rise as a result, which include staffing as well as electricity and floor space consumed by excess capacity.

Storage consolidation (for simplification)

In addition to greater scalability, consolidated storage is easier to manage because it is unified and more flexible. Each administrator can effectively handle more capacity, so staffing levels do not have to rise in proportion with data growth. Better utilization and less overhead capacity also contribute to lower environmental costs. (Refer to *Storage consolidation* in Section #1 for implementation details.)

Storage management software

While consolidation simplifies the storage itself, storage management is a tool for simplifying the management process. It amplifies the work of IT administrators by allowing them to do more with less effort. A storage infrastructure has many components that must interface and work together: files systems, volumes, host bus adapters, cables, switches, routers, storage arrays, RAID groups, LUNs, tape drives and libraries, backup servers, and, of course, the data itself. Add in heterogeneous devices, interoperability issues, a distributed environment – and it is a lot to coordinate. Using individual device managers for every piece of the puzzle can be dizzying. On the other hand, storage management software

[4] See *Top 10 Things You Should Know About Information Lifecycle Management* in **The Clipper Group Explorer** dated May 11, 2004, available at http://www.clipper.com/research/TCG2004041R.pdf.

provides a central point for monitoring, mapping, configuring, reporting, automating, and optimizing. Though product features vary, general categories include:

- **Storage resource management (for simplification)** – Manages and reports on data and storage capacity resources. (See *Storage resource management* in Section #1 for more details on this software category.)
- **SAN management** – Monitors, maps, and configures devices in a storage network, such as switches, host bus adapters, and storage arrays.
- **Provisioning** – Manages and automated the end-to-end process of provisioning capacity, from the storage array through the network to the server and application.

Virtualization

Think of virtualization as *consolidation taken to the extreme*. It applies an abstraction layer on top of physical assets and creates a single, logical entity. It delivers the same benefits as consolidation – simplicity, flexibility, and utilization – only more so. Block virtualization technologies can reside in the storage array (to virtualize itself and possibly other arrays), in network-resident platforms or intelligent switches (to virtualize anything attached to the network), or in servers (as in a volume manager). File virtualization technologies often apply a global namespace over multiple file systems, NAS platforms, and file servers. Clients see one namespace and administrators manage one namespace.

#3 - Data retention for regulatory compliance

Enterprises need data retention policies to meet legal, regulatory, and operational requirements. Regulatory compliance is not the only reason for data retention, but it gets the most publicity nowadays. Sarbanes-Oxley, USA Patriot Act, HIPPA, SEC 17a-4, and other regulations have brought it to the forefront of the corporate conscious. They require archiving data for specified periods, on penalty of serious fines, legal liability, or worse. This applies to electronic records, such as e-mail, instant messages, documents, transactions, and images. Moreover, timely search and retrieval is an important aspect of archiving, so rudimentary backup is insufficient for this purpose.

Tiered storage (as foundation for archiving)

In addition to improving storage economics, tiered storage is a foundational component of archiving. The archiving process moves data from primary to secondary storage.[5] Primary storage has the highest performance and robustness because it hosts production applications directly. Secondary storage is designed to store large amounts of infrequently-accessed data for long periods. While it does not have the performance characteristics of primary storage, it costs much less – even an order of magnitude less per unit of capacity. As secondary storage, enterprises often use tape libraries or storage arrays with low-cost ATA drives. Some data retention requirements stipulate write-once read-many (WORM) media, which is available for disk, tape, and optical. It ensures data integrity by not allowing it to be changed or deleted. (Refer to *Tiered storage* in Section #1 for implementation details.)

Automated data migration (for archiving)

Automated data migration or archiving solutions manage the data migration, retention, and retrieval process. They use descriptive information about data (i.e., metadata) to categorize and move it between storage tiers based on policy. For instance, a policy might require that files not accessed in six months move to secondary storage for long-term archiving. These solutions maintain a link or association between the application and migrated data, so users can still search and access it. Some can even delete data after a specified period, thereby managing the cradle-to-grave lifecycle. Archiving solutions today apply to specific applications, such as e-mail, databases, or file systems. A universal archiving solution that encompasses all forms of data is not (yet) available, so enterprises have to prioritize and decide which data to manage in this way.

Object storage (content-addressed storage)

Object-based storage, also called *content-addressed storage (CAS)*, is a fast-emerging category for centralized, long-term, online storage of fixed content[6]. Think of it as a storage tier optimized for archiving, where the object

[5] Though more than two tiers may be involved.

[6] Fixed content is unchanging data kept for reference purposes.

being sought is a document, form, image, etc., accessed by its content rather than its stored location (as with files in a traditional file system). Each object is associated with metadata and a unique digital identifier. This approach eliminates data duplication and creates a flat, highly-scalable address space. CAS overcomes the scalability limitations of traditional file systems and does away with the application and operating system dependencies that interfere with the long-term usability of data. CAS solutions incorporate other useful data management features such as content searches and WORM.

#4 - Backup and recovery challenges

Backup is the most common means of protecting data, though it is not without its challenges. Growing amounts of data reside on heterogeneous servers, PCs, and handheld devices that number in the tens, hundreds, or thousands. They may be distributed among remote sites around the world. Backup systems and procedures must accommodate all of this complexity. Meanwhile, backup windows are constrained and high failure rates for backup jobs are still too common.

When data corruption, a system failure, or disaster occurs, the only question that matters is how soon and how completely can the business recover? If the answer is unsatisfactory, it is better to resolve the problem sooner than risk unacceptable downtime or data loss.

SAN

A SAN is not only useful for storage consolidation, but it also facilitates backup and restore. A separate, high-speed network dedicated to storage is an excellent mechanism for these processes. It removes backup traffic from the LAN, so it does not interfere with the performance of enterprise applications. The high bandwidth of a SAN makes backup and restore operations faster and its extensibility facilitates electronic vaulting at remote sites for business continuity. A good backup system can provide centralized management and automation, so administrators spend less time with it. In short, better backup is another reason to favor SAN adoption. It may be enough to solve your enterprise's backup challenges. (Refer to *Storage consolidation* in Section #1 for implementation details.)

Disk-based backup

Another major trend is the increasing use of low-cost disk arrays as backup targets because of their advantages over tape in speed and media reliability. The high bandwidth of a disk array and its ability to read and write random, intermittent streams of data make it favorable for fast backups and restores. This helps enterprises cope with strained backup windows and increasingly stringent recoverability objectives. While both tape and disk media can fail, a RAID array inherently stores data redundantly and knows immediately when a failure occurs. With tapes, one does not necessarily know a media failure exists until the tape is needed for restore, which is not a good time to find out. It also requires physical handling of the media. However, tape still delivers the lowest cost per unit of capacity, especially when storing very large quantities of data. Therefore, many advocate a combination of disk and tape for backup. Initial backups are stored on disk for quick recovery, and tape is used for long-term and possibly remote storage.

Snapshot integration

Snapshot copies are useful for non-disruptive backup as well as fast restores from a recent point in time. A snapshot or point-in-time copy takes a "snapshot" of data on disk at an instance in time. It can be a full copy or a space-saving differential copy[7]. Making a copy is fast and, if properly integrated, non-disruptive to the application. The backup system can then use it as the source for a backup job without disrupting user access to the primary data. This means that backups may run during business hours, providing more flexibility for backup windows. Snapshots are also useful for fast recoveries from logical faults, such as data corruption or accidentally deleted files.

Centralized backup (for remote sites)

Backup and restore for remote offices is a challenge. Many enterprises delegate the task to each local site, though the staffing skill levels, equipment quality, and procedural discipline are typically not as high as in the central data center. It can result in a lot of distributed, redundant equipment, high management overhead, and inconsistent data protection. A better

[7] A differential copy keeps an index of the original data plus a log of any changes to the original since the copy was made.

solution is centralized backup, where backup data is sent electronically and automatically from the remote offices to a central data center. From there, the centralized IT staff manages it. This approach is more efficient and consistent and relieves remote offices from the backup burden.

#5 - Disaster recovery and business continuity

Enterprises run on information as much as any other asset. The continuity of business operations depend on information access; without it, activities slow or altogether stop. Moreover, information access depends on effective disaster recovery measures. Backup systems play an important role here, but they do not meet all enterprise requirements for recoverability. Complementary technologies and techniques are also important.

Define recoverability objectives[8]

Before implementing a particular solution, first define your enterprise's requirements for recoverability. Asked differently, how quickly and fully do you need to restore access to data after a failure or disruptive event? The answer may differ by application and user, so a tiered approach may be more sensible than one-size-fits-all. Two metrics characterize recoverability:

- **Recovery point objective (RPO)** – The degree of restoration. In other words, how close to current will the recovered data be?
- **Recovery time objective (RTO)** – The targeted time to restore. In other words, how long will it take to get the system back up and running?

More than just a cost to bear, disaster recovery is an investment in risk mitigation and business productivity.

Implement appropriate recovery technology

In addition to backup, one or more of these technologies may help meet your recoverability requirements:

- **Remote mirroring** – Maintains a complete physical copy at a remote site in real time. As the source data changes, so does the target. Like a spare tire, if the source data becomes unavailable for some reason, a current or nearly current copy is available to resume operations. It protects from local system failures or disasters, such as fires, floods, or electricity outages.
- **Snapshot copy** – Creates a copy of data on disk at a specific point in time. If a remote mirror is like a window with a real-time view, a snapshot copy is like a photograph that captures a scene at a fixed point in time. These copies are useful for non-disruptive backups, quick recoveries from logical faults, and data repurposing. (See *Snapshot integration* in Section #4 for implementation details.)
- **Continuous data protection (CDP)** – Recreates data sets to virtually any prior point in time. If mirroring is like a window and snapshots are like photos, then CDP is like video. It effectively snaps a series of photos for rewinding back to virtually any point. The most recent replica can approach real time, though not as current as synchronous mirroring. Today, this technology applies to a contained set of applications and file systems.
- **Data copy and migration** – Moves data between systems, including servers or storage arrays. This category is admittedly a catchall, though it has distinguishing characteristics. These solutions copy data at a point in time between systems (not within a system, like snapshots). They are useful for data distribution, disaster recovery, centralized backup, equipment upgrades, and capacity load balancing.

#6 - Insufficient application performance

Application performance is defined ultimately by the experience of the end user. The response time and throughput delivered to the person who is using the application to get work done is what matters. If performance lags, so does productivity. Have you ever heard a customer service representative say, "Sorry, the computer is slow today," while he or she tried to update an account or process a transaction? Then you can see how performance makes a difference.

There are many potential reasons for poor

[8] See *Recovery in Perspective – Ensuring Access to Enterprise Data* in **The Clipper Group Explorer** dated January 28, 2005, available at http://www.clipper.com/research/TCG2005003.pdf.

performance, such server, storage, or network issues. Find out where the bottleneck lies. If it is storage, there are ways to boost performance.

Faster storage array

This is like buying a car with more horsepower. Storage arrays on the market vary widely in performance and price, even within the three major categories: high-end "monolithic", midrange "modular", and entry level. Performance depends on attributes that include the number and type of storage processors, amount of cache, RAID type, number of host ports, internal bandwidth and architecture, number and type of disk drives, and the workload profile. If the bottleneck is due to frequent access of a relatively small amount of data, like database tables and indices, consider locking it into cache or placing it on solid-state disk.

Faster connectivity

If the bottleneck is the connectivity between servers and storage, additional connections, faster technology (e.g., 4 GB Fibre Channel), and fewer hops or inter-switch links may solve the problem.

Automated data migration (for better performance)

Databases and file systems are like cargo ships in that they slow and become less stable when overloaded (i.e., too large). Archiving is not only useful for regulatory compliance, but pruning records or files and moving them to a secondary tier also improves performance – perhaps enough to avoid a server or storage upgrade. A streamlined database or file system also speeds up backups, restores, replication, and data migration. (See *Automated data migration* in Section #3 for more details.)

Storage as a Service to the Business

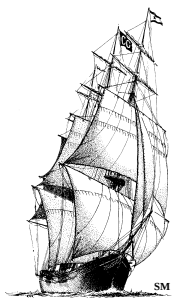
If you take a step back from tactical problem solving, a broader picture of storage evolution emerges. The culmination of the strategies and technologies described above is a shift in how enterprises perceive and manage storage: **More than a box, storage is a service delivered to the business.** This is like the difference between a power plant and a home electrical outlet, or a cellular network and a wireless phone plan. As consumers, we appreciate and experience the service delivered to our doorstep. The infrastructure that provides it, including all of the points of manufacturing and distribution, are behind the scenes and taken for granted. We are not so concerned about the infrastructure's technical specifications, per se, but do notice if something goes wrong along the way and disrupts our service. Enterprises increasingly are seeing storage in a similar way – from the perspective of the applications, processes, and end users that utilize it. What matters is the quality of service delivered.

- ***Is it what the business requires?*** Does it get the data where it needs to be in a timely manner and with appropriate protection and security?
- ***Is it affordable?*** Does it fall within the enterprise's procurement and budgetary constraints?
- ***Is it flexible?*** Does it readily adapt to changing business requirements?

Ideally, the business decides what service levels it needs and the IT department delivers to the specification. The contract or "meeting of the minds" between business and IT is called a *service level agreement* or *SLA*. Making the mental and operational shift toward treating storage as a service will help enterprises more effectively deliver and responsibly consume it.

Conclusion

Choose a solution that addresses a pressing challenge of the day and that builds a solid foundation for tomorrow – by more effectively delivering the right storage services to the business. In other words, it should be both tactical and strategic. The solution should relieve a felt pain, which is always a good thing. Nevertheless, it should also bring the infrastructure one more step in alignment with the precise needs of the business – storage as a service. It is not just a box, anymore!



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Michael Fisch **is Director of Storage and Networking for The Clipper Group.** He brings over nine years of experience in the computer industry working in sales, market analysis and positioning, and engineering. Mr. Fisch worked at EMC Corporation as a marketing program manager focused on service providers and as a competitive market analyst. Before that, he worked in international channel development, manufacturing, and technical support at Extended Systems, Inc. Mr. Fisch earned an MBA from Babson College and a Bachelor's degree in electrical engineering from the University of Idaho.

- ***Reach Michael Fisch via e-mail at mike.fisch@clipper.com or at 781-235-0085 Ext. 211. (Please dial "211" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.