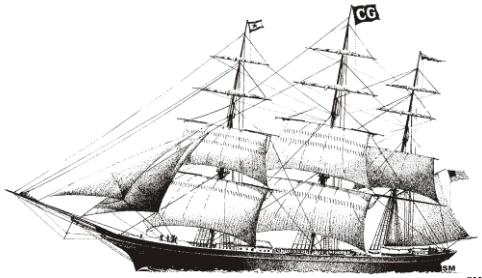
Clipper Notes™
VENDOR-AGNOSTIC EXPLANATIONS AND ADVICE FOR THE INFORMATION TECHNOLOGY BUYER



Navigating Information Technology Horizons℠

Published Since 2007 | Report #TCG2012005 | February 29, 2012

Regaining Control Over Hoards of Enterprise Data — Why You Need an Archive

Analyst: David Reine

Management Summary

For most of us who own our own home, the extra room in our attic or in our basement is sufficient to store the last seven years of our income taxes, as required by regulation, childhood souvenirs, and the clothes that no longer fit (yet!) or those that have gone out of style but are sure to return (like that leisure suit or wide tie, or both). In fact, just in case, we may make a copy of the tax returns and store the original in one place and the copy in another. We have plenty of room. For others, however, the attic and the basement may not exist or long have been filled with not only important papers but also collections of largely useless stuff being held for a rainy day or future need or maybe just enough time to sort through it all. Some folks cannot bear to part with anything that *might have a future value*. Some might be called *hoarders*, as they cannot bring themselves to throw away anything! Unfortunately, this problem is not limited to your neighbor, cousin, or someone you know whose "stuff" has overflowed into a rented storage space. Hoarding can be addictive and expensive. While this often is the stuff of reality TV shows, **hoarding can be an even greater problem for businesses. In this case, we are talking about hoarding data, whether eminently rational or poorly conceived.**

Certainly, this scenario is nothing new to the CIO or information manager at any enterprise or smaller data center. The gathering and preservation of data is out-of-control. For many, the sum of all data being stored is doubling every 18 to 24 months, with no end in sight. How much of this is due to hoarding, either by good intention or default policy? That is a multi-million dollar question in larger enterprises. They might not speak of it as hoarding. Rather it might be called data protection and retention and it might be driven by good policy and automation. *What is the situation at your place of work? Is data being kept (possibly in many identical or near-identical versions) because it is easier to keep it than decide what to do with it?* That probably qualifies as hoarding and it gets worse.

Imagine if you put "piles of stuff" into boxes by the date that they were gathered for storage. Imagine how hard it likely will be to retrieve what you want from a collection of hundreds or thousands of boxes. First, you have to find it. Even if you do a good job of indexing what is in each box (which often is not the case), you still need to retrieve the box and get the desired tidbit to the requestor. *Sound doable?* Maybe, just maybe, but probably not the answer that any user wants to hear. And it gets worse. In the era of the *light speed* of the Internet, *how long might this take? How many indices and contents of boxes will have to be searched before the "right stuff" is found? Will it be done in a meaningful time period?* Probably not. *Will doing this inadequately be cheap?* Again, probably not.

There is a three-word moral to this story – *Hoarding Isn't Archiving!* Even very efficient

IN THIS ISSUE

hoarding of data into a searchable facility with good retrieval isn't archiving. It may be *backup* (to be discussed), but it is not *archiving*.

Archiving is done with a very specific set of intentions and requirements. While most of the archived data may not be retrieved very often, that which is needed usually has a true business value that is time dependent. This paper is about *enterprise archiving*, especially for those businesses that have not deemed archiving to be important enough to do, for whatever reason. Read on to explore the multiple dimensions of *archiving*, *backup/recovery*, *storage*, and *retrieval*, from both technological and economic perspectives.

If Hoarding Isn't Archiving, What Is It?

Once you get beyond the serious hoarding problems often highlighted on reality TV shows, there is something serious to discuss about hoarding within IT organizations. **One "hoards" (i.e., "keeps") something because it might be useful in the future.** We do this all of the time as part of our standard IT practices. **Many of us live in fear that what we value most highly (data of all sorts) will disappear due to hardware failure, software failure, power failure, natural disaster, malicious act, or human error.** If we are talking about a couple thousand digital photos or videos of our family, our concern is very personal and real. We all have deleted something that we wish we had protected better. Multiply that by many, many orders of magnitude and you have an enterprise-sized problem that might affect the ability of an enterprise to continue to operate.

That is why the first priority of any enterprise data center is <u>to protect</u> the enterprise's data. Traditionally, this has been done with backup solutions and multisite locations. The highest priority is given to mission-critical data (needed to be able to operate), followed by business-critical data (can tolerate a small delay in accessing without becoming operationally crippled), followed by less critical data (whose loss might not be noticed). Most archived data, when retrieved, is business-critical, i.e., it needs to available in a reasonable period of time because someone or some process needs (and values) the data.

Data isn't the only thing that is important. The system images (virtual and real), partitions, operating environments, middleware, and applications that are running on an enterprise's many servers also are very important.

Traditional Backup Can Become Hoarding

Traditional backup objectives focus on restoring data and operating environments – after a failure of some sort. The goal is to return to a "working whole" as quickly as possible. This is noble and important but not always economically efficient. **We tend to make multiple copies of everything** from system images, to databases (called snapshots), to email collections, and to data files. This might seem reasonable, **but can approach a level akin to hoarding.**

It is easy to become obsessed with backing up everything on a repeated basis. **The operational motto may be something like "better safe than sorry", or restated, "you can never save too much or too many times".** This sounds good until it comes to managing the hoards of data that have been collected and saved, regardless of choice of media or location. None of this is inexpensive and with data doubling every year or two or three, it only gets more expensive to be a hoarder.

Nonetheless, if you need to be able to recover from a loss as simple as deleting the latest version of a document in preparation for your boss, or as complex as failing over to a remote data center following a true catastrophe, then one of the many backup processes should serve you well (ignoring the costs, of course). You must be prepared. In reality, most backups are done on an amalgamated near-physical basis when logically grouped data on physical disks is backed up in the aggregate. You either restore the whole thing, or look for the pieces needed in the big-blocks of data that have been backed up. These blocks (also called "LUNs"[1]) are like the cartons of stuff that may be stored in your basement or attic. It is easy to store a lot of cartons, with the assumption that what you want can be found in your hoard, *given enough time*. **Thus, there is a risk that backup can become a lot like hoarding, if you don't pay close enough attention to *what, where, why,* and *at what cost* you are "saving something that you might need later". Backup may be akin to hoarding, but neither defines what archiving is.**

Archiving Explored

So what, then, is archiving?

If backup is for that rainy-day need, ar-

[1] LUNs are the short technical name used for "Logical Units" of storage on disks.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

chiving is about data that might need to be used operationally in short order.[2] This difference is sharp. **Archiving is about managing access to working collections of data.** Most commonly, these are files, like documents, spreadsheets, music, digital images, videos, and much more. Sometimes, these are called "business objects".

What creates the need to archive?

It is that they are valued as a collection and as individual pieces within that collection. They need to be retrieved through a meaningful business process. Examples might be a collection of TV programs or video clips, financial reports, tax returns, seismic soundings, medical images, insurance forms, canceled check images, etc. Each of these has a business-driven useful life. Some may be retained for months or years and others may be valuable forever (like images of an ever-changing universe from stellar telescopes).

If these are medical or police records or images, how much time do you really have to get your hands on what you seek? **You don't want to be digging through your virtual boxes to find the right object.** You want it in seconds (or minutes or hours or days), depending on your business requirements.

What is Enterprise Archiving?

If *archiving* is about the orderly storage and retrieval of business objects with time-dependent business value, one might say that doing it at a very large scale (i.e., with large-to-massive quantities, either in number of objects or the digital size of objects, or both) is *enterprise archiving*. Of course, because of the very large scale, enterprise archiving needs to be done in the most economical and rational way, based on the kind of data, the business value, the time dependencies, the likelihood of use, and the available budget. *Thus, hoarding of data that you need to sort through – when you might need it – doesn't cut it.* **Enterprise archiving is orderly, business focused, and cost sensitive (due to the unusually large number of objects and volume of data and long periods of retention).**

It is important to note that archiving is not defined by the kinds of media that may be involved, whether one or many, nor by the application or system software that might deliver it. **Enterprise archiving is defined by the requirements that must be satisfied, both operational and policy-specific. This almost always makes archiving different from backup, because the requirements and policies are not the same.**

It also is important to remember that there is a difference between *just keeping the data* and *keeping it as accessible information for the long-term*. If you retain data for a sufficiently long time, the people and/or the applications that created it may have moved on or even disappeared. In the case of people, they may have retired or just changed jobs. In the case of applications, you may need a version of the application that is no longer supported by a current O/S or even might not be available. There may be no one available who knows what the data means or how to use it. It is important to archive this data with the metadata about how the data was acquired, what it means, how to use it, and how to restore it. You should not have to be a data archeologist in order to retrieve valuable information.

Now, before you tell me that you already do archive your files and databases with your backup software, let me tell you that you do not! Let's explore this further.

The Six Questions of Archiving

Six questions will frame most archiving requirements and help determine the nature of the solution (or solutions) that may be required. These are *what*, *who*, *when*, *where*, *how*, and *why*.

What?

What tends to be the most straightforward of the questions, but by no means does that mean that it is simple. This is about the data to be archived. It is about defining the *scope* of the collection (including what's in it, the nature of the data, the size (or range of sizes) of the data objects, and the rate of object growth), the related *life-cycle characteristics* (such as retention policies and retrieval requirements), and the business value of the collection, now and as it ages.

Who?

There can be many answers to the *who* question. First, whose data is it? That might not be so simple. Often the people who are responsible for gathering the data are different from the people who will use it in the short term

[2] But not necessarily instantaneously (i.e., in less than a second).

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

(days, weeks, or even months). Someone else might be the custodian. How this plays out over time (and who pays for it over the long term of years or decades) adds further to the people involved. From all of these perspectives comes the most important question: *Who is responsible for making the data management and ROI decisions for the collection?*

Where?

Where might be very physical (on a specific storage device or devices) or very locational (at which data center) or very distributed (in many places) or very ethereal (i.e., "somewhere in the cloud"). Exactly where has cost implications, especially in terms of the underlying infrastructure.

When?

When contains two issues, one pertaining to its lifecycle and the other pertaining to its anticipated or expected use(s). These two usually are intertwined. The lifecycle often is tied to the expected use cycle (or mandated retention period). Some data is held because it has to be (like email), just in case it is needed (for litigation). In reality, the vast majority of it may never be accessed again. This is different from data that is subject to operational, business, or economic cycles.

An example of an operational cycle might be medical images. If a follow-up visit to a physician is scheduled for three months after the last visit, then that is an example of an operational cycle. Another might be the publication of quarterly reports for most any business.

Business cycles tend be me more calendar based, although some cycles might be based on hours or minutes. If auditing is done quarterly, that is an example of a business cycle. If the Thanksgiving-to-Christmas season is your period of peak activity, then you may have a once-a-year cycle. If use of archived data can be described as "time dependent" (as in a hot news clip), its reuse life cycle might be measured in minutes, hours, or days.

Sometimes, use of archived data is driven by economic cycles. For example, if the price of oil were to rise to $200 a barrel, certain sources that had been determined to be too expensive to extract when oil was $100 a barrel may now be economically feasible. That's when you may want to look at the seismic data collected years ago.

How?

How is an IT question, primarily, and mostly not one of how the data is to be used. Many of the how questions will be addressed in following sections.

Why?

Primarily, *why* is a question about potentially many sets of rationales used to explain the needs (who/where/when) and the decision on exactly what to do (how). It is important to determine why (at a given point in time, like when setting up an archived collection) and to document this well. The answers to the six questions may change and the *how* and *why* decisions may need to be revisited, both as business requirements change and information technologies evolve. Too often, the folks picking up the pieces never really understand why they have what they have.

Characteristics of *Backup*

From an historical point of view, the IT industry has been doing backups of business data for over five decades. The backup consists of one or more *copies* of the data in order to recover from a data loss or a disaster. This is *additional* data that must be managed by the administrative staff. Originally targeted at tape, backup was performed to protect the enterprise from the accidental or malicious loss of any data over the short term, i.e., typically up to 120 days. Incremental backups were done daily, full backups done weekly, typically to three sets of tapes: the grandfather, the father, and the son, which were overwritten in rotation, to protect the data center from possible backup errors or disasters at the primary site,. Today, because of the amount of data that is typically backed up at any one time, and because of the urgency of immediate recovery, backups are frequently done D2D, or disk-to-disk, in a two-tiered architecture, such as Fibre Channel (FC) disks for primary data and SATA disks for secondary, with D2D2T[3] for a three-tier best practice data protection, with tape data being offline, protecting it from corruption that can occur with online data. Backup is an extremely time-critical function; the backup window has a finite size and the recovery time objective (RTO) is usually very critical, as is the recovery point objective (RPO), i.e., typically a matter of hours. These objectives do not apply as much to archiving as

[3] Disk-to-Disk-to-Tape.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

the data is not being copied, it is being moved, and it is mostly inactive data.

Unfortunately, backup applications often copy duplicate objects as well as unique data. As email has become the preferred means of communication in a mobile workforce, more and more duplicate attachments get replicated to secondary media, over and over and over again. Data deduplication has taken a major bite out of this waste of resources, but it still backs up terabytes of old, rarely used, and obsolete data. Collaboration is another recent phenomenon that has added significantly to the amount of data being backed up unnecessarily.

Characteristics of Good Archiving

An archive is not a backup, and must never be used as such. An archive, sometimes called an *active archive*, is a data store that:

- Is easy to use,
- Receives unchanging data (or, possibly, harvests or migrates it off of the production environment according to enterprise priorities),
- Creates index entries for each item stored for content-based retrieval, and
- Becomes the primary data source from that point forward.

All of this is done to manage the growth of data, to free up critical high-cost storage space, and to protect the record from change. It reduces the TCO of the IT infrastructure and improves the performance of the operating environment.

An archive can store and retrieve individual records for future reference, improve search time with sophisticated search parameters, and assure legal compliance. Archives retain information for a specified retention period, which can be "forever". The data is kept in its original form and cannot be edited or overwritten, in order to meet regulatory requirements, although eventually, it can be deleted, all according to enterprise long-term data retention objectives. One method to ensure the integrity of your historical enterprise data is by applying WORM technology[4] to your archive. The security of your data can be protected by encryption. *How much of your data is duplicate or has not been accessed in over 90 days?*

Archives can be used for a myriad of services, from consolidation to space reclamation, with simplified maintenance and management, as well. They can also be used for analytics and the long-term preservation of historical data. Many kinds of data tend to be unique, such as medical images, and typically are not reducible by data deduplication. What it should not be is actively managed data. Archives consist of data that intentionally has been moved off of high-cost, high-performance storage and onto purpose-built appliances or perhaps, back-to-the-future with tape[5]. Today, high-capacity appliances and tape libraries provide more than enough performance for archiving needs, and that includes eDiscovery and the indexing and searching through terabytes and petabytes of historical data for defending the enterprise against litigation. These appliances may keep recently created or recently used data on disks (as a cache to speed up access or storage). This usually is done transparently. Some solutions are application specific (like email archiving). Others are general-purpose solutions to be applied against user-defined collections of data.

Archives can be deployed in multi-tiered solutions to take advantage of the speed and functionality of SSD, FC, SAS, and the capacity of SATA and Tape in a D2D2T environment. Live data is usually deployed to disk, rarely used data to an appliance or library (disk or tape) or perhaps to the Cloud[6], with generally obsolete data archived to a removable, off-line storage.

An archive enables full text indexing, includes a records management service in order to be able to index for search simplification, and a complete policy management system. It needs to manage a complete metadata stored in an online database by policy covering such fields as subject line, content, keywords, who sent it, who got it, who was copied, record date, and retention date. It must be scalable to meet the existing data needs of your mission- and business-critical applications, as well as the growth

[4] WORM=Write Once Read Many. Certified WORM solutions meet the requirements to manage carefully the retention of data unchanged and who has access to (and accesses) what is stored.

[5] See the issue of Clipper Notes dated December 20, 2010, entitled *In Search of the Long-Term Archiving Solution – Tape Delivers Significant TCO Advantages over Disk*, and available at http://www.clipper.com/research/TCG2010054.pdf.

[6] See the issue of Clipper Notes dated January 25, 2012, entitled *Moving to a Private Cloud? Infrastructure Really Matters!*, and available at http://www.clipper.com/research/TCG2012001.pdf.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

requirements for the future.

Archives are available as bundled solutions or the data center can acquire software applications that can be added to existing configurations to complete an archive stack. The path the IT staff chooses may be different for each enterprise. It is all a matter of TCO for hardware, software, integration, testing, and services.

Compounding Economic Challenges

If archiving and backup could be addressed in isolation from everything else that is going on within the data center and the enterprise, that would be nice … unrealistic, but nice. **However, today's data centers, both enterprise-class and smaller, are faced with a serious problem: *trying to do more with less.* This economic stress is a broad, never-ending challenge that encompasses mission-critical, business-critical, and archived data.**

The Stresses and Demands of Server Virtualization

Many data centers have faced up to the issue of poor server utilization, and laid out a plan to virtualize and consolidate application processing, in order lower the TCO of IT resources. The typical enterprise now has reduced the number of servers throughout their organization and they have improved the efficiency of each by deploying upwards of ten mission- and business-critical applications on each new, multi-core, multi-threaded virtualized server. In fact, the IT staff may have succeeded so well in consolidating servers that they have been able to reduce the server footprint and the number of administrators required to manage the server network.

Unfortunately, unconstrained growth in data along with the inherent nature of virtualized servers has increased the workload (i.e., the real-time demands) on the shared storage environment, and the storage management challenges, as well. **Thus, now is the time to improve the utilization of enterprise storage resources, while addressing the needs for archiving.**

The Stresses and Demands of So Much Data

With hundreds of millions of users surfing the Internet and billions of text messages and images being sent daily, the amount of digital information is exploding with massive growth. The data center needs to maximize the utilization of existing storage resources and grow with the highest density media possible, whether disk or tape. **In fact, the data center needs to virtualize and simplify the use of storage to the point where it does not matter to the user or to the applications where the data resides. However, all potentially valuable data must be preserved and protected**.

In order to cope with this growing problem and protect the enterprise, data centers typically keep multiple copies of all data, both structured and unstructured, in order to guard against a possible disaster, which seems to be more likely than ever, occurring with an amazing frequency of late: earthquakes, hurricanes, and tsunamis, to name a few, not to mention the continued threat of terrorism.

Today, most, if not all enterprises and smaller businesses have some form of backup and recovery deployed throughout their IT infrastructure, creating more and more copies of data, regardless of the frequency that that data is changed or even interrogated, necessitating more and more storage.[7] These backups continue to grow daily, adding to the TCO of the IT infrastructure, despite the fact that disk drives and tape cartridges continue to increase in capacity while their cost remains flat. Why, you might be asking yourself. **Because the data must be protected!**

The Stresses and Demands from Constrained IT Budgets

Finding space to store all of this enterprise data is a problem, but it is not the most significant problem. *Finding efficient and economical ways to store it is!* The capacity of a single disk goes up every couple of years with 1TB disks being replaced by 2TB disks, and now we see the arrival of 3TB disks making their way into enterprise configurations. Unfortunately, the total cost of ownership (TCO) of a disk array continues to go up as a result of any number of factors, the most recent of which is a manufacturing crisis caused by flooding in Thailand, causing panic buying of disk drives around the world.

Because of the continuing rapid and uncontrolled growth of the storage environment, most data centers need to find a way to change the

[7] This ignores the extra storage needed for RAID, which consumes several disks in each RAID group (typically of 8-16 drives) for protecting the group from a failure of a disk or two, which happens more often that you might expect. It also ignores the often-significant "headroom" that often exists for allocated but unused disk space.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

storage paradigm for both structured and unstructured data, while keeping the TCO of the IT infrastructure as low as possible. **The data center needs to implement a virtualized, tiered storage environment for all types of data, where only the most critical data resides on the most costly resources, and rarely used data lives on – perhaps forever – in an archive on the most cost-efficient media, while the rest of the enterprise's data sits somewhere in-between.**

The TCO of the entire IT infrastructure also is rising precipitously because of the poor utilization of IT resources, government regulations, and, simply, the fears of prosecution or finger pointing for not saving and protecting the "important" data. **As a result, many "C"-level executives have decided to save everything – in fact, to hoard all of their data.**

It is no wonder that the enterprise IT budget is drowning in red ink. This should just be a warning; even though the Thai factories are coming back online, who knows when the next natural disaster will strike and what the impact on your budgets will be. **It is imperative that the data center take control of overwhelming and uncontrolled storage growth and, thus, to slow down the forced acquisition of more and more disk arrays and disk drives.[8] Archiving some (or much) of your data is one surefire way to change the storage paradigm and to achieve that goal.**

The Need to Re-Examine Storage Practices

One way to get control of the enterprise storage environment is to reexamine the way that data is being stored within the existing enterprise storage infrastructure. If the IT staff can reclaim storage by removing rarely or never used data from costly high-performance disk drives where it is being hoarded, they can rededicate that space to active, mission-critical data, improving the performance of the involved applications and lowering the TCO. By migrating, or archiving, older or obsolete data to less costly, high-capacity disks or, an even more efficient removable media, such as tape, the IT staff can consolidate its most active data on fewer high cost, Solid-State Disks (SSDs) and Fibre Channel disk arrays, lowering the storage footprint, reducing the number of administrative staff required, and helping to "green" the data center environment by consuming significantly less energy. Add to this the terabytes (or, perhaps, petabytes) of disk storage that can be released back to the virtual disk pool and you may be talking about serious economic savings.

While you are thinking about how much you might save or redirect, here are a few questions for you to ponder.

- *What percentage of your data is obsolete?*
- *What percentage is rarely (or never) used?*
- *How much of your data is duplicate, i.e., multiple, identical images that are being backed up repeatedly?*

What About Data Deduplication?

Data Deduplication is a relatively new method, primarily being used to reduce the amount of data that needs to be backed up, especially in the area of email, where multiple copies of exactly the same attachment are sent throughout your enterprise every day. However, for a lot of your data, deduplication is not sufficient. In many cases, your data is unique, for example, medical records and images such as X-rays and MRIs, geo-physical data for oil exploration, videos, photos, and scanned documents. Retaining this information may be critical to your enterprise in terms of meeting regulatory compliance or protecting the "C" level executives from possible litigation losses, however, **keeping it on the most expensive storage resources may not be the right answer**, if the IT staff needs to reduce the amount of storage being retained in the highest cost storage in order to lower their overall TCO for storing data.

Accordingly, the IT staff may need to implement an automated process to (meaningfully) archive old, but important, enterprise data, thus removing it from high-performance disk arrays and possibly entirely from disk storage. This data may not be needed today, but in years to come, it could become invaluable, like that seismic data for an oil field deemed impractical at $100 a barrel.

Some of you may think that you do this kind of data movement today with your backup software. **However, backup and archiving are two distinctly different processes!**

Back to Backup Versus Archiving

***Backup* enables the IT staff to protect enterprise data by making timely copies and**

[8] Sending data off to "the cloud" just passes the problem along. In the end, you still are paying for it.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

preserving that information in the event of a data center disaster or something less catastrophic but with critical impact. This means not only a secondary copy of data, but perhaps a third or fourth copy as well, for those of us who were brought up on a strategy of keeping a grandfather, father, and son backup set. If you adhere to this method, you will be propagating the original and three copies, as well, not to mention those who create multiple backups and snapshots, and refuse to delete anything. This is hardly a space-saving or economical move for a long-term preservation strategy. The recovery process will restore your enterprise to full operation without too much delay. However, backup does not reduce the amount of data that you are storing; *it increases it*.

On the other hand, ***archiving*** **stores the *original* data, migrated (eventually) to the least costly tier of storage that is acceptable, usually within a multi-tiered environment, making a vital enterprise resource easier to manage and less costly to maintain.** Archiving does not make multiple copies of your data.[9] It frees up your more costly "upper-tier" disk storage that had been dedicated to seldom read (or never read) files or data. Archiving also will improve the performance of your entire IT infrastructure by speeding up access to mission- and business-critical data that is so vital to the success of the enterprise.

In this way, the data center can create a multi-tiered architecture supporting high-cost solid state disks (SSDs) at the highest tier, for critical tables that your applications need immediately, down thru Fibre Channel (FC) and SAS disks needed for mission-critical data, through high-capacity SATA drives for business-critical or high-performance analytics, down to high-capacity or commodity LTO tape with a 30-year life, for long-term data storage[10].

Active archiving does precisely this. It is an intelligent process for moving – from a primary storage space to a secondary space – a precise set of unchanging, inactive, or infrequently-accessed data items, which still have business value and need to restore in a timely, although not necessarily instantaneous manner. Active archiving provides the ability to efficiently search and retrieve this collection, usually through metadata culled when it is stored. However, it the raw data is searchable, that capability may also be available.

Conclusion

Archiving is *not* about backing up constantly changing database information. It is about the long-term preservation of valuable yet static data, images, email, etc., that need to be stored and retrieved in an unchanged state. It is about preserving critical IT resources (and budget), especially for storage asset, hopefully with a very positive net influence on performance and energy. It is about removing the effects of a hoarding mentality from the enterprise data center and replacing it with a set of managed disciplines.

Backup (and deduplication) and archiving are complimentary processes that can be used to protect the enterprise. They are not competing strategies. Together, they can reduce the costs of storage and protection, while improving storage and retrieval processes. This can be done at both enterprise and SMB data centers, alike, with a primary goal to lower the TCO of the IT infrastructure. However, the more archivable data that you have, the more there is to save! **Now that you understand why archiving is important to the enterprise and to the data center, take the time to consider whether you should be archiving, instead of hoarding!**

[9] This does not mean that archives can go unprotected. They too have to be replicated or backed up, depending on the speed required for recovery.

[10] Most customers searching for 10-year-old data can accept a small time delay in accessing their query results, especially given the high-speed performance of today's archiving hardware.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.

About The Clipper Group, Inc.

The Clipper Group, Inc., now in its twentieth year, is an independent publishing and 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

David Reine **is a Senior Contributing Analyst for The Clipper Group.** Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. In 2002, he joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

- ***Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123. (Please dial "123" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosures

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.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

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.

Copyright © 2012 by The Clipper Group, Inc. Reproduction prohibited without advance written permission. All rights reserved.