

White Paper

The Business Value of Intelligent Tiered Storage

**Hitachi Data Systems' Integrated Block
and File Tiering**

By Mark Peters

September, 2009

Contents

Introduction: Tiered Storage Moves from Interesting to Imperative	3
The Essence of Tiering	3
Tiering – Concepts, Value, Challenges, Needs.....	3
Today’s [Dynamic] Hierarchy.....	5
The Business Value	6
Practical Approaches - Hitachi Offerings.....	8
Hardware Platforms	8
Tiering Enablement	9
Advanced Functionalities	9
Bringing it all Together	10
The Bigger Truth	10

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188. This ESG White Paper was developed with the assistance and funding of Hitachi Data Systems. A number of sections of this paper have been adapted and updated from an existing ESG White Paper entitled *Intelligent Tiered Storage*, January 2009.

Introduction: Tiered Storage Moves from Interesting to Imperative

Tiered storage is one of those things that is a means to an end—and not an end in itself. The reason tiered—not just tiered, but *effectively and efficiently* tiered—storage is becoming more crucial (indeed, a prerequisite for many organizations) is a simple matter of economics. After all, if storage were free, then everything would logically be stored on the fastest devices available. Capacity would be unlimited and cost no issue. Yes, there are some practical issues around connectivity and access, but you get the point: the only reason we have long sought the nirvana of an effective storage hierarchy is that storage isn't free and consequently, we need to make choices and allocate data to appropriate levels and types (i.e., costs) of storage.

Now, with the rate of storage growth outstripping the rate of its price decline, tiering and storage economics are back on the agenda—still not because they are inherently desired, but because of what they can deliver: namely, lower costs, better business outcomes, and a pragmatic route to cope with massive storage growth. These are the “economics of necessity.” Of course, once good tiers and tools that make effective tiering feasible (and the economics compelling) exist, then users also start looking at the “economics of opportunity”—in other words, if tiering is something that can be *embraced rather than avoided*, then additional value can be sought and overall efficiency can be increased. Tiered storage can move from something that users do because they *have* to (as little as they can get away with) to something they do because they *want* to.

The Essence of Tiering

The basics of what users want, and why they want it, are really not that difficult. But delivery of a workable solution has proven to be much harder. The “Storage Hierarchy” has been discussed for years and is simply described as putting the right data on the right device at the right time both for data type and for that data's lifecycle requirements. Thus, data moves from one level to another as needed. And it's the “movement” and “as needed” parts of that equation that have proven toughest to provide. Indeed, what we've been searching for is not a Storage Hierarchy per se, as that's always existed. Instead, what's been needed is a *Dynamic Storage Hierarchy*.

Hitachi Data Systems¹ has made significant strides towards addressing this issue and is currently delivering a genuine dynamic tiering capability. Not only that, but it is now able to address both block and file tiering. Doing both is rare enough across the vendor community, doing it in an integrated fashion—as Hitachi can—makes for an extremely valuable proposition. It delivers on the long-promised notions of HSM and ILM.

This paper covers—with a pragmatic and succinct business focus—the generic values of, and approaches to, storage tiering and includes an analysis of what Hitachi can deliver to the increasing number of data centers that want a unified approach to their data whether it is being stored or searched and whether it is files or blocks.

Tiering – Concepts, Value, Challenges, Needs

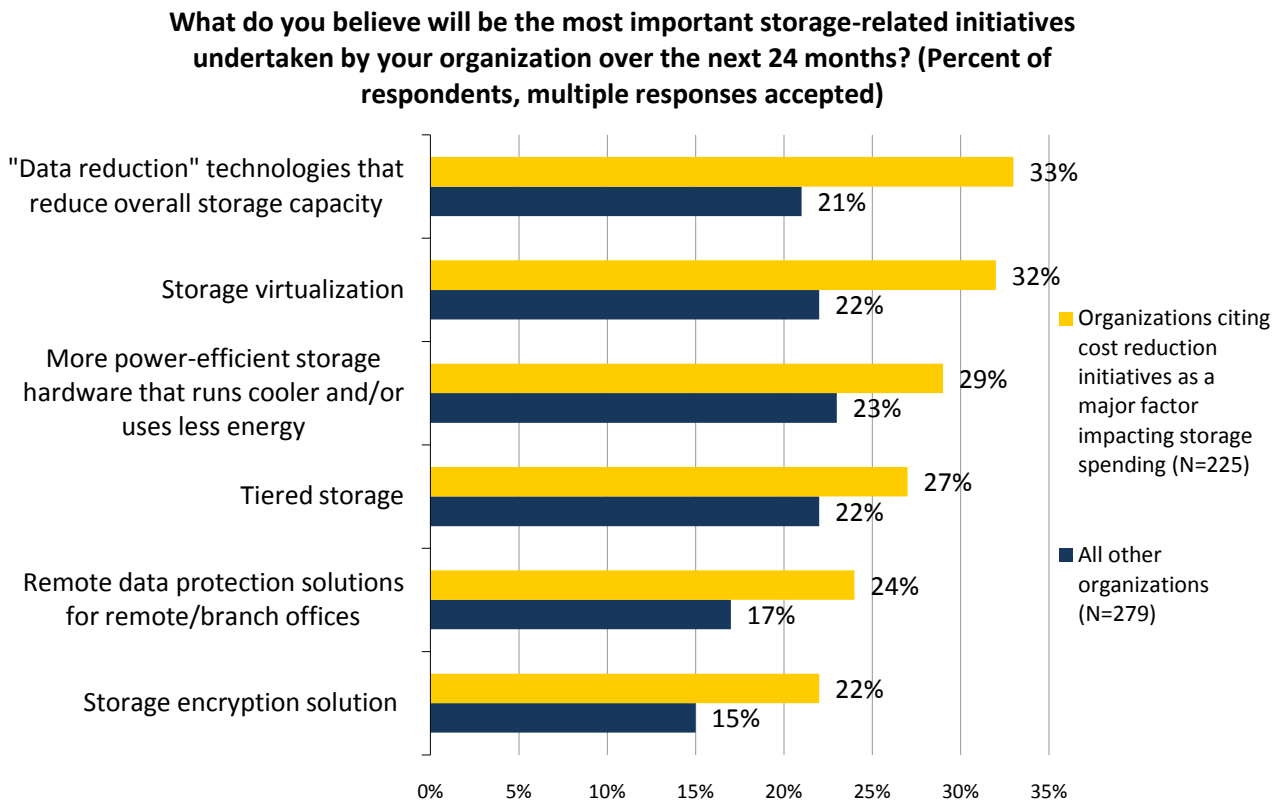
Given that storage tiering is clearly eminently sensible, it might seem surprising that it is not yet common. However, having a sensible concept is a long way from having either sufficient choices and/or capabilities to actually achieve it. One of the major challenges has been that of figuring what data and applications are deserving of what storage repository. Data classification² is essential when building an intelligent tiered storage environment. But who has the time or the tools to actually perform data classification? Most organizations use applications as proxies, but even then it is a mistake to assume that all the data created by a mission-critical application remains mission-critical throughout its lifecycle. For some applications, such as billing, the data's value almost certainly diminishes with time, although it could regain importance later, such as at year end. Just a couple of years ago, SNIA (the Storage Networking Industry Association) reported that 68% of all data is dormant 90 days after its creation. To put this in very plain language, *at least 68% of the data that resides on your most expensive storage and is protected by your most stringent policies and processes is most probably not being used by anyone!*

¹ This paper focuses throughout on Hitachi Data Systems – for readability, the company name is sometimes shown as ‘Hitachi.’

² Qualifying/segmenting data by importance, application, size, backup and replication needs as well as by specific storage and service needs

Knowing this, most users see the value in tiering; recent research by ESG into the topic (see Figure 1) shows tiered storage as a significant initiative for enterprise organizations, ranking highly in general and receiving an even higher response when asked of organizations that had also cited cost reduction as a major factor impacting their storage spending. What is particularly interesting is that tiered storage can also have a positive impact on some other initiatives—for instance, well managed storage tiers will almost certainly place more data on more power-efficient storage and will help ‘data reduction’ through deduplication, content-aware compression, and consolidation.

Figure 1. Top Storage-Related Initiatives by Enterprise Organizations Citing Cost Reduction as a Major Factor Impacting Storage Spending



Source: *Enterprise Storage Priorities Emphasize Information and Infrastructure Efficiency, January 2009.*

If the drive towards tiering is starting, there are still limitations which can reduce the positive impact. No one wants restricted choices and manual intervention; being required to monitor, manage, and migrate data between and within multiple discrete islands not only misses the point, but reduces the potential value. Thus, for example, being able to have some tiering for block-based data, but not for file-based data, is sub-optimal, especially in a world where unstructured data is growing rapidly.

While it's very easy to start with the obvious fact that keeping inactive data on Tier 1 storage is both wasteful and unnecessarily expensive, there is far more to look for in an ideal and advanced tiering implementation:

- Data comes as blocks and files – a tiering solution should be able to handle both
- Maximized ‘intelligence’ is crucial – this means dynamic, policy-based data movement and management
- Supporting a broad range of –preferably heterogeneous—storage platforms and applications (e.g., VMware) keeps the eventual solution flexible and granular
- Sophisticated tools – these should include not only management, but also search capabilities across all tiers

A quick reminder of the storage tiers available today follows.

Today's [Dynamic] Hierarchy

Virtually all data centers employ some form of storage tiering, albeit invariably manually and often by default rather than design. Nonetheless, the idea of a range of devices—with trade-offs for capacity, performance, and price—is second-nature to IT managers. Making these options work fully to their advantage is what *intelligent, dynamic, and automated* tiering is all about. The basic choices are well known, so will only be mentioned briefly here:

Solid State Storage: typically flash-based SSDs today, but the choice is expanding fast into appliances, extended cache devices, and so on. Although solid state looks set to eventually command a large share of the storage market, for the next few years, its use will be focused on serving an increasing proportion of fast and/or high IO demands in combination with high capacity disks for less active duty storage.

Disk Drive Systems: although there are both enterprise and midrange versions, the practical operational differences are gradually coming down to sheer scale rather than function. The choice of disk drives—both protocol and capacity—also covers a broad spectrum, from lower capacity fast Fibre Channel drives to much higher capacity (2 TB are now shipping) slower SATA drives. The arrival of SAS drives—thankfully, more a combination than a compromise of the other two—looks to set a new standard.

Tape & Virtual Tape Libraries: to quote the old adage, rumors of tape's demise have been grossly exaggerated. While very much a backup, archive, and long-term storage media, its per-GB cost structure (and the fact it only consumes power when actually being used) has kept it prominent—even if not popular—in many operations.

Users must choose from these options in order to address their key challenges; ESG research into those key challenges (see Figure 2) reveals a list that looks beguiling simple, but is led by contradictory challenges: on the one hand, there is enormous growth, and on the other, costs that must be managed. Tiering clearly could be a palliative here. There is no absolute formula when it comes to tiering; each data center and organization will be different—their applications, business needs, and budgets will change the outcome in each instance.

Figure 2. Top Storage Challenges



Source: *Enterprise Storage Priorities Emphasize Information and Infrastructure Efficiency*, January 2009.

However, a typical tiering of storage uses and tools (using Hitachi examples) might look something like this:

Primary Production Data: Mission critical applications, perhaps on a Hitachi USP V behind an HNAS, and probably using high speed, highly protected/replicated FC and SAS disks

“Business Class” Production Data – Other applications, using a variety of [heterogeneous] drives, RAID levels and replication, likely packaged in HNAS and AMS devices

Less Critical Data: Snapshots, less critical applications, probably on virtualized storage behind a USP V, and with a large proportion of SATA drives

Archival Data: Low cost, high capacity SATA drives, tape, or VTL for saved snaps, read-only historical data and maybe integrated with the HCAP platform for active archive and data warehouse uses

Beyond the raw physical device choices, there are a couple of other considerations for a truly dynamic hierarchy. Virtualization and migration capabilities are the lubricants that make intelligent and automated tiering possible. These are explained in more detail in terms of Hitachi’s specific implementation in a later section. At its simplest, however, virtualization separates the system view and handling of data from the actual physical infrastructure and migration actually handles the physical moves. This is vital for two reasons:

1. It enables easy migration, even between unlike storage devices, which is crucial for efficient tiering and is the bedrock of the heterogeneous block and file management that Hitachi offers.
2. It enables more efficient provisioning by removing the strict and inefficient utilization restrictions that have resulted in decades of poorly utilized disk systems. Bottom line: it permits a much higher percentage of storage assets to be used for actual data storage, which translates into a better cost structure for users.

The Business Value

Let’s start with the basics—comprehensive and well implemented storage tiering can save money. Saving money is invariably good as long as it is not detrimental elsewhere. Tiering’s value is derived from:

- Having the right data (only) on the right storage device (always) at the right time
- Better utilization of all storage assets (which also means lowering floor space and power expenditure)
- Reduced management resources (despite, and because of, better alignment of data to tiers)
- The potential for more purchasing leverage (Hitachi’s offering specifically embraces and supports heterogeneity, which means users are not committed to wholesale replacement of all non-Hitachi technologies)

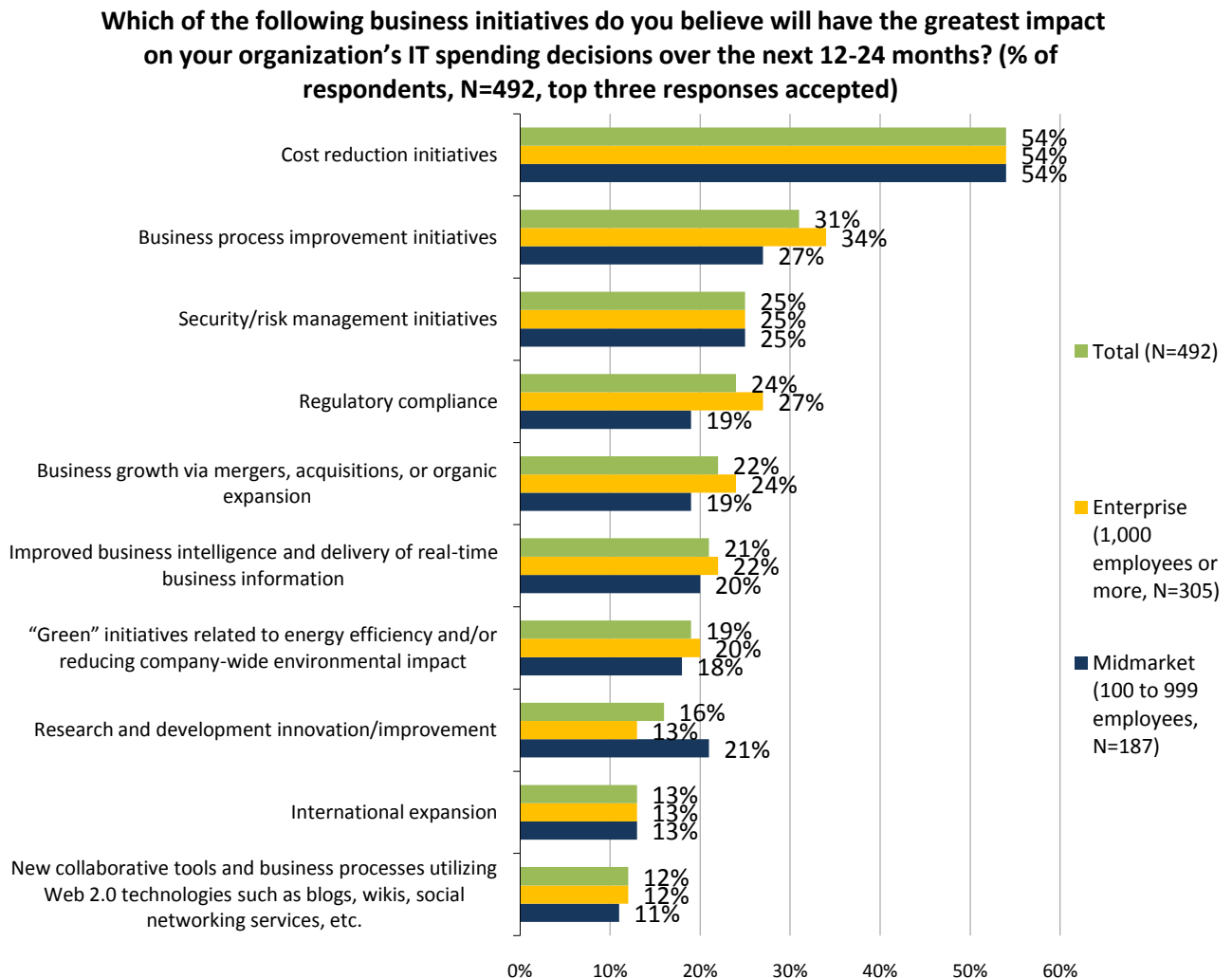
Hitachi has been highly focused on developing a “Practical Automated Dynamic Storage Hierarchy” that works across both block and file storage to reduce costs while maintaining or even improving service levels. The results of its focus (see next section) are both real and impressive; moreover, the overall offering works under a unified management umbrella and provides advanced search capabilities. As previously mentioned, however, everything to do with storage hierarchies and tiering is ultimately a function of the economics that both underpin and surround storage decisions. Before moving to look at the granular Hitachi solutions in more detail, it’s worth understanding what the economics of storage mean in a practical sense. All users are faced with a set of challenges that are disarmingly easy to list: keep driving down costs, minimize complexity, and improve the quality of service that is provided. But how do you do that in an environment of burgeoning capacities; ever growing regulatory, business, and compliance needs; and constant change?

The simple answer appears to be exactly what this paper is discussing: employ storage tiers and data movement tools. But matching such tools to business value might require different responses depending on the situation. Should application data be moved (intelligence)? Is it to be moved temporarily (cached) or permanently (migrated)? What tier should it reside on, and when (dynamic movement)? And clearly, this is not as simple as picking a generic application situation—say, a file containing video—and making assumptions. Quite simply, all video isn’t all going to be treated the same way—a sudden demand for a particular web video might just need caching until the furor

disappears, whereas the rendering or special effects library for a movie might be more suited to a tiered approach. And need will be further refined by industry and use; surveillance recordings are also video, but even in that case, the data storage needs of searching for one perpetrator could be very different from those that are looking to establish overall behavior patterns. And the same issues and decisions could apply across all applications and processes—perhaps there is special tiering for workgroups or individuals (HR at tax time, the CEO’s office, and certain key suppliers or customers?); different parameters might be allowed for VMware OS images; and key tables or directories might be permanently cached while OS images are allowed onto a different tier.

The business value of tiering is obviously tied very much to the specific needs of a business and its particular application usage. To that end, the greater the flexibility of the overall storage tiering offering—across blocks and files, with dynamic intelligence built in, and a wide range of platforms—the more likely it is that the business value delivered can be high. One thing that can be stated clearly, however, is how most organizations are currently measuring that business value. ESG research investigated the business initiatives that will have the greatest impact on storage expenditures in 2009-2010. As the results in Figure 3 show, the number one result, by a very significant margin, is cost reduction. In the current economic climate, it is no surprise to see the dramatic emphasis on cost reduction, with no difference between larger and smaller organizations. What it means is that an efficient, comprehensive, and intelligent storage tiering capability—such as that which Hitachi offers—is absolutely relevant and applicable at the center of today’s IT and business drivers. Indeed, with its ability to deliver “pain free” business value (a.k.a., cost reduction) it has moved from a luxury to a necessity.

Figure 3. Business Initiatives That Will Have the Greatest Impact on 2009-10 IT Spending Decisions



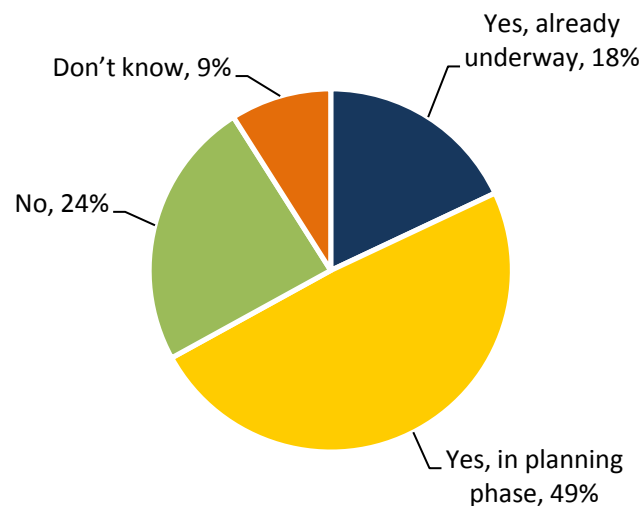
Source: ESG Research Report, 2009 Data Center Spending Intentions Survey, March 2009.

Practical Approaches - Hitachi Offerings

Whatever the approach, and whatever the speed and extent for implementing tiered storage, the user aims and potential benefits are consistent: to simplify management via reduced complexity, drive up utilization, consolidate assets, and thus lower operational cost while maintaining or improving service levels. Hitachi has an impressive array of tools with which to achieve these IT improvements and deliver the business value of cost reduction. The fact that these benefits can be delivered regardless of whether an application uses file or block storage only adds to Hitachi's value proposition in a market (see Figure 4) that ESG research confirms is heading significantly towards a more unified storage approach.

Figure 4. Plans for Unified Storage

Does your organization have any plans to consolidate NAS and SAN storage resources into a unified storage architecture that supports both file-based/NAS and block-based/SAN storage? (Percent of respondents, N=338)



Source: ESG 2008 Enterprise Storage Systems Survey, November 2008.

With two thirds of respondents already either starting or planning a more unified storage architecture, it only makes sense for overall storage management approaches and environmental processes to do the same. But while it makes sense, few vendors are actually making the move. Hitachi, on the other hand, can lay claim to one of the most comprehensive approaches. While full technical descriptions of Hitachi's components are not the point of this paper, what follows is a succinct summary of the three main categories that Hitachi combines into its overall intelligent tiering solution.

Hardware Platforms

Ultimately, any successful tiering strategy will demand a number of varied storage tiers—or tools, if you like—to be available in the toolbox. In this respect, Hitachi not only has a fine range of its own hardware, it can also integrate heterogeneous storage from other vendors into the mix. There is a choice of NAS and SAN platforms, there is solid state and VTL, and the range is from GB to PB—although all can be managed as one from a single console of the Storage Command Suite. Specific devices include the renowned USP V, as well as the SAS based AMS family and the newly extended mid-range Hitachi NAS portfolio. These products—as well as those from, say NetApp, EMC, or IBM that Hitachi virtualizes and manages as one large storage pool—are vital bricks in the overall architecture, but ultimately they are simply bricks. The real differentiation and value comes from the software.

Tiering Enablement

Virtualization: Implementing all these storage tiers could actually be counterproductive without some method of central management. A large part of the answer is storage virtualization, which provides the intelligence to manage all physical devices as a single logical system. Additionally—and crucially, when it comes to intelligent tiering—storage virtualization provides a *logical* management and control layer, which is the key to seamless data movement between different available tiers and devices. While virtualization provides the capability for movement, other data management tools provide the “decision smarts” that drive the dynamic and intelligent automation.

Intelligent File Tiering: This is comprised of Policy Based³ Data Migration and Dynamic Read Caching, which allow for the automatic migration of data to an appropriate tier of storage, optimizing cost and performance based on common file attributes or access patterns. This functionality also extends to third-party file systems and devices visible within the namespace.

Data Migration: Data migration enables organizations to move data *automatically* between different storage tiers based on *user-defined policies*. Policies can be established to demote or promote data based on the performance and protection requirements of the business at any given point in time. Data can be moved at either the file or volume level based upon the needs of the moment, versus storing it once and then walking away.

Dynamic Read Caching: In environments with large, dynamic unstructured data sets, administrators often have trouble predicting when a spike in demand for a particular file will occur. Read caching instantly copies files to a higher performance storage tier for use across physical or virtual servers when a file is requested from a slower tier of storage. While the file is being delivered to the requester, a background copy is made to a high-performance cache tier, ensuring that subsequent accesses will be delivered at maximum speed.

Migration and Caching Combined: Understanding the differences—even though there can be some overlap—between the two features is the route to understanding their combined value. Yes, caching can indeed be viewed as a special kind of tiering, but there remain two copies of the data. So it’s *performance* tiering, whereas genuinely migrated data is more *economic* tiering. Combining the two allows extra flexibility—for example, caching “covers your bets” to be even more aggressive on regular tiering policies. Alternatively, a user may decide to hardly migrate back at all, preferring to generically tier-down anything (say) not used in 60 days, but then to use other tougher or more granular parameters, such as an application or specific user, as the only way to tier back up.

Advanced Functionalities

Hitachi Content Archive Platform (HCAP) is one of the most impressive active archiving systems in the market today in terms of scalability, reliability, and ease-of-management. HCAP is an extremely sophisticated storage system that provides a true single cluster with no theoretical limit on the number of nodes it can support. This level of scalability enables the platform to support massive numbers of files and near-linear performance growth. Because an HCAP cluster is a true single cluster, a 40-node system is as easy to manage as a 4-node system, millions of files are as easy to retain as one hundred files, and hundreds of TBs of capacity are as easy to maintain as one TB. Additionally, by supporting WORM and user-definable retention periods, the HCAP provides all of the characteristics to enable organizations to meet corporate governance and federal compliance mandates.

Hitachi Data Discovery Suite (HDDS) provides a federated search capability across all of a user’s data managed by Hitachi NAS, HCAP, and NetApp filers. After all, even the best tiering in the world would be all-but-useless if you can’t then find anything and get it back! Hitachi understands that the ultimate users of all this will not be IT specialists, but rather will be knowledge-workers who don’t really care about the niceties of SAN and NAS, can’t understand why a Word document would be considered unstructured data, and don’t know the difference between a federated search and a haystack. Business people simply want a few clicks to get them what they are looking for.

³ Hitachi’s policy-based management applications use graphical, rule-based tools that simplify data management by minimizing the need for manual data migration, caching and replication. They provide “what if” analysis tools, policy templates, a scheduler for recurring policies, and integrated data protection safeguards. To formulate and schedule policies that automate Data Migration and Dynamic Read Caching, administrators create quotas and thresholds by selecting from variables that include last access, file name, path, and user identity.

Bringing it all Together

An Integrated Example: Combining all the attributes and functionalities can produce powerful capabilities. To take an example, imagine the challenges associated with lifetime patient record management in the health care industry. To avoid inefficiency and to manage costs appropriately, many different approaches are needed. Hitachi can meet them all in an integrated fashion that can be centrally managed. So, for instance, patient MRIs and PDF records (files, in both senses of the word!) could be stored on an HNAS system, perhaps using FC drives. Users can set policies based on file type so that, for example, all MRIs might be moved to SATA for economic purposes, while all the PDF reports could be migrated to HCAP for long-term retention. With HDDS, users can also conduct ‘*content aware*’ searches and use the results of such searches to trigger data migration across tiers. For example, all reports more than 12 months old and containing the word “arthritis” could be moved to HCAP. When a search is needed to support, say, patient Fred Smith’s hospitalization, HDDS is used to search across the various storage silos with a single query. Looking for records containing “Fred Smith” and the word “arthritis” might turn up appointment records, as well as the images and videos on the HNAS and HCAP systems. And, for the icing on the cake, the query can be conducted via a desktop Vista gadget in plain English by a non-IT specialist and the results packaged up and delivered to that person. This may seem a long way from intelligent tiering, but it is actually the essence of its value – storage itself is only valuable because of access and the ability to get useful stuff done. Migrated and intelligently tiered data is no different; movement, without management, is just another word for hiding things!

Support: As you would expect of an enterprise-class vendor with a top-notch reputation for quality, Hitachi has extensive professional services available around all these products and services. Equally important, the company has completed all the necessary diligence and compatibility testing with popular applications such as VMware, Exchange, Oracle, SharePoint, and so on.

The Bigger Truth

Implementing different tiers of storage is very common. Implementing an Intelligent Tiered Storage environment, however, is still more the exception than the rule—even though it is one of those IT items that is just plain common sense. Users risk almost nothing, yet stand to reap great rewards. Much of the immediate “low hanging fruit” is to be found by moving less active or dynamic data from expensive FC drives to more economical SATA drives, but the Hitachi implementation is rich inasmuch as it also allows users to wrap other external systems—from archive and dedupe appliances to external file systems—into its overall automated “tiering ecosystem” to the benefit of all.

Tiering itself is, of course, only a part of any answer, however well it is done. What is more important for Hitachi is that it has taken its best-in-class Intelligent Tiered Storage and wrapped it in an overall data management architecture. Block or file, SSD or VTL, SAN or NAS, archive or search, you name it. Now it just needs more of its potential customers to know this! Although famed for having the most rock-solid SAN storage systems, there’s a surprisingly broad story around what Hitachi has to offer—it has thousands of virtualized enterprise installations and has been in the NAS business for some three years, where its partner – BlueArc – provides products that combine brute force and brute finesse in equal measure. As to the market success for its offerings, much will depend on Hitachi’s ability to craft a culture of integration and solution into what has been a superbly executing product company. There is, however, a considerable window of opportunity as many of the obvious competitors are faced with a plethora of proprietary and varied platforms that make an architectural overlay extremely difficult.

Intelligent tiered storage is therefore fine, and credible, for Hitachi to claim. While the concepts of tiering are logical, the false-starts of ILM and HSM have tarnished some views. But the potential is now real and IT managers owe it to themselves and their organizations to properly investigate the potential benefits of intelligent tiering in their environments. The ROI can be extremely rapid and compelling—as near as one is likely to get as finding money growing on IT trees! Those who choose to investigate will find that intelligent tiering is not only possible, but can deliver business value via an optimal and dynamic blend of performance, capacity, and consolidation that almost always results in significant cost reduction; those who investigate just a little more will find the Hitachi unified block and file, value-added architectural approach a leading option and highly worthy of consideration.



Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel:508.482.0188 Fax: 508.482.0218 | www.enterprisestrategygroup.com