CLABBY ANALYTICS

Opinion

IBM in the Storage Marketplace: 2012 and Beyond

Introduction

The amount of data storage used by enterprises is doubling every 18 months — consuming an ever-increasing share of usually flat information systems operating budgets. To pay for storage hardware and software, new initiatives are being forestalled— a situation that stymies growth and kills competitive initiatives.

To get storage costs under control, enterprise information technology (IT) managers are currently taking several actions:

1. Increasing the utilization rate of storage (through virtualization, deduplication, performance tuning and other methods);
2. Using storage more wisely (creating storage tiers, using advanced management products to improve systems/storage/network integration while reducing management costs);
3. Automating storage–related activities (like provisioning); and,
4. Employing centralized management tools that improve efficiency

At the same time, these IT managers also are grappling with new user demands such as:

1. Having the ability to store, archive and easily access these growing volumes of structured and unstructured data;
2. Being able to use the wealth of stored corporate data for analytics and to support business decision making; and,
3. Providing access to data and content distribution over large distances but with local access performance characteristics.

Storage Strategies of the Future

The way we see it, future storage strategies will continue to focus on the above-mentioned themes (utilization, wise usage and automation) as a means to cut costs. At the same time, IT and storage managers will see increased pressure from users to provide easier, speedier, and more widespread access to more information — as well to provide business analytics tools to make that information more useful.

To contain costs, and to meet new user demands, IT managers need to build a comprehensive storage strategy that integrates well with enterprise systems and network strategies. And, in order to build this strategy, these managers need to understand where storage is headed in the future. From our perspective, we believe that:

- Today's virtualization activities are the precursor of future cloud deployment. The next logical step after virtualizing/pooling storage resources is to integrate the management of systems/storage/networks in order to seamlessly serve cloud

workflows. To do this, advanced and *integrated systems/storage/network infrastructure and management software* is needed.

- Today's tiering exercises are, for the most part, centered on trying to find the right storage device with the right quality-of-service capabilities to meet the service level requirements of users in the most cost-effective fashion. The big changes ahead in this space will be focused on *figuring out how to capitalize on a slew of new technologies (such as solid-state storage devices and different/more effective ways to use memory) that will significantly change how and where data is tiered.* Further, new storage controllers are being developed that exploit memory in different ways — again changing the way that storage will be tiered as well as influencing storage performance and quality-of-service levels.
- Today's storage-related management activities are still manual in many cases. *Future products will increasingly use analytics, self-learning algorithms, and automated policies to manage data including scheduling back-ups, data migrations, and replication*.
- Today's businesses collect much more information than they can process. And users are demanding easier ways to access and analyze this information to derive business value. *This trend will not only continue, it will accelerate as business analytics applications spread from the control of only a few individuals within select department to widespread use across all departments within an enterprise*.

In this *Research Report*, *Clabby Analytics* takes a look at the state of the storage market today — and we share our thoughts on where storage is headed in over then nest five years. We describe where most of the storage action is taking place today — and then we map IBM's storage products to these activities. We then switch gears and talk about storage futures (this research is based upon discussions with developers in IBM labs). After reading this report, storage manages should have the basics for building a longer term storage deployment and management plan.

To us, storage strategies should focus on exploiting storage technologies and best practices to improve performance and scalability. And, to do this, we believe that enterprises need to understand how various new storage technologies can be used to improve performance/scalability — and how integrated systems/storage/network management environments can be used to reduce storage complexity and drive down storage management costs.

Here's Where We Are Today

Most of the current "action" in the storage marketplace centers on reducing storage costs by improving utilization and simplifying management. The most prominent activities in this regard are:

- Data de-duplication (getting rid of redundant copies of data spread throughout an organization);
- Virtualization (improving the utilization rate of storage systems by finding unused storage and pooling that storage such that it can be used by applications/databases that need storage resources); and,
- Centralized dashboard-based storage management (so that fewer and less skilled personnel can manage storage activities).

If we were looking to get our storage costs under control, these are the activities that we would undertake to reduce/control storage costs. De-duping helps make better use of the storage that is currently owned — and de-dupe best practices ensure that storage deployed in the future is used wisely. Likewise, virtualization helps make better use of storage that has already been acquired — and virtualization best practices ensure that the utilization rates of future storage purchases are maximized. In addition, adopting new storage systems —packaged products that are "cloud-ready" with built-in management of features like thin provisioning and automated storage tiering — will drive down management costs considerably.

Deduplication is a discipline and should be and ongoing practice. Virtualization is the first step toward building a systems/storage/network cloud environment.

How IBM Storage Maps to Current Storage Market Activities

IBM's predominant theme around storage is "Storage Efficiency". IBM describes these best practices for storage efficiency:

- Stop storing so much (through deduplication and compression);
- Move data to the right place (with storage tiering); and,
- Store more with what's on the floor (with storage virtualization and thin provisioning).

IBM recently announced several new storage improvements including enhancements to the Storwize V7000, SONAS, XIV and the DS8000 that offer more capacity, improved scalability and better performance. IBM's Active Cloud Engine, Storage hypervisor, common management GUI and enhancements to Easy Tier build on IBM's theme of storage efficiency by simplifying management, improving data placement and automating storage tasks. We describe these enhancements in greater detail below.

IBM Storwize V7000 Unified

With the Storwize V7000 Unified, IBM has a home-grown unified block/file, SAN/NAS storage array. Also part of this announcement are enhancements to replication, drive options, including 3TB disk drives that provide greater capacity at a lower cost, a fully integrated unified storage management interface (storage pools can be shared among block and file storage) and extended Tivoli Storage FlashCopy Manager support. Also announced was support for Active Cloud Engine (described in detail below).

SONAS

Highlights of SONAS R1.3 include support for network connectivity at speeds greater than 100GB/sec, improved asynchronous replication performance, new drives that store up to 50% more data, and fileset level snapshots for faster back-up. Active Cloud Engine greatly enhances file management capabilities for SONAS by creating a logical file system that can span distance and different media types.

XIV

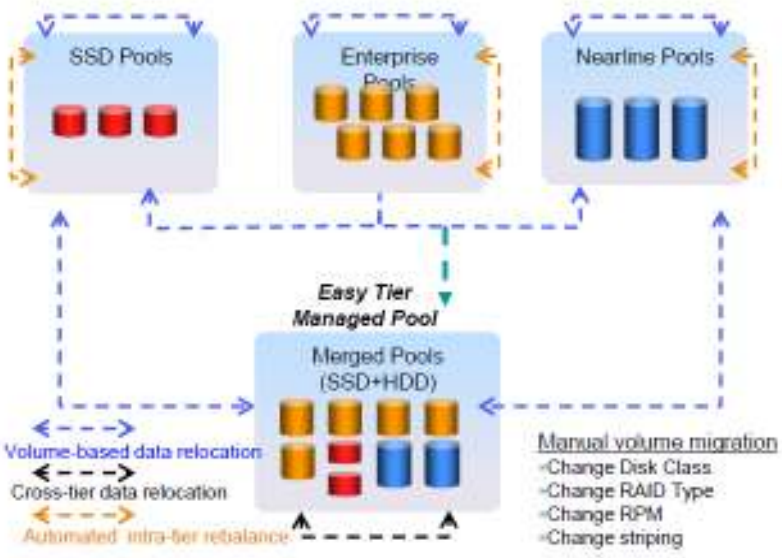
IBM XIV Storage System Gen3, the recently announced high-performance infiniBand model of XIV offers tuning-free, single tier disk storage now with 3 TB drives to enable a 50% capacity increase and even lower cost per TB. XIV's unique ability to use large density drives for demanding applications and handle Big Data is demonstrated through the newly released SPC-2/E benchmark, showing leading enterprise price-performance, and large file processing speed that tops all but the DS8000. XIV also now offers even greater

ease of use with a free iPad app, the IBM XIV Mobile Dashboard, for real-time performance monitoring from anywhere.

DS8000

DS8000 platform enhancements include I/O Priority Manager support for System z volumes which enables the integration of Workload Manager for z/OS (for improved quality of service management), better performance for DB/2 queries (such as index and table scans) and utilities, 3TB nearline SAS drives for almost 4x capacity, support for up to 45% more drives (with only 33% more floor space), Easy Tier (see Figure 1) includes automated performance rebalancing within a single-tier pool, support for both thick and thin provisioning, micro-tiering (for improved migration) and a new bandwidth limiting algorithm, both of which will eliminate unnecessary migration to higher-cost SSD storage.

Figure 1— Easy Tier Allows Automated Data Placement across 3 Tiers (with nearline 3TB drives)



Source: IBM Corporation, October, 2011

Active Cloud Engine

One of the highlights of IBM's recent announcement is Active Cloud Engine, supported on both SONAS and Storwize V7000 Unified. Active Cloud Engine is a policy driven engine that is tightly coupled with the file system and designed for managing large amounts of data both locally and also in geographically dispersed environments. Data managed by Active Cloud Engine can be rapidly searched and shared among users at multiple locations. Files can also be managed and moved to the proper tier, deleted or identified for back-up. Regardless of location, Active Cloud Engine maintains a single view and files that are federated in a global namespace giving users access to files not at their physical location. When users access a file, the current local cached copy is compared to a central a copy to ensure that the local copy is the latest version. Unnecessary replication of files to remote sites is eliminated also, thereby lowering network and storage costs significantly.

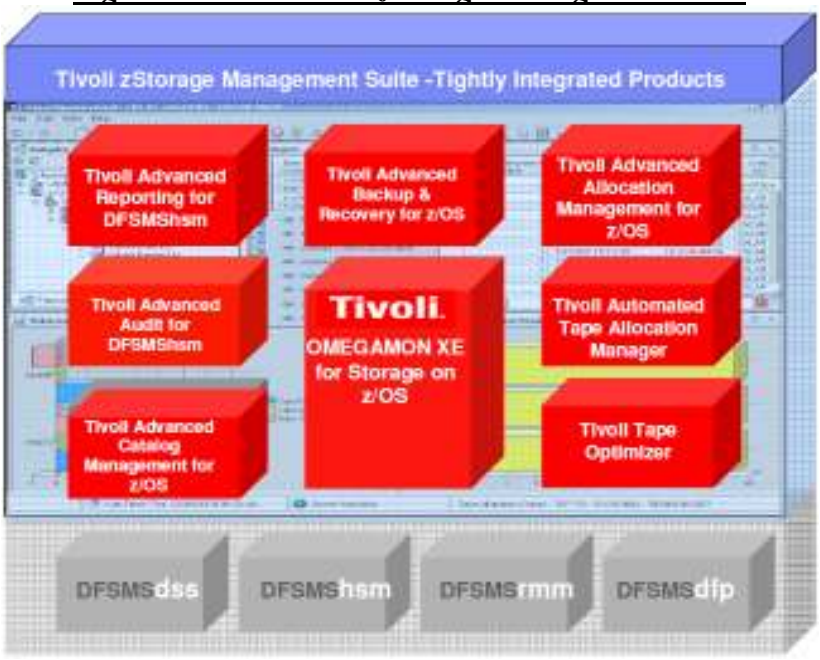
Common GUI

IBM has built on the success of the GUI that was originally designed for the XIV. Now IBM uses this common GUI across IBM XIV, Storwize V7000, Storwize V7000 Unified, SAN Volume Controller, DS8000 and SONAS. The easy-to-use interface allows for real-time monitoring of performance and system changes and proactive capacity planning. In addition the new GUI manages the policies (like file migration) set with Active Cloud Engine. This common GUI is a huge step toward integrating the IBM storage product line, improving storage management efficiency, reducing costs and enabling a single storage administrator to manage multiple platforms. At the same time, policies set using Active Cloud Engine automate storage-related management tasks, making it possible for storage to be managed by a less skilled IT administrator.

System Z Storage Management

The Tivoli zStorage Management Suite is a comprehensive set of products to monitor and manage heterogeneous System z storage. Using Tivoli Enterprise Portal (TEP), seamless integration is provided across the storage suite, and with other IBM/Tivoli management and automation offerings, improving management efficiency and productivity. At the core of the management suite is IBM Tivoli OMEGAMON XE for Storage on z/OS — with OMEGAMON Dashboard Edition (DE) customers can create views of storage information spanning both System z and Distributed storage. Also included are advanced reporting, catalog management, allocation management storage controller-based replication—a no-overhead, instantaneous back-up and recovery solution.

Figure 2 – IBM Tivoli zStorage Management Suite



Source: IBM Corporation, October, 2011

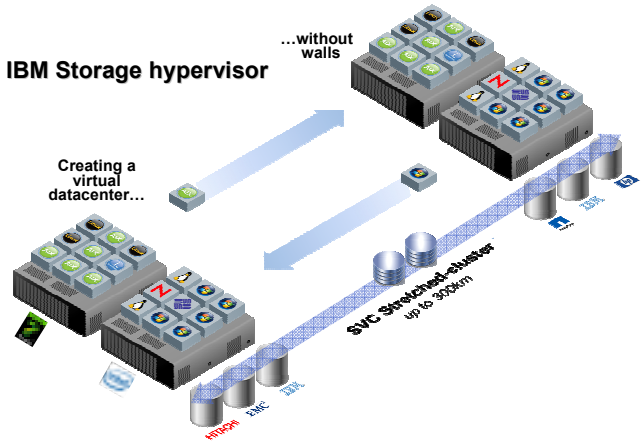
IBM and the "Storage hypervisor"
IBM (along with a few other vendors) is advancing the concept of a storage hypervisor.

For many (maybe most) enterprises, the deployment and exploitation of "server hypervisors" is well underway. A server hypervisor is a piece of code that sits between the operating environment and underlying hardware. This code essentially creates multiple operating environment images (virtual machine images) that the hardware then processes. Server hypervisors have brought new levels of efficiency and mobility to compute environments and are rapidly becoming ubiquitous in their use. IBM's storage hypervisor provides many of the same benefits for the storage environment – pooled physical resources, mobility of virtual volumes, common capabilities regardless of hardware choice, and centralized management.

> ***The point is that the BIG components of an IT budget are tied up in compute resources and storage resources. IT managers have leveraged server hypervisors to make dramatic improvements in the compute part of the budget. Because the benefits are so similar and the IBM Storage hypervisor is so complete, we expect it to play a central role as IT managers complete virtualizing their server resources and move to virtualize their storage resources.***

The IBM Storage hypervisor is both a storage virtualization platform (System Storage SAN Volume Controller) and storage virtualization management (Tivoli Storage Productivity Center). It includes a storage service catalog with highly automated provisioning of virtual storage resources and offers integration with usage-based chargeback. One of the highlights of IBM's most recent announcement is the Storage hypervisor ability to 'stretch' virtual storage resources across physical datacenters up to 300km apart. When combined with virtual server mobility tools like IBM PowerVM Live Partition Mobility and VMware vMotion, clients can create virtual datacenters that span physical locations with the ability to transparently move workloads from one place to another.

Figure 3 – IBM Storage Hypervisor



Source: IBM Corporation, October, 2011

The Future of Storage: A Discussion with IBM's Storage Labs

A few months ago, *Clabby Analytics* attended a briefing provided by Bruce Hillsberg of IBM Storage Systems Research (a key part of IBM's world renowned research organization), who shared his vision of how storage would evolve in the future. IBM Research is a division of IBM that has been put into place to marry customer needs with evolving technology (this is the division that spawned IBM Watson, a next-generation data analytics system). IBM Research's success is based on technological experimentation that drives innovation at all levels – silicon, storage, mainframes, servers, networking and software. As part of the briefing, we learned that IBM led the market in 2010 for the number of patents issued (5896) -- more than the next three (Microsoft, Intel and HP) combined—pointing to IBM's strength in research and technology leadership.

Readers should know that in storage, 40% of IBM's researchers are experimentalists — they are devoted to exploratory projects – experimenting with potentially disruptive technologies that may change the way that data is stored and managed. The other 60% of IBM's researchers/developers actually build products that feed directly into near term (1-5 years) and long term (as many as 10 years) product plans. For example, IBM Storage Research invented IBM Active Cloud Engine, Tivoli Storage Manager, General Parallel File System, SAN Volume Controller, and Linear Tape File System (LTFS) for LTO5 and contributed to IBM Easy Tier, XIV, SONAS and Tivoli Storage Productivity Center.

As for what we learned from Bruce, the following subsections describe some of his vision as to where storage is headed.

A Closer Look at the Evolving Storage Cloud

To build efficient useful storage clouds for global businesses, data must be accessible over long distances and multiple sites must have the ability to contribute data to the cloud. To address this issue, IBM Storage Research continues to enhance General Parallel File System (GPFS), a scalable, highly-available file system designed for high-performance computing. GPFS also features multi-petabyte storage management, shared storage pools, massive namespace support and scalability. GPFS supports clustered file systems and individual clusters can be connected to provide parallel access to data across geographical distances. While GPFS was originally designed for HPC, GPFS is an ideal foundation for today's mainstream applications such as digital media, parallel computing, large web applications, engineering design and business analytics that demand high volume and high performance.

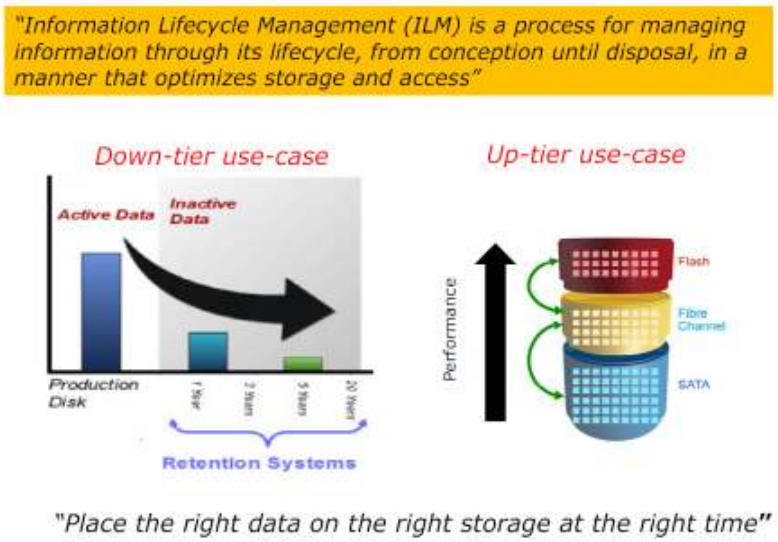
Features of GPFS allow the implementation of "shared nothing" clusters for higher levels of availability, performance and scaling. In "shared nothing" architectures, storage resources are dedicated to each node, allowing huge amounts of data to be gathered from many sources and analyzed in real-time. This distributed architecture which includes a common namespace across computing platforms allows these complex workloads to be run much more efficiently. Although HDFS (Hadoop Distributed File System) is also a "shared nothing" architecture, GPFS has several key advantages. With HDFS, stored data is only available to Hadoop applications, whereas with GPFS there is no limitation. HDFS does not provide high availability, because an HDFS file system instance requires one

unique server, the name node, which will be a single point of failure. If the name node goes down, the file system will be offline. GPFS can be configured so that there are no single points of failure. GPFS provides value-add features such as security replication, snapshots, backup, archiving and data caching. A recent study performed by the Geogrid Project has compared the performance of an HDFS with IBM's GPFS over a variety of workloads. It shows that a suitably optimized cluster file system such as GPFS can match the performance of HDFS for a Map-Reduce workload while outperforming it for the data access patterns of traditional applications. One of the real benefits of future generations of GPFS will be the ability to efficiently manage massive computing environments of up to 1 trillion files.

One of the other attributes that will contribute to the efficiency and performance of compute clouds is virtual server mobility, another concept that IBM Storage research is exploring. Today's clouds are very server-centric, focusing only on CPU utilization which won't address bottlenecks elsewhere. Virtual server mobility will take into account storage and network performance characteristics as well—enabling optimized provisioning, real-time problem isolation and resolution and real-time relocation of processing and storage.

A Closer Look at Foundation Technologies for Information Lifecycle Management (ILM)
Information Lifecycle Management is based on the concept of "using the right storage for the right job" and relies heavily on storage tiering capabilities. In general, more expensive, high performance SSD storage should be reserved for mission critical, newer data and frequently accessed data, while less critical data and reference data should be stored on less expensive HDD. Tape can also factor into the equation for the storage of archival data. An important element of an ILM strategy the automation of the data migration process— this will offer savings both in terms of storage and storage management costs as well as vastly improved performance. This automation can be achieved either by " self-learning" algorithms that look at data access patterns and migrate data accordingly or by setting policies around how the data will be migrated.

Figure 4 – Information Lifecycle Management



Source: IBM Corporation, October, 2011

IBM Storage Research is looking at a range of technologies that will improve the effectiveness of ILM strategies (pushing optimization of solid state up and down the stack — using consumer-class flash for enterprise applications).

SSD's and Storage Class Memory (SCM)
SSD's are data storage devices that use solid-state memory to store data. SSD's retain data in non-volatile memory chips that have no moving parts (so no spin or seek times). As a result, they are more reliable, quieter, and have higher read/write performance. They typically use flash memory that can retain memory even without power. As prices for SSDs drop they are becoming increasingly popular for tiered storage.

Storage Class Memory (SCM), as defined by IBM, is a "new frontier for Solid State Storage", "a new class of data storage/memory devices that blurs the distinction between memory (fast, expensive and volatile) and storage (slow, cheap and non-volatile). SCM promises to be non-volatile, high performance (DRAM-like) and inexpensive (disk-like) and solid-state (no moving parts)".

IBM, through its Solid State Initiative, is experimenting with a range of technologies in this new class of storage.

- Phase-change memory (PCM) or PRAM technologies, according to Wikipedia, use chalcogenide glass "with the application of heat produced by the passage of an electric current, this material can be "switched" between two states, crystalline and amorphous." This capability contributes to improved performance, higher capacity and better scalability.
- Magnetic Racetrack memory or MRAM (Magnetic RAM) – another technology with near DRAM like performance, low power requirements and high capacity that uses electric currents moving through magnetic domain walls.

Archive Systems
Another area of focus for IBM Storage Research is archive systems. Two challenges were identified with respect to archiving data (1) bit preservation and (2) logical preservation. Bit preservation refers to the ability to physically store and preserve the data for a specified period of time. Logical preservation refers to the ability to read and productively use that data for a specified period of time. For example, are there still applications that can access and search the data for relevant information, read it and use it to solve a business problem?

Regulatory requirements (like, for instance, Sarbanes-Oxley) require businesses to have archival strategies in place. Given the ever-growing volume of corporate information, archive systems must be hugely scalable, high capacity and cost effective. For this reason, IBM relies heavily on tape and their leadership in tape technology. Tape is low cost, "green", and tape density is doubling every two years (a goal that IBM Research will continue to pursue). LTO5 and LTFS are two examples of technical advances in tape that were driven by IBM— and are optimized for collaboration and sharing, particularly important in the entertainment industry. LTFS is an open tape format that allows for easy exchange content and has a directory stricture that looks just like disk—making tape much more usable than the tape storage of the past.

Summary Observations

Readers should know that we have a bias when it comes to storage design, deployment, and management. We do not believe in isolated storage strategies — strategies that focus only on storage and not on the integration of systems, networks, and management with storage. Systems, storage, and networks are intertwined. Instead, we believe that in order to maximize the efficiency of an information systems environment, all three elements need to be intertwined into a holistic, integrated strategy.

> ***What this means is that, when it comes to strategic planning, we tend to prefer vendors that offer integrated systems/storage/network solutions over vendors that offer point product solutions. Our reason: vendors that offer converged solutions reduce the integration burden on IT buyers. Converged solutions simplify management and helps reduce operational costs.***

What we found as we examined IBM's line of storage and storage management products is that IBM is a builder of integrated storage, system, and network — and related management products. We especially liked IBM's announcement of a common GUI across storage platforms including IBM Storwize V7000, Storwize V7000 Unified, SAN Volume Controller, XIV, SONAS and the DS8000 is significant because prior to this point, management across these storage subsystems had few commonalities. It also appears to us that IBM is working to improve integration and synergies of storage management capabilities across distributed systems and mainframe environments through TEP, common dashboard management and by adding I/O Priority Manager support for System z volumes in the DS8000.

IBM also continues to deliver on its efficiency promise with automation of storage functions such as tiering and data migration, and policy-setting through Active Cloud Engine and Easy Tier. Automating storage tasks through centralized GUI-based dashboards spanning multiple products will require fewer, and potentially less skilled IT administrators to manage the storage infrastructure (helping enterprises reduce human labor costs).

As for giving users easier access to data, users benefit from IBM's Active Cloud Engine with shared access to data and files across physical distances, improving collaboration and information access for better decision-making.

IBM Storage Research continues to contribute cutting-edge patented technologies to the IBM storage product line. Exploiting storage-class memory will produce dramatic improvements in performance and scalability at a reasonable cost. Enhancements in data archiving and tape technology will give customers flexible options for regulatory compliance, streamlining access and usability of data, while addressing the need to archive growing volumes of data cost effectively.

With IBM's recent announcements that expand functionality, and with some of the information shared by IBM Research about storage futures, IT managers should now have a clearer picture of where storage is headed in the future. Remember, your storage plan should not be siloed — as you build your future storage strategy you should also concentrate on integrating storage with your enterprise's systems and network strategies.

Clabby Analytics
http://www.clabbyanalytics.com
Telephone: 001 (207) 846-6662