IDC
Analyze the Future

WHITE PAPER

50 Years of IBM Innovation

Sponsored by: IBM

Dave Reinsel

August 2006

IN THIS WHITE PAPER

This IDC White Paper focuses on how five decades of IBM storage innovations have helped to drive business productivity and serve as a catalyst for the expansion of computing from a small number of business applications to today's ubiquitous digital world.

Executive Summary

From the earliest writings on cave walls to today's word processors and voice recognition applications, humans have found new and innovative ways to capture data. Capturing, maintaining, managing, and providing access to data has been and will continue to be a primary goal for storage companies like IBM. However, storage is about much more than that, it's also about extracting real business value from that data.

Data can unlock trends to help anticipate future dynamics or events. Data can protect a company from devastating litigation or bring to justice despicable criminals. More than ever before, chronicles of history are preserved in the data that we capture. Of course, in order to calculate the right timing and price for a company acquisition, or to plot the right coordinates to land the space shuttle safely, stored data must be recalled in a timely, uncorrupted, precise manner. In other words, data must be stored efficiently, accessed quickly, preserved flawlessly, and protected indefinitely. These are the chores with which users are faced and these are the chores that IBM has made easier for users through decades of disk storage innovation.

Decades of IBM Innovations

The IBM RAMAC 305 System — The Introduction of Random Access and a Boost for Application Development

In 1956, exactly 50 years ago, IBM introduced the world's first disk storage system, the IBM 350 Disk Storage Unit (see Figure 1) with the IBM RAMAC 305 computer. The groundbreaking storage system was designed not only to capture what seemed to be an unprecedented amount of data (5 million characters), but also to make that data randomly accessible, rather than sequentially. In other words, data could be accessed much more quickly The RAMAC name itself is an acronym for this monumental advancement in how computers access data: Random-Access Method of Accounting and Control.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

FIGURE 1

The IBM RAMAC 305 with the IBM 350 Disk Storage Unit

Source: IBM, 2006

Aside from magnetic tape, which is still a useful, cost-effective storage technology today, sequential storage and access technologies (e.g., punched cards and Mylar tape) were destined to become extinct.

With its 350 Storage system, IBM was addressing an age-old problem: fast access to data. It could literally take hours to find a certain piece of data on punch cards or tape reels. Using RAMAC, specific data could be accessed in roughly 600 milliseconds. RAMAC could stream data at 8,800 characters per second. In addition to faster data access, the RAMAC 305 had other industry-changing ramifications, specifically:

- **Application development.** Prior to the availability of rotating disk storage, application development was a cumbersome process that involved hundreds (or even thousands) of individuals. In addition to code writing, there were many people dedicated to the process of quality assurance and debugging. This process involved the manual loading of punch cards or Mylar tapes to test application logic. With the advent of RAMAC, the entire process of application development became dramatically simpler. Multiple versions of applications could be stored securely, revised, and tested in real time, without delays induced by sequential, and people-intensive, data access.

- **Readily available applications and data.** In the days before products and innovations such as the RAMAC 305, computers were reserved for very specific computational functions in the enterprise. Due to the resources required to develop and test applications, only the largest enterprises could afford both the hardware systems required and the staff of individuals required to do custom software development and testing. Examples of these kinds of enterprises were public and private universities, government agencies, and world-class public and private companies, such as General Motors or Manufacturers Hanover Bank. As enterprises installed disk storage, applications became easier to develop, so they became more prevalent. As new applications emerged, they addressed different kinds of functions in the enterprise and across different enterprises. Applications became more "off the shelf" (in a relative sense), so, additional enterprise

departments and functions could experience the benefits of having access to computing capacity. This drove not only increased demand by very large enterprises, but it also increased demand in the next smaller class of customer (but still large company); companies that could not have had access to this kind of computing and storage capacity and flexibility before

- **Increased level of demand.** Overall, this chain of events translated into market expansion for the industry. "Customers" could now think about new ways to gain efficiency, develop offerings, and create entirely new career paths in the business world.

The IBM 350 was truly the "big bang" that started the storage and IT world on the fast trajectory that it is still on to this day. In today's terms, the price of a RAMAC 305, at $3,200 per month for 5MB, was extraordinary. That equates to over $600,000 per month per gigabyte of storage (of course no one had a gigabyte back then). But at that time, it was a fit for those that had the resources and the unmet needs that were simply not being addressed prior to 1956.

Nearly ten years later, IBM was shipping its 2314 Direct Access Storage Facility. This storage system showed marked improvements in capacity and performance. Access times improved by an order of magnitude, data rates improved by more than 40x, as did maximum capacities. The 2314 leveraging up to eight disk drives of over 25MB per drive. A ninth drive could be integrated for backup or emergency replacement purposes; this was one of the first strategies to emerge to begin addressing persistency in the datacenter. The $175,000 price point of that product, continued to limit adoption of these and other storage systems to larger companies that had the resources to purchase and support the hardware and infrastructure.

From Record Keeping to Revenue Generation

There is no doubt that IBM's advancements in storage made accounting, record-keeping, and scientific computing more efficient. This efficiency delivered a side-benefit. As performance improved in terms of data access and transfer rates and applications became more prevalent, users began to find new ways to leverage their storage. In fact, the focus turned from efficient record keeping to revenue-generating transaction and services.

Customer-facing services (e.g., airline ticketing, ATMs, online banking) require at least three things: availability, fast response times, and accuracy. Once again, IBM innovation helped to address these needs.

Uptime is of utmost importance for companies providing services to paying customers. Usually measured in terms of so many "9's of availability," certain strategies and/or technologies generally are leveraged to guarantee uptime. Redundancy is a key strategy in this endeavor. Dual power supplies, dual controllers, dual ports, just about dual everything is integrated into many of today's high-end storage systems so that a component failure does not bring down the system, to a state where the system is unavailable and non-revenue generating. IBM began integrating redundant arrays of independent disk (RAID technology) in the 1980s as a means to increase data availability.

While storage access times continued to improve through the 1970s and 1980s, IBM introduced high-performance cache into its storage systems to provide customers with even faster response times. More advanced storage controllers, in this case IBM's 3880 Storage Controller, would use advanced algorithms to analyze data access patterns in order to move frequently used data from disk storage into semiconductor storage for high-speed access by the processor. The goal was straightforward: reduce the latency between the data request and the data delivery.

Continuous rapid data access is useless, however, without error-free data. Hence, error-correction and data protection became increasingly important. By 1970, error-correction was more than just something IBM thought about, it was something that it delivered in its 3330 Storage system.

IBM helped to enable a migration from nightly batch processing of data to online transaction processing by integrating a number of innovations in its storage systems. The requirements placed upon storage providers continue to increase as consumers grow accustomed to (and pay for) new services and immediate access to data. Moreover, the profitability of companies providing these services and data is dependent on secure and persistent storage infrastructures.

Today's storage solutions have come a long way in terms of performance, reliability, capacity, and affordability. But, the road is far from over. More data is being captured, stored, and retained than ever before, and this growth in storage will introduce new opportunities for innovation.

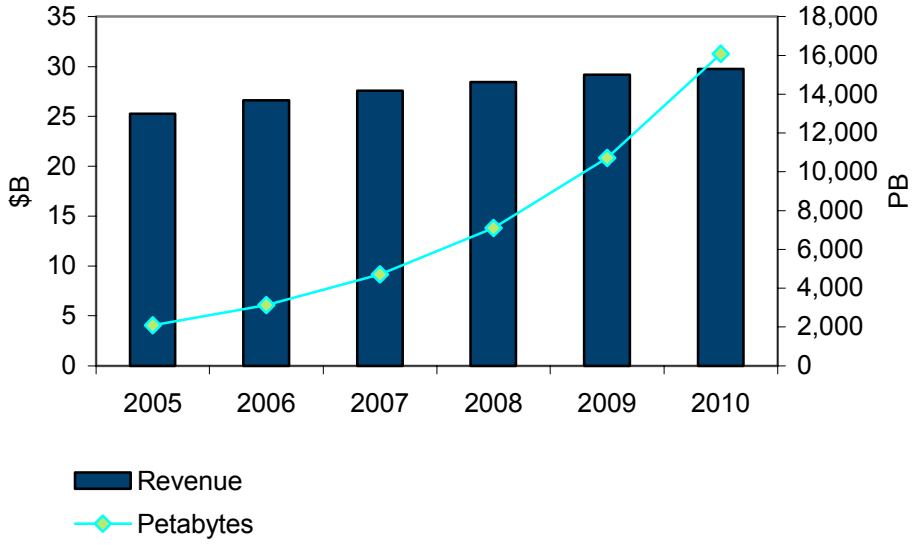
The Disk Storage Market – Billions of Dollars and Thousands of Petabytes

IDC estimates that in 2005, the disk storage systems industry generated over $25 billion in revenue and shipped an estimated 2,075 petabytes of storage. These metrics are expected to increase to nearly $30 billion on shipments of over 16,000 petabytes in 2010. In fact, the industry easily will ship more petabytes of enterprise storage over the next five years than it has in its entire history (see Figure 2).

The growth in storage capacity poses many challenges for many IT managers. From power and space management, to cooling and weight deficiencies, to the overall cost of managing all the required storage, IT managers will be demanding more from storage providers.

FIGURE 2

WW Disk Storage Revenue and Terabyte Forecast, 2005–2010



Source: IDC, 2006

Today's Demands, Tomorrow's Challenges

As IT managers are being asked manage more storage with fewer resources, storage must get cheaper, management must get easier, and efficiency or utilization must increase.

Lower Costs

Over the last few years, high-profile lawsuits and new laws and regulations have compelled CIOs to support compliance initiatives that require data such as emails, instant messaging, and even telephone conversations to be captured, stored, and made available within seconds. Medical institutions are capturing thousands of scans and X-rays of patients that must be kept so many years after the death of a patient. All this data needs to be kept, but on what type of storage?

Data increasingly are being categorized into different tiers of storage associated with different availability and performance requirements. The goal is to identify data that can be managed at a lower tier having a lower cost. IBM has responded to these emerging needs by leveraging lower-cost storage technologies (for example, lower-cost storage devices like SATA HDDs) to provide customers with storage solutions that address compliance, medical, and tiered storage requirements.

Better Management and Efficiency Through Virtualization

Through products like its new DS8000 Turbo (see Figure 3), IBM enables IT managers to manage hundreds of terabytes of data at a small fraction of the cost per gigabyte 50 years ago. Making storage more affordable is certainly a positive

development. However, if the storage is not easy to manage or if it is not used efficiently, then more storage can drive the cost of computing in the wrong direction.

Once again, IBM, through innovation, offers improved storage management and utilization through the introduction of scaleable networked-based storage virtualization. IBM has played a leading role in the development of a new category of storage products that IDC calls networked controllers. Over the next five years, networked controllers and the software that runs on them will emerge as a critical, multibillion-dollar battleground in the storage industry. It will drive changes in storage system architectures, spur significant investments in storage software, and have the potential to disrupt the current competitive environment.

FIGURE 3

IBM DS8000 Disk Storage System, Scales to 320TB

Source: IBM, 2006

IBM's storage virtualization engine, the SAN Volume Controller (SVC) first introduced in 2003 (and now in its fourth generation), makes it possible to extend many of the capabilities of individual high-end storage systems to a pool of diverse storage systems without sacrificing performance or reliability. In addition, because of the efforts of IBM and its business partners, the SVC is deployed in a wide range of companies both large and small. The two organizations below are representative of the reasons why companies seek storage virtualization solutions and the benefits associated with the IBM SVC.

City of Saskatoon, Canada

The city of Saskatoon, Canada has a multi-terabyte storage environment that currently is growing 100% per year. Managing this growth on its existing storage hardware infrastructure, while maintaining required uptime and performance, became increasingly difficult. Add to this the significant chore of migrating data to new storage systems when purchased, maintaining proper device driver versions, and managing various partitions throughout, and it is easy to see what prompted the organization to look for a better solution. Enter the IBM SVC.

Having heard about IBM's SVC, Saskatoon's IT manager traveled to one of IBM's technical seminars in Tucson, Arizona, listened to various presentations, got his questions answered, and caught sight of one IBM's key visions for SVC: SVC presents disk storage to applications and operating systems without any physical characteristic constraints. In other words, Saskatoon's IT manager envisioned a day when he could spend the majority of his time managing the data, as opposed to the hardware on which the data was stored. A couple of months later, Saskatoon had the IBM SVC up and running.

Now that its storage environment is virtualized, Saskatoon's IT manager finds it much easier to manage its disks and other hardware environment. Partitioning its terabytes of storage is accomplished with a few clicks of a button and driver management is simplified significantly given the single SVC control point. These benefits have helped improve staff efficiency and has allowed more time to be spent on other pressing issues, such as: Understanding why storage requirements are increasing 100% per year, identifying opportunities to purge and archive various data, and seeking ways to eliminate some of the junk data management activities."" ""

Colgate-Palmolive

Every three years, the Colgate-Palmolive enterprise data service center faces the chore of renewing, or more likely replacing, its leased storage hardware. In addition, consistent with many large companies, the datacenter has consolidated other data centers into its own (either via acquisitions or internal consolidation initiatives). These consolidation tasks can be quite time consuming and inefficient, especially when trying to keep critical applications up and running with top efficiency. With over 320TB at stake (and increasing 30–40% annually), Colgate-Palmolive's IT managers brought in the IBM SVC to help alleviate its tri-annual headache, as well as to streamline other storage management issues.

As one of the first adopters of IBM's SVC 2–3 years ago, Colgate-Palmolive has benefited from significant IBM support through each SVC version. Today, the service center no longer dreads its lease renewal. Not only are storage upgrades easier, so are server and other device upgrades. Perhaps more important and impressive, is that IBM's SVC has allowed Colgate-Palmolive's data service center to meet its ever-increasing storage requirements with limited growth in budget and flat headcounts.

What is next for Colgate-Palmolive enterprise service center experts? They plan to deploy additional management tools. They would like to see real-time storage usage (what does that mean?), storage on-demand, and more seamless management of

various tiers of storage. (You might want to get rid of these "what's next" sections, they don't add a lot of value to the overall theme of the report).

IBM Storage Innovation Far from Over

IBM has taken storage innovation to a new level with its SVC, but more innovation is needed. The company continues to invest in new technologies in order to meet real company needs today and anticipated needs for the future. Beyond driving management and utilization efficiency through virtualization, IBM will focus on reducing other storage management challenges in the future. By leveraging innovative component designs, IBM expects to provide end users vast improvements in managing failing components. For example, individual companies are increasing their installed base of disk drives to tens or hundred of thousands of drives. Based on today's design life of five years, a company with 100,000 installed and running drives can expect to replace over 50 drives a day. This is a pain point upon which IBM has its eye.

Power and cooling management is another area that demands attention. Today's datacenters are often built on legacy foundations of limited power supply and cooling efficiency. Significant upgrades are on the horizon for some companies. This is another area on which IBM will focus.

Given today's storage issues related to disk drive reliability, power, and cooling, IDC expects IBM to innovate around more reliable storage devices ranging from solid state to future probe storage (I don't know what Probe Storage is), which should provide relief in power consumption, and reduce heat generation. Data protection and security improvements are also absolutely critical moving forward. Driving more intelligence into the storage infrastructure will be another way that IBM will innovate in future storage solutions.

CHALLENGES

Although the storage market is replete with opportunities, Based on IDC research, these are the challenges that IBM will need to address.

- "More with less" will be a consistent theme. Storage must become more intelligent. Today's tasks that are managed by people need to be automated and foolproof. Storage budgets (for hardware and personnel) are typically flat or decreasing.

- Fewer workers in the datacenter means less time to troubleshoot problems or failures. Self-healing technology needs to become more prevalent. Mean time between component failures needs to be extended.

- Today's storage building blocks are reaching a plateau in performance and reliability improvements. This could mean radical changes to storage system designs by leveraging totally new storage building blocks.

- Data protection and security issues and requirements are only going to get more stringent. Iron grid security is way beyond reach today, but is vastly needed now.

- Energy costs and power and cooling deficiencies will compete with storage hardware budgets. IBM must innovate on environmental axes at the same time.

- Improving the efficiency of storing data exacerbates the need for better and smarter data management. Whether it is more intelligent hardware or more sophisticated software, datacenter managers need better tools to manage the amassing terabytes within their storage infrastructure.

CONCLUSION

IBM's decades of innovation have helped the industry's storage infrastructure advance in a multitude of ways. As customers' requirements are brought forward, IBM has responded admirably.

The number of applications dependent upon persistent and flawless storage is increasing at a phenomenal rate. Consumers increasingly rely on edge devices that require access to storage at some point. Examples include mobile phones, car navigation systems, and portable entertainment devices. An increasingly impatient population requires that the storage infrastructure provide continuous and immediate response times.

A company's business must be able to survive catastrophic events, whether it is generated by Mother Nature, a disgruntled employee or a hungry lawyer. Storage is at the backbone of this survival with remote backups, bulletproof security, and intelligent, indexed storage. IBM has been innovating over the last several decades and is primed and positioned to continue this innovation in the future.