

# DATA PROTECTION PROCESSES

## 2005 COMPUTERWORLD HONORS CASE STUDY

### BUSINESS & RELATED SERVICES

DISK-BASED DATA PROTECTION TECHNOLOGIES ENSURE DATA PROTECTION AND RECOVERY IN A TIMELY AND COST-EFFICIENT MANNER SMOOTHING OUT BUSINESS CONTINUITY AND THE COST OF DOWN-TIME BY REDUCING THE DATA VOLUMES BY A FACTOR OF ONE-HUNDRED AND INTRODUCING A UNIQUE PROCESS TO MANAGE DATA RETENTION OVER LONG PERIODS OF TIME. [20055297]

### SUMMARY

Disk-based data protection technologies ensure data protection and recovery in a timely and cost-efficient manner smoothing out business continuity and the cost of down-time by reducing the data volumes by a factor of one-hundred and introducing a unique process to manage data retention over long periods of time.

### APPLICATION

Introduction – Axion Disk-Based Data Protection Project:

This case study describes the development and implementation of a disk-based data protection system to provide backup, restore and remote replication of critical data by using sophisticated data reduction technologies to reduce the scale of data required to provide these functions so significantly that standard IT processing, storage and communications hardware can be utilized to provide the capability. Previously these capabilities required special-purpose solutions that were only economically available to a small number of IT organizations. This system, called Axion, permits a fully integrated data protection and disaster recovery capability that meets both demanding service level objectives and operational cost goals. We also describe the specific implementation of an Axion system to provide protection of the critical data maintained by Morgan Stanley's Institutional Services organization in the New York metropolitan area.

Background – The Growth of Digital Information:

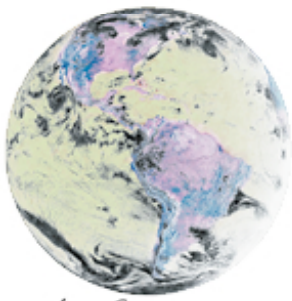
The information that represents the lifeblood of all organizations has historically been represented in many diverse forms. Over the past decade, however, the IT industry has made tremendous strides in providing systems to represent all information in the form of digital data; to communicate, store and process this data effectively and to present the data in structures that permit interpretation, understanding and analysis by knowledge workers from the research lab to the production floor to the shipping dock.

While there have been improvements in all aspects of IT, the most dramatic changes have been in the area of storing digital data; over the past twelve years, the cost of storing data on magnetic computer disks (“disks”) has decreased by a factor of more than one-thousand. In fact, over the half century since magnetic data storage has been available commercially, its costs (adjusted for inflation) have decreased by a factor of more than one-billion.

This economic disruption has been so dramatic, that the University of California, Berkeley recently estimated(1) that 92% of all new information is stored magnetically, primarily on disks, while only 0.01% is stored on paper.

The transformation of disparate information sources to digital forms stored on disks has permitted a level of organizational effectiveness, responsiveness and adaptability that would have been inconceivable only five years ago. It has also freed most workers from routine, low value tasks and enabled a focus on the complex tasks in which human intelligence provides the greatest added value.

This enormous concentration of critical information on a single technology, and often in a single IT system, has brought significant new challenges primarily focused around preserving and protecting the information contained on these disks. The Axion data protection project represents a new use of disk technology to represent, communicate and store the information contained in large data storage environments in a manner that permits effective management of organizational information resources for periods of time that can range from months to years as well as timely restoration of historical data that may be required by operational or



*A Search for New Stories*



Robert Carrigan,  
Chairman of the Chairmen's Committee

Ron Milton,  
Vice-Chairman of the Chairmen's  
Committee

Dan Morrow,  
Chief Historian

regulatory requirements.

#### The Challenge – Information Protection:

Since the information state of any organization is changing on a continuous basis, most organizations use a process called backup to create complete, consistent, point-in-time (PIT) images of the information state on a periodic basis. For most organizations daily PIT images are sufficient to meet both business and regulatory requirements. Since tens, and often hundreds, of these PIT images must be retained and managed to meet these requirements, mere replication of the IT data system for each PIT image is not operationally or economically feasible.

#### Trends – Organizational and Technical Changes Drive New Solutions:

Over the past fifteen years, backup technology has developed by combining a modest amount of data reduction (typically by a factor of five) with the use of low-cost removable digital storage media (typically tape) to provide a workable means of addressing this need. At the same time, three trends have combined to make this approach increasingly problematic.

The first trend is primarily economic; although the cost of both disk storage and tape storage have declined over the period, the disk storage costs have declined approximately five hundred times faster than tape storage costs (driven by device physics, market expansion and manufacturing economies of scale). The cost advantage of tape storage, therefore, has decreased to the point where this cost advantage may be small compared to the operational disadvantages associated with this sequential-access storage medium.

The second trend is primarily structural; organizations, and therefore their IT systems, have become substantially more distributed over the last decade. This means that substantial data must be transferred across extended data networks (often at great cost) to provide an integrated PIT image of the organization's information state. Therefore, even if it were practical to store extremely large amounts of data on a periodic basis in a backup system, in many cases it is not economically feasible to transport this data to the backup system.

The third trend is primarily operational; since the cost of IT systems have declined much more rapidly than the cost of human resources, modern IT system designs are optimized for the effective presentation of information to users rather than the efficient use of processing and storage resources. This has led to a situation in which amount of data that organizations manage has increased much faster than the amount of information they manage since "human friendly" information requires substantially more data for its representation than "machine friendly" information does; consider, for example, the difference between a high-resolution graphic representation (with a very large amount of data) in contrast to a table of numbers (with a small amount of data) both representing the same underlying information.

Taken together, these trends have made it extremely difficult to meet today's operational and regulatory information protection requirements with traditional backup solutions.

The objective of the Axion project was to utilize the advances in processing, networking and storage technologies over the past decade to extend the backup model to meet these new requirements by implementing a backup process that reduced the amount of data that had to be communicated, stored and managed by factors between one-hundred and one-thousand and to deploy this new backup process on cost-effective platforms consisting of standard IT servers and disk storage components.

#### Solution: Axion Disk-Based Data Protection:

This specific deployment of the Axion technology for Morgan Stanley Institutional Services permitted the complete offsite backup of two major data centers on a daily basis as well as a complete replication of each backup to a third data recovery site on a daily basis using existing communication networks. The project includes the capability of completely restoring the information state of the data centers, if necessary, within two days. The project consisted of a four-month evaluation period followed by a four-month implementation period and is now operational.

(1) Varian, Hal et al, How Much Information?, University of California Berkeley, Oct 27, 2003

## **BENEFITS**

Axion Provides Immediate Tangible Benefits:

Axion has permits a shift of capital resources, strategic planning and staff skills from technologies and systems that offered limited performance to a new technology that offers continued growth to meet evolving data protection requirements. It has also permits the operational support of the data protection solution to be integrated into the standard IT support activities rather than as a stand-alone set of skills and expertise.

The Morgan Stanley Axion project has permitted the organization to convert its data protection activities from an unsure, manual and costly process to a highly automated process with extremely little manual intervention while substantially increasing the confidence in meeting all service level requirements for data backup, restore and remote replication. This process of protecting vital information assets was integrated into an overall initiative to improve the confidence and performance in IT service delivery. It has permitted the strategic IT planners, for the first time, to assign the same level of confidence to backup and recovery that they can assume for other IT services. It has also enabled the management of backup data to be driven by complex policies regarding data retention and availability rather than by device characteristics and economic practicality.

The Morgan Stanley Axion project will first be implemented for the most critical shared information resources. As the implementation matures, however, it can be extended to all managed information within the organization providing, for the first time, a complete, managed online archive of the information history of the organization.

In terms of quantifiable performance metrics, the Morgan Stanley Axion project can reduce daily backup times by a factor of two to three and eliminated completely the need for periodic “full backups” that often required over sixty hours to complete. It also reduced routine network utilization for both backup and replication by a factor of greater than ten. Importantly, it was able to provide these operational improvements while reducing backup operations costs by a factor of two.

While each deployment of the Axion technology addresses organization-dependent performance metrics and economic considerations, it has delivered similar dramatic improvements in both service levels and operational costs in over fifty installations across a wide range of industrial, commercial, financial and governmental applications.

Axion has broken the “curse of scale” that has always been associated with managing dynamic information archives. While we do not believe that it will have a dramatic impact on the approach to storing and managing primary (applications-related) data, it will have a tremendous impact on storing and managing secondary (systems-related) data. It will permit the development and maintenance of extensive information records for all organizations.

## **IMPORTANCE**

Significance of Information Technology:

This project is essentially a pure information technology activity. Because of technological limitations, all traditional backup solutions have been obliged to manage data rather than the underlying information. In modern IT systems there are typically hundreds – if not thousands – of times as much data as information and, as a consequence, these systems were limited in scope, uncertain in reliability, operationally disruptive and extremely expensive in both capital and human resources.

By designing the new backup system to reduce the data at its source to the minimum necessary to preserve the information, this project was able to reduce the scale of data required to a degree that backup can become a non-disruptive element of an overall data management strategy and a more balanced element of the IT cost structure.

New Technology Developments:

While the underlying theory upon which the project is based have been know for a considerable period of time, there have been three developments over the past five years that have combined to make this project operationally important at this time. The first is the dramatic reduction in the cost of disk technology that has made it practical to implement the storage for the entire backup archive disk-based – and, therefore, accessible – for the first time. The second is the maturation of data processing techniques such as cryptographic hash algorithms and content-addressed storage structures that provide a computationally efficient means of identifying and removing redundant (non-information-bearing) data objects and of file system structures for dynamically managing large numbers of static data objects. The third is the growing trend toward increasingly information-sparse data structures within IT primary storage which significantly increases the value of data

reduction techniques as a central element in backup information management.

#### Three Key Technologies Make Axion Possible:

The overall approach taken in this project was to identify the need for a new, data efficient, process for data protection and, in particular for backup and restore. This approach led to the development of three key technologies that, combined, provide the structure and performance necessary for making a basic change in the backup process.

The first technology, commonality factoring, was developed by Avamar specifically to convert traditional data structures (files, file systems, databases and BLOBs [Binary Large Objects]) into data objects that expose the redundant occurrences of data that lead to the inefficiency of traditional backup processes. The second technology was a specific implementation of cryptographic hashing algorithms and content-addressed file structures that provided both efficient and computationally practical storage structures for the large numbers of objects created by the commonality factoring process. The third key technology was an efficient means of adding redundancy into the backup archive in a manner that provided an extremely high degree of protection against hardware failures while preserving the efficiency of the storage architecture. Avamar has received US Patents on all three of these key technologies.

As important as these technologies, however, was an implementation of the project within standard, and affordable, hardware configurations and in implementing these technologies in a manner that achieved performance levels that are required to provide significant operational improvements when compared to the best of the traditional backup solutions.

As a result of the focus on both technology and product implementation, this project has been able to deliver significant operational improvements (typically, by factors of five to ten) while reducing the operational costs for the users by a factor of two or more. Several key performance gains have been so significant that they have enabled entirely new capabilities; the most significant of these is the daily replication of full backups over affordable Wide Area Network (WAN) infrastructures.

## ORIGINALITY

#### Exceptional Aspects of Axion:

Data protection systems have traditionally focused on increased bandwidth and increased storage density in an attempt to keep up with the growth of data requiring protection. In fact, the fiber channel SAN technology that is now widely implemented as a primary storage networking technology was originally deployed primarily for data backup.

The Axion solution, on the other hand, concentrated on reducing, at its source, the amount of data necessary to provide data protection and therefore eliminating the need to communicate, process or store this redundant data.

#### Axion Depends Upon New Technologies or Significant Extensions of Existing Technologies:

Central to this approach is an integrated Data Protection Network (dpn) that provides enterprise-wide visibility of all data objects that are currently being protected. Once an object is protected, it is immediately visible to the Axion agents in all client systems so that Axion does not need to store objects multiple times no matter how many instances of the object exist in the enterprise. This strategy requiring that all data objects are accessible at all times is the reason that Axion must be implemented with a disk-based rather than a tape-based archive. Since this approach permits eliminating data redundancies not only across time (i.e., once seen in a client, objects tend to be persistent) but also across clients (i.e. an object seen in one client is very likely to exist in other clients as well – consider e-mails sent to many recipients within the enterprise) it provides data reduction (compression) at an unprecedented scale since earlier techniques could not provide this visibility across both time and clients.

As discussed earlier, Axion is also unique in the manner in which it creates these data objects. Traditional data object definitions (e.g. files, directories, or blocks) are sensitive to user modifications and/or hardware implementations and are, therefore not highly efficient in reducing large-scale data redundancies. They are also difficult to manage in a manner that is independent of both the operating system and the hardware of the client systems. Rather than use these “natural” data objects, Axion uses a process called “commonality factoring” to create data objects that are independent of the client hardware and operating system, efficient in a user modification (versioning) environment and lend themselves to effective management in large numbers (a

typical implementation of Axion will have billions of data objects). Axion also utilizes a technology originally developed for data encryption (cryptographic hash functions) to uniquely identify the data objects in a manner that provides an extremely efficient process for determining whether or not an object exists in the dpn and therefore does not need to be communicated or stored.

Finally, Axion uses a new file system technology, called Content Addressed Storage (CAS) to manage the data objects. CAS technologies have been developed over the past decade, but Axion is the only known implementation of a CAS that can dynamically manage data efficiently enough to be appropriate for use in a practical data protection system.

Although there is a tremendous amount of existing technology in these and related areas, Axion's implementation of these technologies has been granted three US patents over the past year.

Axion Evolved from a Universal Data Protection Network Concept:

The Axion solution implemented for this project evolved from the concept of a universally accessible, network that could be used to protect data on a global basis ("the World Wide Web of Data Protection"). While it has retained the basic design concepts and the scalability necessary to provide such a function, it has been implemented as a specific product to meet the data protection needs of commercial organizations primarily because only these organizations have the network and communications infrastructure that is necessary to support this type process at the current time.

## **SUCCESS**

Rapid Axion Project Implementation:

Axion is implemented as a modular, scalable hardware and software system that is can be integrated into existing Internet Protocol (IP) networks. Typical installation periods range from several hours to several days. Because of the disruptive nature of the technology, Axion has been designed to coexist with tape-based data protection systems during evaluation and transition periods. Depending upon the scale of the implementation we have found that the evaluation and transition period can range from several weeks to several months.

The Morgan Stanley Axion project was implemented for the New York metropolitan area over a period of approximately four months ending in November 2004. The Axion is now operational and the protection of critical data resources is being converted over, in a systematic manner, from a tape-based system to Axion.

In addition to the Morgan Stanley Axion project, we have implemented Axion solutions in approximately fifty other environments ranging from mid-tier enterprises to other IT enterprises on the scale of Morgan Stanley. These other implementations have utilized exactly the same technologies as the Morgan Stanly implementation, but they have typically been configured to meet different specific operational needs. All of these implementations are fully operational and providing active protection of critical data assets at the present time.

Direct and Indirect Benefits:

Axion is currently providing data protection for well over 2,000 system clients in fifty organizations at the present time. Since most of these clients are, themselves, multi-user systems, we estimate that Axion is providing data protection for over 10,000 IT users. It has had the most immediate positive impact on the operational staff who no longer need to dedicate their weekends to backup activity and who have been relieved of the onerous burden of recovering from failed backups (with a 20% to 40% probability) during "off shift" periods. At the present time, we have provided these benefits to over 100 individuals providing backup and restore support in the organizations using Axion.

The individual system clients have also benefited, however, because of the increased levels of protection that Axion provides and because of the ability to restore protected data quickly and efficiently in case of individual user error (the root cause of more than 80% of all organizational data loss).

Axion is Recognized as a Significant Improvement:

The original manager of the Morgan Stanley Axion project described Axion as "... cutting the Gordian knot of backup." because of its dramatic ability to simultaneously change the economic, reliability and communication requirements that had traditionally been associated with backup solutions.

The storage engineer at a large pharmaceutical company said of Axion, "This makes sense – a more intelligent

way to handle backup than just writing the same data over and over again.”

A senior IT operations executive at a financial institution that is currently evaluating Axion said, “I like it – it is a much smarter way to handle backups. Not every piece of data needs to be written out to tape. We solve the problem today with a sledge hammer – we won’t have to with something like Axion.”

Tom Fisher, vice president of information technology for Qualcomm CDMA Technologies said, “By collapsing the time for backups, eliminating a costly two-stage backup process and providing high-performance restore, Avamar Axion clearly justifies its return on investment.”

Curt Van Ripper, programming analyst for the Massachusetts Steamship Authority said, “We selected the Axion E disk-based solution for our Woods Hole headquarters because, once it is fully deployed, it will let us complete daily full backups, including backups of 150 end-user PCs, in less than 1.5 hours. By comparison, with our existing tape-based solution it takes five hours just to backup servers alone. And, equally important, when we add another Axion E and Avamar’s Replicator software at a second site located 30 miles away, we will be able to achieve data redundancy between these two sites, enabling us to fully recover data from any kind of disaster at either site in under two hours.” [Since the time of this quote, the Steamship Authority has fully implemented both backup and replication.]

Arun Taneja, an industry analysis recognized for his storage expertise said, “Most innovations are incremental. Avamar is not. Axion 2.0 can have a major impact on the way data is stored and protected, setting new standards for integrity, speed and scale.”

Axion has been Rapidly Adopted by Enterprises and Users:

Data protection is very likely the most conservative area of IT since its responsibilities are generally considered defensive and not directly contributing to an organization’s competitive position. Despite this, the dramatic economic and operational benefits of Axion have led to its rapid adoption in both large and mid-tier enterprise environments.

Tape-based backup systems are currently implemented in over 100,000 organizations. Many of these organizations have depended upon tape-based backup systems for nearly a decade. Despite both the size and the conservatism of the tape-based backup installed base, we believe that the benefits of Axion disk-based backup solutions will lead to a rapid adoption of the technology across the industry. While it is difficult to predict the exact timing of a technology change as disruptive as this one, we believe that as much as 20% of the existing backup installed base will convert to Axion, or to closely-related disk-based backup solutions, within the next five years. We expect, in other words, that Axion will truly change the backup industry.

Broader Access to Market:

We currently bring Axion to market both through direct relationship between Avamar and using organizations – as we did in the case of Morgan Stanley – and through storage-focused system integrators. We expect to broaden both of these channels as well as to develop relationships with service providers that will base backup services on the Axion technology. We are actively seeking partnerships in all of these areas at the present time.

The Axion technology is of significant interest to IT platform suppliers since it represents a new market opportunity for traditional IT server and disk storage hardware. We are investigating the possibility of forming Axion OEM relationships to significantly expand the market access for the technology, but we have made no OEM commitments at the present time.

## **DIFFICULTY**

Difficulties as a Matter of Choice:

The major difficulties associated with the Axion project arose from a number of specific decisions that we made for the project and can be broadly categorized as technical, economic and motivational.

Technical Difficulties – New Process:

The decision to recast data protection into a data-efficient process in 1999 was made at a time when neither the processing technology nor the disk technology could support practical implementations of the resulting system. There were strong indications that the trends in these technologies would provide the necessary functionality at economically-viable costs but, of course, neither the continuity of these trends nor their timing could be known

with a high degree of certainty.

Some of the core technologies for Axion had been described in the literature at the time the project was initiated and, in particular, cryptographic hash algorithms had been developed extensively by the National Institute of Science and Technology (NIST) for information security applications. The system concept, on the other hand, was entirely new and the requirement to integrate these technologies into a system that could provide a practical solution for a challenge of the scale and dynamics of an enterprise data backup solution had never previously been attempted. Once the basic solution had been converted to software (over 600,000 lines of code) the process of validating both the integrity and the performance across a wide range of representative operational environments represented a unique challenge in itself.

Since the object of the development activities was a product that could be widely deployed in commercial, industrial, and governmental environments, it was necessary to shield system administrators and end users from the significant internal complexity of the Axion product. We accomplished this requirement with two additional key developments; the first was a policy-driven administration and management process and the second was an automated monitoring and status reporting system that reduced the extensive diagnostic information created by Axion to specific, actionable events related to hardware health and system performance status.

#### Economic Difficulties – Completing the Product:

Although the technical complexity of the Axion solution was extraordinarily high, making Axion work in sub-scale or controlled environments was only a part of the challenge. The company anticipated that more than one-half of the resources necessary for completing the project of bringing a viable product to the IT market would be utilized after the development of the technology was essentially complete. As a result, the company had to attract capital investment at the level necessary to complete both of these development activities.

The company has raised over \$50M of capital to date for the Axion project. While long-term market adoption trends are extremely difficult to predict, we believe that this level of investment will prove sufficient to complete the project and to put the Axion product in the position of leading a significant revolution in the data protection market. Although the market is clearly in its “early adopter” phase at the present time, the market is large enough that the reasonable increases in adoption rates predicted by most technology diffusion models should be sufficient to permit the company to support, develop and extend the Axion technology with the operational income from its commercial activities with the Axion product.

#### Motivational Difficulties – Rationalizing Disruption Change:

Although the Axion project provides significant operational, performance and economic advantages compared to existing tape-based systems, it represents a disruptive change to a highly conservative market. Avamar has dedicated significant resources to support early adopters such as Morgan Stanley in understanding the benefits of this change in their own environment.

This process has led to the approach of highly consultative engagements with organizations that are potential adopters of the Axion technology. We have found that, in most cases, it is critical to provide analytical support for performance, operational and economic assessments that are highly tuned to specific organizational environments.

The Morgan Stanley Axion project was one of a relatively small number of projects we have been involved with in which the adopting organization’s project team had significant enough resources and stability to conduct the assessments with only limited support from Avamar. At the time it was implemented, it represented the single largest configuration of the Axion system that was deployed in an operational environment. Since it was not possible to assess full scale Axion performance, the evaluation and assessment was conducted on an Axion implementation that was approximately 10% of the scale required for the New York metropolitan deployment. Morgan Stanley successfully applied the metrics developed during the evaluation to the full-scale deployment.

While an individual project could have been successful with a subset of these three initiatives, the objective of providing a solution that would have a significant, long-term impact on the IT industry required all these elements of the solution.

