



Data Protection

Adapting to the Sea Change

2007 Edition

Free Preview



Mesabi Group

David G. Hill

Mesabi Group research focuses on IT information management with a concentration on storage and storage-management related issues from the data center and beyond.

The Mesabi Group recognizes three trends expanding the scope of storage suppliers within organizations:

1. The move toward **information management** (IM) (not just storage and storage management),
2. The move toward **including all of the enterprise from the data center to the edge of the network** — not just central data centers — as worthy of being managed effectively,
3. The move towards giving **every size organization from enterprises to small, medium size businesses** (SMBs) the necessary IT capabilities that they need to run effectively.

Within that context of IM centered on storage and storage management, the Mesabi Group focuses on two “revolutions” that dramatically change the ability of storage and information managers to zero in on key information in near-real-time.

- The separation of storage (e.g., via storage area networks) as an equal partner with processing and networking in delivering business value
- Information Lifecycle Management (ILM), the proactive, global management of data from creation to destruction

The Mesabi Group focuses on how these revolutions enable the storage infrastructure to serve as a vital component of the IT infrastructure not only to support essential business functions, but also — despite the critics — as an enabler in competitive advantage.

Central to both revolutions is data protection within a comprehensive governance, risk management, and government/regulatory compliance (GRC) perspective. An in breadth perspective of data protection must include existing and emerging technologies for data protection as well as enabling technologies.

The Mesabi Group translates storage technologies into business value; for example, showing how data protection technologies can help business management with the fundamental business functions of governance, risk management, and compliance.

Mesabi Group is managed by **David G. Hill**.

Table of Contents

Executive Summary	1
Key Report Findings	1
Preface.....	5
Business Continuity and the Backup/Restore Process — Never the Twain Shall Meet?	5
The Sea Change in Data Protection	6
What This Report Is — and Is Not — About	7
<i>Chapter One: The Time Has Come for Change</i>	9
What Data Protection Is.....	9
Data Protection Has To Be Placed in the Right Framework	10
Ride the Sea Change in Data Protection.....	11
<i>Chapter Two: Business Continuity: The First Foundation for Data Protection</i> ...	12
Business Continuity and Data Protection	13
Business Continuity Is Not Just Disaster Recovery	13
Disaster Recovery: Let’s Get Physical.....	16
Operational Recovery: Think Logically	17
Disaster Recovery Requires Judgment; Operational Recovery Requires Automation	18
Logical Data Protection Gets Short Shift in Business Continuity.....	19
Logical Problems Feature Prominently in Data Loss or Downtime	20
Logical Data Protection Problems Manifest in a Number of Ways	21
Do Not Neglect Any Facet of Data Protection.....	22
<i>Chapter Three: Data Protection — Where the Problems Lie</i>	23
Data Protection as It Was in the Beginning	23
Typical Data Protection Technology Today Still Leaves a Lot to Be Desired	25
Operational Continuity/Physical: Generally Strong, But Some Improvement Needed	27
Operational Continuity/Logical: More Attention Needs to Be Paid to Logical Data Protection	27
Disaster Continuity/Physical: Done Well, But Cost and Distance Are Issues	28
Disaster Continuity/Logical: The Danger of Being Under Protected May Be Very Real	29
Summing Up Data Protection Challenges by Category	29

Table of Contents

<i>Chapter Four: Data Protection — Setting the Right Objectives</i>	31
How High Is High Enough for Data Availability?	31
SNIA's Data Value Classification: A Point of Departure	32
Do Not Equate Availability with Value	34
Availability Objectives for Operational Recovery and Disaster Recovery Are Not Necessarily the Same.....	35
Availability Is Not the Only Data Protection Objective	36
All Primary Data Protection Objectives Have to Be Met	38
<i>Chapter Five: Data Protection — Getting the Right Degree</i>	40
General Use Classes of Data	40
Tape Is a Special Case	41
Understanding Degrees of Data Protection	41
The Third Degree — Levels of Exposure	42
Mapping Degrees of Protection.....	43
<i>Chapter Six: Information Lifecycle Management Changes the Data Protection Technology Mix</i>	46
Why Data Lifecycle Management Is Not Enough — the Need for Metadata and Management.....	47
ILM Is Deep Into Logical Pools of Storage	47
Logical Storage Pools at a High Level	48
Moving Information across Pools — A Distillation Process	49
Archiving Through a New Lens.....	49
Archiving: The Makeover	50
Protecting Archived Data	51
Data Retention.....	52
Disposition of Data.....	52
Data Retention Required for Both Compliance and Governance.....	53
Creating Data Archive Storage Pools by Data Retention Attributes	55
Security and Privacy	57
Active Archiving and Deep Archiving	58
Active Archiving Requires Active Archive Management	58
Long-term Archiving as Part of an Active Archive	59
ILM Changes the Data Protection Technology Mix.....	60
<i>Chapter Seven: Governance: The Last Piece in the GRC Puzzle</i>	62
Data Governance Must Respond to Changes in the Federal Rules of Civil	

Table of Contents

Procedure	63
Data Knowledge Is a Data Protection Objective of Data Governance ..	63
Litigation Holds	64
Data Auditability Is a Data Protection Objective of Data Governance...	64
ESI in Litigation Hold Must Be Placed in an Active Archive	65
Deciding What Data to Put on Litigation Hold May Be a Challenge.....	65
When Is Good Faith a Safe Harbor?.....	66
The Burden of Inaccessible Data	66
Sharing Responsibility.....	67
The Big Three — Governance, Risk Management, and Compliance — and Data Protection Objectives.....	68
Data Protected Differs by Management Responsibility	69
<i>Chapter Eight: Data Security — an Ongoing Challenge</i>	<i>71</i>
Data Preservation Is Data Good to the Last Bit.....	71
Confidentiality Is a Public Concern about Private Information	72
The Role of Data Availability and Data Responsiveness in Data Security...	73
The Case For and Against Encryption.....	74
The Special Case of Storage Security.....	75
Aside on Process Management.....	76
<i>Chapter Nine: Where Data Protection Technologies Play in the New Model</i>	<i>77</i>
Categorizing Data Protection Products.....	78
Mapping the Base Data Protection Technologies to the ILM Version of the Data Protection Framework	79
<i>Chapter Ten: Back to Basics — Extending the Current Model</i>	<i>82</i>
The Move to Multiple Parity RAID Is a Welcome One	82
Evolving Backup/Restore Software	82
Recovery Management	84
Moving Data Manually and Electronically — the Place of Vaulting and Consolidation.....	84
Remote Office Data Protection	85
Straight-forward Backup Consolidation.....	86
Backup Consolidation as a Byproduct of Server-Storage Consolidation	87
At Your Service — the Role of Service Suppliers	87
<i>Chapter Eleven: When Supporting Actors Play Lead Roles.....</i>	<i>89</i>
WAN Acceleration.....	89

Table of Contents

Data Reduction and Other Space Saving Technologies.....	90
Data Protection Management.....	92
Data Protection Change Management	95
Data Classification.....	95
Looking at Data Classification through Different Lenses.....	96
<i>Chapter Twelve: Disk and Tape — Complementing and Competing with One Another</i>	99
Disk-based Backup.....	99
Speeding up the Backup/Restore Process — Your Mileage May Vary	100
Improving Restore Reliability	101
Keep in Mind	101
Virtual Tape.....	102
Virtual Tape Library.....	102
MAID.....	103
Removable Disk Drives and Disk Media.....	103
Data Protection Appliances	104
Tape Automation.....	105
<i>Chapter Thirteen: High Availability and Low (or No) Data Loss Technologies.</i>	108
Copy Strategies.....	108
Point-in-Time Copy	109
Continuous Data Protection	110
Scheduled-Image Data Protection	111
Replication Strategies.....	112
Mirroring.....	112
Dated Replication — Pay Close Attention.....	116
<i>Chapter Fourteen: Special Requirements for Compliance, Governance, and Data Security</i>	118
The Use of WORM Technology	118
Issues with Physically Destroying WORM Media.....	118
WORM Tape	119
WORM Disk.....	120
Electronic Locking	121
Guaranteeing the Authenticity of Data.....	121
Privacy and Confidentiality	121
Encryption Appliance.....	122

Table of Contents

Compliance Appliance	122
<i>Chapter Fifteen: Tying It All Together — the PRO-Tech Data Protection Model</i>	123
The PRO-Tech Model for Data Protection	124
The PRO-Tech Model — Level 1	125
Process Layer 1	125
Rules Layer 1	126
Order Level 1	127
Technology Provisioning Level 1	128
Tying the PRO-Tech Layers to GRC Business Responsibilities	129
<i>Chapter Sixteen: Summing Up — Redesigning Data Protection</i>	132
Data Protection Is Everyone’s Business	132
Synthesizing the Data Protection Frameworks	132
Guidelines for Data Protection	134
The Challenge Ahead and a Call to Action	135
<i>Chapter Seventeen: Representative Companies</i>	136
Alphabetical Company Listings	142
Mesabi Group Conclusions	173
Glossary	174
Author Profile	183

Figures

Figure 2-1: Overview of Business Continuity	14
Figure 2-2: Business Continuity Is More than Data Protection.....	14
Figure 2-3: Business Continuity Keeps Your Business Running.....	16
Figure 2-4: Causes of Data Loss or Downtime.....	20
Table 2-1: Logical Data Protection Problems and Sources	21
Table 2-2: Data Protection Category Matrix.....	22
Figure 3-1: Data Protection: The Way It Was	24
Figure 3-2: Typical Data Protection Today	26
Table 3-1: Data Protection Challenges by Category.....	30
Figure 4-1: High Availability Depends upon the Entire IT Infrastructure	32
Table 4-1: SNIA Data Value Classification	33
Table 4-2: Operational Recovery and Disaster Recovery Differences	36
Table 4-3: Consequences of Data Loss.....	38
Table 4-4: Summing Up Key Data Protection Objectives.....	39
Table 5-1: Sample Degrees of Data Protection for Application <i>n</i>	43
Figure 6-1: The Storage Pyramid — Tiering and Pooling	48
Figure 6-2: ILM Changes the Logical Topology Storage Look.....	50
Figure 6-3: Data Retention Archive Pools.....	55
Table 6-1: Adding In Archiving to the Data Protection Category Matrix.....	60
Figure 7-1 the Governance Hierarchy.....	62
Table 7-1: Applying the Principles of Data Protection to the GRC Business Responsibilities.....	68
Figure 7-2: Mapping Data Requirements to the GRC Business Responsibilities	70
Table 9-1: Where Active Data Protection Technologies Fit in the Data Protection Framework.....	77
Table 9-2: Base Active Data Protection Technologies for Active Changeable Data	80
Table 9-3: Base Data Protection Technologies for Archived Data	81
Table 11-1: Differentiating Among the Different Types of Data	97
Figure 15-1 the PRO-Tech Model for Data Protection — Level 0.....	124
Figure 15-2a: Process Level 1	126

Figure 15-2b: Rules Level 1	127
Figure 15-2c: Order Level 1	127
Figure 15-2d: Technology Provisioning Level 1	129
Table 15-3: Data Protection: One Size Doesn't Fit All	130
Table 16-1: Data Protection Requirements for Application <i>n</i>	133
Table 17-1: Suppliers of Each Data Protection Technology (1)	138
Table 17-2: Suppliers of Each Data Protection Technology (2)	139
Table 17-3: Suppliers of Each Data Protection Technology (3)	140
Table 17-4: Suppliers of Each Data Protection Technology (4)	141

Executive Summary

Key Report Findings

1. Enterprises must understand that data protection does more than help to ensure business continuity. Failure to meet both compliance and governance requirements for data protection can create an unacceptable exposure for an enterprise.

The traditional purpose of data protection has always been to help ensure business continuity, i.e., to inhibit or correct major disruptions to business processes, within the overall corporate risk management responsibility. While risk management continues to be critically important, corporate governance and corporate compliance are elbowing their way onto center stage to form the governance, risk management, and compliance (GRC) triad. Compliance has attracted noticeable attention because of its role in various regulatory requirements, such as Sarbanes-Oxley (which is not industry specific) and HIPAA (which is). Governance is just starting to attract attention with the tipping point being the changes to the Federal Rules of Civil Procedure (FRCP) for civil litigation.

Although all GRC responsibilities rely on the four basic data protection objectives (preservation, confidentiality, availability, and responsiveness), enterprises must understand that each responsibility requires different answers to the same objective, with the result that the same technology solutions do not necessarily apply to each. Moreover, compliance and governance require businesses to meet a new secondary objective — data auditability. Data auditability requires putting new processes in place. That forces a new — and not necessarily welcome — realization to IT that someone will now — more nearly actually than figuratively — be looking over their shoulder. Finally, data governance as a subset of IT governance and corporate governance requires full content knowledge — not just file and database metadata knowledge. One of the key implications of the data knowledge objective is that business users, as the owners who control access to the content, have to become more fully involved and committed to meeting governance goals. In essence, IT and business users have to learn to run effectively in a three-legged race, which may not be easy on either party.

Failure to completely and accurately carry out all data protection duties now carries a high price; exposing businesses not only to costly disruption of business process, but also to possible legal consequences, such as sanctions and fines.

2. Enterprises must find a way to deal with the growing complexities in understanding and enabling data protection. Failure to do

so can expose an enterprise to an unacceptable level of both business and legal risk.

IT organizations face three sources of complexity when organizing the planning and decision-making process for data protection. The first is the existing data protection infrastructure, which is typically under increasing stress because of continued data growth, heightened demands for better service (such as 24x7 availability), and a lack of commensurate growth in resources. The second is fully understanding the implications and impacts of both compliance and governance that add an additional burden to IT in trying to meet the demands of data protection. The third is dealing with the sheer number of current and emerging technology choices available for data protection.

This report presents a number of concepts and first principles as well as some straightforward frameworks to help IT organizations organize their approach to these three complexities. The PRO-Tech model builds upon the concepts and first principles discussed in the report and provides a starting point for an enterprise to evaluate and examine its data protection strategy in greater depth.

3. Enterprises may be *over-investing in some areas* of data protection, *while exposing their IT assets to unacceptable risk by under-investing in other areas* of data protection.

For example, an enterprise may not understand the importance of logical data protection. An Ontrack study showed that nearly 40% of the causes of data loss or downtime are logical, not physical, problems. Yet enterprises may not have in place a high availability (defined as seconds or minutes of annual downtime) logical recovery approach for critical applications. (The tendency is to think in terms of physical solutions, such as mirroring, which are not the answer to logical data protection problems.)

The report shows enterprises how to think about where they need data protection, as well the degrees of data protection that are required to meet those needs. They can then better determine whether or not they are over-investing or under-investing to be able to meet particular data protection needs. The old bromide “one size does not fit all” applies to how enterprises fulfill their data protection requirements, but not for the basic principles of data protection.

Some large enterprises can afford to have a triad of data centers to ensure a high level of availability in the case of a disaster, whereas other enterprises simply use tape vaulting for disaster recovery, trading lengthy application restores for lower cost. However, all enterprises have to take into account both the need for disaster recovery to a remote site and operational recovery for problems that can be corrected at a local site.

Likewise, all enterprises need to take into account physical problems, such as disk failures, and logical problems, such as database corruption or a computer virus, for both disaster and operational recovery situations. The choice of individual data protection technologies is up to the enterprise — but overall data protection should fit within a common framework and model that applies to all enterprises.

4. Enterprises want “high availability” as part of data protection, yet virtually all use a “low availability” tape solution as part of their data protection strategy. For true data protection, enterprises should use multiple levels of availability in their overall strategies.

In an effort to ensure high availability for critical applications, many enterprises invest in additional, expensive disk storage arrays for an increased degree of physical availability. At the same time, they invest in tape automation solutions that add more levels of data protection, but that only deliver low availability (defined as hours or days to restore a particular pool of data).

In fact, enterprises can segment applications into ones that require high availability and ones that can function with low availability. Those applications, with their accompanying data that can get by with a tape automation solution alone, allow users to avoid the extra investment costs for additional disk storage. Note also that the most critical applications are (and should be) protected by both additional disk arrays and additional tape automation.

Thus, enterprises want and need multiple degrees of data protection. RAID on a production data array provides one degree of physical protection (as the failure of one disk drive can be tolerated without loss of data). A remote mirror can provide a second degree of protection. Where information cannot be lost, a tape solution provides a minimum of one (and generally more, through multiple-generation tape copies) additional degree of protection. Disk does not provide logical data protection; tape does (since the tape is outside the I/O “write” stream that can make logical changes to data). Point-in-time copy capability and its derivatives can provide logical data protection on disk, but require understanding, planning, and investment that many IT organizations have yet to make.

5. IT organizations are going to have to implement active archiving more aggressively. Fixed content stored in active archives has different data protection and data retention requirements than active, frequently-changed data. By implementing Information Lifecycle Management (ILM) and coordinating it with a data protection strategy, enterprises can improve the cost-effectiveness, availability and performance of their storage.

As information in the form of files or records ages, it tends to become fixed data that is unchanging data. That age varies from the

time of creation (e.g., a check entered into the system) to a later time (e.g., closing a transaction in an online transaction processing system). When fixed content data is “distilled” from its active changeable counterparts in an application to an “active archive,” the implications for data protection policies and management are significant.

The traditional backup process is not necessary for fixed content data. A piece of fixed content needs to be replicated after it is captured in an active archive, but no traditional backup process is necessary. Copying the data to a full backup on a regular basis is an unnecessary use of resources since the correct number of data protection copies is already available.

The second major change is the ability to put in place strong data retention policies mandated by governance and compliance requirements. Although data retention policies can be applied to a pool of storage where active changeable data is commingled with fixed content data, data retention management is most effective with a fixed content pool of storage. That is because data retention applies only to fixed content data. An open transaction cannot be disposed of and cannot be considered (at that stage of its lifecycle) to be compliant data, since all compliant data has to be unchangeable.

The migration of data to an active archive will eventually have a significant impact on the active changeable side of the house as well. There will be less data to back up (and restore if necessary), so the burden on the overloaded backup/restore process will be reduced. If critical applications need to be remotely mirrored, the disk space for the remote mirror will be reduced. The upper boundary for fixed content could be as high as 80% or more, but even a movement of 20 to 30% of data could very well have a significant payoff.

6. Focusing on high availability and neglecting the other key objectives of data protection is dangerous.

Too often high availability and data protection are considered synonymous. Data availability is only one of four key objectives for data protection — data preservation, data responsiveness, and data confidentiality are the others. An overemphasis on high availability could lead to underweighting the other objectives. If the necessary amount of data preservation is not in place, high availability of an application will not matter. If the correct controls for data confidentiality are not in place, serious consequences could result. If data responsiveness is not in place, data will not be usable. A sense that all the objectives have to be balanced properly is necessary.

Preface

It is well-known that data protection is a business necessity — yet few agree on exactly what data protection is. And failure to appreciate the full dimensions of the data protection challenge can lead to poor data protection management and costly resource allocation issues. The following example shows some of the difficulties that can arise when enterprises do not have a clear data protection strategy.

Business Continuity and the Backup/Restore Process — Never the Twain Shall Meet?

When asked what words most readily come to mind for “data protection,” the terms “backup/restore” and “business continuity” are likely to top the list. Enterprises clearly understand that all three relate to risk management and that risk management is an essential business task. Very few enterprises, however, understand that improving backup/restore may not improve business continuity. In fact, failure to understand the relationship between the ongoing down-in-the-details task of backup/restore and the global strategy of business continuity may result in unnecessary exposure to risk, under- or over-spending on data protection funding, and wasting of scarce IT administrator resources.

While governance and compliance will also be covered later in the report, let’s start with **business continuity**, the traditional rationale for data protection. Business continuity attempts to prevent any major disruptions to business processes. Thus, business recovery is clearly different from disaster recovery — a concept with which it is often confused. *Disaster recovery* focuses on minimizing the effects of disaster, while *business continuity* focuses both on avoiding unplanned outages (due to either a disaster or an operational problem) in the first place and on minimizing the effects of unplanned outages. Specifically, business continuity emphasizes high availability — defined as restoration of access to applications within seconds or minutes — and resiliency — the ability of applications to continue running despite outages in systems, storage, or underlying software as far as possible. Where high availability and resiliency are not possible, business continuity and disaster recovery share the goal of emphasizing restoration of operational processes as soon as feasible to prevent further loss to the organization.

Now let’s consider **backup/restore**. While backup is performed routinely, restore is only performed when systems are down as a result of an unplanned outage. Inevitably, the focus of backup/restore, like disaster recovery but unlike business continuity, is to minimize the effects of unplanned outages.

Now consider the practical effects of an over-focus on backup/restore rather than business continuity. Any low-to-high availability continuum clearly shows that the backup/restore process with tape is low availability (where low-availability is defined as restoration of data access to applications within hours or days), while technologies such as remote mirroring are high availability. The continued high investment in a low availability backup/restore process in conjunction with tape automation solutions is clearly inconsistent with the desire to move to the higher availability side of the continuum. Moreover, hours of downtime while a restore is taking place can cost customers and threaten the existence of a company. Take, for example, a recent outage of a European discount retailer: had it run longer than two hours, it could have resulted in the loss of millions of euros—a “business-critical situation” (*Progress Fathom: Business Continuity Down to the Details*, June 2005, www.valleyviewventures.com). Yet IT organizations are not likely to replace their current backup/restore processes anytime soon.

The way to avoid the costs and risks of an over-focus on backup/restore is to better understand an enterprise’s overall requirements for data protection. Even though a rip and replace strategy is typically unthinkable, enterprises need to be aware how much and where to place their bets on the data protection roulette wheel today — and those bets will definitely change tomorrow.

The Sea Change in Data Protection

In the last several years, the technology landscape of data protection has fundamentally changed — a true “sea change.” Disk-based backup, compliance/governance technologies, and information life-cycle management (ILM) are examples of the technologies that are affecting where data protection bets should be placed and how much should be bet. The net result is a sea change, a marked transformation.

These new technologies typically reflect new business processes as well. *Disk-based backup* reflects an increasing appreciation of the importance of a business continuity process. *Compliance technologies* reflect the increased importance of meeting regulatory requirements such as Sarbanes-Oxley. *Governance technologies* reflect the increased importance of effectively meeting the changes to the Federal Rules of Civil Procedure. *ILM technologies* reflect a new process that enables finer-grained, more cost-efficient control over an enterprise’s data.

The sea change results in a number of questions for which IT organizations must have answers — about their current data protection infrastructure, and about the direction in which that infrastructure needs to evolve. Among these questions are:

1. How do governance and compliance fit with risk management in presenting a broader picture of data protection than IT typically considers?
2. What is the right target and what are the right objectives for a comprehensive data protection strategy?
3. How are data protection infrastructure holes identified and—if any exist—how are they filled?
4. How are low availability and high availability data protection technologies layered within an overall data protection framework to give sufficient degrees of data protection?
5. How will ILM lead to changes in data protection technologies and strategies?
6. How do all the existing and emerging technologies of the data protection puzzle fit together to help build a roadmap for evolving the data protection infrastructure?

What This Report Is — and Is Not — About

The purpose of this report is to serve as a guide for IT organizations so they are able to more clearly answer these and related questions. This report re-examines the basic principles of data protection in light of all the new demands that are being placed upon the IT infrastructure, and it also looks at how both maturing and emerging data protection hardware and software technologies affect those changes. The framework and model that arises from these basic principles helps put data protection in context to the overall IT infrastructure and helps IT organizations clarify the choices and options that are available to them for data protection.

However, this report is not a buyer's guide — that would require a never-ending encyclopedia! Although representative companies that offer data protection technologies are listed, the suitability — i.e., applying the criteria of scalability, interoperability, resource use, cost, maturity, vendor acceptability, etc. — of each of their products separately and in concert is dependent on the situation and therefore is unique to each reader. What the report does do is to identify and examine the key decisions that should be made and strategies that should be implemented before evaluation of products and services can begin.

Moreover, the report is not a deep dive into the various data protection technologies. It examines current and emerging technologies in relation to an overall framework or approach to data protection. Readers can then better understand how to fit technology options into their overall data protection schema.

Where possible, we conform to the terminology used and directions charted by the Storage Networking Industry Association's Data Protection Initiative (SNIA DPI), so that users do not have to learn new

concepts. Mesabi Group is also working as a member of the SNIA DPI to move the concepts, principles, and technologies of data protection forward. However, since SNIA's work and perspectives on data protection are still evolving there are situations in which this report diverges from SNIA's previous efforts.

This report is not the final word on data protection, but rather intends to arm readers with information that enables them to act more effectively to achieve data protection.

Here is a brief exercise for the reader: before reading further, prepare answers to the following short list of questions:

- What is your view of data protection?
- What are you doing now for data protection?
- What are the issues you currently face regarding data protection?
- What actions, if any, are you planning to improve your data protection processes and infrastructure?

After reading the report, answer the questions again, and compare your answers to the questions before and after. This report will make what is potentially unclear about data protection now, as you start to read the report, obvious after you have finished it.

Chapter One:

The Time Has Come for Change

Studies reveal that data protection — in one form or another — is at the top or near the top of any list of issues facing the management of storage. In the short term, this importance is due to immediate concerns such as “how do I meet regulatory requirements right now.” In the long term, data protection aims to protect the information without which the business cannot function, and which is now a primary source of many enterprises’ competitive advantage. Data protection is therefore a cornerstone of any organization’s management of risk, and risk management is now recognized as one of the fundamental tasks of any enterprise.

Today, data protection is associated primarily with a wide spectrum of IT and business issues:

- Backup and restore
- Disaster recovery
- Business continuity
- High availability
- Data asset preservation
- Compliance
- Governance
- Data privacy
- Data security

Yet today’s IT organizations still tend to focus simply on improving backup/restore processes.

What Data Protection Is

Data protection is the mitigation of the risk of loss of or damage to an enterprise’s data on either a temporary or permanent basis.

Data protection is insurance. Therefore, the aim of data protection is not to maximize profits or revenues, or minimize costs, but to minimize worst-case losses. Like regular insurance, data protection insurance is a necessary cost of the prudent business, and balances the costs of unplanned outages against the costs of the insurance policy. A side-effect of data protection may be more cost-effective use of information assets; but users should not require profits from their data protection solutions, any more than from their life insurance policies on key executives.

Unlike the traditional insurance markets, the data protection market offers no “third-party” insurers (with the possible exception of Lloyd’s of London). Enterprises are “self-insured” today, and should expect to be self-insured tomorrow. Insurance “premiums” are paid internally, in the form of additional hardware, software, and people. One principle remains the same, however — when you pay for data protection insurance, you want to minimize its cost and maximize its value.

“One principle remains the same, however —when you pay for data protection insurance, you want to minimize its cost and maximize its value.”

As we have noted above, data protection seeks to ensure not only the availability of data, but also its confidentiality, privacy, and availability to regulators. This is still insurance — the legal costs of failure to protect confidentiality and privacy, or to fail to supply appropriate information to regulators are high, as are the competitive disadvantages of leaking proprietary information. For example, the attention that is now being turned to “business compliance” has at its heart appropriate protection of data such as e-mails.

Data Protection Has To Be Placed in the Right Framework

IT organizations are actively examining how to improve the data protection function, as shown by an increased interest in disk-based data protection strategies and a number of new replication technologies. Trying to sort through the myriad of choices can be difficult.

The key to choosing any of these strategies and technologies is understanding the overall context, the overall “data protection infrastructure portfolio,” into which individual data protection technologies should fit. Otherwise, what appear to be individually sound decisions may not lead to offering the necessary levels of data protection. Among the problems that can occur are:

- Failure to protect data adequately
- Making the wrong allocation decision (spending too much on areas that do not really require a level of protection and too little on areas that require greater protection)
- Straining the IT administrative resources assigned to data protection even further and with less results than necessary

Without the right framework, enterprises cannot know where to place their longer-term data protection technology investment bets or how much they should place on each bet. And that means that any

framework has to take into account the changing world of data protection technology.

Ride the Sea Change in Data Protection

Change that affects the requirements for data protection is coming from several directions. One of the directions is extending and improving what is already being done. An example of this is disk-to-disk backup.

A second direction is change in the basic way that the movement and storage of information is carried out in an organization. For example, ILM is not only about moving information from one tier of storage to another, but also about managing stored information differently — and a major effect of the difference in information management is in better data protection. Moreover, ILM leads to an overall change in the mix of data protection technologies (e.g., data replication vs. data backup) that are used within an enterprise.

A third direction of change comes about from changing business requirements. A key illustration is a new emphasis on IT business-governance/compliance policies, which require organizations to understand and implement new policies, practices, and procedures as well as possible new hardware and software data protection technologies.

The rest of this report examines the basic principles of data protection in light of these changing business requirements and in light of existing and emerging data protection technologies. The key take-aways that should be kept in mind when reading the rest of this report are:

1. Determine where over-investment and under-investment in data protection technology is taking place, so that your IT organization can direct future investments to shore up the weak spots.
2. Determine what the effects of changing business requirements and technology advances on your enterprise's data protection investment are.
3. Gain a sense of how the major categories of data protection technologies interact, so that you can determine the proper mix and deliver the proper level of service.