



Data Protection

Adapting to the Sea Change

2005 Edition

Report Preview

Full report available for \$495.00 from www.bitpipe.com
or e-mail info@valleyviewventures.com

Mesabi Group



David G. Hill

The Mesabi Group particularly focuses on two “revolutions” that are dramatically changing the ability of storage and information managers to zero in on key data 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 arrival to archiving

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.

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

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 Framework 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: Well Done, 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 Data Protection Objectives Have to Be Met	38
<i>Chapter Five: Data Protection — Getting the Right Degree</i>	39
General Use Classes of Data	39
Tape Is a Special Case	40
Understanding Degrees of Data Protection	40
The Third Degree — Levels of Exposure	41
Mapping Degrees of Protection.....	41
<i>Chapter Six: Information Lifecycle Management Changes the Data Protection Technology Mix</i>	44
Why Data Lifecycle Management Is Not Enough — The Need for Metadata and Management	45
ILM Is Deep Into Logical Pools of Storage	45
Logical Storage Pools at a High Level	46
Moving Information Across Pools — A Distillation Process	47
Archiving Through a New Lens.....	47
Archiving: The Makeover	48
Protecting Archived Data	49
Data Retention.....	50
Disposition of Data	50
Compliance	51
Creating Data Archive Storage Pools by Data Retention Attributes	52
Security and Privacy	54
Active Archiving and Deep Archiving	55
Active Archiving Requires Active Archive Management	56
Long-term Archiving as Part of an Active Archive	57
ILM Changes the Data Protection Technology Mix.....	58
<i>Chapter Seven: Where Data Protection Technologies Play in the New Model</i> .	59
Back to Basics — Extending the Current Model.....	61

Table of Contents

Current RAID Capabilities Are Not Enough.....	61
Evolving Backup/Restore Software.....	61
Better Data Protection through Better Management Reporting and Automation	63
Moving Data Manually and Electronically — the Place of Vaulting	65
At Your Service — the Role of Service Suppliers.....	65
Disk and Tape — Complementing and Competing with One Another	66
Disk-based Backup	67
Virtual Tape	69
Virtual Tape Library	69
MAID	70
Removable Disk Drives and Disk Media	71
Data Protection Appliances.....	71
Tape Automation	72
Getting to the Point.....	74
Point-in-Time Copy	74
Continuous Data Protection	75
Replication Strategies.....	76
Mirroring.....	77
Dated Replication — Pay Close Attention.....	80
Special Requirements for Compliance	81
The Use of WORM Technology.....	82
WORM Tape.....	83
WORM Disk	84
Electronic Locking.....	85
Guaranteeing the Authenticity of Data	85
Privacy and Confidentiality.....	85
Compliance Appliance	85
Mapping the Base Data Protection Technologies to the ILM-Version of the Data Protection Framework	86
<i>Chapter Eight: Summing Up — Redesigning Data Protection.....</i>	<i>89</i>
Data Protection Is Everyone’s Business.....	89
Synthesizing the Data Protection Frameworks.....	89
Guidelines for Data Protection.....	91
The Challenge Ahead and a Call to Action	92
<i>Chapter Nine: Representative Companies.....</i>	<i>93</i>

Table of Contents

Alphabetical Company Listings	96
Mesabi Group Conclusions	116
Glossary	119
Author Profile	123

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	35
Table 4-3: Consequences of Data Loss	38
Table 4-4: Summing Up Key Data Protection Objectives	38
Table 5-1: Sample Degrees of Data Protection for Application <i>n</i>	42
Figure 6-1: The Storage Pyramid — Tiering and Pooling.....	46
Figure 6-2: ILM Changes the Logical Topology Storage Look	48
Figure 6-3: Data Retention Archive Pools	53
Table 6-1: Adding In Archiving to the Data Protection Category Matrix.....	58
Table 7-1: Where Data Protection Technologies Fit in the Data Protection Framework	60
Table 7-2: Base Data Protection Technologies for Active Changeable Data.....	87
Table 7-3: Base Data Protection Technologies for Archived Data	88
Table 8-1: Data Protection Requirements for Application <i>n</i>	90

Table 9-1: Suppliers of Each Data Protection Technology (1).....	94
Table 9-2: Suppliers of Each Data Protection Technology (2).....	95

Executive Summary

Key Report Findings

1. 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.)

2. 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 that applies to all enterprises.

3. 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 in-

vest 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 and 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.

4. Enterprises should implement an overall data protection strategy based on a data protection “framework.” Enterprises are *attacking business continuity, backup and recovery process, and compliance as if they are unrelated problems*, but they really all relate to one another in the context of data protection. The data protection “framework” allows the correct allocation of investment and resources to these three areas, as well as other data protection investments.

Business continuity is a key risk management function of any enterprise. Business continuity is about preventing or ameliorating disruptive impacts on the business that range from threats to survival to productivity drains. One of these disruptive impacts is data loss, and data protection avoids data loss.

Temporary loss of data requires that the data be restored before further disruption can occur. Using *backup and recovery* software is one way of restoring the data.

Compliance data is a special case of data protection to prevent disruptions that would be associated with non-compliance.

Relating all three areas — business continuity, the backup and recovery process, and compliance — through their relationships with data protection is just part of understanding the overall framework of data protection. Understanding that framework is important so that an enterprise can put in place a data protection strategy that takes

these three aspects — as well as many others — into the proper context so that the proper degrees of data protection and the proper levels of investment are in place.

5. 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. 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. Although an enterprise may find it difficult to identify and separate its fixed content data and move it to an active archive, the active changeable side can also benefit when enterprises can find a way to migrate some fixed content data to an active archive. 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. Enterprises should consider their compliance policies in the context of data protection. Compliance is related to data retention, which is part of data protection.

Compliance data is fixed content information in an active archive. Data retention policies can be applied to this active archive. Compliance is simply a more restrictive set of data retention policies, such as chain-of-custody requirements and privacy constraints.

7. 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.

8. Information Lifecycle Management (ILM) actually will play an important role in the IT infrastructure — and data protection is a key part of that role.

Active archives of fixed content require different data protection strategies than for active changeable data. For example, the fixed-content data will use replication upon capture of the data for data protection purposes, and the active changeable data will use a backup/restore process. Moreover, implementation of active archiving requires migration of data between an active changeable pool of storage and an active archive pool of storage. ILM supports both of these tasks, among others.

Preface

It's 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.

Let's start with **business continuity**. 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.

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 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 three years, the technology landscape of data protection has fundamentally changed — a true “sea change.” Disk-based backup, compliance technologies, and information lifecycle management (ILM) are examples of the technologies that are affecting how data protection bets should be placed and by how much. 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. *ILM technologies* reflect a new process that is enabling 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. What is the right target and what are the right objectives for a comprehensive data protection strategy?
2. How are data protection infrastructure holes identified and —if any exist — how are they filled?
3. How are low availability and high availability data protection technologies layered within an overall data protection framework to give sufficient degrees of data protection?
4. How will ILM lead to changes in data protection technologies and strategies?

5. 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 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 dependant 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, conformity to the terminology used and directions charted by the Storage Networking Industry Association (SNIA) have been used so that users do not have to learn new concepts. However, since SNIA's work and perspectives on data protection are still evolving, there are situations in which this report diverges from the current direction that SNIA is taking.

This report is not the final word about data protection, but rather is intended 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 a 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, compare your answers to these 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 restoration
- Disaster recovery
- Business continuity
- High availability
- Data asset preservation
- Compliance
- 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 not only to ensure 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 business emphasis on compliance. IT business-compliance policies require understanding and implementation of 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.