Data Protection — Take a New Look

A Mesabi Group LLC White Paper

January 2005

Mesabi Group LLC
Affiliated with Valley View Ventures
26 Country Lane
Westwood, MA 02090
Telephone: 781-326-0038
www.valleyviewventures.com

Data Protection — Take a New Look

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. Data protection is a cornerstone of any organization's management of risk, and risk management is now recognized as one of the fundamental tasks of any enterprise. 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.

In order to deliver satisfactory levels of data protection, enterprises have to understand 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.

The Role of Data Protection in Business Continuity

Risk management is a key responsibility of any enterprise and business continuity is a key function within risk management. Business continuity is the mitigation of risk caused by interruption to normal enterprise activities and processes.

A key task of any business continuity strategy is data protection; and, by the same token, a key aim of data protection is business continuity. Furthermore, a business continuity strategy and architecture can serve as a good framework into which to fit data protection technologies and strategies. It is comprehensive; it ensures that the needs of other parts of the architecture aside from storage and the business as well as IT are taken into account; and it fully recognizes the crucial role of information storage.

To understand why enterprises may not be receiving the level of data protection that they think they are requires an understanding that business continuity is not only about disaster continuity (more familiarly, thought of as disaster recovery), but also operational continuity — the ability to deal with day-to-day operational problems. The right amount of attention for data protection has to be given to each — and that may not always be the case.

Both operational and disaster continuity require the proper level of both physical (storage device level) and logical (the data itself) data protection. A data item may be flawed although the disk is functioning perfectly; a disk may crash but the same data on a different disk be preserved.

Note that an event is considered a "disaster" only when data processing has to be moved from a primary to a secondary site and when that processing is carried out using a different set of computer hardware (including both servers and storage).

Operational continuity and disaster continuity each need a different mix of data protection technologies to achieve the planned levels of data protection that an enterprise requires. Yet enterprises may not have a clear understanding of the differences between physical and logical data protection — and that may result in a dangerous lack of attention to logical data protection.

A logical data protection problem can affect a key application, whether the application crashes or not. The inability to dispense cash from an automated teller machine or the inability to correctly deliver the right goods to a customer in a timely fashion can affect an enterprise's credibility (and market valuation).

Do Not Neglect Any Facet of Data Protection

No aspect of data protection can afford not to be protected. The target for the IT organization starts with four simple boxes (Table 1). Both operational continuity and disaster continuity have a physical and a logical component to them. Each box has to be considered individually and all four boxes together have to be considered collectively to devise a data protection solution that meets an enterprise's needs.

Table 1: Data Protection Category Matrix

	Operational Continuity	Disaster Continuity
Physical		
Logical		

Source: Mesabi Group, January 2005

Although it seems simple, filling in the matrix is not that easy. The first challenge is in knowing when the levels of data protection are enough. The second challenge is in understanding that the target is moving and knowing how that will affect what needs to go into the matrix to get the right levels of data protection.

Availability Is Not the Only Data Protection Objective

The driving goal is to have *data always available securely, with optimal performance, to authorized users anywhere via any connection on any device*. Availability is certainly critical to obtaining that goal, but from a data protection perspective there are really four objectives that are part of and have to be met in working toward it:

- *Data preservation* — data must be consistent and accurate all the time; data must also be complete within acceptable limits

Telephone: 781-326-0038
26 Country Lane
Westwood, MA 02090
www.valleyviewventures.com

- *Data availability* ─ the ability of I/O requests to reach a storage device and take the appropriate action
- *Data responsiveness* ─ the ability of I/Os to deliver data to an authorized user according to measures of timeliness that are deemed appropriate for an application
- *Data confidentiality* — data is available only to those authorized

Note that data availability is not the same as data preservation. Not all preserved data needs to be immediately accessible. It may take a month to get some historical records back from the tape warehouse for discovery during a legal proceeding, but a month is adequate time. Not all data that needs to be accessed quickly for business intelligence needs to be preserved — in some cases, financials can be quickly reconstituted from sales and other data if the financial spreadsheet is lost.

Job one in data protection is the preservation of digital assets. RPO (recovery point objective) states the amount of time (say seconds, minutes or hours) back to where a recovery is attempted. RPO specifies what the acceptable level of data loss is (in seconds, minutes, hours, or days). RPO should be negotiated between the user and the IT group. Quite frankly, RPO will generally be zero for most applications irrespective of RTO (recovery-time objective) requirements. RTO is the time (for example, seconds, minutes, or hours) that it takes to restore an application to an operational state. For examples, to return to the A/P example, most CFOs would not rate the RTO of accounts payable as being very high at all, but, while they might like to, would probably not agree to any permanent loss of data. (In a litigious society, creditors might object to accounts receivable having an RPO of zero [highly likely] and accounts payable having a non-zero RPO, which means that they might not get paid unless they complain.)

Understanding Degrees of Data Protection

Data protection comes in degrees (also can be thought of as layers). The first degree where data protection can be provided is for the primary copy. The primary copy may or may not have data protection. If it does, that is the first line of defense for operational continuity. Built-in data protection to the primary data copy can help prevent service-level threatening events (such as a single disk failure) from becoming service-level-negative-impacting events.

However, add-in data protection cannot provide disaster continuity protection and the risk-protection diversification that is necessary for operational continuity protection. At least one add-on copy — a full copy of the data that is physically separate and distinct from the original — is necessary.

Call one layer of added-in or added on data protection as one degree of protection. One degree of data protection means that one failure is tolerable; data is recoverable. If a failure should occur, data protection is at zero degrees. Zero degree of data protection means no more failures can be accommodated without total and permanent data loss. This is a level of exposure that IT organizations find unacceptable.

That is why additional degrees of data protection are necessary (Table 2). The question is how many. The minimum number of layers is two. If one failure occurs, the degrees of protection are down to one. Given that technology is not perfect; having only one extra degree of freedom to fall back upon is not advised. So three degrees of data protection is probably a minimum. Each additional layer adds expense, but one or more additional layers may still justify the expense.

Table 2: Sample Degrees of Data Protection for Application *n*

	Operational Continuity	Disaster Continuity
Physical	**Higher Availability** Degree 1: Local replication Degree 2: Remote replication **Lower Availability** Degrees 3-5: Tape/ disk-based backup	**Higher Availability** Degree 1: Remote replication **Lower Availability** Degree 2-4: Vaulted tapes
Logical	**Higher Availability** Degree 1: Point-in-time copy Degree 2: Continuous data protection **Lower Availability** Degree 3-5: Tape/disk-based backup	**Higher Availability:** Degree 1: Remote dated replication **Lower Availability** Degree 2-4: Vaulted tapes

Source: Mesabi Group January 2005

IT administrators should map out the degrees of data protection for each application. The degrees of protection have to be split between higher availability and lower availability degrees. Once the higher availability degrees are exhausted, availability depends upon the lower degree availability options. Note that the term "lower availability" should not be considered a pejorative term, but rather reflect the relative difference between the time-based ability of different technologies to restore information.

ILM Changes the Data Protection Technology Mix

Critical to understanding ILM is that every piece of data becomes fixed (i.e., read-only) at some time during its lifecycle — and that time is typically short as compared to the full length of its lifecycle. Active changeable data reflects a creation and change process where viewing the data at different times would reveal that the data had not stayed the same. At some time change ends. Even an online transaction processing system (OLTP) updating customer records create data that must be "frozen" after a certain period of time. An e-mail is information that is fixed upon capture (as replies do not change the e-mail itself). If an IT organization looks carefully at the data managed under its custodial care, a large percentage of the data will probably be fixed.

Understanding the concept of the bifurcation of production data into two separate and distinct classes — active changeable data and fixed content data — has achieved some measure of mindshare in IT organizations. The term that is on its way to achieving the greatest popularity and acceptance for the storage of fixed content is "archiving." In dealing with archiving, ILM fundamentally divides the storage infrastructure into two separate halves — and active changeable side and an active archive side. The addition of an active archive for fixed content information changes the data protection category matrix (Table 3). The reason is that some of the data protection strategies for active archive are different for active archived information than for active changeable information, such as not necessarily requiring the use of backup/restore software, but rather making dated replicates of the data.

As another example, an application may have data on both the active changeable and the active archive side of the house. That might mean that the RPO and RTO for each side would be different. For example, active transactions in an OLTP may require a different (and probably more stringent) RPO and RTO than closed transactions that are retained for business intelligence purposes.

Table 3: Adding in Archiving to the Data Protection Category Matrix

	Operational Continuity		Disaster Continuity	
	Active Changeable	Active Archive	Active Changeable	Active Archive
Physical				
Logical				

Source: Mesabi Group January 2005

The finer granularity that is expressed in the doubling of the cells in the matrix requires more work on the part of an IT administrator to fill out, but also permits the design of more effective data protection strategies.

Cornucopia of Data Protection Technology Choices

Enterprises have a number of choices for the technologies that can fit into the ILM-version of the framework (Table 4). No single taxonomy for data protection seems suitable. Data protection technologies are not always purely for one task or function; there may be a lot of blending, morphing, blurring, and variations in the data protection functionality that any individual product may contain. The focus is therefore on overall technologies and not specific products. Extract the essence in terms of what function is being performed, where the technology will fit in the framework, and what it adds to the overall degree of protection. Individual products can then be evaluated offline in terms of how they fit one or more of the data protection needs.

Table 4: Where Data Protection Technologies Fit in the Data Protection Framework

	Operational Continuity		Disaster Continuity	
	Active Changeable	**Active Archive**	**Active Changeable**	**Active Archive**
Physical	RAID Cloned Point-in-Time Copy Tape Automation* Virtual Tape Library* Continuous Data Protection Data Protection Appliance*	RAID Dated Replication WORM Tape	Synchronous Remote Mirroring Asynchronous Remote Mirroring Dated Replication Vaulting* Electronic Vaulting*	Dated replication Vaulting
Logical	Point-in-Time copy Tape Automation* Virtual Tape library* Continuous Data Protection Data Protection Appliance*	WORM Disk Guaranteed Uniqueness Electronic Locking Dated Replication WORM Tape Compliance Appliance	Dated Replication Vaulting* Electronic Vaulting*	Dated Replication Vaulting

*Backup/restore software is or might be used in conjunction with this technology.

Source: Mesabi Group January 2005

Storage Suppliers for Key Data Protection Technologies

Storage vendors were asked to list the areas (by specified key words) where they delivered a product or service (Table 5-1 and Table 5-2). Fifty-one vendors who responded to the request are listed in the table. (That number represents nearly 75% of the vendors who were asked to supply inputs.). Each vendor may have its own interpretation of what each keyword actually means.

Table 5-1: Suppliers of Each Data Protection Technology (1)

DP Technologies	Suppliers
General Technologies	
RAID	Adaptec, ADIC, Archivas, BlueArc. Crossroads, EMC, EqualLogic, HDS, HP, IBM, InoStor, LeftHand Networks, NetApp, Spectra Logic, StorageTek, Sun, Unitrends, Xiotech, Zetta Systems
Backup/restore software	Adaptec, Atempo, Avamar, BakBone, Diligent, EMC, ExaGrid, FalconStor, HDS, HP, IBM, InoStor, NetApp, NSI, Revivio, Signiant, Storactive, Sun, Unitrends, VERITAS, XOsoft, Zetta Systems
DP management software	EMC, ExaGrid, HDS, HP, IBM, NetApp, Onaro, OuterBay, StorageTek, SysDM, Unitrends, VERITAS, Zetta Systems
Vaulting	BlueArc, Data Domain, Diligent, EMC, IBM, Iron Mountain, Spectra Logic, SunGard, Unitrends, VERITAS
Electronic vaulting	Archivas, BakBone, BlueArc, Data Domain, EMC, ExaGrid, FalconStor, IBM, Iron Mountain, SunGard, VERITAS
Data protection services	Atempo, BakBone, Cambridge Computer Services, Data Domain, EMC, FalconStor, HDS, IBM, InoStor, Iron Mountain, NetApp, OuterBay, Spectra Logic, StoreAge, StorageTek, SunGard, VERITAS
Disk and Tape Complementing and Competing	
Disk-based backup	Adaptec, ADIC, Avamar, BlueArc, Data Domain, Diligent, EMC, EqualLogic, ExaGrid, FalconStor, HDS, HP, IBM, InoStor, LeftHand Networks, Neartek, NetApp, Overland Storage, Quantum, SEPATON, Signiant, Spectra Logic, Storactive, StorageTek, Unitrends, VERITAS, Xiotech, XOsoft, Zetta Systems
Virtual tape	Crossroads, Data Domain, EMC, Diligent, FalconStor, IBM, InoStor, Neartek, Overland Storage, Quantum, SEPATON, Spectra Logic, StorageTek
Virtual tape library	ADIC, Atempo, BakBone, COPAN Systems, Crossroads, Data Domain, Diligent, EMC, FalconStor, IBM, Neartek, NetApp, Overland Storage, Quantum, SEPATON, Spectra Logic, StorageTek
MAID	COPAN Systems, Spectra Logic
Removable disk media	IBM, Spectra Logic, Unitrends
Data protection appliance	Archivas, Avamar, Crossroads, Data Domain, Diligent, EMC , FalconStor, IBM, InoStor, Mendocino Software, NetApp, Overland Storage, Quantum, Revivio, SEPATON, Spectra Logic, StoreAge, StorageTek, Unitrends, Xiotech
Tape automation	ADIC, BakBone, Diligent, EMC, HP, IBM, InoStor, NetApp, Overland Storage, Qualstar, Quantum, Sony Electronics, Spectra Logic, StorageTek, Sun, Unitrends

Source: Mesabi Group January 2005

Table 5-2: Suppliers of Each Data Protection Technology (2)

DP Technologies	Suppliers
Point-in-time-related	
Point-in-time copy	Adaptec, Atempo, Avamar, BakBone, BlueArc, Crossroads, EMC, EqualLogic, FalconStor, HDS, HP, IBM, InoStor, LeftHand Networks, NetApp, NSI, Signiant, StoreAge, StorageTek, Sun, Unitrends, VERITAS, Xiotech, Zetta Systems
Continuous data protection	BlueArc, Crossroads, Data Domain, ExaGrid, FalconStor, HDS, IBM, InoStor, LeftHand Networks, Mendocino Software, NetApp, NSI, OuterBay, Revivio, Storactive, StoreAge, StorageTek, TimeSpring, VERITAS, Xiotech, XOsoft, Zetta Systems
Replication-related	
Asynchronous remote mirroring	Adaptec, Avamar, BlueArc, EMC, FalconStor, HDS, HP, IBM, InoStor, Kashya, LeftHand Networks, NetApp, NSI, Permabit, Revivio, Signiant, StoreAge, Sun, Unitrends, VERITAS, Xiotech
Synchronous remote mirroring	BlueArc, EMC, FalconStor, HDS, HP, IBM, Kashya, NetApp, Revivio, StoreAge, Sun, VERITAS, Xiotech, XOsoft
Dated replication - local	Avamar, Crossroads, EMC, EqualLogic, ExaGrid, FalconStor, HDS, HP, IBM, InoStor, Kashya, LeftHand Networks, NetApp, Permabit, Revivio, StoreAge, StorageTek, Topio, VERITAS, XOsoft, Zetta Systems
Dated replication - remote	Avamar, Crossroads, EMC, EqualLogic, ExaGrid, FalconStor, HDS, HP, IBM, InoStor, Kashya, LeftHand Networks, NetApp, Permabit, Revivio, Signiant, StoreAge, StorageTek, Topio, Unitrends, VERITAS, Xiotech, XOsoft, Zetta Systems
Compliance-related	
WORM disk	Archivas, EMC, FalconStor, HDS, IBM, NetApp, Permabit, Princeton Softech, Spectra Logic
WORM tape	ADIC, BlueArc, HP, IBM, Overland Storage, Qualstar, Quantum, Princeton Softech, Sony Electronics, Spectra Logic, StorageTek
Compliance software	Archivas, EMC, HDS, HDS, IBM, NetApp, OuterBay, Permabit, Princeton Softech, Revivio, VERITAS
Compliance appliance	Archivas, Crossroads, EMC, HP, IBM, InoStor, NetApp, Permabit, Revivio, Spectra Logic, Sun

Source: Mesabi Group January 2005

Mesabi Group LLC Conclusions

Business as usual for data protection is not an option for IT organizations. For each application, IT organizations have to figure out their:

- Data protection objectives
- Necessary degrees of data protection
- The data protection technologies that can best be used to fill each box in the data protection category framework

That is a real challenge, but the rewards — attaining the proper level of data protection, wisely using scarce data protection budget dollars, and getting the best out of everyone who has a role in data protection — is worth the effort.

Mesabi Group LLC
26 Country Lane
Westwood, MA 02090
USA
Telephone: 781 326 0038
www.valleyviewventures.com

Mesabi Group LLC is affiliated with Valley View Ventures (V3). V3 provides thought leadership and sound advice to both vendors and users of information technology from leading independent industry analysts.

This document is the result of research performed by Mesabi Group LLC, which was not sponsored or underwritten. Mesabi Group LLC believes its findings are objective and represent the best analysis available at the time of publication.

Email the analyst:
david.hill@valleyviewventures.com