

# Data Protection and Recovery in the Small and Mid-sized Business (SMB)

---

*An Outlook Report from Storage Strategies NOW*

By Deni Connor, Patrick H. Corrigan and James E. Bagley

Intern: Emily Hernandez

October 11, 2010

Storage Strategies NOW

8815 Mountain Path Circle

Austin, Texas 78759

Note: The information and recommendations made by Storage Strategies NOW, Inc. are based upon public information and sources and may also include personal opinions both of Storage Strategies NOW and others, all of which we believe are accurate and reliable. As market conditions change however and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. Storage Strategies NOW, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.

This report is purchased by Geminare, who understands and agrees that the report is furnished solely for its internal use only and may not be furnished in whole or in part to any other person other than its directors, officers and employees, without the prior written consent of Storage Strategies NOW.

Copyright 2010. All rights reserved. Storage Strategies NOW, Inc.

**Sponsor**



## Table of Contents

Sponsor .....	2
Introduction .....	5
The Small and Medium Business Market .....	5
Size of Market by Revenue and IT Spending .....	5
Importance of Data Protection and Business Continuity Software .....	5
SMB Unique Requirements .....	6
Growing Data Retention Demands .....	6
Technology Availability .....	6
US SMB Businesses and Revenue by Size (SBA, 2007 data) .....	6
The North American Industrial Classification System (NAICS) .....	7
How To Reach the SMB Market .....	8
SMB Sectors Requiring Large Amounts of Data .....	8
Energy exploration and operations for oil and natural gas .....	8
Mining operations other than oil and gas .....	8
Motion picture and video production .....	8
Data processing, hosting and related services .....	8
Software publishers .....	9
The financial industry .....	9
Legal services .....	9
Accounting, tax preparation, bookkeeping and payroll services .....	9
Architectural, engineering and related services .....	10
Computer systems design and related services .....	10
Research and development in physics, engineering and life sciences .....	10
Healthcare .....	10
Managed Service Providers (MSPs) .....	11
Data Protection Technologies .....	12
Backup to Tape .....	12
Virtual Tape Library (VTL) .....	13
Disk-to-Disk-to-Tape (D2D2T) .....	13
Backup to Fixed Disk or Array .....	13
Backup to Removable Disk .....	13
On-Line Backup .....	14
Online Backup Issues .....	15
Methods for Backing up Data .....	15

File Synchronization .....16

Remote Data Replication .....16

Images, Clones and Snapshot Images.....17

Continuous Data Protection and Near Continuous Data Protection.....18

Agent vs. Agentless Backup.....18

Windows Volume Shadow Copy Service (VSS).....18

Encryption and Password Protection of Backup Media .....19

    Tape Drive-based Encryption.....19

    Encryption Issues .....19

    Backup Data Compression .....19

Data Deduplication ..... 20

    File Mode and Block Mode..... 20

    In-Line or Post-Processing Deduplication ..... 20

    Source or Target Deduplication..... 20

    The Downsides of Data Deduplication ..... 21

Application-Specific Backup .....21

Virtual Machine (VM) Backup ..... 22

Backing Up Virtual Machines ..... 22

Hypervisor-specific Backup Methods ..... 23

- Microsoft Hyper-V ..... 23
- KVM, VirtualBox, Xen, XenServer and Others ..... 23

Tips and Best Practices for Effective Backups ..... 24

Customer Name: Sprott Asset Management ..... 25

Vendor Name: Geminare ..... 27

Tables of Geminare Features ..... 29

## **Introduction**

The data protection and recovery space is exploding as more businesses recognize that protecting their assets – their information -- is key to business survival. Small and mid-sized businesses are a market that has been underserved by data protection software, appliances and online backup services until the last few years. Yet, these organizations have the same needs as large enterprises to protect their data. SMBs, unlike large enterprises, though are faced with a number of unique challenges.

Providing full-time dedicated IT resources may be beyond their means and paying for that IT help and for the software to manage their data may quickly overwhelm them. They often turn to managed service providers or value-added resellers to manage their infrastructures or to supplement the IT skills they have.

Now, there are many software packages, appliances, target arrays (which have integrated snapshot and replication capabilities) and services available to SMB customers that provide data protection and recovery. This survey addresses most of them.

## **The Small and Medium Business Market**

For this survey we analyzed companies with at least one paid employee but less than one thousand employees. In the United States alone, there are approximately 5.75 million firms in this category and only about 13,000 firms with one thousand or more employees. Worldwide, SSG-NOW estimates there are more than eight million firms in this category. In addition, many governmental units, be they departments of larger organizations or typical municipalities, have similar IT requirements of the SMB.

The SMB market can't be characterized solely by the number of employees an organization has. We talked to many SMBs that while they have few employees, have storage capacities under management that may surprise the casual observer. Their ability to consume storage varies widely from 500GB at the low-end to 100TB at the high-end. In some instances such as video post-production, the data can grow into the petabyte range just in the manufacture of a single movie. The amount of data growth SMBs are experiencing is growing at a pace that doubles every 18 months.

### **Size of Market by Revenue and IT Spending**

All US companies had revenue of about \$30 trillion in 2007. SMBs accounted for \$13 trillion in revenue that year, which is the most recent data set available. Despite the effects of global recession, worldwide IT spending by SMBs was about \$575 billion in 2009 and is estimated to grow to \$630 billion by 2014.

### **Importance of Data Protection and Business Continuity Software**

Data protection has become the highest priority for IT spending in the SMB according to surveys conducted in 2010. This represents a shift as SMB executives realize how computer-centric their organizations have become. In recent times, data protection was viewed as expensive insurance against events that could not easily be predicted and costs of data loss were unknown. But organizations of all sizes now realize that the loss of access to data directly affects their ability to operate. The recommended allocation of IT budget to this critical function ranges from 5% to 10%, depending on the type of business. Worldwide, SMB spending on data protection is estimated at \$30 billion to \$60 billion in 2010.

### SMB Unique Requirements

SMBs have unique requirements and challenges when compared to their larger counterparts. First, IT resources are minimal and often performed by the proprietor or staff members that have many other responsibilities in the firm. We talked to one SMB, whose administrator was also responsible for human resources and finance, as well as a myriad of other miscellaneous responsibilities.

Further, while infrastructure is often limited to a number of desktop or laptop clients and perhaps a few dozen servers, technology available to these organizations is second to none and the ability to adopt new equipment, often at lower cost and better performance, is usually easier than in large organizations that are not as nimble because they need to move technology forward en masse. Purchasing decisions are likely to be quicker due to fewer people involved. And one data loss experience is usually enough to justify acquisition of data protection and business continuity products. With the reliance upon technology within virtually every business endeavor, data loss experiences happen at an ever increasing rate.

### Growing Data Retention Demands

One thing SMBs have in common with their larger counterparts is the explosive growth in data storage requirements. Certain business segments have higher capacity requirements, for example, healthcare providers, law firms and financial organizations. But all organizations have data retention requirements for accounting information and increasing governmental reporting demands. The data must be retained and available for long periods, often, as in the case of electronic health records, forever.

### Technology Availability

Changing technology can be rapidly adopted by SMBs. Simple tape backup systems, the norm of a decade ago, are now being replaced by low-cost drive arrays and even small organizations are adopting virtualization, replication, mirroring and deduplication technologies. Low-cost bandwidth supplied by cable and telecommunication providers allows the delivery of cloud-based data protection services to home and small offices. The use case of local data storage appliances that automatically back up to cloud storage are becoming widely deployed as organizations realize that offsite data retention can be transparent and automatic to their operations, as opposed to an expensive, problem-prone effort.

### US SMB Businesses and Revenue by Size (SBA, 2007 data)

Size	Firms	Establishments	Employees	Revenue (x1000)
<b>Total</b>	<b>6,049,655</b>	<b>7,705,018</b>	<b>120,604,265</b>	<b>\$29,746,741,904</b>
<b>0-4</b>	3,705,275	3,710,700	6,139,463	\$1,434,680,823
<b>5-9</b>	1,060,250	1,073,875	6,974,591	\$1,144,930,232
<b>10-14</b>	425,914	444,721	4,981,758	\$791,709,665
<b>15-19</b>	218,928	237,689	3,674,424	\$603,788,766
<b>20-24</b>	134,254	152,547	2,928,296	\$489,530,870
<b>25-29</b>	89,643	106,623	2,405,637	\$402,007,359
<b>30-34</b>	64,753	81,086	2,063,987	\$364,392,992

Size	Firms	Establishments	Employees	Revenue (x1000)
<b>35-39</b>	47,641	62,878	1,754,582	\$304,339,758
<b>40-44</b>	38,221	51,847	1,600,913	\$293,476,569
<b>45-49</b>	29,705	43,325	1,391,754	\$249,407,544
<b>50-74</b>	86,364	139,864	5,195,105	\$979,545,562
<b>75-99</b>	41,810	85,215	3,582,686	\$710,220,323
<b>100-149</b>	39,316	102,135	4,749,055	\$967,245,234
<b>150-199</b>	18,620	66,602	3,205,201	\$674,337,913
<b>200-299</b>	17,780	87,923	4,309,143	\$897,848,746
<b>300-399</b>	8,155	55,515	2,808,347	\$595,711,397
<b>400-499</b>	4,715	43,678	2,101,982	\$476,906,931
<b>500-749</b>	6,094	71,702	3,695,682	\$800,475,934
<b>750-999</b>	2,970	45,990	2,561,972	\$636,199,229
<b>0-999</b>	<b>6,040,408</b>	<b>6,663,915</b>	<b>66,124,578</b>	<b>12,816,755,847</b>

## The North American Industrial Classification System (NAICS)

The NAICS is maintained by the Census Bureau as a way to classify businesses into sectors. The following are major classifications with subsectors defined under each. Sectors shown in bold are included in this report as containing SMBs with the largest amount of data per employee.

### Sector Description

11	Forestry, Fishing, Hunting and Agriculture Support
<b>21</b>	<b>Mining</b>
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
<b>51</b>	<b>Information</b>
<b>52</b>	<b>Finance and Insurance</b>
53	Real Estate and Rental and Leasing
<b>54</b>	<b>Professional, Scientific and Technical Services</b>
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
<b>62</b>	<b>Health Care and Social Assistance</b>
71	Arts, Entertainment and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
99	Unclassified

## How To Reach the SMB Market

With upwards of \$60 billion in annual data protection spending, many technology companies are specifically targeting the SMBs with data protection and business continuity products. Due to the large number of organizations, marketing to millions of SMBs is distinctly different than targeting the Global 5000 enterprises. Mass media marketing by software-as-a-service providers MozyPro and Carbonite is an example. Much of the delivery of technology to this market is performed by managed service providers (MSPs), effectively, outsourced IT groups. Data protection software is often bundled with storage systems, ranging from small network attached storage (NAS) filers to large, purpose-built systems with all of the advanced features used by large enterprises. How do these systems apply to SMB's? The answer is simple. While the SMB market is defined as fewer than 1,000 employees, the amount of storage that needs protection varies wildly based on the type of business. As a result, vertical marketing, along with channel partner recruitment, is a critical factor in reaching the gold in the SMB market. A small radiology practice may only have a dozen employees but will often be managing terabytes of data that doubles every 18 months and needs to be retained forever.

## SMB Sectors Requiring Large Amounts of Data

As we focus our research on the businesses with fewer than 1,000 employees, we find a disconnect between the amount of storage being protected and the number of employees. This is because the sector the business serves is a bigger predictor of storage requirements than the number of employees. We find that the small business segments that have large amounts of data include energy exploration and extraction, engineering, healthcare practices, law firms and motion picture/video production, to mention a few.

### Energy exploration and operations for oil and natural gas

6,430 firms are involved in oil and natural gas extraction (NAICS 211111). 1,950 firms perform oil and natural gas drilling (NAICS 213111). 6,880 firms are involved in support services for oil and gas operations (NAICS 213112). These SMBs consume large amounts of data in the analysis of geologic information, engineering and equipment design data, well design and mapping, production volume and management of flows and depletion. Regulations require significant data retention periods, in many cases, permanent retention, as well as continuous reporting to various governmental entities for safety, revenue, environmental and mapping purposes.

### Mining operations other than oil and gas

4,440 firms are involved in mineral extraction other than oil and gas (NAICS 212). Another 680 firms are involved in support of mining operations (NAICS 213113, 213114, 213115). These SMBs have similar data set analysis, retention and reporting requirements to the oil and gas industries.

### Motion picture and video production

Motion picture and video production, perhaps surprisingly, is dominated by SMBs. Some 12,300 firms have fewer than 1,000 employees while only 45 firms employ more than 1,000 (NAICS 51211). Another 2,015 firms are involved in post production (NAICS 51219), which has the highest storage requirement per employee because these firms are dealing with editing the full content. The advent of high-definition video and three dimensional productions has multiplied the amount of storage required during production and post production. Additionally the cutover from analog to nearly 100% digital content has hit this sector like a tidal wave.

### Data processing, hosting and related services

7,280 firms in this sector are SMBs (NAICS 5182). These include the Managed Service Providers who, in many cases, are the IT departments for the majority of SMBs. These organizations are further critical to the data protection market as resellers, recommenders or providers of data protections equipment and services. The growing amount of cloud-based storage services are accessed through this sector.

## Software publishers

6,000 software publishers (NAICS 5112) with fewer than 1,000 employees have special data protection requirements including revision control and maintenance of large test case databases.

## The financial industry

The financial industry, while including many huge banking institutions, is made up of many thousands of SMBs. Due to the transactional data protection, as well as retention requirements, these organizations all have very large data sets that must be protected from data loss, corruption and theft.

### *Credit intermediation and related services*

This large segment includes all banking, savings and lending services. Approximately 67,500 firms fall into the SMB category (NAICS 522). Transactional data is critical to all companies. It is not acceptable to have any loss of data in this area. Inability to access transactional data may be acceptable for brief periods, but loss of data is fatal. In addition, all data must be retained for indefinite periods, and most firms keep all transactional data permanently. In addition, support data such as e-mails and other communications are subject to regulatory immutability -- that is, any communication must be maintained in case of inquiry or litigation. Finally, all documentation related to lending has come under additional regulation during the last year. This creates a tremendous data protection requirement for all firms in this sector.

### *Securities intermediation and related services*

54,500 firms involved in the brokerage of stocks, bonds and other financial instruments are included in this sector (NAICS 523). Since the Sarbanes-Oxley Act of the early 2000s, and recent enactments under extended financial regulation, transactions and communications are under increasingly strict regulatory control. Additional scrutiny extends to executive compensation and communications regarding all security transactions. This creates additional stress in this sector in terms of data protection and retentions.

### *Insurance carriers and related services*

67,000 SMBs are involved in the insurance business (NAICS 524). From simple insurance agents to large-scale fiduciary activity, the sector has come under significant new reporting and financial allocation requirements. The new regulations in health insurance create additional reporting and data management requirements. Unless significant changes to existing legislation occurs, this segment will be adding storage and protection at levels of an order of magnitude over prior periods. Large data sets also include research functions and actuarial data.

### *Funds, trusts and other financial vehicles*

2,100 SMBs are involved in the management of trusts, mutual funds and other financial instruments (NAICS 525). These firms have similar transactional, data retention and regulatory reporting requirements.

## Legal services

185,000 SMBs are involved in legal services (NAICS 5411). Data protection and security are extremely critical in law firms. Extensive access to online research has replaced the traditional law library. Recent enhancements to tool sets related to electronic discovery and immutable archiving have increased the amount of data managed by law firms.

## Accounting, tax preparation, bookkeeping and payroll services

106,200 SMBs are involved in accounting and payroll services (NAICS 5412). Data protection is critical to these organizations that have regulatory requirements for long-term record retention.

### **Architectural, engineering and related services**

100,000 SMBs are involved in architectural and civil engineering (NAICS 5413). Permanent data retention and fast access to the data is critical in this field. Often multiple offices of these firms require simultaneous access to this information. Computer aided design (CAD) data represents very large data sets with access and revision control as major application requirements.

### **Computer systems design and related services**

99,600 SMBs are involved in computer systems design and services (NAICS 5415). These firms have extensive data requirements for CAD files, computer program source files, test and simulation data and development support data.

### **Research and development in physics, engineering and life sciences**

10,900 SMBs perform research in these areas (NAICS 54171). Huge databases including genomes, pharmaceuticals and theoretical simulations are required for basic research. This data is often modified and updated with associated revision control and results derivatives. Data protection is critical to the field, with many regulatory aspects related to field testing and trial results.

### **Healthcare**

The American Reinvestment of Recovery Act of 2009 (ARRA), often referred to as the 'stimulus plan' created a fund in excess of \$35 billion to fund new technology for healthcare providers of all types. Along with the large source of funding came new requirements for data retention and security of data against breach of personal data. The Affordable Care Act of 2010 added a number of new regulations that directly affect information technology in this sector. New diagnostic equipment generates huge data sets that must be retained within electronic health records (EHR) permanently. The net result is an exponential increase in the amount of data that must be retained and protected.

#### ***Offices of physicians***

190,500 SMBs make up the vast majority of physician practices in the US (NAICS 6211). These firms are under the same regulations, incentives and data retention requirements as those of the hospital system, generally without the benefit of information technology employees. In addition to an array of specialized medical equipment that generates large amounts of data, physicians are becoming more computer-centric in all areas, including EHR, billing and even prescription writing. A major requirement of EHR compliance under the ARRA is computerized prescription order entry (CPOE) which will require automation far beyond the simple scribbling of a prescription onto a piece of paper and sending it off with the patient.

#### ***Outpatient care centers other than family planning and substance abuse***

Some 8,200 firms are involved in this sector (NAICS 62149) which includes surgical, HMO, dialysis and emergency care centers.

#### ***Medical and diagnostic laboratories***

7,500 medical and diagnostic laboratories are SMBs (NAICS 6215). Huge data sets are generated by these organizations and are subject to the same type of regulatory considerations as all organizations involved in EHR generation.

#### ***Acute care hospitals***

We do not consider the 4,000 Acute Care Hospitals to be in the SMB market (NAICS 622). While some individual hospitals may fall into the category, even small hospitals are generally managed by larger organizations with IT staffing and centralized support.

### **Managed Service Providers (MSPs)**

Managed Service Providers are major resources for IT support to the SMBs. Ranging from a few employees to large regional and national entities, MSPs provide hardware and software recommendations, resale of equipment and software and often provide hosted data center support. They are a major channel for data protection services to the SMB market.

## Data Protection Technologies

### Backup to Tape

The traditional method of system backup has been file-by-file backup to tape. Typically a tape rotation scheme is used that provides backups at different points in time. Tape has suffered from a number of problems:

- A multiplicity of tape and data formats. Numerous tape formats of varying capacities have been and are being used. These formats are incompatible with each other, so moving from one tape technology over time or to one tape library to another can be an expensive process. In addition, backup software vendors have often used their own proprietary logical data formats (similar to the way different word processors, such as Word and WordPerfect, use different formats), which further compound the problem. In both cases you must have the same type of tape drive and often the same software to restore a tape. The recent development of the Linear Tape File System (LTFS), a standardized file system for LTO-5<sup>1</sup> tape, should help alleviate the compatibility issues to some degree, assuming the standard is widely accepted.
- Reliability issues. Over the years tape has suffered from reliability issues, both with drives and media. It is not uncommon for a tape drive to require repair or replacement within three years. Although newer tape technologies, such as LTO, have improved tape reliability, media issues are still all too common. In addition, tape drive read/write heads must be cleaned on a regular basis to maintain reliability. Tape, libraries add additional mechanical components that can fail as well and require replacement.
- Cost. The cost per megabyte of tape media has dropped considerably over the years, but it has not kept pace with the drop in the cost of disk media. In addition, the cost of the tape drive itself is relatively high. Internet pricing on LTO-4 drives (800GB native capacity) is about \$2,500-\$4,000, while LTO-5 drives (1.5TB native capacity) sell for about \$3,500-\$5,000.
- Performance. Although tape drive performance and tape capacity have both increased significantly in recent years, the amount of data most organizations need to back up has increased dramatically as well. Even with faster backups, many organizations cannot perform full backups to tape in their available backup window without using multiple tape drives and multiple backup servers, further increasing the cost and complexity of tape backup.
- Recovery. Recovering data from tape can be time-consuming. For effective recovery tapes must be labeled properly and the backup system must maintain a catalog of tapes in a database. If the database is lost or corrupted tapes must be re-cataloged, which can itself be a very time-consuming process. Since most tape rotation schemes include some offsite storage of tapes, if the data that needs to be recovered is on a tape stored off site that tape must be retrieved to recover the data. Since tape is a linear format, accessing and restoring a file or files usually takes

---

1

Linear Tape Open (LTO) is a tape format created by the LTO Consortium, which was initiated by Seagate, HP and IBM. LTO is an open standard created in the late nineteen nineties as an alternative to the numerous proprietary tape formats then in existence. LTO-5 is the latest incarnation of the standard. LTO-5 tape cartridges have a native capacity of 1.5 TB. Linear Tape File System (LTFS) is a standardized file system for LTO-5 and above. Data written in LTFS format can be used independently of any particular storage application. Since LTO is an open standard, LTO drives and media are available from many manufacturers.

significantly longer than restoring the same file or files from disk. Automated tape libraries and bar coding of tapes can alleviate some of these issues, but automated libraries add additional mechanical and electronic components that can fail, so in some circumstances they can create additional problems.

In spite of these issues many SMBs still use tape backup. In some organizations it is the only backup method employed, while in others it is used in addition to or in conjunction with another backup method.

### **Virtual Tape Library (VTL)**

Virtual Tape Libraries solve one of the major problems of tape – the difficulty of completing a backup within the available backup window. A VTL appears to the system to which it is connected as a tape library with multiple tapes. This means that an organization can use their existing legacy tape backup software to back up to a much faster disk-based systems. With a VTL, the virtual tapes are stored on the system for a period of time to allow file restorations, if necessary. Sophisticated VTLs can also export data to tape for archiving purposes. Vendors of VTLs include IBM, SEPATON, Quantum, FalconStor Software, Data Domain, Overland Storage and Hitachi Data Systems. Employing a VTL might make sense for SMBs who are trying to extend the life of their existing backup software, but a disk-to-disk-to-tape approach<sup>2</sup> (see below) probably makes more sense if software is being upgraded or currently supports disk-to-disk-to-tape.

### **Disk-to-Disk-to-Tape (D2D2T)**

This approach uses disk storage as a form of cache. The backup first goes to disk and is then sent to tape. Like a VTL, this system helps alleviate the backup window problem because writing to disk is significantly faster than writing to tape. Once the data is written to disk, write access to the system can be safely restored. Unlike a VTL only the most recent backups are stored on disk, so mounting a tape is required to restore data from an older backup set. D2D2T also helps with the issue known as “shoe shining.” If the rate of data transfer is below that of the tape drives’ write speed, the drive must stop the tape, partially rewind it, then restart. This not only affects performance, but it can shorten tape life and cause additional debris build up on the tape drive’s read/write head, requiring more frequent cleaning and increased wear on the head. D2D2T can also be employed by VTLs for long-term archiving of data.

### **Backup to Fixed Disk or Array**

This approach uses fixed hard disks or disk arrays as the backup medium. When this approach is used the backup media is typically network attached or attached to a separate backup server. When a failure-resistant RAID drive array is used there is a degree of protection against data loss due to backup media failure that does not exist with tape and removable disk approaches. This approach has the advantage of speed of backup and recovery, but does not have the media portability of other approaches. Also, multiple backups are typically available without having to mount backup media. The major disadvantage here is that if there is a catastrophic failure of the backup system then all backups are lost. This problem can be minimized by replicating the backups to an off-site system, however.

### **Backup to Removable Disk**

This approach uses removable disks in a manner similar to tape. One or more backup sets are written to multiple set removable disks, which are then periodically rotated using a scheme similar to a tape rotation scheme. With this approach the cost of a tape drive is eliminated and the speed of backup and restore is increased. Hard disks still cost more than tape, however. Also, they are more susceptible to damage from

---

2

VTL is a form of disk-to-disk-to-tape, but it is usually not recognized as such by backup software. Most backup software sees VTL as an actual tape library.

dropping than tape and their ability to retain data while sitting on the shelf is still relatively unknown, although a spokesperson for one vendor said the shelf life should be at least five years, and periodic refreshing by powering up and rereading and rewriting the data should extend the data retention period another five years. One vendor of cartridge systems claims thirty years archival storage. Both tape and disk appear to be susceptible to damage from temperature extremes, but hard disks appear to be less susceptible to damage from high humidity than tape.

There are many ways to mount removable disks:

- External disk drives using USB, FireWire or eSATA interfaces.
- Internal cartridge dock and cartridges, such as the RDX system developed by ProStor Systems. The dock is installed in a 5 1/4" drive bay. Internal RDX docks use an USB or SATA interface.
- External cartridge dock and cartridges, such as RDX. External RDX docks use a USB interface.
- Internal tray-less hot-swap rack. This device allows the swapping of bare SATA drives and requires an available hot-swap SATA port.
- External tray-less hot-swap rack. This typically requires a USB or eSATA port.

The tray-less drives are the least expensive, since the racks for them only cost approximately \$20-\$75 and you are not paying for a case or cartridge for each drive. They are, however, the most susceptible to damage from dropping and static electricity. The cartridge systems are probably least susceptible to damage. The damage resistance of the standard external drives is difficult to determine and to a great extent depends on the construction of the enclosure.

## On-Line Backup

Increased Internet access speeds, combined with ever decreasing disk storage costs have made across-the-Internet backup viable. Known as both online backup and cloud backup, the use of these services has increased dramatically over the last few years. Numerous companies are providing online backup services, software and even dedicated backup appliances. Some systems combine online backup with more traditional disk-to-tape or disk-to-disk backup. Some online systems provide for maintaining multiple versions or revisions of files and some do not.

Because of the low transfer speed of online backup when compared with disk-to-disk or disk-to-tape, most organizations do not rely on it for primary backup. This is not true in all cases, however. Some backup systems, for example, provide for online mounting of virtual machine images, allowing users to access their server resources while local virtual machines are being rebuilt.

On-line backup is usually used in conjunction with some method of local backup. Increasingly, backup systems that provide local backup are providing online backup as well.

Most online services have a fixed base monthly or yearly cost plus data transfer and storage costs. Low-end services can cost as little as \$4-5 base monthly while the base cost of some services can be in the hundreds of dollars per month. Transfer costs and storage costs can vary from a low of about \$0.15 per gigabyte to \$3.00 per gigabyte or more. Some vendors charge for data transfer and some do not. Also, the types of services provided vary as well. For example, some services are strictly backup and restore, while others provide shared remote access and/or remote drive mapping, so that multiple users can access online data as they would from local storage. Some provide remote application support as well. Some backup services compress and deduplicate your data before uploading to reduce network traffic, data transfer costs and storage costs.

## Online Backup Issues

The following issues should be considered when selecting an online backup service:

1. Data transfer rate. When large amounts of data need to be backed up, high-speed Internet connections are required.
2. Security. Most online backup services provide 256-bit SSL connections, but in some cases secure connections are optional. Also, some service providers encrypt your data and some do not.
3. Protection of your data. Some online providers have redundant data sites, while some store all your data in a single location. It is important to know how your provider protects your data.
4. Retention policies. Can you set a policy for retention of multiple versions of your data? How flexible can your retention policy be? Can you set different policies for different classes of data? What is the service provider's retention policy if a billing issue or dispute should arise? Is your data immediately deleted? Is there a grace period before access is cut off, and an additional grace period before data is deleted?
5. Emergency data access. How do you access your data if the systems being backed up are unavailable? Are there alternate access methods? What if you need a large amount of data quickly? Some services can arrange to ship your data to you on disk, if necessary. Also, is the data stored in a proprietary format or can it be accessed by multiple applications?
6. Appropriateness of service. Are the services provided optimal for your organization? For example, if you would like online access to a virtual machine image in an emergency, can your software and online service provide that?
7. Costs versus benefits. Price per gigabyte of data stored or transferred is not the only measurement of online service costs and benefits. Make sure the services provided fit your organizations needs in a cost-effective fashion.

## Methods for Backing up Data

There are several means for backing up and protecting data.

### Traditional File-based Backup

The traditional file-based backup approach backs up a system's files and directories, along with file attributes, as discrete items. Some systems can back up directory (Active Directory, eDirectory, etc.) information as well, usually as a separate backup. The big advantage of this approach is that it is easy to restore a file or group of files, or a directory object or objects, relatively quickly and easily from any available backup medium. Most file backup systems maintain a catalog of the files and directory entries of all tapes (or other media) in the backup rotation. As tapes are overwritten those entries are removed from the database.

File backup doesn't lend itself to quick "bare metal" recovery, so a number of backup software vendors have

provided add-ons that perform disk imaging of the basic system, including boot sector and operating system. This approach, although greatly improved in recent years, has been problematic, especially if the recovery image has not been kept up to date or if a system was being restored to a different server or dissimilar hardware.

Another problem with traditional file backup systems is that if the catalog becomes unavailable, due to problems with the backup server for example, backup media needs to be reimported into a new catalog, which can be a time-consuming process with multiple sets of backup media.

## File Synchronization

File synchronization refers to the periodic or continuous copying of files and directories from a source location to one or more destination locations in order to maintain duplicate file sets. This technique is often used to make sure the most recent versions of files are available elsewhere if a primary system fails. When implemented with a versioning system, this approach can maintain multiple revisions of files. File synchronization, with or without versioning, is often used in cloud (on-line) backup systems. It is also used between systems within an organization, commonly between sites to make sure data is quickly available in case of a site-related disaster. File synchronization is often used in addition to traditional backup systems since it can provide immediate access to data. Most file synchronization approaches are unidirectional, meaning they synchronize in one direction only. Bidirectional or multi-directional approaches also exist, but they are much more complex to implement and often require manual intervention to avoid version conflicts. When updating files that have previously been replicated some programs re-replicate entire files while some use delta encoding to only replicate file changes. Delta encoding can significantly reduce both network traffic and replication time. Data compression and data deduplication can also be employed to optimize performance across WAN links.

## Remote Data Replication

Remote data replication is the process of duplicating data between remote sites. With replication data is written to both a local, or primary, storage system and one or more remote, or secondary, storage systems. It is usually employed to guarantee data currency and availability in the event of a site disaster. Remote data replication can be conducted across the Internet or private networks.

Remote data replication can be synchronous, asynchronous, semi-synchronous or point-in-time.

Synchronous replication assures that each write operation is completed to both primary and secondary storage before a host system or application is notified that the operation is complete. This method assures that identical data is written to both primary and secondary storage, but, because of the timing issues involved, it can definitely affect application performance. Effective synchronous replication requires extremely reliable, high-speed networks. Typically Fibre Channel over IP is used. Synchronous replication is usually employed where real-time replication with the highest level of reliability is a greater concern than cost. This method is often used by financial institutions where the loss of even a few minutes of data can cost millions of dollars.

Because of network performance requirements, synchronous replication over long distances typically employs Fiber Channel over IP with channel extenders. As distance increases, latency also increases, which can affect application performance. Typically, distances of less than 150-200 miles are recommended, but under some circumstances greater distances can be achieved.

With asynchronous replication, data is written to primary storage and then to secondary storage sometime later. The host system or application is notified that the operation is complete when the write to the primary system is complete. Data is then passed to secondary storage when network bandwidth is available. Typically this is within seconds or less, but sometimes can be several hours. Asynchronous replication is a good choice when relatively slow or unreliable networks are employed.

With semi-synchronous replication a transaction is considered to be complete when it is acknowledged by the primary storage system and the secondary storage system has received the data into memory or to a log file. The actual write to secondary storage is performed asynchronously. This results in better performance than a synchronous system, but it does increase the chance of failure of the secondary system write.

Point-in-time replication uses snapshots to periodically update data changes, usually on a scheduled basis. This is the least reliable approach, but can be more effectively performed over low-speed links.

Asynchronous, semi-synchronous and point-in-time replication can span virtually any distance, and so are therefore good choices when storage systems are great distances apart. Because these approaches do not require immediate write acknowledgment from secondary storage they also create less of a potential performance impact on the host.

### **Images, Clones and Snapshot Images**

Another method of backup is to replicate a disk or volume to another device. The methods to do this are known as imaging, cloning and snapshotting. The descriptions here are representative and do not reflect all methods used by various software vendors to create images, clones or snapshots.

Imaging software creates a replica of a disk, volume or multiple volumes as a file or set of files that can be used to restore a system to its state at the time the image was created. An image file is similar in function to CD/DVD ISO file. There are no standards for disk and volume image file formats and most are proprietary to a particular software package. Older imaging software only allowed the restoration of complete images, but many current systems allow the restoration of specific files and folders.

Cloning creates replicas of disks, including bootable replicas of system disks. While imaging requires restoring the image file to a disk, a clone can be used as is in place of a failed disk.

Snapshotting is a term that refers to the process of capturing the state of a system at a particular point in time. Disk imaging and cloning are both forms of snapshotting. There are two primary forms of snapshots—full and differential. A full snapshot captures an entire volume, disk or system, while a differential snapshot only captures changes made since the last full snapshot. By creating and maintaining multiple differential snapshots along with a full snapshot a system can be restored to different points in time.

Early image and cloning software, as well as some current software, require the system that is being imaged to be shut down and booted with a floppy disk, CD or USB device that hosts the imaging software in order to create or restore the image or clone. A number of current products, however, allow imaging or cloning of a live system. In the Windows environment most products use Microsoft's Volume Snapshot Service or Volume Shadow Copy Service (VSS) for this function. VSS is a set of services that are designed to provide consistent copies of Windows systems and applications such as Microsoft SQL Server and Exchange.

There are also live imaging systems for Macintosh OS and Linux as well. Apple's Time Machine, included with Macintosh OS X, can be used to create bootable backups, and there are several third-party products that do this as well. For Linux, Acronis Backup and Recovery 10 and the open source package Mondo Rescue can be used for live imaging.

## Continuous Data Protection and Near Continuous Data Protection

When data is written to disk a continuous data protection system saves that new or updated data to a backup system. A near continuous data protection system will capture changed data every few seconds or at pre-defined intervals instead of immediately upon disk write. For most purposes the effect of the two approaches is the same—data can be restored from nearly any point in time. Both approaches can have some effect on system performance and both generally consume more backup media space than more traditional approaches. Some CDP packages allow administrators to set ‘event-driven’ points such as the monthly closing of the books.

## Agent vs. Agentless Backup

When the backup server or service is not running on the system being backed up some method of data transfer must be employed. This can be accomplished by installing a special piece of software, an agent, which is written to specifically communicate with the backup system, or by using software that is already installed on the computer. This often means using standard communication protocols such as CIFS (SMB) or NFS. The agentless approach usually simplifies the rollout of a backup system and can also reduce overall costs. Agents, on the other hand, can often provide better communication between the backup server and client, allowing, for example, a client to tell the server about changes that need to be implemented in the backup. In agentless systems, as well as some agent-based systems, backup control is generally handled at the backup server. Agents are also used for application backup. An agent can make sure a database is in a consistent state for backup, for example.

## Windows Volume Shadow Copy Service (VSS)

Volume Shadow Copy Service (VSS) is a set of services that are designed to provide consistent copies of Windows systems and Windows applications such as Microsoft SQL Server and Exchange. VSS has been included with Windows since Windows Server 2003. VSS allows the backup of open files, locked files and open databases. Backups created with VSS are called shadow copies. VSS can back up full volumes and, with the use of application-aware components, back up specific applications, such as Microsoft SQL Server and Exchange.

For volumes, VSS can create clones, or complete volume copies and differential copies, which are copies of data changed since the last full or clone backup. For databases such as SQL Server and Exchange, VSS be used to create full backups, copy backups, incremental backups and differential backups.

- A full backup includes all selected databases but deletes transaction log files older than the start of the backup.
- A copy backup does not delete log files and will consume more disk space, but it does allow the ability to restore data from points in time prior to the backup, if that data is in the transaction logs.
- An incremental backup only backs up database changes since the last full or incremental backup and then deletes logs older than the start of the backup. When using incremental backups, to restore a database, you must have a full or copy backup and all subsequent incrementals. Generally, differential backups are preferred over incremental backups.
- A differential backup only backs up changes since the last full or copy backup but it does not delete pre-backup logs. When using differential backups you only need the full or copy backup and the last differential.

Some backup systems provide for transaction log backup through VSS as well.

A VSS requester, which is usually a component of the backup software, starts the creation of the backup, or shadow copy. A VSS writer, usually using copy on write, will make sure the data being backed up is in a consistent state. A VSS provider creates the copy.

Most current software that backs up Windows uses VSS to some degree.

## Encryption and Password Protection of Backup Media

Encryption and password protection are often used for media that will be physically transported from one site to another or will be stored in an unsecured location. In some industries legal and/or regulatory compliance may require encryption of such media. Accidental disclosure of personal health records or financial data can have severe repercussions, even if specific laws or regulations are not violated.

Some backup programs, such as older versions of Symantec Backup Exec and EMC Networker, for example, provide password protection but not encryption. This makes unauthorized restoration of data difficult but not impossible.

## Advanced Encryption Standard (AES)

Advanced Encryption Standard (AES) is an encryption standard adopted by the U.S. government as Federal Information Processing Standard (FIPS) 197 in 2001. AES is the encryption standard used by most enterprise-level backup systems. AES supports key sizes of 128, 192 and 256 bits.

## Tape Drive-based Encryption

As of version 4, the Linear Tape Open (LTO) tape format supports hardware-based compression at the tape drive. Although encryption is available for LTO-4 and LTO-5 tape drives, it is not implemented in all drives, so if it is used both the backup drive and restore drive, if different, must support encryption.

## Encryption Issues

There are several issues to look for when you decide to encrypt data.

- **Performance.** Encryption uses CPU cycles, so it will affect performance of the system doing the encryption.
- **Key Management.** In simplest terms, an encryption key is a piece of information that determines the output of a cryptographic process or algorithm. Once data is encrypted with a particular key the appropriate key (with AES it is the same key) is required for decryption. When encryption is used for backup systems it is absolutely critical to make sure the key is available when data restoration is necessary. Effective key management procedures must be in place to make sure keys are properly generated, stored, used and replaced if necessary. Keys and key management procedures must be stored and backed up outside the systems to which they apply so that they are available in an emergency.
- **Encryption and Backup Data Compression.** If both compression and encryption are used on a backup system, the data should be compressed before it is encrypted. If software encryption is used then compression should be disabled on the backup device. If LTO hardware encryption is being employed then both compression and encryption can be performed by the tape drive.

## Backup Data Compression

Data compression is the process of encoding data so it uses less media space. Standard compression algorithms usually operating on the bit level by removing redundant bits of data and replacing them with codes that can be used to restore that data on read. Data compression is supported by most backup systems. Compression can be

provided by the backup application, the tape drive, or, if backing up to disk, the operating system of the backup disk. Application-based compression is sometimes proprietary, so the compressed data can only be read by that software. Tape drive-based compression is transparent to the backup software and does not affect readability of the tape. Current Windows operating systems provide for transparent or on-write compression. Transparent compression is not native to any of the current production Linux file systems. Compression and decompression both affect system performance, since the processes use CPU cycles, RAM and disk space.

## Data Deduplication

Data deduplication eliminates redundant data to reduce storage requirements. Pointers are used to reference the single unique instance of the data retained on the storage system. Depending on the type of data, deduplication can significantly reduce storage requirements. For example, an e-mail system might maintain copies of a file attachment in multiple users' mailboxes. With data deduplication only one copy is maintained. Currently deduplication is used primarily for backup and archiving systems. Although data deduplication can be used on primary file systems, system overhead, lack of standards and lack of direct operating system support<sup>3</sup> make this less attractive.

## File Mode and Block Mode

There are two primary modes of data deduplication -- file mode and block mode. File mode looks for duplicate files while block mode looks for duplicate blocks of data within files.

Block deduplication can be either fixed block deduplication or variable block deduplication. Fixed block deduplication looks for identical data blocks, while variable block deduplication uses more intelligent, thus more processor-intensive, algorithms to look for identical data within blocks.

The effectiveness of the three modes varies with the type of data being stored. Generally, file deduplication is the least effective in terms of data reduction but has the least system overhead, variable block deduplication is the most effective, but with the greatest system overhead, and fixed block deduplication falls somewhere in the middle.

## In-Line or Post-Processing Deduplication

In a backup or archiving environment, deduplication can be applied in two ways-- in-line or post-process. In-line deduplication operates as data is being written to a target device. If a new block (or file) is the same as an existing block the new block is not written to the storage device. Instead, a pointer is set to the existing block (or file). With post-process deduplication, data is written to disk as it is received and then analyzed and deduplicated after the fact. In-line deduplication uses RAM instead of disk space, but it can affect performance while data is being written to disk. Post-process deduplication requires data to be written to disk, reread, compared with existing data and then processed if necessary. Overall, post-processing creates more system overhead, but it can be scheduled so as to not affect backup performance.

## Source or Target Deduplication

Data duplication can occur at or near the data source--the system being backed up--or at or near the target--the backup system. Source deduplication deduplicates data either on the source system itself or on a separate system near the source, such as a dedicated deduplication appliance. Source deduplication can reduce network traffic, and this could be significant in a WAN environment. Target deduplication operates on the target system itself or on a separate system near the target. One advantage of target deduplication is that it can deduplicate

---

3

Sun Microsystems' (now Oracle) ZFS file system includes block deduplication support. ZFS is supported on current versions of Oracle Solaris, OpenIndiana (formerly OpenSolaris) and FreeBSD.

data from multiple sources and can potentially provide greater overall data reduction. Both source and target deduplication can be used in the same backup environment, but usually at a significant cost.

A hybrid form of deduplication, sometimes called client backup deduplication, performs the deduplication on the source system and then compares the result to data stored on the target. If identical data is already stored on the target then the data is not transferred and the appropriate pointers are created to reference the existing data. This approach prevents duplicate data from being transferred across the network and can potentially reduce impact on network performance. This approach is especially effective when backing up data from multiple similar systems, such as client PCs.

### **The Downsides of Data Deduplication**

The deduplication process uses system resources and, depending on how and where it is implemented, can affect system and network performance. Since there are currently no industry standards for deduplication, the system used to dedupe the data must be used to reconstitute it, creating a high degree of vendor lock-in. Another major issue is that cost-effective deduplication requires an investment in hardware, software and implementation services. The costs, however, are often offset by reductions in the cost of backup time and backup media.

### **Application-Specific Backup**

The major issue with backing up database applications is that the files must be “quiesced” in order to be backed up in a consistent, synchronized state. This can be accomplished a number of ways<sup>4</sup>:

- Shut down the database before the backup and restart it after. This is the simplest method and can easily be done with scripting, but it makes the database unavailable during backup.
- Lock and flush (write all pending updates to disk) the tables before the backup and unlock after. A read-only lock allows other clients to query the database but not update it. This still has the problem of the database not being updatable during backup.
- Export data as SQL statements. This preserves the table data, but each table must be individually restored to the database server.
- Use application-specific APIs or utilities. For example, most approaches to backing up Microsoft SQL Server and Microsoft Exchange use Microsoft’s Volume Shadow Copy Service (VSS) and a Virtual Backup Device Interface (VDI) for that particular application. Oracle backups typically utilize Oracle Recovery Manager (RMAN). Oracle also provides a VSS writer to allow Oracle database backup with VSS. PostgreSQL provides pgdump and pgdumpall. Zmanda provides Zmanda Recovery Manager for MySQL (ZRM) in both community and commercial versions. These utilities and APIs will lock the database to prevent updating during backup. Application write requests are written to a buffer and the database is updated after backup. Read requests can also access the buffer during backup so those requests will be able to access any data written while the database is locked.
- Use database transaction logging. A transaction log is a record of all changes made to a database. Transaction logs are commonly used to restore data that was deleted, modified or corrupted after the most recent backup.

---

4

This is not an all-inclusive list.

- Use third-party application-specific software to back up an application. For example, Zmanda Recovery Manager (ZRM) will back up MySQL databases. Zmanda also provides ZRM agents for using ZRM with other backup products. Such third-party products might use proprietary techniques or use some of the approaches outlined here.

Please note that if a particular backup system doesn't support these methods directly they can often be implemented through scripting and scheduling.

## Virtual Machine (VM) Backup

Virtual machines can be backed up in a number of ways:

- Back up a virtual machine as if it were a physical machine. This means install a backup agent or backup software in the VM.
- Back up the virtual machine from the host as a file system object or objects.
- Back up the virtual machine as a bootable copy or snapshot. Depending on the VM, host and backup application, this may be done with or without agent software in the VM itself. Some systems allow the creation of base snapshots and incremental snapshots, which only store changes since the last snapshot. This approach allows restoration to any point in time at which a snapshot was taken.
- Use live migration. Live migration allows you to copy a running VM to a different physical machine without shutting down the machine being migrated. Most current VM managers support live migration.
- Use continuous live migration. This technique uses a combination of live migration and checkpointing to replicate a VM on a continuous basis. Remus, a continuous live migration utility included with current releases of Xen, is an example of this.
- Some combination of the above. For example, at least one backup system creates an image backup of the entire VM and then backs up the file system from that image. This approach provides the image for disaster recovery and the file system backup for restoration of individual files.

## Backing Up Virtual Machines

The widespread deployment of virtual machines (VMs) on microprocessor-based systems has created a whole new set of backup issues. Different virtual machine managers, or hypervisors, have different levels of support for backup. Likewise, different backup systems take different approaches to VM backups.

One purpose of virtualization is better utilization of hardware resources. This means that VM hosts are typically running closer to maximum CPU, memory and I/O capacity than physical (non-VM) servers. Backup also tends to use a lot of CPU, memory and I/O resources, so it can impact the performance of a busy VM host. In addition, backing up multiple VMs means encountering most or all of the issues encountered when backing up multiple physical servers, including scheduling issues and completing backups within the available backup window.

VM backups should make sure that the entire VM as well as individual files and directories can be restored quickly and easily. Depending on the hypervisor being backed up, a variety of backup techniques may be necessary to achieve this goal.

There are a number of options for backing up VMs:

- Run the backup from the VM host or a proxy and back up the files that contain the VM and its definitions from the host. In most cases the VM must be shut down during backup.
- Back up the contents of the VM as if it were a physical machine. This approach usually requires an agent running in the guest VM. Database applications must be quiesced using the same techniques used for applications on physical servers.
- “Hot” snapshot - create an image of the VM while it is running. Since the backup application is running on the VM host or a proxy, there must be some degree of coordination between the backup system and database applications running in the VM to make sure they are in a consistent, stable state.
- A combination of approaches. Some backup applications will use a combination of the above

### Hypervisor-specific Backup Methods

Specific hypervisor software use different methods to backup up virtual machines.

- **VMware** -- Software to back up VMware commonly uses either the VMware Consolidated Backup (VCB) application or the more recent VMware vStorage APIs for Data Protection. VCB is a standalone application that can be called by other backup software. The vStorage APIs, however, are effectively file system drivers that allow access to VMware from Windows or Linux applications. Applications can also access VMware’s proprietary VMFS file system from within the VMware ESX service console. Safely backing up database applications usually requires an agent running in the server within the VM.
- **Microsoft Hyper-V** -- Backup applications for Hyper-V typically use VSS. VSS allows VMs to be backed up from the Hyper-V host. It also allows VSS-aware applications within a VM to be properly quiesced for backup.
- **KVM, VirtualBox, Xen, XenServer and Others** -- These hypervisors generally require command-line utilities, scripts or third-party software for backup.

## Tips and Best Practices for Effective Backups

Given the variety of backup requirements and methods it is nearly impossible to outline a set of best practices that will meet all or even most situations. You should, however, create a data protection plan that fits the needs of your organization. Creating this plan will require a cost/benefit analysis. For example, every step closer to zero downtime will increase costs, so you need weigh the costs of additional anti-downtime measures versus the benefits of the incremental protection provided. Your plan should consider the following:

- **Minimization of Downtime** - You need to consider the consequences of both short-term and long-term downtime. Some downtime, even if it is only a few minutes, is inevitable. Even systems that have been sold as non-stop systems have failed. If nothing else, most systems must be shut down for maintenance, at least occasionally. You need to determine what is the maximum unplanned downtime you can afford. You then need to weigh that against the costs of implementing the systems and services necessary to achieve your desired level of uptime.
- **Access to Data in an Emergency** - You back up data so that it is available in an emergency. This could mean a disk failure, data corruption, accidental data deletion or any number of other causes. This means having the data available and having the means to restore or access it. This requires, among other things, making sure backed up data is both immediately available and stored off-site.
- **Long-term Storage of Archived Data** - Make sure that your data protection plan provides for archival storage of data that must be kept for legal, financial or business reasons.

An effective data protection plan should include some method of local backup, such as disk-to-tape, disk-to-disk-to-tape or disk-to-disk. This system should provide for maintaining multiple backup sets or restore points. This gives you protection against the failure of a single backup media set or allows you to restore a file from an earlier point in time. Your media rotation scheme should also include storing some of your backup sets off-site. Your local backup system should back up all critical data as well as provide the ability to fully restore critical systems, such as servers. The ability to perform a “bare metal” restore (restoring a complete system, including system files, directly from backup) is a plus. Many backup programs offer this feature at varying levels of cost and complexity.

Effective use of server virtualization can help minimize downtime. Many backup systems create bootable VM images, and with some systems it is easy to fail-over to a backup image. Some systems will also begin the rebuild of the primary VM and allow it to run while it is being rebuilt, transferring data from the backup as needed. Other than a possible reduction in performance, this process is transparent to users and applications.

Online backup can give you additional protection against deletion or corruption of critical data. In most environments online backup is used for critical data files that cannot be easily recreated. Depending on requirements and the online system employed some organizations back up all data online.

Applications such as databases and email systems usually have specific backup requirements to make sure they are properly quiesced so they can be backed up in consistent, synchronized state. Make sure your backup routine provides for this.

When multiple systems are backed up redundant data is usually backed up as well. Effective use of data deduplication in the backup process can save backup media space, backup time and LAN/WAN bandwidth. One downside of deduplication is that there are no standard formats, so you are reliant on a specific vendor’s system.

## Customer Name: Sprott Asset Management

Type of Organization: Financial services

Number of Employees: 120 employees; 2 IT staff

Location: Toronto, Ontario, Canada in one location.

Environment: A Windows shop with Exchange, file, Blackberry Enterprise Servers and database servers.

Contact name: Dan Elder, vice president of Information Technology and Infrastructure

Amount and type of data protected: 750GB, transactional data from hedge and mutual funds.

### Challenges:

- Merging Infrastructures – Two companies combined and Elder was brought in to consolidate two existing infrastructures, creating a new network with new hardware and software.
- Continuous Uptime – The new infrastructure required that trading databases and Blackberry Enterprise Servers be up at all times.
- TCO Management – Cost management meant implementing a business continuity server replication solution without deploying more server infrastructure.
- Determining Appropriate BC Services – Solution options initially included replication to co-location facility.

### Solution involved:

- Cloud Environment – Using an enterprise data center, negated the need for a co-location facility.
- Replication Software – Using Geminare's Cloud Recovery server replication, the same replication software (CA's Replication and High Availability) Elder had chosen.
- Data Replication – Replicating 750GB of data continuously to Geminare using 100MB fiber-optic.
- Data Archiving – Archiving Exchange e-mail to an offsite location.
- On-Site Backup – Using disk-to-disk-to-tape backup onsite with Symantec Backup Exec to do full backups nightly and transporting weekly and monthly backups offsite; using Dell PowerVault MD3000 as backup target.
- Solution Assurance – Testing of the replication services regularly.

In event of site disaster, Elder has a comprehensive business continuity plan including Citrix for remote user access should the building be inaccessible, but still running, and Geminare's server replication should one or multiple servers become unavailable—in this event users would rely on virtual hard drives accessing the Cloud environment through a VPN.

### Benefits:

- “The hardest thing to set up with the Geminare service was setting up the VPN connection.”
- “There were nights at other places where I didn't sleep that well, worrying about backup and data protection. Now, I don't think about it that much. With Geminare, it's just set it and forget it. That I like. When we did the failover on the file server it worked like it should have.”

Using Geminare’s Cloud Recovery server replication offering, Elder was able to overcome each of the initial business challenges. A Cloud environment running real-time server replication and supporting a wide range of applications allowed for the merging of two infrastructures and the continuous uptime of their mission-critical servers—all without any new capital expenses such as additional server infrastructure or virtualization.

## Vendor Name: Geminare

Product Name: Cloud Recovery, Cloud Storage Assurance and iCloudRecovery

Link to website: [www.geminare.com](http://www.geminare.com)

Link to data sheet: [http://geminare.com/site\\_english/public\\_product\\_solutions.asp](http://geminare.com/site_english/public_product_solutions.asp)

[http://geminare.com/site\\_english/public\\_product\\_cloudstorage.asp](http://geminare.com/site_english/public_product_cloudstorage.asp)

[http://geminare.com/site\\_english/public\\_product\\_icloudrecovery.asp](http://geminare.com/site_english/public_product_icloudrecovery.asp)

Software, Virtual Appliance, Online, Target Array: Online hosted service

Product Description: Geminare Cloud Recovery is a cloud-based service for replication of Windows, Linux, UNIX servers providing continuous block-level replication in real-time to a replica server in a Cloud environment.

It provides replication to Amazon S3 and EC2, EMC's Atmos Partners including AT&T and Pier1, Rackspace, Nirvanix and Iron Mountain's Archive Service, as well as several others. When a server failure occurs, the user traffic is automatically redirected to a replica server in the cloud and business continues unabated.

The replication capability provided by Cloud Recovery is block-based and allows for automatic failback when the file or application server is restored. After an initial synchronization between the local server and the replica server, the software detects block-level changes to files and data and copies the changes to the replica server in real-time. Data is compressed and encrypted in-flight to further protect customer data. Cloud Recovery protects not only physical servers, but also any virtual servers deployed in a customer's environment.

Geminare Cloud Recovery also supports running on Microsoft Hyper-V, VMware ESX and vSphere, Citrix XenServer and Xen. In addition, Geminare Cloud Recovery and Server Replication supports Exchange, SQL Server, SharePoint, Oracle, BlackBerry, Dynamics and File Servers. It supports the real-time replication of mission-critical applications to similar or identical hardware hosted in the public or private Cloud. A key capability that is unique among offerings is the ability to failback Replica servers in minutes while not disrupting the current production operations, a process that normally requires a bare metal recovery.

Cloud Storage Assurance 2.0 adds an indexing and archiving engine to the replication service that lets users build a verifiable audit trail for their files and data. For every file or e-mail replicated into the cloud, a metadata reference or key is created. This metadata guarantees the authenticity of all archived data. Any files within the cloud that are altered can be easily identified through mismatched keys. This capability lets customers meet their compliance and eDiscovery requirements easily and affordably. Cloud Storage Assurance also provides data migration capability from one cloud to another so that customers can change their service or cloud provider at will. On-the-fly encryption and compression capabilities ensure data stored in an offsite cloud remain completely protected and secure.

A remote management application is also available for Cloud Recovery. Called iCloudRecovery, the iPhone application allows customers, as well as MSPs, VARs or OEMs to initiate "one-button" failovers or failbacks and to manage the server replication environment through any iPhone-enabled handheld device. Customers have full visibility into their replica cloud environment at anytime and from anywhere.

Channel: Geminare's products and services are available from MSPs, hosted service providers, VARs and corporate/government distributors. Geminare's channel-only focus allows customers to utilize the technology through their trusted MSPs, VARs and OEMs which is key. It offers them disaster recovery capabilities without a large capital investment. With Geminare's offering neither the customer nor the partner are required to invest in hardware, software or staffing. In addition, its subscription-based service and easy deployment – typically less than 24 hours – makes it a boon for customers that have strained IT resources.

Cost: Cloud Recovery costs \$399-\$499 per server per month and \$1 per month per gigabyte stored.

## Tables of Geminare Features

	Product Name	Software/Hardware-based Appliance/Virtual Appliance/Online-Target Array	Media Support (Disk, Tape)	Server/Desktop/Laptop support	Backup type (Full, Incremental, Synthetic, other)	Starting cost range	Price per server range	Per agent	Price per GB /TB range	Price per laptop range	Per desktop range	Notes
Geminare	Cloud Recovery Server Replication, Cloud Storage Assurance	O	D	S,D,L	F / other		\$366-\$499/mo.		\$1/GB/mo	\$1/GB/mo	\$1/GB/mo	

## Virtualization Support

	Virtualization supported	Microsoft Hyper-V	Image-Level Backup	File-level recovery	Block-level replication	VMware ESX and ESXi	vStorage API support	Image-level backup	File-level recovery	Block-level replication	VMware vSphere	vStorage API support	Image-level backup	File-level recovery	Block-level replication	Citrix XenServer	Image-level backup	File-level recovery	Block-level backup	Xen	Solaris Zones	Solaris Logical Domains	Other
Geminare	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	VirtualBoX, QEMU, KVM

## Server operating systems supported

	Windows Essential Business Server 2008	Windows Small Business Server 2008	Windows Small Business Server 2008 R2	Windows Server 2003	Windows Server 2008	Windows Server 2008R2	SUSE	Red Hat	Oracle Enterprise Linux	Asianux/RedFlag	FreeBSD	Macintosh OS X	Novell NetWare	UNIX	AIX	HP-UX	Solaris	Other
Geminare	X	X	X	X	X	X	X	X	X					X	X		X	

### Client operating systems supported

	Windows 2003/2008/2008R2	Vista/XP/Windows 7	Storage Server 2003/2008	SUSE	Red Hat	Novell OES Linux	Asianux/RedFlag	Debian/Ubuntu	Macintosh OS X	Novell NetWare	AIX	HP-UX	Solaris	Notes
Geminare	X	X	X	X	X	X	X	X	X	X	X	X	X	

### Applications supported

	Exchange	SQL Server	SharePoint	Oracle	Sybase	MySQL	Other
Geminare	X	X	X	X		X	Third party

### Agents

	Optional	Agent Price	Bare Metal Restore	Client Encryption	Media Server Encrivation	SQL Server	Exchange Server	Active Directory	Virtual Tape	Desktop	Laptop	NDMP	Notes
Geminare		\$0											

### Snapshots and Bare Metal Recovery

	Snapshots Supported	Price	Any point-in-time	# of snapshots	Hardware-agnostic	Bare Metal Recovery	Price	Bootable CD	Recover to dissimilar hardware	Recover to identical hardware	Recover to virtual environments	File and folder recovery	Notes
	X	\$0	X	∞		X	\$0	X	X	X	X	X	
	X	\$0	X	∞	X	X	\$0*	X	X	X	X	X	*\$299 to recover to dissimilar hardware
	X	\$0	X	∞	X	X	\$0	X	X	X	X	X	
	X	\$0	X	X*	X	X	\$0	X	X	X	X	X	*Limits imposed by NetApp Snapshots and SnapVault, Microsoft VSS, Engenio
	X	\$0	X	∞	X	X	\$0	X	X	X	X	X	
	X	\$0	X	∞	X	X	\$0	X	X	X	X	X	
	X*	\$0	X	∞	X	X	\$0	X	X	X	X	X	*Restore files, volumes, databases or system
	X	\$0	X	64	X	X	\$0	X	X	X	X	X	
	X	\$0	X	∞	X	X	\$950/domain	X	X	X	X	X	
	X	\$0	X	∞	X	X	Unknown	X	X	X	X	X	
	X	\$0	X	200+	X	X	\$0-\$199*		X	X	X	X	*Free for up to 50 PCs;\$199 for perpetual 50 PC pack; \$199 perpetual license per server
	X	\$0	X	200+	X	X	\$0-\$199*		X	X	X	X	*Free for up to 50 PCs;\$199 for perpetual 50 PC pack; \$199 perpetual license per server
Geminare								X	X	X	X	X	Replica identical to source

### Deduplication

	Deduplication	Price	Source	Inline	Target	Post-processing	Appliance	Raw Capacity	Logical capacity	Maximum throughput
Geminare										

### Continuous Data Protection

	Continuous Data Protection	Price	Any point-in-time	# of snapshots supported	Microsoft Exchange	Microsoft SQL Sever	Microsoft SharePoint	Virtualization	# of servers supported	# of laptops/desktops supported	Notes
Geminare	X		X	8	X	X	X	X	8	8	

### Replication and File Versioning

	Replication	Price	Local	Remote	Asynchronous over IP	Bi-directional	Heterogeneous	Many-to-one	One-to-many	Block-level	File-level	Bandwidth reduction	Compression	Policy-based	File versioning	Price	Number of versions	Notes
Geminare	X	\$0		X	X		X		X	X	X	X	X	X	X	\$\$1/GB/mo	8	

### Management Console

	Management Console	Monitoring	Analytics	Reporting	CLI	Scripting	Web-based	GUI	Remote console	Notes
Geminare	X	X	X	X	X	X	X	X	X	

Channel

	Direct	MSP	Hosted service provider	Online/cloud	VAR	Corporate/Government Distributor	Other
Geminare			X	X	X	X	