

The Three C's of Data Protection

symform

Praerit Garg

1/1/2011

ABOUT THE AUTHOR

Praerit Garg, President and Co-founder

Praerit Garg is the President and Co-founder of Symform. Prior to Symform, Praerit was a Senior Director in Microsoft's Server and Tools division where he built and managed the Dynamic Systems Platform & Tools team. Under his leadership, the team grew from zero to over 70 people and drove a broad industry initiative to expand the W3C's XML standards to incorporate modeling enhancements called the Service Modeling Language (SML). The team also delivered technology and solutions across several Microsoft products, including the distributed modeling system in Visual Studio 2007, Group Policy tools in Windows Vista/Server 2008, Server Manager in Windows Server 2008, and Desired Configuration Management in System Center Configuration Manager 2007. Praerit also led the acquisition and integration of a software company developing Group Policy tools.

Earlier in his 12+ year tenure at Microsoft he held various engineering leadership positions in Windows, ranging from a software developer driving US and European security certifications of Windows NT 3.51/4.0 products, to Group Program Manager for Windows Security delivering security infrastructure, including Active Directory, Kerberos, PKI, EFS and Group Policy in Windows 2000, XP and 2003 releases.

Praerit is a co-inventor on 14 U.S. and international patents and a co-author of IETF RFC 3645, Web Services WS-Trust and WS-Secure Conversation specifications. He holds an M.S. in Computer Science from Purdue University and a B.E. (with Honors) from Birla Institute of Technology & Science, Pilani, India.

The Three C's of Data Protection

Comprehensive, Convenient, and Cost-effective

The ever-increasing digitization of information – documents, customer records, employee records, financial records, photos, music, etc. – is forcing companies to store more and more data. The expansion of data in our lives seems inexorable and the Internet is certainly the driving force. More and more, we rely on data to run our businesses. Today, protecting our data means protecting our livelihoods.

The U.S. Bureau of Labor Statistics reports that the majority of small-to-medium businesses (SMBs) never recover from a catastrophic data loss. Yet most businesses have no data-protection plan.

Unlike the loss of physical assets such as buildings and equipment, which can be replaced relatively quickly through insurance payouts, lost data offers very little recourse. It is no surprise that data protection is now a top concern for businesses. It can mean the difference between being in business or not.

A data-protection solution must succeed across three dimensions in order to meet the needs of businesses:

Comprehensive – It must address all facets of data protection – human errors, hardware or software failures, and disasters such as theft, fire, flooding, etc.

Convenience – It must be *set-and-forget*. Businesses are strapped for resources, so any solution that requires constant care will not be effective.

Cost-effective – It must fit within the budget. Businesses have a very limited budget for IT overall and data protection is only one part of an IT budget. Any solution needs to have a cost that doesn't change dramatically month to month –particularly with ever-increasing amounts of data.

Current data protection solutions

Onsite disk backups – According to research from International Data Corp. (www.idc.com), **58% of businesses do only local backup**. Local backups are a critical first step in any good data-protection plan. There are several well-known backup applications on the market. Both the Windows and Mac operating systems have native applications and there are several third-party applications from vendors like StorageCraft and Symantec. Local backup provides adequate protection against common data-loss scenarios, such as human errors, hardware failures, etc. However, local backups alone are not a complete data-protection solution. They do not protect against theft, natural disasters such as flooding or fire, or multiple hardware failures that could be caused by something as simple as a power surge.

Tape rotations – Research by IDC indicates that about 16% of businesses use tapes for backup. Historically, tapes have been a popular backup medium because they are portable and inexpensive. This meant that a business was able to get both onsite and offsite data protection using a tape-rotation strategy. That said, industry data show that 50–70% of tape-based backups cannot be restored. With 1-2 TB USB hard drives now available for \$100 or less, tapes are nearly, if not already, obsolete.

External hard drives – These are a good replacement for tapes. They provide high capacity and are fast, inexpensive, and portable. Like tapes, external hard drives address some of the limitations of local backups. They can be taken offsite, which means that if a process is set up to regularly rotate external hard drives, you not only have a local backup but also one that is offsite and may be just a few days old. A key challenge with this solution (and tapes) is the high degree of human involvement required, which, by its nature, is error-prone. Some person or persons in an organization must take the responsibility for the process and follow it diligently. In business, there is always a shortage of human resources and this task is not something that is core to the day-to-day functioning of the business. As a result, it rarely happens. Even when it does, the manual transportation and handling process has risks – dropping and damaging the drive, losing it somewhere on the way, theft, etc.

Data-center-based online backups – Increasingly, data-center-based online backups are becoming a choice for a *set-and-forget* data-protection solution. With the ubiquity of Internet access and ever-growing bandwidth, online backups are becoming an attractive alternative. Not only is the data backed up offsite, it is done so automatically with no constant human involvement.

Despite the attractiveness of this solution, IDC’s research suggests that only 10% of businesses are using online backup solutions. This is surprising given that online backup solutions have been around for over ten years. Several factors may explain their slow adoption in the market:

Cost – Storing data in data centers is prohibitively expensive – as much as several dollars per gigabyte per month. This adds up quickly for an SMB – several hundred dollars per month for just a few hundred gigabytes of data. In contrast, a disk-rotation solution using a couple of 1-TB external hard drives can be implemented for a *one-time* cost of less than \$200!

Security – A significant concern with online solutions has been the security of the data centers. Copies of sensitive data are sitting in some remote data center – what kind of security does the facility have? Who has access to the facility and how much can these unknown individuals be trusted? Additionally, the stored data is co-mingled with the data from other companies – often competitors. What kinds of data isolation, access controls, and protective measures are in place to ensure there is no breach? None of these issues are relevant in the disk-rotation solution because the disks are under the control of the business.

Time to initial backup – With limited upload bandwidth, companies with large data sources may be disinclined to adopt a solution if it takes *several weeks* to get that first backup uploaded. Online vendors must pay substantially for data center bandwidth, so they tend to “throttle” incoming traffic across multiple clients to manage their costs. By comparison, backup to an external hard drive is blazingly fast.

Time to restore – Given the rarity of disasters, even if a company does overcome the hurdle of time-to-initial-backup (a one-time event), restoring can become an even greater challenge. Again, given limited bandwidth and back-end throttling by the vendors, a restore could take many days for companies with substantial amounts of data. By comparison, it is much quicker to restore data from a local drive, or bring back one that was rotated offsite (of course, assuming it was properly transported and handled).

Maturity of backup software – One of the most problematic aspects of data protection is doing a restore after a data loss. Anyone with data-protection experience will be quick to point out that backup is easy, restore is hard. The true strength and quality of backup software is only evident when a restore is necessary. This makes selection of the right backup software critical. Most online backup services require use of their own backup software. These new applications do not have the maturity of local backup solutions that have existed for years. Switching to untested, unproven backup solutions creates significant friction.

Support – When a disaster happens, the last thing a business owner wants to do is call an 800 number and be placed on hold. Businesses need immediate, local help from someone who understands their systems and who can start the recovery process right away. Most online backup solutions are impersonal web- and phone-based services. This makes gaining customer trust increasingly difficult.

Current online solutions address only a subset of these issues, and with varying degrees of effectiveness:

Cost – This has been decreasing as the price of storage hardware has decreased. That said, online storage services remain orders of magnitude more expensive than local storage. For example, you can easily buy a 1-TB USB drive for \$100, yet the cost to backup 1 TB of data using an online service can be as much as \$500 *per month*. This is because the cost of hardware is only a small fraction of the overall costs of running a data center. The capital and operational expenditures required to build and run a data center account for as much as 82% of the fully loaded costs. In addition, data centers must be over-provisioned in order to handle potential demand and there must be at least two data centers to ensure geographical redundancy to protect against data center downtime. This increases the overall costs that ultimately must be paid by customers.

Security – Most solutions now encrypt the data on a customer's computer prior to sending it to an offsite facility. This means, however, that customers now must manage their encryption keys. Losing these keys could render the data irrecoverable. Creating and managing keys is yet another point of friction.

Time to backup – A few solutions enable customers to have an onsite backup to a dedicated device which then trickles the data to their back-end data centers over time. Some also allow mailing in (e.g. via FedEx) a hard drive with the initial backup to their data center for fast upload. Each of these options typically costs thousands of additional dollars.

Time to restore – Some solutions offer overnight mailing of DVDs or hard drives to enable a quicker restore. This requires an additional fee.

Maturity of backup software – Most online solutions do not address this issue. In only one case where the vendor is providing an onsite-plus-offsite solution, have we seen the use of industry-standard backup software.

Support – Most online solutions do not have a strong local-channel model. This is due to the high cost of goods, as discussed earlier. This shrinks the target market significantly and leaves very little margin for the channel to be motivated to provide local sales and the necessary support.

To summarize, let's evaluate the current data protection solutions across the three dimensions outlined above:

Comprehensive – Does the solution adequately cover all aspects of the data-protection problem?

Convenient – How much effort needs to be expended regularly to achieve the necessary data protection with this solution? Is it really set-and-forget?

Cost-effective – Is the solution affordable to an SMB?

Solution	Comprehensive	Convenient	Cost-effective
Onsite disk backup	No	Yes	Yes
Tape rotation	No	No	Yes
Disk rotation	Yes	No	Yes
Data center-based online backups	Some	Yes	No

Unfortunately, none of the current solutions on the market succeed at being comprehensive, convenient, and cost-effective. The best you can achieve is two out of three C's.

Creating a comprehensive, convenient, and cost-effective solution

The Symform Cooperative Storage Cloud™ takes the best attributes of each solution listed above and combines them into one solution:

- Like disk rotation and online solutions, it is **comprehensive**.
- Like onsite disk backup and online solutions, it is **convenient**.
- Like onsite disk backup and disk rotation, it is **cost-effective**.

Here's an easy way to think about it: imagine a disk-backup-and-rotation solution *without the need to rotate disks or store them off-site*.

- You use your favorite backup software – we support them all.
- Instead of using two disks for rotation, you use disks on an existing server or add USB drives, *once*.

- You configure your local backup to use one of those disks. The other is a “spare.”
- Using the power of the Internet and the innovative Symform Cooperative Storage Cloud technology, you effectively *trade* the “spare” disk for a vastly more reliable and secure virtual backup drive in the storage cloud.
- Your local backups are automatically mirrored to this virtual drive in the Cooperative Storage Cloud.
- And it only costs a small flat monthly fee to do this irrespective of the amount of data.

The result is a data-protection solution that is comprehensive, convenient, *and* cost-effective.

Achieving the Three C's

Comprehensive – The solution is comprehensive because it addresses all dimensions of data protection:

Onsite local disk backup provides fast, efficient restore capability for the most common data-loss scenarios – human error, data corruption, primary hardware failure, etc. Backups are done using *any* backup software desired, e.g., built-in backup in Windows, StorageCraft ShadowProtect, Symantec BackupExec, etc. Local backups are on the physical premises and are as protected as the live data. Data can be encrypted using any standard encryption technology. Some backup software includes built-in compression and encryption capability.

Using the Symform software, local backups are automatically mirrored into a secured virtual disk in the Symform Cooperative Storage Cloud. This provides offsite protection against disasters – theft, flooding, fire, etc. Symform encrypts the data *locally* (prior to mirroring to the cloud) using a federally certified, military-grade encryption algorithm – 256-bit AES. This ensures that no business data leaves the site without proven, exhaustive protection.

Every block of 64 MB is encrypted using a 256-bit random key. This means that even in the highly unlikely event that such a key is compromised only one block of data may be at risk. File and associated block information, including all block keys, are stored securely in the Symform Cloud Control. Only properly authorized and authenticated Symform software is able to retrieve file and block information from the Symform Cloud Control. The information is always protected using SSL in transit.

Symform software itself must authenticate with the Symform Cloud Control using a large random key to gain access to file and block information including keys that were used to encrypt its blocks. =.

This means that only Symform software at the customer site and the Symform Cloud Control know the authenticating keys used for storing and accessing customer specific file and block meta-data. Furthermore, in the event of a disaster, a brand-new installation must be performed

to recover the customer data from the Symform Cooperative Storage Cloud. This requires a new random key that can be obtained only by the trusted user account with appropriate authority by doing a restore operation in the Symform Configuration UI. Restore operation renders the old keys useless eliminating any potential risk associated with lost keys. This new key is provided directly to the authenticated Symform software. This approach enables a highly secure yet fully automated key-management solution.

The encrypted data blocks are *redundantly dispersed* to thousands of other randomly selected participating systems running at other customer sites in the Symform Cooperative Storage Cloud. This results in unparalleled security, availability, durability and speed. This is done as follows:

- Each 64-MB block of encrypted data is divided up into 64 1-MB fragments.
- 32 1-MB parity fragments are added to make a total of 96 1-MB fragments for every 64-MB encrypted block. (Parity fragments are generated using the industry-standard Reed Solomon encoding scheme which makes *any* 64 out of the 96 fragments sufficient for recreating the block.)
- These 96 fragments (Symform RAID 96™) are then sent to 96 randomly selected computers operating within the Symform Cooperative Storage Cloud.

Unparalleled Security – Dispersing the encrypted fragments to random location implies that there is no one place where the entire data set is stored outside of the premises. In order to breach this security, 96 random computers would have to be discovered across the Internet and contacted for every 64-MB block. Each block would then need to be decrypted using a random 256-bit key, which can only be obtained by first breaching the Symform Cloud Control. This process would have to be repeated for every block for the entire file to be re-assembled. *This is truly superior to any other data-security solution in the market.*

High Availability – Using Symform RAID 96™ means that as many as 33 systems (each storing one fragment of the block) must fail at the *same time* for the block to be inaccessible at that instance. The probability that 33 out of the given 96 happen to fail at the same time is vanishingly small.

Strong durability – With 32 parity fragments for every 64 original (Symform RAID 96™), the system has sufficient redundancy to protect against any type of failure. As a comparison, RAID 5 has only one parity fragment for every four original, and even so, it is regarded as a highly robust data storage system.

Blazing Speed – Taking a 64-MB block and transforming it into 96 1-MB fragments – each of which go to different locations on the Internet – enables the Symform Cooperative Storage Cloud to achieve very high levels of parallelism during uploads and downloads. Assuming sufficient bandwidth, the net effect of this is equivalent to 64 MB of data getting transferred in roughly the same amount of time as it would take to transfer 1 MB of data to a server in a data center. That is potentially a 64X increase in speed compared to traditional data centers!

Convenient – The solution is truly set-and-forget with several convenient attributes:

Simple to set up – The system requires a five-minute download and installation of the Symform software.

True set-and-forget – Once the software is set up, it never needs to be touched again unless there is a human error, hardware or software failure, or a disaster. It runs in the background and automatically mirrors the local backups (or any set of selected files and folders) into the storage cloud according to the configuration defined during set-up.

Highly secure, yet no keys to manage – The Symform Cooperative Storage Cloud solution is architected to be secure without compromising convenience. As discussed earlier, each block of data is encrypted using a random 256-bit key. From security perspective, this means that no two blocks are encrypted using the same key and having a key for one doesn't mean you can decrypt another block. What is even more important is the fact that none of these keys needs to be stored and managed. They are stored securely in the Symform Cloud Control and made available only to authenticated and trusted Symform software running over an SSL-protected, secure channel – never to a human.

Always-available dashboard – Symform customers get an online dashboard to monitor their synchronization status.

Regular email reports – Symform customers receive a regular email report providing them the status of synchronization.

World-Class Support – In the event of a problem, Symform support is available by phone and email with fast turn-around times. Customers also have the option of getting local onsite support from a Symform authorized service provider.

Locally restored within a few hours – The unique mirroring technology built into the Symform Cooperative Storage Cloud enables customers to host a hot standby of data at a secondary location. This allows users to recover and be operational within a few hours after a disaster. Customers who do not have access to a secondary location can implement this feature by working with a Symform authorized service provider.

Multiple options for a large initial backup – The Symform Cooperative Storage Cloud is distributed across the Internet so there is no one data center where you must send initial backups. The options for initial backup are:

- Upload into the storage cloud from the data originating site. This needs to be done only once. The time required to will depend on the site's bandwidth.
- If higher bandwidth is available at a secondary site, simply create a copy and perform the upload from both locations. Symform's global deduplication service will ensure that no block is uploaded twice. Again, customers who do not have access to a secondary location can implement this feature by working with a Symform authorized service provider.
- Send the drive to Symform and we can perform the seeding on your behalf.

- Elect not to upload the baseline backup at all, but simply deposit it once in a safe offsite facility and use the Symform Cooperative Storage Cloud to protect subsequent incremental backups only.

Cost-effective – The solution is extremely affordable and creates an immediate ROI relative to the alternatives. For the price of a couple of large hard drives plus an economical flat monthly fee, businesses get a comprehensive, convenient, *and* cost-effective data-protection solution. *The best part is that you can use as much storage as needed to achieve comprehensive data protection.* No more per-gigabyte fees, and no more unpredictable increases in expenses every year.

Call to Action

Visit www.symform.com to start your no risk, no commitment 30-day free trial. You can stop worrying about data loss forever.