InfoWorld

SPECIAL REPORT

Disaster Recovery Deep Dive

Preserve your data when disaster strikes

Smart strategies for data protection

Disaster recovery and business continuity may be IT's most important responsibilities. The key is to create smart policies collaboratively—and then line up the appropriate technologies to match

By Matt Prigge

DATA IS LIKE OXYGEN. Without it, most modern organizations would quickly find themselves gasping for breath; if they went too long without their customer, transaction, or production data, they'd die.

The threats to data are everywhere, from natural disasters and malware outbreaks to simple human error, like accidental deletions. Yet surprisingly few IT organizations take the time to construct the policies and infrastructure needed to protect the data organizations need to survive. The result? Poorly planned, ad hoc recovery and continuity methodologies that no one fully understands and which are often obsolete before they're even implemented.

Thriving in today's data-centric environment requires a renewed focus on disaster recovery and business continuity procedures and policies. It's not just an IT problem. Data loss affects every part of every organization; because of that, every stakeholder—from the executive suite to facilities and maintenance—needs to be involved in the decisions surrounding data ownership.

This type of collaborative approach is neither easy nor inexpensive. But it will allow both IT and business stakeholders to be on the same page about where the organization's disaster recovery (DR) and business continuity (BC) dollars are being spent, and why. More important, it will ensure everyone knows what to expect in the event of a disaster and can respond appropriately.

WHAT ARE DR AND BC?

Disaster recovery and business continuity have come to mean different things to different people. But the core concepts are the same: Deploying the correct policies, procedures and layers of technology to allow organizations to continue to function in the event disaster strikes. This can range from something as simple as having easily accessible backups in case of hardware failure to nearly instant fail-over to a secondary data center when your primary one is taken out by a cataclysmic event.

Generally speaking, DR is a series of procedures and policies that when properly implemented allow your organization to recover from a data disaster. How long it takes you to recover, and how stale the data will be once you're back online, will depend on the recovery time objectives (RTOs) and recovery point objectives (RPOs) decided upon by your organization's stakeholders. DR's job is not to provide consistent up-time, but to ensure you can recover your data in the event that it's lost, inaccessible, or compromised.

A comprehensive disaster recovery plan would also include anything else required to get your environment back on its feet, including how and where backup sets are stored, how they can be retrieved, what hardware is needed to access them, where to obtain alternate hardware if the primary systems are inaccessible, whether installation media for applications would be required, and so on.

Business continuity systems aim to mitigate or neutralize the effects of a disaster by minimizing business interruption. They typically rely on high availability (HA) hardware—redundant systems operating in real time and with real data, so that if one fails the others can pick up the slack almost instantaneously with little or no effect on business processes. An HA event (such as a cluster node failing over) might introduce a minute's worth of

service interruption, whereas recovering a backup onto a new piece of hardware could take hours or even days—especially if replacement hardware isn't readily available.

But having HA systems in place does not ensure disaster won't strike. There will always be failure vectors like data corruption, software glitches, power spikes, and storms that can skip right on past every redundant system to despoil your data and ruin your day. That's the most important thing to remember about HA: It only lowers your probability of experiencing disaster, it doesn't make you immune.

Business continuity ties together the availability goals of HA with the recoverability goals of DR. A carefully conceived BC design can allow you to recover from a full-blown disaster with limited downtime and near-zero data loss. It is by far the most involved and expensive approach to take, but many enterprises have concluded that data is so vital to their day-to-day operations BC is too important not to pursue.

BUILDING A POLICY

There are four main stages to building a coherent BC/DR policy for your organization. The first task is to define a scope that details which applications will be covered by the policy. Second, you need to define the risks you're attempting to protect your infrastructure against. The third task is to define the precise requirements of the policy in terms of RTO and RPO for the risk categories you've defined. Only after you've completed those steps can you determine the technological building blocks you'll need to deliver on the policies you've defined.

Throughout this process, it's vital to involve business stakeholders in the decision-making process—however nontechnical they may be. In fact, the less technical those stakeholders are, the more critical it is that they be involved. A failure to set downtime and recovery expectations properly can have devastating, career-ending consequences.

SCOPE

When considering the scope of a BC/DR plan, it's important to start by looking at the user-facing applications you are attempting to protect. If you start from the perspective of the application and then determine everything it needs to survive, you'll find all the infrastructure and data associated with it. If you start from the infrastructure and work up to the application, it's easy to blind yourself to things the application needs in order to run. For example, you could create a flawless backup of your application and database clusters, but still forget to back up a file share that contains the application client, along with all of its hard-to-reproduce configuration data.

Though it makes a great deal of sense to separate less mission-critical applications from those that the business depends upon heavily, it is increasingly common to see BC/DR policy scopes that extend to nearly all of the core IT infrastructure. This is primarily as a result of the popularity of server virtualization, network convergence, and centralization of primary storage resources. In such environments, a BC/DR policy that seeks to protect a mission-critical application served by a small collection of virtual machines may well end up including nearly every major resource in the data center within its scope. This is not necessarily a bad thing, but it can drive up the costs significantly.

RISKS

After a scope for the policy has been set, the next step is to determine which risks the policy will attempt to address. IT architectures are vulnerable to a dizzying array of risks, ranging from disk failure all the way up to Hurricane Katrina-level natural disasters.

Typically, almost everyone will want to plan for common risks such as hardware failure, but some organizations may choose not to plan for less-probable risks such as regional disasters or pandemic outbreaks. It's not unheard of to see significant investments made in an effort to mitigate disasters that the business's core operations—specifically human resources—could not withstand in any case.

Knowing which risks you are and are not protecting the organization's data against is critical to thorough planning and implementation. No matter what your organization chooses to do, the most important thing is that everyone from the C-level executives on down is on board with the decisions made.

REQUIREMENTS

After you've set your scope and defined the risks, it's time to put hard numbers on how quickly your BC/DR policy dictates you recover. At the very least, this should

include targets for recovery times and recovery points. However, your organization may not decide to place the same RTO/RPO requirements on every kind of disaster.

For example, your organization may require a 30-minute RTO and 15-minute RPO for a particularly critical database application in the face of hardware and software failures, but may be satisfied with longer RTO/RPOs if the entire office building is destroyed by a fire and employees are unable to return to work.

Of course, we'd all like to have near-zero recovery time and data loss. While this is absolutely possible with technology available today, it also may be prohibitively expensive. Thus, defining policy requirements is a balancing act between the relative cost of providing ultra-low RTO/RPO and the cost to the business of downtime or data loss.

However, it's important to avoid getting caught up in discussions of how much recovery options will cost until you've determined what the organization truly needs in order to remain viable. Too often, huge price tags will dissuade IT from even offering aggressive RTO/RPO solutions for fear of being laughed out of the corner office. If in the final analysis budgets won't support more expensive options, a collective decision can be made to scale back the requirements, but it will be made based on a full understanding of the situation.

BUILDING BLOCKS

After you have a solid idea of the services you need to protect, the risks you need to protect them from, and your recovery objectives, you can figure out the technological building blocks you'll need to deliver on the policy. Due to the sheer number of different solutions available, there is really no one right way to protect any given infrastructure.

The trick to designing a good BC/DR solution is to treat each component as a layer of protection, each with different capabilities that combine to provide a comprehensive solution.

SOFTWARE BUILDING BLOCKS

The heart of any BC/DR strategy is the software that makes it tick. Generally speaking, this will involve at least one kind of backup software for disaster recovery, but may also leverage application-level backup/replication and even full-blown host-based replication software to deliver on business continuity goals.

Agent-Based Backup—Traditionally, the most common software component you'll see is the agent-based backup utility. This software generally utilizes one or more centralized server installations to draw backup data from software agents installed within the individual servers you want to protect, then stores the backups on tape, direct-attached disk (DAS), or other storage hardware.

If disaster strikes and data is lost, the data can be restored to the way it was at the time of the last backup. This works reasonably well for situations where you're protecting against data corruption or deletion and there is a relatively long RPO (usually 24 hours or more). However, when risks include wholesale loss of a server containing a substantial amount of data, traditional agent-based backup is unlikely to meet more aggressive RTO requirements, since restoring a server from bare metal can require a significant amount of time.

But don't write off agent-based backup utilities just yet. Despite their stodgy reputation, modern backup agents may offer advanced features such as image-based backups and client side deduplication, which can both shorten recovery objectives and decrease the time needed to create backup sets.

Image-based backups capture the contents of an entire server at once instead of file by file and restore the machine exactly as it was, making backups easier to capture and restore. Client-side deduplication lets you use cheap centralized backup storage and distribute the dedupe load, making backup over a WAN faster because there's less data to pass down the wire. These more advanced forms of agent-based backup are useful for organizations who need to shorten their RTO/RPO but don't have the virtualization stacks in place to make backup more seamless.

Virtualized Backup—Virtualized infrastructures have significantly more backup options available to them. Because virtualization introduces an abstraction layer between the operating system and the underlying server and storage hardware, it is far easier to both obtain consistent image-based backups and to restore those backups onto different hardware configurations should the need ever arise.

In fact, image-based virtualization backup software can also be leveraged to perform business continuity duties in addition to disaster recovery. Instead of storing backups on dedicated storage media like tape or network-attached storage, the backups can be stored live on a backup virtualization cluster– a collection of virtualized hosts located in a secondary location that can be turned on at a moment's notice. Though it's difficult to achieve RPOs of under 30 minutes with this approach, it can be an exceptionally cheap way to offer aggressive RTO when more advanced hardware-based approaches like SAN-to-SAN replication (see below) aren't feasible.

Host-Based Replication—You can achieve a similarly aggressive RTO for physical servers by using host-based software that replicates data from your physical production servers to dedicated recovery servers, which can be either physical or virtual. However, that solution requires an expensive one-to-one ratio between your protected assets and recovery assets—you'd essentially be paying for two of everything. This kind of software is also typically not dual-use, which means you'd still need to field separate disaster recovery software to provide archival backups. However, in situations where operating systems and applications can be virtualized and have not been, it is often possible to utilize host-based software to replicate many physical servers onto a single virtualization host– handily avoiding the need for a duplication of physical resources.

Application-Level Backup—Certain types of applications come with their own tools for providing disaster recovery and business continuity. For example, you could configure a database engine to capture full backups and snapshots of transaction logs each night to a secondary storage device for DR, while replicating transaction logs to a server in a secondary data center for BC. Though these solutions can provide a cheap way to provide very stringent RTO/RPO, they are also one-off solutions that will only apply to applications that support them. If you've got more than one critical database with its own DR and BC processes, you'd have to manage each of them separately, increasing the complexity of your BC/DR approach and creating a potentially huge source of headaches.

Keep it Simple!—No matter what mix of software you choose, try to keep it as simple as possible. Though you can achieve a great deal by carefully layering several different kinds of backup software and adding a healthy dose of customized scripting, the more complexity you introduce into the equation, the more likely you are to encounter backup failures or—worse—put your data at even greater risk.

HARDWARE BUILDING BLOCKS

No matter what kind of approach you use, hardware will play an essential role. With disaster recovery the hardware's job is to store backup data and provide data retention/archival capabilities. In BC applications, the role of hardware can much more varied—ranging from virtualization hosts with their own inboard storage, to shared network attached storage (NAS) systems, and synchronized storage area networks (SANs) hundreds of miles apart. The public cloud can also be leveraged to play a substantial role in both DR and BC.

Tape Backup—When most people think of DR-related backups, the first thing that comes to mind is usually tape backup. Though much maligned, tape backup systems have remained relevant due to sequential read and write performance that can be two to three times faster than disk, as well as well as their long shelf life and the relative durability. Though disk-based backup has become much more popular recently, disaster recovery approaches that require backup media to be shipped off site for archival storage still generally rely on tape.

Disk-Based Backup—However, tape suffers from a number of substantial limitations, primarily because it's a sequential medium. Reading random chunks of data off different parts of the tape can be very time consuming. Backups that rely heavily on updating or referencing previous backup file data are much better suited to on-line spinning disk. But even where disk is being utilized as the first backup tier, off-site archives are still frequently stored on tape—forming the oft-used disk-to-disk-to-tape model.

Disk backup media can take a number of different forms. In the simplest application, it can be a backup server with a large amount of direct-attached storage. In situations where the backup software is running within a

virtualization environment or is on a system that stores backup sets elsewhere, NAS is frequently used.

NAS and VTL—In large installations requiring significant amounts of throughput and long periods of data retention, it's much more common to see dedicated, purpose-built NAS backup appliances. These devices come in two basic flavors: one that acts as an arbitrarily large NAS and is accessible via a variety of file-sharing protocols; and virtual tape libraries (VTLs) that emulate iSCSI or Fibre-Channel-based tape libraries. VTLs can be useful in situations where legacy backup software expects to have a real tape drive attached, but tapes themselves aren't desirable. By using a VTL, you can provide a deduplicated disk-based backup option while still allowing the legacy backup software to function.

For most organizations that need to store large amounts of data for a significant period of time, however, a NAS backup appliance is much more likely to be the right answer. The primary reason for this is that nearly every enterprise-grade backup appliance will implement some kind of deduplication to remove portions of the data set that are exact copies of each other—like, for example, the same company-wide email sent to every employee's in-box. Deduping the data can significantly reduce the amount of storage required by each backup set, allowing organizations to increase the amount of historical backup data they can maintain.

There are a wide variety of deduplication products, some of which are more efficient than others, but they generally utilize one of two approaches: at-rest deduplication and inline deduplication. In the first instance, data is backed up onto the appliance as-is, and then deduplicated after the backup is complete. In the second, data is deduplicated as it is being written onto the device.

Typically, devices that leverage at-rest deduplication take less time to backup and restore data because they don't have to dedupe and/or rehydrate data as it moves on and off the device. However, because they must store at least one back up in its un-deduplicated state, they also require more raw storage compared to inline deduplication devices.

The type of deduplication device that's best for a given application really depends upon how it will be used. For instance, if you plan to store backups of a virtualized environment on an NAS, you may want to use an at-rest deduplication platform due to its typically faster read and write performance. Some virtualization backup software can restore a virtual machine while that virtual machine's backup data resides on the NAS—yielding extremely short RTO.

In this kind of instant restore model, a virtualization host can mount the backup image directly from the NAS, start the VM, and the live-migrate it back to the original primary storage platform—essentially eliminating the time it would normally take to restore the data back to primary storage in a traditional backup. In that sense, what is normally considered a DR resource can be leveraged to supply some level of BC functionality—especially if the NAS is located at a remote data center along with dedicated virtualization capacity.

SAN Replication— However, most BC approaches rely on a dedicated warm site—a data center housed in a remote location that contains all of the infrastructure necessary to keep the organization running if the primary data center goes down. That site is almost always coupled with some type of storage replication, usually a secondary SAN that's replicating data from the production SAN.

Because it essentially doubles your organization's investment in storage hardware and requires large amounts of low-latency bandwidth to connect both sites, this type of replication can be very expensive, but it is widely recognized as the most comprehensive business continuity approach available.

There are two kinds of SAN-to-SAN replication: synchronous and asynchronous. With synchronous replication, data is kept in precise synchronization between the production SAN and recovery SAN. By contrast, asynchronous replication captures snapshots of your data at specified intervals; if the primary storage system fails, you can roll back to a recent snapshot to minimize the data loss.

At first glance, it would appear that synchronous replication is better. Because data on the recovery SAN is always identical to that on the production SAN, the recovery time is exactly zero. The problem? If that production data becomes corrupted, the corrupted data will be replicated on the recovery SAN as well. (That's why disaster recovery is an essential component of even

the most robust business continuity system—you always want to have a reliable backup set.) With asynchronous replication, you can roll back to the most recent snapshot of data preceding the corruption event.

If you have stringent RTO requirements, you probably need synchronous replication to meet them. But you'll also need to make sure your DR resources are aligned correctly in case you need them. Depending on the storage platform you choose you may be able to adopt a hybrid approach, taking snapshots of the SAN production data and storing them on the recovery side SAN—effectively providing you with zero RPO as well as the ability to roll back to an earlier data set without having to restore from backups.

Cloud-based DR and BC—In scenarios where a large amount of Internet bandwidth is available, the public cloud can also provide a viable option for both DR and BC goals. In DR applications, cloud-based backup can provide a replacement for tape media or NAS replication in order to provide off-site backup protection. However, managing the security of backup data being placed in the cloud and ensuring that the backup data can be quickly and easily brought back on site can be a challenge.

The public cloud can also fill the role of a fully deployed hot recovery data center. Through the use of application-specific backup software or host-based replication tools, entire server infrastructures can be protected by storage and computing resources located in the cloud.

Whether you decide to opt for cloud-based DR or BC depends on a variety of factors, including the type of IT resources your organization has, whether you have the skill set in house to navigate the fundamentally different nature of cloud services, how much bandwidth your data requires, and how closely you must adhere to government regulations regarding information security.

TESTING AND COMPLIANCE

No matter what solution you choose for disaster recovery or business continuity, building a consistent testing and compliance regimen—and sticking to it—is vital. Without frequent testing, you can never be certain the data you thought was backed up can actually be restored, or that your secondary data center will come on line when the Big One hits.

This is where many organizations fail. Overworked and understaffed IT orgs may feel they lack the time and resources required for frequent testing of DR and BC. But that may also be due to poorly designed processes or a lack of sufficient automation. Making an extra effort to streamline these processes will help not only in testing these solutions but also in implementing them when a real disaster strikes.

In the ideal scenario, organizations will deploy their DR or BC solutions as part of their day-to-day operations—allowing them to both leverage their investments and test them at the same time. For example, you might use the DR system for a mission-critical database to populate a read-only database platform that's used for generating reports. If a problem arises with a report, that could be a sign your DR system isn't working as it should.

Or you might routinely transition workloads between a production data center and a recovery data center—spreading the load across both data centers when your need more performance while also demonstrating that your fail-over methodology is functioning properly.

Backups can and do fail silently. If you do not test them, you won't find out until it's too late—a situation nobody wants to find themselves in. You'll also have an excellent opportunity to evaluate whether you're meeting your organization's written RTO/RPO requirements, and to lobby for changes if you're not.

BC/DR: A BUSINESS NECESSITY

As businesses become increasingly dependent on data to survive, the importance of disaster recovery and business continuity measures cannot be overstated. Only through thoughtful and collaborative planning that involves all levels of the organization—both technical and non-technical—can you field a comprehensive disaster recovery and business continuity strategy that appropriately sets expectations and adequately protects the lifeblood of the business.