BUSINESS CONTINUITY FOR MISSION-CRITICAL APPLICATIONS

By

Jon Toigo
Managing Principal, Toigo Partners International
Chairman, Data Management Institute

SUMMARY

Unplanned interruption events, aka “disasters,” hit virtually all data centers at one time or another. While the preponderance of annual downtime results from interruptions that have a limited or localized scope of impact, IT planners must also prepare for the possibility of a catastrophic event with a broader geographical footprint. Such disasters cannot be circumvented simply by using high availability configurations in servers or storage. What is needed, especially for mission-critical applications and databases, are strategies that can help organizations prevail in the wake of “big footprint” disasters, but that can also be implemented in a more limited way in response to interruption events with a more limited impact profile. Technologies exist for creating a modular and flexible disaster recovery (DR) strategy; the trick is to be able to manage and orchestrate these technologies so that the right recovery services are provided to the right data and so that appropriate combinations of services can be brought to bear in response to the interruption event itself. DataCore Software’s storage platform provides several capabilities for data protection and disaster recovery that are well-suited to today’s most mission-critical databases and applications.

INTRODUCTION

Depending on the survey one reads, disaster recovery is first or second on the list of priorities among IT managers. This is heartening, since it reflects a bit of rejection of the canard advanced by some vendors in the industry that server clustering and storage mirroring (collectively referred to as high availability -- or HA -- architecture) are the only protection that a company needs to survive and surmount an unplanned interruption of their mission-critical databases and applications, and that disaster recovery or business continuity planning are obsolete.

As a foundation for their claim that "HA trumps DR," vendors taking this position usually cite statistics showing that up to 95% of annual data center downtime does NOT result from disasters with a broad geographical footprint – affecting the data center facility broadly or the geographical region where it is located – but is attributable instead to interruption events that are highly localized in their impact. These localized events may include application glitches, minor hardware failures, viruses and malware, and security breaches by hackers. Though often unstated, the 95% figure also includes the 30 to 40% of annual downtime resulting from scheduled maintenance. Regardless of the reality behind the statistics, many vendors recommend that the majority of effort in disaster recovery or business continuity planning be focused on high availability – that is, to techniques and methods for working around these types of limited scope scenarios.

The problem with "playing the odds" is that the 5% of disasters that have a more catastrophic result – the loss of a data center, or of milieu supports required to make the data center purposeful (power, communications, transportation, financial services, etc.) -- will likely wipe out a firm completely that has not prepared to respond to them. Percentage-wise, these big "D" disasters may not seem to account for much of the expected annual downtime in a firm. However, they appear to be happening with greater frequency.

Simply put, organizations cannot afford protracted downtime, or lengthy interruptions in access to data, whether they accrue to logical or localized impact disasters or "big D" disasters. Some firms go to great lengths to quantify the amount of downtime they can tolerate, establishing recovery time objectives. With respect to databases, recovery point objectives may be determined to reflect how many transactions the organization can afford to lose. The real parameter for assessing successful recovery is time to data: how long it takes for all recovery services to be implemented so that the application or database is back in operation. As a rule of thumb, a company denied access to its mission-critical data for longer than a few weeks is likely to be out of business within a year.

This reality may not be a pleasant one to confront, but it is part of the mandate associated with IT stewardship in most firms. Managing risk requires business-savvy business continuity planning.

The good news is that the approach for developing an effective disaster prevention and disaster recovery capability is fairly well understood. The tenets are simple:

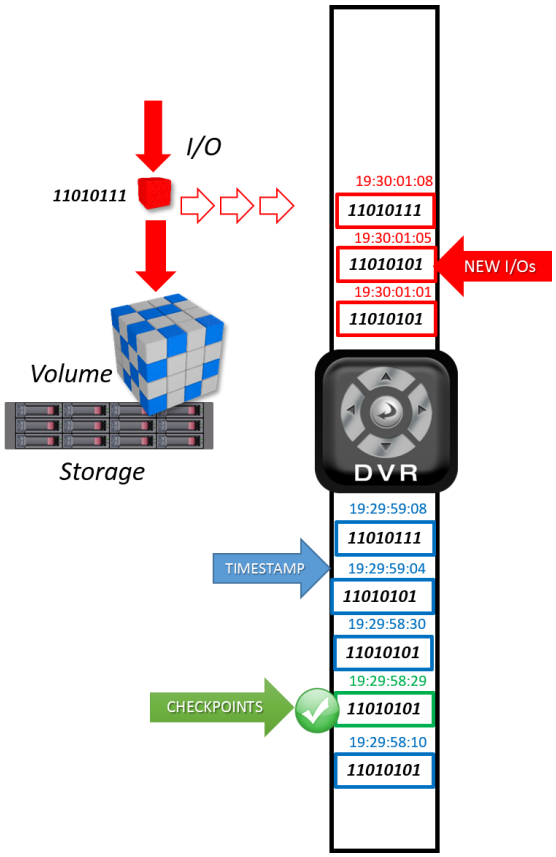
1. Identify the assets you need to protect and determine their criticality with reference to the business process that the asset serves. This is another way of saying that not all applications, databases, infrastructure and data are mission-critical, or in need of "always on" protection. Those that are critical and that may require high availability services need to be identified so that appropriate and cost-effective provisions can be made to protect them.
2. Select strategies for protecting data first. Without data, there can be no recovery. There are many techniques for protecting data, so you need to select those techniques that can be implemented most efficiently in your shop, that can be managed and tested (preferably on an ad hoc basis), and that map to the nature of the data being protected (including its access frequency, update frequency and useful life).
3. Select strategies for protecting applications and databases and their infrastructure components. Objectives for recovery should drive your choices, while available budget will likely pare down the available options. There is no perfect solution to the challenge of formulating a recovery strategy, but common sense criteria – like the ability to manage the recovery strategies centrally and test them non-disruptively and on an ad hoc basis -- should guide your selection process.
4. Failback is as important as failover. To avoid prolonging a disaster, or creating a second one, ensure that your continuity strategy includes automated support for failing back into a production mode once the disaster conditions subside.
5. Test everything as often as possible, both to spot gaps in your plans created by change over time in the business, the IT infrastructure and application environment, the storage infrastructure and the cadre of recovery personnel, and also to rehearse recovery teams in the recovery process itself. That way, if and when a disaster occurs, everyone will know their own job and how it relates to others.
6. Make the DR capability pay dividends in other areas such as data storage cost-containment, regulatory compliance, archive and data preservation, security, maintenance downtime reduction, etc. This will help to demonstrate the benefits of the DR capability to the organization, beyond serving as so much more insurance, and can help to maintain Front Office sponsorship of the program and its budgetary requirements.

These six tenets will produce an effective recovery capability. Without adhering to them, your best disaster recovery plan is an up to date resume.

PROTECTING DATA FROM LOGICAL AND PHYSICAL THREATS

One way to think about organizing a disaster recovery capability is to consider threats to data and operations using two categories: logical and physical. Logical threats include software glitches, human error, viruses and malware, and deliberate data vandalism by hackers.

While certain types of logical threats, such as denial of service attacks, may have a broad footprint, most are fairly localized. A transaction may be compromised or a file containing data may be corrupted. Recovery from these events would be best served by a continuous data protection (CDP) capability that records every write I/O (or change) with a time stamp so that a database, file or storage volume can be "rolled back to a point before the corruption occurred. This technology would also be useful for dealing with certain physical errors, such as non-recoverable bit errors caused by interconnect noise or component level errors on storage media that can corrupt individual data objects or, in some cases, render entire RAID groups unreadable.

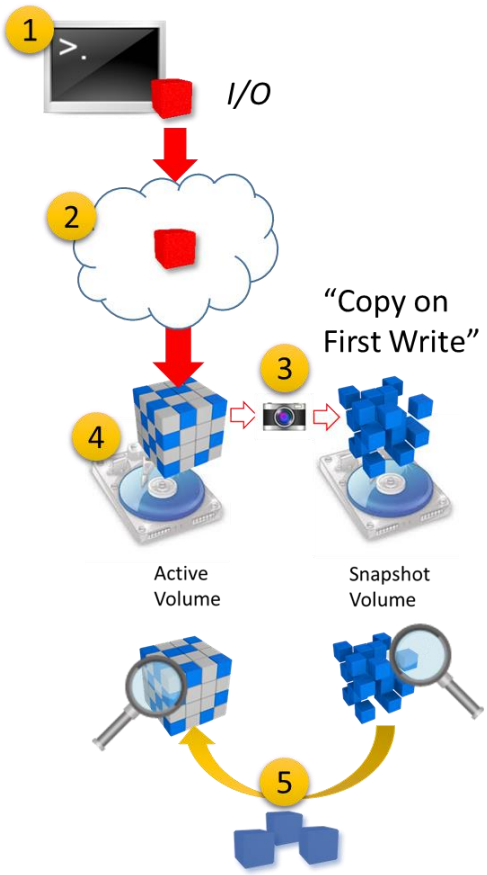


CDP technology takes many forms, with most approaches taking the form of sequential snapshots. The problem with snapshots or similar point-in-time mirror splitting techniques is that they require the quiescing of the database or application writing the data long enough to make the timestamped copy. Needed is a technology that can create CDP entries without interfering without application or database interruption.

DataCore Software provides write buffering in a non-disruptive way that provides the ability to roll back write history to a particular point in time. CDP buffers can be used to create a rollback volume that can be mapped to a specific server or application if its production volume becomes corrupt. This is actually a more efficient approach than more disruptive approaches, such as point-in-time mirror splitting or volume snapshot, preferred by some hardware and operating system vendors.

Of course, CDP isn't the answer for every kind of logical failure. It works when you simply want to return to a point in time and continue processing. But if the database or filestore has been damaged in a way that requires the replacement of blocks of data, you need a copy.

The challenge that every database admin knows is that making a copy requires the quiescing of the database or application that generates data – long enough to copy the data. In many applications, the copy process is extensive and takes quite a while. With DataCore, two options exist for making copies of changed blocks that can be used for fast restoral.

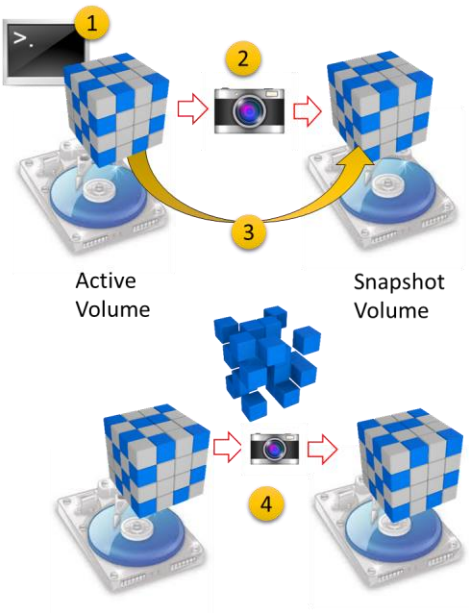


DataCore provides technology for both incremental snapshots and snap clones that can be stored near the active volume and accessed rapidly to restore from certain types of faults. With DataCore incremental snapshots, when a chunk of data is about to be written to a block location (step 1 in the illustration), this write operation is paused in a buffer (step 2) long enough so that the chunk that is about to be changed can be copied to a storage volume designated as a snapshot repository (step 3). The write then proceeds (step 4) with minimal disruption to the application or database

This Copy on First Write approach establishes over time a series of snapshots that can be used to restore damaged or deleted block data in the active volume. As shown in step 5, restore involves an automated comparison of blocks on the active volume and blocks in the snapshot volume and only those blocks that are needed for recovery are shipped to update the active volume. It's a wonderful timesaver.

In some cases, especially in some in-memory databases, incremental snapshots alone are not deemed sufficient protection against a logical fault. For reasons of both data protection and maintenance, users may elect to set up a snap clone.

A snap clone is an image of the entire dataset that is being protected. First, the application or database must be quiesced. Then the entire storage volume is snapped to the snapshot volume. Then, the application or database can be restarted while the data corresponding to the pointers in the snapshot is copied and migrated over to the snapshot volume, creating a clone of the original. The migration is transparent to the application and happens in the background, managed by DataCore so as not to impose any application or database latency.

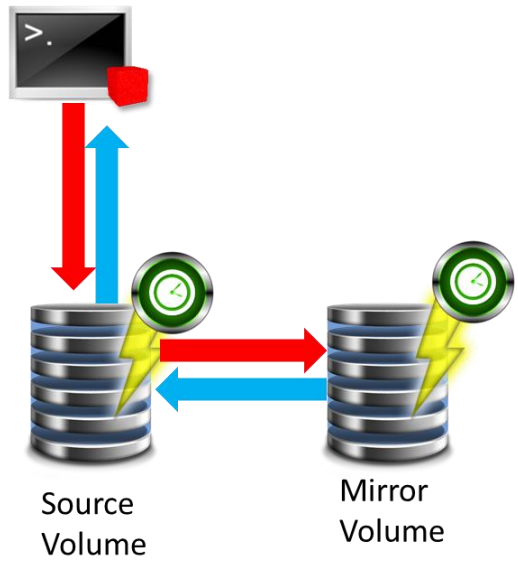


Subsequently, as shown in step 4, the clone can be kept up to date using incremental snapshots. The clone can be used as an alternative source volume in an emergency, or, for databases, it can be used to test patches and reorgs before they are applied to the production database. As with incremental snaps, a logical fault in the active volume can be quickly diagnosed by DataCore so that only the repair blocks that are needed will be shipped to the active volume to repair a logical error. Again, a great timesaver – and one that can be accelerated even more by DataCore's Adaptive Parallel I/O capability

In addition to logical threats, planners must also consider physical threats to data that run the gamut from localized equipment failures (such as failed disk drives, failed server power supplies, failed storage arrays or switches, etc.) to more data center wide faults including power outages, public network outages, cloud service provider outages, facility outages whether due to internal problems (pipe leaks, fires, etc.) and even "milieu-level" events such as nuclear or chemical events, terrorism, etc.

Protecting your data from any of these physical threats generally requires that you make a copy of the data and store the copy in a manner that does not expose it to the same threat that might compromise the original. Some DR experts recommend a 3-2-1 approach: make three copies of the data, on at least two different types of media, and store one copy off-site at a distance sufficient to keep it out of harm's way.

DataCore has a clean strategy for making three copies of data, including a remote copy that involves a mixture of local mirroring and remote replication. The local mirroring functionality is different from snapshot-based approaches in several ways. For one, a snapshot volume is usually in the same node as an active volume, while a mirror volume is usually located behind a separate node to avoid single points of failure.



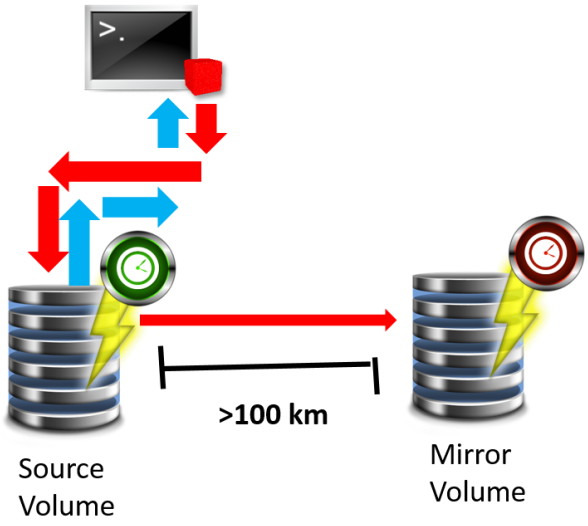
Secondly, a mirror is usually set up to operate in a mode best described as a two-phase commit. That is, data from the application is written to the target volume, then DataCore executes a mirror of the data write to the mirror volume. This write is acknowledged when it is made, then the application is cleared to send more data I/O. DataCore's Adaptive Parallel I/O helps to get the I/O moving through infrastructure at top speed, and DataCore's virtualized infrastructure guarantees fast performance on the first and second (mirrored) writes.

In the event of a catastrophic failure in the first node that effects access to the source volume, the application is automatically directed to the mirrored volume so that processing can continue. This configuration also enables maintenance work to be performed on the source volume and nodal hardware without interrupting operations. Moreover, DataCore enables testing of the failover configuration and buffers I/O to facilitate fail back, which is an innovation worth having.

If the impact of a disaster event expands beyond a particular node of server or storage hardware, it might be necessary to recover data on storage placed some distance away. DataCore's mirroring is extensible up to between 80 and 100 km (about 50 miles). It works well over MPLS and other Metropolitan Area Networks. Even at that distance, the data on the source and mirror volumes will remain synchronized.

However, in recent years, we have seen disaster events with an even wider destructive radius. That disaster potential requires mirror data to be separated from its source by greater than 100 km. Because of distance induced latency, an asynchronous, rather than synchronous, replication technology is required.

DataCore provides everything you need to do asynchronous replication over extended distance, across a Wide Area Network connection for example. Asynchronous replication is similar to synchronous mirroring, except that application writes are not delayed until a write to the remote volume is acknowledged. This would introduce too much latency into the performance of applications or databases that are generating I/O. Instead, DataCore handles the delivery of the remote writes and alerts if an error occurs.



Many companies use local mirroring and remote asynchronous mirroring in combination – a strategy sometimes called "multi-hop" mirroring. This is a particularly effective strategy if you conceive of data protection as a set of layered services that can be leveraged in different ways in response to different interruption events. The important thing is that all of the strategies are managed centrally and are easy to associate with storage volumes created using DataCore.

Another advantage, besides centralized management and orchestration of services, is the workload and hardware neutrality of the solution. DataCore will provide storage and services to any workload, whether virtualized or not. And DataCore can deliver protective services to any workload, regardless of the hypervisor it may operate with. Most backup products are not hypervisor agnostic.

From a hardware perspective, DataCore's technology eliminates the most of the challenges that exist with hardware-centric mirroring. Brand X array vendor often provided mirroring functionality, but only if the customer had an identical Brand X array to serve as the mirror volume. With DataCore, hardware is virtualized. Data can be replicated to any storage volume that is created and presented by DataCore.

OTHER ADVANTAGES

DataCore's technology reduces the complexity of building a business-savvy business continuity strategy. It delivers a set of services that can be applied to the storage volumes to which application and database I/O are directed. Different services can be applied to different volumes based on the attributes of the data (access requirements) and the criticality of the business process that the application or database serves (time to data requirements).

DataCore data protection services are easily deployed and may be modified at any time. They leverage the extraordinary capabilities of DataCore infrastructure including Adaptive Parallel I/O for the industry's leading I/O speeds, intelligent adaptive caching, and automated infrastructure load balancing and thin provisioning. Moreover, DataCore can replace a multiplicity of software products you are currently using for synchronous mirroring, incremental snapshots, CDP and drive cloning, or, if you prefer, it can accelerate and augment the functionality of products that you wish to keep.

On the hardware side, DataCore takes the challenges out of mirroring and replication in a heterogeneous hardware environment. It's replication and mirroring capabilities are truly "any to any." This enables you to keep gear in service longer and to take advantage of less expensive hardware options.

DataCore also supports multiple storage topologies and can replicate across un-like infrastructure. DataCore is already used extensively by server manufacturers to create best of breed hyper-converged infrastructure appliances, but DataCore's foundational technology is also used in thousands of companies to provide robust and scalable SAN infrastructure. Data protection services can span these different infrastructures in ways limited only by the creativity and requirements of the user.

Perhaps most important is DataCore's ability to provide ad hoc testing capabilities. At any time, you can fail a mirror over, or test a remote asynchronous volume, or check a clone volume. Given that testing is the long tail cost of business continuity and disaster recovery planning, the value of ad hoc testing cannot be understated.

For a high performance, comprehensive, easily deployed and managed, multi-layer data protection capability, DataCore Software can provide the solution that most firms are seeking. If you are undertaking a business continuity planning project, DataCore Software is worth your careful consideration.