White Paper

Considerations for a Comprehensive Data Protection Architecture—and NetApp's Approach to Delivering It

By Jason Buffington, Senior Analyst; and Monya Keane, Research Analyst

November 2015

This ESG White Paper was commissioned by NetApp and is distributed under license from ESG.

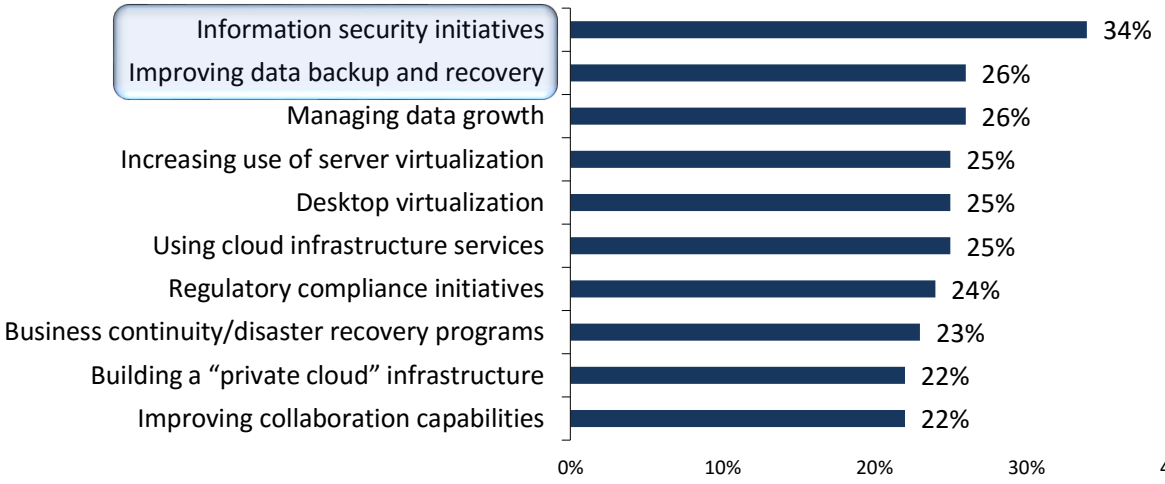
Contents

Introduction

Organizations continue to be increasingly dependent on their IT resources. As a result, IT teams tend to embark on myriad IT transformations every year that relate to securing and protecting data (see Figure 1).[1]

Figure 1. Top Ten IT Priorities for 2015

Top 10 most important IT priorities over the next 12 months. (Percent of respondents, N=601, ten responses accepted)



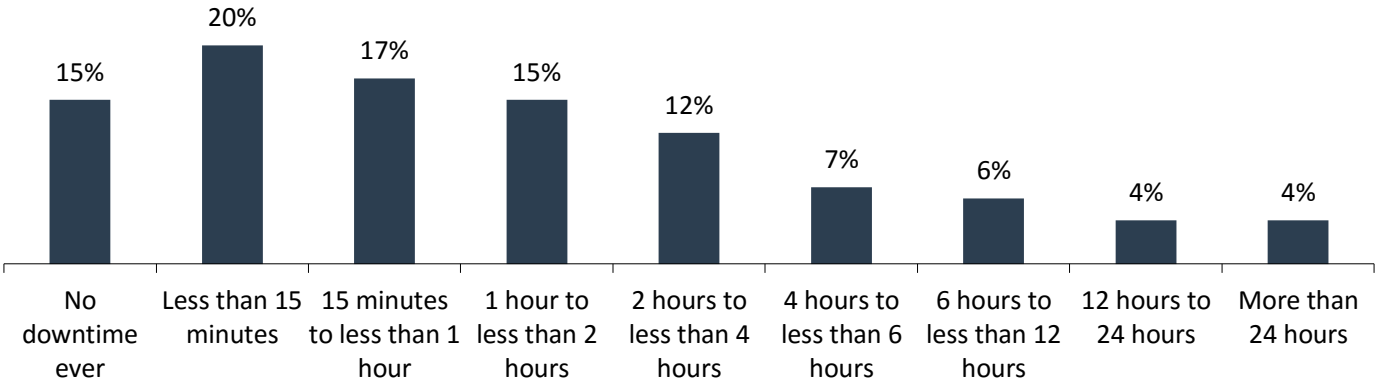
Source: Enterprise Strategy Group, 2015.

Many production-enhancing initiatives appear in the responses in Figure 1, but the two most commonly cited IT priorities both relate to protecting data and infrastructure in some way. Additionally, tactical-level (i.e., "must-have") backup-related priorities are accompanied by strategic-level priorities as well—priorities pertaining to business continuity and disaster recovery (BC/DR) to ensure IT resource availability.

Because of organizations' great dependence on IT, today's servers often have such low downtime tolerance profiles (see Figure 2) that traditional backup, even when well implemented, rarely satisfies their availability requirements.[2]

Figure 2. Percentage of Production Servers/Services that Fall Within Each of the Intended Recovery Times

Considering all of your organization's production applications/workloads (including both "high-priority" and "normal" workloads), approximately what percentage of these production servers/services fall within each of the intended (i.e., target or "desired" recovery time RTO/SLA versus what your organization has actually delivered) recovery times listed below? (Mean, N=391)



Source: Enterprise Strategy Group, 2015.

[1] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.

[2] Source: ESG Research Report, *The Evolving Business Continuity and Disaster Recovery Landscape*, to be published.

It's interesting to note that a combined 52% of servers at the organizations surveyed by ESG have a downtime tolerance of less than one hour. Another one-fourth (27%) have a downtime tolerance between one and four hours.[3]

Historically, meeting those tight SLAs would necessitate whatever heroics the IT organization could cost-justify. But those heroic measures were typically applied only to the top 5-10% of those server infrastructures.

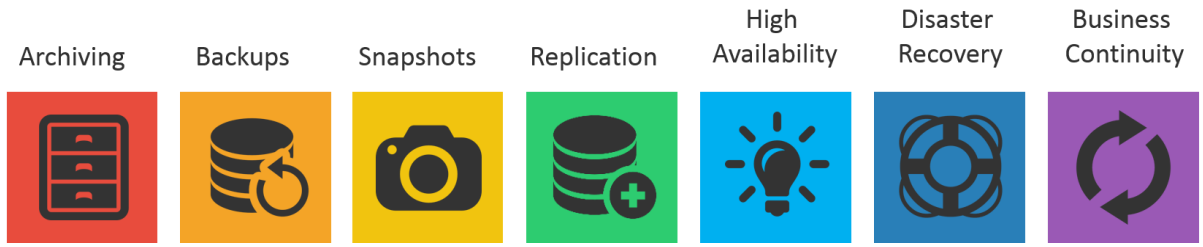
In 2015, with so many organizations operating under extremely high-uptime SLAs for so many servers/applications, even "heroics" are no longer sufficient. In most organizations today, availability and agility architectures have to be the norm instead.

With so many organizations operating under extremely high-uptime SLAs for so many servers/applications, even "heroics" are no longer sufficient. In most organizations today, availability and agility architectures have to be the norm instead.

Plan for a Comprehensive Data Protection Strategy

Recognizing that their business units' requirements for SLAs are often unachievable with backup alone, IT organizations should plan for a more comprehensive approach to data protection (see Figure 3).

Figure 3. The Spectrum of Data Protection



Source: Enterprise Strategy Group, 2015.

The illustration tells two stories:

1. Each data protection initiative is complementary to the others. Backups alone are not enough, but they certainly continue to be a mainstay requirement.
2. Backup should absolutely be supplemented with snapshotting, replication, and availability technologies that allow for a broader range of agility.

Even distilling the data protection spectrum to its true core—backups plus snapshots plus replication—there are key differences that highlight the complementary nature of the approaches:

- **Backups** are partial and full copies of data that *reside outside of the production stack*—often in a retention-optimized storage pool—providing restoration capabilities from a myriad of points in time.
- **Snapshots** are incremental versions of the data *within the production storage system* itself that have the ability to revert to a relatively recent point in time almost immediately.
- **Replication** creates an immediately usable copy of the current (or near-current) data *at an alternate location* for BC/DR or availability purposes as well as non-data protection business reasons, such as test/dev or analytics.

Similar distinctions can be made for the archival and availability mechanisms, which have varying methods of data movement, storage, and usability characteristics. The key point is that they are complementary to one another and should not be considered replacements or evolutions of their adjacent "colors."

Also of importance is the assertion that data protection should be accomplished as a single strategy using multiple methods. A comprehensive approach really should involve combining data protection mechanisms *within the primary infrastructure*, not simply as an after-the-fact "add-on."

[3] ibid.

Considering NetApp's Integrated Data Protection Portfolio

One company that has always invested in such a built-in approach to data protection is NetApp. For example, to meet the universal need for traditional backup, NetApp has partnered with leading backup software vendors, forming integration-centric relationships with Commvault, Veritas, Veeam, and Catalogics. But as IT decision makers are awakening to the reality that backups alone are insufficient for today's IT resiliency requirements, they are rediscovering storage-centric or storage-enabled protection scenarios that now integrate for a more holistic and agile data protection and recovery experience; namely snapshots and replication.

Snapshots and Replication

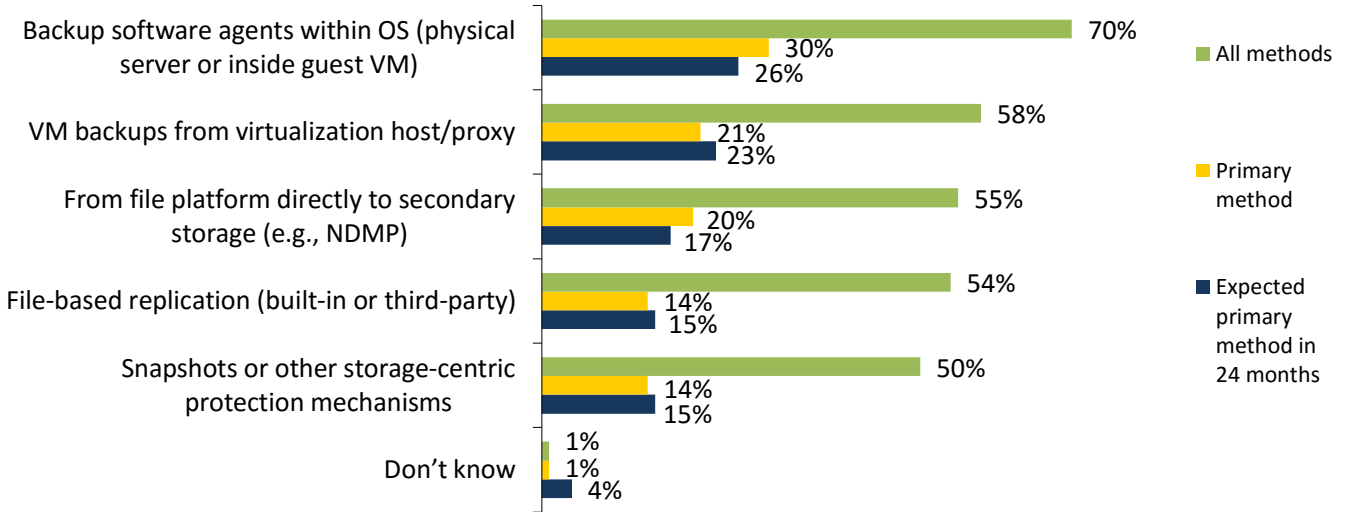
As mentioned, many organizations require a level of *recovery* that is not accomplishable via a *restore* from backup. However, two common approaches for very quick recovery do exist—snapshotting and replication:

- **Snapshots**, typically derived from within primary storage, allow for an almost immediate (i.e., within seconds) reversion of data to a previous but relatively recent state within the same primary storage device.
- **Replication** provides a second copy that is nearly immediately usable from a secondary set of storage somewhere else within a private, hybrid, or cloud infrastructure.

As is the case with backups, snapshotting and replication serve as complementary components of a broader data protection strategy. But both differ from backups by rebalancing longer-term retention with near-immediate usability; the data is in a native state as opposed to being in some kind of backup pool or optimized secondary container. As Figure 4 shows, both methods are often used as a supplement to backup within file/storage services.[4]

Figure 4. Methods Used to Backup File Servers and NAS Platforms: Today and 24 Months from Now

Which of the following methods are used to back up file servers and NAS platforms? What is the primary way of backing up file servers/NAS? How do you expect this to change—if at all—over the next 24 months?



Source: Enterprise Strategy Group, 2015.

Clearly, many organizations recognize the importance of adding a degree of agility and availability beyond what backup alone can offer. They are supplementing their backups with technologies including snapshotting and replication, and ESG research indicates that the usage of these alternative approaches to data protection will grow even more prevalent over the next two years.

[4] Source: ESG Research Report, *Data Protection Personas and Methods*, February 2015.

NetApp's Answer Is Snapshots, Which Enable SnapMirror, SnapVault, and More

Arguably, no IT name is more synonymous with snapshot technology than NetApp. After all, the company coined the term "snapshot" and is the holder of the official U.S. registered trademark.

In addition to partnering with traditional backup software offerings, NetApp's ***SnapVault*** creates transportable snapshots, effectively offering a block-based alternative to backups. SnapVault is only one of the derivative data protection capabilities that come from the NetApp platform. Other storage systems provide what might be perceived as "equivalent" snapshotting technology, but even NetApp's original file systems were developed from the ground up with snapshot enablement *truly* built in. Today, NetApp snapshots are not only agile and efficient, but they also support several other data protection objectives thanks to their architectural portability.

Efficient replication of data via NetApp ***SnapMirror*** was one early innovation made possible by NetApp's pioneering snapshot technology. With it, different NetApp storage solutions synchronize data transparently across vast distances while preserving the data's integrity and enabling additional levels of agility. The SnapMirror for SVM feature enables replication at the level of Storage Virtual Machine (SVM), NetApp's construct for a virtualized storage container spanning all nodes in a NetApp FAS cluster. By setting data protection policies at the SVM level, customers can easily manage SLAs at tenant-level for shared environments, or manage replication for thousands of LUNs and volumes by consolidating them into a small number of SVMs.

As mentioned, snapshotting and replication both store data in a native, nearly immediately usable state. They are excellent examples of alternatives to traditional restoration from backup as a means to achieve levels of availability (speed) and agility (within primary storage or from a secondary site) in ways that backup alone simply cannot. As such, it makes sense that SnapMirror and SnapVault use a unified mechanism for replication and transport.

Organizations leveraging a combination of snapshots and replicas from NetApp should see higher levels of agility during restoration as well as higher availability than is possible with backups alone. Consequently, the IT organizations using these methods will be more able to achieve the SLAs that their business units require.

Availability

Although a range of data recovery, restoration, and retrieval approaches exist, all are reactive. And all usually have to be manually invoked to help the business units resume productivity. When considering the comprehensive data protection and availability approach depicted in Figure 3, note that the four left-side technologies are reactive, while the right three are proactive. Those right-side HA and BC/DR technologies, approaches, and methods can help to reduce downtime to *nearly* zero.

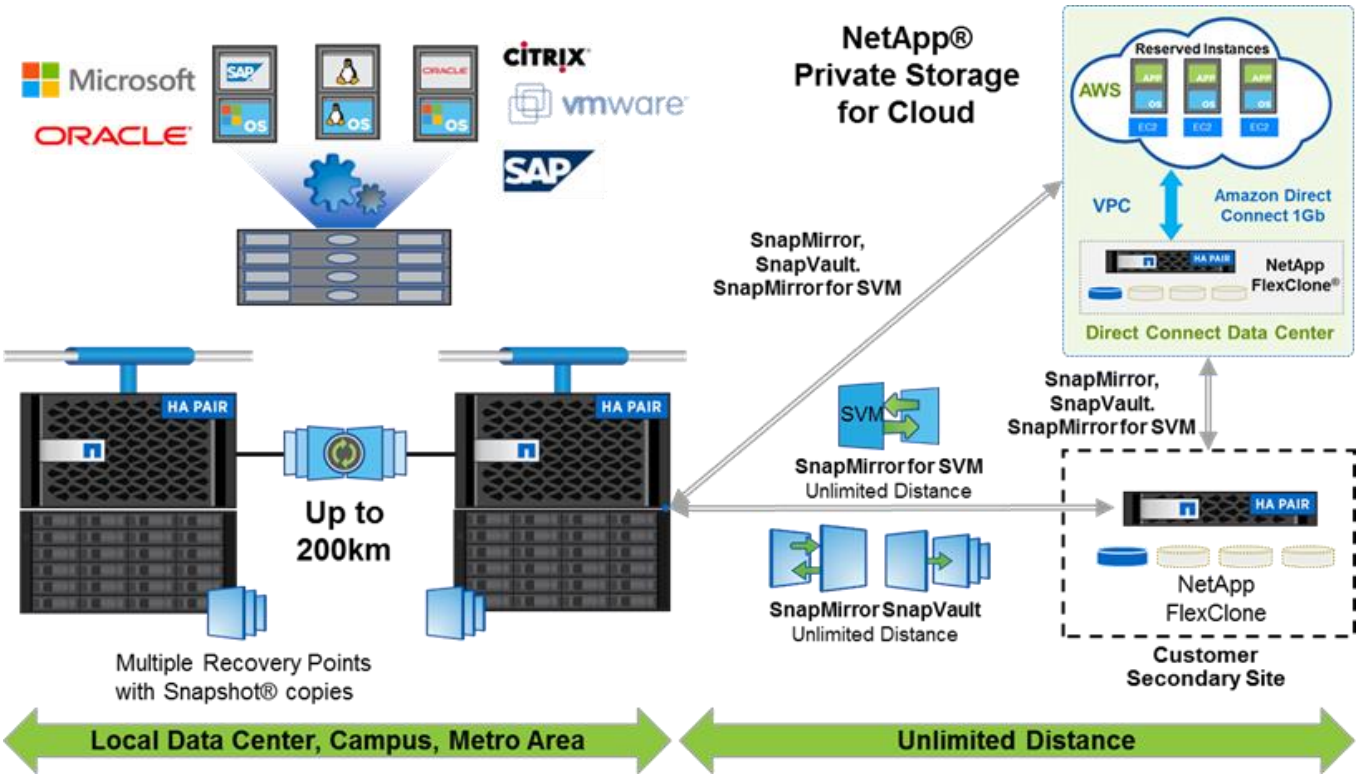
NetApp's Answer Is MetroCluster

NetApp's approach for availability (HA) and IT resiliency (BC/DR) is powered by the data portability within NetApp snapshotting and replication. NetApp also has added another layer on top—in the form of a technology called ***MetroCluster*** (see Figure 5).

True availability, particularly in the context of BC/DR, has to be accomplishable across geographic boundaries, with availability being the primary driver. Synchronous replication using MetroCluster adds a BC component that protects data beyond the data center with zero data loss and zero downtime. MetroCluster combines array-based clustering with synchronous replication up to 200 kilometers away. And, because it is part of clustered Data ONTAP, it doesn't require an external device to manage it. After MetroCluster is set up, it no longer requires ongoing configuration whenever IT admins add, remove, or resize LUNs or volumes.

Consequently, all changes made on one side of the cluster are automatically replicated on the other. (Even snapshot copies created on one side are automatically created on the other.) It is another example of seamless integration, and it has significant benefits. Specifically, in the event of a switchover to the remote site, it might be necessary to restore some data from an earlier snapshot copy: It will just "be there" as if nothing had happened.

Figure 5. NetApp MetroCluster for Availability Enablement within a Data Protection Strategy



Source: NetApp, 2015.

SAN and NAS protocols are both supported, as are deduplication, compression, and all the other features of clustered Data ONTAP (i.e., local non-disruptive operations, upgrades, and data mobility). As Figure 55 shows, MetroCluster integrates seamlessly with backup and long-distance asynchronous replication to provide a comprehensive protection solution providing both near-term failover and ongoing business continuity. BC is the ultimate goal of all data protection—ensuring accessibility of data and systems to maintain and bolster business productivity.

While manually invoking snapshots and replicas to resume productivity will meet the "durable IT" goals that traditional backup cannot (i.e., between one and four hours), only by leveraging proactive availability technologies can one hope to meet SLAs of less than 15 minutes. So, considering the NetApp technologies discussed so far and the SLA requirements shown in Figure 2:

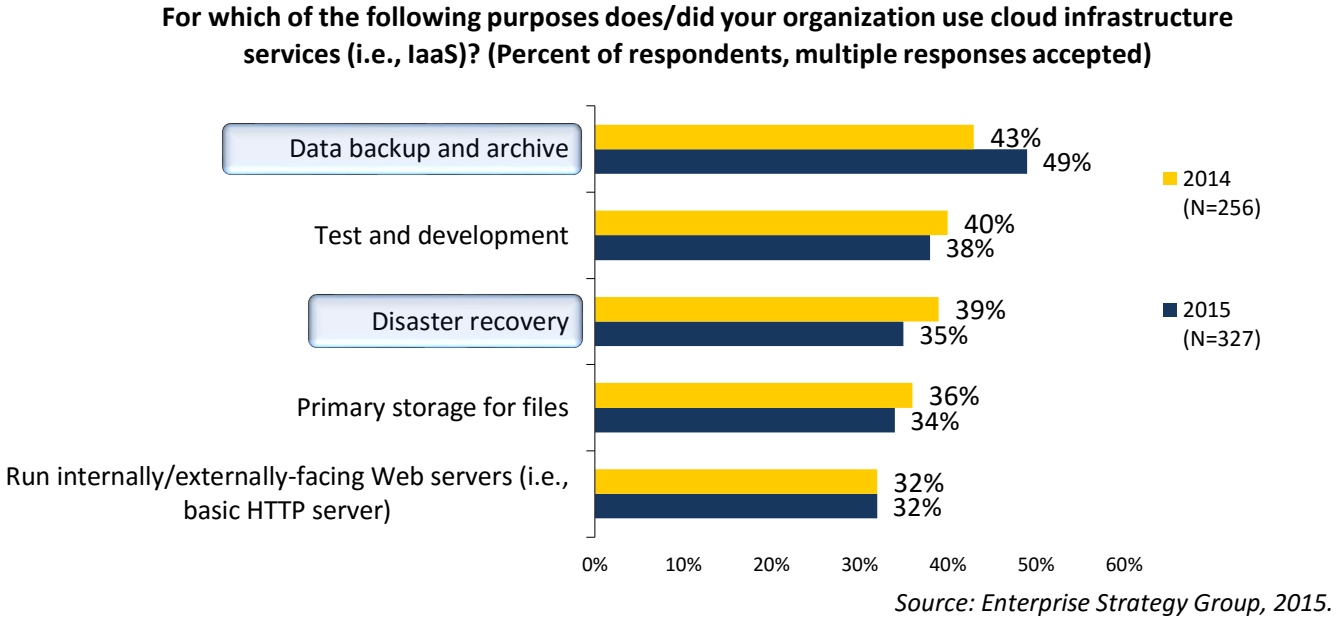
- For servers and services with a downtime tolerance of less than 15 minutes, a proactive approach such as MetroCluster is most appropriate.
- For servers and services with a downtime tolerance of between one and four hours, snapshots and replicas are ideal.
- For servers with longer downtime tolerances, backups may be considered to be the only required level of protection (bearing in mind that backups should be used to protect all the data within an infrastructure).

Every Data Protection Strategy Should Include the Cloud

Organizations of all sizes can benefit from cloud-based services that offer an alternative cost model, (likely) higher reliability, and the peace of mind that comes with having a secondary site. The only question remaining is which "cloud" to use: a cloud that is self-managed, a public cloud, or a hybrid approach with multiple providers/multiple facilities.

In any case, "the cloud" should be a part of every IT transformation conversation, particularly any discussion of enhancing a data protection availability and agility strategy. Not surprisingly, two of the top three most-cited reported usage scenarios for organizations using cloud-based infrastructure *are data protection centric* (see Figure 6).[5]

Figure 6. Top Five Purposes for Using Cloud Infrastructure Services



Source: Enterprise Strategy Group, 2015.

NetApp's Answer Is AltaVault

Certainly for those environments with NetApp filers in both their private and public/hybrid clouds, the same SnapMirror and SnapVault technologies described earlier apply. But for heterogeneous environments involving multiple storage solutions or myriad backup software offerings, NetApp has invested (through acquisition) in what is arguably one of the most seamless ways to immediately add cloud storage to any IT infrastructure.

NetApp ***AltaVault***, formerly known as Riverbed SteelStore, is a physical or virtual appliance that provides local storage nearly immediately synchronized to the public cloud of one's choice. There are two core tenets of the AltaVault strategy that organizations should actively consider as part of a broader data protection strategy:

- To provide maximum flexibility, NetApp offers AltaVault appliances in both physical appliance and virtual appliance form factors for both VMware and Hyper-V environments.
- While many other storage hardware and data protection software vendors continue to struggle with the best approach for cloud connectivity, AltaVault presents itself simply as network-attached storage (NAS), with either CIFS or NFS shares that almost every data protection software knows how to leverage.

Those tenets are enhanced by Riverbed's pioneering approach to WAN optimization, allowing organizations to very efficiently replicate data to the public cloud of their choice.

[5] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.

Modernizing Protection When You Modernize Production

Often, data protection is considered an "add-on" to whatever IT infrastructure is already in place. NetApp, however, has always been mindful of data protection enablement. After pioneering snapshot technology, and then building on it with offerings associated with replicas, clones, and availability clusters, NetApp continues to innovate its approaches to integrated data protection.

What to Consider When Planning Your Next Data Protection Strategy

ESG recommends considering three factors when evaluating your holistic approach to IT infrastructure, including your data protection methods and strategy:

- **Quantify the SLAs for your server infrastructure as they relate to possible impacts to business productivity.**

  While your servers' exact SLAs may vary from those depicted in Figure 2, you may be surprised to learn how many of your servers require "better than backup" to achieve the availability levels the business units are depending on.

- **Plan for a comprehensive data protection strategy.**

  As Figure 3 showed, a single data protection strategy should include not only reliable backup and recovery, but also snapshotting, replication, and availability technologies to achieve the availability and agility that today's organizations require.

- **Solve for availability as the primary goal.**

  All of the protection methods described earlier (backups, snapshots, replication, and clustering) are tied to their own levels of downtime tolerance, with progressively declining RPO and RTO values. Although not every server requires a "zero-RTO, always-on" architecture, your strategy should be grounded in the single goal of ensuring that availability per server is higher than the tolerable downtime impact.

The Bigger Truth

Although "backup" (offline copies) will invariably always be the basis of data protection, most organizations are realizing that backups alone are simply insufficient. Organizations of all sizes are dependent on their data, and no one can afford the downtime and data loss that legacy backups require. Instead, consider a Good, Better, Best model:

- **Good** means using a modern backup solution that supports contemporary workloads and integrates with the larger environments.
- **Better** means complementing backups with snapshots and replication to achieve significantly higher SLAs for the business units and their data.
- **Best** means proactively heightening availability through clustering and BC/DR processes, instead of relying on reactionary recovery mechanisms alone.

But the advice to add diversity of data protection mechanisms should be not be misinterpreted as a recommendation to create additional silos of information through myriad disconnected methods. Instead, data protection is really only achievable through a single strategy that is enacted by interoperable parts. One further way to simplify and enhance the encompassing approach to data protection is to consider ways of gaining that IT resilience as "built in" instead of "bolted on." Certainly, when it comes to snapshots, as well as storage-based replication and the broader range of data protection initiatives, one company to watch or perhaps rediscover is NetApp.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com