Converged Infrastructure and Data Protection Should Be One Conversation, and Hitachi Is Delivering It

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

May 2014

This ESG White Paper was commissioned by Hitachi Data Systems and is distributed under license from ESG.

© 2014 by The Enterprise Strategy Group, Inc. All Rights Reserved.
Introduction: Virtualization and Protection Must Be One Conversation

IT modernization is a journey, not a destination. “Virtualization” is the means, but sometimes, it’s also an obstacle. IT organizations that are proudly on their IT modernization journeys and embracing virtualization too often discover that as they proceed from 20% virtualized to a modern converged infrastructure (see sidebar), their ability to protect data using legacy methods diminishes.

Other industry terms and configurations further complicate any explanation of what a highly virtualized ecosystem looks like. For example:

- A **software-defined data center (SDDC)** refers to an IT infrastructure in which virtualization mechanisms have abstracted the hardware components—making the stack agile enough so that IT professionals aren’t bound by physical components that have to be routinely configured and managed individually.

- The **“private cloud” versus the “public cloud”** can mean different things to different people based on the layers of the stack over which they have administrative control. For instance, dedicated/exclusively controlled virtualization hosts within a third-party service provider could be considered a “private” cloud residing within a “public” data center.

As a highly virtualized infrastructure matures, its systems invariably create a highly converged architecture that ESG refers to as an integrated computing platform (ICP).

The **good** news is that ICPs aren’t meant only for the largest of enterprises. They are available in configurations that enable midmarket organizations to continue their virtualization modernization journeys as well. (Of course, enterprises do enjoy a variety of converged system options.)

The **bad** news is that as the density and efficiency of a converged stack increases, the importance of adequate data protection grows. However, the ability to use legacy approaches diminishes, going from simply being no longer suitable to negatively affecting the production environment. That bad outcome happens when an organization modernizes its production infrastructure without modernizing its data protection mechanisms.

Virtualization Trends Are Affecting Data Protection Trends

In an agile IT infrastructure—one underpinned by a highly virtualized infrastructure such as a converged system—yesterday’s backup solutions are inadequate. In part, it’s because modern virtualized IT infrastructures tend to demand rather strict SLAs, and to meet those SLAs, IT teams and managers need to combine the capabilities of snapshots, replication, and backup. In other words, they need to create a unified architecture offering a range of data protection capabilities.

Today, some IT environments are already heavily virtualized, with 61% or more of their potential “candidate” x86 servers having undergone virtualization. But a much larger percentage of the organizations surveyed by ESG seem still to be on the journey—they are closer to the “20% to 60% virtualized” point in terms of progress (see Figure 1).¹ According to these survey results, at least, virtualization still has plenty of room to grow and mature into a truly ubiquitous IT infrastructure component.

---

On the other hand, those environments were less virtualized even just a few years ago (perhaps less than 20%), and any virtualization-related activities they were pursuing mainly centered on consolidating servers.

In those days, the “old, traditional” methods of data protection were still viable. But as server environments move from 20%, to 50%, to well over 60% virtualized—often at a large scale and typically best achieved within converged systems—organizations need to think differently about data protection and availability.

Specifically, they need to be thinking about how to improve their levels of data protection agility in:

- **Server virtualization**—Server virtualization boosts the agility levels of hosts in an environment. It also demands protection methods that are different from the methods suitable for decentralized physical machines.

- **Storage virtualization**—One hallmark of storage virtualization is the added agility it yields. But a storage virtualization project can also be a stimulus for an IT organization to investigate some of the newer, very compelling data protection solutions available (exemplified perhaps nowhere better than within the Hitachi Data Systems storage portfolio).

With converged systems being designed to remove the I/O bottlenecks between compute, storage, and networking, an ICP becomes the epitome of an agile infrastructure that can transform and accelerate business operations. But with so much riding on the converged system, its data protection strategy requires a combination of snapshotting, plus replication, plus virtualization-savvy backup. All of those activities, particularly when combined, deliver IT agility. This paper explores the various technical and business requirements for protecting converged systems and other highly virtualized environments, and it analyzes elements of the HDS portfolio as viable answers to what those environments need.

**The Challenges of Legacy Backup Methods**

As mentioned, legacy data protection approaches for converged systems can be quite problematic in an enterprise-scale, heavily virtualized environment. As Figure 2 shows, concerns and worries tend to center on recoverability and less-than-ideal visibility into the status of the environment. One-fifth (21%) of the IT pros responding to an ESG survey cited the actual recoverability of data as their primary virtual server data protection challenge, and four of
the top five challenges reported related to visibility (i.e., validating success, identifying bottlenecks, and troubleshooting).²

**Figure 2. Top Five Challenges Protecting Organization’s Virtual Server Environment**

Which of the following would you characterize as challenges for protecting your organization’s virtual server environment? Which would you consider to be your organization’s primary virtual server data protection challenge? (Percent of respondents, N=325)

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Primary virtual server data protection challenge</th>
<th>All virtual server data protection challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recoverability of data</td>
<td>21%</td>
<td>41%</td>
</tr>
<tr>
<td>Validating backup success</td>
<td>11%</td>
<td>44%</td>
</tr>
<tr>
<td>Validating recovery success</td>
<td>10%</td>
<td>44%</td>
</tr>
<tr>
<td>Identifying factors impacting the performance of backup operations</td>
<td>9%</td>
<td>27%</td>
</tr>
<tr>
<td>Simplified views across virtual infrastructure</td>
<td>9%</td>
<td>27%</td>
</tr>
</tbody>
</table>


Here’s another reason why legacy backup isn’t working: Under the old model, if you had ten applications to protect on ten servers, the backup process had access to ten sets of CPU, memory, and network connections. In a virtualized model, those ten applications plus the hypervisor share the *same* CPU, memory, and network connections—really squeezing the availability of those resources during a backup. The amount of data to protect may not have changed, but the amount of hardware-resource headroom did. Virtualization absolutely improves hardware utilization, but the downside is less resource availability for a resource-intensive process such as backup.

**Combining Snapshots, Replication, and Backup**

A modern infrastructure, particularly a converged system that is empowering whole farms of production services, demands modern data protection featuring a combination of snapshotting, replication, and backup capabilities. Each of those activities has its own purpose and is not a replacement for the others:

- **Snapshotting** enables an IT team to bring data back online *quickly*. It is a speed-related, availability-first protection mechanism.
- **Replication** shores up an organization’s data protection resiliency and agility-centric goals. It is a distance-based protection mechanism.
- **Backup** is a means of retaining data or preserving iterations of applications, servers, and datasets. It is a version-centric protection mechanism.

² Ibid.
IT organizations seem to be well aware of the distinctions between, and merits of, each activity. According to ESG research, less than 10% of organizations with highly virtualized environments do VM backups only. The other 90%-plus engage in some combination of snapshotting, distance replication, and backup (see Figure 3).³

**What Should Virtualization Protection Look Like?**

Anyone (or almost anyone) can reliably back up a simple VM. The real question is, how agile and granular will the subsequent restore capability be? Recoverability is priority number one, and with that fact in mind, it’s wise to look for a data protection architecture that:

- **Is designed and optimized for protecting a modern infrastructure.** Just as the best virtualization platforms are converged systems because of their integration and manageability (as well as performance), the best data protection solutions are easy to use and offer integrated manageability. That is, they allow an IT organization to integrate backup tools with replication and with the primary storage’s snapshot capabilities. In this way, an IT admin can invoke near-instant recoveries where a given VM “lives” as well as where that VM is being protected. And because you are integrating backup, snaps, and replication, you should look for a common user interface—a single lens through which you can manage backup, snapshots, and replication.

- **Offers equivalent protection capabilities for VMware and Hyper-V hypervisors.** You will almost surely have both of those virtualization platforms in place (either now or in the near future, and probably sooner than you think). It is imperative to seek out a protection solution that works with both VMware VADP and Microsoft VSS in such a way that virtualized application servers (e.g., Microsoft Exchange and SQL Server) get what they need to (1) be protected, and (2) be restored with some level of application awareness.

- **Protects a range of virtualized applications and IT services.** VMs are not just nameless containers filled with ones and zeroes. Application awareness is vital. Every transactional application is different, and some may require particular protection-related finesse that not all data protection solutions can provide.

³ Ibid.
• Provides great agility in rapid-restore operations. Remember, the biggest reason we make backups is so that we can restore that data if needed. How fast will the solution you are considering restore a VM? Will it allow an administrator to restore a VM from the backup storage pool without first recovering the virtual machine disks (VMDKs) or virtual hard disks (VHDs) to a production host? You need a data protection solution designed for quick recoverability.

• Enables an admin to move a VM transparently to a more suitable host, even as the VM continues to run.

Virtualization as a method and converged systems as an architecture make several aspects of IT easier. But ensuring protection recoverability isn’t always one of them, at least not without modern protection tools. Using older/legacy methods to try to protect highly virtualized environments puts an organization at risk of missing out on the optimized features offered by the latest solutions.

It also costs more, considering that legacy-product licensing models don’t line up with how modern IT infrastructures are architected. (Organizations would benefit from choosing a modern data protection solution that is licensed per the amount of data being protected, rather than the number of agents needed. After all, a single virtualized server may require a dozen or more agent licenses.)

The Highest Levels of Availability and Agility Come Through Automation

Data protection and availability are achieved through automation. In fact, the biggest challenge to any data protection strategy is, most often, the human factor—marked by a lack of execution, a lack of environment awareness, inconsistent application of policies, and the simple fact that tired, overworked people make mistakes.

Those vulnerabilities can manifest early in a data protection strategic effort as policies are applied inconsistently or without taking the needs of various business units and application owners into account. Perhaps the most dangerous aspect is that human-factor-related risks can affect a company’s ability to recover. The right people don’t know the right things to do when it matters most. Imagine:

You’re in the midst of a BC/DR crisis; your primary data center has just gone down. You’ve failed over to your secondary site, but you must restore operations by bringing your primary center back up.

Instead of making an IT administrator run around searching for the procedure binder containing instructions specifying which servers should be lit up first, automation serves as an “easy button.”

The orchestration and previously defined scripting just “kick in” as the right infrastructure components come up in the correct order, and you’re back online again.

In the second scenario, the human effort centers on defining the strategy, scripting the strategy, and then making the decision to execute on it. Humans still lead the charge, but they are not solely responsible for every detail that could be missed. When it comes to the human factor in BC/DR activities, the last human step should be to “push the easy button.” That step always should be done by a human being, someone capable of making an intelligent, informed decision that the primary environment actually is down.

Just as converged systems are unblocked in performance due to their integration and are accelerated by comprehensive management tools, data recovery is unblocked through automation and orchestration, which are based on predefined strategies by the business stakeholders.
Storage-enabled Protection Capabilities Make a Real Difference

Backups, snapshotting, replication, and application-awareness functionality are software-centric concepts. But storage hardware-enabled data protection capabilities can be equally important. They add agility that can contribute significantly to a data protection strategy. Unfortunately, storage-centric protection mechanisms are often perceived of as unattainable in piecemeal virtualization platforms. That’s yet another reason that converged systems offer a superior infrastructure for virtualization of both servers and storage.

With storage virtualization, data is no longer bound to the original “container” it was written to—similar to the way compute-centric virtualization allows VMs to be portable. When an underlying storage container is portable, additional levels of availability, agility, and protection all become possible.

The Hitachi Approach to Protecting a Modern Infrastructure

When it comes to storage technology innovation to support IT strategy enablement, a name that frequently comes to mind is Hitachi Data Systems. The HDS storage portfolio has long boasted a robust, diverse set of data protection capabilities. This portfolio, as it relates to the protection of highly virtualized environments, includes the vendor’s own hardware capabilities plus a combination of software-based solutions.

The Next Generation in Enterprise Storage—Hitachi Virtual Storage Platform G1000 (VSP G1000)

Perhaps most interesting of all are the data availability and agility features in the Hitachi Data Systems storage portfolio. With Hitachi VSP G1000, storage virtualization is providing more agility and availability for the storage that underpins enterprise environments.

In compute virtualization, physical metal is abstracted into VMs, which then become portable and easy to move to a different location, restart somewhere else, replicate as needed, etc. What VMware and Hyper-V do for servers, Hitachi VSP G1000 does for storage (see Figure 4).

As Figure 4 shows, commonalities are evident in how HDS’ storage virtualization (right) looks compared with VMware or Hyper-V-style server virtualization (left). In server virtualization, instead of having a server and its applications living only in a physical box, a VM is portable, agile, and can be protected as a whole. Storage virtualization is similar. Instead of data residing in a physical box and living only in that box, the storage container is encapsulated. The data repository becomes portable, agile, and like VMs, it can be protected “as a whole.”

Figure 4. How Compute Virtualization Compares with Hitachi Global Storage Virtualization

Virtual machines changed the way we view data centers.
CouldHitachi Virtual Storage machines do the same for data protection?

Source: HDS, 2014.
In this case, the storage container—a LUN or array—is virtualized. Virtual storage becomes a portable, easily replicated container bringing lots of agility benefits, including on-the-fly restart. This robust storage virtualization supplements the traditional backup, snapshot, and replication tools available to anyone, providing extra—and otherwise unattainable—agility.

**HDS Unified Compute Platform Director for Agility and Availability**

Hitachi Unified Compute Platform is an enterprise-class converged solution leveraging Hitachi mission-critical compute and storage components, preconfigured for easy ordering and rapid deployment.

HDS customers can leverage the newest release of the Hitachi Unified Compute Platform Director management tool to remove complexity as they attempt to orchestrate VMware hosts and VMware Site Recovery Manager (SRM) in combination with HDS replication. Unified Compute Platform Director automates the setup process, provides easy configuration for recovery, and offers unified monitoring of the now-resilient infrastructure, as Figure 5 shows.

![Figure 5. HDS Unified Computing Platform Architecture for Disaster Recovery Scenarios](image)

As stated earlier, removing the human factor anywhere that it would be less than effective is a key contributor in a data protection success story. Hitachi has an answer, and it comes in the form of Hitachi Unified Compute Platform solutions and Unified Compute Platform Director software. They provide visibility, orchestration power, and a way to leverage an underlying VMware infrastructure so that the human factor in the equation can do what it does best—planning and overseeing the protection strategy—and infrastructure can do what it does best—namely, executing on that strategy.

**Partnership with CommVault for Backups and Snapshot Management**

For years, HDS has chosen to partner with CommVault in offering HDS customers CommVault Simpana data protection software tools—referred to by HDS as the Hitachi Data Protection Suite, powered by CommVault—as a supplement to the Hitachi storage portfolio’s native protection capabilities. HDS has now announced that it will be offering CommVault’s **IntelliSnap** management toolkit outside of the Simpana full-featured product to manage HDS snapshots. It is one of the first times CommVault has enabled any storage vendor to utilize its IntelliSnap framework as a standalone management tool. This is an effort by HDS to help its customers who want to leverage HDS’s native snapshotting capability—essentially, the CommVault management UI offers these users a new alternative for managing the HDS snaps.

In August 2013, ESG Lab published the results of its assessment of Hitachi Unified Compute Platform. Testers focused on validating the platform’s ease of management, the speed and reliability of its operations, and the granular efficiency of the Hitachi converged infrastructure cell architecture. Read a summary here.
Partnership with Symantec for NetBackup

In addition, HDS recently announced a partnership with Symantec to offer NetBackup to its largest enterprise customers and those familiar/comfortable with the Symantec platform. Although Symantec competitor CommVault has been and continues to be a solid OEM partner with Hitachi, this step from HDS regarding reselling “native” Symantec NetBackup gives HDS customers even more choice in large-scale traditional backup mechanisms. And as part of the NetBackup solution, Symantec V-Ray technology delivers visibility for true granular file recovery, deduplication, backup load balancing, and automated protection for any new VMs that come online.

Whatever software the customer prefers—from CommVault or Symantec—HDS can provide that software and the expertise to maximize the software features in combination with HDS storage solutions. These are products from two longstanding innovation leaders in enterprise backup now available in the HDS ecosystem (each with notable capabilities in protecting diverse physical- and virtual-infrastructure IT environments). And adding snapshot integration offers even better recovery options to environments of all sizes.

The Bigger Truth

The modern software-defined data center, particularly as it is often implemented in private cloud environments, provides levels of agility and availability that were unachievable even a few years ago. By leveraging modern building blocks such as a converged infrastructure and compute and storage platforms that take a broader enterprise-enabling view, large IT organizations can accomplish even greater things.

That said, ensuring protection and availability within a highly agile, abstracted infrastructure can be a challenge requiring a different approach to data protection. Simply running backup software on VMs is, to put it plainly, not enough. Instead, private cloud infrastructure deserves an orchestrated combination of backup, plus snapshots, plus replication—including a hybrid use of best-of-breed software solutions and embedded storage intelligence to really achieve those goals.

As important as combining software- and storage-centric protection and recovery capabilities are, it is even more important that the evolution of one’s production infrastructure (presumably and eventually into a converged architecture) and one’s protection mechanisms should be developed in parallel. A modern infrastructure that is not adequately protected, or worse is hindered, by a legacy backup solution can undo many of the benefits that the virtualization journey brings.

With all of that in mind, it really should come as no surprise to see an enterprise IT leader such as Hitachi Data Systems bringing storage virtualization and converged infrastructure solutions to market. HDS is offering its customers a choice when it comes to protection software, while providing comprehensive protection that utilizes backup, plus snapshots, plus replication, within a single modernization strategy.