Enterprise Strategy Group | Getting to the bigger truth.™

White Paper

Four Data Protection Imperatives for 2015 and HP's Answer to Each

By Jason Buffington, Senior Analyst

March 2015

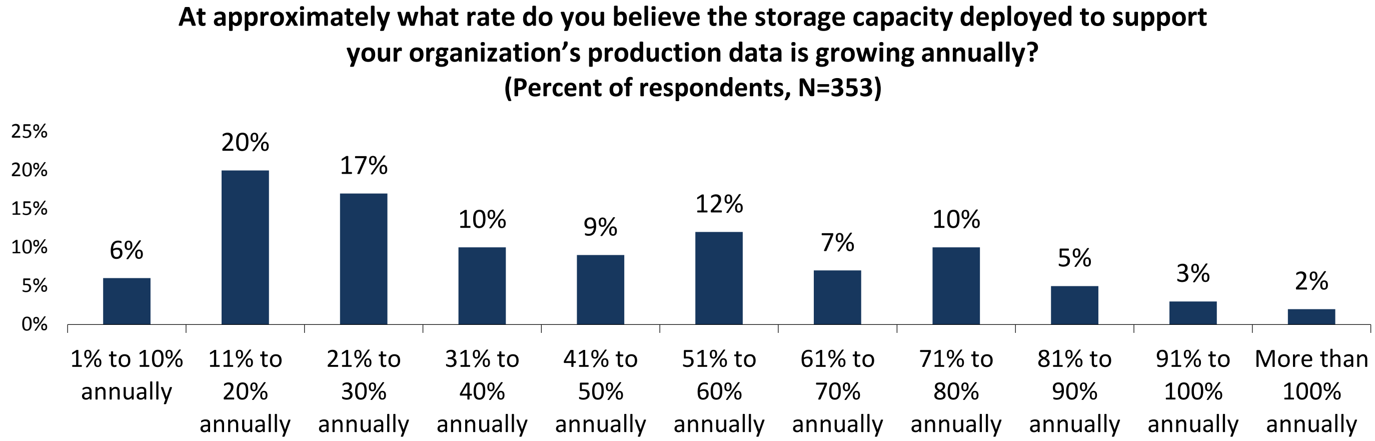
This ESG White Paper was commissioned by Hewlett-Packard and is distributed under license from ESG.

Contents

Introduction

We have never been more dependent on our data—and our data continues to grow at rates that are far exceeding our IT budgets. According to recent ESG research, primary (production) storage is growing at approximately 38% year-over-year (see Figure 1).[1] In the same report, ESG found that secondary/tertiary storage is growing at nearly the same rate (42%). Unfortunately, IT budgets are not increasing at nearly the same pace.

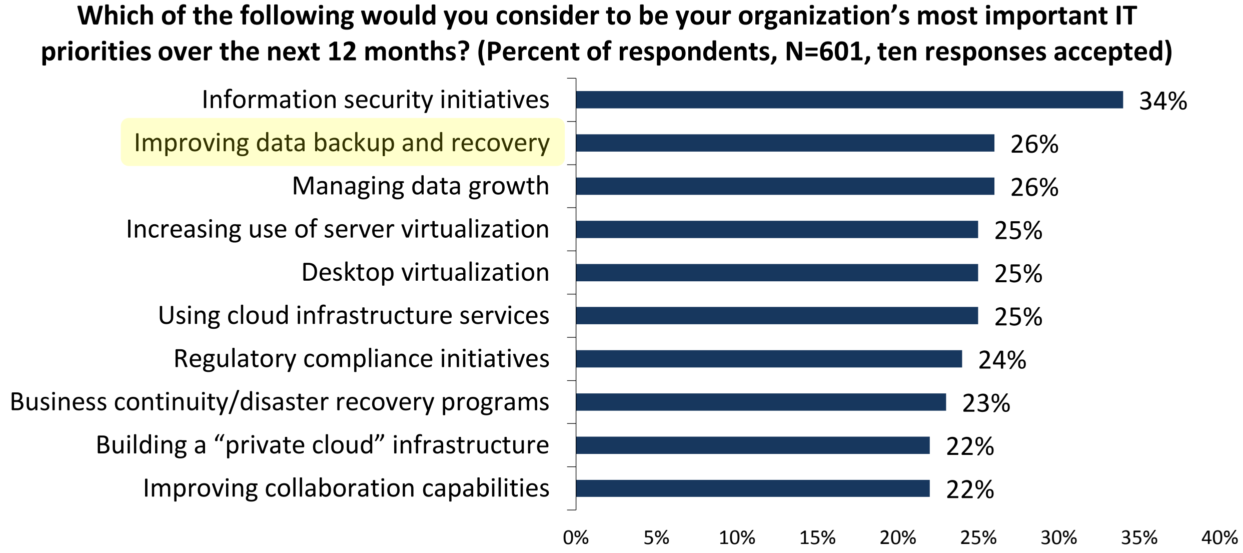
Figure 1. Annual Growth Rate of Storage Capacity for Production Data



Source: Enterprise Strategy Group, 2015.

The reality is that IT organizations cannot continue to do what they have always done, if for no reason other than storage growth. But storage growth isn't the only aspect of IT that is driving the need to do "better," not just "more." According to ESG's *2015 IT Spending Intentions Survey*, improving data backup is of especially high priority, second only to information security (see Figure 2).[2] In fact, data backup has been a top-three most-cited priority in IT spending among both enterprises and midmarket organizations for the past five years in a row.

Figure 2. Top Ten Most Important IT Priorities for 2015



Source: Enterprise Strategy Group, 2015.

[1] Source: ESG Research Report, *Backup and Archiving Convergence Trends*, April 2014.
[2] Source: ESG Research Report, *2015 IT Spending Intentions Survey*, February 2015.

With the top two IT trends being the *security* of data and the *preservation/recoverability* of data, it is apparent that IT cannot focus on simply delivering service as usual. It must instead deliver "better" IT that accommodates increased storage growth, increased server virtualization, distribution of data in remote offices, etc., while ensuring access of the data to the organization at ever-heightening standards of service.

Four Improvements that IT Must Make in 2015

Considering the need to do *better* (data protection) for *more* (workloads and data) with *less* (budget), here are four key areas that IT can and should investigate and implement in 2015:

- Reduce the burdens of backup (time and dollars).
- Improve how remote office data is protected.
- Integrate data protection with core workloads.
- Remember to never disrupt in the name of progress.

Reduce the Burdens of Backup (Time and Dollars)

Although being able to restore the organization's data is absolutely essential, the actual processes and infrastructure necessary for backup do not have to be cumbersome. Often, the cumbersomeness of data protection is a result of antiquated methods or solution components. By investing in more modern data protection mechanisms, organizations of all sizes can save significant amounts of time and money—particularly by:

- **Reducing the backup window through acceleration and optimization.** Legacy approaches to backup are often invasive to the production servers by sending far too much data, but modern agent technologies can intelligently discern at a sub-file or object level what has changed—and only send those granular changes. In doing so, production server performance is far less hindered, and the backup window shrinks dramatically. That lets the production servers do more of what they were meant to (serve users), and it results in better ROI for the IT infrastructure as a whole due to having to move less data.
- **Reducing the data protection infrastructure through deduplication.** As perhaps the single most effective investment to mitigate growing storage costs and data protection sprawl, investing in deduplication for protection storage will reap huge dividends. As discussed earlier, environments without modern deduplication are seeing their secondary/tertiary storage grow at nearly the same rate as primary storage (around 40% YoY), and while some of that growth is logical for storing the new data, far too much of the protection storage is consumed by protecting stagnant primary data. Deploying modern deduplication not only saves capital expenditures (CapEx) in secondary storage and the related operating expense (OpEx) of managing that ballooning storage, but also can further reduce the backup window (thus saving additional time and I/O impact) if the backup software agents can intelligently discern what is already protected in the duplicated storage and discard redundant data.

Improve How Remote Office Data Is Protected

Historically, remote offices were underprotected, often due to complexity or cost. Due to minimal WAN connections, it was typically not viable to pull remote office or branch office (ROBO) data from the remote sites back to the corporate data center's primary backup solution. Consequently, either the same backup solution was deployed in some offices, adding complexity of distributed IT management, or a mediocre backup/copy utility was used in the ROBOs to ensure at least minimal protection. And although many ROBOs still don't have the bandwidth that they'd like to have for near-LAN performance to their corporate data centers, other technologies have reduced the need to do so.

Deduplication can make ROBO centralized backups a reality. The same intelligent deduplication technology that ensures minimal production server performance issues (and faster backup windows) during backups can also be used on ROBO servers wanting to be backed up to their corporate data center. This actually can be achieved in two ways:

- **Backup agents with intelligent deduplication** (often referred to as source-side or client-side deduplication) will first discern what data has changed within the production data sets and then further identify what unique data blocks are not yet part of the deduplicated storage pool. Only those unique and granular changes are sent, thereby making centralized backups of ROBOs actually possible.
- **Deduplicated storage solutions can also be deployed in the ROBOs**, particularly larger regional offices, as either physical or virtual appliances. These solutions enable a complete data protection and recovery experience within the ROBO, and then replicate their granular and unique (deduplicated) blocks from the small ROBO deduplication storage appliance to the data center's centralized deduplication storage appliance. As a best practice, having the same deduplication storage technology throughout distributed enterprises can be particularly beneficial if the data can be deduplicated within each ROBO and remain deduplicated throughout its replication and preservation in consolidated storage.

By utilizing the same modern data protection mechanisms at ROBOs that are in the data center, IT can save money in multiple ways. Specifically, it can:

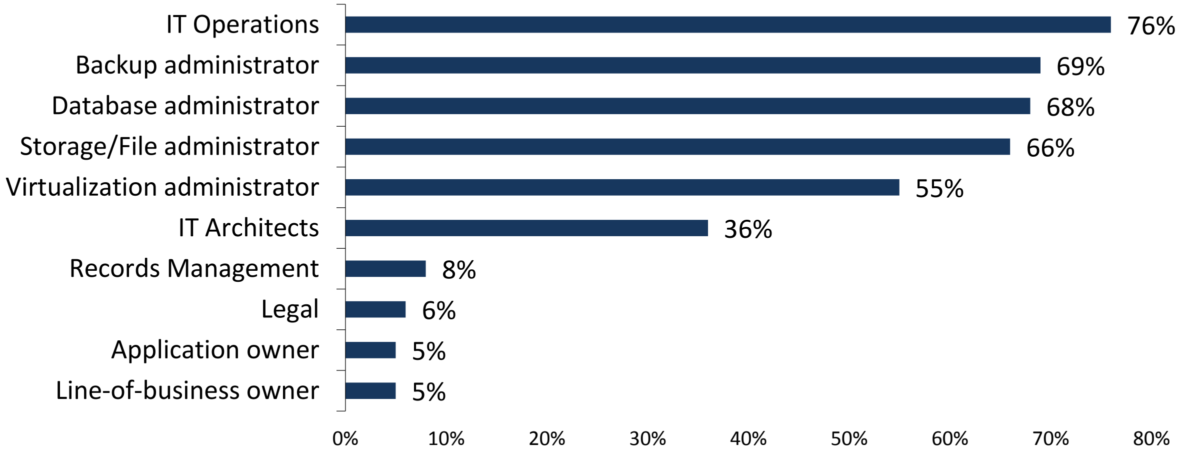
- Reduce WAN costs by sending only granular and unique deduplicated data.
- Reduce management costs by being able to manage all solutions within a single framework.
- Reduce storage costs by first deduplicating within the branch, and then sending to a globally deduplicated tertiary repository within the data center.
- Reduce CapEx costs by optionally implementing virtualized deduplication appliances within the same hypervisor framework that the virtualized production servers operate within.

Integrate Data Protection with Core Workloads

No one knows the needs of the application and its end-users, as well as the protection/recovery requirements of the application's data, better than the workload administrators (e.g., the database administrators, storage administrators, and vAdmins). In a recent ESG survey of the IT roles involved in data protection strategy and operation, several non-"backup administrator" roles were found to be extremely active in the daily operation of backups and in the selection of the tools themselves (see Figure 3).[3]

Figure 3. Organizational Roles Involved in Any Aspect of Data Protection Processes

Which organizational roles (specific groups or unique individuals) are involved with any aspect of your organization's various data protection processes and operations on a day-to-day basis? (Percent of respondents, N=272, multiple responses accepted)



Source: Enterprise Strategy Group, 2015.

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

By involving the workload owners in the data protection process, the recoverability of the data will be heightened, and the costs associated with protecting those workloads will almost surely go down.

By partnering with workload owners, "rogue backups" will diminish. Many of today's workload owners are already doing some form of copy/replication/backup unbeknownst to the central IT operations team or the backup administrators:

- DBAs are doing database dumps, log shipping, or other replication tasks, perhaps to expensive primary storage (because it's the only storage that they have access to).
- vAdmins are doing VM exports, replicas, or hypervisor-based snapshots.
- File/storage admins are habitually invoking any number of secondary copy mechanisms or retaining snapshots for inordinate amounts of time, also consuming expensive primary storage.

These "rogue" examples are in addition to the backups that IT is doing, and it is all due to a historic disconnect between the workload owners and the data protection specialists. By partnering, the workload owners will not be consuming expensive capacity, and the overall redundant footprint across the organization will diminish—thereby saving CapEx and the OpEx of no longer duplicating the effort of protection.

Utilize data protection hardware that enables workload protection. While rogue-IT methods of protection often consume expensive primary storage, secondary storage that is designed for data protection (such as a modern deduplication storage solution) often provides native access from those applications directly to the deduplicated storage. It ensures that not only are the applications being protected to reliable and durable storage, but also the data is stored in its optimum (deduplicated) state within the central repository that the rest of the organization's data resides in.

Utilize data protection software that ensures workload-protection via monitoring. While traditional backup software may be best for IT operations or backup specialists, those tools are not typically embraced by application owners such as DBAs or vAdmins. Instead, by using data protection tools that can be integrated into those workload frameworks (e.g., Oracle RMAN, a vCenter plug-in, or a SysCtr management pack), workload administrators can manage their backups and restores in a UI that they are familiar with, while the data protection specialist can have monitoring or auditing insights to ensure that the organization's data is protected, even if those individuals are no longer the ones enacting the backups or restores.

Remember Never to Disrupt in the Name of Progress

Several ways exist to lessen the impact on IT that legacy backup approaches often incur. Executing on these suggestions might help to reduce operational expenses:

Remember that host-based backups impact servers less than agent-based backups do. There are only a few business justifications for using legacy agent-based methods to protect the inside of virtual machines (VMs) the same way that older physical servers were protected. Agent-based approaches use a significantly higher amount of CPU, memory, and I/O during backups than a host-based method requires. For most organizations, the modern host-based methods that protect VMs from the outside (while still enabling granular restores from within VM data sets) are more than adequate. This does not mean that installing an agent or other "assistance" module within the VM to quiesce a database or catalog data isn't useful, but those agents should be for data protection management, not the actual transmission of data. Reducing the need for agents per server also reduces change-control efforts and upgrade-related labor costs as new agents or patches are released.

Look for data protection mechanisms that do not incur downtime. Whether for upgrading firmware within storage appliances, adding incremental storage capacity, or completing other common maintenance tasks, it is unacceptable to incur a lack of protection due to maintenance to the data protection toolset. Instead, look for solutions that can be updated or expanded without shutting down data protection tasks or hindering the continued protection and assured recovery of the production environment.

One Company to Consider: Hewlett-Packard

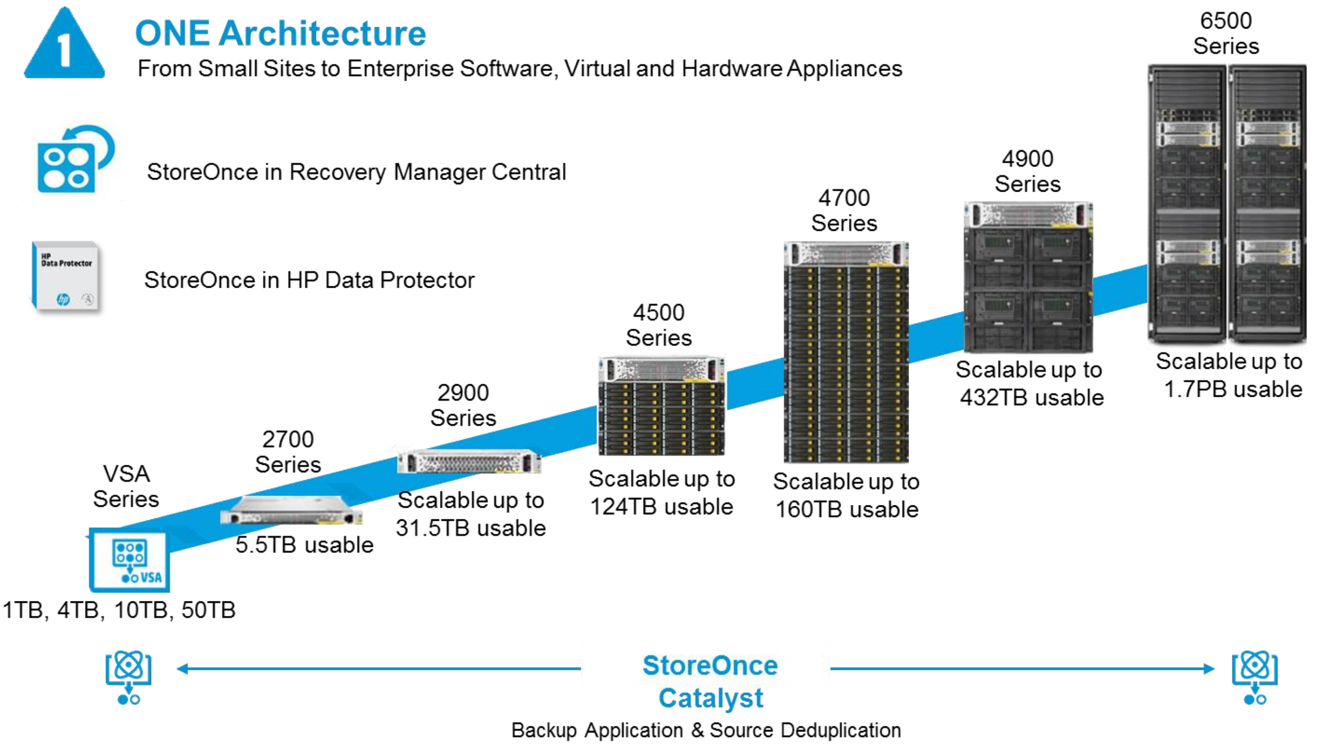
HP offers a comprehensive portfolio of offerings in data protection, including:

- HP servers—standalone, rack-mount, and blade-server configurations.
- Primary storage products such as HP 3PAR StoreServ tier-1 storage platform, the HP Modular Smart Array (MSA), the HP StoreVirtual Storage (scale-out storage for virtual environments).
- HP Data Protector backup software and HP Consolidated Archive software.
- HP StoreOnce disk backup and deduplication storage.
- HP StoreOnce Recovery Manager Central.
- StoreOnce VSA.
- HP StoreEver tape drives and libraries.
- HP Helion Cloud services.

In the context of deduplication, three key offerings are important to understand:

One, **HP StoreOnce—**deduplication storage solutions, including virtual storage appliances (VSA), small appliances for midsize organizations and regional offices, and truly enterprise-class protection storage platforms of up to 1.7PB of usable capacity in HP's 6500 solution (see Figure 4). One noteworthy mention is the single architecture across the StoreOnce family, including the virtual appliance and the deduplication logic within the HP Data Protector software, such that deduplication data moves between devices in its most optimized state.

Figure 4. The HP StoreOnce Product Line


1
ONE Architecture
From Small Sites to Enterprise Software, Virtual and Hardware Appliances
StoreOnce in Recovery Manager Central
HP Data Protector
StoreOnce in HP Data Protector
VSA Series
VSA
1TB, 4TB, 10TB, 50TB
2700 Series
5.5TB usable
2900 Series
Scalable up to 31.5TB usable
4500 Series
Scalable up to 124TB usable
4700 Series
Scalable up to 160TB usable
4900 Series
Scalable up to 432TB usable
6500 Series
Scalable up to 1.7PB usable
StoreOnce Catalyst
Backup Application & Source Deduplication

Source: HP, 2015.

Two, **HP StoreOnce Catalyst—**the underlying deduplication technology within and across the StoreOnce appliances, as well as backup software such as HP Data Protector, and third-party backup software leveraging the Catalyst APIs to better integrate with HP StoreOnce appliances.

Three, **HP StoreOnce Recovery Manager Central (RMC)—**Convergence between primary storage and backup with the ability to back up and restore snapshots stored on a 3PAR StoreServ array extremely quickly. With those offerings in mind, one can briefly revisit ESG's four prescriptive mandates for better protection in 2015.

How HP StoreOnce Reduces the Burdens of Backup

A key to reducing the burdens of backup is optimizing how changed data is gathered, transmitted, stored, and replicated via optimization and deduplication, as stated earlier. And that is the whole point of HP Catalyst technology. Catalyst-enabled backup technology discerns either at the production server (via agent) or within the backup server what data has changed (or not), as well as which data fragments are unique versus what is already stored within the deduplicated storage pool. By combining those capabilities with HP StoreOnce's assertions of being among the fastest deduplication appliances for both ingest/backup and restore available in the market, organizations can gain several business benefits:

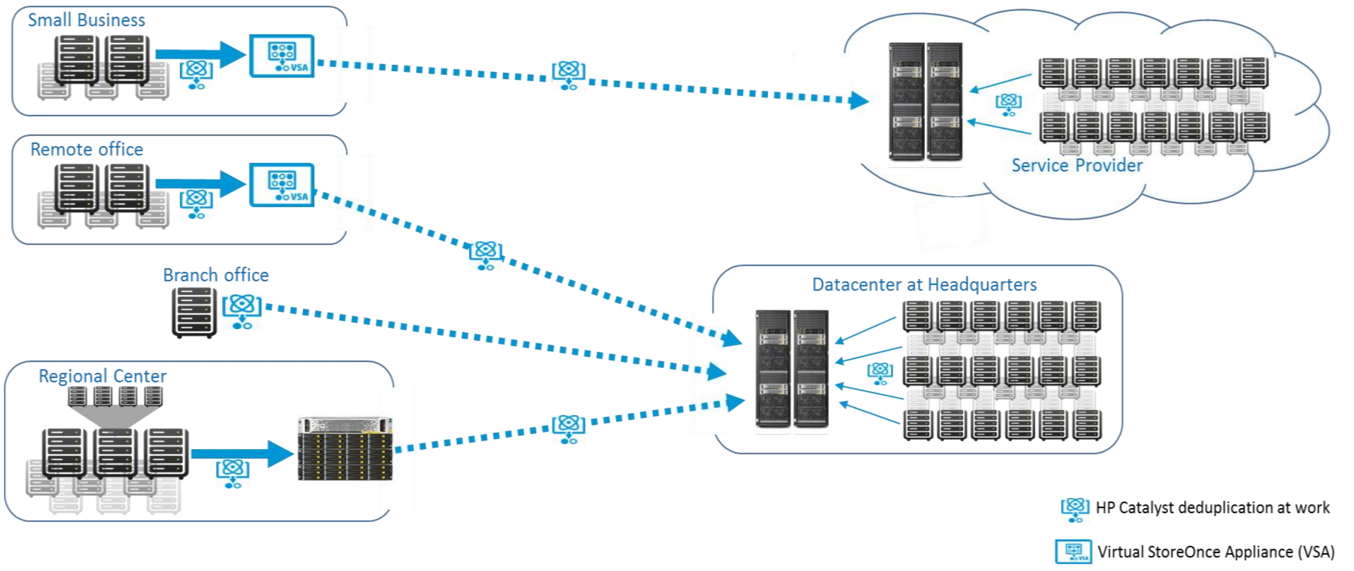
- **The backup window shrinks**, resulting in less impact to production resources and more assurance that the data is protected more frequently.
- **Secondary storage requirements decrease** due to deduplication technology, which reduces the operational costs associated with managing secondary storage.

Beyond traditional backup, HP is further optimizing its data protection capabilities through StoreOnce RMC, which copies snapshots and snapshot differentials directly from primary (production) storage to HP StoreOnce (protection) storage. Once within the storage solution, further deduplication and synthetic full backups are generated; all of which is done without a traditional backup agent/job and with improved SLAs and reduced backup windows being the result.

How HP StoreOnce Improves ROBO Protection

The importance of protecting all of an organization's data, regardless of where it resides, cannot be overstated. HP Catalyst and StoreOnce have several permutations that enable centralized backup of remote offices, including Catalyst-accelerated backups for smaller offices, virtual StoreOnce appliances that replicate for midsize offices, and right-sized physical appliances for larger remote offices (see Figure 5).

Figure 5. HP StoreOnce Scenarios for Remote Offices


Small Business
VSA
Remote office
VSA
Branch office
Regional Center
Service Provider
Datacenter at Headquarters
HP Catalyst deduplication at work
Virtual StoreOnce Appliance (VSA)

Source: Enterprise Strategy Group, 2015.

In all cases, the data is first deduplicated within the remote office, and then stored locally in an on-premises StoreOnce repository, including either a physical appliance or a virtual StoreOnce VSA. Then the unique data is transmitted from the ROBO to the data center, while remaining in its optimized state, which offers business benefits:

- **Network costs remain flat while the organization gains centralized backup** due to the consistent use of Catalyst deduplication across the enterprise.
- **Operational costs of remotely managing separate backup solutions are greatly reduced** by using the same data protection mechanisms (software and hardware) across the enterprise.

How HP StoreOnce Integrates with Core Workloads

Recently published ESG research shows a significant influence on data protection strategy, as well as ongoing operations, by the IT professionals who know the data and workloads best—the workload owners.[4] With this in mind, it behooves IT organizations to evolve the role of the data protection specialist from being the operator of backup functions to being the enabler of data protection. HP Catalyst, StoreOnce Backup, and StoreOnce RMC provide that enablement by being able to integrate with the application platforms themselves, including Catalyst plug-in technology, to allow DBAs and vAdmins to protect their own data directly to HP StoreOnce deduplicated repositories. By combining that enablement and distribution of tasks with the oversight capabilities provided to the data protection administrator, organizations can be assured that data protection tasks are ongoing even as the daily operators change:

- **Application owners should consume less-expensive (lower CapEx) storage for their own rogue backups** and instead leverage enterprise-consolidated deduplicated storage.
- **Data protection specialists can focus on data protection architecture and oversight** instead of daily backup configurations and monitoring, resulting in reduced management costs.

How HP StoreOnce Ensures Fewer Disruptions

Production should never be hindered by protection. Beyond the reduced impact on production servers, hosts, storage, and networks due to optimized data transmission from Catalyst-accelerated backups, the HP StoreOnce infrastructure plays its own part to ensure mitigated infrastructure impact: It allows for upgrades, StoreOnce storage-node failures, and, with Federated Catalyst, even multi-generational usage (e.g., with HP StoreOnce B6200 and 6500) in parallel without stopping production data from being protected. That provides operational benefits:

- **Production servers and networks are less affected** due to having to send less data via HP Catalyst agent technology, resulting in less CapEx investments to offset backup requirements.
- **Upgrades to HP StoreOnce enterprise platforms can be done without impacting protection** of production resources, including upgrading firmware or even expanding B-series node couplets for scale-out and scale-up of the data center's deduplication solution.

[4] ibid.

The Bigger Truth

An important reality for IT in 2015 is that although data protection continues to be a high priority due to security concerns and the evolving landscape of production data to be protected, all of those data protection improvements must be done with a budget that isn't even capable of maintaining the status quo due to storage growth. IT must get smarter in terms of how it provides data protection, with at least four imperatives:

- Reduce the burdens of backup (time and dollars).
- Improve how remote office data is protected.
- Integrate data protection with core workloads.
- Remember never to disrupt in the name of progress.

Each of these imperatives should reflect a near-immediate gain in protection assurance and recovery agility while providing business benefits in reduction of CapEx and OpEx across the IT infrastructure. One company that has all of the pieces to enact these imperatives is HP.

HP Document No. 4AA5-7436ENW

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