White Paper

EMC Best Practices in Data Protection Monitoring

By Jason Buffington, Senior Analyst

February 2013

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

Contents

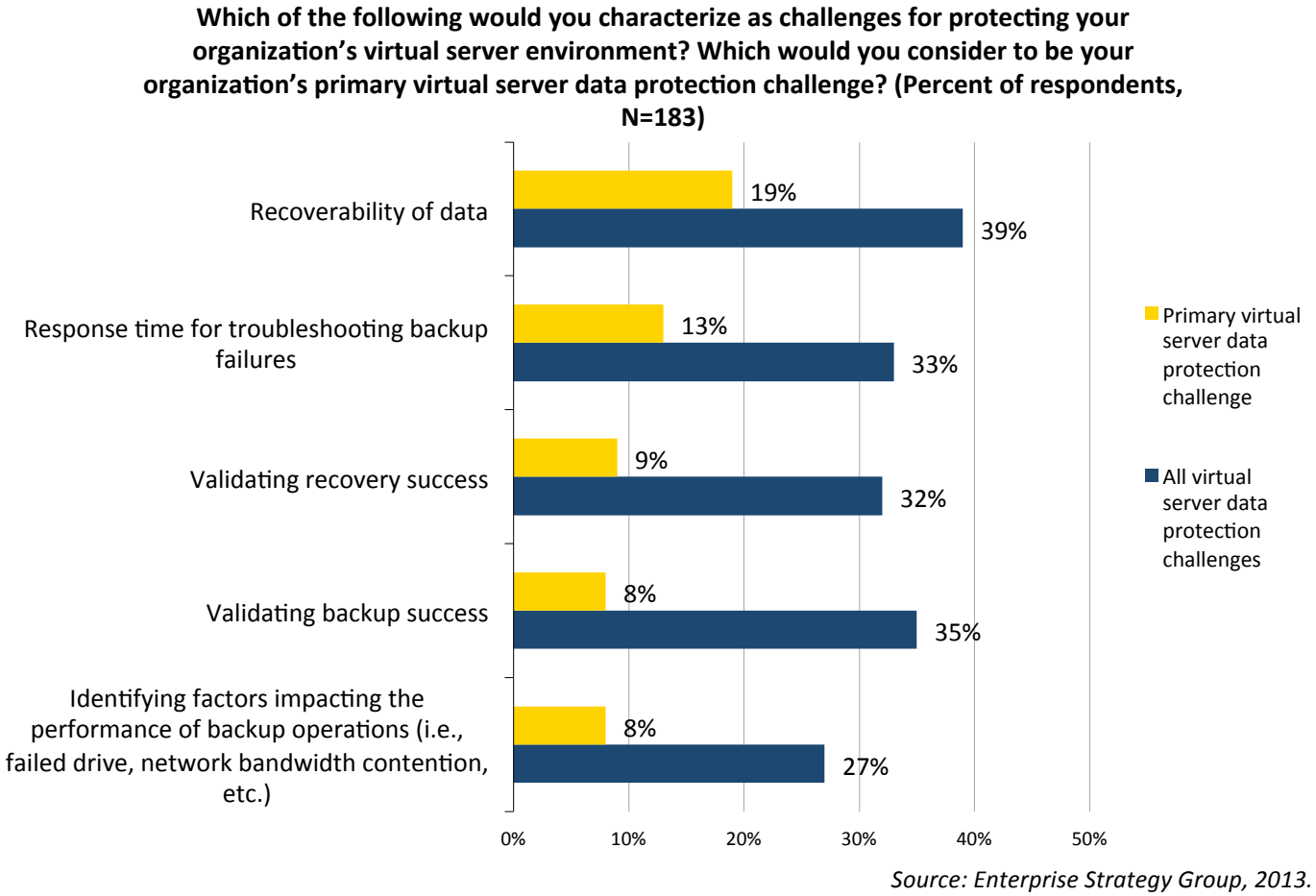
Introduction

If you're reading this white paper, then quite possibly you're in the business of providing data protection, i.e., backup, replication, deduplication, and possibly IT-as-a-Service (ITaaS). You may be a member of an IT department within a large enterprise providing backups to constituent business units. You may be a systems integrator providing Backup-as-a-Service (BaaS)—where your clients own their own backup infrastructure, but you manage it for them. Or you may work for a cloud-based service provider delivering BaaS or Disaster Recovery-as-a-Service (DRaaS).

Regardless of whether you consider your domain to be a private cloud, a hybrid cloud, or a public cloud, you are providing data protection as a service. It should not come as a surprise that, according to ESG's *2013 IT Spending Intentions Survey*, improving data backup and recovery was the number-two most important IT priority selected by respondents (27%), just two percentage points behind the top priority, information security initiatives.[1]

As server virtualization increases (and arguably becomes commoditized), minor nuisances and issues related to using legacy protection methods to protect a handful of virtual machines (VMs) grow into mission-impacting challenges when applied across an enterprise-wide virtualization infrastructure (see Figure 1).[2]

Figure 1. Top Five Challenges in Protecting a Virtual Server Environment



Source: Enterprise Strategy Group, 2013.

[1] Source: ESG Research Report, *2013 IT Spending Intentions Survey*, January 2013.
[2] Source: ESG Research Survey, *Virtual Server Data Protection*, September 2011.

It Is All About Visibility

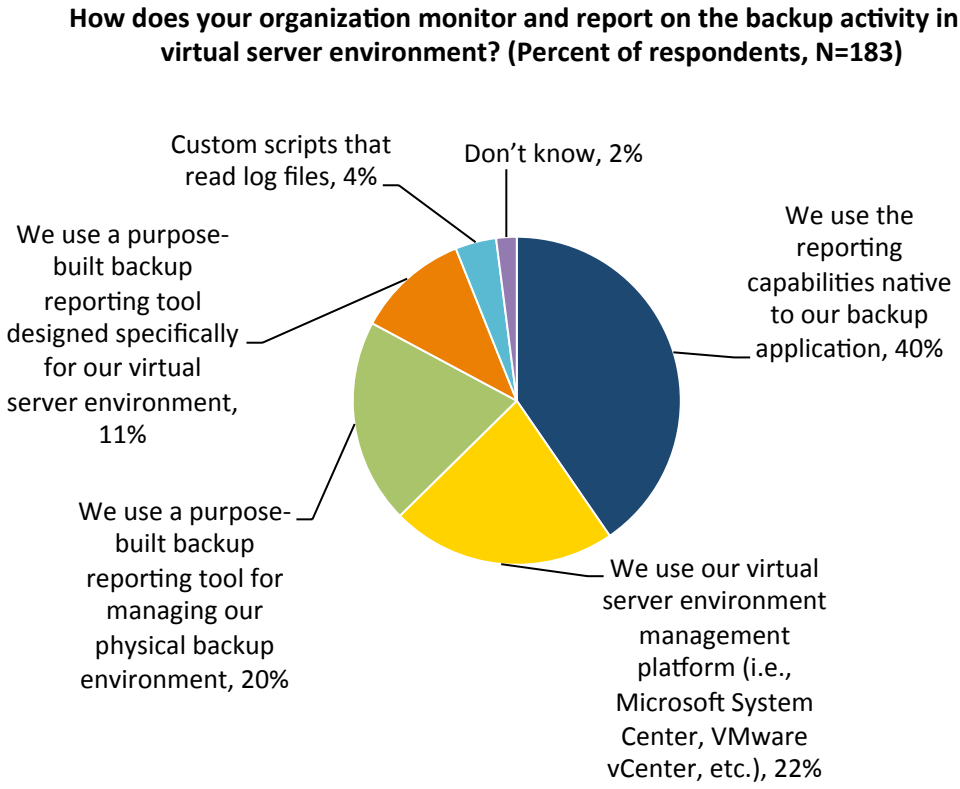
Aside from the top challenge of recoverability, the next most common challenges all have to do with visibility:

- Do you know whether your backups were successful?
- If your backups were unsuccessful, how effective are you in troubleshooting the errors or identifying the bottlenecks?
- Even if you can you secure your SLAs, do you still need to be more productive/proactive in resolving issues quickly?
- Assuming that your backups are good, are you able to validate and ensure recoverability?

Almost every challenge listed in Figure 1 equates back to one of those questions. In a modern virtualized environment of any size, visibility is often the greatest data protection-related challenge. Arguably, without reliable reporting and insight into your data protection infrastructure, you don't have a plan, just a hope.

Several approaches exist to help you gain that necessary insight into your data protection status (see Figure 2).[3]

Figure 2. How Organizations Monitor and Report Backup Activity in a Virtual Server Environment



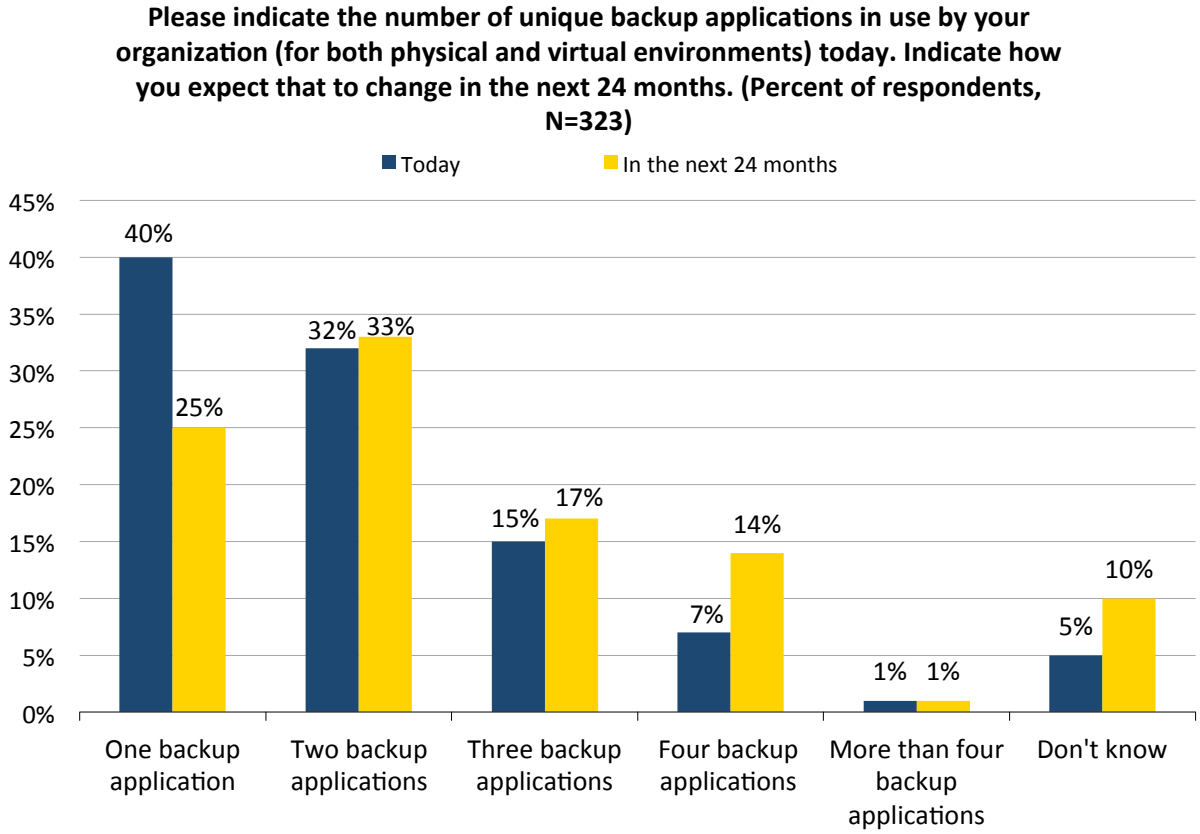
Source: Enterprise Strategy Group, 2013.

[3] Source: ESG Research Survey, *Virtual Server Data Protection*, September 2011.

The two most logical approaches for gaining the visibility come from the backup application itself (40%), and from the server virtualization management platform (22%).[4]

Of course, leveraging the server hypervisor's management platform works only as well as the backup application(s) integrated with the hypervisor(s). And in those parenthetical plural "(s)"s lies the problem: Today's modern infrastructure seldom has a single backup application, nor a single server hypervisor. In fact, a modern protection infrastructure is likely to span the physical world, the virtual world, hosts, network switches, disk and tape storage, and multiple protection technologies including backup managers and various replication solutions. Determining faults, detecting trends, and securing end-to-end visibility are significant challenges.

Figure 3. Number of Unique Backup Applications Used



Source: Enterprise Strategy Group, 2013.

As Figure 3 shows, only 40% of respondents' environments use a single backup application today—with only 25% planning on using a single backup application within 24 months. Said another way, most organizations are not solidifying on fewer backup applications; they are diversifying their strategy to reduce costs or improve recoverability and agility through a broader data protection toolset. And the larger the organization is, the more likely it is that they have multiple data protection tools and server hypervisors, or even more server platforms in general.

In ESG's *Trends in Data Protection Modernization* research report, respondents discussing both physical and virtual backups reported that only 85% of backups were completing successfully, and only 80% of recoveries were completing within pre-established RPO/RTO service level agreements.[5] The challenges listed in Figure 1 apply to physical environments as well as virtual environments, and as Figure 3 shows, those challenges are exacerbated as the toolsets and hypervisors diverge.

[4] Source: ESG Research Report, *Trends in Data Protection Modernization*, August 2012.
[5] Ibid.

What Should a Backup or DR Administrator Do?

The only presumptions for the remainder of this paper are:

1. You are in the business of offering data protection as a service, either to internal stakeholders and business units or to subscribing clients. As such, you are subject to service level agreements (SLAs), perhaps written and well documented, or perhaps simply understood or presumed.
2. Your world is large-scale (again, either internal or external clients), which very likely includes a high utilization in virtualized infrastructure, multiple data protection tools, and perhaps multiple vendors' server hypervisors, physical servers, and storage solutions.

If those generalities do not match your situation, it may be because you are working in a small environment with a single hypervisor, running on a single storage and server platform architecture, being protected by a single data protection utility.

But for the rest of us, as a general rule, manual aggregation of data is not the answer—*because the goal isn't just reporting; it is also the management of data protection.*

A good data protection management technology (that is, a "data protection manager" or DPM) has a few key characteristics, including:

- Interoperable tools with a common SLA taxonomy.
- A capability to discover dynamic workloads, including protection automation, and particularly, virtualized components.
- Flexibility in monitoring, reporting, and auditing.

Interoperability

As stated earlier, ESG's latest data protection modernization research indicates that 60% of respondents use more than one data protection tool today, and 75% plan on using more than one moving forward. As much as success rates for backups and recoveries are still surprisingly lower than one might expect, the primary driver for using multiple technologies is cost.

Certainly, many IT organizations are looking to augment traditional backup and recovery tools with hardware-based snapshots or replication technologies, and they are looking to add cloud-based repositories to complement their on-premises solution(s). Overall, ESG agrees with the logic of those approaches. However, multiple tools should not necessitate multiple interfaces for ongoing maintenance and management because that approach can introduce the human factor that inevitably will "break" a data protection strategy.

Unified View of Status

If the only goal of a unified view was to get a single report on the "protection readiness" of a known list of production resources, then numerous scripts, methods, or tools could aggregate the equivalent of "green checkmark vs. red x" for those resources. But your goals should be broader. Health is not a binary status condition. Health is qualifiable and quantifiable, relative and subjective. As such, a unified view in a DPM—a view across backup and replication solutions—must be able to inform its users at a high level about overall status (green/red), and about trends of success, frequency of protection, effectiveness of data flow/movement, backup capacity availability, etc. Being able to successfully monitor trends in resource utilization (ideally forecasting when a "cliff fall" will happen) allows proactive rather than reactive management of assets.

Common Taxonomy

What is the difference between a *job*, a *policy*, a *task*, or a *process*? What is different among a *source*, a *client*, a *workload*, a *node,* and a *server*? If you only used one data protection tool, you'd only care about the terms and activities that your tool used. But for the rest of us, it is daunting to correlate the events each backup tool invokes

to the actual production resources each solution is responsible for protecting. Add to that complexity several health statistics tied to frequency and speed, which also vary between days versus minutes and MB per second versus GB per hour (respectively), and health monitoring aggregation becomes exponentially more complex.

A DPM should not only glean health and statistical information from a variety of data protection tools, but also facilitate a common taxonomy so that you can apply a single mindset and assessment of your entire infrastructure, regardless of which tool is protecting which parts.

Discovery of Dynamic Workloads

Murphy's Law dictates that if something can go wrong, it will. In the world of data protection, one of the most common calamities is a user asking IT to restore data from a server that IT wasn't aware of (and therefore was not backing up). Today's business units (internally) and subscribers (externally) are very agile in IT provisioning. In a virtual world, we are no longer held back by long hardware purchasing cycles, custom operating systems, application rollouts, and other matters.

It is frankly almost "too easy" for people to create new machines that might start out as test or development resources, then quickly find their way into production use. This situation is especially true in highly virtualized environments, and it is even truer in environments that have adopted the self-service provisioning capabilities of private cloud architectures. This activity causes peaks and troughs of resource consumption across dynamically present VMs, which makes tracking utilization problematic.

If your data protection strategy still requires business units, clients, and other stakeholders to manually inform you (IT) of their newly provisioned servers (so that you can back them up), then you are doomed. Instead, a contemporary DPM solution must use multiple means of discovering new computing assets that may require data protection. These methods can be extended from the discoverability tools in the underlying data protection tools, or even be discovered by the DPM and then be automated for protection via one of those data protection tools.

A DPM's discovery strategy solves two problems:

- It ensures that every production server is protected without having to be manually notified to do so (because they would otherwise be missed, resulting in some servers never being backed up).
- It ensures that every data protection tool doesn't "automatically protect" every server discovered individually by those various tools. If a new server is protected by multiple technologies, it can cause space consumption and resource constraints, or even jeopardize that server's ability to recover successfully.

Virtualization Awareness and Agility

The scenarios and challenges described above are even more prevalent in virtualized environments. Virtualized servers are not only easier to provision, but also easier to move. As such, it is extremely important to have a holistic data protection strategy that recognizes the difference between a VM that was created (and should be protected) versus a VM that was moved to a new host (requiring its protection policies and previous data to follow it).

This setup requires a high degree of engineering and interoperability between the server hypervisors, the private cloud management platforms, the various data protection tools, and the DPM. Collectively, the result should be that new virtualized servers are protected, and that the protection follows them as they potentially move between storage pools, hosts, or virtualization clusters. In addition, because virtualized servers are still servers, the DPM should also ensure that the proper policies are coordinated between the virtualization protection tool(s) that see each VM as a workload or data object and the physical protection tool(s) that see new servers being brought onto the network.

Private Cloud SLAs

One area of anticipated evolution in virtualization protection in 2013 and beyond will be SLA awareness within private clouds. Within most private cloud architectures, machines and resources are not presumed to be equal. Instead, most provisioning systems have various SLA concepts in place, perhaps including "gold, silver, bronze"; "high impact, medium impact, low impact"; or simply "critical, important, normal" for provisioned resources and services. These private cloud SLAs often result in different tiers of storage performance, server resource consumption, etc.

The challenge is that although the private cloud may enjoy prioritized views of its resources, many generic data protection tools have one of two extremes for their protection strategies: "all or nothing."[6]

- Some auto-discovery methods simply discover every new VM regardless of private cloud SLA tiering, then apply the same generic protection policy to them.
- Other backup tools have no awareness of virtualization discovery, and rely on manual configuration per VM or service for data protection.

Both extremes will result in gaps in your data protection strategy. ESG expects that as private cloud management tiers and their underlying hypervisors mature, data protection innovators will apply different data protection policies (for frequency, RPO/RTO, retention, distance, availability) to those provisioned virtual servers and services in accordance with their private cloud SLAs.

Flexibility in Monitoring, Reporting, and Auditing

For the purpose of this discussion, monitoring, reporting, and auditing (MRA) describe three kinds of (mistakenly perceived as interchangeable) views of your data protection status:

- **Monitoring**—related to offering ongoing operational information for IT administrators responsible for backups to help the admins ensure that those backups (and requested recoveries) are completing successfully, and/or providing insight into failed events for troubleshooting. Protection monitoring has to become more real-time to support SLA windows. A scheduled/timeboxed notification won't suffice: You have already failed your SLA by the time the report is in your inbox.
- **Reporting**—recurring output of data protection status to interested stakeholders such as business units or subscription clients whose data is being protected.
- **Auditing**—ad hoc reporting of overall data protection status, recovery capabilities, and last known protection and recovery events in an effort to ensure compliance with corporate, government, or industry mandates related to recoverability or retention.

This description of these varied aspects of MRA is intended to help emphasize that a script or other aggregation of point-in-time status information across backup tools is not enough. Different audiences exist for the different aspects of MRA, including people responsible for backups and recoveries, people whose data is being protected, and people who inspect recovery capabilities, then assure others on the organization's ability to recover.

Each audience requires different levels of information about and insight into the overall data protection infrastructure, and each audience has different actions to take, based on what they see within their view. This is an area that makes DPM solutions invaluable within large enterprises and cloud environments: the ability to provide tactical aggregation of health for ongoing operations, and to provide different lenses that cater to the needs of interested second- or third-party teams who don't want the technical minutia but need actionable status information.

[6] Source: Jason Buffington, SearchDataBackup.techtarget.com, *Private cloud backup needs to get better*, December 2012.

EMC Data Protection Advisor v6

EMC is one company that understands enabling IT-as-a-Service and the complexities of managing enterprise-wide (and varied) data protection technologies. In large enterprises that rely on VMware-powered virtualization infrastructures, EMC is especially well suited to help with its combination of Avamar, NetWorker, and Data Domain. But many environments have more than one hypervisor, more than just EMC storage, and more data protection tools than just those from EMC. Understanding this fact, EMC also offers EMC Data Protection Advisor (DPA) as a manager for data protection tools across large enterprises and subscribed clients.

Interoperability

As discussed earlier, a key aspect of a manager for data protection is interoperability with a range of data protection tools. Table 1 shows the list of backup products, tape library products, and replication technologies that are manageable by DPA v6.[7]

Table 1. Summary of the Backup and Replication Tools EMC Supports with DPA

Backup Applications	Tape Libraries	Replication Technologies	B2D Platforms
Avamar	Sun	Symmetrix DMX, VMAX, VMAXe:	EMC Data Domain
NetWorker	ADIC	• RecoverPoint	EMC Disk Library
NetBackup	IBM	• SRDF	Fujitsu CentricStor
Backup Exec	HP	• TimeFinder	HP Virtual Library Systems
PureDisk	Fujitsu	• VPLEX	NetApp NearStore
TSM		CLARiiON/VNX:	
Data Protector		• RecoverPoint	
CommVault		• MirrorView	
ArcServ		• SnapView	
Oracle RMAN		• SAN Copy	

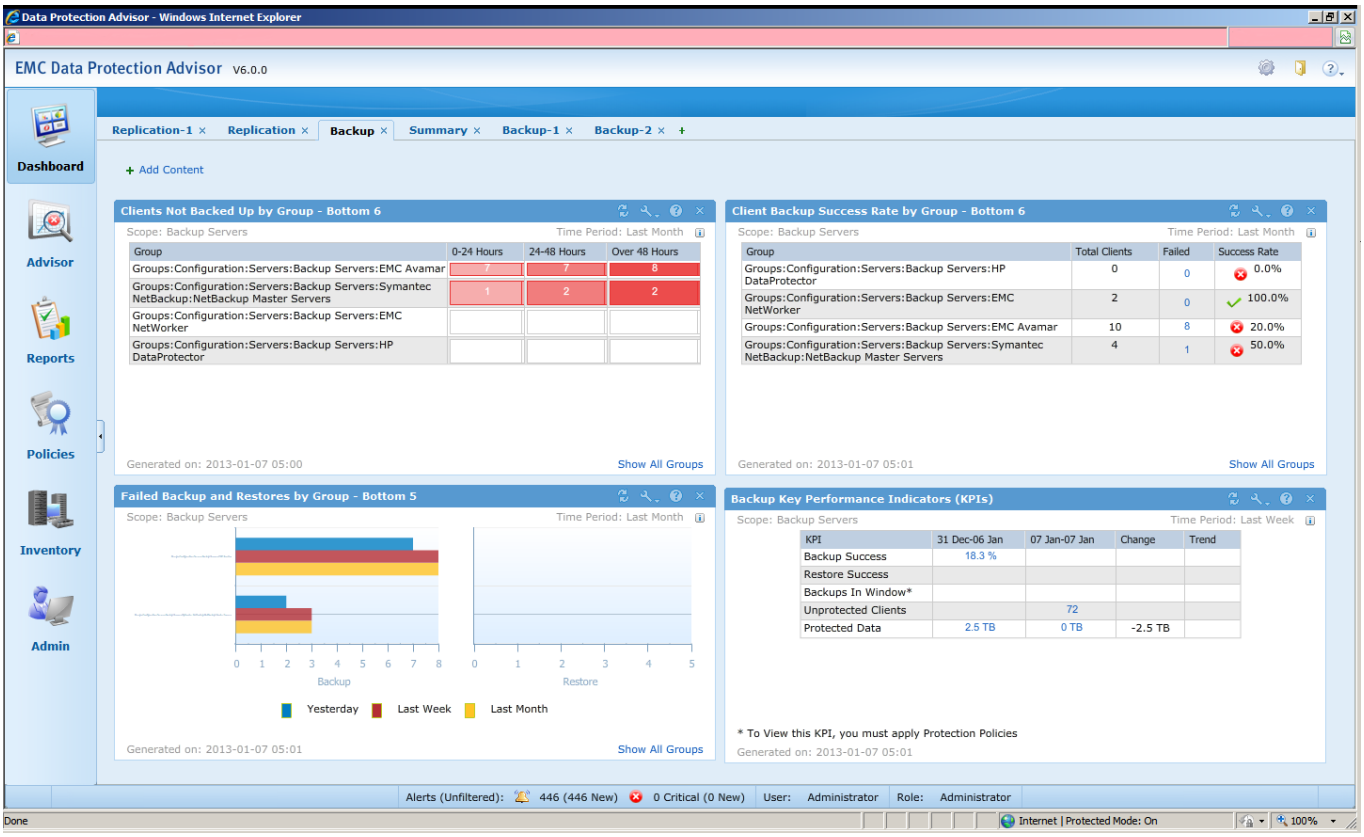
Source: EMC, 2013.

[7] A complete list of supported replication software, backup applications, file server platforms, backup-to-disk platforms, tape libraries, database platforms, and EMC Data Protection Advisor server platforms is available at: http://www.emc.com/collateral/software/data-sheet/h1767-data-protection-advisor-ds.pdf.

Visibility

As stated earlier, most of the top reported challenges in data protection (particularly in virtualized environments) relate to visibility: verifying successes, identifying bottlenecks, and troubleshooting errors. This is a primary area of focus for DPA v6, with the goal being to clarify how well your data protection infrastructure is working (see Figure 4).

Figure 4. Screenshot of EMC DPA v6 Dashboard



Source: EMC, 2013.

Notice that the dashboard does not reference a specific data protection tool, nor does it use terms/naming conventions that are obviously specific to one tool or another. (It should be noted, though, that the dashboard does let administrators dig deeper into various areas to obtain additional details and access different or customized views.)

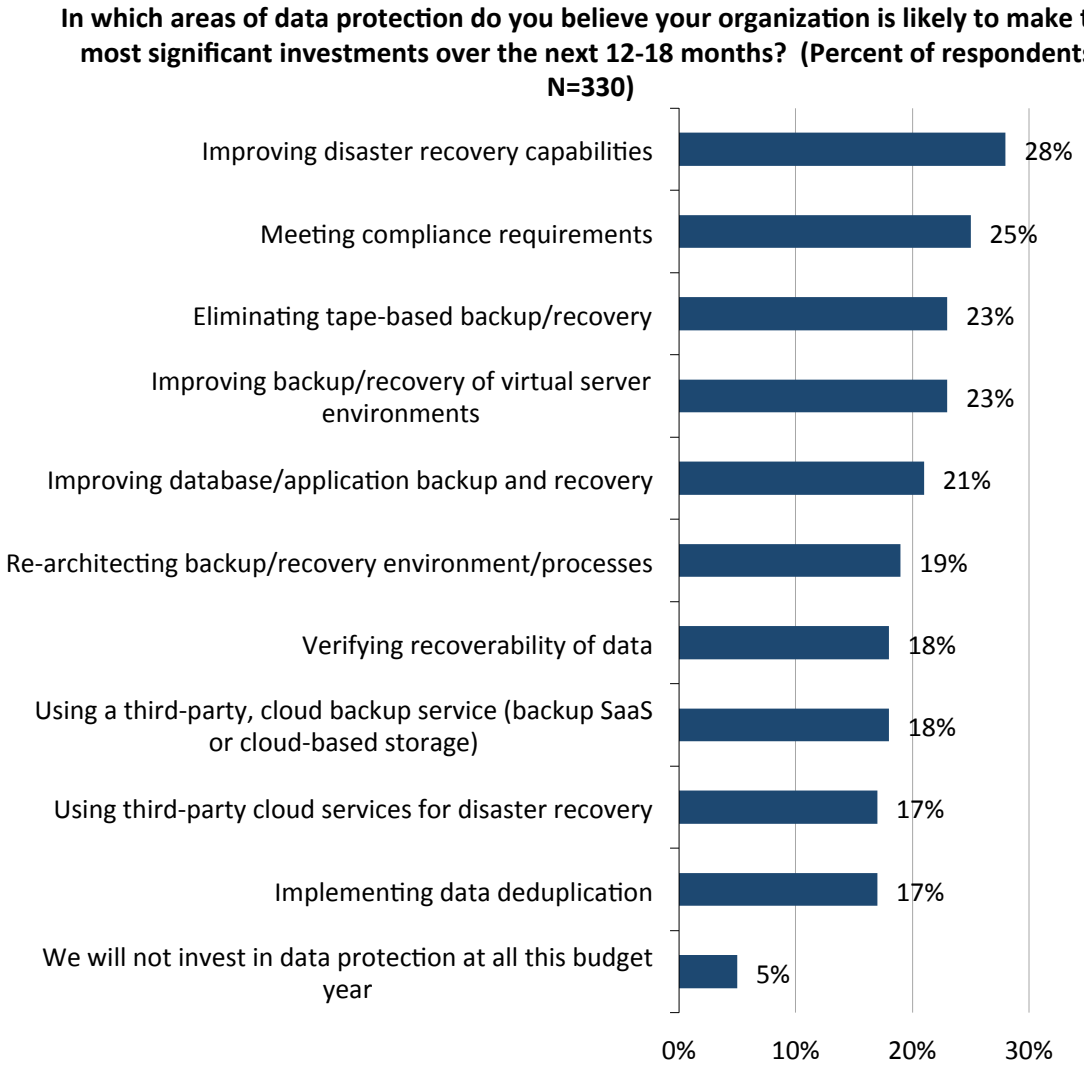
The top-level view (in this case) raises awareness of departments that are experiencing alerts, key performance indicators for both backups and replication, and overall health indicators.

Auditability

Similar to what we see from an aggregation viewpoint in the screenshot in Figure 4, it is also notable that the information displayed was based on one user's needs and permissions. Other users, such as stakeholders whose data is being protected or auditors who need to understand capabilities versus implementation details, will see what they need to—based on their needs, job roles, and thereby the DPA permissions.

When ESG asked IT professionals what data protection areas were tied to planned investments in 2012, the top two answers related to disaster recovery and regulatory compliance (see Figure 5).[8] This fact emphasizes how your overall data protection strategy needs to include not only a unified view of traditional backup solutions, but also a view of replication technologies, and it must include support for auditors to provide oversight.

Figure 5. Top Areas of Investment in Data Protection in 2012



Source: Enterprise Strategy Group, 2013.

Specifically, DPA correlates otherwise siloed data protection tools' status information, the recoverability conditions, and the various workload information from sources such as server hypervisors, Exchange e-mail servers, and databases such as Microsoft SQL or Oracle—to provide a single view of assurance related to protection and recoverability.

[8] Source: ESG Research Report, *Trends in Data Protection Modernization*, August 2012.

The Bigger Truth

Whether you are backing up a private cloud across your enterprise or several smaller environments on behalf of a service provider, one of the most significant keys to success for enabling modern data protection is *visibility*.

Without visibility that crosses server/storage boundaries, hypervisors and data protection silos, and vendor product lines, your data protection infrastructure is guaranteed to miss backing up some servers, while potentially duplicating the protection of others. The clarity gained by a unified view is necessary for proactively reporting protection to stakeholders and compliance to auditors. It is also required to understand overall data protection (backup and replication) health, performance, and reliability.

It is often said that "backup is hard, and recovery even harder." So why would you do it with a blindfold on? To take the blindfold off and move from data protection hopes to data protection strategies, you need a manager for your data protection tools—one that understands the dynamic nature of today's workloads, while adapting the presentation of its aggregated insight to different audiences who are all interested in ensuring that the data is recoverable.

To address all of those needs, EMC has released Data Protection Advisor v6 as a means to enable service-based delivery of backup and replication. With support for the major backup software and tape technologies, along with EMC's own storage and replication offerings, DPA correlates what might otherwise be managed by different folks with varying skills. The result is a comprehensive and heterogeneous data protection management system that supports not only large enterprises, but also multi-tenant service-provider models.

In a perfect, no-calorie, never rainy-day world, everyone would use a single, all-powerful, all-capable data protection tool that protected everything. But for the rest of us, the norm is multiple backup tools protecting varied workloads across multiple hypervisors that are supported across servers and storage from a variety of vendors. And although all of those pieces may be necessary for now, it doesn't mean that you need numerous lenses to understand how they are being protected. DPA v6 intends to be that single pane of glass for making the complicated world of data protection less so.

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