

## I D C   W O R K B O O K

---

# Protecting Data in the Context of Enterprise Infrastructure Consolidation and Data Growth: The Rise of New Technologies

August 2008

Adapted from *Storage Virtualization Comes of Age, Following Server Virtualization* by Laura DuBois and Richard L. Villars, IDC #204175

Sponsored by Data Domain

---

*Today, enterprises are caught between the continued growth in the amount of data they create, store, and depend on and the need to consolidate IT resources to reduce costs and create a more "green" IT environment. But while organizations have worked to reduce IT footprint, they must balance the need to improve resource utilization with more efficient protection and recovery capabilities, all in the face of increasing capacity pressures. This IDC Workbook discusses consolidation at the datacenter, server, and storage system levels, including the use of virtualization. This paper also discusses how deduplication can help improve consolidation of resources and reduce storage footprint, as well as improve data protection and disaster recovery under heavy workload conditions. A worksheet to determine organizational readiness for deduplication is included.*

## Data Consolidation and Disaster Recovery: The Growing Need

According to IDC research, the amount of global digital data created and stored on a worldwide basis has increased over 3,000% in the past three years. But while 70% of the data is generated by individuals (consumers and corporate workers), in the near future, enterprises will have the responsibility of protecting and managing 85% of the data generated. Paradigm shifts in the way that content is created, managed, protected, and made available — leveraging the Internet and Web 2.0 — are driving new approaches to storage and data protection. Globalization, Web access, new techniques for data gathering and analysis, digital communications, conversion from paper to digital processes, and increased regulatory and legal requirements all have added to this tsunami of information. While some of this data comes from traditional IT applications — business processes, customer information, and so forth — much of it comes from nontraditional sources, such as VoIP packets, surveillance videos, and real-time sensor information.

The challenge is that CIOs and IT staff not only must find a way to collect and manage this digital data but also must effectively store and protect this valuable resource. IT organizations are expected to ensure that data is recovered quickly after a data-loss event. The business environment is requiring ever-shorter downtimes, and the pressure is tremendous to recover information and make it available again as soon as possible. Increased customer and partner demands, globalization, regulatory compliance, and simply heavier reliance on data to improve business all add up to a data storage and recovery imperative.

As the value of information assets has grown, so too have IT operational costs. In particular, IDC research shows that over the next five years, the expense to power and cool the worldwide installed base of servers is projected to grow four times compared with the growth rate for new server spending. With the cost of energy rising, companies will be hard-pressed to provide not only the necessary space for information creation and protection but also the power and cooling resources.

Costs are only part of the challenge, however. Enterprises also are paying more attention to the environmental aspects of the increasing demand for storage capacity and its power demands. Because of growing concerns on the part of the public and shareholders, more companies are looking for ways to minimize the environmental impact of their datacenter operations. This has created the "green computing" movement in corporate IT. As organizations build new datacenters, they are looking for ways to be more environmentally friendly, including meeting the Leadership in Energy and Environmental Design (LEED) Green Building Rating System set forth by the U.S. Green Building Council.

This means that companies are facing the daunting task of dealing with two opposing forces — the need to increase capacity to keep up with data growth while also shrinking the footprint.

Meanwhile, as companies try to minimize the number of datacenters either as a result of rising real estate costs (or reducing available real estate) or the reduced availability of power (in the case of many major metropolitan cities), many firms are experiencing an increase in the number of remote office and branch locations. With expansion at the edge and consolidation at the core, firms must take into account available datacenter space, adequate infrastructure, and available power and cooling.

To shrink the IT footprint, two technologies have garnered a lot of attention — virtualization and data deduplication. While server and storage virtualization can help address initial or immediate capital constraints, they don't address the continuous need to maximize utilization of resources. Data deduplication technologies can augment and improve resource utilization during and beyond consolidation efforts, as well as yield energy conservation objectives.

Virtualization and deduplication, while at different stages of adoption, go hand in hand. Deduplication can drive further consolidation and footprint reduction within virtualized environments. Protection of virtual machines can mean the backup of redundant data within the images themselves. Using deduplication, companies can optimize storage requirements for backup of virtual machine images to reduce the capacity required for their protection. For firms still working on consolidation, data deduplication is a technology with immediate benefits and can be deployed and redeployed to meet changing datacenter needs.

## Minimizing Resource Use Through Consolidation

Centralization is a resurgent theme for datacenters. Organizations are closing smaller sites and moving the assets to an existing or a new enterprise-class datacenter. The scope of this consolidation appears to be accelerating. While some customers have indicated plans to consolidate 20 sites to 6 or 7 central sites, IDC is starting to see even more aggressive closures, from 60–70 sites to 4 or fewer sites. Standard systems consolidation is happening simultaneously, helping to lower the previous double-digit growth rates of servers to single digits.

Organizations are focusing their consolidation efforts on the following two areas:

- **Minimizing the datacenter footprint.** Better managing the datacenter is the most visible approach to consolidation. As companies look to cut costs and migrate toward green IT, they are consolidating facilities at a rapid pace.
- **Reducing the number of servers.** Typically, organizations have consolidated servers in the following two ways:
  - Using blade servers to increase workload per surface area
  - Migrating from many direct-attached storage devices to using fewer, more efficient storage area network (SAN) devices

Both approaches reduce physical space and power and cooling–related challenges and costs. Blade servers ease management of multiple workloads and are often associated directly with consolidation projects. SANs have made it possible to share storage resources across a network to utilize resources regardless of their location.

Blade servers and SANs used in consolidation projects are typically associated with server virtualization, enabling the use of a single resource for multiple applications. Server virtualization permits moving software from system to system without concern for differences in configuration because the virtual machine (VM) environment insulates the application software from the server's actual physical details. IT staff can adjust the level of deployment of virtual machines on each server for optimum resource utilization. This reduces unused capacity and ongoing operational costs because management of the virtual environment is more easily automated and because server recovery procedures are simplified. Importantly, data can be distributed and stored at multiple locations, making it less likely that information will be completely lost during an outage.

In addition to consolidation, these approaches can further help organizations become "virtual enterprises," with the flexibility to expand or contract computing and storage resources when and where needed.

## **Better Data Protection and Faster Recovery: The Role of Deduplication**

Even with consolidation through server virtualization and the use of shared SAN storage, however, enterprises still need effective data protection and rapid recovery in the case of a data-loss event. The implication of consolidation and virtualization for storage means that the role of data protection must change. Storage, backup, and recovery are strategic considerations in the design of consolidation and virtualization projects and cannot be an afterthought or a postprocess. As organizations consolidate IT, they have looked to data replication as a way to ensure data protection and recovery. Data replication software includes software designed to create image copies of volumes or files via techniques such as clones, mirrors, and snapshots.

Replication may be storage system, server, fabric, or appliance based and may occur locally or between remote sites, potentially separated by long distances. Replication software is often used in conjunction with backup software to improve data protection performance.

Unfortunately, data protection via replication is in direct conflict with the goals of consolidation, virtualization, and green IT. Historically, data replication has increased demands for storage space. It has not been uncommon for a large firm to have between five and nine copies of a piece of data. Further, with increased dependence on distributed computing, moving vast amounts of replicated data across a network for either storage or recovery takes time. Often, firms have relied on removable tape media, which has raised security or recoverability concerns. In an era of server and storage system consolidation, newer methods must be employed to reduce storage demands, speed up recovery times in a networked environment, provide for an electronic offsite copy for disaster recovery, and eliminate the physical security risks of removable media.

Data deduplication technology is increasingly a must-have technology for data management and protection, especially in the context of replication of data between datacenters or remote offices to enable rapid disaster recovery. IDC research suggests that deduplication is being evaluated and deployed with greater and greater frequency. Deduplication replaces duplicate data with references to a shared copy to save storage space — a key benefit given today's data protection and archiving demands.

Specifically, deduplication refers to any algorithm that searches for duplicate data objects and discards those duplicates. When duplicate data is detected, it is not retained; instead, a "data pointer" is modified so that the storage system references an exact copy of the data object already stored on disk.

This elimination of redundant copies of data provides the following benefits to organizations:

- More efficient use of storage capacity and, as a result, a potential reduction in the amount of storage customers need to purchase and manage
- Local recoverability of data, better management of media, and consolidation at remote sites, with an added benefit of faster data backups plus replication to core enterprise datacenter (edge-to-core consolidation)
- Ability to store more data on disk-based storage systems (versus tape or optical storage), thus enabling faster data retrieval/restore and longer retention on disk
- Efficient and reliable disaster recovery/business continuity; easier to get data offsite using the wide area network than using tapes on trucks; enables faster time to DR readiness and faster recovery in the event of a disaster
- Green IT; reduces requirements for power, cooling, and storage footprint

Currently, IT organizations can implement data deduplication for backup or archive data in two ways — an appliance-based approach or a software-based approach. The appliance-based approach offers the following benefits:

- Support for multiple workloads, not only backup data but also archive data, and ultimately primary data
- Tight integration of all components and easy installation and management
- Elimination of any impact that deduplication may have on application server and wide area network performance
- The ability to integrate with different datacenter and remote and branch locations using a single architectural approach versus reliance on different architectures for the edge and the core

The software-based approach provides flexibility to define server and storage specifications to manage performance. In addition, it offers improved backup window times and optimized LAN network traffic because deduplication can be done at the client level instead of the server level.

Both approaches can address the requirements for data protection and recovery. While still a relatively new technology, deduplication is tested and growing in acceptance for supporting consolidation, virtualization, and green IT initiatives. Organizations should select the deduplication option best suited for their IT environment based on individual needs.

## Worksheet Section

### *Is Deduplication Right for Your Consolidation Initiatives?*

As previously noted, organizations face a two-sided challenge: store and manage increasing amounts of data and simultaneously reduce system footprint — and in a way that enables enterprises to meet recovery requirements. Deduplication is a technique that can dramatically improve storage system performance, minimize capacity requirements, and speed up data recovery. However, companies should not adopt deduplication without a proper needs and usage audit. Like any enabling technology, deduplication must be applied appropriately or risk becoming just another "technology for technology's sake."

#### **How to Use This Worksheet**

The following worksheet can help organizations determine where deduplication fits within their consolidation initiatives, as well as the level of need for deduplication. Each answer choice to the questions below provides a point value. After answering each question, add up all the points to determine an overall score. Match the points to the point range to determine whether deduplication is a future consideration, a "nice-to-have" technology, a "should-have" technology, or a "must-have" technology.

These questions are purposely more general in nature, making this worksheet a flexible tool that can be used multiple times. The worksheet can be used to assess overall needs, as well as to determine need specifically at the datacenter, server, and storage system levels.

#### **Deduplication: Scoping Your Organization's Needs**

##### 1. How do you see your organization's disk storage capacity growth over the next 12 months?

Shrinking – 1 point

Limited growth (less than 10%) – 2 points

Score \_\_\_\_\_

Moderate growth (between 11% and 35%) – 3 points

Significant growth (between 36% and 50%) – 4 points

Overwhelming growth (greater than 51%) – 5 points

##### 2. What is your budget situation for the next year?

Budget growth (greater than 5%) funding approved – 1 point

Budget growth (between 3% and 4%) funding approved – 2 points

Budget growth (between 1% and 3%) funding approved – 3 points

Budget will stay the same – 4 points

Score \_\_\_\_\_

Significant capacity growth, funding to be determined – 5 points

Don't know how we will be able to fund what we need – 6 points



7. What are your consolidation plans for the next year?

No plans – 1 point

Minor consolidation – 2 points

Score \_\_\_\_\_

Pilot consolidation program(s) – 3 points

Major consolidation initiatives – 4 points

8. What are your server capacity needs for the next year?

We have excess capacity – 1 point

We have enough capacity to meet our needs – 2 points

Score \_\_\_\_\_

We will need some additional capacity – 3 points

We will need a lot of additional capacity – 4 points

9. What are your storage capacity plans for the next year?

We have excess capacity – 1 point

We have enough capacity to meet our needs – 2 points

Score \_\_\_\_\_

We will need some additional capacity – 3 points

We will need a lot of additional capacity – 4 points

10. What is your target retention period on disk before migrating backup or archive data to tape?

Only one day, immediate move to tape – 1 point

1 week – 2 points

Score \_\_\_\_\_

2 weeks – 3 points

1 month – 4 points

1 quarter or more – 5 points

11. What is your average full-backup policy?

Full backup weekly, plus daily incrementals – 2 points

Full backup weekly with no incrementals – 3 points

Score \_\_\_\_\_

Full backup every two to six days with no incrementals – 4 points

Full backup nightly with no incrementals – 5 points

12. How critical is immediate recovery of data to your organization?

No need – 1 point

Needed, but not business critical – 2 points

Score \_\_\_\_\_

Required, business critical – 3 points

Mission critical, our business depends on it – 4 points

13. How critical is having complete data redundancy?

No need – 1 point

Nice, but not business critical – 2 points

Score \_\_\_\_\_

We could survive without it – 3 points

Our business depends on it – 4 points

14. What are your wide area networking plans for the next year?

Our current network will meet our needs – 1 point

We will need minimal growth in capacity/bandwidth – 2 points

Score \_\_\_\_\_

We will need moderate growth in capacity/bandwidth – 3 points

We will need significant growth in capacity/bandwidth – 4 points

15. How many sites are you protecting in your distributed enterprise?

1-5 – 1 point

5-10 – 2 points

Score \_\_\_\_\_

10-20 – 3 points

20 or more – 4 points

If you are using tape to protect any of these sites – 5 points

16. What are your virtual computing plans for the next year?

No plans – 1 point

Pilot program – 2 points

Score \_\_\_\_\_

Slow departmental phase-in of virtual computing – 3 points

Major consolidation/expansion using virtual computing – 4 points

Total Score \_\_\_\_\_

## **Score Key**

16–24 points – **Future consideration.** While you do not need deduplication technology yet, you likely will and should look to better understand this technology.

25–39 points – **Nice to have.** While deduplication is not critical, your business could benefit from the technology. It's worth looking into.

40–54 points – **Should have.** Deduplication could provide measurable benefits to your organization, particularly as you grow and/or undertake consolidation projects. Perhaps a pilot project is in order?

55–70 points – **Must have.** Deduplication is a critical tool to increase the success of your consolidation and growth plans. You will see immediate results by adopting deduplication technology as part of your consolidation initiatives.

## **Conclusion and Essential Guidance**

Deduplication is a rapidly emerging technology that has the potential to help organizations improve enterprise storage and recovery capacity and performance. While deduplication could provide benefits to essentially any organization, some are more ready than others for this technology. By completing the worksheet, organizations can determine whether they should immediately invest in deduplication or take a wait-and-see approach.

It should be noted, however, that deduplication is a complementary technology to an organization's consolidation strategies, namely virtualization and improving traditional storage, backup, and recovery solutions. Deduplication can make these areas more efficient and cost-effective by reducing redundant data. As such, deduplication should be an essential component of an organization's consolidation, storage, and recovery initiatives.

Because IDC believes that deduplication will become an essential IT tool in the future, organizations should complete the worksheet periodically, perhaps as often as every quarter, to help assess needs as the nature of business, and associated data storage and recovery requirements, changes.

---

### A B O U T   T H I S   P U B L I C A T I O N

This publication was produced by IDC Go-to-Market Services. The opinion, analysis, and research results presented here-in are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDG Go-to-Market Services makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

### C O P Y R I G H T   A N D   R E S T R I C T I O N S

Any IDC information or reference to IDC that is to be used in advertising, press releases, or promotional materials requires prior written approval from IDC. For permission requests contact the GMS information line at 508-988-7610 or [gms@idc.com](mailto:gms@idc.com). Translation and/or localization of this document requires an additional license from IDC.

For more information on IDC, visit [www.idc.com](http://www.idc.com). For more information on IDC GMS, visit [www.idc.com/gms](http://www.idc.com/gms).

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 [www.idc.com](http://www.idc.com)