



Richard Villars

Vice President, Storage Systems & Executive Strategies

The Information Explosion

February 2011

The creation, organization, and distribution of data stored in files are driving much of the growth in organizations' overall storage capacity today. As data and information continue to grow, organizations are responding by making corresponding investments in their storage environments. According to IDC, annual spending on file-based storage solutions increased from \$12.1 billion to \$15.6 billion between 2007 and 2010. For organizations in industries such as biosciences, healthcare, entertainment, design, and business services, the information explosion is particularly pressing. Consequently, the efficient storage, organization, long-term retention, and timely retrieval of files are at the core of these businesses and are driving demand for file-based storage solutions as well as the need for better information management strategies.

The following questions were posed by Hitachi Data Systems to Richard Villars, vice president of IDC's Storage Systems & Executive Strategies, on behalf of Hitachi Data Systems' customers.

Q. What is driving the explosion of information?

A. There are really three different trends driving the growth of information. The broadest trend is related to regulatory and ediscovery issues. Organizations need to keep information longer. Whether the information is in the form of PowerPoint presentations, images, videos, or product design specs, organizations recognize that they have to archive the information. Such data may be needed for legal purposes later on, or organizations may want to use it to mine for additional data in the future. Essentially, information volumes are growing because organizations recognize that information is valuable at a later date if they can find it and access it.

Another important, though more targeted trend is what IDC refers to as the digitization of many industries. For example, the media and entertainment industries are digitizing as they convert to high definition, support distribution to mobile devices, and move to digital recording. The healthcare industry is transitioning to electronic x-rays and electronic medical records, and life sciences is making huge investments in genomic sequencing. Less obvious examples include legal services where there is the growing importance of ediscovery within the overall legal process and the need to vault, archive, and review the information that is associated with discovery.

The third trend is that resolutions of data keep on changing in response to technical advances. For example, just think of a simple consumer case of everyone moving from a 2 megapixel camera to a 10 megapixel camera. The picture is the same, but the photo file is actually four to five times larger. As organizations move to ever more sophisticated and rich content, whether images, PowerPoint slides, or design documents, the files themselves are becoming significantly bigger.

The combination of these trends is forcing datacenters to respond by installing much more storage capacity to manage unstructured data rather than mission-critical transaction-oriented data or the backups associated with mission-critical data. The storage device may be cheaper, but power consumption, datacenter floor space, and the cost of managing and organizing all of this file-based data are growing concerns.

Q. How is the information explosion affecting the business at organizations?

- A. Clearly the information explosion is driving an often surprising growth in storage capacity outside the traditional IT organization. In hospitals, for example, electronic records or medical x-ray systems are consuming four to five times as much capacity as the back-office transaction system. In this situation, the backup of back-office transaction information is not nearly as big a problem as backing up and restoring all the files.

At many organizations, however, unstructured data is still seen as a secondary asset and has yet to be properly addressed in the disaster recovery plan. Even without the legal mandates, such data is absolutely critical; if something were to happen to that data, that would affect whether the business can survive.

Finally, even if an organization has few problems saving all the data and buying all the storage capacity it needs, the issue of finding the data later remains a real challenge. The fear is that the data can be saved, but then it is eventually lost when the organization needs it. It's still somewhere on the system, but it just can't be located. This is actually the worst thing that can happen because it means that an organization is wasting its storage capacity.

Q. What are companies doing about it?

- A. Companies need to do three things in response to the information explosion and the associated issues such data growth can cause. First of all, a lot of the unstructured data is very widely dispersed in the organization — it's in file servers and in individual offices, departments, and so on. This dispersal makes the management problem related to the efficient utilization of storage capacity much worse. The first step an organization can take is to centralize the management of information. If that is not an appropriate strategy, organizations may want to consider establishing one data pool. At a minimum, information needs to be more rationally structured so that there are more shared resources and common access points.

Second, organizations should evaluate technologies such as scale-out NAS appliances that accommodate storage, content, and information growth without the constant need of having to throw out old boxes and replace them with new boxes. This traditional "replacement" strategy is expensive in terms of equipment as well as migration costs.

Third, organizations must recognize that they have to better classify their information. The idea behind classifying information at the time of creation is to have a clear record as to what information exists so that at a later date it can be found and used effectively.

Q. Can you have one platform for all data?

- A. When you are considering some of the large Web environments — such as Google or YouTube — the goal is to have one environment. What's key in this case is that this doesn't mean one box. When you are thinking in terms of one platform, it's not just a question of buying the biggest NAS box. Having one platform for all data really means that an organization has one system for organizing and managing the data. An organization can use multiple hardware devices if for no other reason than it wants to take advantage of SATA disk drives to do low-cost tiering.

For example, an organization with high-performance information that requires quick access may want to deploy a storage system that's designed to support that. On the other hand, the organization has archived data that may never get looked at again for the 10 to 50 years that the data will be stored. The organization may want to put that data on a low-cost system even if there's a performance penalty. It needs a common, but modular software platform for organizing the information and moving the information between tiers as needed. As IDC sees it, a single platform is actually a critical requirement because a single platform is more about the management of data and not about the actual box in which data is sitting.

Q. What advice do you provide clients?

A. When IDC advises companies, we tell them first to take a quick inventory of existing or potential under-the-radar data drivers. In many organizations right now, one of the most significant drivers of unstructured content that is often overlooked by the IT department is the explosion of collaboration technologies like SharePoint. That's why it's important to take an inventory of what is in IT as well as what is in the business units, right down to the departmental and workgroup levels. The goal of any inventory is to identify those things that are going to be the big data generators over the next three to four years.

Next, we advise companies to rationalize their file environments. Organizations should try to eliminate those dispersed independent systems as much as possible. Such dispersed systems are risky and cause additional administrative overhead. It's important to look for a solution that allows an organization to have a common data store and also enables the ability to be intelligent about where data is placed.

The key to a successful storage and data management strategy is to recognize that not all content is the same but that all content is ultimately valuable. Thus, organizations need a system that allows them to intelligently place, manipulate, and organize data for different points in its life cycle — from the moment of creation to the time when the data is finally disposed. A sound information management strategy plus a foundation of products that allow an organization to act on that management strategy is critical for an organization that wants to take advantage of the information explosion as opposed to being overwhelmed by it.

A B O U T T H I S A N A L Y S T

As vice president, Storage Systems & Executive Strategies, Richard Villars is a senior member of IDC's Information Infrastructure research team, which assesses the development and adoption of storage solutions for rapidly growing information assets. He develops IDC's viewpoints on the evolution of storage networking infrastructure and next-generation storage technologies. He advises clients on the impact of tiered storage, bladed servers, virtualization, and regulatory compliance on organizations' storage and information management practices.

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 herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC 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. Translation and/or localization of this document requires an additional license from IDC.

For more information on IDC, visit www.idc.com. For more information on GMS, visit www.idc.com/gms.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com