

White Paper

Extracting the Value of Big Data with HP StoreAll Storage and Autonomy

By Terri McClure, Senior Analyst and Katey Wood, Analyst

December 2012

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

Contents

Hyperscale Storage and Real-time Intelligence for Big Data.....	3
Big Data Demands Scale	4
Add an Integrated Intelligence Approach	5
Archive Big Data.....	5
Accelerate Analytics with Meaning-based Retrieval	5
The Bigger Truth	7

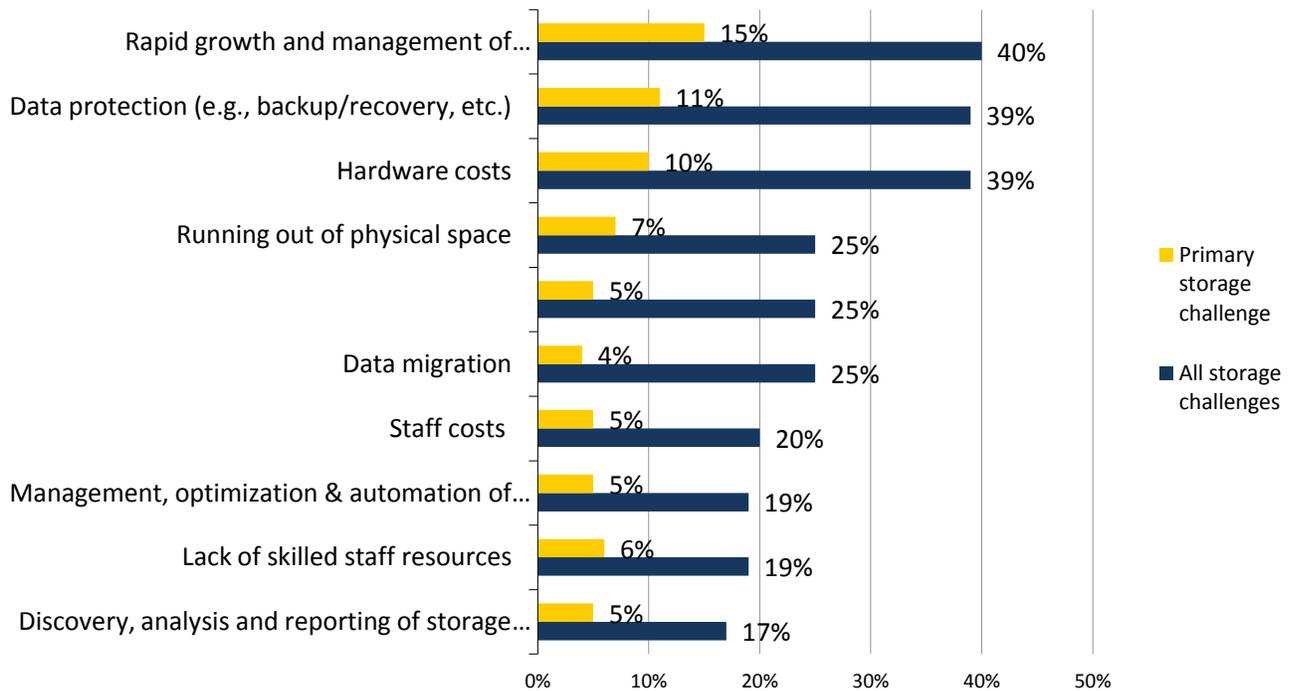
All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of The Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at 508.482.0188.

Hyperscale Storage and Real-time Intelligence for Big Data

Organizations large and small have been struggling with the storage and management of unstructured data for quite some time. In fact, managing unstructured data was chosen by 40% of the storage professionals surveyed for a recent ESG research survey as one of their biggest storage challenges, with 15% identifying it as their primary storage challenge, placing it at the top of the list this year (see Figure 1).¹

Figure 1. Top Ten Biggest Storage Challenges

In general, what would you say are your organization’s biggest challenges in terms of its storage environment? Which would you characterize as the primary storage challenge for your organization? (Percent of respondents, N=418)



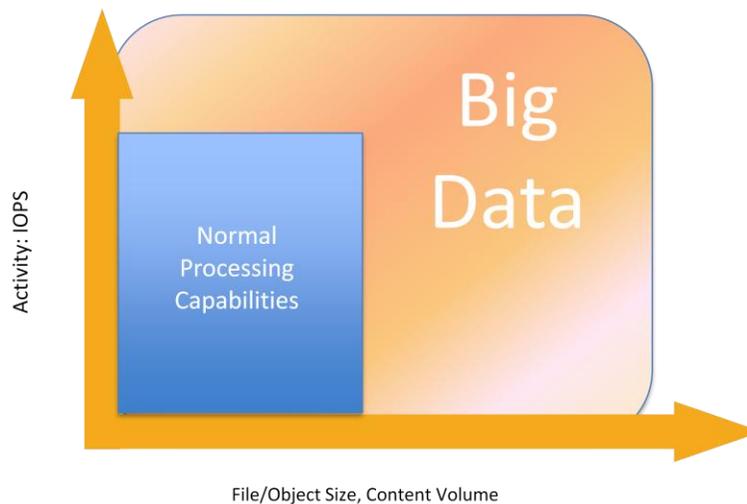
Source: Enterprise Strategy Group, 2012.

Organizations are currently storing more data than ever before, and with the advent of “big data”—or data sets that exceed the boundaries and sizes of normal processing capabilities (for instance, from the standpoint of IOPS, file size, or just plain magnitude), that force users to take a nontraditional approach to storage (see Figure 2)—the challenges of managing unstructured data are magnified. But this also presents IT with an opportunity.

¹ Source: ESG Research Report, [2012 Storage Market Survey](#), November 2012.

Figure 2. ESG's Big Data Definition

Data sets that exceed the boundaries and sizes of normal processing capabilities, forcing you to take a non-traditional approach



Source: Enterprise Strategy Group, 2012.

The essence of the challenge can be broken down into two parts. The first is to *store massive amounts of data economically and efficiently*. Once organizations have successfully accomplished that, the second challenge is to *leverage the data for competitive advantage*. That means that IT organizations need a way to both understand this unstructured data, regardless of format, and then leverage it to extract true business value based on the meaning of the unstructured data.

Big Data Demands Scale

The first piece of the puzzle is the implementation of a highly scalable multi-protocol storage pool—one that can scale performance with capacity—in other words, a nontraditional storage system that meets the demands of big data because it handles more throughput, capacity, and/or IOPS than a traditional system can. That's where [HP StoreAll Storage](#) comes in. It is a system that meets the new needs defined by today's big data environments:

- Able to scale-out to add capacity and performance into the multi-PB range
- Simple and automated
- Flexible to suite multiple use cases
- Intelligent
- Economic

In a bit more depth: Today's big data applications really need storage systems that can scale both capacity and performance. And that scale needs to be modular, so the system grows with you in a manageable and predictable pay-as-you-grow fashion, allowing you to ride commodity cost savings curves. And to be economic, the system must be able to intelligently tier data to match the cost of storing data with your organization's performance and protection needs. Managing the system needs to be as simple and automated as possible in order to avoid creating an IT administrative burden, with support for storing either files or next generation objects. And in this big data era,

let's not forget that the system needs to be readily accessible at all times. That means non-disruptive upgrades and rolling migrations are now a necessity, not a luxury, so that organizations can support everything up to and including lease rollovers without downtime.

But storing data is only the first part of the puzzle. Now that they've collected all this unstructured data in one place, one highly scalable storage system, organizations need to consider how it can now be leveraged for competitive advantage. And that requires a layer of intelligence.

Add an Integrated Intelligence Approach

Actually, a lot starts with the intelligence of the underlying information platform—and having the ability to increase performance and capacity in a scale-out fashion gives you horsepower to integrate applications such as meaning-based index and search—and that is what HP has developed leveraging the Autonomy IDOL-powered suite of applications, including Autonomy Consolidated Archive. New upgrades to HP's storage assets, in tandem with the seamless integration of the Autonomy IDOL-powered software platform, enable intelligent information access and retrieval based on meaning-based analysis of the relevance and importance of the data even at hyperscale. This combination is crucial for big data applications requiring real-time updates, as well as more cost-efficient and intelligent retrieval from long-term storage.

HP's approach rests on a two-pronged storage solution strategy: one prong around archive and one around analytics. First, and as a basis for these solutions, is the StoreAll storage platform, which offers significant differences from traditional scale-out options through its object visibility with new interfaces.

Archive Big Data

On the archive front, Autonomy Consolidated Archive integration through the StoreAll Retention API can be used to extend intelligent information retention and retrieval based on meaning, for archiving at scale. As data volumes increase, archiving can be an onerous task riddled with inefficiencies.

The integration of StoreAll with Autonomy's Consolidated Archive and retention capabilities offers a more scalable repository for content to be archived to the storage platform and retained longer term at greater cost efficiencies.

This extends the Autonomy Consolidated Archive's meaning-based capabilities for e-discovery, legal hold preservation, data management, and other internal use cases requiring intelligent retention and retrieval. E-discovery and searches are based on indexes managed by Autonomy IDOL.

Benefits include:

- Better ability to make decisions around retention and disposition—letting customers understand the business value or risk of data by examining its contents and managing it appropriately for optimal systems performance and business security.
- Faster querying of long-term, large scale archives—crucial to efficient internal investigations and meeting e-discovery production deadlines.
- Ability to create storage management profile on consumption and future capacity planning.

Accelerate Analytics with Meaning-based Retrieval

Another Autonomy-based solution HP is rolling out with StoreAll tackles analytics. StoreAll Express Query is a new embedded metadata database from HP Labs created to significantly improve filename scan speeds on data across the platform—HP claims it runs up to 100,000 times faster, enabling rapid retrieval from the repository for real-time data analytics. Also on the analytics front, HP has Autonomy IDOL connector to Express Query, which can be effectively leveraged in dynamic data environments to identify new data to be analyzed and indexed based on meaning in near real-time, increasing intelligence and more efficiently extracting value, particularly for rapidly updated applications.

Extraction for Real-Time Analytics

A number of compelling benefits can be gained once you have all the pieces in place—such as extraction for Real-Time Analytics:

- Autonomy's IDOL engine can index data in parallel with StoreAll, for example from a production system, in order to access it in real time.
- The HP StoreAll platform makes Autonomy more efficient by accelerating the rate upon which new data is identified and handing it over to the IDOL platform to be indexed based on meaning in order to power everything from universal search to e-discovery and legal hold.
- HP Express Query enables real-time insight without impacting performance from accessing underlying storage.
- Autonomy IDOL can then understand the meaning of data more rapidly, and send it to the appropriate application for analytics functions, getting data into the analytics pipeline sooner.

This is critical for leveraging intelligence from applications based on huge volumes of data generated by mobile devices, social networking, internet sites, or other large-scale, real-time data problems.

The seamless integration of HP StoreAll with Autonomy's Consolidated Archive and retention capabilities also offers a scalable repository for content to be archived to the storage platform and retained longer term at greater cost efficiencies.

The Bigger Truth

Living in a big data world requires new thinking and a new approach to storing, managing, and leveraging information. It starts with a highly scalable storage platform that consolidates data into one place where it can be stored and managed efficiently. With the right storage platform—one that supports today's file formats and tomorrow's object stores, and that adds performance and capacity—new layers of intelligence can be added. Having big data is just the start—putting it to use is what drives competitive advantage. That's the secret sauce to creating business value through IT. There is a tremendous amount of data that companies have been storing and maintaining without extracting true, actionable business value. To get there, organizations need *visibility* into their information assets and a way to intelligently store them for the right period of time. The fastest way to get there is through an integrated solution, such as HP's Autonomy-powered StoreAll platform—itsself a scalable, intelligent big data platform. When combined with Autonomy Consolidated Archive for policy-based retention and disposition, or IDOL and Express Query for more efficient analytics and faster business responsiveness, the storage platform becomes a powerful business acceleration tool.



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

HP Publication Number 4AA4-4553ENW