



A DataGravity White Paper:

# Data-Aware Storage: Introducing a New Era of Data Management

## CONTENTS

Living in a Data-Driven World .....	pg 2
Evolving From Blocks to Value .....	pg 3
Introducing Data-Aware Storage .....	pg 4
The Four Pillars of Data-Aware Storage .....	pg 6
The Benefits of Data-Aware Storage .....	pg 7
Moving Beyond Traditional Storage to High-Value Data Management .....	pg 8
About Data Gravity .....	pg 9

### DataGravity, Inc

100 Innovative Way, Suite 3410  
Nashua, NH 03062

P 603.943.8500  
info@datagravity.com  
datagravity.com



DATAGRAVITY



@DATAGRAVITYINC



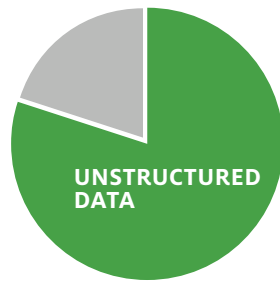
DATAGRAVITY

## LIVING IN A DATA-DRIVEN WORLD

Whatever your favorite metaphor may be – mountain, tsunami, landslide, or avalanche – it’s an inarguable fact that we live in a world that is creating *and storing* more data than most of us can imagine. From presentations and emails, to videos and photos, to documents of every type, we create *2.5 exabytes* of data *every day* – and that volume shows no signs of abating.

The biggest contributor to that growth is *unstructured data* – the data that lives outside the rows and columns of classic database systems. Definitions of unstructured data vary, but put simply, it is data with little or no metadata or classification attributes. It comprises documents, presentations, emails, videos, images, and many other file types. Experts believe that 80 percent of all corporate data is unstructured, and the volume of that unstructured data is growing 60-80 percent annually.<sup>1,2</sup>

### Unstructured Data Growing Exponentially



**80%**

IT'S ESTIMATED<sup>3</sup> 80% OF ALL CORPORATE DATA IS UNSTRUCTURED.



**60-80%**

UNSTRUCTURED DATA IS GROWING<sup>3</sup> 60-80% ANNUALLY.

Not surprisingly, this unconstrained growth has created significant manageability problems. Facing petabytes of data, exasperated companies are failing to tap into the latent value of these assets. It’s simply too hard to find what you’re looking for. And that translates into unacceptably high (and rising) costs, lost opportunity, and suboptimal corporate performance.

According to Gartner, professionals from storage admins to business analysts are in search of better ways to manage their unstructured data. The catalyst is not only lowering expenses and potential risk, but also a desire to harness information for better decision-making. Here’s how Gartner outlines the problem of harnessing information in its report:



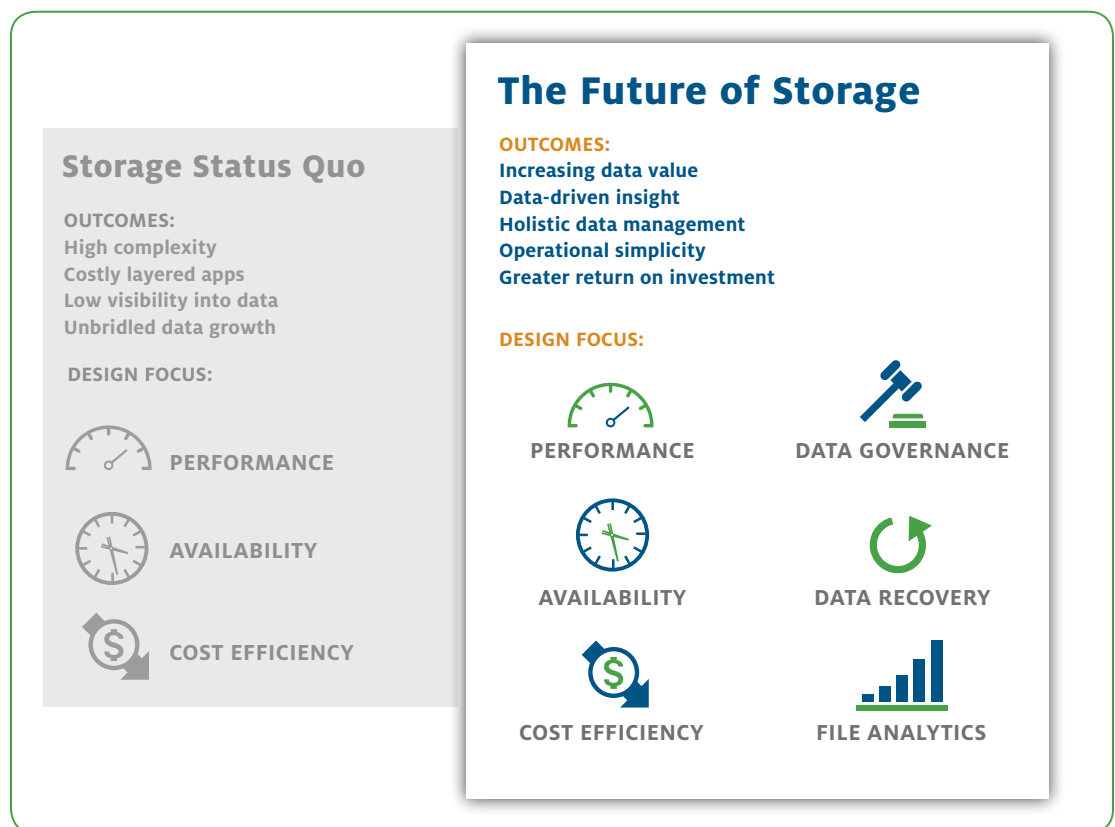
A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com

*“Organizations are struggling with ever-growing and uncontrolled data perplexity. Progressive and cost-conscious organizations are realizing they need a better understanding of their data to effectively manage their growing storage environments as well as to facilitate business use of data that has previously been ‘dark.’”<sup>3</sup>*

Just possessing all of that data creates significant risk exposures as well. For instance, in litigation you face the high costs of finding relevant information during discovery phases. Breaches and unauthorized access to confidential or private/sensitive information can create unpleasant headlines, unwelcome regulatory scrutiny, and unacceptable costs. Better methods are needed to tap the potential value of all this data and neutralize the risks associated with managing and protecting this data.

## EVOLVING FROM BLOCKS TO VALUE



To date, business has viewed storage as a necessary infrastructure component, one that must be managed carefully to achieve high performance and maximize availability and reliability at a low cost. Those are certainly worthy and crucial metrics, but they miss the core issues underlying all of that data growth: What will you *do* with that data, after all? We must find ways to transform the mere “better, faster, cheaper” storage of bytes and blocks into a value-producing proposition that enables the company to manage, govern, and leverage this rich asset.

That starts by confronting the reality: We cannot continue to expand our “as-is” storage infrastructure forever – throwing more, bigger, or faster boxes at the problem still doesn’t address the core issue.

Status quo storage only stores data. IT organizations must subsequently use layered applications to make sense of the data – to perform backups, enforce governance, enable searches, and other important functions. Since those applications operate *outside* of the storage device, they use an “install and crawl” post-processing regimen that drags down primary storage performance and siphones off production cycles to non-production tasks. Or they insert themselves into the network between the server and storage in an attempt to correlate all the traffic and changes.

Either way, the result is an increasingly costly and unmanageable framework with significant overhead, complexity, and latency. Ironically, the one resource through which all data passes, namely storage, knows practically nothing about the data it stores.

While accelerating access and reducing costs are important, they’re not enough. We want to derive value from the data that we store – to enhance productivity, enable data-informed decision making, and at the very least, reduce data management overhead. That means not only storing the data, but adding intelligence within the storage infrastructure to capture content and metadata, from which to track, index, correlate and surface data insights.

Moving forward, the storage paradigm will shift, from one that revolves around capacity and speeds, to one that’s predicated on a return-on-investment (ROI) derived from data-driven insights that are impossible to identify using existing architectures, layered applications, and equipment.

## INTRODUCING DATA-AWARE STORAGE

In response, a new storage architecture has emerged: *data-aware storage*. Data-aware storage combines state-of-the-art primary storage with integrated data services to create an entirely new class of storage that gathers information at the point of storage to provide unprecedented intelligence and insights. This insight not only serves IT administrators, but also other key stakeholders throughout the organization, including line-of-business users and governance and compliance professionals.

### **Data-Aware Storage dey-tuh uh-wair stawr-ij**

*An integrated hardware/software platform that not only stores and protects the data, but also understands that data, enabling search, discovery, and visualization of the data, and its related attributes, including people, content, time, and activities.*

To date, IT professionals have had to devote significant resources and efforts to understand the data they are storing, protect it, and derive value from it – even to simply understand who is accessing it and how, what data has changed, and what data must be protected. Unfortunately, those time- and resource-intensive efforts are happening *after* the data has been stored using siloed, serial processes – data protection, search/analytics, security, and more.

### SMART THINKING



Increase the ROI of your storage investment by gaining data insights directly from the storage



A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com

## SMART THINKING

Capturing data about the data at the time and place it's created enables rich insights.



## Data-Aware Storage Value Proposition

- **Storage Visualization** See all of your data in real time through intuitive click-and-drill visualizations.
- **File Analysis** Get a 360-degree view of data: consumption, activity, and demographics.
- **Information Discovery** Data-aware storage supports searches and discovery across people, content, and time.
- **Activity Tracking** See who interacted with what data and when.
- **Data Protection** Protect against physical or virtual data loss.

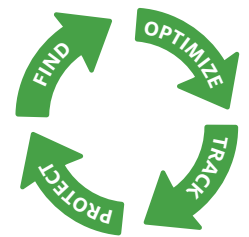
By contrast, a data-aware storage architecture includes an intelligence engine that silently and efficiently protects data, tracks activity, extracts insights and content, and puts idle resources to use – *mining the data at the time and place at which it is stored* – without any impact on production I/O. Beyond the basics of reliably storing data, key strategic objectives can be met, making valuable data more accessible, reducing dormant data, securing data from misuse, and protecting it from loss.

*“This...could literally revolutionize what we think of as ‘storage,’ bringing a new era of data management and a new relationship between IT and the business.”*

Stephen Foskett, Pack Rat Blog<sup>4</sup>

## Data-Aware Storage – The Data Flow

- **Optimize data as it is written** Store and serve data via NFS, CIFS, and iSCSI, including virtualized environments, while performing inline data compression and targeted deduplication.
- **Track data as it is used** Track data creation, changes, deletions, and user-access activities throughout the data lifecycle.
- **Protect data as it is stored** Continuous data protection through fault-isolated, complete catalogs of data changes and activities.
- **Find data as it is needed** Automatically create a searchable data repository, enriched with file and activity metadata.



A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com

# THE FOUR PILLARS OF DATA-AWARE STORAGE



## State-of-the-Art Primary Storage

At its core, data-aware storage does not compromise on its key purpose of storing data for high performance access while maintaining data integrity and availability. Data-aware storage appliances should embrace state-of-the-art technologies for primary storage. They should support block and file protocols, and natively support today's virtualized environments. They should be flash-optimized, and contain advanced features expected in storage such as thin-provisioning, clones, in-line compression and targeted deduplication. To be effective, data-aware storage needs to be meaningfully affordable to acquire and operate, not for the exclusive use of the world's largest, best-funded organizations, but for mainstream use across organizations of all sizes.



## Integrated Search and Discovery

A well-designed data-aware storage platform provides exponentially greater simplicity for users – from IT to line-of-business - who want to explore and visualize information. This means having a user interface that enables users to explore the data – a concept that is unprecedented in storage platforms. Those users want and need the ability to correlate data across people, time, activities, and content types. With data-aware storage, all content is fully indexed, enabling rapid keyword search. In addition, data-aware storage not only stores data and associated metadata, but also tracks activities by timeframe and user. This enables a new class of complex queries. For instance, you can find archived assets for a marketing campaign or identify internal content experts and potential collaborators by searching on content, timeframe, and author.

That requires high-caliber horsepower and an architecture that builds near real-time indexes from hundreds of file types as the data is stored. The outdated alternative is to have layers of applications, agents, and spiders crawl through your *already-stored* data to rediscover and index your data. This creates new management headaches with additional licensing and maintenance demands. By contrast, data-aware storage builds and updates indexes as data is created and modified in the primary storage appliance without impacting the production environment.



## Enhanced Data Governance

More than ever, companies are facing strong pressure to be more effective stewards of their data. Recent headlines about major data thefts provide a clear sense of the scope of damage that may result from a failure to protect data fully.

*“Data-aware storage will provide a quantum shift in the way enterprises utilize their stored data assets and ultimately lead to a significant competitive advantage.”*

Ron Powell, BeyeNETWORK <sup>5</sup>

Data-aware storage provides near real-time data-activity tracking so you can see who has accessed which content and when. Who created the content? Who read or updated it? Who changed user permissions? What *kinds* of data are in those files? Simple analytic views and powerful visualizations make it easy to proactively govern data, and address questions from auditors and regulators. For instance, a real-time internal audit may detect personally identifiable information (PII) that is needlessly and dangerously exposed to viewing or theft. By catching that vulnerability prior to a formal review, the company can sidestep potential sanctions.



## Instant Recovery and Zero-Impact Protection

In typical storage architectures, backup and recovery functions are managed by software applications that overlay the storage hardware to read data that was just stored – and store a copy elsewhere. This is an inherently inefficient process. Snapshot processes can provide rapid recovery of data, but are typically stored with the production data, so vulnerabilities remain. Data-aware storage combines the simplicity and instant recovery of snapshots with the fault isolation and granular protection of traditional backups.

Data-aware storage protects the data *as it is created*. It provides granular-level restores – approaching real-time, that can be far simpler and efficient than a restore from a snapshot. Since the data-aware storage architecture already knows the versions and catalog, restorations to the production system are instantaneous, whether you are restoring an entire LUN, file share or VM, individual blocks within a file share or VM, or specific files within a file share or VM.

## THE BENEFITS OF DATA-AWARE STORAGE

A data-aware storage platform not only serves and protects data more simply, but also *surfaces actionable insights and issues* that streamline data management, reduce business risks, and increase user productivity.

*“If we were data-aware, we would be able to find malware and audit, perform forensics, secure, and protect our data at a much finer-grained resolution.”*

Edward Haletky, Data-Aware Services: Oh, the Places We Could Go! <sup>6</sup>

- **Streamlined Data Management** – Using enterprise-proven hardware, data-aware software architectures improve operational efficiency and visibility by seamlessly consolidating and automating workflows across storage, protection, governance, and search and discovery processes. That reduces operational overhead to near-zero while giving IT and business users unprecedented visibility and protection of file systems, file shares, and virtualized environments. What’s more, the financial profile points to lower CAPEX and OPEX and lower total cost of ownership.
- **Lowered Business Risks** – Data-aware storage provides advanced insights, so you can better govern your data from the point of creation. Identify dormant data or prevalent file types



A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com



### SMART THINKING

Understanding what is stored can help you find, protect and recover your data more efficiently.

and prolific users to help enforce data retention and deletion policies. Find and protect sensitive data, and track access in order to reduce your risk exposures.

- **Increased User Productivity** – By providing a means to easily discover information within the storage array and correlate this information with people, time, and content, data-aware storage improves productivity and collaboration while promoting improved decision-making. In addition, data-aware storage can consolidate independent processes to create seamless workflows that reduce IT overhead, management, and maintenance costs.



## Data-Aware Storage Unlocks the Latent Value of Your Data

- Achieve unprecedented data visibility and protection
- Identify dormant data, prolific file types, and users
- Track data access, find and protect sensitive data, and assess exposure risk
- Support governance and compliance initiatives
- Find valuable data and content assets more easily
- Identify subject-matter experts, key collaborators, and content authors
- Find and recover data instantly

## MOVING BEYOND TRADITIONAL STORAGE TO NEW DATA INSIGHTS

We live in a data-driven world, and companies everywhere are struggling to benefit from their growing unstructured information assets. It is imperative to unlock the value embedded in this information by embracing a new storage software architecture that does more than just create more and larger data repositories.

In a data-driven world, data-aware storage represents a quantum shift in the way data insights are exposed and exploited. By utilizing an innovative storage architecture and integrated capabilities to protect, organize, and explore data, companies can transform storage from a tactical resource to a strategic asset that provides far greater value back to the business.

Forward-thinking organizations are rethinking the role of storage. No longer just a “bit bucket,” data-aware storage can be empowered to extract insights and business value from the content-rich corporate resources that are stored. Through built-in intelligence, data-aware storage architectures play an increasingly prominent role in the organization’s ability to rein in data growth and more fully leverage the value of the assets it is rapidly accumulating.

*“Purchasing new storage technology is no longer a simple storage decision, but an organizational decision based around improving productivity, lowering cost, and reducing risk.”*

ESG Lab Review, DataGravity Discovery Series<sup>7</sup>



A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com



*“DataGravity has the ability to not only replace our current storage vendor, but also enhance how we run our law practice by providing insights into data we couldn’t easily search before.”*

Chris Berube IT Manager, Law Offices of Joe Bornstein

## ABOUT DATAGRAVITY

Organizations with fast-growing, file-rich environments need smarter solutions to strategically manage and benefit from their growing data resources. DataGravity Discovery Series data-aware storage platforms streamline data management, reduce business risks, and increase user productivity.

In contrast to classic storage systems, DataGravity not only serves and protects data more simply, but also surfaces insights and issues embedded in the data through an advanced user interface designed specifically for data and file analytics.

Unlike traditional storage and layered data management and governance applications, the DataGravity system transforms multiple discrete processes into a single automated workflow that stores, tracks, protects, and searches unstructured data cost-effectively.

For more information, visit [www.datagravity.com](http://www.datagravity.com)



A DataGravity White Paper:  
**Data-Aware Storage:  
Introducing a New Era  
of Data Management**

info@datagravity.com  
datagravity.com

---

<sup>1</sup> [http://wikibon.org/wiki/v/The\\_Growth\\_and\\_Management\\_of\\_Unstructured\\_Data](http://wikibon.org/wiki/v/The_Growth_and_Management_of_Unstructured_Data)

<sup>2</sup> <http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy>

<sup>3</sup> Gartner, Inc report entitled “Market Guide for File Analysis Software” - September 32, 2014

<sup>4</sup> Stephen Foskett, Pack Rat Blog, August 19, 2014 - <http://blog.fosketts.net/2014/08/19/datagravity/>

<sup>5</sup> BeyeNETWORK, Ron Powell Blog, August 25, 2014 - [http://www.b-eye-network.com/blogs/powell/archives/2014/08/datagravity\\_pro.php](http://www.b-eye-network.com/blogs/powell/archives/2014/08/datagravity_pro.php) The Virtualization

<sup>6</sup> Practice Blog, September 3, 2014 - <http://www.virtualizationpractice.com/data-aware-services-oh-places-go-28659/>

<sup>7</sup> ESG Lab Review - DataGravity Discovery Series Authors: Aviv Kaufmann, ESG Lab Analyst and Mike Leone, ESG Lab Analyst - August 2014 - <http://datagravity.com/sites/default/files/resource-files/ESG-Lab-Review-DataGravity-Discovery-Series-Aug-2014.pdf>