Executive summary

Storing, preserving, and protecting digital assets is proving to be one of the most significant challenges facing IT today. Growth of unstructured data – data not stored in conventional database systems – is accelerating dramatically, fueled by high-definition everything and new data hungry applications that depend on unstructured data.

Research firm IDC predicts many organizations will experience up to 80 percent yearly growth in unstructured data.¹ At this rate, many IT groups will be fast approaching petabyte-scale storage capacity in the near future. Conventional storage systems were simply not designed to cost-effectively manage this much data, and, when demand crosses the petabyte threshold, public cloud storage services can also get very costly, very quickly.

Object-based storage systems are designed to serve petabyte-scale storage and beyond. These systems can be ideal for unstructured data and are being used by web-scale cloud storage service providers to support their own backend infrastructure needs.

IBM® Cloud Object Storage (COS) is a leading software platform for web-scale storage.² The company’s software-based object storage has been developed for exabyte-scale deployment, and is already in use by multiple customers who each have greater than 100PB of unstructured storage capacity in production.

What sets IBM COS apart from other object-based storage implementations, however, is a focus on delivering robust solutions, for many of the most demanding petabyte-scale applications in use today. IBM has teamed with technology vendors and certified content repository, public cloud services, as well as enterprise collaboration, backup, and active archive solutions to create tried and tested implementations in the company’s integration lab. This approach enables IBM to provide customers with consistent support and successful deployment of these critical applications.
Part one
Growth-on-growth

Growth of unstructured data – data not stored in conventional database systems – is accelerating dramatically. It’s not just that more individuals and machines are generating more data. Today’s unstructured data is substantially larger, more complex, and more dense than before, requiring more capacity and more management resources to store and protect.

Up until recently, unstructured data consisted primarily of text-based documents, like presentation files and email. Their size was measured in kilobytes. As use of digital audio, images, and video became more common in business, file sizes grew from megabytes to gigabytes. With extensive use of high-definition video and applications like genomic sequencing and seismic data analysis, vast amounts of data are being generated, and it’s not uncommon to see file sizes measured in terabytes. This trend is compounding the growth of unstructured data.

The growth-on-growth phenomenon is challenging IT organizations of all shapes and sizes. Even well-funded IT groups are finding the rapid rate of growth to be practically unmanageable.

80 percent growth is common

Recent analysis by research firm IDC, found that unstructured data now constitutes up to 80 percent of all data in a typical organization. Growth rates of 60 to 80 percent are common, with some data-intensive organizations witnessing over 100 percent growth per year. This rate is increasing (IDC, January 2015).

This rapid rate of growth means many organizations now face the prospect of operating their storage infrastructure at web-scale: scale of the size and complexity only previously experienced by cloud service providers, like Amazon, Facebook, and Google. As the organization’s data storage needs cross the petabyte threshold, a completely new approach to storage infrastructure and data management is needed.

Operating IT at web-scale

Operating at web-scale means finding ways to scale capacity and capabilities, without a corresponding increase in IT budget and resources. However, achieving web-scale does not happen overnight. Despite declining infrastructure costs, from greater use of commodity hardware, organizations face tremendous budget and resource challenges. This frequently leads to compromises in what data to keep online and for how long, how many staff to hire to manage the growing storage environment, and even data availability.

As stewards of critical digital assets, IT organizations have the responsibility to help ensure data is always available and protected against potential loss or compromise. This can be challenging in a world where cybercrime is pervasive and the cost of downtime is crippling. Overprovisioning storage and data replication are common approaches to solving data availability, durability, and security challenges. But at web-scale, these techniques can compound existing problems and quickly become unsupportable.

Unstructured data growth also has a knock-on affect on the ability of IT to respond quickly to the organization’s need for business applications. Application development agility and flexibility are prized in today’s fast-paced and highly competitive business world, but resources consumed in managing data growth distract IT teams from rapidly deploying new and innovative solutions to improve the business.

Figure 1: Diagram showing the growth in unstructured data
Part two
Web-scale storage architecture
The criteria for a new web-scale storage architecture to satisfy the need for unfettered growth of unstructured data reads like a storage administrator's ultimate wish list. Storage must be flexible to support multiple applications simultaneously; cost-effective, with the ability to meet the demands of unconstrained data growth; secure, to protect data from inadvertent or malicious corruption and always-on, eliminating the potential for application downtime. And if possible, it should work alongside existing storage components to avert the need for expensive “rip and replace” upgrading of the storage infrastructure.

Public Cloud storage
Public cloud offerings enable IT teams to quickly apply storage resources to address demand for unconstrained growth. These services offer flexibility and ease of use for storing unstructured data, with the simplicity of an off-premises infrastructure. It is a compelling value proposition. However, as storage capacity approaches the petabyte threshold, the cost of these services starts to become significant. In addition, many are lacking in enterprise-level security and have an inability to control where data lives, something IT organizations might want to consider when thinking about public cloud storage as an alternative.

Traditional SAN and NAS
Traditional SAN and NAS storage systems are trying to rise to the challenge of unstructured data growth. These solutions can provide a vast array of built-in and bolt-on management capabilities. However, they are often hampered by storage architectures that were developed before web-scale capacity was a requirement.

Open-source software
Multiple open-source, software-defined storage projects exist that promise to deliver capabilities necessary to meet the challenge of unstructured data growth. These projects benefit from licensing that makes a solution effectively free-of-charge. But as compelling as open-source solutions are on paper, they inevitably lack maturity and are often missing the tools IT needs to enable ease of management. To make these systems functional, IT groups must dedicate resources to tailor the solution to suit a particular environment. This can distract from the overall goal of making IT more responsive to the needs of the organization.

Object storage
There are many new storage solutions available from established and emerging vendors that claim to deliver the capabilities needed to address the challenge of unstructured data growth. These storage systems are predominantly object-based. They offer a variety of data management capabilities and interfaces to give IT organizations the flexibility they need. The object storage architecture was developed with web-scale infrastructures in mind. It is designed to provide flexibility, cost-effectiveness, security, and availability, making it well suited for hosting unstructured data.
Part three
IBM COS

IBM is a market leader in object storage systems. The company’s software-based storage platform has been demonstrated in many of the world’s largest data repositories, with multiple customers having greater than 100PB storage capacity in production environments. IBM COS system is designed for deployment on a variety of hardware platforms, giving customers flexibility and choice for their hardware infrastructure. A sophisticated, simple management system helps enable an administrator to control tens of petabytes of storage capacity and support an always-on architecture that verifies data is available independent of failure, expansion, upgrade, or relocation.

The IBM COS System incorporates built-in, carrier-grade security capabilities at the core of the data management process. This helps ensure data is protected from compromise, whether accidental or malicious. IBM COS delivers these capabilities in a cost-effective storage platform that becomes more efficient as the system grows.

A solution-based approach to object storage

IBM recognizes that delivering a highly reliable, secure, and scalable storage platform is not enough on its own to address the storage challenges faced by IT organizations. That’s why the company has teamed with a number of industry-leading technology vendors to develop and deliver integrated solutions that help IT organizations address specific elements of their digital application portfolio and deliver real business value.

Content repository

Content repositories preserve and protect an organization’s essential digital content. Repository selection is based on a number of key factors, including reliability, petabyte scalability, security, and ease of management. A content repository solution must be able to deliver high data durability and provide built-in site fault tolerance, without the need for expensive replication. Software-based content repository storage solutions tend to offer the most flexible deployment, supporting a variety of hardware platforms.

Figure 3: Content repository typical for image, audio and video data
IBM COS helps deliver the capabilities organizations need in a content repository. The software-based storage solution integrates with content management solutions, such as Microsoft’s SharePoint and Integrated Rule-Oriented Data System (iRODS), to provide an integrated content repository solution.

In the media and entertainment market, organizations use media asset management (MAM) and digital asset management (DAM) software to provide content management. Many market-leading vendors of MAM and DAM solutions are qualified to run with the IBM COS System. IBM works closely with software vendors to perform joint certification, providing confidence to customers that their chosen content repository solution will deliver the reliability, scalability, security, and ease of management they need.

A large European media broadcasting company chose the IBM COS System for their next-generation cloud-based content repository to preserve and deliver key digital assets. This company’s television broadcasts, sporting events, news shows, and other events are stored on the IBM COS System. Media is distributed from the IBM COS System, in a variety of formats, to subscribers around the world. This company also uses the IBM COS System as a repository for other types of unstructured data, including call recordings and analytics data, which helps this company manage their infrastructure more efficiently.

**Private Cloud Storage-as-a-Service (STaaS)**

Private Cloud Storage-as-a-Service (STaaS) enables organizations to deliver the storage flexibility and agility needed by internal users and customers, with the additional security and control of a private cloud environment. As with storage provided by a public cloud service provider, these systems use multi-tenancy to serve the needs of many internal organizations from a common infrastructure. It is essential that data stored by one user not be compromised by another.

IBM COS integrates with leading cloud management and orchestration platforms, including Citrix CloudStack, OpenStack, and VMware, to provide private cloud STaaS capabilities.

Encryption of data at rest and data in motion helps provide a fundamental assurance of the confidentiality, availability, and integrity of user data for multi-tenant environments. IBM COS also certifies with backup, archive, and cloud gateway solutions helping enable private cloud customers to offer fully integrated solutions to their customers.

Private cloud also offers IT organizations a chance to consolidate multiple business solutions on a single platform. This approach can be cost-effective to manage and provide the scalability, reliability, and security needed by private cloud users. The IBM COS System is robust and capable of supporting a consolidated private cloud, offering a more seamless path to grow to petabyte-scale operations, and beyond.
Hightail (formerly YouSendIt) is one of the leading web-based file sharing services. The company chose IBM to deliver file-based storage and delivery services to their customers worldwide. Hightail has over 45 million subscribers in nearly 200 countries around the world. Their service must be reliable and available around the clock. The company chose IBM for their built-in reliability and high-availability, and because IBM COS is able to perform maintenance, upgrades, expansions, and relocations without interrupting service to users. IBM COS helps enable Hightail to achieve their performance and reliability goals without expensive additional copies of data, translating into huge infrastructure savings.

**Enterprise collaboration**

Enterprise collaboration software is designed to improve productivity and enables business to be more agile and competitive. These solutions let workers more easily and securely access and share digital content no matter where they are in the world. Concerns about security often rule out use of public cloud for collaboration solutions. For enterprise users, an internal private cloud solution that facilitates secured sharing of business-critical content at web-scale is often preferred.

The IBM COS System integrates with many global file sharing gateway and synchronization solutions to enable a geographically dispersed workforce to cost-effectively collaborate on business-critical content in a protected environment. IBM has integrated with cloud gateway vendors like Nasuni, Panzura, and Avere, that enable distributed file sharing on a private cloud infrastructure. The company has also teamed with vendors of a number of file synch and share technologies that enable synchronization of data between multiple devices and safeguard sharing between many users. These vendors are fully integrated with the IBM COS System.

A global marketing and advertising firm with a large number of employees producing rich media content and advertising campaigns for customers in multiple offices around the world share data across a globally dispersed IBM COS system. The IBM COS solution allows this company's employees to collaborate on campaigns no matter of their location, enabling this company to apply the talent of their entire organization to help meet the marketing needs of their customers.
Backup
Protecting critical digital information from potential loss, whether accidental or malicious, is one of the most important responsibilities of the IT organization. However, in many cases, the data protection processes are difficult to manage and the level of protection may fall short of what the organization desires. Keeping backup data online for faster recovery in the event of a failure is a key feature of today’s backup applications. Integrating these capabilities with a web-scale storage infrastructure that incorporates site fault tolerance for disaster recovery makes sure critical recovery data is available when needed.

IBM has teamed with many leading backup software vendors. These solutions are fully integrated with the IBM COS System and designed to deliver a more reliable, security rich, and cost-effective backup and recovery infrastructure. In addition, IBM is integrating with backup applications that support RESTful interfaces.

Integrating with these data protection solutions provides IT organizations a more cost-effective, reliable, and security-enhanced online repository for backup data, with built-in site fault tolerance and disaster recovery.

Data protection is one of the primary concerns of organizations storing data in a public cloud service. A leading UK cloud service provider offers their customers protected backup storage using the IBM COS System. They chose IBM COS to make sure that their customers are able to back up their data to the cloud in a more protected and cost-effective manner to help alleviate the worry about whether someone else can gain access to their critical information.

Figure 6: Phone backup protection
**Active archive**

As organizations seek to mine more and more corporate data for competitive insight and market advantage, storage systems are being tasked with preserving archival digital content almost indefinitely. An active archive provides a repository for corporate unstructured data that is online, accessible, and available whenever needed. These systems must provide massive scalability, with the capability to easily upgrade, relocate, and refresh the underlying hardware while the system is online.

The IBM COS System is designed to deliver all the necessary capabilities for an active archive.

IBM COS integrates with a number of commercially available archive to give customers a robust, fully integrated solution.

The IBM COS System supports easy refresh of the underlying hardware, without interrupting data access and without the need to make costly copies of the data.

The Center for Data Intensive Sciences at the University of Chicago had a massive data storage challenge. They needed to build an active archive solution that could store tens of petabytes of genomics research data and allow researchers and medical professionals from around the world to easily gain access to this information. And they had to bring the system online in just a few months’ time. They chose IBM COS for their Center for Data Intensive Science, which has developed a large-scale biomedical commons for the scientific research community. The multi-petabyte system was deployed in just a few hours and was receiving genomics data very quickly. Bringing data at massive scale together for researchers around the world is helping enable collaboration and transform the study of cancer research.

---

*Figure 7*: IBM COS’s model for active archive photo and video sharing
IBM COS System
The IBM COS System was designed from the beginning to provide web-scale storage that is easier to manage and more economical to deploy. Adopting IBM COS enables capacity to increase in line with requirements, while cost-effectively helping meet the availability and security needs.

Tried and tested scalability
IBM COS software is tried and tested at web-scale. Multiple IBM customers have production deployments that exceed 100PB capacity and use a shared-nothing architecture that will scale to exabytes. These deployments are cost-effective without compromising reliability, availability, or manageability. Unlike many legacy systems that rely on a centralized metadata server, IBM COS uses a single global namespace that offers a unified, single point of management and access at web-scale.

IBM COS’s tried and tested scalability helps IT organizations to eliminate storage silos forced on the infrastructure by the limits of other solutions. Data is consolidated in a private cloud-based storage system.

Configurable reliability and availability
IBM COS gives IT organizations the ability to tune system reliability and availability settings for the needs of an application. Demanding applications can benefit from nine nines of reliability. Tuning the environment for more typical levels of reliability and availability helps deliver economic savings. IBM COS gives organizations the ability to choose the combination of reliability, availability, and economic efficiency that best suits their application and user requirements. Built-in integrity checking and self-repair features help ensure data durability is maintained.

Software-defined storage flexibility
The IBM COS System is a software-defined storage solution and requires no proprietary hardware. Using IBM COS software with virtually any number of industry standard hardware systems can help an IT organization to significantly reduce per-terabyte storage costs, allowing for the rapid and economical increase in storage capacity.

Reduced total cost of ownership
IBM COS software protects an organization’s digital assets using a unique information dispersal algorithm (IDA). Instead of storing multiple copies, the IDA encodes the data and disperses it across the system. This reduces physical storage requirements by more than 50 percent. IBM COS is able to maintain availability with only a single copy of that data, even if multiple hard drives fail, servers fail, or even during a site outage.

While the exact reduction in raw capacity requirements is a function of the specific system configuration, IBM customers report total cost of ownership (TCO) savings of up to 80 percent, compared to existing storage systems, and significant savings relative to the use of public cloud storage. Total savings reflect not only reduced acquisition cost, but also savings on data center power, cooling, and floor space.

15x increase in management productivity
IBM COS software enables an administrator to manage 15 times as much storage capacity compared to traditional storage systems. IBM COS is designed for continuous availability and allows administrators to perform software upgrades, hardware maintenance, storage capacity expansion, hardware refresh, and even physical relocation of the storage system without requiring downtime.

Data-at-rest encryption
IBM COS software incorporates built-in, data-at-rest encryption, with no need for separate key management.

A variety of authentication and access control options are also available for both data access and access to the management system.

System configuration
IBM COS is composed of three types of components. An IBM® Cloud Object Storage Manager provides a single pane of glass for configuration, administration, monitoring, and fault management. A set of IBM® COS Accesser® nodes handle storage IO requests, encrypting and encoding data as it is written to the IBM COS System. And, a set of IBM® Cloud Object Storage Slicestor® storage nodes to store the data.
Industry-based solutions
IBM COS solutions address the challenge of unstructured data growth in many industries.

Financial services
In the financial services sector, trust and compliance are two of the most important issues companies face. The IBM COS System provides exceptional data protection for customer data, helping financial services firms to offer excellent service in a highly competitive and mobile environment. IBM COS provides a security-enhanced and scalable storage solution to help these organizations eliminate data vulnerabilities and meet increasingly stringent compliance regulations.

Media and communications
Media and communications companies can rise and fall on the experience they deliver to customers. Companies are under constant pressure to produce, deliver, and preserve new content and meet demand for new innovations like 4K video resolution and on-demand video services. Whether the organization is dealing with feature-length movies, high-definition digital images, or multimedia documents, media and communications companies are able to house all of their production content in IBM COS’s reliable and security-enhanced content storage solutions. IBM COS’s massively scalable storage system helps enable these companies to meet today’s production demands and support tomorrow’s business goals.

Figure 8: Example of IBM COS configuration showing the three components
**Life science and healthcare**
Life sciences and healthcare organizations are utilizing technology and data to drive continuous improvement in people's lives and solve many of the medical industry's most difficult challenges. IBM COS delivers innovative storage solutions to these organizations to support the development of new medical breakthroughs. With IBM COS's always-on storage solutions, life science and healthcare organizations can have more confidence in the security of their data and are able to focus their time and attention on what they do best.

**Cloud service providers**
In the demanding technology world of cloud service providers, scalability is the key to driving business growth. Delivering competitive enterprise and consumer solutions depends on a service provider's ability to cost-effectively store, analyze, and access data at web-scale with confidence. IBM COS Solutions provide an easy to manage and easy to scale storage platform for exceptional service expansion and agility. With IBM COS, cloud service providers are able to concentrate their energy on developing and managing robust technology services, rather than worrying about their storage infrastructure.

**Oil and gas**
Seismic data produced for the oil and gas industry is responsible for massive growth in unstructured data. This data is frequently very difficult and costly to generate. As energy consumption around the globe increases, the need to securely store, manage, and distribute seismic data is critical to exploring and exploiting new oil and gas reserves. IBM's globally accessible storage solutions help enable exploration and discovery teams to cost-effectively manage their data to make smarter decisions and faster discoveries.

**Government**
Organizations that serve the public good are constantly challenged with the growth and preservation of unstructured data. Whether supporting military missions, scientific discovery, or civilian interests, these organizations depend on their ability to manage huge amounts of unstructured data. IBM COS Solutions help enable government agencies to efficiently store and access sensitive data, and help protect it against potential unauthorized access. The IBM COS System helps enable government and public service organizations to cost-effectively scale their unstructured data storage to meet the demands of their mission and preserve critical digital assets.

**Integrated partner solutions**
IBM teams closely with leading technology vendors to provide high quality, customer-focused solutions. Partners are often multi-billion dollar organizations with a global reach, but many are smaller companies producing innovative technologies that help solve critical challenges.

IBM is investing in technology integration labs and personnel to verify that these solutions are truly integrated and perform as expected for customers. IBM, teaming with its technology partners, develops procedures for integrated support and exchange engineering and technical knowledge. Integration labs provide a strong focus on performance and satisfying customer demand for throughput, operations per second, and latency. Customers expect solutions to work, and IBM, collaborating with its technology partners, works hard to deliver on those expectations.
Part four

Conclusion

IBM COS is an industry-leading innovative, time tested web-scale object-based storage system available today. Many of IBM’s customers have over 100PB of unstructured data storage deployed in production environments. These same customers anticipate satisfying their need for exabyte-scale capacity on the same platform. But delivering a highly reliable, secure, and scalable storage platform is, on its own, not enough for many of today’s most demanding IT organizations. These organizations need specific, targeted solutions that address very real business needs for content repositories, public cloud services, enterprise collaboration, data protection, and active archives.

IBM has teamed with many industry-leading technology vendors to provide solutions that incorporate the IBM COS System. Through the company's integration labs, partners are able to certify their products for use with the IBM COS System to give joint customers a level of confidence that these solutions will not only work well together in a production environment, but also will be jointly supported by IBM and the solution vendor.

Unstructured data growth poses a very real challenge to companies and public service agencies across a broad spectrum of market sectors. It is becoming imperative that these organizations learn how to address the challenge of storing, securing, and accessing this data if they are to reap its benefits. IBM has tried and tested technology, experience, and solutions that customers can rely on as their unstructured data continues to grow to exabyte-scale and beyond.

About IBM Cloud Object Storage

IBM Cloud Object Storage provides organizations the flexibility, scale and simplicity required to store, manage and access today’s rapidly growing unstructured data in a hybrid cloud environment. Relied upon by some of the world’s largest repositories, our proven solutions turn storage challenges into business advantage by reducing storage costs while reliably supporting both traditional and emerging cloud-born workloads for enterprise mobile, social, analytics and cognitive computing. IBM Cloud Object Storage is built on technology from object storage leader Cleversafe, acquired by IBM in 2015.

For more information

Please call: 312-423-6640 or email: sales@cleversafe.com or visit the Cleversafe website: www.cleversafe.com.

To learn more about IBM Cloud computing, please visit http://www.ibm.com/cloud-computing/infrastructure/object-storage/
IBM

© Copyright IBM Corporation 2016

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
July 2016

IBM, the IBM logo, ibm.com and Cleversafe are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at: ibm.com/legal/copytrade.shtml.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other product, company or service names may be trademarks or service marks of others.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on the specific configurations and operating conditions. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM product and programs. THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed or misappropriated or can result in damage to or misuse of your systems, including to attack others. No IT system or product should be considered completely secure and no single product or security measure can be completely effective in preventing improper access. IBM systems and products are designed to be part of a comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM does not warrant that systems and products are immune from the malicious or illegal conduct of any party.


