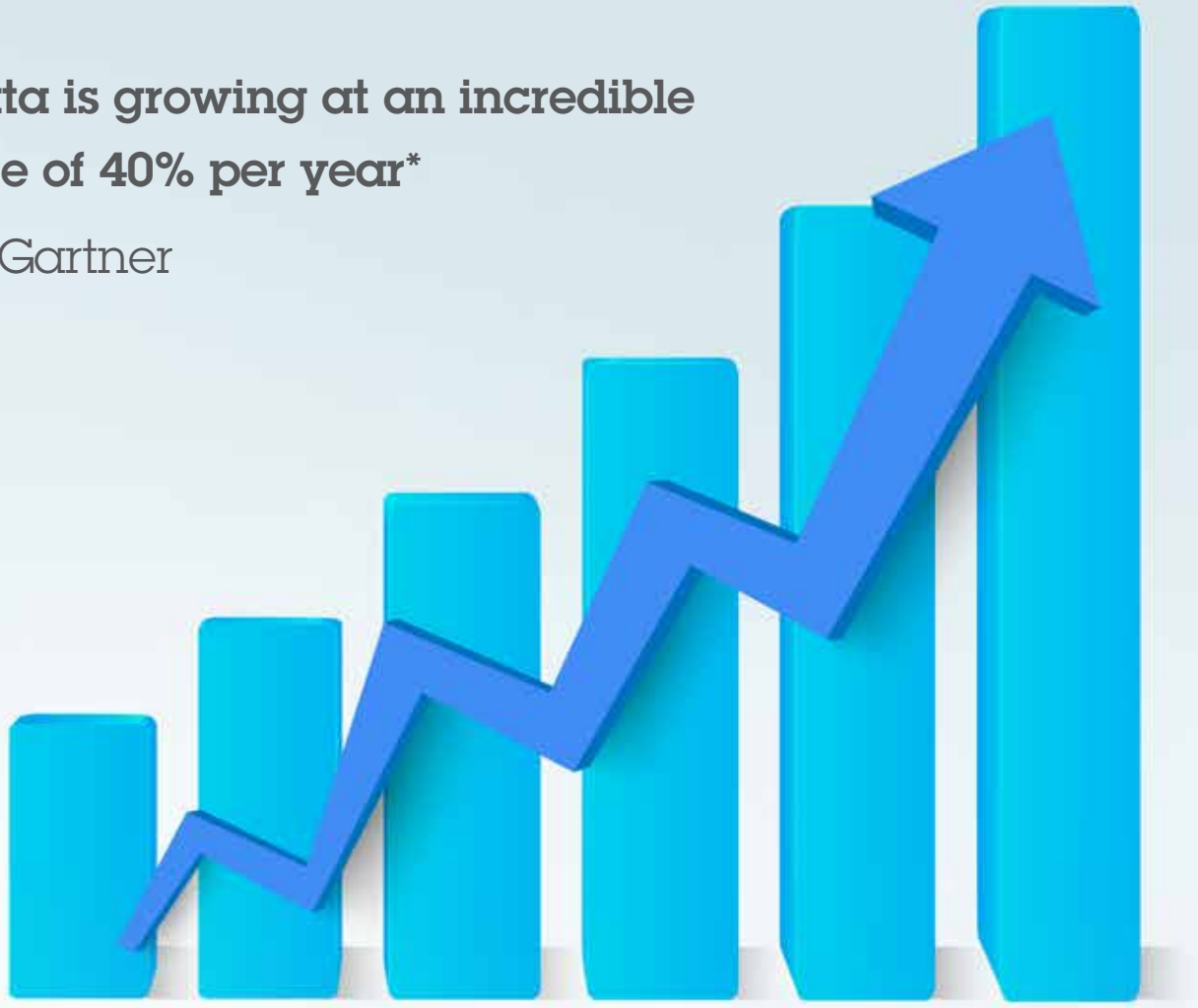
WHITE PAPER

Efficiently and Effectively Managing Data Growth

Data is growing at an incredible rate of 40% per year*

— Gartner



Secure Infrastructure & Services
48 South Service Road, Melville, NY 11747
Tel: 855-772-3447
www.siasmsp.com

Business Partner IBM
Managed Services
Infrastructure
Platform

SECURE
Infrastructure & Services

Contents

Introduction

According to Gartner group, an average IT shop's data grows at the incredible rate of 40% per year. This means that most organization's storage needs will roughly double every 2-2.5 years, causing a number of issues with system performance, storage administration, complexity, and scalability.

But what's driving this furious pace of organizational data growth? And what strategies can organizations take to efficiently and effectively manage the massive data growth that's expected over the next several years? How can they grow their application data at this furious pace while maintaining (or improving) system performance, keeping their storage costs under control, and reducing the complexities associated with data growth?

This white paper examines what creates massive data growth in an organization's data center, the real effects data growth has on organizations, and the cloud-based strategies companies can employ to get data growth under control. It provides a template for thinking about massive data growth and how migrating to the cloud can address those issues. We outline the specific problems organizations face with data growth, along with the cloud-based solutions for overcoming those problems.

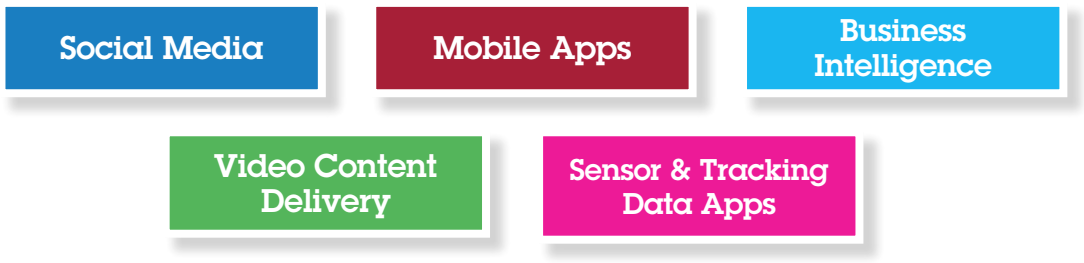
Why is data growing so fast?

Although every organization has its own special causes for massive data increases, significant data growth can generally be tracked to the following causes.

New Applications require more storage

Data growth follows application and device growth. Storage needs were modest prior to the year 2000, when organizations generally only had four types of data to worry about: production and financial data (ERP, MRP, and financial systems); network file server data; email; and Web site data. Backups and disaster recovery plans were simple. Full system backups fit on a few low-capacity tapes, and most disaster recovery plans consisted of a bare metal restore from a backup tape. There were few options for high availability switchovers or continuous backup and recovery software.

Fueled by new devices such as smartphones, tablets, and wearables, organizations are installing many new applications that exponentially use more data than traditional applications. In 2017, a typical organization will process massive amounts of data for applications that didn't exist at the turn of the century, including:



Data growth will only continue to increase as we move further into the next wave of Internet usage. where data retention, analytics, and new devices will require more storage to record, manage, track, and report on organizational data.

Organic business growth & growth through acquisitions

Organizational growth also causes data sets to rapidly double or triple. Acquisitions, mergers, and partnerships frequently require more storage, as consolidated data is shifted, duplicated, and converted from acquired companies to the acquiring organization's equipment. Companies are also gathering more information on customer buying and usage habits, histories, and preferences. Company growth creates more data and more data creates company growth.

The Data Multiplier effect

By itself, organizational data creates a data multiplier effect, where stored information grows just by existing, even without adding new devices, applications, or business.

Production data sets and applications are frequently duplicated for backup, high availability, and disaster recovery using the following technologies:

- **Disk based backup at geographically remote locations**
- **Clustering technologies**
- **Near-continuous data protection and reovery services**
- **High availability and disaster recovery software**

Programming needs also contribute to data growth. Production data sets are frequently duplicated or triplicated for development, testing, quality control, and integration with other systems.

Data multiplier effects also show up in IBM i logical files, SQL views, queries, or business intelligence applications, where data is sorted, selected, massaged, and recombined into different configurations for processing and analysis.

Regulatory requirements causing expanded data retention

Industry and government regulations such as the Health Insurance and Accountability Act of 1996 (HIPAA), Sarbanes-Oxley (SOX), and the Payment Card Industry Data Security Standard (PCI DSS) are requiring ever expanding amounts of historical data to be retained for longer periods of time. Regulations and auditors routinely require companies to store and analyze more metadata (data about data) than ever before, including log files, audit journals, access histories, application changes, database change detail, and other data that serves auditing rather than business needs. System log files are also exported to centralized syslog servers for Security Information and Event Management (SIEM) analysis and reporting, which duplicates event and log files.

Unstructured Data

Unstructured data has exploded. Content on demand is sold as its own product. Video is recorded from multiple cameras in all areas of a customer's premises. Video and images are used for marketing products, historical documentation, product demonstration, customer service, corporate presentations such as Webinars, and much more. And the continued development of big data applications adds exponentially more data to an organization's storage.

How data growth affects an organization

Data growth has a cost for storing, handling, backing up, and safeguarding higher amounts of data each year on both old and new systems. Typical costs of data growth include:

Increased storage costs on internal equipment

As older disk units and Storage Area Networks (SANs) remain active, they require continued support for aging hardware, including annual hardware maintenance costs for old disk drives and SANs. Maintenance costs on older equipment increase every year. Drives and parts become more difficult to obtain as manufacturers reduce and eliminate support for older equipment. Older disk drives are also inefficient. They are slower and contain less capacity than more recent drives.

Data redundancy also has a cost. Most data arrays contain unused disk that is dedicated to redundancy, including additional drives for RAID sets, mirrored disk, and hot spare drives that do nothing more than wait for a live drive to fail before being forced into usage.

Increasing data administration costs

Keeping older storage active while adding modern storage systems such as SANs and object storage systems increases the need for administrative tasks. Administering old storage requires personnel with different skill sets for managing a variety of SANs, disk arrays, and conventional hard drives.

Keeping older disk solutions and servers increases internal data center costs, where servers, disk arrays, and older SANs are physically located in an internal or external data center, consuming precious rack space and driving up other data center costs such as rent and electricity.

Increasing complexity and risk

Managing different types of data and storage systems results in increased complexity, where applications and users must have access to storage on several different arrays and SANs spread out through the corporate network. Users seldom access data belonging to just one storage unit.

Increasing processing time and processing windows

Batch processing time increases as data sets grow larger and batch programs process more data. Larger files also increase the size of data maintenance windows, as purge and reorganization programs and file refreshes reorganize larger amounts of records.

Application response time increases as files gets larger. This can be mitigated somewhat by data archiving and the strategic use of SQL views or IBM i logical files, but there will still be processing time needed to maintain and update those views.

Service level impacts

As data growth affects system processing, it can become more difficult to meet Service Level Agreements (SLAs) regarding uptime and response time. Additional processing resources may be needed to keep up with SLA obligations.

Lack of scalability

Large growth in siloed data on specialized file and application servers may obsolete storage infrastructures quicker. The maximum hard drive capabilities in arrays may not be enough to handle storage requirements and older arrays often need replacement. Massive data growth makes it more difficult to scale for future growth.

Impact on disaster recovery and high availability initiatives

Disaster recovery and high availability planning becomes more difficult as data increases. With multiple Windows, IBM i, Linux, and UNIX systems, organizations must have backup, HA, and DR plans for each critical application and its data. The data multiplier effect increases an organization's backup, disaster recovery and high availability storage costs.

RAID vulnerabilities

After replacing a failed RAID drive, the system must rebuild drive data to reestablish the complete RAID set it belongs to. It can take several hours or a day or longer to rebuild a failed RAID drive, depending on drive capacity, speed, and the amount of data stored on the drive. Rebuild time leaves a system vulnerable to multi-drive failures where two or more drive failures can completely destroy an entire RAID set.

Hitting the limits of traditional file systems

Rapid data growth is creating problems for traditional file systems, with their emphasis on blocks and files stored in hierarchical folders. New applications require storage for large amounts of non-hierarchical data including video and specialized images. Some of the issues facing traditional file systems include:

- **Additional folders and sub-folders must be created for data organization, creating more complex hierarchies**
- **Searches take longer over larger folder structures with complex hierarchies, affecting performance**

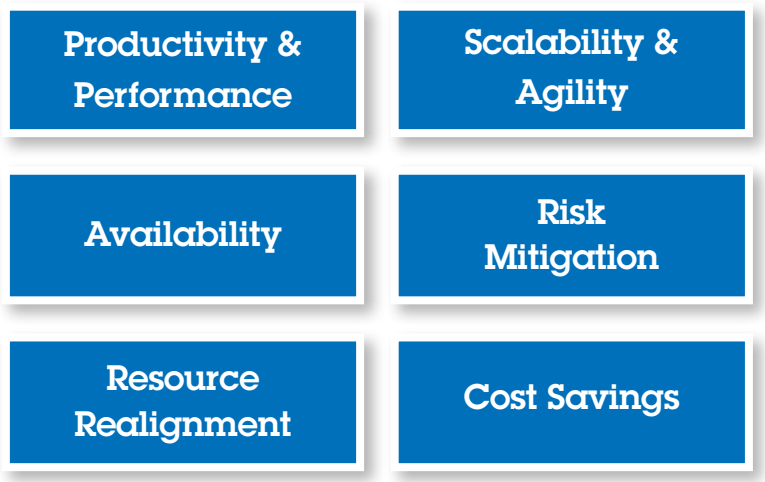
- Most traditional file systems have limits to the number of directories and files they can store or search
- It's difficult to store and easily find earlier versions of files and objects, as traditional file systems only store one copy of a file with the same name

With requirements to quickly index, search, and deliver large amounts of unstructured data in a non-hierarchical manner, many organizations will soon hit the limits of what their traditional file systems can accomplish.

The benefits of effectively managing data growth through cloud migration

While data growth may be an IT reality, it can be managed to decrease costs and increase benefits. Many of the problems listed above can be solved by migrating existing servers and storage to a cloud-based environment. Figure 1 shows a list of benefits organizations can realize by migrating servers and data from internally maintained data centers to the cloud-based data growth strategies listed in this white paper.

Here's how each of the benefits listed in figure 1 can be realized when moving internally managed servers and data into a cloud environment.



Productivity and Performance

- Most cloud vendors use faster and more modern storage, which allows for:

- Faster response times over using older storage technology
- Faster batch processing by using faster storage solutions to process orders and service customers
- Decreased production maintenance windows, creating increased up time for critical systems. Storage-based maintenance completes in a shorter time.

- Faster search times and storage structures that are more compatible with unstructured data, provide customers faster access to filtered information
- Cloud providers offer many reporting capabilities that make it easier to monitor and maintain SLA requirements for valued customers.

Scalability and Agility

• When additional storage is needed, customers notify their cloud provider who can quickly scale up storage to meet expanded needs.

• Expedited access to increased storage allows customers to move quickly when new projects need additional data storage

• Cloud-based storage provides a more flexible storage infrastructure for future growth

Availability

• Storage is always available through cloud-based IT recovery strategies, such as cloud-based backup, disaster recovery, high availability, and hardware replication

• Data archiving is available in the cloud. In many cloud storage implementations, archiving is part of the storage architecture. Cloud storage solutions such as object storage provide different options for handling frequently accessed data, long lived but less frequently accessed data, and long-term archived data.

Risk Mitigation

• Cloud storage mitigates disaster recovery risk with cloud-based IT recovery services, including cloud backup, disaster recovery, high availability, and hardware replication services. Rather than providing these services on-site in their own data center, organizations can shift the risks for disaster recovery to the cloud provider.

Resource Realignment

• Cloud migration shifts many server and infrastructure management tasks from the organization's IT department to the cloud vendor. The cloud vendor designs, plans, purchases, and maintains the server and storage infrastructure, leaving organizational IT resources more time to focus on revenue and bottom line-enhancing activities rather than infrastructure management.

Cost Savings

• Cost of storage and associated servers becomes part of an organization's operating budget, which doesn't require management approval to implement. Storage becomes an operating expense rather than a capital expense.

• Data storage is now managed by the cloud service provider, cutting down on storage administration costs.

• Customers no longer have to bear the expense of housing multiple servers on-site or at an external data center. Data center costs become part of a cloud system contract.

• Customers no longer have to purchase and maintain redundant storage for raid sets, mirrored storage, and hot spare disk drives. Customers only contract for the storage they need.

Strategies for efficiently managing data growth

There are several strategies that help organizations reap the benefits of managing data growth. The strategies listed here solve massive data growth problems by migrating internal servers and data to the cloud, where organizations can enjoy the benefits listed in figure 1. Each strategy can be employed individually or in tandem with other strategies for increased benefits, efficiencies, and reduced costs.

Cloud-based data archiving

Strategy benefits: Improved availability, costs savings, availability

Data archiving systems use business rules to identify and move inactive data to separate storage devices for long term retention and retrieval. Archived data is indexed and has search capabilities for ready retrieval, but it is segregated to avoid many of the performance issues identified above. Data archiving provides performance benefits while meeting data retention requirements and the need to keep old data for business reasons.

By archiving inactive data to the cloud, IT departments can retain inactive data and make it available to their customers anywhere, without maintaining individual servers and data storage arrays in their corporate data centers. As listed in the Object Storage strategy, many cloud storage systems provide data archiving as part of their architecture.

Cloud-based Storage consolidation

Strategy benefits: Improved availability, costs savings, availability

In siloed corporate networks, each file and application server has its own storage media or disk array. Much of the on-board storage is never used, as siloed disk drives are set aside for RAID sets, mirrored drives, or hot spares, and the disk utilization percentage for siloed media may only approach 40-80%.

When moving to the cloud, storage from several different network servers can be consolidated, managed, and maintained by the cloud service provider, not by an organization's IT staff. By its nature, cloud storage is all about storage consolidation and all cloud storage is consolidated storage.

Storage consolidation increases performance while reducing administrative complexity and costs. It centralizes storage resources and administration, replacing in-house data silos with enterprise-level cloud-based storage that has faster access times than the replaced units.

Cloud-based consolidated storage can host storage for many file and application servers, providing centralized storage for all users and applications at a higher utilization rate than using organization-owned siloed servers. Cloud storage providers can provide hot failover capability for servers in the event of a storage failure, planned maintenance, or upgrade, allowing for storage continuity. Using a product like IBM PowerHA, companies can easily switch storage inside the vendor's cloud environment, allowing systems to continue processing data during a storage hardware shutdown.

Administration is more efficient and less costly in a consolidated storage environment. Rather than administering several different arrays attached to many different servers, each server or user can access their own centralized storage space in the cloud. Consolidated storage allows organizations to significantly reduce the number and types of storage media they use, as well as the skill sets needed to manage storage.

Object Storage

Strategy benefits: Improved availability, costs savings, availability

Object storage treats all forms of files, unstructured data, and system data and resources as a collection of distributed objects belonging to a single virtual pool. Each object contains its actual data (payload), along with meta-data about the object.

Because all data and resources are treated as objects, object storage systems overcome the limits of the "blocks and files" traditional file systems described above, providing improved system performance and administrative benefits. Object storage flattens out directory structures, storing object information in a flat namespace that can scale upwards to trillions of objects stored using hundreds of petabytes in capacity. This eliminates hierarchical storage limitations and allows for faster search, retrieval, and access to all objects and data. Object storage is more efficient when accessing data.

Many cloud storage providers including Amazon Simple Storage Service (S3) and Microsoft Azure Storage, use object storage architecture, and it is the dominant storage model in the cloud industry. Some vendors also use free and open source object storage alternatives such as OpenStack Swift and Ceph for cloud storage.

Many object storage providers incorporate data archiving into their architecture, offering customer options for accessing frequently accessed data, long lived but less frequently accessed data, and long-term archived data.

Cloud-based IT Recovery Services

Strategy benefits: Improved availability, costs savings, availability

Many vendors offer cloud-based solutions for IT recovery services such as cloud-based backups, disaster recovery, high availability, and hardware replication services.

Rather than backing up data to tape, cloud-based backup services use disk to reduce backup and restore time and increase production system up-time. They do this by creating synthetic backups where an incremental backup is combined with previous backups to synthetically create a full backup for faster restores. Because they rely on incremental backups, synthetic backups are faster than traditional tape backups; they save money by eliminating backup tape purchasing, handling, and storage; and they provide quicker restores because they can be used to restore data to a specific point-in-time.

Cloud-based disaster recovery uses synthetic backups to restore a system from a specific point in time. Organizations pay for a minimally-sized DR partition or virtual machine (VM), residing in the cloud in waiting mode.

When a disaster is declared, the cloud vendor expands the organization's small standby DR partition/VM to match the size needed for their restored system. An organization only pays for the DR partition or VM resources it needs as it needs them. They don't have to pay for an exact mirrored resource that they will seldom or never use, which results in lower costs than traditional DR.

Organizations moving to a cloud-based high availability system (HA) experience several advantages over running a full server based HA system in a remote data center. Cloud-based HA eliminates hardware and storage costs. It also reduces network costs, licensing fees, and maintenance and administration costs.

Hardware-based replication uses SAN-to-SAN replication in the cloud to create a synchronized copy of a production system on another SAN that resides elsewhere in the cloud. The hardware-based replica is a duplicated standby system for the production system, one that is ready to be powered on and used if a disaster disables production.

For More Information

There are a number of options for efficiently and effectively managing data growth, by migrating an organization's servers and data storage from an in-house data center to a cloud environment.

There are many Secure Infrastructure & Services (SIAS) business partners who can help organizations use new cloud hosting and storage technologies to improve and reduce the cost of their existing storage solution, or to create a new solution that is appropriate to their company, no matter what the company's size and budget.

If you have any questions about improving your data management setup, please visit SIAS on our business partner Web site or call SIAS at xxx-xxx-xxxx for the name of a business partner who can help you analyze your particular situation and create a customized data storage management plan for your organization. An SIAS business partner can help you implement a solid strategy that allows your data to easily grow now and in the future.

About Secure Infrastructure & Services

Secure Infrastructure and Services is a cloud service provider specializing in IBM Power Systems, x86 servers, and all other major environments. Through our Infrastructure-as-a-Service (IaaS) model, we provide exceptional levels of secure, reliable and cost-effective computing resources for small and mid-sized businesses.

Visit us at: siasmsp.com

SECURE Infrastructure & Services

Secure Infrastructure & Services
48 South Service Road, Melville, NY 11747
Tel: 855-772-3447 www.siasmsp.com

Business Partner IBM Managed Services Infrastructure Platform