FUTURE-PROOFING STORAGE

Modernizing Infrastructure for Data Growth Across Hybrid, Edge, and Cloud Ecosystems

RESEARCH BY:

Andrew Smith
Research Manager,
Cloud Infrastructure Services, IDC

Natalya Yezhkova
Research Vice President,
Infrastructure Systems, Platforms and Technologies Group, IDC

Ashish Nadkarni
Group Vice President,
Infrastructure Systems, Platforms and Technologies Group, IDC

March 2021 | IDC Doc. #US47565321

Navigating this White Paper

Click on titles or page numbers to navigate to each section.

Executive Summary 3

Situation Overview 5

- Data Visibility and Insight Are the Foundations of a Data-Driven Business 5
- Modern Workloads Rely on Digital Infrastructure Capabilities to Manage Massive Data Sets 6
- Management Inefficiencies Inhibit Enterprise Visibility and Insight into Data 8
- Addressing Complex Challenges of Enterprise Data Management 10
- Consider Your Organization at the Beginning of Its Data Management Journey 14

Essential Guidance for IT Buyers and Decision Makers 18

Conclusion 20

About the Analysts 22

Message from the Sponsor 23

Executive Summary

Enterprise infrastructure ecosystems, and the data sets and workloads they enable, have become more complex and more critical to today's data-driven businesses than ever before. In large part this is due to the unlimited potential and perceived value of enterprise data. In many cases, the expectations of generating value from data do not align with the reality that enterprises operate on fixed budgets, have limited resources, and rely on suboptimal infrastructure systems and tools.

This analysis explores this disconnect and the challenges many enterprises face in creating and implementing a strategy to manage massive, distributed data sets and the infrastructure requirements of critical workloads based on these data sets. We offer recommendations to buyers in terms of effective planning, implementation, and purchase decisions to help future-proof their enterprise data management strategy based on the key assumption that their storage environments are becoming increasingly complex and distributed.

This IDC white paper presents empirical data gathered from two Seagate Technology-commissioned surveys: the 2021 Cloud Data Storage & Infrastructure Trends survey and the 2021 Storage Systems & Infrastructure Trends survey.

Future-proof your enterprise data management strategy based on the assumption that storage environments are becoming increasingly complex and distributed.

Both surveys were commissioned by Seagate Technology and conducted in the first quarter of 2021:

- The **2021 Cloud Data Storage & Infrastructure Trends** survey, commissioned by Seagate Technology, focused on public cloud infrastructure services and data transfer. The survey included 1,050 United States–based respondents in organizations with 1,000 or more employees. All respondents belong to organizations managing massive data sets, which are defined as sets containing 2 petabytes (PB) or more. Respondents are a mix of IT and storage managers, administrators, and executives responsible for managing their organizations' data storage infrastructure and/or cloud services. Respondents span a range of industry verticals, including financial services, manufacturing, government, retail, and healthcare.
- The **2021 Storage Systems & Infrastructure Trends** survey, commissioned by Seagate Technology, focused on traditional storage and infrastructure systems purchasing, deployment, and operation. The survey included 400 United States–based respondents in organizations with 1,000 or more employees. All respondents belong to organizations managing massive data sets, which are defined as sets containing 2PB or more. Respondents are a mix of IT and storage managers, administrators, and executives responsible for purchasing and/or managing data storage infrastructure and/or cloud services for their organizations. Respondents span a range of industry verticals, including financial services, manufacturing, government, retail, and healthcare.

This survey data provides unique insights into the challenges associated with infrastructure and storage implementation for massive data sets and associated workloads. It also provides insight into challenges associated with data management, migration, and movement across internally managed datacenter, cloud, edge, and endpoint environments. From a subjective standpoint, this white paper analyzes the value that data visibility and insight offer to modern, data-driven businesses and the importance of effective data management in order to generate business opportunity.

Situation Overview

Data Visibility and Insight Are the Foundations of a Data-Driven Business

By 2022, IDC predicts that 65% of global GDP will be digitized, driving $6.8 trillion of direct digital transformation (DX) investments for 2020–2023. Furthermore, by 2022, 46% of an enterprise's products and services will be digital or digitally delivered, increasing businesses' reliance on infrastructure resources (compute, storage, and networking) to support more than just the traditional business applications.

Timely access to infrastructure resources — both shared and dedicated — will be imperative to sustaining the scalable, secure, and compliant data-driven business models of the future. This means enterprises will increasingly rely on access to resilient and trusted infrastructure at the physical and logical levels, regardless of operational location (e.g., internally managed datacenters, cloud, or edge).

We recommend that enterprises consider the following details when making infrastructure purchases and implementations:

- Digital infrastructure is no longer limited to traditional central enterprise services nor to discrete cloud datacenters. Digital infrastructure includes all assets and resources that enable the movement of workloads, applications, and/or code. Digital infrastructure will be foundational for delivering modern, satisfactory customer experiences.
- Digital infrastructure should embed intelligence/automation into business operations and should support ongoing innovation out to the digital edge of the enterprise. Successful digital strategy should transform enterprise infrastructure to eliminate data and application silos, break down technology-imposed barriers, and extend support beyond traditional or existing IT tools and applications.

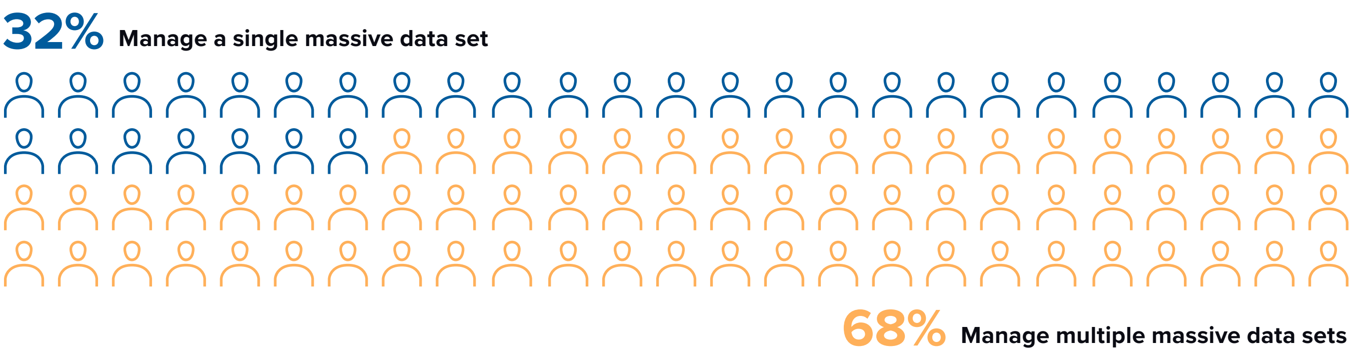
Modern Workloads Rely on Digital Infrastructure Capabilities to Manage Massive Data Sets

More than half of enterprises surveyed are managing massive data sets (2PB or more) that serve multiple critical workloads. The top three massive data sets managed by enterprises today include data warehouses, transactional data, and digital content repositories (see Figure 1).

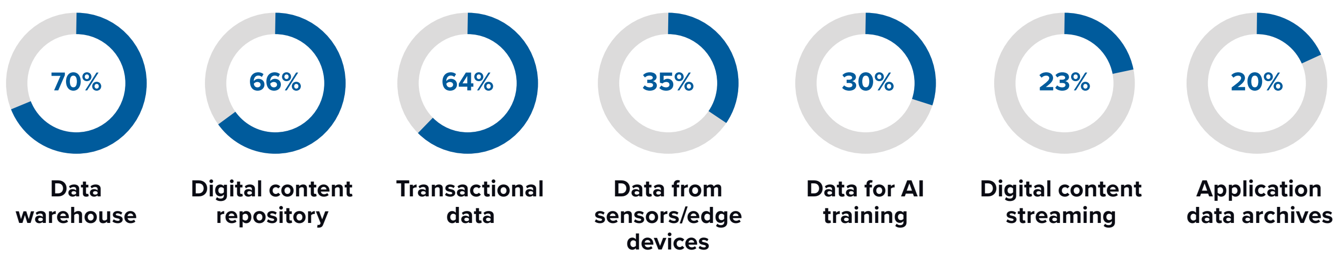
FIGURE 1

Over Half of Enterprises Surveyed Manage Multiple Massive Data Sets (Larger than 2PB in Volume), Which Enable a Range of Workloads

Q. Does your organization have massive data set(s) (2PB or more) associated with a single workload (e.g., AI training or media streaming) or serving multiple workloads?



Q. Which of the following describes massive data set(s) that your organization manages?



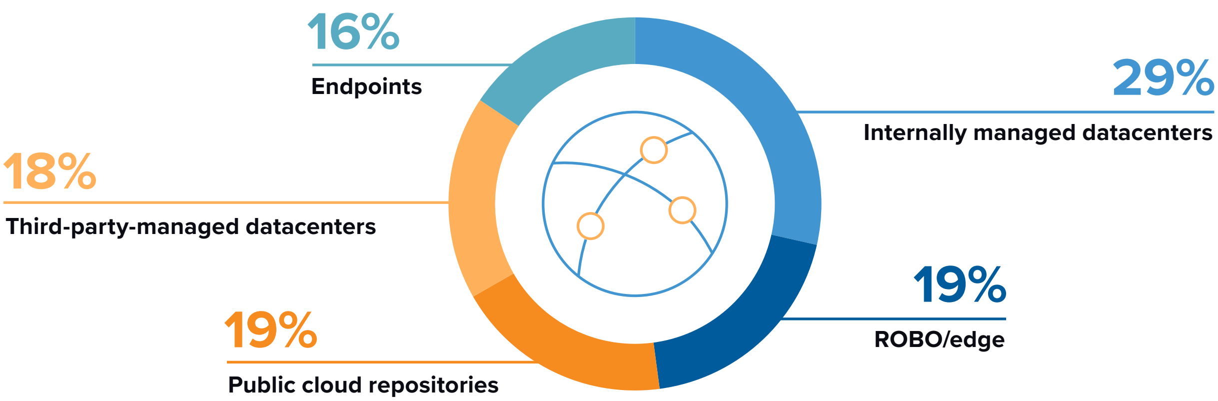
Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

These massive data sets are not static: 99% of respondents indicated that they move data among different storage locations, utilizing network or physical data transfer. As a result, enterprise data is distributed across multiple storage locations, with approximately 29% residing in internally managed datacenters, 19% in remote office/branch office (ROBO)/edge locations, 19% in public cloud repositories, 18% in third-party-managed datacenters, and 16% on endpoints (see Figure 2).

FIGURE 2

Enterprise Data Is Distributed Across Multiple Storage Locations

Q. Approximately what percentage of your organization's enterprise data is stored in the following locations?



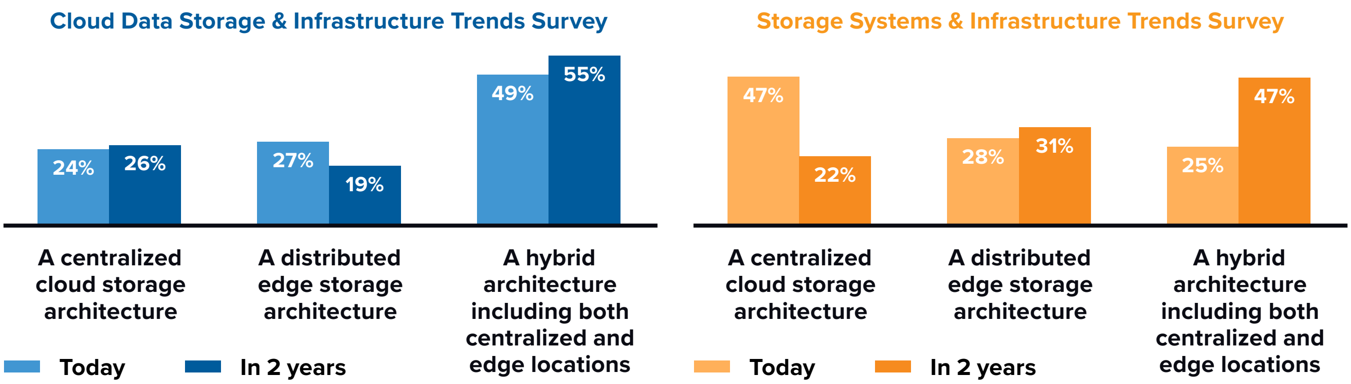
Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

This distribution will vary by workload, and is also expected to change over time. Enterprises indicated that in two years, they intend to store and manage more of their massive data sets in hybrid architectures (a combination of centralized and edge locations); see Figure 3.

FIGURE 3

Both Surveys Show Preference for Storing and Managing Massive Data Sets in a Hybrid Architecture

Q. Which option best describes your organization's approach to storing and managing massive data sets, today and in two years?



Source:
IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021; IDC Storage Systems & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021

This analysis will further explore this shift in the way storage for massive data sets is moved and managed, including the challenges many enterprises face with their storage infrastructure from technical, operational, security, and cost standpoints.

Management Inefficiencies Inhibit Enterprise Visibility and Insight into Data

Enterprises are gathering more data than ever before in order to deliver on the promises of digital transformation and fuel novel AI- and ML-based solutions and business initiatives. That simply means enterprises need to store more data, and they need to know exactly where the data must be placed (e.g., internally managed datacenters, edge locations, public cloud repositories) in order to meet requirements for data access, security, and cost.

IDC's Enterprise Storage Systems Installed Base Forecast provides a snapshot of how IDC expects storage-installed base capacity to expand over the next five years by location:

- **Public cloud:** The installed base in the public cloud is estimated to grow at a five-year compound annual growth rate (CAGR) of 35.9% and reach 2,872 exabytes (EB). Note there are 1,000PB in an exabyte.
- **Private cloud:** The installed base in the private cloud is estimated to grow at a five-year CAGR of 31.6% and reach 392EB.
- **Traditional IT:** The installed base in traditional IT (non-cloud) environments is estimated to grow at a five-year CAGR of 8% and reach 620EB.

All three of these installed bases are huge in volume. As a result, we expect enterprises to manage more massive data sets than ever before, spanning across internally managed datacenters, edge, and cloud.

Consequently, CEOs, CTOs, CIOs, and other enterprise decision makers will be challenged to address the following data storage issues:

- **Pace of growth of massive data sets:** On average, enterprises indicated to IDC that they expect their stored data to grow 30% annually. Using this growth rate as a general guide, we can assume that an organization managing 50PB of data today will store upwards of 65PB of data the following year. The challenge with this data growth trajectory is that spending on IT infrastructure is only expected to grow in the single digits (and may even remain flat for many enterprises). Furthermore, many enterprises are already dealing with inflated spending on storage infrastructure. Of respondents who serve massive data sets on premises, 86% indicated that their spend on raw storage capacity exceeds spending on storage management software and support contracts. Almost 20% of respondents estimated that the annual total cost of ownership of their storage solutions/services is over $1 million. This juxtaposition — data growth significantly outpacing IT infrastructure spend —

puts many organizations in a precarious position: How can enterprises effectively store and protect growing volumes of data without adequate resources? They will be challenged to strike a balance between the promise of generating value from massive data sets and the reality of running their businesses efficiently.

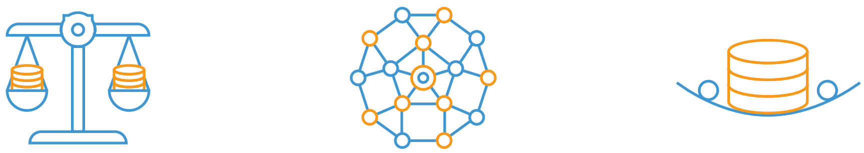
- **Finding and hiring the right talent:** Our survey results indicate that finding and hiring the right talent is one of the top two challenges enterprises face when it comes to their data storage in both public cloud and operational (ROBO/edge) locations. As enterprises expand their storage footprints to new locations, workloads, and applications, they will continue to face people skills and training shortages.
- **Data management complexity:** Our survey results show that storage management complexity is the number 1 challenge facing enterprises when it comes to their data storage, both on premises and in operational (ROBO/edge) locations. In on-premises environments, enterprises will be tasked with simplifying their storage environments and eliminating data silos. In edge locations, enterprises must ensure they are not creating additional data silos or management overlap/inefficiency as their data estate sprawls across environments.
- **Operationalizing advanced analytics use cases:** Our survey results show that enterprises are challenged to overcome a lack of analytics capabilities, specifically for their storage in operational (ROBO/edge) locations. Enterprises with massive data sets require access to analytics capabilities in order to operationalize workflows that leverage deep learning and training of algorithms, for example. If organizations don't establish an infrastructure with the means to connect or move massive data sets to the related analytics platforms and tools needed, they simply will not be able to operationalize use cases for deep learning, machine learning, artificial intelligence, etc.
- **Data security, compliance, and sovereignty:** Security and compliance are perennial data management challenges. Our survey data indicates that when enterprises evaluate alternative storage solutions/services, their number 1 consideration is finding an offering with improved security and compliance capabilities. And local data security requirements are just one part of the equation; increasingly, multinational organizations must also plan for security and compliance regulations unique to the data sovereignty requirements of the individual nation-states or countries in which they operate. As data moves across borders, enterprises must manage multiple dimensions of security, compliance, and user regulations. This complexity creates uncertainty and is a major reason that 80% of enterprises surveyed said they are concerned with their ability to comply with existing privacy laws (e.g., CCPA, GDPR).

These storage and data management complexities may seem overwhelming. The intent of enumerating such a list isn't meant to demoralize, but rather to provide context for enterprises regarding the current and future challenges associated with data storage environments. The following sections of this analysis attempt to compartmentalize these challenges and provide recommendations backed by survey data. The idea is to help enterprises address and preempt challenges associated with managing infrastructure and storage for massive data sets.

Addressing Complex Challenges of Enterprise Data Management

Our survey results indicate that limited performance, complexity of management, and lack of integration with cloud are three of the leading technical shortcomings of current storage infrastructure solutions. But these broader challenges are likely just the “tip of the iceberg,” underpinned by a host of additional variables that may be people-, process-, or technology-related. These challenges can be generally assigned to one of three principles of data management, which this analysis will define in order to provide a better understanding of why modern data management initiatives can be so complex.

The three principles of modern enterprise data management are:



1. **Data and its workloads will constantly move in search of equilibrium.**
2. **Data management ecosystems must be resilient and extensible.**
3. **Data has gravity.**

Each of these principles is explored in depth below.

1 Data and its workloads will constantly move in search of equilibrium.

Workloads and their associated data should be expected to migrate among infrastructure locations (e.g., internally managed datacenter, edge, cloud) without residing in a single location permanently. Stored data associated with many workloads can fluctuate between deployment locations until it reaches a state of equilibrium, which in this case is defined as an optimized balance between cost and performance.

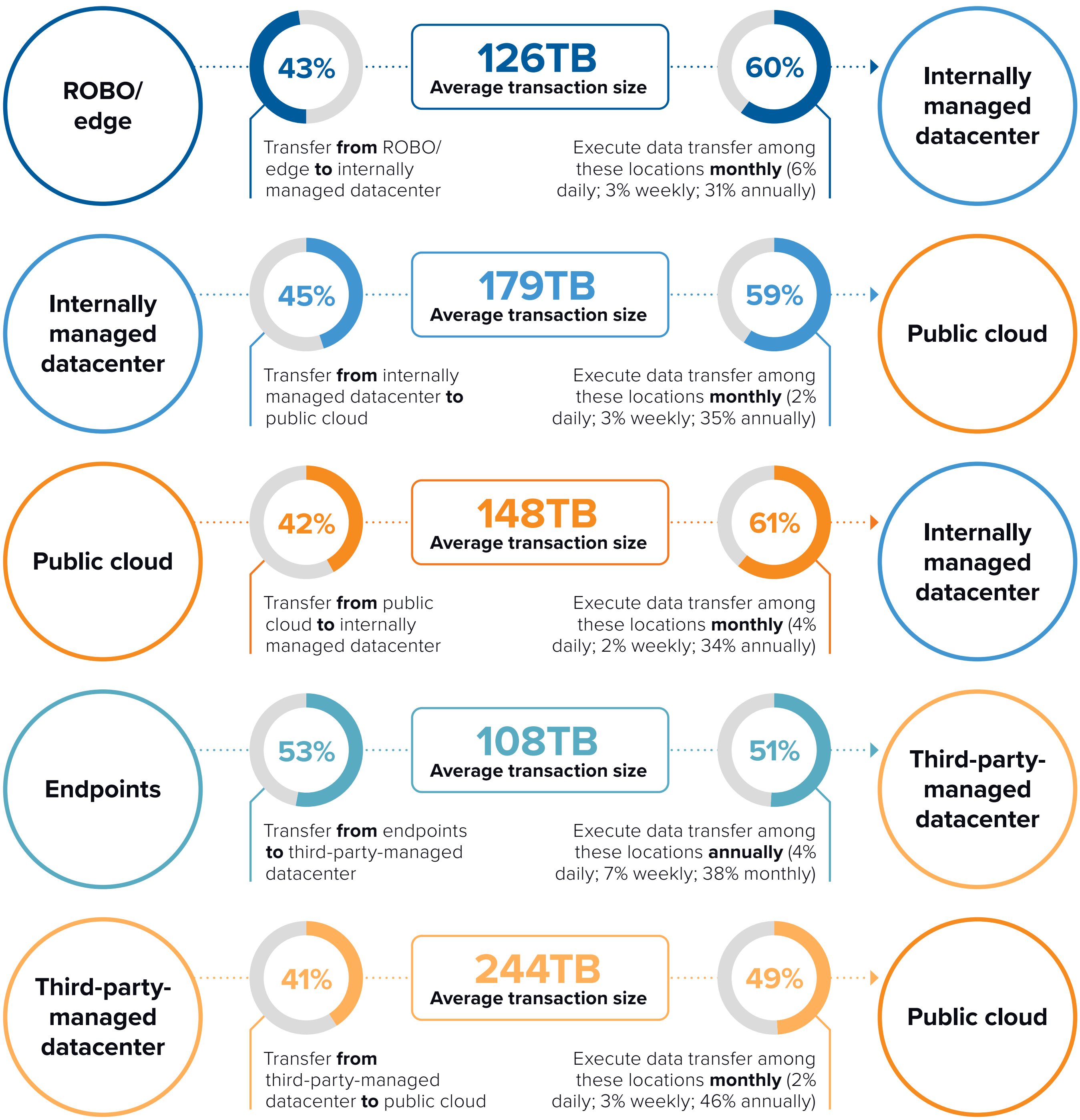
Our survey findings show that 31% of respondents think they are paying too much for storage infrastructure currently used for massive data sets. In many cases, spending inefficiency will result in data movement in the name of cost or performance requirements. As this balance between cost and performance changes over time, enterprises should expect their data and workloads to also change in search of new equilibrium.

This movement of data is further illustrated by our survey results: Almost all respondents indicated that they transfer data among locations (e.g., from core datacenters to public cloud). On average, the amount of data transferred in a single transaction is over 140TB. In terms of frequency of transfer, the majority of respondents indicated that they transfer data on a monthly basis among locations.

The only exception in data transfer frequency was from endpoints to third-party-managed datacenters: 51% of respondents execute this transfer annually. Figure 4 illustrates the scope of data transfer and frequency among select storage locations.

FIGURE 4

Enterprises Are Regularly Transferring Large Sets of Stored Data Among Storage Locations



Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

This high prevalence of data movement across locations reinforces the fact that applications and workloads seldom remain static. It also highlights that in many cases, moving massive data sets requires alternative solutions to traditional wired or wireless networks: 78% of survey respondents indicated that they use physical data transport/ migration solutions because data transfer over their networks could no longer keep up with capacity. In terms of physical data transport capacity, on average, the total capacity associated with their organization's physical data transfer services is 473TB.

2 Data management ecosystems must be resilient and extensible.

Enterprises should look for storage infrastructure and management platforms that enable agile and consistent scaling of resources (in this case, storage and compute), utilizing on-demand, workload-specific components. Such platforms will help enterprises ensure they are using their resources effectively and, in the case of a services offering, paying only for the resources they actually use.

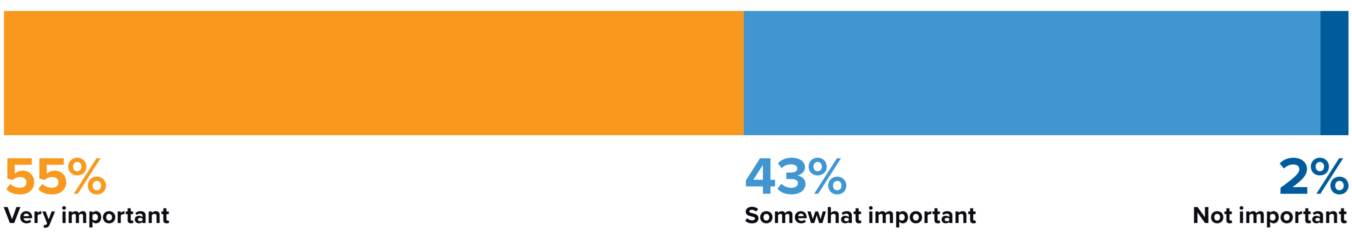
The concept of extensibility also goes beyond resource scaling to include the adjacent applications and services that enterprises can leverage. This is most prominent in public cloud storage deployments, where over half of respondents (55%) indicated that having access to native cloud services for their stored data is "very important" (on a scale that also included "somewhat important" and "not important"); see Figure 5.

FIGURE 5

Application Access Is a Key Element of Data Management, and Highly Important to Enterprises

Q. How important is it that your organization's stored data has access to native public cloud services?

Legend:
(% of respondents that indicated importance from very to not important when evaluating / considering migrating data to cloud.)



Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

An extensible data management platform that offers access to adjacent services gives enterprises the flexibility to move data to where workloads reside (e.g., migrate data to a public cloud database) or to move applications to where data is stored (e.g., moving compute and analytics tools to an on-premises archive), in a "best of both worlds" scenario.

As mentioned previously, data security and compliance are ongoing challenges for all enterprises, regardless of their infrastructure model. The more important fact is that security remains a major catalyst for data movement. Survey respondents indicated that security is one of the leading reasons data is moved out of public cloud environments and repatriated back on premises. With increasingly complex data management requirements come increasingly stringent data security and compliance needs to ensure appropriate access, visibility, and control of sensitive data.

A resilient and extensive data management ecosystem should be designed to accommodate data movement for security and compliance reasons while also providing the technical capabilities needed to protect, replicate, and restore data in the event of a malicious attack or recover data in the event of data loss due to user error or unplanned disasters. These elements will help minimize the need for data movement across locations due to gaps in security/compliance/data protection.

3 Data has gravity.

As storage associated with massive data sets (defined as 2PB+ in volume for the purposes of our surveys) continues to grow, so will its gravitational force on other elements within the IT universe. This is the key concept behind data gravity, inspired by how Earth's gravity causes objects with larger mass to draw smaller objects toward their center. Imagine an identical concept applied to an enterprise's stored data, where workloads with the largest volumes of stored data exhibit the largest mass within their "universe," attracting applications, services, and other infrastructure resources into their orbit.

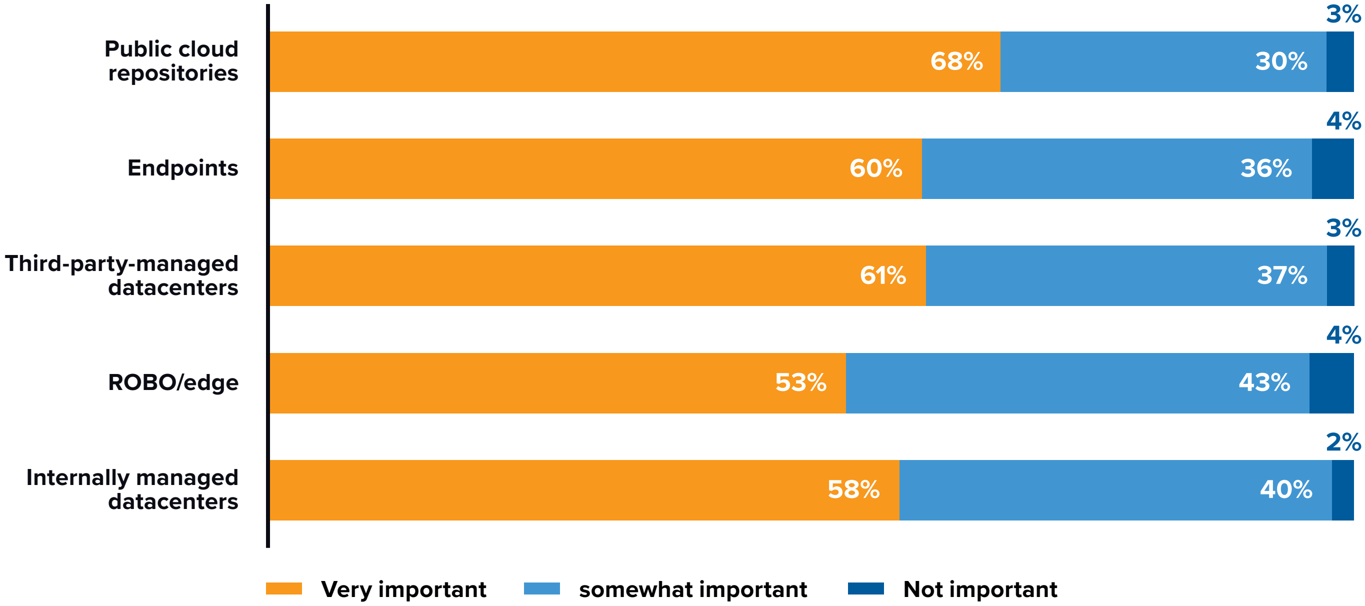
As these massive stored data sets grow, it can become harder to detach applications and services from the data on which they rely. As a result, applications and services end up having to move to the data sets, or remain near the data sets, in order to fully operate. In some cases, these massive data sets risk becoming "black holes," trapping stored data, applications, and services in a single location, unless IT environments are architected to allow the migration and management of stored data, along with the applications and services that rely on it, regardless of operational location.

One way to mitigate the impact of data gravity is to ensure that stored data is collocated adjacent to applications regardless of location. Our survey results indicate that enterprises feel the same way: More than half of survey respondents believe collocation of data with adjacent applications is very important, regardless of the operational location (see Figure 6).

FIGURE 6

Ensuring Applications Have Access to Data, Regardless of Location, Can Help Reduce the Impact of Data Gravity

Q. For enterprise data stored in the following locations, how important is it that this data is collocated adjacent to applications?



Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

Modern enterprise data management strategies should be established with the concept of data gravity in mind and ensure that no single data set exerts uncontrollable force on the rest of the IT and application ecosystem.

Consider Your Organization at the Beginning of Its Data Management Journey

The challenge of data management is that it's a moving target, requiring agility and adaptation on the part of enterprises as they craft strategies that can mature over time and react to changing variables in technology, industry, and macroeconomics.
In other words, enterprise data management is a journey, not a destination. The three principles of modern data management outlined above can help enterprises begin to strategize solutions for their unique data management needs and contextualize what data management means for their organization and their overall data-driven business strategy. IDC advises organizations to consider the recommendations on the following pages as they embark on their data management journey.

Future-proof your data management strategy by preparing for mass data storage across internally managed datacenter, edge, and cloud environments.

The scale of data and its limitless potential to the business have made enterprises averse to vendor lock-in at the infrastructure level as well as adoption of any proprietary storage architectures that cannot be extended to new use cases, applications, or workloads. Despite the risks associated with using inefficient or legacy storage architectures to manage massive data sets, only 45% of enterprises surveyed indicated that the storage systems used for their mass data sets are optimized for mass storage. This suboptimal design can result in enterprises struggling with performance, management complexity, an inability to integrate across storage environments (e.g., edge, cloud), and the need to deliver forklift upgrades (migrating data from one storage platform to another) in order to effectively modernize their environments. Clearly, these systems are not designed for the long term.

In many cases, this suboptimal design also has a financial cost that can further jeopardize the enterprise data management strategy. Infrastructure management complexity and performance constraints often lead to inaccurate storage capacity or usage forecasts as well as billing complexity. This is particularly prevalent in the cloud, where inaccurate planning or unforeseen events can result in charges for data access (e.g., puts, gets, egress) in order to move among environments. Survey data indicates that reducing overall cost is the number 2 reason organizations are moving data out of public cloud (security is number 1; see Figure 7).

FIGURE 7

Security, Cost Reduction, and Reduced Management Complexity Are the Top Three Reasons Organizations Move Data Out of Public Cloud

Q. What are the top drivers influencing your organization to move data out of public cloud?

1

Security

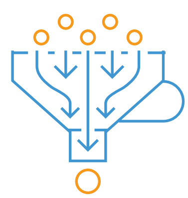
51% of respondents ranked security as their top driver for moving data out of public cloud.

2

Reducing overall costs of public cloud services

36% of respondents ranked reducing overall costs of public cloud services as their second top drivers for moving data out of public cloud.

3



Reducing management complexity

34% of respondents ranked reducing management complexity as their top third drivers for moving data out of public cloud.

Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

Almost all survey respondents (99%) said they incur planned or unplanned egress fees (at least annually). And these fees are material; respondents estimated that on average, egress fees account for 6% of their public cloud storage bill (see Figure 8). It is imperative that enterprises create a data management strategy that accommodates the need for mass storage regardless of environment and provides flexibility and transparency in terms of data access and billing. Without this flexibility and transparency, many data management initiatives will be doomed to fail due to unmet service levels and excessive spending.

FIGURE 8

Cloud Egress Fees Happen Regularly, and Can Add Up

99%
of cloud storage users incur planned or unplanned egress fees

41%
incur egress fees at a frequency of two to three times per year

6%
of an organization's cloud storage bill can be attributed to egress fees, on average

Source: IDC Cloud Data Storage & Infrastructure Trends Survey, sponsored by Seagate Technology, January 2021.

Implement data management initiatives using a workload-centric approach.

Enterprises need a data management strategy to provide flexible, organized access to all data, of every type, regardless of where it lives. Any data management modernization effort should define and deploy an information architecture that provides an open, extensible foundation, with choice and flexibility, capable of communicating with other data services while adhering to the requirements of specific workloads. Below are some examples of workload requirements that may help inform data management needs.

→ **Storage requirements:** Perhaps the most obvious but also most relevant to this analysis are the various storage requirements associated with massive data sets. Identify mixed need for file, block, and object storage protocols. Measure performance and accessibility requirements of individual workloads based on a scale of "hot" (having data that is accessed daily/monthly) to "cold" (having data that is accessed annually). Identify the need for long-term data retention associated with specific workloads, as this may result in unique storage requirements (e.g., immutable object buckets).

- **Compute requirements:** It is often assumed that all modern workloads will be containerized. Far from it: Many workloads will be designed to run on bare metal or virtualized. Many modern workloads are also being optimized to utilize accelerators and/or functional offload coprocessors (also known as SmartNICs). These variables will drastically change the overall compute and storage requirements and level of customization needed on a workload-by-workload basis.
- **Data security and protection requirements:** Data protection and security requirements may vary considerably based on workloads. Ensure that necessary levels of data encryption, policy- and role-based user access to data, auditing, data replication, copy management, backup, and object/file restore exist depending on workload need. For example, an archiving workload may require stringent policies for data immutability, retention, and erasure, whereas a transactional data workload may require more strict protection of personally identifiable information (PII) and extremely fast recovery capabilities (e.g., Recovery Time Objective [RTO]/Recovery Point Objective [RPO]).
- **Deployment location requirements:** Given the distributed nature of massive data sets, organizations should plan for the need to migrate and manage data across multiple environments (e.g., internally managed datacenter, edge, cloud). This may require solutions for data transfer, application migration, load balancing and orchestration, and storage gateways to ensure that data can be moved to the optimal location depending on the workload requirements.
- **Resource scale:** Ensure that your data workload operates efficiently. Some workloads may have little to no variability in terms of storage and compute resources, while others may spike due to usage patterns. These characteristics may require unique operational profiles to ensure uptime, efficient operation, and cost management.
- **Cost profile:** Many workloads have unique cost profiles. From an infrastructure perspective, this may be measured simply by the compute and storage resources needed to support or run the workload. From a business perspective, the cost of a workload might be measured more on its perceived value, its opportunity cost, or the value of the business outcomes the workload supports or enables. These unique views on cost may help provide additional ways to implement data management strategy.

Essential Guidance for IT Buyers and Decision Makers

Many technology buyers are rightly confused about the process of building, buying, and managing their own data management and infrastructure stack designed to accommodate the needs of mass storage architectures. They have to define use cases, launch business initiatives, and hire or train data scientists and application developers — while also avoiding the constraints posed by their legacy infrastructure in terms of performance and cost. In an attempt to move forward, enterprises may find themselves tempted to create temporary solutions that are limited in scope. But these stopgaps typically lead to suboptimal performance and cost over time.

IDC advises IT buyers and decision makers to take a more measured approach to solution deployment wherever possible, with consideration for the following buyer criteria:

- **Start with business outcomes.** Organizations must begin by tying together service constraints and use cases to identify business outcomes that benefit from an investment in their storage infrastructure. They must seek to quantify and measure the benefits of such investments. These business outcomes need not be all-encompassing; a business outcome could simply be improved data management as measured by a reduction in employee hours spent on manual storage tasks (e.g., data provisioning, migration). Proving this outcome could act as a stepping stone for future infrastructure projects and initiatives.
- **Choose the right reference stack.** Several vendors and service providers have put out reference stacks for implementing a data infrastructure. Many of these are open in nature, allowing a modular plug-and-play experience, and can be consumed as a pay-as-you-go service for capex-friendly implementation. This is an important consideration, as data infrastructure investments can get expensive quickly. IT benefits to keep in mind when examining reference stacks are reduced costs, data and application availability, effective infrastructure utilization and consolidation, and, where possible, a single interoperable application delivery platform.

- **Decide whether to build, buy, or employ a mix of both.** The "build versus buy" decision should be based on use cases, in-house skill sets, training, custom (open source or proprietary) versus off-the-shelf culture, and appetite for infrastructure investment (typically in the form of capital expenditure). If an enterprise embarks on a build-heavy approach, ensure that solutions are piloted, proofs of concept are carried out, and the necessary levels of support and training are available from vendors or their partner ecosystems.

If taking a services-led approach, we recommend focusing on three key areas.

First, ensure that the infrastructure provider is committed to a cloud-like customer experience, characterized by extensive automation of storage-related tasks, integrated workflows for ease of use, guaranteed uptime and durability service-level agreements, nondisruptive improvements to service quality/performance over time, integrated tools for data movement and migration, and seamless access to adjacent services within the service provider ecosystem.

Second, ensure the availability of multiple consumption models from the vendor in terms of services delivered as opex as well as capex.

Third, measure the service provider's ability to provide infrastructure services that meet or exceed on-premises performance (measured by latency, accessibility, IOPS, etc.). When specifically asked about storage-related professional services, our respondents indicated that the highest utilized services are those for workforce training, data migration to cloud, consulting for storage-related product/service deployments, and services for multicloud data management.

Finally, remember that this doesn't have to be a binary decision: Enterprises often leverage services for select use cases or applications while also managing their own infrastructure and deployments in other areas.

Conclusion

Modern, effective data management strategies should be designed with mass storage and data movement as key tenets. Not every enterprise manages multiple massive data sets, but many already do. And given the pace of digitization of business and the importance placed on the value of enterprise data and data gathering, many organizations will find themselves managing massive data sets in the near future.

Preparing for this reality will help ensure the long-term success of enterprise data management strategies, which in turn enable the success of digital infrastructure and business initiatives. This is the ideal scenario. As much of the survey data shows, enterprises face a host of challenges.

Despite these challenges, IDC recommends embarking on your data management journey with the following concepts in mind:

- **Don't let a single workload or operational location dictate the movement of storage or data resources.** Remember that data has gravity, and infrastructure environments must be architected to efficiently move storage resources as needed. If not, individual workloads or massive data sets may end up exerting uncontrollable force on storage resources.
- **Enable your workloads to move in search of equilibrium.** Define equilibrium as the ideal balance of cost and performance, and seek flexible, transparent storage solutions/services that can help strike this balance. Prioritize scalability of resource consumption, flexibility of data placement, and automation of data migration/management as key enabling features. Assume that stored data and data sets associated with critical business workloads and applications will always need to move among environments (internally managed datacenter, edge, cloud) in search of this equilibrium, and architect solutions that enable this movement.
- **Adopt storage solutions and services that help your organization do more with less.** Assume that your organization's data growth will significantly outpace your storage budget. The fact that data management solutions should be designed with cost savings in mind is obvious, but how these cost savings can be achieved or the specific features/functions of a service that enable long-term cost savings can be harder to define. Seek to adopt solutions/services that offer a mix of purchase

models (spanning both capex and opex). Look for a solution with a proven track record of price reductions over time in terms of dollar per GB month for stored capacity or reductions in overall subscription costs. Prioritize platforms that can automate or eliminate manual infrastructure tasks such as storage provisioning and placement or long-term resource planning. Look for transparent billing models with limited fee structures (e.g., egress, early deletion, data transfer).

- **Place a premium on data resilience.** Data resilience is an umbrella term encompassing data security, compliance, and data protection features. Security and data protection are some of the most challenging aspects of data management — and some of the most influential features that enterprises consider when making a purchase decision for solutions/services. Ensure that data resilience requirements are clearly defined for your digital infrastructure and workload requirements, and outline acceptable tolerances for uncertainty regarding malicious attacks and unplanned disasters, as these occurrences cannot be avoided.

An effective data management strategy will help organizations adapt to the challenges associated with exponential data growth and digital business transformation. These challenges are persistent and complex; there will always be new data types, security threats, regulatory compliance laws, and applications to accommodate. But IDC believes that the survey results and concepts outlined within this paper can be leveraged by enterprises to design a data management strategy that is modern, extensible, and flexible enough to be "future-proof" in the face of known and unknown data demands.

About the Analysts

Andrew Smith
Research Manager,
Cloud Infrastructure Services, IDC

Andrew Smith is a research manager within IDC's Enterprise Infrastructure Practice. Andrew's research focuses on public cloud infrastructure-as-a-service platforms and solutions, with specific focus on storage services. Andrew contributes to market sizing and forecast efforts across IDC's Public Cloud IaaS segments, as well as adjacent markets like multicloud data management, data protection as a service, and public cloud cold storage.

More about Andrew Smith

Natalya Yezhkova
Research Vice President,
Infrastructure Systems, Platforms and Technologies Group, IDC

Natalya Yezhkova is a research vice president within IDC's Enterprise Infrastructure Practice. Natalya's research coverage includes derivative data products, forecasting of IT infrastructure markets, custom data modeling, and infrastructure buyer research. On the derivative data product side, Natalya manages IDC's Cloud IT Infrastructure Tracker, Server and Storage Workloads Trackers, Infrastructure Buyer Tracker, and the infrastructure portion of the Cognitive/AI Tracker. Natalya also manages Enterprise Infrastructure Market Pulse, quarterly primary research focused on end-user purchasing plans for infrastructure products and adoption of technologies shaping the infrastructure market.

More about Natalya Yezhkova

Ashish Nadkarni
Group Vice President,
Infrastructure Systems, Platforms and Technologies Group, IDC

Ashish Nadkarni is group vice president within IDC's Worldwide Infrastructure Practice. He leads a team of analysts who engage in delivering qualitative and quantitative research on computing, storage, and data management infrastructure platforms and technologies, via syndicated research programs (subscription services), data products (IDC Trackers), and custom engagements. Ashish's vision for his team is to take a holistic, forwarding-looking, and long-term view on emerging as well as established infrastructure-related areas in the datacenter, in the cloud and at the edge.

More about Ashish Nadkarni

Message from the Sponsor

Introducing Lyve: The Edge-to-Cloud Mass Storage Platform, by Seagate Technology.

Lyve is the Edge-to-Cloud Mass Storage Platform, a comprehensive suite of services, hardware, and software built to help the enterprise overcome the cost and complexity of managing data at scale.

Seagate Technology crafts the datasphere, helping to maximize humanity's potential by innovating world-class, precision-engineered data storage and management solutions with a focus on sustainable partnerships. A global technology leader for more than 40 years, the company has shipped over 3 billion terabytes of data capacity.

Learn more by visiting www.seagate.com/lyve

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. 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 Custom Solutions 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.

IDC | ANALYZE THE FUTURE

idc.com

@idc

March 2021 | Doc. #US47565321