IDC

Data Deluge: Why Every Enterprise Needs a Cold Storage Strategy

Increasingly Massive Unstructured Data Sets Challenge Legacy Data Management

RESEARCH BY:

Andrew Smith
Research Manager
Cloud Infrastructure Services, IDC

Navigating This InfoBrief

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

Quantifying the Data Deluge

The quantity of stored data within enterprises is forecast to expand significantly, regardless of location: 64% of enterprises expect the raw storage capacity deployed on dedicated infrastructure to increase in the next year, and 60% expect the raw storage capacity deployed on cloud IaaS to increase during that time.

The total enterprise storage systems installed base is expected to grow at a 30.9% CAGR between 2020–2025, totaling 5.5 zettabytes (ZB) – that's over five million petabytes.

How does this data growth break down by location?

40%

Public cloud
CAGR to 4.2ZB

32%

Dedicated (private) cloud
CAGR to 509 Exabytes

8%

Traditional IT (non-cloud)
CAGR to 733 Exabytes

Note: Installed base is raw capacity deployed; public cloud includes IaaS/PaaS/SaaS and digital services providers like Facebook, Twitter, Netflix
Source: IDC's Storage Systems Installed Base, 2019-2025
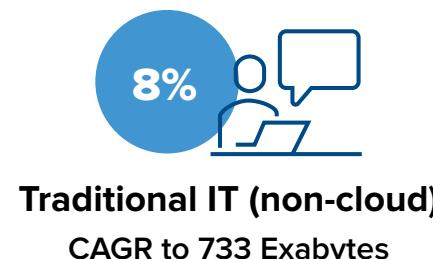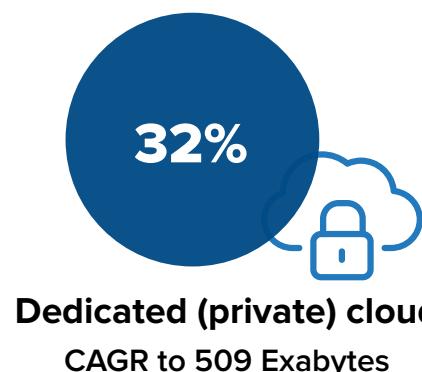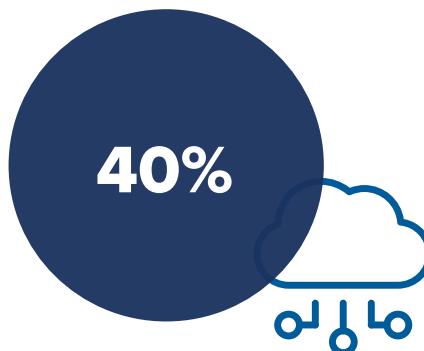
Data management is a foundational element of the enterprise IT infrastructure whether deployed on or off premises, in traditional non-cloud environments, or in public or private clouds. An effective data management strategy helps organizations enrich their data with tagging and classification tools, ensuring information can be accessed, organized, and retained according to the needs of the business.

What Are the Primary Causes of the Data Deluge?

The factors driving data growth are largely consistent across dedicated storage and cloud storage environments.

The top three capacity drivers are:

▶ Implementation of new data-driven projects (e.g., machine learning or artificial intelligence)

▶ Increased amounts of data generated by existing sources (e.g., increased resolution of video)

▶ General increase in digital data

Dedicated Storage Drivers
n = 308

TOP THREE

1. Increased amounts of data generated by existing sources (e.g., increased resolution of video)
2. General increase in digital data
3. Implementation of new data-driven projects such as machine learning or artificial intelligence
4. New sources of the data (e.g., sensors)
5. Implementation of data protection/ disaster recovery strategies

IaaS Storage Drivers
n = 384

1. Implementation of new data-driven projects such as machine learning or artificial intelligence
2. Increased amounts of data generated by existing sources (e.g., increased resolution of video)
3. General increase in digital data
4. Necessity to store data for longer periods of time (driven by corporate policies or regulatory requirements)
5. New sources of the data (e.g., sensors)
5. Workload migration from on-premises/ dedicated infrastructure to IaaS

Source: *IDC 2020Q3: IT Infrastructure for Storage Survey*, October 2020

Mind the Gap: Enterprise Data Growth vs. IT Budgets

Enterprises are generating and storing more data than ever before in support of modern digital business initiatives. IDC estimates that on average, enterprises can expect their storage capacity to grow 30% annually.

But while the perceived value of enterprise data may be limitless, IT budgets are not. Infrastructure spending will struggle to keep up with the demand for more storage and larger data sets. How do enterprises increasingly do more with less when it comes to data management?

Growth of Storage Installed Base Capacity Outpaces IT Infrastructure Spending

■ Storage Installed Base (Exabytes)   ■ IT Infrastructure Spending

Storage Installed Base (Exabytes):
- 2020: 1,418
- 2021: 1,829
- 2022: 2,399
- 2023: 3,212
- 2024: 4,247
- 2025: 5,451

IT Infrastructure Spending:
- 2020: $244,731
- 2021: $269,592
- 2022: $307,497
- 2023: $348,305
- 2024: $395,922
- 2025: $447,791
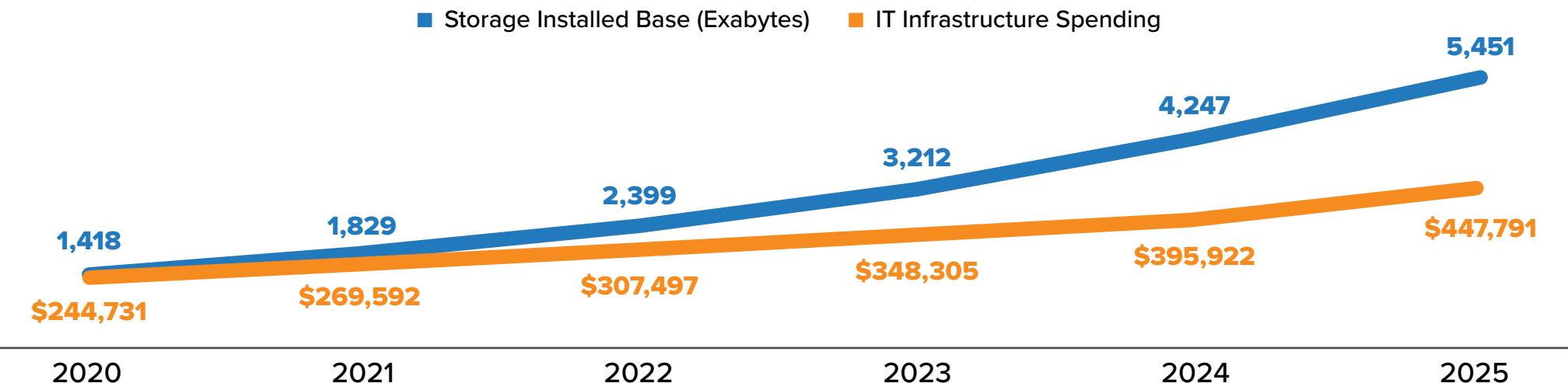
Sources: IDC's Storage Systems Installed Base, 2019–2025; *IDC WW ICT Spending Guide*

What Is a Cloud "Cold Storage" Solution?

Increasingly, enterprises are making deliberate trade-offs between storage access, performance, and price in order to effectively manage their insatiable appetite for data.

Cold storage solutions and services have become a valuable part of modern enterprise data management strategy by enabling the long-term storage of less active data, either in a standalone fashion (basic services offering storage capacity) or as one component of a data management platform.

As cold storage solutions mature, they continue to add advanced functions like data tiering, classification, and compliance as they seek to increase their value and relevance to enterprises.

Cold Storage Solutions: Key Use Cases

► Preventing extended periods of service downtime associated with traditional offline, offsite, and/or removable media

► Enabling cost-effective and timely access to new data sets (generated via artificial intelligence/machine learning algorithms, big data initiatives, device sensors, etc.)

► Delivering access to integrated or adjacent data services such as tagging, index and search, data visualization and management, and data analytics

Where Cold Storage Solutions Fit Within the Data Management Ecosystem

Think of enterprise data on a spectrum of "hot" to "cold," where data accessed frequently (hourly, daily) is considered hot and data accessed less frequently (weekly, monthly, annually) ranges from warm to cold.

In many enterprises, hot data is just the tip of the data management iceberg. Warm or cold data comprises most of the capacity stored.

As enterprise workloads become more demanding and require increasingly massive quantities of data for long-term analytics and compliance purposes, organizations will be pressured to manage and retain the 90% of data likely considered warm/cold.

**10%
Hot**

**30%
Warm**

**60%
Cold**
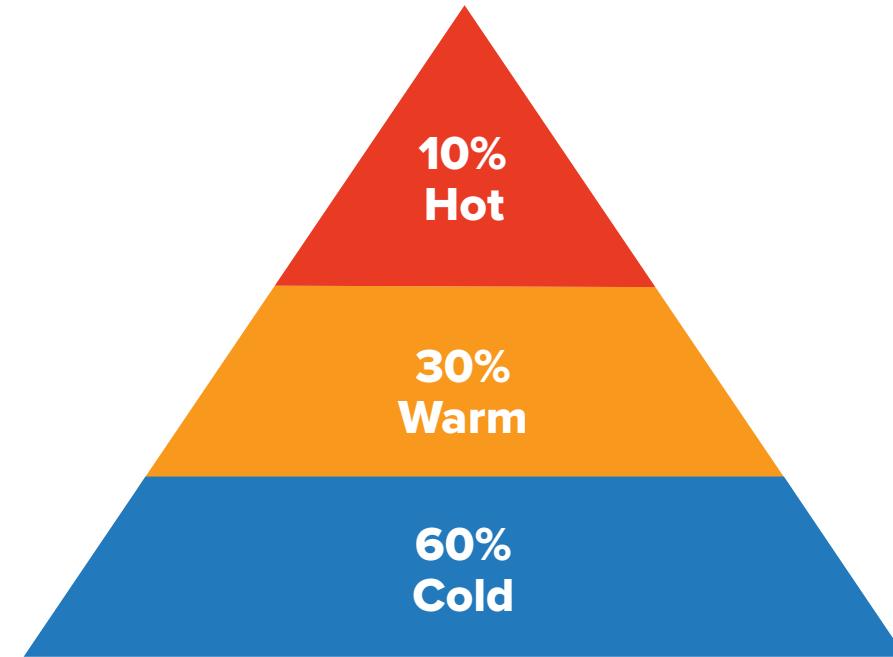
Many cold storage solutions serve the warm/cold segments of the enterprise data ecosystem, ensuring cost-efficient storage, timely access, security, compliance, and long-term retention of massive data sets.

Key Benefits and Challenges of Cold Storage Solutions

Benefits:

▸ **Low cost:** Solutions are built to achieve low costs, often prioritizing cost reductions over time as well

▸ **Extended retention:** Designed with long-term data retention in mind

▸ **Efficiency:** Solutions and services are built to consume the fewest resources in terms of power/cooling necessary

▸ **Low touch:** High level of data management and storage automation to achieve efficiency and minimal human intervention

▸ **Access to adjacent data services:** Built to offer access to adjacent/embedded services native to the ecosystem (e.g., analytics tools to produce reports and dashboards; security and compliance tools to trigger alerts or enforce access policies; self-service data retrieval and backup capabilities)

Challenges:

▸ **Longer data access times:** These solutions rarely deliver millisecond response time for data

▸ **Minimum storage duration:** Many solutions enforce a minimum storage duration for files/objects – typically measured in months or years

▸ **Costs associated with data access:** Many service providers have built additional billing structures around data access (e.g., egress)

▸ **Lack of automation:** The scale and cost profiles required for cold storage can often overshadow the need for automation and advanced data management tools which allow more active use of cold data

Cold Storage Solutions Provide More Than Low-Cost Data Management

Cold storage solutions are designed with low cost in mind, but cost is only one piece of the puzzle. Modern solutions also enable automated, timely access to data that might otherwise be discarded, lost, or deleted.

Key Purchasing Criteria for IaaS Storage

Q. When purchasing storage delivered as IaaS, what is the single most important purchase criterion?

Criterion	Percentage
Workload requirements (e.g., performance, availability, scalability, etc.)	14%
Storage platform features (snapshots, inline data reduction, quality of service, etc.)	14%
Security/trust delivered by provider	13%
Advanced security features	13%
Need for a specific storage architecture (e.g., multi controller array, software-defined storage, hyperconverged infrastructure, scale-out NAS, etc.)	12%
Performance of storage resources	10%
Price	10%
Need for a specific data organization and access method (e.g., block, file, or object)	6%
A vendor should be on our "approved" vendor list	4%
Tiering of storage resources/pricing	4%

n = 582 Source: *IDC 2020Q3: IT Infrastructure for Storage Survey,* October 2020

Buyer prioritization of "workload requirements" and "storage platform features" as key purchasing criteria are indicative of a shift in the perceived value of storage — moving from a focus on cost toward prioritizing workload performance, security, and the breadth of enterprise-grade storage features provided natively.

Building the Business Case: Improved Cloud Data Security and Compliance

Security and compliance remain top concerns of cloud storage buyers.

IDC data shows that 75% of enterprises expect secondary storage capacity to increase due to privacy regulations (e.g., CCPA, GDPR, and HIPAA). Security and compliance also are the top two reasons why enterprises host workloads in private cloud instead of public cloud.

Most cold storage solutions address these concerns with data security features like encryption, 11+ 9's of durability, geo-redundant replication and erasure coding, integrated data management and life-cycle policies, and long-term data preservation and deletion policies. Many of these features go beyond the baseline data management and protection offered in standard cloud storage services.

28%

Security

Top Concern with Public Cloud

47% **44%**

Data exposure Malware

Top Public Cloud Security Concerns

57% **52%**

Security Regulatory constraints or internal policy

Top Reasons to Host on Private Cloud vs. Public Cloud

Source: IDC, *IaaSView 2020*
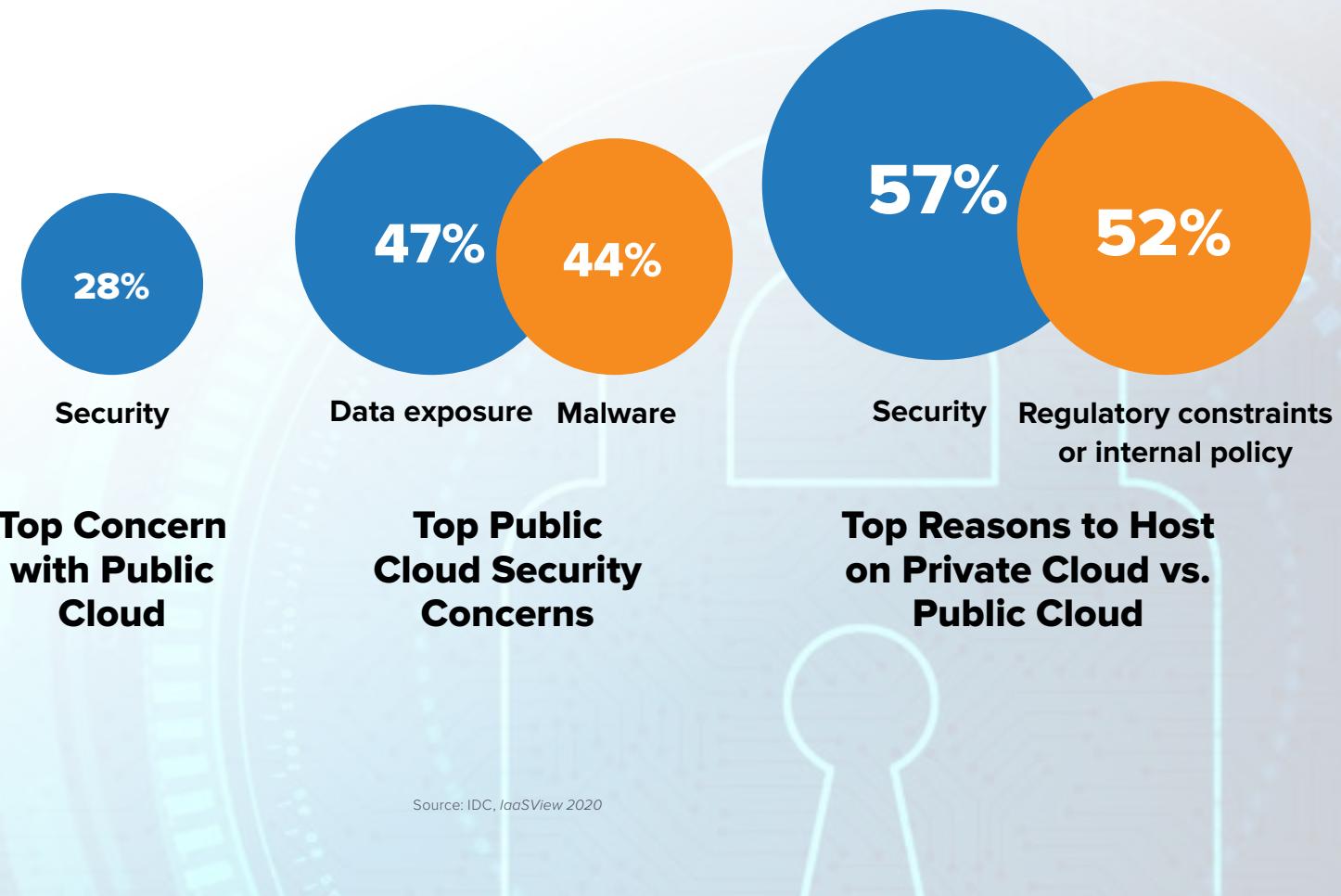
Building the Business Case: Enabling Long-Term Data Retention

The perceived value of data will cause many enterprises to rethink their data management strategy to deliver maximum retention (7+ years). It is impossible to know what data will be valuable for future business endeavors, or how data will be leveraged for future enterprise applications/workloads.

Many organizations will take a conservative approach to data deletion and disposal, preferring to keep as much data as possible, for as long as possible. This leads to increased storage durations and complexity. Cold storage solutions are architected with long-term data retention as a core tenet – providing tools for search, discovery, classification, audit, and deletion of data. With these tools, enterprises can be confident their long-term storage remains accessible when needed, and can be monitored and disposed of efficiently and compliantly.

Enterprise Data Retention Rates

61% of enterprises retain data for at least **six months**

46% retain data for at least a **year**

18% retain data for more than **seven years**
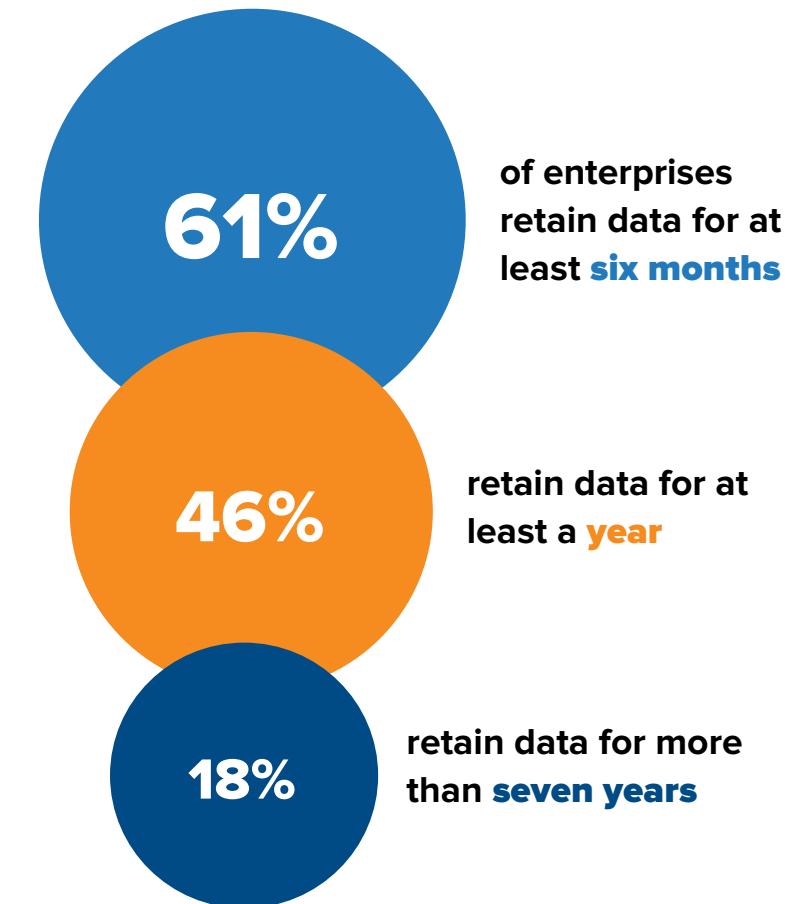
n = 1,007 Source: IDC survey data

Building the Business Case: Ensuring Data Movement

For many enterprises, data migration is a journey, not a destination. Cold storage services are an integral part of this journey, providing data management capabilities that efficiently tier data to the most cost-effective storage environments.

These services also help control the movement of that data via metadata tagging and classification, advanced compression and deduplication, and modern data protection and backup features.

Workload Repatriation: Primary Drivers

Q: What are the primary drivers for repatriating workloads to on-premises infrastructure or off-premises private cloud infrastructure?

Driver	Percentage
Consolidation of IT resources	62%
Better total cost of ownership (TCO)	56%
Concerns about data security	55%
Better return on investment (ROI)	53%
Compliance with regulations or corporate policy	52%
Performance issues	51%
Moving from pilot project running in public cloud to mainstream	47%
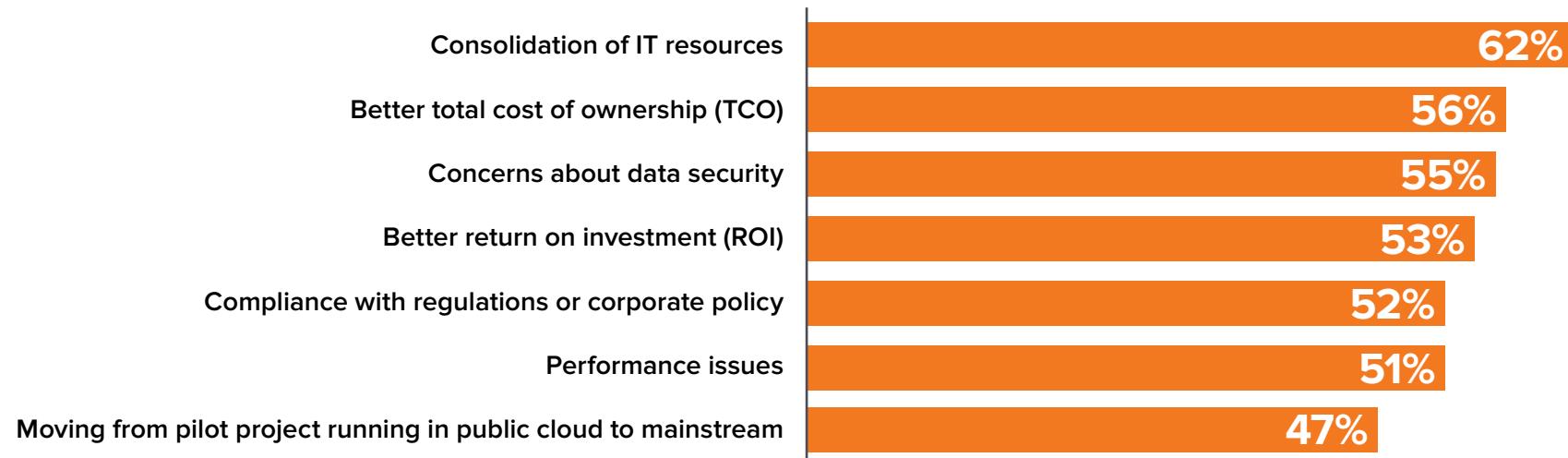
n = 510 Source: *Public Cloud Infrastructure Workloads Survey,* IDC, December 2019

About 50% of the applications are expected to stay in place over the next year.

The primary drivers of workload movement to and from cloud include:

- IT resource consolidation
- Better TCO
- Data security

Essential Guidance

If your organization is considering a cold storage solution,
IDC recommends the following evaluation criteria:

Expertise with Media Types

Using a range of storage media types (HDD, tape, optical, flash) should be seen as a differentiator. Innovations in storage media (e.g., molecular/DNA-based storage, flash-based storage) may result in significant cost savings in the long term.

High Levels of Automation

Extensive automation of data management-related tasks, integrated workflows, access to cloud storage APIs, and nondisruptive service improvements, all reduce human interaction and cost.

Access to Adjacent Services

Analytics, reporting, search capabilities, and seamless access to adjacent services (e.g., databases, security apps, collaborative apps) help ensure the platform keeps up with future business requirements.

Multiple Deployment Models

Finding a provider with a range of deployment flexibility (on premises, in the cloud) can be highly advantageous for enterprises with complex data management and security requirements.

About the Analyst

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 multi-cloud data management, data protection as a service, and public cloud cold storage.

More about Andrew Smith

Message from the Sponsor

Yesterday, data was simply stored and managed. Today, data is an essential differentiator.

At Quantum, we believe it's time to shift the focus from accumulating data to making it work harder for you and your business. Data is massively growing, widely distributed, unstructured, and gaining value at every turn. Your video and unstructured data not only needs to be fully protected, it needs solutions that enrich the data to unlock actionable insights that drive business. Your difference is in your data.

Learn more about our solutions to solve the data deluge with a cold storage strategy

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. As a premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets, IDC's Custom Solutions group helps clients plan, market, sell and succeed in the global marketplace. We create actionable market intelligence and influential content marketing programs that yield measurable results.

IDC

 @idc   @idc   idc.com

ST02370A-v01