A White Paper, sponsored by Dell Technologies and NVIDIA

IDC

High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises

RESEARCH BY:

Eric Burgener
Research Vice President,
Infrastructure Systems, Platforms
and Technologies Group, IDC

John Rydning
Research Vice President,
Global Datasphere, IDC

July 2022 | IDC Doc. #US49359722

Table of Contents

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

Introduction

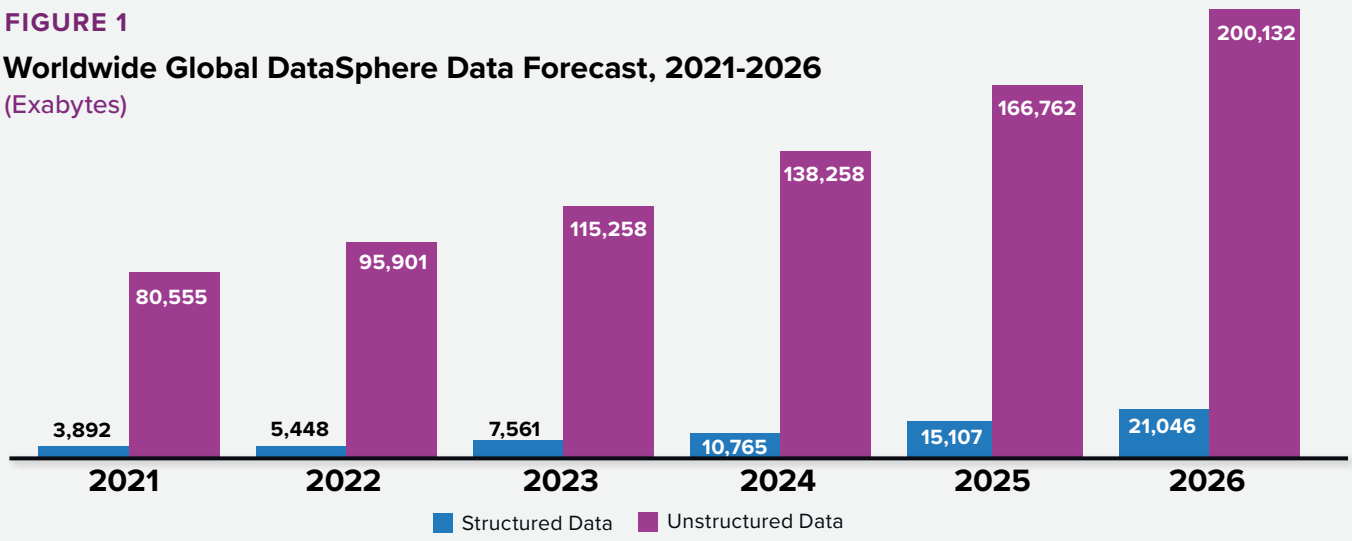
In 2018, enterprises began to embark on their digital transformation journeys, with the ultimate objective of becoming digital-first organizations. Digital-first organizations move to data-driven business models and thrive by delivering exceptional digital products to customers who expect digital-first experiences. IDC expects that a large number of enterprises will be achieving their digital-first goals in the 2023-to-2025 time frame.

To drive the insights that come from data-driven business models, IT organizations are capturing, storing, protecting, and analyzing more data than ever before. More data and improved algorithms result in better insights. And data growth at the edge, core, and cloud, fed by data mobility, big data and analytics, cloud, and social media requirements, is running rampant. IDC's Global DataSphere, which forecasts the amount of data that will be created on an annual basis, predicts that over the next five years, data will grow at a compound annual growth rate (CAGR) of 21.2% to reach more than 221,000 exabytes (an exabyte is 1,000 petabytes) by 2026. That data will include both structured and unstructured data, but unstructured data overwhelmingly dominates, accounting for more than 90% of the data created each year (see **Figure 1**).

FIGURE 1

Worldwide Global DataSphere Data Forecast, 2021-2026

(Exabytes)



Source: IDC WW Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2022-2026

Unstructured data includes video, photos and images, voice, and documents as well as other productivity content. The types of activities driving unstructured data creation include entertainment (primarily streaming media and gaming), the Internet of Things, non-entertainment images, productivity, social media, and voice. By 2026, entertainment and non-entertainment image data alone will make up more than 56% of unstructured data generation. The developing metaverse is expected to be a huge driver of data growth over the next five years, and its emergence has forecasters shifting predictions of future data growth strongly upward.

While it is still early days for the metaverse, IT leaders at forward-thinking companies must accelerate their journey now or risk playing catch-up with more agile competitors down the road.

As a concept, the metaverse is still a new idea, but there is no doubt it is providing innovative new options for socializing, buying things, learning, playing games, and achieving human fulfillment outside of the real world. What is abundantly clear is that the foundational technologies of the metaverse—technologies like augmented and virtual reality—generate massive amounts of data, and trends toward higher-resolution 3D design collaboration and rendering, increasingly rich digital experiences, and the vast appetite for data inputs to artificial intelligence, machine learning, and/or deep learning applications will all act as significant drivers to the growth of primarily unstructured data.

Digital transformation, with its imperative for enterprises to move to more data-centric business models, requires IT organizations to manage much larger amounts of data than they did in the past. New approaches now open to digital-first enterprises—like digital twins, AI-driven real-time decision making, and evolving forms of ecommerce that enhance customers' participation in virtual worlds—will strongly drive unstructured data growth as well. Handling this unstructured data deluge puts significant demands on both storage and compute infrastructure. Newer scale-out storage architectures, along with accelerated compute, will need to become part of the IT infrastructure of digitally transformed enterprises.

Storage Requirements for Digital-First Enterprises

While enterprises will need to maintain many legacy workloads going forward, the next-generation applications being developed and deployed as part of digital transformation efforts in general look very different. Built around microservices architectures and deployed in containers, most of these workloads will be expected to operate in a cloud-native manner, even when deployed on premises. To support these new cloud-native applications, storage infrastructure must support persistent storage in containers and be interoperable in Kubernetes environments. The coexistence of legacy workloads along with cloud-native applications requires careful planning and support by IT personnel.

In IDC's *Future Enterprise Resiliency and Spending Survey*, Wave 10, October 2021, cybersecurity and recovery investments were expected to provide the greatest strategic advantage for digital business success in 2022.

Increasing concerns around data privacy and the rise in ransomware attacks make security a top priority during the storage infrastructure refresh that more than two-thirds of IT organizations expect to encounter during their digital transformation efforts. As many of these next-generation applications will be mission-critical and customer-facing, they will also need to meet extremely high goals for data availability.

In IDC's *Modernized Infrastructure Survey*, March 2020, 68.2% of enterprises planned to modernize their storage infrastructure as part of their digital transformation efforts.

Big data analytics workloads have a voracious appetite for data, and most enterprises will have to manage petabyte-size data sets over the next several years. Many enterprises are already there. But merely storing the data at the edge, core, or cloud is not enough; it needs to be readily available for analysis, with the performance necessary to meet response time requirements. An increasing percentage of next-generation workloads (like those associated with the metaverse) will be real-time with each successive year, driving a need for not only very low latency but also very high concurrency against very large data sets.

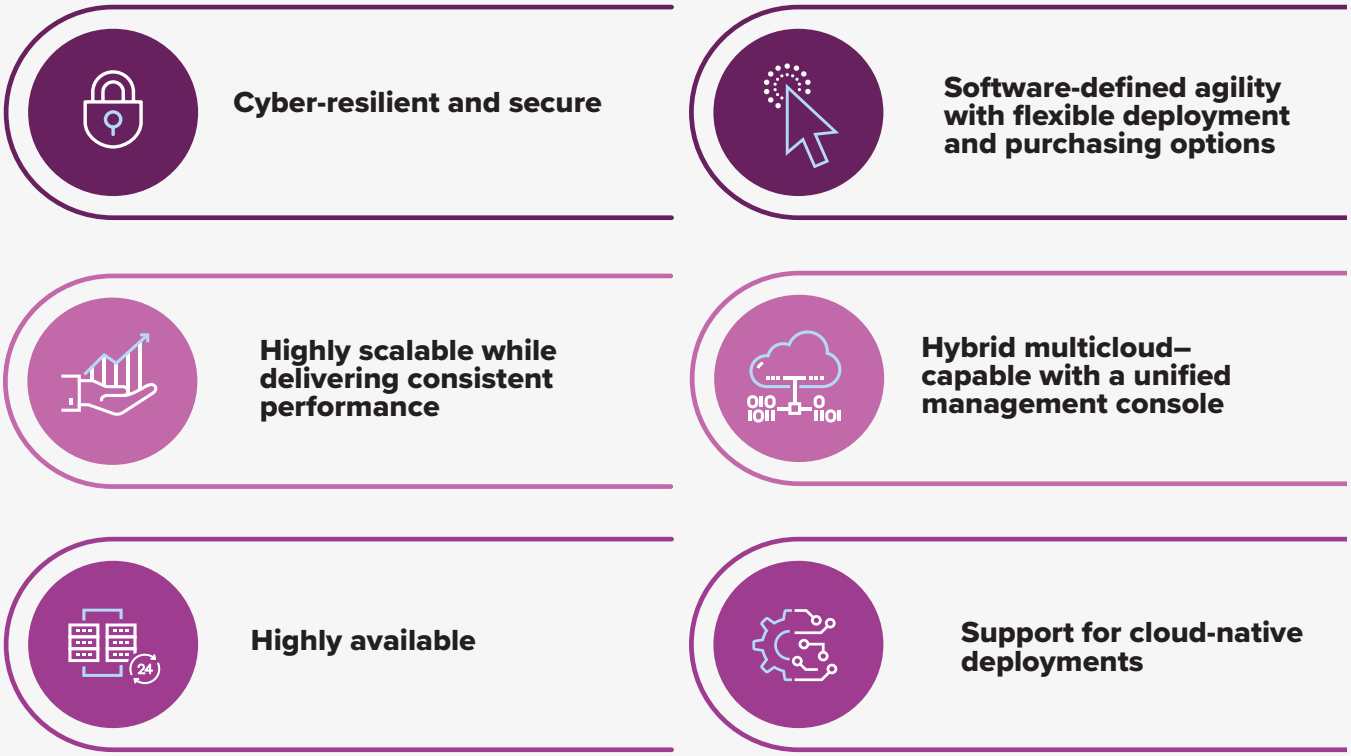
In IDC's *Storage Infrastructure Portfolio Survey*, September 2020, 62.5% of enterprises expected that they would need to support more real-time response requirements in the next one to two years.

With an increasingly dynamic business climate and the need to support hybrid multicloud–based IT infrastructures, flexibility and agility are critical. In crafting this type of hybrid IT infrastructure, enterprises are looking for software-defined storage infrastructure that supports a variety of different on- and off-premises deployment models. Software-defined solutions provide hardware deployment flexibility, are easier to manage than more traditional hardware-defined approaches, and offer better economics. But organizations are also looking for purchasing flexibility and need to be able to choose among various capital and operating expenditure options.

The more mature digital-first IT organizations know the importance of a unified management plane that spans on- and off-premises deployment models, providing comprehensive visibility for monitoring, managing, and troubleshooting distributed infrastructure. Most enterprises believe that the ability to deploy the same storage infrastructure software both on premises and in the top public cloud providers makes hybrid multicloud environments easier to manage (see **Figure 2**).

FIGURE 2

Storage Requirements in the Digitally Transformed IT Organization



Source: IDC, 2022

With environment, social, and governance (ESG) considerations beginning to impact IT infrastructure decisions, IT managers are also looking at green initiatives to improve storage infrastructure efficiency. Factors like storage density, energy and floor-space consumption, and server CPU utilization affect not only carbon footprint but also overall costs, and improving infrastructure efficiency is another key reason IT managers are looking to modernize their storage infrastructure during digital transformation.

Accelerated compute will be a key part of most enterprises' infrastructure modernization projects. Legacy storage systems have had trouble keeping up with the general-purpose CPU advancements over the last decade. To use the additional performance of accelerated compute requires newer storage architectures that can provide low latency, handle massive parallelism, and offer integration points that help to more fully utilize and improve the efficiency of graphics processing units (GPUs), application-specific integrated circuits (ASICs), field-programmable gate arrays (FPGAs), and other accelerated compute options.

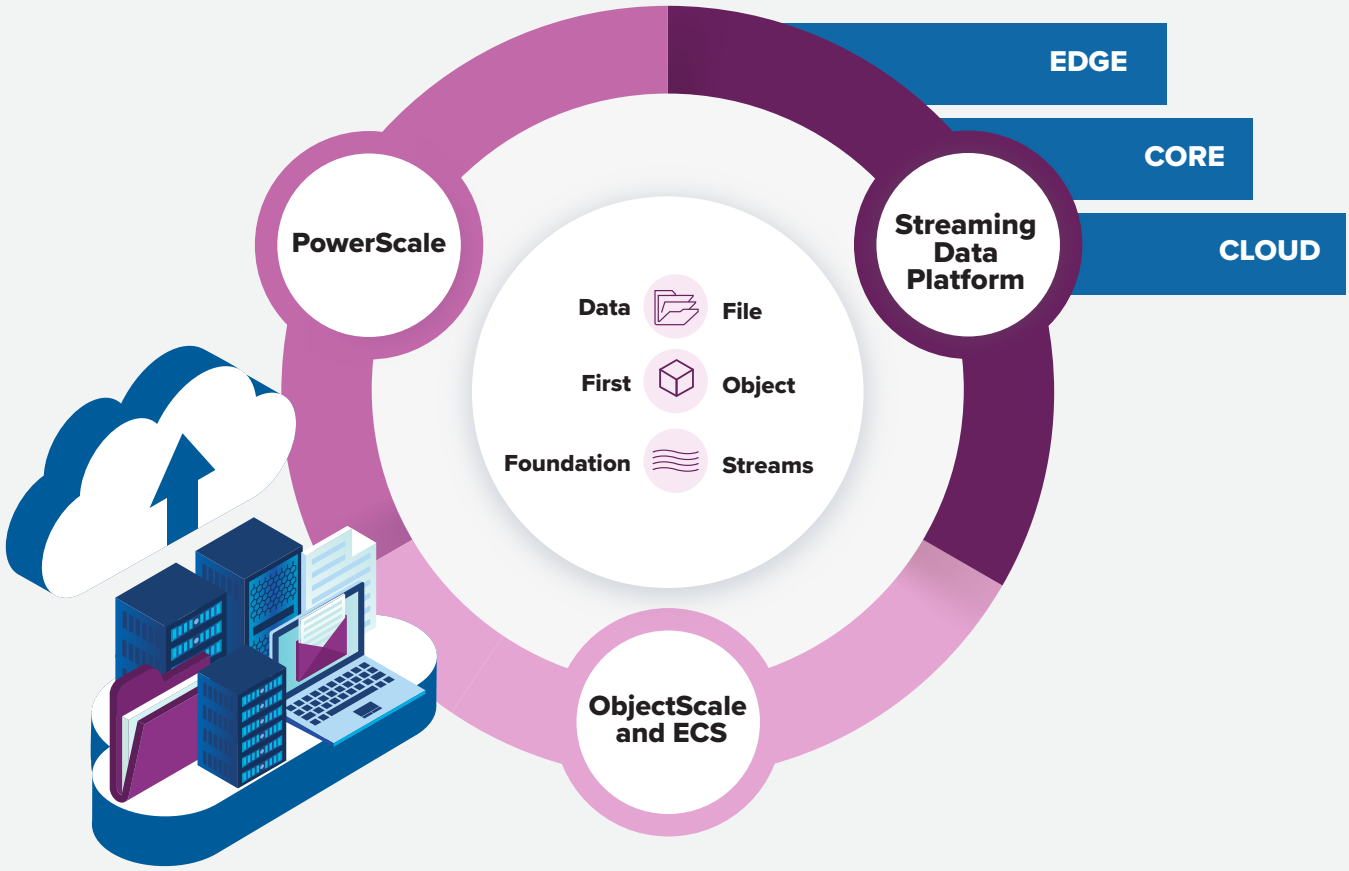
Dell Unstructured Data Storage Portfolio

Dell is a $92 billion enterprise infrastructure provider that has been a longtime market-share-by-revenue leader in both servers and storage. Dell has a comprehensive enterprise storage portfolio that meets both structured and unstructured storage needs; includes both scale-up and scale-out architectures; spans entry-level, midrange, and high-end offerings; features edge, core, and cloud-based platforms; and supports excellent integration into virtualized infrastructure as well as hybrid multicloud environments.

Dell's unstructured data storage portfolio includes Dell PowerScale, Dell ObjectScale, Dell ECS, and Dell Streaming Data Platform (SDP). Dell PowerScale is a highly scalable native file storage platform with enterprise-class availability and storage management capabilities that supports a broad range of access methods for highly efficient workload consolidation. Dell ObjectScale (software-defined option) and Dell ECS (appliance option) are highly scalable native object storage platforms, both supporting modern workloads such as artificial intelligence, analytics, and media content as well as more traditional backup and archive object storage workloads. Dell SDP is built around Pravega, a cloud-native streaming infrastructure specifically designed for continuously generated and unbounded data, and can leverage PowerScale, ObjectScale, and ECS at its core (see **Figure 3**, next page).

FIGURE 3

Dell's Unstructured Data Portfolio Covers a Range of Unstructured Data Storage Needs



Source: Dell Technologies, 2022

A variety of mixed media types are supported between the two object platforms (ObjectScale and ECS), giving customers the flexibility to put performance and cost-effective high capacity where it is needed, and they both are built around software-defined, scale-out architectures that can scale to tens of petabytes and beyond. Kubernetes support is built into these systems, and they can be deployed in on- or off-premises locations in storage clusters that support all-flash, hybrid, and archive nodes. All platforms can be purchased outright or acquired through Dell Technologies APEX, the vendor's portfolio of as-a-service offerings that provides pay-as-you-go options for core, edge, colocation, and public cloud–based deployments. The APEX Console provides a unified management plane that lets customers choose and configure cloud and infrastructure subscriptions to match their business needs, (see **Table 1**, next page).

TABLE 1

Mapping Dell PowerScale and Dell Object Platforms (ObjectScale and ECS) to Digital-First Storage Requirements

Requirement	Dell PowerScale	Dell Object Platforms
Cyber resiliency and security	OneFS Zero Trust Model for security, 256-bit encryption, FIPS compliance, SmartLock for immutability, ransomware attack prevention, detection, and recovery, air gap protection	Highly distributed erasure coding, 256-bit encryption, versioning, resource isolation, global identity and access management, ObjectLock for immutability
High performance and scalability	Low latency with NVMe-based all-flash "F" nodes, QoS, scale up to 252 nodes, tens of petabytes and almost 1TB/sec of bandwidth under a single unified namespace	Support for flash and spinning disk media, hundreds of petabytes of capacity, and multiple TBs/sec of bandwidth (no cluster node count limit)
High availability	Erasure coding or file replicas, transparent recovery from device and node failures, snapshots, and replication, SyncIQ for simple DR failover/failback	Geographically distributed erasure coding, transparent recovery from device and node failures, snapshots, and replication
Software-defined deployment flexibility	Deploy as an appliance, on selected Intel x86-based servers, in virtual machines or containers, and in the public cloud	Deploy as an appliance, on selected Intel x86-based servers, and in containers
Multicloud–capable with unified management	Automated tiering to S3 storage in on- or off-premises locations through SmartSync, SmartPools, CloudPools; support for Amazon, Azure and Google public cloud deployment, supported in APEX Console	Participates in SmartPools, CloudPools; APEX Console support planned
Support for cloud-native deployments	CSI support, Kubernetes support, REST APIs for integration into management, orchestration and directory tools, container-based deployment	COSI support planned, Kubernetes support, deploy on Red Hat OpenShift or VMware Tanzu, REST APIs for integration into management, orchestration and directory tools; container-based deployment

Source: IDC, 2022

While the Dell PowerScale, ObjectScale, ECS, and SDP configurations include features that map directly to the requirements of digitally transforming organizations, the platforms include capabilities that deliver additional benefits as well.

Dell PowerScale includes at-scale optimizations for in-line data protection, in-line data reduction capabilities (compression, deduplication) that improve cost efficiency, multiple replication topologies, nondisruptive multigenerational technology refresh, and a simple "snap to object" feature that makes it simple to create a comprehensive recovery copy of an entire namespace. Additional performance features include Network File System (NFS) over remote direct memory access (RDMA) host connection and support for NVIDIA's GPUDirect storage API. Recently released quad-level cell (QLC) drive support delivers twice the cluster capacity, twice the node density, and lower costs without any increase in power, cooling, or floor-space requirements.

The Dell Validated Design for VDI with NVIDIA Omniverse

NVIDIA is a leading supplier of accelerated compute technology deployed across enterprises, government agencies, and research laboratories alike. NVIDIA Omniverse is an easily extensible platform for 3D design collaboration and scalable multi-GPU, real-time, true-to-reality simulation that is fueling the efforts of many organizations playing in the metaverse.

The powerful NVIDIA A40 GPUs need high-performance storage to keep them operating efficiently, and Dell's partnership with NVIDIA has produced NVIDIA-Certified Systems that deliver on this requirement. The Dell Validated Design for virtual desktop integration (VDI) with NVIDIA Omniverse and NVIDIA RTX Virtual Workstation (vWS) software uses a VxRail-based virtual workstation environment with VMware Horizon and PowerScale as the core media repository, and is validated with Autodesk Maya 3D animal and visual effects software.

Dell Object platforms include different erasure coding scheme options, a comprehensive S3 API implementation, intelligent workload sizing, native multi-tenancy with resource isolation and secure access, nondisruptive technology refresh, and advanced retention, indexing, replication, and reporting features that help to ensure data compliance. In the future, Dell will support a nondisruptive upgrade to ObjectScale from the Dell ECS object storage platform that has been shipping since 2014.

All systems support a number of different device-type capacities, allowing customers to select the mix that best meets their performance, capacity, and cost requirements.

The primary use cases for Dell PowerScale include media and entertainment, healthcare (e.g., DNA sequencing), research, technology, telecommunications, autonomous driving, agriculture, public safety, general-purpose file sharing, and a variety of artificial intelligence–based, performance-intensive, and other technical computing workloads across a variety of industries.

The primary use cases for Dell ObjectScale include cloud-native web and mobile applications, test/dev, consolidated data lakes, artificial intelligence and machine learning, data analytics, the IoT, and backup and archive. ECS use cases generally exclude cloud-native applications but are otherwise similar. Both storage platforms can be used with the emerging, more performance-oriented big data analytics across many industries, with ECS in particular supporting those workloads through available NVMe support options.

Dell unstructured data storage platforms offer the performance, availability, scalability, and comprehensive enterprise-class management capabilities needed for both legacy and modern workloads that will be deployed as part of digital transformation.

Challenges and Opportunities

As enterprises go through digital transformation, deciding where to place workloads is a key consideration. Given that most enterprises will be using hybrid multicloud environments going forward, IT infrastructure choices need to support the flexibility necessary to accommodate both older and newer workloads in both on- and off-premises locations. The challenge for vendors is to evolve with the times, providing the set of capabilities required to support the massive unstructured data growth being driven by the burgeoning metaverse era. Vendors that do flexibly support those capabilities are providing the foundation on which their customers can effectively differentiate themselves from their own competitors.

Conclusion

There will be an explosion in unstructured data over the next five years, driven by digital transformation, the metaverse, and a number of other data-centric trends. Capturing, storing, protecting, and analyzing this data while meeting evolving business requirements will be increasingly difficult for legacy storage infrastructure, and enterprises are looking to newer, more software-defined scale-out designs. The storage infrastructure platforms of the future need to be cyber-resilient and secure, able to deliver performance at higher levels of scale than ever before, highly available with quick recovery, and software-defined for flexibility. They also need to support hybrid multi-public-cloud environments that are managed by a unified management console and provide the necessary support for cloud-native applications. Vendors like Dell that offer a broad, proven unstructured data storage portfolio, unified hybrid cloud management, and flexible consumption models, along with integration into accelerated compute technology from leaders like NVIDIA, provide the foundation on which successful digital-first enterprises can be built.

About the Analysts

Eric Burgener
Research Vice President, Infrastructure Systems, Platforms and Technologies Group, IDC

Eric Burgener is Research Vice President within IDC's Enterprise Infrastructure Practice. Eric's core research coverage includes Storage Systems, Software and Solutions, quarterly trackers, end-user research as well as advisory services and consulting programs. Based on his background coverage on enterprise storage, Eric's research includes a particular emphasis on flash-optimized arrays, emerging persistent memory technologies, and software-defined storage. He is an active participant in the IT Buyers Research Program at IDC and blogs throughout the year on the topic of Infrastructure and Data Management.

More about Eric Burgener

John Rydning
Research Vice President, Global Datasphere, IDC

John Rydning is responsible for the Global DataSphere forecast, which measures the amount of data created by year and also for the Global StorageSphere forecast, which is a measure of the installed base of storage capacity worldwide, and the amount of data stored in any given year. Additionally, John leads insightful research that explores key trends, use cases, technologies, and other factors shaping both the Global DataSphere and StorageSphere.

More about John Rydning

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.