

White Paper

The U.S. Datasphere: Consumers Flocking to Cloud

Sponsored by: Seagate

John F. Gantz
January 2019

David Reinsel

John Rydning

EXECUTIVE SUMMARY

The Global Datasphere, a measure of how much new data is created and replicated each year, will grow by more than five times over the next seven years. The total amount of new data created in 2025 is forecast to increase to 175ZB from 33ZB in 2018.

The major drivers of this growth are largely consistent across the world's various regions but occur at different rates. Entertainment data and video surveillance footage have long been (and continue to be) significant drivers of the Global Datasphere. However, signals from the Internet of Things (IoT) devices, metadata (vital for analytics, contextualization, and artificial intelligence [AI]), and productivity data are showing even faster growth in today's increasingly digitized world.

Nevertheless, amid the similarities across various regions, there are subtle differences. These differences are based on technology adoption and digital transformation across a region's population of consumers and enterprises.

For example, the U.S. Datasphere is growing more slowly than many other regions. The U.S. Datasphere is 21% of the Global Datasphere in 2018 but only 17.5% in 2025, growing from 6.9ZB to 30.6ZB, a 23.6% CAGR for 2018-2025. Like other regions, much of the growth will be driven by signals from IoT devices, metadata, and video surveillance. However, growth in entertainment-driven data is slowing as productivity-related data accelerates.

Consumers and businesses in the United States are integrating technologies to drive efficiencies and productivity. The desire to be connected and provided with real-time data is accelerating cloud adoption and data creation by servers in the core and at the edge. In fact, the percentage of data in the U.S. Datasphere emanating from or replicated in the edge will grow by more than seven times – from 15% to 25% of the total U.S. Datasphere – as IoT devices increasingly drive processing and analytics closer to the point of origin of the data itself.

Data is at the heart of this digital world, and we are increasingly becoming an information-driven economy. Deeper and more incisive insights are being extracted from data so that we can create a new world of smarter products, better customer experiences, and self-learning and always improving digital services. Data is also the heartbeat of modern user experiences and services built using next-generation technologies such as cognitive, IoT, AI, and machine learning (ML).

Although not all data created will require permanent storage, the amount of data stored in the United States will grow from 0.9ZB in 2018 to 3.4ZB in 2025. Between now and then, the use of public cloud

storage is poised to increase aggressively. In fact, by 2025, 55% of the data stored by organizations and consumers will take place in the public cloud, compared with 49% globally.

This unprecedented data growth combined with the pressures of deriving value from data for digital transformation will create imperatives for organizations across all regions over the next decade to develop appropriate data storage, management, and capitalization strategies that bring about a new level of engagement with people using data-informed services and products.

METHODOLOGY

IDC has been studying the size and nature of the Global Datasphere – all the data created and replicated in one year – for more than a decade. The data creation numbers are driven by IDC forecasts of installed devices and their data creation or capture capacities across more than 70 categories. The analysis takes into account duty cycles and compression techniques. IDC also sizes the amount of data that is stored. It is driven by IDC's ongoing market analysis of the storage market in more than 80 countries.

IN THIS WHITE PAPER

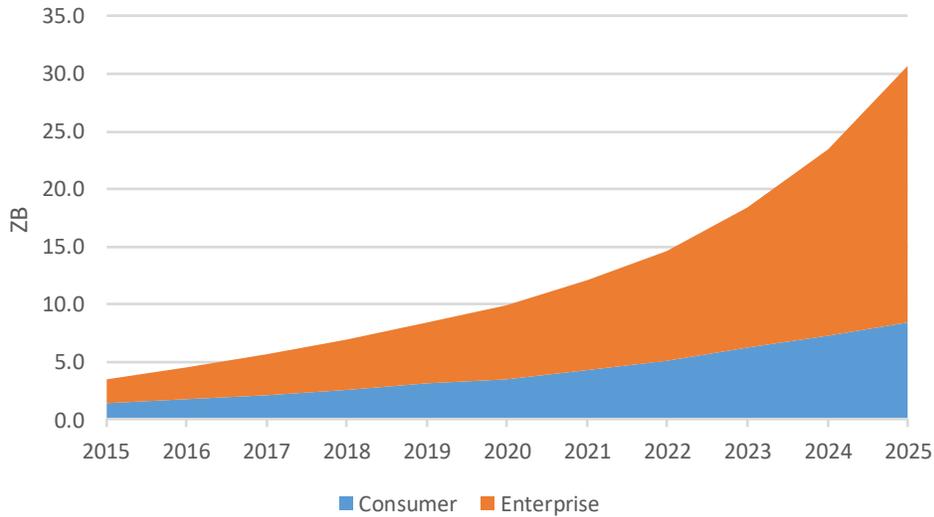
This White Paper is a regional companion document to *The Digitization of the World – from Edge to Core* (IDC #US44413318, October 2018). It summarizes the trends and dynamics of the Datasphere and of data storage pertaining to the United States.

CREATING DATA IN THE DATASPHERE

A decade of growth of the U.S. Datasphere is shown in Figure 1, which also shows the changing share of the Datasphere generated by consumers and enterprises. The enterprise Datasphere grows from 60% of the U.S. Datasphere in 2015 to 72% in 2025.

FIGURE 1

U.S. Datasphere: Originator Segmentation, 2015-2025



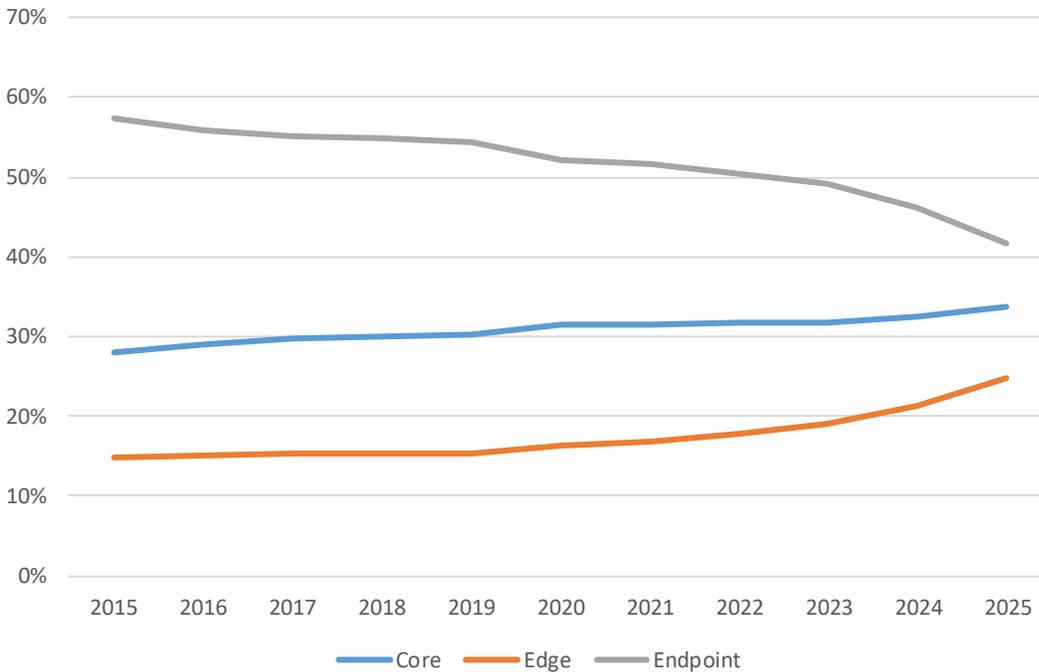
Source: Data Age 2025, sponsored by Seagate, November 2018

Growth of the enterprise share of the Datasphere is driven by several dynamics, including the growth of big data and analytics, which in turn is compelling longer data retention time periods; the proliferation of applications and IoT devices and sensors that act as systems of engagements with people and collect data; the completion of the migrations of analog to digital music, TV, and movies combined with more people streaming rather than storing entertainment media; and the migration of consumer data storage from local devices to the cloud.

Figure 2 shows the U.S. Datasphere based on where data originates or is replicated. Basically, while most data will originate at endpoints, more and more will be replicated, transferred, or backed up at the edge. Some of the data replicated at the edge will, in turn, be replicated at the core.

FIGURE 2

U.S. Datasphere: Location, 2015-2025



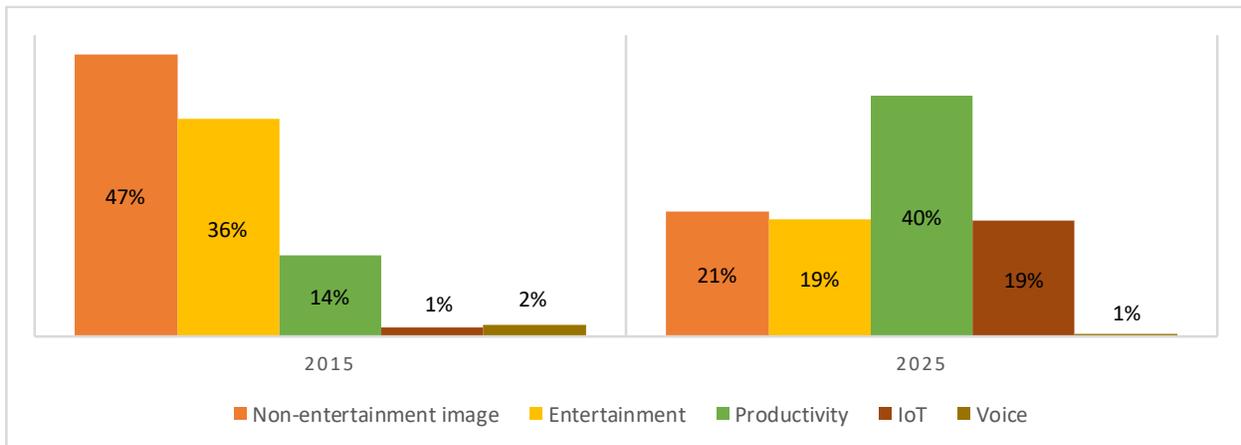
Source: Data Age 2025, sponsored by Seagate, November 2018

The *type* of data in the U.S. Datasphere will change dramatically, as shown in Figure 3. While entertainment data will grow by a factor of 4.7 from 2015 to 2025, it will be outpaced by growth in productivity (big data and metadata) and IoT such that entertainment data will fall from 36% of the Datasphere in 2015 to less than 20% in 2025. Growth of the image sector, which includes surveillance video, medical imaging, and entertainment-related content, slows as the conversion from analog to digital in the media and entertainment industry slows considerably. One of the top 5 drivers of data creation, complementing the growth to cloud, as well as analytics and AI supporting the drive toward more efficiency and productivity, comes from servers. The amount of server data creation is expected to grow at a 33% CAGR for 2018-2025.

There are other types of data that present challenges (and opportunities). For example, mobile data continues to grow at the same pace as the Global Datasphere today (28%); hypercritical data that can affect human lives or property – like telemetry of self-driving cars, real-time medical imaging, and antiterrorism facial recognition – is growing 32% annually; and data touched by artificial intelligence is increasing at a rapid CAGR of 68% for 2015-2025. Use of data in life-critical areas such as automated cars, defense, and healthcare will become viable only if the systems are in place to make sure the data is secure, available, accurate, and contextualized.

FIGURE 3

U.S. Datasphere: Data Type Share, 2015 and 2025



Source: Data Age 2025, sponsored by Seagate, November 2018

One example of growth in productivity data is how Union Pacific, a U.S. railroad company, turned to using machine vision technology to inspect rail cars. The system imaging area, a structure built over the railroad tracks housing cameras, lasers, radars, and sensors, can scan a mile-long train moving at 70mph capturing 50,000 photos per second, creating 30GB of data every few seconds. The photos are stitched together to form 3D images of the train, which are analyzed near real time to inspect for defects that need to be repaired. Union Pacific has also begun testing a sensor installed on rail cars that can collect roughly 40,000 measurements of vibration and movement per second. Data collected by the sensors are transmitted to data storage locations for further analysis.

The digitization of oil fields has helped make the United States one of the largest and most productive crude oil producers in the world. In one of many examples, Apache Corp., an oil and gas exploration and production company, installed data aggregation boxes on onshore rig sites in North America. The boxes run physics-based analysis of daily drilling raw data to make better data-driven drilling decisions to reduce drilling costs. Apache also built a remote operations center in Elk City, Oklahoma. The remote operations center collects and analyzes data from well pads and tank farms spread across Texas and Oklahoma to reduce production downtime and improve the efficiency of trucking and hauling operations. The center is manned by one person during normal working hours.

Storing Data from the Datasphere

Much of the data in the Datasphere will not be stored in any permanent way. In fact, total U.S. storage capacity relative to the size of the U.S. Datasphere in 2018 was 20%. By 2025, it will be a little more than 15% of the U.S. Datasphere. This is because most of the data in the Datasphere will evaporate after use – such as digital TV signals not stored on DVRs, multiplayer gaming uploads and downloads, IoT sensor signals that don't send off alarms, and surveillance images that are quickly overwritten with new surveillance images.

The value of data is increasing, and it is imperative for companies to understand the value of the data that they store. Data is the lifeblood of the data age – it can support customer-facing activities,

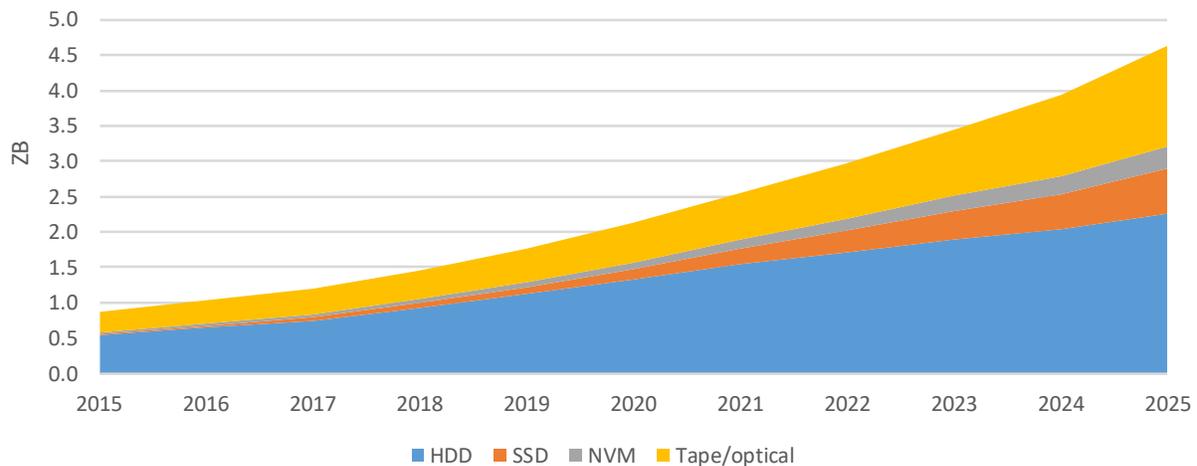
operations, research and development, retention of intellectual property, and financial and employee records. Increasingly, data is leveraged in automation, AI, and IoT. Data can also be sold, creating new data-as-a-service opportunities and revenue streams. Hence companies must manage their data properly.

With the value of data in mind, companies, as well as governments and countries, should take note of how much data and storage capacity they have. For example, in 2018, the United States had 29.1% of the world's 5.0ZB installed base of storage capacity¹. This share of storage capacity decreases to 28.5% in 2025, when the world is expected to have an installed storage capacity of roughly 16.5ZB.

Figure 4 shows growth of U.S. installed storage capacity that will be available from 2015 to 2025 by storage type. Perhaps surprisingly, tape and optical will continue be important storage technologies, even as movie music distribution migrates from media such as DVDs and CDs to streaming. Archival storage and backup on tape and optical systems will still be needed years from now.

FIGURE 4

U.S. Datasphere: Installed Storage by Media Type, 2015-2025



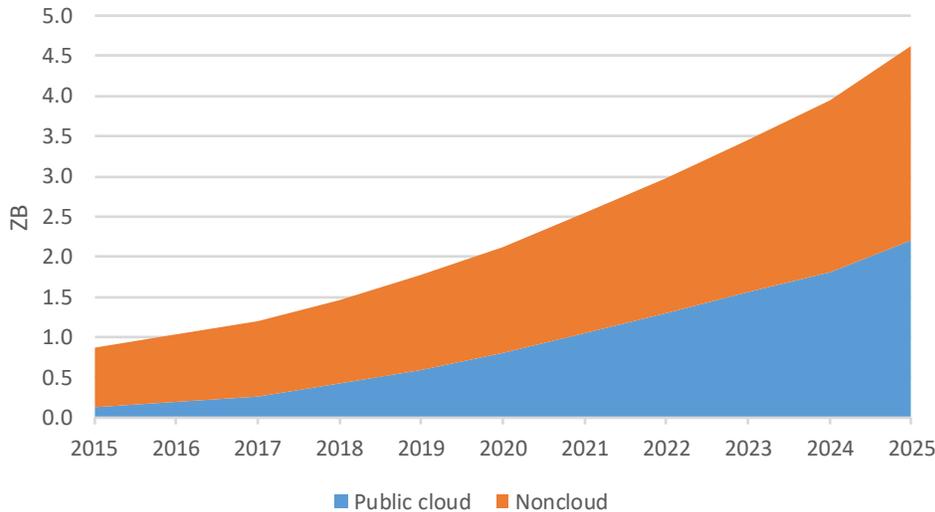
Source: Data Age 2025, sponsored by Seagate, November 2018

There *will* however be a major change in the storage environment as storage migrates from the enterprise and end-user devices to the cloud. Public cloud installed bytes will grow from 15% in 2015 to 48% in 2025 (see Figure 5). In fact, IDC believes that public cloud and hybrid on-premise and cloud adoption by U.S. organizations is already impacting enterprise storage infrastructure strategies and is entering a phase of maturation and common use. Nevertheless, 49% of U.S. organizations plan to expand the use of cloud storage over the next two years.

¹ Installed base of storage capacity is the sum of used and unused bytes across hard disk drives, flash memory, and tape and optical media.

FIGURE 5

U.S. Datasphere: Installed Public Cloud Storage, 2015-2025



Note: Noncloud equals all HDD, flash, tape, and optical bytes not residing in a public cloud infrastructure.

Source: Data Age 2025, sponsored by Seagate, November 2018

IDC GUIDANCE

The nine times growth of the U.S. Datasphere from 2015 to 2025 by itself will create challenges for enterprises – in management, security, storage, and utilization. The United States is in the throes of digital transformation – using new technologies and applications to transform its businesses.

As a developed economy, the United States must deal with rules of compliance, legacy information technology and organizations, and global competition.

But the challenges (and opportunities) won't come from data growth alone. Consider just three aspects common across all regions:

- **Security:** On a global basis, IDC estimates that in 2018, 56% of the data in the Datasphere required some kind of security protection, from simple protection of account information to full "lockdown" protection of bank deposits, critical infrastructure, and user identities. By 2025, that 56% will increase to 66%. And yet the percentage of the data needing protection that *is* protected will barely move from 45% to 50%. At this rate, unprotected data in the United States that *needs* protection will grow faster than its own Datasphere between now and 2025 and account for a full one-third of the Datasphere, or 10.2ZB, by 2025.
- **Real-time data:** Again on a global basis, the real-time percentage of the Datasphere will grow from 12% in 2015 to 29% in 2025, driven by growth of IoT. Apply these percentages to the U.S. Datasphere and you get better than 20 times growth in real-time data by 2025, or 18% of the world's real-time data. This will not only drive automation to edge computers that aren't

already in place but also introduce more interrupt-driven traffic into IT organizations as they begin to inherit responsibility for computing – once the province of operational organizations.

- **Data fragmentation:** As multicloud infrastructure proliferates and as organizations develop a continuum of IT services and applications from edge to core to cloud, the data is highly fragmented across multiple platforms and applications making it harder to identify, classify, manage, secure, and utilize this data. Organizations need to evaluate the complete data pipeline and develop a data-first strategy to mitigate risks.

And it won't just be enterprises facing challenges of the Datasphere. By 2025, the percentage of the global population interacting with data will approach 75% – and will surely be higher in the United States. Gigabytes per day per capita – and it *is* gigabytes – will grow at 21% a year between now and 2025. For example, in 2025, the average connected consumer in the world can expect to have one digital data engagement every 18 seconds, which translates into almost 5,000 interactions per day.

CONCLUSION

The Datasphere is large, dynamic, and complex and increasingly intersecting the physical world – a far cry from the past century where data was something kept in records and files, analyzed over time (if at all), and assisting with running – but not running – factories, automobiles, home appliances, or toll systems. Data growth and the growing value attached to it are changing the U.S. consumer and business landscape, and data is shaping how consumers, governments, emergency service providers, and businesses work.

Web inventor Sir Tim Berners-Lee says that data is a precious thing and will last longer than the systems themselves.

Businesses are using data to reach new markets, better serve existing customers, streamline operations, improve productivity, and even create new revenue streams selling it. Data may not be on a balance sheet, but data is a company's most valuable intangible asset, which can lend a competitive edge in digital transformation. One has to just remember Caesars Entertainment's (one of the world's largest casino entertainment companies) bankruptcy: the data it had collected and analyzed became its most valuable asset (worth \$1 billion) for creditors.

Companies looking to be relevant between now and 2025 will need to understand the role data plays in their organization and how the Datasphere will evolve during that period. They will need to embrace their role as data guardians, leverage the cloud, and take a global approach to their data. But organizations need to make data management and innovation priorities to remain competitive in the digital era as digital disruptors across all verticals are making data-driven innovation their key priority.

Consumers are building more and deeper connections to data and accessing products and services more easily – and at the time and place of their choosing. They are also benefitting from productivity improvements and advances in medical technology, enjoying new forms of entertainment, and living in smarter homes and cities. They are also beginning to expect products and services that are smarter and that learn over time to provide accurate, personalized experiences without breaching data privacy sentiments.

In the digital era, model airplanes have morphed into self-navigating, video-capturing rescue drones; 100,000s of books can be downloaded onto a single digital device smaller than a paperback book; and vacuum cleaners can vacuum a home, navigate around obstacles, and return to their own charging

stations all on their own. Ultimately, in this new age of data, organizations will need to become data savvy to optimize opportunities and mitigate risks using innovative technologies and become data thrivers to gain a competitive advantage or remain relevant in the data-driven economy.

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.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2019 IDC. Reproduction without written permission is completely forbidden.

