

## White Paper

# The China Datasphere: Primed to Be the Largest Datasphere by 2025

Sponsored by: Seagate

David Reinsel  
January 2019

Lianfeng Wu

John F. Gantz

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.

The China Datasphere is the fastest-growing regional Datasphere and is outpacing the Global Datasphere by 3% annually on average. In 2018, China's Datasphere was 23.4% of the Global Datasphere, or 7.6ZB. This will grow to 48.6ZB in 2025 and emerge as the largest Datasphere in the world, at 27.8% of the Global Datasphere. Similar to worldwide dynamics, the China Datasphere will be driven by growth of signals from IoT devices, metadata, entertainment-related data, cloud, and edge computing. The percentage of data in the China Datasphere emanating from or replicated in the edge will nearly double – from 13% to 23% of its total Datasphere – as IoT devices increasingly drive processing and analytics closer to the point of origin of the data itself and as smart infrastructures like buildings and bridges and Smart Cities leverage edge infrastructure and computing to enable a real-time world.

While video surveillance is a common driver of content within the Global Datasphere, it is especially significant to China based on a nationwide endeavor called Project Dazzling Snow. Beginning as an experiment to bring new levels of intelligence and security to China's northwest and other rural communities, Dazzling Snow is now expanding across the country under the banner of infrastructure modernization, improved security and intelligence, and supporting a nationwide strategy of Smart City construction. The infrastructure necessary to support this large ambition around massive intelligent video surveillance and analytics is a boon to technology providers across the full spectrum of IT from storage to compute, networking, and software. While this type of data doesn't feed growth of public

cloud, it certainly fuels growth in enterprise infrastructures of all types and is reflected in the fast growth of China's enterprise Datasphere.

Data is at the heart of this digital world and we are increasingly becoming an information economy. The value is moving to data so that we can create a new world of smarter products, better customer experiences, and self-learning and always improving digital services. In fact, 32% of China's organizations executing on digital transformation initiatives have put data capitalization as the top priority to progress. Data is also the heartbeat of modern user experiences and services built using next-generation technologies such as cognitive, IoT, AI, and machine learning.

Between 2018 and 2025, the use of public cloud storage is poised to increase aggressively. In fact, by 2025, 32% of data storage in China will take place in the public cloud. This growth is partly due to the increasing number of internet users in the country. IDC estimates that in 2017, about 56% of China's population of 1.4 billion used the internet. This is expected to increase by nearly 10% over the next five years – an increase of over 200 million users<sup>1</sup>. This growth not only places pressure to increase networking and cloud infrastructures but also raises the amount of data being created in the China Datasphere.

This unprecedented data growth combined with the pressures of deriving value from data for digital transformation will create imperatives for IT and business organizations in China and across all regions globally over the next decade to develop a fitting data storage, management, and capitalization strategy and drive a new level of engagement with consumers 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 China.

## CREATING DATA IN THE DATASPHERE

---

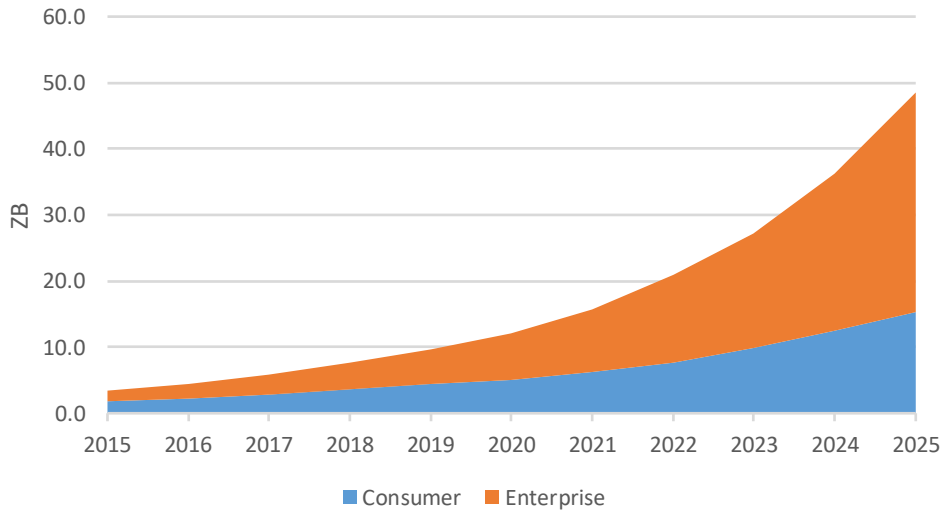
A decade of growth of the China Datasphere is shown in Figure 1, which also shows the changing share of the Datasphere generated by consumers and enterprises. The enterprise Datasphere will grow from 49% of the China Datasphere in 2015 to 69% in 2025.

---

<sup>1</sup> IDC's 4Q18 New Media Market Model

**FIGURE 1**

**China Datasphere: Originator Segmentation, 2015-2025**



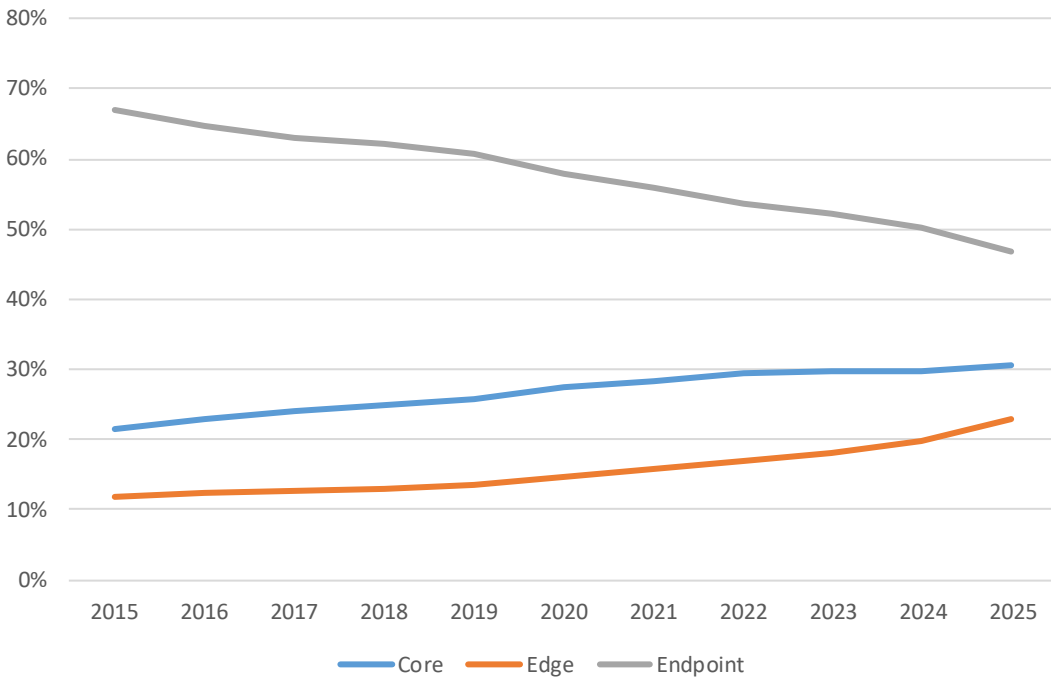
Source: Data Age 2025, sponsored by Seagate, November 2018

Growth of the enterprise share of the Datasphere is driven by several dynamics, including the growth of big data and analytics; the migration of consumer data storage from local devices to the cloud; AI application in healthcare, Smart City, autonomous vehicles, and so forth; the proliferation of applications, edge devices, and IoT sensors that act as systems of engagements with customers and collect data (including the previously mentioned Dazzling Snow initiative); the completion of the migrations of analog to digital TV; and data retention for test and dev, analytics, or compliance purposes.

Figure 2 shows the China 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 that data replicated at the edge will, in turn, be replicated at the core.

## FIGURE 2

### China Datasphere: Location, 2015-2025



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

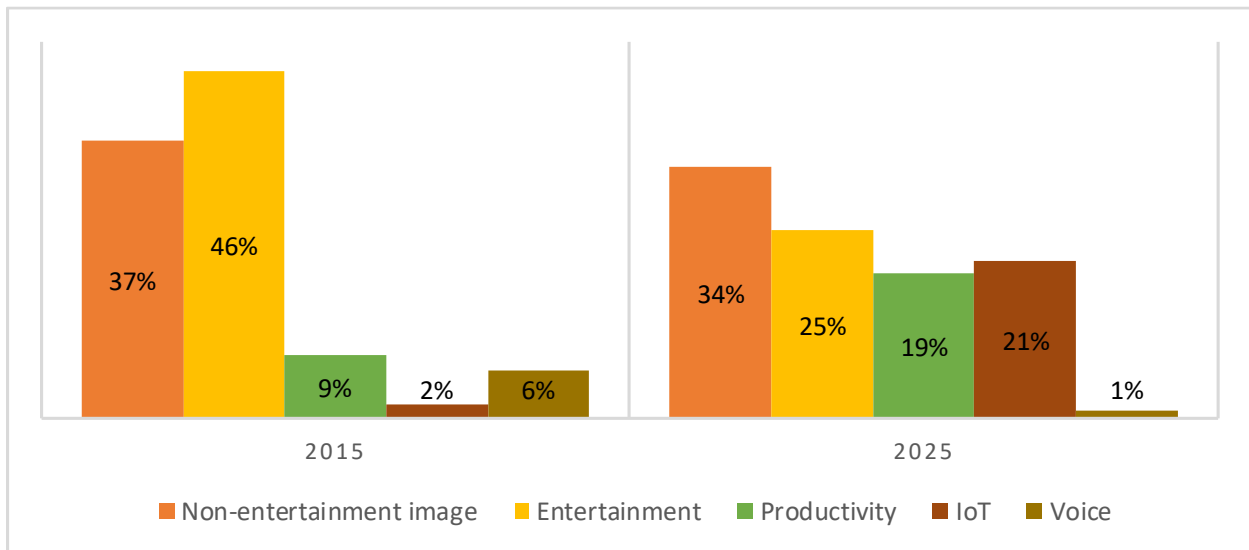
The *type* of data in the China Datasphere will shift in significant ways, as shown in Figure 3. While entertainment data (data from digital TV, online videos, music, and games) will grow by a factor of 7.8 from 2015 to 2025, it will be outpaced by growth in productivity (big data and metadata) and IoT such that it will fall from more than 46% of the Datasphere to a quarter. Growth of the image sector (which includes both non-entertainment- and entertainment-related content) slows as the conversion from analog to digital in media and entertainment winds down.

Non-entertainment image data in China, which includes surveillance, digital cameras and camera phones, scanners, and medical imaging, has a 2015-2025 CAGR of 29% compared with a global rate of 24%. At 29%, however, the non-entertainment image segment in China is still growing slightly lower than the overall China Datasphere, at 31% over the same 10 years.

There are other types of data that present challenges (and opportunities). For example, mobile data on a global level 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, video analytics, and so forth – 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**

**China Datasphere: Data Type Share, 2015 and 2025**



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

One example of growth in productivity data is that of a Shanghai-based airline company. This company built a big data lab to consolidate all aviation and airline-related data, including customer, agency, marketing, quick access recorder (QAR), airplane, and itinerary map data. By analyzing all of this data, the company was able to improve its marketing activity, services satisfaction, and flight cost control and to create more value. This data-driven campaign helped the airline company combine data and 3rd Platform technologies to build a "business-driven big data" platform.

Another example comes from a unified traffic command center use case in East China. This company began to build a new collaborative and dispatch system to cover the entire province in March 2017. The system is based on a cloud platform, which supports functions like intelligent detection, coordination, and statistical analysis, as well as data about video surveillance, mileage, road conditions, and weather. It consolidates 17 traffic controlling units and nearly 50 road centers. One hundred percent of highway vehicle accidents and rescue events are handled on this platform, with more than 300 incidents being handled daily. The result has been a significant improvement in the intelligent management of expressway operations and services.

**Storing Data from the Datasphere**

Much of the data in the Datasphere will not be stored in any permanent way. In fact, the total China installed storage capacity compared with its own Datasphere in 2018 was just over 12%. By 2025, it will be little less than 7%. 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 get recorded over.

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,

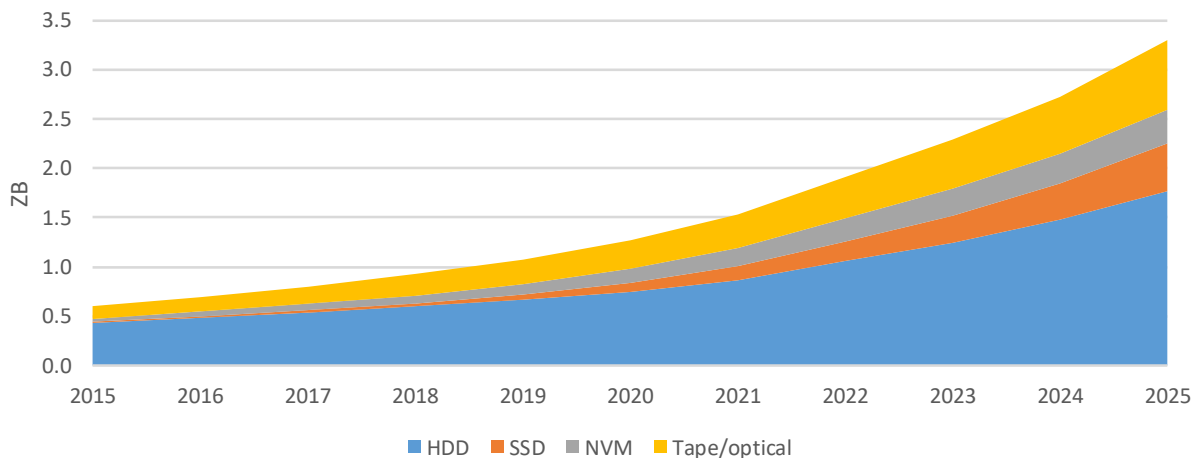
operations, research and development, retention of intellectual property, and financial and employee records. Increasingly, data is leveraged in automation, AI, and IoT. Data can also be sold, creating new data-as-a-service opportunities and revenue streams. Hence companies must manage their data properly.

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, China had 18.3% of the world's 5.0ZB installed base of storage capacity<sup>2</sup>. This share of storage capacity increases to 19.8% in 2025, when the world is expected to have an installed storage capacity of roughly 16.5ZB.

Figure 4 shows growth of China's installed storage capacity that will be available from 2015 to 2025 by storage type. Perhaps surprisingly, tape and optical will continue to have an important role, even as movie viewing and music listening migrate to streaming from DVDs and CDs. Archival storage and backup on tape and optical systems will still be needed for years to come.

## FIGURE 4

### China 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 3% in 2015 to 28% in 2025 of all the storage bytes (excluding public cloud bytes, but including all other enterprise, PC, phone, and server bytes across all media types – HDDs, flash, tape, and optical) that exist in China (see Figure 5).

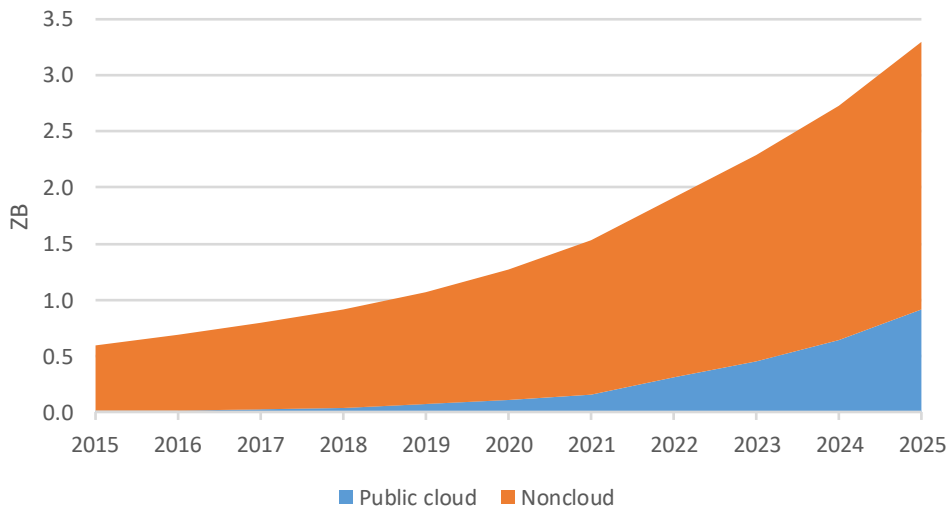
In fact, IDC believes that cloud adoption is entering the next phase and enterprises are moving away from an "all in" or public cloud-only adoption model to a much more nuanced and best-fit approach – making multicloud and hybrid cloud infrastructures a reality. In fact, more than 80% of China's

<sup>2</sup> Installed base of storage capacity is the sum of used and unused bytes across hard disk drives, flash memory, and tape and optical media.

enterprises admit to having multicloud strategies by 2020 but only a handful are confident about their multicloud environments being well designed or orchestrated, meaning that they have complete visibility of data to ensure effective analytics or information governance.

## FIGURE 5

### China 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

China is in the throes of digital transformation – using new technologies and applications to transform its businesses. The 14 times growth of the China Datasphere from 2015 to 2025 by itself will create challenges for enterprises – in management, security, storage, and utilization. As a developing economy, China must also consider rules of compliance, how to transform 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 China 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 16.2 ZB, by 2025.
- **Real-time data:** 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 China and

you get better than 34 times growth in real-time data by 2025, or 28% 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 (where signals or inputs must be handled in real time) into IT organizations as they begin to inherit responsibility for computing – once the duty 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.

In fact, IDC's 2018 China multicloud and storage research on enterprises revealed that the top 4 priorities for organizations are managing data growth, creating information governance strategies, focusing on analytics, and leveraging cloud for data storage, backup, and disaster recovery.

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 China. 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

---

Data growth and the growing value attached to data are changing China's consumer and business landscape region, and data is shaping how consumers, governments, emergency service providers, and businesses work.

Businesses are using data to reach new markets, better serve existing customers, streamline operations, 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 create a competitive edge in digital transformation. One need only look at internet and mobile payment platform giants like WeChat (by Tencent) and Alipay (by Alibaba) and the competitive threat these companies pose to traditional banks. By leveraging the massive amount of data that these companies capture on their customers, these tech giants have the potential to create personalized financial vehicles tailored to the specific behaviors and preferences of their customers. According to Agustin Carstens, head of the Bank for International Settlements, these tech companies "may have better information on customers' spending and lifestyles, which might make it easier to judge the risks of providing a loan."

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.

Take the use case of a China-based rubber manufacturing company as an example. This company created an industrial artificial intelligence brain platform to support its core manufacturing process. Intelligent control of the viscosity rate of rubber helps effectively control the processes of mixing, rolling, extrusion, and vulcanization of rubber to obtain good physical and mechanical properties of tire products. By using big data and analytics to optimize its model iteration of manufacturing process



parameters, the company was able to establish ideal process parameters for various rubber tire types and configurations. At the same time, the index data of the refining process and the multidimensional data chain of natural rubber warehousing tests were used to support the quality control of natural rubber suppliers.

Consumers are building more and deeper connections to digital data and accessing products and services more easily – and at the time and place of their choosing. They are also benefitting from 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 learn over time and provide accurate, personalized experiences without breaching data privacy sentiments.

The Datasphere is large, dynamic, and complex and increasingly intersecting and driving 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. 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, businesses are not just technology or software businesses first, but they are also information businesses first and will need to become data savvy to optimize the opportunities and mitigate risks using technologies and become data thrivers and gain a competitive advantage 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.

