

White Paper

The APJxC Datasphere: Enterprise Acceleration to Meet Demand

Sponsored by: Seagate

David Reinsel
January 2019

Kar Leong Tew

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 Internet of Things (IoT) devices, metadata (vital for analytics, contextualization, 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 Asia/Pacific, including Japan, but excluding China (APJxC) Datasphere is one of the fastest-growing Datasphere regions. It is growing at a 2018-2025 CAGR of 28.3% compared with a worldwide rate of 27.2% and a U.S. growth rate of 23.6% over the same time period. Data created in the APJxC regions will increase from 5.9ZB in 2018 to 33.8ZB in 2025 (18.2% and 19.3% of the Global Datasphere, respectively).

While much of the growth will be driven by video surveillance, signals from IoT devices, and metadata (like other regions), entertainment-related data is growing at a faster pace than most other regions, expanding at a rate of 23% compared with a global average of 20% for 2018-2025. IDC research also shows that the number of online users participating in entertainment-related activities is growing faster than the global rate – 7.2% compared with 4.6% for 2017-2022¹. The demand for connectivity and online content will compel enterprises and governments to invest in traditional enterprise and cloud infrastructures.

These entertainment engagements increasingly happen on the go on mobile devices. Hence the percentage of data in the APJxC Datasphere emanating from or replicated in the edge will increase

¹ IDC's 4Q18 New Media Market Model

from 12% to 20% of the region's total Datasphere – as data is delivered to endpoints and as IoT devices increasingly drive processing and analytics closer to the point of origin of the data itself.

The value of data is increasing as is the realization that we can create a new world of smarter products, better customer experiences, and self-learning and digital services that are continually improving. In fact, in a recent survey of over 400 APJxC enterprises, over 60% had a method of quantifying the value of data. Of those 264 organizations, 43% were evaluating the revenue associated with their data, ahead of the regions surveyed and significantly higher than the 37% average of all regions.

Although not all of this data created will require permanent storage, storage utilization in APJxC will grow from 0.5ZB in 2018 to 2.3ZB in 2025. Between now and then, the APJxC use of public cloud storage is poised to increase aggressively. In fact, by 2025, 58% of data storage will take place in the public cloud compared with 22% in 2018. One of the growth drivers influencing public cloud usage is that, as a region, the APJxC countries have the fastest-growing number of internet users (growing 5.9% on average, nearly 2% faster than the worldwide average), which drives use of the cloud not only by users but also by enterprises as they race to keep up with user demands.

This unprecedented data growth combined with the pressures of deriving value from data for digital transformation will create imperatives for IT and business organizations across all regions over the next decade. Enterprises must develop a fitting data storage and 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

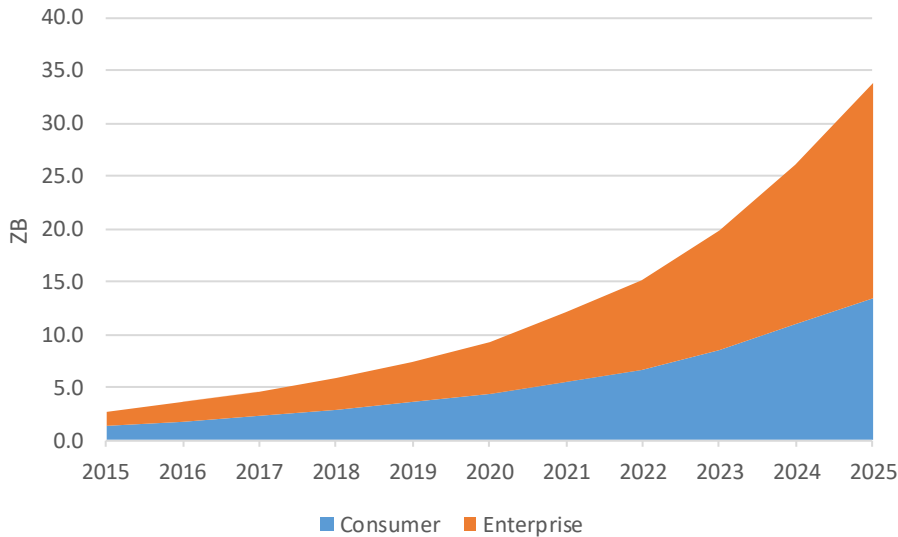
This White Paper is a regional companion document to *The Digitization of the World: from Edge to Core* (IDC #US44413318, October 2018). It summarizes the trends and dynamics of the Datasphere and of data storage as they relate to APJxC.

CREATING DATA IN THE DATASPHERE

A decade of growth of the APJxC 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 50% of the APJxC Datasphere in 2015 to 61% in 2025.

FIGURE 1

APJxC Datasphere: Originator Segmentation, 2015-2025



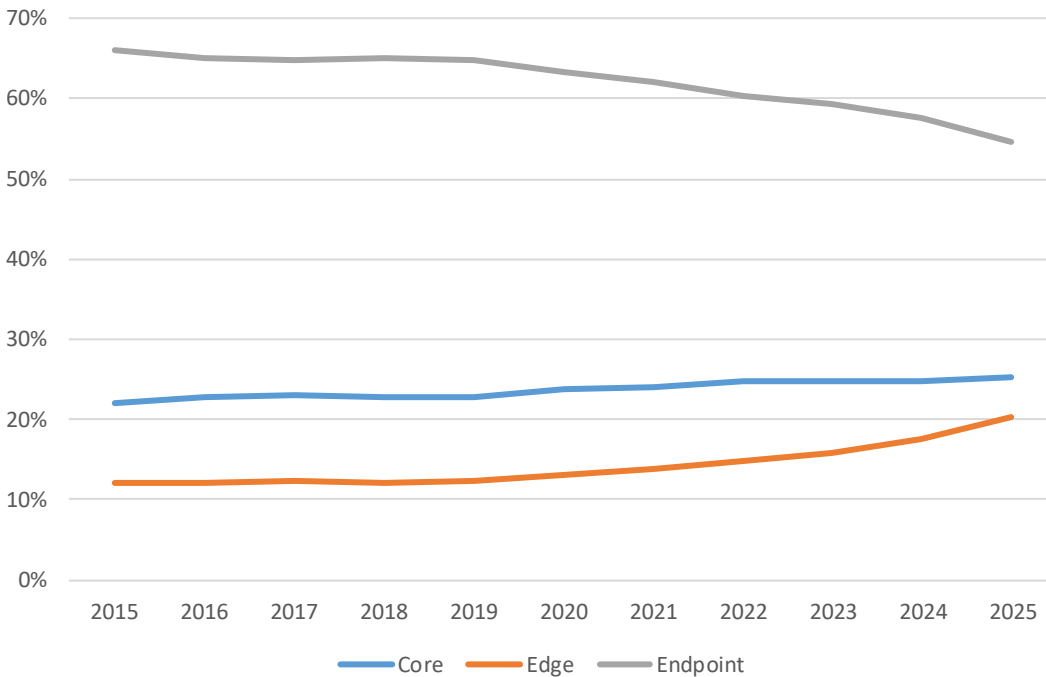
Source: Data Age 2025, sponsored by Seagate, November 2018

Similar to other regions, growth of the enterprise share of the APJxC Datasphere is driven by several dynamics, including the completion of the migrations of analog to digital TV and movies; the growth of big data and analytics; the migration of consumer data storage from local devices to the cloud; the proliferation of applications, edge devices, and IoT sensors that act as systems of engagements with customers and collect data; and data retention for test/dev, analytics, or compliance purposes.

Figure 2 shows the APJxC 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. The edge becomes an important intermediary (between the core and the endpoint) where analytics happen, and results are stored to fuel real-time decisions and latency-sensitive services that are delivered in to personal streams of life or business workflows. Milliseconds matter in a real-time, data-driven world, and many times, the edge will help shorten the delivery of insight by eliminating the need to process data at the core.

FIGURE 2

APJxC Datasphere: Location, 2015-2025



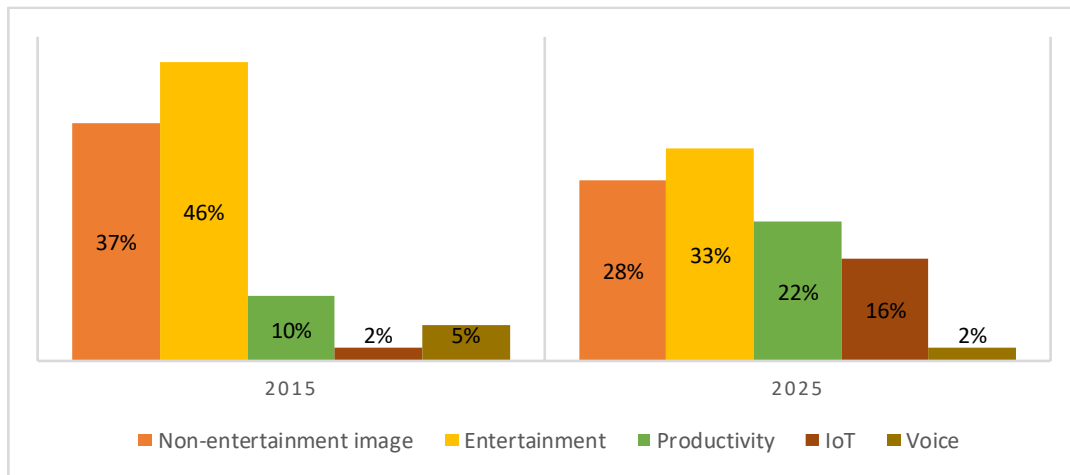
Source: Data Age 2025, sponsored by Seagate, November 2018

The *type* of data in the APJxC Datasphere will shift significantly, as shown in Figure 3. While entertainment data will grow by a factor of 9.2 times from 2015 to 2025, it will be outpaced by growth in productivity (big data and metadata) and IoT. Entertainment-related data will fall from 46% of the Datasphere to less than 33% by 2025. The declining share of image-related content is largely due to the slowdown in converting analog data into digital data because the media and entertainment industry now captures most data digitally. Also compounding the share shifts are the increasing IoT connections and the always-on nature of IoT and machine-to-machine data capture and communication, resulting in tremendous growth through 2025 and beyond.

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

APJxC 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 Tokyo-based Plaid is enabling companies to perform real-time customer data analysis via its tool called Karte. The company introduced its customer tracking service in March 2015 and to date has analyzed more than 2 billion online customers, representing more than \$5 billion in ecommerce. In 2018, the company transformed its Karte tool into a customer experience platform. The newly upgraded tool essentially allows companies to visualize their online customers' behaviors. It allows companies to measure a customer's experience in real time visually and allows them to identify and support those customers that might require help, increasing transaction close rates. The tool also allows company administrators to map out a user's path across a website to refine and reinforce website design elements.

Another data-intensive example is how big data is being used as a tool to address sustainable access to and reduce waste of fresh water in India. Central India has an impending water crisis ahead of it, with up to 40% of the renewable surface water diminishing. Bangalore is also facing a water crisis as its water table's accessible depth has shrunk from 10-12m to 76-91m over the past two decades. Finally, water being wasted due to leaks and pipe breaks adds to the severity of the water shortage.

Various big data initiatives have been implemented to start the war against water crisis in India, including TetherBox Technologies' Elemento Aqua, which leverages real-time sensors to monitor lake oxygen and contamination levels (and responds by pumping oxygen into lakes when necessary); SmarterHomes Technologies, which leverages real-time sensors to detect water leaks and responds by shutting off the water supply within a home when detected; and the Bangalore Water Supply and Sewerage Board, which leverages a data visualization dashboard that monitors real-time data to understand the flow and distribution of water, helping identify leaks and abuse. Data is at the heart of all these big data initiatives being used to help mitigate a water crisis in India.

Storing Data from the Datasphere

Much of the data in the Global Datasphere will not be stored in any permanent way. In fact, the APJxC installed storage capacity compared with the APJxC Datasphere in 2018 was 16%. By 2025, it will be

a little more than 10%. 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. Regardless of how much data in the Datasphere evaporates, the opportunity to store more data persists as more data is created, consumed, and analyzed. The installed base of storage capacity is expected to increase 20% per year on average from 2018 to 2025.

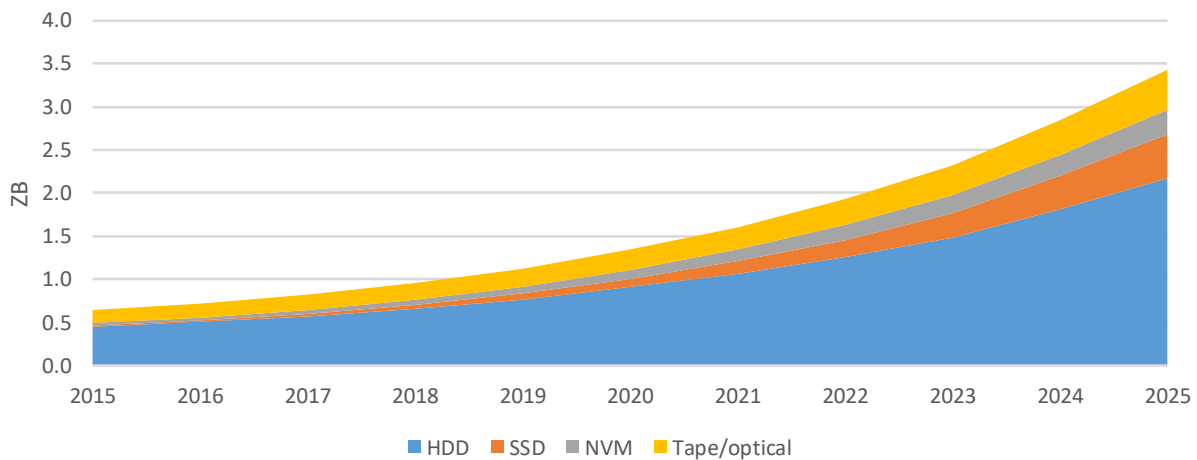
The value of data is increasing, and it is imperative for companies to understand the value of the data that they store. Data is the lifeblood of the data age – it can support customer-facing activities, operations, research and development, retention of intellectual property, and house financial and 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, APJxC countries had 18.9% of the world's 5.0ZB installed base of storage capacity². This share of storage capacity increases to 20.6% in 2025, when the world is expected to have an installed storage capacity of roughly 16.5ZB.

Figure 4 shows growth of installed storage capacity that will be available in the APJxC countries from 2015 to 2025 by storage type. Perhaps surprisingly, tape and optical 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

APJxC Datasphere: Installed Storage by Media Type, 2015-2025



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

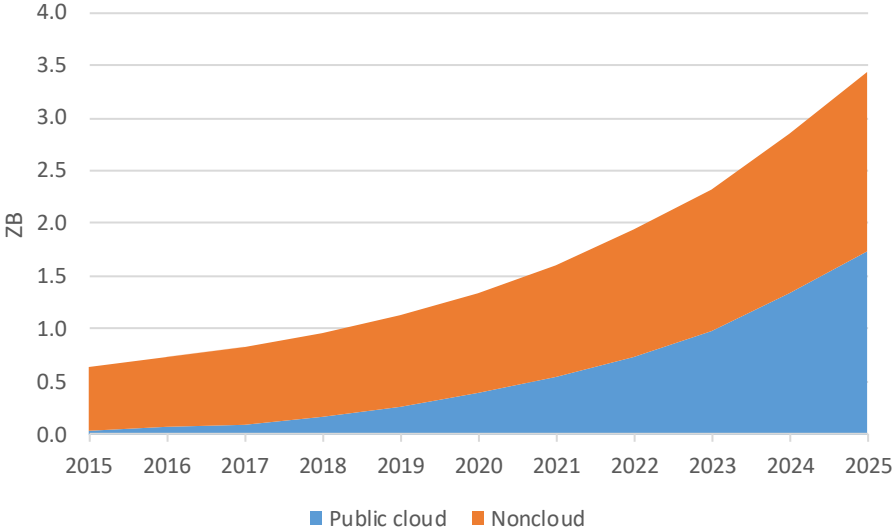
² Installed base of storage capacity is the sum of used and unused bytes across hard disk drives, flash memory, and tape and optical media.

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 6% in 2015 to 51% in 2025 (see Figure 5). In fact, IDC believes that enterprises are entering the next chapter of digital transformation with the proliferation of application deployment models and locations. IDC predicts that, by 2021, over 65% of enterprises in APJxC will use multiple cloud services and platforms, with more than one-third of these organizations having established mechanisms to operate their multicloud environments. Visibility and integrity of the data asset across the multiple clouds will be the focus to ensure effective information governance and management. This won't happen automatically or by chance. Companies must hire skilled data scientists and provide existing employees with proper training.

AIS, Thailand's largest mobile phone operator, established its Analytics Center of Excellence, collaborating with technology partners and universities to develop an intensive training program to groom employees at the AIS headquarters and regional offices to become data engineers, data analysts, and data scientists. Traditional business intelligence tools were replaced by modern analytics software, while the adoption of big data and analytics and machine learning was accelerated through workshops and boot camps, which tackled real-life business cases including mobile and fixed broadband churn prediction, upselling and cross-selling different products, and customer acquisition. Within one year, AIS was able to demonstrate how an organization with 13,000 employees can become fully data driven. The number of data scientists increased from 0 to 100 through internal development rather than recruiting new employees. The most impressive change in the organization is that "now, data analytics has become an essential part of AIS' business, and data is being loved by everyone in AIS."

FIGURE 5

APJxC 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 13 times growth of the APJxC Datasphere from 2015 to 2025 by itself will create challenges for enterprises – in management, security, storage, and utilization. The APJxC countries are in the throes of digital transformation – using new technologies and applications to transform their businesses.

The developed countries in the APJxC region must deal with rules of compliance, legacy information technology and organizations, and global competition. The emerging countries have an opportunity to integrate privacy and compliance technology from the start.

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, the amount of data requiring some kind of security protection increases 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 APJxC regions that *needs* protection will grow faster than its own Datasphere between now and 2025 and account for a full one-third of its Datasphere, or 11.3ZB, by 2025.
- **Real-time data:** Again on a global basis, the percentage of real-time data in the Datasphere will grow from 12% in 2015 to 29% in 2025, driven by growth of IoT. Apply these percentages to APJxC and real-time data increases 30 times by 2025, or 19% 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.

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 APJxC region. 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 the APJxC consumer and business landscape, 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 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.

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

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. IDC believes that leading-edge organizations that put data at the heart of their transformation will define the future market dynamics of their sectors as well as the adjacent sectors.

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. Businesses must become data savvy to optimize the opportunities and mitigate risks using technologies and become data thrivers to 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.

