# Priority Boarding for SAP Data Hub: Dell Technologies and Red Hat Partner to Feed the Intelligent Core

Introduction: Red Hat and Dell Technologies' Approach to Creating a Stable Foundation for the Intelligent Core for SAP Data Hub

Digital transformation programs are increasingly dependent on a strong "intelligent core" — an IT stack able to extract insights and trigger actions from a mix of structured and unstructured data. For example, a freight cargo insurer that wants to price its contracts based on weather may want to connect its internal sales and risk management systems to real-time data from weather forecasters.

Figure 1
The Role of the Intelligent Core in a Digital Platform
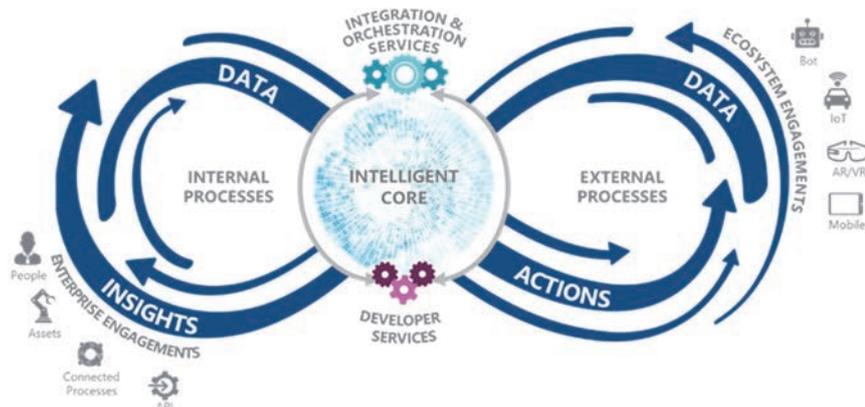
Source: IDC, 2019

Setting up and monitoring data flows across a myriad of sources (from Hadoop to sensor data) is the goal of SAP Data Hub, a rapidly growing container-based data orchestration solution developed by SAP.

Examples of enterprise use cases for SAP Data Hub include integrating real-time data from IoT sensors with traditional data warehouses and databases or managing web log flows into Big Data pools. SAP developers and applications architects are looking to generate value from data and not spend time implementing and managing the infrastructure foundation. Many will do this on-premises, but the underlying infrastructure foundation has to be rock solid, optimized, and efficiently managed.

IDC's view is that pre-tested, preconfigured architectures that ensure container management compatibility, storage performance, and scalability often make the difference between success or failure in such projects. This requires an ecosystem of technology partnerships that run deep and offer a strong, battle-hardened support structure to ensure return on investment over the lifespan of the solution.

Pre-tested architectures that ensure container management compatibility, storage performance, and ability to scale make the difference between success and failure in SAP Data Hub projects.

The end goal of SAP Data Hub is feeding the intelligent core with extremely rich datasets to generate unedited insight.

In This Partner Spotlight

This IDC Partner Spotlight describes the multipartner architecture proposition that Red Hat and Dell Technologies have developed around a tested reference architecture for SAP Data Hub. The architecture includes Dell Technologies' compute, storage, and integration capabilities, combined with Red Hat's OpenShift and CEPH storage software. It is aimed at simplifying and shortening deployment of SAP Data Hub in the enterprise customer base. This report is targeted at IT decision makers involved in the purchase and implementation of SAP Data Hub, including data scientists, enterprise architects, and high-level IT infrastructure managers.

Situation Overview

Business programs linked to digital transformation are typically centered around very specific use cases such as predictive maintenance in physical supply chains or omni-channel operations in marketing. In those scenarios, combinations of technology are deployed to enhance or even revolutionize the way processes are optimized and use cases delivered.

Process improvements are impossible without real-time insights from internal assets and correlated external data sources. IDC believes that virtually all new use cases require an "intelligent core" — an IT stack able to extract intelligence and actions from a mix of structured and unstructured data. To this end, IDC is seeing accelerated growth in investments in this area: an estimated $90 billion has been spent globally in 2019 on hardware, software, and cloud services supporting analytics workloads and this will grow at 13% CAGR into 2023 (source: *IDC WW Semiannual Big Data and Analytics Spending Guide, 2H18*).

SAP Data Hub — Context and Usage Scenarios

Demand for intelligent core capabilities is felt strongly across the SAP customer base. SAP has responded to this need from its customers with a series of investments and acquisitions. Today, SAP brings to the market a rich on- and off-premises stack including SaaS applications targeting business buyers (e.g., SuccessFactors), SAP S/4HANA business suite, SAP HANA database, developer tools, and a vast analytics and machine-learning portfolio.

SAP Data Hub is a key component of this latter toolset. The data orchestration software solution was first announced by SAP in 2017, and since then SAP has launched multiple updates (most recently in version 2.6 in August 2019). The end goal of SAP Data Hub is to provide architects, data scientists, and database administrators with an on-premises tool to integrate data repositories and data flows from multiple sources — feeding the intelligent core with extremely rich datasets to generate unedited insight.

Unstructured "Big Data" repositories (e.g., Hadoop files), raw IoT sensor feeds, web logs, cloud storage formats (e.g., AWS S3), as well as structured datasets sitting in SQL systems like SAP HANA, can be ingested through SAP Data Hub.

Using the graphical user interface (GUI) dashboard or a variety of coding options, the data expert can then:

1. Gain full visibility of the data landscape on- and off-premises.

2. Manage and modify dataflows and dependencies between datasets.

3. Create new insight and data models using templated "operators" with ready-to-use tasks.

4. Manage policies regarding data access, hierarchy, etc.

Key usage scenarios listed by SAP include:

- Integration of IoT sensor data into back-end systems (e.g., warehouse environmental data connected into SAP supply chain management applications).

- Streamlining of multisource data ingestion for machine-learning environments (e.g., for predicting customer churn in a B2C business).

- Policy management and governance around data pipelines.

IDC has noted that SAP Data Hub customer references and use case examples have been increasing rapidly since the second half of 2018. European SAP customers publicly referencing deployment include PWC Germany (tax analytics), ATB Financials, Kaeser Compressors (connecting IoT data to service management for improved ticket handling), and Hartmann Group (embedding web data into sales forecasting for healthcare products).

Technology and Procurement Implications of SAP Data Hub

From a technology standpoint, SAP Data Hub is built as a package that runs directly on Kubernetes container management systems installed on physical systems. SAP Data Hub is therefore highly dependent on the stability of the Kubernetes container environment underneath (see Figure 2).

Figure 2
Technology Stack Requirements for SAP Data Hub

Source: SAP, IDC, 2019

SAP Data Hub is highly dependent on the stability of the Kubernetes container environment underneath.
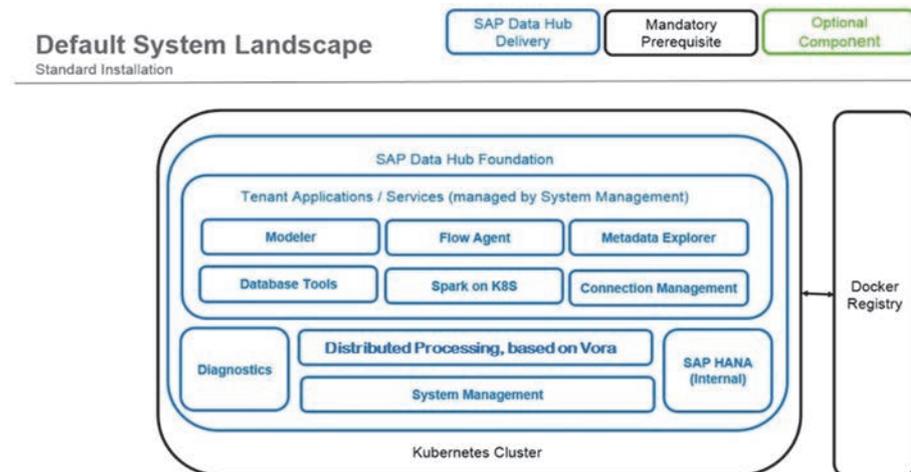
If Kubernetes is the first "landing dock" for SAP Data Hub, the underlying infrastructure (including elastic storage and compute management functions) represents the "big harbor" that the containerized application needs to run on. IDC believes the ability to deliver a performant and mostly automated infrastructure is a must for any SAP Data Hub deployment.

Implementation Process for SAP Data Hub

The buying process for SAP Data Hub includes multiple personas, each with a different need and role to play. Figure 3 displays a typical implementation process flow, including key internal and external stakeholders.

Figure 3

Illustrative SAP Data Hub Implementation Process and Related Challenges
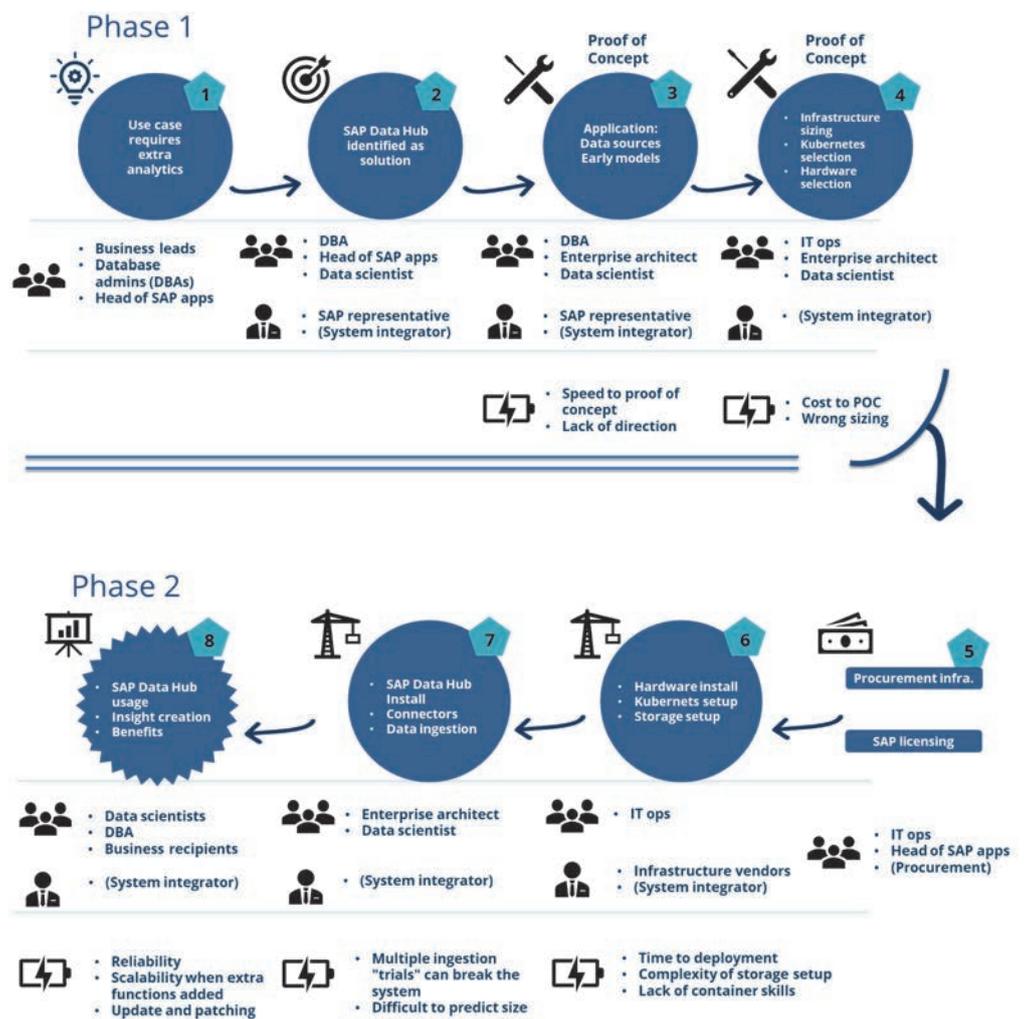
Source: IDC, 2019; stakeholders in ( ) are potentially in play

SAP Data Hub Needs a "Priority Boarding" Lane

SAP Data Hub is a powerful tool to create new insights from a dizzying array of data repositories, regardless of their format and location. However, IDC surveys show that, after data privacy concerns, the most common challenge when implementing AI and intelligence systems is the lack of necessary skills or resources

IDC maintains that infrastructure configuration is one of the major hurdles to SAP Data Hub deployments, impacting multiple stages of the process.

to manage or even simply deploy the operating stack (*IDC European Cognitive Solution Survey*, n = 350). IDC maintains that infrastructure configuration is one of the major hurdles, impacting Steps 4, 6, 7, and 8 of the implementation process for SAP Data Hub, as described in Figure 3.

In particular:

- Infrastructure must be fine-tuned from storage to container management to support the extra pressure of mixed data sources.

- Not all enterprise buyers have a full command of Docker containers — this is especially true for the stakeholders impacted by SAP Data Hub who are not classic DevOps.

- Predicting capacity growth at ramp-up, both at proof-of-concept and production stage, is key to avoiding hiccups.

- Physical installation and standing up the container environment can become a long trial-and-error process in a "do-it-yourself" environment.

In IDC's opinion, SAP customers looking to overcome some of these challenges and reap the benefits of an SAP Data Hub solution should put in place a "priority boarding" lane for SAP Data Hub.

This will enable customers to streamline and bullet-proof some of the more operational steps described, enabling data scientists and business recipients to "board the plane" quickly and dedicate their full attention to creating strong models (Step 3) and generating business benefits (Step 8).

Addressing the Market Needs — The Red Hat and Dell Technologies Reference Architecture for SAP Data Hub

Dell Technologies and Red Hat have an established methodology for platforming SAP workloads. Realizing the market need for "priority boarding" to SAP Data Hub, they came together in 2019 and now jointly offer a certified Reference Architecture for SAP Data Hub. The Reference Architecture is built on Red Hat OpenShift Container Platform (OCP), Red Hat Enterprise Linux 7, and CEPH Storage, combined and fine-tuned on compute-intensive and storage-intensive Dell EMC PowerEdge servers.

By leveraging the reference architecture approach, customers can fast-track infrastructure configuration and procurement processes. This means:

- **Certainty of performance.** Red Hat OpenShift Container Platform (where SAP Data Hub containers run) and CEPH storage (where data is ingested) have been stress-tested by Dell Technologies and Red Hat on those specific configurations. The partnership makes it easier for buyers to get support from either of the vendors, as well as SAP.

- **Reduced time to deployment.** Sizing is simplified, removing some headaches for IT operations teams. Red Hat subscriptions may be purchased through Dell Technologies when procuring hardware. In

Organizations should put in place a "priority boarding" lane for SAP Data Hub to get data scientists and business recipients "on the plane" quickly to create strong use cases and models.

addition, if required, Dell Technologies can deliver racked-and-stacked hardware to the customer within three weeks.

- **Clarity in infrastructure cost.** With a pre-tested approach to sizing, buyers can easily predict how much the infrastructure software and hardware will cost.

In other words, the Reference Architecture acts as a catalyst in multiple phases of the implementation. Architects and data scientists can use it as a blueprint to select Kubernetes versions and size hardware. After that point, procurement and IT operations can contact Dell Technologies (directly or through a potential service partner) to validate or get a second opinion on the sizing, which is invaluable and a recommended step (Step 4).

Procurement of infrastructure hardware and software (Step 5) is quicker when using the Reference Architecture, whereas an SAP Data Hub license purchase must happen directly with SAP or an SAP reseller. Hardware is delivered in the correct configuration and cabled in the datacenter, minimizing configuration time (Step 6). Red Hat OpenShift Container Platform and CEPH can be safely installed on the hardware, before the SAP Data Hub container is deployed. Having a strong, tested foundation in place helps minimize the pressure on the system once the data scientists start cranking up the data flows (Step 7).

The Market Opportunity: Segments and Buyers

IDC believes typical buyer segments interested in this solution are large midmarket companies (1,000+ employees) or multinational enterprises that have a solid SAP installed base. These organizations are very likely to have already deployed SAP HANA databases and to have (or be in the process of building) strong analytics and data science expertise. Roles that will be involved in the buying process for this type of architecture include enterprise architects, data scientists, and IT operations teams. SAP Data Hub is a container-based application, but it sits "deep" in the data management stack, so it is unlikely that typical front-end DevOps teams will be major stakeholders in its deployment.

It is important to note that this is a Reference Architecture, and not a set of fixed configurations. Due to the variety of SAP Data Hub implementation scenarios, sizes, and possible data sources, it is virtually impossible to have click-to-buy hardware and software configurations. IDC believes reference architectures enabling adaptation to customer needs are the best way to shorten deployment on-premises, while lowering requirements in terms of in-house hardware skills.

Unique Elements of the Reference Architecture

IDC maintains that this Reference Architecture has a number of unique elements:

- **Containers on a tested base.** SAP's own application development is quickly standardizing on Docker containers. Having Red Hat OpenShift as the foundation ensures enterprise-grade support for Kubernetes as well as deep engineering partnership with SAP. For example, some containers used by SAP Data Hub have special requirements with regards to privileged user permissions. SAP and Red Hat work jointly on performing validation of the

SAP's own application development is quickly standardizing on Docker containers — forcing buyers to plan enterprise-grade Kubernetes soon.

latest software versions to ensure configuration best practices are available to customers.

- **Storage finesse.** Due to data size, a typical application bottleneck of SAP Data Hub is storage bandwidth and throughput. Dell Technologies has a lot to bring to the table here, and the storage-intensive servers and fast network (that are part of the Reference Architecture) combine with Red Hat CEPH open source storage software to tackle that.

- **Steady flow of compute.** Many SAP Data Hub environments shift dynamically between different types of data sources. This puts a heavy strain on memory and CPU allocation, as the stack needs to dynamically move from memory-heavy to throughput-heavy workloads. Within Red Hat OpenShift, CPU and memory bottlenecks can be addressed by the built-in dynamic resource management capabilities. This is augmented by the broad range of options for both types of resources in the Dell PowerEdge server line, helping to achieve a steady flow.

- **Management automation to simplify operations.** Red Hat OpenShift Container Platform offers a large set of features to automate steps in container management, including tasks such as patching, cloning containers, and templating their hardware footprint.

The Dell Technologies and Red Hat Strategic Partnership

The Red Hat and Dell Technologies Reference Architecture for SAP Data Hub is the first reference architecture available on the market for container-based SAP applications, which is a testament to the speed of innovation that Red Hat and Dell Technologies ignited together.

The two vendors are market leaders in their respective datacenter and enterprise open source software fields. They also have a broader, long-standing partnership including:

- A co-innovation lab that certifies and tests each other's products.

- Long-standing coordination on Red Hat OpenShift. Dell Technologies was one of the early datacenter hardware providers to build expertise on Red Hat OpenShift, dating back to 2014.

- Dell Technologies' proactive approach to open source communities. Dell Technologies' software engineering team is closely cooperating with open source communities, especially around driver optimization. As a result, Dell hardware can be fully exploited "upstream" in the open source communities. Also, all hardware enhancements can be leveraged from day one in the newest Red Hat Enterprise Linux releases.

Face-to-face collaboration enabled Dell and Red Hat to quickly stand up the Reference Architecture and to create an environment where customers themselves can test it.

It is clear to IDC that Dell Technologies, Red Hat, and SAP have demonstrated their ability to collaborate and co-create to provide a very specific outcome to a specific set of customers' requirements, in a very short time frame.

In November 2018, the Dell Technologies and Red Hat teams based in Walldorf decided to tackle the need to fast-start SAP Data Hub for their customers. Involving

the SAP product teams, they quickly set up a working environment in the Dell Technologies Global SAP Center of Excellence in Walldorf. This face-to-face collaboration was crucial, as it enabled the teams to rapidly stand up and document the Reference Architecture and create an environment where customers themselves could run proofs of concept of the architecture.

During the project, the Dell Technologies team was mainly responsible for hardware housing, providing and designing the hardware and network architecture, and setting up the operating system. The Red Hat team performed the Red Hat OpenShift and CEPH setup, fine-tuning container management settings and the "landing" of the SAP Data Hub application, provided by SAP.

The project moved fast, and the teams were able to complete the infrastructure deployment in March 2019, before Red Hat proceeded to deploy software components, including Red Hat OpenShift and CEPH. As part of the project, SAP provided some examples of applications to test the Reference Architecture. Dell Technologies and Red Hat settled on a heavy workload around facial recognition to stress-test the system. This is the workload currently used to showcase the Reference Architecture.

The final significant milestone was making the Reference Architecture public in May 2019, through a demo of the solution at SAP SAPPHIRE Now — SAP's annual customer and partner conference. Subsequently, Dell Technologies and Red Hat jointly started pushing an ambitious go-to-market motion to promote the Reference Architecture, showcasing it at SAP TechEd in October 2019 and embarking on a training schedule dedicated to system integrators and delivery service partners in EMEA.

Future Outlook

IDC believes the Reference Architecture for SAP Data Hub is a very strong base for a future road map of cooperation between the two suppliers around the SAP ecosystem. In the short term, IDC expects continuous updates to the Reference Architecture following advancements of SAP Data Hub. As implementation partners and customers deploy into the field, the workload repository will continue to grow and so will the competitive advantage for the partnering suppliers.

In the mid term, IDC maintains that other container-based SAP applications are likely targets for new solution sets, as the vendors continuously ramp up their partnership. Future areas of cooperation are likely to include SAP Edge computing workloads (a strength in the Dell Technologies portfolio) and orchestrated hybrid deployments of SAP analytics and machine-learning tools based on Kubernetes (a pillar to Red Hat's OpenShift strategy).

Challenges and Opportunities for Dell Technologies and Red Hat

For Dell Technologies and Red Hat to exploit the full potential of the Reference Architecture, they will have to overcome two tactical challenges:

- **The purchasing flow** for infrastructure (hardware and software) and SAP Data Hub itself is not unified. This does not impede market success, as SAP HANA growth over the past few years has confirmed. However, it will

The Reference Architecture for SAP Data Hub is a very strong base for a future road map of cooperation between the two suppliers around the SAP ecosystem.

require the partnering vendors to properly prepare customers and potentially recommend implementation partners.

- **Broad range of use cases requires sustained investment** in engineering skills to update the Reference Architecture. After a quick barrage of releases, SAP Data Hub is now seeing ever expanding use case scenarios. Red Hat and Dell Technologies have a very deep bench of engineering firepower — a lot of it will be required to keep up with customer needs for adjustments on the Reference Architecture.

On the positive side, opportunities for growth around SAP Data Hub are large, as customer demand ramps up:

- IDC is seeing **strong demand** from SAP customers for SAP Data Hub design, implementation, and support from system integrators and other services companies (e.g., business consultancies). The providers should be able to see quick growth driven by partners in large accounts needing implementation support.

- **SAP S/4HANA migration** will provide an opportunity boost. A growing number of SAP customers sitting on very old ERP versions are kicking off two- to three-year plans to migrate to the latest SAP S/4HANA suite on- or off-premises. These programs will lead to SAP HANA database deployments and at the same time provide a trigger to revise or set up an analytics strategy for the SAP landscape — fueling demand for SAP Data Hub through 2021 and beyond.

SAP S/4HANA migration programs will trigger buyers to revise or set up an analytics strategy for their SAP landscape — fueling demand for SAP Data Hub through 2021.

Advice for End Users

IDC advises companies with plans to add SAP Data Hub to their on-premises IT stack to have a "priority boarding" strategy from the start. IDC recommends that companies consider the Reference Architecture approach as a way to shorten time to market and shield themselves from the unpredictability of SAP Data Hub workloads. When assessing infrastructure vendors to support this workload, key elements to probe include:

- Depth of partnerships and history of joint testing.

- Intellectual property in Kubernetes.

- Ability to bring relevant storage and networking intellectual property to the table.

- Experience and willingness to provide fact-based feedback to buyer sizing requests.

- Ability to industrialize packaging and delivery of the hardware stack to remove friction.

Conclusion

With products such as SAP Data Hub, SAP is clearly expanding its solution set to cover the intelligent core of a digital platform, moving beyond pure enterprise applications. SAP Data Hub is the precursor of this expanding solution set, increasingly developed on a container basis. It will require buyers to implement a strong, well-tested infrastructure foundation to "land" it safely. Tested, industrialized approaches such as the Red Hat and Dell Technologies Reference Architecture will go a long way in easing deployment and quickening time to value for businesses across various sectors.

About Red Hat

Red Hat is a global provider of enterprise open source software and platform solutions, using a community-powered approach to deliver reliable and high-performing Red Hat Enterprise Linux (RHEL), hybrid cloud, container and Kubernetes technologies, as well as automation and management tools.

About Dell Technologies

Dell Technologies is one of the largest technology companies in the world, with revenue of $90.6 billion and 157,000 employees in the latest full fiscal year (FY19, which ended February 1, 2019). The company operates three major divisions (Dell Client Solutions, Dell EMC Infrastructure Solutions, and VMware) and owns additional brands in the cloud, infrastructure software, and security space. Dell Technologies was by far the largest provider of enterprise storage systems in CY2Q19 globally with a 30% market share and tied at number 1 in the server market in CY2Q19 with a 19% market share, according to IDC's WW Server and Enterprise Storage Trackers.

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.

IDC UK

5th Floor, Ealing Cross,
85 Uxbridge Road
London
W5 5TH, United Kingdom
44.208.987.7100
Twitter: @IDC
idc-community.com
www.idc.com