White Paper

Mission-Critical Infrastructure for the Data-Driven Enterprise

Sponsored by: Hewlett Packard Enterprise and Intel Corp.

Peter Rutten Ashish Nadkarni
April 2019

EXECUTIVE SUMMARY

The role played by modern analytics environments at data-driven enterprises is clear. By leveraging Big Data and analytics (BDA), lines of business can analyze growing amounts of data from the core to the edge. BDA helps modern enterprises compete more effectively in the digital economy via the analysis of data coming from core enterprise applications and emerging initiatives such as the Internet of Things, robotics, next-generation security, and next-generation supply chain automation.

IDC predicts that IT spend on Big Data and analytics will reach $274.3 billion by 2022. Much of this spend is going toward modernizing the application landscape and, therefore, the IT infrastructure that hosts these applications. While the move to cloud continues, IDC is, at the same time, seeing a return to on-premise on private cloud, referred to as "repatriation," for transactional and analytics applications. From a data perspective, a move to a future-ready real-time enterprise includes the converging of business-centric transaction processing and data-centric analytics systems. To speed up analytics, businesses are deploying in-memory and memory-centric databases, infusing data analytics platforms with high-performance technologies, and using a highly available and secure conduit for data movement between the various application tiers.

The HPE Superdome Flex server is designed to host critical enterprise workloads with evolving memory-centric demands. Specific examples include conventional Oracle and Microsoft SQL Server and in-memory database platforms such as SAP HANA, Oracle Database In-Memory, and Microsoft SQL Server with in-memory capabilities, as well as in-memory high-performance computing (HPC) and artificial intelligence (AI) applications. In addition, because of its high levels of reliability, it is well suited for Unix-to-Linux migrations. When deployed in mission-critical environments, these applications benefit greatly from the near-linear compute, memory, and I/O scalability; extreme availability; and simplified management capabilities of Superdome Flex.

IDC believes that Superdome Flex sets a high standard for mission-critical servers for data-driven enterprises and is worthy of consideration by firms embarking on a journey to modernize their applications and infrastructure and, crucially, to unlock the value of their data in a timely manner.

SITUATION OVERVIEW

Modern analytics environments provide the crucial underpinning for firms transforming themselves into data-driven enterprises. Big Data and analytics enable modern enterprises to compete more effectively in the digital economy via analysis of data coming from core enterprise applications and emerging initiatives such as the Internet of Things, robotics, next-generation security, and next-generation supply chain automation. As firms seek to create and deliver digital offerings and experiences, insights obtained from data are paramount in making key business decisions.

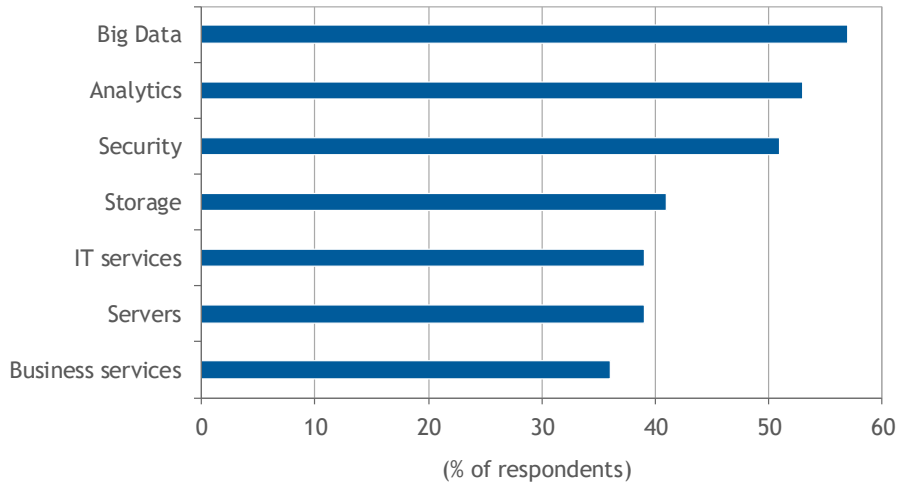
Data is the new basis of competitive advantage. Big Data and analytics made their way to the top of the technology investments list in 2017 and continue to be areas of increased spend (see Figures 1-3). IDC finds that businesses are realizing that accelerating analytics and, ultimately, unlocking the value of data in real time (i.e., real-time Big Data and analytics) are crucial to their ability to lead in the digital economy (see *IDC Survey Spotlight: Big Data and Analytics Go Mainstream,* IDC #US42388517, March 2017). Furthermore:

- As data continues to pervade organizations at all levels at an ever-increasing pace, organizations are challenged to handle the volume, velocity, and veracity of data as leadership strives to derive value from the data and drive business impact in a real-time fashion.
- Generating data intelligence requires the analysis of vast quantities of diverse data, either structured or unstructured and generated by humans or by machines, to uncover patterns and pursue breakthrough ideas.
- AI is the next phase of Big Data and analytics with AI training and inferencing becoming an integrated aspect of organizations' analytical capabilities.

FIGURE 1

Technology Spending, 2017

Q. For each of the technologies listed, do you expect your organization's 2017 spending to increase compared with 2016?



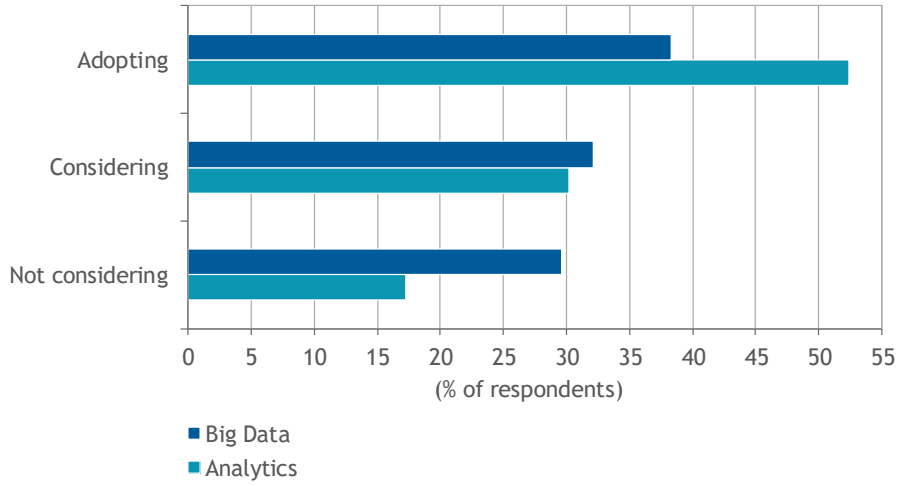
n = 3,602

Source: IDC's *Industry IT and Communications Survey,* April 2017

FIGURE 2

Big Data and Analytics Adoption

Q. *At what stage is your organization today in the deployment of each of the following?*



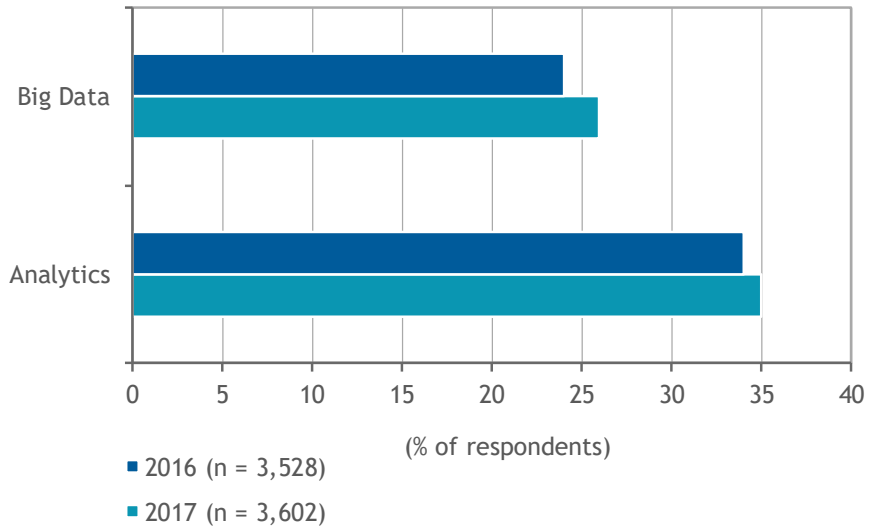
n = 3,602

Source: IDC's *Industry IT and Communications Survey,* April 2017

FIGURE 3

Big Data and Analytics Priority, 2016 and 2017

Q. *Thinking about your organization's top priorities in the coming year, please rank the following initiatives in order of importance from 1 to 6, where 1 is most important and 6 is least important to your organization. (Big Data, analytics)*



Base = percentage of respondents who ranked Big Data and analytics 1 or 2

Source: IDC's *Industry IT and Communications Survey,* 2016 and April 2017

Unlocking the intelligence from data in real time requires a modern application and data management environment. The IT infrastructure that hosts these applications and data management platforms serves as a critical foundation layer. The move to a real-time enterprise includes:

- Converging business-centric transaction processing and data-centric analytics systems to increase the quality and timeliness of insight (i.e., systems of record, engagement, and insight)
- Deploying in-memory databases for low-latency response times as part of the application environment
- Infusing data analytics platforms with high-performance technologies to optimize application performance for large data sets
- Using a highly available and secure conduit for data movement between the various application tiers
- Implementing an appropriate data persistence tier that can support the storing, securing, and fast access of rapidly changing data sets
- Preparing for increasingly AI-infused applications and the related data movement and data processing requirements, for example, by introducing hardware acceleration

The IDC research mentioned previously shows that Big Data and analytics initiatives will continue to become more mainstream across many different applications. Using BDA to optimize business processes was the most popular choice among respondents, followed closely by helping improve security. These three focus areas have continued to be top priorities into 2019. Customer-facing activities and product life-cycle activities were the next most popular uses for BDA within organizations. 84% of respondents have some BDA initiatives implemented within their organizations. These findings support the notion that BDA has hit mainstream adoption and will continue to grow strongly.

IDC believes that to support business agility needs as well as to manage the resources and skills shortage, more and more businesses will choose to deploy enterprise infrastructure platforms that are designed specifically for Big Data and analytics.

The Role of Mission-Critical Platforms for Modern Data-Centric Applications

Firms are increasingly deploying x86-based mission-critical platforms that scale up to optimize their BDA-centric applications and IT infrastructure transformation. IDC believes that there continues to be a misconception in the market about the merits of scale-up server infrastructure.

The Benefits of Scaling Up

When web infrastructure, collaborative workloads, and application development burst onto the scene some 20 years ago, scaling horizontally became de rigueur. Next, virtualization and cloudification caused the scale-out paradigm to become even more dominant. Along the way, scale-up systems became somewhat misunderstood, even as they were aggressively being modernized.

Business processing, decision support, and Big Data and analytics have never fared well on horizontally scaled environments. These are demanding workloads that require maximum resources to process multiple terabytes of data. And when these resources — lots of processors that are close together and lots of flat RAM that is globally addressable for in-memory computing — are packaged in a single system, the benefits, compared with scale-out environments, are significant.

The large memory footprints of scale-up systems allow for large and growing databases to be completely held in-memory, eliminating the latencies of disk access. Latency is also much reduced because of the use of interconnects that allow for dynamically scaling rather than the complex and extensive networks needed to connect nodes in a scale-out environment. Power consumption and cooling costs are significantly lower, as are software licensing costs.

Scale-up systems are also suitable for consolidation projects as they are easier to implement and more efficient to manage and operate than scale-out clusters. They take up a smaller footprint and provide greater reliability and availability (see the Extreme Availability section). In terms of economics, many of today's scale-up systems are nothing like legacy scale-up systems — they leverage the same commodity components (memory, processors, and storage) that scale-out servers are built with rather than the expensive and proprietary components from the past. The idea that scale-up systems are too costly is simply no longer valid, especially if they are available on a consumption model basis, as many are.

Fault Tolerance and Next-Generation Capabilities

As noted previously, modern scale-up systems leverage commodity hardware, and one trend in scale-up platforms has been standardization on x86. This standardization on x86 for modern scale-up systems has led to another misconception, which is the idea that x86-based platforms cannot be as reliable as the Unix-based systems of the past. This, too, is a myth. In the past five years, IDC has seen scale-up x86-based servers evolve to the highest levels of availability, meeting service levels that are demanded by mission-critical workloads for the data-driven enterprise. They deliver extreme reliability, availability, and serviceability (RAS), even reaching availability level 4 (AL4) in IDC's High Availability framework.

IDC classifies servers in four levels of availability, with the highest being availability level 4, or "fault tolerance." At this level, the combination of multiple hardware and software components allows a near-instantaneous failover to alternate hardware/software resources so that business processing continues as before without interruption. In short, modern x86-based scale-up platforms have joined the ranks of the mainframe-like fault-tolerant category. This trend is in sync with the increasing desire for zero downtime in today's always-on world.

At the same time, IDC noted that fault-tolerant platforms have been becoming increasingly suitable for the modern datacenter with capabilities such as mobile apps, cloud, APIs, open source software, and next-generation application development.

A Move to the Cloud — And Back

Cloud adoption has been accelerating, and with IaaS and PaaS maturing, businesses are moving some of their critical business applications to IaaS and PaaS environments. While public cloud IaaS and PaaS continue to be popular, businesses are also investing more in private cloud (both on-premise and off-premise) as this allows them to manage a dedicated system for their mission-critical applications and core data that require above-average security and performance while maintaining full control over that system, including its cost.

As a result, businesses are now increasingly managing multicloud environments, often with various cloud service providers. Yet operating these multiple clouds is still difficult and enterprises often choose to gain more interoperability via a private cloud. According to *Cloud Repatriation Accelerates in a Multicloud World* (IDC #US44185818, July 2018), 80% of businesses interviewed report repatriating some workloads from public cloud environments. Respondents said they expect to move 50% of their

public cloud applications to hosted private or on-premise locations over the next two years. Note that this does not mean that they will consume 50% less public cloud.

IDC does not expect growth in public cloud adoption to slow, but a significant portion of businesses will leverage a private cloud to modernize their large installed base of noncloud applications. Many of these noncloud applications will be their mission-critical applications including core databases, business processing, or apps that execute Big Data and analytics and AI.

HPE Superdome Flex

HPE Superdome Flex is built upon decades of experience and innovation on scale-up platforms from HPE and Silicon Graphics International (SGI), which HPE acquired in 2016. The Superdome product line was introduced in 2000 and has since gone through multiple transformations to address emerging market needs. Born on RISC, then transitioned to Itanium, and for many years running the HP-UX operating system (OS) with the reputation of a relentless and reliable workhorse, HPE ultimately introduced Superdome X running Linux and Windows. Superdome X addressed the market trend toward standardizing on x86 architectures for mission-critical workloads, delivering a level of reliability previously unseen on standard platforms. With the SGI acquisition in 2016, HPE gained access to SGI's decades-long experience with building some of the most powerful high-performance scale-up systems in the industry, including the SGI UV 300, which HPE started selling as HPE MC990 X.

Superdome Flex is born out of HPE's efforts to combine the best of Superdome X and MC990 X to deliver a modular, standards-based mission-critical system with maximum flexibility, performance, and reliability. According to HPE, the platform has seen strong adoption since its introduction, covering businesses in all geographies and spanning industries from telecommunications to banking, manufacturing, public sector, education, and more. HPE reports that large systems (up to 32 sockets) are running in several production environments, although notably, the majority of shipped units are 4- and 8-socket systems, reflecting buyer desires to equip for growth. HPE also reports that the ramp of Superdome Flex has been faster than that of its predecessors Superdome X and MC990 X.

Noteworthy Characteristics of HPE Superdome Flex

Superdome Flex features a next-generation multisocket, multicore x86 architecture and is built with Memory-Driven Computing principles. It now features the second-generation Intel Xeon Scalable processors code-named Cascade Lake.

Unmatched Scale and Flexibility

Designed to address the needs of environments of all sizes, Superdome Flex has a unique modular design that enables firms to start small and scale up seamlessly as their needs grow. Utilizing a 5U 4-socket chassis "building block," the platform scales from 4 sockets to 32 sockets and from 768GB to 48TB of shared memory in a single system. The new 128GB DDR4 DIMMs available with the next-generation system use the latest 16Gb DRAM technology, doubling the density of the DRAM die and allowing for better performance and improved cost. The chassis are connected via a high-bandwidth, ultralow-latency fabric through cabling.

The platform scales on Intel Xeon Scalable Gold and Platinum processors, both the first-generation processor, code-named Skylake, and the new second-generation processor. Intel's reference design for these processors uses the new UltraPath Interconnect (UPI) that normally limits scaling to 8 sockets. Most vendors using these processors base their server designs on this "glueless" interconnect limit accordingly, but unlike these systems, HPE Superdome Flex utilizes a chassis-based architecture and

eight-generation HPE ccNUMA (cache coherent Non-Uniform Memory Architecture) hardware technology that enables Superdome Flex to scale beyond the capabilities of Intel — from 4 sockets to 32 sockets in a single system.

This means that businesses can also scale up more cost efficiently on Superdome Flex, without requiring the pricier Intel Xeon Scalable processors. HPE has developed a diverse product lineup based on the Gold and Platinum processors for various speeds, cache sizes, and core counts, enabling businesses with a low entry point and helping them avoid overprovisioning.

It's noteworthy that Superdome Flex can help businesses avoid the premium-priced high-memory SKUs from Intel. Depending on the amount of memory required per socket, there are three types of Intel Xeon Scalable processor SKUs. The base level allows for up to 1TB per socket, the M level allows for 1-2TB per socket, and the L level, only available with platforms featuring second-generation Intel Xeon Scalable processors, allows for more than 2TB per socket. If a customer wants a large total memory capacity, Superdome Flex can scale up compute to reach that capacity, therefore avoiding the extra cost associated with the M or L SKUs. Other vendors are limited to the 8-socket compute size and therefore limited to scaling memory within those sockets and forced to use the pricier SKUs to attain the same amount of memory. Spreading memory across more sockets also increases the memory bandwidth available to work on the large data sets rather than restricting access to large amounts of memory behind a single socket.

Support for HPE Persistent Memory

Superdome Flex supports HPE Persistent Memory in 128, 256, and 512GB capacities, featuring Intel Optane DC Persistent Memory. Businesses have the choice, depending on the requirements of their workloads, to run their system either with all DRAM or with a mixture of DRAM and HPE Persistent Memory. HPE Persistent Memory will be available on Superdome Flex in what Intel calls the "app direct" mode while supporting direct processor load/store access, with speed characteristics that are slower to access than DRAM (especially for writes due to the persistent property supported) but faster than SSDs. One of the use cases for HPE Persistent Memory is SAP HANA database restarts, for example, after shutting down the system for maintenance. The main column store (data tables) reside in the persistent memory. Because data does not need to be loaded from storage when restarting the system, the time to bring SAP HANA back up is significantly reduced. HPE states that it has completed several successful tests with persistent memory for this purpose.

Unbounded I/O

When fully configured, Superdome Flex supports 128 Gen3 PCIe card slots in a 32-processor system that can be used for external storage connectivity, hardware accelerators like GPUs (up to 16 NVIDIA Tesla or 8 NVIDIA Quadro GPUs), 32Gb Fibre Channel cards, Intel Omni-Path, Mellanox InfiniBand, and other peripherals. As HPE does not modify Linux for Superdome Flex, compatibility can be expected with any peripherals running under standard SUSE Linux Enterprise Server (SLES), Red Hat Enterprise Linux (RHEL), and Oracle Linux distributions. Along with the compute capabilities, a highly scalable I/O subsystem enables the deployment of HPC software (which often requires high IOPS and low-latency bandwidth access to storage or accelerator cards).

Extreme Availability

Next-generation HPE mission-critical platforms such as Superdome Flex are designed to provide "Unix on RISC"-like RAS at a system level, which can be augmented further by using clustering technologies; HPE is a key player in the AL4 market. The Superdome Flex predecessor, Superdome X, is included in the AL4 market, and as HPE Superdome Flex inherits the Superdome X RAS framework, IDC expects it to be classified at this level.

Superdome Flex features many RAS capabilities not present in other x86 servers — a key differentiator for the platform. Some of those capabilities are:

- **Firmware First:** This approach ensures error containment at the firmware level, including memory, CPU, or I/O channel errors, before any interruption can occur at the operating system layer. Firmware First covers correctable errors and uncorrectable errors and gives firmware the ability to collect error data and diagnose faults even when the system processors have limited functionality.
- **Analysis Engine:** This feature reduces human error through predictive fault handling. It monitors resources continuously, predicts hardware faults, and initiates self-repair without operator assistance.
- **Self-healing capabilities:** When faults do occur, Superdome Flex provides several mechanisms to avoid unplanned downtime, including disabling failed or failing components during boot and attempting recovery on failed or failing components during runtime.
- **Processor RAS:** Superdome Flex servers use the latest-generation Intel Xeon Scalable processors. These processors include extensive capabilities for detecting, correcting, and reporting hard and soft errors. Because these RAS capabilities require firmware support from the platform, they are often not supported in other industry-standard servers. Superdome Flex implements the RAS functionality provided in Xeon Scalable series processors, including corrupt data containment, PCIe live error containment, poison error containment, processor interconnect fault resiliency, and advanced MCA recovery.
- **Memory RAS:** Superdome Flex servers use several technologies for enhancing the reliability of memory, including proactive memory scrubbing and Advanced Double Device Data Correction (ADDDC), which HPE enhanced with specific firmware and hardware algorithms to substantially reduce memory outage rates over other x86 offerings.
- **Platform RAS:** Superdome Flex uses a fabric interconnect scheme featuring adaptive routing capabilities. The system routes traffic down the optimal latency path for performance and provides the ability to route traffic around failing or failed links in the fabric and while the system is running.
- **Application-level RAS:** Superdome Flex supports Serviceguard for Linux to enable software failover and five-nines availability.

Multiple Standard Operating Environments

Superdome Flex supports multiple standard operating environments and virtualization technologies, including SUSE, Red Hat, Windows Server, Oracle Linux, VMware, Oracle VM, and KVM. The platform runs on standard, unmodified Linux, which means it supports all the certified stacks from Red Hat and SUSE, including containers and container management software such as Docker and Kubernetes.

Simplified User Experience

Superdome Flex provides a simplified management experience by supporting HPE-specific tools such as HPE OneView, Insight Remote Support, and Proactive Care, as well as the open source Redfish API and OpenStack.

Target Use Cases and Workloads

Superdome Flex is designed for mission-critical workloads, in-memory databases, data analytics, and in-memory high-performance computing. Like its predecessors, it continues to target SAP HANA, Oracle databases, Microsoft SQL Server, in-memory high-performance computing, Unix-to-Linux migrations, and now also AI workloads.

SAP HANA

SAP has made its in-memory database, SAP HANA, the foundation of the entire environment for combined analytical and transactional processing under SAP S/4HANA. Superdome Flex's abundant memory and modular architecture make it particularly optimized for SAP HANA environments. With Superdome Flex, HPE is offering the greatest scalability on the market. It scales seamlessly from 4 sockets to 32 sockets in 4-socket increments as a single system for SAP HANA use cases. Superdome Flex has recently been SAP certified at 28 sockets/21TB and 32 sockets/24TB for SAP HANA, spanning SoH/S/4HANA and BWoH/BW/4HANA workloads and scale-up and scale-out configurations.

Oracle

Oracle continues to innovate on its core data management product. Oracle Database 18c adds new functionality and enhancements to features previously introduced in Oracle Database 12c including multitenant architecture, in-memory column store, and native database sharding.

Oracle 18c can be configured as a scale-up database or scale-out database using clustering via Oracle RAC. By deploying 18c as a scale-up database on Superdome Flex, businesses can increase their database performance per core, resulting in TCO savings from lower licensing costs. Scaling up also simplifies deployment, management, and consolidation. In addition, businesses can add in-memory options for real-time workloads. Because of its ample compute resources and memory abundance, businesses can leverage Superdome Flex to run a mix of transactional and analytic workloads on the same Oracle database simultaneously.

Microsoft SQL Server

Continuing to innovate on its venerable SQL Server database, Microsoft has brought the enterprise capabilities of SQL Server to Linux as well as Windows and Docker containers. In addition to now standard features like advanced analytics and machine intelligence, SQL Server 2017 provides exceptional performance and security. Built-in features enable faster transactions with in-memory OLTP and faster analytics with In-Memory ColumnStore. Polybase enables easy querying across the SQL Server and data stored in Hadoop. Superdome Flex is ideal for critical SQL Server workloads on bare metal or virtualized server deployments. It is also suitable for database consolidation and migration initiatives, where the target database is SQL Server, and for cases where businesses need reliability levels that they can't achieve with other industry-standard servers for their critical SQL Server workloads.

Unix-to-Linux Migration

Like its predecessor, Superdome Flex is ideal for firms that want to standardize on x86-based compute infrastructure but do not want to compromise on performance or RAS. With support for standard operating environments and virtualization technologies, firms get a wide set of options for migrating their mission-critical databases and workloads from Unix systems.

In-Memory High-Performance Computing

This use case is notably one in which Superdome Flex benefits from SGI's heritage and strong leadership. Superdome Flex equips scientific, engineering, and other technical computing environments with the ability to solve complex, data-intensive problems holistically at extreme scalability with "single-system simplicity." These types of problems are often challenging to distribute across multiple nodes in an HPC cluster and benefit from "fat" nodes (more processors and memory). They include CAE, genomics, fraud detection and prevention, and large data visualization, among others.

Artificial Intelligence

HPE has a strong platform for AI with its Apollo 6500, but Superdome Flex provides a different take on AI workload processing, which is the end-to-end workflow acceleration that can be achieved in a single system. For example, organizations can outfit Superdome Flex with abundant Ethernet for ingesting data, have CPUs and/or GPUs run AI training or inference on the data sets, keep all the data in memory, and have a number of (unmodified) applications in a workflow pipeline so that they can pipe the data from stage to stage through an in-memory file system.

Another benefit for AI is the system's huge memory capacity within a single OS. The memory footprint of accelerators can be limiting (32GB for the latest GPUs), making it difficult to process, for example, large numbers of very large, high-resolution images at speed. With the Superdome Flex's terabytes of memory, these restrictions are less noticeable. HPE is focusing on areas around genomics, analytics (graph and Big Data), and risk management (Monte Carlo simulations for FSI) and illustrating how HPE Superdome Flex can unlock new capabilities in HPC and AI with Memory-Driven Computing.

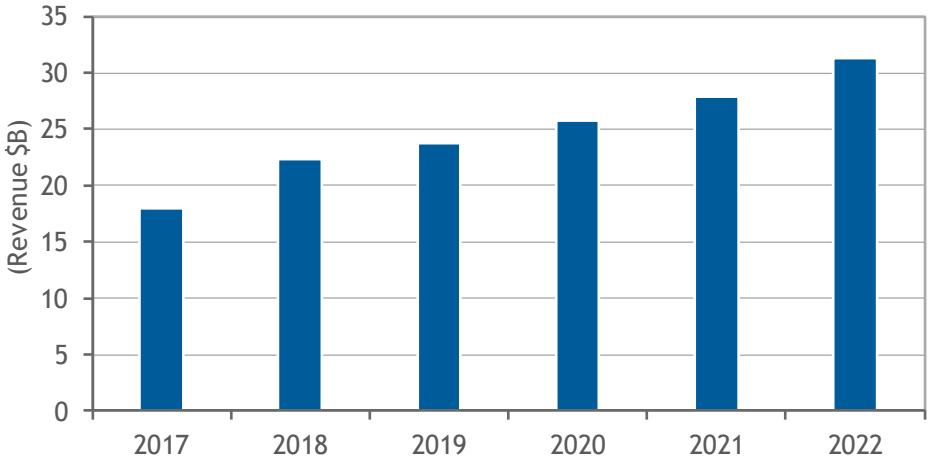
It should be noted that Intel has done a lot of work to make the Skylake processor significantly more performant for AI inferencing without changing any hardware, just through software libraries and framework optimizations, including Caffe2. On top of this, with Cascade Lake, Intel is introducing a new AVX-512 extension called Vector Neural Network Instructions, which Intel markets as "DL Boost," that accelerates AI inferencing on the processor.

CHALLENGES/OPPORTUNITIES FOR HPE

IDC believes that firms are converging their systems of record, engagement, and insight as they advance on their journey to become data-driven enterprises. As a part of this journey, many firms are also standardizing on x86-based infrastructure, even for very data-intensive, mission-critical workloads. This is driving healthy growth for the Big Data and analytics infrastructure market (see Figure 4). The move to the cloud for certain workloads continues unabated, but IDC is also seeing a distinct repatriation of workloads back on-premise, notably mission-critical workloads. These workloads run on private clouds as organizations develop a multicloud data-processing approach, and mission-critical infrastructure will play an important role in these private clouds. Indeed, IDC believes that businesses are only going to spend more money on critical infrastructure platforms that enable them to accelerate their data-driven journey.

FIGURE 4

Worldwide Big Data Infrastructure Revenue, 2017–2022



Note: For more information, see *IDC's Worldwide Semiannual Big Data and Analytics Spending Guide Taxonomy, 1H18* (IDC #US44889518, March 2019).

Source: IDC, 2019

With so much of a focus on scale-out architectures, it is easy to take for granted the importance of mission-critical scale-up platforms in data-driven enterprises. Similarly, it is also easy to overlook the fact that such platforms can be based on x86 architectures. IDC believes that a growing appetite for scale-up x86-based platforms such as Superdome Flex will continue to command traction in firms that require:

- Scale-up multisocket design for high-performance scaling
- Security, availability, and reliability for mission-critical deployments
- Flexible and modular design for opex-friendly deployments
- Optimizations for in-memory databases and real-time analytics applications
- Support for an open standards-based and cloud-ready design for deploying hybrid IT

Superdome Flex is a powerful scale-up x86 platform that combines the best of HPE's mission-critical reliability and SGI's scalable technology. It has been optimized for high-end performance at scale, in-memory databases, and a range of high-availability features throughout the platform — both hardware and software. It can handle the most demanding workloads quickly and without interruption. Because of its scale-up architecture, Superdome Flex also provides TCO efficiencies that, after a decade of x86 server sprawl and soaring opex in the datacenter, are in high demand.

For HPE, the opportunity lies in providing all the elements of a modern infrastructure environment in which fault tolerance truly matters. It is about positioning Superdome Flex as a platform that:

- Is flexible and powerful enough to handle the massive and growing amounts of data moving through a modern business
- Provides the ability to analyze data from the core to the edge in real time with an optimized in-memory design

- Is modular and cloud ready and is the right fit for any business of any size that is pursuing a traditional private cloud or hybrid IT design
- Is well-equipped for applications that require an AI inferencing component

Enterprises are also embracing a world in which app developers would want a vibrant open source ecosystem on which they can develop complex, stateful apps that depend on the hardware to maintain their state, sometimes in multiple stages. Stateful apps expect the hardware to not fail, and stateful apps in many industries may have compliance requirements that mean they cannot fail.

Here, HPE should also enlist the developer community and ensure that Superdome Flex stays open and remains developer friendly. HPE should also make sure that its systems are capable of sustaining state without performance downgrades through superior compute, fabric, and storage components.

CONCLUSION

While traditionally mission-critical systems have represented a smaller part of the server market, they are well poised to grow in new areas as next-generation data analytics, in-memory databases, and AI inferencing and the expansion into the HPC space increase the market demand for this type of platform. With Superdome Flex, HPE reaffirms its commitment to delivering a high-end x86-based system for mission-critical workloads running standard operating environments, and with the latest release, HPE demonstrates that it continues to ensure that the platform is ready for emerging workloads such as AI inferencing. Superdome Flex differentiates itself from most of the competition by being a modular and flexible x86-based mission-critical platform. HPE should no longer have to convince prospective customers of the RAS features of Superdome Flex for hybrid IT deployments; the next stage for the mission-critical platforms is to demonstrate how well they can execute AI inferencing, which is becoming an integral part of Big Data and analytics. HPE is well-positioned to be the vendor that shows the market how these dynamics can play well together.

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