Publication date:
17 March 2021
Author:
Wang Shen

Next-generation Full-stack Data Center: Fully Upgrade Computing Power to Address Data Upsurge

Brought to you by Informa Tech

Digital Business Upsurge

The development of Information and Communication Technology (ICT) has led the physical world to begin its wide-range and long-term digitalization journey. The digital connectivity market entered an explosive period, with the shipment of connected devices rapidly increase to 35 billion by 2030 compared with only 6 billion in 2014. From industrial connectivity, city management, autonomous driving to digital payment, large volume of data is generated and collected all the time. Take city management as an example. In a city with a population of tens of millions, there are hundreds of thousands of monitoring devices, thousands of traffic lights, electromagnetic sensors, intelligent inspection systems, parking lot sensors, traffic dispatch sensors, and connected vehicle sensors and other equipment combined to generate PB-level data every day, and it is expected to increase one or two times within 5 years. In addition, in the AI training for vehicle autonomous driving, the data used for training daily can reach dozens of terabytes. For connected factories with a high level of digitization and automation, PB-level data is generated daily. The data from digital applications is not only amazing in scale, but also has complex sources and different formats, forming a huge data pool with rapid circulation.

In the digital era, data becomes a new key resource to drive the development of future scenarios. Quick data mining to capture high-value information is required to generate positive feedback for future applications. The data center plays as a node for massive data flow and collection, with its compute, storage, and networking capabilities serving as the basis of data processing. The advanced computing power brought about by the upgrade of chips and servers, the storage devices with faster read and write speeds, and the high data throughput supported by high-speed network enable the efficient filter and extraction of massive data. In response to the pressure of digital business explosion, the next-generation data center urgently needs to take the computing power and computing efficiency improvement as its core goals, and fully upgrade its compute, storage, and networking capabilities.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Full Upgrade of Data Center Computing Power

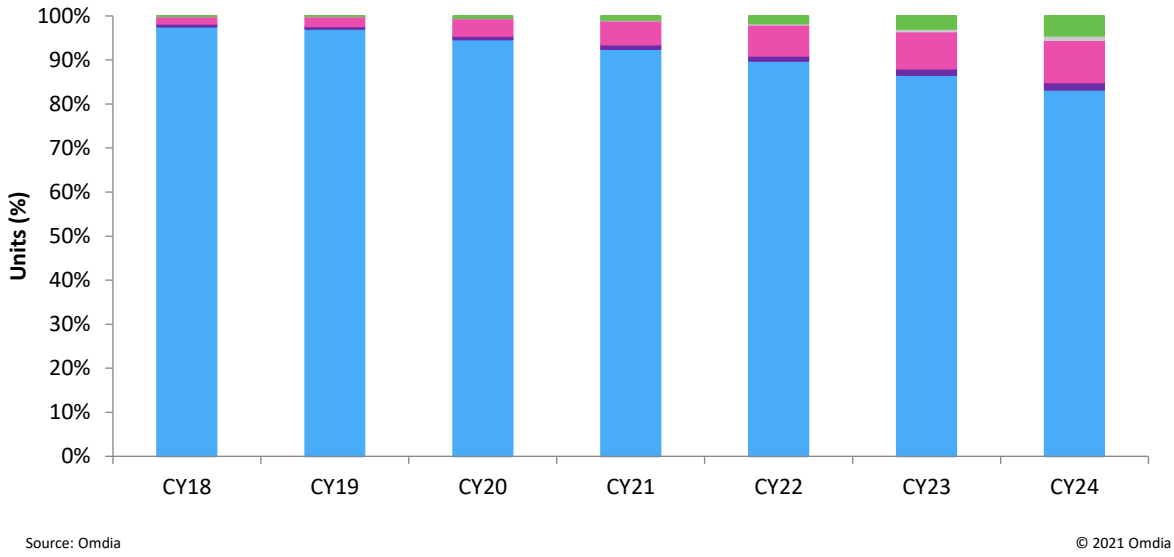
Computing Power Upgrade

Numerous enterprise business scenarios have generated a large amount of unstructured raw data, with the scale rapidly increasing from the GB level to TB or even PB. Compared with the structured data, it is much more difficult to process, and the management and storage are much more expensive. Computing power in traditional data centers is unable to effectively extract more value from data, resulting in a low utilization efficiency of the massive data. This highlighted problem has put a lot of pressure on data center's computing power. Given these massive amounts of unstructured data, it is inevitable to increase the computing power density of data center so that to efficiently reduce volume and conduct data mining and at the same time bring down unit energy consumption with controllable costs. The key lies in the rational allocation of computing resources to deal with the improved performance of chips and racks, and the subsequent cluster scaling.

More and more business scenarios are requiring high concurrency, low latency, heavy load, and elastic expansion of the computing power. With these scenarios having different focuses and specific requirements for power consumption, computing power, and cost, the computing equipment deployment in data centers requires greater flexibility. For this reason, the computing power tends to be heterogeneous and diversified, resulting in native applications based on different computing architecture. For example, for AI applications, to efficiently extract key information from unstructured data such as images and voices, AI model is required to perform fast logical reasoning, fitting simulations, and regular analysis, and give feedback in ultra-low latency. Due to the extremely high requirements of AI training for high-bandwidth parallel computing capabilities, a dedicated AI processor should be added to assist. While in other scenarios, for example, the fast processing of image or video-type data requires the embedded GPU, and low-power computing tasks require Arm architecture to improve computing power and energy efficiency ratio. Native applications based on different computing architectures have become the focus of innovation. In 2020, only about 5% of data center servers deploy dedicated processors. As the applications of new technologies such as AI become mature, computing power of data center will continue to be more diversified, resulting in servers using dedicated processors accounting for about 18% in 2024.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Figure 1: Computing Power Diversification: Market Share of Various Processor Types in Data Center Servers



Source: Omdia

© 2021 Omdia

Source: Omdia

In the long term, with the expansion of computing power sharing and separated silos of different applications broken down, it is necessary to further get rid of the previous business binding, flexibly provide platform-based computing services for different scenarios, forming horizontal hardware layer, data forwarding layer and business logic layer, etc. The intelligent allocation of computing resources according to business needs and quick preparation of data center for new projects make the new business model - Data Center as a Service (DCaaS) a reality.

Storage Upgrade

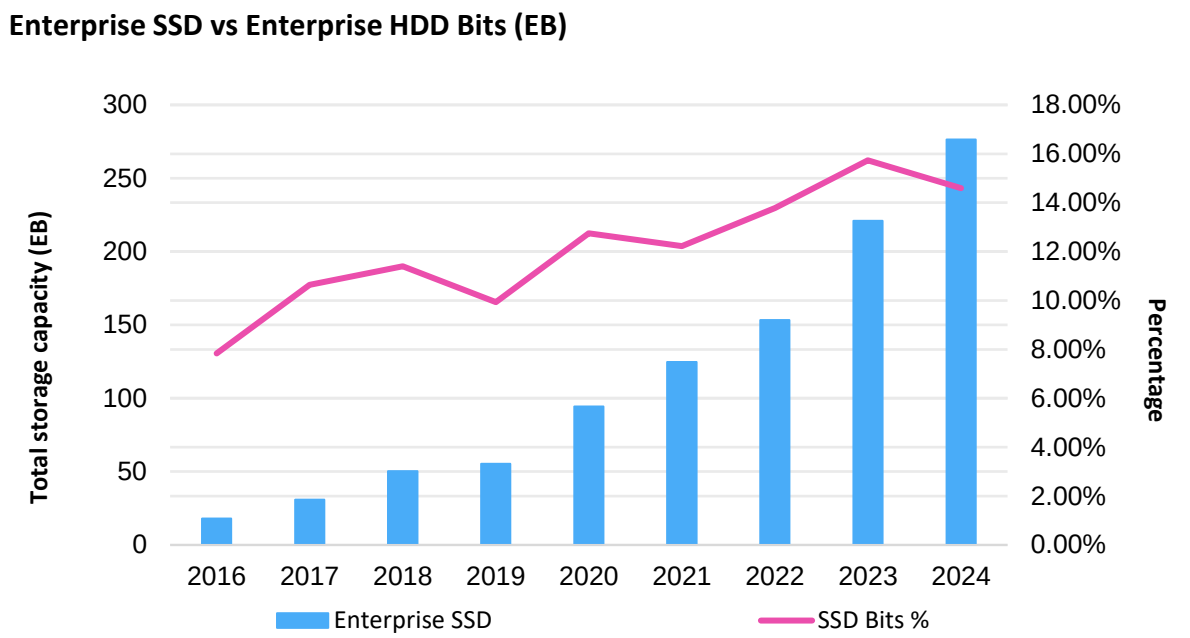
The increase in computing power efficiency has a direct relation to storage performance. In traditional data centers, storage and business scenarios are often deeply bound. The original purpose is to reduce response latency and improve storage performance, so that storage devices could better meet application requirements. However, with business scenarios becoming more granular and complex, silos restricting data flow across scenarios are emerged. These silos cause the inflexibility in allocating storage resources according to the needs for computing power of different scenarios, resulting in the increase of cost and resource waste. Therefore, in the next-generation data center, like computing equipment, storage should also be upgraded at the same time, with its goals including high-speed performance, resource optimization, and intelligent management.

For applications with high demand, read and write performance and interface bandwidth of the storage are often bottlenecks determining the upper limit of computing power. As a result, upgrading storage devices and transmission channels will be an important means to increase computing power for the next-generation data centers. The first step is to be more or even completely flash memory-based in high-computing power business scenarios. The server could

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

access external SSD storage devices through network channels, break the bottlenecks of performance and latency, and achieve the PCIe bus level of massive data throughput rate, making the running speed of core system and database sufficient to support the future upgrade requirements. This trend has quickly become a reality. Figure 2 highlights the rapid adoption of flash memory by enterprise applications.

Figure 2: The Shift to Flash Memory: Proportion of SSD Continues to Grow in Enterprise Applications



© 2021 Omdia

Source: Omdia

In operator's data centers, the increasing new business scenarios have hugely different storage requirements. In traditional data centers, storage equipment is purchased and managed completely based on the specific business requirement, which creates many silos. This deeply-bound architecture makes it impossible to share data across businesses, which not only causes unnecessary resource occupation and waste of cost, but also brings troubles to future upgrade and expansion, unable to flexibly realize the elastic scaling requirements of the operators' enterprise business. Therefore, in view of the differences in security, cost, and performance sensitivity of each application, the next-generation data center needs to change the traditional model of storage-application binding, deeply decouple compute and storage, and integrate different types of storage into a unified resource pool through network virtualization, so that to flexibly mix computing power and storage according to performance or capacity requirements of different applications. For scenarios with large fluctuation in storage requirements, data could be stored by accessing the single storage pool no matter what their storage requirements are. High-speed flash memory

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

resources in the storage resource pool can be allocated for applications such as high-frequency transactions in need of computing power. While for applications such as multimedia paying attention to capacity, storage devices with lower cost and larger capacity could be used. For historical archive data, it can be stored in mediums such as Blu-ray for a long time.

In the integrated storage resource pool, data scheduling and management efficiency determines the utilization efficiency of the resource pool. Relying on manual management is no longer realistic. Therefore, in the next-generation data center, achieving full-scenario automated management and operation & maintenance by AI technology so that to realize the "autonomous driving" in storage management is essential. The automated management system should quickly and automatically identify the characteristics of data once it receives a data request, and then direct the request to the storage device in the resource pool according to its importance and throughput requirements. Furthermore, AI can be used to predict the storage requirements and intentions of the business, and perform accurate distribution in advance to ensure the storage latency and mobility for different applications and greatly reduce the processing time and storage space of the core business.

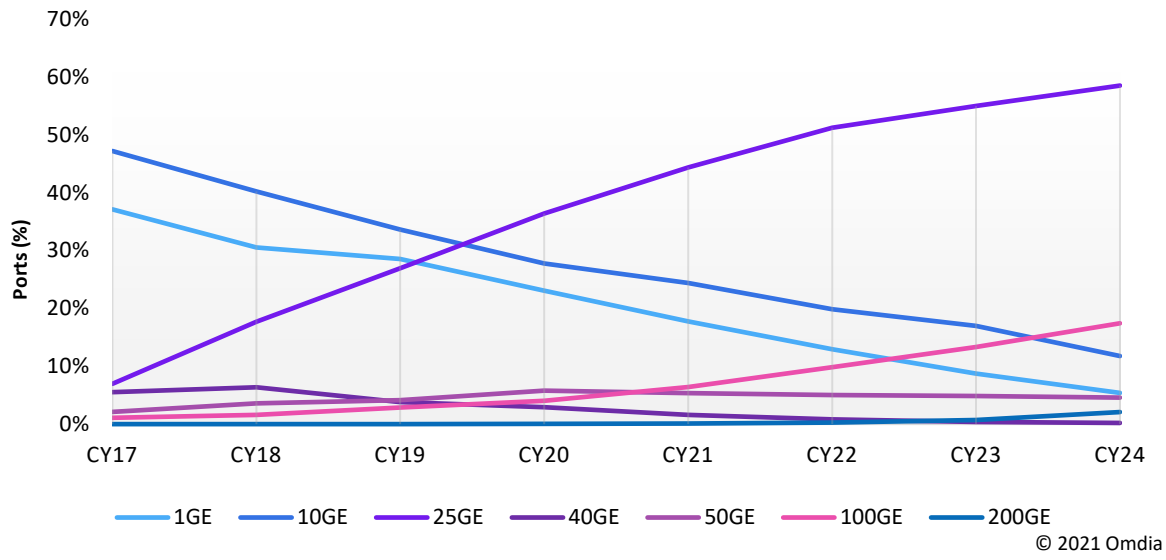
In addition, with the breaking down of storage silos across businesses, data security issues will become more prominent. Apart from ensuring data flows, an intelligent storage management system also bears greater security responsibilities. Therefore, it is also necessary to enable more granular data security management by identifying data patterns. For the most important high-frequency data, it is crucial to fully guarantee its disaster tolerance; for non-immediate low-frequency data, hot backup is required to give full play to the computing power of data center and improve its cost-effectiveness and safety.

Network Upgrade

In a data center with integrated computing and storage pools, the internal network is responsible for the data throughput between storage and servers, which is another decisive factor for computing power. Greater server performance, higher computing power density and the requirement for low latency push the network to quickly enter a high bandwidth era, and it will quickly become the mainstream in the next few years. In the communication process between server and storage, even in the high-bandwidth network, network packet loss is still the main reason for latency and becomes a bottleneck restricting the performance of computing power. Therefore, in the next-generation data center, it is necessary to make all-round improvements in the architecture and communication protocol of the data transmission network between IT equipment to address the shortcomings of traditional network.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Figure 3: The Shift to High Bandwidth: Advanced Server Performance and Computing Density Lead to Growing Demand for Network Bandwidth



Source: Omdia

Traditional data centers often adopt all these three network types: Ethernet, FC (Fiber Channel) storage private network, and IB (InfiniBand) computing private network. While the three independent network transmission protocols co-exist, but their different interfaces and operation & maintenance methods hinder the free flow of data in storage, compute, and networking, and go against the reduction of operation & maintenance costs, which unable to meet the network performance and cost requirements of future scenarios.

For reasons above, in the next-generation data center where storage and computing power are highly decoupled, the network needs to be more unified and open. Thanks to the large-scale adoption of higher-speed SSD storage and NVMe protocol in the next-generation data center, the IP protocol-based Ethernet can achieve better performance, enable much higher bandwidth than that of the FC network, and widely support IP routers for data exchange, with more open interfaces, reduced maintenance difficulty and lower cost. In scenarios using IP Ethernet for data transmission, the probability of packet loss can be minimized through network and algorithm optimization. Ultimately, zero packet loss in data transmission and ultra-low latency are realized, enabling the IP network to completely replace and integrate FC private network and IB private network, so as to become an ideal carrier for the next-generation data network and realize full-scale application in the next-generation data center.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Summary

With the upsurge of digital business generating massive amounts of data all the time, the in-depth adoption of AI provides high-quality solutions for the application of data in various industries. Given that the data is large in scale - reaching hundreds of terabytes or even petabytes, fast flowing and highly unstructured, and data mining and value extraction need to be completed within a short period of time, the AI training and processing of this data pattern put tremendous pressure on the current computing power of data center. With compute-storage-networking being the hub of data processing, computing power upgrade is a must to meet future business needs.

Future upgrade is not only about the data processing efficiency. In the next-generation data center aiming at achieving sustainability, continuous evolution towards low cost, easy maintenance and high security is also required, leading to the shifts to high-density and integrated computing, flash memory storage and resource-pooling, all-IP network and so on. These are beneficial to the granular management of digital services, which further improve the friendliness of resource utilization efficiency and operation & maintenance. More importantly, the roles that compute, storage and networking played in digital business complement each other, and the backwardness of one often limit the performance of the other two. Thus, the upgrade of the three need to be highly synchronized to promote the free flow and maximize the value of massive data in the next-generation data center.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Appendix

Author

Wang Shen

Senior Consultant, Global Enterprise & IT

askananalyst@omdia.com

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Get in touch

www.omdia.com
askananalyst@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision-makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa Tech, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help your company identify future trends and opportunities.

About [partner company]

Up to 150 words about the company and its offerings. Use only factual, non-marketing-led language. Omdia retains editorial control over this section. If this section is used, be sure to include "In partnership with" box on cover.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the “Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together “Informa Tech”) and represent data, research, opinions or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an “as-is” and “as-available” basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness or correctness of the information, opinions and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees and agents, disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial or other decisions based on or made in reliance of the Omdia Materials.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.