Translation

The following annual white paper by a Chinese state-run think tank analyzes the computing power landscape, in China and globally, as of late 2022. The report begins by summarizing global developments in compute over the past year, and then provides statistics on China's compute industry. The authors find that, despite disruptions caused by the COVID-19 pandemic, China's computing power growth outpaced the world average in 2022.

Title
White Paper on China's Computing Power Development Index (2022)
中国算力发展指数白皮书（2022 年）

Author
China Academy of Information and Communications Technology (CAICT; 中国信息通信研究院; 中国信通院). CAICT is a think tank under China's Ministry of Industry and Information Technology (MIIT; 工业和信息化部; 工信部).

Source
CAICT website, November 5, 2022.

The Chinese source text is available online at:
http://www.caict.ac.cn/kxyj/qwfb/bps/202211/P020221105727522653499.pdf
An archived version of the Chinese source text is available online at: https://perma.cc/4GW2-UWDR
U.S. $1 ≈ 7.2 Chinese Yuan Renminbi (RMB), as of May 13, 2024.

Translation Date	Translator	Editor
May 13, 2024	Etcetera Language Group, Inc.	Ben Murphy, CSET Translation Manager

No. 202215

White Paper on China's Computing Power Development Index (2022)[1]

[1] Translator's note: The China Academy of Information and Communications Technology (CAICT) published its first *White Paper on China's Computing Power Development Index* in 2021. An English translation of the 2021 version is available on CSET's website at: https://cset.georgetown.edu/publication/white-paper-on-chinas-computing-power-development-index/.

China Academy of Information and Communications Technology (CAICT)

November 2022

Copyright Statement

Preface

The importance of computing power ("compute") has been elevated to new heights. As a new productive force (新的生产力) in the digital economy era, compute plays an important role in driving S&T progress, the digital transformation of industries, and economic and social development. While compute development is facing challenges from application diversification and the imbalance between supply and demand, the rise of new fields such as artificial intelligence (AI), digital twins, and the metaverse is driving the rapid growth of compute, the diversified innovation of computing technology, and the accelerated redrawing of the industrial landscape.

2021 was the opening year of the "14th Five-Year Plan" and the starting year of China's construction of the new pattern of development (新发展格局). In the face of the complex and grave international environment, multiple pandemic outbreaks domestically, and many other tests, China's compute development has achieved steady improvement. Overall, the following characteristics can be seen:

Compute scale continues to expand, and intelligent compute has become the main driving force. On the infrastructure side, China is rapidly deploying data centers and intelligent computing centers, with infrastructure compute reaching 140 exaflops (EFLOPS) in 2021, ranking China second in the world. The number of data center racks in use exceeded 5.2 million standard racks, and the number of intelligent computing centers already in operation approached 20, with over 20 more centers under construction. **On the computing devices side**, China shipped more than 19.6 million general purpose servers and 500,000 AI servers in the past six years, and the total scale of compute grew at a 50% rate to reach 202 EFLOPS, accounting for 33% of the global total. Intelligent compute (智能算力) in particular has maintained steady and rapid growth, with a growth rate reaching 85%.

The compute industry is booming, and compute innovation ability is constantly improving. China has formed a computing industry with a relatively complete system, enormous scale and volume, and active innovation. In terms of size, China's computing industry accounts for more than 20% of the electronic information manufacturing industry, with over 2,300 large-scale enterprises,[2] and China's market share in whole machines (整机) continues to climb. An industrial ecosystem has formed covering the underlying hardware and software, whole-machine systems, and applications, and a number of advanced computing technology innovations have emerged. Breakthroughs and in-depth application continue to be achieved in areas such as computing

[2] Translator's note: The term "large-scale enterprises" (规模以上企业) refers to companies that have 20 million or more Chinese yuan Renminbi (RMB) in annual operating income.

microchips, computing systems, and computing software, and innovations are being made in computing platforms and system acceleration, with multiple breakthroughs in cutting-edge computing technology.

The development environment continues to improve, and industry empowerment (赋能) benefits are becoming increasingly visible. China's network infrastructure capabilities have improved steadily. Inter-provincial outlet bandwidth is continuing to grow, 5G network construction continues to advance, and the computing-network collaboration (算网协同) system is developing rapidly. Compute investment continues to expand, and the "East-West Compute Transfer"[3] project has brought about a marked strengthening of investment in the western region. Data output is growing rapidly, and the open sharing of data resources is accelerating. China's compute demand for consumer and industrial applications is growing rapidly. The internet remains the largest industry in terms of compute demand, accounting for nearly 50%. Telecommunications and finance are the biggest traditional industries for compute applications, while compute demand in the manufacturing industry has relatively large potential for growth.

As compute boosts economic growth, the pace of development is accelerating everywhere. The scale of China's compute industry, as represented by computers, reached 2.6 trillion Chinese yuan Renminbi (RMB) in 2021, directly and indirectly driving RMB 2.2 trillion and 8.2 trillion in economic output, respectively. The Beijing-Tianjin-Hebei, Yangtze River Delta Economic Zone, Guangdong-Hong Kong-Macao Greater Bay Area, and Chengdu-Chongqing Economic Circle (成渝双城经济圈) regions of China have leading levels of compute development, with Guangdong, Beijing, Jiangsu, Zhejiang, Shanghai, and Shandong in the first tier. The compute development of core provinces in the central and western regions is gathering steam, but it still faces problems such as technology industry weakness, a development environment that urgently needs improvement, and limited demand for compute.

Building off the 2021 white paper, the 2022 version strengthens the evaluation of compute supply levels, and establishes the China Computing Power Development Index 2.0 from five dimensions: compute scale, compute industry, compute technology, compute environment, and compute applications. We added two dimensions—compute industry and compute technology—to more objectively evaluate the compute development level at the current stage in China as a whole and in its provinces and

[3] Translator's note: "East-West Compute Transfer" (“东数西算”; literally, "eastern data, western compute") refers to an initiative to build computing power in western China, where land and electricity are relatively cheap, in support of data centers located along China's densely populated and developed east coast.

cities. The hope is that it will provide a reference for localities to promote the compute technology industry, infrastructure construction, and compute application development.

This white paper still has many shortcomings, and criticism and corrections from the public are sincerely requested.

Table of Contents

I. Global compute has entered a new period of rapid development. 10

(i) The scale of compute is growing steadily 10

(ii) The compute industry is flourishing 12

(iii) Diverse innovation in compute technology 14

(iv) Compute's empowerment effect continues to deepen 17

(v) Compute competition continues to intensify 19

II. China has opened a new chapter of compute-empowered digital economy...... 21

(i) Compute scale continues to expand, and intelligent compute has become a growth driver. 21

(ii) The supply level has risen significantly, and advanced computing innovations are emerging 24

(iii) Optimization of the development environment continues, with network systems opening up new data arteries 25

(iv) The transformation to intelligentization and the transition to digitalization (智改数转) are deepening under the continuous stimulation of empowerment effects 28

(v) Compute boosts economic growth by helping the digital economy achieve quantitative and qualitative growth 31

III. Evaluation of the China Computing Development Index 33

(i) Basis for index establishment 35

(ii) Index system establishment 38

(iii) Evaluation of China's compute development 39

(iv) Relationship between the compute development index and the economy 46

IV. Accelerating the consolidation of compute's basic foundation, and stimulating the engine of digital economy development 47

(i) Solidify compute's foundation and guide the layout of infrastructure 48

(ii) Persist in being innovation-driven, and promote core technology research and development 49

(iii) Build the industry system to improve the supply of computing products ... 49

(iv) Build the development environment, and optimize network data services 50
(v) Enrich integration-based application, and strengthen compute's industry-empowering effect........ 50
(vi) Strengthen international cooperation and promote the construction of the Belt and Road Initiative 51
Appendix 1: Framework for calculating the compute index 51
(i) Compute scale sub-index calculation method 52
(ii) Compute industry sub-index calculation method 52
(iii) Compute technology sub-index calculation method 53
(iv) Compute environment sub-index calculation method 54
(v) Compute application sub-index calculation method........ 54
Appendix 2: Data Sources........ 55

List of Figures

Figure 1 Computing power development research system........ 10
Figure 2 Size and growth of global compute 11
Figure 3 Framework of the advanced computing technology industry system 15
Figure 4 Relationship between the scale of compute and GDP globally in 2021 19
Figure 5 Distribution of global compute in 2021 20
Figure 6 Scale and growth rate of China's compute........ 23
Figure 7 Internal structure of China's compute 24
Figure 8 Development of mobile communication base stations in China 27
Figure 9 IT hardware, software, and services spending in China 28
Figure 10 Distribution of compute applications among industries in China 30
Figure 11 Growth in mobile internet traffic and monthly DOU, 2016-2021 31
Figure 12 Relationship between the scale of compute, GDP, and the digital economy, globally and in China, 2016-2021........ 32
Figure 13 China Computing Power Development Index 2.0........ 34
Figure 14 Compute development index of some Chinese provinces in 2021........ 40

Figure 15 Compute development index of some Chinese provinces in 2021 41
Figure 16 Infrastructure compute scale of some Chinese provinces in 2021 42
Figure 17 Compute industry sub-index levels of some Chinese provinces in 2021...... 43
Figure 18 Computer technology sub-index levels of some Chinese provinces in 2021 44
Figure 19 Compute environment sub-index levels of some Chinese provinces in 2021 45
Figure 20 Compute application sub-index levels in some Chinese provinces in 2021 46
Figure 21 Relationship between the computing power development index and GDP.......... 47

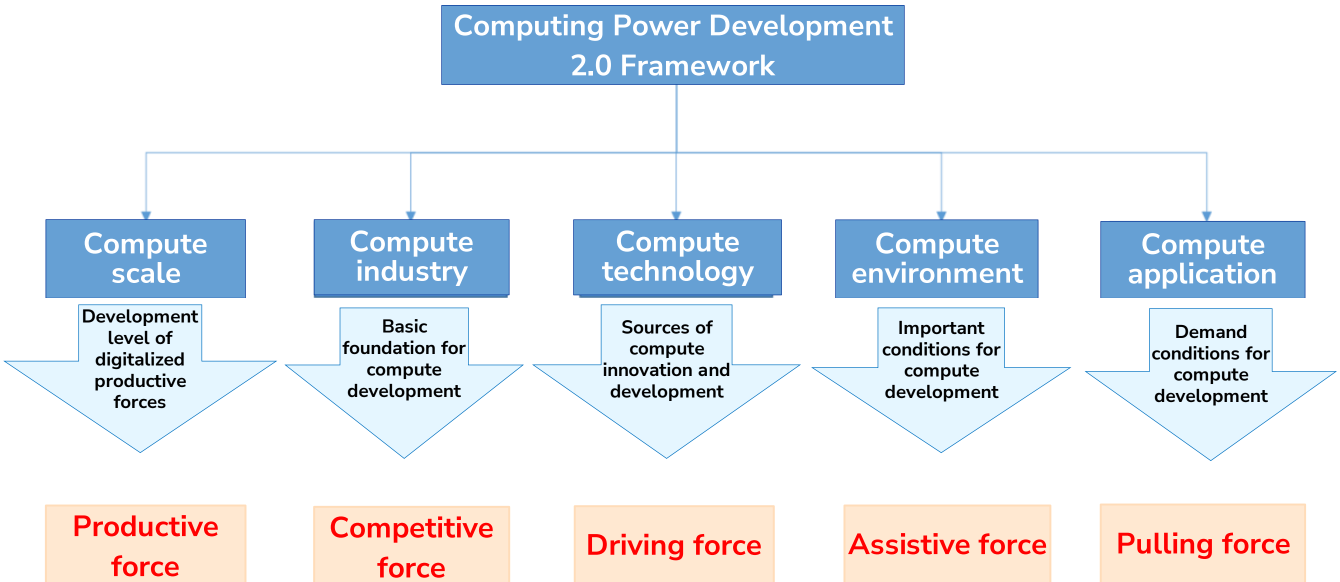
List of Tables

Table 1 China Computing Power Development Index system 38

Compute is a new productive force (新生产力) in the digital economy era, but it has already become a core force driving digital economy development and a solid foundation supporting its development. Compute plays an important role in driving S&T progress, promoting the digital transformation of industries, and supporting economic and social development. According to calculations in the *White Paper on China's Computing Power Development Index (2021)*, every 1 RMB invested in compute leads to 3 to 4 RMB of economic output.

The current challenges in compute development come from the diversification of applications and the imbalance between supply and demand. On one hand, in the "intelligent everything" era, the rapid rise of emerging technologies, the explosive growth of massive data, and the ever-growing diversity of application scenarios have stimulated a thousand-fold increase in compute and accelerated its diversified upgrading. On the other hand, the improvement of compute is facing multi-dimensional challenges, and there is still a huge gap when it comes to transforming chips into compute. The compute growth that can be brought by upgrading existing computing technology is only on the order of a few fold per year, leaving a big gap between supply and demand. Whether at the hardware level or the architecture level, computing technology development is in urgent need of transformation.

Based on the connotations and characteristics of compute, the 2021 white paper mainly established an overall computing power development framework from three perspectives: compute scale, compute environment, and compute application. On the basis of the 2021 white paper, this 2022 version further improves the computing power development research system from five dimensions: compute scale, compute industry, compute technology, compute environment, and compute application. As shown in Figure 1, two dimensions—compute industry and compute technology—have been added to strengthen the evaluation of compute supply levels, highlighting the cornerstone roles that compute technology and the compute industry play in the development of compute. In the digital economy era, compute scale is an important indicator of the level of digitalized productive force (生产力) development of countries and regions; the compute industry, as the basic foundation for the development of compute, is a competitive force; compute technology, as the source of compute innovation and development, is a driving force; the compute environment is an important condition for the development of compute, and is an assistive force; and compute application, which reflects the demand situation for compute development, is a pulling force. Compute technology, represented by advanced computing, and the compute industry provide solid support for compute development at scale; the compute environment provides fertile soil for the growth of compute and the development of compute technology and the compute industry; compute application has a pulling effect on the growth of compute scale and the compute industry, and drives compute technology upgrading. These five major elements promote each other and develop synergistically.



Source: CAICT

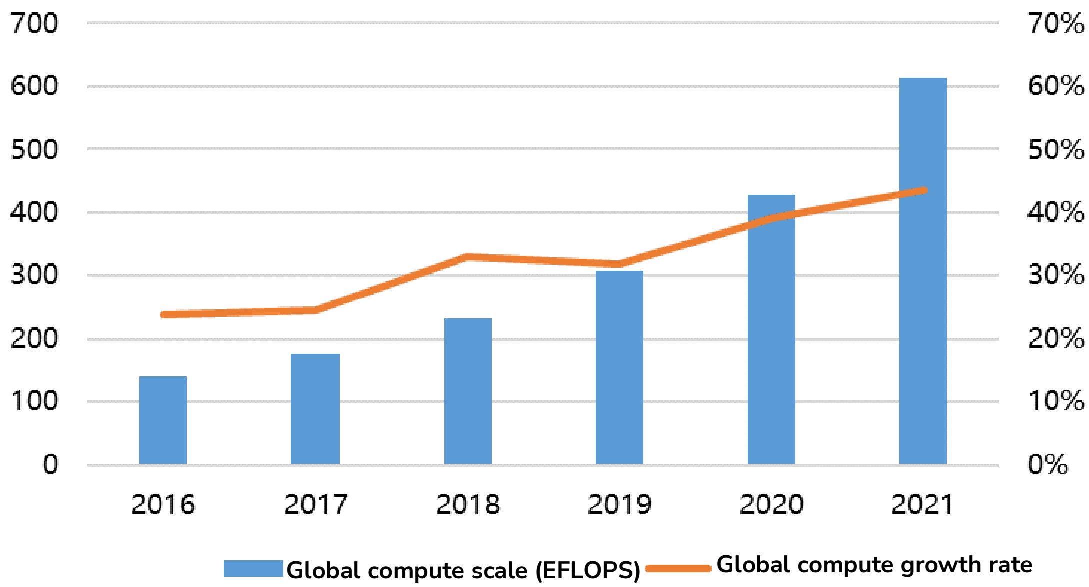
Figure 1 Computing power development research system

I. Global compute has entered a new period of rapid development.

Compute is the new productive force of the digital economy era. In the era of intelligent everything, the demand for computing is increasing by hundreds and thousands of times. The rapid rise of emerging fields such as scientific research, AI, digital twins, and the metaverse has promoted rapid growth in global compute and driven diverse innovations in computing technology and products, driving the restructuring and reshaping of the industrial landscape. Compute has become a new engine for the development of the global digital economy and a new focus of strategic competition among countries.

(i) The scale of compute is growing steadily

Global compute has maintained rapid and steady growth. Against the backdrop of the "everything sensing, everything connected, and everything smart" digital economy era, the global total amount of data and scale of compute continue to show high-speed growth. According to *National Data Resources Survey Report* (国家数据资源调查报告) figures, the global data total was 67 zettabyes (ZB; 10^{21} bytes) in 2021, with the growth rate over the latest three years exceeding 26%. According to CAICT's calculations, total global compute grew 44% in 2021 to reach 615 EFLOPS, of which basic compute[4] (in FP32[5]) was 369 EFLOPS, intelligent compute[6] (converted to FP32) was 232 EFLOPS, and supercomputing compute[7] (converted to FP32) was 14 EFLOPS. Based on forecasts by Huawei Global Industry Vision (GIV), humanity will usher in the yottabyte (YB; 10^{24} bytes) data era in 2030. Global compute will reach 56 zettaFLOPS (ZFLOPS), with an average annual growth rate of 65%, of which basic compute will reach 3.3 ZFLOPS, with an average annual growth rate of 27%; intelligent compute will reach 52.5 ZFLOPS, with average annual growth over 80%; and supercomputing compute will reach 0.2 ZFLOPS, with average annual growth over 34%.



Sources: CAICT, International Data Corporation (IDC), Gartner, TOP500

Figure 2 Size and growth of global compute

Demand diversification is accelerating compute diversification and upgrading. Diversified intelligent scenarios require diversified compute, and the rapid rise of new fields such as AI, scientific research, and the metaverse is placing greater requirements on compute. Take the metaverse for example. Intel predicts that the metaverse will

[4] The scale of basic compute is estimated based on the total compute of servers worldwide over the last 6 years. Global basic server compute = $\sum_{\text{last 6 years}}$ (annual scale of server shipments*average server compute in the current year).

[5] FP32 is single-precision floating point data format, FP16 is half-precision floating point format, and FP64 is double-precision floating point format.

[6] The scale of intelligent compute is estimated based on the total compute of AI servers worldwide over the last 6 years. Global intelligent compute = $\sum_{\text{last 6 years}}$ (annual scale of AI server shipments*average AI server compute in the current year).

[7] The scale of supercomputing compute is estimated mainly based on TOP500 global supercomputer data, and by referring to relevant data from supercomputer manufacturers.

require a thousand-fold increase in compute, and NVIDIA believes that compute for real-time rendering in immersive experience still falls short by a million-fold. **In terms of basic compute**, cloud computing growth has basically recovered to the pre-pandemic level. According to International Data Corporation (IDC) data, the global cloud computing infrastructure as a service (IaaS) market grew to US$91.35 billion in 2021, up 35.64% year-on-year, and the IaaS + platform as a service (PaaS) market totaled US$159.6 billion, up 37.08% year-on-year. Cloud computing will become the mainstream general purpose computing model in the future, providing basic support for new technologies such as big data, AI, and 5G, as well as a wellspring of impetus for the digital transformation and intelligent upgrading of industries. **In terms of intelligent compute**, there is a contradiction between the current demand for massive and complex data processing on the one hand and the supply of non-diversified compute (单一算力) on the other. More than 80% of the world's rapidly expanding data is unstructured data (text, images, voice, video, etc.), and as Moore's Law and Dennard scaling slow down, the annual performance improvement of chips represented by central processing units (CPUs) is no more than 15%, which cannot meet the demand for processing unstructured data, such as video and images. There is thus an urgent need for diversified intelligent compute. **In terms of supercomputing compute**, the law of a thousand-fold increase every decade is still continuing. The performance of the new Frontier supercomputer of Oak Ridge National Laboratory (ORNL) in the United States scores 1.102 EFLOPS in the Linpack benchmark test, surpassing Japan's Fugaku and becoming the world's first publicly confirmed quintillion floating-point operations per second (FLOPS) supercomputer. The world's supercomputers have officially entered the exascale computing era.

(ii) The compute industry is flourishing

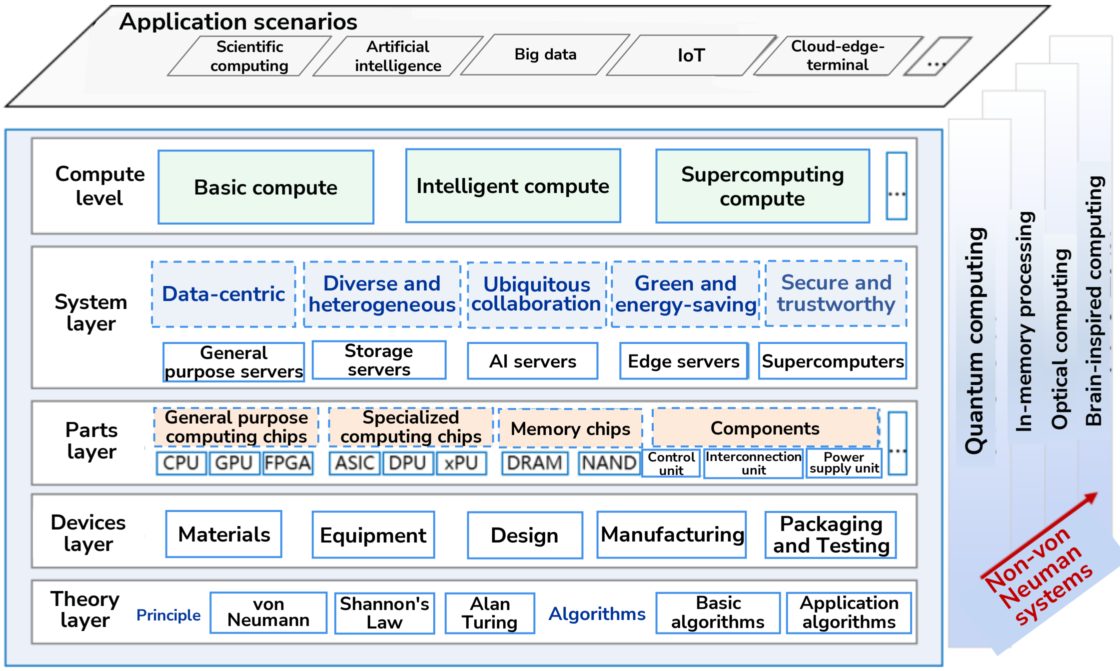
Benefiting from the rapid economic recovery, the global server market continues to grow. Whole machines: Global server market shipments and sales in 2021 amounted to 13.539 million units and US$99.22 billion, respectively, up 6.9% and 6.4% year-on-year. Hewlett Packard Enterprise (HPE)/H3C (新华三) topped the global server market with a 15.6% market share, while Dell, Inspur (浪潮), Lenovo (联想), and Huawei ranked second through fifth, with market shares of 15.4%, 8.9%, 6.4%, and 1.9%, respectively. **Chips**: The server chip market has long been dominated by the x86 architecture. Intel and AMD had market shares of 81% and 16%, respectively. As Intel's dominant position in servers weakens, AMD's market share will continue to increase. In addition, ARM server chip products have been rising gradually. NVIDIA, Amazon, Huawei, Alibaba, and other foreign and domestic giants have launched self-developed ARM server CPUs, and ARM server market share is also expected to continue to improve to nearly 10% by 2024, becoming an important complementary force of general purpose compute.

Increasing model complexity and training data scale are driving rapid growth of AI server demand. Whole machines: According to IDC statistics, the global AI server market reached US$15.6 billion in 2021 with year-on-year growth of 39.1%, exceeding the 22.5% growth rate of the overall global AI market (including hardware, software, and services), and becoming a driver of overall AI market growth. In the global AI server market in 2021, Inspur, Dell, and HPE were the top three with market shares of 20.9%, 13.0%, and 9.2%, respectively, and the combined market share of the three vendors was 43.1%. Huawei and IBM ranked fourth and fifth, with market shares of 5.8% and 4.1%, respectively. **Chips**: Traditional chip giants have accelerated improvement of their AI chip product systems and continue to advance construction of all-round capacity as they strive for dominance of the diverse compute ecosystem. Intel released Habana Gaudi2, representing a new generation of high-performance deep learning AI training processors, with a processing speed twice that of the previous generation. AMD, after completing the acquisition of Xilinx, plans to incorporate Xilinx's field-programmable gate array (FPGA) AI engine into its CPUs. Using Taiwan Semiconductor Manufacturing Company's (TSMC's) 4nm process, NVIDIA has launched a graphics processing unit (GPU) with a new architecture that integrates 80 billion transistors, dramatically increasing AI computing speed.

The era of exascale supercomputing has arrived, and supercomputing equipment makers have accelerated the pace of industrialization. In terms of whole machines, supercomputing equipment manufacturers have strengthened industrial integration and layout. Among those with computers on the TOP500 list, Lenovo is currently the largest supercomputer manufacturer, with a total of 161 units, accounting for 32.2% of the global total. HPE has 96 units on the list, accounting for 19.2% of the total and ranking it second. Inspur, Atos, and Sugon (曙光) are third through fifth with 50, 42, and 36 units, respectively, accounting for 10%, 8.4%, and 7.2%. After acquiring Cray for US$1.3 billion, HPE launched Frontier, the first exascale supercomputer, and expects to launch a supercomputer with performance of more than 2 EFLOPS in 2023. France's Atos launched the new BullSequana XH3000 supercomputer, which will achieve EFLOPS-level traditional digital simulation and 10-EFLOPS-level AI accelerated computing. **In terms of chips,** CPUs are still dominated by Intel and AMD, with up to 388 (77.6%) of the supercomputers on the TOP500 list using Intel CPUs, and 93 using AMD processors. In addition, heterogeneous computing chips are increasingly used in supercomputers. A total of 168 supercomputers in the TOP500 ranking use accelerator or coprocessor technology, of which 154 use NVIDIA chips and 8 use AMD chips.

(iii) Diverse innovation in compute technology

Advanced computing is laying a firm foundation for compute development. In the era of intelligent connection of everything, the massive flood of data and exploding demand from diverse applications have driven the exponential growth of compute scale and the continuous adjustment of the compute structure. A single technology upgrading path would be hard pressed to keep pace with the requirements for high-quality development of compute. A new wave of advanced computing technology development featuring diversification and fusion is thus being ushered in. Given the information processing needs of scenarios characterized by massive data, real-time response, ubiquity and diversity, environmental friendliness and security, etc., the formation of computing technologies and products with greater compute, higher energy efficiency, and greater diversity and flexibility, through integration-based innovation and disruptive reconfiguration of computational theory, computing devices, computing components, and computing systems, will help achieve the performance enhancement of single-point computation and the efficient utilization of computing systems, thereby meeting the compute challenges and ending the post-Moore's Law-era crisis in compute. On the one hand, advanced computing, as a new driving force for technological innovation, has further pushed forward the evolution of classical computing technology using silicon-based semiconductors, with systematic thinking gradually changing chip design ideas and resulting in diverse computing architectures. On the other hand, breakthroughs are being made in disruptive computing technologies such as quantum computing, in-memory processing (存算一体), optical computing, and brain-inspired computing, pushing non-classical computing from theory into practice. The continuous development of advanced computing technology will effectively boost the existing scale of compute, significantly reduce its cost, and improve its utilization efficiency, subjecting the compute system to disruptive changes.



Source: CAICT

Figure 3 Framework of the advanced computing technology industry system

Moore's Law continues to evolve. First, advanced processes continue to improve. Samsung announced in mid-2022 that 3nm had officially entered mass production. Using a new gate-all-around (GAA) structure transistor to replace the traditional fin field-effect transistor (finFET) technology, performance is improved by about 23% compared to the 5nm node, power consumption is reduced by about 45%, and the area is reduced by 16%. Competition is underway to develop advanced processes with nodes of 2nm and below. This multi-dimensional research effort is mainly focused on the directions of GAA devices, high-numerical-aperture extreme ultraviolet photolithography (high-NA EUV lithography), two-dimensional materials, and other new structures, new equipment, and new materials. TSMC has launched a 1.4nm process layout, and Intel plans to enter the era of angstrom-level manufacturing in 2024, continuing to extend the lifespan of Moore's Law. **Second**, chiplets have opened up a new evolutionary path. Relying on advanced packaging technologies such as 2.5D and 3D, chiplets can realize three-dimensional integration between different processes and different types of chips, becoming an effective solution to address the difficulty of advanced process design and ballooning costs. With Intel, AMD, ARM, and other industry giants promoting the unification of chiplet interconnection standards, chiplets are expected to change the existing rules of the industry, affecting the future upgrading path of chips.

Computing chip breakthroughs are accelerating. While the performance of

CPUs, GPUs, and other general purpose chips continues to improve, the development of specialized computing chips is continuing. With AI chips having entered the commercial implementation stage, chip solutions covering all scenarios have emerged in the industry. On the cloud-based training side, NVIDIA GPUs have a dominant share of the market, while on the cloud-based inference side, there is diversified development of high-performance chip architectures. On the terminal side, scenarios are highly fragmented, but a number of products have been put into commercial service in fields such as automated driving, video surveillance, and smart home. Data processing unit (DPU) chips have become a new hotspot for the industry to chase after. With rapid growth of data-intensive types of demand placing new requirements on existing cloud-terminal computing architectures, there is active innovation in DPU chips focused on data acceleration and processing, as well as various types of resource management in the cloud, and this has become an important driver of system performance improvement. At present, chip vendors including NVIDIA, Intel, and Marvell, cloud service providers like Amazon and Alibaba Cloud, and emerging enterprises such as Corigine (芯启源) and Yusur Tech (中科驭数) have all developed DPU products.

Heterogeneous computing has become a mainstream model. With the significantly increasing presence of heterogeneous computing in mobile internet, AI, high-performance computing, and other typical applications, diversified, cross-system processor collaboration has become an important means of improving the parallelism and energy efficiency of computing. This mainly involves reconstruction in the two dimensions of hardware architecture and hardware-software integration and coordination. As hardware architectures break away from CPU-centered systems, there has been a gradual deepening in the application dimension, from chips and nodes to system-level partition heterogeneity, as well as exploration in taking computing architecture from the typical model of “control chip + various types of specialized acceleration chips” to new systems such as multi-engine separation (多擎分立), which promise to achieve larger-scale, multi-system, efficient parallel task scheduling. In software collaboration, cross-domain unification and flexible deployment are important directions. By integrating compilers, programming languages, acceleration libraries, development tools, etc., with the help of a unified heterogeneous software platform, they provide programming models and application programming interfaces to different underlying computing architectures, achieving the unified management and scheduling of diverse and heterogeneous types of compute. Typical representatives including Intel’s OneAPI, NVIDIA’s CUDA, and Huawei’s Beiming Diversified Computing Convergence Architecture (北冥多样性计算融合架构).

The application of ubiquitous collaborative computing is deepening. Ubiquitous collaboration is a broader concept of computing system innovation. Center-edge collaboration in particular has been applied in many scenarios, and has gradually deepened from local data preprocessing to cloud-edge collaborative support for application computing processes. AI training is thus shifting from cloud-based to cloud-edge collaboration, and inference is moving from cloud- and terminal-based to edge-terminal collaboration. At present, efforts are focused on cloud-edge-terminal ubiquitous computing architectures and edge-side compute realization. In the compute network, cloud, edge, and terminals together constitute a multi-layered and three-dimensional ubiquitous computing architecture, which in turn constitutes a new type of infrastructure for the compute network through deep integration with the network. On the edge side, compute realization is subject to scenario constraints, and there are large differences in function definitions and performance requirements. It is currently in the preliminary stage of sorting out the general requirements and clarifying the architecture. As the commercial scale of cloud-edge-terminal collaboration, edge computing, multi-device collaboration, and other multi-dimensional collaboration systems continues to grow, edge computing's share of deployments will continue to increase.

Exploration has begun on industrializing cutting-edge technologies. Interdisciplinary cross-fertilization between computing technology and mathematics, physics, biology, etc., has resulted in active innovation of disruptive cutting-edge computing technologies such as in-memory processing, optical computing, and quantum computing, which have become important directions for future exploration. Industrialization is starting to be explored in some of these fields. In-memory processing architecture, which realizes computation in the memory unit, is expected to overcome the "memory wall" bottleneck. A number of startups such Witmem (知存科技) have emerged in the industry, and related products have been applied in Internet of Things (IoT) and wearable devices. Optical computing uses refraction, interference, and other optical characteristics of optical devices to carry out computation. Trials of product prototypes have begun in data centers, commercial products are expected to be launched within two years, and ecosystem construction has become the focus of future development and breakthroughs. Quantum computing has demonstrated compute superiority in solving specific problems such as random circuit sampling and boson sampling, and research institutes are trying to apply it in encryption and decryption, chemical simulation, drug development, and other scenarios.

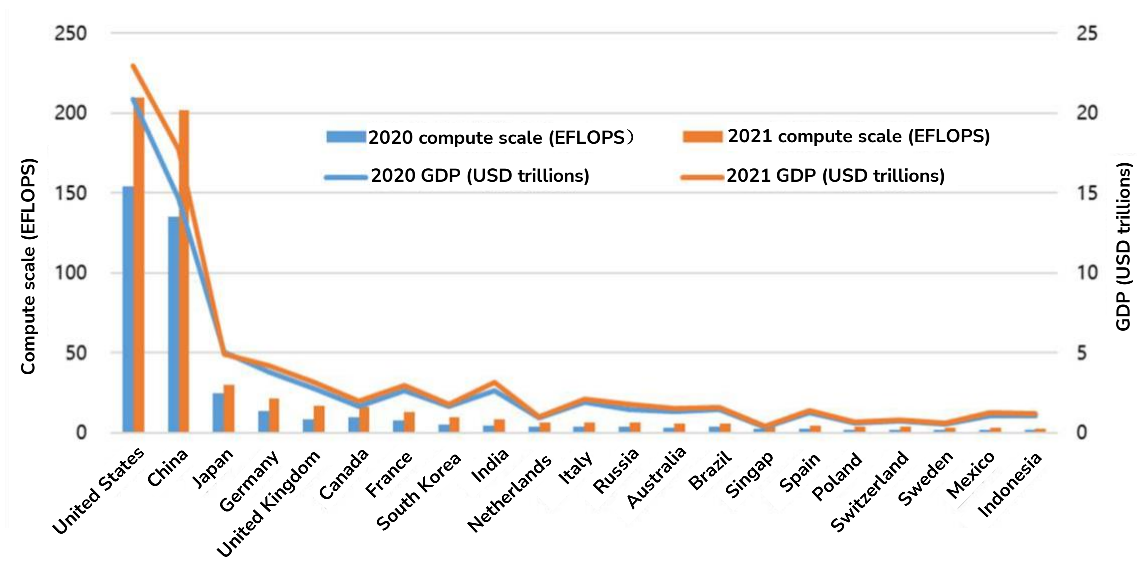
(iv) Compute's empowerment effect continues to deepen

Compute has a significant driving effect on digital economy development and GDP. On one hand, it accelerates the innovation-based development of IT industries such as electronic information manufacturing, software and information technology services, internet, and communications. On the other hand, it boosts the digital transformation and upgrading of traditional industries such as manufacturing, transportation, retail, etc., bringing extended benefits such as industrial output value growth, production efficiency enhancement, business model innovation, user experience optimization, and so on.

Compute is driving digital industrialization. As an important underlying support for the core industries of the digital economy, compute is currently driving upstream and downstream production chain development. This is especially obvious in the fields of integrated circuits, servers, and cloud computing. Integrated circuits: Global sales of computing and memory-related integrated circuits in 2021 was about US$200 billion, an increase of more than 20% over the previous year. Servers: Investment in data center infrastructure continued to rise in 2021, with global server market shipments and sales of 13.539 million units and US$99.22 billion, respectively, up 6.9% and 6.4% year-on-year. Cloud computing: Cloud-native technologies continue to be implemented, driven by the digital transformation of industries enabled by compute. This in turn is driving all-round improvement in technical architectures, application performance, and cloud-derived benefits. The global cloud computing market thus maintained its high growth and rapid innovation trends, reaching US$408.6 billion in 2021, a year-on-year increase of 29.0%.

Compute has become the key engine behind the steady growth of industrial digitalization. Continuous investment in compute has provided the original impetus for the digital transformation of industry, and established a solid foundation for achieving productivity improvement, service capability optimization, and business model innovation. The manufacturing industry's digital transformation in particular has depended on compute to a high degree, and the improvement in production efficiency has been relatively significant. Investment in compute represented by cloud computing, edge computing, and intelligent computing has helped create a highly synergistic smart manufacturing ecosystem. International manufacturing giant Siemens successfully built the first factory based entirely on the digital enterprise concept in 2021, using digital twin technology in planning, analysis, simulation, testing, verification, and other processes. With the added support of strong compute, it achieved a 50% increase in replenishment speed, a 40% increase in efficiency of use of space, a 30% increase in mass production flexibility, and a 20% increase in productivity.

The scale of compute in countries around the world shows a positive correlation with the level of economic development. At present, with the continuous consolidation of the compute base, compute's role in supporting the development of the digital economy and the digital transformation of thousands of industries is becoming more and more prominent, and compute has become an important indicator for measuring the degree of economic and social development of a region. In particular, compute has a significant driving effect on digital economy development and GDP. Global compute grew by 44% in 2021, and the digital economy and nominal GDP grew by 15.6% and 13%, respectively. The scale of compute in countries around the world is closely related to the level of economic development: The higher the level of economic development, the larger the scale of compute. In 2021, seventeen of the top 20 countries ranked in terms of compute were among the 20 top-ranking economies globally, and the top five rankings were the same. The UK, the Netherlands, Italy, Australia, and Singapore improved their compute rankings compared to 2020.



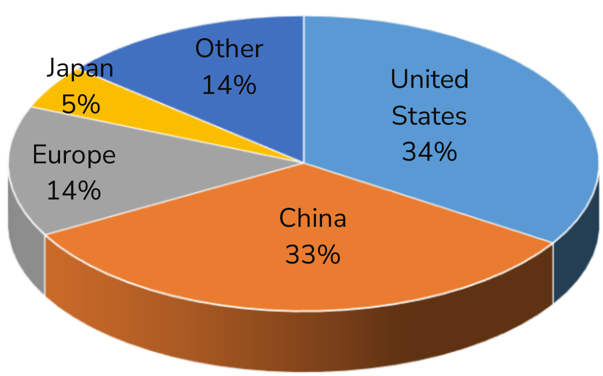
Compute ranking	2020	1	2	3	4	6	5	7	8	9	11	13	10	14	12	16	15	17	18	19	20	21
	2021	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
GDP ranking	2020	1	2	3	4	5	9	7	10	6	17	8	11	13	12	35	14	21	18	22	15	16
	2021	1	2	3	4	5	9	7	10	6	17	8	11	13	12	35	14	21	20	22	15	16

Sources: CAICT, IDC, Gartner, World Bank

Figure 4 Relationship between the scale of compute and GDP globally in 2021

(v) Compute competition continues to intensify

Major countries and regions around the world are deepening their compute development paths. Compute has become an important way for countries to seize development opportunities, and major countries and regions around the world are stepping up their strategic deployment processes. The United States attaches great importance to the development of emerging technologies in compute. By updating technology lists and leading the directions of technological innovation, the United States has continued to consolidate its global leadership in compute technology. In February 2022, the U.S. White House released a new version of the *Critical and Emerging Technologies List* covering fourteen key technologies, such as advanced computing, and five emerging technology areas. Under advanced computing, it lists six subfields: supercomputing, edge computing, cloud computing, data storage, computational architectures, and data processing and analysis. Japan formulates quantum and AI technology development strategies at the national level. For quantum technology, it has set up eight "quantum technology innovation bases" and established the "Quantum STrategic industry Alliance for Revolution" (Q-STAR) to promote the R&D and industrialization of quantum technology. In its *Vision of Quantum Future Society*, Japan proposed that the first Japanese-made quantum computer will be built by 2022, and the goal is for the number of quantum technology users to reach 10 million by 2030. In terms of AI technology, Japan released the *AI Strategy 2022*, which accelerates the development of the AI industry and technology by focusing on five strategic objectives: human resources, industrial competitiveness, technology systems, international cooperation, and crisis response. The EU has continuously increased investment in compute infrastructure construction and key technology R&D. In September 2021 the EU said it planned to invest US$177 billion in data infrastructure, 5G, quantum computing, and other fields; in February 2022 the EU released the *European Chips Act* with an investment of more than 43 billion euros, proposing to focus on chip technologies such as new generation processors and AI and edge computing chips, and to develop FinFETs, GAA transistors, fully depleted silicon on insulator (FD-SOI) technology, and other semiconductor process technologies, so as to strengthen the EU's competitive advantages in high-end chip design and semiconductor production processes, and increase the EU's chip production capacity from 10% of the world's total at present to 20% by 2030.



Sources: CAICT, IDC, Gartner, TOP500

Figure 5 Distribution of global compute in 2021

Global compute competition is intensifying. In terms of compute levels, the shares of the United States, China, Europe, and Japan in global compute are 34%, 33%, 14% and 5%, respectively. As for basic compute, the United States and China are the first echelon in global competition, and the gap between China and the United States is shrinking continuously. The United States ranks first in share of global basic compute, with 37%, while China ranks second with a 26% share; in intelligent compute, China and the United States are in the lead, with 45% and 28% shares of global compute, respectively; in supercomputing, the United States, Japan, and China have obvious advantages when it comes to indicators of overall performance, with 48%, 22%, and 18% shares of total compute, respectively, and China is in the lead in terms of absolute quantity. **From a global point of view**, in the current complex and forbidding international environment, countries around the world have accelerated the construction of local compute industries, reshaping the competitive pattern of global production and supply chains, and posing new challenges for compute technology innovation and industrial ecosystems. European countries, the United States, Japan, and other developed countries have issued relevant policies to further protect the security of their local production and supply chains, in order to maintain their technological advantages in computing chip design, key raw materials, semiconductor manufacturing equipment, etc., and to consolidate and improve the global competitiveness of their enterprises. Every country around the world should start from its own national conditions, development stage, and resource endowment, comprehensively draw on the useful experiences of other countries in the international arena, and figure out a feasible path for its own compute development, while actively promoting international cooperation.

II. China has opened a new chapter of compute-empowered digital economy

2021 was the opening year of the "14 Five-Year Plan" and the starting year of China's construction of the new pattern of development (新发展格局). While faced with a complex and grave international environment and multiple tests such as multiple pandemic outbreaks domestically, China has taken full advantage of its super-large-scale market to achieve significant increases in compute scale and supply levels. The industry-empowering benefits are increasingly visible, and the development environment is being continuously optimized, helping to promote the vigorous development of the digital economy.

(i) Compute scale continues to expand, and intelligent compute has become a growth driver.

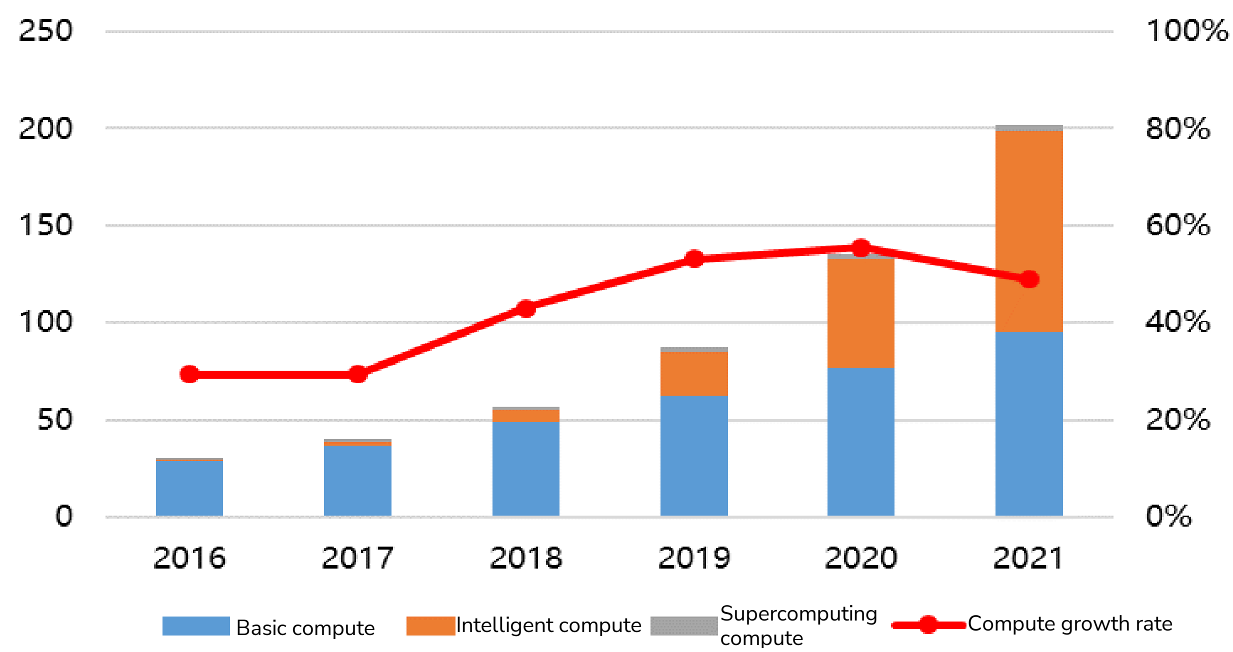
On the infrastructure side, deployment of data centers, intelligent computing centers, and supercomputing centers has accelerated. With the deployment of national hubs and nodes of the nationwide integrated compute network (全国一体化算力网络国家枢纽节点) and the advancement of the "East-West Compute Transfer" project, China's compute infrastructure construction and application have maintained rapid development. According to data from the Ministry of Industry and Information Technology (MIIT), the scale of China's infrastructure compute has reached 140 EFLOPS, ranking second in the world. **First, data center scale has increased significantly.** By the end of 2021, the total scale of data center racks in use in China exceeded 5.2 million standard racks, with an average availability rate (上架率) of more than 55%, and the scale of in-use data center servers was 19 million, with a storage capacity of 800 EB (1 EB = 1024 PB). Power usage effectiveness (PUE) continues to improve. The PUE of advanced green data centers in the industry has been reduced to about 1.1, reaching the world's advanced level. **Second, the deployment of intelligent computing centers has accelerated.** According to statistics from the Intelligent Computing Power Alliance (ICPA; 人工智能算力产业生态联盟), as of June 2022, nearly 20 AI computing centers had been put into operation nationwide, and more than 20 additional AI computing centers were under construction. Localities rely on smart computing centers both to provide enterprises with inclusive compute to support local scientific research innovation and talent cultivation, and to combine with local industrial characteristics to accelerate AI application innovation and aggregate AI industrial ecosystems. For example, the Wuhan Artificial intelligence Computing Center has incubated a succession of large models such as Zidong Taichu (紫东.太初) and Wuhan.LuoJia (武汉.LuoJia), accelerating the promotion of AI in multi-modal interaction, remote sensing, and other areas of application. **Third, the supercomputing commercialization process continues to speed up.** China's supercomputing has entered an application demand-oriented development stage, and a number of supercomputers developed by server suppliers to provide commercial compute services were ranked near the top of China's high-performance computing (HPC) TOP100 in 2021. At the same time, many domestic supercomputing centers have strengthened the reform of commercialization operations, bringing in professional supercomputing commercialization and operation companies and using cloud service concepts and methods to provide supercomputing resources.

On the equipment supply side, China's compute scale continues to grow. CAICT estimates that the total compute of China's computing devices reached 202 EFLOPS in 2021, accounting for about 33% of the global total, with a growth rate of more than 50% for two consecutive years, which was higher than the global growth rate. **Basic compute has grown steadily**. The scale of basic compute[8] grew 24% to 95

---

[8] The scale of basic compute is estimated based on the total compute of servers in China over the last 6 years. China's basic server compute = $\sum_{\text{last 6 years}}$ (annual scale of server shipments*average server compute in the current

EFLOPS in 2021, accounting for 47% of China's compute. General purpose server shipments reached 3.749 million units, representing a year-on-year increase of 7%, and cumulative shipments over six years reached 19.6 million units. **Intelligent compute has grown rapidly**. The scale of intelligent compute[9] reached 104 EFLOPS in 2021, with a growth rate of 85%. It accounted for 50% of the compute in China, and has become the driving force for rapid compute growth. AI server shipments in particular saw year-on-year growth of 59% to reach 230,000 units, and cumulative six-year shipments exceeded 500,000 units. **Supercomputing compute has continued to improve.** Supercomputing compute scale[10] grew at a rate of 30% to reach 3 EFLOPS. In particular, among China's high-performance computers in the TOP100, the top-ranking computer's performance in 2021 was 1.34 times that of the previous year. Lenovo, Inspur, and Sugon were the top three in China.



Sources: CAICT, IDC, Gartner, TOP500, HPC TOP100

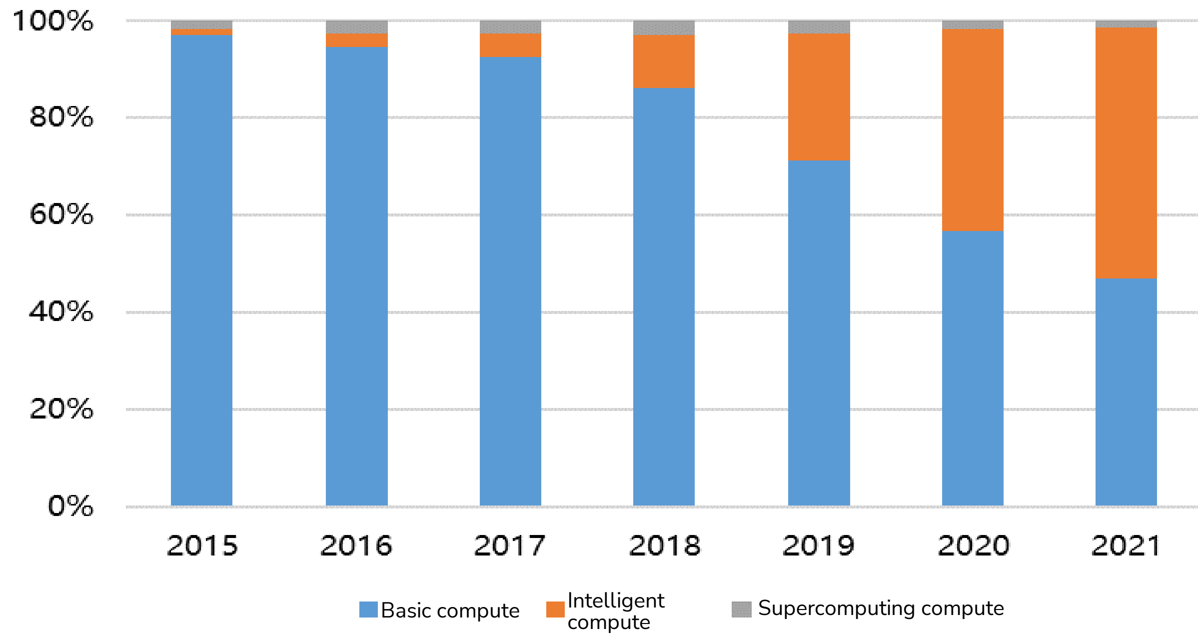
Figure 6 Scale and growth rate of China's compute

---

year).

[9] The scale of intelligent compute is estimated based on the total compute of AI servers in China over the last 6 years. China's intelligent compute = $\sum_{\text{last 6 years}}$ (annual scale of AI server shipments*average AI server compute in the current year).

[10] The scale of supercomputing compute is estimated mainly based on TOP500 global supercomputer and China HPC TOP100 data, and by referring to relevant data from supercomputer manufacturers.



Source: CAICT

Figure 7 Internal structure of China's compute

(ii) The supply level has risen significantly, and advanced computing innovations are emerging

The compute industry is accelerating its growth and upgrading. After years of development, China has formed a computing industry with a relatively complete system, enormous scale and volume, and active innovation, and its importance in the global industrial division of labor continues to rise. At present, China's computing industry accounts for about 20% of the electronic information manufacturing industry, with more than 2,300 large-scale enterprises, and the new industry development pattern of "innovation and breakthroughs, compatibility and inclusion" is being accelerated. **First**, [China's] market share in whole machines is rising. In the general purpose computing field, according to IDC data, Inspur, H3C, Dell, Lenovo, and Huawei hold the top five spots in China's server market, and the combined market share of domestic brands is nearly 75%. In the field of general purpose computing, according to IDC data, Inspur, H3C, Enginetech (安擎), and Huawei hold the top three places in the AI server market in China, and the combined market share of domestic brands is 85%. In the high-performance computing field, China's supercomputing systems have maintained global leadership in terms of both share and the total installed capacity of manufacturers. **Second**, the industry ecosystem has continued to improve. Domestic chips have begun to achieve significant scale. Large-scale application of CPUs based on x86, ARM, and independent architectures continues to deepen, and the iterative optimization of AI chips from Baidu (百度), Cambricon (寒武纪), and others is accelerating. Application of domestic operating systems is gradually penetrating the financial, telecommunications, medical, and other industries. Computing industry ecosystems such as the Kunpeng (鲲鹏) ecosystem and PKS system[11] are being continually perfected, covering key aspects such as the underlying hardware and software, whole-machine systems, and applications.

Compute innovation ability is constantly improving. In 2021, China had over 30,000 invention patent applications in the computing field, and a number of innovations have emerged in advanced computing. **First**, breakthroughs continue to be made in basic software and hardware. Internet vendors have accelerated their independent research on server chips and AI chips: Alibaba launched the Yitian 710 (倚天 710) CPU chip, which has been deployed at scale in Alibaba Cloud data centers; Baidu AI Cloud (百度智能云) and Kunlunxin (昆仑芯) jointly launched second-generation Kunlunxin cloud servers equipped with 2nd-generation Kunlun Core AI chips (昆仑芯 2 代 AI 芯片), improving overall performance by 2-3 times over that of the previous generation. The open-source operating system openEuler has strengthened its updating and iteration, further improving its security, ease of use, and ecosystem capabilities. It has a cumulative installed base of over a million sets. **Second**, the layout of emerging computing platforms is accelerating. New computing system structures and systems, new storage systems, and computing systems based on domain-specific software-hardware collaboration have become hot directions for innovation. Huawei and other companies have launched a diversified computing convergence architecture. The architecture builds a diversified computing software stack incorporating a programming language, compiler, acceleration libraries, and development frameworks, thus reducing the difficulty of developing and deploying diversified compute while enhancing application performance. **Third**, breakthroughs are being made in cutting-edge computing technologies. SWQSIM, a quantum computing simulator developed by a team that included Zhejiang Lab (之江实验室), is based on the Sunway (神威) supercomputer and can provide sustained computational performance of 4.4 EFLOPS. Its research achievements won the 2021 Association for Computing Machinery (ACM) Gordon Bell Prize. The “Jiuzhang 2” (九章二号) optical quantum computer prototype developed by a team that included the University of Science and Technology of China has certain programming capabilities giving it application potential in areas such as graph theory and quantum chemistry.

(iii) Optimization of the development environment continues, with network

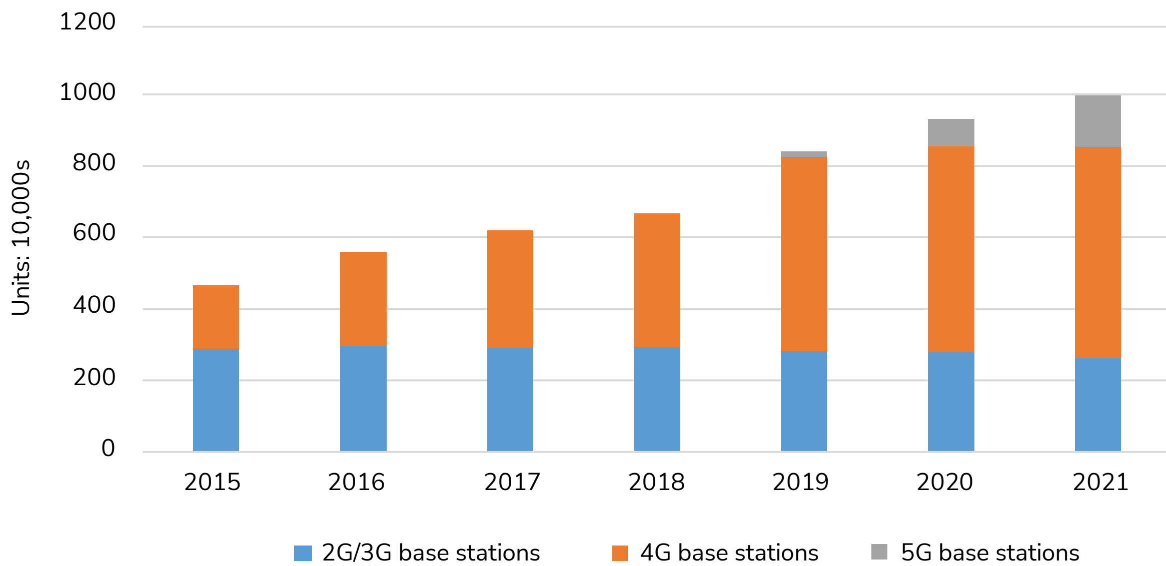
[11] Translator's note: "PKS" is short for "Phytium + Kylin + Security" (飞腾+麒麟+安全). It refers to a computing ecosystem comprised of Phytium processors, manufactured by Chinese provider Founder (方正), the Kylin operating system, developed by researchers at China's National University of Defense Technology (NUDT; 国防科技大学), and Chinese homegrown cybersecurity architecture.

systems opening up new data arteries

The "dual gigabit" network[12] infrastructure has provided solid assurance for compute development. With construction of the nationwide integrated compute network (全国一体化算力网络) as the backdrop, the integration of compute facilities and network facilities has further deepened. The low-latency, high-bandwidth, interconnected network has become the basis for connecting compute facilities in the eastern and western regions, and strengthening compute collaboration in various regions, thereby supporting the spatial layout optimization of China's compute facilities. The capacity of China's network infrastructure has steadily improved, and the structures of the internet backbone network and metropolitan area networks have been continuously optimized. Construction of a number of "East-West Compute Transfer" optical fiber trunk lines (干线光缆) has begun, which will promote the construction of an integrated big data center system and industrial internet big data center system. Inter-provincial outlet bandwidth continues to expand. The average internet inter-provincial outlet bandwidth of provinces reached 42 terabytes per second (Tbps) in 2021, with a growth rate of over 90%. 5G construction has led development. As of the end of 2021, the total number of mobile communication base stations nationwide had reached 9.96 million, of which 5.9 million were 4G base stations, and the cumulative number of 5G base stations completed exceeded 1.42 million, covering all prefecture-level cities and 98% of counties and districts. There were 10.1 5G base stations for every 10,000 people, with a coverage rate of 24%, an increase of 11 percentage points. As the network infrastructure continues to improve, the boundaries between cloud, edge, and terminal have blurred, and the ability to achieve synergy between compute and networks has gradually strengthened, thereby supporting a diversified compute development system.

---

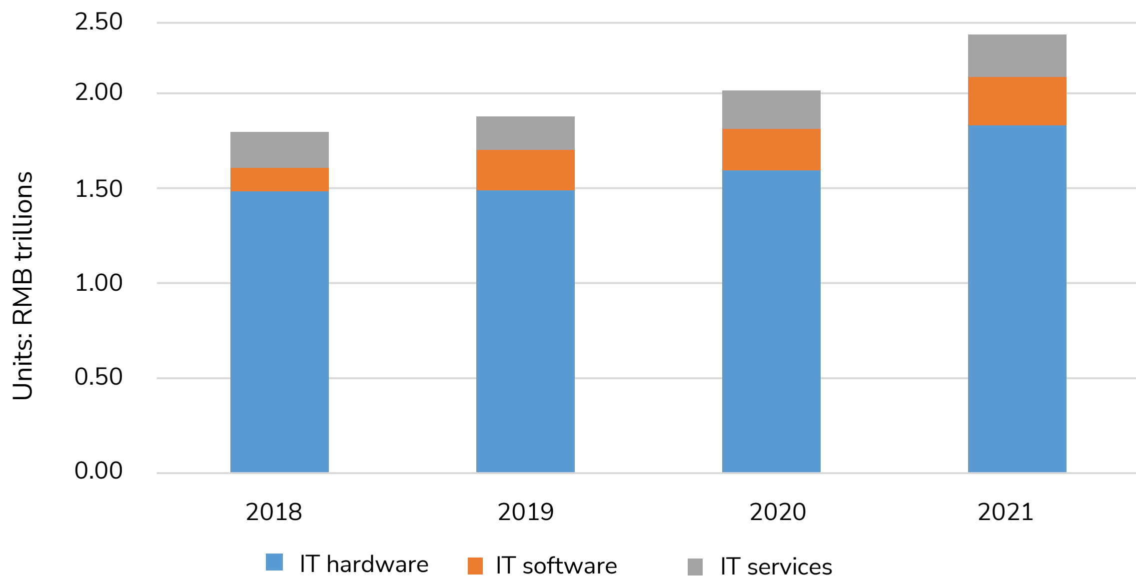
[12] Translator's note: The "dual gigabit" networks (“双千兆”网络) refer to 5G mobile networks and fifth generation fixed networks (F5G; 千兆光网).



Source: MIIT

Figure 8 Development of mobile communication base stations in China

Continuous growth of compute investment has injected impetus for the sustained rapid and steady growth of compute. Cloud computing, big data, AI, and other new generation information technologies are accelerating their penetration of, and integration with, various economic and social fields. The industry digital transformation process has been accelerated and upgraded, a deep fusion of informatization (信息化) and industrialization has developed, and the proportion of enterprises that have digitalized continues to increase. This has driven the continuous growth of China's expenditures in the fields of computing hardware, software, and services, which has in turn provided a strong impetus for compute development. IDC data show that IT spending in China grew 14% year-on-year to reach RMB 2.3 trillion in 2021, driving China's strong economic recovery. At the same time, driven by China's "East-West Compute Transfer" project, compute investment showed a trend of shifting from east to west. Western region investment has increased significantly, with eight compute hub regions becoming hot spots for investment. The compute investment spatial layout is being optimized to promote the intensification, scale expansion, and green development of China's compute facilities.



Sources: CAICT, IDC

Figure 9 IT hardware, software, and services spending in China

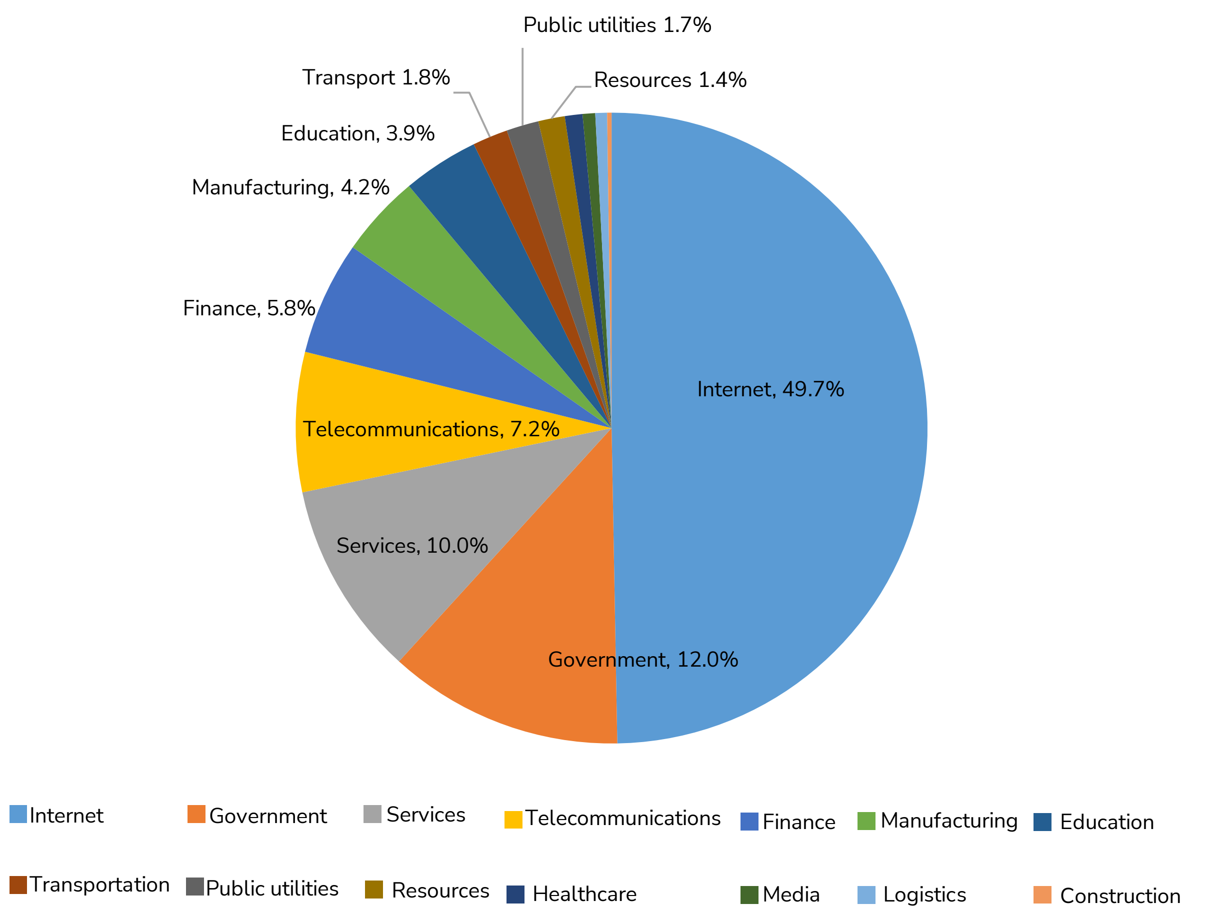
The increasingly perfected data resource system is opening up compute development "main arteries." Data is the new means of production in the digital economy era and the basis for compute development. Rapid growth of data production, continuously increasing data processing requirements, and the open circulation of data resources have spurred growing demand for the development and application of compute, indirectly becoming boosters of compute development. China's data production is currently growing rapidly, and according to data from the *National Data Resources Survey Report*, from 2017 to 2021, China's data production grew from 2.3ZB to 6.6ZB, accounting for 9.9% of the global share and ranking it second in the world. The pace of data resource opening and sharing has been accelerating. According to the data in *Digital China Development Report (2021)*, between 2017 and 2021, the number of provincial-level public data open platforms across the country increased from 5 to 24, and the number of publicly available effective datasets increased from 8,398 to nearly 250,000. Local governments are actively exploring data governance rules, incubating data markets, and facilitating the flow of data for trade, development, and utilization, thereby helping to unleash the value of data factors of production.

(iv) The transformation to intelligentization and the transition to digitalization (智改数转) are deepening under the continuous stimulation of empowerment effects

With the continuous enhancement of China's compute support capacity, new business formats (新业态) and new models of compute application are emerging at an accelerating pace. On the one hand, compute has been deeply integrated into the internet, telecommunication, finance, manufacturing, and other industries, and has provided support for the intelligentization and digital transformation of all kinds of industries and fields. On the other hand, virtual reality and the metaverse are expected to become an important lever for boosting information consumption, or will become a driving force to propel the next stage of compute growth.

Compute is injecting strong momentum for the digital transformation of industries. With regard to application fields, the application of compute in China is expanding from traditional fields such as the internet and e-government services to traditional industries like telecommunications, finance, manufacturing, and education. The internet industry, with its constantly increasing data processing and model training requirements, is the industry with the largest demand for compute, accounting for 50% of overall use. The government industry, with its increasing investment in digital government, safe cities (平安城市), and other fields, ranks second in terms of compute share. The service, telecommunications, financial, manufacturing, education, and transportation industries rank from third to eighth. Telecommunications and finance in particular are traditional industries in China with extensive compute application, and enterprises in these industries have high degrees of digitalization. As for the manufacturing industry, as the industrial internet continues to mature, manufacturing application scenarios will proliferate, and its potential for increased compute demand is relatively great. **In terms of support capacity,** compute application scenarios are being extended, from virus research, geological exploration, aerospace, and other S&T exploration, to environmental monitoring, precision marketing, intelligent scheduling, and other fields, which is energizing data factor of production-driven innovation. Compute continues to play an important support role in epidemic prevention and control. More than 55.6 billion Telecom Big Data Itinerary Card[13] user inquiries have been accumulated, with the number doubling every six months on average, strongly supporting epidemiological surveys of personnel (人员流调) and the resumption of work and production.

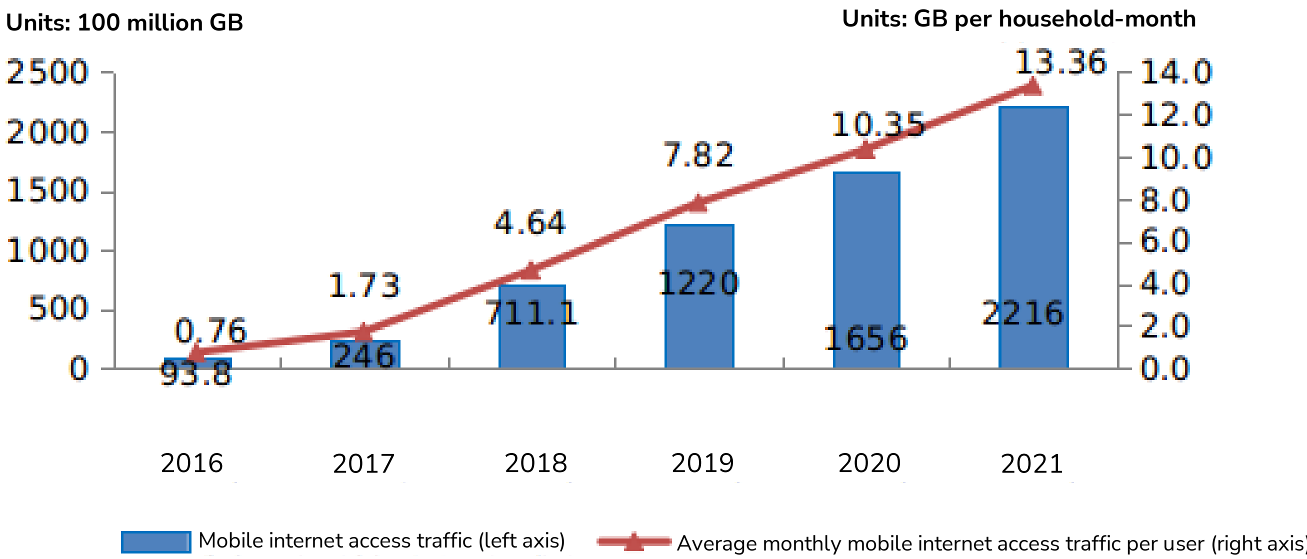
[13] Translator's note: The "Telecom Big Data Itinerary Card" or "Itinerary Card" (通信大数据行程卡; 通信行程卡; 行程卡) is a Chinese smartphone app that digitally verifies users' travel histories for COVID-19 monitoring purposes.



Sources: CAICT, IDC

Figure 10 Distribution of compute applications among industries in China

Compute is helping promote the continued expansion and upgrading of information consumption. First, the scale of mobile data traffic consumption continues to expand. Compute is what actually handles the massive loads of mobile internet data, and the growth of data traffic is the core driver of rapid growth in compute scale. Mobile internet applications such as smartphones, remote work, online meetings, and mobile payments have driven the growth of back-end compute infrastructure, greatly promoting the booming growth of compute. China's mobile internet traffic saw rapid growth in 2021, with traffic increasing 33.9% from the previous year to reach 221.6 billion GB, of which mobile internet traffic was 212.5 billion GB, up 35.5% and accounting for 95.9% of the total. The average handset data traffic per user per month (DOU) continued to improve, increasing 29.2% over the previous year to an average of 13.36 GB per month, and DOU reached 14.72 GB in December, a record high. **Second, the number of 5G and IoT users is increasing rapidly.** With construction at scale of 5G and IoT, the spread of compute from the cloud to the edge and terminals has accelerated, and edge computing ability has continued to grow. This is promoting the popularization of emerging applications such as HD content, video production and broadcasting, augmented reality (AR) navigation, cloud-based games, smart homes, and so on, which in turn is promoting the growth of 5G and IoT user numbers. In 2021, the total number of mobile phone accounts in China was 1.643 billion, up 48.75 million for the full year, and the penetration rate was 116.3 per 100 people, up 3.4 from the end of the previous year. There were 1.069 billion 4G mobile phone users and 355 million 5G mobile phone users, together accounting for 86.7% of all mobile phone users. The scale of cellular IoT use has continued to expand. The three basic telecommunications enterprises[14] had developed 1.399 billion cellular IoT users, representing a net increase of 264 million over the previous year.



Source: MIIT

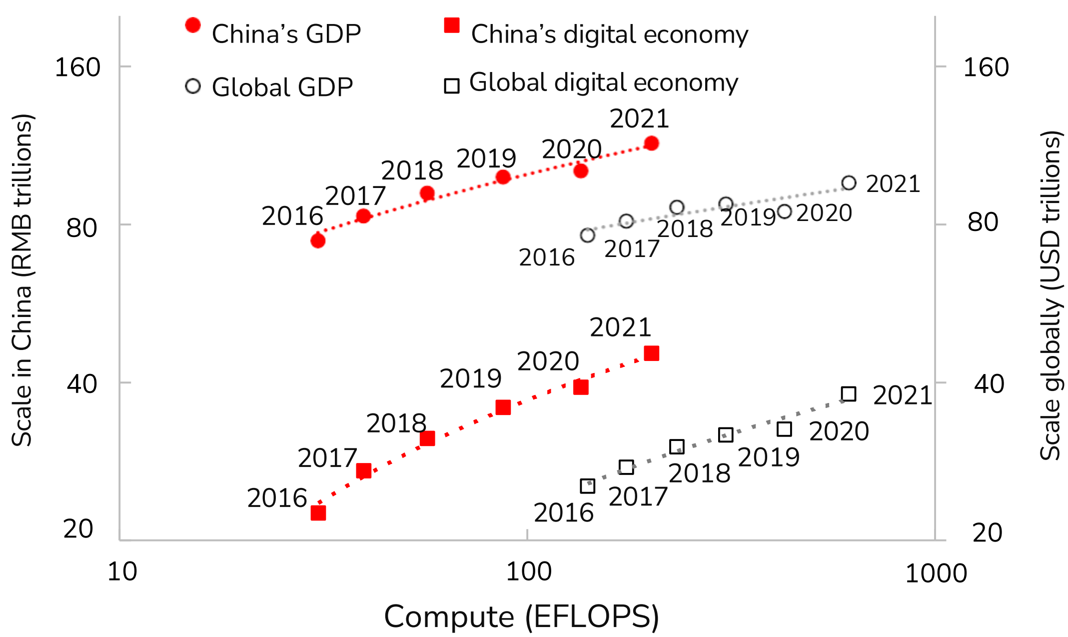
Figure 11 Growth in mobile internet traffic and monthly DOU, 2016-2021

(v) Compute boosts economic growth by helping the digital economy achieve quantitative and qualitative growth

Compute development has boosted the quantity and quality of China's digital economy. In recent years, China has pushed forward the integrated development of compute and applications, promoted the compute-based enabling of thousands of industries, and facilitated the in-depth integration of digital technology in the real economy, building new strengths in digital economy development and expanding new

[14] Translator's note: The term "three basic telecommunications enterprises" (三家基础电信企业) refers to the three state-owned companies that dominate the Chinese telecom services market: China Mobile (中国移动), China Telecom (中国电信), and China Unicom (中国联通).

economic growth points (增长点). **Digital industrialization**: China is experiencing a shift from quantitative expansion to qualitative improvement. Value added through digital industrialization in China was RMB 8.35 trillion in 2021, with nominal growth of 11.9% year-on-year, accounting for 18.3% of the digital economy and 7.3% of GDP. Compute is an important foundational support for the core industries of the digital economy, and its pulling effect on the upstream software and hardware industries is increasingly evident. In the electronic information manufacturing industry, nationwide value added by large-scale enterprises increased by 15.7% in 2021 compared with the previous year, the highest growth rate in nearly a decade. The software and IT services industry maintained faster growth, with nationwide revenue by large-scale enterprises increasing 17.7% year-on-year to RMB 9.5 trillion. **Industry digitalization:** The scale of industry digitalization saw nominal growth of 17.2% year-on-year growth to reach RMB 37.18 trillion, accounting for 81.7% of the digital economy and 32.5% of GDP. The digital transformation of industries, led by manufacturing, has played a key role in the growth of the digital economy. China has incubated over 150 large industrial internet platforms, which serve more than 1.6 million industrial enterprises, and the industrial internet has entered a period of rapid growth in integration and application.



Source: CAICT

Figure 12 Relationship between the scale of compute, GDP, and the digital economy, globally and in China, 2016-2021

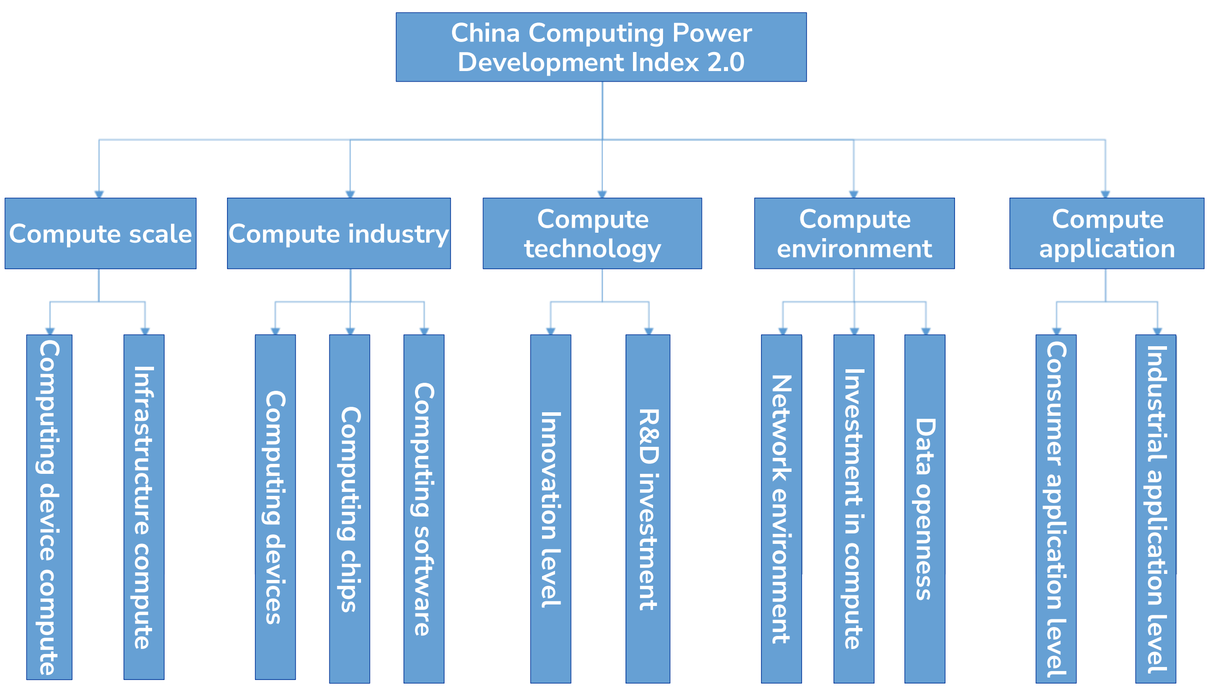
The pulling effect of compute development on China's GDP growth is striking. On the one hand, the scale of compute and the level of economic development show a significant positive correlation. Compute has become the latest key productive force, following heat and electricity. Under its effects, productivity has been released like never before, and the larger the scale of compute, the higher the level of economic development. Statistics show that in 2021, China's compute scale grew by 50%, the digital economy grew by 16%, and GDP grew by 12.8% in nominal terms. Compared with the world as a whole, the contribution of China's compute to GDP growth stands out. In the 2016-2021 period, China's compute grew by an average of 46% per year, the digital economy grew by 15%, and GDP grew by 9%. Globally, compute grew by an average of 34% per year, the digital economy grew by 8%, and GDP grew by 4%. **On the other hand**, the compute industry's role in energizing China's economy, society, and industries is growing stronger by the day. Calculated through the national input-output table model, the scale of the compute industry, as represented by computers, reached RMB 2.6 trillion in 2021, accounting for about 20% of the electronic information manufacturing industry, directly driving RMB 2.2 trillion in total economic output, and indirectly driving RMB 8.2 trillion. With the "East-West Compute Transfer" project fully underway, compute resources have been elevated to the level of water, electricity, gas, and other basic resources. As China's compute infrastructure construction continues to speed up, compute's role in boosting China's economic development will be further enhanced in the future.

III. Evaluation of the China Computing Development Index

The "14th Five-Year Plan" clarifies the direction of China's compute industry development in the next five years by clearly indicating the need to accelerate construction of a nationwide integrated big data center system, strengthen the intelligent scheduling of compute, build a number of national hubs and big data center clusters, and build EFLOPS-scale and 10-EFLOPS-scale supercomputing centers. China has accelerated the formulation of compute plans, and a series of documents have been issued to launch implementation of the "East-West Compute Transfer" project, including *Implementation Plan for Nationwide Integrated Big Data Center Collaborative Innovation System Computing Power Hubs*, *Three-Year Action Plan for the Development of New Data Centers (2021-2023)*, *"14th Five-Year" Plan for the Development of the Information and Telecommunications Industries,"* *"14th Five-Year" Plan for the Development of the Digital Economy*, and *"14th Five-Year" Plan for National Informatization*. The project focuses on building a new information infrastructure system based on new generation information and telecommunications networks, with data and computing facilities as the core, and integrated infrastructure as the main focus. It has bolstered efforts to tackle core computing technologies, promote supply-demand linking (供需对接) of computing resources, and incubate new industries, new business formats, and new modes of computing, and has achieved positive results.

Driven by both demand and policies, all parts of the country are vigorously promoting the development of the compute technology industry, infrastructure construction, and compute application. In order to comprehensively sort through and objectively evaluate the compute development situation in China, and arrive at a more scientific and concrete understanding of China's compute, CAICT has established the China Computing Power Development Index, taking into consideration the characteristics and key influencing factors of compute development. Our institute has further improved the China Computing Power Development Index to evaluate the level of compute development comprehensively and objectively in each province of China, thereby providing effective support for the formulation of compute development policies for the country as a whole and for individual provinces.

Based on the compute development research system, this white paper establishes the China Computing Power Development Index 2.0 from five dimensions: compute scale, compute industry, compute technology, compute environment, and compute application. Compared with the Computing Power Development Index 1.0, the China Computing Power Development Index 2.0 adds two primary indicators—compute industry and compute technology—to strengthen the evaluation of compute supply levels. The China Computing Power Development Index 2.0 is shown in Figure 13.



Source: CAICT

Figure 13 China Computing Power Development Index 2.0

In addition, the secondary indicators in the China Computing Power Development Index 2.0 have also been increased and adjusted accordingly. For compute scale, in addition to taking into account computing device compute (server compute scale, AI server compute scale, and supercomputing compute scale), the infrastructure compute dimension has been added. It mainly takes into account the scale of compute centrally deployed in regional data centers and intelligent computing centers. The added dimension of compute industry mainly takes into account the factors of production of the computing system equipment industry and related basic hardware and software industries. The added dimension of computing technology mainly takes into account R&D investment and innovation level (intellectual property rights) in computing-related technologies. For the compute environment dimension, data openness is added as a factor in addition to network environment and compute investment. Data openness mainly reflects urgent requirements for compute development. The open circulation of data directly and indirectly promotes the growth of compute. The compute application dimension still includes the factors of consumer application and industrial application.

(i) Basis for index establishment

This white paper selects relevant indicators from five dimensions—compute scale, compute industry, compute technology, compute environment, and compute application—to establish the China Computing Power Development Index 2.0. It is based on analyzing the compute development situations globally and in China, synthesizing research on compute measurement and related index systems by IDC,[15] Roland Berger,[16] Huawei, Inspur, and other domestic and foreign institutions and enterprises, and fully seeking expert opinions. The aim is to evaluate the development of compute comprehensively and objectively in China and analyze the current level of compute development in different localities. The indicators for the China Computing Power Development Index 2.0 were selected according to the principles of scientific soundness, representativeness, and independence, taking into account the characteristics and key influencing factors of compute development, as well as the accessibility and comparability of data at the provincial level.

Dimension 1: Compute scale. This is measured mainly based on the two aspects of computing device compute and infrastructure compute. Computing device compute

[15]The *2021-2022 Global Computing Power Index Evaluation Report* jointly released by IDC, Inspur, and Tsinghua University carries out evaluation mainly with four types of indicators focusing on computing capability, computing efficiency, application level, and infrastructure support.
[16]*Ubiquitous Computing Power: The Cornerstone of an Intelligent Society*, jointly published by Roland Berger and Huawei, provides a system of indicators for global compute measurement, and carries out estimation of the overall compute of countries around the world based mainly on cloud, edge, and terminal computing.

is mainly based on the market distribution of compute devices in different regions in the past six years, and measuring basic compute, intelligent compute, and supercomputing compute scale from the three major product categories of general purpose servers, AI servers, and supercomputers, respectively. Basic compute mainly focuses on the scale of server compute in each region, using single-precision floating-point format (FP32) computing capability to measure compute performance.[17] Intelligent compute is mainly focused on the regional scale of AI server compute, using mainstream half-precision floating-point computation format (FP16) computing capability to measure compute performance. Supercomputing compute is mainly based on the internationally well-known TOP500 ranking and China's TOP100 ranking of high-performance computer performance, and referring to the relevant data of supercomputing manufacturers, while adopting double-precision floating-point format (FP64) computing capability to measure supercomputing compute performance. Infrastructure compute is mainly based on CAICT statistics on the computing power of data centers and intelligent computing centers. Compute scale calculations are uniformly converted to compute in single-precision floating-point (FP32) format for statistical purposes.

Dimension 2: Compute industry. The compute industry is measured mainly based on three aspects: computing devices, computing chips, and computing software. The compute industry covers key links in the production chain—devices, chips, and software. The compute industry is the basic foundation for the development of compute. Computing devices mainly focuses on the output of servers and other complete machines, which reflects the computer manufacturing capacity of each region. Computing devices are the actual mainstay carriers of compute loads. Computing chips mainly focuses on the output of integrated circuits such as microprocessors and memory, which reflects the chip production and supply capacity of each region. Computing chips are the foundation and core of compute realization. Computing software focuses mostly on software business revenue, which mainly reflects the level of development of the software and IT service industries in each region. Computing software is key to compute's industry-enabling effect.

Dimension 3: Compute technology. Compute technology is measured mainly based on the levels of compute innovation and R&D investment. Compute technology innovation is the power source of compute development, and enterprises continue to accelerate R&D investment and patent layouts. Compute innovation level mainly focuses on the number of computing invention patent applications and the number of

[17] The general purpose compute of servers is evaluated using single-precision floating-point (FP32) computing capability, where server compute = number of processor chips*number of single-precision floating-point operations performed per clock cycle*processor clock rate*number of processor cores.

computing invention patents granted in each region, reflecting the market value and competitiveness of computing R&D achievements in each region. Compute R&D investment mainly focuses on R&D expenditures in the computer manufacturing industry in each region, which captures the scale and level of S&T investment in compute fields in each region.

Dimension 4: Compute environment. The compute environment is measured mainly based on the network environment, compute investment, and data openness. Continuous optimization of the network environment provides solid support for compute development, while large-scale open data circulation and IT investment generate direct and indirect driving effects on the growth of compute. The network environment focuses on inter-provincial internet outlet bandwidth and 5G mobile base station coverage. It mainly reflects the service capability of each region regarding data transmission to and from other regions in China for business such as data and internet, as well as the edge computing support capability of the region. Compute investment strength focuses on the investment situation for computing hardware, software, and services. AI, autonomous driving, smart cities, and other emerging applications drive the development of compute, spurring the growth of computing hardware, software, and service expenditures. For data openness, we adopt the China Open Data Index[18], which comprehensively evaluates the level of data openness in each region from the dimensions of readiness, platform layer, data layer, and utilization layer. It mainly reflects urgent requirements for compute development.

Dimension 5: Compute application. Measurement of compute application is based mainly on the level of consumer application and industrial application. Compute promotes the development of consumer and industrial applications, while consumer and industrial applications have a pulling effect on the growth of compute. Consumer application level mainly focuses on average monthly mobile internet traffic. There is a significant correlation between mobile internet access traffic and the scale of compute. Compute handles the massive loads of mobile internet data, and the growth of data traffic is the core driver of the rapid growth in compute scale. Mobile internet applications such as smartphones, remote work, online meetings, and mobile payments have driven the growth of back-end compute infrastructure, greatly promoting the booming growth of compute. Industrial application level mainly focuses on industry digitalization. It reflects the application of compute in the internet, manufacturing, finance, and other fields. The industrial internet, fusion between informatization and industrialization (两化融合), smart manufacturing, Internet of Vehicles (IoV), the platform economy, and other new integration-based industries,

[18] Refer to the “2021 China Open Data Index” and *China Local Government Data Openness Report* by the Digital and Mobile Governance Lab (DMG Lab; 数字与移动治理实验室) at Fudan University.

models, and business formats provide broad space for compute development.

(ii) Index system establishment

To obtain a composite index of China's computing power development, the different stages of the evaluation process, including determination of weights, assignment of values, and calculation of scores, were performed for each indicator in accordance with scientific research and analysis methods. The process of forming the composite index can be divided into four stages:

1. Index system formation: On the above basis for establishing the index system, experts' opinions were sought, the current situation in terms of evaluating China's compute development was sorted through, and the characteristics and key influencing factors of compute development were factored in to build the computing development index system from the five dimensions of compute scale, compute industry, compute technology, compute environment, and compute application. The index involves five primary indicators, including compute scale and compute industry, twelve secondary indicators, such as computing device compute and infrastructure compute, and 16 tertiary indicators, such as basic compute scale and intelligent compute scale.

Table 1 China Computing Power Development Index system

Level 1 metrics	Level 2 metrics	Level 3 metrics	Units
Computing power scale	Computing device compute	Basic compute (server compute) scale	EFLOPS
		Intelligent compute (AI server compute) scale	EFLOPS
		Supercomputing compute (supercomputer compute) scale	EFLOPS
	Infrastructure compute	Compute scale of data centers and intelligent computing centers	EFLOPS
Compute industry	Computing devices	Computing device output	10,000 units
	Computing chips	Integrated circuit output	10,000 pieces
	Computing software	Software business revenue	RMB 100 million
Compute technology	Innovation level	Computing invention patent applications	Number granted
		Computing invention patents granted	Number granted
	R&D investment	Computer manufacturing industry R&D expenditures	RMB 100 million
Compute environment	Network environment	Inter-provincial internet outlet bandwidth	Tbps
		5G coverage rate	%

Level 1 metrics	Level 2 metrics	Level 3 metrics	Units
	Investment in compute	IT expenditures	RMB 100 million
	Data openness	China Open Data Index	/
Compute application	Consumer application level	Average monthly mobile internet traffic	EB
	Industry application level	Scale of industry digitalization	RMB 100 million

Source: CAICT

2. Determination of indicator weights: The analytic hierarchy process (AHP), which is based on expert scoring methods, was used to obtain the relative weights of all the primary, secondary, and tertiary indicators in the evaluation index system.

3. Assigning values to the indicators according to the actual situations of regions: The actual values of each indicator were obtained according to the actual compute development situations in 31 provinces[19], and the data were standardized to obtain the assigned values of each indicator.

4. Calculating the composite index: Lastly, based on the specific assigned values of indicators and corresponding weights, the composite index of each region is finally generated.

(iii) Evaluation of China's compute development

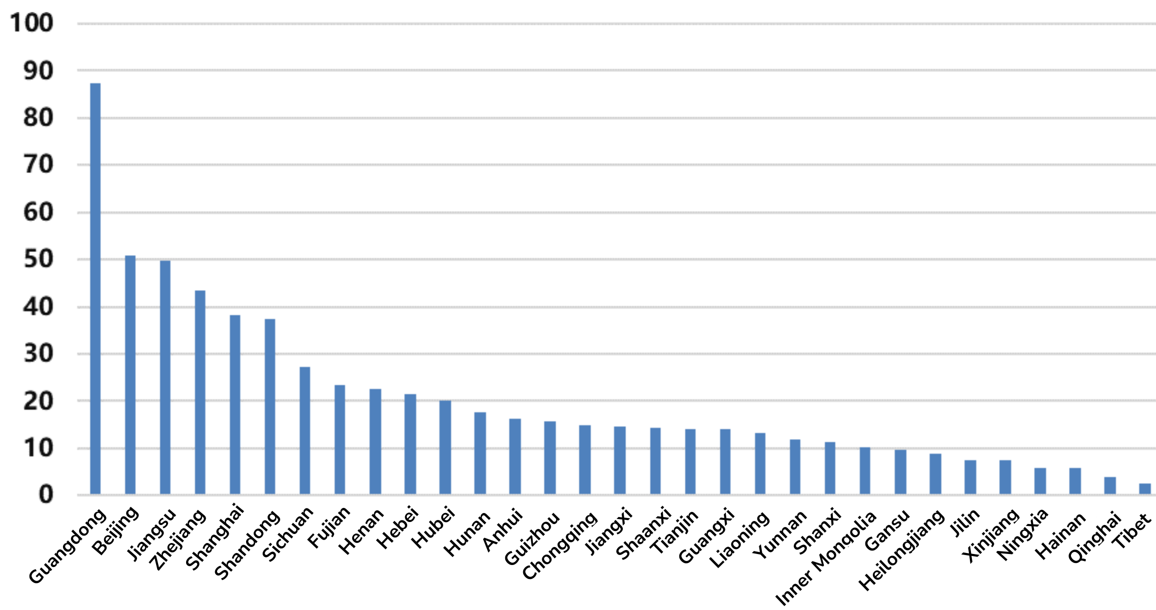
1. Compute development index

Beijing-Tianjin-Hebei, the Yangtze River Delta Economic Zone, the Guangdong-Hong Kong-Macao Greater Bay Area, and the Chengdu-Chongqing Economic Circle are at leading levels. Overall, the compute development index is generally higher in Guangdong, Beijing, Shanghai, and their neighboring provinces. The six highest ranking are Guangdong, Beijing, Jiangsu, Zhejiang, Shanghai, and Shandong, with compute development index values over 35. Sichuan, Fujian, Henan, and Hebei are ranked from seventh to tenth, with compute development index scores of over 20. Beijing, Shanghai, and Guangzhou, as well as their neighboring provinces, have seized upon opportunities for compute development. They have continuously strengthened innovation in key technologies for advanced computing, enhanced the supply capacity of the compute industry, and accelerated the construction of compute infrastructure, thereby creating a favorable environment for the development of compute and actively promoting the extension of compute innovations and

[19] Due to data availability and data continuity limitations, this report's calculations do not include China's Hong Kong, Macao, and Taiwan regions.

applications. Their overall development index numbers are thus generally higher.

The compute development of core provinces in the central and western regions is gathering steam, but still faces many weaknesses. Compute demand in Beijing, Shanghai, Guangzhou, and neighboring provinces is strong. Land and energy being scarce, it is difficult to develop compute on a large scale. Supply has fallen short of demand, and compute development is subject to certain limitations. The central and western provinces have sufficient resources, and with the comprehensive advancement of the national "East-West Compute Transfer" project, the core provinces of Guizhou, Inner Mongolia, Gansu, and Ningxia have greater potential for compute development, but there are still many development bottlenecks, including insufficient drivers of technological innovation, weak industrial foundation, development environments in need of optimization, little demand for compute, and other problems.

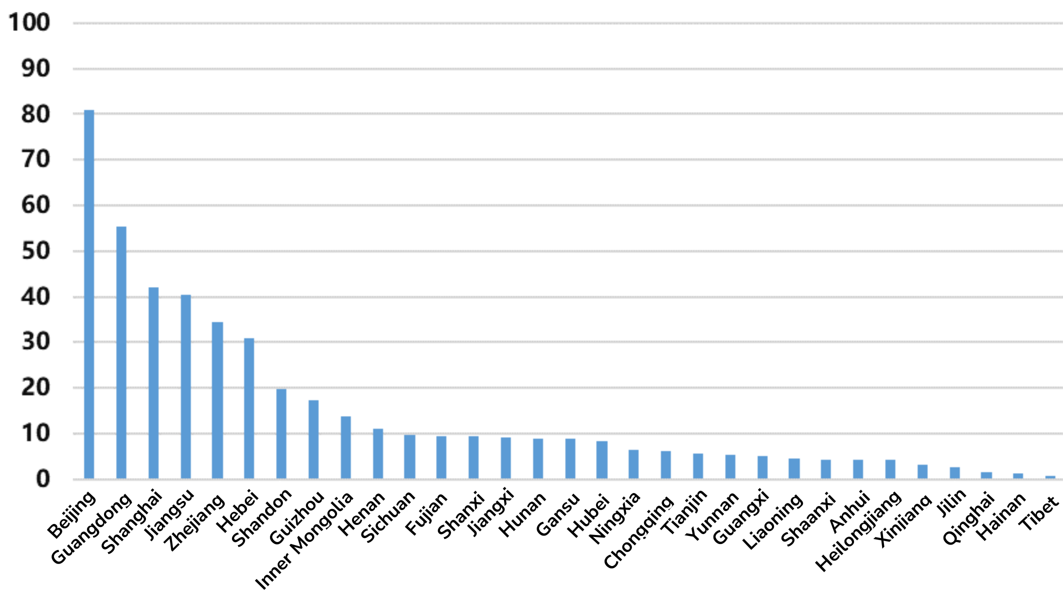


Source: CAICT

Figure 14 Compute development index of some Chinese provinces in 2021

2. Compute scale sub-index

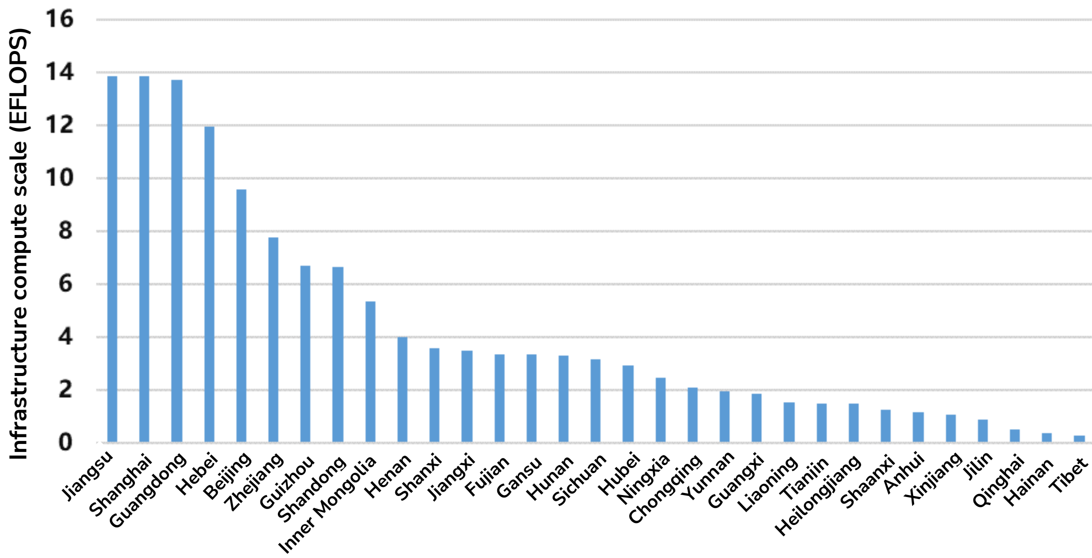
The compute scale sub-index scores of Beijing, Shanghai, Guangzhou, and their surrounding provinces and regions are particularly high, and western region hubs are improving. As in 2020, hotspots Beijing, Shanghai, and Guangzhou remain relatively high in terms of compute scale index. Beijing, Guangdong, and Shanghai are the top three regions, and Jiangsu, Zhejiang, Hebei, and Shandong make up the first echelon below them, with a compute scale index of 20 or more. Provinces and regions around the three hotspots of Beijing, Shanghai, and Guangzhou have been gradually taking up the spillover demand from those regions, and places such as Jiangsu, Zhejiang, and Hebei are gradually narrowing the gap with them. Guizhou, Inner Mongolia, Henan, Sichuan, Fujian, Shanxi, Jiangxi, Hunan, Gansu, Hubei, Ningxia, Chongqing, and Tianjin are in the second tier within the top 20 in the compute scale index. Guizhou, Inner Mongolia, Gansu, Ningxia and other western "East-West Compute Transfer" hubs have strengthened their layout of compute, and their compute scale index numbers have increased significantly.



Source: CAICT

Figure 15 Compute development index of some Chinese provinces in 2021

Compute scale has been steadily rising everywhere. In terms of the distribution of computing device compute, Beijing, Guangdong, Zhejiang, Shanghai, and Jiangsu are the top five in the server and AI server markets, with market shares totaling 75% and 90%, respectively. In terms of supercomputing compute, Jiangsu, Tianjin, Shandong, Beijing, Guangdong, Sichuan, Henan, Hunan, and Shanghai are the top-ranking provinces. With strong demand for compute in the eastern region, computing device compute mainly comes from the eastern provinces, accounting for nearly 90%. **In terms of the distribution of infrastructure compute**, the top ten provinces in China's infrastructure compute scale in 2021 were Jiangsu, Shanghai, Guangdong, Hebei, Beijing, Zhejiang, Guizhou, Shandong, Inner Mongolia, and Henan, all of which exceeded 4 EFLOPS. In Shanghai, Guangdong, Jiangsu, and Hebei in particular, infrastructure compute was over 12 EFLOPS. Accumulation in earlier stages has given Beijing, Shanghai, Guangzhou, and their neighboring provinces a clear advantage in infrastructure compute scale, but due to limited power and land resources and tightening of policies, the central and western regions are catching up at accelerating rates.



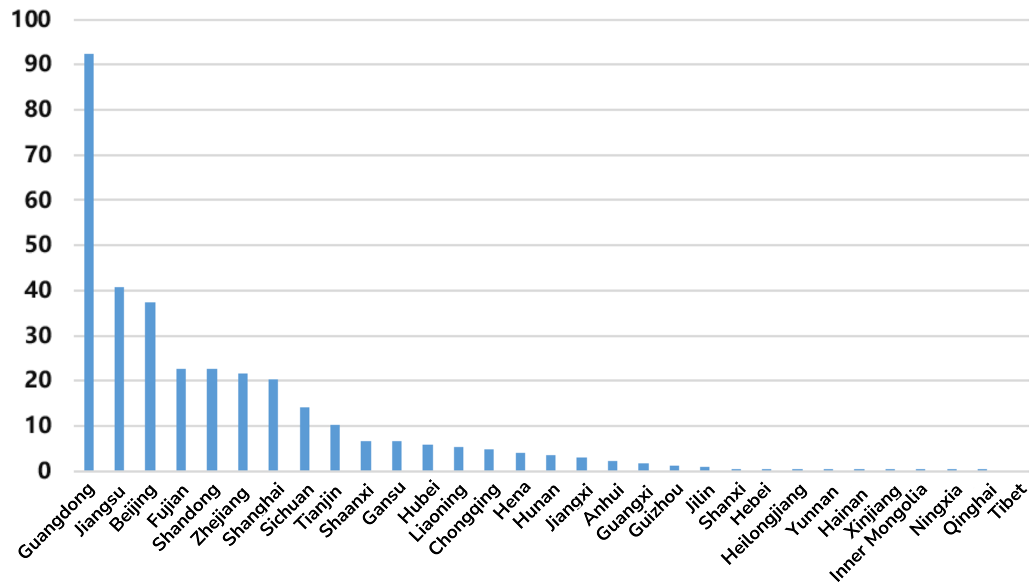
Source: CAICT

Figure 16 Infrastructure compute scale of some Chinese provinces in 2021

3. Compute industry sub-index

China's compute industry has formed a development pattern of three poles and multiple points, and the industry's leading and driving role has become clearer. Overall, compute industry development levels are relatively high in the Guangdong-Hong Kong-Macao, Yangtze River Delta, and Beijing-Tianjin-Hebei regions, represented by Guangdong, Jiangsu, and Beijing, with compute industry sub-index numbers over 30, forming three growth poles. Fujian, Shandong, Zhejiang, Shanghai, Sichuan, Tianjin, and Shaanxi rank among the top ten, forming a new pattern of compute industry development supported by multiple points. **Computing devices**: Guangdong, Fujian, Shandong, Jiangsu, and Tianjin, which lead in the production of computing equipment such as servers (excluding microcomputers), make up the top five. Guangdong is the absolute leader, with computing device production exceeding 9 million units in 2021, close to half of the country's output. An electronic information industry belt has formed on the east bank of the Pearl River, and it is driving the regional industry's coordinated development. **Computing chips**: Jiangsu, Gansu, Guangdong, Shanghai, Zhejiang, and Beijing are ahead in computing chip production. Jiangsu in particular has reached a production level of 118.6 billion integrated circuit (IC) chips, accounting for nearly 30% of the nationwide total. Linkages have formed with Shanghai, Zhejiang, and Anhui, driving IC industry development in the Yangtze River Delta region. **Computing software**: Beijing, Guangdong, Jiangsu, Zhejiang, and

Shandong are at the forefront in software business revenue, with Beijing ranking first at RMB 186.61 billion in 2021, generating strong radiating and driving effects on compute industry development in other provinces in China.

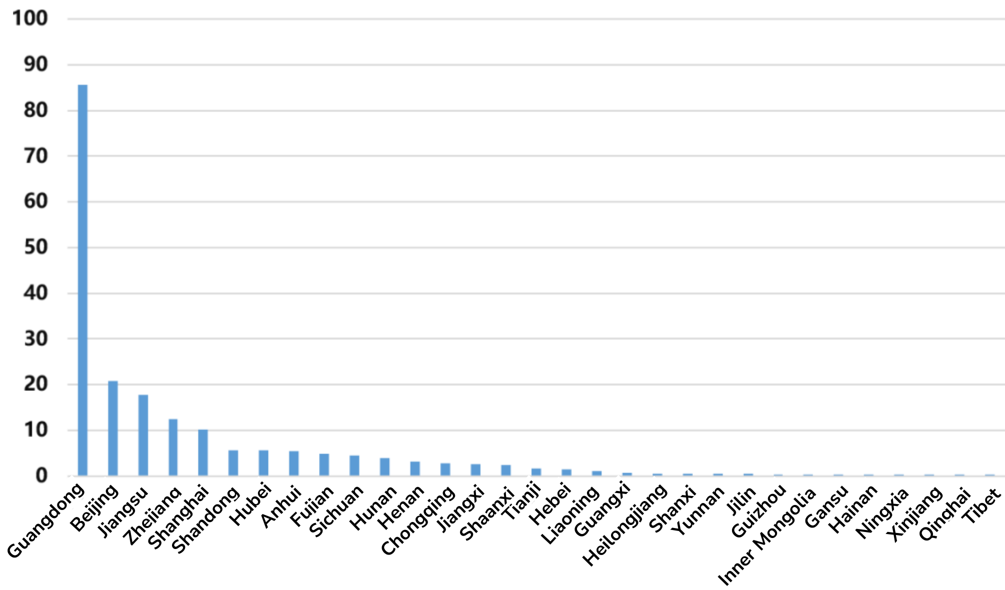


Source: CAICT

Figure 17 Compute industry sub-index levels of some Chinese provinces in 2021

4. Compute technology sub-index

Eastern provinces have higher compute technology sub-index scores, and are in the lead in compute innovation and R&D investment. Overall, Guangdong is at the top in terms of the compute technology sub-index, and its levels of compute innovation and R&D investment are the highest in the country. Eastern provinces have leading levels of compute technology, with Beijing, Jiangsu, Zhejiang, Shanghai, Shandong, and Fujian in the top ten, along with Hubei, Anhui, and Sichuan. **Compute innovation level**: Guangdong, Beijing, and Shanghai are the top three. They lead in the numbers of computing invention patent applications and invention patents granted, with a cumulative national share exceeding 70%. Guangdong in particular had 55,000 computing invention patent applications and 13,000 patents granted over the past five years, accounting for nearly half of the total. Zhejiang, Jiangsu, Shandong, Henan, Anhui, Hunan, and Hubei fill out the top ten. **Compute R&D investment**: The top ten in computer manufacturing industry R&D spending are Guangdong, Jiangsu, Zhejiang, Shanghai, Fujian, Hubei, Anhui, Sichuan, Shandong, and Beijing. In Guangdong in particular, annual R&D spending in the computer, communications, and other electronic equipment manufacturing industries exceeds RMB 100 billion, ahead of other provinces by a wide margin.

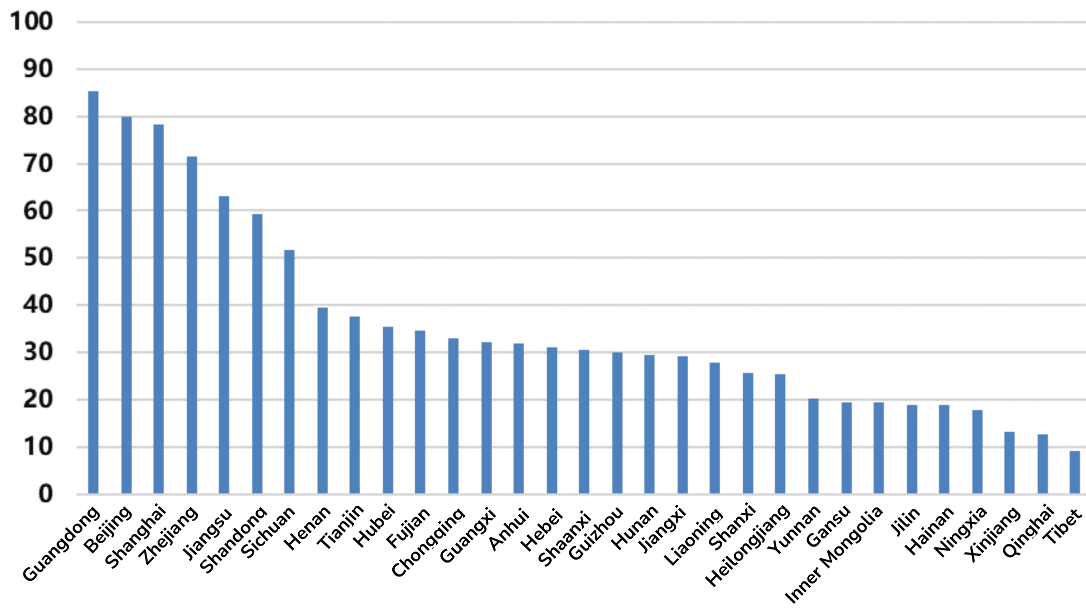


Source: CAICT

Figure 18 Computer technology sub-index levels of some Chinese provinces in 2021

5. Compute environment sub-index

The four major urban agglomerations—Beijing-Tianjin-Hebei, the Yangtze River Delta Economic Zone, the Guangdong-Hong Kong-Macao Greater Bay Area, and the Chengdu-Chongqing Economic Circle—have higher compute environment index levels. Overall, the environment for compute development in the provinces continues to improve, the computing network environment continues to be refined, and investment in compute continues to grow. The six highest ranked are Guangdong, Beijing, Shanghai, Zhejiang, Jiangsu, and Shandong, with compute environment index levels of over 60, while Sichuan, Henan, Tianjin, and Hubei fill out the top ten. **In compute network environment,** the top five are Shanghai, Jiangsu, Guangdong, Zhejiang, and Beijing, and Shandong, with relevant index levels of over 70. Jiangsu, Zhejiang, and Guangdong lead the country in inter-provincial internet outlet bandwidth, and Shanghai, Beijing, and Tianjin lead the nation in 5G coverage, with rates of over 35%. **In compute investment,** Beijing, Guangdong, Shanghai, Jiangsu, and Zhejiang are the top five, with index levels of over 60. Beijing and Guangdong in particular have IT hardware, software, and services expenditures of over RMB 260 billion, putting them ahead of other provinces and regions in compute investment strength. In terms of the degree of data openness, taking into account China Open Data Index data, the places with higher data openness index scores are mainly concentrated in the southeastern coastal region. Zhejiang, Shanghai, Shandong, Guizhou, and Guangdong make up the top five, with index values over 50. Zhejiang ranks first in readiness and data, Shanghai ranks first in platforms, and Shandong ranks first in utilization. Guizhou and Sichuan provinces in the west have excellent data openness.

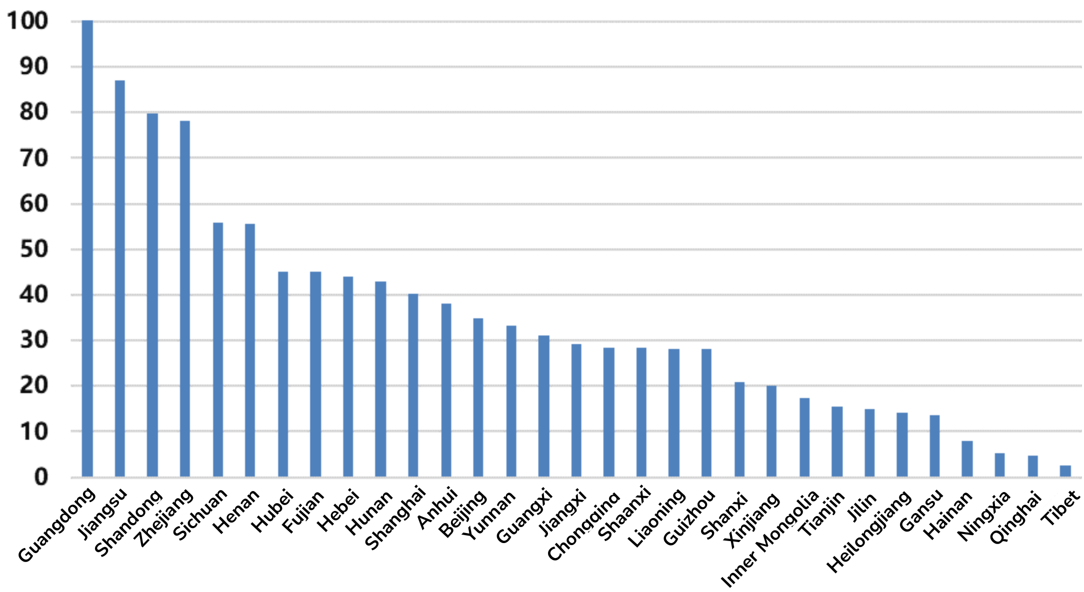


Source: CAICT

Figure 19　Compute environment sub-index levels of some Chinese provinces in 2021

6. Compute application sub-index

China's eastern coastal provinces generally have higher compute application index levels. Overall, the eastern provinces of Guangdong, Jiangsu, Shandong, Zhejiang, Fujian, and Hebei have relatively high compute application index levels. Guangdong, Jiangsu, Shandong, and Zhejiang are ranked as the top four. The central provinces of Henan, Hubei, and Hunan, and the western province of Sichuan, are also among the top ten, with compute application index levels over 40. **In terms of consumer application level,** Guangdong, Jiangsu, Zhejiang, Shandong, Sichuan, and Henan are the top six. They lead the country in mobile data traffic consumption, and their compute demand for mobile internet applications is relatively great. With average monthly mobile internet traffic in excess of 1.3 EB, they have consumer application level indexes of over 70. **In terms of industrial application level,** compute produces a strong impetus for the sustained and healthy development of industrial digitalization in provinces, and the pulling effect on the digital transformation of industries is especially obvious. Guangdong, Jiangsu, Shandong, Zhejiang, Fujian, and Shanghai are the top six by industrial application level index, with index values over 55. Guangdong in particular leads in industrial digitalization development, with the scale of industrial digitalization exceeding RMB 4 trillion. Industrial digitalization exceeds RMB 2 trillion in Jiangsu, Shandong, Zhejiang, Fujian, and Shanghai.

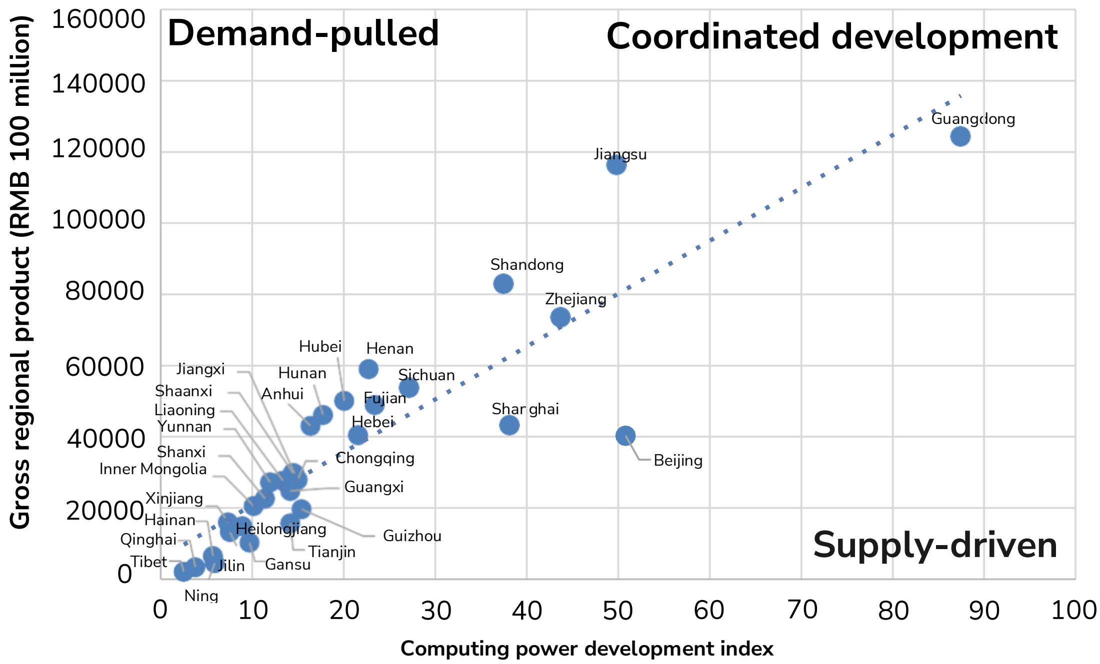


Source: CAICT

Figure 20 Compute application sub-index levels in some Chinese provinces in 2021

(iv) Relationship between the compute development index and the economy

A significant positive correlation is evident between the compute development index levels of provinces and the sizes of their economies. Compute has a strong driving effect on the economic development of provinces, and the provinces with larger digital economies and gross regional product in 2021 also had higher levels of compute development. For every 1-point rise in the computing power development index, the digital economy grows by about RMB 76.2 billion (about 0.167% of the national digital economy), and gross regional product grows by approximately RMB 148.5 billion (representing about 0.130% of national GDP). Increases in the compute development index are determined by increases in the development levels of compute scale, compute industry, compute technology, compute environment, and compute application, as coordinated development of supply and demand is achieved.



Source: CAICT

Figure 21 Relationship between the computing power development index and GDP

Provinces are mainly divided into three types according to compute development types: supply-driven, demand-pulled, and coordinated development. First is the supply-driven type represented by Beijing and Shanghai. Having made sustained improvements in the supply capacity of the compute industry for many years, and stepped up investment in large-scale compute infrastructure construction, these places have larger-scale compute and better compute environments. This not only serves the compute application of these provinces, but also provides compute support for consumer and industrial application in other provinces. **Second** is the demand-pulled type represented by Jiangsu, Shandong, Henan, Sichuan, Hubei, Fujian, and Hunan, where local compute demand is strong and the levels of consumer and industrial application of compute are high. For these provinces, compute generates new momentum and opens up new space for sustainable and healthy digital economy development and economic growth. **Third** is the coordinated development type represented by Hebei, Zhejiang, and Chongqing. The digital economy and compute are being developing in a coordinated fashion. Coordinated development layouts have gradually formed in the Beijing-Tianjin-Hebei, Yangtze River Delta, Guangdong-Hong Kong-Macao Greater Bay Area, and Chengdu-Chongqing Economic Circle regions.

IV. Accelerating the consolidation of compute's basic foundation, and stimulating the engine of digital economy development

At present, with implementation of national and local compute development plans under the “14th Five-Year Plan” accelerating, compute has become the new productive force (新的生产力) of the digital economy. The next step is to comprehensively implement the decisions and deployments of the Chinese Communist Party (CCP) Central Committee and the State Council. Firmly rooted in building a manufacturing powerhouse, cyber powerhouse,[20] and digital China, we should firmly grasp the wave of development based on industry digitalization and intelligentization, deeply grasp the features and basic principles of compute development, continue to grow the scale of compute, stimulate innovation-driven vitality, continue to optimize the development environment, strengthen application-empowering effects, deepen opening up to the outside and foreign cooperation, and strive to build a new pattern for China’s compute development, thereby providing strong support for the vigorous development of the digital economy.

(i) Solidify compute’s foundation and guide the layout of infrastructure

While focusing on strengthening support for digital transformation, intelligence-based upgrading, and integrated innovation, and adhering to the principle of taking the lead at a reasonable pace (适度超前), we should: (1) Accelerate the construction of computing infrastructure such as data centers, intelligent computing centers, and supercomputing centers, using construction to drive usage and usage to drive construction, while promoting the continuous improvement of compute infrastructure, and striving to build a new type of information infrastructure system based on new generation telecommunications networks, with data and compute facilities at its core and focused on breakthroughs in integrated infrastructure. (2) Accelerate the construction of the national integrated big data center system, strengthen the overall intelligent scheduling of compute, build several national hubs and nodes and big data center clusters, and build exascale and 10-EFLOPS-level supercomputing centers. And (3) continue to promote the green and low-carbon development of compute infrastructure, coordinate the construction of green and intelligent compute infrastructure, promote the green upgrading of traditional compute infrastructure in an orderly manner, and accelerate the creation of a multi-level compute infrastructure system characterized by data-network synergies, data-cloud synergies, cloud-edge synergies, and green intelligence.

---

[20] Translator's note: This translation renders the Chinese word 强国 qiángguó—which literally means "strong nation"—in English as "powerhouse," as in the phrases "manufacturing powerhouse" (制造强国) and "cyber powerhouse" (网络强国). For a more thorough discussion in English of the Chinese word qiángguó, see: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/lexicon-wangluo-qiangguo/.

(ii) Persist in being innovation-driven, and promote core technology research and development

While giving full play to the advantages provided by the new structure for leveraging national capabilities (新型举国体制) and the depth and breadth of China's super-large market, and staying closely aligned with the requirements of S&T self-reliance (自立自强), we should: (1) Create a collaborative innovation-based hardware and software ecosystem with compute at its core, and use diversified and systematic innovation to drive product chain upgrading. (2) Strengthen technological innovation related to advanced computing, promote the tackling of key technologies and R&D of important products in fields such as high-end chips, computing systems, and software tools, with the focus on filling in gaps and strengthening weak links. (3) Enhance basic research and multi-path exploration, and step up strategic layout work in cutting-edge fields such as in-memory computing, quantum computing, and brain-inspired computing, so as to build competitive advantages for future development. (4) Encourage computing enterprises to continue improving their independent innovation (自主创新) and intellectual property layout capabilities, thereby enhancing their core competitiveness. And (5) strengthen industry-academia-research institute-user (产学研用) collaboration mechanisms, optimize the allocation of innovation resources in the compute industry, strengthen the training of high-end talents in computing fields, and use high-quality enterprises and high-caliber industrial clusters and industrial innovation platforms as the vehicles for the all-round training and recruitment of innovation-oriented and interdisciplinary talents.

(iii) Build the industry system to improve the supply of computing products

To enhance the supply capacity of the compute industry, we should: (1) Accelerate the cultivation and growth of the advanced computing industry, promote technology integration and product innovation oriented towards diverse application scenarios, enhance the competitive advantages of computing products such as computing devices, computing chips, and computing software, and promote the development of the industry towards the middle and high ends of the global value chain. (2) Build a gradient-based system for incubating advanced computing enterprises, and guide small and medium-sized enterprises to become more "professional, meticulous, specialized, and innovative" (专精特新) while making leading enterprises in advanced computing bigger and stronger, so as to build a new pattern of development characterized by accommodative development between large, medium, and small enterprises, and collaborative development between the upstream and downstream of the production chain. And (3) optimize the layout of the advanced computing industry in each region, promote the development of industrial agglomeration and clustering, improve the quality and level of development of existing parks, and form a compute industry system with a rational regional layout and strong radiating and driving influence.

(iv) Build the development environment, and optimize network data services

We should: (1) Continuously promote the structural optimization and expansion of key links of the internet backbone network and metropolitan area networks, steadily promote 5G network construction, and continuously improve the quality of network services. (2) Enhance the support capacity of new types of compute networks, optimize the interconnection capacity of regional compute, and promote data interaction across networks, regions, and enterprises, thereby supporting the requirements of high-frequency, real-time interactive business. (3) Explore construction of an integrated compute network scheduling system, and strengthen cross-industry, cross-regional, and cross-level compute resource scheduling. (4) Channel social capital[21] to participate in the construction of compute infrastructure and the development of the compute technology industry, guide financial institutions to increase support for key areas and weak links of compute, encourage qualified financial institutions and enterprises to issue green bonds, and support qualified enterprises in going public for financing. (5) Deepen the development and utilization of public data resources, and accelerate promotion of regional data sharing and opening, government and enterprise data integration and application, and other general purpose data circulation facilities and platforms. (6) Accelerate full-process application of data, and build scenarios for the standardized development and utilization of data in various industries and fields, so as to enhance the value of data resources. And (7) strengthen security management throughout the life cycle, including data collection, aggregation, storage, circulation, and application.

(v) Enrich integration-based application, and strengthen compute's industry-empowering effect

We should: (1) Deeply uncover the integrated application of compute in new information consumption, smart cities, smart manufacturing, industrial internet, IoV, and other scenarios, and perfect supply-demand linking for compute. (2) Strengthen the application and promotion of compute, give full play to the enabling effect of compute on various industries such as manufacturing, finance, education, and

[21] Translator's note: The Chinese term 社会资本, translated literally as "social capital," refers to any source of funding outside of government budget outlays. This term encompasses investment by private individuals and private institutions. However, investment from state-funded entities such as state-owned enterprises (SOEs), including state-run banks, also falls under the umbrella of "social capital."

healthcare, create benchmarks for application in thousands of industries, and promote the formation of general purpose standard models in key areas. And (3) encourage the strengthening of innovation in advanced computing system solutions and industrial applications, and promote the expansion of heterogeneous computing, intelligent computing, cloud computing, and other technologies in vertical fields, so as to accelerate the digital transformation of traditional industries and promote the high-quality development of the real economy.

(vi) Strengthen international cooperation and promote the construction of the Belt and Road Initiative

We should: (1) Strengthen cooperation with "Belt and Road Initiative" (BRI)[22] countries in the areas of compute infrastructure, compute technology industry, digital transformation, etc., and create partnerships of mutual trust and benefit, inclusiveness, innovation, and win-win situations, so as to expand the broad development space for digital trade, and build a compute community of common destiny (算力命运共同体) for the BRI countries. (2) Further optimize the business environment, promote fair competition, strengthen intellectual property protection, encourage more foreign-funded enterprises to enter the Chinese market, attract high-level talents at home and abroad, and encourage domestic enterprises to actively expand in overseas markets. And (3) continuously deepen and expand international exchanges and cooperation, taking universities, scientific research institutes, and leading S&T enterprises as the mainstays, actively promote international exchanges and cooperation in compute fields through academic conferences, international forums, academic communities, project cooperation, and other ways, and promote the flow of technological innovation factors of production in the international arena, so as to create a favorable international environment for the development of compute in China.

Appendix 1: Framework for calculating the compute index

Based on the China Computing Power Development Index 2.0, the compute index includes five dimensions: compute scale, compute industry, compute technology, compute environment, and compute application. Compute scale is measured mainly based on the two aspects of computing device compute and infrastructure compute. The compute industry is measured mainly based on three aspects: computing devices, computing chips, and computing software. Compute technology is measured mainly based on compute innovation level and R&D investment. The compute environment is measured mainly based on the network environment, compute investment, and data

[22] Translator's note: The "Belt and Road Initiative" ("一带一路"), abbreviated BRI, refers to the Silk Road Economic Belt (丝绸之路经济带) and the 21st Century Maritime Silk Road (21世纪海上丝绸之路).

openness. Compute application is measured mainly based on the levels of consumer application and industrial application.

(i) Compute scale sub-index calculation method

The compute scale sub-index is calculated by weighting the values of secondary indicators—computing device compute and infrastructure compute. Standardized processing is performed separately on computing device compute and infrastructure compute to obtain the assigned values of each indicator.

1. Computing device compute. Reflects the basic compute provided by each region based on general purpose servers, intelligent compute provided based on AI servers, and supercomputing compute provided based on supercomputers. It is measured mainly by using tertiary indicators—scale of basic compute, scale of intelligent compute, and scale of supercomputing compute—and is calculated through standardized processing and weighting of values.

(1) Basic compute. Reflects the compute of each region provided by CPU chip-based servers, and is measured mainly using an indicator of the scale of server compute.

Scale of server compute = $\sum_{\text{last 6 years}}$ (annual scale of server shipments*average server compute in the current year).

(2) Intelligent compute. Reflects the compute of each region for AI training and inference provided by AI servers based on GPUs, FPGAs, application-specific integrated circuits (ASICs), and other accelerator chips, and is measured mainly using an indicator of the scale of AI server compute.

Scale of AI server compute = $\sum_{\text{last 6 years}}$ (annual scale of AI server shipments*average AI server compute in the current year).

(3) Supercomputing compute. Reflects the compute of each region provided by high-performance computing clusters based on supercomputers, etc., and is measured mainly using an indicator of the scale of supercomputing compute.

Scale of supercomputing compute = ∑ supercomputer compute

2. Infrastructure compute. Reflects the compute scale of data centers and intelligent computing centers in each region.

Scale of infrastructure compute = ∑ compute of data centers + ∑ compute of intelligent computing centers

(ii) Compute industry sub-index calculation method

The compute industry sub-index is calculated by weighting the values of secondary indicators—computing devices, computing chips, and computing software. Standardized processing is performed separately on computing devices, computing chips, and computing software to obtain the assigned values of each indicator.

1. Computing devices. Reflects the production and manufacturing capacity of each region in computing devices such as servers, AI servers, and supercomputers. It is measured mainly using the computing device output indicator.

Computing device output = sum of the production of computing devices such as servers, AI servers, and supercomputers

2. Computing chips. Reflects the production and manufacturing capacity of each region in microprocessors, memory, and other ICs. It is measured mainly using the IC output indicator.

IC output = sum of the production of microprocessors, memory, and other ICs

3. Computing software. Reflects the development level of the software and IT service industries in each region, and is measured mainly by the software business revenue indicator.

Software business revenue = sum of revenue in the software products, information system integration services, IT consulting services, data processing and operation services, embedded system software, and IC design businesses

(iii) Compute technology sub-index calculation method

The compute technology sub-index is calculated by weighting the values of secondary indicators—innovation level and R&D investment. Standardized processing is performed separately on innovation level and R&D investment to obtain the assigned values of each indicator.

1. Innovation level. The compute innovation level reflects the market value and competitiveness of each region in terms of computing R&D achievements. It is measured mainly by the numbers of computing invention patent applications and computing invention patents granted by region, and is calculated by standardized processing and weighting of values.

2. R&D investment. Reflects the scale and level of S&T investment in compute fields in each region, and is measured mainly using the R&D expenditures of the computer manufacturing industry in each region.

R&D investment = sum of R&D expenditures by large-scale industrial enterprises in the computer manufacturing industry in each region

(iv) Compute environment sub-index calculation method

The compute environment sub-index is calculated by weighting the values of secondary indicators—network environment, compute investment, and data openness. Standardized processing is performed separately on network environment and compute investment to obtain the assigned values of each indicator.

1. Network environment. Reflects the service capability of each region in data transmission to other regions in China for businesses such as data and internet, as well as the edge computing support capability of each region. It is measured mainly using tertiary indicators such as inter-provincial internet outlet bandwidth and 5G coverage rates, and is calculated by standardized processing and weighting of values.

Inter-provincial internet outlet bandwidth = the sum of the metropolitan area network outlet bandwidths of all operators

5G coverage ratio = number of 5G base stations divided by the number of 4G base stations

2. Compute investment. Reflects each region's investment in compute, and is measured mainly using an indicator of the scale of IT expenditures.

Scale of IT expenditures = sum of each region's investment in IT hardware, software, and services

3. Data openness. Reflects the data openness level of regional governments. It mainly uses the China Open Data Index to comprehensively evaluate the data openness level in each region from the dimensions of readiness, platform layer, data layer, and utilization layer.

(v) Compute application sub-index calculation method

The compute application sub-index is calculated by weighting the values of secondary indicators—consumer application level and industrial application level. Standardized processing is performed separately on consumer application level and industrial application level to obtain the assigned values of each indicator.

1. Consumer application level. Reflects each region's level of compute application in consumer fields such as mobile internet, and is measured mainly using an indicator of average monthly mobile internet traffic.

Average monthly mobile internet traffic = number of mobile phone users in each province*average handset data traffic per user per month

2. Industrial application level. Reflects each region's level of compute application

in traditional industrial fields such as agriculture, industry, and services, and is measured mainly using an indicator of the scale of industry digitalization.

Scale of industry digitalization = each region's increased output and efficiency gains (value added) from the integration and penetration of information and communications technology (ICT) products and services in other fields

Appendix 2: Data Sources

1. Basic data, including population data, economic value added, industrial value added, national input-output tables, computing device output, IC output, software business revenue, and R&D expenditures come from the relevant data of the National Bureau of Statistics and provincial statistics departments.

2. 5G base station numbers, mobile phone user numbers, average handset data traffic per user per month, and inter-provincial internet outlet bandwidth for China and each province come from MIIT statistical data.

3. Server and AI server shipment quantities globally and in China come from IDC and Gartner statistical data, and are used to calculate and assess basic compute and intelligent compute globally and in China.

4. Data on the scale of supercomputing compute globally and in China come from the internationally well-known TOP500 ranking, the TOP100 ranking of Chinese high-performance computers, and data provided by relevant firms.

5. Expenditures on compute hardware, software, and services in each of China's provinces come from the National Bureau of Statistics and relevant IDC statistical data, and are used to assess investment in compute in each of China's provinces.

6. Data for the data openness index of Chinese provinces comes from the "2021 China Open Data Index" and *China Local Government Data Openness Report* by the Digital and Mobile Governance Lab at Fudan University.

7. Data for calculating the number of invention patent applications and the number of invention patents granted in China and by province come from the Innojoy Patent Database.