INSIGHT

China's HPC Activities: Key Updates and Observations

Robert Sorensen Steve Conway
Earl C. Joseph, Ph.D.

IDC OPINION

A number of Chinese HPC activities have been in the news of late, and taken in total, these announcements demonstrate that China is entering a new phase of HPC development with the dramatically increased presence of Chinese HPC vendors on the TOP500 list, plans for two new Chinese 100 petaflops systems this year, and a new program for developing an exascale system by 2020. The Chinese are moving forward with a clear two-pronged approach:

- Large, and highly visible, Chinese government-funded programs will continue to be the main drivers of Chinese R&D in HPC technology, securing government-funded labs as the center of gravity for leading-edge Chinese HPC hardware and software development.
- IDC believes that the collection of commercial Chinese HPC vendors will likely continue to build on its success in the domestic market to expand its presence overseas, and we look for it to enhance its ability to address a wide class of users with increasingly capable commercial-grade software.

IN THIS INSIGHT

This IDC Insight discusses China's HPC activities. A number of Chinese HPC activities have been in the news of late, and taken in total, these announcements demonstrate that China is entering a new phase of HPC development with the dramatically increased presence of Chinese HPC vendors on the TOP500 list, plans for two new Chinese 100 petaflops systems this year, and a new program for developing an exascale system by 2020. Chinese HPC development strategies emphasize the use of Chinese indigenous HPC technology that, if realized, would represent a significant milestone in China's efforts to reduce, if not eliminate, reliance on foreign – and typically United States – technology, the culmination of over a decade's work by China. The next major hurdle facing the Chinese HPC development community is to advance its software capabilities to match its growing hardware prowess, and IDC expects to see increased emphasis in HPC-related algorithm and application development in the coming years.

SITUATION OVERVIEW

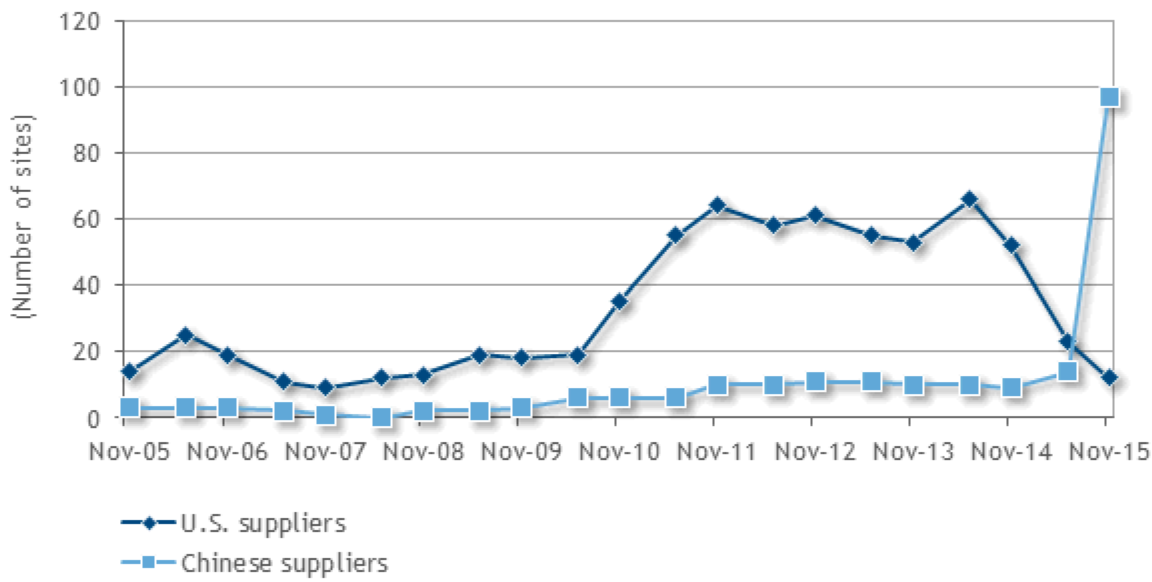
A number of Chinese HPC activities have been in the news of late, and taken in total, these announcements demonstrate that China is entering a new phase of HPC development with a dramatically increased presence of Chinese HPC vendors on the TOP500 list, plans for two new Chinese 100 petaflops systems this year, and a new program for developing an exascale system by 2020. IDC analysts in both the United States and China are monitoring these and other developments in the Chinese HPC sector and offer key updates and observations in the sections that follow.

The TOP500 List: China's Dramatic Advancement

The most recent TOP500 list from November of 2015 illustrated a watershed event in China's HPC composition, from both a user and a vendor perspective. As shown in Figure 1, the number of TOP500 HPCs in China grew almost three times from 37 to 109 in the six months between the June 2015 and the November 2015 lists.

FIGURE 1

TOP500 List of Chinese HPC Sites for U.S./Chinese Vendors, 2005-2015



Source: TOP500 list and IDC, 2016

Perhaps more impressive, however, was the increase in the share of Chinese HPC vendors that made the same list. In November 2015, Chinese HPC suppliers held 97 slots on the overall TOP500 list, representing almost 90% of all TOP500 HPCs in China, and almost 20% of all TOP500 systems worldwide, making China the second-largest HPC supplier nation after only the United States.

Prior to this, Chinese vendors have typically had only about 10 HPCs on the list – almost all in China – whereas U.S. vendors normally had 50 to 60 systems installed in the Chinese user base compared with only 12 on this most recent list.

For the sixth consecutive list – spanning the past three years – China's Tianhe-2, a supercomputer developed by China's National University of Defense Technology (NUDT), retained its position as the world's number 1 machine, but the bulk of the gains on the overall list were shared among the three main Chinese commercial HPC suppliers:

- Lenovo, which went from zero systems on the list six months ago to 24 on the new list, relied heavily on server base acquired from IBM for its TOP500 HPC entries this time around. Indeed, all were IBM-derived ThinkServer or NeXtScale servers. Of these 24, 21 went to either

ISPs, Internet retail companies, or telecommunications firms, and only three are installed at the more traditional HPC research sites.

- Sugon shipped 28 of its 49 HPC systems to ISPs or telecommunication firms, but it also benefitted from an aggressive Chinese government buying binge, placing 16 systems at various unspecified Chinese government facilities.
- For its part, Inspur grew at the highest rate, growing from 1 to 15 entries on the list; it installed 14 HPC systems at ISP sites in the past six months.

Finally, according to the list, Chinese firms placed only a handful of systems into the Chinese commercial sector and collectively had one system installed outside China, at a university in Denmark.

IDC believes that China's impressive presence on the most recent TOP500 list is admirable, but it does not yet make China a world leader in the development and use of HPC systems for established technical computing applications. That would require the Chinese vendors to gain market share in simulation-driven HPC vertical sectors such as aerospace, oil and gas, the financial sector, or even the growing high-performance data analytics market. These firms also need to boost their export potential, if for no other reason than to gain exposure to a strong base of international users, who can help drive next-generation HPC architectures and requirements.

On the other hand, the Chinese vendors are clearly expanding into high-potential proximity markets for HPCs, notably the Internet and telecommunications sectors, whose leading firms require large scale-out systems, and indeed, these Chinese HPC vendors have made a strong initial foray into a major new growth market for HPC vendors worldwide. This development can only help their efforts to grow their businesses.

China may also be concentrating on driving HPC capabilities to the Internet and telecommunications spaces as a way to help support Beijing's ambitious Internet of Things agenda that will require large amounts of computational capabilities to support the collection and analysis of vast amounts of real-time and standing data for applications, including food safety, healthcare, transportation, and pollution control.

Two 100 Petaflops HPC Systems to Be Announced Soon

According to the Chinese government officials and various news stories, China will officially announce the installation of two 100 petaflops systems in 2016: first will be an upgrade to the world's largest HPC, the Tianhe-2 from NUDT, and the second will be the ShenWei-BlueLight NG built by the Jiangnan Research Institute that has reportedly already been delivered at the end of 2015.

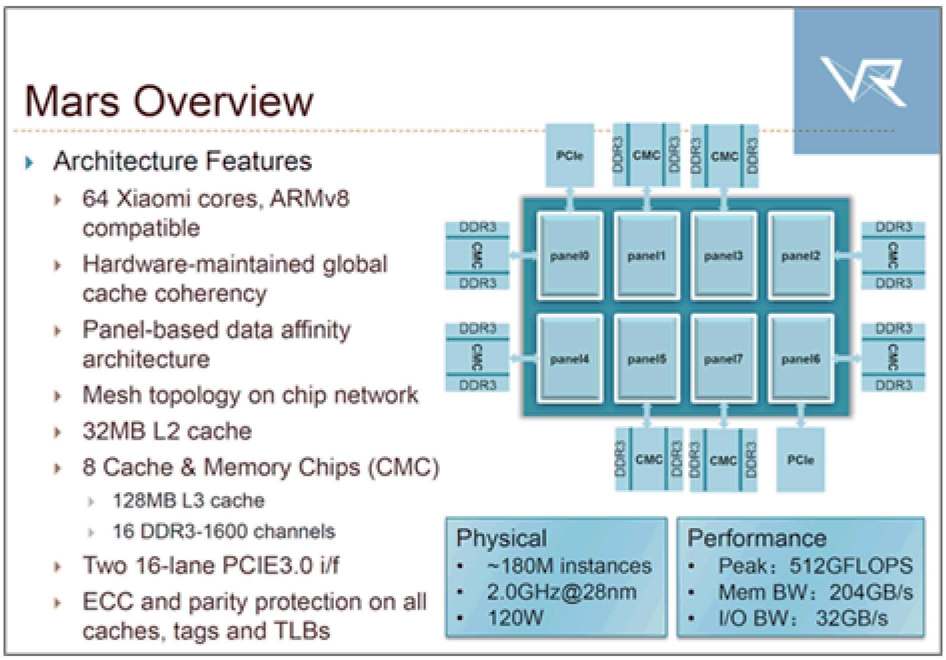
Should both of the machines live up to their announced performance claims, it would signify a significant leap forward in China's efforts to develop the full set of capabilities needed to build HPCs without relying on foreign — and typically United States — technology, and it would be the culmination of over a decadelong work by China to develop an indigenous processor capability. For example, the new Tianhe-2 follow-on, which is scheduled to debut at ISC 2016 in June, will feature a new chip supplied by Phytium Technology, a Chinese processor and ASIC maker founded in 2012 in Guangzhou. Phytium will reportedly supply its 64-core ARM-based Mars processors on PCI Express cards to replace the existing Xeon Phi cards currently used in the Tianhe-2 (see Figure 2). In addition:

- Each Mars chips has a peak performance of 512 gigaflops, and for the initial Tianhe-2 follow-on configuration, there will be 48,000 Mars chips devices installed for a total of over three million ARM cores. These chips combined with the existing complement of Xeon chips will yield an estimated performance on the LINPACK benchmark that should exceed 100 petaflops.

- Furthermore, plans call for additional upgrades to the Tianhe-2 follow-on that at its full deployment could consist of 32,000 Xeons, 32,000 ShenWei processors, and 96,000 Phytium accelerator cards, yielding a final system configuration capable of peak speeds in the range of 200-300 petaflops.
- In addition to the aforementioned processors, the Tianhe-2 follow-on will reportedly also incorporate a number of NUDT-developed FeiTeng processors, as has been done in previous Tianhe models. For those systems, the NUDT chips were used not as computational engines but were used in a front-end service system for job scheduling and access control, and that will likely be the case here as well.

FIGURE 2

Tianhe-2 Upgrade: Mars ARM Processor



Source: Phytium Technology, 2016

The original plan called for the Tianhe-2 follow-on to use upgraded United States-based Intel Xeon Phi processors, but NUDT needed to redesign the system last year after the U.S. government stopped shipment of those parts to NUDT, citing ties between NUDT and the Chinese military R&D infrastructure. In addition:

- Recent export control actions by the U.S. government that restricted the export of U.S. microprocessors to specific Chinese entities such as NUDT judged to have military connections may have provided the additional impetus needed for the Chinese government to dedicate a sizeable part of its future HPC development efforts based on domestically produced chips.

- China's processor developments are sponsored by the State High-Tech Development Plan, known as the 863 Program, which is funded and administered by the Chinese government to stimulate the development of advanced technologies for the purpose of strengthening China's homegrown industries and reducing or eliminating dependence on foreign interests.

The second Chinese machine, which may have already reached the 100 petaflops mark late in 2015, was developed by the Jiangnan Institute of Computer Technology in Wuxi, near Shanghai, and uses another Chinese designed and manufactured processor, the ShenWei. The chip is most likely a follow-on part to the ShenWei SW1600 that currently powers the Sunway BlueLight in operation at the National Supercomputing Center in Jinan and that ranked 86 in the TOP500 list published in July 2015. In addition:

- The SW1600 reportedly operates at 1.1 GHz, achieves 140 gigaflops performance from its 16-core RISC architecture, and is believed by many to be based on the Alpha RISC processor designed and built by the United States–based Digital Equipment Corporation.

A Chinese Exascale System in 2020?

Chinese officials recently announced that they were embarking on a development plan to field an exascale computer by 2020. To achieve this goal, China is simultaneously funding research into novel high-performance interconnects with 3D chip packaging, silicon photonics, and on-chip networks. The programming model for the heterogeneous computer will emphasize ease in writing programs and exploitation of performance of heterogeneous architectures.

With the bulk of the work likely being conducted by NUDT, the project will proceed in two main phases. Phase 1 — which is slated to run between 2016 and 2018 — will center on the development of prototype systems for verification of exascale computer technologies. During that time, Chinese researchers will explore possible exascale computer architectures, interconnects that can support more than 10,000 nodes, and energy efficiency technologies, as power demand is considered by Chinese planners to be one of the biggest obstacles for exascale systems. Further:

- The exascale prototype will be about 512 nodes, offering 5–10 teraflops per node, 10–20 gigaflops per watt, point-to-point bandwidth greater than 200Gbps, and support MPI latencies of less than 1.5µs. Development will also include system software and three typical HPC applications that will be used to verify the effectiveness of the design (see Figure 3).

FIGURE 3

Chinese Exascale Program: Phase 1

Key Technologies

- Multi-objective optimized architecture that balances performance, energy usage, programmability, reliability, and cost
- Energy-efficient computing nodes
  - 50-100 teraflops/node
  - 30+ gigaflops/W
- High-performance processors/accelerator design
  - 20 teraflops/chip
  - 40+ gigaflops/W
  - Supporting multiple models

Source: IDC, 2016

Phase 2, which is scheduled to culminate in 2020, will concentrate on the development of an energy-efficient computing node and a scheme for high-performance processor/accelerator design for use in the finished system. Ultimately, program planners expect their exascale system to operate within a 35MW power limit. The Chinese are taking a somewhat conservative approach to their system performance stating that the system may have a peak performance of 1 exaflops or more, but overall LINPACK efficiency is targeted for as low as 60%. Although there have been no announcements about specific hardware technology for the exascale project, it is almost certain that the Chinese will use indigenous technology where they can, including the processors and interconnect schemes. In addition:

- A major goal of the exascale program is technology transfer. Chinese planners said that China will work to field high-end domain-oriented servers based on exascale system technologies. These servers will take advantage of the advances at the node, the interconnect level, scalable I/O, storage, energy savings, reliability, and application software for inclusion in a wide range of server systems (see Figure 4).

FIGURE 4

Chinese Exascale Program: Phase 2

Exascale Computer System Development

- Exascale in peak
- LINPACK efficiency >60%
- 10 petabytes memory
- Exabyte storage
- 30 gigaflops/W efficiency
- Interconnect >500Gbps
- Large-scale system management and resource scheduling
- Ease-to-use parallel programming and software development
- System monitoring and fault tolerance
- Support for larger-scale applications

Source: IDC, 2016

Chinese HPC Software Skills Are the Next Challenge

Impressive performance claims aside, there is some concern within China that these systems, like their predecessors, may not be used to their fullest extent because of a combination of the overall lagging skills of Chinese HPC programmers.

Speaking about the Tianhe-2, one Chinese Academy of Science official noted that although the system is at the world's frontier in terms of calculation capacity, the functional capability of the supercomputer is still behind its counterpart systems in the United States and Japan and noted that the Tianhe-2 is like "a giant with a super body but without the software to support its thinking soul." In addition:

- Others within China indicate that the Tianhe-2 is running far from full capacity, and officials at the Guangzhou Supercomputing Center where the machine is housed are repurposing parts of the system and are looking for all types of workloads to fill the excess capacity.
- The chief designer of the Tianhe-2 states that the software development skills necessary to take full advantage of the Tianhe-2 processing potential are lacking and that China is still behind in software as high-efficiency software development depends on the overall scientific and technological level of the nation. Software talent is a major issue around the world.

Experts within China note that future efforts should focus more on application development to support fusion simulation, CFD for aircraft design, drug discovery, digital rendering, structural analysis for large machinery, and simulation of electromagnetic environment. Further:

- Additional areas ripe for advancement center on the ability to develop software that can effectively manage scale-out HPC systems with more than 300,000 cores while achieving more than 30% overall system efficiency.

One particular bright spot is that the HPC field has been gaining increased recognition among the younger generations in China, especially among students majoring in computer science and engineering. This may be because of a combination of China's high ranking on the TOP500 list in recent years, the result of the ongoing supercomputer race with other nations, and competition among the large supercomputing centers across provinces within China, mostly for bragging rights. Further:

- This year's International Supercomputer Student Challenge contest, sponsored by Inspur and organized by Asia Supercomputer Community (ASC), demonstrated the promising future of HPC in China. Among the 16 university teams in the final, 8 were from Chinese universities. Both the champion and the silver medal winners were claimed by Chinese universities.
- This contest has been held for six years and it's been growing in terms of the number of participating teams and the level of difficulty of the problems students are given to solve.

FUTURE OUTLOOK

Clearly, Chinese HPC makers have demonstrated that they can assemble world-class HPCs from a pure peak performance standpoint that are comparable to the best that the rest of the world has to offer. In addition, their decade-long effort to develop indigenous processors and related hardware technology seems to be finally paying off, and Chinese efforts will likely continue in that area. Ultimately, the next major hurdle facing the Chinese HPC community is to develop the needed software skills to match their growing hardware prowess.

To that end, the Chinese are moving forward with a clear two-pronged approach:

- Large, and highly visible, Chinese government-funded programs will continue to be the main drivers of Chinese R&D in HPC technology, securing government-funded labs as the center of gravity for leading-edge Chinese HPC hardware and software development.
- Meanwhile, the collection of commercial Chinese HPC vendors will likely continue to build on its success in the domestic market to expand its presence overseas while looking to increase its ability to address a wider class of users with increasingly capable commercial-grade software.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2016 IDC. Reproduction is forbidden unless authorized. All rights reserved.