



MARKET NOTE

Inspur Walks with Hyperscalers to Propel Artificial Intelligence Application Adoptions

Jessie Danqing Cai

Chwee Kan Chua

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: Inspur Walks with Hyperscalers to Propel Artificial Intelligence Application Adoptions

This IDC Market Note analyzes the Inspur Analyst Conference held in Jinan, China, from September 13 to 15, 2017.

Key Takeaways

- The fast-growing demand for cognitive/artificial intelligence (AI) is clearly felt at the platform layer of the AI ecosystem, in which hyperscalers, including Baidu, Alibaba, and Tencent (BAT) as well as Facebook, LinkedIn, Amazon, and Google (FLAG), are the key players.
- Cognitive/AI provides huge opportunity for Chinese tech companies, such as Inspur, to ramp up around hyperscalers and grow their global presence.
- Adoption of AI for many traditional companies remains a challenge because of the lack of technology know-how and uncertainty of short-term ROI. It is imperative to address the adoption barriers cohesively, so that the AI ecosystem can thrive as a whole.

Source: IDC, 2017

IN THIS MARKET NOTE

Inspur held its first analyst conference, Inspire Intelligence World, in Jinan, China, from September 13 to 15, 2017. At the event, Inspur detailed its product portfolio geared toward a future powered by artificial intelligence (AI). The product portfolio consists of four tiers: reconfigurable server system, cluster management software, performance optimization tools, and open source deep learning frameworks, such as Caffe-MPI. At the same time, Inspur will invest in end-to-end industry-specific AI solutions targeting a larger partner ecosystem.

The four-tiered architecture has been built on top of years of research on cognitive/AI, and it was verified through close collaboration with top Internet companies in China — Baidu, Alibaba, and Tencent (BAT). An almost 400% revenue growth with these hyperscalers was recorded in the past year, further strengthening the company's leadership position in China's server market. Inspur understands that the fast-paced development of AI applications is the primary driver of the demand surge, and that AI will promise bigger opportunities beyond boundaries. For example, the company's smart computing platform has taken off at the U.S. market, with the sales revenue expected to hit US\$617 million in 2017.

According to IDC's Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide, the worldwide hardware market for AI solutions is growing at a five-year compound annual growth rate (CAGR) of 61%. Smart computing solution providers, such as Inspur, will continue to see a strong market demand, given that they innovate together with the front-runners of AI application development, address the adoption inhibitors cohesively, and cultivate a partner ecosystem progressively. IDC sees the key implications of the Inspur Analyst Conference as follows:

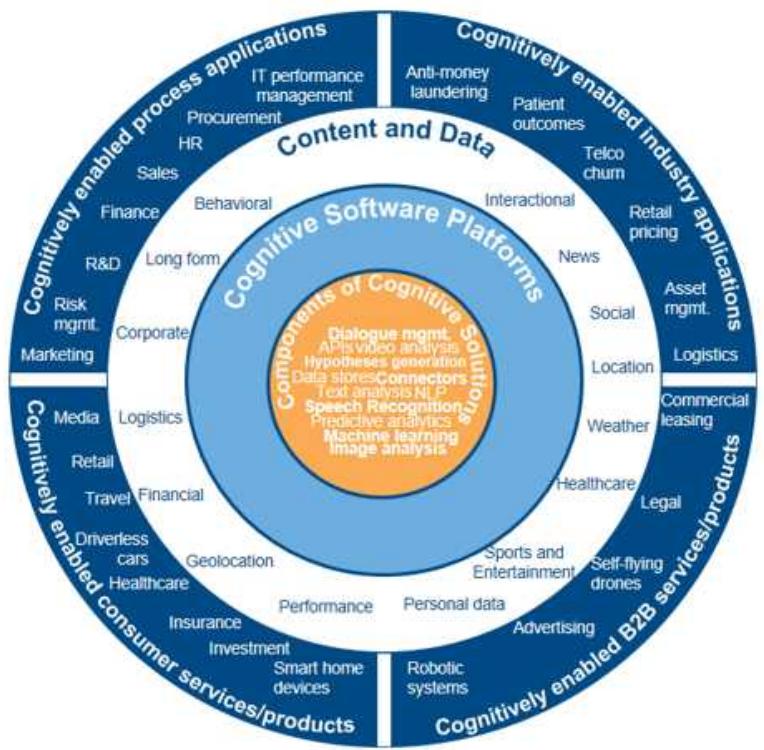
- The fast-growing demand for AI computing infrastructure is clearly felt at the platform layer of the cognitive/AI ecosystem, in which hyperscalers, including the BAT trio and Facebook, LinkedIn, Amazon, and Google (FLAG), are the key players.
- AI provides huge opportunity for Chinese tech companies, such as Inspur, to ramp up around hyperscalers and grow their global presence.
- For many traditional companies, the adoption of AI remains a challenge because of the lack of technology know-how and uncertainty of short-term ROI. It is imperative to address the adoption barriers cohesively, so that the cognitive/AI ecosystem can thrive as a whole.

IDC'S POINT OF VIEW

IDC defines cognitive/artificial intelligence systems as a family of technologies that uses deep natural language processing and understanding to answer questions and provide recommendations and direction. The system hypothesizes and formulates possible answers based on the available pattern, can be trained through the ingestion of vast amounts of data and content, and automatically adapts to and learns from successes and failures. There are four layers in the ecosystem wrapping around AI: the solution component layer, the platform layer, the data and content layer, and the application layer, as depicted in the Figure 2.

FIGURE 2

The Cognitive/AI Solution Ecosystem



Source: IDC, 2017

Developing and Accelerating the AI Computing Solution with Hyperscalers

Cloud platform providers are best positioned in the cognitive/AI ecosystem to scale, and the major players are called hyperscalers. Being digital natives, hyperscalers were among the first to exploit data to understand customer behavior, uplift experience, and achieve upsell by using machine learning and deep learning. Since then, other businesses have followed suit and embraced the digital revolution jointly enabled by big data availability, software innovations, and hardware acceleration options.

In a recent IDC cognitive/AI adoption survey conducted across three continents, 48% of the companies were found out to have adopted or planned to adopt AI within the immediate two-year time frame, showing that many companies have dropped the wait-and-see approach when faced with emerging technologies. This bodes well for hyperscalers and cloud infrastructure solution suppliers, such as Inspur. Inspur has stayed ahead of the curve through years of close collaboration with Tencent, Baidu, and Alibaba on AI-related initiatives, such as driverless cars, intelligence video surveillance, and smart city. The company has consolidated learning and developed a software-defined reconfigurable server and storage system specialized to support AI training and inference, in which performance can be accelerated, pooled, and reconfigured. Its AGX-2 product, for example, supports 8 GPUs, NVLink, a 2U form factor, and 100Gbps InfiniBand. Users can choose to utilize either 4 or 8 GPUs depending on the workload requirement.

The co-development model brought in large revenue for Inspur not only in the domestic market, as it was also able to zero start in the U.S. market and accelerate rapidly. For now, Inspur touts a turnkey AI computing solution covering heterogenous server system, software-defined storage, managed service for resource optimization, software to facilitate AI model training, and a customized multinode deep learning application framework. Inspur aims to become the business partner of choice for companies adopting cognitive/AI and transfer best practices gathered by working with hyperscalers. To a large extent, the product portfolios of Inspur well address one of the biggest obstacles organizations face in adopting AI: the lack of skilled resources, according to IDC's 2017 APEJ Cognitive/AI Adoption Survey.

Leading Server Sales in the Data-Rich Sectors of China

Inspur's smarting computing solution powers close to 90% of China's active cognitive/AI applications running on BAT platforms. These applications target primarily companies in data-rich sectors, such as government, telco, energy, manufacturing, and banking; for many of them, Inspur has been the IT infrastructure solution supplier since the age of mainframes. Use cases of fraud detection, risk assessment, public safety surveillance, rich media processing and rendering, and smart sensor network were shared at the conference, showcasing that Inspur's solution was able to achieve business results, including up to 10 times performance boost and 50% total cost reduction — these capabilities are in high demand as a previous IDC study found out that 40.5% of the organizations in Asia/Pacific will prioritize real-time decision support in the next 12–24 months.

Massive data load and mission-critical workflow are the two things in common among the customer success stories reported at the conference. IDC sees both traits are crucial to propel AI adoptions in traditional businesses. A clear value proposition to solve complex problems faster than ever possible resonates well in a world constantly searching for new revenue, competitive edge, and organic growth.

Increasingly, IDC foresees that companies will come to appreciate AI capabilities, as compared with traditional modeling means, to remove or reduce repetitive manual effort for commonplace business processes. For instance, an IDC cognitive/AI adoption survey shows that customer service and support are the business processes expecting the most immediate business value in the coming two years, and companies at the higher level of information DX maturity are more likely to lead the change.

An AI-First Innovation and Development Strategy That Has Paid Off

Inspur has put in place a consistent innovation strategy that addresses the computational challenges of AI at both hardware and software layers. Apart from the development of customized products for hyperscalers, Inspur has worked on the following to address the differences of requirement between hyperscalers and traditional businesses, so as to help traditional businesses maximize their ROI on AI computing infrastructure:

- GPU/FPGA virtualization and pooling to allow high-performance computing (HPC) resources to be better utilized by smaller companies, and effectively reduce the amount of investment needed for such companies to start exploring AI.
- Made modification to the Caffe deep learning framework, so that it works in a multinode environment. The result product, called Caffe-MPI, delivered nearly two times higher performance than TensorFlow when tested on a standard ImageNet data set.
- Released a cluster management software called AIStation to facilitate data preparation, analysis, and model training which supports popular ML/DL frameworks, such as Caffe, Microsoft CNTK, Google TensorFlow, and more.

- Committed to work on end-to-end AI solutions catered to industries such as healthcare, security, and financial services.

To keep abreast of latest innovations, Inspur watches closely and makes active contributions to the open source communities, including OpenStack and Open19. In addition, Inspur co-creates with leading IC manufacturers, such as Intel, NVIDIA, IBM, and so forth, through various initiatives to further elevate the computing capability at the core and reduce latency in the network — both are fundamental for the continued advancement of cognitive/AI technologies. The open and diversified product innovation and development strategy has helped Inspur deepen its industry penetration, expand into international markets, and stay ahead of changes.

Bigger Opportunity Awaits in the Mass Market

While Inspur's success with hyperscalers is hugely exciting, it has longer-term goals to conquer the mass market and become the business partner of choice for many companies embarking on AI applications. It is a task not short of challenges. For one, AI application adoption is still in the early days. There are various adoption barriers: lack of technical know-how, lack of clarity of vendor capabilities, uncertainty of short-term ROI, shortage of amenable data for model building, and increased pricing of GPUs, just to name a few.

It has become imperative to develop and leverage heavy-lifting options, such as AI as a service, to lower down such barriers, so that even SMBs can make agile use of existing data assets and explore the capabilities of AI without making a heavy up-front investment. It is only then can an AI ecosystem thrive as a whole, with consistent demand that drives the healthy consumption of HPC resources.

Nevertheless, IDC sees that the time has come for AI to not only be proven as possible but also unavoidable. Companies, such as Inspur, are presented with massive opportunities, and they need to stay innovative and open and address the adoption inhibitors cohesively to ensure that the ecosystem flourishes from the application layer.

LEARN MORE

Related research

- *Cognitive/Artificial Intelligence Asia/Pacific (Excluding Japan) 2017 Survey* (IDC #AP42736917, August 2017)
- *Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending by Industry Market 2016-2020 Forecast* (IDC #US42749817, June 2017)
- *IDC FutureScape: Worldwide Big Data, Analytics, and Cognitive/AI 2017 Predictions – APEJ Implications* (IDC #AP42148017, January 2017)
- *IDC's Worldwide Semiannual Cognitive/Artificial Intelligence Systems Spending Guide Taxonomy, 2016* (IDC #US41895216, November 2016)

Synopsis

This IDC Market Note analyzes the Inspur Analyst Conference held in Jinan, China, from September 13 to 15, 2017. During the event, Inspur detailed its product portfolio geared toward a future powered by artificial intelligence (AI), and shared its successful product innovation strategy by working closely with the hyperscalers of China and the United States. Future opportunities would require the cognitive/AI ecosystem to flourish from the application layer.

About IDC

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

IDC Asia/Pacific Headquarters (Singapore)

80 Anson Road, #38-00

Singapore 079907

65.6226.0330

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 2017 IDC. Reproduction is forbidden unless authorized. All rights reserved.

