IDC PERSPECTIVE

NISI Group: NetApp's ONTAP Reference Architecture Is Critical for Deep Learning Success

Glen Duncan Mikhail Jaura

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: NISI Group: NetApp's ONTAP Reference Architecture Is Critical for Deep Learning Success

The NISI Group (NISI) is a medical technology (medtech) business headquartered in Hong Kong, specializing in the development, manufacturing, and sales of surgical and diagnostic innovations. NISI engages NetApp through a local partner to provide a storage solution to support the company's deep learning software as part of its research and development (R&D) processes. NISI utilizes NetApp's ONTAP Artificial Intelligence (AI) reference architecture for deep learning and AI workloads.

Key Takeaways

- IT infrastructure providers, particularly those that offer solutions to the start-up and small and medium-sized enterprise (SME) market, should ensure that documentation and whitelisting are coherent and easily accessible to customers.
- A key learning from NISI is that even smaller organizations may require datacenter colocation services for power- and cooling-hungry applications.
- Migrating some services to a cloud platform will benefit deep learning developers by increasing their solution training speeds, through a cluster of GPUs and CPUs performing sophisticated matrix operations.

Recommended Actions

- When deploying deep learning solutions, organizations should ensure that the hardware solutions provided by different vendors are tightly integrated and a thorough AI reference infrastructure is used.
- Deep learning systems place considerable strain on on-premises datacenters' power and cooling infrastructure. Hence, for the long term, organizations should consider partnering with a datacenter colocation services vendor.
- Both on-premises and cloud datacenter architectures provide specific advantages in security, scalability, and performance. Hence, organizations should aim to achieve the right balance with a multicloud datacenter architecture.

Source: IDC, 2020

May 2020, IDC #AP44534719

SITUATION OVERVIEW

Overview

The NISI Group (NISI) is a medical technology (medtech) business headquartered in Hong Kong, specializing in the development, manufacture, and sales of surgical and diagnostic innovations. Its corporate mission is to become a world leader in non-invasive surgical innovations, pushing the limits of cutting-edge technologies to revolutionize the current practice of surgical procedures and drive accessibility. Currently, NISI is developing innovative technologies for the medical device market to benefit patients, practitioners, and relevant stakeholders. It was established in early 2012 under its parent company, NISI (HK) Ltd. NISI works with international experts to develop its products and these partners include accredited university institutions and technology vendors that have expertise in precision machinery, networked control systems, surgical robotic devices, and minimally invasive and endoscopic surgical procedures. The company's footprint spans over 50,000 square feet across two locations: its Hong Kong Cyberport headquarters, which handles corporate, research and development (R&D), clinical affairs, and intellectual property (IP) management; and its China office in Dongguan's Songshan Lake, which serves as a manufacturing facility and handles the firm's engineered products and its R&D operations.

As a start-up, NISI is still in the early stages of development. Once it is ready for the market, it intends to deploy a deep learning solution through a software-as-a-medical-device (SaMD) model, as well as several types of medical devices. In early 2018, it initiated an R&D project that focuses on developing a deep learning solution, which is part of an AI software solution that will support some of its medical devices, but this is still under development. In addition, NISI also uses medical imaging collected from real clinical settings before being annotated and labeled by expert surgeons or clinicians. The data collected is then used as training references, as well as bases through which deep learning algorithms are generated. Software based on the algorithms is then installed on medical devices to perform diagnosis or support clinical practitioners in their decision making. NISI's business model and core competencies are based upon the deep learning solution becoming increasingly accurate over time, as it is exposed to more cases. Data is fed offline into the database and this will then be used for the creation of new generations of products. In the future, data from medical sites over the cloud may feed directly into the database in real time.

NISI's R&D programs have been supported by Hong Kong's Innovative and Technology Commission (ITC). So far, it has secured a total of HK$58 million (US$7.45 million) of government funding. In 2019, it strengthened its senior management and advisory board by hiring industry experts and experienced professionals to facilitate the production and commercialization of products. NISI has also received an ISO 13485 certification for designing and manufacturing its medical devices and completed its human feasibility study in Queen Mary Hospital with its clinical partner, the University of Hong Kong, for the testing of a new robotic endoscope device.

In 2020, NISI will move onto the next stage and prepare for its multicenter clinical trials for its robotic endoscopic device in collaboration with hospitals in China. This regulatory pathway is an important step for NISI toward product commercialization.

Top Investment Priorities

Through a local partner, NISI engaged NetApp to provide a storage solution supporting its deep learning software as part of its R&D processes. The start-up utilizes NetApp's ONTAP artificial intelligence (AI) reference architecture for deep learning and AI workloads. A reference architecture is an infrastructure that can deliver sustained high input/output (I/O) throughput at low latencies to exploit the graphics processing units' (GPUs') I/O parallelism and nondisruptively scale compute and storage systems to achieve faster training times. NISI's design is implemented using NetApp AFF A300, an all-flash storage targeted toward midrange customers, underpinned by

predictive AI-based analytics to offer speed, performance, and IT agility for business applications. In addition, the ONTAP AI reference architecture and infrastructure are utilized in conjunction with a NVIDIA DGX-1, a system that deploys high-performance computing nodes, allowing NISI to run its deep learning simulations. Moreover, for NISI to connect the NetApp solution with the NVIDIA DGX-1, NetApp sourced and resold Mellanox SN2100 100Gb Ethernet Switch Systems. Currently, NISI's datacenter infrastructure is on-premises in its headquarters and its redundancy protocol includes a Veeam Backup and Replication software system with tape backup. This specialized IT infrastructure setup reflects the start-up's early stages of development in the support of its R&D processes. In the coming years, NISI will enter the market. Should it experience revenue growth, it will correspondingly expand its R&D IT infrastructure assets to support the operations of a growing and larger organization that utilizes a hybrid cloud architecture.

Best Practices

NISI uses several vendors for IT infrastructure and services. In late 2018, it first used an NVIDIA DGX-1 as a high-performance computing node to help with its survey on research courses. The initial training data set was not too large; hence, the simulation was done entirely on the NVIDIA DGX-1. However, considering that NISI was collecting more labeled medical images for training its new network and sourcing data from several hospitals and clinical centers, this meant that its data volumes were growing exponentially. Aside from the images, NISI also needed to collect videos showing the times when the images were taken. These metrics enhanced not only its overall system and performance but also its data storage requirements. NISI required larger storage and more data throughput to facilitate its deep learning process. Its previous storage network throughput speeds clocked around one gigabit per second, which was simply not enough for its deep learning solution.

Considering NISI's limited resources as an early-stage start-up, NetApp was able to provide the company with several cost-effective and high-tier solutions. In conjunction with the NVIDIA DGX-1 already deployed in NISI, NetApp provided ONTAP AI, a reference architecture system to carry out deep learning and AI workloads, implemented using NetApp AFF A300, an all-flash storage that uses predicted AI to improve performance, speed, and IT agility for end users. Meanwhile, facilitating the connection between the NetApp solutions and NVIDIA DGX-1 computing node was carried through Mellanox SN2100 100Gb Ethernet switches, which were sourced and resold by NetApp. All these components were optimized by the reference architecture.

Consequently, NISI was able to migrate data from DGX-1 to the new NetApp storage quickly. These technologies have created seamless and transparent operations for the technology team and AI researchers. Furthermore, NISI now has spare computing performance and storage bandwidth capabilities. With this, the start-up is currently migrating some of its business networks to the NetApp solutions, such as its enterprise resource planning (ERP) system.

ADVICE FOR THE TECHNOLOGY BUYER

For the IT Infrastructure Vendor

IT infrastructure providers, particularly those that offer solutions to the start-up and small and medium-sized enterprise (SME) market, should ensure that documentation and white paper are coherent and easily accessible by customers. Other than the previously mentioned criteria, NISI's decision to choose NetApp's solutions came down to customer experience, as it has experienced some vendors lacking transparency or providing too little detail about their solutions. Hence, it was concerned about implementing other services. Besides, several reseller partners are part of NetApp's ecosystem. NetApp led the outside communication and overall project management, whereas resellers and distributors help along the way. This creates a single point of contact for

NISI, in which solutions are installed and implemented coherently, well organized, and arranged for seamless operations. Meanwhile regarding the technical capabilities split between NetApp and its resellers, NetApp's engineers provided all the designs, whereas the heavy lifting, including the physical installation and routing of the devices, was carried out by its partners. In addition, the reference architecture installation guide for NetApp's ONTAP solution was detailed so thoroughly it included very basic commands that customers must type in, thus making installation and deployment a smooth process for NISI. The start-up also encountered difficulties in setting up the network interface on the NVIDIA DGX-1. However, the problem was solved after referring to one of NetApp's verified architecture documents. Ultimately, IT infrastructure vendors should ensure that information about their services is as transparent as possible to customers to minimize the point of contact for customers when dealing with several vendors in the same ecosystem and guarantee that installation guides are easy to follow and well detailed.

For Start-Ups and Other SMEs

A key learning from NISI is that even smaller organizations may require datacenter colocation services for power- and cooling-hungry applications. Although this was not initially a priority for the start-up due to limited resources, it will be something that NISI will be forced to consider as it grows. Organizations should also think about other metrics besides hardware performance, such as power availability and storage. In NISI's case, supporting the deep learning process is almost a 24/7 commitment, with new data fed into the system and algorithm development research, requiring power-intensive resources, such as the NetApp and NVIDIA solutions. Utilizing a small on-premises datacenter places significant demand on limited power and cooling resources. Other start-ups and SMEs with similar applications may also want to consider datacenter colocation services. By leasing rack space, tenants have access to sophisticated security protocols and power and cooling services. When datacenters are on-premises, organizations must consider handling their security affairs and ensure that they have enough power and cooling to make sure that their datacenter functions well and does not overheat. Moreover, dealing with confidential patient information is also a concern for NISI, driving its future need to outsource its datacenter setup to colocation services to take advantage of advanced physical security features and network and cloud connectivity. Large enterprises including technology giants also use trusted colocation vendors to handle core and noncore business operations. Generally, these facilities are more secure compared with on-premises sites, as they have many layers of security to protect tenants from security breaches.

For Cloud SPs

SMEs and start-ups, such as NISI, should assess their level of redundancy, especially when dealing with sensitive information. The world is full of uncertainties and exposed to various natural and man-made disasters. Events such as the COVID-19 virus outbreak can disrupt business operations. Hence, businesses must have a disaster recovery (DR) protocol to ensure that in the event of data loss, there will be business continuity, especially for critical business operations. NISI currently uses a Veeam Backup and Replication solution, but this may be insufficient to protect all its data in case of a disaster on its premises or in Hong Kong. As NISI scales up in size and receives more capital investment, it may consider using a colocation datacenter with two sets of redundant servers. This setup ensures that even if one datacenter goes offline, another site will support the business continuity of critical operations. NISI is also considering using cloud SPs, as migrating some services to cloud will benefit deep learning developers by increasing its solution training speeds, through a cluster of GPUs and CPUs performing sophisticated matrix operations. It will also support scalability by taking advantage of various on-demand resources available through the cloud, as well as flexibility through frameworks such as Apache MXNet, TensorFlow, and Microsoft Cognitive Toolkit. These allow customers to package libraries of deep learning algorithms for mobile, internet, and connected services.

IDC'S POINT OF VIEW

IDC considers the tight integration of hardware solutions provided by NetApp, NVIDIA, and Mellanox through the ONTAP AI reference infrastructure as a key reason for the implementation success of NISI's deep learning solution. This reference infrastructure was pivotal for seamless efficient setup, configuration, and troubleshooting throughout the deployment process. It was also a key differentiator for NISI, compared with other competitor solutions, and fundamental to its procurement decision-making processes and general satisfaction with the solution.

IDC agrees that NISI will be continuously challenged with making financial resource allocation decisions, considering its financial constraints as a start-up company. There comes a point when there is only so much a small organization can purchase, so the decision must be executed at the right time. As a start-up in its early stages, NISI's priority is to focus on developing its solutions and commercializing its products. As part of the medical industry, NISI must do further clinical trials and seek regulatory approval from governments before its products hit the shelf. After considering the pathway, the timeline for selling products may take years ahead. However, through the development of an innovative and revolutionary solution, NISI may get recognized by the medical community. This will grow its data feed, allowing its deep learning solution to become smarter and handle more cases with various scenarios.

Currently, NISI's IT infrastructure contains excess storage and processing capacity, partly because of the efficiency of the NetApp-led ecosystem solution. However, this will likely change as the business grows its operations. NISI should consider using a cloud SP such as Amazon Web Services (AWS) to migrate some of its data and noncritical operations. Through this platform, more speed and flexibility will be added in its business operations. Moreover, using a colocation services vendor will remove its power capacity constraints, improve its data security, and solve concerns around limited cooling. With these initiatives, NISI can preempt future operations disruption. Lastly, a DR protocol supported by additional redundancy sites is crucial for businesses dealing with sensitive data. These solutions should be considered for businesses such as NISI as they experience growth in their organization.

LEARN MORE

Related Research

- *IDC FutureScape: Worldwide Datacenter 2020 Predictions — APEJ Implications* (IDC #AP44535019, January 2020)

Methodology

This IDC Perspective utilizes a case study qualitative research design, including a one-on-one interview with an introduction from NetApp Asia/Pacific (AP), as its research method. A one-hour group call was held with YH Cheng, head of Research and Technology of NISI (HK) Limited. Cheng's primary role in the is in information technology, as the firm has a small-scale IT team that includes an IT manager and a senior systems engineer in charge of running daily operations. Cheng's role involves controlling an on-premises datacenter in NISI's Hong Kong office. This caters to business operations such as the ERP system, data storage, and data sharing. The technology team is also involved in selection of computing resources for the company's R&D in deep learning, medical imaging, recognition, and classification. NetApp AP's deployment team was also involved in the call, providing some detail to help understand some of the technical details associated with the solution. After the interview, the interview recording was transcribed and a

thematic content analysis was then carried out to identify key insights and themes and build the narrative for this report.

Synopsis

The NISI Group (NISI) is a start-up that specializes in the development, manufacturing, and sales of surgical and diagnostic innovations. This IDC Perspective discusses the key learnings from a deep learning infrastructure deployment case concerning the company's software and hardware solutions that support the R&D of its products and algorithms. These insights were gleaned from a one-hour group call interview conducted with YH Cheng, Head of Research and Technology.

According to Dr. Glen Duncan, associate research director, Datacenters, IDC Asia/Pacific (excluding Japan) (APEJ), "NISI is one of the first Asia/Pacific (AP) cases to come through demonstrating the importance of close collaboration and integration between hardware partners. It also shows the critical need for a robust AI reference infrastructure to ensure deployment success when implementing deep learning solutions."

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