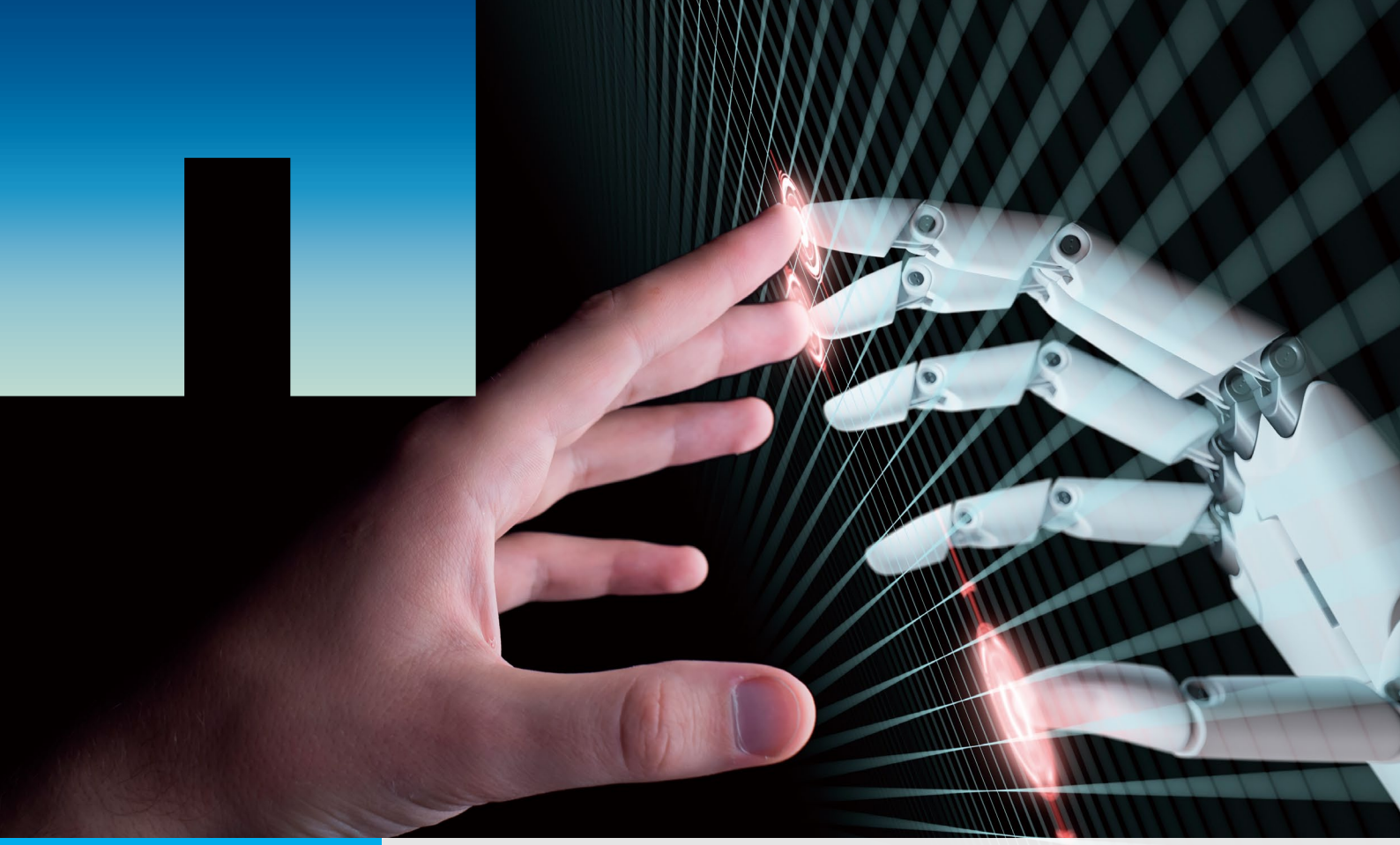


Customer story

IT service company

ITOCHU Techno-Solutions Corporation

Technical Solution Center

ITOCHU Techno-Solutions Corporation

ITOCHU Techno-Solutions has built a large-scale AI verification environment using the NetApp ONTAP AI integrated AI system. The company is developing an image library using Docker containers and wants to realize an advanced AI verification environment that can quickly respond to various verification requests.

The AI integrated system “NetApp ONTAP AI” adopted to “AI_LAB,” the largest-scale AI verification environment in Japan, which accelerates the utilization of AI/deep learning in business

ITOCHU Techno-Solutions (CTC) has recently set up AI_LAB, Japan’s largest verification environment dedicated to AI within a “Technical Solution Center (TSC),” a multi-vendor comprehensive verification facility. NetApp's AI integrated system “NetApp ONTAP AI” has been adopted as the core system. This system integrates an NVIDIA® DGX-1™ cluster (4 nodes / 32 GPUs) with petaflops-class computing performance and the “NetApp AFF A800” high-end all-flash array. , Having conducted various AI/deep learning verification tests that include cloud integration, it is now making a significant contribution to the utilization of AI for business by client companies.

DGX-1/32 GPU usage rate

95% or Higher

Integrate to public clouds

Easily

Contact us

NetApp®

"NVIDIA recommends NAS as storage for the DGX-1. We have every confidence in the NetApp AFF series as the high-performance storage that can handle the NFS protocol and with no hesitation chose NetApp ONTAP AI as the initial environment for AI_LAB."

ITOCHU Techno-Solutions Corporation
TSC Verification Support Section, TSC Department, Data Center Service Head Office
Tomofumi Arakawa

Challenges

Newly establishing a verification facility "AI_LAB" specializing in AI development

The "Technical Solution Center (TSC)" operated by ITOCHU Techno-Solutions (CTC) is a comprehensive facility that allows client companies to evaluate and verify systems to be introduced using actual equipment. It boasts Japan's largest verification facilities for a multi-vendor environment and offers full technical support from experienced and dedicated engineers. Tomofumi Arakawa of the TSC Verification Support Section, TSC Department, Data Center Service Head Office, explains as follows:

"TSC conducts approximately 1,200 performance evaluations and functional tests annually using the latest system products. In the last few years, opportunities to verify data and service integrations between the TSC system and the public cloud are growing exponentially, as companies take advantage of connecting to Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). There has also been a remarkable increase in the number of AI/deep learning verification projects."

The growing need for verification of hybrid cloud and AI/deep learning has driven CTC to create this new verification environment.

"AI_LAB was launched in late 2018 to verify large-scale AI utilization and to build learning models. It provides an integrated environment comprising a GPU cluster that can use up to 32 GPUs and high-end all-flash storage and can link with the major public cloud services. The response has been great, and we're always fully booked at least 3 months in advance."(Arakawa)

The integrated system NetApp ONTAP AI, jointly developed by NetApp and NVIDIA, was adopted for the AI-dedicated verification environment AI_LAB.

Solutions

Adoption of NetApp ONTAP AI, jointly developed by NetApp and NVIDIA

The NetApp ONTAP AI integrated system integrates the NVIDIA DGX-1 AI, incorporating eight Tesla V100s and the NetApp AFF A800 high-end all-flash array and forming a reference architecture which is optimized for AI/deep learning applications.

In the configuration adopted for CTC's AI_LAB, the NVIDIA DGX-1 cluster with 4 nodes/32 GPUs and one NetApp AFF A800 system (HA pair) provided with a high-speed connection by two 100 GbE switches (each with four 100 GbE links). Tomokazu Tanimoto of the TSC Verification Support Section, TSC

ITOCHU Techno-Solutions Corporation
Data Center Service Head Office
TSC Verification Support Section, TSC Department
Tomofumi Arakawa

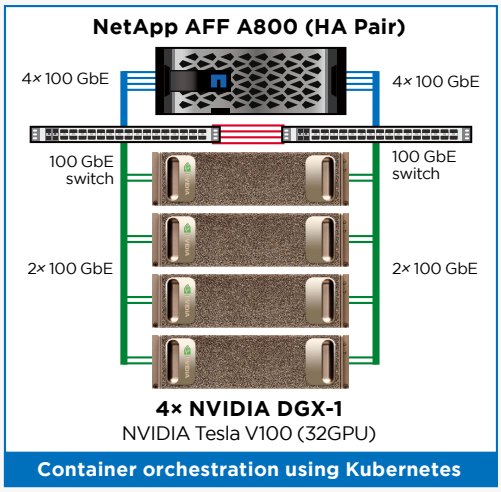
Department, Data Center Service Head Office, adds the following:

"The NetApp AFF A800 equipped with NVMe SSD can supply training data to the GPUs at an extremely high rate. This makes it possible to take full advantage of the petaflops computational performance of the DGX-1 cluster (4 nodes / 32 GPUs). We now have the best possible environment to meet the advanced verification requirements of AI/deep learning."

In the performance verification of NetApp ONTAP AI conducted by Arakawa, Tanimoto and their colleagues they achieved a throughput of up to 25 GB/s and latency of 500 microseconds or less while constantly maintaining utilization rate of at least 95% on all 32 GPUs of the DGX-1 cluster.

"The majority of AI/deep learning training data is made up of unstructured data such as text, images, and videos. Because of

NetApp®



Adopted by CTC for AI_LAB

NetApp ONTAP AI

AI integrated system jointly developed by NetApp and NVIDIA

Get superior performance from the NVIDIA DGX-1 with NetApp AFF A800

- Constantly maintains 95% or higher usage rate over 32GPU
- Max Throughput: 25GB/s
- Latency 500 microseconds or less
- Compatible with AWS, Azure, GCP, etc.

this, NAS is used as storage for the DGX-1. We have every confidence in the NetApp AFF series as the high-performance storage that can handle the NFS protocol and did not hesitate to choose NetApp ONTAP AI as the initial environment for AI_LAB" (Arakawa).

"CTC and NetApp have been allies for 25 years. This means that many of our engineers are already familiar with NetApp products and we have a great deal of knowledge within the company. Knowing that we would be able to handle their system with confidence, was a great advantage." (Tanimoto)

Data Fabric technology for cloud integration

AI_LAB, which runs NetApp ONTAP AI (NVIDIA DGX-1 / NetApp AFF A800), has another major advantage. This is integration with public cloud services such as AWS, Azure, and GCP.

As Arakawa explains, "Customers can easily build an AI/deep learning data pipeline by connecting the on-premises verification environment in AI_LAB to the public cloud. This can be used not just for things like verification of cooperative processing using GPUs on the cloud side and hierarchical management of data that incorporates cloud storage, but also for verification of the return on investment of on-premises and cloud AI environments."

NetApp AFF A800's ONTAP storage OS consistently provides robust and simple data management functions from on-premises to the cloud. Furthermore, seamless use of data in the hybrid cloud is enabled by NetApp Data Fabric technologies such as Cloud Volumes ONTAP, SnapMirror, and Cloud Sync.

ITOCHU Techno-Solutions Corporation
Data Center Service Head Office
TSC Verification Support Section, TSC Department
Tomokazu Tanimoto

Tanimoto explains, "For example, Cloud Volumes ONTAP makes it possible for SnapMirror to synchronize data between AI_LAB and AWS volumes. In addition, using Cloud Sync, we can create a verification process that uses Redshift to analyze backup data stored in Amazon S3."

Container orchestration using Kubernetes

"AI_LAB provides an environment with the capability to meet our customers' every need, from large-scale verification that makes full use of 32 GPUs to small-scale tests with 1 GPU. Right now, we're using Cloud Shell to manage the deployment and scheduling of AI frameworks, but we're working to create an automated environment that is easier for customers to use." (Arakawa)

The challenge for Arakawa and his colleagues is to combine container technology with NetApp ONTAP AI. This would be a mechanism that could freely configure an image library converted into a Docker container using the container orchestration tool Kubernetes.

Arakawa expressed his goals as, "Our objective is to create an environment in which the customer can use the necessary data and computational resources immediately on demand by simply selecting a container registered in the image library. By automating and speeding up the setup, we can maximize the time our customers spend on verification and increase the efficiency of AI_LAB resource utilization."

The storage OS ONTAP is also praised for its high level of compatibility with container environments. One example of this is the provision of NetApp Trident, a tool that makes it easy to create the "persistent volumes" that are essential for handling applications and databases in a container environment.

Tanimoto says, "While using NetApp AFF A800, as it is as a data storage destination in a container environment, it will be possible to do things like move applications that incorporate the learning model into the cloud and run them there."

NetApp®

Benefits

Achieving better business outcomes using AI/deep learning

While the company is committed to developing the AI skills of its own engineers, CTC is also involved in projects to develop engineers with advanced IT skills through industry-academia collaboration. An example of this is the free provision of an AI_LAB environment where high-performance GPUs are available for the "Reiwa Year Data Analysis Competition," held by the Joint Association Study group of Management Science (JASMAC).

Arakawa explains, "The application of AI/deep learning to business and social infrastructure is expanding dramatically. We think it will be meaningful if AI_LAB can contribute to the fostering of AI human resources and the development of society."

There are new business challenges in a variety of fields, such as the detection of equipment failure signs and devices using IoT, support for diagnosis through analysis of medical images, inquiry responses using virtual assistants, and materials informatics that efficiently search for new materials. Arakawa concludes with the following:

"With AI_LAB, we intend to further strengthen AI verification environments and services to provide the link between the challenges faced by our customers and their business results. For NetApp, we look forward to the evolution of technologies for realizing cutting-edge storage products and data fabrics. We also hope that cooperation in the AI/deep learning space will take our alliance to a new stage."

Solution components

NetApp Products

NetApp ONTAP AI

NetApp AFF A800
High-end all-flash array

For details

https://www.netapp.com/us/products/ontap-ai.aspx

Contact us

+1 877 263 8277

NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify the management of applications and data across the cloud and on-premises environments to accelerate digital transformation. Together with its partners, NetApp empowers global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation, and optimize their operations. For details, visit www.netapp.com.
#DataDriven