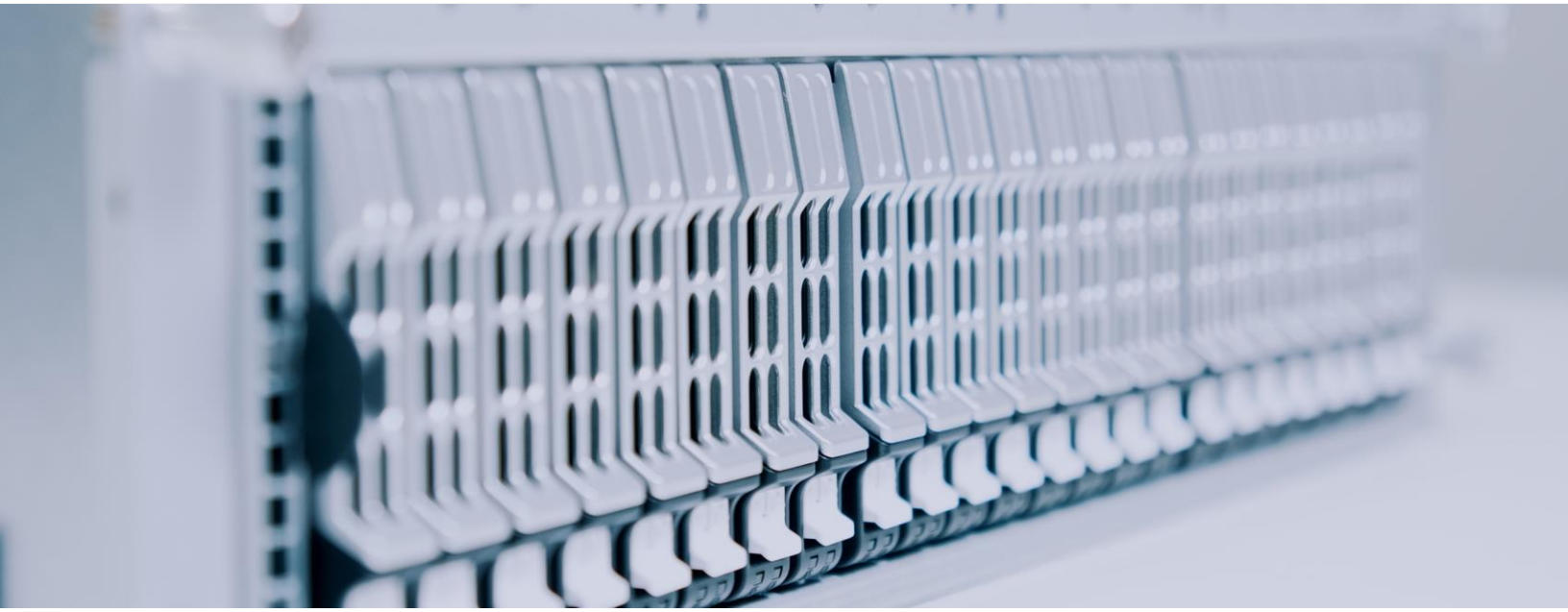
Supermicro Contributes to the MN-3 Supercomputer Earning #1 on Green500 list

Sophisticated Systems Supporting Deep Learning Projects at Preferred Networks Earn International Recognition



INDUSTRY

Government Research

Supermicro and Preferred Networks (PFN) collaborated to develop the most efficient supercomputer anywhere on earth, earning the #1 position on the Green500 list. This supercomputer, the MN-3, is comprised of Intel® Xeon® Scalable processors and MN-Core™ boards developed by Preferred Networks. The Green500 list is compiled semi-annually and recognizes the most efficient supercomputer in the world based on real-world scientific application benchmarks. The MN-3 delivers over 21 GFlops of computing power per watt of electricity.


The GREEN 500 CERTIFICATE

MN-3 - MN-Core Server, Xeon 8260M 24C 2.4GHz, MN-Core, RoCEv2/MN-Core DirectConnect

Preferred Networks, Japan

is ranked

No. 1 in the Green500

among the World's TOP500 Supercomputers

with 21.11 GFlops/Watt Linpack Power-Efficiency

on the Green500 List published at ISC 2020 Digital Conference, June 22nd, 2020

Congratulations from the Green500 Editors

Wu-chun Feng
Virginia Tech

Kirk Cameron
Virginia Tech

When looking at the performance per watt of electricity, the MN-3 is ranked #1 in the world in the Green500 ranking. The MN-3 delivers a whopping 21.11 Gigaflops of performance per watt on a benchmark run that showed a total of 1.62 Petaflops performance. In recognition of this engineering effort between Preferred Networks and Supermicro, the MN-3 was awarded the "Number 1 in the Green500" certificate at the annual (virtual in 2020) ISC conference in June, 2020.

You might ask, how was this world-record-breaking system designed?

Overview

Preferred Networks (PFN) is a leading technology company based in Japan that strives to solve complex problems that require deep learning, robotics, and other latest technologies. Many issues are difficult to address with existing technologies, requiring PFN to innovate with state-of-the-art technology to solve them.

By working with a wide range of partners across multiple industries, PFN can address a variety of challenges that are brought to them. Using cutting-edge technologies, PFN innovates through both hardware and software, combined with industry expertise.

Challenges

PFN recently faced a challenge where existing technology was not able to solve particular machine/deep learning (DL) applications with the agreed-upon Service Level Agreements (SLAs) for their customers. PFN was also concerned with the energy usage that such a large system would require. PFN needed to develop technology that would drastically speed up their DL training applications that are used in a variety of domains. Some of the projects that PFN specializes in have to do with computer vision applications. Applications that are designed to recognize a face, scene, or object require a training phase, where thousands to millions of images are used to train the application. The massive amounts of images are typically stored or archived on an external storage system. For performance reasons, the images are first brought into solid-state drive (SSD) drives from an image database before the processing and training take place.

Besides the required performance requirements, PFN was trying to understand and plan for the overall energy footprint that a new supercomputer would require. With increasing power costs, a large system of the size that PFN was going to need, the operating costs of both the power and associated cooling would exceed the budget that was allocated. Therefore, the energy efficiency of the new solution would have to be designed into the system, and not become an afterthought. PFN decided that partnering with a known supplier whose expertise included power-efficient designs would be critical to this project. Both Supermicro and PFN were excited that the MN-3 supercomputer is now recognized as the most efficient supercomputer globally in the June, 2020 Green500 rankings.

The performance of existing accelerators was not keeping up with the demands of their customers, and PFN needed a faster, more optimized solution. PFN determined that they would have to architect and design an ASIC that met their demanding requirements where an off the shelf solution did not exist. A joint project team between

PFN and Kobe University was assembled and created the MN-Core, which is designed for the training phase in deep learning.

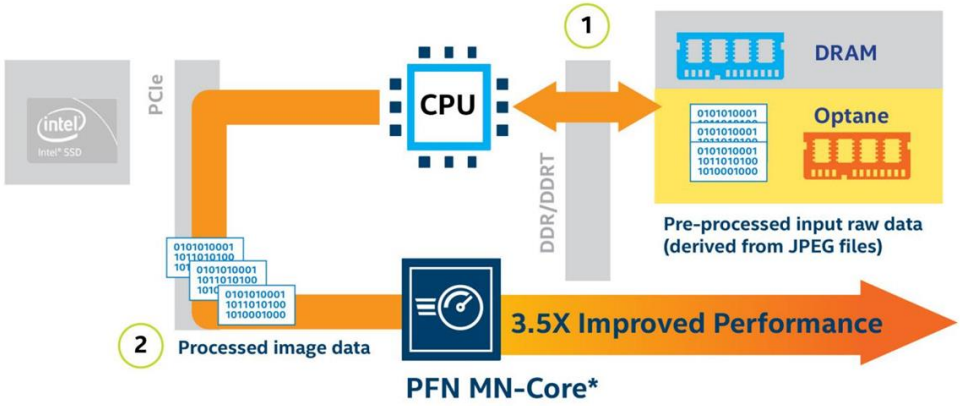
The engineers were able to get a high number of teraflops of computing power by designing an ASIC that is highly optimized for deep learning, and the team was able to create a processor that was not only fast, but energy efficient as well. During the design phase, the performance per watt of computing power was particularly important, not only because of the cost of power, but as more wattage is used, the cooling requirement goes up as well. PFN achieved its design goal and was able to benchmark the MN-Core board above 1 Teraflop (1,000,000,000,000 floating-point operations) per watt in actual applications.

Solution

As the new ASIC was being designed, the engineers realized that they would have to find a partner to work with to house the MN-Core board. There were several considerations that Preferred Networks required in a partner:

- Experience working on demanding projects that pushed the engineering team
- Design expertise and flexibility with system layout and configuration
- Deep understanding of the mechanical and power requirements
- Expertise in power-efficient system designs
- Ability to meet tight deadlines

Initially, PFN contacted several global suppliers who would be able to supply PFN with the required knowledge and product design expertise. After an exhaustive selection process, PFN chose Supermicro as their partner because they were able to leverage Supermicro's expertise in creating a custom solution for housing the MN-Core board along with two Intel® Xeon® Platinum processors. The mechanical solution required out-of-the-box thinking by Supermicro -- the ability to fit not only two CPUs, four MN-Core boards, but also up to 6TB of DDR4 memory. Each server would also have to house the Intel® Optane™ persistent memory modules, each containing 3TB of memory. Image AA below shows the overall architecture and workflow of an MN-Core server, which includes Intel Xeon scalable processors, Intel® SSDs, Intel® Optane™ persistent memory, and the MN-Core board, all integrated on to a Supermicro motherboard.

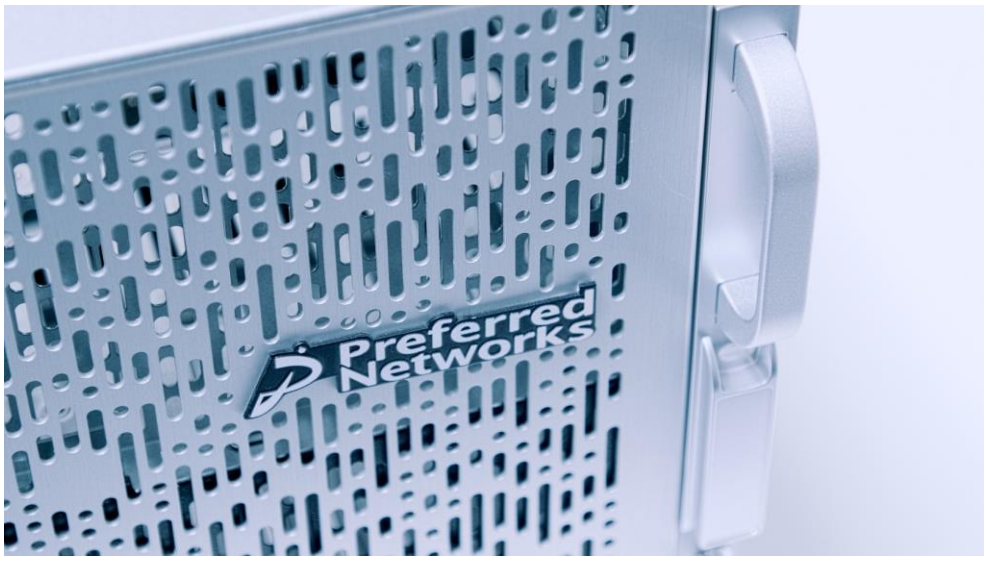


Overall architecture and workflow of an MN-Core board.

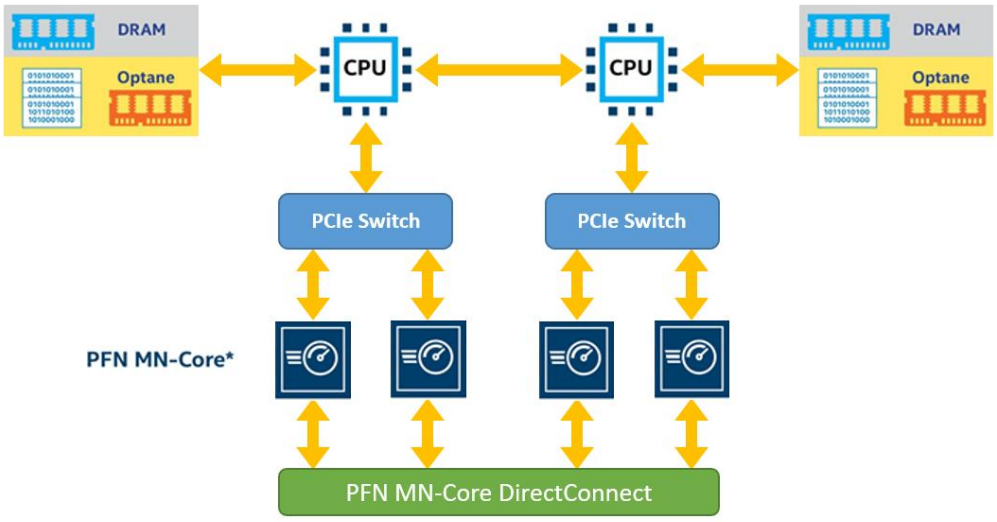
Specifically, the solution that PFN designed is based on the Supermicro GPU server that contains the X11DPG-OT-CPU motherboard. This advanced system is ideal for housing multiple GPUs or accelerators, as well as the interconnects that enable ultra-fast communication between GPUs. The PFN customized server is shown below, containing all this technology, and is housed in a 7U high rack-mounted unit.

The PFN Customized server

In the block diagram below, note that the PFN customized server contains four of the MN-Core boards. Each of these boards is directly connected to the PCIex16 lanes to communicate with the CPUs. The MN-Core boards are also connected to the MN-Core Direct Connect board, which allows for high-speed communication between the individual MN-Core boards. PFN partnered with Supermicro to develop this customized server, which addresses a wide range of applications that require ultra-fast communications. The image below is a block diagram of the MN-Core server communication paths between hardware components



The PFN Customized server detail



Block diagram of PFN solution

PRODUCTS

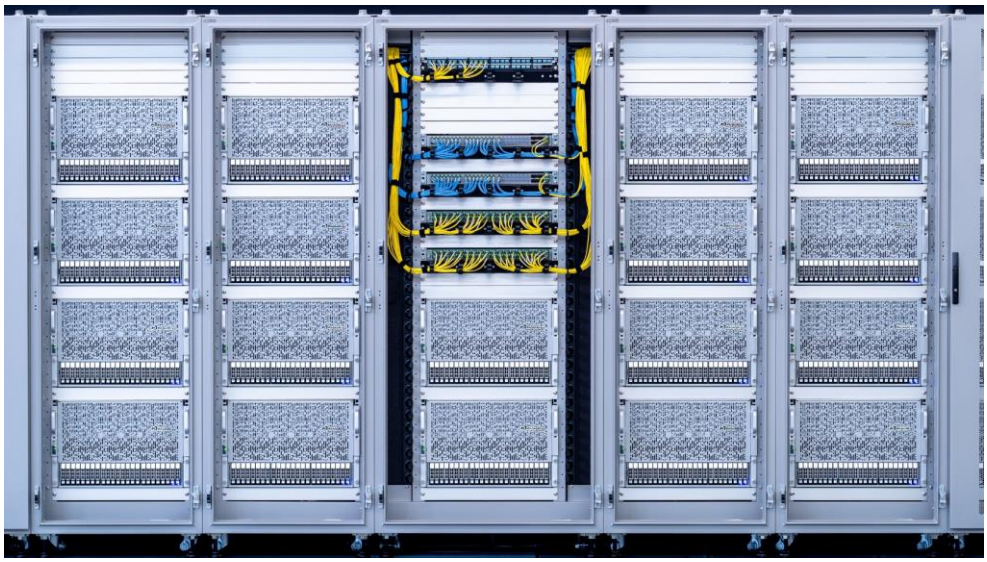
40 servers, four interconnect nodes, and five 100GbE switches, a total of 2080 cores, based on Supermicro GPU server that contains the X11DPG-OT-CPU motherboard.

Benefits

Once the server was designed, PFN then built a supercomputer, called the MN-3, that would be able to serve many constituents at once. The initial cluster would consist of 40 servers, four interconnect nodes, and five 100GbE switches. There was a total of 2080 cores, which delivered a fantastic 1,621.1 TFlops of performance while using just 77 kW of power. This supercomputer is ranked # 393 in the world Top500 list, based on the LINPAC benchmark as of June 2020. The image below shows a small section of the Preferred Networks MN-3 Supercomputer.

"We are very pleased to have partnered with Supermicro, who worked with us very closely to build MN-3, which was recognized as the world's most energy-efficiency supercomputer," stated Yusuke Doi, VP of Computing Infrastructure at Preferred Networks. "We can deliver outstanding performance while using a fraction of the power that was previously required for such a large supercomputer."

–Yusuke Doi, VP of Computing Infrastructure at Preferred Networks



A portion of the MN-3 supercomputer located in Japan Agency for Marine-Earth Science and Technology (JAMSTEC) 's Simulator Building at Yokohama Institute for Earth Sciences

In summary, PFN partnered with Supermicro to develop a server that accelerated DL algorithms. The industry expertise of Preferred Networks and the ability to design a custom ASIC combined with the systems expertise of Supermicro will enable PFN to create new applications that address their customers' most pressing requirements. The MN-3 supercomputer leads the entire industry in energy efficiency, resulting in reduced operating costs while accelerating innovation.

SUPERMICRO

Supermicro is a global leader in high performance, green computing server technology and innovation. We provide our global customers with application-optimized servers and workstations customized with blade, storage, and GPU solutions. Our products offer proven reliability, superior design, and one of the industry's broadest array of product configurations, to fit all computational need.

For more information, visit
https://www.supermicro.com