# Osaka University Creates AI Supercomputer with Supermicro SuperBlades Utilizing 3ʳᵈ Gen Intel® Xeon® Scalable Processors

*New Features Enable Faster Results and AI Research*



## Introduction

Osaka University is a world-renowned university known for leading-edge research in a wide range of disciplines. As a 90-year-old institution, Osaka University serves a wide range of researchers, students, and external experts, making available a supercomputer system that runs state-of-the-art software on the latest CPUs and GPUs.

## Challenges

The Cybermedia Center at Osaka University offers high-performance systems and storage to diverse users, both internal to the university and external researchers. The need for faster processing and a reduced time to solution for a wide range of disciplines was a key factor when Osaka University looked to upgrade their system. Osaka University is expanding its research areas, including high-performance data analysis, Artificial Intelligence (AI), Deep Learning (DL), as well as more traditional High-Performance Computing (HPC), which required the fastest system to be constructed in a balanced way for high performance and the lowest TCO. This fact

**INDUSTRY**

University Research and Industry Collaboration

**CHALLENGES**

- Faster processing and reduced time to solution
- Respond quick for new application requests
- Ability to take advantage of new CPU instructions

## SOLUTION

- Supermicro SuperBlade
- Supermicro GPU systems
- 3rd Gen Intel® Xeon® Scalable Processors

## BENEFITS

- Faster simulations
- AI Workloads with the latest GPUs and PCI-E 4.0
- Workload Matching

## PRODUCTS

- Compute Nodes: SuperBlade® 8U

  SBI-420P blade

  SBE-820 enclosure, 4 enclosures per rack

  InfiniBand HDR 200
- GPU nodes:

  SYS-420GP-TNAR
- Front end Nodes:

  SYS-620U-TNR

meant that the CPUs had to be fast with an increasing number of cores compared to previous CPU generations, but communication with other devices such as GPUs and storage had to be increased as well.

HPC applications always need the latest technologies such as processing speeds and other system capabilities for research into diverse fields such as genome research, medical, weather prediction, earth sciences, and manufacturing simulations. With an increasing demand for performance for a range of innovative technologies, Osaka University needed a new system that could address the demands of the researchers. Thus, the SQUID (Supercomputer for Quest to Unsolved Interdisciplinary Datascience) system was developed to meet these new challenges.

Osaka University offers a wide range of software to its users that is thoroughly tested and qualified for new systems. When moving from one generation to the next with Intel products, the binary compatibility makes this qualification relatively easy. Although applications may not be entirely optimized for a new instruction set, the application should run without any issues.  Re-compiling with updated compilers will increase the optimization amounts when a new CPU is introduced, containing new and more performant instructions.  Osaka University expects the FLOPS/Watt to be significantly improved for reducing the carbon footprint with the new system. Users will get massive improvement with their application performance by changing the parallelization tuning to include the latest Intel® Xeon® Scalable processor's instructions.

## Solution

Osaka University worked closely with NEC Corporation and Supermicro to design and implement a new supercomputing class system based on a combination of Supermicro SuperBlade® systems and 4U 8 GPU NVIDIA NVlink ® systems, which is built with the 3rd Gen Intel® Xeon® Scalable Processors. These systems were benchmarked with the new X12 SuperBlade and exceeded the expectations of the system architects and early users. The performance, cost, and power usage of the latest 3rd Gen Intel® Xeon® Scalable Processors have many advantages compared to previous generations of Intel® Xeon® Scalable Processors.

"Researchers and scientists in Osaka University are now working on their research projects in the hope of bringing open innovation to our society. Furthermore, the Cybermedia Center at Osaka University is in charge of delivering high performance to researchers in Japan to advance scientific research as a national joint-use facility. The new supercomputing system named SQUID, scheduled to begin operations in May 2021, was designed and implemented to facilitate researchers and scientists to quest to unsolved interdisciplinary data science. The test drive conducted on SQUID before its commencement of operation showed us that the Supermicro servers with Intel CPUs provided more performance than was expected. We are excited to leverage the latest generation of Intel and Supermicro's solutions for advanced scientific research projects necessitating HPC (High-Performance Computing) and HPDA (High-Performance Data Analysis)." -  Dr. Susumu Date, Cybermedia Center, Osaka University.

The SQUID system is a 27 Rack cluster that contains 1520 SuperBlades, 42 nodes of the 4U 8 GPU NVLink systems, and was integrated and tested by the Supermicro Rack Integration Services team.

- The systems use the 3rd Gen Intel® Xeon® Scalable Processors, specifically the Platinum 8368, running at 2.4GHz, 38 Cores, TDP 270W. Though this high TDP processor, in many cases, can be air-cooled, the most effective cooling solution for stable computing is to use liquid cooling. Supermicro Direct Liquid Cooling solution is the most cost-effective method to reduce the carbon footprint and PUE of their data center.

- The cluster is constructed from  Supermicro X12 SuperBlades consisting of 1520 blades (nodes) in 80, 8U enclosures with 3rd Gen Intel Xeon Scalable Processors for high-performance computing. All of the chassis are integrated with the direct liquid cooling system.

- A complete rack integrated solution has direct liquid cooling to provide optimized thermal efficiency for 76 nodes in a 42U rack with InfiniBand HDR 200 connectivity between nodes.

- The Supermicro Rack Integration Services team installed and configured the systems into 42U racks and tested the systems to ensure multi-rack level confidence upon delivery.

- The GPU rack for the heterogeneous HPC/HPDA needs also has direct liquid cooling to its CPU and GPU systems with A100 GPUs and NVSwitch™ modules, with each rack having 48 GPUs in six SYS-420GP-TNAR servers.

- The 2U Ultra server, SYS-620U-TNR, with direct liquid cooling, is used as the front end node and is equipped with the Quadro RTX6000 card and hardware RAID configuration.

The total deployment is 27 complete racks which were all assembled, integrated, and tested at the Supermicro factory and then directly shipped to Osaka University. Below is a table that summarizes the different nodes in the Squid supercomputing system.



*SuperBlade with DLC*



*GPU System with DLC*

| Product | Qty | 3rd Gen Intel® Xeon® Scalable processors | # of Sockets | Memory Per Node | Direct Liquid Cooling Component |
|---|---|---|---|---|---|
| **Compute node:** SuperBlade® 8U SBI-420P blade SBE-820 enclosure, 4 enclosures. per rack | 1520 blades 80 enclosures | 3rd Gen Intel® Xeon® Scalable processors, Platinum 8368, 2.4GHz, 38 Core, TDP 270W | Dual | DDR4-3200 256GB per node | CPU |
| **GPU node:** SYS-420GP-TNAR | 42 nodes | 3rd Gen Intel® Xeon® Scalable processor, Platinum 8368, 2.4GHz, 38 Core, TDP 270W | Dual | DDR4-3200 512GB per node | CPU, GPU, and NVSwitch |
| **Front end node:** SYS-620U-TNR | 10 nodes | 3rd Gen Intel® Xeon® Scalable processor, Platinum 8368, 2.4GHz, 38 Core, TDP 270W | Dual | DDR4-3200 256GB per node | CPU |

## 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
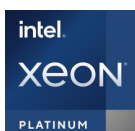https://www.supermicro.com

## Benefits

The research community's popularity and scientific needs that Osaka University serves required that more performance be added to the computing system. For many researchers, the run queues, which can be thought of as a backlog of applications waiting for their turn to run with the needed resources for their research, were growing longer, based on more complex simulations and the absence of GPU specific workloads. Critical studies that used the latest systems technology were delayed, and users were looking elsewhere to run their applications.

The arrival of the new systems from Supermicro, based on the 3rd Gen Intel® Xeon® Scalable Processors, the X12 SuperServer solution includes the latest GPUs and has benefited users in the following ways:

- Faster time to completion of simulations. With the new system, more simulations of similar resolution to be run in a given amount of time, and more research can be accomplished.
- Increased resolution or fidelity with existing applications. Faster CPUs enable more complex simulations to be run in a given time frame, with the addition of more physics or finer meshes.
- Workload matching – With the use of the new GPUs, applications in AI, ML, and DL, has been accelerated significantly, enabling researchers to develop new algorithms that were not possible previously.

The new Supermicro SuperBlade systems containing 3rd Gen Intel® Xeon® Scalable Processors enable Osaka University to conduct more research much faster than previous HPC systems would allow. Entire applications are now able to use the whole system, employing over 120,000 cores simultaneously. PCI-E 4.0 also enables fast communication with the latest GPUs, further enhancing and speeding up research applications.

**3RD GEN INTEL® XEON® SCALABLE PROCESSORS**

3rd Gen Intel® Xeon® Scalable processors, Platinum 8368

PCI-E 4.0

Support for up to 12 TB memory per socket

Further Resources, please visit:

Supermicro X12 Products:
https://www.supermicro.com/en/products/x12?utm_source=Supermicro%20Homepage&utm_medium=Homepage&utm_campaign=X12%20post%20launch

Supermicro SuperBlade Products -
https://www.supermicro.com/en/products/superblade

Supermicro GPU Systems - https://www.supermicro.com/en/products/GPU

Supermicro Rack Integration Services -
https://www.supermicro.com/en/solutions/rack-integration

SQUID system, Cybermedia Center, Osaka University - http://www.hpc.cmc.osaka-u.ac.jp/en/squid/