

TSUBAME4.0 Specifications

Center for Information Infrastructure, Institute of Science Tokyo

Overview

TSUBAME4.0, which have been operational since April 2024, boasts 20 times the peak computing performance of its predecessor, accelerating AI-driven cutting-edge research and convergence science. TSUBAME4.0 has been installed in a renovated building on the Suzukakedai campus.



System Total: 240 Compute Nodes

Peak Performance: 66.841 PFlops (FP64), 952.441 PFlops (FP16)

Memory: 180.0 TiB, 221.184 TB/s

Node Local SSD: 460.8 TB, Read 1,632 GB, Write 648 GB/s

Storage: HPE ClusterStor E1000

HDD 44. PB, 280 GB/s

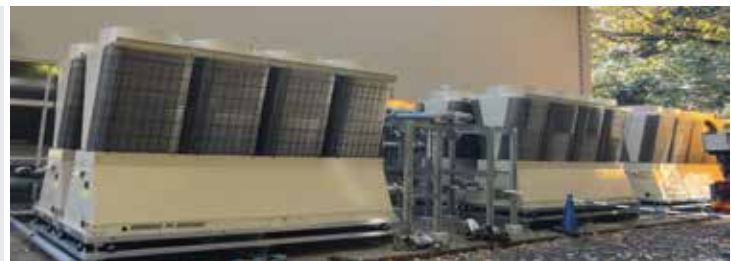
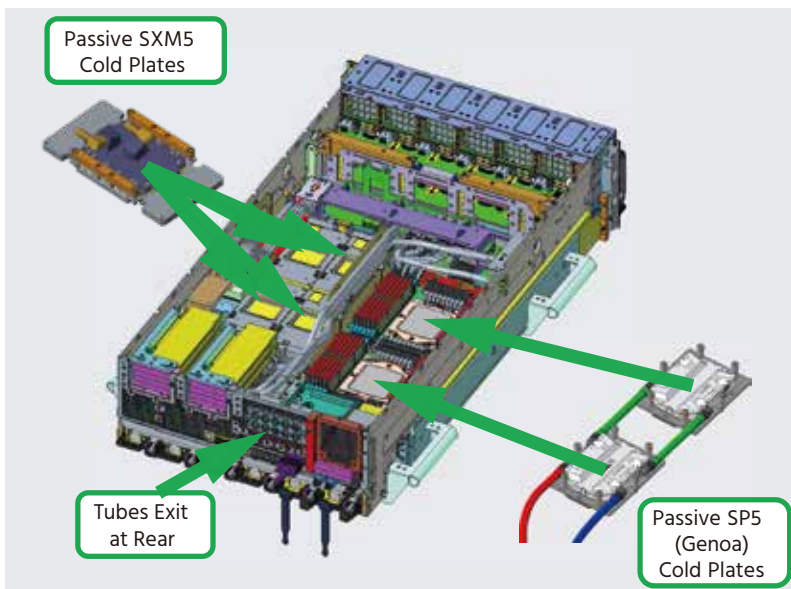
SSD 327 TB, 325 GB/s

Network: Full-bisection Fat Tree type InfiniBand



Implementation and Cooling

CPUs and GPUs are directly water-cooled. Other components are cooled by air, which is generated at the rear door of the rack. Inlet water temperature is expected to be ~20°C.

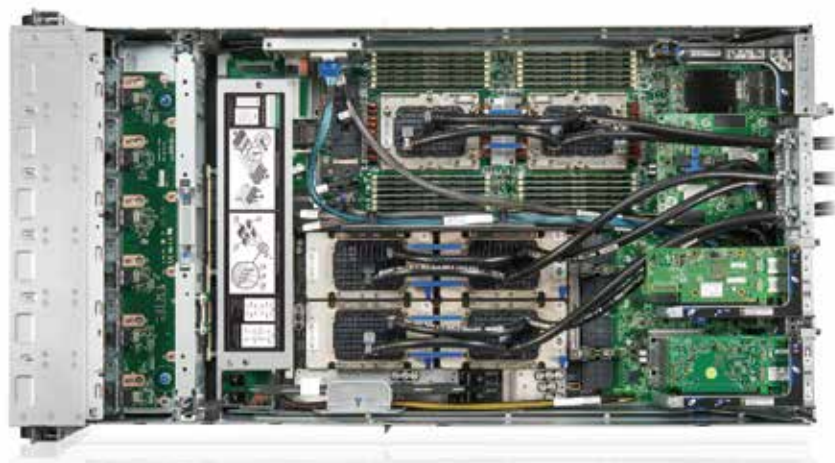


Chilling Units
(Cooling Capacity 200 kW/unit)



Compute Node

TSUBAME4.0 compute nodes are designed to accommodate various workloads including simulations, big-data analysis, AI applications, and also their convergence, which require combination of high performance processors, memory/storage and network. These components are installed in a 4U rack-mounted node using water/air hybrid cooling.



Nvidia H100 SXM5 94GB HBM2e (Hopper), 960 units in total

Node Specifications

HPE Cray XD665 Server

CPU: AMD EPYC 9654 (2.4GHz, 360W) ×2 sockets
96 cores per socket, total 192 cores per node.

GPU: NVIDIA H100 SXM5 94GB HBM2e ×4
Hopper GPU, 94GB, 2395.87GB/s Memory,
33.5TFlops(FP64), 66.9TFlops(FP64 Tensor),
66.9TFlops(FP32), 494.7TFlops(TF32 Tensor),
989.4TFlops(FP16/BF16 Tensor), 1,978.9TOps(INT8 Tensor)

Mem: 768GiB (DDR5-4800 32GiB ECC Registered module ×24)

SSD: Samsung PM9A3 (1.92TB NVMe U.2)
sequential read 6.8GB/s, sequential write 2.7GB/s.

NW: InfiniBand NDR200 (200Gbps) ×4

Predecessors

TSUBAME4.0 continues accelerated computing with GPUs



TSUBAME1.2 with Tesla S1070 (Tesla GT200)



TSUBAME2.0 with Tesla M2050 (Fermi)

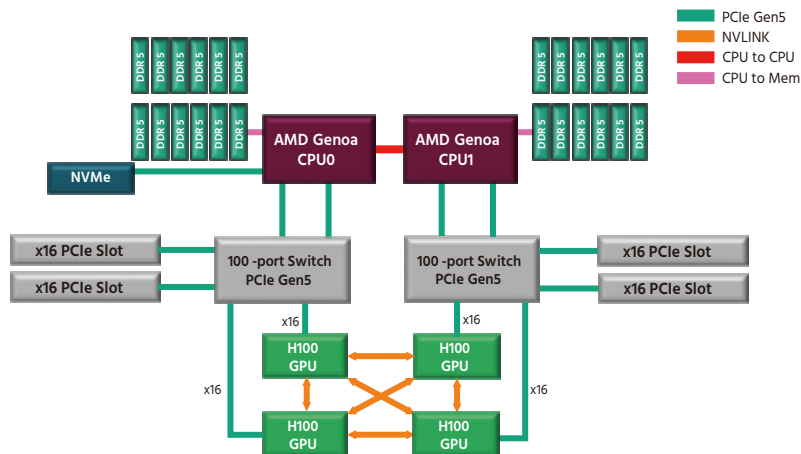


TSUBAME2.5 with Tesla K20X (Kepler)



TSUBAME3.0 with Tesla P100 (Pascal)

Block Diagram



Available Software

Compilers / Programming Languages / Debuggers

Intel oneAPI, NVIDIA HPC SDK, CUDA, GNU Compiler Collection, g++, Python, Ruby, Perl, PHP, R, Arm Forge

Linux Container

Aptainer

Applications / Tools

ANSYS, ABAQUS, ABAQUS CAE, AMBER, COMSOL Multiphysics, Gaussian, GaussView, Materials Studio, Discovery Studio, Mathematica, MATLAB, VASP, Schrödinger Small-Molecule Drug Discovery Suite, AlphaFold, CP2K, GAMESS, GROMACS, Hadoop, LAMMPS, NAMD, OpenFOAM, ParaView, POV-Ray, PyTorch, Quantum ESPRESSO, TensorFlow, Tinker, VESTA, VisIt, vmd, etc.

Green: Available for all users

Blue: Available for academic users

Black: Available for Science Tokyo users

Red: License required

Published by Center for Information Infrastructure, Institute of Science Tokyo

2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, JAPAN

TEL: +81-3-5734-2087 E-mail: tsubame@gsic.titech.ac.jp

<https://www.t4.gsic.titech.ac.jp/>

Version: 2024.10.31

