



Pegasus – Big memory supercomputer

Center for Computational Sciences
University of Tsukuba



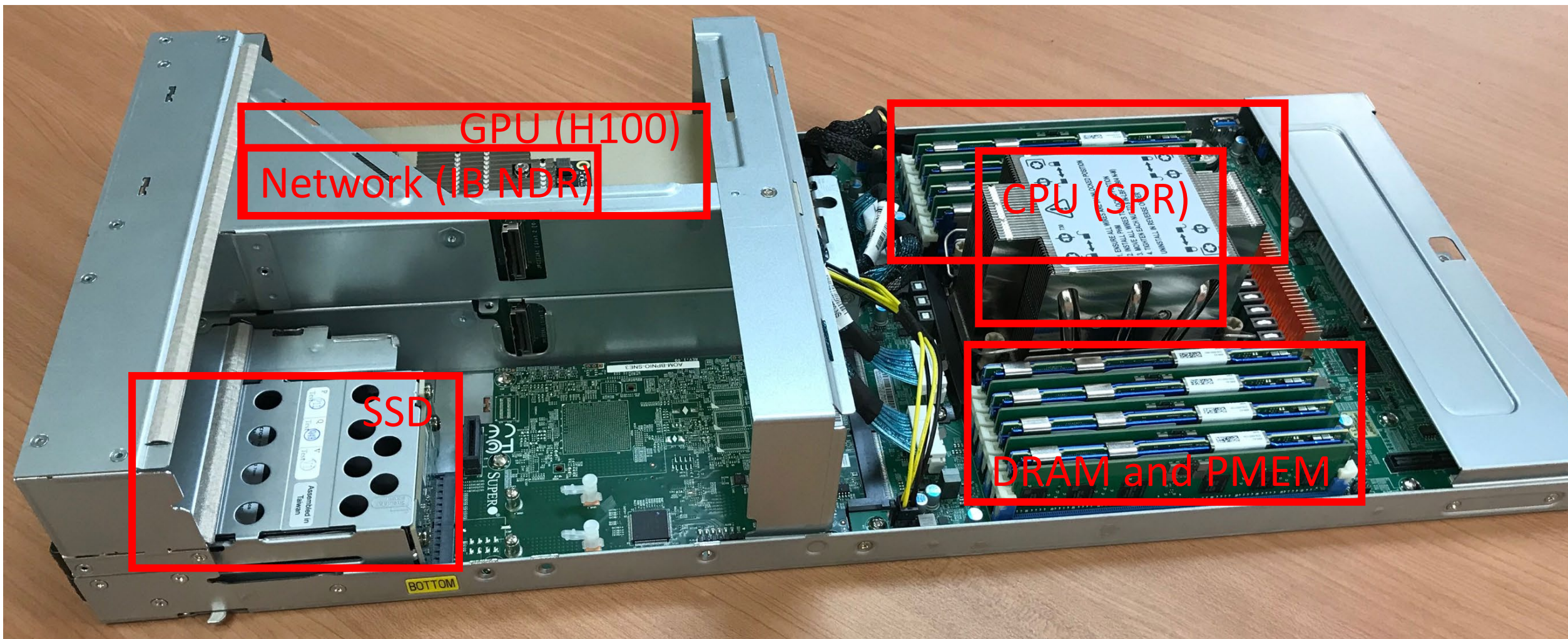
Background

- Data-driven and AI-driven Science requires large memory size and storage performance, but memory capacity per CPU core decreases
- Introduces Persistent Memory in compute nodes to accelerate large-scale data analysis and big data for better cost performance, power consumption, and application performance
- Fosters new fields of large-scale data analysis, new applications of big data AI, and system software research

Pegasus – Big memory supercomputer

System name	Pegasus
Total performance	6.5 PFlops
Total memory size	255 TiByte (15 TiByte DDR5 + 240 TiByte Persistent Memory)
Number of nodes	120
Interconnects	Full bisection fat-tree network interconnected by the NVIDIA Quantum-2 InfiniBand platform
Parallel file system	7.1PB DDN EXAScaler (40 GB/s)

Compute Node (Blade)



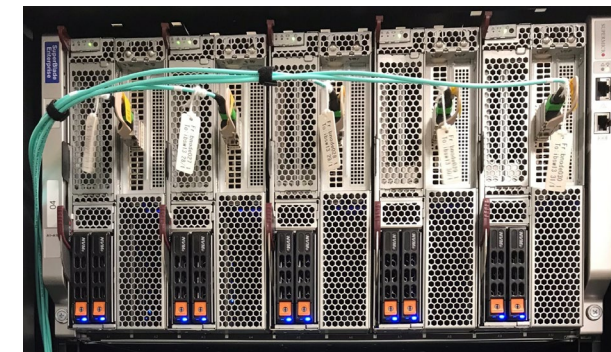
Specification (1/2)

- Compute nodes (NEC LX 102Bk-6) x 120

CPU	Intel Xeon Platinum 8468 (codenamed Sapphire Rapids) 2.1GHz/48c (3.2256 TFlops)
GPU	NVIDIA H100 Tensor Core GPU with PCIe (51 TFlops in FP64 Tensor Core)
Memory	128GiB DDR5-4800 (282 GB/s)
Persistent memory	2TiB Intel Optane persistent memory 300 series
SSD	2 x 3.2TB NVMe SSD (7 GB/s)
Networking	NVIDIA Quantum-2 InfiniBand platform (200 Gbps)

- Login nodes (NEC LX 124Rk-2) x 3
 - 2 x Intel Xeon Platinum 8468 (2.1GHz/48c)
 - 256 GiB DDR5 Memory, NVMe SSD, InfiniBand x 2, 100GbE

NEC LX B1000E Blade Enclosure



NEC LX 102Bk-6



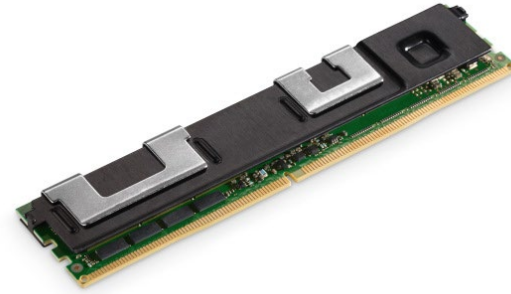
NEC LX 124Rk-2

Specification (2/2)

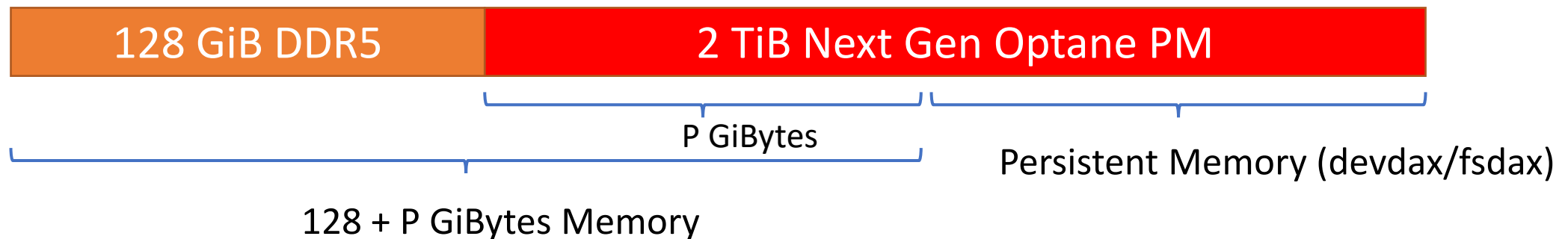
- Parallel File System (DDN ES200NV/ES7990X/SS9012)
 - DDN EXAScaler (Lustre)
 - MDS/MDT (**4.2 billion inodes**)
 - Active/Standby MDS
 - 1.92 TB NVMe SSD x 11 (8D + 2P + 1HS)
 - InfiniBand HDR100 x 4
 - OSS/OST (**7.1 PF available**)
 - 4 x Active/Active OSSs
 - 18 TB 7,200rpm NL-SAS x 534 ((33drives x 8pools + 3I
 - 8D + 2P Declustered RAID
 - InfiniBand HDR100 x 8



Persistent Memory



- Persistent memory is configured as **Filesystem DAX** (fsdax) by default
 - DAX (Direct Access) mounted XFS is available at **/pmem**
- Pmem mode can be configured by the environment variables
 - MV_SIZE – use Pmem for **extended memory** and specify the **memory size** in GiB or ALL
 - USE_DEVDAX – use **Device DAX** (devdax) instead of fsdax



Software Component

- Ubuntu
- Intel oneAPI (C++/C/Fortran, oneMKL, MPI, VTune, Trace Analyzer&Collector)
- NVIDIA HPC SDK (C++/C/Fortran/Cuda, cuBLAS, cuTENSOR, cuFFT, ..., Open MPI, NVSHMEM, NCCL, Profilers, Debugger)
- Open Source SDK (GNU Compilers, Python, PMDK, Open MPI)
- Tensorflow, Keras, PyTorch, ...
- JupyterHub, TensorBoard, Nextcloud, Gfarm