



Building Korea's first university-based supercomputing center with Backend.AI at Sungkyunkwan University

See how Sungkyunkwan University transforms
HPC infrastructure with Backend.AI



Building Korea's first university-based supercomputing center with Backend.AI at Sungkyunkwan University

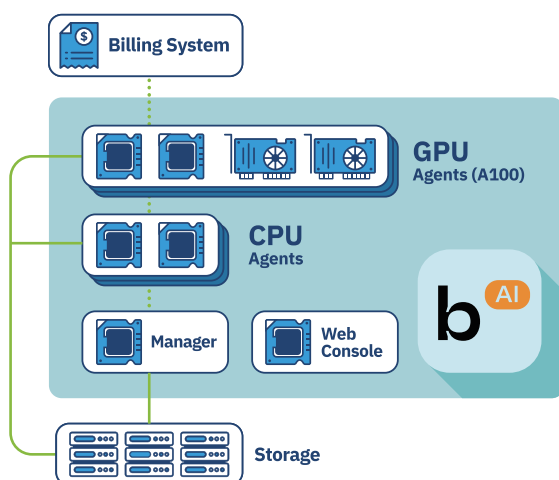
"Running a university supercomputing center prevents redundant spending and, with ongoing upgrades, will greatly boost research and hands-on learning across campus."

SKKU
Director, Supercomputing Center
Hyoungkee Choi

Sungkyunkwan University (SKKU) is widely recognized for its strong capabilities in computer science and engineering. SKKU aimed to build an independent high-performance computing (HPC) infrastructure capable of processing large-scale computations efficiently. The centralized infrastructure is equipped with multiple accelerators including 40+ NVIDIA A100 GPUs, high-density compute nodes with numerous CPU cores per node, and offers a storage capacity of approximately 2000 terabytes, sufficient for handling intensive data-driven research workloads. Opened in 2022, SKKU became the first university in Korea to operate a supercomputing center.

A way to manage scalable HPC infrastructure

While the university secured groundbreaking hardware, managing such high-end computing resources posed significant challenges. Budget constraints and the complexities of resource allocation across diverse research groups made it essential to monitor usage. SKKU faced hurdles in operating and maintaining such as cost duplication, operational inefficiencies, and a lack of flexible resource management. Furthermore, providing fair access to computing power for individual researchers required a scalable approach to resource procurement, usage monitoring, and cost allocation.



Backend.AI: Centralizing resources for easy management

To meet these challenges, SKKU adopted Backend.AI to streamline GPU cluster management. The platform automated resource allocation, usage tracking, and billing tailored to each researcher's consumption, while its session-based architecture let researchers deploy their preferred machine learning frameworks on demand.

The center operates as open research infrastructure, supporting university members alongside partners and approved external researchers. Shared resource pool with per-user quotas keeps idle capacity low, adapts to diverse research demands, gives admin fine-grained policy control, and aligns with a usage-based billing policy: researchers use what they need and pay only for what they used, at a cost below commercial cloud services. For time-critical work, Backend.AI's resource group provides guaranteed capacity for priority projects.