NAVER

intel

UP TO 6.37X boost

for inference performance with Intel® Distribution of OpenVINO™ toolkit.[1]

"We look forward to continued collaboration, working closely with Intel to optimize our AI models and exploring other data types and Intel® Deep Learning Boost."

Bado Lee, OCR Leader, NAVER Corporation

Low-latency Machine-learning Inference on Industry-standard Servers

As NAVER expands its services, including AI as a Service, it is turning to Intel to help fine-tune its machine-learning hardware and software stack. Performance is key to meeting service-level agreements (SLA), but cost efficiency, resource utilization and power consumption are important considerations. Using GPUs for inference provided great performance, but consumed too much power and left the company's CPU-based servers standing idle. Tests showed the performance benefits that result from combining 2nd Gen Intel® Xeon® Scalable processors with a ML framework optimized for Intel® architecture. PyTorch with Intel® Extensions for PyTorch and the Intel® Distribution of OpenVINO™ toolkit demonstrated that NAVER could meet its SLAs while driving up data center efficiency.

Products and Solutions
2nd Gen Intel® Xeon® Scalable processors
Intel® Distribution of OpenVINO™ toolkit
Intel® Extension for PyTorch*

Industry
Internet

Organization Size
1,001-5,000

Country
South Korea

Learn more
Case Study