ARTIFICIAL INTELLIGENCE AT ANY SIZE

Artificial intelligence (AI) and, in particular, deep learning (DL) are rapidly transforming entire industries and scientific disciplines. If your toe isn't already in the AI/DL water, it might feel like you're already too late. You're not — because Cray can help. How? By giving you the tools and support to guide you to success whether you're just starting out or ready for production AI.

Cray delivers a powerful range of accelerated systems with NVIDIA GPUs for machine learning, deep learning and extreme HPC. You can even use different Cray platforms together to enable complete multi-stage workflows. We work with leading deep learning packages on systems that scale beyond typical GPU density.

GPU-Accelerated Computing

The Cray® CS-Storm™ cluster supercomputer is a flexible, open and integrated system for machine learning and extreme HPC computing, and the foundation for organizations that want to take advantage of GPU accelerators in their production workflows. Our CS-Storm series includes support for NVIDIA Volta and NVIDIA Pascal™ GPU architecture accelerators — up to 10 NVIDIA® Tesla® V100, P40 or P100 PCIe accelerators (CS-Storm 500GT), and four or eight NVIDIA Tesla V100 or P100 SXM2 accelerators utilizing the NVIDIA® NVLink™ high-speed interconnect (CS-Storm 500NX).

Test, Launch & Grow Your Deep Learning Initiatives

Cray Accel AI solutions leverage the enterprise-ready, industry-standards-based, fully scalable Cray CS-Storm accelerated cluster supercomputer and come with a robust deep learning environment from Bright Computing that includes TensorFlow™, MXNet, Caffe2, Chainer, Microsoft Cognitive Toolkit and more. Choose from three fast-start configurations: the Accel AI Quick Start for initial deep learning trials, the Accel AI Cluster Starter Kit for deep learning exploration and small proof-of-concept projects, or the Accel AI Deep Learning System for production-level deep learning training and inference.

Storage to Handle the Largest Problems

ClusterStor™ storage systems solve the most demanding ML/DL/AI challenges in high-performance environments. Use cases with enormous datasets, like autonomous driving, require up to 500 PB and performance up to 1,000 GB/s. Those requirements, paired with the diverse data types being collected and analyzed, can quickly tax the capabilities of most storage systems. The Lustre®-based ClusterStor L300N model with NXD easily exceeds those requirements and provides linear scaling without degradation in performance. Plus, the ClusterStor system's NXD flash acceleration technology manages random and mixed I/Os, so you can manage data and serve it up when and how you need it.