Boost AI efficiency
AI infrastructure software portfolio from HPE
AI and HPC are converging. Are you ready for what’s next?

Artificial intelligence (AI) is driven by data. Large data sets are required to train AI, which in turn helps you derive valuable insights from massive amounts of data.

As AI adoption expands, data scientists naturally want to run AI workloads on the most powerful technology available, which often means high-performance computing (HPC). The expansion of AI—and thus, HPC—into a greater variety of companies is driving demand for a software ecosystem with simplified programming and management. As a leader in the HPC space, HPE offers you a choice of stable, resilient AI software and solutions that will drive efficiencies and better business insights with less complexity—no matter your AI journey.

HPE AI infrastructure software portfolio

HPE offers a comprehensive portfolio of software solutions for AI so you can choose the right combination of software and deployment methods to fit your needs, from best-in-class offers in each category. As AI requirements evolve, software remains an area of constant change, and HPE will continue to evaluate the options on the market to offer the best software solutions available.
AI deployment software

AI software may be packaged as a container or as an executable on bare metal. Because AI resources are in high demand, deployment decisions need to be made with the goal of optimizing performance from GPU and CPU investments. The deployment model you choose should be based on your performance requirements, workload demands, and user preferences.

Containers
Typical user: Data scientists who want to run machine learning (ML) and deep learning (DL) workloads quickly and easily.

Decision factors
- **Portability:** You can take your containerized software and run it as a service on any Linux OS.
- **Performance:** There is about a 5% performance delta for containers versus bare metal, which is negligible for most workloads.
- **Repository:** You can use public container repositories such as NVIDIA® NGC.

Bare metal
Typical user: Data scientists managing combined HPC and AI environments with no need for special AI workflows.

Decision factors
- **Performance:** Need the best performance available.
- **HPC background:** Prefer running AI jobs the same way as HPC jobs.
- **User interface:** Prefer the speed and control of command line interface (CLI) to scripts and graphical user interfaces (GUIs).
- **Security:** Addresses the perception of containers not being secure.
AI-ready configurations from HPE

Depending on how you want to run your workloads, HPE offers several tested and proven AI-ready configurations.

### Containers: HPE BlueData EPIC

To scale quickly, data science teams need to adopt new tools and techniques that will allow them to get better results and quickly deliver more insights to the business. **BlueData EPIC software** transforms the infrastructure consumption experience for AI, enabling IT to create distributed environments for ML, data science, and analytics in minutes. Plus, you can offer a self-service experience with the data and tools your data science teams need, while providing enterprise-grade security and reducing costs.

Deploying BlueData EPIC software with GPU-enabled servers—such as HPE Apollo and HPE ProLiant servers with NVIDIA Tesla® or NVIDIA Quadro® GPUs—lets you turn your infrastructure into a **GPU-as-a-Service (GPUaaS)** solution. BlueData can consolidate GPUs from multiple servers and make them available for multiple applications—for on-demand and elastic provisioning of containerized GPU resources with just a few mouse clicks. This enables you to deliver GPUaaS in an on-premises deployment to increase business agility, optimize GPU utilization, and realize significant cost savings.

---

1 HPE Performance Cluster Manager is not supported on Ubuntu.
Containers: NVIDIA NGC

To get started with AI quickly without the need to invest in specialized software, you can tap into the NVIDIA NGC DL ecosystem. Developed by NVIDIA, NGC allows developers to access a DL software stack at no cost, to establish a development environment suitable for DL. NGC consists of software pre-packaged in containers and optimized for NVIDIA GPUs which you can readily run using Kubernetes or another container orchestration tool. This solution is available on NGC-ready platforms. This validation program was developed to provide a repeatable way to roll out AI and HPC applications from development to production. It consists of GPU-enabled systems that pass NGC-ready validation tests with NGC software. HPE offers several NGC-ready platforms, including HPE Apollo 6500 Gen10 Systems, and HPE ProLiant DL380 Gen10 Servers.

Bare Metal: Bright Data Science

As AI and HPC converge, software developers are embracing algorithms for HPC workloads, driving a preference for bare-metal deployment due to the ease of code modification. Rapid software installation for AI is offered to HPE partners with Bright Computing. This purpose-built, integrated hardware and software solution is based on GPU-enabled HPE HPC systems combined with Bright Cluster Manager for Data Science, enabling you to develop DL applications quickly by using Bright Cluster Manager with Bright Data Science add-on.

Bright Data Science add-on offers leading AI/DL frameworks that have been ported to run on multiple Linux OSs and can be installed quickly and easily using Bright Cluster Manager. Bright validates and updates the necessary software components for the DL environment so developers can focus on the application. In addition, Bright avoids potential container security issues by enabling Linux OS security to protect files and user access. Bright Cluster Manager and Data Science add-on are supported by Bright Computing.
HPE-enabled systems

Accelerate your data’s value with a proven, practical approach to AI
HPE understands AI. We’re putting AI into action to help you derive value from all your data, using AI to create new experiences, drive smarter operations, and enable breakthrough innovation. Our proven, practical approach, validated solutions and partners, AI-optimized infrastructure, and turnkey AI software platform—that you can consume as a service—reduce complexity and help you realize the value of data sooner.

HPE Apollo 6500 Gen10 System
The HPE Apollo 6500 Gen10 System provides unprecedented AI performance with support for up to eight industry-leading NVIDIA Tesla V100 GPUs with NVIDIA NVLink technology for fast GPU interconnect, high-bandwidth fabric, and a configurable GPU topology to match your AI workloads.

HPE Apollo 2000 Gen10 System
The HPE Apollo 2000 Gen10 System is designed as an enterprise-level, density-optimized, cost-effective 2U shared-infrastructure chassis that supports two to four GPUs. It allows you to mix and match servers—HPE ProLiant XL170r Gen10 for general purpose and HPE ProLiant XL190r Gen10 for AI workloads requiring up to two NVIDIA Tesla V100 PCIe GPUs per server and four GPUs in a 2U chassis.

HPE ProLiant DL380 Gen10 Server
The HPE ProLiant DL380 Gen10 Server, the industry’s most trusted compute platform, delivers maximum AI utility with the powerful acceleration of up to seven NVIDIA T4 GPUs, the world’s first GPU for accelerating mainstream enterprise service, plus validation as an NGC-ready platform.

1 NVLink provides industry-leading performance with dedicated GPU-to-GPU communication.
Why HPE?

AI can help turn data into insight, action, and value. HPE is the trusted, global partner you need to navigate this fast-changing field and realize the promise of AI everywhere.

Find out more today

Together, HPE and NVIDIA make it easier for you to unlock valuable insights from data with AI solutions customized for your workloads. Contact your HPE or authorized representative to find out more.

Learn more at hpe.com/info/deep-learning