

# HPE GREENLAKE FOR HPC DELIVERS AN INNOVATIVE AND FLEXIBLE ENTERPRISE HPC/AI SOLUTION

## EXECUTIVE SUMMARY

A decade ago, supercomputing was almost exclusively used by government and academic researchers to solve enormous and complex scientific problems. Today, the use of high-performance computing (HPC) has broadened significantly to include use by federal agencies and a growing number of industry segments. Combined with artificial intelligence, machine learning, data analytics, and other technologies—often deployed at the edge—HPC impacts everything from space exploration, drug discovery, and weather prediction to financial modeling, automotive design, and systems engineering.

At the corporate level, digital transformation is driving data-intensive workloads to operate at an unprecedented scale. Corporations need high-performance computing (HPC) to unlock the potential of data and efficiently run ever-growing models, simulations, artificial intelligence (AI), and analytics workloads.

Data continues to grow at an exponential rate, even though corporations already produce and store massive amounts of it. To manage and extract business intelligence from the mountains of existing data, companies are deploying AI-supported analytic platforms that are driving demand for HPC even higher.

A unique symbiotic link exists when AI and HPC operate within a shared infrastructure. HPC helps AI process and inference data faster. Conversely, AI helps HPC manage and monitor its processes and workloads more efficiently. Of key importance, however, is how and where HPC and AI jobs are run. HPC workloads can be run on-premises, in a colocation facility, in the public cloud, or as a mixture of services.

As one of the world's largest providers of HPC systems, HPE holds a significant leadership position in the market and offers comprehensive end-to-end solutions accompanied by expertise, support, and services. It is critically important that HPE has made a long-range corporate commitment to support and develop HPC as a service and make it available to more users with accompanying energy sustainability.

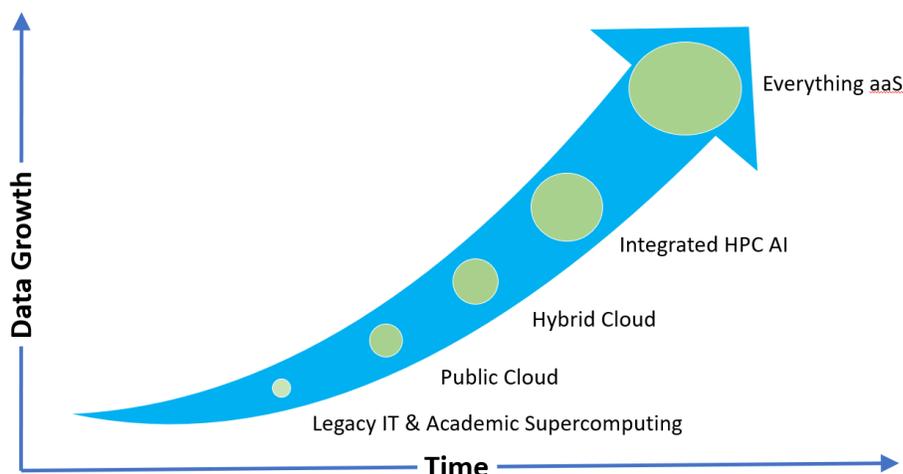
After a careful review, Moor Insights & Strategy (MI&S) has concluded that, for enterprises that want to provision an integrated AI and HPC system, HPE GreenLake for HPC offers an innovative platform that is flexible, scalable, and provides control of HPC solutions with a cloud service consumption model on- or off-premises.

This paper examines the environment, industry trends, and types of service providers to determine the best method to run the technology.

## THE TRANSFORMATION OF HPC

The synergistic nature of HPC and AI has dramatically impacted research, business, and our daily lives. It has transformed business, fueled scientific discoveries, and forever changed the world of computing. It has given us more precise weather forecasts, autonomous vehicles, highly accurate diagnostic medical procedures, instant credit approval, 360-degree fraud detection, national defense, and thousands of other applications too numerous to mention.

FIGURE 1: EVOLUTION AND CONVERGENCE - HPC AND AI TODAY



Source: Moor Insights & Strategy

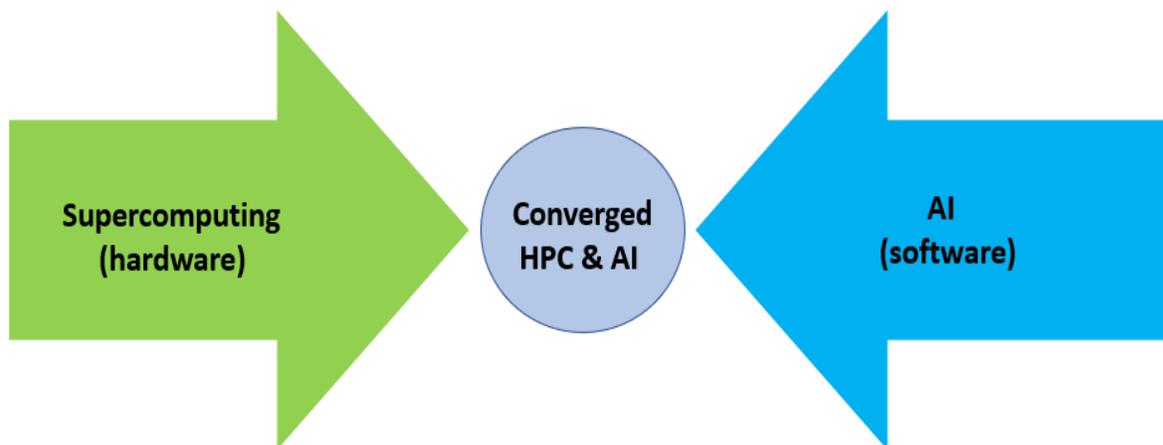
It is likely we will see increased usage of high-performance computing (HPC) as more organizations begin to use it to underpin both operational processes and strategic decision-making. Whether that involves large-scale, complex simulations, data analytics, or sophisticated artificial intelligence (AI), these workloads all demand dedicated HPC infrastructure.

Clusters of computers connected to a high-speed network and data storage system help power today's AI. The clusters have co-processors that include AMD and Intel central processing units (CPUs) and typically NVIDIA graphics processing units (GPUs) that work in parallel as workload accelerators to solve computationally complex problems and perform data analysis at speeds far beyond the capability of traditional computers.

### *ENTERPRISE NAYSAYERS*

Despite numerous demonstrated business advantages of HPC, some companies have shied away from using it. Their rationale includes beliefs that HPC requires too much capital or extensive workforce retraining will be required or that HPC is incompatible with enterprise applications. A thorough technical investigation will show these are erroneous assumptions. HPC can be provisioned with very flexible enterprise-grade solutions with cost-effective scaling options, varying configurations, and a variety of support options for large and small applications.

**FIGURE 2: THE PARTNERSHIP OF CONVERGED AI AND HPC**



*Source: Moor Insights & Strategy*

HPC and AI process enterprise data originating from unstructured and structured files broadly scattered across a company's internal and external data landscape. The technologies work together to inference and extract meaningful insights from these widely dispersed data sources and make it possible for a business to create competitive advantages, improve customer service, spur innovation, and achieve critical business goals.

The combination of AI and HPC creates a uniquely powerful data compute environment. It provides efficiencies to reduce compute time, lower costs, increase application flexibility, and tighten collaboration between resources. Running both AI and advanced analytics on HPC makes the system an order of magnitude more powerful.

### *HPC BENEFITS FOR AI*

HPC manages complex resources needed for AI development, which is much different from traditional software because of AI's size, dynamic behavior, nonlinear scaling, and unpredictable growth. In addition, the mountains of data used by AI are drawn from multiple sources and in different forms. Even after deployment, AI still requires continuous intervention and updating. HPC connects, facilitates, and accelerates these functions, allowing AI to operate and scale more efficiently.

HPC's scalable architecture increases AI's efficiency through simultaneous compute, storage, interconnects, and software. HPC can reduce AI processing bottlenecks and latency by accelerating complex data environments using GPU parallel processing power.

HPC clusters use CPU (software) and GPU (hardware) co-processors that work in parallel. A GPU has more cores than a CPU, and although its cores are smaller and less powerful than CPU cores, there are thousands of them. GPU architecture is designed to handle a massive number of parallel computations. A CPU, on the other hand, can rapidly run complex serial operations but since it only has a few cores it can't handle high-volume operations very well.

GPU acceleration is done at the hardware level and accomplished by moving computationally complex tasks from the CPU to the GPU for rapid resolution using its powerful parallel processing capability.

High-quality AI model predictions and classification are dependent on GPU hardware that enable the use of massive amounts of training data when required. GPU accelerated pre-trained models allow faster insights to be obtained.

GPUs used by HPC are key to AI's documented wide-ranging success in every industry. The technology can be used for model training as well as deep learning inference. GPUs accelerate compute with high throughput and by performing thousands of operations simultaneously. AI analyzes and interprets real-world events that are too difficult to see, understand, or measure.

Once integrated into an organization, HPC changes how a company and its management think, learn, collaborate, and engage the world.

### *AI BENEFITS FOR HPC*

AI makes HPC operate more efficiently by optimizing its workflows and improving model-targeting accuracy, which reduces the number of simulation runs. In addition, AI provides HPC with resource scheduling and workload distribution across available computing resources to maximize resource usage and monitor storage health, networking, and servers for proper configurations and failure predictions.

### *KEEPING PACE*

According to Jensen Huang, CEO of Nvidia, the complexity of AI models increases by up to 10x each year. That exceeds the speed of improvements for any computing technology—CPUs, GPUs, or networking. Compensating for the growth in model complexity requires a combination of many types of accelerated computing and continuous development of new types of parallel computing.

With AI model complexity setting the pace, HPC must have a continually evolving history to match it. An HPC service provider must have superior technical expertise to maintain the constant technological evolution necessary to match AI's ever-changing complexity. Its constancy is maintained because each HPC technical improvement establishes a platform on which the next improvement is built.

## HPC AND AI GOING FORWARD

As mentioned earlier, AI, compute, and mountains of processed data have impacted nearly every aspect of our daily lives. And the unprecedented exponential growth of data volume experienced over the past decade is expected to continue unabated for many years.

According to the *Harvard Business Review*, AI will add \$13 trillion to the global economy over the next decade.<sup>1</sup> AI growth translates into a corresponding growth for HPC that will likely push it out to a broader enterprise market to facilitate and expand its democratization.

---

<sup>1</sup> Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. *Harvard Business Review* 97 (4): 62–73.

HPC can manage future complex enterprise problems that will soon range from petascale to [exascale](#), making it the logical candidate to be the standard for next-generation computing. New advancements will also occur at the hardware's microarchitectural level, where improvements will speed up interconnections and increase computing densities, among other things.

While the use of HPC is expected to increase, some companies will likely resist revising their infrastructure. Still, competitive pressures will force most laggards to adopt some form of HPC to replace the limited capabilities of legacy IT and cloud.

The exponential growth of data and workload complexity will accelerate the speed and methodologies businesses use to make decisions, evaluate customer interactions, develop products, and create strategies. Legacy technology cannot match HPC's speed in performing these tasks which will create a market shift toward alternative solutions such as HPC.

Over the next decade, we could see the beginning of the next evolution in computing as quantum technology matures and larger quantum machines begin using millions of error-corrected qubits. Problems that are impossible today will become possible with the combined computational power of exascale HPC, quantum computers, and AI. Researchers are already contemplating the benefits and capabilities of such a powerful architecture.<sup>2</sup> The capabilities of such a powerful combination are almost incomprehensible.

## HPC AND THE PUBLIC CLOUD

Public cloud became popular when enterprises needed a more efficient data service than what was provided by legacy IT. Public cloud provides a pay-as-you-go consumption model and handles data with on-demand servers, compute, storage, and flexible network capacity that can be increased or decreased as needed.

Over the years, because public cloud offered so much flexibility and ease of use, many companies adopted a "cloud-first" or a "cloud-only" strategy. Instead of using non-cloud solutions for new applications, the cloud became the de facto solution for new data services. Today, almost every enterprise uses cloud services, and many have multiple clouds.

---

<sup>2</sup> High-Performance Computing with Quantum Processing, <https://dl.acm.org/doi/pdf/10.1145/3007651>

Although cloud has its advantages, it is not always the best choice, especially for applications with stringent requirements for data sovereignty, security, compliance, data transfer, and latency. All the major cloud service providers—AWS, Microsoft Azure, and Google Cloud—offer some form of HPC.

As with traditional cloud service advertising, HPC cloud service providers advertise that deploying HPC workloads in the cloud is easy and frees up data teams for more productive work.

It is a common tactic for cloud providers to highlight three advantages of cloud service:

1. There is no need to worry about managing infrastructure because the cloud does it for you.
2. Cloud is a monthly consumption model, and no capital is needed.
3. Cloud provides the flexibility to scale up or down as needed.

While those reasons may be valid for traditional data, they don't necessarily hold for HPC applications that involve complex configurations for clusters and parallel file systems. Moving HPC workloads into a public cloud is complicated. HPC workload efficiency depends on low latency data movement, powerful processing, and high availability of server resources.

HPC can be complex. Customers testing HPC in the public cloud are learning that the cloud was not built for this. Achieving the expected performance and predictable cost can be a challenge for several important reasons:

- Total cost of ownership (TCO) is always a consideration, and a large capital outlay is risky.
- Skills to operate, tune, and manage HPC are hard to come by and harder to afford.

Here are a few additional issues that could be challenging for HPC on public clouds:

- Clouds have preloaded libraries. It is challenging to add needed custom features and new libraries.
- Workloads still need to be configured, even on HPC-supported clouds.
- Remember that standard cloud support is minimal and premium support is expensive.

- It costs nothing to move data into the cloud, but datasets grow rapidly, and data egress charges will apply if data repatriation is needed.
- AI and HPC encounter occasional difficulties on a cloud service. AI's massive datasets can run as a cloud layer but not without excessive back and forth data uploads and downloads.

We have determined that none of these problems exist with HPE GreenLake for HPC. HPE is designed to efficiently manage massive datasets, and it can provide trained resources for support and fully manage the entire HPC environment.

HPE provides expert support to handle performance tuning, capacity planning, lifecycle management, firmware updating, and patch management while monitoring critical KPIs of performance, uptime, time to resolution, and ticket status.

## HPE CLOUD WITH COLOCATION

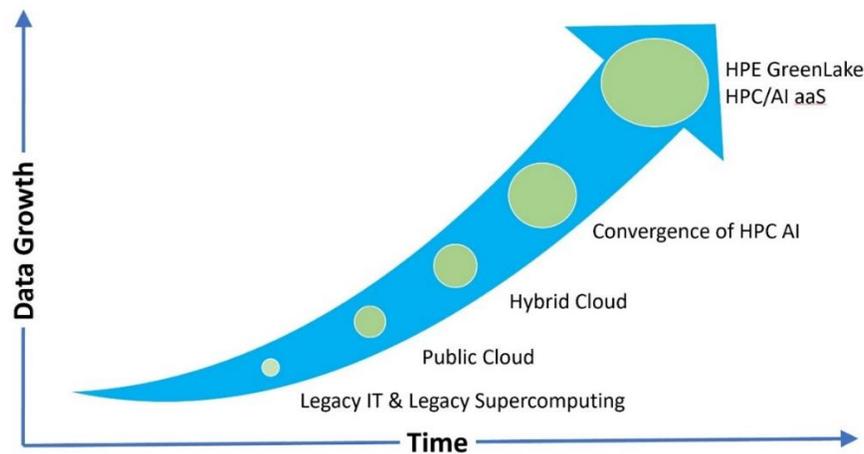
Based on a number of benefits, such as customer support, reliability, and configuration options, MI&S believes HPE GreenLake for HPC is an excellent solution that enables an enterprise to manage and support an AI/HPC environment now and well into the future.

HPE GreenLake provides an HPC solution on- or off-premises, where it can be consumed as a simple, flexible, and powerful service. It brings cloud attributes to HPC, like consumption-based economics and flexible managed infrastructure.

HPE continues to invest and innovate in HPE GreenLake to provide customers an easy-to-use platform to radically transform and modernize their organization. It recently announced new platform updates that include:

- Convergence of Aruba Central, a cloud-native, AI-powered network management solution, with the HPE GreenLake platform. Now, more than 120,000 Aruba networking customers with nearly 2 million devices to manage and 2 million API calls per day can use the HPE GreenLake platform to order services on-demand and manage their assets.
- A new, unified operational experience that provides a simplified view and access to all cloud services, spanning the entire HPE portfolio, with single sign-on access, security, compliance, elasticity, and data protection.

FIGURE 3: HPE GREENLAKE FOR HPC ENTERPRISE SOLUTION



Source: Moor Insights & Strategy

Some HPE clients look at cloud migration as a must-have solution needed to complete its digital transformation strategy. Public cloud is important to many business-centric applications which is why HPE recognized that a hybrid IT environment is the model of choice for many customers.

HPE GreenLake with colocation delivers cloud-like self-service to collocated infrastructures for a number of workloads, including compute, storage, networking, and virtual desktop infrastructure. It has consumption-based pricing for an as-a-Service solution.

HPE GreenLake also provides a unified view of usage, cost performance, and compliance across a customer's entire IT estate. HPE will manage the solution for the customer if needed.

HPE manages and supports its off-premises services such as HPC through colocation relationships with three high-quality companies:

- **Equinix** has over 220 data centers in 60+ markets on five continents and connects enterprises, networks, services, and innovators from Mumbai to Dublin, Singapore to Chicago, and Helsinki to San José. Thousands of enterprises use its interconnect infrastructure.
- **CyrusOne** is a colocation data company with more than 40 enterprise-class facilities across three continents and more than 4 million square feet of total net rentable square footage. It is headquartered in Dallas, Texas.

- **Digital Realty** is one of the largest global colocation facility providers. It operates PlatformDIGITAL, a global platform of more than 280 data centers across 24 countries. The company has a strong presence in EMEA, where it is the market leader. It provides workload-centric and industry-centric workload architectures.

## A SUMMARY OF HPE GREENLAKE FEATURES

### Fully integrated, modern HPC platform

- Flexibility to design your own HPC infrastructure solution or standardize the service with pre-sized configurations offering self-service and managed for you.
- Pre-integrated small, medium, and large platforms with storage and the accompanying software stack. This includes the cloud management platform and tools to run HPC workloads, queues, and containers.

### Cloud experience for apps and data

- Bring cloud experience to HPC applications that cannot be moved to the cloud because of security, cost, compliance, and timing constraints.
- Self-service management portal to provision HPC resources and manage workloads (including containers and orchestration), resulting in greater efficiency.
- Expand HPC usage on-demand, then scale back when requirements reduce and monitor with analytic tools.

### Expertise when needed

- HPC expertise to design the right environments for your workloads, configure and tune your custom workloads, and then deploy your platform, integrate with your existing IT ecosystem, and manage it for you.

### Better economics

- Scale and deploy AI on a high-performance platform that you can share with your other workloads and pay as a monthly service, with no up-front costs to procure the infrastructure.
- Provision a base level of always-on 'reserved' capacity. The remainder is available to access on-demand, with cloud-style pay-as-you-use pricing.

### Sustainability

- More energy-efficient CPUs/GPUs suited for data and compute-intensive workloads.

- Sustainable manufacturing and components to allow a high level of recycling.
- Cooling technology that adjusts with the number of HPC servers used and fits with existing enterprise infrastructure to avoid large up-front investment.

## CONCLUSION

HPE studies show that the HPE GreenLake platform can deliver important business outcomes, providing 75% faster HPC project deployments and 40% lower Capex for HPC solutions.

With HPE GreenLake, customers can choose an on-premises HPC solution with all the flexibility, scalability, and utility-like consumption of the cloud. Customers can increase their agility with pay-per-use pricing and pre-installed buffer capacity that is ready to provision when needs grow.

Whether demand for HPC resources spikes suddenly or grows steadily, customers are always ready to meet new needs. Customers can design their own HPC infrastructure solution within HPE GreenLake using industry-leading HPE technologies. Alternatively, customers can standardize their service with pre-sized configurations that offer self-service and are managed for them. With access to the latest technology available in the market, HPE can also buy out—and recycle—customers' existing infrastructure, helping them meet sustainability targets.

HPE continuously evolves HPE GreenLake for HPC service. Recent enhancements to HPE GreenLake for HPC service boost application performance for larger data-intensive and AI workloads including modeling, simulation, and deep learning workloads.

HPE and NVIDIA deliver an industry-leading portfolio of optimized AI solutions and consumption flexibility combined with breakthrough HPC technologies and AI innovations. For example, HPE GreenLake for High-Performance Computing includes support for NVIDIA NVLink and NVIDIA A100, A40, A30 Tensor Core GPUs in increments of 2-4-8 accelerators, ready to quote, price, and deliver.

More information:

<https://www.hpe.com/us/en/newsroom/fact-sheet/hpe-greenlake-for-high-performance-computing.html>

<https://www.hpe.com/us/en/greenlake/hpc.html>

## IMPORTANT INFORMATION ABOUT THIS PAPER

### *CONTRIBUTORS*

[Paul Smith-Goodson](#), Vice President and Principal Analyst, Quantum Computing, AI/ML/DL

[Patrick Moorhead](#), CEO, Founder and Chief Analyst

### *PUBLISHER*

[Patrick Moorhead](#), CEO, Founder and Chief Analyst at [Moor Insights & Strategy](#)

### *INQUIRIES*

[Contact us](#) if you would like to discuss this report, and Moor Insights & Strategy will respond promptly.

### *CITATIONS*

This paper can be cited by accredited press and analysts but must be cited in-context, displaying author's name, author's title, and "Moor Insights & Strategy". Non-press and non-analysts must receive prior written permission by Moor Insights & Strategy for any citations.

### *LICENSING*

This document, including any supporting materials, is owned by Moor Insights & Strategy. This publication may not be reproduced, distributed, or shared in any form without Moor Insights & Strategy's prior written permission.

### *DISCLOSURES*

HPE commissioned this paper. Moor Insights & Strategy provides research, analysis, advising, and consulting to many high-tech companies mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

### *DISCLAIMER*

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions, and typographical errors. Moor Insights & Strategy disclaims all warranties as to the accuracy, completeness, or adequacy of such information and shall have no liability for errors, omissions, or inadequacies in such information. This document consists of the opinions of Moor Insights & Strategy and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice.

Moor Insights & Strategy provides forecasts and forward-looking statements as directional indicators and not as precise predictions of future events. While our forecasts and forward-looking statements represent our current judgment on what the future holds, they are subject to risks and uncertainties that could cause actual results to differ materially. You are cautioned not to place undue reliance on these forecasts and forward-looking statements, which reflect our opinions only as of the date of publication for this document. Please keep in mind that we are not obligating ourselves to revise or publicly release the results of any revision to these forecasts and forward-looking statements in light of new information or future events.

©2022 Moor Insights & Strategy. Company and product names are used for informational purposes only and may be trademarks of their respective owners.