HPE GreenLake

Accelerate innovation at supercomputing speed

Rapidly train, tune, and deploy large-scale AI on demand with HPE GreenLake for Large Language Models

More complex models and larger data sets for both artificial intelligence (AI) and modeling and simulation (MOD/SIM) are pushing enterprises to a level of computing that — until just a few years ago — was reserved for supercomputing sites on a national level.

With compute requirements for large-scale AI models doubling every 10.7 months,[1] enterprises need access to supercomputing power to stay ahead. And they often need access very quickly, to take advantage of unexpected opportunities and get to market faster. For example:

• After a successful proof of concept (POC) for generative AI, the CEO wants to scale the use case into global production — within a week. However, the required compute capacity outstrips any available buffers in on-premises capabilities.

• Shortly before the launch of a complex product — such as a car, plane, drug, financial product, or similar — a partially failed compliance test mandates an urgent, simulation-based redesign. But there isn't enough on-premises capacity to perform the simulation in time to make the launch — and beat the competition to market.

[1] "Compute Trends Across Three Eras of Machine Learning," arXiv, March 2022.

For these and many other scenarios, enterprises can benefit from readily available supercomputing resources. But the traditional options for gaining access to supercomputing resources for large-scale AI or simulation workloads running 24x7 do not work for most enterprises. For most, traditional supercomputers simply don't fit into their data centers, and other on-demand, multi-tenant cloud solutions can pose challenges when running single jobs for strong scaling workloads requiring hundreds or thousands of processing units.

HPE GreenLake for Large Language Models enables enterprises to rapidly train, tune, and deploy large-scale AI using a sustainable supercomputing platform that combines HPE's AI software and market-leading supercomputers.

With this cloud-native offering, you can tap into supercomputing power on demand for large AI models or large-scale simulations through any web browser to accelerate time-to-market for new products and key AI initiatives. It provides nearly instant access to supercomputers that are owned and operated by HPE housed in sustainable colocation facilities, where cost-effective hydropower is abundant, and cooling is free 80% of the year.

HPE GreenLake for LLMs removes all traditional access barriers so you can get the power you need, when you need it.
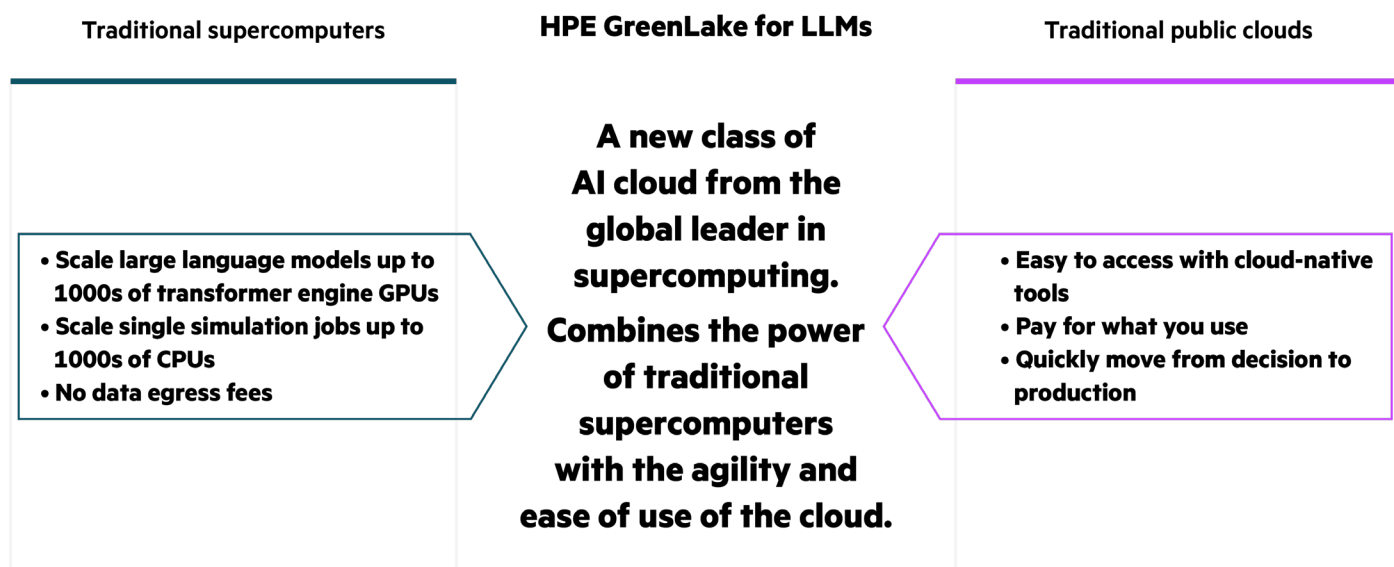
Traditional supercomputers	HPE GreenLake for LLMs	Traditional public clouds
• Scale large language models up to 1000s of transformer engine GPUs • Scale single simulation jobs up to 1000s of CPUs • No data egress fees	**A new class of AI cloud from the global leader in supercomputing.** **Combines the power of traditional supercomputers with the agility and ease of use of the cloud.**	• Easy to access with cloud-native tools • Pay for what you use • Quickly move from decision to production

Figure 1. HPE GreenLake for LLM combines the best of both worlds

Timely and cost-effective access to the latest supercomputing technologies will help you accelerate your next phase of digital transformation. And you can make the journey with a partner who has a nearly 50-year track record of delivering leading supercomputing solutions around the globe.

50+
years of experience

#1
supercomputer worldwide[2]

#1
supercomputer in the EU[3]

1st
to break the exascale barrier[4]

0
data egress fees

6 of the top 10
greenest supercomputers in the world[5, 6]

[2] "A Supercomputing Journey Inspired by Curiosity," HPE, accessed April 2023.

[3] Frontier supercomputer, powered by HPE Cray EX235a. "TOP500," November 2022.

[4] LUMI supercomputer, powered by HPE Cray EX235a. "TOP500," November 2022.

[5] "US supercomputer breaks exascale barrier," World Nuclear News, June 2022.

[6] "GREEN500 List," TOP500, June 2023.

Example workloads

Use HPE GreenLake for LLM for problems that require hundreds or thousands of processing units working in concert to solve in the shortest time possible:

- Development, training, fine-tuning, inference, and serving of large language models

- Development and training of machine learning (ML) models

- Build and market AI applications to integrate into workflows and unlock value.

- Future applications include climate modeling, drug discovery, manufacturing, and more

Be ready for what comes next with the ability to (super) power the next stage of your digital transformation with HPE GreenLake for LLMs. Getting started is easy:

- Start with a one-hour virtual and collaborative workshop, set up by your HPE enterprise account manager.

- Define the optimal solution to meet your time-critical business needs with your experts and our experts working together.

- Get access to your supercomputer — optimized for your specific challenge — in our AI cloud.

- Empower your engineers and scientists with access to supercomputing power using the tools they use today and know — supercomputing expertise is optional, but not required.

Get started today

If you need the power to break boundaries, take advantage of our multi-tenant cloud service built on HPE Cray technology, performance, engineering, and expertise and our AI portfolio.

Learn more at
HPE.com/HPE-greenlake-large-language-models

Visit **HPE GreenLake** ▭

💬 Chat now (sales)

Hewlett Packard Enterprise