HYPERION RESEARCH

EMAIL:
Info@hyperionres.com

Hyperion Research HPC Site Checkup

A HYPERION RESEARCH SPECIAL ANALYTICAL SERVICE

The Hyperion Research HPC Site Checkup aims to assist HPC/AI data center leadership successfully navigate the rapidly evolving world of advanced computing driven by a combination of expanding computational demands of traditional workloads, new options for big data and AI-centric computing, cloud-based opportunities, and a host of growing hardware and software options. This unique engagement seeks to ensure that HPC end user sites are well informed of new opportunities in technology, vendors, policies, and practices specific to their site, vertical, and region through an assessment of specific site capabilities compared with relevant counterparts, and with an emphasis on the development of a successful HPC systems and facility strategic plan.

Overview

In order to characterize site-specific strengths and weaknesses and to generate a strategic plan for going forward, Hyperion Research performs an HPC/AI site evaluation, primarily through interactions with key HPC technical and end user personnel to gather critical information including but not limited to:

- Type and scale of on-premises and cloud hardware infrastructure (CPUs, GPUs, storage, network)
- On-premises and cloud-based software complements
- Workload composition and related end use prioritization
- Incorporation of new technologies and new types of workloads like AI/LLMs
- Requirements for HPC technical staff, cloud support personnel, and SME end users
- Financing (e.g., CAPEX, OPEX) and cloud budgets
- Facilities
- Creating a sustainability agenda

This data is used to assess the relative capabilities of the HPC site under consideration with counterpart sites, highlighting areas of leading or lagging abilities as well as emphasizing unaddressed (or under addressed) opportunities. The Checkup also offers recommendations for ways to move forward, ensuring better use of existing and new technologies across the HPC/AI ecosystem with an eye towards maximizing workload capacities, minimizing operating costs, and futureproofing a strategic HPC technology roadmap.

Deliverables include a PowerPoint summary of key findings and a one-hour webinar to present and discuss the results; additional options may include a formal Word document and an on-site presentation.

Select Key Questions Addressed

- How does your site compare with other HPC/AI data centers and what changes can be made to build on current strengths and/or reduce weaknesses?
- What key technologies, processes, and/or organizational structures should be considered for your HPC/AI site or related personnel complements?
- What are some of the emerging workloads or related enabling technologies that will impact your site's overall composition in the next few years and what steps need to be taken today to address those prospects?
- What technologies, practices, or policies are ripe for sunsetting to ensure your site operates at the leading-edge within your vertical?
- What are the larger shifting trends within the HPC/AI sector that could significantly accelerate or threaten any near-term strategic plans?