

White Paper

Outcome as a Service (OaaS) for Accurate, High-fidelity Complex Modeling and Simulation Workloads

Sponsored by: Viridien (formerly CGG)

Mark Nossokoff and Jaclyn Ludema

July 2024

HYPERION RESEARCH OPINION

A primary motivator for users to turn to HPC resources in the cloud has been to provision temporary capacity for jobs that surge and exceed the available utilization of on-premises systems. While that is still a driver for consuming HPC cloud resources, an increasing number of complex, computationally challenging jobs are capable of being run entirely in the cloud, particularly with the focused approach taken by HPC-as-a-service (HPCaaS) cloud providers.

But HPCaaS only goes so far. While the term properly and appropriately defines the type and scale of services available to users, the engineering and science produced is the ultimate end goal. Prioritizing the end results, or outcomes, and assembling an HPC solution to achieve a specific outcome is leading to the emergence of a new type of service: Outcome as a Service (OaaS).

HPCaaS is one of several options that exist for users to access HPC cloud-based resources. Options are based on the strengths of a particular cloud service provider (CSP). Strengths can range from:

- Availability of the most recent CPU or GPU technologies
- Access to custom hardware, and software
- Domain-area expertise in various business and research areas
- Special tools that assist users to optimize utilization of HPC cloud resources
- Ability to quickly scale resources

HPCaaS is a more focused approach towards provisioning some combination of, or all the above, attributes. Delivery mechanisms can range from bare-metal, single-tenant configurations that users can deploy with their own preferred operating environment to fully virtualized environments that include a containerized operating environment, application code licensing, domain-area expertise and assistance to migrate existing on-premises codes to cloud-based infrastructure.

OaaS provides all the benefits of HPCaaS and aims to layer above Software-as-a-service (SaaS), Infrastructure-as-a-service (IaaS), and Platform-as-a-service (PaaS), to additionally provide an emphasis on cloud productivity (e.g., optimizing the amount of work to be done over a fixed amount of time) while also minimizing expense for the optimal amount of work. Viridien (formerly CGG) is an example of a CSP looking to bring OaaS to its HPCaaS offerings. Leveraging decades of leadership and expertise in ingesting raw data and delivering valuable insights in the geosciences sector, Viridien is now globally providing its HPC research, design, and operational know-how to other data-intensive sectors via cloud services with the same analytical challenges faced by the geosciences sector.

SITUATION OVERVIEW

HPC Cloud Market Snapshot

HPC in the cloud has been experiencing accelerating growth over the last few years. Consumption of HPC public cloud resources is projected to grow to \$14B in 2027, with several verticals embracing the use of cloud resources more warmly than others. Biosciences and CAE/Manufacturing sites are the largest verticals relative to spending, while CAE/Manufacturing, Economics/Financial, and Geosciences are expected to experience the largest cloud spending growth. See Table 1.

TABLE 1

HPC Cloud Forecast by Vertical 2021-2027

(\$M)	2021	2022	2023	2024	2025	2026	2027	CAGR 22-27
BioSciences	\$1,439	\$1,557	\$1,847	\$2,120	\$2,361	\$2,534	\$2,821	12.6%
CAE/Manufacturing	\$957	\$1,289	\$1,587	\$1,917	\$2,263	\$2,637	\$3,017	18.5%
Chemical Engineering	\$128	\$154	\$189	\$226	\$267	\$308	\$267	11.7%
DCC & Distribution	\$289	\$347	\$425	\$510	\$599	\$693	\$732	16.1%
Economics/Financial	\$315	\$398	\$512	\$627	\$752	\$889	\$1,041	21.2%
EDA	\$372	\$443	\$538	\$639	\$744	\$853	\$975	17.1%
Geosciences	\$328	\$403	\$505	\$621	\$753	\$898	\$1,069	21.6%
Mechanical Design	\$24	\$27	\$33	\$37	\$41	\$45	\$45	10.6%
Defense	\$391	\$471	\$578	\$697	\$821	\$953	\$1,091	18.3%
Government Lab	\$364	\$443	\$548	\$704	\$881	\$1,094	\$1,351	25.0%
University/Academic	\$242	\$276	\$322	\$368	\$411	\$467	\$549	14.7%
Weather	\$140	\$184	\$246	\$322	\$412	\$541	\$675	29.7%
Other	\$111	\$141	\$187	\$242	\$307	\$384	\$436	25.3%
Total	\$5,100	\$6,132	\$7,516	\$9,031	\$10,610	\$12,297	\$14,069	18.1%

Source: Hyperion Research, 2023

Additionally, recent studies indicate that approximately 60% of all HPC sites surveyed indicate using the cloud, with over 70% of respondents in the industry sector employing cloud resources. While that reflects a large percentage of sites leveraging the cloud, this is not fully reflected in the amount of

workload runtime users run in the cloud. 29% of workload runtimes were run in the cloud for those sites who indicated using the cloud, while 37% of the workload runtimes of the industry sites surveyed were run there. And there is still ample room for growth.

Drivers and Barriers to Adoption of HPC Cloud Resources

Drivers for HPC Cloud Resource Adoption

Several items are consistently cited that are driving the increased utilization of HPC cloud resources:

- **Cost-effectiveness:** Users whose workloads run infrequently without stringent completion-time requirements and without heavy data movement needs often find running workloads in the cloud more cost-effective than on-premises.
- **Extra capacity for surge workloads:** Many HPC users have access to adequate on-premises HPC infrastructure. There are times when the resources aren't immediately available for critical modeling and simulation runs on-site, while the required resources are immediately available in the cloud.
- **Access to resources not available on-premises:** In many cases, users are now keeping on-premises machines up to 5 years, causing them to miss a generation or two of the latest technology. Cloud resources allow them to access the latest innovations for critical workloads, as well as evaluate new technology before acquiring it on-site.

Barriers for HPC Cloud Resource Adoption

Users have also been consistent with items that are barriers for them to migrate their HPC workloads to the cloud. The top barriers are:

- **Cost-effectiveness:** Interestingly, cost-effectiveness is also cited as the top barrier, in addition to the top driver. While cloud utilization may look more cost-effective when running test cases in the cloud, users have found costs from running their workloads at scale in the cloud far exceed what the test case projected. In large part, costs were associated with unexpected charges for data movement within and data egress from the cloud.
- **Security:** While this item is seen to be a decreasing limitation for adoption of HPC cloud services, it remains a sizable concern for many sites. Many users view their data as critical IP to their businesses and are concerned about bad actors gaining access to it, as well as potential inadvertent access in virtualized, multi-tenant environments. A related issue is also compliance with regulations or requirements for data to remain in certain geographic areas.
- **Data locality:** Data creation is occurring in a wide variety of places, including at the edge (e.g., smart city sensors, real-time data capture during high-performance auto racing, large arrays of astronomy antennae, manufacturing line visualization quality control) and in multiple locations. Running the computing analysis, whether it be modeling/simulation or AI training/inference, near the data creation can be more cost-effective and more performant than moving data to the compute in the cloud.
- **Challenges in migrating application code to the cloud:** Many legacy on-premises HPC codes may have challenges when migrating to the cloud. Engineers, scientists, and researchers are loathe to crack open stable code merely to migrate to new technology or operating environments, regardless of potential benefits. They want to focus on their respective areas of domain expertise and not have to worry about the underlying infrastructure.

Trends in HPCaaS

Broad Range of Cloud Service Models

There are multiple types of services available to users for cloud-based HPC resources. The common element between the various types of services is a consumption-based, "pay as you go" OPEX business model. Various examples include:

- **Full service CSPs:** Cloud service providers offering a wide range of capabilities across many different markets to numerous different companies.
- **Industry-focused:** Cloud providers offering resources focused on a specific industry or vertical, such as energy or media and entertainment.
- **Capability-focused:** Cloud providers delivering resources and optimized sets of resources for several markets, such as an HPC cloud supporting multiple industry verticals.
- **Private clouds:** Infrastructure placed on-premises at a single organization but managed by the system provider and billed based on consumption OPEX, as opposed to up-front CAPEX.
- **Increased emphasis on results:** While the resources that run scientific workloads are critical in producing the output, what really matters is the output itself. Scientists, researchers and engineers ultimately don't care how their jobs get run. They want to know they can run as many jobs as they can within an optimized budget as quickly as possible and have reliable, repeatable, secure outcomes.

Each of the above can also support various service models, ranging from bare metal systems which a single-tenant user can deploy with their own preferred operating environment to fully virtualized with the CSP providing everything, including operating environment, application code licensing, and domain-area expertise and assistance.

Hybrid Cloud, Multi-cloud, and Native Cloud

Utilization of hybrid cloud can provide the best of both on-premises and cloud worlds. It affords users the capability to augment their on-premises HPC capabilities with specific resource requirements in the cloud with the lowest possible risk.

The use of multiple cloud suppliers can provide even greater flexibility. Different CSPs have different strengths, and the ability to have portable code that can run on multiple CSP platforms allows users the capability to select best-of-breed for a wide range of workloads.

Native cloud deployments are growing, particularly with newer startups and newer technologies. New businesses often have limited budgets for capital investments and the cloud provides the capability for businesses to grow until they reach a scale where it may be feasible to bring resources on-premises. Cloud providers are also increasingly able to provide both HPC operational expertise as well as domain-specific know-how in optimizing codes for the cloud.

Sustainability

Sustainability, energy efficiency, and reducing carbon footprints have become top-of-mind strategic requirements for HPC data center managers worldwide, driven in large part as a response to the current geopolitical climate and trends in HPC utilization. The rapidly increasing demand for HPC resources for AI/ML/HPDA workloads, coupled with new CPUs/GPUs/xPUs requiring substantially larger power requirements than previous iterations, are placing a growing financial strain of high energy costs on HPC sites. Additionally, other factors, such as new government regulations on energy-

efficient datacenters, and sustainable business practice priorities, are causing many HPC center decision-makers to more highly prioritize sustainability and energy efficiency in operations and procurement plans.

Emerging Technologies

As new technologies and innovations occur, users want to evaluate the technology prior to deciding what to procure on-premises. Many times, CSPs are first to market with these technologies (e.g., most recent high-end GPUs), and may also receive priority access to them in supply-constrained environments.

Still other emerging technologies, such as AI and quantum computing, require very high capital investments that make it economically unfeasible for users to bring the technologies on-premises for evaluation or code development. Vendors developing these advanced technologies may also only initially provide access via cloud-based capabilities.

Viridien's Approach to HPCaaS and OaaS

One reason cost-effectiveness is viewed as both a driver and a barrier to cloud adoption in the HPC market is the absence or presence of cloud optimization expertise. Cloud adoption with the right expertise can result in sizable cost savings, but without proper expertise it can easily result in exorbitant monthly fees. Under-optimizing provisioned cloud resources, mismanaging workload scaling efforts, and overprovisioning can quickly make the cloud OPEX model difficult to manage. Hiring an internal cloud resources management team can be expensive and is often out of reach for small to medium-scale businesses. A growing appeal of HPCaaS offerings for many users is access to cloud resource expertise with relevant experience and demonstrated success in both HPC architectures and related at-scale computational and data management challenges.

Viridien has a 70-year history of designing computing solutions at scale for seismic imaging. As a global HPC leader in the geosciences sector, it offers an array of data, products, services and solutions, including customized, optimized, and sustainable end-to-end HPCaaS solutions. While many HPCaaS providers offer technical support and domain-specific expertise, Viridien adds value to their HPCaaS offerings by committing to specific engineering and scientific end results with OaaS. Viridien bases its OaaS business model on the commitment to design and deliver cost-effective solutions that produces the customer's desired outcomes. Viridien's approach is built on decades of experience in multi-parameter optimization, positioning Viridien as a global leader in the geoscience market.

Services provided within Viridien's HPCaaS model include:

- Expertise in computationally heavy workloads
- Infrastructure hosting, currently supporting over 7 million processing job submissions per week
- Collaboration with existing staff members on algorithm optimization
- Scaling existing workloads for production cloud environments
- Design of HPC cloud solutions that meet the environmental, social and governance (ESG) goals of a business.

Viridien's OaaS model is focused on customizing solutions for industry sector customers with specific business goals they want to achieve. Viridien looks at the HPC needs of a business that is challenged with multiple parameters that need to be optimized concurrently to achieve a global optimal solution.

Their solution strategy involves optimizing infrastructure, operations, cooling, software, and code simultaneously to find the most fiscally advantageous and environmentally friendly end-to-end solution that meets business goals.

The Viridien AI and HPC labs across the globe are a crucial component to their OaaS solution strategy. The labs work in partnership with technology vendors to develop proofs of concepts (POCs) of new technologies as they function within a global HPC solution. By testing new technologies prior to implementation, Viridien can expertly recommend solutions that will maximize return on investments. Viridien also has a culture committed to routinely researching and understanding the latest technologies to maintain a strong awareness of the best options available to their customers.

The commitment to researching new technologies has resulted in Viridien becoming early adopters and experts of energy efficient and sustainable technology. For example, thirteen years ago, Viridien implemented oil immersion cooling at an industrial scale for their on-premises systems. Their newest areas of sustainability research are code development to optimize energy use, and waste heat reuse capabilities.

From Industry-focused to Capability-focused

Viridien has successfully implemented its HPC solutions strategy within the geoscience field for decades. Viridien experts' deep understanding of multi-parameter optimization and iterative evaluation of new technology is seen as a competitive advantage in the HPC seismic imaging field. Upon review of other HPC use cases outside of the geoscience sector, Viridien recognized a need for multi-parameter optimization in a variety of industry sectors. With this insight, Viridien shifted from an industry-focused to a capability-focused OaaS provider and opened its OaaS solutions to new industry sectors, providing its expertise in HPC and AI to a broader audience. An example of the success of this business strategy is the collaboration between Viridien and BioSimulytics.

Customer Success Story: BioSimulytics

BioSimulytics is a technology company supporting the pharmaceutical industry by combining AI-based neural network technology and HPC to more accurately predict new drug molecule structures and speed up drug development. BioSimulytics wanted to move their BioSim Molecule to Medicine (M2M) platform to the cloud to provide the convenience, performance, and cost benefits of cloud to their customers. Viridien worked with BioSimulytics to optimize their algorithms and scale their solution on bespoke and optimized HPC cloud resources.

BioSimulytics gained the following by benefiting from Viridien's OaaS:

- Improved ROI for their customers
- Provided optional localized data hosting
- Improved speed of application (500x their previous public cloud solution)
- Expanded their business to new customers

FUTURE OUTLOOK

New applications and use cases for cloud HPC resources have been developed consistently throughout the past several years. Key among these new applications are AI workloads, including training, inferencing, and large language models (LLMs). Businesses and research groups alike are interested in gaining the benefits offered by cloud resources, including access to the latest technology without the large upfront costs of ownership, additional capacity for surge workloads, the outsourcing of energy cost and sustainability concerns, and supplemental HPC and domain-area expertise.

For many businesses, the opportunities and advantages of moving workloads to HPC cloud resources can feel out of reach due to a lack of in-house expertise, possible disruptions and downtime required for migration, or concerns with cost-effectiveness. As CSPs continually expand their offerings, the knowledge gap between businesses and potential cloud solutions continues to widen. HPCaaS options, particularly those that offer end-to-end outcome-specific solutions and subject matter expertise, are expected to continue expanding to fill that knowledge gap.

OaaS should be considered a viable option for both new HPC users of the cloud and existing HPC cloud resource users. With customizable cloud platforms and the ability to design them from the ground up, as well as build around users' existing infrastructure, software, and business practices, Viridien is a strong partner candidate to consider for AI and high-fidelity, complex modeling and simulation workloads.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology, and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). Hyperion Research provides thought leadership and practical guidance for users, vendors, and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA

612.812.5798

www.HyperionResearch.com and www.hpcuserforum.com

Copyright Notice

Copyright 2024 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit www.HyperionResearch.com to learn more. Please contact 612.812.5798 and/or email info@hyperionres.com for information on reprints, additional copies, web rights, or quoting permission.