Innovate without limits with HPC as a Service

Introduction

How can we accelerate time to insight? That's a constant challenge for many organizations relying on High Performance Computing (HPC) clusters to process, store, and analyze vast amounts of data. The faster you get the results, the sooner you can review them, and turn that insight into action to advance your research goals. The pressure is always on, and the only way to stay ahead of the curve is to ensure that maximal time, effort, and resource can be allocated to these objectives, and the time spent maintaining and updating the vital computing capacity minimized. On top of these challenges, many organizations deal with complex and often sensitive data which can place additional demands and restrictions on the HPC infrastructure, and multiply these challenges.

This whitepaper discusses these challenges and the limitations traditional approaches to HPC can present when it comes to procuring resources, rapidly responding to changing requirements, and optimizing the computing infrastructure throughout its lifecycle.

It then outlines how new, more agile service-based models can give organizations the flexibility they need to leverage new opportunities and take their innovation journey to the next level.

The need for speed

Many organizations using HPC are operating under constant pressure to find ways to speed up research and meet the increasing demand for shorter product development cycles (PDC) to keep ahead of the competition[1]. From car and airplane design to oil field exploration, financial risk assessment, genome mapping, weather forecasting and medical research, operators in diverse industries share the same challenge of how to optimize their IT infrastructure to deliver the computing power needed to support their research goals without exceeding the budget.

Those organizations need real-time, highly reliable IT infrastructure to process, store, and analyze massive amounts of data. And when it comes to immediate and continual access to aggregated data, location is everything. HPC systems can be deployed at the edge, in the cloud, or on-premises, with all options associated with their own benefits, depending on the application. Ultimately though, the key is processing the data where it is created and not having to transfer it to another remote location for processing. Delays in accessing the information or poor performance are simply out of the question when every nanosecond counts. This becomes more prevalent when the size of the organization grows, for example, when corporations have multiple design centers around the world, all collaborating and utilizing the same set of data to make decisions.

Constraints of cloud environments

Thanks to the ease of use, deployment and simple scalability, cloud-based HPC is providing many organizations with the flexibility they need, and the market is seeing rapid growth[2]. However, this development is partly restrained by concerns over the safety and legitimate use of the data[3]. This is not surprising, since a large number of organizations have HPC applications and data that isn't suitable for the public cloud. This could be because of the sheer size of the data sets, like in the case of seismic data: uploading the data to the cloud can simply be too time-consuming to consider. There can also be security concerns regarding proprietary algorithms or data, making some sensitive workloads unsuitable for storing in the cloud. Indeed, research shows that only 11% of organizations say they don't have applications or data that can't be placed on a public cloud[4].

Still, this doesn't mean that organizations would need to miss out on the benefits associated with cloud environments. Alternative approaches exist to bridge the gap between the freedom provided by the cloud and the security and control of an on-premises solution.

Indeed, factors such as application availability and performance requirements, and data governance and sovereignty regulations mean organizations should consider on-premises technologies that can deliver the benefits of a cloud operating model while keeping the business fully in control of the infrastructure supporting these applications.

These solutions can deliver the best of both worlds: cloud-based economics with on premise resources – meaning they combine their performance and security requirements with the flexibility that comes with embracing a hybrid approach.

Towards a service-based model

Fortunately, HPC as a Service (HPCaaS) has emerged as a viable model that allows organizations to tap into the power of supercomputing through their existing resources. On-premises Lenovo TruScale for HPC, for example, enables customers to leverage their internal infrastructure through a cloud-like experience, while allowing HPC schedulers to access and provision resources as needed. This drives a faster time to answer for researchers working on solving their greatest challenges.

This is critical as on-demand resource consumption provides more efficient and scalable computing resources to meet workload requirements. Here, a service-based consumption model can deliver additional capacity to speed up innovation as workloads evolve. Many leading providers offer tailored rack-to-cloud level solutions to cater for the needs of multiple sectors, including academia, pharma, manufacturing, and healthcare.

Put simply, HPCaaS can expedite and facilitate access to HPC on-premises resources. It maximizes HPC cluster resources and controls shared access through a simple, pay-per use model. With technology that is installed at the customer data center, HPC workload managers can dynamically access and provision resources on-demand. Some solutions, such as Lenovo TruScale for HPC, further enhance access through additional cluster resources that can be obtained through 'push-button' access, which allows businesses to utilize additional computing, storage and acceleration capabilities as needed.

This way, organizations can seamlessly grow their HPC environments as and when new workloads, grants, or research opportunities come in. This approach helps secure in-flight project resource provisioning as well as emerging opportunities that require additional resources in a fast delivery mode.

Ensuring data security and sovereignty

Importantly, HPCaaS also ensures that organizations can adhere to any security, legal or strict data sovereignty requirements to which they are bound. And there's no complexity relating to spinning up instances to the cloud or doubt as to the location of data.

The on-premises HPC capability provides seamless, protected access to data without any changes to job scripts, or potential risks associated with public cloud exposure.

Additionally, with Lenovo TruScale for HPC, dedicated professionals ensure that data is secure from installation to utilization and through decommissioning. These capabilities are delivered in a cost-effective offering, without any data egress or ingress charges associated with use.

Financial flexibility with HPCaaS

The global HPCaaS market size is projected to reach $17 billion by 2026, a growth of 13.3% from 2019 to 2026 . This is because HPCaaS offers ease of deployment, scalability, and predictability of costs, using an established service.

Unsurprisingly, research shows that 48% of IT leaders say their organization would prefer to purchase on-premises data center infrastructure through a consumption-based model[6].

The financial flexibility the aaS approach offers is a key factor in driving the adoption. Instead of having to secure millions of dollars in upfront capital expenditure (CAPEX) and undertake the often-long procurement approval process, organizations can

benefit from immediate access to resources via a much lower ongoing monthly fee. Moreover, their hardware is no longer a depreciating asset on their books, as the risk is absorbed by the technology provider. Similarly, they don't have to worry about the disposal of those assets. And should the time come when more compute is required, the racks are automatically refreshed so organizations always benefit from the latest hardware that can contribute to better productivity improvements, increase competitive advantage, and get their products to market faster.

This means HPCaaS can enable a greater number of companies of all sizes to tap into the power of supercomputing. Put this together and the HPCaaS model becomes an increasingly attractive proposition for organizations.

With HPCaaS, organizations can:

- Pay for what they consume with no hidden costs

- Easily scale HPC clusters by removing resource limitations

- Run applications at scale without architectural bottlenecks

- Translate capital investments into operational expenses, avoiding long and intensive budget cycles and procurement delays

- Easily manage budgets and visualize consumption and billing

- Access specialized hardware

- Obtain greater protection and security

Lenovo

Dedicated support on standby

HPC teams cannot afford performance problems, delays, or downtime. Support, therefore, must be proactive, carried out by technical specialists that work closely with the customer and have a deep understanding of their environment.

With Lenovo, organizations have an HPC technical account manager or system admin as part of their HPCaaS contract as the single point of contact. Whether on-site, working remotely or a mixture of both, the support professionals can quickly pinpoint and resolve any issues, and ensure the HPC environment remains running optimally 24/7.

This involves the use of unique solutions, such as Lenovo Neptune direct-water cooling technology, to optimize operation. Neptune allows the support teams to tackle the problem of traditional data center cooling in which every component inside the system produces heat, which can cause shutdowns, slower performance, and data loss. Neptune enables high performance components to run at lower temperatures, boosting performance without driving up energy costs.

However, Lenovo goes way beyond specialized technical support. TruScale's end-to-end service for HPC includes initial consultation, analysis and configuring the right environment, through to ongoing water-cooling assessment and maintenance services to billing and administration. Lenovo supports the organization throughout their entire HPC journey.

This includes helping customers focus on sustainability in their data center. Lenovo can help them reduce their carbon footprint, reusing the heat and energy from the compute-intensive equipment to power air conditioning units, lights or in some cases even homes.

Lenovo's holistic approach to energy efficiency encompasses hardware, software, and infrastructure, allowing organizations to save up to 40% energy compared to a traditional air-cooled data center[7].

Looking to the future with Lenovo TruScale for HPC

Lenovo TruScale for HPC is always evolving and Lenovo is working to fulfil the growing demand for innovation in the HPC environment. Looking to the future, this will mean expanding the TruScale family of offerings to provide even greater flexibility and transparency for customers who want to consume HPC in the way that best suits them.

As the global leader in High Performance Computing, Lenovo is working closely with key industry partners to develop, integrate, and deploy the technologies of exascale-level computing to organizations of all sizes.

Lenovo
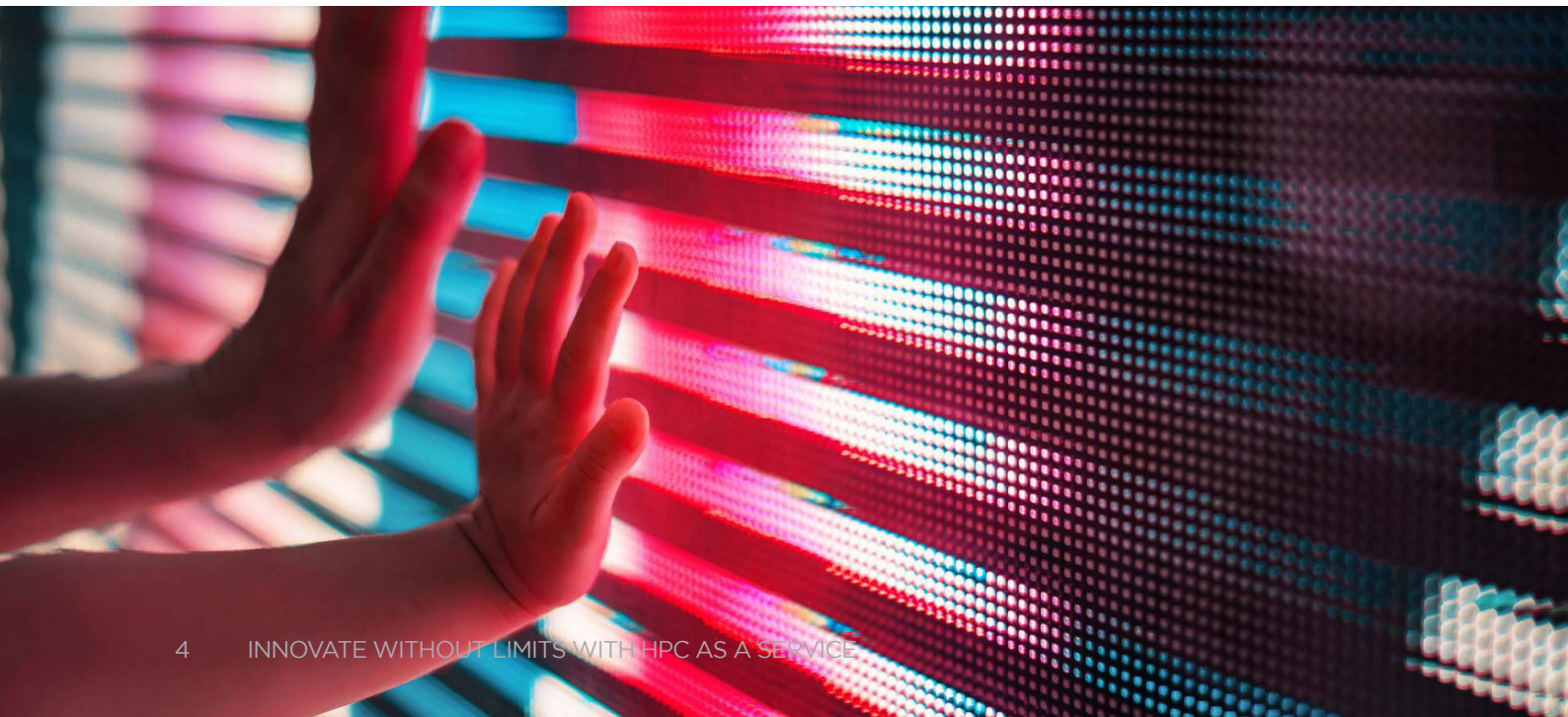
Optimizing HPC resources

HPCaaS helps organizations maximize their return on investment in HPC and accelerate time to value. It also gives the flexibility to add provisioned resources rapidly, without being hindered by procurement delays or supply chain disruptions – something which has never been more important.

Moreover, HPCaaS makes consuming HPC more transparent and affordable compared to traditional solutions. As every organization has unique needs and requirements, working with a leading provider will help achieve a HPC solution that is tailored to support the evolving needs of the workload, workflow, and workforce.

- Global HPC revenue is expected to increase from \$42 billion in 2021 to \$71.78 billion in 2030[8].

- Lenovo is the number-one Supercomputer provider in the world according to Top500.org

- Award winning HPC solutions including ISC Vendor Showdown and multiple HPCWire Awards

- #1 in reliability for x86 and a leader in performance benchmarks

lenovo.com

[1]   https://www.mordorintelligence.com/industry-reports/high-performance-computing-market
[2]   https://www.mordorintelligence.com/industry-reports/cloud-high-performance-computing-hpc-market
[3]   https://www.marketsandmarkets.com/Market-Reports/Quantum-High-Performance-Computing-Market-631.html#:~:text=Driver%3A%20Increasing%20de-mand%20for%20high,6.7%25%20between%202022%20and%202027
[4]   https://www.intersect360.com/report/hpc-technology-survey-2021-cloud-computing/
[5]   https://www.alliedmarketresearch.com/high-performance-computing-as-a-service-market#:~:text=High%2DPerformance%20Computing%20(HPC)%20as%20a%20Service%20Market%20Statistics,13.3%25%20from%202019%20to%202026
[6]   https://research.esg-global.com/reportaction/DataCenterInfrastructureSpendingBrief/Marketing
[7]   https://www.lenovo.com/gb/en/data-center/solutions/hpc?orgRef=https%253A%252F%252Fwww.google.com%252F
[8]   https://www.prnewswire.co.uk/news-releases/high-performance-computing-market-size-to-reach-usd-71-78-billion-in-2030-rising-prominence-of-cloud-computing-along-with-digitalization-is-one-of-the-key-factors-driving-hpc-market-revenue-growth-says-emergen-research-859281191.html