RED OAK CONSULTING REPORT:

Incorporating the Cloud into the HPC Mix

Red Oak Consulting
Expert advice, exceptional delivery

ROMS MANAGED SERVICES

Contents

Introduction

In the ever-evolving landscape of technology, two remarkable innovations have significantly transformed the way computational tasks are performed and managed: High Performance Computing (HPC) and Cloud Computing. HPC refers to the use of clusters of computers (supercomputers) to solve large scale, complex problems that require significant computational heft, while public Cloud Computing offers on-demand access to a pool of shared computing resources over the internet. The intersection of these two technologies has opened new possibilities revolutionising industries and research fields alike. The ability extends to either having on-premise in isolation, or the use of Cloud be it private or public or a hybrid of both.

What is High Performance Computing?

Over the span of recent computing history, high-performance computing has consistently shown its utility to accurately model and predict a wide range of physical properties and phenomena. Many of these models have had an incalculable impact on aspects of our world, contributing to economic growth and improving the quality of our lives. However, the demand for large-scale problem-solving has driven the development of supercomputers. The term originated during the 1960s when Seymour Cray, an electrical engineer, embarked on a mission to develop the fastest computer globally.

Today HPC refers to the utilisation of parallel processing techniques to perform computational tasks, sometimes at an extraordinary scale and speed using clusters of servers (also called nodes) optimised for numerical computations and connected by fast and highly efficient networks. These tasks involve complex computational simulations of physical processes, data analysis, visualisation, and other operations that cannot be effectively undertaken by individual systems. HPC servers consist of multiple processors working in tandem, on large amounts of data concurrently and connected with other systems enabling them to work cooperatively on the same problem.

At the heart of every HPC cluster lies the scheduler, a tool that orchestrates available resources, optimally allocating jobs to different resources (CPUs and GPUs).

These HPC solutions are deployed in self-managed datacentres (often on-premise), close to edge networks (where much of modern data is collected), or even within Cloud environments. HPC finds applications in diverse domains including weather forecasting, molecular modelling, financial simulations, and elementary particle research. These substantial and rapidly growing computational demands have presented an increasing challenge for organisations and researchers, in terms of maintaining infrastructure, managing costs, and ensuring accessibility.

What constitutes a supercomputer?

A supercomputer is an amalgamation of hundreds to thousands of compute nodes working collaboratively on tasks. While the term supercomputer used to mean singular, fast machines, today's 'supercomputers' are built using clusters of servers equipped with the highest performance central processing units (CPUs) manufactured.

By leveraging HPC, data- and compute, intensive problems can be tackled with greater scientific fidelity, often in shorter timeframes, radically improving user productivity.

Increased productivity

made possible

The Applications of HPC

In an era marked by rapid development and deployment of new technologies, organisations are generating immense volumes of data that require rapid, real or near-real-time processing to produce actionable insights. HPC's capabilities are beneficial across a wide spectrum of problems and industries including science, healthcare, academia, genomics, financial services and engineering. Additionally, supercomputing is powering the advances in AI and deep learning frameworks which are rapidly coming to the market.

HPC has become an overloaded term. So, when people talk about HPC, they can mean two quite different things. First, there's what is referred to as classical high-performance computing, where all or significant fractions of the processors in a system are working on the same problem which has been distributed across the cluster.

Then there's high throughput computing (HTC), where many entirely independent calculations are run on the same system at the same time. A classic example of HTC would be Monte Carlo or Black Scholes method for finance, where effectively millions of simulations with slightly different inputs are performed to achieve a probabilistic result.

Traditional HPC Paradigm

Traditionally, HPC has predominantly operated as an on-premise resource, employing symmetric multiprocessing (SMP) machines in clusters of Linux-based servers to deliver the necessary computational muscle for demanding scale-out workloads. This approach stems from the high demands that HPC workloads place on hardware and supporting infrastructure. Thus, many organisations opt to retain local HPC facilities.

Certain applications might require low-latency networks access to high-performance storage when operating at significant scale, 10 – 100s of nodes, or other HPC technologies that could impede application performance in the Cloud. It's not just the bandwidth but the speed of the slowest message being exchanged with another machine that is important. Factors such as data volumes, software licensing constraints, and other challenges might curtail the feasibility of transitioning HPC to the Cloud.

Hosting HPC

Leading Cloud providers wield colossal scale and capabilities, eclipsing those of conventional HPC centres (even the very largest). They are now also routinely able to deploy the latest technology in flexible configurations immediately with general availability of a new processor, contrasting the long lead times and procurement processes tied to acquiring new hardware for everyone else.

In many respects, the future of HPC is synonymous with the Cloud. Cloud providers offer comprehensive services, enhanced resilience, and unmatched scalability. However, privacy concerns might warrant on-premise solutions. In such scenarios, two avenues emerge: Private Cloud and Hybrid Cloud.

Selecting the appropriate hosting solution holds the potential to enhance robustness and dependability, curtail expenses via scalability advantages, elevate connectivity, reinforce security, and optimise efficiency. Furthermore, it contributes to resource preservation, aligning with the prevailing sustainability objectives embraced by a majority of establishments nowadays.

There exists a continuum of options for a hosting solution for research computing systems and services. The main difference between self-hosting, colocation and managed services in the context of datacentre provision is the amount of control you have over the server and surrounding infrastructure.

What are the options?

On-Premise

With an on-premise infrastructure, organisations have full control of their HPC operation and retain the responsibility of maintenance for both the server and datacentre environment. The HPC operation will be run from a location owned by the organisation. This option can be referred to as running HPC as a fixed asset or fixed resource. All costs, maintenance and human resources needed to administer the system and support users where they lie within the organisation.

Co-location

With co-located hosting, users own the server, but it is operated in a data centre owned by a third-party provider. Organisations are responsible for maintaining the server while the provider is responsible for providing power, cooling and external connectivity. Contracts are usual fixed term for committed capacity.

With managed service hosting, the options range from co-location plus a remote hands capability all the way through to combinations around renting a server or indeed a cluster from a provider who is responsible for physically maintaining it. They provide power, cooling and internet connectivity as well as manage the hardware and potentially the software on an organisation's behalf (very similar to a public Cloud arrangement). Costs are usually fixed term with contractual commitments around capacity and additional marginal costs to provide the service wrapper.

The Cloud

With Cloud, the environment is fully virtualised with a choice of services. These services range from the infrastructure (which is conceptually closest to a managed hosting environment), through to Platform and Software. The management of this Cloud service can be undertaken by internal staff or outsourced to a managed service partner.

One of the key advantages of the Cloud, is that within the pay-as-you-go system, organisations only need to apportion budgets a year at a time, enabling them to invest elsewhere with the allocated operational HPC budget, where a guaranteed return on investment can be realised.

With Cloud, a guaranteed return on investment can be realised

Hybrid

The hybrid route is often one which organisations take before fully moving into the Cloud. A hybrid environment facilitates exception workloads, such as when Covid vaccination research had to be conducted in a matter of months, rather than years. It allows for organisations to utilise the Cloud when there's critical high demand, rather than ringfencing the entire on-premise HPC, which impacts or halts other projects.

The allure of Hybrid Cloud could be seen as a temporary stance before migrating to full-fledged Cloud solutions. After all, going hybrid could be seen as a business's safety net. Having the capability to utilise the Cloud, whilst still have the security of the known on-premise service. Access to Cloud resources is an insurance policy against the rapid evolution of AI and related services which cannot be met by current resources.

HPC in the Cloud (what's so different about it)

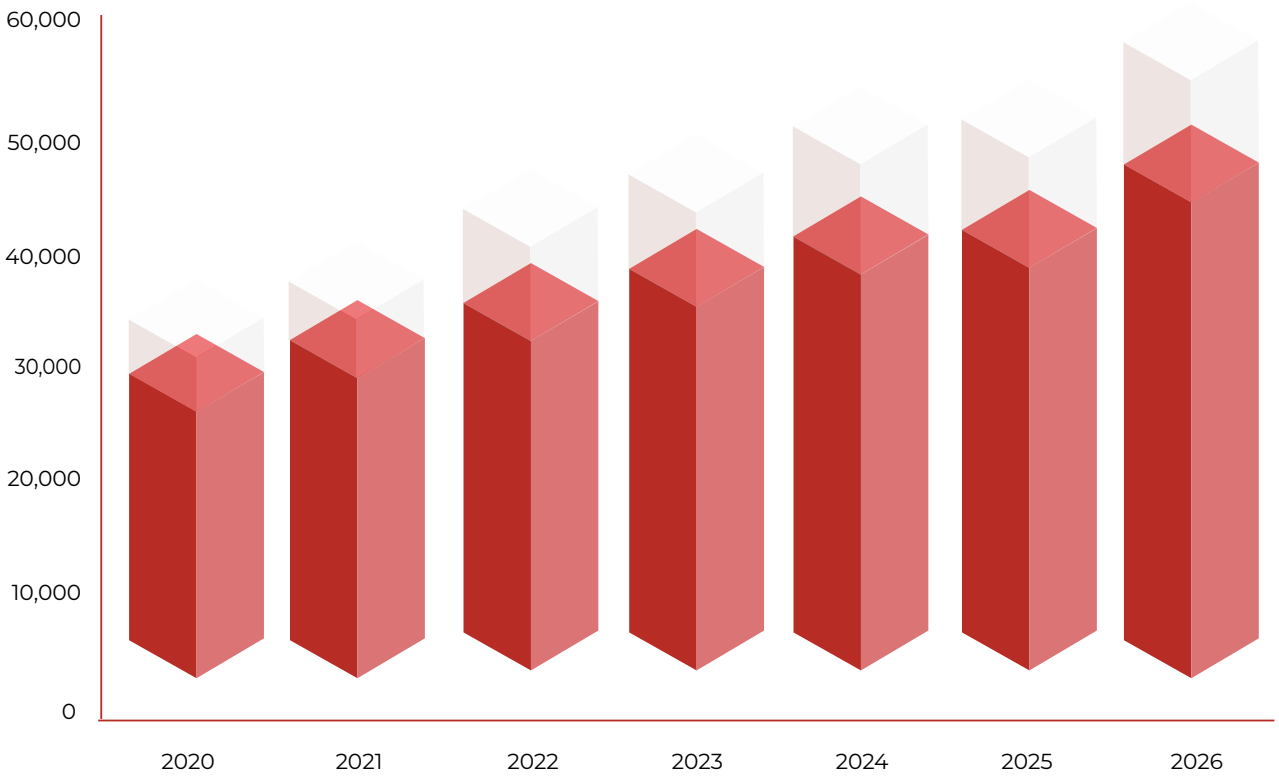
Numerous major Cloud providers now offer specialised HPC services and capabilities, enabling the successful construction and deployment of intricate HPC systems within the public Cloud. By harnessing a diverse range of compute instance types, high memory-to-core ratios, accelerators, GPUs, and access to high-bandwidth, low-latency network connections and storage, organisations can capitalise on the public Cloud's potential for HPC workloads.

Adoption of HPC in the Cloud is radically changing how workloads are managed

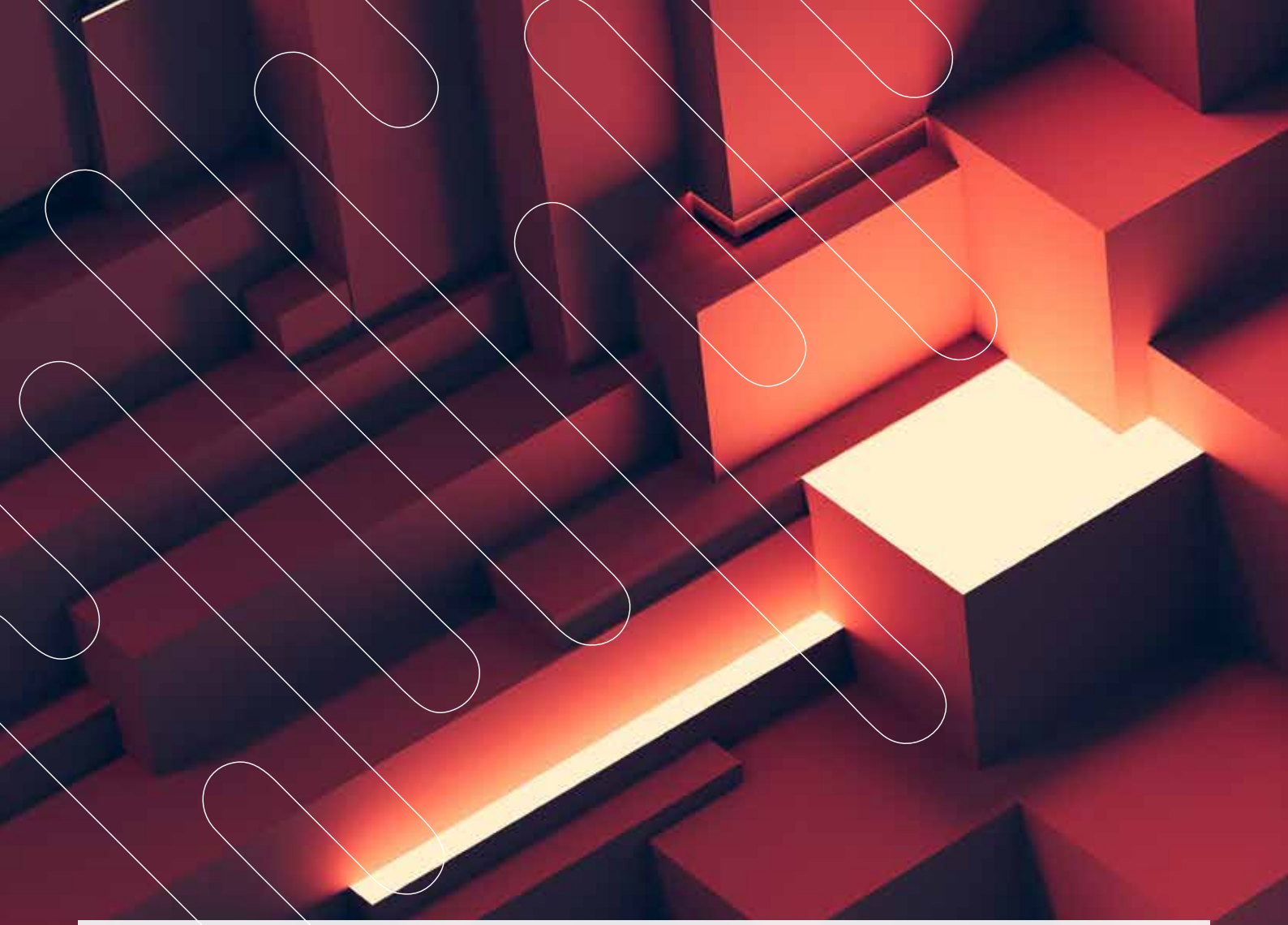
Indeed, adoption of HPC in the Cloud is radically changing how workloads are managed, breaking free from fixed infrastructure and location constraints, providing on-demand resources whilst managing costs and offering more flexibility for running HPC tasks.

A hybrid environment for HPC refers to integrating distinct computing environments, such as a on-premise cluster / private Cloud and a public Clouds capability. A hybrid environment means you can balance the use of capital-intensive resources with Cloud-based resources to service exception needs (sometimes called Cloud bursting). However facilitating uniform application deployment and seamless information sharing is challenging as is understanding where the most economic place to run a workload.

Container technology is a way to encapsulate complex programs with their dependencies in one environment thus making it more portable. This addresses one of the key requirements for compute mobility as well as delivering flexible solutions that reduce administrative overheads and which can ultimately deliver cost efficiencies. By employing containers to deploy HPC applications and workloads in the Cloud, users are unshackled from specific HPC systems or Cloud providers.



Total HPC Market: On-Prem and Cloud



Understanding the value of HPC to your organisation and how Cloud HPC might fit

Traditional HPC infrastructures are expensive to build, operate, and maintain. They are also effectively a fixed resource, exhibiting an almost boom and bust cycle as far as the resources available to users are concerned. Careful management of the available capacity, via proactive capacity management and use of the schedulers ability to prioritise jobs, can mitigate the fixed nature of the resource. This ensures that high value, high importance work is prioritised over less urgent workloads, but it cannot entirely alleviate the issues that arise from an over contended fixed resource.

How Cloud addresses the value in HPC

Cloud-based HPC solutions allows a business to make value-based decisions. It allows businesses to avoid these high upfront costs by buying access to resources from Cloud providers on a pay-as-you-go basis. Cloud providers such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) have built massive data centres worldwide to host their own SaaS and PaaS requirements whilst also servicing their IaaS customers.

Scalability for all

This democratises access to HPC capacity, making it accessible to smaller organisations and research groups when they need it most. Cloud computing provides a solution to the scalability challenge in HPC as well as the need for flexibility and agility. Researchers and organisations can quickly scale up their computational resources during peak demand and scale down during periods of lower usage. This elasticity ensures that resources are utilised optimally, reducing costs and enhancing efficiency generating results faster and insights quicker.

That said, capacity in the Cloud while it may seem effectively limitless is not. Careful planning can also be needed to get the most out of running HPC workloads in the Cloud. A hybrid approach addresses the limitations of both the more traditional on-premise and the pure Cloud approach.

This approach offers flexibility while maintaining data sovereignty suitable for industries with stringent security and compliance requirements.

Challenges - are there any?

In the present business landscape, the vast majority of enterprises and organisations have integrated their enterprise IT operations into the public Cloud to some degree. Recent surveys indicate that over 90% of businesses have made this choice, solidifying the public Cloud's pivotal role in their IT infrastructure and the delivery of technology and services to their users and customers.

Indication shows over 90%

of businesses use an element of Cloud

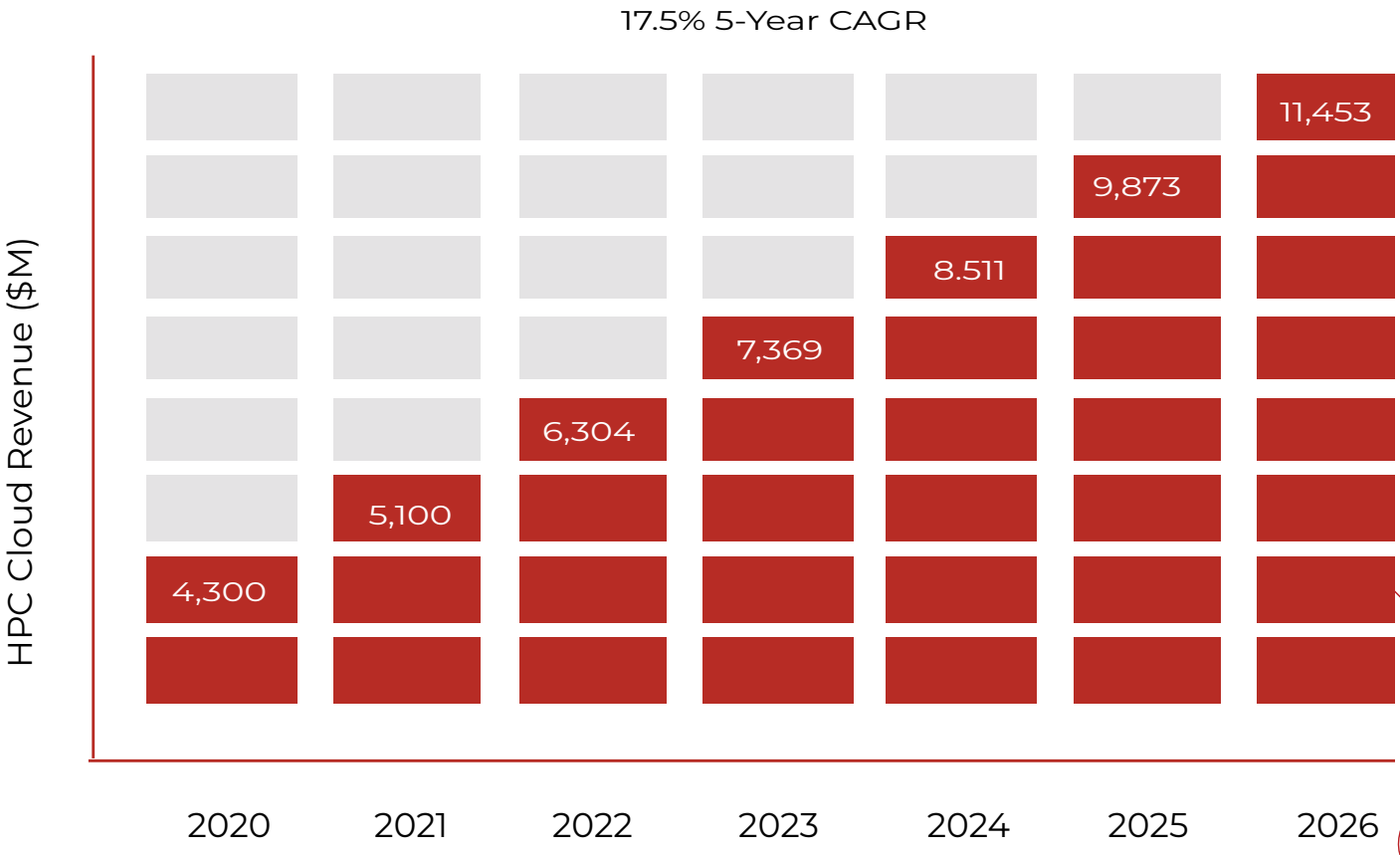
While bringing HPC to the Cloud brings some significant advantages, challenges remain. Data transfer between on-premise systems and the Cloud can introduce significant latency, and once it is in the Cloud, the principles of data gravity and inertia tend to take over. Cloud also provides a diverse set of options for storage which because of the varying costs involved may necessitate a change in how workflows are run for optimal cost. Not all Cloud solutions are ideally suited to supporting the largest scale of HPC for capability problems and subtle differences in the performance of parallel applications in virtualised environments is areas that requires careful consideration.

Looking ahead, advancements in network technologies, specialised processors, and containerisation as well as enhanced price competition will likely address many of these challenges. The growth in HPC and Cloud computing will further, reduce barriers to adoption and improve access to high-performance resources, empowering more industries to leverage advanced computing capabilities.

Reduce barriers to adoption

and improve access to high performance resources

Nevertheless, for organisations using HPC applications the decision of where to run their HPC workloads is largely dependent upon scale. The smaller the scale, the overhead in maintaining and operating your own self-managed HPC systems and services is proportionally larger. This choice brings forth a host of questions and challenges that demand resolution before making the shift from local HPC deployment within an internal data centre to embracing the public Cloud.



HPC Cloud Forecast

Benefits of Cloud-based HPC

Migrating HPC workloads to the Cloud introduces several advantages for organisations grappling with dynamic user community needs. This approach allows organisations to reap the following benefits:

- Elasticity for bursting HPC workloads into the Cloud
- Enhanced capacity and capability for HPC workloads
- Flexibility to tailor hardware resources to individual HPC jobs
- Testing and benchmarking of new hardware in isolation
- Economical deployment and time savings on expensive ISV software licenses (usually in the ISVs own Cloud based service)
- Easy and cost-effective archiving of data
- Provision of performance (eg parallel) file systems as a service, when required

Considerations Before Moving to Cloud-based HPC

Nonetheless, this approach isn't without its challenges. Organisations must grapple with a range of concerns, such as:

- Huge range of VMs can make delivering optimal performance difficult in Cloud environments
- Persistence of data in the Cloud due to data gravity
- Incurred data egress costs when retrieving from the Cloud
- Considerations of ISV license conditions and network access
- Establishment of robust networking security for Cloud operations – the threat levels within private Cloud environments tend to be lower than those in public
- Business rules needed, if maintaining a hybrid-environment, to steer jobs to the most appropriate service.
- Exploration of hybrid HPC models for optimised outcomes

Weighing the Pros and Cons

The decision to embrace Cloud-based HPC hinges on a nuanced evaluation of its advantages and drawbacks. Frequently, a comprehensive shift of all HPC workloads to the Cloud might not be the most appealing course, due to factors like existing hardware investments and the need to gain familiarity with HPC in the Cloud before fully committing.

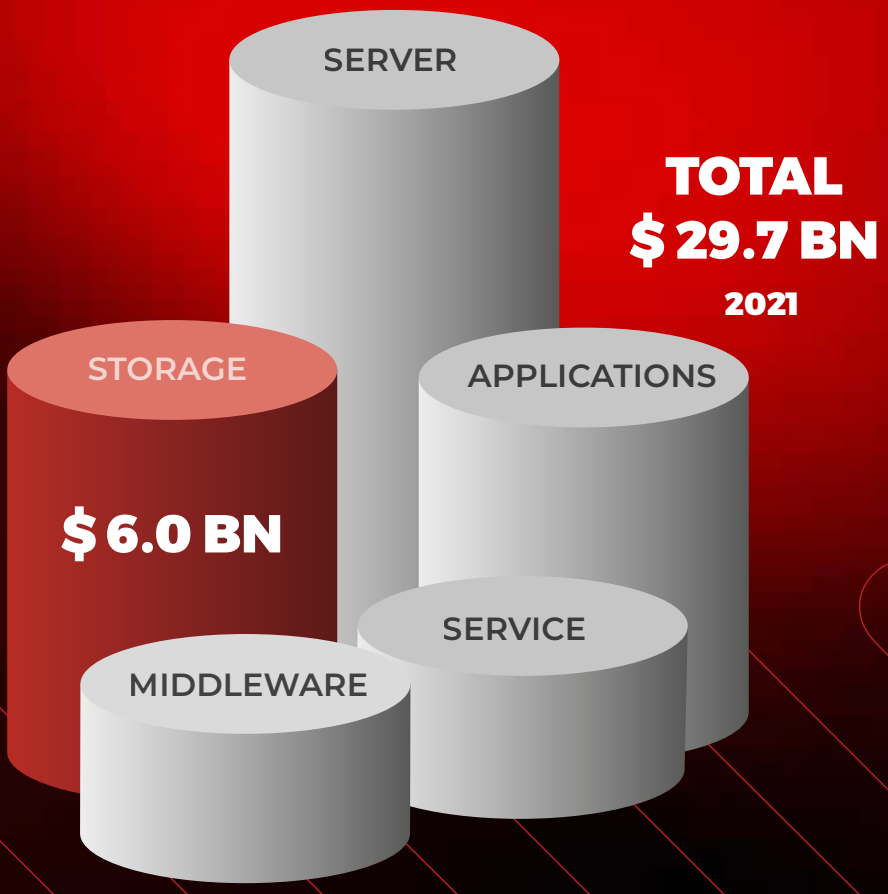
A practical approach often involves a “Hybrid HPC” strategy—moving select HPC workloads to the Cloud that are most suited for Cloud-based resources, while retaining others within the local on-premise HPC cluster. The preceding sections have delved into critical factors such as data management, data gravity, egress costs (moving data to the Cloud), networking, and ISV licenses.

These considerations must guide the strategic allocation of workloads between the Cloud and on-premise settings. Subsequent sections underscore specific aspects that hold relevance within the realm of Hybrid HPC.

While extending an existing on-premise HPC cluster into the Cloud may seem enticing, this approach introduces complexities. Moreover, integrating HPC jobs and data across these environments poses challenges due to varying network bandwidths. Thus, a dedicated Cloud-based HPC cluster often emerges as a more viable solution.

Local data centres usually rely on shared file systems that facilitate streamlined data access for HPC jobs. Cloud-based HPC environments often have constrained network bandwidths, necessitating innovative approaches. A two-step approach involving caching major Cloud-based data and delivering input files for individual jobs directly to the Cloud is often pragmatic - known as data staging or pre-staging; making sure the data is in the right location before the job starts.

Whether an HPC system resides on-premise or in the Cloud, user authentication must remain consistent. Achieving this means establishing an authentication environment in the Cloud and synchronising it with the on-premise authentication service, ensuring users can access both settings with unified credentials.



HPC STORAGE GROWTH

The Benefits of Cloud Computing: What about Cloud makes it ‘good’?

In today’s fast-paced digital landscape, Cloud computing has emerged as a transformative force, reshaping how businesses and individuals interact with technology. Cloud computing offers a multitude of benefits that revolutionise traditional IT operations, enabling greater flexibility, efficiency, and strategic value. As organisations navigate an increasingly complex technological terrain, understanding these advantages is crucial for harnessing the full potential of Cloud computing.

The hallmark of Cloud computing is its flexibility, empowering users to tailor their digital experiences to their exact needs. The ability to scale services seamlessly allows organisations to expand or contract resources in response to fluctuating workloads. Whether handling a sudden surge in demand or downsizing during quieter periods, Cloud infrastructure adapts with remarkable agility.

Moreover, the spectrum of storage options provided by Cloud computing is diverse, catering to a range of security and operational requirements and cost sensitivities. Enterprises can opt for public, private, or hybrid storage offerings, granting them the autonomy to safeguard sensitive data while maximising efficiency.

Cloud computing also introduces a spectrum of control choices through as-a-service models such as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). This empowers organisations to align their level of control with their strategic goals. The availability of prebuilt tools and features offers a versatile toolkit for crafting tailor-made solutions that meet specific business demands.

With emphasis on security, public Cloud computing integrates advanced features such as encryption mechanisms, and API keys to fortify data protection. This holistic approach to flexibility lays the foundation for a dynamic and secure digital ecosystem.

Developing and deploying applications in the Cloud can drastically reduce development cycles, enabling organisations to swiftly bring their offerings to market. Furthermore, the inherent resiliency and reliability of Cloud infrastructure, characterised by networked backups whilst delivering a robust, disaster recovery, helps to mitigate the risk of data loss due to hardware failures.

By utilising remote resources, organisations circumvent the expense of investing in physical servers and equipment. The utility pay structure, akin to paying for the resources consumed, allows for efficient resource allocation and minimises wastage.

In the competitive landscape of modern business, strategic value is a defining factor for sustainable growth. Cloud services provide organisations with a transformative competitive advantage by delivering the latest and most innovative technologies. By entrusting Cloud service providers (CSPs) with the management of underlying infrastructure, enterprises can streamline their operations and allocate resources towards high-value activities, such as application development and strategic planning.

Regular updates and enhancements from service providers ensure that organisations remain at the forefront of technological progress. This perpetually refreshed technology stack is instrumental in maintaining operational excellence and competitive prowess.

Most significantly, Cloud computing provides organisations with a distinct competitive edge. Organisations can pivot more swiftly than competitors shackled by the burdens of managing intricate infrastructure. This freedom to innovate, experiment, and adapt, positions enterprises to exploit emerging opportunities with real flexibility.

The ROMS Story

To meet the rapid increase in demand for computer processing power of high-end research, we have introduced a three-tiered, fully managed service solution to help organisations move their research systems into the realm of Cloud-based supercomputers.

Designed to deliver complete peace of mind to any organisation requiring the seamless reliable functionality of a HPC service. From the transition of the HPC solution through to incident logging, cost and usage analysis, Red Oak Managed Services (ROMS) is tailored to a business's requirements, especially when critical issues require speed, and deep-level expertise.

Within the bespoke packages – Silver, Gold and Platinum – ROMS offers three distinct areas of support, namely HPC system admin support, business-oriented reporting and cost & usage analysis. All of which are designed to help organisations running multiple HPC workflows greater knowledge of their systems, and a guarantee to get the best productivity, flexibility and value for users and the business.

ROMS MANAGED SERVICES

Support is provided via a team of highly skilled HPC specialists with backgrounds in both academia and the corporate sector as users, engineers, solutions architects or software support agents.

Speaking of his experience with ROMS, Nigel Cowler, digital Cloud platform owner for Lloyds Register, said:

> **“I highly recommend ROMS for their exceptional HPC support services. From the moment we engaged with them, their professionalism and expertise were evident.**
>
> **“They demonstrated a deep understanding of our needs and provided tailored solutions that perfectly aligned with our requirements. Their level of engagement and commitment to delivering high-quality results matched our expectations.**
>
> **“Red Oak Consultancy has proven to be a reliable and knowledgeable partner in the realm of HPC, and we continue to benefit from their valuable support.”**

Once you’ve successfully migrated your HPC function to the Cloud, what comes next? This pioneering move marks just the initial phase of your journey with HPC in the Cloud. As with any technology, maintaining functionality remains crucial. It’s vital to apply the appropriate level of oversight, expertise, and skill to ensure the uninterrupted operation of a dependable HPC solution.

As the foremost authorities in HPC and Cloud Computing in the UK, our team has played a pivotal role in driving success for numerous businesses. We excel at navigating the vast spectrum of HPC. Our expert guidance and top-tier implementation continue to serve as the cornerstone of achievement for a multitude of companies and institutions, both domestically and internationally.

Now, we are taking our expertise to the next level. Building upon our knowledge and experience, we’re venturing further with an ongoing HPC service support. Why leave anything to chance? At Red Oak Consulting, we are renowned for our skill, professionalism, integrity, and honesty. Our comprehensive understanding and expertise enable us to ensure a consistent level of managed support of exceptional quality for all our clients.

ROMS MANAGED SERVICES

Conclusion

Every successful research organisation needs a scalable infrastructure that can support data transformation, deliver fast and secure file transfer, and end-to-end visibility of all the data that flows through its ecosystem. Universities and research bodies are no exception to this, although their security requirements may be greater.

The advent of Cloud solutions is an inevitable development as the spread of computer and internet technologies shapes society, including the diversification of the information sources, the ability of users to access information from anywhere, and the shifting of the communication environment from traditional fixed locations to the virtual world. In the information age we live in, universities are the most important determinants of social development. The most important things a university should do are education and research, and the role of the Cloud in this will continue to be a fundamental and increasingly central to this. Having the right partners and expertise in place will ensure the transition is secure, resilient and permanent.

info@redoakconsulting.co.uk
sales@redoakconsulting.co.uk

36 Pure Offices
Cheltenham Office Park
Hatherley Lane, Cheltenham GL51 6SH
United Kingdom

01242 806 188
01420 565 682

www.redoakconsulting.com

Follow us on Twitter, Instagram, Facebook

Red Oak Consulting
Expert advice, exceptional delivery

ROMS MANAGED SERVICES