

# Enabling AI: High-Performance Compute



# Building Secure, Sovereign HPC Infrastructure for Enterprise AI

Deploying enterprise AI is a key priority for security and sovereignty conscious CIOs and CTOs. With the data and throughput-hungry workloads of AI, high-performance computing (HPC) environments are one of the essential infrastructure elements that must be designed and implemented.

HPC is now a key enabler for enterprise AI.

Industries from defence and universities to financial services and healthcare are deploying AI to improve efficiency and transform their use of technology. Many HPC workloads are designed primarily for CPUs and are not optimized for GPUs (except in AI-related tasks). Moreover, not all AI workloads necessarily require a GPU.

The question enterprises now face is, how to best architect and deploy HPC for accelerated AI workloads?



# The Case for On-Prem: Optimal HPC Without Compromise

Cloud has dominated enterprise IT solutions for traditional business workloads due to its flexibility and scalability. But with HPC, cloud is not the default. HPC environments demand GPUs, ultra-fast fabric and high-throughput storage, which on-prem is better placed to deliver.

## Economics

Data egress fees make cloud solutions uneconomic for many organisations scaling HPC. Whether you're deploying HPC for the first time or refreshing your technology, cost is a deciding factor.

On-prem is more cost-effective given the vast amounts of data handled by accelerated computing workloads. By creating on-prem infrastructure, you can bend TCO with higher utilisation and thinner "hot" tiers.

In addition, choosing processors with higher core density, like the AMD EPYC™ 9004 series processor, improve performance per £ and mean fewer servers are needed, cutting down on your server and licensing costs.

## Sovereign compliance

When sensitive data is embedded within your organisation, or when systems are mission critical, sovereignty matters. For organisations in defence, research, financial services, healthcare or other regulated markets, demonstrating compliance is a must for infrastructure choices.

With on-prem, sovereign compliance is a given. Rather than risk sensitive data being shared in inadvertently, on-prem provides control and compliance over where the data resides, how it's secured and, critically, who has access to it.

## Performance optimisation

With the throughput-hungry workloads of HPC, optimising performance is a must. I/O bottlenecks can't be a blocker. Data gravity and low latency make on-prem the go-to solution for meeting the demands of HPC and accelerated AI computing. You have complete control over the hardware, interconnectivity and software stack, and can dedicate resources to specific workloads.

## 'Plug and play': Overcoming facility constraints

On-prem is the recommended option, but the reality is that many facilities lack the capabilities required for HPC and accelerated computing. Power density is often the limiting factor. While AI-class racks can run at over 100kW/rack today, many legacy data centres top out at single-digit kW/rack and cannot cope without redesign. Rack capacity determines your capabilities.

When your existing facilities fall short, how do you stand up HPC on-prem?

Modular data centres give you HPC capability without creating a complete data centre. 'Plug and play'. The portable solution with pre-built components allows rapid deployment on a hard-standing area with power and comms connectivity. This approach de-risks your implementation timescales and realises the value of HPC faster.

# Designing AI and HPC architecture

Architecture is the cornerstone for deploying HPC at scale. While compute is essential, the other tiers are equally important. Ultrafast storage through NVMe and parallel file systems. Efficient cold archives for long-term storage to manage cost and energy consumption. And the fabric for interconnectivity across your network. Each tier needs to be engineered to your bespoke requirements to get the most efficient solution.

## Sustainable HPC environments

The power consumption of HPC raises questions about its sustainability. GPUs can consume up to 1200W per card, with the potential for this figure to rise in the future. With the addition of cooling requirements, the power demands rapidly mount. But with strategic architecture decisions, there are options for building sustainable AI infrastructure, whether it's for a new investment in HPC or a refresh of existing infrastructure.

The starting point is to audit your infrastructure and understand your sustainability goals. Using our technology expertise and relationships with hardware vendors, such as AMD, we can advise you on an architecture that delivers a sustainable HPC environment.

For example, AMD EPYC™ processors advance energy-efficient x86 servers, allowing you to reduce energy costs and create a sustainable HPC environment. By using accelerators in all the GPUs in tandem with the CPUs, it's possible to shrink traditional HPC environments to a smaller footprint.

An AI-HPC readiness review can also uncover opportunities for efficiency gains in the cooling architecture. For some organisations, moving from air-cooled to liquid-cooling for the HPC technology stack could deliver substantial energy savings.

# HPC for AI: An end-to-end solution

Deploying HPC is a complex programme but a critical part of delivering value from AI. As you consider how to approach HPC for your enterprise, you'll need each of these elements to create the end-to-end solution.

## Audit

The starting point is to be clear on your business goals for AI and the target timelines. Auditing your infrastructure and facilities is also crucial for informing which solutions are appropriate for your HPC environment.

## Consultancy

Get advice on the strategy and possible options for your HPC deployment. Consultants can recommend how to integrate into your existing data centre facilities, identify co-location facilities or introduce new ways to deliver HPC architectures, such as modular data centres.

## Architecture design

Work with architects who can design production-ready on-prem infrastructure that optimises performance using leading HPC and accelerated computing vendors. It's critical to design a solution with processors that meet the demands of the compute intensive workloads of HPC and AI, such as AMD EPYC™ 9004 (CPUs) processors.

AI isn't a one-size-fits-all solution, which is why consultancy and design guidance, from SCC and AMD, as a partner are critical.

## Hardware procurement

Choosing a vendor-agnostic partner means you can get a bespoke solution using multiple vendor capabilities and established relationships.

## Programme management

Deploying HPC on time, within budget and to quality demands dedicated programme management. Experienced programme managers can plan dependencies, manage risks and apply insights from previous implementations.

# Where to start with your HPC programme

Before contacting vendors or redesigning your architecture, there are five key questions to answer about your HPC deployment:

## 1) What are your goals and timelines for delivering value from AI?

When HPC is the enabler for enterprise AI, knowing the timelines you have for AI sets the goals for the HPC deployment.

Clarify your AI objective - it shapes the next steps. Is it for productivity tools, in-house development, predictive analytics, or something else? Remember, not everything labeled 'AI' truly is.

## 2) What are the capabilities and constraints of your existing infrastructure?

The power density and rack capacity in your infrastructure will determine what can be deployed within your current facilities.

## 3) How important is sustainability to your organisation?

Deploying a sustainable HPC environment is possible, it will depend on your appetite for prioritising sustainability and choice of facilities.

## 4) What performance requirements does your enterprise AI have for HPC?

Understanding the use cases for your AI and the expected workload is a key factor for architecting your infrastructure.

## 5) What alternative facilities do you have available with power and connectivity?

Knowing what other facilities you have access to can determine the available options for your HPC environment.

# Why SCC?

## Expertise

Our team of experts are up to date with the latest AI and HPC technologies to ensure our customers receive the best possible advice and service. From consultancy and architecture to procurement and programme management, you receive insights and recommendations tailored to your business.

## Infrastructure

Our partnerships and alliances with the world's leading technology vendors make us vendor-neutral, allowing us to offer multi-vendor solutions to our customers. We provide consultancy on the most suitable migration-ready infrastructure and virtualised solutions.

## Experience

Our customers have trusted us to deliver transformational projects across a wide range of technologies for nearly 50 years. In addition to in-depth experience with AI infrastructure, data centre deployments and hardware procurement, we also have capabilities in deep-archive tape storage.

## Ready to Accelerate Your AI Strategy?

Ensure your infrastructure can handle the demands of AI-driven workloads. Our AI-HPC Readiness Review will help you:

- ✓ Assess your current HPC and GPU capabilities
- ✓ Identify gaps for AI scalability and performance
- ✓ Get expert recommendations for architecture and deployment.

[Book your review today and take the first step toward building a future-ready AI environment.](#)