White Paper, sponsored by Axians
APRIL 2024

The Essentials of Successful HPC Project Implementation

Sustainable Domain-Specific HPC Expertise, Hybrid-Cloud Experience, and Software Offerings

Madhumitha Sathish
Research Manager,
High Performance Computing

Peter Rutten
Research Vice President,
Performance Intensive Computing

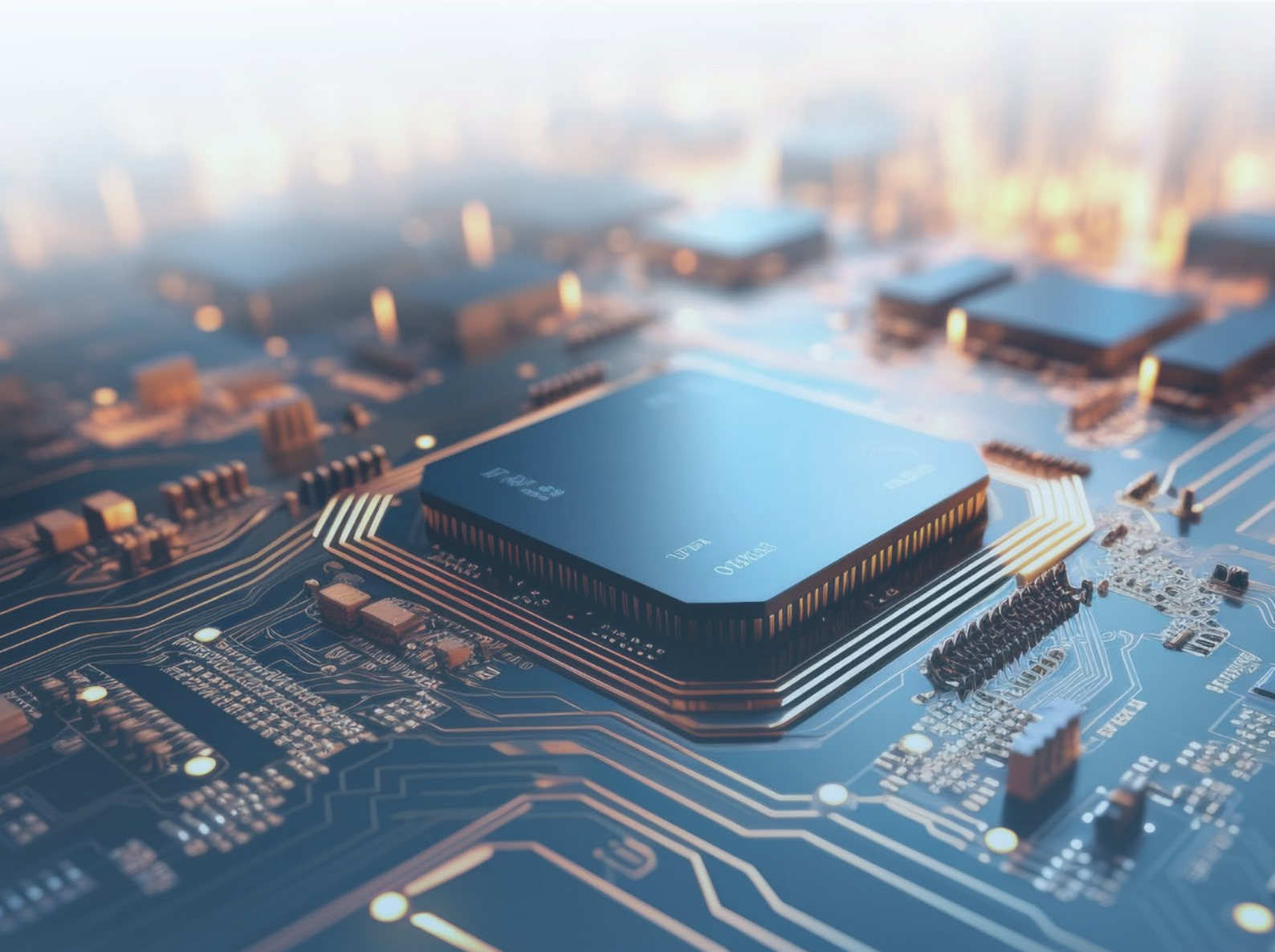


Table of Contents

CLICK BELOW TO NAVIGATE TO EACH SECTION IN THIS DOCUMENT.

Executive Summary

Countries, worldwide, view high-performance computing (HPC) and AI as top national priorities and as means to keep their economies, science bases, and defenses competitive. With HPC going mainstream, enterprises have adopted HPC and AI as strategic and indispensable components of their capabilities and are therefore expected to have sustained budgets and funding for related workloads.

The number of HPC use cases across vertical industries is expanding, driving the growth of the overall HPC market. This can largely be attributed to the convergence of HPC, AI, big data analytics, and quantum computing.

HPC is being adopted in the cloud, marking a shift in the deployment paradigm, although some end users still prefer to have their HPC on premises due to the low-latency, manageable-cost, and high-security requirements of certain HPC workloads. Demand for converged infrastructure for HPC, AI, big data analytics, and even quantum computing is increasing. Organizations seek mature end-to-end high-performance solutions that ease complex integrations, improve business agility, and address sustainability concerns.

Vendors are advised to tier their offerings based on client type (national research laboratory, corporate business, and academic institution), infrastructure locality (on-premises, cloud, edge, and hybrid), and required workflows. Vendors with proven track records of building HPC environments for a diverse spectrum of end users and with demonstrable capabilities in on-premises and cloud infrastructure HPC projects and managed services are likely to have a competitive edge when competing for contracts in this competitive market.

Overview

HPC best practices date back to the 1960s — from research laboratories involved in high-priority national security projects that required customized hardware (e.g., Cray supercomputers) to achieve the scale and performance previously targeted. Since then, the HPC community has delivered revolutionary advances in science and technology in areas such as aircraft design, digital twins, chip design, weather and climate predictions, nuclear energy, and vaccine and drug design.

IDC sees HPC as a constant evolution that will continue to revolutionize many technologies, particularly in such fields as generative AI and quantum computing. HPC is expected to solve many of humanity's greatest problems and challenges at an accelerated pace. HPC is used to progressively increase the accuracy and speed of resolution to current challenges, including:

- Designing a new hypersonic aircraft that can travel at Mach 10 and a greener aircraft that requires less fuel to make possible net-zero carbon emissions by 2050
- Forecasting the weather to higher degrees of accuracy (>90% for up to 5 days)
- Modeling a pandemic spread (e.g., flu, COVID-19, malaria, and Ebola) with higher accuracy using multiple parameters (e.g., vaccinated vs. non-vaccinated, age, and genetic predisposition) and enabling the creation of vaccines very quickly (e.g., a COVID-19 vaccine in less than a year — a process that normally takes 5–10 years)
- Simulating NASA's landing of a vehicle on Mars, the Moon, or Venus for deeper space exploration

Banking and financial institutions use HPC in fraud detection and protection, real-time risk management, building and executing complex financial models, algorithmic trading, and high-frequency trading. HPC has become a technology enabler for smart features in every industry vertical — fully autonomous vehicles, smart factories, smart ports, smart groceries, intelligent traffic control, and dynamic routing, among others.

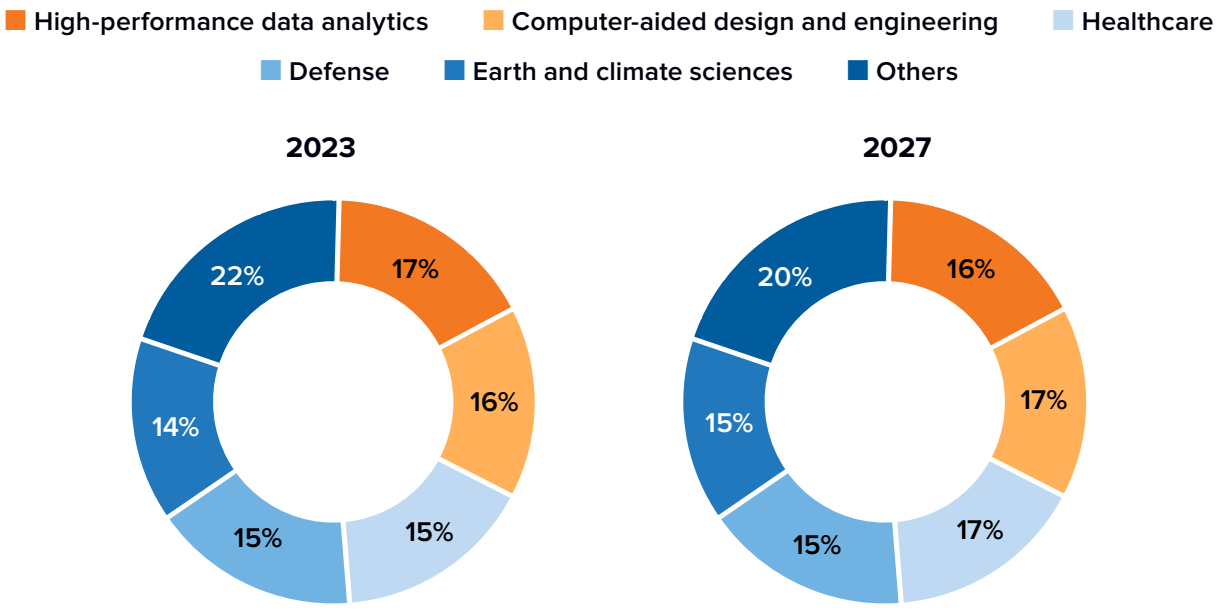
HPC Goes Mainstream with Ever-Expanding HPC Use Cases

IDC research shows that organizations — from corporate businesses to national & government research labs and academic institutions — view HPC as important, strategic, and indispensable. This implies that HPC has gone mainstream, with organizations adopting new strategies to increase their organizational intelligence through HPC and AI-driven initiatives.

Six decades of HPC has led to the foundation of infrastructure for generative AI and several other performance-intensive computing technologies. With HPC going mainstream, the traditional HPC community (federal governments around the world, national research laboratories, and academic institutions) is being joined by corporate (for-profit) businesses undergoing digital transformation. The number of HPC use cases continues to expand, driving the growth of the HPC market. This can largely be attributed to the convergence of HPC, AI, big data analytics, and even quantum computing.

FIGURE 1

HPC Server Market by Use-Case Category



Source: *Worldwide High-Performance Computing Server Forecast, 2023–2027: Enterprise Will Overtake HPC Labs* (IDC #US50525123)
Note: Other segments include finance and insurance services, space sciences, IT, and chemical engineering.

What to Expect from a Technology Supplier

CAPEX/OPEX Expenditure Analysis of All Deployment Models

For many organizations, HPC deployments on premises and in the cloud are of a heterogeneous and hybrid nature; they retain their on-premises setups while migrating some of the less sensitive workloads to the cloud and the edge. Typically, performance trumps all for HPC, but cost flexibility follows. The costing of short-term and long-term HPC purchases involves detailed total-cost-of-ownership (TCO) and ROI analysis based on the cost per core per time unit (hours, minutes, and seconds) and corresponding CAPEX/OPEX expenditure analysis, covering on premises, cloud, and edge. Such calculations are central to decision-making.

Consolidated HPC-AI Infrastructure

Technology suppliers should strive to offer infrastructure that can accommodate a combination of HPC and AI computing workflows. Several top-tier suppliers have already embarked on this course at various stages of maturity in on-premises, cloud, and edge implementations.

Expanding Application Use Cases

Performance is based on application use cases. Technology providers should build as many application benchmarks as possible for traditional and mainstream applications (e.g., relating to aircraft design, oil and gas, engineering, bioinformatics, and pharmaceuticals) and for newer ones entering the market, such as those around financial services and smart cities. It is worth noting that application use cases define performance requirements.

Streamlining the Data-Management Pipeline

Streamlining the entire data pipeline while qualifying input/output (I/O) requirements at every processing stage of the workload, especially when dealing with unstructured data, is very important. In addition, data analytics capability is a critical component, and technology suppliers should prioritize improving this to enable faster insights from data (i.e., in real time, including data in motion) with higher predictive accuracy.

Simplifying HPC Software Integrability

Suppliers should have a comprehensive software portfolio, with software that is easy to integrate with other software tools (both open source and proprietary), as HPC infrastructure is highly heterogeneous and hybrid in its implementation and workflows.

HPC Trends

Organizations Striving to Improve Business Outcomes by Building Enterprise Intelligence

IDC's Future of Intelligence (FOI) practice defines enterprise intelligence as an organization's capacity to learn, combined with its ability to synthesize the information it needs to learn, and to apply insights at scale by fostering an enterprise-wide data culture. As organizations strive to improve business outcomes by building enterprise intelligence, they are experimenting with new technologies, such as high-performance computing (HPC), as they become more accessible and mainstream.

HPC — Hybrid and Heterogeneous

Traditionally, HPC infrastructure has been placed on premises to keep the compute and storage hardware close by for optimization, ease of management, and compliance regarding restricted data. However, the deployment paradigm has shifted, with HPC being migrated to the cloud.

Leading HPC vendors are building hybrid and converged HPC, AI, big data, and quantum computing infrastructures on premises, in the cloud, and at the edge. In the cloud and at the edge, the priorities are greater business agility, more sustainable solutions, ease of use, and flexibility to help ease the complexity of integration. With data being freed from on-premises silos and migrated to cloud, an analysis of the entire data pipeline is needed, looking at data management and data logistics, to move toward real-time data analytics capability for faster insights.

Customers that have on-premises HPC infrastructure would benefit from establishing access to compute infrastructure in the cloud for immediate deployment while keeping an eye on cost, performance, scalability, and task-completion timelines. This awareness will help them plan and apportion budgets for both short-term and long-term compute and storage requirements in areas with a shortage of adequate infrastructure or skilled labor.

Because HPC is inherently hybrid and heterogeneous, managing such complex infrastructure with long-term planning using decades of experience from a skilled team is crucial. It is typical for HPC to use diverse compute architectures and multitiered high-performance storage with varying I/O requirements. Moreover, several generations of network topology (e.g., InfiniBand and Ethernet) are also meshed in the same network interconnect, linked by gateways. Given these conditions, sustained high operational performance, security, and availability are predominantly the priority metrics.

By 2025, 50% of enterprises will have broken with the public cloud migration trend and will primarily deploy HPC on premises or with a managed services or colocation provider.

The number of workloads migrating to various cloud environments continues to expand, but a significant portion of the HPC end-user community prefers to remain largely on premises, with the largest clusters in self-owned/operated facilities. Many workloads remain on premises due to the low-latency, manageable-cost, and high-security requirements of certain performance-intensive computing workloads, such as those conducted using HPC.

IDC research reveals that, owing to concerns over data compliance and security and the cost of additional cloud compute and storage, organizations expect only a smaller percentage of the HPC environment to be deployed in the public cloud in the next 12 months. The research also revealed that the integration of public cloud into an organization's HPC environment is complex and that the data used for HPC cannot move freely between the organization's on-premises location/colocation and public cloud for cost, performance, and security reasons. Furthermore, bare-metal–like performance is always preferred with dedicated/private cloud access for enhanced security, as organizations usually require higher security on their HPC workloads than is the case for their general-purpose workloads.

Although public cloud is currently the largest deployment scenario for HPC across many industry sectors (manufacturing, construction, oil & gas, healthcare, higher-education, IT services, utilities, and telecommunications, among others), in the government and finance sectors, self-managed on-premises datacenters still dominate. Research reveals that, in the coming years, organizations in the government and finance sectors are likely to expand their HPC deployments in managed on-premises and colocation datacenters (as a service) and with managed service providers at the expense of public cloud and self-managed on-premises datacenter deployments. This will also be true of organizations in several other sectors — IT services, cloud services, manufacturing, and oil and gas. We expect this trend to continue.

By 2025, global sustainability mandates will have driven 80% of institutions and businesses to use renewable energy sources for all HPC and AI deployments.

It is not just the digital-first strategies of businesses around the world that have led to an increase in power consumption by IT systems; it is also the need for performance-intensive computing (PIC) to remain competitive. These PIC workloads, which run on compute-intensive infrastructure, are power hungry, as are their cooling systems. IT infrastructure that enables HPC and AI deployments will continue to consume a lot of energy.

Governments around the world are formulating and implementing strategies to tackle climate change. Large energy consumers are being watchful and are proactively aligning with global interests and ambitious targets set by their respective governments to curb greenhouse gas (GHG) emissions. These include organizations that rely heavily on datacenters or build and manage datacenters to enable HPC and AI deployments. Some of the measures being adopted at the corporate and IT infrastructure levels include energy-efficient hardware design, efficient cooling, resource optimization, and the modernization of legacy hardware and applications. The use of renewable energy sources stands out.

By 2028, the need for sufficient and affordable computational power for HPC and AI workloads will have led to breakthroughs in processing technologies, breaking the monopoly of accelerator vendors.

Workload acceleration using coprocessors has greatly driven up the average selling price (ASP) of a datacenter or cloud server. Today, the ASP of an accelerated server is approximately three times higher than the ASP of a nonaccelerated server, and the gap is expected to increase in the coming years.

Organizations will likely become more focused on returns on investment from coprocessing technologies, and cloud service providers are already leading the way in opting for alternatives — or even developing their own. In addition, competitors are in every corner of the market, from silicon incumbents to start-ups eager to disrupt leading vendors' market dominance, whether with lower pricing for a comparable — or possibly greater and more targeted — performance or with a much lower cost for a lesser but adequate performance.

Given the amount of talent and resources that competitors are pouring into developing compelling alternatives, IDC expects the market to be disrupted in the next five years and for it to become more competitive, thus leading to lower ASPs across the board.

All AI and HPC use cases are centered on complex mathematical problems. As businesses' reliance on these performance-intensive computing use cases to drive differentiation increases, their requirements for the accuracy and/or predictability of outcomes will also increase. In tandem, advances in analog and quantum computing are leading to both computing system types becoming more capable and versatile in solving an increasingly diverse array of complex mathematical problems.

An increase in demand for quantum computing integration into HPC environments and classic hybrid computing environments is likely to prevail.

IDC expects systems vendors to continue collaborating with technology enablers, increasing the number of options available to businesses for their performance-intensive computing problems. Many of these options are available today but are not yet fully mature. More such options will be born from technological breakthroughs, such as in next-generation quantum computing systems. It is therefore noteworthy to understand that investing in quantum computing technology and quantum skills development will not result in the demise of classic compute technology or the workforce.

Because quantum computing will only be suitable for running specific problem types, the technology will exist in hybrid quantum-classic compute environments. Quantum and classic hardware vendors will continue to collaboratively develop quantum and quantum-inspired computing systems dedicated to solving HPC problems. Enterprises will continue to identify potential HPC use cases that are suitable to run using quantum computing technologies.

Considering Axians' Proposition

VINCI Energies, a group contributing to the environmental transition by helping to set trends in the digital landscape and energy sector, offers ICT solutions under the brand name of Axians. Axians is a €3.6 billion network with a presence in 37 countries around the world. Axians supports its customers — private-sector businesses, public-sector entities, operators, and service providers — with their development of infrastructure and digital solutions. To this end, Axians offers a comprehensive range of ICT solutions and services spanning business applications and data analytics, enterprise networks and digital workspaces, datacenters and cloud services, telecommunications infrastructure, and cybersecurity.

Through VINCI Energies' acquisition of a major portion of its IT services business in 2022 from Kontron, the group has been bolstering its position in the Europe ICT market under its Axians brand. Axians, an already established digital transformation solution provider with several strategic partnerships and deep industry know-how, has integrated the former S&T IT services business. VINCI Energies' business units in DACH (Germany, Austria, and Switzerland) and Central and Eastern Europe (Poland, Croatia, Czech Republic, Serbia, Slovakia, North Macedonia, Albania, and Montenegro) now operate under the Axians brand name.

Axians has delivered, among others, the following notable projects:

- Axians Austria designed and developed the addHelix platform, using AI algorithms to increase the flexibility and agility of logistics services.
- Axians Poland supported Cargotec Poland, a leading provider of cargo and load handling solutions, in the implementation of cloud-based ERP Infor LN, including integration with the national e-invoicing system.
- Axians and Cisco are digitalizing and expanding enterprises' Internet of Things (IOT) in Germany and the Netherlands with Cisco Private 5G.

- Axians has helped improve the quality of life in Aix-en-Provence through the deployment of smart city projects, including IoT.
- A collaborative project on sustainable building design by VINCI Energies, DIANE, uptown Basel, Quantum Basel, and D-Wave leveraged quantum computing in a quantum proof of concept (Quoc) for optimizing the design of heating, ventilation, and air conditioning (HVAC) systems for complex buildings.

A major portion of Axians' turnover is from its Cloud and Datacenter Infrastructure business, which has been delivering classic cloud and datacenter projects for years.

Service offerings under Axians' Cloud and Datacenter Infrastructure business include digital transformation consulting services, managed services, software solutions, modernizing on-premises datacenters, building cloud environments (including private cloud and multicloud deployments), and agile solutions and IT services for the consolidation of infrastructure in hybrid cloud. The company's innovation prowess is evident through several notable projects delivered via strategic partnerships with leading companies across a wide range of vertical markets, including finance, healthcare, utilities, communications (IoT and smart cities), education, and logistics, and in several niche markets, such as cybersecurity and HPC.

Axians has partnered with Hewlett Packard Enterprise (HPE) to drive innovation and deliver projects in the HPC space. The following section outlines examples of HPC projects delivered by Axians and discusses Axians's association with the University of Zagreb University Computing Centre (SRCE) as an HPC partner and systems integrator. The capabilities of Axians were instrumental in delivering Croatia's first petascale supercomputer, called Spek, which is a part of an important national project called the Croatian Scientific and Educational Cloud (HR-ZOO), which was coordinated by the University of Zagreb University Computing Centre.

The supercomputer, based on HPE Cray technology, was constructed by Axians with 100% direct liquid cooling. As per Axians, the supercomputer has a total of 8,384 processor cores and 81 graphics processors, with 32TB of RAM. Together, this provides a performance of 1.25PFLOPS. The storage is based on an HPE ClusterStor 1000 system with a usable capacity of 580TB and a throughput of 238GBps for writes and 450GBps for reads. Most servers are housed in a single HPE Cray EX2500 cabinet fully integrated into the supercomputer.

IDC interviewed senior managers at the University of Zagreb University Computing Centre to gain a perspective on the HPC capabilities and expertise of Axians.

SRCE was founded in 1971 as a central Croatian institution for the utilization of information technology in science, higher education, and other areas. Within its Advanced Computing service, SRCE provides computing infrastructure for solving resource-demanding computational problems for the academic and scientific community in Croatia. More than 60 different institutions, 600 users, and 280 scientific projects in Croatia use these advanced computing resources, including planning, developing, and improving e-infrastructure and digital services. SRCE also runs a number of digital services for the community and provides courses, workshops, and gatherings for the professional community to improve digital competencies.

The acquisition of a new supercomputer, which replaced ageing computer cluster Isabella, was planned through EU structural funds within the HR-ZOO project. The value of the project was around €26 million. HR-ZOO provides free advanced compute, network, and storage resources to the Croatian scientific and academic community and a flexible environment for collaboration for every scientist, researcher, and teacher, including the possibility to transfer, store, and process large amounts of data. This collaborative environment features specialized support for advanced IT. The project included building next-generation e-infrastructure, the supercomputer, and a data cloud — all fundamental components of national research and innovation in support of modern and multidisciplinary science and state-of-the-art research and education in the Republic of Croatia.

The project involved building datacenters and procuring HPC and cloud resources — processes handled by two ICT companies, with one being Axians. The vendor helped build the datacenter and delivered the HPC resources. Axians ensured integration with HPE Cray and the installation of HPC resources, water cooling,

"Basically, we have built an up-to-date national e-infrastructure for our academic and scientific community. This was a strategic project for us because these advanced ICT resources and services, including HPC, serve as an underlying foundation on a national level and will be used by different scientific and academic institutions and end users."

Ivan Maric,
Director of the University of Zagreb University Computing Centre and HR-ZOO project manager

and all other technical systems needed for HPC. SRCE's procurement of compute and storage hardware was via public tender and public procurement. Everything procured was ultimately HPE based.

SRCE's experience with Axians has been very positive. HPC was a prime component of the infrastructure that SRCE was building. SRCE did not have the in-house expertise to build an energy efficient HPC system. Axians addressed the need for a proven partner. Axians demonstrated its competencies in the HPC realm, including in cooling systems and ICT know-how.

SRCE acknowledges that, because the project was complex, SRCE required close collaboration with its partners — under a single-team model, rather than as two separate entities. Cooperation with Axians gave SRCE exactly that. Regular meetings with a team from Axians — whose members had no conflicts of interest from the initial preparation phase — resulted in the successful implementation of HPC and the expected project outcomes. SRCE stated that Axians met all criteria; it built the racks, installed the cooling systems and pipelines, benchmarked the software, provided compilers and batching systems, and ensured sustained power for 1.25PFLOPS.

End users include researchers, scientists, and academics from diversified scientific fields in Croatia. The biggest consumers of HPC resources, involving over 100 different scientific applications, are from the areas of computational chemistry, biology, physics, and AI. The supercomputer contributes significantly to science and research in SRCE and other institutions in Croatia.

According to SRCE, most end users are scientists who are well versed in computing. With its partners, SRCE is developing and finetuning software in areas such as computational chemistry, AI, and computational fluid dynamics. Other areas of software development also offer opportunities for the institution.

Infrastructure modernization has enabled easy access to SRCE's Advanced Computing Service, which offers rapid mathematical calculations. For example, an earthquake simulation using seismological data, which used to take 14 days to perform, now takes a day or even just a few hours. Axians' expertise has thus contributed notably to the academic and scientific community of Croatia, further strengthening Axians' position as an innovative HPC player and systems integrator in the Croatian ICT market.

SRCE recommends the involvement of HPC players like Axians in the very early stages of such projects. Axians' proven expertise in the integration of HPE Cray, HPC, liquid cooling systems, and systems integration made the vendor a strong partner in SRCE's HPC project — one able to overcome prominent challenges such as installing a cooling system in winter.

SRCE has a four-year maintenance services contract with Axians. During this contract, the institution expects the transfer of knowledge from Axians to continue, which will help SRCE in its position as a broker for the IT vendor and research community in Croatia and will decrease the ICT concerns of scientists and researchers. SRCE acknowledges that Axians' input into the HPC project was on budget and on time.

"Axians is a very solid and proven partner with strong quality expertise in different areas of HPC, which is very complex. Axians provided all the professional help, especially with these new technologies. Their expertise was very beneficial to us because we didn't know every aspect of building an energy efficient HPC system. The collaboration with the Axians team was commendable. During the implementation of this HPC project, it was like there were no two companies; we were like a single team trying to get this implementation done. The knowledge and expertise they showed in our project make them a very good partner for any similar project in Europe or elsewhere."

Ivan Maric,
Director of the University of Zagreb's University Computing Centre and HR-ZOO project manager

"Axians addressed everything we specified in the tender call and fulfilled all the requirements. The project was accomplished very successfully."

Emir Imamagic,
Head of Advanced Computing, SRCE

Conclusion

To maintain and sustain their competitiveness in R&D technology at the global level, many governments around the world have marked HPC as a top national priority. As such, funding and budgets will be sustained and may potentially increase. Academic institutions will continue to drive demand for compute-intensive infrastructures, with state funding being consistent or expanding to accelerate research.

With a significant and increasing number of organizations embracing digital strategies and navigating the digital landscape, the prioritization of technology investments is crucial for enhancing efficiency, maintaining competitiveness, driving innovation, and identifying new revenue-generating opportunities. As enterprises across industries consider increasing technology investments not to miss out on opportunities and to keep pace with the rapidly evolving market landscape, the adoption of HPC, AI, big data analytics, and quantum computing is becoming inevitable. As businesses increasingly rely on performance-intensive compute use cases to drive differentiation, their collaboration with ICT vendors with proven capabilities across a wide array of use cases and several vertical industries will be key to establishing a strong presence in the HPC-AI realm.

Regarding government-funded organizations and institutions, since the selection of ICT vendors (including HPC and AI vendors) is conducted via public tender or public procurement, the strategic efforts of vendors to bolster their positions in regional markets will be key to their success in this market. Proven track records in successfully implementing projects that align with regional or global sustainability mandates and strong partnerships with national research labs will provide a competitive edge.

Systems integrators operating in the high-performance computing market can ascend the value chain and better serve organizations of all types — whether academic institutions, national labs, or corporations — by incorporating managed services into their offerings. By providing ongoing support and managing infrastructure, maintenance, and optimization services, a systems integrator can ensure that clients' HPC systems operate at peak efficiency and reliability. Offering managed services will not only enhance the integrator's value proposition; it will also foster deeper partnerships with clients through continued collaboration. Vendors should therefore work strategically toward innovation, enhancing cloud capabilities, and moving up the value chain to offer a complete software stack supported by managed services.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology, IT benchmarking and sourcing, and industry opportunities and trends in over 110 countries. IDC's analysis and insight help IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives. Founded in 1964, IDC is a wholly owned subsidiary of International Data Group (IDG, Inc.)

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. As a premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets, IDC's Custom Solutions group helps clients plan, market, sell, and succeed in the global marketplace. We create actionable market intelligence and influential content marketing programs that yield measurable results.

IDC UK
5th Floor, Ealing Cross, 85 Uxbridge Road, London, W5 5TH, United Kingdom
T 44.208.987.7100