VERTIV™

E-book

Strategic imperatives: AI's impact on colocation data center trends



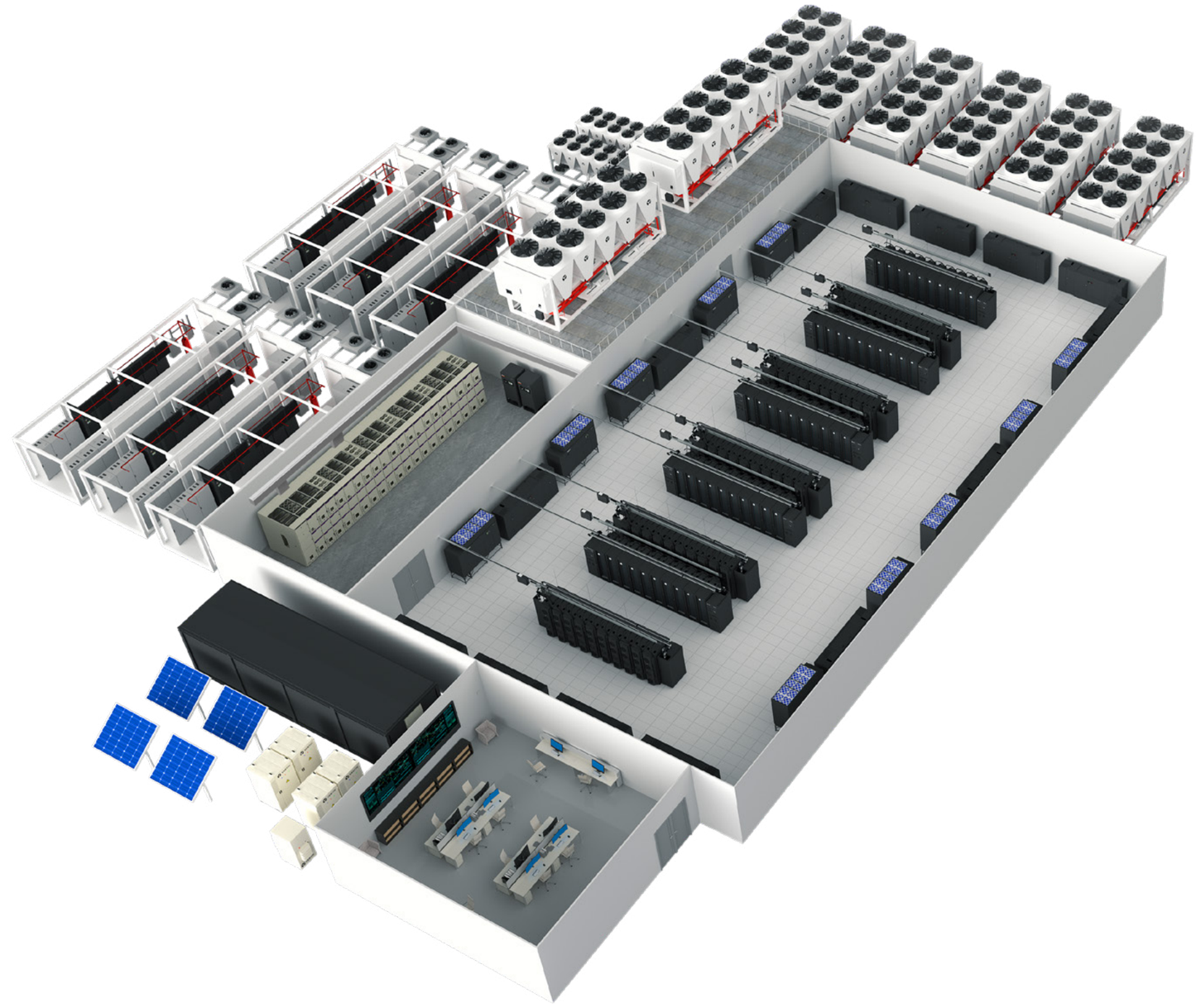
Table of Contents

The Five Vertiv Imperatives 05
AI's Impact on the Data Center
Vertiv Imperatives and AI Adoption
- Imperative 1: Be Transformative
- Imperative 2: Be Efficient
- Imperative 3: Be First
- Imperative 4: Be Confident
- Imperative 5: Be Future Ready

Using Imperatives to Shape Data Center Strategy 13
Strategic imperatives in cloud and colocation data center design

Streamlining the Path to AI Adoption and Growth 15
Rethinking Infrastructure Design – Principles for Success
Vertiv™ 360AI: Transforming Principles into Solutions
- AI-ready Reference Designs
- AI-ready Prefabricated Solutions

Selecting the Right Partner



The Vertiv imperatives

Recent rapid changes from AI to an evolving energy grid have presented data center operators with both unique challenges and remarkable opportunities. To help them manage this change, Vertiv has defined five imperatives to guide response to and adoption of new technology.

Prioritizing the imperatives of transformation, efficiency, speed, confidence, and future readiness can help enterprise data center owners and operators better position themselves to exploit the benefits of new technology and hopefully side-step the pitfalls.

While the imperatives apply to a wide range of challenges, this ebook focuses on how they can guide owners and operators to address artificial intelligence's (AI) transformative impact on data center design and overall capacity.



The Five Vertiv Imperatives

1. Be transformative

2. Be efficient

3. Be first

4. Be confident

5. Be future-ready

While these imperatives apply to a wide range of challenges, this ebook focuses on how they can guide owners and operators to address the transformative impact of artificial intelligence (AI). In leveraging these imperatives, data centers can build on their foundations that embraces the complexities of AI and delivers scalable solutions across different enterprise verticals.

AI's impact on the data center

AI encompasses various subfields and technologies, including machine learning, natural language processing and others shown in the accompanying table. These different types of AI are interconnected and often build upon each other to enable more advanced AI systems.

AI types	Definition	Sample use cases
Machine learning	Algorithms that can learn from and improve through experience without human programming.	• Personalization and recommendations • Fraud detection • Image recognition
Predictive analytics	The use of historical and current data to benchmark and forecast future trends, events, and outcomes.	• Demand and inventory forecasting • Sales and revenue forecasting • Staff optimization
Natural language processing	AI systems that can understand, interpret, and generate human language.	• Customer service • Automated document processing • Market intelligence
Computer vision	AI systems that understand and interpret visual information.	• Facial recognition • Medical Imaging • Autonomous vehicles
Robotics	AI systems that improve the performance and functionality of robotic systems.	• Material handling • Industrial automation • Robot-assisted surgery
Expert systems	AI systems that emulate the human decision-making process and apply expertise to solve complex problems.	• Medical diagnosis • Product engineering and design • Compliance management
Generative AI	AI systems that can create new content (e.g., text, images, videos, and code) in response to prompts.	• Content creation and marketing • Software development • Customer service agents

Table 1. Some types of AI in use today.

While many of these technologies have been used by enterprises for years, the release of publicly accessible generative AI (GenAI) products in 2022 catalyzed AI adoption in a way other types had not. According to McKinsey, which surveys business leaders on their use of AI annually, "AI adoption has increased dramatically in [2024], after years of little meaningful change."[1] The percentage of organizations using **AI jumped from 55% in 2023 to 72% in 2024**, with the use of generative AI accounting for most of that growth.

Now, data center directors and managers are facing critical decisions about how they will support their organizations' AI initiatives today and in the future. In the report, "AI: Enterprises Are Active But Cautious," Uptime Institute Intelligence notes: "The strategies and spending decisions of these enterprises will be a critical determinant in shaping the future role of AI and the evolution of the data center in the decade ahead." [2]

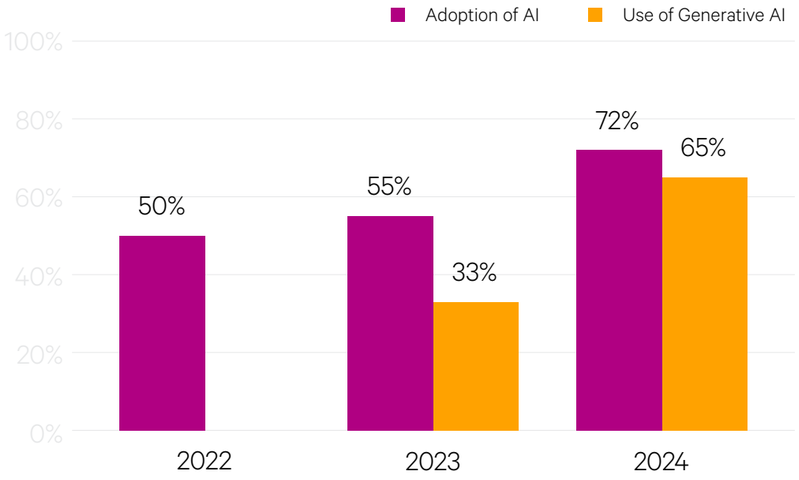


Figure 1. Percentage of organizations that have adopted AI in at least one function. The use of generative AI (GenAI) almost doubled between 2023 and 2024. Source: McKinsey.[1]

The same report makes clear that, despite executive-level enthusiasm for AI, "enterprises remain uncertain about where and how they will deploy AI." On-premises, colocation, and cloud hosting options all offer potential benefits and challenges with the right choice for each organization depending on factors such as size, data sensitivity, tolerance for latency, performance requirements and existing infrastructure. However, as AI continues to evolve, directors and managers will need to adapt the designs and capabilities of their data centers regardless of the chosen deployment model.

Vertiv imperatives and AI adoption

AI workloads are considerably more power-intensive than traditional data center workloads with power demand characterized by pulse loads during periods of high computational activity. AI hardware is also physically denser, consuming more power per unit area than traditional CPUs, which concentrates heat generation and often requires new approaches to heat removal.

Enterprises today must take steps to safeguard the AI systems they are hosting and make sure they have the infrastructure required to manage these dynamic and ultra-dense workloads. Vertiv imperatives are presented as tools for establishing priorities and identifying solutions that address the data center challenges introduced by AI.

Imperative #1: Be transformative

According to a global survey of executives by Deloitte, 79% of respondents said they expect GenAI to drive substantial organizational transformation in less than three years.[3]

Before that transformation can take place, it may be necessary to upgrade some data center systems. This could mean increasing system capacity across the power train, adopting cooling technologies designed specifically to remove heat from high-density racks and clusters, and making more efficient use of data center white space.

Whether retrofitting or building new, data center infrastructure could be required to support rack power densities exceeding 100kW. Rack weight is also an issue, with some AI systems requiring more than double the weight of conventional racks. Achieving these goals will require a holistic approach to data center infrastructure that utilizes designs developed collaboratively by computing and infrastructure providers, as is now happening with Vertiv and NVIDIA.[4]

> "The implications of AI's advancement in data centers extend beyond technical challenges; they offer unprecedented opportunities for growth and innovation. Done right, organizations can create opportunities to achieve higher efficiency, improved performance, and competitive edges in their respective fields."
>
> **— Martin Olsen,**
> *Senior Vice President, Global Product Strategy, Vertiv*

Imperative #2: Be efficient

AI can improve organizational efficiency by automating routine processes, speeding the analysis of large data sets, and streamlining forecasting and training. McKinsey's 2024 AI survey found that "the most frequently adopted use cases for gen AI tended to favor applications that focused on automating tasks to improve efficiency." [5]

Optimizing energy efficiency can be achieved through design approaches that minimize stranded power, leverage new approaches to cooling, implement management and control strategies that minimize energy consumption, and improve the circular economy within the data center.

> "One of our roles today is making the power that is there more available and enabling it is used more efficiently at the data center. You need to make sure you use the power you have effectively."
>
> **— Stephen Liang,**
> *CTO, Vertiv*

Imperative #3: Be first

Many enterprises see AI as a strategic technology that can create a competitive advantage by being first to market with AI-powered solutions. According to Uptime Intelligence a common trait among the enterprises that have been most aggressive in adopting AI is the belief that "continued success is based on, or protected by, technological innovation."

That can put pressure on data center directors to compress timelines for adopting AI compute. The traditional approach of designing and managing power and cooling systems independently will prove inadequate in meeting business expectations. An integrated approach to critical infrastructure that brings together power and cooling design can mitigate risk and accelerate deployment speed.

Assessing multi-vendor qualifications, dependencies, and technical specifications can also delay AI design and implementation. Providers that can supply complete solutions may be better positioned to support fast-moving projects. However, that approach still needs to be underpinned by industry-wide standards and best practices, which are still fairly nascent in the AI arena.

> "The element that will make a data center operator truly ready for the future will be finding partners that not only have broad knowledge of emerging technologies, but also how these technologies can work together to meet future demands. Without that partner, I think data center operators risk watching the future pass them by."
>
> **— Martin Olsen,**
> *Senior Vice President, Global Product Strategy, Vertiv*

Imperative #4: Be confident

If AI is to play the transformative role many expect, enterprises need to be confident in the ability to support and scale their AI solutions reliably. The Deloitte "3Q 2024 State of Generative AI Survey" found that positive value generation led two of three surveyed organizations to increase AI investments. However, nearly 70% said their organization had moved 30% or fewer of generative AI investments into production.[6] Concerns related to governance, security, and data center limitations were cited as key obstacles to advancing initiatives.

Thoughtful planning can help enterprises have the right data center infrastructure in place to support their AI strategy. Finding the right balance between risk and cost, anticipating the impact of rising rack densities, and preparing for the risk tolerance of future business-critical use cases should all be considered in long-term planning. Remote management systems that enable fast and efficient troubleshooting and well-defined service and maintenance plans will be essential to enable and reinforce the reliability of AI systems.

> "The future of data center maintenance is here—smarter, more efficient, and more reliable than ever... Data centers can anticipate risks and benchmark assets, leading to improved risk management and enhanced availability. An advanced AI model can gather data from many assets and provide composite health scores across units and systems. These reports can empower data center stakeholders to make more proactive and informed decisions, potentially achieving unprecedented levels of operational efficiency and resilience."
>
> **— Michael O'Keefe,**
> *Senior Vice President of Global Services, Vertiv*



Imperative #5: Be future-ready

The density and performance of AI platforms are growing rapidly. Today's "high" density racks will become normalized as they are eclipsed in the near future by even higher-density racks. Plus, new form factors such as NVIDIA's SuperPODS[7,8] will push densities even higher than is possible with current architectures.

Preparing the data center for a power-dense future requires making infrastructure investments that mitigate the cost of upgrades, can grow in large blocks as well as incrementally and pave the way for self-healing through industry standard APIs. It's also important to put in place systems that can meet current and future responsible business challenges by using power efficiently, enabling support for alternative energy sources, capturing and reusing heat, and minimizing e-waste.

> "With AI driving rack densities into three- and four-digit kWs, the need for advanced and scalable solutions to power and cool those racks, minimize their environmental footprint, and empower these emerging AI factories has never been higher."
>
> **— Giordano Albertazzi,**
> *CEO, Vertiv*

> "We understand that our actions today have a direct impact on the environment tomorrow, and we're dedicated to making sustainability a fundamental part of everything we do. Our innovation focuses on reducing energy consumption, minimizing waste, enhancing efficiency, and promoting sustainability. We closely track metrics such as energy use, water use, efficiency services, life cycle carbon emissions, support for alternative energy, and adherence to a circular economy to continuously monitor and improve our offerings."
>
> **— Dan Rapp,**
> *Senior Director, Global ESG and Environmental Affairs, Vertiv*

Imperatives	Business value	Data center requirement
Be transformative	Transform products, services, and customer experiences.	• Cross-functional expertise • Holistic solution design
Be efficient	Streamline processes and eliminate unnecessary costs.	• Extend the value and function of existing systems. • Close-coupled systems that minimize space and energy requirements • Connect enterprise-wide systems through unified monitoring and management platforms
Be first	Increase speed to market; accelerate innovation; create competitive advantage.	• Designs that reduce time and risk • Pre-configured systems that reduce deployment times
Be confident	Rely on AI systems to perform reliably when needed.	• Proven solutions supported by comprehensive services • Robust management and control strategy for power, cooling and IT systems, including remote management capabilities
Be future-ready	Expand AI use cases and prepare for continued evolution.	• Interoperable, upgradable, and scalable solutions that can adapt to change and utilize standard APIs to enable self-healing

Table 2. The Vertiv strategic imperatives can be used by enterprises and data centers to maximize and leverage AI for their development and strategic growth.

Strategic imperatives in cloud and colocation data center design

Cloud and colocation environments are accelerating AI infrastructure deployment for regulated, compute-intensive clients. In healthcare, for example, John Hopkins Hospital is using AI, to improve emergency department admissions, resulting in 38% faster bed assignments.[9] Visa's colocation-hosted fraud models analyzes 180 risk attributes, cutting false positives by 85%.[10] Recursion, a biotech firm, uses AI to process biological image data for drug discovery.[11] These examples point to a growing reliance on scalable, compliant, and high-performance cloud and colocation strategies tailored to AI workloads.

The shift is reshaping infrastructure requirements. Enterprises are moving away from traditional builds and toward rack-scale colocation deployments to support AI workloads. Systems like NVIDIA's GB300 NVL72 and Blackwell DGX SuperPODs are pushing power densities beyond 120kW per rack, requiring direct-to-chip liquid cooling with 4-to-make-3 power redundancy configurations. Leading colocation providers now deploy pre-engineered pods capable of delivering up to 2.6 MW of IT capacity and design modular blocks with integrated power and cooling to support AI clusters.[12] These strategies enable faster tenant rollouts, optimized capacity usage, and reduced overprovisioning.

Hybrid shift underway

98% of enterprises are adopting a hybrid IT model, moving generative AI workloads from the public cloud to colocation environments.[13] Colocation facilities offering pre-engineered pods, integrated cooling, and rapid power onboarding can reduce deployment cycles by up to 50%. These capabilities allow infrastructure teams to compress time-to-service while mitigating design risks, such as thermal overload, uneven load distribution, and regulatory compliance gaps, especially in sectors where time-to-market is critical.



Security drives decisions

Security and compliance remain top priorities for colocation tenants deploying AI. Healthcare systems require HIPAA-compliant infrastructure with encrypted storage, private connectivity, and validated recovery protocols. Financial firms need low-latency, physically isolated racks to enable secure transaction processing and AI-driven fraud detection. Government workloads demand FedRAMP- and FISMA-certified environments with sovereign hosting. In response, colocation providers offer audit-ready infrastructure with isolated cages, fault-tolerant designs, Zero Trust segmentation, and support for sovereign zones and regional compliance mandates

Align for return

To maximize the return on AI infrastructure investments, cloud builders and colocation users need strategies that align with workload profiles, power and cooling needs, and long-term capacity planning. The table below outlines key deployment priorities, such as rack density, power delivery, cooling integration, and security protocols to help teams design scalable, efficient, and resilient AI infrastructure across cloud and colocation environments.

Strategic Imperatives in cloud and colocation services

Imperatives	Use cases	Data center requirement
Be transformative	• Deliver custom-built environments optimized for AI rack densities and liquid cooling • Enable scale without requiring on-prem expansion	• Cross-functional expertise across IT, facilities, and power teams • Purpose-built designs for density and advanced cooling
Be efficient	• Provide right-sized, high-efficiency capacity for unpredictable AI workload profiles • Maximize footprint utilization per rack	• Modular pods and liquid-ready capacity that scale with demand • Close-coupled cooling systems that reduce energy and space usage
Be first	• Enable faster go-live timelines for GPU clusters • Support rapid regional deployment and AI cloud expansion	• Pre-engineered infrastructure to reduce design and deployment cycles • Factory-integrated systems that simplify site installation
Be confident	• Offer secure, audit-ready infrastructure for regulated workloads • Maintain SLA-backed uptime for real-time AI services	• Dedicated, audit-ready spaces with physical isolation options • Integrated monitoring and control for power, cooling, and network availability
Be future-ready	• Support evolving AI form factors and mixed training/inference models • Integrate with public cloud and orchestration platforms	• Interoperable and upgradeable systems that accelerate deployment and simplify expansion across regions or workloads • Scalable designs that support next-gen AI infrastructure, from GPU-dense racks to liquid-cooled clusters

Table 3. Examples in reimagining a data center infrastructure design with AI use cases and the Vertiv strategic imperatives

Streamlining the path to AI adoption and growth

Rethinking infrastructure design – principles for success

It isn't just legacy data center technologies that need to evolve to meet the requirements of AI. The design strategies and processes that have been employed for decades also need to be updated. To meet these challenges, Vertiv has developed AI-specific design principles to meet new workload and density requirements:

- **Design power and cooling holistically:**
  A holistic approach to infrastructure is required to meet AI's simultaneous power and cooling demands. By employing highly efficient integrated technologies like direct-to-chip liquid cooling alongside advanced power infrastructure, holistically designed solutions enhance overall efficiency, enable scalability, and validate that AI workloads are not throttled or slowed by infrastructure limitations. Management and control systems, including remote management capabilities, can help improve efficiency and resiliency and provide an integrated 'single pane of glass' for holistic management. Out-of-band management of AI platforms is required to provide continuous access to AI infrastructure if the production network goes offline, enabling remote troubleshooting and issue resolution. It also enhances security by isolating management traffic from production networks to reduce the attack surface.
- **Make effective use of available power:**
  AI is projected to create unprecedented growth in data center power consumption. AI racks must use every watt of power as efficiently as possible, necessitating designs that eliminate stranded power by aligning AI clusters to data center capacity blocks and leveraging the latest advances in equipment efficiency. Real-time monitoring and optimization of power distribution through out-of-band management helps eliminate inefficiencies and optimize resource usage.
- **Balance TCO, redundancy, and blast radius:**
  Maximizing the value of AI infrastructure requires a careful analysis of total costs, redundancy, and the potential scope of damage that could occur in the event of a failure (blast radius). Achieving the proper balance optimizes capital investment, risk management, scalability, and reliability. In the event of system failures, remote out-of-band management can reduce recovery times from hours to minutes.
- **Prepare for AI workload surges:**
  AI workloads can have significant variances in their resource requirements, leading to dynamic computing demand. Infrastructure must be designed to accommodate dynamic workloads through buffer capacity and the use of advanced system-level controls.
- **Leverage liquid- and air-cooling technologies:**
  Combining liquid and air cooling technologies allows the strengths of each technology to complement each other. This results in a solution that is flexible to address varying cooling needs across workloads and is efficient and scalable.
- **Design for the future:**
  Anyone designing for AI today must have an eye on the future. While AI is delivering value across multiple industries, it is still in the earliest phases. Vertiv is shaping a new era in which the computing capacity of a 1 MW data center will be packed into a single rack.

Vertiv™ 360AI: Turning principles into solutions

These principles form the foundation for Vertiv™ 360AI, a comprehensive portfolio of infrastructure solutions that support the demands of AI and high-performance computing workloads. The portfolio offers a streamlined approach to powering and cooling AI systems of various sizes, from edge inferencing to full AI data centers. Bringing together power, cooling, enclosures and structures, digitized management, and lifecycle service into holistic solutions, Vertiv 360AI solutions are designed to streamline interoperability, reduce deployment times, and optimize space and efficiency.

With Vertiv 360AI, data center directors and managers have the flexibility to take a phased approach to AI support by implementing liquid cooling technologies that work with existing air-cooling systems. They can also take a holistic and streamlined approach to retrofits and new builds using Vertiv 360AI Reference Designs or prefabricated modular solutions.

Learn more about VERTIV. 360AI

AI-ready reference designs

Vertiv and its strategic AI partner NVIDIA have co-developed a library of reference designs that simplify and accelerate the deployment of AI racks, clusters, and factories. Using the reference designs, enterprises can shave months off deployment times and enable compatibility and interoperability across systems, reducing the cost and complexity of deploying AI. Vertiv has developed reference designs for retrofitting existing data centers with single-rack and multi-rack AI clusters and for virtually any size new data center.

Explore available reference designs



Vertiv™ CoolChip CDU | 70-1350 kW

Figure 2. Vertiv Reference Designs are available for single-rack and multi-rack data center retrofits, and for new data center builds.

AI-ready prefabricated modular (PFM) solutions

The Vertiv™ 360AI portfolio also includes prefabricated, modular data center solutions that can be scaled to meet the requirements of any AI initiative and be configured to support the platforms of leading AI computing providers.

By bringing together the quality and process efficiency enabled by offsite fabrication with best-in-class AI-ready technologies, prefabricated solutions can reduce the time to deploy AI infrastructure by up to 50%. Solutions are designed to support high-density computing with integrated liquid cooling infrastructure and high-efficiency power protection and distribution. Data center modules arrive onsite ready for assembly in prefabricated units, including the building enclosure and all building systems.

Prefabricated solutions provide single-source accountability from initial consultation through configuration, fabrication, installation, commissioning, and lifecycle services. All aspects of solution development are managed to streamline processes, reduce customer resource demands, and provide cost and schedule efficiencies.

Selecting the right partner

Across industries, AI is creating opportunities to accelerate innovation, improve customer experiences, increase speed-to-market, and enhance organizational efficiency. Choosing the right infrastructure partner is an essential step in addressing the power and cooling complexities introduced by AI.

Vertiv is a recognized leader in critical infrastructure, with proven expertise in high-performance computing. As the market share leader in data center thermal management and large power protection and distribution systems, it offers the most complete portfolio of power, cooling, and service solutions specifically designed to address the challenges arising from accelerated compute. The Vertiv™ 360AI suite includes end-to-end solutions for retrofits and new builds of any scale, from rack solutions to entire prefabricated modular data centers.

Vertiv is co-developing solutions with partners like Intel and NVIDIA to remove the barriers to the adoption of AI and other accelerated compute applications, and leads the industry in addressing the unique architectural requirements of a new breed of AI factories.

References

1. Singla, A., Sukharevsky, A., Yee, L., Chui, M., & Hall, B. (2024). The state of AI in early 2024: Gen AI adoption spikes and starts to generate value. QuantumBlack, AI by McKinsey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
2. Lawrence, A. (2024). AI: Enterprises are active but cautious. Uptime Intelligence. https://intelligence.uptimeinstitute.com/resource/ai-enterprises-are-active-cautious
3. Deloitte. (2024). New Deloitte survey finds expectations for Gen AI remain high, but many are feeling pressure to quickly realize value while managing risks. https://www.deloitte.com/global/en/about/press-room/gen-ai-survey.html
4. Vertiv. (2024). Vertiv's Nvidia partnership leads to breakthroughs in technology that support AI's future. https://www.vertiv.com/en-us/solutions/ai-hub/vertivs-nvidia-partnership-leads-to-breakthroughs-in-technology-that-support-ais-future/
5. Heimes, H. (2024). Gen AI in corporate functions: Looking beyond efficiency gains. McKinsey. https://www.mckinsey.com/capabilities/operations/our-insights/gen-ai-in-corporate-functions-looking-beyond-efficiency-gains
6. Rowan, J., Ammanath, B., Perricos, C., Sniderman, B., & Jarvis, D. (2024). Now decides next: Moving from potential to performance. Deloitte. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-state-of-gen-ai-q3.pdf
7. Merritt, R. (2021). What Is a Pod? What Is a Cluster? Nvidia Corp. https://blogs.nvidia.com/blog/ what-is-a-cluster-pod/
8. Vertiv. (2024). Vertiv™ 360AI High Density Reference Design for AI (AI Reference Design #020). https://www.vertiv.com/4a91c0/globalassets/campaigns/ai-hub/vertiv-reference-design-020-ds-en-na-2024-gr-00109-web.pdf
9. Keating, M. (2021). Capacity command center celebrates 5 years of improving patient safety, access. Johns Hopkins Medicine. https://www.hopkinsmedicine.org/news/articles/2021/03/capacity-command-center-celebrates-5-years-of-improving-patient-safety-access
10. Visa. (2024). Visa announces Generative AI-powered fraud solution to combat account attacks. https://usa.visa.com/about-visa/newsroom/press-releases.releaseId.20661.html
11. Lin, F. (2024). Recursion's fast-track road to therapeutics using AI-based maps of biology. Genetic Engineering and Biotechnology News. https://www.genengnews.com/topics/artificial-intelligence/recursions-fast-track-road-to-therapeutics-using-ai-based-maps-of-biology/
12. Vertiv. (2025). Vertiv Reference Design 006. https://www.vertiv.com/494fff/globalassets/documents/ai-hub-reference-design/vertiv-360ai-reference-design-006.pdf
13. Kerner, S. M. (2024). Data Center Trends: Industry Report Reveals Shift Towards Hybrid IT, Colocation. Data Center Knowledge. https://www.datacenterknowledge.com/cloud/data-center-trends-industry-report-reveals-shift-towards-hybrid-it-colocation



Vertiv.com | **Vertiv Headquarters, 505 N Cleveland Ave, Westerville, OH 43082, USA**

© 2025 Vertiv Group Corp. All rights reserved. Vertiv™ and the Vertiv logo are trademarks or registered trademarks of Vertiv Group Corp. All other names and logos referred to are trade names, trademarks or registered trademarks of their respective owners. While every precaution has been taken to ensure accuracy and completeness here, Vertiv Group Corp. assumes no responsibility, and disclaims all liability, for damages resulting from use of this information or for any errors or omissions. Specifications, rebates and other promotional offers are subject to change at Vertiv's sole discretion upon notice.

SL-80076 Rev.1-07/25