ISG Provider Lens®

AI-ready Infrastructure Solutions

A report comparing vendor strengths and differences across AI-ready Infrastructure Solutions

Customized report courtesy of:

T Systems

QUADRANT REPORT | JULY 2026 | GLOBAL

Table of Contents

Report Author: Sonam Chawla

AI-ready infrastructure leadership now hinges on governed scale, cost discipline and operations

This ISG Provider Lens® AI-ready Infrastructure Solutions study evaluates providers' evolving capability landscape worldwide for 2026 as enterprises scale AI from experimentation into sustained operations. It examines how infrastructure models are adapting to support AI training, fine-tuning and inference across on-premises, hosted and distributed environments. The research covers the following two quadrants: Integrated AI Infrastructure Systems and GPU as a Service (GPUaaS), with a focus on infrastructure architecture, delivery models and operational readiness required to support production-grade AI at scale.

Market Context

Global demand for AI infrastructure is changing rapidly and unevenly. Different AI activities, such as model training, fine-tuning and inference, are growing at different rates, creating uneven pressure on compute, storage and networks. Unlike earlier cloud adoption, AI exacerbates these imbalances. Access to GPUs, power availability and network performance are no longer minor considerations; they now directly determine whether AI systems can be built and scaled. This pressure is felt most strongly by large multinational companies, which must secure limited AI capacity across multiple regions while maintaining consistent operations.

These challenges are compounded by growing energy and sustainability constraints. AI systems consume large amounts of power and require advanced cooling, while enterprises are expected to track and report their environmental impact. As a result, AI infrastructure can only scale where power, cooling and sustainability requirements are met. This means that even regions with strong cloud infrastructure may not be ready

AI infrastructure is shifting from capacity planning to control and repeatable economics.

for large-scale AI unless energy systems are in place. Infrastructure decisions, therefore, depend more on local energy policies and long-term capacity planning, not just on cost or performance.

To manage uncertainty and changing demand, many enterprises are turning to consumption-based infrastructure models. However, AI makes this more complex than in the past. AI usage often spikes during experiments, then stabilizes in production, with frequent model updates. Without strong governance, costs can rise quickly and become hard to predict. For global organizations, these challenges are amplified by regional data protection and compliance variations. This increases the need for infrastructure models that are consistent across regions yet flexible enough to meet local requirements.

At the same time, AI infrastructure readiness varies significantly by region. North America and parts of East Asia have better access to advanced chips and large cloud ecosystems, while other regions still face shortages in supply, power or skilled talent. As a result, global enterprises cannot rely on a single deployment model. They often need a mix of centralized AI platforms and localized execution close to data sources or regulatory boundaries.

Together, these factors indicate that enterprises must design AI infrastructure with flexibility. The quadrant results in this study are intended to showcase provider capabilities, not to serve as fixed models for replication. AI-ready infrastructure has moved from being a supporting layer to becoming a key limiting factor in AI success. As AI is used in regulated and business-critical areas, infrastructure choices now affect long-term cost, reliability and control. This study is timely because it evaluates how providers are addressing these challenges and supporting stable, production-ready AI operations rather than short-term experiments driven only by access to capacity.

Enterprise Priorities

As AI moves from experimentation into real business use, enterprises are becoming more selective and disciplined in their infrastructure decisions. Early efforts focused mainly on gaining access to AI capacity. The initial urgency to get GPUs has been replaced by a need to stabilize utilization, control spend leakage and align infrastructure capacity with AI lifecycle stages. Today, the focus has shifted to developing infrastructure that can be managed reliably, scaled safely and controlled financially over time. Enterprises no longer treat AI infrastructure as a temporary foundation for pilots, but as a long-term operating platform for business-critical workloads.

Cost predictability has become a central concern. Many enterprises have learned that AI infrastructure costs can rise quickly due to low utilization, idle capacity or poorly managed experimentation. As a result, buyers now prioritize infrastructure models that provide clear visibility into usage and spend. They increasingly expect cost controls, chargeback mechanisms and predictable pricing behavior across development and production environments, rather than relying on best-effort estimates.

Operational stability is another growing priority. As AI systems are embedded into core business processes, enterprises expect the same levels of reliability they require from other critical platforms. This includes tested recovery procedures, predictable performance under load and clear responsibility models for ongoing operations. Infrastructure that cannot demonstrate production-grade resilience is less attractive, even if it offers high performance or rapid initial deployment.

Governance is also rising on the enterprise agenda. Organizations want AI infrastructure that supports security, compliance and audit needs by design, not through manual workarounds. Buyers increasingly evaluate whether access controls, data-handling rules and monitoring capabilities are built into the infrastructure layer itself. For global enterprises, consistency across regions is critical, as fragmented governance models increase risk and management effort.

Buying behavior is changing. Enterprises are more cautious about large upfront commitments and more demanding during evaluations. They rely more heavily on proof-of-capability exercises, production-like pilots and clear operating documentation. Decision-making often starts centrally but now includes regional validation to ensure

infrastructure models operate under local power, data and regulatory conditions.

Overall, enterprise priorities are shifting toward infrastructure that can support AI at scale in a controlled, predictable and sustainable way, rather than solutions optimized only for rapid deployment or short-term access to capacity.

Key enterprise priorities emerging across global buyers include:

- Infrastructure that unifies AI experimentation and production under a single economic and governance framework to reduce friction between teams
- Demonstrable utilization controls and financial transparency over headline performance metrics
- Operational resilience equivalent to traditional enterprise-critical platforms
- Governance embedded in the infrastructure layer, not as an add-on control layer

Provider Dynamics

Providers are evolving their infrastructure offerings as enterprise expectations shift from short-term AI experimentation to long-term, production use. Market success increasingly depends on how well providers help enterprises manage complexity, control costs and operate AI environments reliably over time, rather than simply offering access to advanced infrastructure.

Integrated AI Infrastructure Systems

In this quadrant, providers focus on simplifying how AI infrastructure is built and operated. Leading providers offer integrated systems that bring together compute, storage, networking and management tools into a single, coordinated setup, enabling rapid deployment of AI environments and consistent operations across locations. Instead of relying on custom configurations for each deployment, leading providers emphasize standard architectures that can be reused and scaled.

A key differentiator is the depth of built-in automation and operational support. More advanced offerings include automated provisioning, monitoring and lifecycle management to accelerate the move from pilots to production and to update or expand AI environments without disrupting existing workloads. Less mature providers rely more on manual integration and customer-managed processes, increasing operational risk and slowing scaling.

Governance and resilience are also key differentiators. Leading providers embed security, access controls and compliance directly into the infrastructure, rather than treating them as add-on features. They can demonstrate how AI systems are monitored, controlled and recovered in real production scenarios. Providers that lack such depth often struggle to meet enterprise requirements as AI systems become business-critical.

GPU as a Service (GPUaaS)

In the GPUaaS quadrant, competition has shifted beyond simply offering GPU capacity. While access remains important, enterprises now prioritize how efficiently that capacity is managed. Leading providers enable customers to manage demand fluctuations more smoothly through intelligent scheduling, shared capacity models and improved visibility into usage and performance, reducing waste and improving cost predictability.

Service reliability is a major differentiator. Providers with mature offerings demonstrate strong operational support, clear service commitments and proactive capacity planning. This is especially important as enterprises run production workloads that cannot tolerate frequent disruptions. Less mature GPUaaS models often resemble basic cloud services, with limited support for sustained, high-intensity AI workloads.

Cost transparency is becoming a deciding factor. Leading providers offer clear insights into how costs change across training, fine-tuning and inference stages, helping enterprises manage spending more effectively. Providers that cannot explain or control cost behavior face rising skepticism, even if their performance is attractive.

Across both quadrants, a recurring gap is the lack of alignment between infrastructure delivery and enterprise operating needs. Providers that can demonstrate standardized operations, strong governance and predictable economics are better positioned as enterprises look to scale AI responsibly and sustainably across regions.

Outlook

Over the next one to two years, AI infrastructure will continue to move toward more stable, controlled and repeatable operating models. As AI becomes part of daily business operations, enterprises will prioritize reliability, governance and long-term cost control over rapid capacity expansion. Infrastructure choices made now will increasingly shape how successfully organizations can scale AI without incurring operational or financial risk.

For enterprises, the biggest challenge will be avoiding infrastructure decisions that suit early pilots but break down at scale. Models that lack strong usage controls, cost visibility or recovery planning may seem attractive in the short term but can become difficult and expensive to fix later. Enterprises should treat AI infrastructure as a core platform, with clear ownership, operating procedures and performance expectations.

Providers, in turn, will face growing pressure to prove operational maturity, not just technical capability. As client expectations rise, gaps in automation, governance and support readiness will become more visible. Providers that focus only on expanding capacity, without improving how AI infrastructure is operated and governed, risk falling behind as AI workloads become business-critical.

Looking ahead, leadership in AI-ready infrastructure will increasingly be defined by the ability to deliver consistent operations across regions, predictable cost behavior and built in governance. Enterprises should monitor how providers invest in standardization, automation and lifecycle management, while providers must continue to evolve their offerings to support AI as a long-term operating environment rather than a short-term performance race.

Successful AI infrastructures enforce production discipline early through standardized architectures, auditable controls and predictable economics, treating AI platforms as long-term operational assets rather than experimental capacity pools.

Provider Positioning

Page 1 of 4

	Integrated AI Infrastructure Systems	GPU as a Service (GPUaaS)
Advantech	Contender	Not In
Alibaba Cloud	Not In	Market Challenger
ASRock Rack	Contender	Not In
Asus	Leader	Not In
AWS	Not In	Leader
Bull	Product Challenger	Not In
Cirrascale	Not In	Product Challenger
Cisco	Leader	Not In
CoreWeave	Not In	Leader
Crusoe Cloud	Not In	Leader
Dell Technologies	Leader	Not In

Provider Positioning

Page 2 of 4

	Integrated AI Infrastructure Systems	GPU as a Service (GPUaaS)
DigitalOcean	Not In	Contender
Exoscale	Not In	Product Challenger
FluidStack	Not In	Product Challenger
Fujitsu	Contender	Not In
Gcore	Not In	Rising Star ★
Gigabyte	Product Challenger	Not In
Google Cloud	Not In	Leader
H3C	Product Challenger	Not In
HPE	Leader	Not In
Huawei	Product Challenger	Not In
Hyperstack	Not In	Contender

Provider Positioning

Page 3 of 4

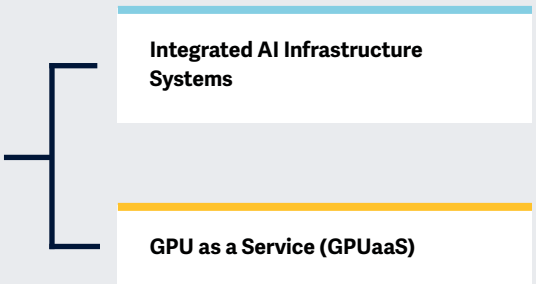
	Integrated AI Infrastructure Systems	GPU as a Service (GPUaaS)
IBM	Not In	Leader
Inspur	Product Challenger	Not In
Lambda	Not In	Leader
Lenovo	Leader	Not In
Microsoft	Not In	Leader
MiTAC	Product Challenger	Not In
Nebius	Not In	Leader
Oracle	Not In	Leader
OVHcloud	Not In	Rising Star ★
Runpod	Not In	Product Challenger
Scaleway	Not In	Contender

Provider Positioning

Page 4 of 4

	Integrated AI Infrastructure Systems	GPU as a Service (GPUaaS)
Sify Technologies	Not In	Product Challenger
Sugon	Contender	Not In
Supermicro	Leader	Not In
Tata Communications	Not In	Product Challenger
Together AI	Not In	Product Challenger
T-Systems	Not In	Leader
UnitedLayer	Not In	Product Challenger
Vultr	Not In	Contender

This study focuses on what **ISG perceives** as **most critical** in 2026 for **AI-ready Infrastructure Solutions**



Simplified Illustration Source: ISG 2026

Definition

AI has moved from experimentation to an operational backbone, redefining enterprise needs for compute, storage and networking. As AI models grow in scale, complexity and autonomy, infrastructure has become a key constraint and a differentiator of AI success. Enterprises are no longer debating whether to invest in AI, but how to build scalable, efficient, economically viable foundations for training, inference and deployment.

The market is shifting beyond general-purpose infrastructure toward solutions that maximize performance, energy efficiency and ease of management. Integrated AI infrastructure systems simplify deployment, orchestration and lifecycle management of complex AI stacks across on-premises, hosted and edge environments.

Consumption-based models, such as GPU as a service (GPUaaS), are gaining traction, allowing enterprises to access high-performance accelerators without long-term capital commitments while addressing demand volatility, skills shortages and capacity constraints. These models also introduce new considerations around pricing transparency, SLAs, data sovereignty and operational governance.

At the same time, enterprises face rising pressure to balance performance with cost clarity, power consumption, compliance and resilience. Fragmented vendor offerings, architectures and delivery models complicate buyers' ability to select solutions aligned with their AI strategies, maturity levels and regulatory environments. The ISG Provider Lens® AI-ready Infrastructure Solutions study helps address these challenges by evaluating providers across two interrelated quadrants — Integrated AI Infrastructure Systems and GPU as a Service (GPUaaS). The study offers a structured, standardized comparison of vendor capabilities, GTM models and enterprise relevance, enabling decision-makers to identify the most suitable partners for building robust, future-proof AI infrastructure foundations.

Scope of the Report

This ISG Provider Lens® quadrant report covers the following two quadrants for services/ solutions: Integrated AI Infrastructure Systems and GPU as a Service (GPUaaS)

This ISG Provider Lens® study offers business and IT decision-makers:

- Transparency on the strengths and weaknesses of relevant providers
- A differentiated positioning of providers by segments (quadrants)
- Focus on the global market

Our study serves as the basis for important decision-making by covering providers' positioning, key relationships and go-to-market considerations. ISG advisors and enterprise clients also use information from these reports to evaluate their existing vendor relationships and potential engagements.

Provider Classifications

The provider position reflects the suitability of providers for a defined market segment (quadrant). Without further additions, the position always applies to all company sizes classes and industries. In case the service requirements from enterprise customers differ and the spectrum of providers operating in the local market is sufficiently wide, a further differentiation of the providers by performance is made according to the target group for products and services. In doing so, ISG either considers the industry requirements or the number of employees, as well as the corporate structures of customers and positions providers according to their focus area. As a result, ISG differentiates them, if necessary, into two client target groups that are defined as follows:

- **Midmarket:** Companies with 100 to 4,999 employees or revenues between $20 million and $999 million with central headquarters in the respective country, usually privately owned.
- **Large Accounts:** Multinational companies with more than 5,000 employees or revenue above $1 billion, with activities worldwide and globally distributed decision-making structures.

The ISG Provider Lens® quadrants are created using an evaluation matrix containing four segments (Leader, Product & Market Challenger and Contender), and the providers are positioned accordingly. Each ISG Provider Lens® quadrant may include a service provider(s) which ISG believes has strong potential to move into the Leader quadrant. This type of provider can be classified as a Rising Star.

- **Number of providers in each quadrant:** ISG rates and positions the most relevant providers according to the scope of the report for each quadrant and limits the maximum of providers per quadrant to 25 (exceptions are possible).

Provider Classifications: Quadrant Key

Product Challengers offer a product and service portfolio that reflect excellent service and technology stacks. These providers and vendors deliver an unmatched broad and deep range of capabilities. They show evidence of investing to enhance their market presence and competitive strengths.

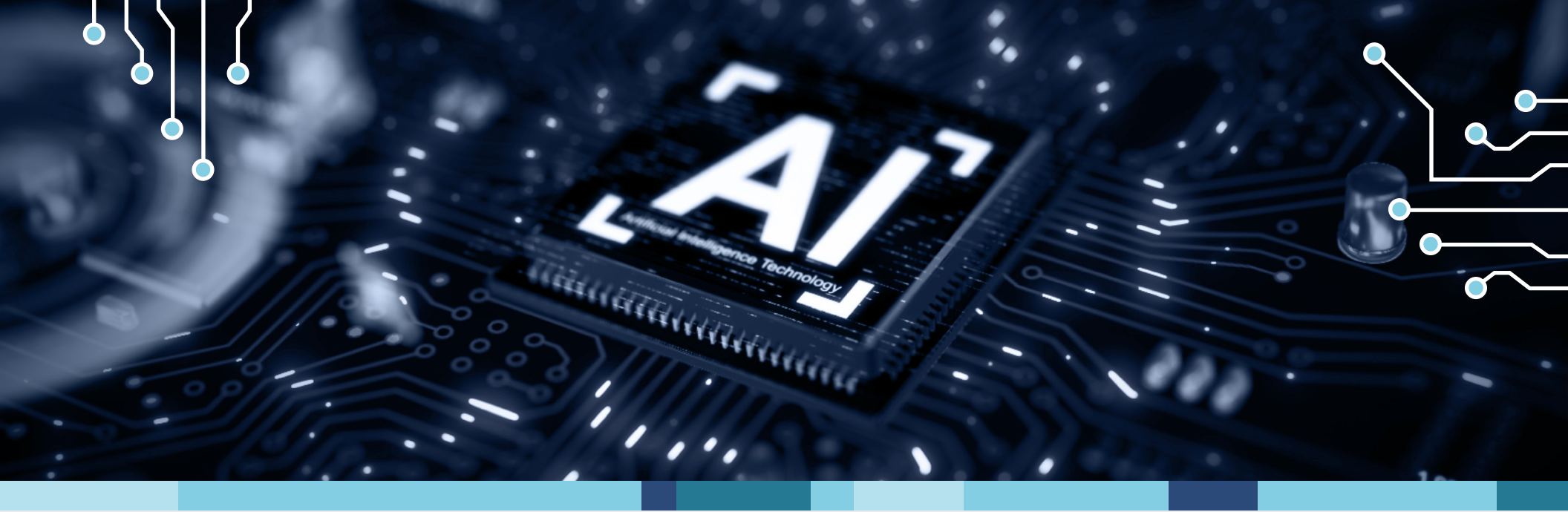
Leaders have a comprehensive product and service offering, a strong market presence and established competitive position. The product portfolios and competitive strategies of Leaders are strongly positioned to win business in the markets covered by the study. The Leaders also represent innovative strength and competitive stability.

Contenders offer services and products meeting the evaluation criteria that qualifies them to be included in the IPL quadrant. These promising service providers or vendors show evidence of rapidly investing in products/ services and follow sensible market approach with a goal of becoming a Product or Market Challenger within 12 to 18 months.

Market Challengers have a strong presence in the market and offer a significant edge over other vendors and providers based on competitive strength. Often, Market Challengers are the established and well-known vendors in the regions or vertical markets covered in the study.

★ **Rising Stars** have promising portfolios or the market experience to become a Leader, including the required roadmap and adequate focus on key market trends and customer requirements. Rising Stars also have excellent management and understanding of the local market in the studied region. These vendors and service providers give evidence of significant progress toward their goals in the last 12 months. ISG expects Rising Stars to reach the Leader quadrant within the next 12 to 24 months if they continue their delivery of above-average market impact and strength of innovation.

Not in means the service provider or vendor was not included in this quadrant. Among the possible reasons for this designation: ISG could not obtain enough information to position the company; the company does not provide the relevant service or solution as defined for each quadrant of a study; or the company did not meet the eligibility criteria for the study quadrant. Omission from the quadrant does not imply that the service provider or vendor does not offer or plan to offer this service or solution.



Integrated AI Infrastructure Systems

Who Should Read This Section

This report is valuable for providers offering integrated AI infrastructure systems globally to understand their market position and for enterprises looking to evaluate these providers. Providers are assessed on their ability to deliver proprietary, rack-scale AI systems with end-to-end responsibility across compute, storage, networking and the AI lifecycle. Enterprises engage them for validated platforms that support development, deployment and inference across on premises, hybrid and edge environments, enabling consistent performance and governed, enterprise-wide AI deployment.

CIOs and CTOs

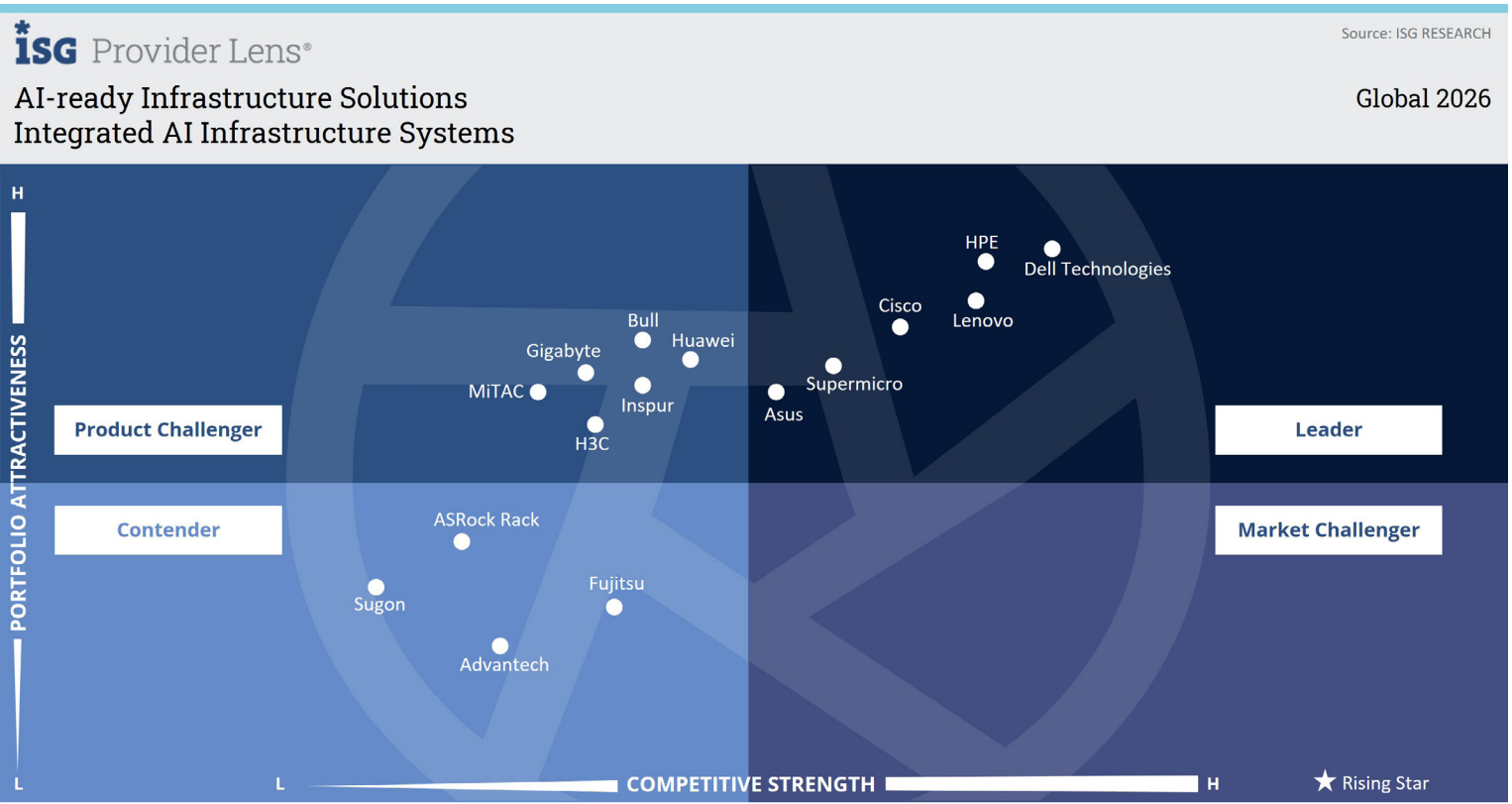
Can use this report to define their enterprise AI infrastructure vision and evaluate platform originators that are capable of supporting large-scale AI and GenAI workloads. The insights from the report help them assess architectural coherence, vendor maturity and roadmap credibility, enabling long-term scalability, governance and return on AI infrastructure investments.

VPs and heads of infrastructure and data center operations

Should read this report to evaluate providers delivering integrated AI infrastructure platforms with end-to-end system accountability. By understanding provider strengths in integration, orchestration, security governance and global delivery, infrastructure leaders can select solutions that reduce operational complexity, ensure predictable performance and support evolving AI workloads without fragmenting the data center environment.

Head of AI platforms and AI engineering

Will gain value from this report as they enable scalable, production-grade AI development across the enterprise. The insights help them identify providers [SB1.1][MR1.2]with validated, enterprise-scale systems for advanced training, inference and GenAI workloads, accelerating experimentation, improving reliable deployment and aligning AI innovation with infrastructure capabilities.



This quadrant evaluates vendors delivering **integrated rack-scale AI platforms,** focusing on system-level design, deployment flexibility and the ability to provide consistent, **enterprise-ready AI infrastructure** across on-premises and edge environments.

Sonam Chawla

Definition

This quadrant evaluates platform originators that design, manufacture and ship proprietary rack-scale AI systems with system-level responsibility. The integrated enterprise-scale AI infrastructure unifies compute, storage, networking and AI lifecycle enablement. They support model development, training, deployment and inference across on-premises, hybrid and edge environments. Evaluation centers on architectural coherence, modularity, orchestration, data pipeline efficiency, security governance and deployment flexibility. The quadrant reflects vendor maturity in delivering repeatable, validated AI platforms with consistent performance, compliance, global delivery and credible roadmaps for next-GenAI system designs. Demonstrated HPC implementations acts as a supplementary evidence of system scalability and engineering maturity when aligned to integrated AI platform architectures and enterprise AI workloads.

This quadrant excludes vendors that provide only component hardware or unmanaged infrastructure without system-level responsibility.

Eligibility Criteria

1. **Offer commercially available, integrated AI infrastructure systems** that combine validated compute accelerators, high-performance storage, networking and aligned lifecycle management capabilities — demonstrated HPC maturity is a plus
2. **Provide orchestration and automation capabilities,** either natively or through certified partner integrations, to support AI workload scheduling, monitoring and cluster management
3. **Demonstrate production-grade deployments,** backed by customer references, benchmarked AI performance and evidence of operational use
4. **Support flexible deployment models,** including on-premises, hybrid and edge environments, with unified or federated management
5. **Meet enterprise security and compliance standards** with industry certifications, such as ISO 27001 and SOC 2 Type II, and built-in controls for encryption, access management and audit logging
6. **Deliver managed or fully supported AI platforms** with system-level responsibility
7. **Maintain ecosystem partnerships and support capabilities,** including with silicon vendors, software providers and global services organizations
8. **Provide documented use cases and references,** demonstrating applicability across industries, including regulated environments

Observations

The Integrated AI Infrastructure Systems quadrant reflects the increasing formalization of rack-scale and cluster-scale AI architectures into enterprise-oriented offerings. While rack-scale computing and AI factory concepts have long existed in hyperscale and high-performance computing (HPC) environments, infrastructure vendors are now packaging these capabilities as pre-integrated systems that combine accelerated compute, high-performance networking, storage and system-level responsibility for enterprise AI workloads. This shift is driven by rising model complexity, increased power density and the need for scalable, production-grade infrastructure for large-scale training and inference.

In response, some providers emphasize vertically integrated, platform-led AI systems with clear ownership and lifecycle alignment, while others focus on engineered hardware systems, modular building blocks or performance-optimized designs closely aligned to accelerator ecosystems.

Network-centric architectures and validated AI POD designs have also gained prominence, highlighting the critical role of fabric performance and architectural coherence in scaling distributed AI workloads across racks and clusters. These approaches reflect differing interpretations of system-level responsibility, ranging from single-vendor accountability to partner-centric, ecosystem-led models.

Market activity is driven primarily by ecosystem collaboration rather than large-scale acquisitions. Co-engineering efforts among infrastructure vendors, accelerator providers and power and cooling specialists are accelerating the delivery of rack-scale AI systems and supporting higher densities.

From the 20 companies assessed for this study, 16 qualified for this quadrant, with six being Leaders.

ASUS

ASUS demonstrates strong capability in delivering performance-optimized, rack-scale AI infrastructure through its ESC GPU server portfolio and AI Factory designs, with expertise in GPU density and liquid-cooling integration for training-intensive workloads.

Cisco

Cisco demonstrates strong relevance in integrated AI infrastructure through its network-centric AI system designs, combining high-performance networking, validated AI reference architectures and ecosystem partnerships to support scalable, rack-scale AI deployments.

Dell Technologies

Dell Technologies delivers end-to-end AI services, from strategy, data preparation, to implementation, deployment/testing, to operation/scale, closing gaps in data readiness, integration, performance and governance to turn infrastructure into operational AI.

Hewlett Packard Enterprise

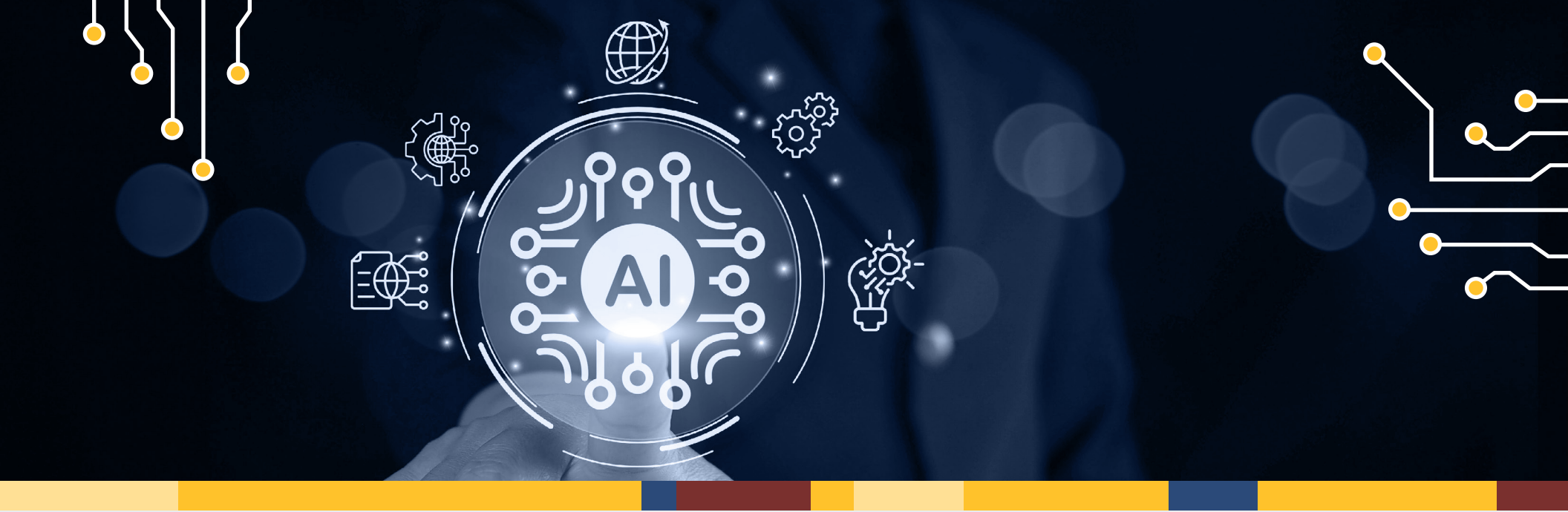
HPE demonstrates strong leadership in integrated AI infrastructure through its HPC-derived AI systems, combining advanced compute, high-performance interconnects and lifecycle services to support large-scale, production-grade AI workloads.

Lenovo

Lenovo demonstrates strong capability in delivering pre-configured, rack-scale AI systems through its ThinkSystem-based rack-scale AI and Secure AI Factory offerings. It also combines accelerated compute, integrated infrastructure and security-by-design to support enterprise AI deployment.

Supermicro

Supermicro demonstrates strong capability in delivering modular, rack-scale AI infrastructure systems through its GPU-optimized server portfolio and building-block architecture, enabling rapid deployment of high-density AI platforms.



GPU as a Service (GPUaaS)

Who Should Read This Section

This report is valuable for providers offering GPUaaS globally to understand their market position and for enterprises looking to evaluate these providers. Providers are assessed on their ability to deliver elastic, on-demand access to high-performance GPU compute for AI, ML and data-intensive workloads across public, private and hybrid environments. Enterprises use these services to consume modern GPU architectures at scale, avoiding upfront investment while supporting use cases from prototyping to distributed training with predictable performance, flexible consumption and MLOps integration.

Technology and digital leaders, including CIOs and CTOs

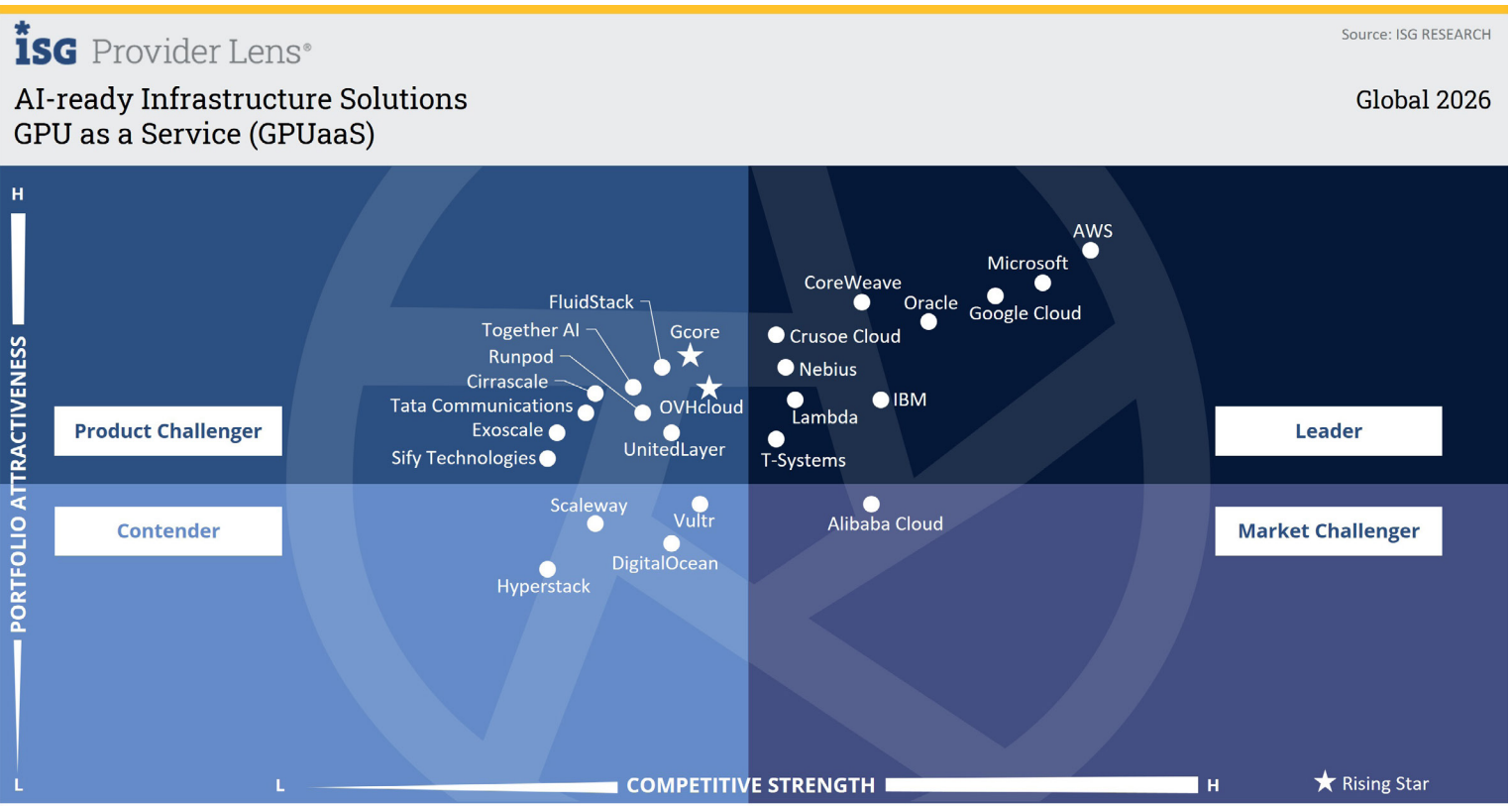
Should read this report to evaluate GPUaaS providers that offer rapid, flexible access to high-performance AI compute without infrastructure ownership. The insights help leaders assess enterprise-grade GPU platforms, consumption models and vendor maturity, supporting informed decisions that accelerate AI adoption while maintaining cost control, governance and alignment with cloud and hybrid strategies.

Chief data officers (CDOs)

Will benefit from this report as it ensures data and analytics teams have timely access to scalable GPU resources for advanced analytics and AI initiatives. The report highlights GPUaaS providers delivering reliable, on-demand performance with strong security and governance, enabling evolving data pipelines and AI workloads without infrastructure bottlenecks or lengthy procurement cycles.

Heads of AI and ML

Should read this report to identify GPUaaS providers that support rapid experimentation and scalable AI workloads. As teams focus on shortening development cycles and operationalize models faster, the report can be valuable as it evaluates platforms offering elastic GPU access, deployment flexibility and consistent performance across cloud and hybrid environments, enabling faster insights, higher productivity and enterprise-grade AI execution.



This quadrant assesses the evolving **GPUaaS landscape,** comparing providers that enable **on-demand, scalable GPU** compute for AI and ML workloads and highlighting differentiation across **enterprise readiness orchestration depth and speed-to-GPU.**

Sonam Chawla

GPU as a Service (GPUaaS)

Definition

This quadrant evaluates infrastructure providers delivering GPUaaS, including elastic, on-demand access to high-performance GPU compute for AI, ML and data-intensive workloads. Providers deliver on-demand access to modern GPU architectures through public, private or hybrid environments, allowing enterprises to avoid capital investment in dedicated hardware. They support flexible workloads, from rapid prototyping to large-scale distributed training, through virtualization, cluster orchestration and multi-tenant capabilities. Evaluation focuses on service breadth, performance consistency, orchestration maturity, AI development and MLOps integration, pricing and consumption flexibility and security. Enterprise relevance is determined by a provider's ability to deliver predictable performance, transparent commercial models and reliable access to modern GPU architectures at scale.

The quadrant reflects enterprises' preference to consume AI infrastructure as a scalable service rather than own it and excludes vendors offering only generic compute or storage services without dedicated, AI-oriented GPU capabilities and proven enterprise deployments.

Eligibility Criteria

1. **Provide commercially available GPUaaS offerings** to external enterprise or professional customers, with the provider owning or operating the underlying GPU infrastructure
2. **Support flexible consumption models,** including on-demand usage, reserved capacity and burst or short-term access
3. **Provide streamlined dashboards,** with one-click deployment and integration to popular AI tools enterprises already use for model development and management
4. **Demonstrate geographic availability and data residency options** aligned to enterprise latency, sovereignty and regulatory requirements
5. **Meet enterprise security and compliance expectations,** with encryption, tenant isolation and recognized certifications, such as ISO 27001 and SOC 2, where applicable
6. **Present a compelling portfolio of customer success,** including case studies across multiple industries with measurable AI outcomes, complemented by responsive 24/7 support teams
7. **Demonstrate transparent sustainability and efficiency practices,** including energy efficiency initiatives, renewable energy sourcing and reporting on infrastructure footprint

Observations

The GPUaaS market continues to gain strategic importance as organizations accelerate investments in GenAI, large-scale model training and inference-driven workloads. Demand remains shaped by the need for flexible access to high-performance computing (HPC), ongoing constraints in advanced GPU availability and the growing diversity of enterprise AI use cases across industries.

The quadrant shows greater clarity in provider positioning. Hyperscale cloud providers continue to anchor enterprise and regulated workloads, leveraging mature global infrastructure, security frameworks and tightly integrated cloud services. In parallel, neocloud GPU providers have strengthened visibility by focusing on AI-native use cases, rapid provisioning cycles and early access to new GPU generations, particularly for GPU-centric and time-sensitive workloads.

Buyer behavior has evolved accordingly. Enterprises increasingly adopt multi-provider GPU strategies, selecting platforms based on workload intent, access speed, architectural flexibility and regional or sovereignty considerations, rather than relying on a single consumption model. This shift has contributed to a more segmented GPUaaS landscape, spanning enterprise-grade, sovereign and speed-optimized infrastructure approaches.

Selective M&A activity over the past year reflects a maturing market. Notably, neocloud providers have pursued targeted acquisitions to enhance AI-native capabilities beyond infrastructure. For example, CoreWeave's acquisition of Weights and Biases strengthened its observability and developer tooling ecosystem, signaling a broader move by specialized GPU providers toward higher-value software integration.

From the 35 companies assessed for this study, 25 qualified for this quadrant, with 10 being Leaders and two Rising Stars.

aws

AWS drives AI-ready infrastructure adoption through unmatched global GPU scale, deep NVIDIA and custom silicon integration and a full-stack offering across EC2, UltraClusters, networking and managed AI services, enabling secure, large-scale AI deployments.

CoreWeave

CoreWeave is a specialized AI cloud provider focused on large-scale training infrastructure, offering early access to advanced NVIDIA GPUs, deep orchestration via Kubernetes and Slurm and strong execution for hyperscale and frontier AI workloads.

Crusoe Cloud

Crusoe Cloud operates an energy-first AI infrastructure model, building purpose-designed AI data centers and a GPU cloud platform optimized for cost and power efficiency, primarily serving large-scale training and inference workloads.

Google Cloud

Google Cloud delivers a broad, end-to-end AI platform anchored by Vertex AI and deep data integration. This approach reduces operational complexity and supports scalable AI adoption, positioning it beyond raw GPU provisioning toward full lifecycle AI enablement.

IBM

IBM is an enterprise-focused AI infrastructure provider integrating GPU acceleration with the watsonx platform and a strong hybrid cloud foundation. Its emphasis on governance, sovereignty and deep enterprise integration supports trusted AI adoption in regulated industries.

Lambda

Lambda combines AI-native infrastructure design with enterprise-grade execution, delivering large, dedicated clusters and rapid access to new hardware. This balance supports both frontier research and production-scale AI training requirements.

Microsoft

Microsoft delivers GPUaaS as part of a full-stack AI platform, combining NVIDIA, AMD and custom silicon with Azure AI, Kubernetes and data services. Its global footprint, confidential computing and sovereign cloud options support enterprise-grade AI at scale.

Nebius

Nebius leads with AI-native infrastructure, Soperator/SUNK orchestration and sovereign European AI deployments. Long-term hyperscaler agreements and NVIDIA's backing validate its ability to deliver large-scale AI training and inference.

Oracle

Oracle provides high-performance GPU compute with NVIDIA and AMD accelerators, supporting large-scale training and inference. With extensive global regions and sovereign cloud models, it enables secure, scalable AI deployments across industries.

T-Systems

T-Systems demonstrates strong capability in production-grade GPU deployments. Its Industrial AI Cloud supports large GPU clusters based on NVIDIA reference architectures, optimized for high-performance AI training, digital twins and industrial simulation workloads.

Gcore (Rising Star)

Gcore (Rising Star) is an AI-focused cloud provider delivering GPUaaS across cloud, hybrid and edge via a unified control plane. It focuses on consistent performance, flexible deployment and compliance-ready designs, making it well-suited for enterprises running AI workloads.

OVHcloud

OVHcloud (Rising Star) is a European cloud provider focused on sovereign, privacy-first AI infrastructure. Its vertically integrated hardware, energy-efficient data centers and open-source stack make it well-suited for transparent, compliant AI deployments in regulated environments.

T-Systems

Leader

"T-Systems differentiates its GPUaaS portfolio through a sovereign, enterprise-grade approach combining scalable GPU capacity, strong compliance controls and integrated T Cloud platforms for production AI workloads in Europe."

Sonam Chawla

Overview

T-Systems is headquartered in Frankfurt, Germany. It has more than 25,000 employees across 26 countries. In FY25, the company generated €4.1 billion in revenue. T-Systems delivers GPUaaS through its T Cloud portfolio, combining T Cloud Public for on-demand GPU consumption with the Industrial AI Cloud for large-scale, production AI workloads. It emphasizes sovereign, EU-based infrastructure, compliance with European regulatory frameworks and integration with enterprise cloud, networking and managed services capabilities. T-Systems' GPUaaS offering supports AI training, inference, simulation and digital twin use cases.

Strengths

Sovereign and regulatory-aligned GPU infrastructure: T-Systems positions GPUaaS as a trusted alternative for European enterprises and public sector organizations seeking compliance with GDPR, the EU AI Act and national data-sovereignty requirements. GPU workloads are delivered from EU-based data centers, supported by enterprise-grade security certifications and operational controls.

Industrial AI Cloud latest-generation GPUs: T-Systems operates Germany's first Industrial AI Cloud equipped with approximately 10,000 NVIDIA GPUs of the latest Blackwell generation, including DGX B200 and RTX Pro 6000 systems. This infrastructure enables high-performance AI model training, real-time AI inference and digital twin workloads, covering diverse industrial use cases. Its large-scale capacity and innovative hardware address growing AI compute demand in Europe and enhance AI innovation speed.

Dual model GPUaaS delivery: T-Systems' GPUaaS portfolio is anchored in its dual-model approach, combining flexible, consumption-based GPU access via T Cloud Public with high-capacity, reserved GPU deployments through the Industrial AI Cloud. This allows it to address a broad spectrum of enterprise AI requirements, ranging from experimentation and inference workloads to large-scale training and simulation use cases.

Caution

T-Systems should further strengthen its position in the global GPUaaS market by expanding its GPU footprint beyond Germany. A broader multi-region presence would enhance comparability with global peers and better align with enterprise expectations for distributed access, scalability and resilience.

Appendix

The ISG Provider Lens® 2026 – AI-ready Infrastructure Solutions study analyzes the relevant providers in the global market, based on a multi-phased research and analysis process, and positions these providers based on the ISG Research methodology.

Study Sponsor:
Heiko Henkes

Lead Author:
Sonam Chawla

Editor:
Indrani Raha

Research Analyst:
Mamtha R

Data Analyst:
Akshay Rathore

Project Manager:
Padma Mohapatra

Information Services Group Inc. is solely responsible for the content of this report. Unless otherwise cited, all content, including illustrations, research, conclusions, assertions and positions contained in this report were developed by, and are the sole property of, Information Services Group Inc.

The research and analysis presented in this report includes research from the ISG Provider Lens® program, ongoing ISG Research programs, interviews with ISG advisors, briefings with service providers and analysis of publicly available market information from multiple sources. The data collected for this report represent information that ISG believes to be current as of July 2026 for providers that actively participated and for providers that did not. ISG recognizes that many mergers and acquisitions may have occurred since then, but this report does not reflect these changes.

All revenue references are in U.S. dollars ($US) unless noted otherwise.

The study was conducted in the following steps:

1. Definition of AI-ready Infrastructure Solutions market
2. Use of questionnaire-based surveys of service providers/ vendor across all trend topics
3. Interactive discussions with service providers/vendors on capabilities and use cases
4. Leverage ISG's internal databases and advisor knowledge & experience (wherever applicable)
5. Detailed analysis and evaluation of services and service documentation based on the facts & figures received from providers and other sources
6. Use of the following key evaluation criteria:
   * Strategy and vision
   * Innovation
   * Brand awareness and presence in the market
   * Sales and partner landscape
   * Breadth and depth of portfolio of services offered
   * Technology advancements

Author

Sonam Chawla
Senior Lead Analyst

Sonam Chawla is a senior lead analyst at ISG, specializing in authoring Provider Lens® studies on the Google, Oracle and Salesforce ecosystems and AI-ready infrastructure solutions. With around eight years of experience in the market research industry, she has developed strong expertise in insight generation, market analysis, secondary research, report writing, blog creation and company analysis.

Sonam's key areas of interest include hyperscalers, cloud, AI-ready infrastructure, infrastructure technology, digital workplaces and enterprise collaboration. In her current role, Sonam also contributes to the research process by authoring Focal Points, providing valuable insights into regional and global market trends. Additionally, she manages custom engagement requests from providers and advisors.

Before taking on this role, Sonam worked as a research analyst, where she was responsible for developing syndicated research reports and providing consulting services for various research projects.

Research Analyst

Mamtha R
Research Specialist

Mamtha is a Research Specialist at ISG, supporting ISG Provider Lens® studies across AI-ready infrastructure, Telecom, Media and Entertainment, and Enterprise Managed Network Services. She collaborates with Lead Analysts on research, provider assessments, market insights, and the development of global summary reports and focal points.

Prior to joining ISG, Mamtha gained over nine years of experience in data collation, secondary research, and analysis. She conducted market and workforce research, analyzed industry trends, and delivered strategic insights for the CPG/FMCG sector, leveraging consumer data to identify market opportunities and buying behaviors.

Study Sponsor

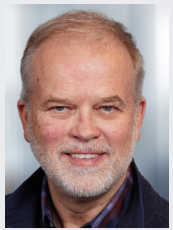
Heiko Henkes

Director & Principal Analyst, Global IPL Content Lead

Heiko Henkes serves as Managing Director and Principal Analyst at ISG, where he oversees the Global ISG Provider Lens® (IPL) Program for all IT Outsourcing (ITO) studies alongside his pivotal role in the global IPL division as strategic program manager and thought leader for IPL Lead Analysts. Additionally, Henkes heads the Star of Excellence, ISG's global customer experience initiative, steering program design and its integration with IPL and ISG's sourcing practice.

His expertise lies in guiding companies through IT-based business model transformations, leveraging his deep understanding of continuous transformation, IT competencies, sustainable business strategies, and change management in a Cloud-AI-driven business landscape. Henkes is renowned for his contributions as a keynote speaker on digital innovation, where he shares insights on leveraging technology for business growth and transformation.

IPL Product Owner



Jan Erik Aase

Partner and Global Head – ISG Provider Lens®/ISG Research

Mr. Aase brings extensive experience in the implementation and research of service integration and management of both IT and business processes;. With over 35 years of experience, he is highly skilled at analyzing vendor governance trends and methodologies, identifying inefficiencies in current processes, and advising the industry.

Jan Erik has experience on all four sides of the sourcing and vendor governance lifecycle - as a client, an industry analyst, a service provider and an advisor. Now as a partner and global head of ISG Provider Lens®, he is very well positioned to assess and report on the state of the industry and make recommendations for both enterprises and service provider clients.

ISG Provider Lens®

The ISG Provider Lens® Quadrant research series is the only service provider evaluation of its kind to combine empirical, data-driven research and market analysis with the real-world experience and observations of ISG's global advisory team. Enterprises will find a wealth of detailed data and market analysis to help guide their selection of appropriate sourcing partners. ISG advisors use the reports to validate their own market knowledge and make recommendations to ISG's enterprise clients. The research currently covers providers offering their services across multiple geographies globally.

For more information about ISG Provider Lens® research, please visit this webpage.

ISG Research™

ISG Research™ provides subscription research, advisory consulting and executive event services focused on market trends and disruptive technologies driving change in business computing. ISG Research™ delivers guidance that helps businesses accelerate growth and create more value.

ISG offers research specifically about providers to state and local governments (including counties and cities) and higher education institutions. Visit: Public Sector.

For more information about ISG Research™ subscriptions, please email contact@isg-one.com, call +1.203.454.3900, or visit research.isg-one.com.

ISG

ISG (Information Services Group) (Nasdaq: III) is a leading global AI-centered technology research and advisory firm. A trusted partner to more than 900 clients, including 75 of the world's top 100 enterprises, ISG is a long-time leader in technology and business services sourcing that is now at the forefront of leveraging AI to help organizations achieve operational excellence and faster growth.

The firm, founded in 2006, is known for its proprietary market data, in-depth knowledge of provider ecosystems, and the expertise of its 1,600 professionals worldwide working together to help clients maximize the value of their technology investments.

For more information, visit isg-one.com.

ISG Provider Lens®

JULY, 2026

REPORT: AI-READY INFRASTRUCTURE SOLUTIONS