



# Navigating AI Infrastructure: the Backbone of the AI-Driven Era

September 2024: Complimentary Abstract / Table of Contents

Market Report  
Cloud and Infrastructure Services



# Our research offerings

This report is included in the following research program(s):

## Cloud and Infrastructure Services

- ▶ Advanced SciTech
- ▶ Amazon Web Services (AWS)
- ▶ Application Services
- ▶ Artificial Intelligence (AI)
- ▶ Asset and Wealth Management
- ▶ Banking and Financial Services Business Process
- ▶ Banking and Financial Services Information Technology
- ▶ Catalyst™
- ▶ Clinical Development Technology
- ▶ Cloud and Infrastructure
- ▶ Contingent Staffing
- ▶ Contingent Workforce Management
- ▶ Customer Experience Management Services
- ▶ CX Excellence
- ▶ CXM Technology
- ▶ Cybersecurity
- ▶ Cyber Threat Detection and Response
- ▶ Data and Analytics
- ▶ Digital Adoption Platforms
- ▶ Digital Services
- ▶ Digital Workplace
- ▶ Employee Experience Management (EXM) Platforms
- ▶ Employer of Record (EOR)
- ▶ Engineering Research and Development
- ▶ Enterprise Platform Services
- ▶ Exponential Technologies
- ▶ Finance and Accounting
- ▶ Financial Crime and Compliance
- ▶ Financial Services Technology (FinTech)
- ▶ Forces & Foresight
- ▶ GBS Talent Excellence
- ▶ Global Business Services
- ▶ Google Cloud
- ▶ HealthTech
- ▶ Human Resources
- ▶ Insurance Business Process
- ▶ Insurance Information Technology
- ▶ Insurance Technology (InsurTech)
- ▶ Insurance Third-Party Administration (TPA) Services
- ▶ Intelligent Document Processing
- ▶ Interactive Experience (IX) Services
- ▶ IT Services Excellence
- ▶ IT Talent Excellence
- ▶ Life Sciences Business Process
- ▶ Life Sciences Commercial Technologies
- ▶ Life Sciences Information Technology
- ▶ Locations Insider™
- ▶ Marketing Services
- ▶ Market Vista™
- ▶ Microsoft Azure
- ▶ Microsoft Business Application Services
- ▶ Modern Application Development (MAD)
- ▶ Mortgage Operations
- ▶ Multi-country Payroll
- ▶ Network Services and 5G
- ▶ Oracle Services
- ▶ Outsourcing Excellence
- ▶ Payer and Provider Business Process
- ▶ Payer and Provider Information Technology
- ▶ Price Genius – AMS Solution and Pricing Tool
- ▶ Pricing Analytics as a Service
- ▶ Process Intelligence
- ▶ Process Orchestration
- ▶ Procurement and Supply Chain
- ▶ Recruitment
- ▶ Retail and CPG
- ▶ Retirement Technologies
- ▶ Revenue Cycle Management
- ▶ Rewards and Recognition
- ▶ SAP Services
- ▶ Service Optimization Technologies
- ▶ Software Product Engineering Services
- ▶ Supply Chain Management (SCM) Services
- ▶ Sustainability Technology and Services
- ▶ Talent Genius™
- ▶ Technology Skills and Talent
- ▶ Trust and Safety
- ▶ Value and Quality Assurance (VQA)

If you want to learn whether your organization has a membership agreement or request information on pricing and membership options, please contact us at [info@everestgrp.com](mailto:info@everestgrp.com)

Learn more about  
our custom research capabilities

Benchmarking

Contract assessment

Peer analysis

Market intelligence

Tracking: providers, locations, risk,  
technologies

Locations: costs, skills, sustainability,  
portfolios

# Contents

5	<b>Introduction and overview</b>	24	Emerging trends shaping enabling components of AI Infrastructure
6	Research methodology	25	Implications for enterprises, technology providers, and service providers
7	Introduction	26	<b>Enterprise playbook for AI infrastructure adoption</b>
8	Phases of AI infrastructure adoption	27	Guide to navigating the AI infrastructure playbook
9	Overview of the infrastructure layer	28	Step 1: Assess and analyze
10	Market view	29	Step 2: Align and augment
12	<b>Understanding the current state of AI infrastructure</b>	30	<b>Infrastructure offerings from AI providers</b>
13	AI stack overview	31	Shortlisting criteria
14	Components of the infrastructure layer	32	Cloud platform and data center providers
15	Infrastructure layer   Foundational elements: Compute hardware	33	Akamai
16	Infrastructure layer   Foundational elements: Storage	34	AWS
17	Infrastructure layer   Foundational elements: Connectivity	35	Azure
18	Infrastructure layer   Enabling elements: Cloud platforms	36	Cloudflare
19	Infrastructure layer   Enabling elements: Data centers	37	CoreWeave
20	Challenges with current AI infrastructure	38	Cyxtera Technologies
21	<b>Emerging trends shaping up AI infrastructure landscape</b>	39	Equinix
22	Emerging trends shaping enterprise adoption of AI	40	GCP
23	Emerging trends shaping foundational components of AI Infrastructure		

Copyright © 2024 Everest Global, Inc.

We encourage you to share these materials internally in accordance with your license. Sharing these materials outside your organization in any form – electronic, written, or verbal – is prohibited unless you obtain the express, prior, and written consent of Everest Global, Inc. It is your organization's responsibility to maintain the confidentiality of these materials in accordance with your license of them.

For more information on this and other research published by Everest Group, please contact us:

**Yugal Joshi**, Partner

**Mukesh Ranjan**, Vice President

**Zachariah Chirayil**, Practice Director

**Praharsh Srivastava**, Senior Analyst

**Rachita Rao**, Senior Analyst

# Contents

32	Cloud platform and data center providers (continued)
41	IBM
42	Oracle
43	Hardware providers
44	AMD
45	Arm
46	Broadcom
47	Cerebras
48	Cisco
49	Dell
50	HPE
51	Intel
52	Lenovo
53	NVIDIA
54	Appendix
55	Glossary
56	Research calendar

# Introduction

The Artificial Intelligence (AI) boom has driven significant advances across various sectors, such as precision medicine in healthcare, algorithmic trading in finance, autonomous vehicles in automotive, and intelligent networking solutions in telecommunications.

However, with this surge in AI applications, enterprises are finding the traditional IT infrastructure insufficient for handling the high computational demands and vast data processing needs of AI workloads. These limitations are prompting enterprises to invest in specialized AI infrastructure to address the challenges by reducing latency, speeding up inference, accelerating model training, and improving application scalability. This dedicated infrastructure is crucial for the optimal functioning of AI applications, effectively forming the backbone of AI deployment.

As the demand for AI grows, enterprises recognize an increasing need for advanced compute hardware, high storage capabilities, enhanced connectivity, and robust cloud platforms and data centers. Without these specialized infrastructure components, AI applications cannot reach their full potential, underscoring the critical role of AI infrastructure in supporting and advancing enterprise AI capabilities.

In this report, we provide an outlook on the global AI infrastructure market along with enterprise concerns and challenges, adoption, recent developments, and implications for enterprises and providers. This report also offers an enterprise playbook for AI infrastructure adoption. Additionally, this report provides a fact-pack view of the top 20 AI infrastructure providers.

The analysis is based on Everest Group's primary reachouts, surveys, client reference checks, enterprise interactions, and ongoing analysis of the AI infrastructure market.

## Scope of this report

**Geography:** Global

**Industry:** All industries

**Technology:** AI infrastructure

# Overview and abbreviated summary of key messages

This report examines the global AI infrastructure market, covering enterprise concerns, challenges, adoption trends, recent developments, and implications for enterprises and providers. Additionally, it includes an enterprise playbook for AI infrastructure adoption and a fact-pack on the top 20 providers.

## Some of the findings in this report, among others, are:

**The infrastructure layer forms the base for all AI applications, supplying the necessary computing power for AI workloads**

- The selected infrastructure determines the performance, security, scalability, and latency of applications
- Enterprises are heavily investing in AI infrastructure, leading providers to increase their investments to meet demand

**The AI technology stack is complex and includes foundational and enabling components**

- Foundational components include compute hardware, storage, and connectivity, while enabling components encompass cloud platforms and data centers
- There is a compelling need for substantial upgrades and expansions in IT infrastructure to meet the soaring demands of AI workloads

**Emerging trends in AI adoption, foundational, and enabling components used are shaping the AI infrastructure landscape**

- New trends are shaping AI infrastructure by adopting edge-to-cloud and RAG stack technologies to improve efficiency, innovation, and security
- Enterprises are shifting to specialized AI applications, preferring hybrid cloud for better performance and security

**Enterprises need to adopt a thought through approach towards their AI infrastructure to ensure they are future ready while being cognizant of the current needs**

- Identify key AI infrastructure investments based on an assessment, keeping the evaluation parameters in mind
- Use the three Ps<sup>1</sup> to align your infrastructure with key investment areas, keeping the 5S<sup>2</sup> in mind

**Many technology providers are developing infrastructure solutions for AI workloads**

- Most cloud and data center providers are using NVIDIA GPUs and investing in AI startups for AI workloads, with GCP gaining an edge by producing its own TPUs
- NVIDIA leads in the chip industry, but new providers such as Cerebras are offering unique products

<sup>1</sup> Ps: Provision, power, and platform

<sup>2</sup> 5S: Scalability, sustainability, security, simplicity, and sustainability

# This study offers distinct chapters providing a deep dive into key aspects of AI infrastructure market; below are four charts to illustrate the depth of the report

## The infrastructure layer is the foundation for all AI applications

The infrastructure selected determines:

- Power and performance of the applications** Utilizing high-performance GPUs or TPUs accelerates model training and inference, while distributed infrastructure architectures facilitate parallel processing, significantly reducing computation time and enhancing overall efficiency
- Security** Secure infrastructure is paramount for safeguarding sensitive AI data. Implementing robust security measures, such as encrypted communication channels, ensures the confidentiality and integrity of data both in transit and at rest, mitigating the risk of unauthorized access or data breaches
- Scalability** Optimal infrastructure enables seamless scaling by dynamically allocating resources as the workload fluctuates. This ensures consistent performance even during peak usage periods, enhancing the reliability and availability of AI applications
- Latency** Low-latency network connections facilitate rapid data exchange and enable real-time responsiveness, crucial for applications requiring instantaneous decision-making capabilities for model inferencing

Current landscape of IT Infrastructure layer for AI

Cloud providers



Co-location providers



Hardware providers



## Limitations of current IT infrastructure

- High demands of advanced AI systems** often **outstrip existing hardware**, hindering model training and deployment.
- Scarce specialized chips** hinder AI growth as production is concentrated among a few manufacturers.
- The high costs and computational demands** of large AI models limit accessibility for businesses with budget constraints or security concerns.
- Centralized AI infrastructure struggles to handle the **demands of large datasets** and real-time applications, while also limiting control over data and raising **security concerns**.
- On-demand scaling is very expensive for enterprises due to **GPU scarcity** and high demand.
- Inefficient storage solutions** hinder training speed and effectiveness, affecting enterprises, especially those with heterogeneous architectures and diverse hardware configurations.
- Unsustainable energy demands** of current AI models within existing infrastructure require urgent development of supportive infrastructure for scalable deployment.

## Framework for AI infrastructure adoption

STEP 1:  
**Assess and analyze**

Identify key investments for AI infrastructure based on the assessment

Keep the parameters in mind while evaluating



STEP 2:  
**Align and augment**

Use the three Ps to align your infrastructure in the key areas of investment

Keep the 5S in mind

## Top 20 providers in the AI infrastructure market

Cloud platform and data center providers

- Akamai
- AWS
- Azure
- Cloudflare
- CoreWeave
- Cyxtera Technologies
- Equinix
- Google Cloud Platforms (GCP)
- IBM
- Oracle

Hardware providers

- AMD
- Arm
- Broadcom
- Cerebras
- Cisco
- Dell
- HPE
- Intel
- Lenovo
- NVIDIA

# Research calendar

## Cloud and Infrastructure Services

	Published	Current release	Planned
Reports title	Release date		
Google Cloud Services Specialists PEAK Matrix® Assessment 2024			February 2024
IT Services CXO Insights: Key Issues for 2024			March 2024
Cloud Modernization: Maximize Your ROI in Cloud			April 2024
Mainframe Services PEAK Matrix® Assessment 2024			April 2024
Generative AI – Review of Adobe Summit 2024			May 2024
Mainframe Solutions: Review of Broadcom's WatchTower Platform			July 2024
<a href="#">Navigating AI Infrastructure: the Backbone of the AI-Driven Era</a>			September 2024
FinOps Cloud Cost Management Products PEAK Matrix® Assessment 2024			Q3 2024
AWS Services Specialists PEAK Matrix® Assessment 2024			Q3 2024
AWS Services PEAK Matrix® Assessment 2024			Q4 2024
Microsoft Azure Services PEAK Matrix® Assessment 2024			Q4 2024
AI-led Network Transformation for Businesses			Q4 2024
Google Cloud Services PEAK Matrix® Assessment 2024			Q4 2024
Cloud and Infrastructure State of the Market 2024			Q4 2024

Note: [Click](#) to see a list of all of our published Cloud and Infrastructure Services reports



# Stay connected

Dallas (Headquarters)  
info@everestgrp.com  
+1-214-451-3000

Bangalore  
india@everestgrp.com  
+91-80-61463500

Delhi  
india@everestgrp.com  
+91-124-496-1000

London  
unitedkingdom@everestgrp.com  
+44-207-129-1318

Toronto  
canada@everestgrp.com  
+1-214-451-3000

Website  
everestgrp.com

Blog  
everestgrp.com/blog

Follow us on



Everest Group is a leading research firm helping business leaders make confident decisions. We guide clients through today's market challenges and strengthen their strategies by applying contextualized problem-solving to their unique situations. This drives maximized operational and financial performance and transformative experiences. Our deep expertise and tenacious research focused on technology, business processes, and engineering through the lenses of talent, sustainability, and sourcing delivers precise and action-oriented guidance. Find further details and in-depth content at [www.everestgrp.com](http://www.everestgrp.com).

## Notice and disclaimers

**Important information. Please review this notice carefully and in its entirety. Through your access, you agree to Everest Group's terms of use.**

Everest Group's Terms of Use, available at [www.everestgrp.com/terms-of-use/](http://www.everestgrp.com/terms-of-use/), is hereby incorporated by reference as if fully reproduced herein. Parts of these terms are pasted below for convenience; please refer to the link above for the full version of the Terms of Use.

Everest Group is not registered as an investment adviser or research analyst with the U.S. Securities and Exchange Commission, the Financial Industry Regulatory Authority (FINRA), or any state or foreign securities regulatory authority. For the avoidance of doubt, Everest Group is not providing any advice concerning securities as defined by the law or any regulatory entity or an analysis of equity securities as defined by the law or any regulatory entity.

All Everest Group Products and/or Services are for informational purposes only and are provided "as is" without any warranty of any kind. You understand and expressly agree that you assume the entire risk as to your use and any reliance upon any Product or Service. Everest Group is not a legal, tax, financial, or investment advisor, and nothing provided by Everest Group is legal, tax, financial, or investment advice. Nothing Everest Group provides is an offer to sell or a solicitation of an offer to purchase any securities or instruments from any entity. Nothing from Everest Group may be used or relied upon in evaluating the merits of any investment. Do not base any investment decisions, in whole or part, on anything provided by Everest Group.

Products and/or Services represent research opinions or viewpoints, not representations or statements of fact. Accessing, using, or receiving a grant of access to an Everest Group Product and/or Service does not constitute any recommendation by Everest Group that recipient (1) take any action or refrain from taking any action or (2) enter into a particular transaction. Nothing from Everest Group will be relied upon or interpreted as a promise or representation as to past, present, or future performance of a business or a market. The information contained in any Everest Group Product and/or Service is as of the date prepared, and Everest Group has no duty or obligation to update or revise the information or documentation. Everest Group may have obtained information that appears in its Products and/or Services from the parties mentioned therein, public sources, or third-party sources, including information related to financials, estimates, and/or forecasts. Everest Group has not audited such information and assumes no responsibility for independently verifying such information as Everest Group has relied on such information being complete and accurate in all respects. Note, companies mentioned in Products and/or Services may be customers of Everest Group or have interacted with Everest Group in some other way, including, without limitation, participating in Everest Group research activities.