ANTRIKSH CLOUD PRIVATE LIMITED

Business Data Insights on AI & Cloud Growth

Global Data Center Critical IT Power Growth

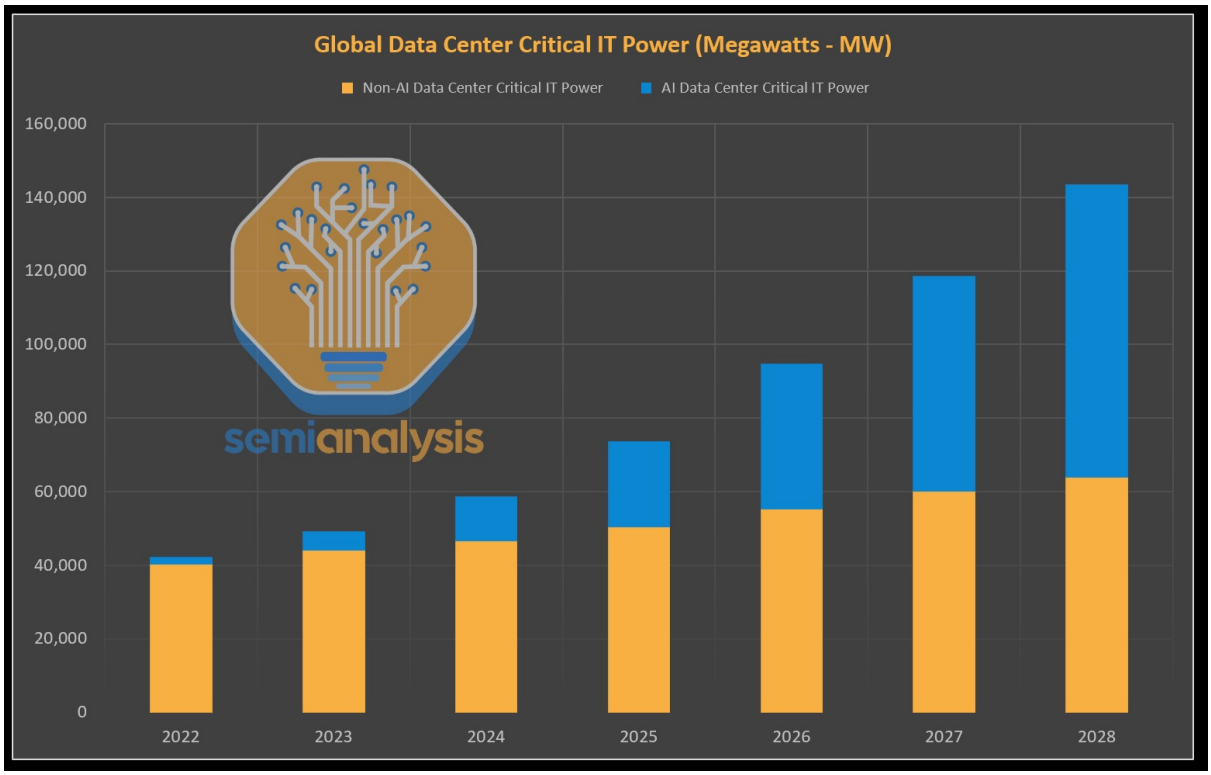
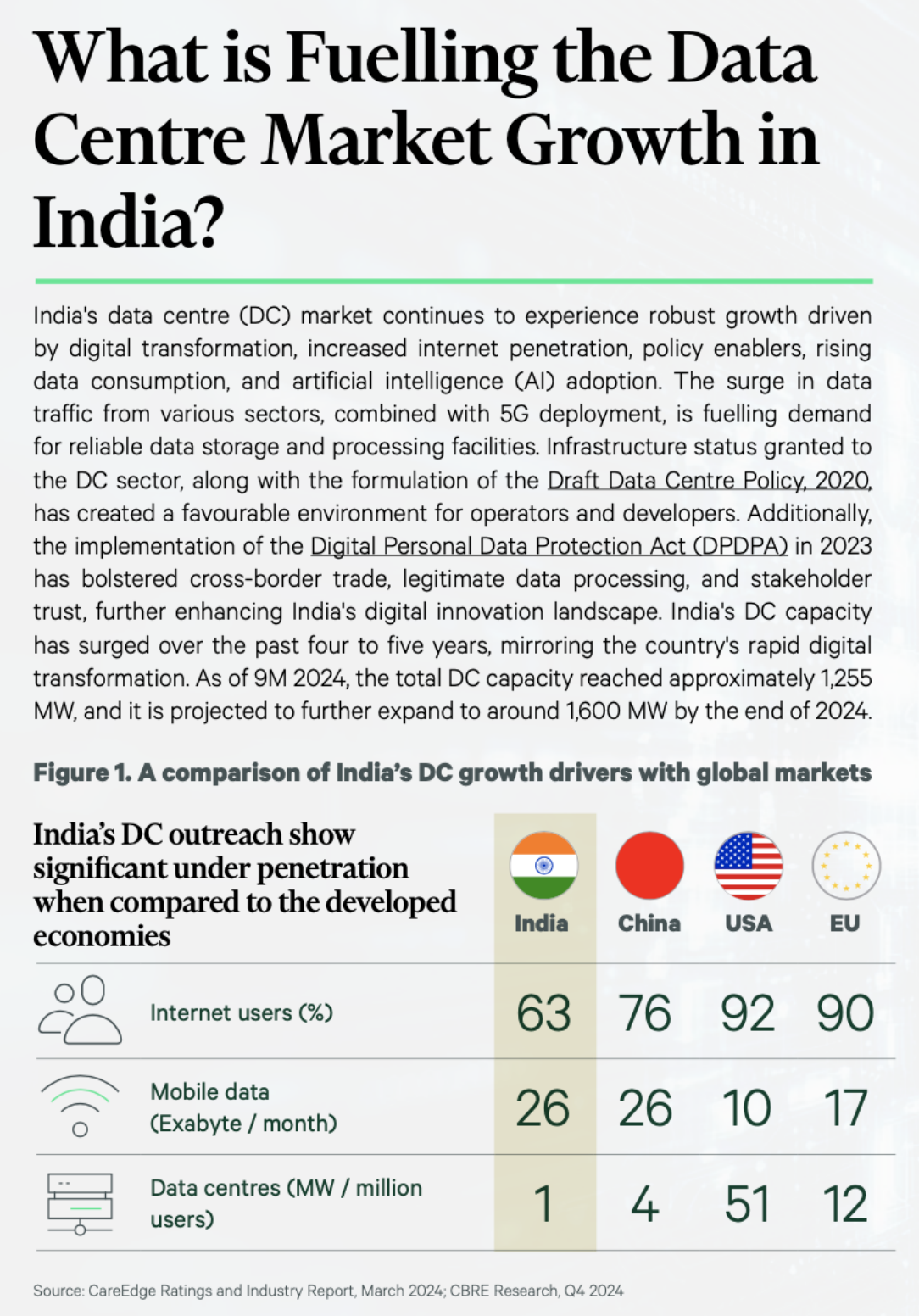


Figure 1: Growth of AI Data Center Power Consumption. AI-driven workloads are projected to significantly increase power demands in data centers.

Key Takeaway: AI Data Centers are driving a significant portion of power demand growth, with projections showing a steep rise through 2028.

India's Data Center Market Growth Drivers



What is Fuelling the Data Centre Market Growth in India?

India's data centre (DC) market continues to experience robust growth driven by digital transformation, increased internet penetration, policy enablers, rising data consumption, and artificial intelligence (AI) adoption. The surge in data traffic from various sectors, combined with 5G deployment, is fuelling demand for reliable data storage and processing facilities. Infrastructure status granted to the DC sector, along with the formulation of the Draft Data Centre Policy, 2020, has created a favourable environment for operators and developers. Additionally, the implementation of the Digital Personal Data Protection Act (DPDPA) in 2023 has bolstered cross-border trade, legitimate data processing, and stakeholder trust, further enhancing India's digital innovation landscape. India's DC capacity has surged over the past four to five years, mirroring the country's rapid digital transformation. As of 9M 2024, the total DC capacity reached approximately 1,255 MW, and it is projected to further expand to around 1,600 MW by the end of 2024.

Figure 1. A comparison of India's DC growth drivers with global markets

India's DC outreach show significant under penetration when compared to the developed economies	India	China	USA	EU
Internet users (%)	63	76	92	90
Mobile data (Exabyte / month)	26	26	10	17
Data centres (MW / million users)	1	4	51	12

Source: CareEdge Ratings and Industry Report, March 2024; CBRE Research, Q4 2024

Figure 2: Factors driving India's data center market expansion.

Key Takeaway: Indian Infrastructure is highly underdeveloped. Above image shows that our consumption is the highest at 26 Eb/month per the Data Centers are capacity is only 1MW/ Million.

AI's Role in Transforming India's Data Center Market

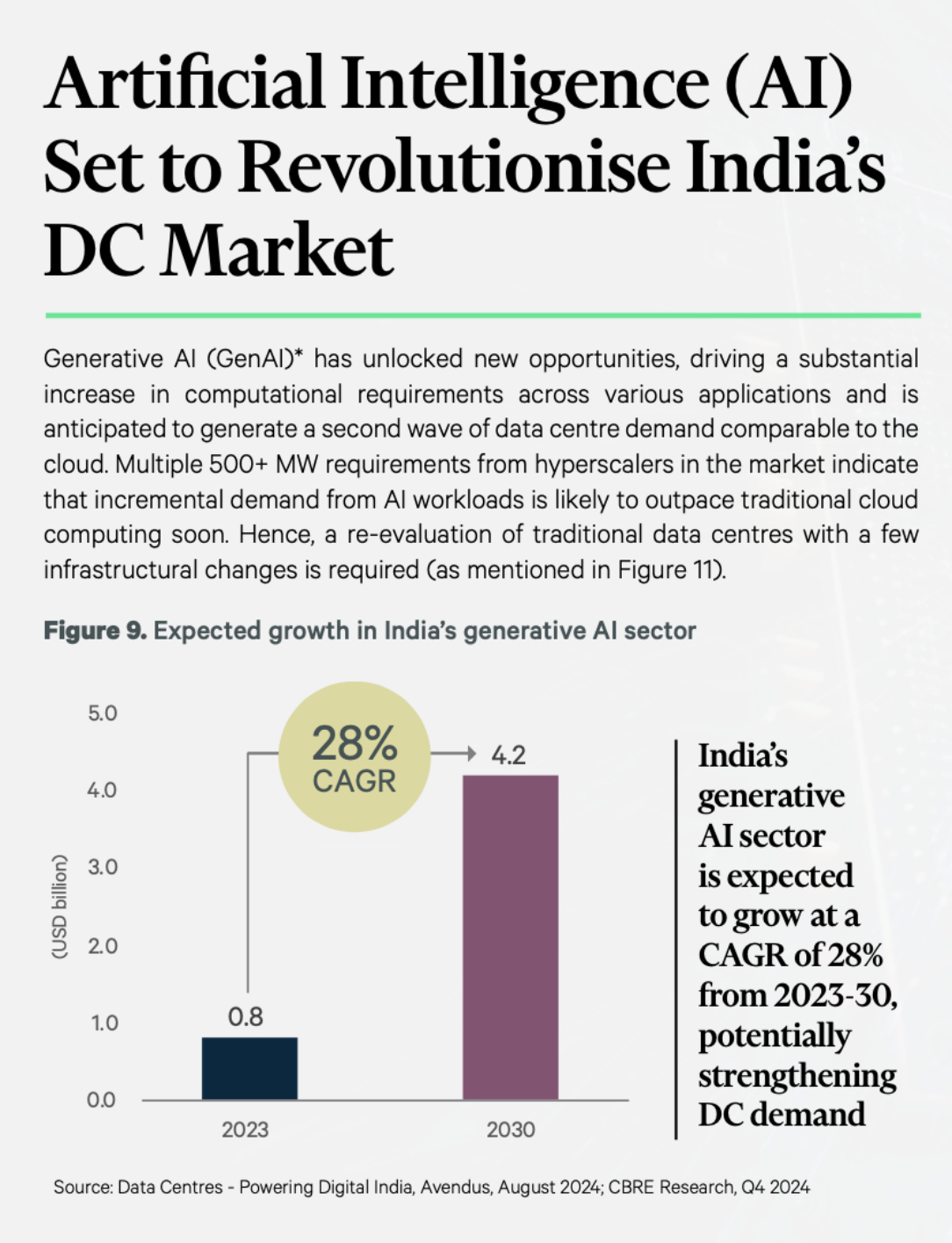


Figure 3: AI-driven demand for India's data center sector, projected 28% CAGR growth.

Key Takeaway: Generative AI is driving increased computational demand, with India's AI data center sector projected to grow at 28% CAGR from 2023 to 2030.

AI-Driven Capital Expenditures

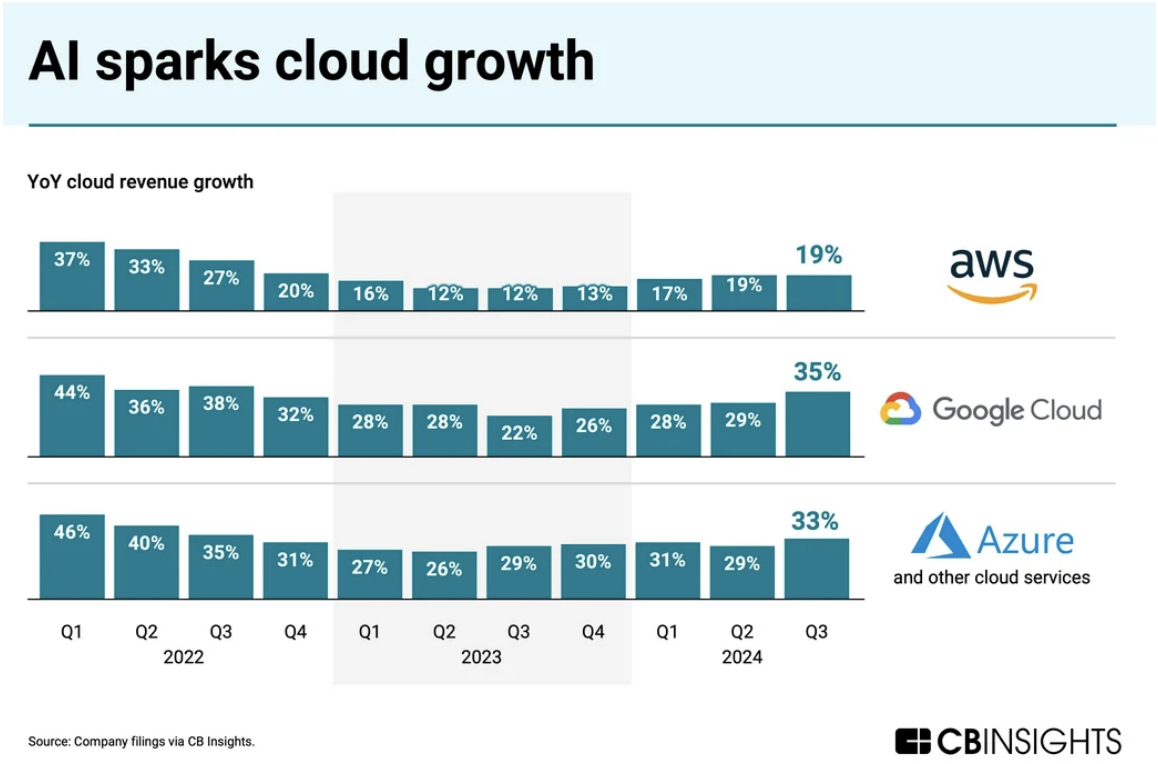


Figure 4: Big tech's AI-fueled CapEx spending. Amazon, Microsoft, Alphabet, and Meta are increasing their capital expenditures significantly to support AI infrastructure.

Key Takeaway: AI-driven investments are accelerating CapEx spending, with companies pushing past $50 billion per quarter to scale AI infrastructure.

Worldwide Cloud Infrastructure Services Revenue

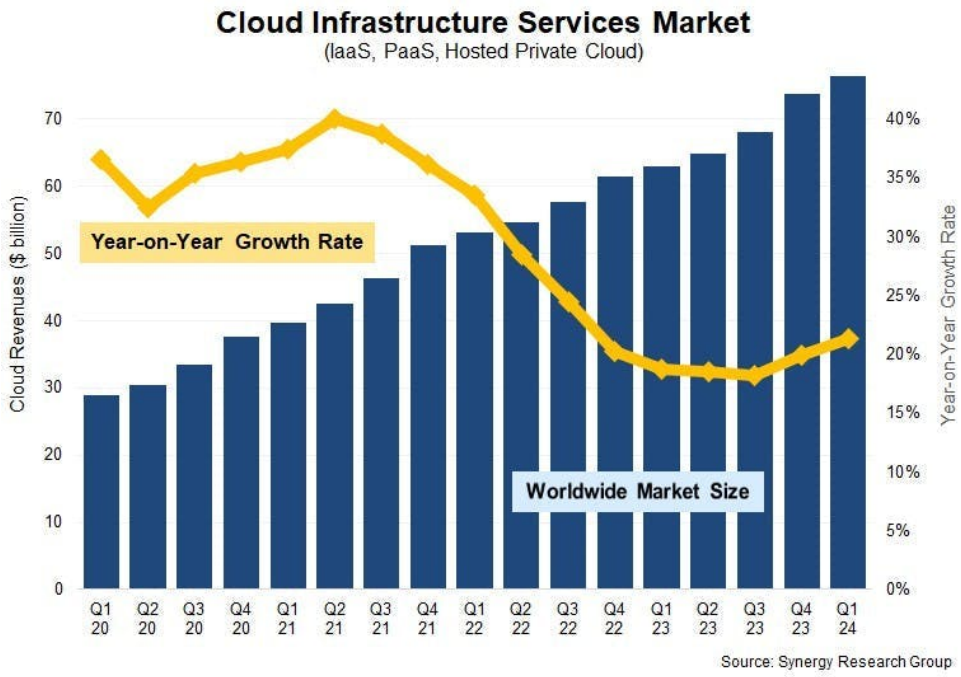


Figure 5: Quarterly revenue growth in cloud infrastructure services from 2018 to 2023, highlighting steady increases and surpassing $74 billion in Q4 2023.

Key Takeaway: The market has grown steadily over the past years, indicating strong demand for cloud computing resources.

Cloud Revenue Growth Trends

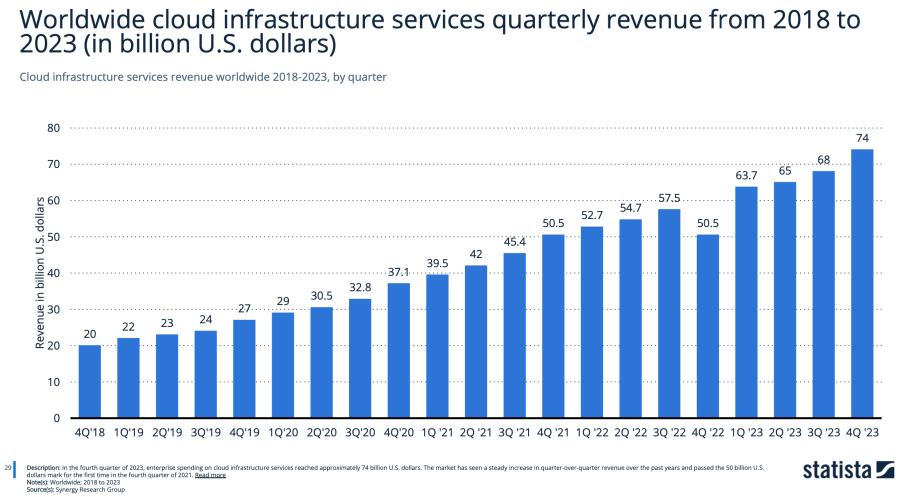


Figure 6: Year-over-year growth in cloud revenue for AWS, Google Cloud, and Azure. Google Cloud leads with 35% growth, followed by Azure at 33% and AWS at 19%.

Key Takeaway: Cloud adoption, especially AI-driven workloads, has reignited growth in cloud infrastructure, with Google Cloud showing the highest growth rate.

AI Infrastructure Cost Breakdown

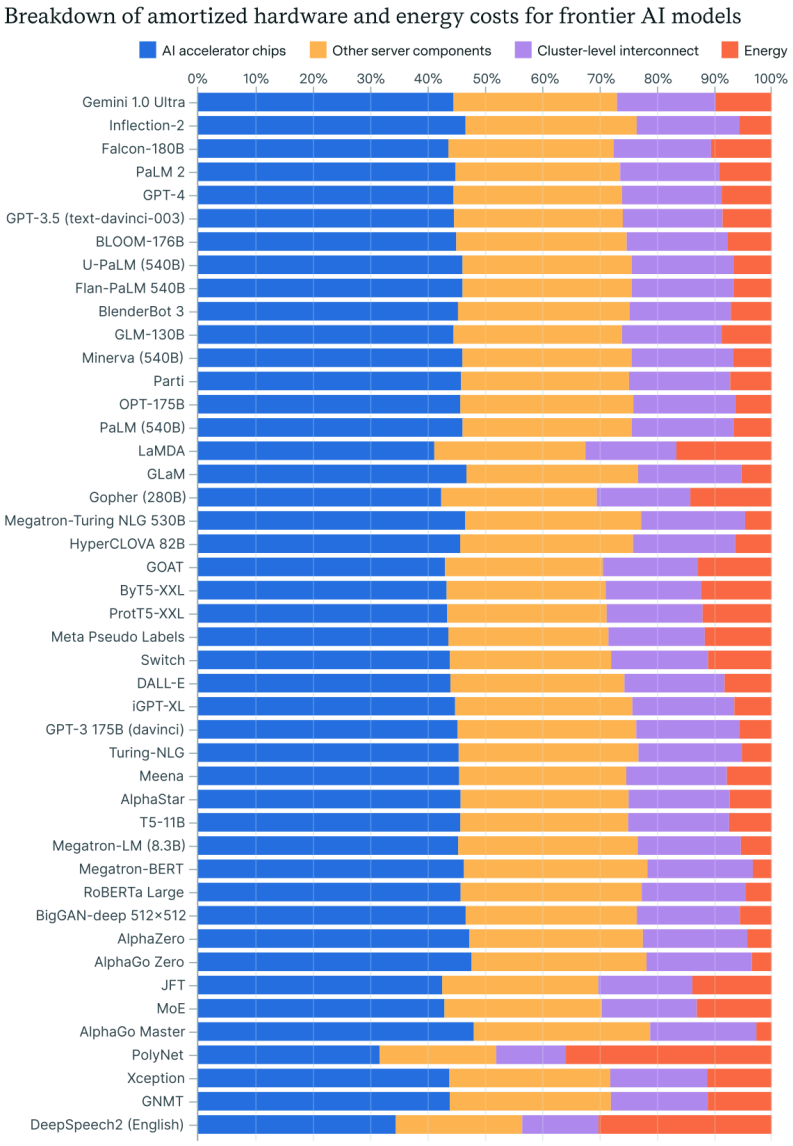


Figure 5: The percentage of the amortized hardware CapEx + energy estimates made up by different hardware and energy costs. Note that the breakdown across models is approximate. Cluster-level interconnect is assumed to be a constant 19% fraction of the cluster CapEx, and the proportion of server components is based on only three comparisons between NVIDIA DGX server prices and single GPU prices (see Appendix A.1 for details). The energy costs are more specific, varying with the number of training chip-hours and the hardware (see Appendix A.4).

Figure 7: Cost distribution of AI model deployment, including AI accelerators, interconnects, energy consumption, and server components.

Key Takeaway: AI accelerator chips form the largest cost component, while energy costs also play a substantial role in total infrastructure expenses.

Increasing Compute Needs for AI

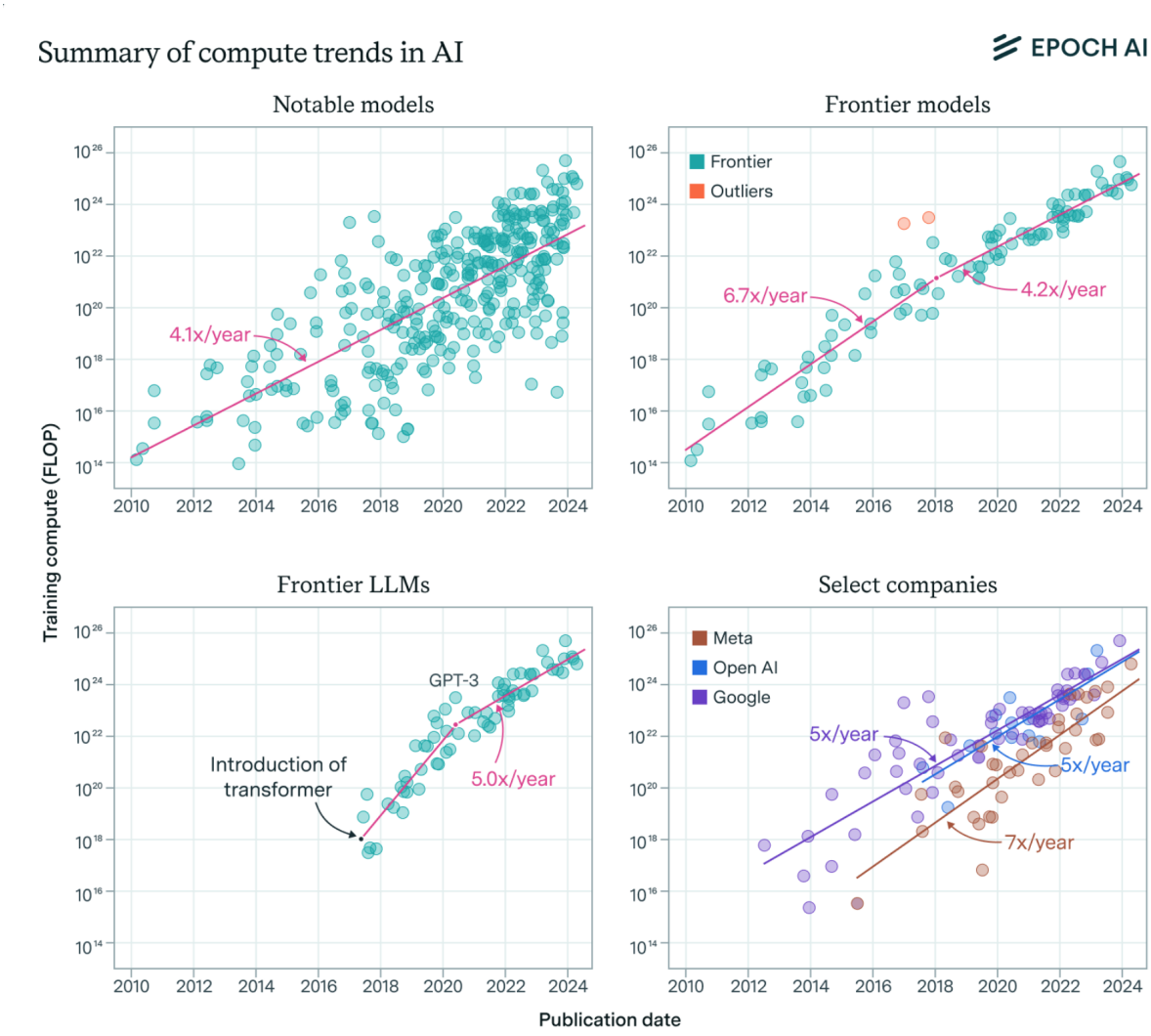


Figure 8: The rapid increase in AI compute requirements, growing exponentially with each new AI model.

Key Takeaway: AI models are consuming increasingly more compute resources, growing at a rate of 4-7x per year, accelerating demand for AI data centers.

Cloud Provider Competitive Positioning

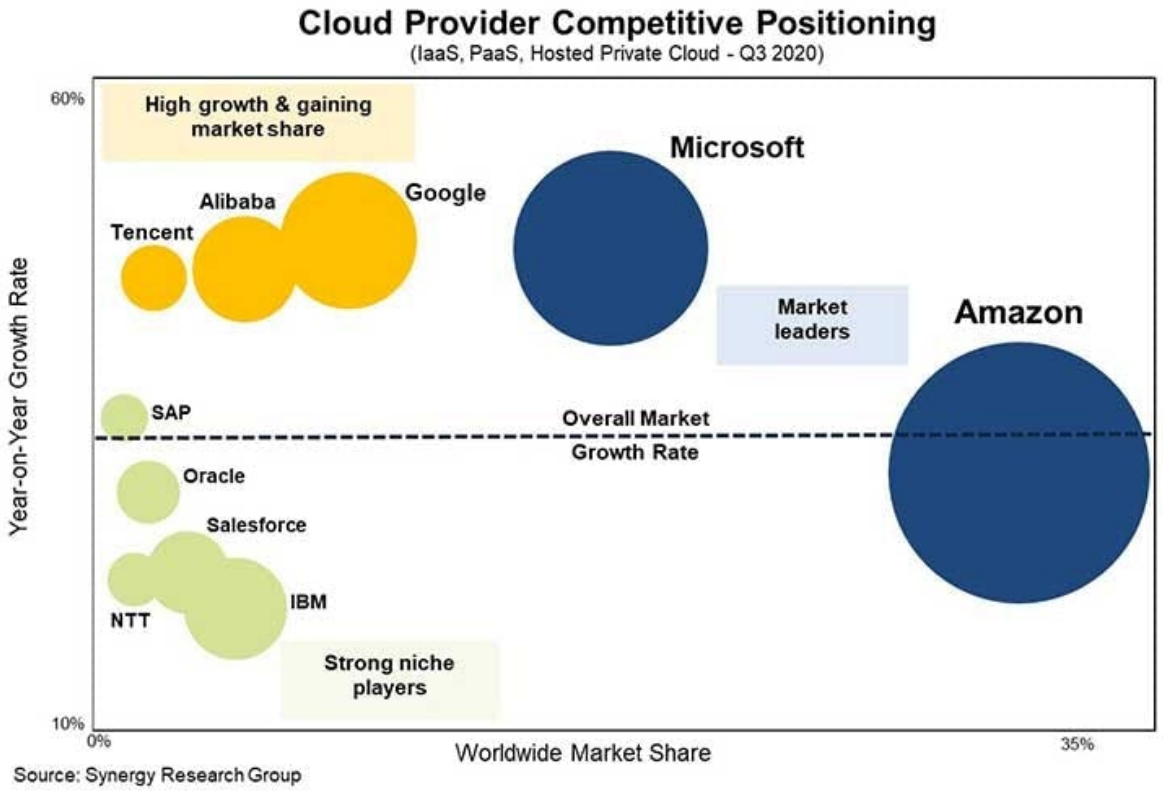


Figure 9: Competitive positioning of cloud providers. Amazon and Microsoft dominate in market share, while Google, Alibaba, and Tencent show high growth rates.

Key Takeaway: Amazon and Microsoft lead the cloud market, while Google and Alibaba are catching up with higher growth rates.

Data Center Capital Expenditures Forecast

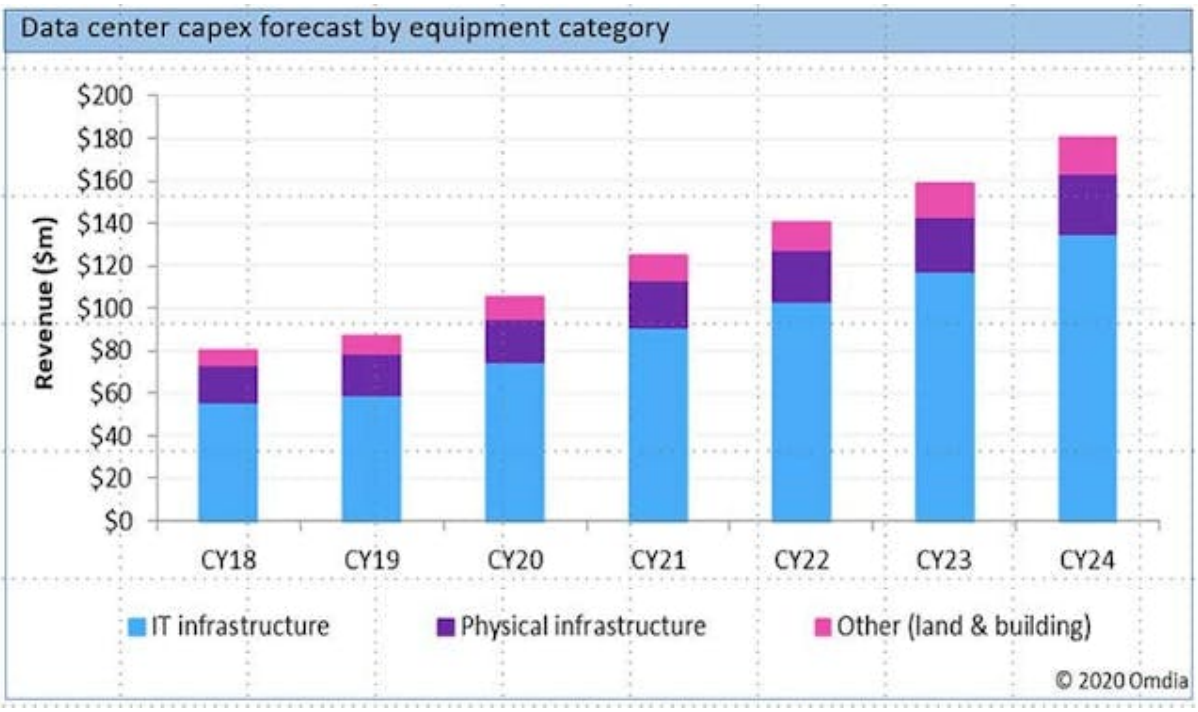


Figure 10: Forecast of data center CapEx by equipment category, highlighting increasing investments in IT infrastructure.

Key Takeaway: IT infrastructure investments dominate CapEx, signaling strong demand for data center expansion.

Cloud Market Growth Trends

Figure 11: Year-over-year growth trends in PaaS, IaaS, and SaaS segments, showing a slowdown but still strong growth.

Key Takeaway: IaaS continues to lead cloud market growth, followed by PaaS, with SaaS showing slower growth.

Global Cloud Infrastructure Market Expansion

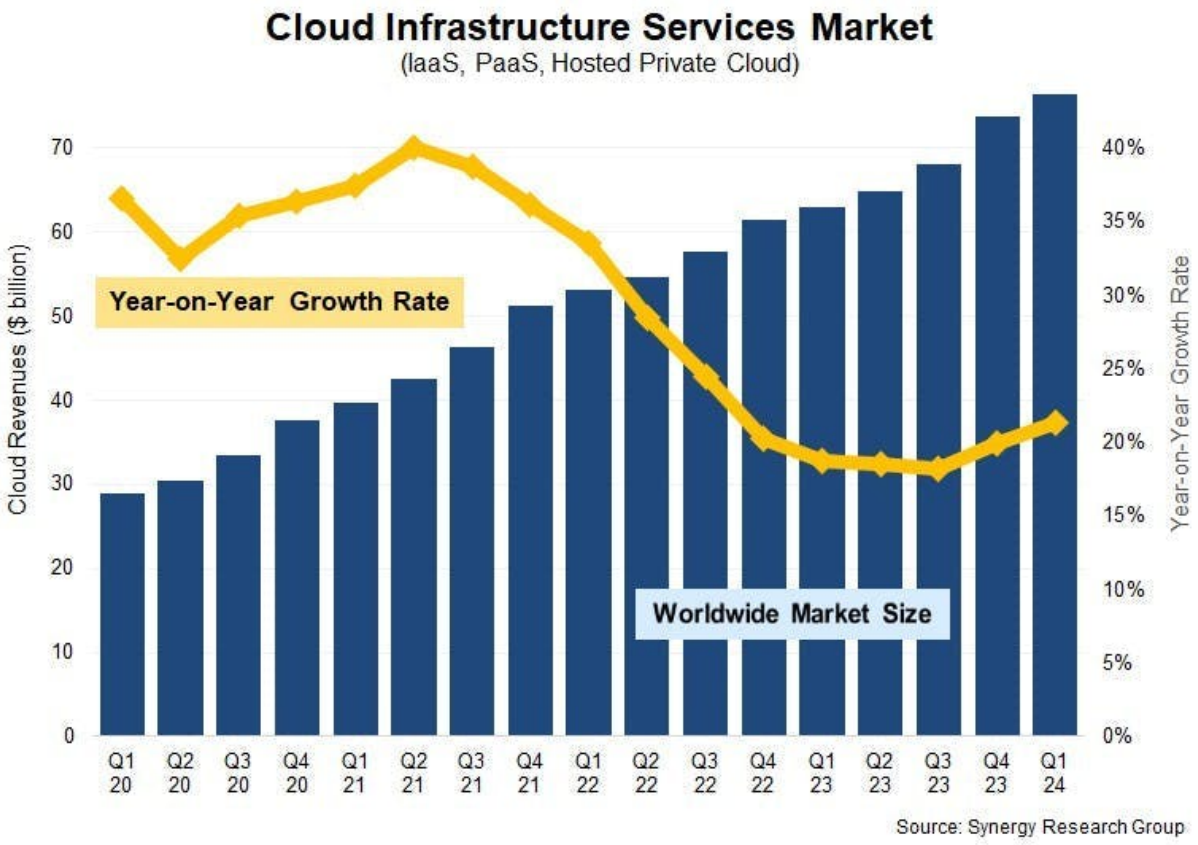


Figure 12: Worldwide cloud infrastructure market size and year-over-year growth trends.

Key Takeaway: Despite fluctuating growth rates, cloud infrastructure services continue to expand significantly.

Top LLM Cloud Providers

Figure 13: Top 15 Cloud Providers for Large Language Model (LLM) training and inference.

Key Takeaway: The LLM cloud market is expanding beyond traditional hyperscalers (AWS, Azure, Google Cloud) to specialized providers like Lambda Labs and CoreWeave, optimizing for AI workloads.

Compute Scaling Improves Performance

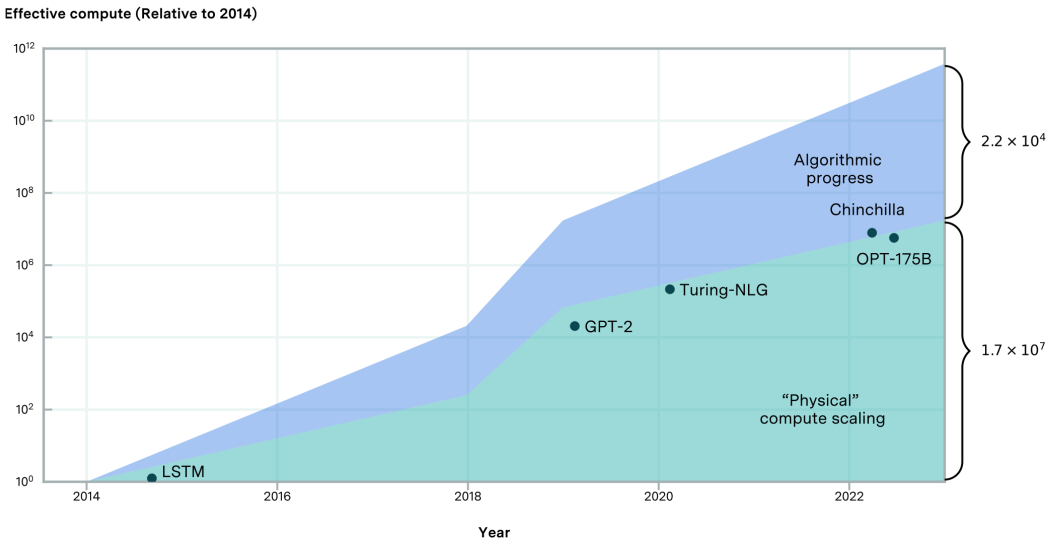


Figure 5: A stylized illustration of the relative contribution of compute scaling and algorithmic progress to effective compute. The physical compute contribution is estimated from the doubling times in Sevilla et al. 2022, and the algorithmic progress contribution is based on the aggregated doubling time estimate from the top 10 models in cross validation (see section 3.1). We further plot the physical training compute values for several notable models (e.g. GPT-2) in their publication years.

Figure 14: The role of algorithmic progress and compute scaling in improving AI model performance.

Key Takeaway: AI model efficiency is improving due to both hardware advancements and algorithmic optimizations, significantly increasing compute effectiveness.

APAC Data Center Market Growth

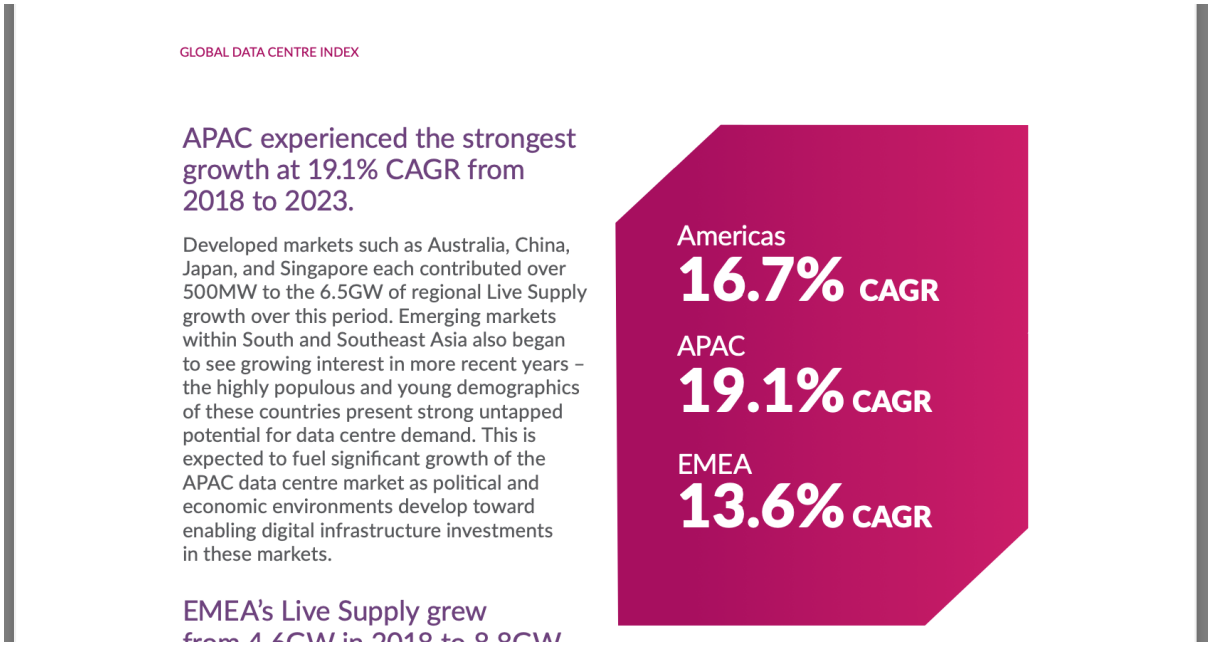


Figure 15: APAC experienced the strongest growth at 19.1% CAGR from 2018 to 2023.

Key Takeaway: APAC is the fastest-growing data center region, outpacing other global markets, driven by high digital demand and cloud infrastructure investments.

AI Data Center Value Chain

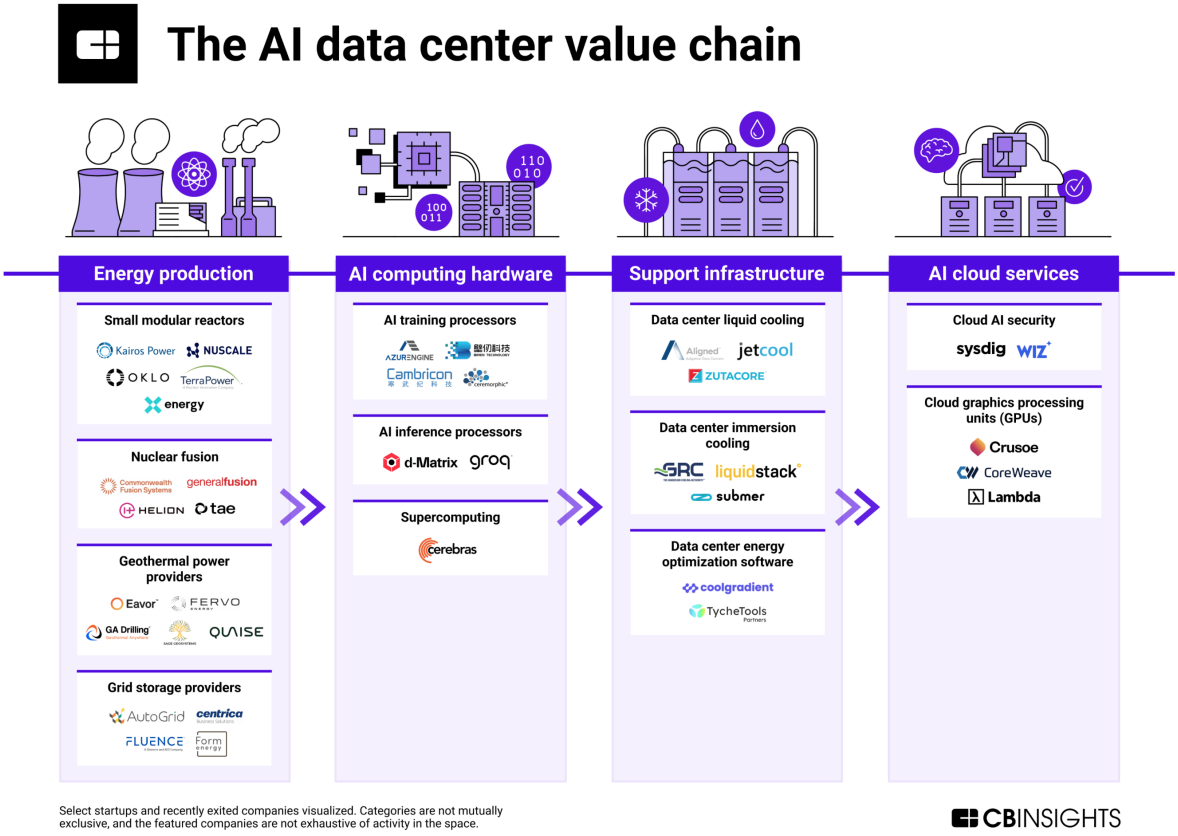


Figure 16: The AI data center value chain covering energy, computing hardware, infrastructure, and cloud services.

Key Takeaway: AI data centers require specialized infrastructure across multiple layers, from power sources to AI-optimized cloud services.

SWOT Analysis of GPU Cloud Providers

Prospects for GPU cloud providers

A simple SWOT analysis provides a summary of the prospects for these emerging companies (see **Table 1**).

Table 1. Analysis of GPU cloud providers' competitive position

Strengths	Weaknesses
GPU capacity in multiple locations at reasonable cost. Ability to focus on AI capabilities and cost reductions without being encumbered by broad product portfolios. Range of customizations and optimizations, with a focus on core AI proposition. No data egress fees — a common and costly charge imposed by hyperscalers, which hinders cost predictability.	Narrow product portfolio may deter customers that want a comprehensive suite of services, especially where application integration is required. Little product differentiation between providers considering they sell products developed by a third party.
Opportunities	**Threats**
The rise of AI-driven applications has increased demand for GPUs, presenting a growth opportunity. With growing concerns over cloud costs, particularly data egress fees, enterprises are looking to squeeze costs.	Hyperscalers may cut their prices, undercutting the GPU providers and removing their key differentiation. As supply chain issues resolve, future buyers of hardware may be able to sell GPU services more cheaply than current providers. Any disruption in the GPU supply chain or a market shift could severely impact GPU cloud providers' revenues. If revenue declines, GPU investments may not make a return. A lack of other revenue streams to offset losses could have a detrimental financial impact. Reliance on private equity or venture capital may hinder rapid growth and scale.

UPTIME INSTITUTE 2024 uptime INTELLIGENCE

Figure 17: SWOT Analysis of GPU cloud providers, highlighting strengths, weaknesses, opportunities, and threats.

Key Takeaway: While GPU cloud providers present a strong growth opportunity, hyperscalers' pricing pressure remains a key challenge.

Cost Breakdown of AI Training and Experiments

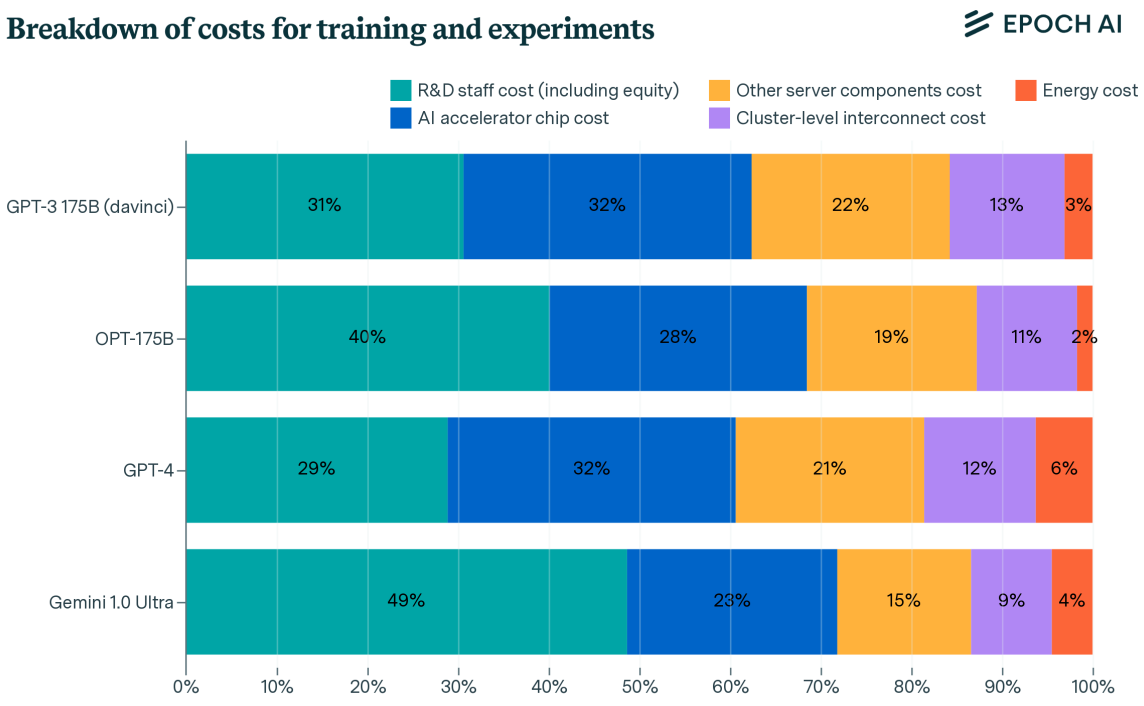


Figure 18: Breakdown of costs for training AI models, including R&D, AI chips, server components, interconnect, and energy.

Key Takeaway: AI training costs are driven by compute resources, particularly GPU accelerators, while energy costs remain a significant factor.

AI vs Human Performance Across Capabilities

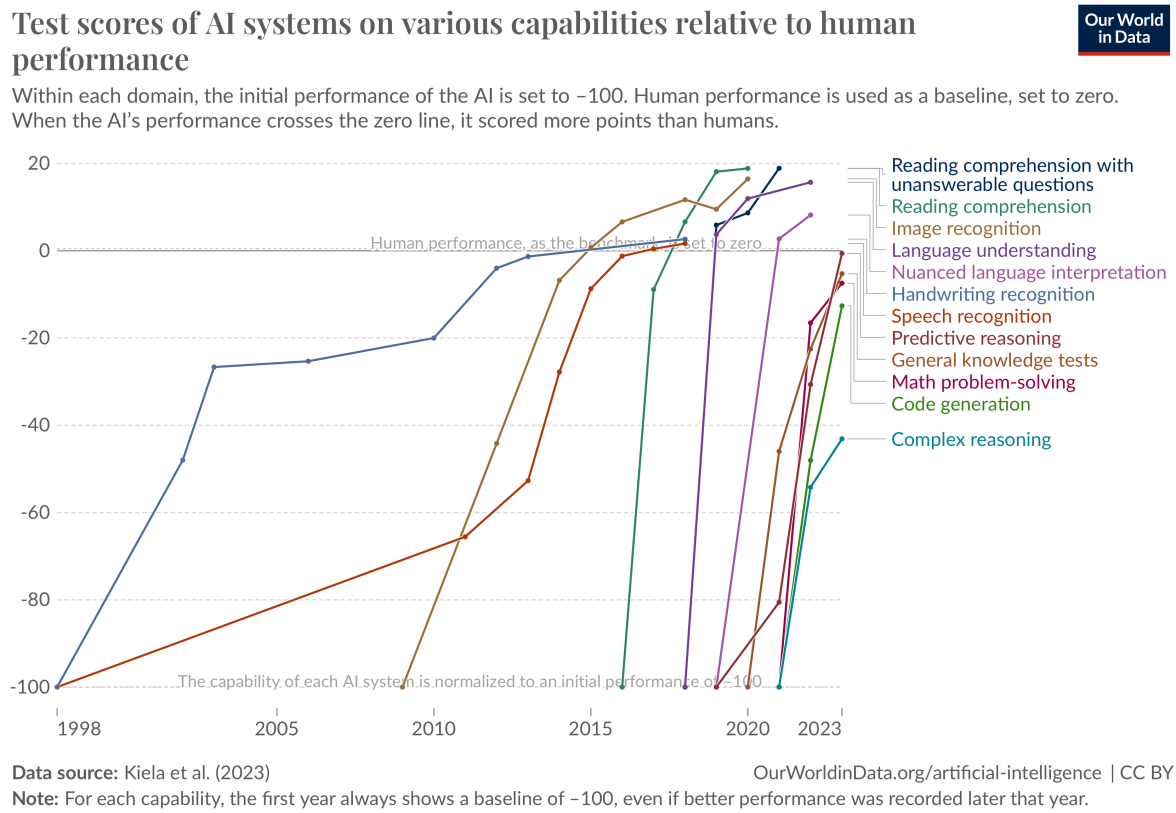


Figure 19: Test scores of AI systems on various capabilities compared to human performance.

Key Takeaway: AI models are now outperforming humans in several cognitive tasks, such as math problem-solving and code generation.

Data Center Power Infrastructure Schematic

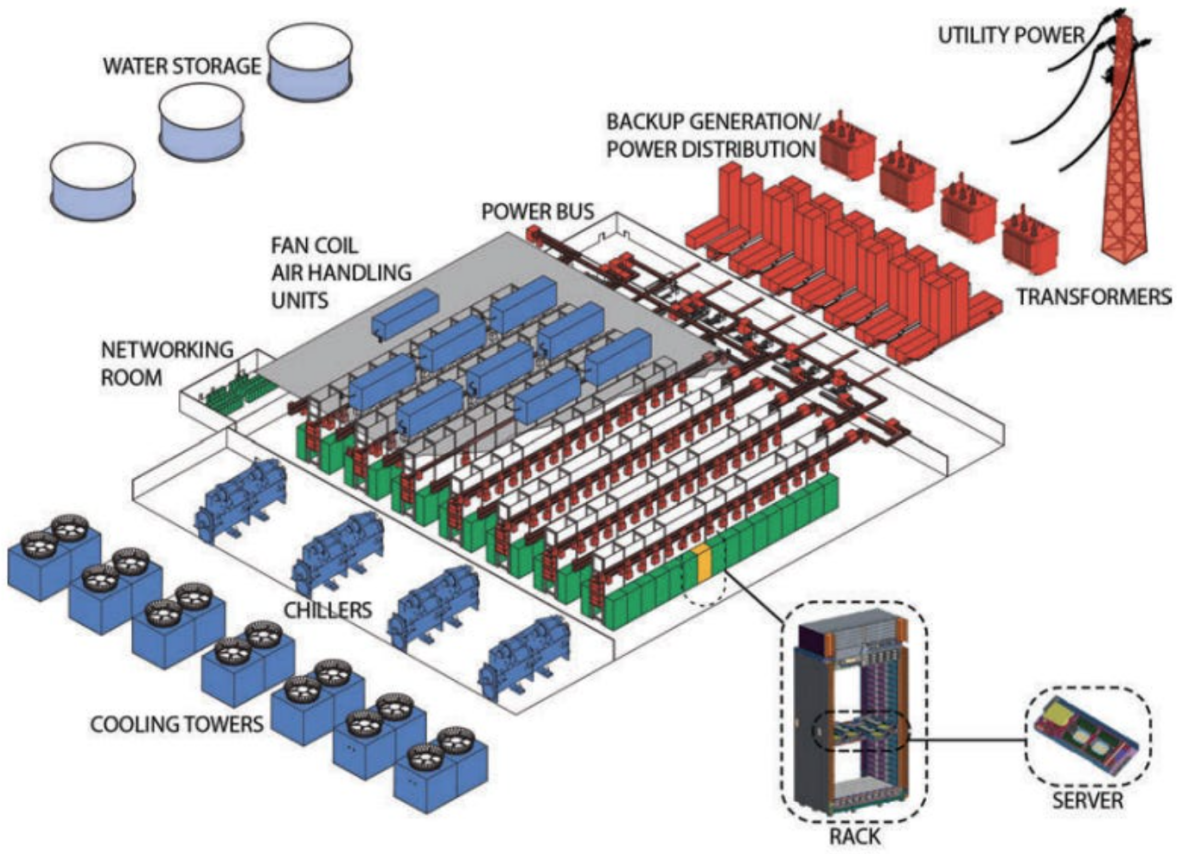


Figure 20: A power distribution schematic for a modern data center, including backup power, cooling, and networking.

Key Takeaway: Efficient power management is crucial for modern AI data centers, requiring robust infrastructure for cooling and redundancy.

Revenue of Leading Data Center Markets

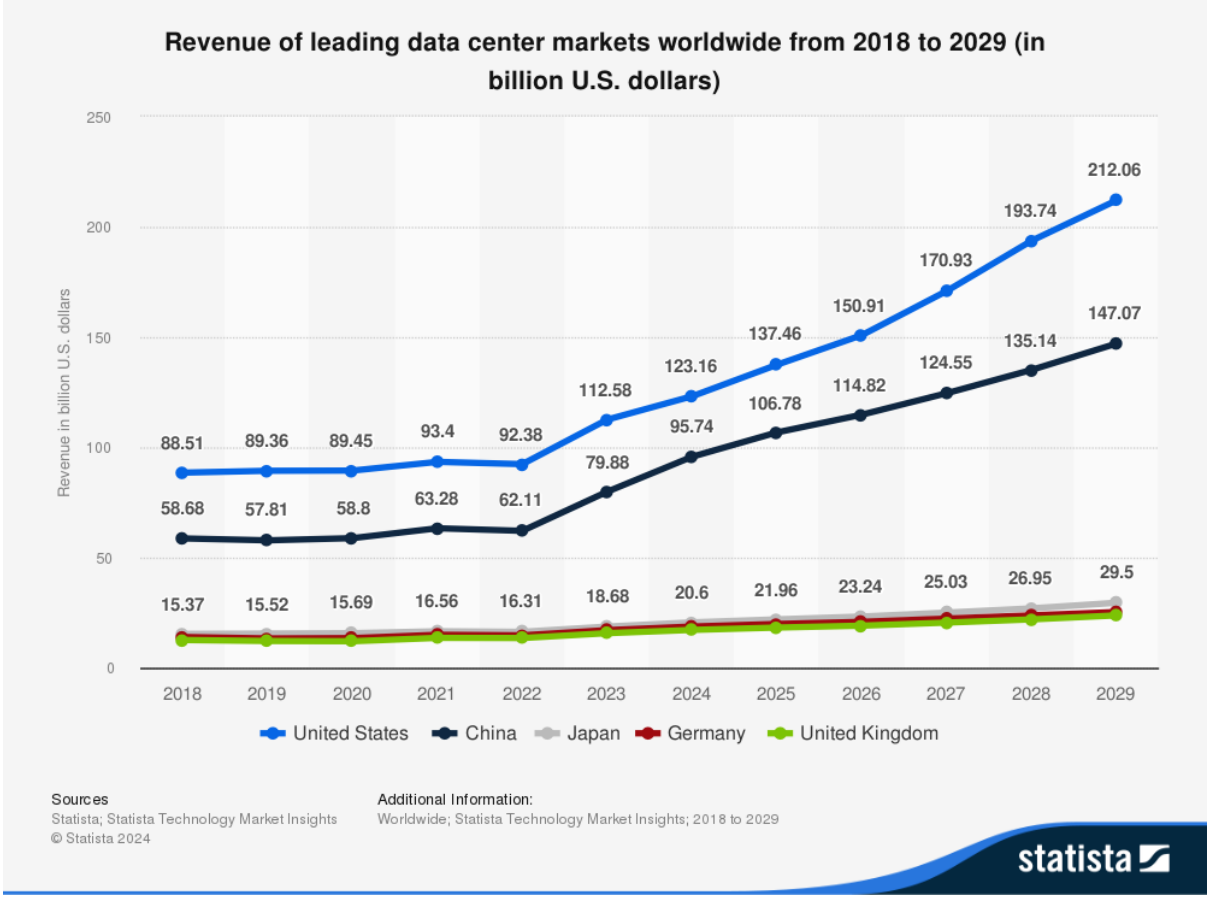


Figure 21: Projected revenue of major data center markets from 2018 to 2029, showing strong global expansion.

Key Takeaway: Data center markets are on a steady growth trajectory, with the U.S. and China leading in revenue.

Data Center Construction Costs by Market

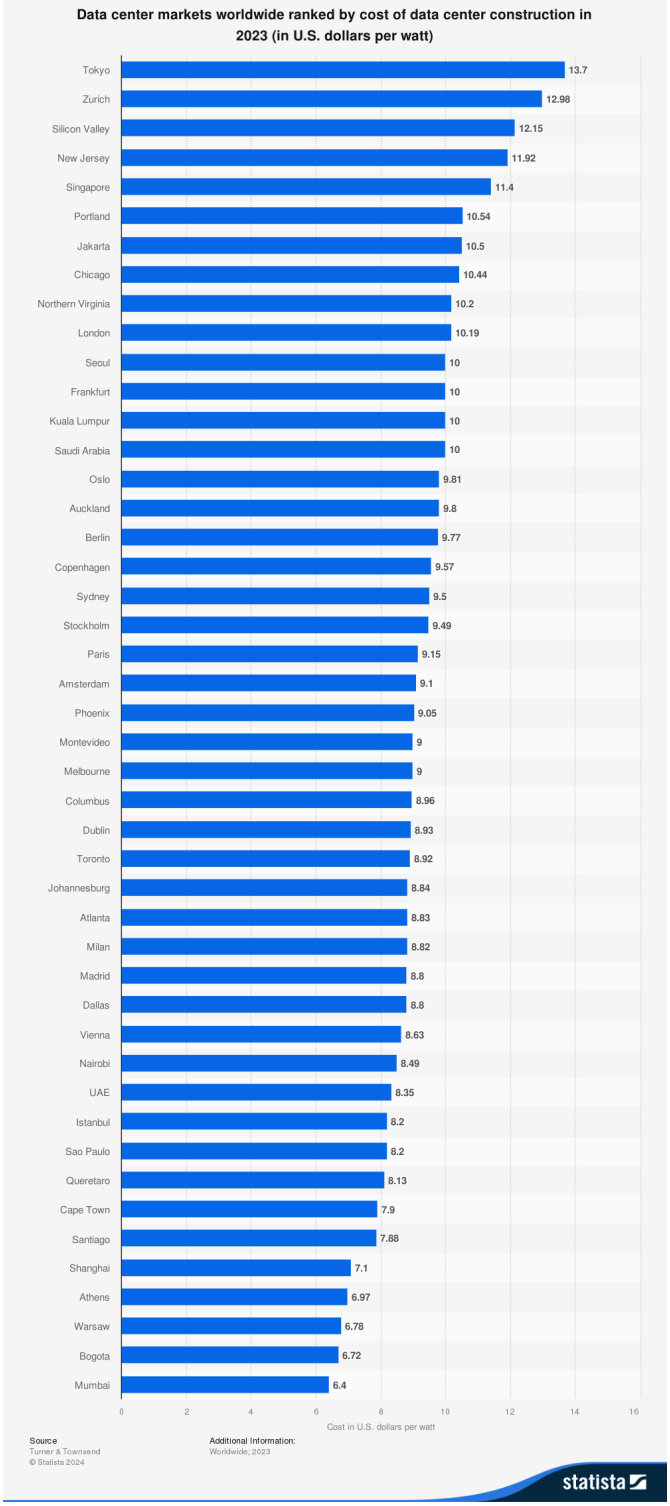


Figure 22: Global ranking of data center markets by cost of construction per watt in 2023.

Key Takeaway: Mumbai is the cheapest Market by cost of construction

Enterprise Spending on Cloud and Data Centers

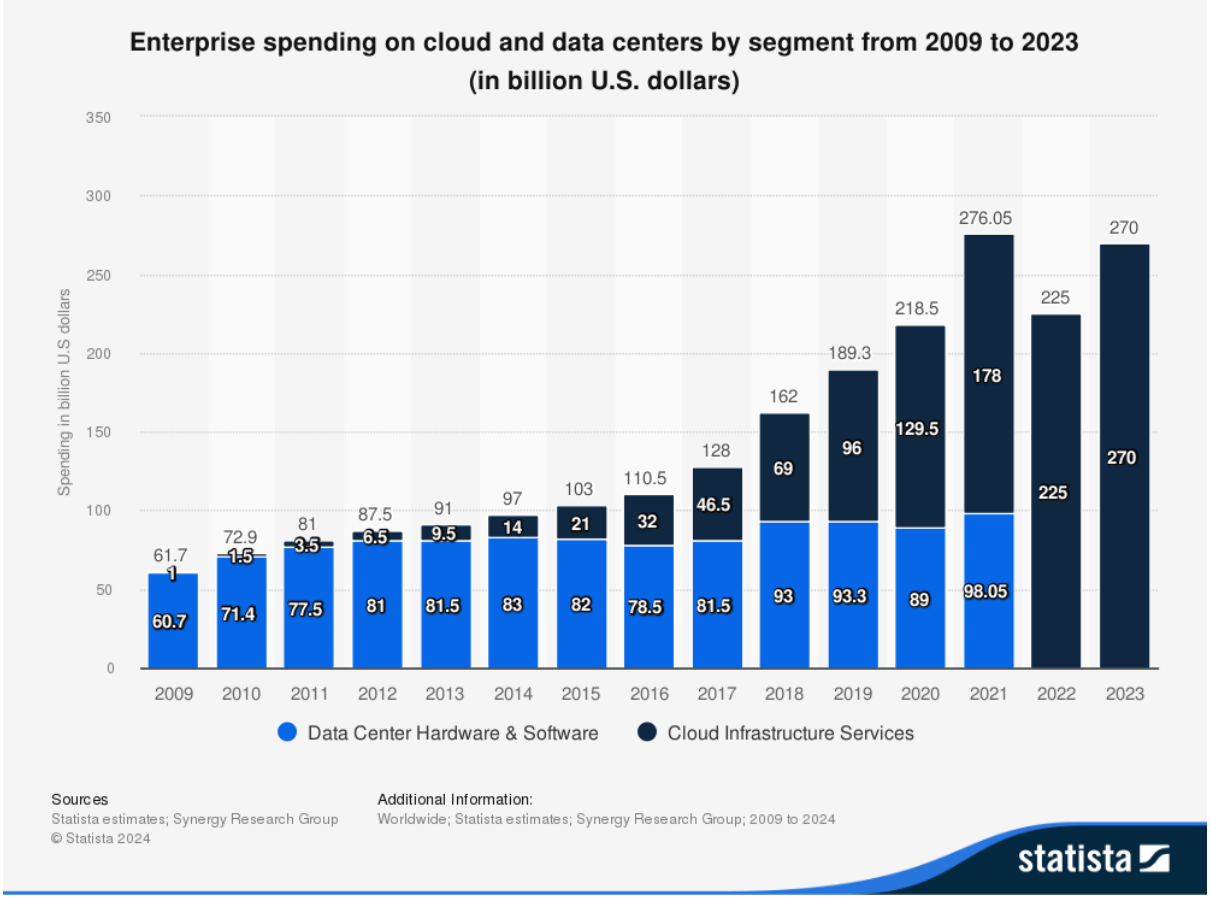


Figure 23: Enterprise spending on cloud and data centers from 2009 to 2023.

Key Takeaway: Cloud infrastructure services have overtaken hardware and software investments, indicating a major shift towards cloud-based solutions.

Worldwide Cloud Infrastructure Revenue Growth

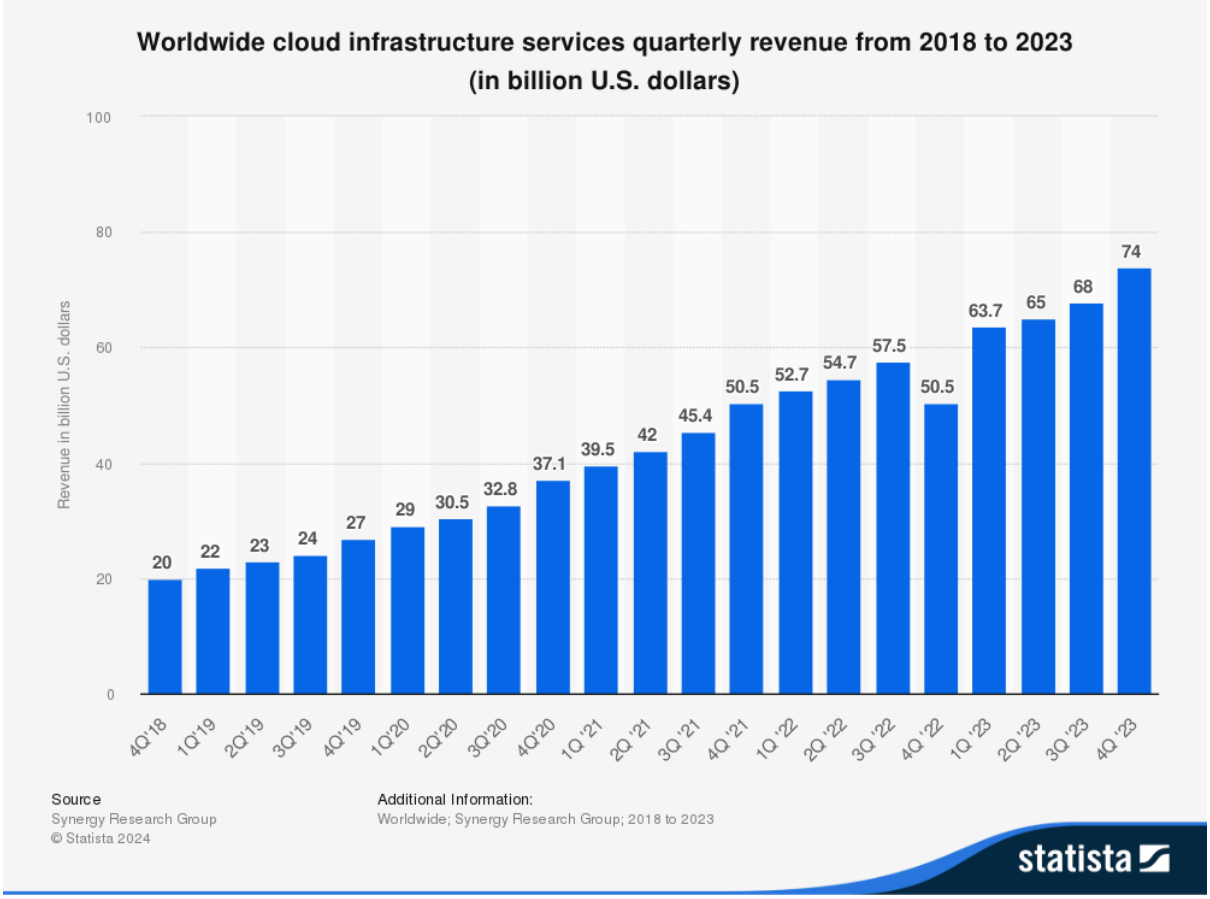


Figure 24: Cloud infrastructure revenue growth from 2018 to 2023.

Key Takeaway: The cloud market has shown consistent quarterly growth, reflecting strong enterprise adoption of cloud services.

Number of Data Centers in India (2024)

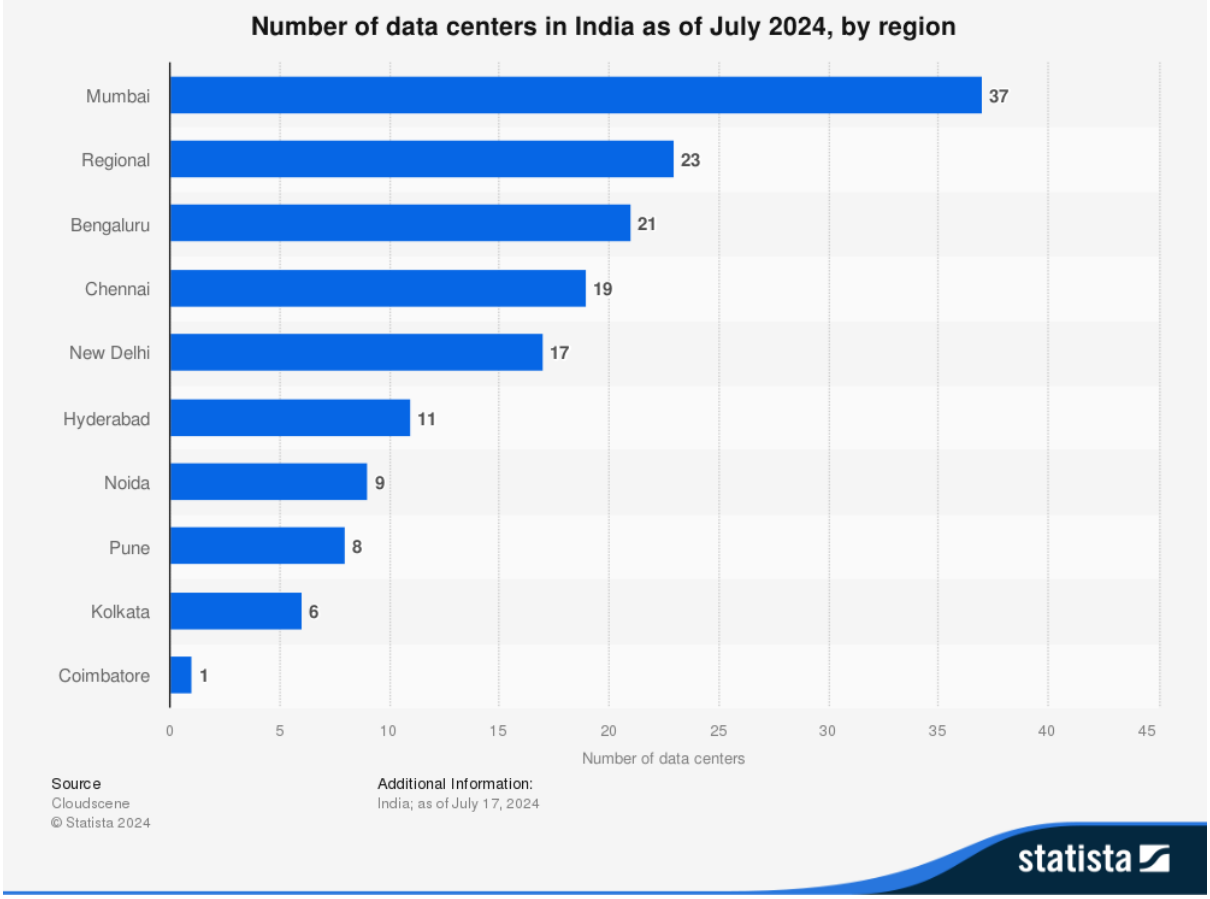


Figure 25: Regional distribution of data centers in India as of 2024.

Key Takeaway: Mumbai leads India's data center landscape, followed by Bengaluru and Chennai, indicating regional hubs for cloud infrastructure, Because of Undersea Fiber Optical Cable Landing Station there.

Leading Countries by Number of Data Centers (2024)

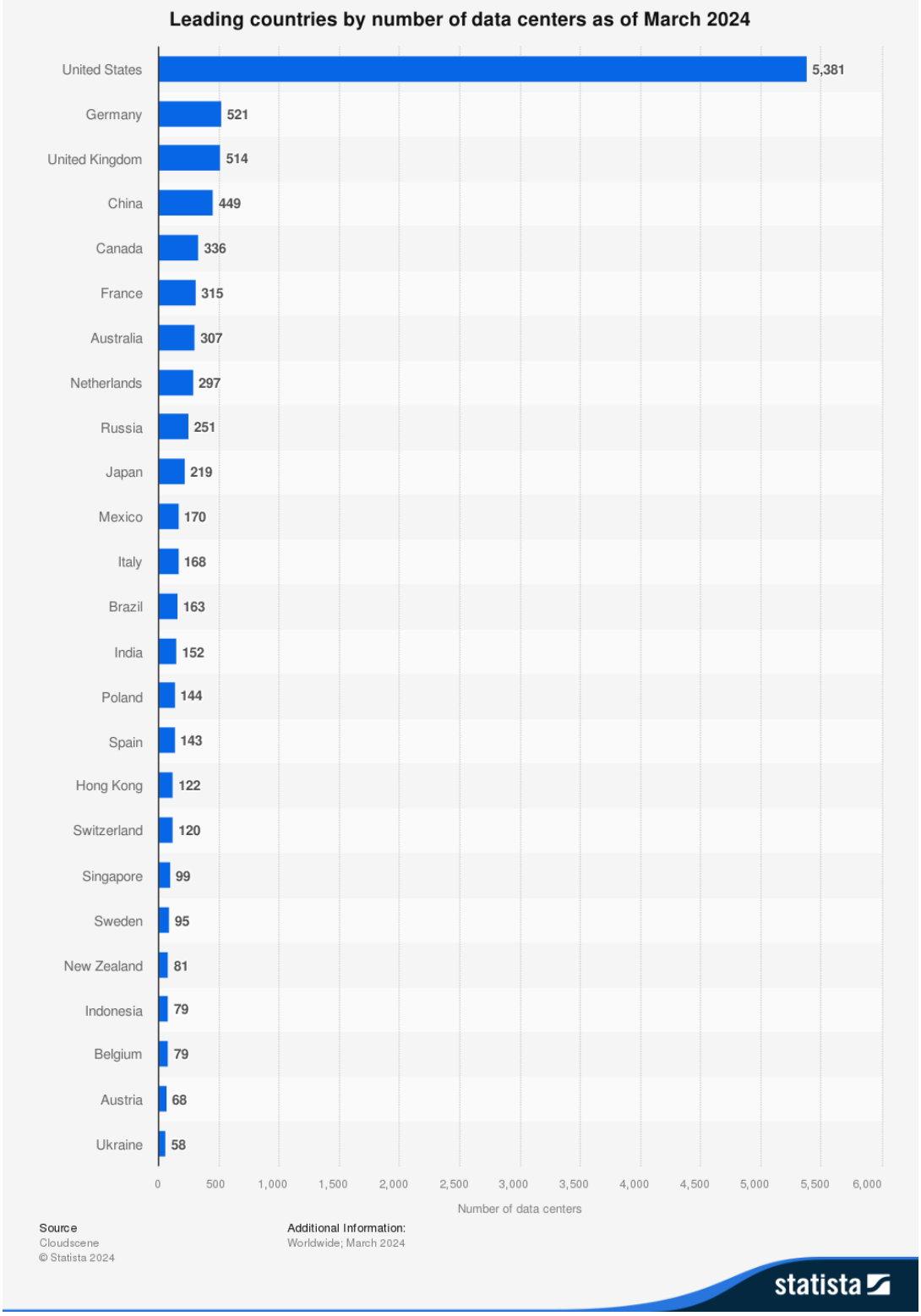


Figure 26: Global ranking of countries by the number of data centers.

Key Takeaway: The United States dominates the global data center market, far ahead of other countries. Indian needs to rapidly catch up for sovereignity reasons.

Top Countries by Computing Power

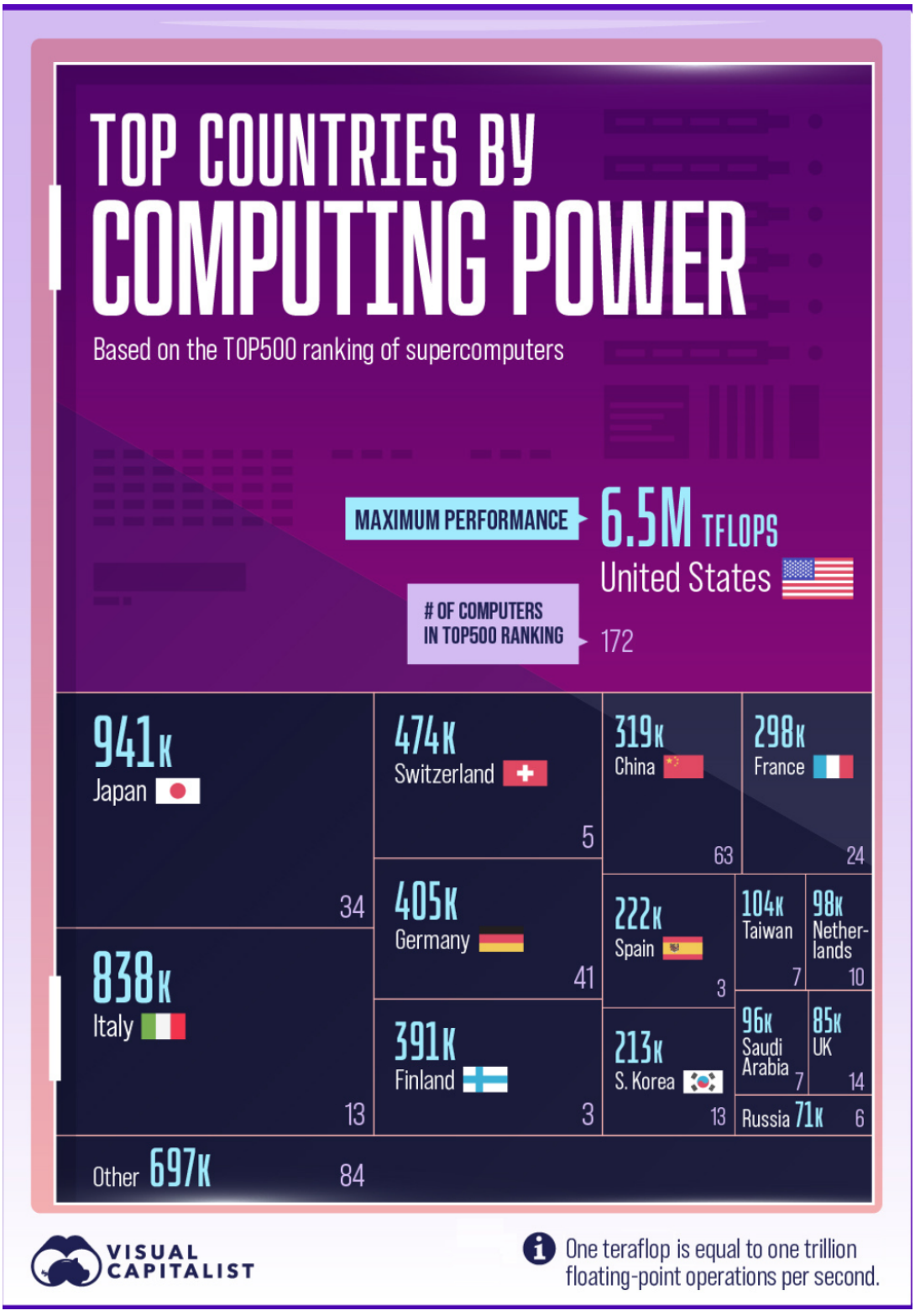


Figure 27: Top countries by computing power based on supercomputer rankings.

Key Takeaway: The United States leads in computing power, followed by Japan and Italy, reflecting their investments in HPC (High-Performance Computing).

Market Share of Cloud Infrastructure Providers

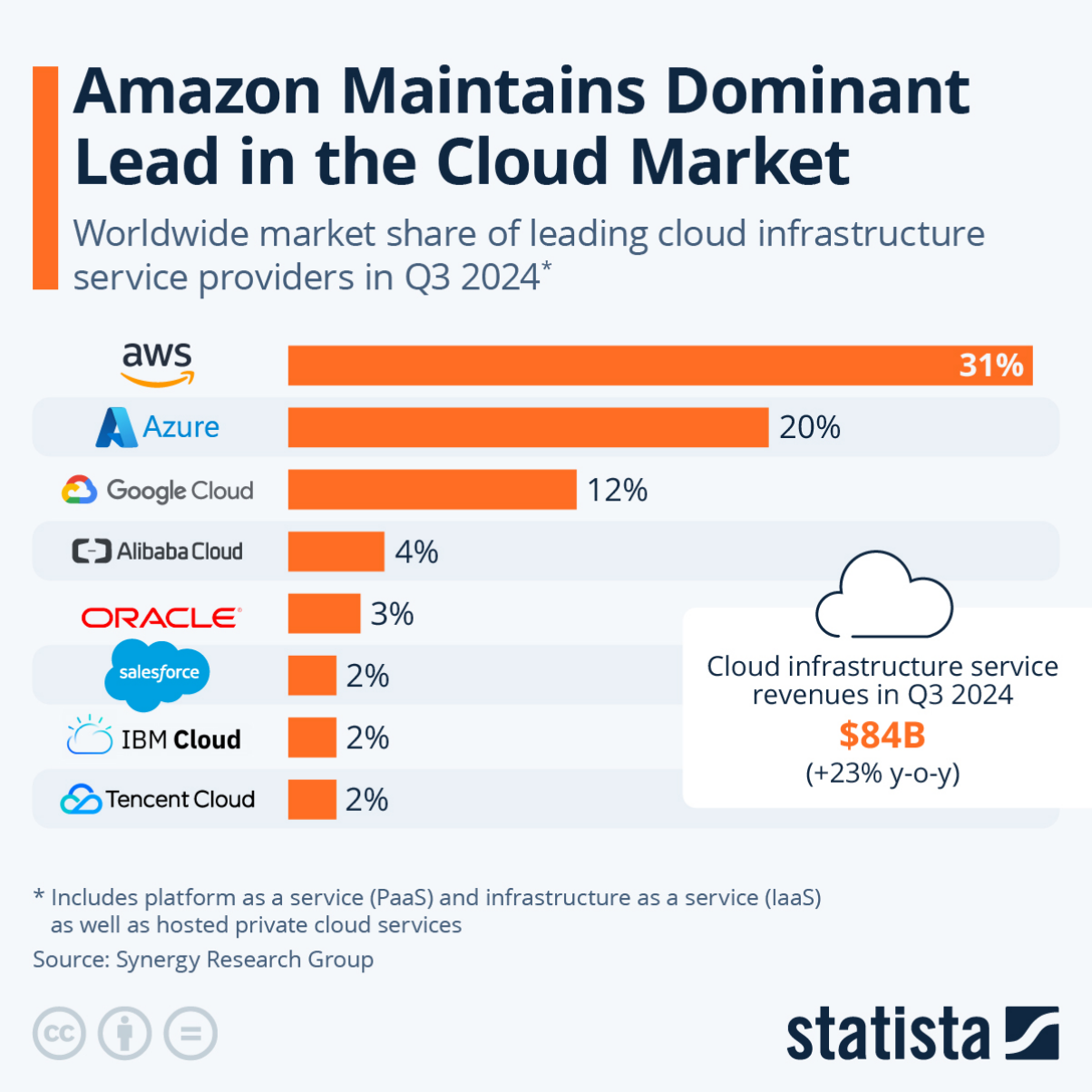


Figure 28: Market share of top cloud providers in Q3 2024.

Key Takeaway: Amazon Web Services (AWS) maintains its lead at 31% market share, followed by Azure and Google Cloud.

Energy Production by Country

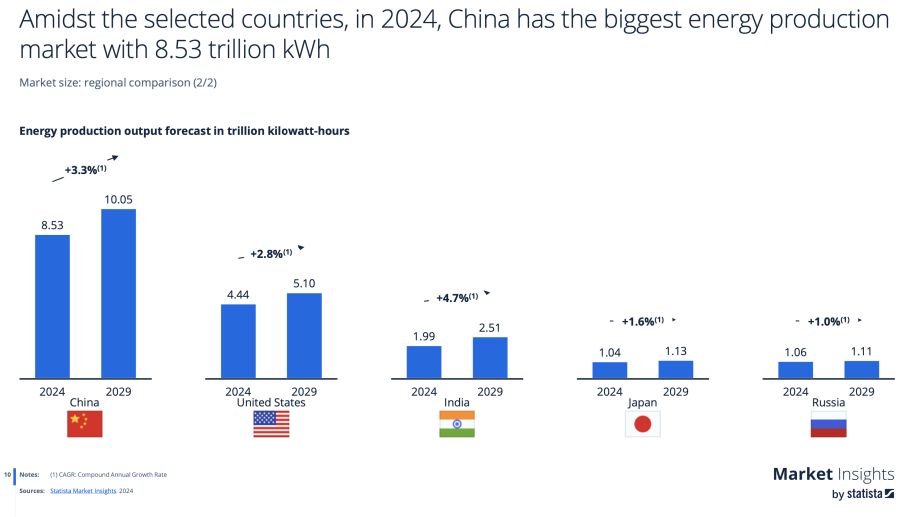


Figure 29: Energy production forecast for major economies by 2029.

Key Takeaway: China and the US will remain dominant energy producers, critical for supporting cloud and AI infrastructure growth. India has the highest growth rate.

AI Driving Demand for Data Centers

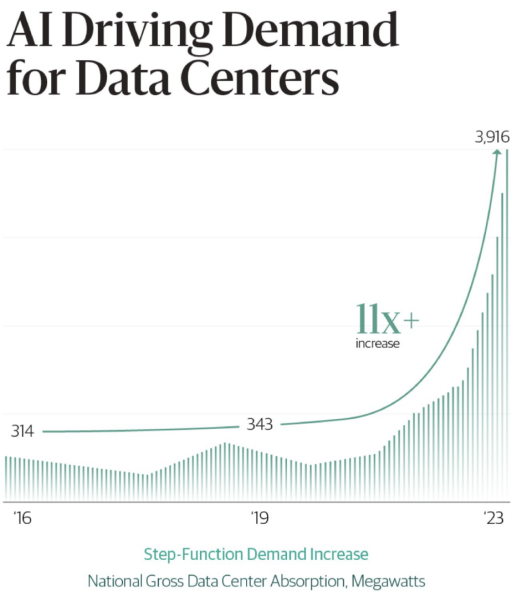


Figure 30: The impact of AI on data center demand growth.

Key Takeaway: AI workloads are driving exponential growth in data center capacity, with an 11x increase since 2016.

Server and Smart NIC Shipments

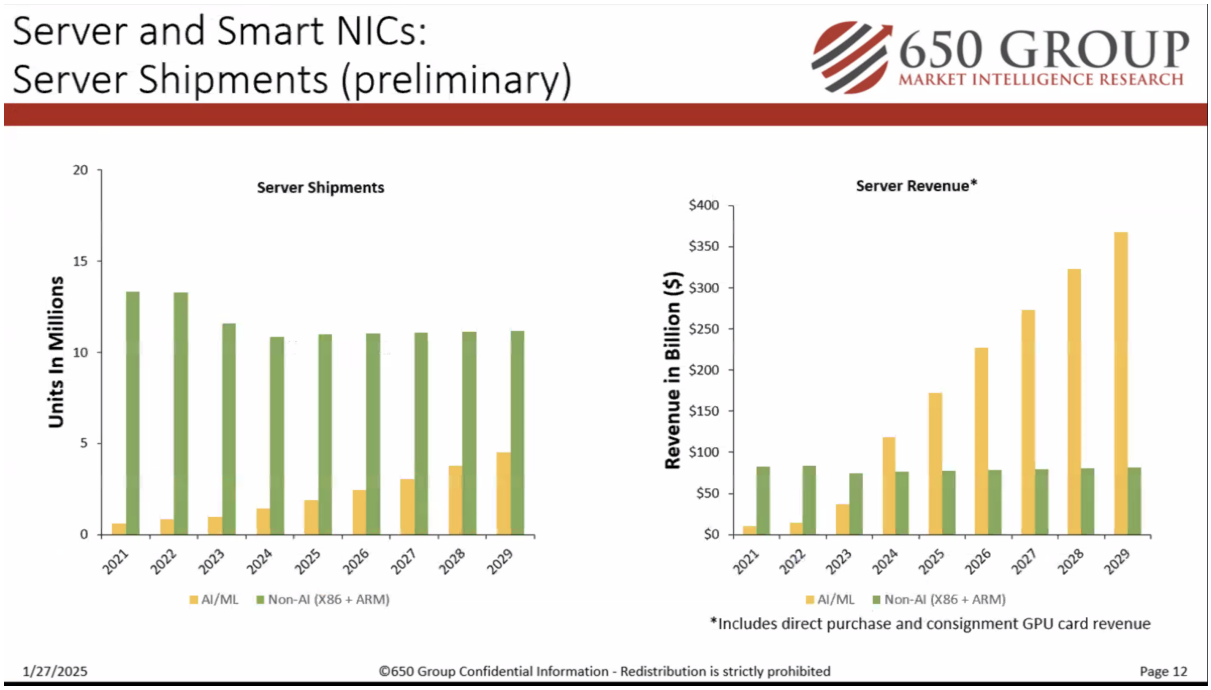


Figure 31: AI-based server shipments vs. traditional non-AI servers and revenue growth.

Key Takeaway: AI-driven server demand is accelerating, with revenue surpassing traditional non-AI server segments.

Data Center Equipment Market Shift

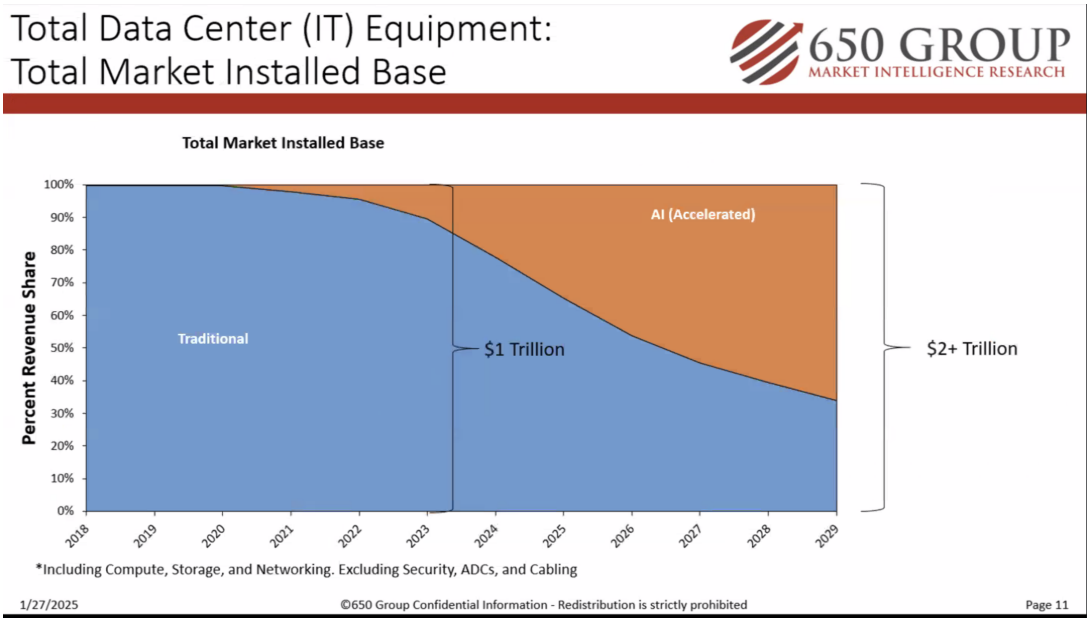


Figure 32: Shift from traditional data center infrastructure to AI-accelerated workloads.

Key Takeaway: AI workloads are replacing traditional compute, expected to drive over $2 trillion in investments by 2029.

North America Data Center Power Demand

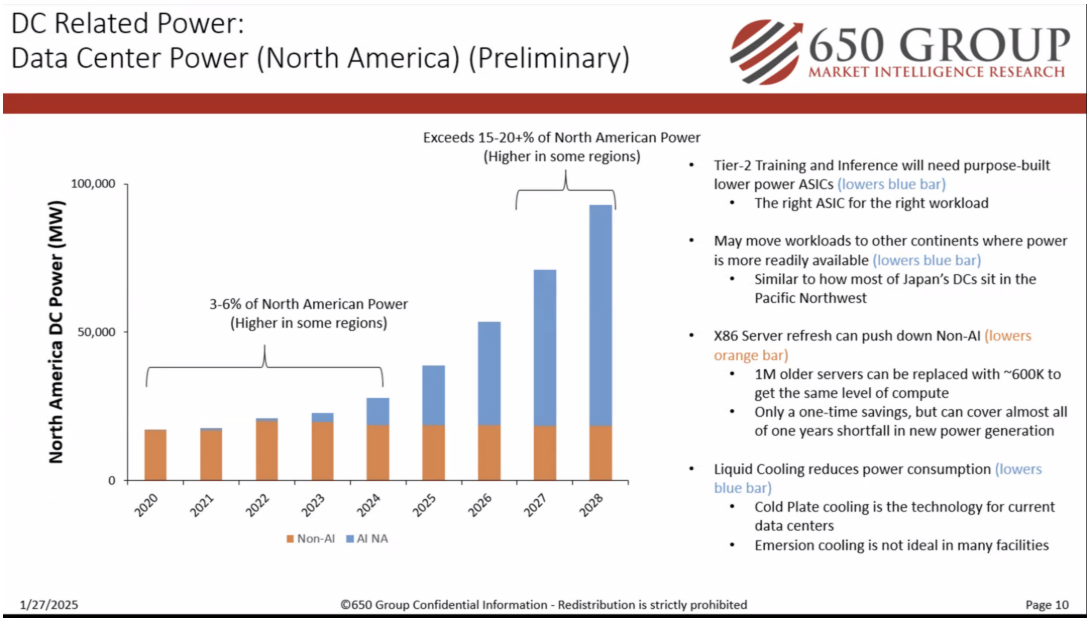


Figure 33: Power demand for North American data centers, with AI driving exponential consumption.

Key Takeaway: AI data centers could exceed 20% of North America's total power consumption by 2028.

Data Center Semiconductor Revenue Growth

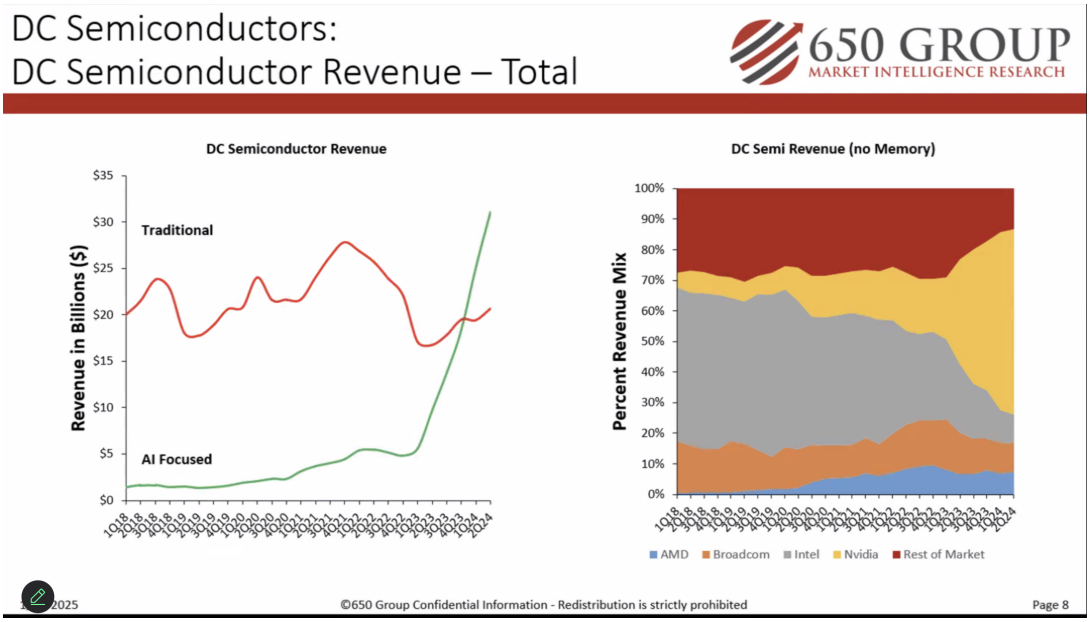


Figure 34: AI-focused semiconductor revenue surpassing traditional semiconductor sales in data centers.

Key Takeaway: AI-specific chips are overtaking general-purpose semiconductors, with NVIDIA leading the market.

Cloud Capital Intensity and Revenue Growth

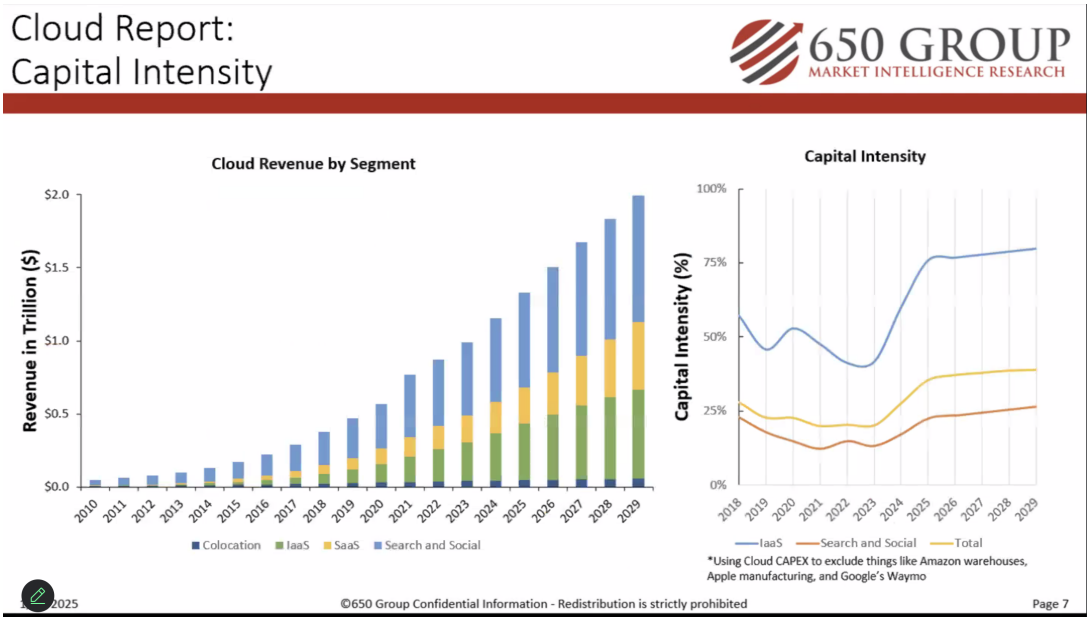


Figure 35: Rising capital expenditures in cloud infrastructure, dominated by AI investments.

Key Takeaway: AI workloads are driving increased capital intensity in cloud data centers, requiring continued investments.

Data Center Construction Costs

How Much Does it Cost to Build a Data Center?

Data center build costs are best characterized on a **relative** basis, as measured by square footage (sqft) and megawatts (MW). This is because there are various types of data centers, such as retail and wholesale, as well as sizes of facilities, including 1-megawatt and 50-megawatt, where scale provides a build cost advantage.

Square Footage (sqft)

Below is an illustrative example of the total development cost components for a **greenfield** data center, based on **gross square footage**, which generally ranges between **$625 to $1,135 per gross sqft**. The two key cost groupings are those related to the **powered shell** and **data center improvements**.

Data Center – Total Development Costs Example

Cost Component	Low Cost/sqft	High Cost/sqft
Land	$25	$75
Building Shell	$80	$160
Subtotal – Powered Shell	***$105***	***$235***
Electrical Systems	$280	$460
HVAC / Mechanical / Cooling	$125	$215
Fire Suppression	$15	$25
Building Fit-Out	$100	$200
Subtotal – Data Center Improvements	***$520***	***$900***
Total Development Costs	**$625**	**$1,135**

Figure 36: Estimated cost per square foot for building a data center.

Key Takeaway: Data center costs vary significantly, ranging from $625 to $1,135 per square foot, depending on infrastructure.

Data Center Equipment Lifespan

How Long Does Data Center Equipment Last?

Data center equipment, including generators, power distribution units (PDUs), uninterruptible power supplies (UPS), switchgear / transformers, and HVAC / cooling systems all have useful lives of 20+ years. As a reference point, Dupont Fabros, which was acquired by Digital Realty (NYSE: DLR), historically disclosed the useful lives over which the major components of its data center properties were depreciated:

Data Center Component	Category	Useful Life
Power Distribution Units	Electrical Infrastructure	20 years
Uninterrupted Power Supply	Electrical Infrastructure	25 years
Switchgear / Transformers	Electrical Infrastructure	30 years
Heating, Ventilating and Air Conditioning	Mechanical Infrastructure	20 years
Chiller Pumps / Building Automation	Mechanical Infrastructure	25 years
Chilled Water Storage and Pipes	Mechanical Infrastructure	30 years
Fire Protection	Other	40 years
Security Systems	Other	20 years

Notably, one component of a data center that needs to be replaced much more frequently are batteries, which safeguard facilities from power disruptions. In the event of a power outage, batteries supply power first, followed by backup generators, which take about a minute to turn on. Typically, batteries have a useful life and are replaced every 5 to 6 years.

Figure 37: Average lifespan of major data center equipment.

Key Takeaway: Core data center components, such as power systems and cooling infrastructure, have a 20-30 year lifespan.

Modern AI Stack for Generative AI

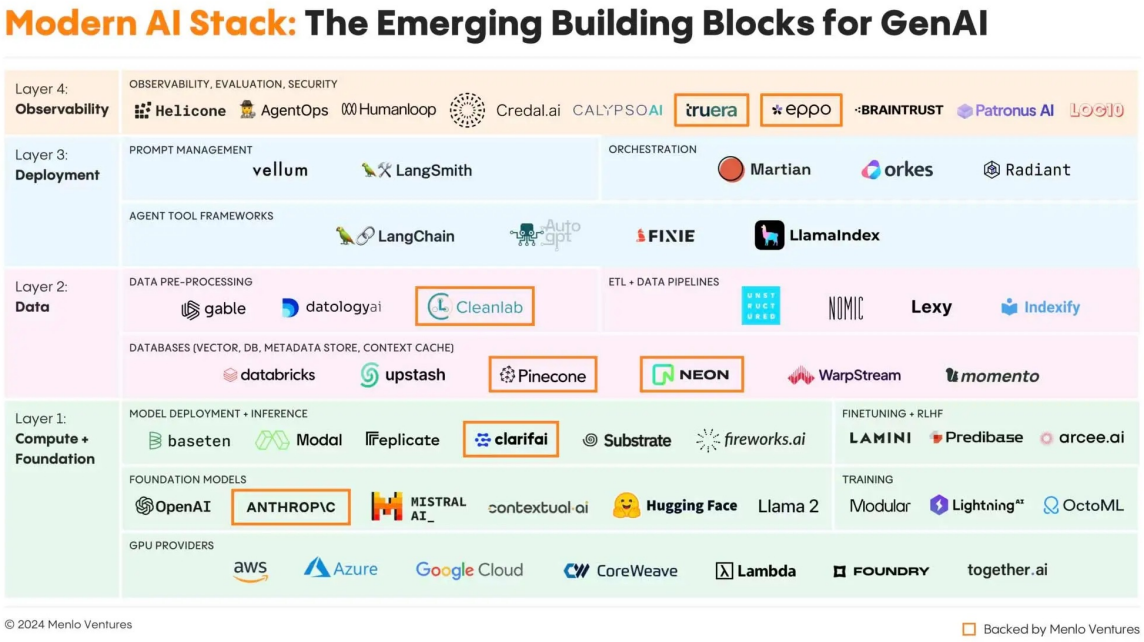


Figure 38: The Modern AI Stack: Key infrastructure components for GenAI development.

Key Takeaway: AI deployment is structured into multiple layers, from compute and data processing to observability and security, showcasing an evolving AI ecosystem.

AI GPU Clusters in India - Market Overview

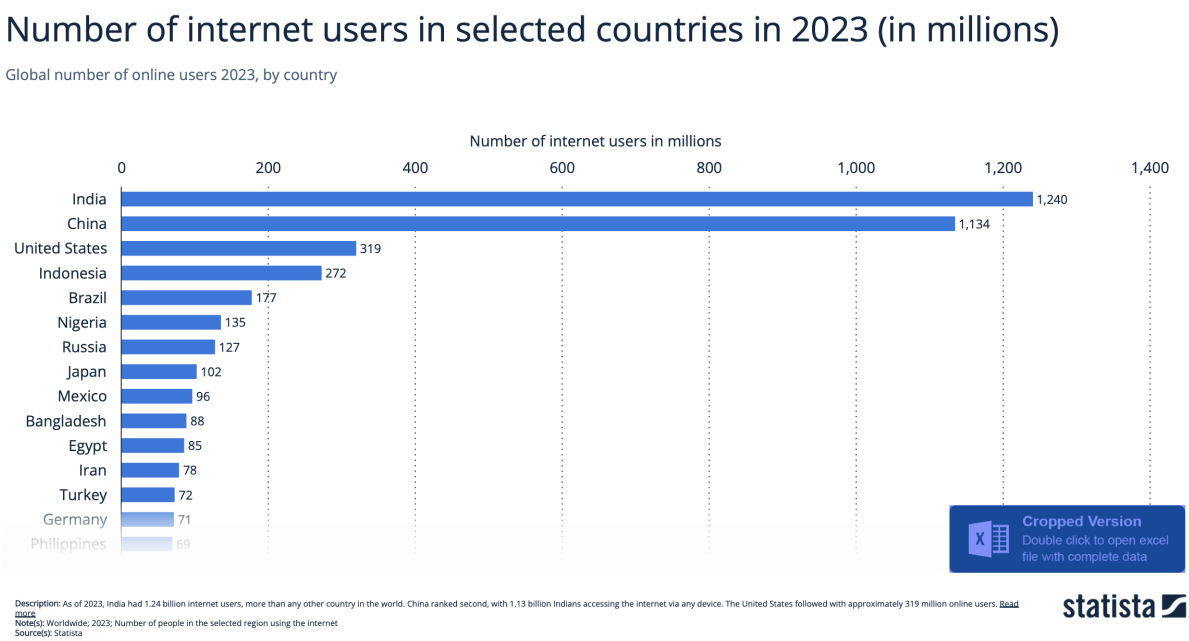


Figure 39: India's AI GPU cluster market growth potential.

Key Takeaway: India is emerging as a key market for AI GPU clusters, with increased cloud and data center investments driving adoption.

Declining Vacancy Levels in Data Centers

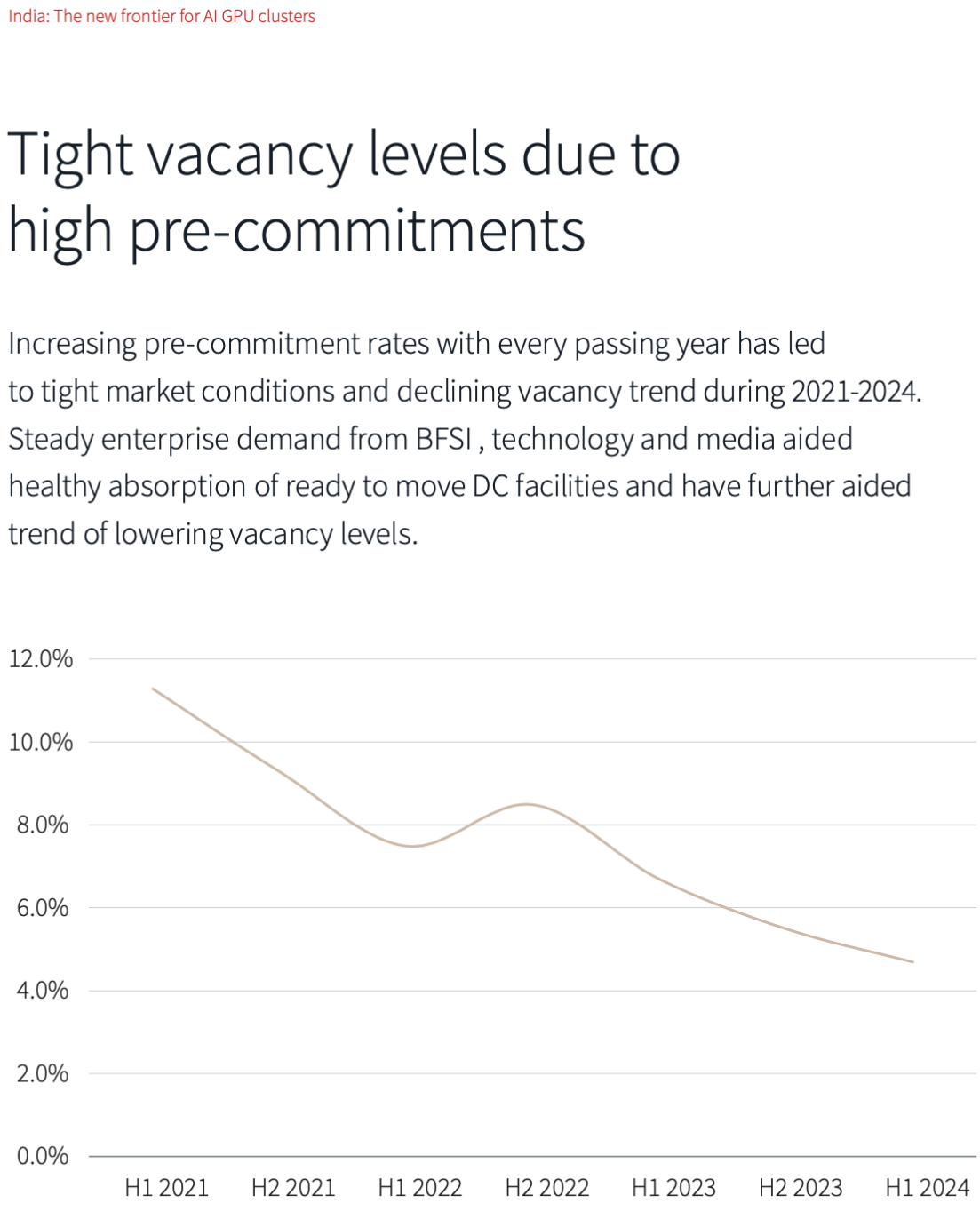


Figure 40: Data center vacancy trends in India (2021-2024).

Key Takeaway: Rising pre-commitments from BFSI, tech, and media firms have resulted in declining vacancy levels, signifying strong demand.

Cooling Technologies for High-Density AI Data Centers

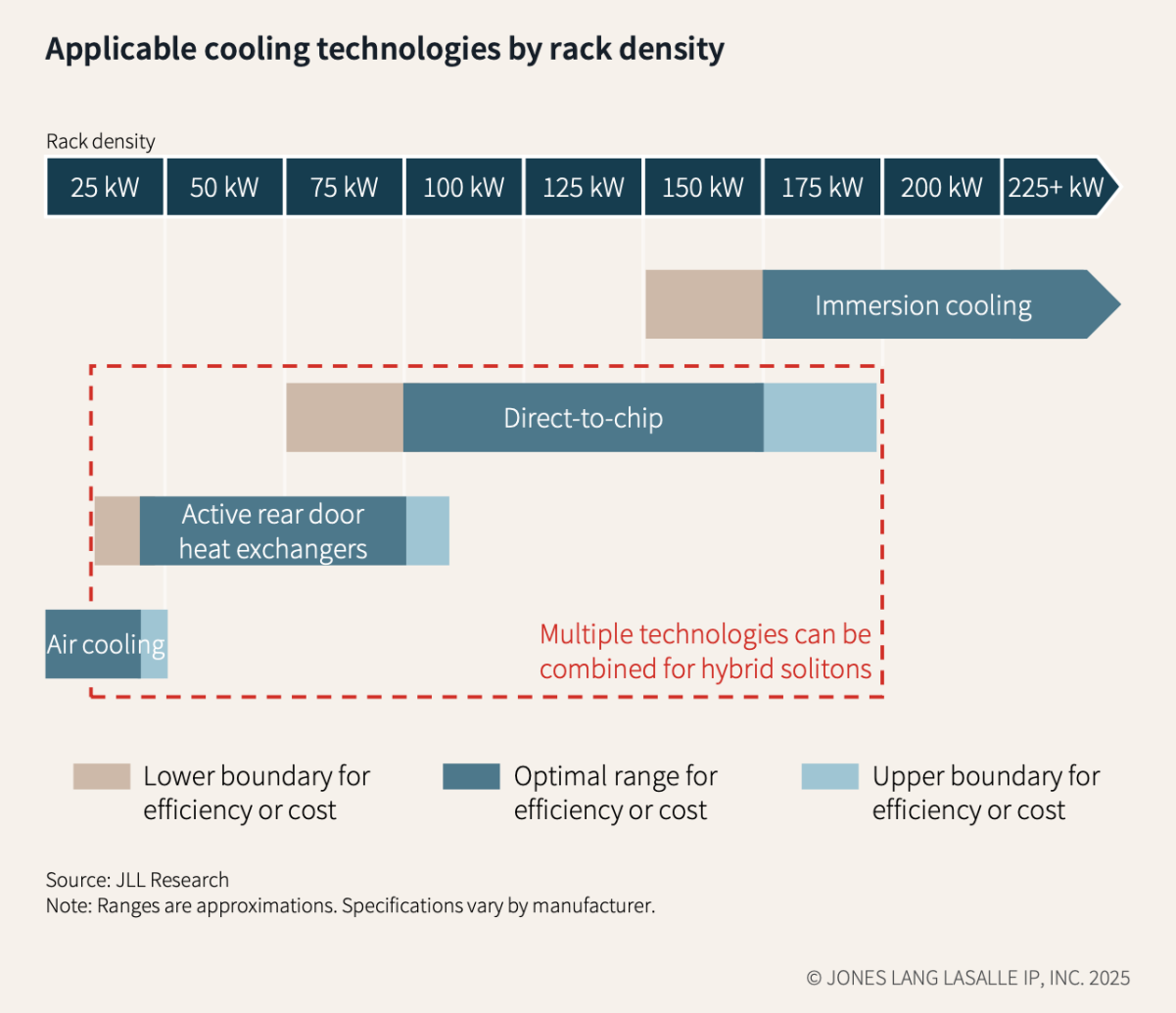


Figure 41: Cooling technologies for AI and HPC data centers based on rack density.

Key Takeaway: Advanced cooling techniques such as immersion cooling and direct-to-chip are becoming essential for high-density AI workloads.