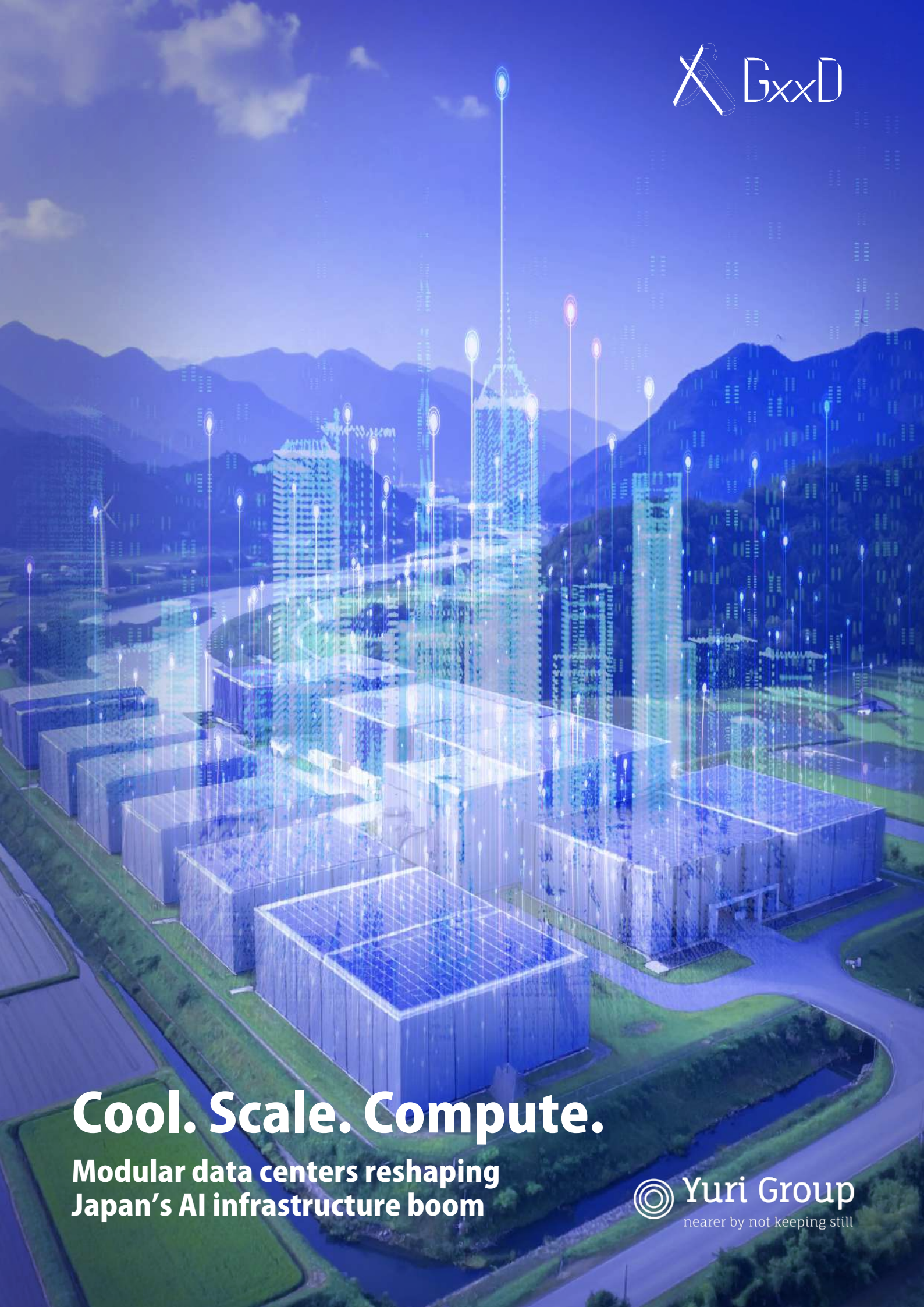

GxxD
Cool. Scale. Compute.
Modular data centers reshaping
Japan's AI infrastructure boom
Yuri Group
nearer by not keeping still

The global AI boom has triggered a stampede for new data centers — and most major IT firms want them yesterday. That's good news for one Japanese company. While traditional data centers take around three years to build, Getworks Co. can deliver a custom facility in just 4–6 weeks.

The company brings the industrial trend of modularization to the data center market. Rather than constructing entire facilities on-site, Getworks assembles most of the data center in a factory, packages it into container units, and ships them to client locations for rapid deployment.

Crucially, Getworks does not see this as a niche product. Its leadership argues that modular, container-style data centers are becoming a necessity, particularly in an era of fast-evolving GPU technologies. Conventional data centers — large, permanent buildings requiring years of planning, grid connection, and community approval — are often outdated by the time they go live. Modular centers offer a faster, more adaptable solution.

Speed is a competitive edge. Japan is experiencing a data center boom driven by both local and global forces: growing cloud adoption, soaring internet traffic, and a government-led push for digital transformation. Digital infrastructure investment in Japan is expected to at least double by 2030.

Power is another factor. Data centers are energy-intensive, not only for computing but also for cooling. While the industry norm favors centralized gigawatt-scale power sources such as thermal and nuclear plants, Getworks is betting on a different model: decentralized, renewables-powered systems. Its container data centers are designed to run on clean energy — a strategy that aligns with Japan's policy goals to develop rural data hubs and reduce the load on congested urban grids. The rural hubs are expected to be used for AI training and are likely to woo domestic players, rather than hyperscalers.

> TRADITIONAL DATA CENTER BUILDINGS CAN BE INFLEXIBLE AND SLOW TO DEPLOY IN EMERGENCIES, WHEREAS CONTAINERIZED FACILITIES OFFER RAPID DEPLOYMENT AND ADAPTABILITY. GETWORKS HAS BEGUN RECEIVING INQUIRIES FOR PROJECTS ON THE SCALE OF SEVERAL HUNDRED MEGAWATTS. TO MEET THIS SURGE IN DEMAND, WE ARE INVESTING HEAVILY IN EXPANDING PRODUCTION CAPABILITIES.
>
> Getworks Business Integration Manager, Hayashi Ryutaro

Getworks is currently the only Japanese firm building and operating container-based data centers in the country. According to Yuri Group estimates, these modular centers could make up 5–10% of total data center capacity in Japan by 2030.

Game Changer

Perhaps the most powerful force behind data center expansion today is generative AI. At the heart of this trend are GPU (graphics processing unit) servers. In contrast to CPU (central processing unit) based servers, which are the standard for general-purpose computing, GPUs offer much higher performance for compute-intensive workloads like deep learning, simulations, and image processing. GPU servers are significantly larger than CPU servers and require significantly more power and cooling.

Historically, GPU adoption has been driven by research institutions, universities, and AI startups focused on model development and training. But with the explosion of generative AI since 2023, demand has surged across sectors — from IT to pharmaceuticals, and from advanced manufacturing to scientific research.

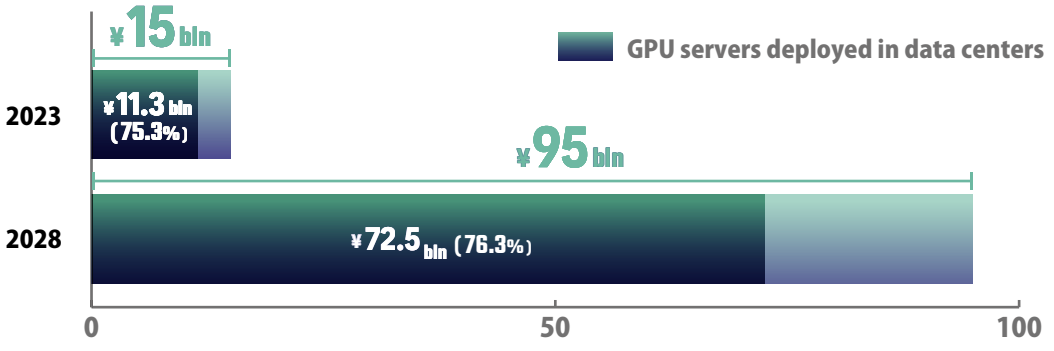
JAPAN'S GPU SERVER MARKET IS EXPECTED TO GROW 44.7% ANNUALLY BETWEEN FY2023 AND FY2028, SOARING FROM ¥15 BILLION TO ¥95 BILLION IN SHIPMENT VALUE. DATA CENTER DEPLOYMENTS WILL ACCOUNT FOR 76.3% OF THAT MARKET BY 2028, REFLECTING THE STRATEGIC ROLE OF GPUS IN ENTERPRISE AND CLOUD INFRASTRUCTURE.

Use cases now include building large language models (LLMs), drug discovery, robotics, and autonomous driving. This rush has led to global chip shortages and long lead times, yet the upward trajectory remains clear: more firms are adopting GPU-based AI tools, and few are likely to scale without data center-level infrastructure.

That's because GPU servers require specialized conditions: high power capacity, advanced cooling systems, and noise-isolated spaces. These constraints often exceed what in-house server rooms can handle, pushing demand toward dedicated data centers. At the same time, cloud providers are scaling up their GPU-enabled offerings, creating additional demand for colocation-ready, AI-optimized facilities.

The result: the data center segment of the GPU server market is set to remain the dominant growth driver — and a critical infrastructure priority for Japan's digital economy.

Japan GPU Total Server Market Size



Source: Fuji Chimera Research

Growth Impact

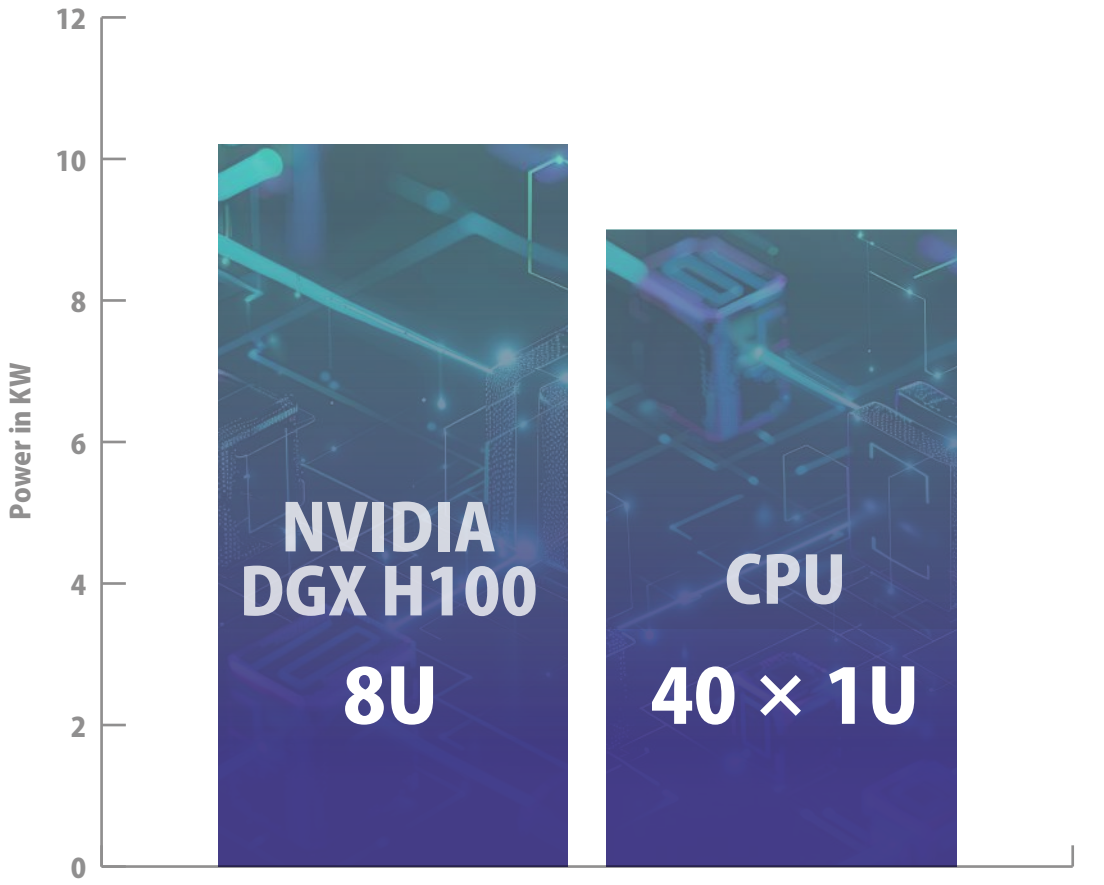
As GPU deployments accelerate, power requirements are rising just as fast — and GPU servers are emerging as a significant source of demand in next-generation data centers.

Today, a GPU server cluster requires approximately 67 MW of dedicated power capacity. Within just a few years, that figure is expected to jump to 212 MW for a single hyperscale deployment consisting of 5 to 8 data center buildings, according to IDC forecasts. Importantly, this figure refers only to GPU servers — it excludes additional energy needs for cooling systems, networking gear, or other supporting infrastructure.

According to the above forecast, power demand for GPU-based AI servers alone will rise more than threefold (3.2x) over the coming years.

To put this into perspective, an NVIDIA DGX H100 system — a commonly used GPU server — occupies 8 rack units (8U) and consumes up to 10.2 kW at peak load. By contrast, a standard CPU-based server may occupy just one rack unit (1U) and consumes significantly less energy. Even with 40 stacked 1U CPU servers, the total power draw remains under 10 kW. In other words, one 8U GPU server consumes around the same amount of power as 40 CPU servers — but in a much denser, hotter form.

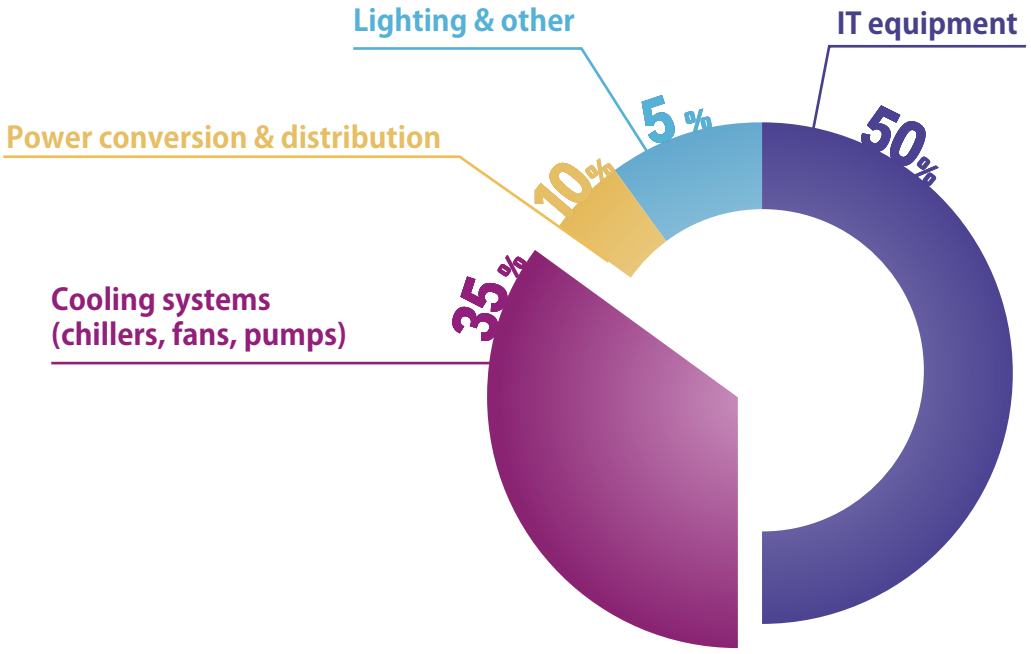
5x increase in power and space requirements



Source: Telecommunication magazine, June 2024

This dramatic increase in power density is forcing data center operators to rethink their infrastructure strategies, especially in Japan where land, grid capacity, and cooling options are constrained.

Energy consumption breakdown in a typical data center



Source: Ministry of the Environment

Bottleneck

In a typical data center, 30–40% of total energy goes to cooling systems — including chillers, fans, and pumps. In a 1 MW facility, that means 300–400 kW just for cooling. And as GPU power loads rise, so too does the heat that must be managed.

Chillers are often the biggest energy consumer in traditional, air-cooled systems. Improving their efficiency is critical to lowering Power Usage Effectiveness (PUE) and overall energy costs.

To meet the demands of modern AI infrastructure, technologies like liquid cooling and immersion cooling are gaining global traction. These methods offer far more efficient heat transfer — and are increasingly seen as essential in high-density deployments.

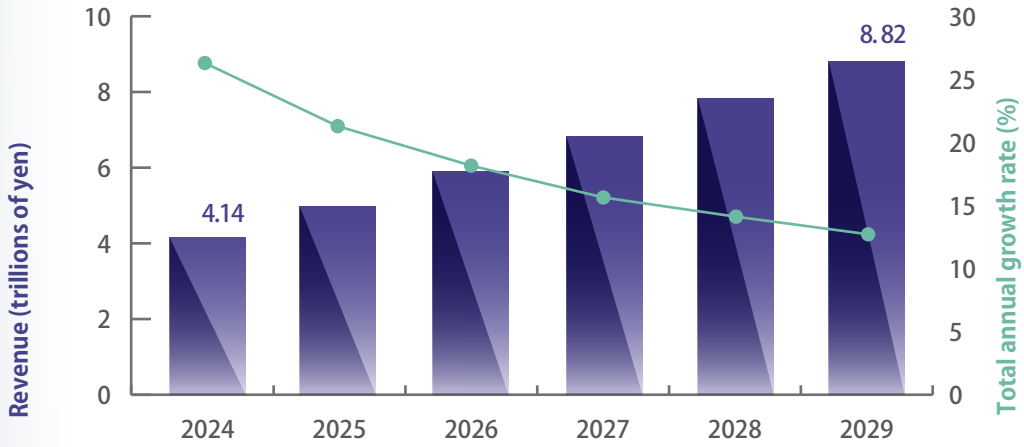
Yet Japan lags in adoption. Despite the benefits, liquid cooling uptake remains limited. One reason: most chillers produce water at very low temperatures (below 10℃), which isn't always necessary for server cooling. If not carefully managed, this can lead to condensation and hardware failure. Designing systems to avoid condensation is complex — and a major barrier to broader rollout.

Another challenge is lack of standardization. While multiple vendors now offer water-cooled server models, they differ in required water pressure and cooling specs depending on the CPU or configuration. For data center operators, this complicates infrastructure planning and slows down adoption.

Every Cloud

Japan's public cloud services market grew by 26% in 2024, reaching ¥4.14 trillion, according to IDC Japan. This growth is expected to continue, with a projected CAGR of 16.3% from 2024 to 2029, pushing the market to ¥8.82 trillion by the end of the forecast period.

Japan's Public Cloud Services Market



Source: IDC Japan, 2/2025

A major catalyst is the rise of Generative AI. In 2024, cloud products and services incorporating GenAI saw significant sales increases. While the shift of easily portable systems — such as web apps and packaged software — has largely plateaued, there has been a marked acceleration in migrating legacy and custom-built systems to the cloud.

Looking ahead, the public cloud — and by extension, data center demand — will continue to be driven by:

- **Ongoing migration of complex systems**
- **Adoption of cloud-native architectures**
- **Integration of GenAI functionalities**
- **Investment in digital transformation initiatives**

Already, SaaS offerings with GenAI features are being deployed in productivity-focused use cases: language translation, document summarization, keyword extraction, content generation, and programming assistance.

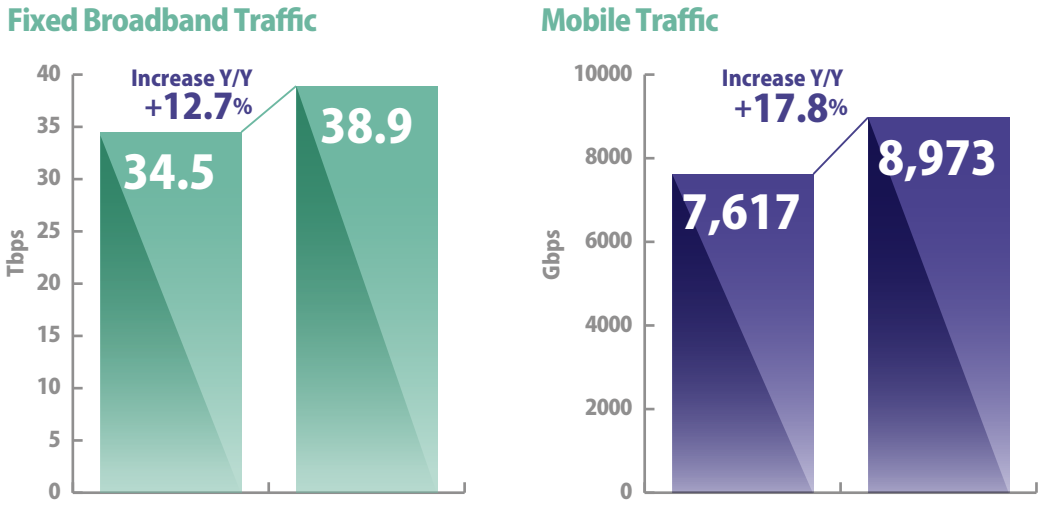
That said, adoption in core business processes remains limited. Challenges include data preparation, workflow standardization, and visualization. In 2024, many early adopters began proof-of-concept trials to explore solutions.

One area of notable momentum is the development of AI agents — autonomous systems that make decisions and take actions by interfacing with external software. These agents are designed to reduce manual input, automate workflows, and enhance productivity.

As these technologies mature, AI agents are expected to play a central role in driving GenAI adoption across business functions, expanding the scope of automation beyond back-office tasks.

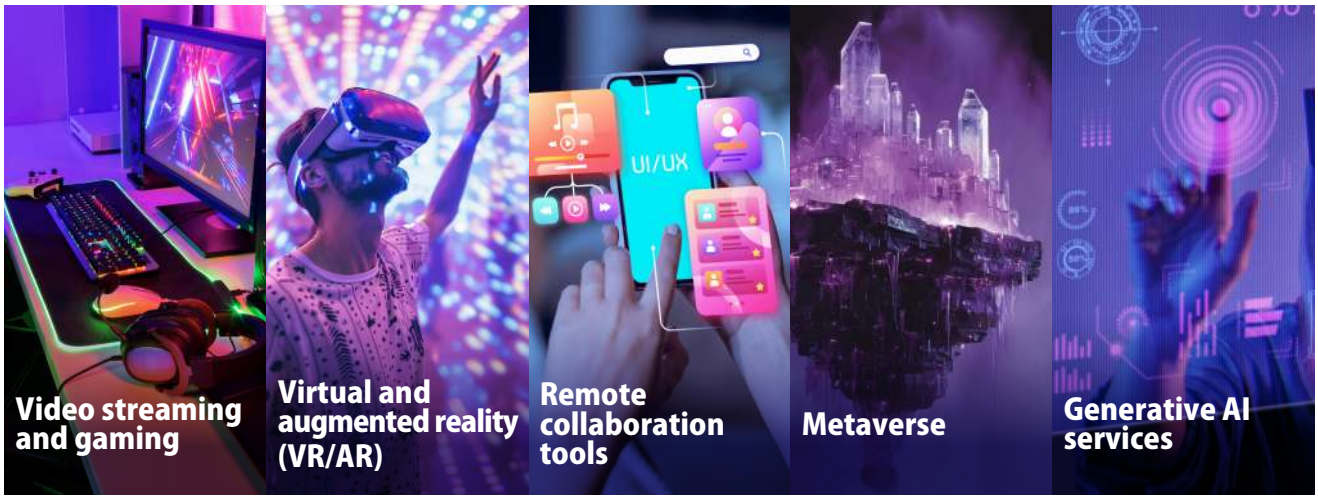
Internet Traffic

Alongside GenAI, a quieter — but equally important — force driving data center demand is the steady rise in internet traffic, which year on year was in the double digits for both fixed broadband and mobile traffic at the end of 2024.



Source: Ministry of Internal Affair and Communication

This surge reflects the growth of bandwidth-heavy services including:



All of these are compute-intensive, often requiring deployment across both hyperscale and edge data centers. The result: growing demand for high-performance CPUs and GPUs, faster storage, and ultra-low-latency networks. Mobile use cases like AR/VR and live video require edge computing to reduce latency. In parallel, core data centers must scale up to support cloud workloads, AI model training, and long-term data storage.

To keep pace, Japan's data center operators are:

- **Expanding existing facilities**
- **Investing in new campuses**
- **Upgrading interconnect bandwidth**

While Tokyo and Osaka remain dominant hubs, there's growing momentum in regional areas where land, power, and cooling infrastructure are more readily available — and where the government is encouraging digital decentralization.

Digital Transformation

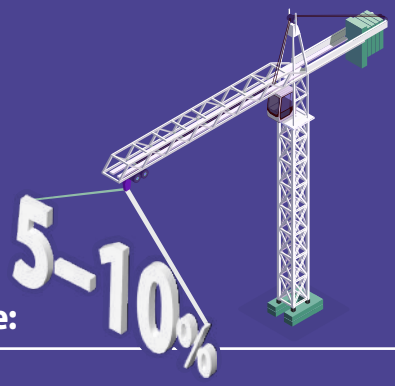
60% of Japan's mission-critical computer systems are now 20 years or older, making them well behind the technological curve. There has been increasing clamour to address that issue since the early 2020s, with a growing number of Japanese companies viewing upgrades to digital systems as less of a technology issue, and more a core management priority.

Yet, despite an ongoing government drive to promote what it calls DX (digital transformation), that shift in mindset is only now starting to show up in corporate budgets. According to Fuji Chimera Research, total DX-related investment in Japan is projected to nearly double by FY2030, reaching ¥8.35 trillion.

That growth will continue as companies overhaul legacy systems, adopt cloud and AI tools, and modernize customer-facing platforms. All of these digital upgrades increase Japan's total demand for compute power and, by extension, for data centers.

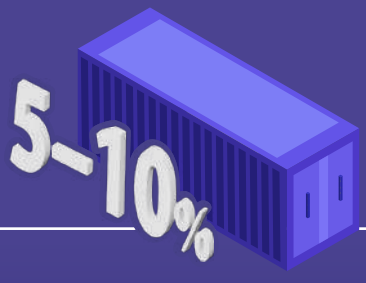
Data Center Market Snapshot

Estimated DX spending in Japan by FY2030: **¥8.35 trn**



Estimated share allocated to data center infrastructure:

Yuri Group's estimate for the share that will go to container-based solutions:



Resulting market opportunity for containerized data centers:

>¥83.5 bln by 2030

Fast and Nimble

> ELECTRICITY SHORTAGES ARE INCREASINGLY BECOMING A REALITY — A CHALLENGE MADE MORE URGENT BY THE RAPID GROWTH IN GPU SERVER DEPLOYMENTS, WHICH DEMAND FAR HIGHER ENERGY LOADS THAN TRADITIONAL IT SYSTEMS.
>
> Getworks Business Integration Manager, Hayashi Ryutaro

Among the Japanese firms capitalizing on AI infrastructure demand is Getworks, a developer specializing in custom container-based data centers. Founded in 2002, the company is not a hyperscaler — large-scale data center clusters typically operated by tech giants like Google Cloud and Amazon Web Services — but its solutions are scalable, fast to deploy, and tailored for high-density GPU workloads.

Getworks began focusing on modular container data centers in 2013, positioning itself ahead of the curve as AI workloads surged and energy constraints tightened. The company can complete a full 1 MW-class deployment in as little as four months — far faster than the three to five years typically required for conventional data center construction and at a fraction of the cost.

Container vs Brick and Mortar

This speed is increasingly valuable in an AI-driven market where server specs, especially for GPUs, evolve rapidly. Traditional facilities often require retrofits to handle the heat, power density, and cooling demands of today's generative AI systems. Getworks, by contrast, designs its containers for liquid cooling, custom airflow, and high-performance computing from day one.

A single 20-foot container can house up to 32 liquid-cooled GPU servers, and Getworks is scaling up production to meet growing demand — including interest from hyperscale clients seeking rapid-deployment solutions

Decentralized Compute

Beyond speed, Getworks' container model helps address Japan's geographic and energy challenges. These include a highly mountainous terrain and densely packed urban communities in the center of the country that are separated by large distances from energy production sites in rural areas. The main sources of renewable energy are, for example, in the far north in Hokkaido and in the far southwest of the country in Kyushu – far from the manufacturing bases and population centers of Tokyo and Osaka.

The top-down, centralized nature of the national power grid means that electricity supply is also at risk of disruption from factors such as Japan's frequent and often powerful earthquakes, grid congestion related to – for example – oversupply of solar during the summer months, and heavy demand from dense urban centers putting strain on supply. By distributing smaller-scale facilities across regions, Getworks aims to reduce these risks for companies in need of reliable outsourced compute power.

This decentralized approach also allows for deployment in rural or infrastructure-limited locations where large, grid-dependent campuses are not feasible. Getworks also designs for renewable energy integration, enabling siting near surplus solar, biomass, or other distributed power sources.

Green Leaning

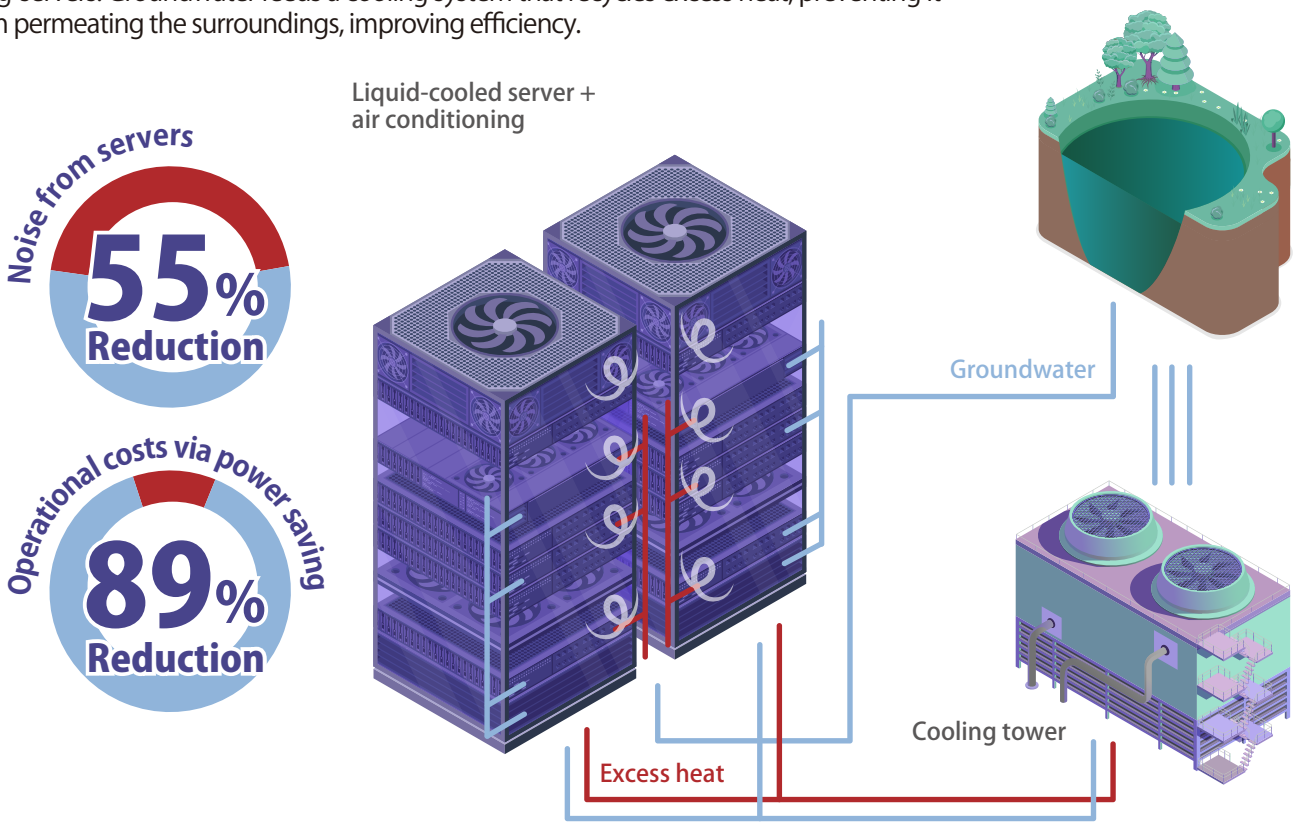
> We believe that energy efficiency and responsible sourcing are no longer optional. While fossil fuels or nuclear energy may offer short-term relief, they are not scalable or sustainable solutions in the long run. In many cases, grid-based power is simply not available in sufficient quantities to support new data center builds — especially in rural or infrastructure-limited locations. That is why we see a strategic advantage in renewable energy.
>
> Getworks Business Integration Manager, Hayashi Ryutaro

Beyond contributing to decarbonization goals, Getworks argue that renewables offer a practical solution for powering high-density computing where traditional grid infrastructure falls short. The company has made it a core policy to prioritize renewable energy as a primary power source — not just for environmental reasons, but as a business-critical resource that enables stable growth in energy-constrained markets

In line with this philosophy, Getworks incorporates energy-efficient cooling systems, including free-air cooling, liquid cooling, and well water-based heat exchange — a technique rarely used in Japan's modular DC sector.

Liquid Cooling Container Solution

The rack-type layout allows for air conditioning units to be placed close by the heat generating-servers. Groundwater feeds a cooling system that recycles excess heat, preventing it from permeating the surroundings, improving efficiency.



Source: Getworks

The U.S. leads the world in the number of data centers, with its total surging past 5,000 facilities last year. That's more than double the combined total of EU states, which is still gargantuan in comparison to Japan's 219 facilities. Still, there are a number of factors that indicate that the local market for data centers will grow quickly and that Getworks will be one of the players worth paying attention to.

Implications

- Getworks offers a rare combination of speed, flexibility, and specialization in a market where hyperscaler data center builds can take 3–5 years — giving it a fast-mover advantage in a slow-moving industry.
- Japan's AI and cloud adoption is accelerating, with public cloud spending expected to more than double by 2029 and GPU server demand growing at over 40% annually — driving structural demand for new data center capacity.
- Unlike legacy facilities, Getworks' modular data centers are engineered for high power density and liquid cooling from the start — ideal for GenAI workloads that push conventional infrastructure to its limits.
- Its container-based design supports flexible siting and rapid deployment, allowing Getworks to serve customers in regions where real estate, grid access, or seismic risk constrain traditional data centers.
- The company's green-leaning strategy — pairing compute infrastructure with local renewable energy sources — aligns with Japan's decarbonization goals and growing pressure for sustainable digital infrastructure.
- Getworks is expanding its production footprint and reports growing interest from hyperscale clients, signaling potential to move from niche provider to key player in Japan's AI infrastructure landscape.
- Japan's fragmented mid-market offers opportunities for Getworks to partner with cloud service providers, enterprise AI users, and local governments — all seeking agile infrastructure to meet rising demand.
- The firm's proven track record, low deployment timelines, and specialization in GPU-ready design give it a unique value proposition that's difficult for legacy builders or overseas entrants to replicate.
- As AI adoption scales and latency-sensitive use cases multiply, Japan's edge and regional data center markets are set for growth — and Getworks is structurally positioned to be a significant player in that expansion with an asset-light, innovation-led business model.

The GxxD reports series covers the megatrends, business models, and innovations at the crossover between digital and clean energy. With stories on areas as diverse as AI, the CO_2 economy, robotics, and fusion, we share intelligence that anyone with an eye on Japan will want to explore further.

For further details about the GxxD project:

info@yuri-group.co.jp

www.yuri-group.co.jp/gxxd

linkedin.com/company/yurigr/

Yuri Group

nearer by not keeping still

Hulic Ochanomizu Bldg. 3F, 2-3-11, Surugadai,
Kanda, Chiyoda-ku, Tokyo, 101-0062

Writer: **George Hoffman**

Editor: **Yuriy Humber**

Designer: **Memi Fee** – memiart.com