Tech Dossier

Empyrion Digital: **Leading Tomorrow's AI-Ready Data Centers**

In this report:

- > The Golden Age after the Cloud: How AI Is Reshaping the Data Center Ecosystem
- > Requirements of Future Data Centers: "Power, Density, Cooling, Location"
- > The War Against Heat: Liquid Cooling, an Essential Weapon for AI Data Centers
- > Empyrion Digital: "Setting a New Standard for the Korean Data Center Market with AI-Ready and Green-by-Design Infrastructure"
- > **Checklist** AI Data Center Selection Guide: 5 Essential Factors

Empyrion Digital:
Leading Tomorrow's AI-Ready Data Centers

The exponential growth of the AI industry is driving significant changes in the IT infrastructure market. As AI training and inference tasks require increasingly more data and computing power, storage, server, and network technologies are evolving to meet this surging demand. As AI workloads continue to expand, acquiring GPUs—a critical component of AI infrastructure—now involves wait times ranging from several months to a year.[1]

Despite significant investments in AI infrastructure, many organizations continue to face challenges in using it effectively. For instance, some companies in South Korea have gone as far as shutting down air conditioning in entire buildings to meet the power demands of their GPU servers, while others hold back from fully leveraging their infrastructure due to soaring electricity costs. To address these challenges, organizations are increasingly adopting AI-ready data centers.

The Golden Age after the Cloud: How AI Is Reshaping the Data Center Ecosystem

Demand for high-performance computing (HPC) is skyrocketing. High-performance computing resources are especially vital for HPC workloads, including AI initiatives, scientific modeling, engineering simulations, financial analytics, and product innovation. These workloads require the processing of massive amounts

1 https://www.tomshardware.com/pc-components/gpus/nvidias-h100-ai-gpu-shortages-ease-as-lead-times-drop-from-up-to-four-months-to-8-12-weeks

of data and the execution of complex computations that exceed the capabilities of traditional IT infrastructure. Consequently, specialized data centers purpose-built for high-performance workloads have become indispensable.

There are three primary approaches to building a data center. The first is constructing your own on-premises data center. This option provides dedicated infrastructure customized for HPC workloads, offering greater performance control along with advantages in data security and regulatory compliance. However, on-premises data centers involve significant upfront costs and demand substantial resources for maintenance and upgrades. They also lack flexibility when scaling resources to meet evolving demands.

The second option leverages public cloud infrastructure. Recently, public cloud service providers have started offering AI-specific hardware, such as GPUs and TPUs, as a service, which can significantly reduce initial setup costs. While this option reduces upfront capital expenditure, long-term financial viability remains a concern due to potentially high data transmission and processing fees. Reliance on external infrastructure can restrict access to high-performance computing resources during peak demand. Data security and regulatory compliance requirements may also limit cloud deployment options.

Colocation data centers present a third solution, combining the advantages of on-premises infrastructure with the flexibility of the cloud. Through colocation services, organizations retain control of their hardware in a purpose-built data center facility. The data center operator handles all aspects of physical security, power supply, cooling, and network infrastructure management—minimizing operational overhead while delivering considerable cost advantages over public cloud solutions.

Rapid adoption of AI and HPC workloads has catalyzed significant growth throughout the data center industry. AI servers demand substantially higher power density than conventional servers, requiring facility upgrades, extensive power infrastructure modifications, and even new data center construction. On-premises infrastructure frequently falls short of AI performance requirements, compelling organizations—including those with established data centers—to explore colocation options. Cloud computing offers an alternative solution yet presents limitations for complex computational tasks. These constraints lead organizations to favor colocation facilities, accelerating the demand for colocation services.

Gartner predicts that generative AI projects will boost data center system revenue by more than 10% in 2024.[2] This represents a substantial increase from the 4% growth rate in 2023.

According to the “Global Data Center Trends 2024” report by CBRE[3], data centers across multiple global regions are operating near capacity. The report highlights record-low vacancy rates in North American markets. In Chicago, the vacancy rate dropped dramatically from 6.7% to 2.4% year-over-year. Virginia, a major data center hub, saw its vacancy rate nearly halve from 1.8% to 0.9%, despite an 18% increase in supply. While Asia-Pacific markets experienced a slight uptick in vacancy rates—from 13.5% to 16%—due to new facility completions, Singapore maintains exceptionally tight supply with vacancy rates below 1%.

2 https://www.gartner.com/en/newsroom/press-releases/2024-04-16-gartner-forecast-worldwide-it-spending-to-grow-8-percent-in-2024

3 https://www.cbre.com/insights/reports/global-data-center-trends-2024

Requirements of Future Data Centers: "Power, Density, Cooling, Location"

AI is reshaping both the data center industry and its underlying technologies. What key features separate traditional and AI-ready data centers? Industry experts[4] identify several key characteristics of AI-ready facilities: robust power infrastructure, high-density racks, advanced cooling systems, and strategic site selection. These fundamental attributes serve as critical evaluation criteria for organizations deploying AI workloads.

Power infrastructure stands as a fundamental component of AI data centers. AI models, particularly deep learning applications, perform massive data processing and complex mathematical calculations. This processing relies on high-performance hardware such as GPUs and TPUs, which consume enormous amounts of power. A stable, high-capacity power supply remains essential for AI-ready data centers.

Specific data illustrates this power demand: As of December 2023, the total power capacity of 150 data centers in Korea reached 1,986MW, equivalent to the power consumption of more than two 1,000MW nuclear power plants.[5] Data centers equipped for AI operations consume six times more power than traditional data centers, necessitating a complete redesign of building and wiring structures for stable operation. Recognizing these power demands, emerging data center providers entering the market integrate eco-friendly principles into their foundational design to optimize power consumption and minimize their environmental impacts. This approach enhances energy efficiency while reducing operational costs and ensuring sustainability.

High-density racks represent another critical requirement. In technical terms, these are specialized rack systems that accommodate increased computing power and higher power consumption within a defined space, surpassing the capabilities of conventional racks. Such advanced configurations enable dense

4 https://www.cbre.com.sg/insights/articles/navigating-data-centres-as-demand-for-ai-surges-what-does-this-mean-for-data-centres

5 https://www.nars.go.kr/report/view.do?cmsCode=CM0018&brdSeq=44746

integration of servers and high-performance equipment, allowing organizations to operate substantially more computing resources than traditional rack designs. The resulting concentrated deployment maximizes computing power within a given footprint, though it requires significantly higher power capacity.

Prior to the generative AI era, enterprise and gaming companies—traditionally primary data center customers—used 4kW to 6kW per rack, while cloud service providers required approximately 10kW per rack. However, the average power demand of AI companies now far exceeds 20kW. A single NVIDIA DGX H100 system, widely deployed in AI projects, consumes more than 10kW. Multiple DGX H100 systems in one rack can push power density beyond 40kW.

In response, AI-ready data center operators have developed specialized systems to effectively accommodate and manage high-density racks. Beyond expanding power supply capacity, they deliver comprehensive solutions encompassing efficient space utilization, thermal management, and overall infrastructure optimization.

Advanced cooling technology represents the third critical factor. AI operations involve continuous high-powered processing that generates significant heat. This makes thermal management essential to prevent system overheating, which can lead to both performance degradation and system failures. To maintain stable operations, systems must be protected from high temperatures that can significantly reduce the lifespan of costly AI servers and risk severe equipment damage. Given these challenges, implementing efficient cooling systems becomes crucial, serving dual purposes: protecting valuable hardware investments while optimizing overall energy efficiency to reduce facility power consumption.

A stable power supply and thermal management technology have emerged as core competencies for modern data centers. Traditional air-cooling methods—circulating cold air and expelling hot air—can manage heat densities of up to 20kW per rack. However, deployments of high-performance AI processors, such as NVIDIA's specialized chips, demand enhanced cooling capabilities.

This requirement has accelerated the adoption of liquid cooling technologies alongside existing air-cooling systems. Liquid cooling solutions excel at managing higher heat densities, proving essential for data centers optimized for AI workloads.

Location strategy represents the final critical consideration. While traditional data centers prioritize proximity to end-user locations for low-latency performance, AI workloads present different requirements.

The AI training process remains relatively latency-tolerant. Minor variations in data transfer speeds minimally impact training operations. Although AI inference demonstrates higher latency sensitivity, the required data doesn't necessarily demand local storage.

The landscape of AI inference has evolved beyond centralized data centers to encompass various edge computing devices, such as IoT endpoints and mobile devices, as well as cloud infrastructure. This shift has transformed location selection criteria for AI-ready data centers, offering greater geographic flexibility to both operators and clients. This evolution enables prioritization of other critical factors, particularly power availability and cooling efficiency.

The War Against Heat: Liquid Cooling, an Essential Weapon for AI Data Centers

Liquid cooling technology stands at the forefront of data center innovation. The shift from air cooling to liquid cooling solutions offers superior heat management, enhanced efficiency, and reduced noise levels in data center operations.

The industry currently explores three primary liquid cooling technologies. Rear-Door Heat exchangers (RDHx) represent the first approach. This solution incorporates heat exchangers mounted on individual server rack backs for enhanced cooling efficiency. Advanced implementations can replace entire cage or rack cooling units with liquid cooling systems. This modular approach enables targeted upgrades at the rack or cage level rather than requiring facility-wide modifications. Industry analysis[6] indicates RDHx systems can support power densities from 20kW to 50kW per rack.

Direct-to-chip cooling offers the second approach, delivering superior performance compared to RDHx with support for power densities from 20kW to 100kW. The system employs small heat-absorbing components attached directly to processors such as CPUs and GPUs, circulating coolant for heat dissipation. While this approach requires hardware modifications and complex implementation, NVIDIA's recent announcement[7] of direct-to-chip cooling support has heightened industry interest.

Immersion cooling presents the third approach, submerging computing equipment in dielectric coolant for direct heat transfer. This method has already proven successful in cryptocurrency mining operations, supporting power densities from 50kW to 250kW per rack according to industry analysts[8]. However, implementation challenges include increased weight loads, potential fluid leakage risks, limited operational data from production environments, and varied regulatory compliance requirements across regions.

Research across global markets shows immersion cooling technology will play a fundamental role in future data center operations. However, significant challenges currently limit widespread adoption. Current horizontal rack designs for immersion cooling reduce optimal space utilization in data centers, while

6 https://www.networkworld.com/article/2076039/data-centers-warm-up-to-liquid-cooling.html

7 https://blogs.nvidia.com/blog/hot-chips-2024/

8 https://www.networkworld.com/article/2076039/data-centers-warm-up-to-liquid-cooling.html

Figure 1 | Front view of Empyrion Digital's KR1 data center in Korea

comprehensive liquid detection and mitigation systems must be integrated during initial facility design. This cooling approach also faces two major adoption barriers: unproven long-term reliability of submerged chips and ambiguous manufacturer warranty terms for liquid-cooled components.

Most critically, the substantial weight of liquid-filled racks demands enhanced seismic design requirements, necessitating purpose-built facilities. This structural challenge makes immersion cooling a future-focused solution better suited for new construction projects rather than an immediate option for existing data centers, particularly in established markets like Korea.

Empyrion Digital: "Setting a New Standard for the Korean Data Center Market with AI-Ready and Green-by-Design Infrastructure"

Since its inception in 2021, Empyrion Digital has invested in five data centers across its platform in Singapore, South Korea, Japan, Taiwan and Thailand. Our strategic presence in these markets enables us to capitalize on Asia's digital transformation growth while serving a diverse customer base.

Empyrion Digital specializes in carrier-neutral, AI-ready data centers with sustainable designs, primarily serving hyperscalers, enterprises, and telecommunications companies. As a new entrant unencumbered by legacy infrastructure, the company is uniquely positioned to support emerging technologies like AI without the burden of legacy system upgrades. By the end of 2025, Empyrion Digital projects its pan-Asian platform capacity to exceed 170MW of IT load.

The company's KR1 facility, currently under construction in Seoul's Yangjae district, is slated for completion in Q3 2025. Located 9km from Pangyo Technology Hub and 12km from Seoul's central business district, the 30,714㎡ facility is purpose-built for AI workloads, attracting significant interest from enterprises, cloud providers and AI startups.

A key differentiator for Empyrion Digital lies in its technical team's deep expertise in data center design and operations. The KR1 design phase incorporated expertise from industry veterans who previously led hyperscale data center design and operations at major global technology companies, including Naver, Yahoo and Meta. This wealth of experience has been instrumental in developing KR1's architecture to deliver both exceptional performance and environmental sustainability.

“Digital transformation has accelerated across industries since COVID-19, with companies increasingly adopting cloud and AI solutions to maintain competitive advantage,” notes Yongsuk Choi, Chief Strategy Officer – Product and Infrastructure Officer of Empyrion Digital. “This shift demands scalable, secure infrastructure capable of supporting AI operations. Empyrion Digital meets these market demands by delivering purpose-built facilities specifically designed to accommodate AI workloads.”

Traditional data center construction timelines often result in facilities that reflect technology requirements from two to three years prior. However, KR1 incorporates AI-era specifications from its initial design phase. Choi said, “Empyrion Digital anticipates enterprise and hyperscaler requirements, delivering high-performance environments optimized for AI and HPC workloads.” He also added, “Our facilities provide flexible support for emerging technology adoption, including generative AI initiatives.”

Sustainability stands as a core strategic pillar for Empyrion Digital. Data center efficiency is primarily measured by Power Usage Effectiveness (PUE)—the ratio between total facility power consumption (including cooling systems, elevators, security, lighting, and fire suppression equipment) and IT equipment power usage. A PUE closer to 1.0 indicates higher energy efficiency.

While the global average PUE for data centers stands at 1.58,[9] Empyrion Digital facilities achieve a PUE of 1.3 or lower. This efficiency translates to approximately 30% reduction in annual power costs for customers. Legacy facilities often struggle with higher PUE due to aging infrastructure, whereas new facilities can optimize energy efficiency from inception through modern design and equipment.

The company has also optimized Water Usage Effectiveness (WUE), achieving 25% to 30% reduction in water consumption compared to industry standards. Although water costs aren't directly billed to customers like electricity, improved water efficiency supports customers' ESG initiatives. Traditional cooling tower systems lose significant water through evaporation annually. Empyrion Digital addresses this through closed-loop systems, reducing water consumption by approximately 30%.

The facility design features a seven-meter (23-foot) floor height to optimize cooling efficiency. This increased height enhances natural convection, allowing hot air to rise and exit more effectively. The design also supports direct-to-chip liquid cooling technology, chosen for its proven effectiveness and practical implementation in modern data centers

Empyrion Digital leads the industry with its high-power rack capacity. Supporting high-density racks goes beyond simply increasing power input - it requires fundamental innovations in data center technology. The key lies in stable power delivery and efficient heat management. While current AI workloads typically demand over 20kW per rack, this facility delivers 40kW per rack without modifications, while maintaining optimal PUE.

The 7th and 8th floors feature direct-to-chip liquid cooling (DLC) systems with

9 https://www.kharn.kr/news/article.html?no=19486

advanced coolant distribution units (CDU). This cooling infrastructure is essential for next-generation AI hardware like NVIDIA's NV72 GB200, which demands up to 120kW per rack.

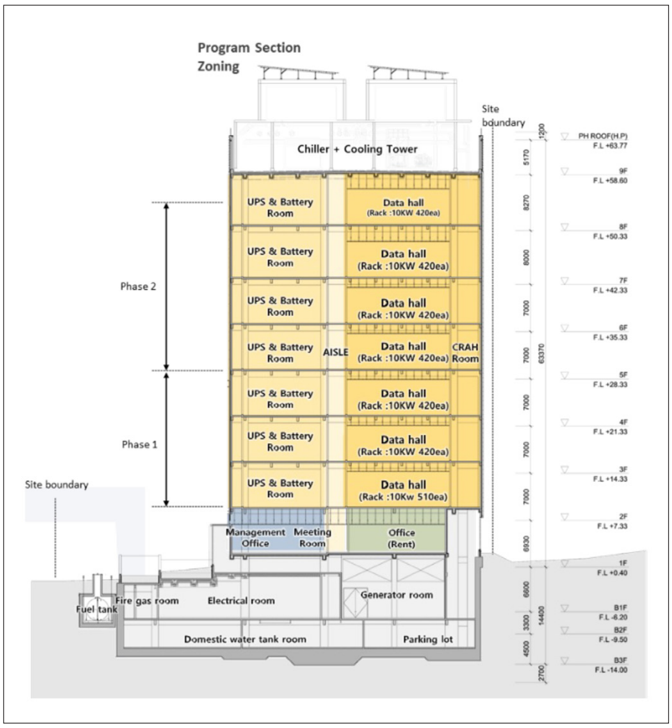


Figure 2 | Empyrion Digital Data Center Structure

Implementing DLC in existing data centers poses significant challenges. It requires extensive modifications including coolant piping installation and three-phase power infrastructure. That's why this facility incorporated DLC capabilities from the initial design phase - a strategic decision that not only ensures seamless support for high-density computing needs but also sets it apart as a market leader in advanced data center solutions.

"While some organizations hesitate to migrate from their existing data centers due to operational disruption concerns, we're seeing a significant shift in customer priorities," notes Choi. "An increasing number of customers are turning to Empyrion Digital to reduce long-term costs and strengthen their ESG initiatives."

"Interestingly, we're also observing a 'reverse cloud migration' trend. As data storage and processing demands grow, some companies face escalating cloud costs, leading them to return to data centers. Many are now adopting hybrid cloud strategies to optimize data management." he explains.

The facility demonstrates its future-ready design in two key areas: incorporating state-of-the-art technologies while maximizing both energy and water efficiency. This focus on sustainability is becoming increasingly crucial as environmental regulations tighten globally. Major players, including hyperscalers and enterprise customers, are now prioritizing sustainable digital infrastructure in their transformation strategies - a trend that's reshaping the data center market landscape.

Checklist

AI Data Center Selection Guide : 5 Essential Factors

1. Power Efficiency (PUE)

Power Usage Effectiveness (PUE) indicates a data center's energy efficiency - the lower the number, the more efficient the facility. The global industry average PUE hovers around 1.58. PUE significantly impacts electricity costs and overall operational expenses. For instance, if IT equipment consumes $1,000 in electricity, a data center with a PUE of 1.8 will incur total electricity costs of $1,800, while one with a PUE of 1.3 will result in $1,300 in charges.

For AI and High-Performance Computing (HPC) projects, Total Cost of Ownership (TCO) calculations should include all operational costs beyond the Monthly Recurring Charges (MRC). Advance power cost simulation is essential for accurate budgeting. Organizations prioritizing ESG initiatives should pay particular attention to PUE metrics as a key performance indicator.

2. Cooling System

The cooling infrastructure in a data center is not just about having cutting-edge technology - it's a core system that directly impacts operational costs. AI-ready data centers typically employ hybrid cooling solutions that combine air and liquid cooling methods. When evaluating facilities, verify cooling capabilities specifically engineered for high-density AI server racks. For data centers using air cooling and cooling tower systems, examine what additional efficiency measures have been implemented. A comprehensive understanding of these cooling factors enables more efficient and cost-effective operations.

3. Scalability

The data hall - where servers, storage, and network equipment are installed - is crucial for scalability. A single, large-scale data hall provides operational advantages - equipment can be placed flexibly, space remains available for seamless expansion, and network configurations can be optimized efficiently. Conversely, fragmented data halls with multiple small rooms can create challenges in equipment relocation, space utilization, and often require additional network equipment.

For large-scale infrastructure deployments, consider the data hall's physical layout, particularly column-free designs that optimize space usage and airflow efficiency.

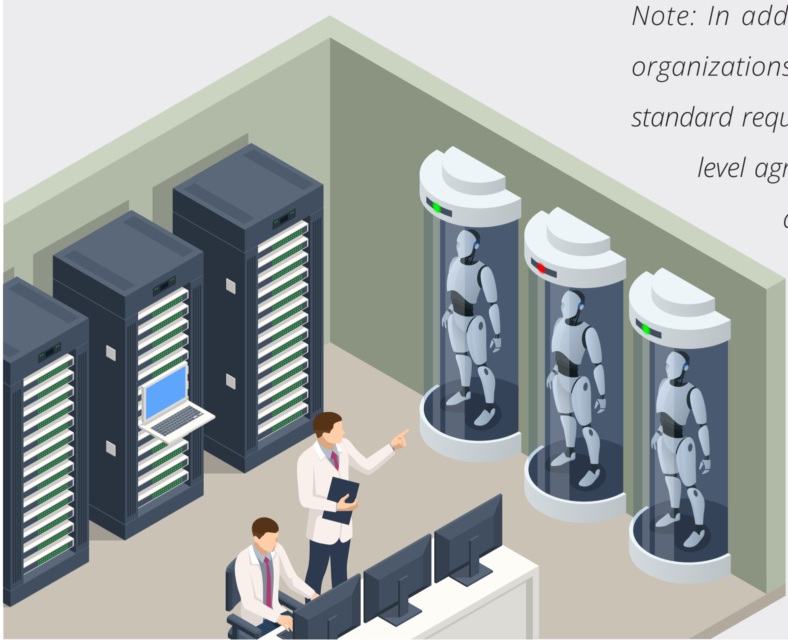
4. Location

Data center location is primarily driven by customer requirements. Companies generally prefer locations near their headquarters or main offices for easy access by operations teams, unless they specifically need a separate disaster recovery (DR) facility in a different region. While urban data centers may command higher costs compared to suburban locations, they offer significant advantages in terms of workforce availability and network connectivity. However, regulatory restrictions in metropolitan areas may limit future large-scale data center construction. Organizations planning to leverage urban data centers should factor these potential expansion constraints into their long-term strategy.

5. High-Density Rack Support

AI workloads are driving increased demand for high-density racks ranging from 10kW to 20kW or higher. Traditional data centers typically provided 2kW to 4kW per rack, with some offering up to 10kW, but today's requirements far exceed these specifications. Upgrading power capacity in existing facilities involves complex processes including power authority approvals, substation capacity expansion, and power line replacements. Therefore, when selecting a data center, stable power delivery per rack must be thoroughly evaluated as a first step, and verifying overall power delivery capabilities should be a top priority.

An important consideration is that "high-density" definitions vary significantly among data center providers. While some providers market 10kW racks as high-density solutions, the AI industry typically requires 20kW per rack to support current workloads. Given these varying standards, organizations should carefully evaluate and confirm the actual power capacity available at their prospective data center before making commitments.



Note: In addition to the key factors above, organizations should also consider industry-standard requirements such as security, service level agreements (SLAs), legal compliance, and disaster recovery capabilities. Although these features are generally standardized across providers, they remain essential components of a thorough data center evaluation process.