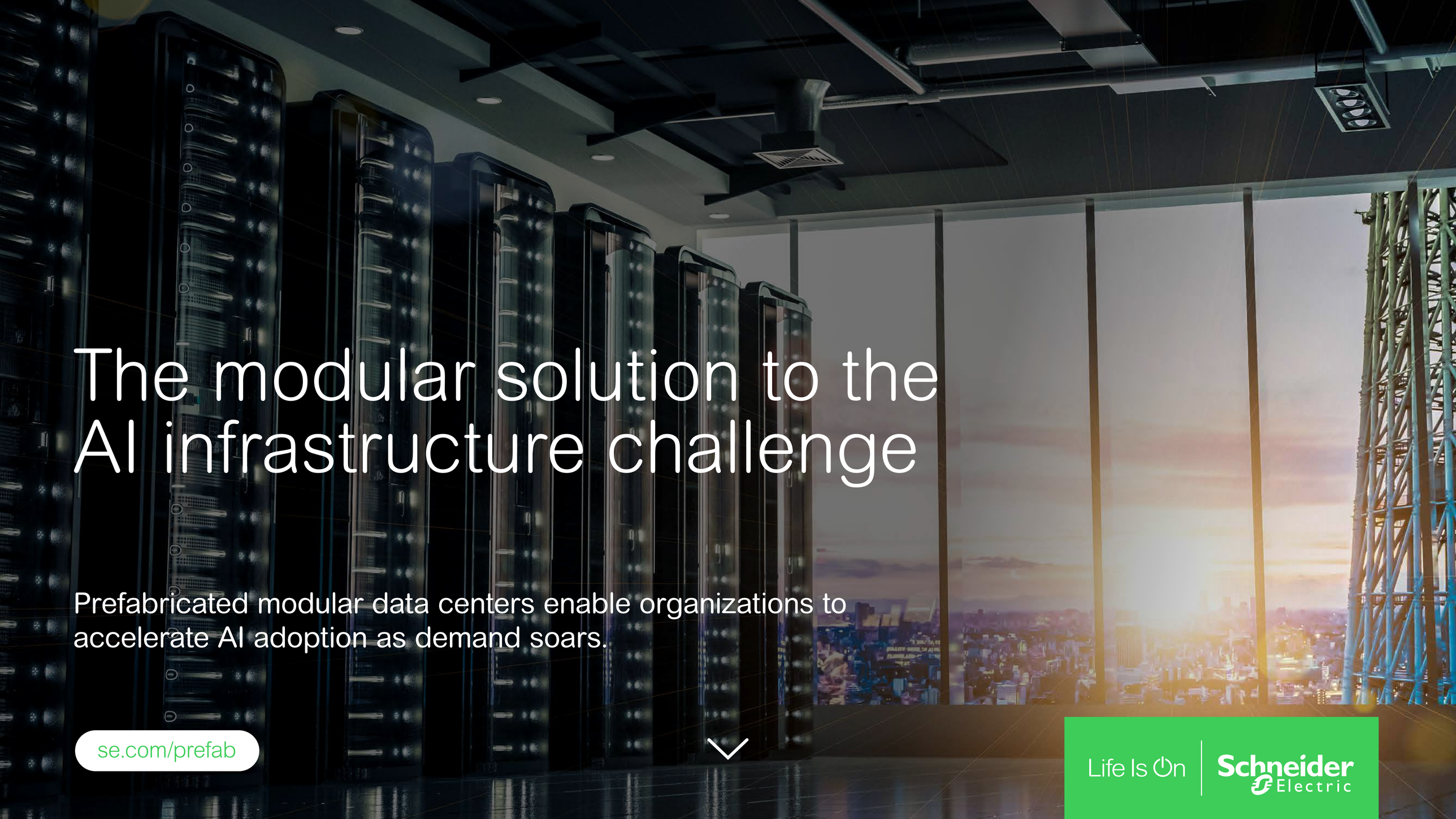


The modular solution to the AI infrastructure challenge

Prefabricated modular data centers enable organizations to accelerate AI adoption as demand soars.

se.com/prefab

Life Is On | Schneider Electric

Table of contents

The promise of generative AI

The promise of generative AI

The introduction of generative artificial intelligence (Gen AI) opens possibilities for organizations of all sizes by converting complicated processes into tasks that last seconds or minutes.

Despite the potential benefits, Gen AI poses serious challenges. Previous innovations, such as electricity, required the building of infrastructure for adoption, but AI can be put to use immediately through internet-connected machines. And there's the conundrum: Easy availability creates the biggest challenge to AI adoption. While anyone with a computer or smartphone can access a Gen AI engine, there isn't enough data center infrastructure to support the expected rush to adopt AI in the coming years.

This means significant investment in data center capacity is necessary. AI is very data-intensive, driving massive demand for compute and energy. The technology places unprecedented demands on existing data center infrastructure at a time when data centers are under pressure to decarbonize to combat climate change. Data center owners and operators need quick solutions.

Easy availability creates the biggest challenge to AI adoption.

Explosive Gen AI demand

Explosive Gen AI demand

The introduction of Gen AI tools such as ChatGPT, Jasper, and Microsoft Copilot has fueled massive demand for the technology. Consider these projections:

Gen AI will reach

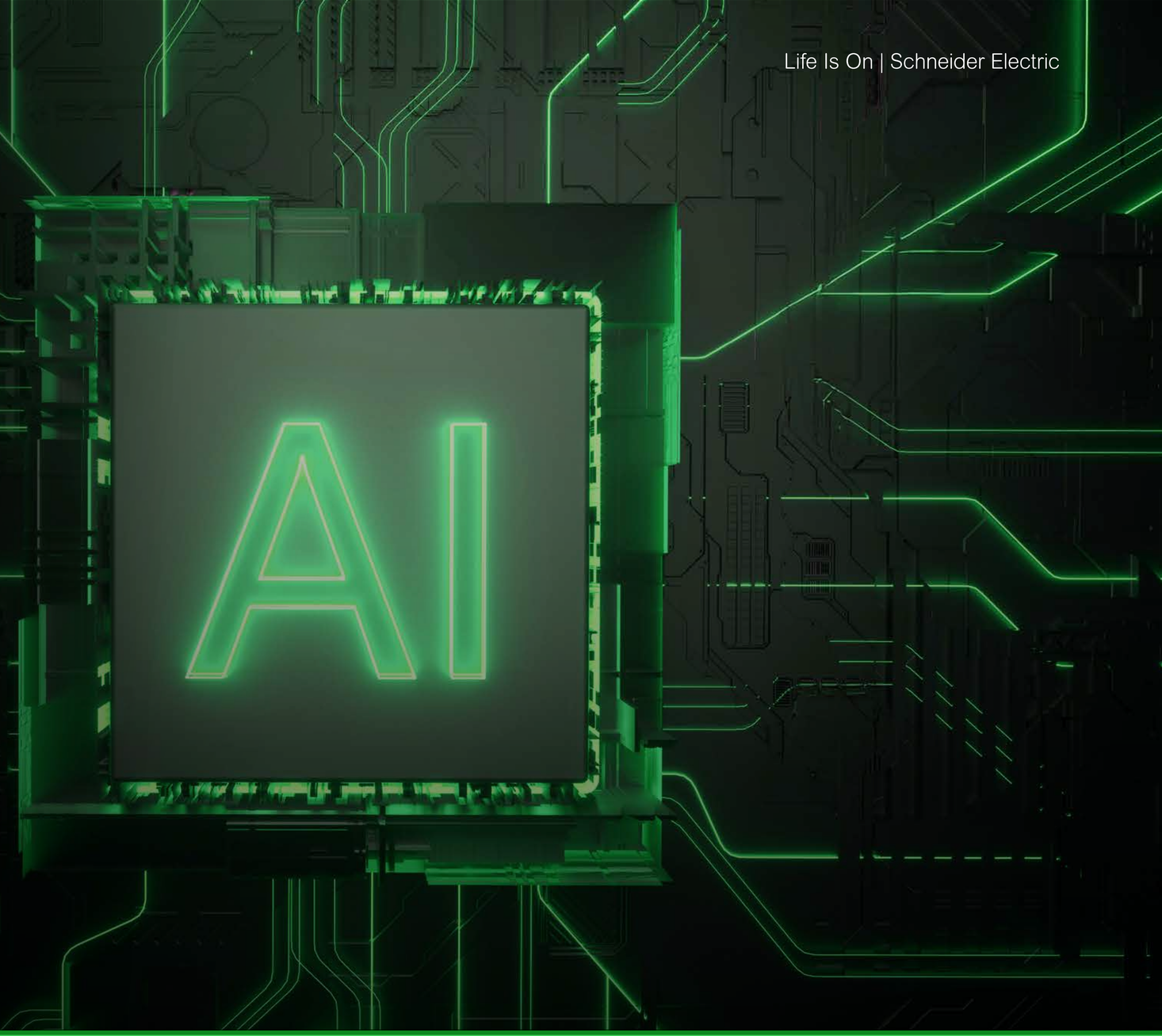
$1.3 trillion

within a decade,
from $40 billion in 2022.

AI will grow at an annual rate of

33%

between 2023 and 2030.

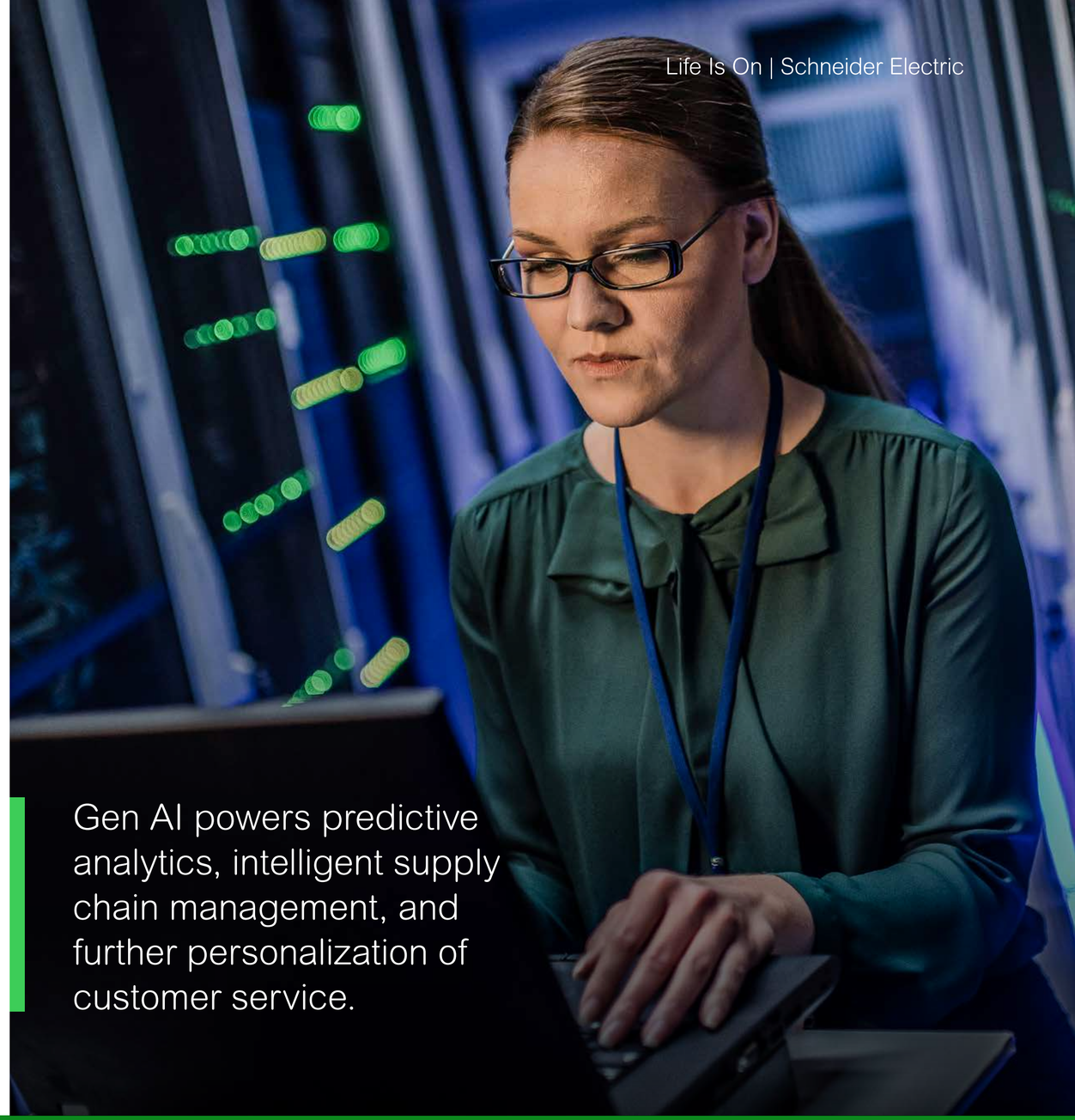


Initial growth will be driven by training infrastructure and then shifting gradually to inference systems.

Gen AI delivers multiple use cases for just about every industry, enabling organizations to rethink how to accomplish tasks and redesign processes. It powers predictive analytics, intelligent supply chain management, and further personalization of customer service.

- Gen AI can generate code, accelerate application development, and place new functionality in users' hands.
- AI-powered chatbots and virtual assistants reduce operational costs by handling customer questions.
- AI assistants can answer targeted questions to help users quickly compile slide presentations and reports.
- AI significantly reduces preparation for presentations.

And many more use cases will emerge.



Gen AI powers predictive analytics, intelligent supply chain management, and further personalization of customer service.

Massive power requirements

As already noted, Gen AI comes at a cost by generating unprecedented demand for compute and electric power. The following statistics from Goldman Sachs Research illustrate why:

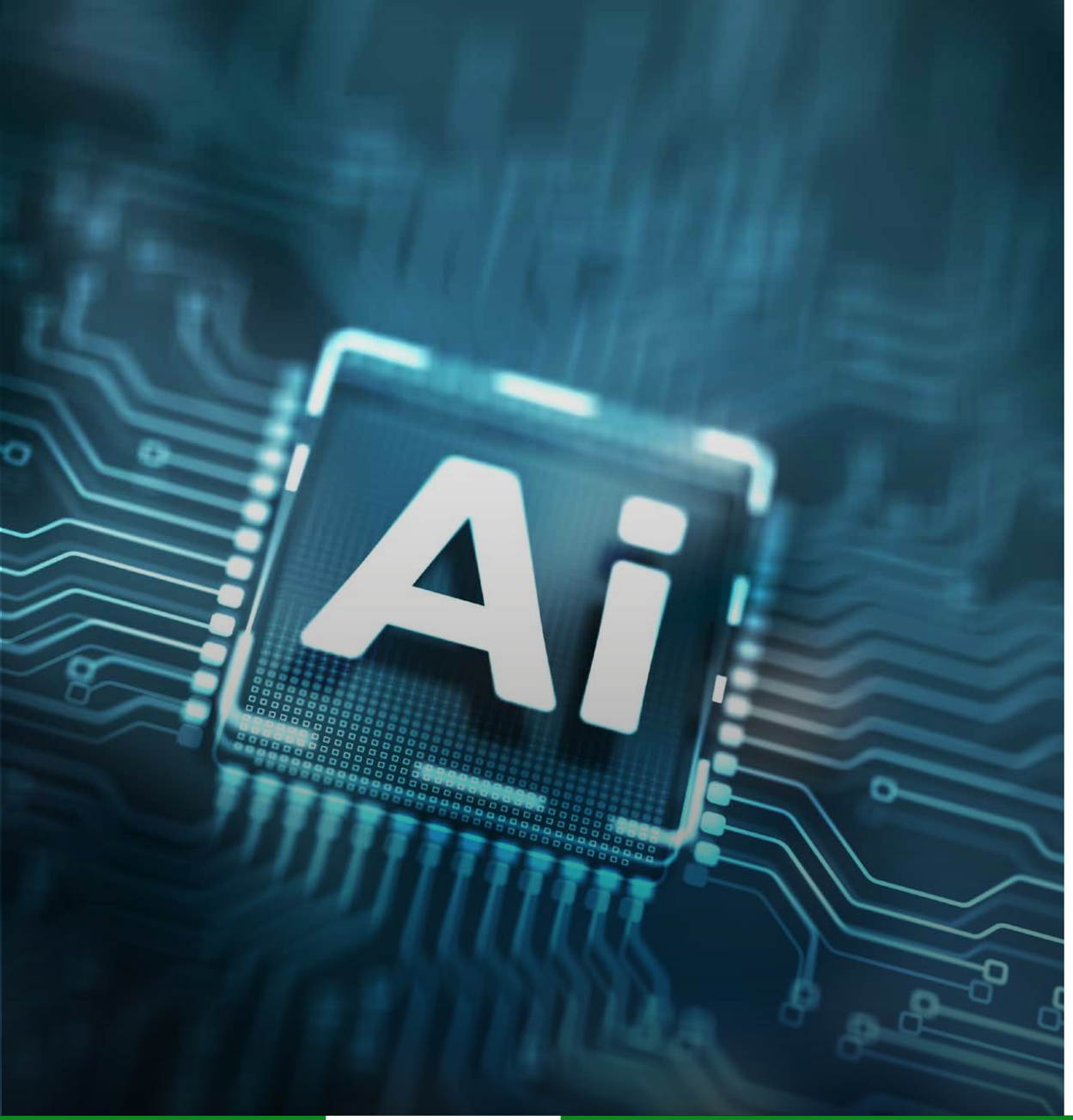
A ChatGPT query uses nearly **10 times the power of a Google search.**

Data center power demand will **grow 160% by 2030.**

For years, data centers maintained a steady demand for power despite rising workloads. However, with efficiency gains slowing and the AI revolution accelerating, Goldman Sachs projects a 160% surge in data center power consumption by 2030.

Data centers consume nearly 2% of the world's electricity, which will double by 2030, Goldman Sachs Research predicts. As a result, electricity demand will soar, especially in the U.S. and Europe. “Along the way, the carbon dioxide emissions of data centers may more than double between 2022 and 2030.”

To accommodate this rate of growth, a rethinking of the data center is urgently needed, not only to provide the required infrastructure for Gen AI but also solutions for power management and sustainability.



Gen AI training and inference

Gen AI's insatiable hunger for data is at the heart of soaring demand for data center computational resources and electricity. Gen AI training models consume mindboggling volumes of data to deliver meaningful results.

And because Gen AI models operate in dynamic environments where new data is introduced constantly, the need to feed data to models becomes almost constant. This is necessary because the more data Gen AI models consume, the smarter they get. As a result, the quality of the results improves, providing more targeted actionable data.

The second primary function of Gen AI – inference – also contributes to Gen AI's hunger for compute and electricity. Inference is what leads to meaningful results. In a process that recalls a human's ability to infer from listening to another,

AI models infer by ingesting data. The more data they ingest and analyze, the more accurate the inference. AI learns to recognize images, text, and other information to solve problems and complete the tasks that organizations entrust to the technology.

Pause-restart workloads

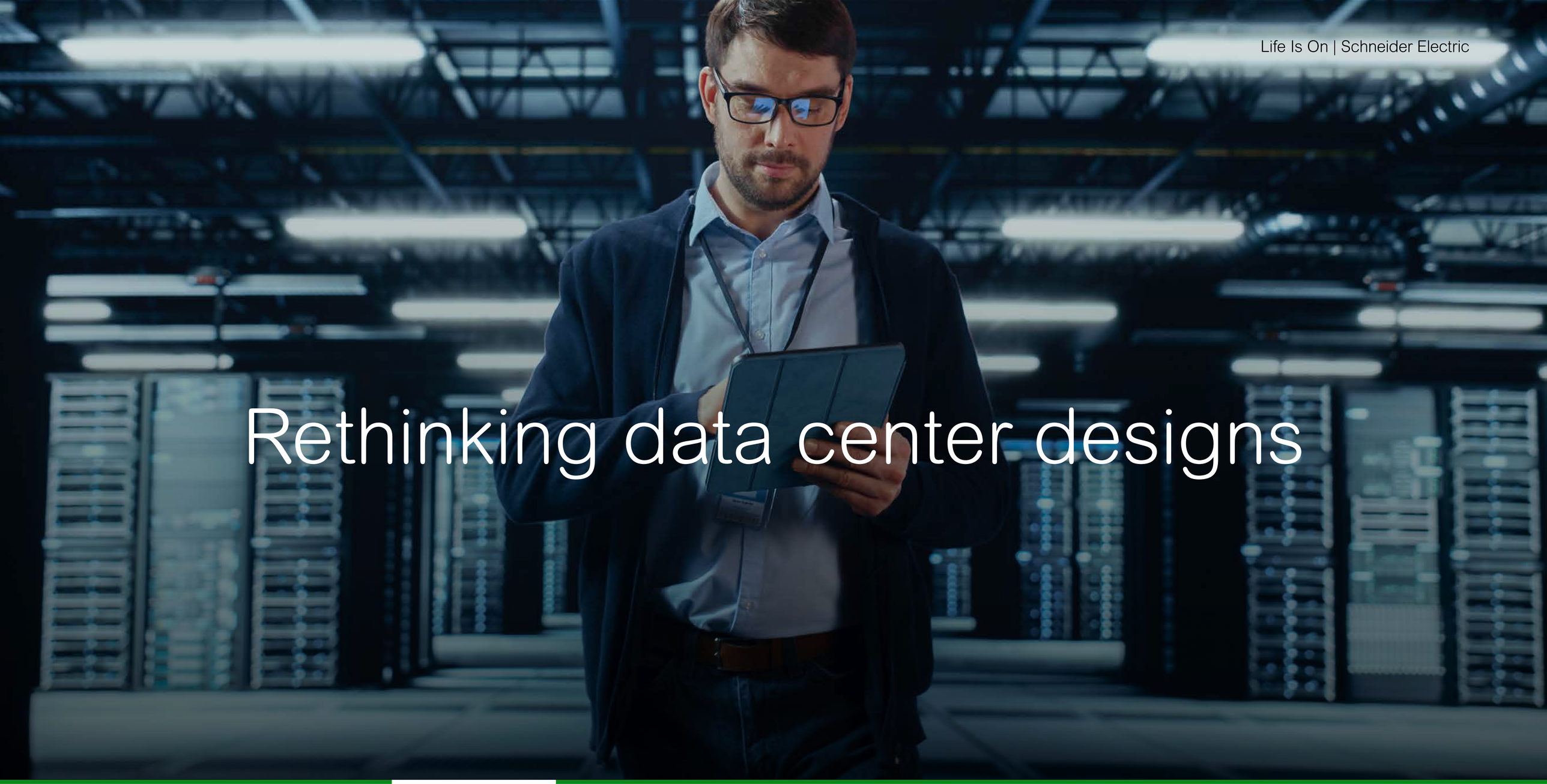
AI creates new types of workloads in the data center. Besides consuming far more resources, these workloads require:

- **Lightning-fast networking**
- **High levels of flexibility and scalability**
- **New approaches to cooling and power management**

The most significant difference from traditional data center workloads is that Gen AI workloads pause and restart frequently for training models and inference. In other words, they only run when needed. But when they restart, they resume with improved efficiency toward their goals.

The pause-start nature of AI workloads causes fluctuations in compute far more dramatic than those of traditional workloads. Fluctuations of 50% to 60% can occur in minutes for AI training, compared to variations of 10% to 15% over several hours with regular workloads.

Regardless of these fluctuations, data centers need uninterruptible power to support AI workloads, not only for when they are running but also for the continuous protection of networking and storage resources. Cybersecurity systems and uninterruptible power supplies (UPS) run constantly.



Rethinking data center designs

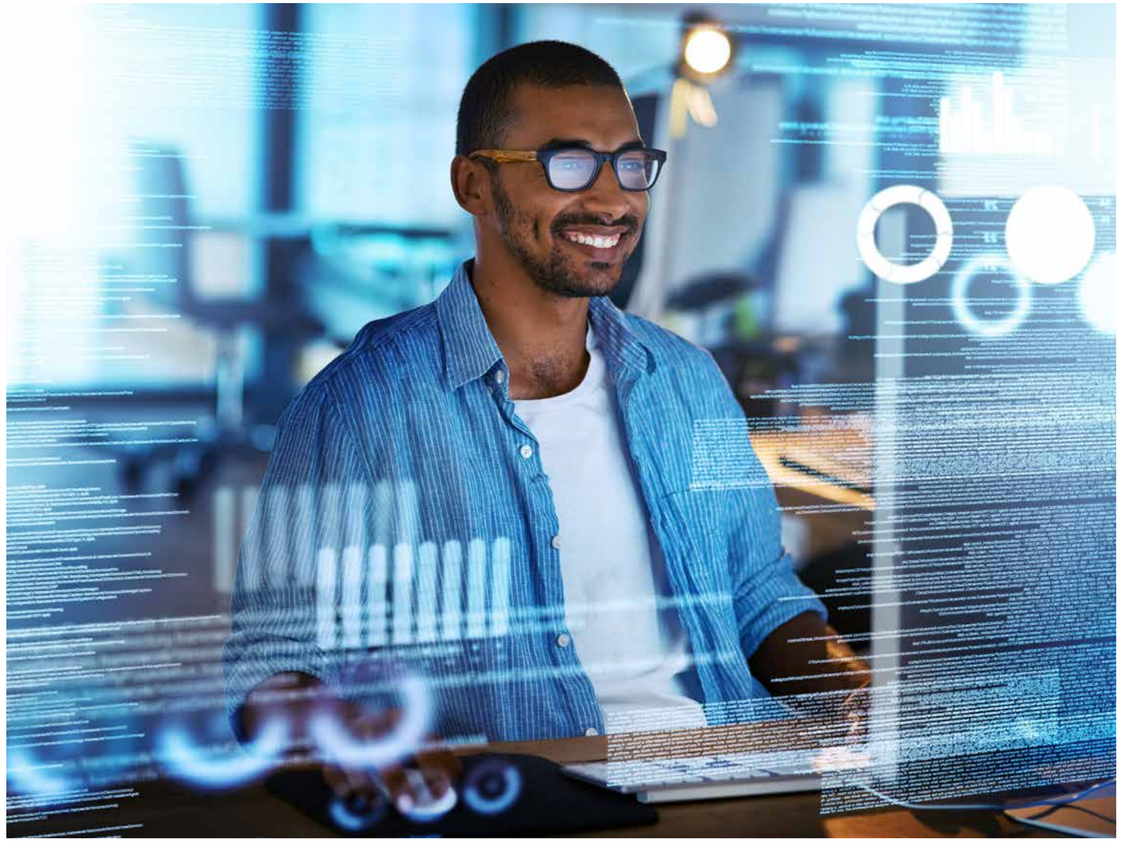
Rethinking data center designs

Besides generating urgent demand for new infrastructure, Gen AI requires specialized infrastructure such as servers equipped with GPUs and high-density chipsets with processors placed very close together. These infrastructure changes are among the reasons new approaches to data center design are needed. Design modifications affect physical layout, power consumption and management, cooling, and rack density.

Existing, traditional data centers will require retrofitting the space, which isn't always practical or cost-effective. Building new data centers is expensive, complex, and time-consuming. So, a third alternative is needed to provide infrastructure for Gen AI workloads quickly. A viable option involves deploying purpose-built modular data centers at the edge – or anywhere AI infrastructure is needed. A significant amount of work is underway to develop modular designs for AI.

The designs must take into account three primary needs:

- Much higher rack densities
- New cooling requirements
- Changes in power management



Higher rack densities

The biggest immediate demand for compute from Gen AI workloads comes from model training. Not long ago, typical rack density ranged from 7 kW to 10 kW. More recently, data-intensive applications, like the ones that perform data mining and analytics and share information across multiple sites, have driven rack densities to 20 kW or higher.

The increase in rack density has been gradual – until now. With Gen AI's explosive growth, skyrocketing compute demand is pushing rack densities to 50 kW and higher. Racks are now being designed to handle 80 kW to 100 kW+ to support Gen AI training models. Inference is less data-intensive, requiring racks of up to 20 kW.

The dramatic jump in power requirements partly explains why new data center designs must account for higher-density racks. A benefit of higher-density racks is a reduction in data center footprint. However, designs also have to consider a substantial increase in energy requirements and the massive amounts of heat generated by chipsets that are purpose-built for AI. These chipsets place multiple processors close together, generating a lot of heat.

Racks are now being designed to handle 80 kW to 100 kW+ to support Gen AI training models.

Power management

Data centers are under pressure to meet sustainability goals, whether from regulations in countries and regions trying to counter the effects of climate change or from customers wanting to hit their decarbonization targets. The drive to sustainability affects all aspects of data center planning, design, and operation, from construction materials to equipment to power management strategies.

For power management, data centers can leverage more efficient hardware, such as uninterruptible power supplies (UPSs), that consume less energy while providing more capacity. Power management systems such as DCIM monitor energy consumption and temperatures, delivering insights on optimizing power use.

In some cases, data centers can increase their use of renewables by generating energy locally with solar panels and storing it on-site.

Reinventing cooling

Air cooling has been the standard method of dissipating heat in data center aisles, but it is inadequate for handling the additional heat generated by AI chipsets. To address this, data center designers are turning to another method – liquid cooling. Liquid cooling is an evolution of an existing technology for a new environment.

Liquid cooling systems enable optimal operating temperatures. They are much more effective than air cooling in high-density environments, making them more suitable for AI chipsets and servers. Liquid cooling offers multiple benefits:

More Effective
It is up to 20 times more effective than air cooling, because liquids are better thermal conductors than air.

Quieter
It is much quieter than the fans and other equipment required for air cooling.

Less Energy
Liquid cooling systems use considerably less energy than air cooling, which helps improve the data center's power usage effectiveness (PUE).

Cost Savings
Liquid cooling helps defray the added costs and environmental impact of higher-density computing.

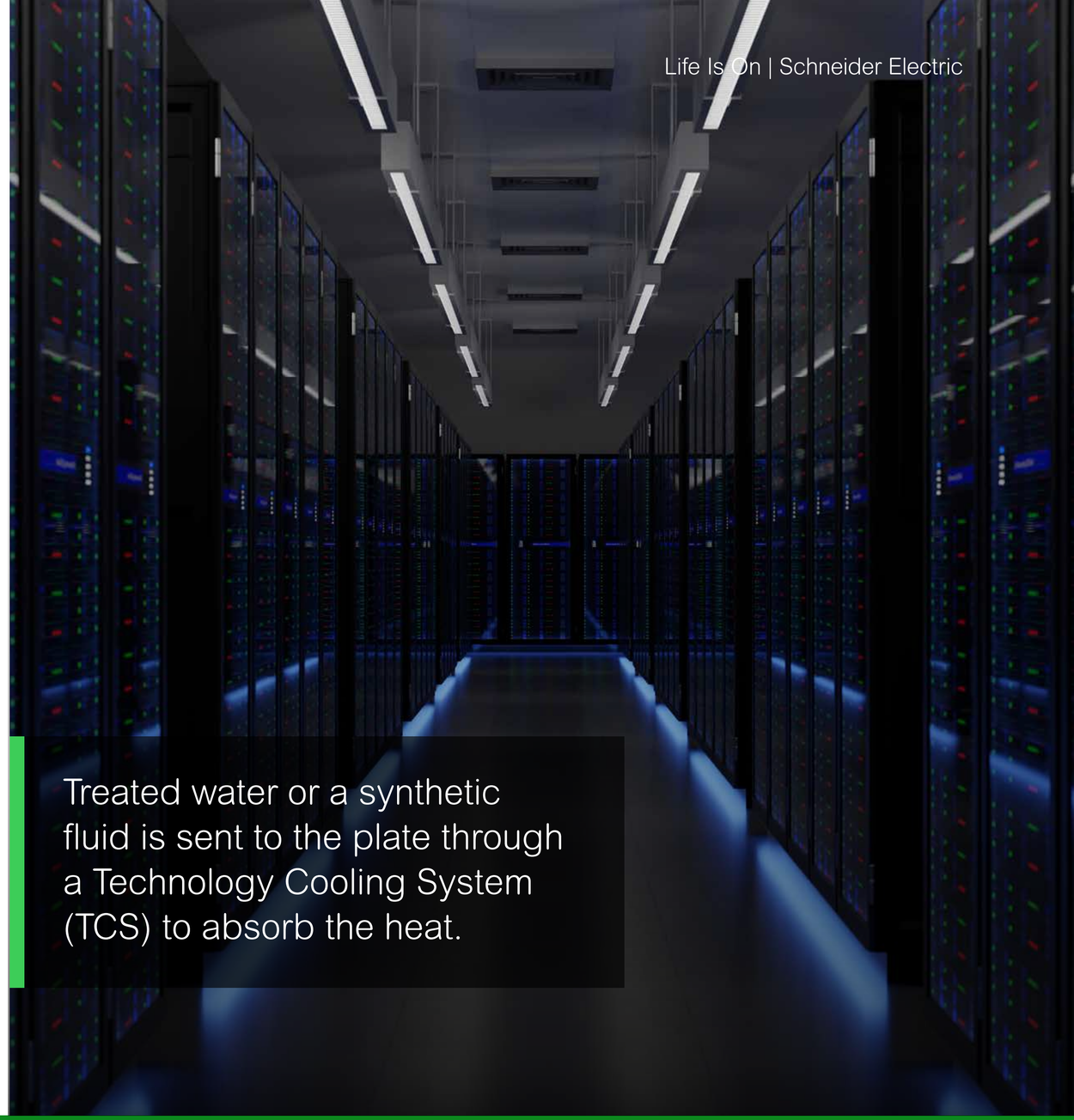
How liquid cooling works

One type of liquid cooling involves attaching cold plates to processors. Treated water or a synthetic fluid is sent to the plate through a Technology Cooling System (TCS) to absorb the heat.

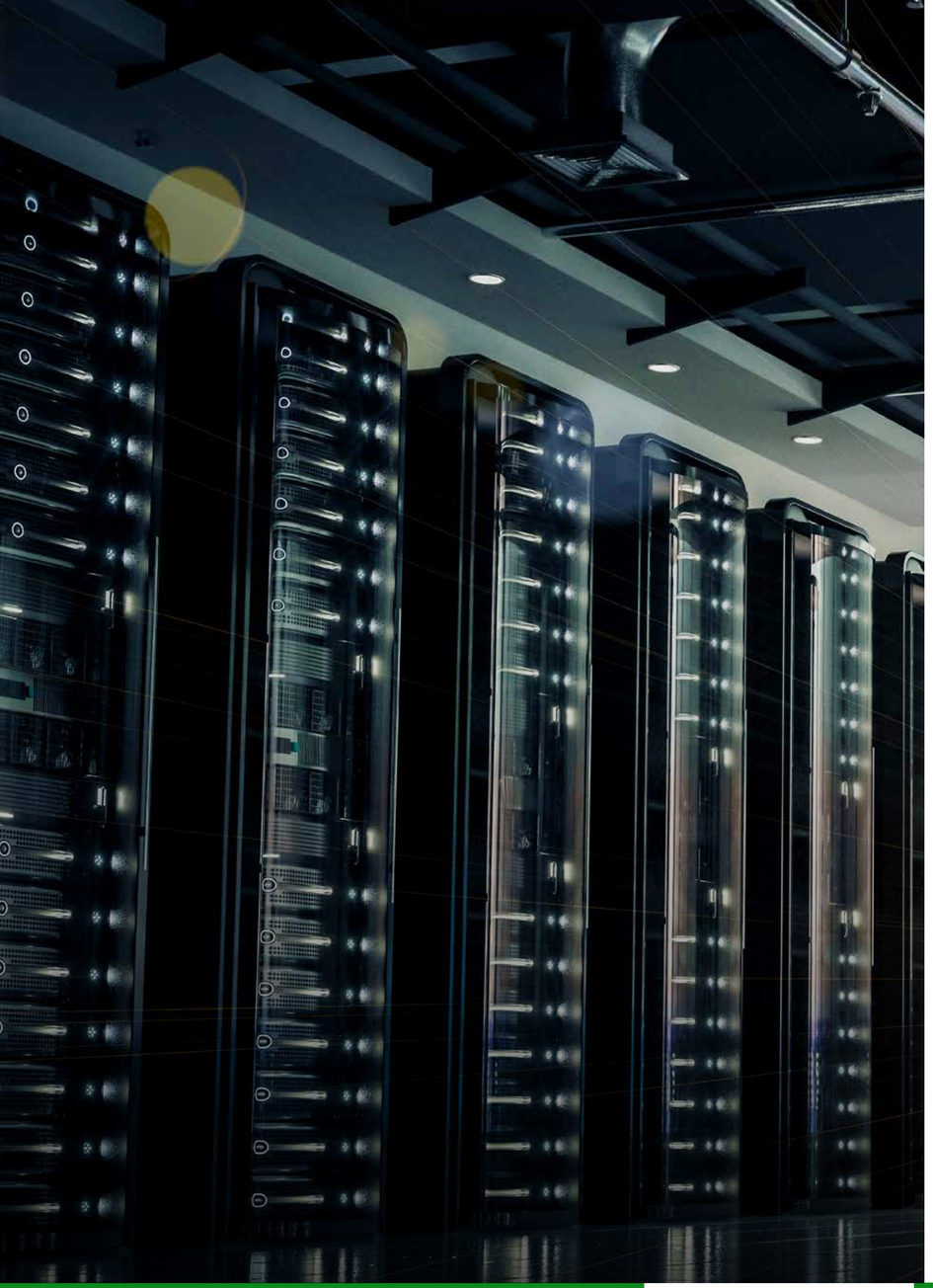
The liquid is then routed to a coolant distribution unit (CDU) with a heat exchanger to disperse the heat. Liquid from the secondary loop is then piped to a chiller unit and back to the CDU. Pipes and manifolds convey the liquid through the racks and out of the building.

A second liquid cooling method is immersion cooling. A server or entire rack is placed in a chassis filled with dielectric fluid to moderate the hardware's temperature. The dielectric fluid (a liquid that doesn't conduct electricity) then removes the heat from the hardware.

High-density AI clusters and liquid cooling are pushing the boundaries of data centers. This new wave of thermal management is required to manage and perform at this new demand that is transitioning cooling from air to liquid.



Treated water or a synthetic fluid is sent to the plate through a Technology Cooling System (TCS) to absorb the heat.



Existing tech, new designs

Liquid cooling will not completely replace air cooling, but will likely handle up to 80% of heat dissipation in AI data centers. Processors for networking and functions such as video GPU will continue to use air cooling systems for the most part. This means a hybrid approach is necessary to cool infrastructure supporting Gen AI workloads.

Liquid cooling is far from new. It is used in most cars, and in the 1960s, IBM used it for its System 360 computers. However, applying it to AI data centers will require new designs to manage airflow in high-density racks.

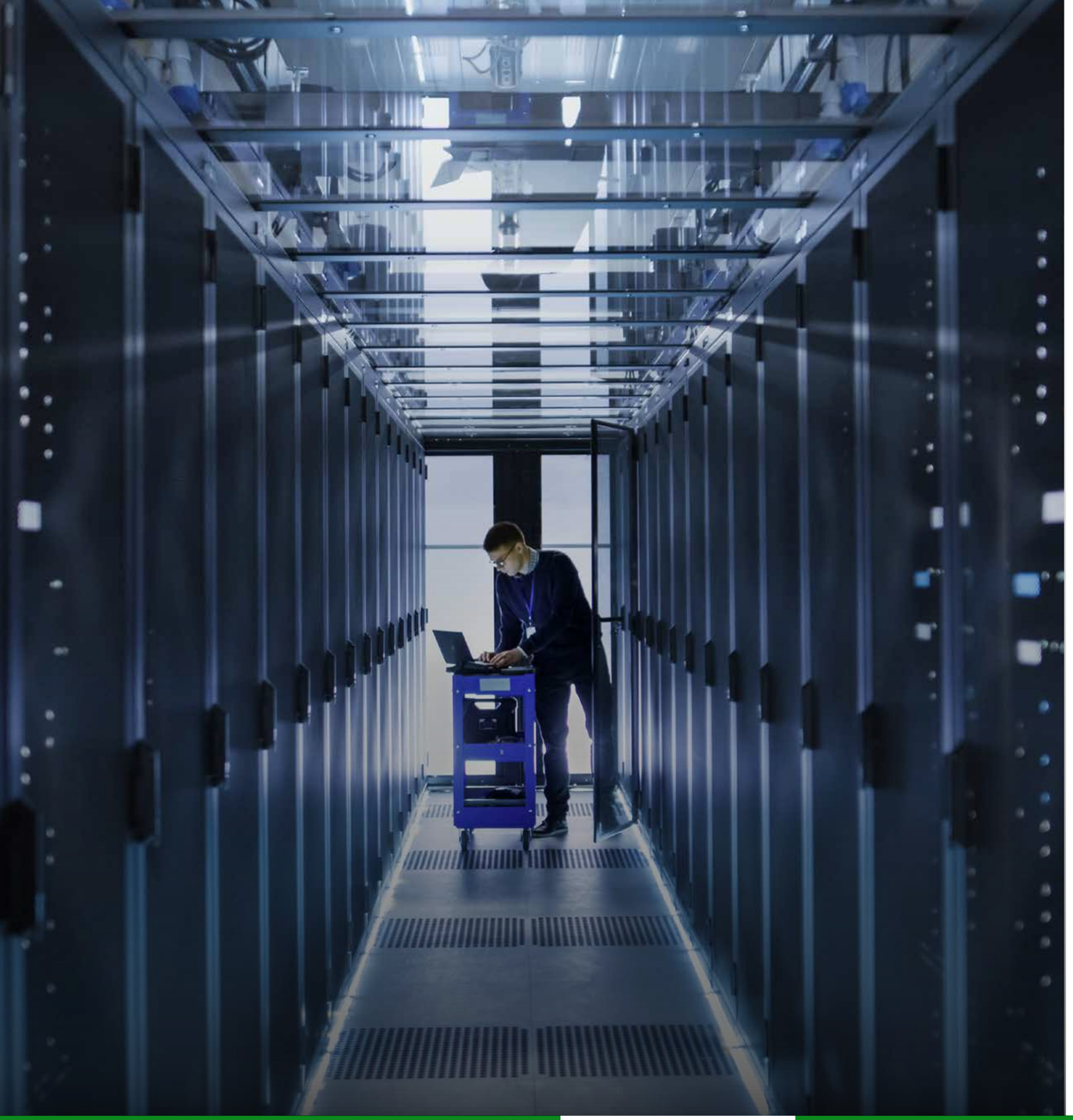
Designs must account for equipment such as water chillers, pipes, and manifolds to transmit the liquid. Modular data centers purpose-built for AI, whether deployed at the edge or elsewhere, must accommodate liquid cooling systems. As such, modular data centers will require added physical space for the equipment.

Modular data centers and sustainability

Prefabricated modular data centers that are purpose-built for AI can support sustainability strategies in the following ways:

- Data center modernization typically requires knocking down walls and expensive construction, which is not the case with modular data centers.
- Equipment used in modular units can be sent to the factory for refurbishing and reuse or recycling.
- Remote monitoring and management systems provide visibility and security, minimizing truck rolls and helping reduce carbon emissions.
- In locations with strict power allocation, modular data centers make upgrading equipment with more efficient systems easier.

Schneider Electric offers the most complete single-source prefabricated solutions for data centers. Since Schneider Electric manufactures 90% of the parts, it reduces waste, maintains quality, and optimizes efficiency.



Data center scalability

As AI use cases increase, scalability is essential. Right-sizing infrastructure has always been challenging, requiring some guesswork that often results in overcapacity or scalability restraints. When a data center reaches capacity, scaling typically means having to expand the facility or investing in new construction. However, the modular approach overcomes these requirements.

Rather than guessing how much space it may need in the future, an organization that leverages the modular approach must invest only in the space it currently needs. When it is time to scale, the organization can add more modules as necessary. Right-sizing the infrastructure becomes a much more efficient and precise endeavor.

Modular AI data centers

Modular AI data centers address challenges

An effective solution to the AI infrastructure challenge is to deploy purpose-built modular data centers designed specifically for AI applications. Modular in design, these data centers accelerate deployment time. New designs developed specifically for AI support much higher rack densities by leveraging cooling innovations to dissipate processor heat. The new designs address major infrastructure obstacles organizations are facing:

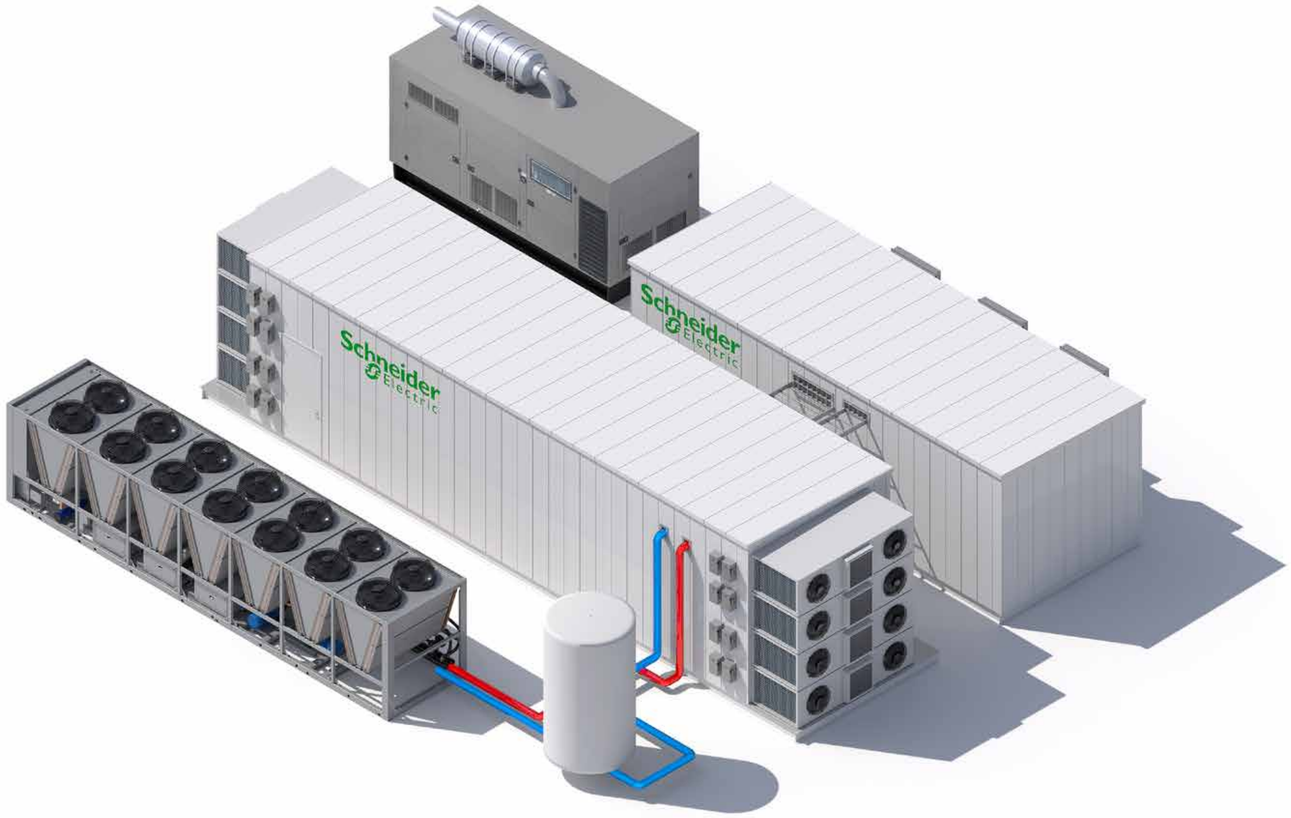
- Diminishing access to reliable power that can be scaled while securing uptime
- Scarcity of skills to design, build, and operate additional AI capacity
- A rush to design and build AI-ready data centers quickly to capture market demand
- Pressure to drive efficiencies in power and water while scaling for AI

Schneider Electric has been working on the AI infrastructure challenge by creating new modular units that can be scaled into repeatable clusters for deployment wherever organizations need rapid, adaptable, scalable infrastructure to achieve strategic AI goals. Companies investing in the modular data center approach gain quick entry into the world of possibilities Gen AI creates. They leverage a forward-thinking approach that delivers the flexibility and scalability they need.

Modular AI data center reference design

Modular AI data centers replicate standard data centers' infrastructure, security, cooling, and power backup in a portable, modular design. Schneider's new AI-focused modular designs include hybrid cooling solutions in an all-in-one IT solution. The designs can be scaled into repeatable clusters for implementation wherever organizations need rapid deployment and scalable infrastructure for AI workloads. AI Data Center features include:

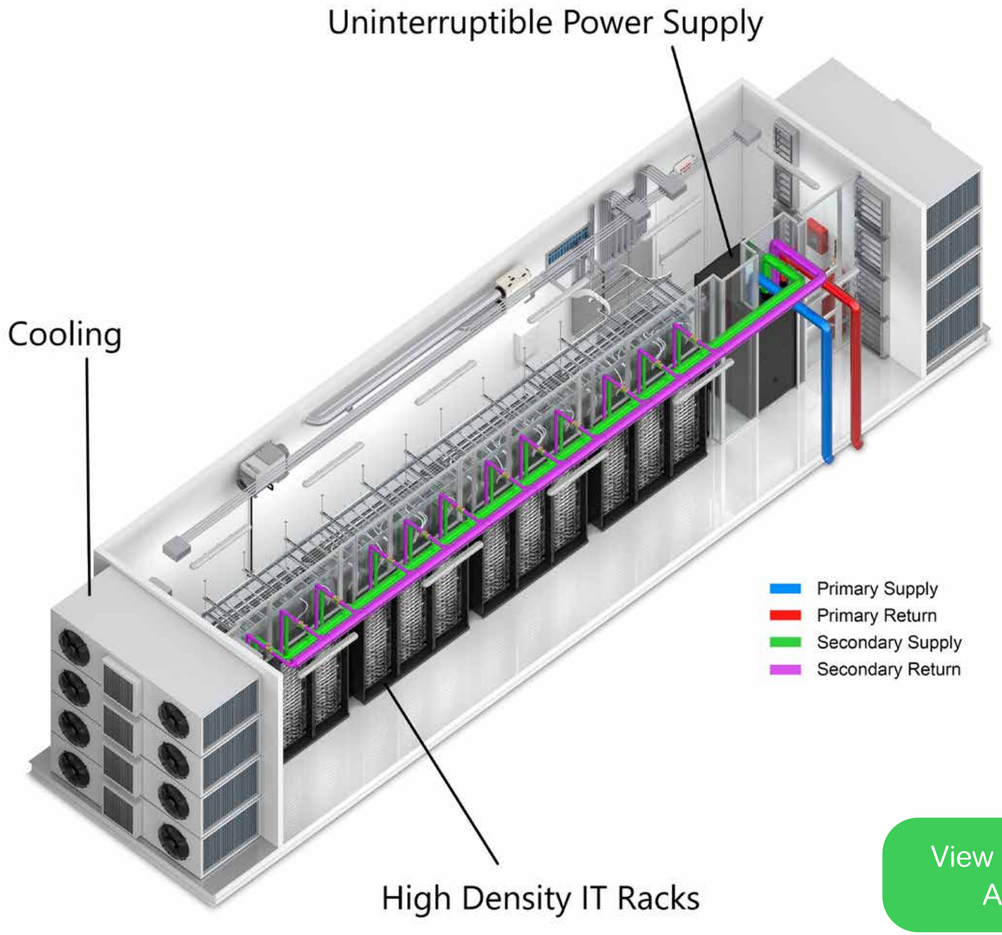
- Support for data-intensive AI training and inference workloads with high-density racks
- Up to 12 racks, each supporting 80 kW
- 1 MW UPS for power backup and protection
- Innovative, holistic cooling designs that combine liquid and air cooling



Your solution starts here

Start here to explore Schneider Electric's comprehensive end-to-end solutions designed to support cutting-edge artificial intelligence (AI) applications. This modular AI data center reference design provides the infrastructure needed to address critical challenges such as extreme power and thermal management, rapid load fluctuations, and the integration of advanced IT rack solutions, all while ensuring ease of scalability.

The growing demands of AI are driving the adoption of innovative technologies like liquid cooling systems, server racks with integrated liquid cooling manifolds, and other forward-thinking solutions. This reference design highlights key hardware advancements and demonstrates how modular AI data centers can enable organizations to deploy adaptable, scalable, and repeatable clusters. By leveraging these solutions, organizations can harness the full potential of AI technology to stay competitive in a rapidly evolving landscape.



View Modular Data Center AI Reference Here

Innovative cooling solutions

Modular AI Data Centers support unique cooling designs that distribute water or synthetic fluid to lower the temperatures of AI chips. Liquid cooling provides 80% of heat dissipation. The remaining 20%, which handles heat dissipation for non-AI processors, is accomplished with air cooling.

The AI-focused modular units are built with an accompanying water chiller supporting air and liquid cooling loops. Water chillers are often used to cool buildings by circulating a refrigerant through a closed loop that absorbs heat inside the buildings and ejects it to the exterior. In this case, they work to dissipate heat in modular units.

AI-focused modular units are built with an accompanying water chiller supporting air and liquid cooling loops.

Proven track record

Thanks to our vast experience in modular infrastructure solutions, Schneider has the number one market share in North America. We have a track record of excelling in delivering prefabricated modular data centers that are purpose-built for an organization's specific needs, including data-intensive AI workloads. Schneider handles everything from design to manufacturing to installation and maintenance services.

Schneider also provides 24/7 support through our EcoCare plan, which lets you identify issues that can cause outages if they are not caught early enough. With 24/7 proactive monitoring and alarming, EcoCare enables peak performance in operating modular data center infrastructure.

With seven dedicated modular data center facilities globally and multiple partners, Schneider Electric is the most local of global companies, ensuring we are as close as possible to our customers.

EcoCare enables peak performance in operating modular data center infrastructure.

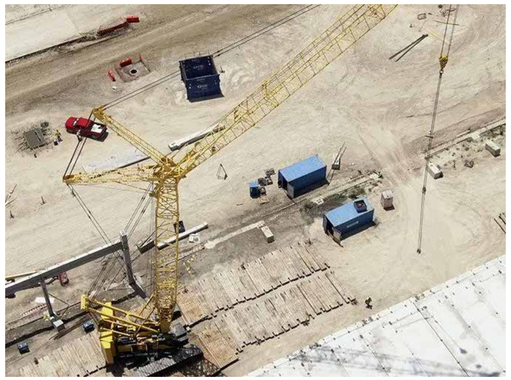
How Compass Datacenters scaled for AI

How Compass Datacenters scaled for AI

Compass Datacenters designs and constructs data centers for some of the world's largest hyperscalers and cloud providers worldwide. The biggest challenge these organizations face is how to scale for AI workloads while continuing to meet steady demand from online entertainment, cloud computing, and mobile apps. The pressure to build quickly is enormous, but requirements around land development, power allocation, local permits, design, planning, and construction get in the way.

Compass tackled these issues with a modular, scalable approach. The company worked with Schneider Electric to pre-engineer, build, and test power distribution and cooling components at the factory, which saved up to 50% of operational costs by rightsizing for actual demand. Modules can be added as demand increases.

In another cost-saving move, Schneider helped Compass optimize the supply chain by accelerating planning and deployment. This shortened the process to 10 months or less from the industry standard of two to three years. As a result, Compass saves up to 30% in total cost of ownership over conventionally built data centers.



"Speed to market is the most important factor for our client base, and it's our number one value-add. Schneider Electric makes it easier for us to make an impact by being a totally integrated supply chain partner."

—Chris Crosby, Founder & CEO
Compass Datacenters

Why customers choose Schneider Electric

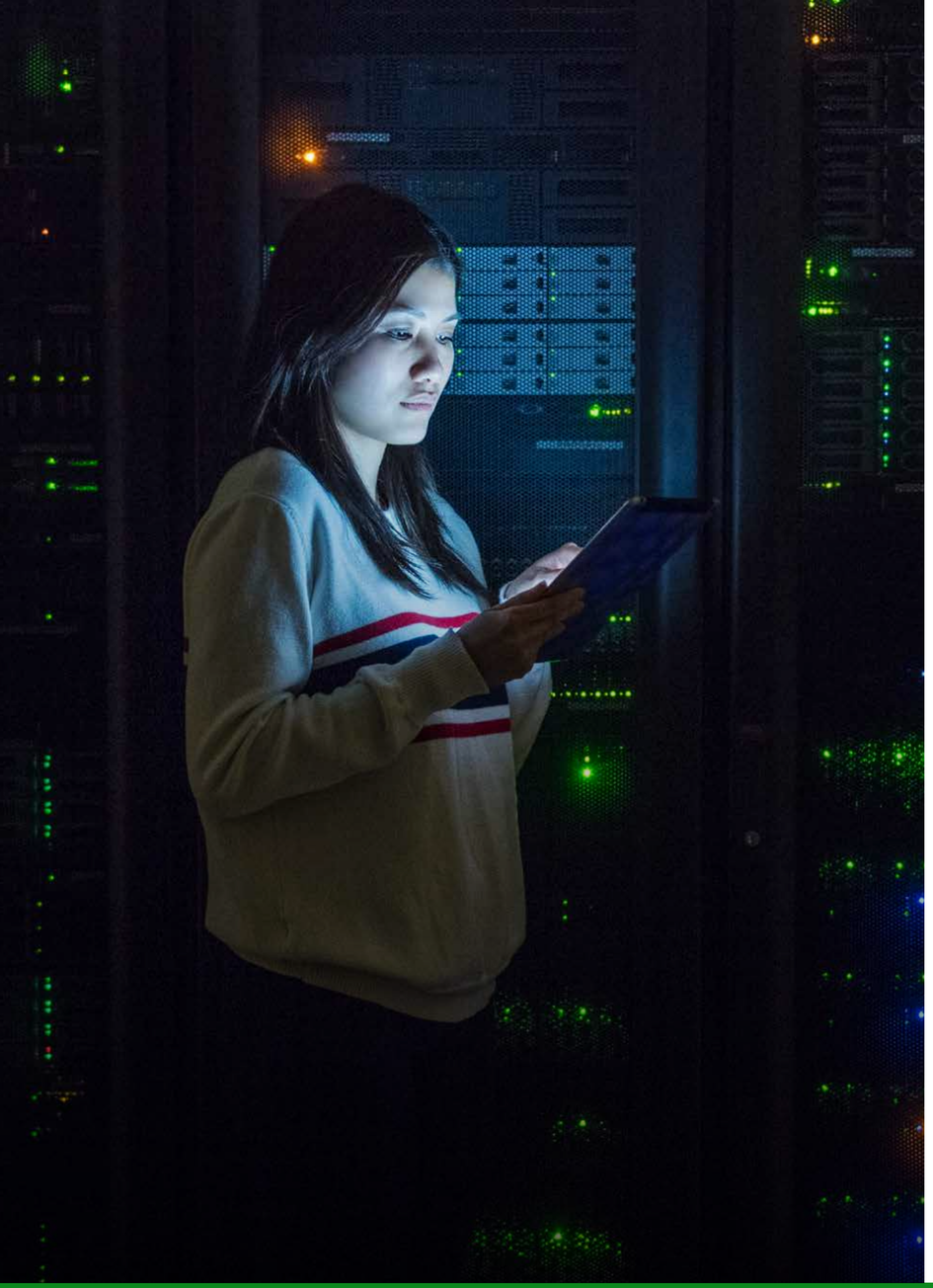
Like Compass Datacenters, many owners and operators choose Schneider Electric as their partner of choice to accelerate AI data center deployment. Modular data centers are at least 30% faster to deploy than traditional facilities.

The world's 10 leading cloud and service providers trust Schneider Electric solutions. We have the largest, most global solution portfolio, supported by local technical experts specializing in data center power, cooling, and sustainability.

With seven dedicated modular data center facilities around the world, and numerous partner facilities, we provide design excellence closer to our customers.

- **End-to-end solutions for all AI training and inference workloads**
- **Innovative solutions and unparalleled support**
- **Sustainability planning and strategy**
- **Investment in R&D, manufacturing capacity, and solution architect coverage**

Conclusion: Modular makes sense



Modular makes sense

Generative AI is about to revolutionize how organizations in almost every sector operate, transforming workflows and driving efficiencies that remove costs and accelerate production and go-to-market strategies. However, the data-intensive nature of AI requires a rapid expansion of infrastructure to support training and inference models.

With Schneider's Modular AI Data Centers, customers get a prefabricated, modular approach that immediately addresses scaling requirements, as in the case of Compass Datacenters, and helps to future-proof the business.

At the same time, purpose-built modular data centers support sustainability strategies in various ways, including leveraging more efficient, recyclable equipment, using fewer construction materials, and using more efficient processors and cooling solutions. With Schneider Electric's Modular AI Data Centers, organizations can reap the benefits of Gen AI faster to boost their competitiveness in the digital future.

View the modular AI data center reference design.

ANSI Design 1N **IEC Design 1N**

Life Is On | Schneider Electric

To learn more about Schneider Electric
prefabricated modular data centers, visit

se.com/prefab

Schneider Electric
35 rue Joseph Monier
92500 Rueil-Malmaison, France
Tel : +33 (0)1 41 29 70 00

998-23832850