# AI Workload Deployment in Data Centers: Retrofit, Outsource or Build New?

## A strategic framework for decision-making

A Data Center Research and Strategy Report

Patrick Donovan

Schneider Electric

# KEY TAKEAWAYS

## When to retrofit data centers for AI workloads

1. AI workloads often (but not always) require a new class of data center infrastructure. This includes high-density racks, direct liquid cooling, and upgraded power delivery systems which most existing facilities do not have nor were they designed to support.

2. Retrofitting existing data centers offers a strategic opportunity to deploy AI at lower cost compared to building new facilities or outsourcing to the cloud. And it may be faster than building new.

3. A structured decision framework (such as the one presented in this report) should be used for evaluating whether to build new, outsource (cloud or colo), or retrofit, factoring in TCO, cash flow, time-to-market, and other strategic priorities.

4. Evaluating retrofit feasibility begins with characterizing the AI workloads — defining their performance targets and operational profile — and determining the corresponding IT and physical infrastructure requirements.

5. Eight strategic factors (e.g., latency, data sovereignty, scalability, staffing, cash flow preference, and supply chain) heavily influence the decision on where to host AI workloads.

6. Retrofitting makes sense when existing facilities have solid fundamentals (e.g., adequate space, power, cooling, and structural integrity), and the organization seeks control and compliance without the full cost and timeline of new construction.

7. A five-step site evaluation process helps determine retrofit feasibility by assessing workload needs, current capacities, location relevance, operational requirements, and upgrade options.

8. Successful retrofit projects require careful planning. Use of validated reference designs, digital design tools, prefab modules, and encouraging close collaboration between IT and facilities teams is recommended.

RATE THIS REPORT

★★★★★

LET'S GET STARTED →

Many data centers, designed for conventional CPU-based workloads, are ill-equipped to handle the unprecedented demands of Generative AI. Training and agentic AI inference workloads come with extreme rack power densities.

## This requires a shift away from traditional power and cooling architectures.

These power-hungry AI workloads demand a new class of infrastructure.

It is defined by high-density racks, direct liquid cooling, and upgraded high-voltage AC (or even DC) power delivery to efficiently and reliably support the immense power and cooling needs, as well as the weight of the specialized hardware they run on. See **Figure 1**.

Not all AI workloads are energy intensive, however. Smaller AI inference models, for example, may behave like traditional IT having power, cooling, and space requirements that fit well within an existing site's capabilities.

## Impact on design and operations

| High power density | Increasingly too hot for air | Big & heavy |
|---|---|---|
| **Power** | **Cooling** | **Racks** |
| » Possibly more power capacity from utility<br>» Higher voltage and amperage distribution within IT space<br>» Power system engineering studies; new Operations & Maintenance (O&M) procedures<br>» Increased risk of upstream breaker trips<br>» Higher fault current potential in IT space | » Direct-to-chip or immersion liquid cooled servers<br>» CDUs, rack manifolds, and potentially rear door heat exchangers and new water piping are needed<br>» New O&M procedures & cooling controls | » Wider, deeper, taller<br>» Increased footprint<br>» Increased density of power and network cabling<br>» 2-3x heavier<br>» Hard flooring or structurally-reinforced raised flooring |

**Figure 1** - Summary of possible impacts of AI workloads on data center power, cooling, and rack systems

For owners of existing data center portfolios, deploying high density AI workloads presents both a significant challenge and a strategic opportunity. For colocation companies, supporting tenant AI workloads is a necessity. For enterprise companies, deploying AI across their business processes, offers, and services is equally important for being innovative and remaining competitive. Taking advantage of existing assets to potentially accelerate AI deployments at a lower cost is the opportunity.

While much attention has been put on the construction of new facilities and "AI factories", executives are looking at existing facilities with excess capacity. **Can they improve their return on capital by retrofitting existing data centers to support heavy, energy-dense AI infrastructure?** Is it faster and/or more cost effective than building new or outsourcing? This report helps answer these questions.

Note, this report is not about the specific changes needed in your physical infrastructure systems to support high-density AI workloads. We have targeted white papers and reports addressing those topics (click here).

This report describes a decision framework for determining if retrofitting makes sense and offers advice for executing a retrofit/modernization project more effectively (**see Figure 2**).

**Figure 2** - Topics covered in this report

A framework for deciding whether to build new, outsource, or retrofit
*Go there* ▶

5 steps to determine site suitability ▶

AI retrofit project tips ▶

# Deciding whether to build new, outsource, or retrofit

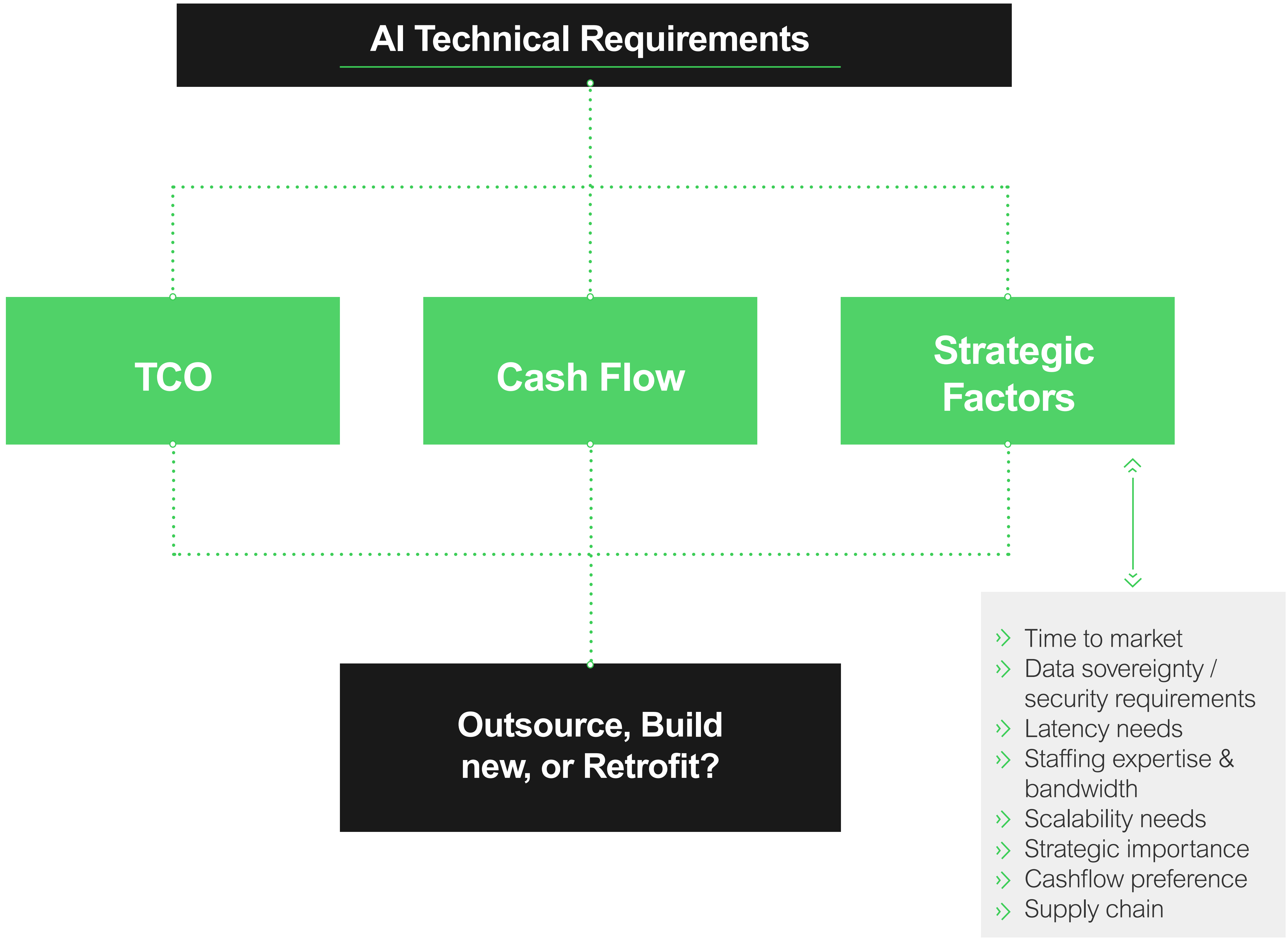
When faced with supporting or adopting AI applications, the fundamental initial decision to make if you're an enterprise company is whether to:

- Outsource to the cloud
- Outsource to colocation providers
- Build new
- Retrofit/modernize existing sites

A colocation data center service provider, on the other hand, would be looking at either building new or retrofitting/modernizing existing sites.

The decision of where to host AI workloads is complex. It impacts capital/operating expenditures, operational flexibility, performance, regulatory compliance, sustainability goals, time-to-value, and so on. This section presents a structured framework[1] to guide executives and project managers through this strategic decision-making process, focusing on the interplay between technical requirements, economic constraints, and long-term business alignment.

**Figure 3** summarizes the decision framework and influence factors for where to host the AI workloads.




The first two "influence factor" categories, total cost of ownership (TCO) and cash flow, are quantitative. Time-to-market and the other strategic factors consist of common business preferences and constraints that affect the decision qualitatively. Note, **Figure 3** illustrates common strategic factors, but this may not represent a complete list for you. Some of the strategic factors may eliminate certain alternatives altogether, while other factors can heavily influence the decision, depending on the business objectives and priorities of the decision maker(s).

## Understand AI technical requirements first

Before making infrastructure decisions, it's necessary to first characterize and understand the AI workloads being deployed. Step 1 of the five-step site evaluation process described in the next section goes into this in detail. To quantify costs and understand how much influence the strategic factors will have in your decision, you need to know what you are going to deploy or need to support. This includes the:

- AI application performance (e.g., model type/complexity/size, latency, number of concurrent users, scalability) and required service levels
- IT compute, storage, and networking requirements to achieve that performance
- Physical infrastructure requirements (i.e., power, cooling, space, weight) and grid capacity to enable and maintain that IT performance

# Analyze total cost of ownership (TCO)

A critical first step is the financial analysis of alternatives: would it be more sound to retrofit, build or outsource the AI workloads to a colocation provider or the cloud?

Taking on any AI application or workload, regardless of whether the data center is in the cloud, at a colocation site, or an end user's location, incurs the same types of capital expenses (CapEx) and operating expenses (OpEx). A colocation or cloud services provider has some economies of scale and negotiating power that allows them to build/lease and operate their data center for less (i.e. lower cost/watt to build, better electricity rates, better bandwidth rates) compared to an end-user, but it's important to view the costs from the end-user's perspective when evaluating any potential savings.

For an end-user building new or retrofitting a data center, the cost analysis should presume the following major costs will be incurred:

- Capex of:
  - AI IT (compute, storage, networking)
  - Design / engineering, project implementation
  - Data center physical infrastructure systems
  - Any necessary network and structural upgrades (retrofit only)
- Real estate – lease or buy
- Cost for bringing or upgrading the network to the site
- Utility connections/upgrades - Cost associated with bringing or upgrading electrical, water, gas utilities, etc. to the site
- Energy and network bandwidth
- Facilities staff (for new, additional staff for retrofit)
- Maintenance (for retrofit, include any additional maintenance costs related to AI workloads)

For outsourcing to a colocation company, the pricing structure to tenants varies from one vendor to the next, although in general, is based on monthly recurring charges. Some charge primarily by space consumed while others charge by kilowatt reserved or kilowatt used. Some also offer discounts as the reserved capacity increases. For an enterprise considering using a colocation vendor to host their AI workloads, consider the following costs:

- Upfront CapEx for AI IT (compute, storage, networking, racks)
- Charge for the space – often done on a per rack or per cage basis
- Charge for circuits – a 60-amp circuit costs less than a 100-amp circuit, some are moving to "pay per use" instead of "pay per circuit"
- Charge for bandwidth – a T1 line costs less than a T3 line
- Charge for remote hands – an extra fee for basic IT technician services
- Charge for reserving future space – different cost models exist for reserving space for future growth

When an end user is running AI applications in the public cloud, the pricing structure is very different from a colocation model. Instead of paying for physical space, power, and circuits, you're charged for the virtual resources they consume. Here's a break-down of the typical pricing structure for AI workloads in the public cloud.

- Charge for compute (GPU, CPU, and memory resources of virtual machines) based on a per second or per hour basis. (Models based on per API call or per token have emerged as well.)
- Charge for data storage typically on a per gigabyte per month basis
- Charge for data transfer out of provider's network (egress) based on volume. Ingress is usually free.
- Charge for any managed services used

These capital and operating expenses can then be compared cumulatively across the four scenarios (e.g., own, outsource, etc) starting at year "zero" through the time horizon appropriate for your business (likely 5 or 10 years ). Also, don't forget to consider any opportunity costs with any of these scenarios. For example, building and operating a data center consumes resources that may otherwise be used on other business initiatives.

# Analyze cash flow

The cash flow model (year over year capex and opex) for owning (build new or retrofitting) a data center and outsourcing (to colo or cloud) are very different. Performing a cash flow analysis helps you understand and compare the capital and operational expenses over time for each scenario. It goes beyond simple profitability and can better show the financial viability of each option.

Generally speaking:

- "Owning" cash flow model has lower TCO, higher upfront and occasional on-going capital expenses (maintenance/replacements)
- "Outsourcing" cash flow model has higher TCO, but lower to no capital expenses, so it preserves cash flow

Schneider Electric White Paper 171, [Considerations for Owning vs. Outsourcing Data Center Physical Infrastructure](), includes a quantitative analysis of building a new data center vs. outsourcing everything to a retail colocation provider.

For AI factories, however, the IT equipment becomes the dominant cost factor with some estimates putting the AI IT equipment at 55%-65% of the total site capex.[2]

Businesses may have a strong preference towards one of these two models. In one case, the overall TCO is lower, but there's an upfront capital expense followed by fluctuating capital and operating expenses yearly. In the cloud scenario, and to a lesser degree the colo scenario, there's a higher TCO, but no large capital expenditure upfront, and smooth predictable operating expenses. Although, remember, in the case of colo, there's a significant capex for the AI IT. A preference for predictable, steady operating expenses may trump any potential long-term savings for some businesses. Cash flow model preference is a key factor to consider.

## Strategic factors

We list 8 factors below (there may be more) that exert influence or dominate the decision-making process on where to put your AI workloads. Some, in fact, may be the determining factor depending on your situation and business priorities. These factors should be discussed, prioritized, and ultimately agreed upon by the project team and its stakeholders.

### 1. Time to market

Business needs drive the urgency of how fast you need to deploy new AI workloads. Cloud is likely the fastest way to get up and running with AI services as the facility, IT, and service is already in place and operational. Colocation would be the next fastest assuming the facility, power capacity, and related infrastructure is available for you to move your AI cluster into the space. Retrofitting an existing site would be the next fastest with building new being the slowest to market.

A common strategy is a hybrid approach that involves using the cloud to get up and running as soon as possible and run pilot programs to test the viability of new AI applications. And then, once validated, deploy them more broadly on premise or through a colocation partner.

### 2. Data sovereignty, privacy, & security

If the applications demand strict data sovereignty/security this may necessitate deployment in an on-premise (i.e., owned by the enterprise) or local colocation data center due to limitations in cloud environments. Highly sensitive data, such as personally identifiable information (PII) or classified government data, often requires data localization, mandating that it must be stored and processed within specific geographic borders (data residency) to comply with laws like GDPR or industry-specific regulations (data sovereignty). This doesn't necessarily rule out the cloud as an option, but it does require careful selection of more regional services and having strong contractual assurances to maintain data residency and control under foreign laws.

### 3. Latency

This refers to the time delay between an AI system receiving a request (input) and delivering a response (output). It is a measure of the system's responsiveness and is affected primarily by the network and compute speeds. For applications requiring real-time processing, response, and ultra-low latency, edge or on-premise data centers are likely necessary…i.e., on site where the application is being used. Examples include surgical assist systems, robotic surgeries, and complex manufacturing processes like semiconductor fabrication. Less severe latency requirements may mean deploying the workloads in the broader region via a colocation company or your own regional data center.

### 4. Staffing expertise/ bandwidth

Operating and maintaining AI infrastructure can be unique, requiring new skill sets and collaboration both in the IT operations space as well as facilities. Sourcing from the cloud offloads all the work of managing the physical IT stack, and the foundational software stack, as well as the supporting power and cooling infrastructure. Going to a colo also reduces the burden on your staff, but the IT software and hardware stacks remain your responsibility. Remote hands services may help reduce that work, however. When considering building new or retrofitting, it is important to evaluate your facilities and IT staff coverage and capabilities, perform a gap analysis, and understand what it will take to fill the gap. Uptime Institute projects more than 300,000 new data center professionals are needed. Education and training pipelines provide only a fraction of what's required.

## 5. Scalability needs

Cloud offers the greatest flexibility and speed for scaling AI workloads up or down on demand, of course, making them ideal for unpredictable or rapidly growing needs. In contrast, colocation provides more control and cost efficiency for steady-state or moderately scaling workloads, especially when paired with owned hardware. Building or retrofitting a private data center can support massive scale, but it requires significant upfront investment and longer lead times, making it more suitable for organizations with long-term scalability requirements. When planned for in advance, prefabricated, integrated modular infrastructure solutions make it easier for enterprises and colocation providers to scale over time.

Ultimately, the need for rapid, elastic scalability often tips the balance toward cloud. Some, as mentioned earlier, may pursue a hybrid strategy where they deploy AI in the cloud, at colocation facilities, and on-premise to give them increased flexibility for scaling.


## 7. Cash flow model preference

Organizations with a preference for preserving cash flow or minimizing large upfront investments typically favor the cloud. This offers a pay-as-you-go model and shifts costs to OpEx, allowing for more financial flexibility and easier budgeting. Colocation sits in the middle — it requires capital investment in the IT hardware but avoids the full cost of building, operating, and maintaining a data center.

In contrast, building or retrofitting a private data center demands significant upfront CapEx and extended payback periods, making it less attractive to cash flow-conscious organizations, such as retail and hospitality/food service where margins can be thin. Therefore, strong cash flow sensitivity generally steers decisions toward cloud or colo rather than private data center investments.

## 6. Strategic importance

This factor addresses how critical the workload is to the organization's core business, intellectual property (IP), or competitive advantage. If the AI workload is strategically vital — such as proprietary models involving IP, sensitive data, or mission-critical applications (financial, healthcare, etc.) — organizations may prefer deploying in a private data center or a colocation facility to maintain maximum control, security, and compliance.

Conversely, for less strategically sensitive workloads, the cloud may be more attractive due to its agility and lower operational burden. The more strategically important the AI workload — ones that have strong economic, or operational resilience impact — the more likely it is to be deployed in an environment where the organization has greater oversight and control.

## 8. Supply chain

The availability and timing of critical components - such as accelerator chips, high-bandwidth networking gear, cooling systems, and power delivery infrastructure - can dictate both cost and speed of deployment. Cloud providers often mitigate these risks by securing hardware at scale and offering immediate access, making them attractive when supply chains are tight or unpredictable. Colocation facilities provide a middle ground, offering ready-made infrastructure but still requiring you to source and manage specialized hardware, which may be delayed by shortages.

Building or retrofitting a private data center introduces the most exposure to supply chain volatility, as you bear the full burden of procuring scarce equipment, coordinating delivery schedules, and integrating systems. In short, strong supply chain constraints often push organizations toward cloud or colo options, while stable and predictable supply chains make private deployments more feasible.

**Table 1** creates and summarizes a decision matrix for the 8 strategic factors.

| Strategic Factor | Cloud | Colocation | Build New | Retrofit |
|---|---|---|---|---|
| Time to market | **Fastest:** immediate access to infrastructure and services | **Moderate:** quick if facility capacity is available; hardware install required | **Slowest:** long design, permitting, and construction cycle | **Slower:** depends on site readiness and retrofit complexity |
| Data sovereignty security | **Limited:** requires careful regional service selection and contractual assurances | **Strong:** workloads can be placed in regional/local facilities under your control | **Strongest:** full enterprise ownership and control | **Strong:** like new build but within existing footprint |
| Latency | **Variable:** depends on region/provider; not ideal for ultra-low latency | **Good:** regional deployments possible; closer to end users | **Excellent:** on-site proximity allows lowest latency | **Excellent:** same as new build if facility is at or near point of use of AI application |
| Staffing expertise / Bandwidth | **Lowest need:** provider manages infra, software stack, and facility | **Moderate:** provider manages facility; you manage IT hardware/software | **Highest:** enterprise must staff for facility + IT + AI infrastructure | **High:** similar to new build, but facility quirks may add complexity |
| Scalability | **Elastic:** near-infinite, on-demand scalability | **Good:** scalable within hardware owned and colo capacity | **High:** can support massive scale but requires planning, time, and CapEx | **Moderate:** scaling limited by physical site constraints. Prefab may help |
| Strategic importance | **Lower:** less control, better for non-core or lower sensitivity workloads | **Strong:** balance of control and external support; good for critical apps. | **Strongest:** maximum control for IP-sensitive or mission-critical workloads | **Strong:** significant control but may inherit facility limitations |
| Cash flow preference | **Best:** pay-as-you-go OpEx, preserves cash flow | **Middle:** hardware CapEx, but avoids facility investment | **Weakest:** heavy CapEx upfront with long ROI | **Weak:** substantial CapEx, though likely much less than new build |
| Supply chain | **Low exposure:** providers buffer against hardware availability risks | **Medium:** facility ready, but you must secure specialized hardware | **High:** full responsibility for sourcing facility + IT components | **High:** like new build; existing facility helps but sourcing still on you |

## So, in this context, when does retrofitting an existing site for AI workloads make sense?

Retrofitting an existing data center becomes the right option when an organization already has a facility with solid fundamentals—adequate space, bulk power and cooling capacity, and structural integrity—that can be upgraded more quickly or cost-effectively than building new. It's often chosen when:

- Moderate time-to-market is acceptable. Retrofitting is faster than a ground-up build, but slower than a cloud or colocation deployment.
- Regulatory or data sovereignty requirements demand on-premise control, but there's no appetite for the full CapEx and timeline of new construction.
- Existing assets can be leveraged (e.g., real estate, power feeds, chilled water capacity all readily available) to reduce CapEx cost compared to building new. This would be the case when AI workloads are closer to traditional rack power densities (e.g., many inference workloads will still be less than 20 kW/rack) where direct liquid cooling is not required.
- Workloads are strategically important enough to warrant control and security, but not so latency-critical or fast growing that only a purpose-built site will do.
- Staffing expertise exists in-house (or can be supplemented by software and services) to manage the unique quirks of a retrofitted facility.
- Cash flow can support CapEx, but leadership prefers a small incremental investment than a full greenfield build.

In short, retrofitting is most attractive when you need strong control and compliance as with a new (private) build, but you want to shorten timelines and lower costs by making use of existing available infrastructure.

**Our next section explains a five-step process for determining how suitable an existing site is for hosting new AI workloads.**

# 5 steps to determine site suitability

A key task is evaluating the existing site's suitability for hosting AI workloads. **Figure 4** lays out five steps for this evaluation that will determine the basic feasibility and scope of the project.

The output of this evaluation serves as a starting basis for estimating the cost and time to complete the retrofit project.

**Figure 4**

A five-step framework for determining the feasibility of retrofitting an existing site.

1 **Determine AI workload requirements**
What are you deploying?

2 **Determine current excess capacities**
Do you have enough power, cooling, and space?

3 **Evaluate site in context of its location**
Does the site warrant continued investment?

4 **Evaluate new operational requirements**
Can you address changes to processes and human resources?

5 **Consider options to increase available capacity & suitability**
What else can you do to increase site fitness?

Understanding the suitability of an existing site and the changes required will serve to inform whether to build new, outsource, or move forward with retrofitting. We recommend, in the interest of time and efficiency, that an end user go through these five steps in parallel to the decision framework discussed in the first section. You may quickly be able to rule out, or in, certain options before getting very far in the decision framework.

# Step 1 – Determine AI workload requirements

Determining the suitability of a given site for hosting AI workloads ideally starts first with understanding what the site needs to support. As with traditional IT equipment, AI server, storage, and networking equipment specifications determine the power, cooling, and space capacity needs, as well as floor loading requirements. Although we hear plenty about AI's high-power density and exhaustion of grid resources, individual AI deployments can span a wide ride range of needs for power, cooling, and space.

The amount of impact on a given data center's design depends on several factors including:

- AI model type (e.g., training vs. inference), size, number of concurrent users, complexity, and software architecture
- Latency, accuracy, and throughput requirements
- Scalability needs

By impact on design, we're referring to how disruptive or different the new workloads are from traditional, lower density IT that the existing site is presumably designed for.

**Figure 5**, in a general sense, illustrates the relative difference in impact on data center design between different workload types.

**Figure 5**
Relative level of impact of different workload types on data center design


It is important, however, that just because AI inferencing might have much less impact than LLM training would, it does not imply it won't be challenging and disruptive to start hosting it in an existing site. The power density, heat output, and weight might still be beyond what the existing site was designed to handle.

However, we think in most cases this will not mean dramatic changes will be needed. For example, for many enterprise data centers deploying AI inference workloads, traditional air cooling - perhaps with rear door heat exchangers or other close-coupled cooling - will suffice. Note, air-cooled servers are much taller than liquid cooled. See White Paper 210, Direct Liquid Cooling System Challenges in Data Centers, for more information on retrofitting cooling solutions to support AI.

Enterprise data center retrofit project teams should work with their IT Infrastructure & Operations (I&O) group, managed service providers, and hardware vendors to understand the physical compute, storage, and networking requirements for the new AI workloads.

Colocation companies would obviously work with their tenants' internal IT teams along with any relevant external vendors like consulting engineering (CE) firms and hardware vendors.

AI drives shift in solution integration levels leading to a unique AI ecosystem. To cope with the additional complexity of AI clusters, server vendors (and even chip manufacturers) have moved to a higher level of design and integration. Thus, all components of an AI rack and even the larger pod or cluster are designed, built, and tested together as a total solution. This includes the full software stack as well as the networking cables and switches. In other words, the decision-making power for everything in the rack is much more with the equipment hardware vendors than it is with end users. So, to accurately understand the physical infrastructure requirements, it is important to work directly with AI server vendors, chip makers, and/or neocloud providers (if involved).

However, you might be in a position where you don't yet know what the specific AI workloads will be. One could imagine being in an early planning stage as a colocation company trying to get a sense for how much new AI tenant business can be had using the excess capacity of an existing site. In that case, it makes sense to move on and perform Steps 2 through 5 below.

# Step 2 - Determine current excess capacities of the existing site

The next step is to figure out how much capacity is unused and available at the site, as well as what you will need in the future. This requires having a solid understanding of the site's capabilities in terms of:

- **Power and cooling capacity (kW)**
- **Rack power density (kW/rack)**
- **Space capacity (sq ft or sq m)**
- **Building structural capacity and capability**
- **Future growth capability**

Determining available capacity requires an understanding not only of the bulk power and cooling, but the power distribution and also air distribution. If you have sufficient bulk power and cooling, but insufficient distribution, there are often "low hanging fruit" changes that can increase that capacity a bit, such as implementing air containment or upgrading the rack-level power distribution.

But these are likely insufficient measures to support higher power AI workloads. In those cases, higher voltage and current carrying capacities will be needed and may require new transformers, higher amperage breakers, rPDUs, RPPs, or busway and larger water piping loops.

Another form of needed capacity is the available space, tied to the density capability. The higher the density deployed, the more physical space that becomes available for incoming IT equipment.

The building structural capacity and capabilities include, for example, floor loading capacity. While traditional data centers might be designed for 150-200 pounds per square foot (PSF), AI racks can easily exceed 400-500 PSF.[3] Raised floor systems might need reinforcing or even removal and replacement with solid concrete slabs in high-density areas.

Once the building envelope is understood, then evaluate the potential pathways to get the necessary power, water, and networking from the source to the loads. Consider, for example, the ceiling's ability to support a heavier load when using overhead trays for power and networking. Figure out how any new water piping will get to where it needs to be. All this can be a major structural undertaking and disruptive to existing workloads. It may require very creative solutions.[4]

You should also understand what the planned/expected future load growth is for that site beyond the new AI workloads to see how that might impact what your true available capacity is.

White Paper 177, "Determining the Power, Cooling, and Space Capacities when Consolidating Data Centers" explains in much more detail how to do this using a standardized approach.

Having the specific AI workload requirements along with the current capacities of the existing site, a gap analysis can now be performed to understand what modernization investments would be needed.

Note, there are third-party services available from vendors who can do this capacity assessment for you if you lack the experience and bandwidth to do it yourself. Schneider Electric partner and global IT service management company, World Wide Technology, for example, offers a Facilities Infrastructure AI Readiness Assessment.

# Step 3 – Evaluate the site in the context of its location

It's also important to reconsider the existing site's location to ensure it really makes sense to invest in modernizing or expanding the site for AI. **Table 2** lists some key aspects to consider.

**Table 2**
Key considerations for the site before deciding to modernize or invest in a retrofit project to support AI workloads.

| Location aspects | Considerations |
|---|---|
| **Age of site** | An older site will likely require more modernization than a younger one. |
| **Relevancy to served market** | Markets and demographics change over time. It is possible that the local market for the site's traditional IT services has diminished. It may be an opportunity to relocate those workloads to increase capacity for AI.<br>Is the local market optimal for AI services or perhaps, in the case of training particularly, it does not matter? |
| **Regulatory/tax landscape & public attitudes** | Is the local business climate suitable for continued or increased investment? |
| **Capacity for growth** | Is there enough properly zoned land?<br>Is there sufficient grid capacity to support the planned load? If not, what must happen to reach sufficiency?<br>If grid capacity is constrained, are there feasible options to deploy onsite prime power?[5]<br>How would a load increase affect tariffs, rate structures, and operations (e.g., mandatory load flexibility or curtailments)? |

# Step 4 – Evaluate new operational requirements

Adding high power density AI workloads to an existing data center brings new operations and maintenance procedures that require, in turn, new skillsets and staff bandwidth. Introducing liquid cooling systems brings unique equipment and controls unfamiliar to most data center operators. High power densities could bring higher available fault currents and arc flash risks into the IT space requiring new methods of procedure (MOPs) and emergency operating procedures (EOPs).

The heavy weight of AI gear will affect change management operations. If you offer "remote hands" services, understanding the unique aspects of AI physical compute, storage and networking is mandatory.

Consideration of these new operational requirements is important when deciding whether to retrofit or not. Does the site have sufficient expertise, staff bandwidth, and operational maturity (training, safety, CM, MOPs/SOPs/EOPs management) to operate and maintain AI infrastructure? We recommend performing a gap analysis to determine what would be needed. Consider operations and maintenance service contracts from 3rd parties to fill any gaps.

# Step 5 – Consider options to increase available capacities & suitability

The last step in determining the suitability of an existing site for hosting new AI workloads, is to consider how that site's IT capacity and suitability could be increased if needed. There may be options that, although they require additional investment, may positively impact the business case for retrofit. The following are a list of possibilities worth considering:

| Improvement | Description |
|---|---|
| **Reduce existing IT load, data center consolidation** | Relocate some or all traditional IT loads to other sites or cloud to free up capacity. |
| **Compact IT workloads** | Virtualize more to reduce physical workload and increase rack density to increase available space. |
| **Optimize cooling** | Improve airflow management (containment, blanking panels, etc.), raise set points, and use intelligent cooling controls to increase available cooling capacity. |
| **Prefabricated data center modules** | Consider using a prefabricated, modular approach to add more power, cooling and IT space outside of the existing building footprint. |

Modernization of a data center involves significant CapEx, it is almost always less than building a new facility. These updates should bring long-term energy savings, operational efficiency gains, improved reliability, lower maintenance costs, improved sustainability, and so on throughout the remaining life of the site.[6] These benefits should be factored into any business case for retrofitting or modernizing an existing site to support AI workloads.

# AI retrofit project tips

Assuming you have sites suitable for a retrofit to support new AI workloads, this section offers practical tips to consider once a decision to move forward with the project. The advice aims to help minimize project cost and time, as well as downtime risk for the existing live site.

It bears repeating: retrofitting live data centers is a technically complex and operationally sensitive process.

## Choose experienced CE firms and infrastructure vendors with deep portfolios

The obvious guidance here is to partner with consulting engineering firms and infrastructure vendors with proven experience in AI infrastructure deployments similar in scale and type to your project. Evaluate their experience with retrofits and not just new builds given the challenges of deploying in live sites originally designed for low density IT. The uniqueness of AI equipment and networking, as well as its power and cooling requirements makes "proven experience" more critical.

## Start with a reference design

One easy way to benefit from "proven experience" is start your early planning and design process with an AI reference design. These are high-level conceptual plans for how to build out an AI pod or cluster. Schneider Electric's reference designs include one-line electrical diagrams, piping diagrams, dimensioned floor layouts, and a bill of materials for major components.

We create effective AI designs that are tested, and then validated by both AI IT equipment vendors and the power and cooling vendors together. These designs provide an effective starting point for detailed design leading to faster, less error-prone deployments. Note, some vendors have created reference designs for AI liquid cooling control systems as well.

## Use digital design tools

When rack power densities are below 10 kW and there is excess bulk power and cooling capacity, adding IT equipment might be straightforward, without needing as much scrutiny and verification. Point-in-time power and cooling measurements might be used along with common power distribution components and existing cooling units that you're familiar with.

**This more manual, "eye-ball" retrofit design approach will not suffice for large high density AI training clusters.** An AI cluster drawing hundreds of kilowatts presents more serious consequences if you make a design mistake (i.e., not knowing actual peak to average power draws, being unsure of what loads are on which circuits, etc.). You can't afford to have unknowns and uncertainties with the design. Also, because AI cluster designs are so unique (e.g., non-standard high amperage rPDUs/busway, use of liquid cooling, etc.), there is greater uncertainty about how the cluster will perform on startup.

Prior to any changes, we recommend performing safety and technical studies including:

- capacity analysis
- protection coordination
- arc flash study
- short-circuit and device evaluation

Note, for larger projects involving an increase in grid capacity, a grid impact study may be needed so that the utility can confirm its ability to serve.

Using electrical design (a.k.a., power system engineering) software tools can simplify the data collection and calculations.

After the assessment, expect changes to the electric network to meet the needs of proposed AI clusters. In this case, electrical design software tools help you:

- select optimal electrical equipment
- prevent electrical faults
- develop effective methods of procedure
- implement proper safety protocols when working on and servicing the electrical network in the IT space

Existing data centers with digitalized single-line diagrams (iSLDs) will be able to simplify the assessment process described above. Accurate, intelligent, iSLDs reduce the time and expertise needed to collect data and perform the calculations. An iSLD is a more advanced single-line diagram stored and managed in specialized software that includes advanced functionality, awareness of the devices' characteristics and their operating behavior. It creates a digital twin of the physical electrical network. In essence, this one software platform can be used to design the electrical network, create and maintain the SLD, and perform all technical studies and safety assessments. See White Paper 281, How Modern DCIM Addresses CIO Management Challenges within Distributed, Hybrid IT Environments, for more on this topic.

For those designing large AI training clusters or "AI factories," emerging electrical design tools offer a digital twin platform to design and simulate AI cluster power requirements.[7] Once operational, this virtual replica can then be combined with real-time power system data and analytics to provide performance tracking, energy optimization, and "what-if" scenario planning.

## Leverage prefabricated data center modules for speed & capacity

Prefabricated modules are standardized and pre-engineered, built and tested in a controlled factory environment. This allows for simultaneous off-site manufacturing and on-site preparation. It provides opportunities for improved reliability and abbreviated timelines to bring new capacity online. For a retrofit project, this means you can reduce onsite work, lower the risk of service impacts to your live facility, and accelerate your time-to-market for new AI services.

Assuming there is enough bulk power capacity from the grid and outdoor space, ISO container-based prefab modules that are "AI ready" could be faster and an easier way to add additional power, liquid cooling, and space capacity to an existing site.

Some vendors also offer AI rack-ready prefabricated pod frames for the IT space that allow you to deliver AI infrastructure faster and easier, and supports all of the cabling, power, and liquid cooling piping and manifolds independent of the building.

## Use digital services to address staffing challenges

We've noted the shortage in educated data center technicians and utilities personnel. Good talent is an ongoing challenge and we can expect it to worsen. It's why hardware vendors and service providers are working to deliver digital services including remote monitoring. Using predictive analytics and AI algorithms, these modern condition-based maintenance tools can simplify and reduce O&M costs while improving reliability and responsiveness when service is eventually needed.

## Future proof the site to the extent possible

AI technologies and their IT stacks are evolving: market leader NVIDIA has reinvented the hardware cadence to deliver a new generation each year. These infrastructure updates represent massive compute efficiency gains while AI rack power densities continue to climb dramatically. So, we recommend using the retrofit or modernization project as an opportunity to future proof the site as much as possible.

Future proofing involves planning for even higher power densities, flexible power sources, and modular, spacious designs to adapt to rapid hardware innovations and increasing AI demands.[8] Consider these step-ups to begin:

- If in the U.S. and using 120V/208V in the IT space, move to a 415V distribution system.
- Work with your utility company early to plan for additional bulk power capacity and to understand what impacts increased demands may have on tariff structures, curtailments, and so on.
- Engage in direct conversations with hardware providers on how power requirements will evolve and impact your AI model size and applications. Design to those power specifications with flexibility to grow.
- Design spaces and pathways with the ability to add taller, wider, deeper rack enclosures; and use larger diameter water piping and higher capacity electrical conduit and network cabling than needed on "day 1."
- Remove raised flooring (for concrete slab) and adopt pod frames with overhead trays for supporting cabling, power, and piping.
- Invest in staff development and education around AI technology.
- Build effective supply chain partnerships to leverage key vendors and suppliers.

## Facilitate close cooperation between facilities and IT throughout the project lifecycle

Retrofitting or modernizing an existing live facility requires careful planning and coordination between multiple external and internal stakeholders. That cooperation and planning should begin at the early planning phase of the project to develop good teamwork and address domain needs and concerns are accounted for and addressed. Internally, its crucial for both IT and facilities – two groups traditionally siloed from each other – to work together. This is critical when the construction, installation, and testing and commissioning take place.

The high-power density of AI along with novel liquid cooling systems force greater facilities involvement in the IT space. Some of the key topics to work out between them include:

- developing solutions for electrical, mechanical piping, and network cabling pathways throughout the building.
- deciding if and when hot work is acceptable, maintaining safety.
- how to avoid impacting IT service levels by coordinating work schedules based on maintenance windows, scheduled downtime, IT service requirements, and tenant workload schedules.
- how to test & commission the installation, for example, without interrupting existing IT services.
- managing staff knowledge, expectations, and stress that comes naturally from disruption of routine operations and maintenance procedures when novel tech is being installed.

# NEXT STEPS

If pursuing a retrofit/modernization project for existing sites, here are next steps to consider:

## 1 Engage qualified vendors for an AI readiness assessment.

Begin by quantifying your site's current power, cooling, and space capacities versus the demands of AI workloads. Schneider and partners like World Wide Technology offer AI infrastructure readiness assessments to benchmark feasibility and identify upgrade needs efficiently.

## 2 Request a consultation with vendors on reference designs and digital twin tools.

Seek out validated AI reference designs and CE firms using digital design and power system engineering tools (iSLDs) that help model retrofit options, simulate grid impacts, and avoid design errors

## 3 Explore prefabricated and modular retrofit solutions.

Consider prefabricated, AI-ready modules modules that include containerized power and liquid cooling systems to add density or expand capacity quickly with minimal disruption to live operations.

# Endnotes

1 The framework presented here is adapted from Schneider Electric White Paper 171, "Considerations for Owning vs. Outsourcing Data Center Physical Infrastructure".

2 https://nucleuswealth.com/blog/the-economics-of-data-centres?utm_source=chatgpt.com

3 https://www.wwt.com/article/glossary-of-key-aspects-related-to-data-center-readiness-for-ai-workloads

4 https://www.digitalrealty.co.uk/resources/articles/integrating-ai-with-legacy-infrastruc-ture#:~:text=For%20example%2C%20many%20data%20centre,a%20solid%20concrete%20slab%20floor

5 Navigating Data Center Energy Constraints: Considerations for On-Site Prime Power

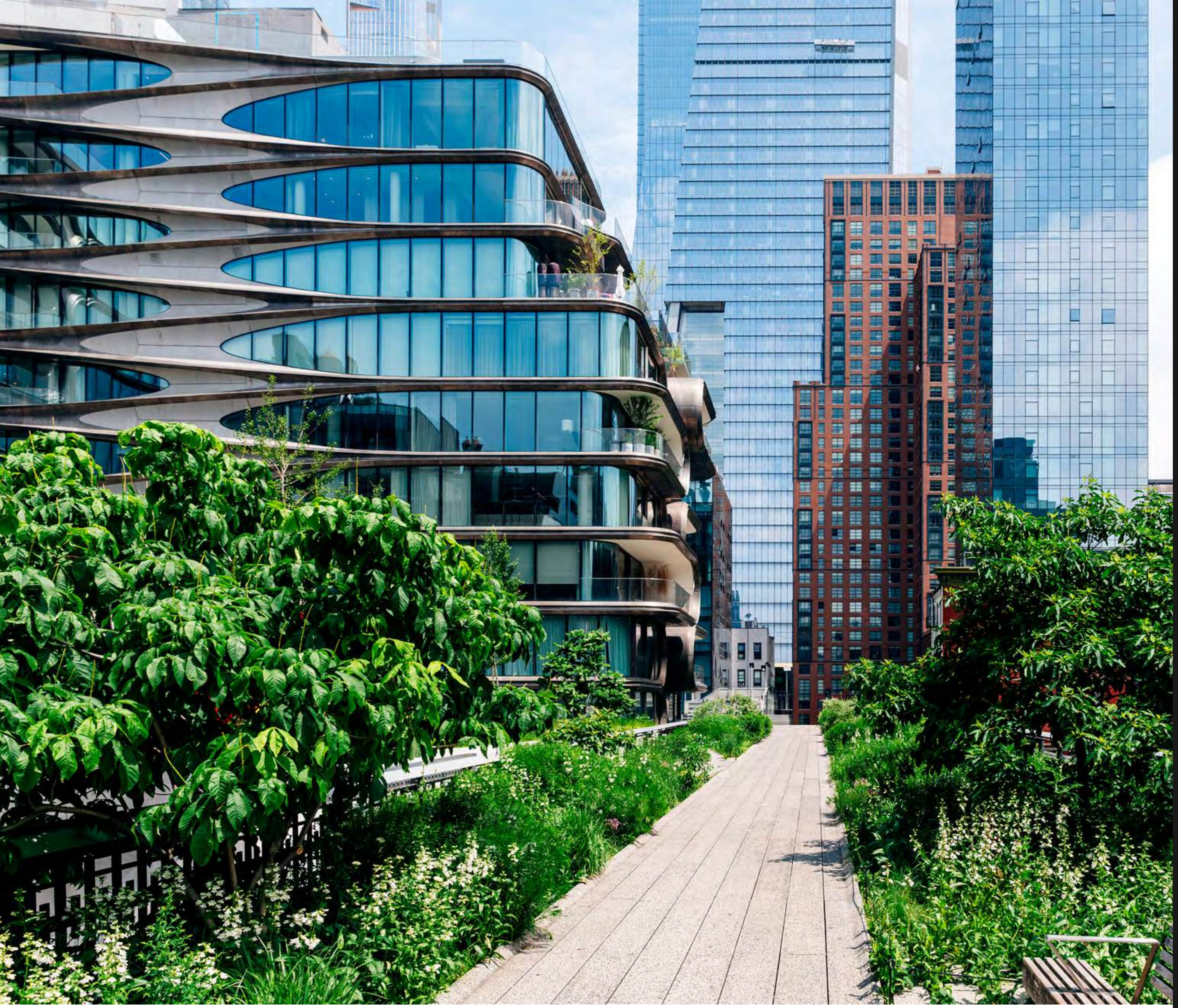
6 https://avidsolutionsinc.com/data-center-modernization-cuts-energy-costs-by-40-study-shows/

7 ETAP and Schneider Electric Unveil World's First Digital Twin to Simulate AI Factory Power Requirements from Grid to Chip Level Using NVIDIA Omniverse

8 Modernizing Legacy Data Centers for the AI Revolution with Schneider Electric's Steven Carlini | Data Center Frontier

**Note:** Internet links can become obsolete over time. The referenced links were available at the time this paper was written but may no longer be available now.

Page 3



## About the Author

**Patrick Donovan** is a Senior Research Analyst for the Data Center Research and Strategy Group at Schneider Electric. He has over 30 years of experience developing and supporting critical power and cooling systems for Schneider Electric's Secure Power Business unit including several award-winning power protection, efficiency, and availability solutions. An author of numerous white papers, industry articles, and technology assessments, Patrick's research on data center physical infrastructure technologies and markets offers guidance and advice on best practices for planning, designing, and operation of data center facilities.

## Unlock more insight here

- How 6 AI Attributes Change Data Center Design
- Retrofitting Existing Power Systems for AI Clusters
- Navigating Liquid Cooling Architectures for Data Centers with AI Workloads
- Optimizing AI Infrastructure: The Critical Role of Liquid Cooling
- Generative AI Inferencing Ramp-up: A CIOs Guide to Physical Infrastructure Considerations

THIS DOCUMENT IS TO BE CONSIDERED AS AN OPINION PAPER PRESENTING GENERAL AND NON-BINDING INFORMATION ON A PARTICULAR SUBJECT. THE ANALYSIS, HYPOTHESIS AND CONCLUSIONS PRESENTED THEREIN ARE PROVIDED AS IS WITH ALL FAULTS AND WITHOUT ANY REPRESENTATION OR WARRANTY OF ANY KIND OR NATURE EITHER EXPRESS, IMPLIED OR OTHERWISE.

THE COMPANIES WHO CONTRIBUTE TO THIS DOCUMENT WILL CONTINUE TO DEVELOP, MANUFACTURE, SELL, AND SERVICE OR OTHERWISE DO THEIR BUSINESS INDEPENDENTLY. THE STATEMENTS, DECLARATIONS OR CONCLUSIONS MADE IN THIS DOCUMENT SHALL IN NO WAY CREATE ANY JOINT RESPONSIBILITY OR LIABILITY OR ANY FORM OF COOPERATION OR JOINT VENTURE BETWEEN THEM.

© 2026 Schneider Electric. All rights reserved.

Schneider Electric