

INTELLIGENCE UPDATE

# Rethinking thermal storage as a capacity tool



Dr. Rand Talib 4 Jun 2026

Modern IT systems, particularly AI workloads, are pushing rack densities well beyond legacy infrastructure limits, while simultaneously increasing load volatility. This, in turn, increases both cooling demand and cooling variability. For large sites with tens of megawatts of design load, a further consideration is that the availability of electrical power capacity has become the primary capacity constraint. This is renewing interest in thermal energy storage (TES), not simply as an energy efficiency and resiliency measure, but as a way to shave peak cooling demand to unlock additional IT capacity within the substation power envelope.

The concept is not new. Commercial buildings have used chilled water and ice storage systems for decades to shift cooling power consumption to off-peak periods. Some data centers have already invested in similar capabilities in suitable locations, but adoption has been limited. What is changing is that the business case has become more compelling. AI training workloads, for example, introduce large transient peaks on top of already densified compute environments, while utilities across major data center markets struggle with interconnection delays, limited transmission capacity and rising demand charges.

Together, these conditions are making cooling flexibility more valuable. Rather than focusing only on reducing total energy consumption to lower the bill, many operators are increasingly interested in lowering total peak electrical power requirements by shifting cooling energy consumption.

## Why thermal storage matters now

Traditional enterprise data centers typically operated with relatively stable cooling loads. Mechanical systems (such as compressors, pumps and fans) were sized for peak demand and operated continuously with limited variation. AI compute-heavy environments are changing that assumption.

GPU clusters supporting model training can drive rapid fluctuations in cooling demand, particularly during high-intensity computational periods. In the standard engineering approach, infrastructure is sized for maximum load in design day conditions, operating at 100% load capacity at the highest ambient temperature assumed. However, in most cases, these are only short-duration, infrequent peaks, leaving portions of the electrical and cooling system substantially underutilized for most of their operating cycle.

TES provides a way to buffer these peaks. Cooling capacity can be generated during periods of lower IT demand and/or when temperatures are lower, to be stored for later use. During peak events, stored cooling can supplement or partially replace chiller operation, reducing the maximum of instantaneous electrical demand from the cooling plant. Depending on temperature set points, system design and climatic conditions, mechanical systems can represent more than 30% of the IT load at peak operation. TES can substantially lower this power overhead by spreading out cooling capacity more evenly.

This TES capability is becoming increasingly relevant in regions where power availability constrains data center growth. In U.S. markets such as Northern Virginia, Texas, and European hubs including Dublin and Frankfurt, data center capacity growth is constrained by limited grid capacity and interconnection timelines that often extend several years. Reducing peak cooling demand can free up more of the available electrical capacity for IT equipment.

The value proposition is therefore shifting. Historically, TES was primarily a tool for optimizing utility cost. Increasingly, it is being viewed as a strategy for managing capacity.

## How TES fits into data center cooling systems

TES is typically integrated at the chilled water plant level, upstream of CRAHs, CRACs or liquid cooling distribution systems. Most deployments use either chilled water storage or ice-based storage systems. Less commonly, thermal storage systems may use a different proprietary phase-change material. This report focuses on chilled water and ice systems:

- **Chilled water TES.** A traditional form of thermal storage in data center applications, that uses large, thermally stratified tanks in which colder water remains at the bottom while warmer return water accumulates at the top. In conventional chilled water systems, water is typically stored at approximately 39-44°F (4-7°C), with supply-to-return temperature differentials often reaching 18°F (10°C). This relatively large temperature differential allows a significant amount of cooling energy to be stored. Discharge is nearly instantaneous, making chilled water TES well suited for emergency and backup cooling applications. Its primary limitation is physical footprint, as the tanks require substantial space and can be difficult to deploy in retrofit environments. At elevated facility water temperatures, such as 63°F (17°C) and above, the available temperature differential is often smaller. Achieving the same cooling capacity may therefore require larger storage volumes.
- **Ice-based TES systems.** An alternative to chilled water TES is ice systems. Ice systems freeze water in storage cells during charging periods and are typically connected to the facility through a secondary glycol-water loop, often using ethylene or propylene glycol mixtures for freeze protection. Ice has roughly 3.5 times the thermal density of chilled water by volume, enabling a smaller footprint. The most significant trade-off is a 12- to 15-minute discharge lag before chilled water reaches design temperature. This disqualifies ice as a sole backup source but makes it well-suited for scheduled peak shaving alongside existing chillers. Producing ice is also a more energy intense

procedure.

**Table 1** summarizes the primary differences between chilled water and ice-based TES configurations.

Table 1 Comparing chilled water and ice-based thermal energy storage

Parameter	Chilled water thermal energy storage	Ice-based thermal energy storage
Storage medium	Stratified chilled water	Ice or proprietary phase-change material
Footprint	Large tanks	Smaller modular systems
Response time	Near-instant	Delayed discharge response approximately 10–15 min
Retrofit suitability	More difficult	Easier
Best use case	Backup and load support	Peak demand reduction
Loop configuration	Often integrated directly into chilled water loop	Commonly uses a secondary glycol loop

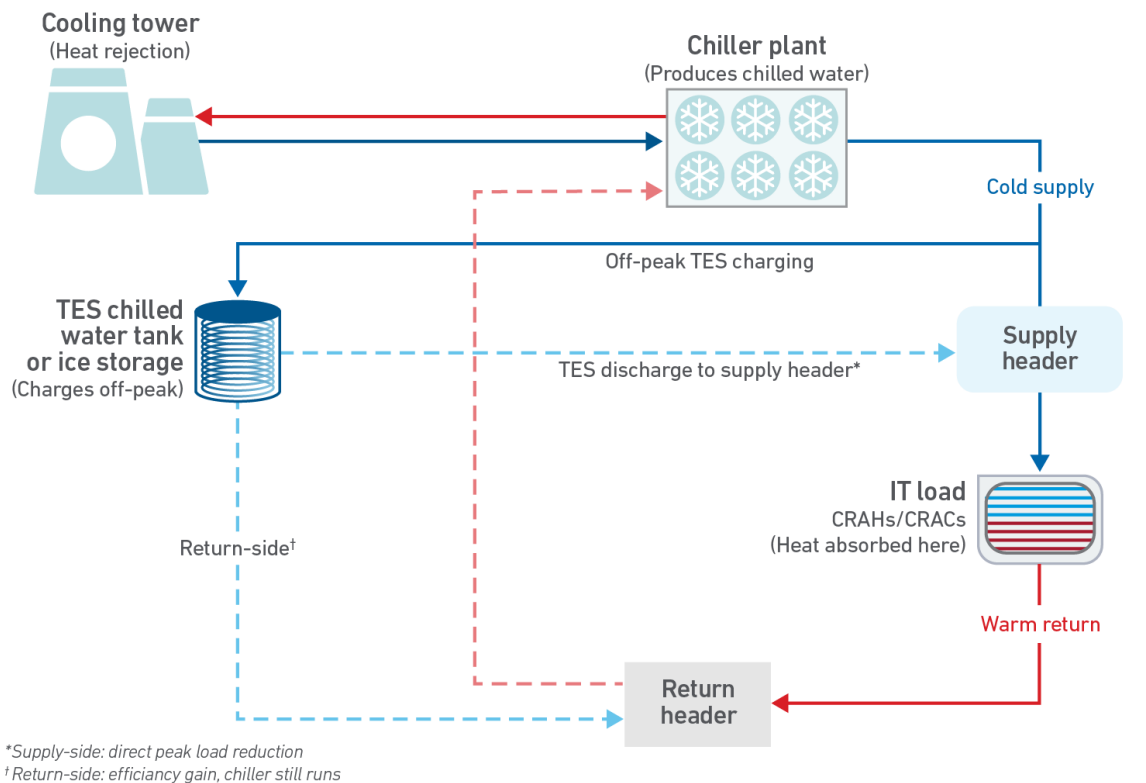
UPTIME INSTITUTE 2026



Most data center deployments favor a parallel configuration, where TES operates alongside the chiller plant rather than replacing it. Chillers, storage, or a combination of both serve the cooling load depending on conditions, preserving design redundancy while giving operators flexibility to reduce peak chiller demand when it matters most.

As shown in **Figure 1**, supply-side integration is the preferred connection point: during discharge, pre-cooled water enters the distribution loop directly, delivering an immediate reduction in peak electrical load. Return-side integration is possible and can improve chiller efficiency by pre-cooling return water before it re-enters the plant. However, it does not reduce chiller operation during peak periods and is better viewed as an efficiency measure rather than a capacity one.

Figure 1 TES parallel configuration, chilled water loop integration



UPTIME INSTITUTE 2026

uptime  
INTELLIGENCE

## Considerations for adoption

Several factors are increasing interest in thermal energy storage among data center operators:

- Technology evolution.** Thermal storage itself is not new, but the technology is evolving. Vendors, such as Nostromo Energy, are developing modular ice-based systems designed for integration with existing chilled water plants, making TES more practical for retrofit deployments and incremental capacity expansion. At the same time, the US Department of Energy's National Renewable Energy Laboratory (NREL) is investigating underground thermal energy storage (UTES), which uses subsurface geology for longer-duration cold storage. Together, these developments are expanding the role of TES beyond utility cost savings and toward capacity management, cooling flexibility and grid interaction.
- Economics.** TES is most attractive in regions with strong time-of-use pricing, high demand charges or limited power availability. Facilities with flat electricity pricing or ample utility capacity may see limited financial benefit.
- Operational complexity.** Effective deployment requires coordination between chiller sequencing, storage state and real-time cooling demand. Ice-based systems may also introduce secondary glycol loops that require additional maintenance and monitoring.

### The Uptime Intelligence View

Although thermal storage is not new, its role in data centers is changing as operators contend with AI-driven cooling variability and growing power constraints. Rather than functioning primarily as an energy efficiency tool, TES is increasingly being evaluated as a

means to shift cooling demand and increase usable IT capacity within fixed electrical limits. The most significant recent development is not the underlying technology itself, but the operating environment around it. AI workloads, constrained utility infrastructure and emerging modular storage products are creating conditions that may support broader adoption than thermal storage has historically seen in data centers.

## ABOUT THE AUTHOR

---



### Dr. Rand Talib

4 Jun 2026

Dr. Rand Talib is a Research Analyst at Uptime Institute with expertise in energy analysis, building performance modeling, and sustainability. Dr. Talib holds a Ph.D. in Civil Engineering with a concentration in building systems and energy efficiency. Her background blends academic research and real-world consulting, with a strong foundation in machine learning, energy audits, and high-performance infrastructure systems.

[rtalib@uptimeinstitute.com](mailto:rtalib@uptimeinstitute.com)

## **About Uptime Institute**

Uptime Institute is the Global Digital Infrastructure Authority. With over 4,000 awards issued in over 122 countries around the globe, and over 1,100 currently active projects in 80+ countries, Uptime has helped tens of thousands of companies optimize critical IT assets while managing costs, resources, and efficiency. For over 30 years, the company has established industry-leading benchmarks for data center performance, resilience, sustainability, and efficiency, which provide customers assurance that their digital infrastructure can perform across a wide array of operating conditions at a level consistent with their individual business needs. Uptime's Tier Standard is the IT industry's most trusted and adopted global standard for the design, construction, and operation of data centers.

Offerings include the organization's Tier Standard and Certifications, Management & Operations reviews and assessments including SCIRA-FSI financial sector risk assessment, the Sustainability Assessment, and a broad range of additional risk management, performance, availability, and related offerings. Uptime Education training programs have been successfully completed by over 100,000 data center professionals, such as the much-valued ATD (Accredited Tier Designer) and AOS (Accredited Operations Specialist). The Uptime Education curriculum has been expanded by the acquisition of CNet Training Ltd. In 2023.

Uptime Institute is headquartered in New York, NY, with offices in London, Sao Paulo, Dubai, Riyadh, and Singapore, and full-time Uptime professionals based in over thirty-four countries around the world.

For more information, visit [www.uptimeinstitute.com](http://www.uptimeinstitute.com)