PROCEEDINGS, 50th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 10-12, 2025
SGP-TR-229

Reducing Data Center Peak Cooling Demand and Energy Costs With Underground Thermal Energy Storage (UTES)

Jeffrey A. Winick[1], Kevin Kitz[2], and Andy Adams[1]

[1]U.S. Department of Energy, Golden, Colorado 80401

[2]KitzWorks LLC, Boise, Idaho

jeff.winick@ee.doe.gov

Keywords: Data Centers, Grid, AI, UTES, RTES, BTES, closed loop, VPP

ABSTRACT

By recent estimates, data center energy demands are projected to consume between 6.7% and 12% of U.S. annual electricity generation by the year 2028, driven primarily by expanded demands from cloud services, big data analytics, and Artificial Intelligence (AI) (Shehabi et al., 2024). As much as 40% of data center total energy consumption are loads associated with the site infrastructure cooling systems, and these are often highly water consumptive (Aljbour et al., 2024). For energy system planners, this presents significant challenges to meeting and managing the anticipated loads, and especially the peak loads of projected data center deployments.

Geothermal technologies offer two unique solutions to these challenges: 1) by serving loads through the deployment of new conventional and/or next-generation geothermal power technologies such as EGS and 2) through an often-overlooked opportunity to reduce data center peak cooling loads. The latter is the focus of this paper which explores Cold Underground Thermal Energy Storage ("Cold UTES") as an emerging industrial-scale geothermal cooling solution. This cooling solution is energy efficient, non-water-consumptive, and utilizes long duration energy storage (LDES) on both diurnal and seasonal time scales. Cold UTES has the potential to also function as a virtual power plant (VPP). The US Department of Energy's Geothermal Technologies Office is supporting R&D to understand the grid and system-wide value, costs, and impacts of deploying this emergent cooling solution at scale.

1. INTRODUCTION

In response to transformations in technologies like artificial intelligence (AI), data center expansion, new domestic manufacturing, and electrification in different sectors, the United States is returning to a period of rising electricity demand, with total energy demand potentially growing ~15-20% in the next decade (Figure 1; DOE, 2025).

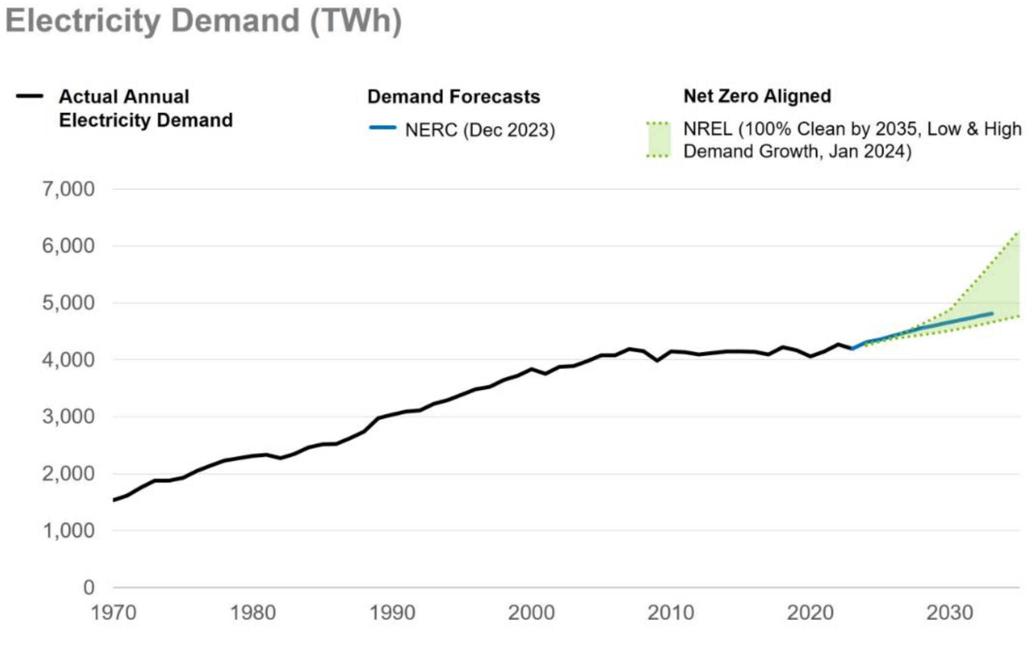


Figure 1: U.S. Electricity Demand (1970-2035; DOE, 2025)

Rapid adoption of Artificial Intelligence (AI) has driven large increases in data center energy consumption with exponential near-term growth projections. A recent report by Lawrence Berkely National Laboratory provides estimates of total U.S. data center electricity use, including servers, storage, network equipment, and infrastructure from 2014 through 2028 (Shehabi et al., 2024). The analysis indicates electricity consumption of US data centers is currently growing at an accelerating rate, with a compound annual growth rate of 7% from 2014 to 2018, increasing to 18% between 2018 and 2023, and then ranging from 13% to 27% between 2023 and 2028. This report found that data center energy demands are projected to consume between 6.7% and 12% of US annual electricity generation by the year 2028 (Figure 2).

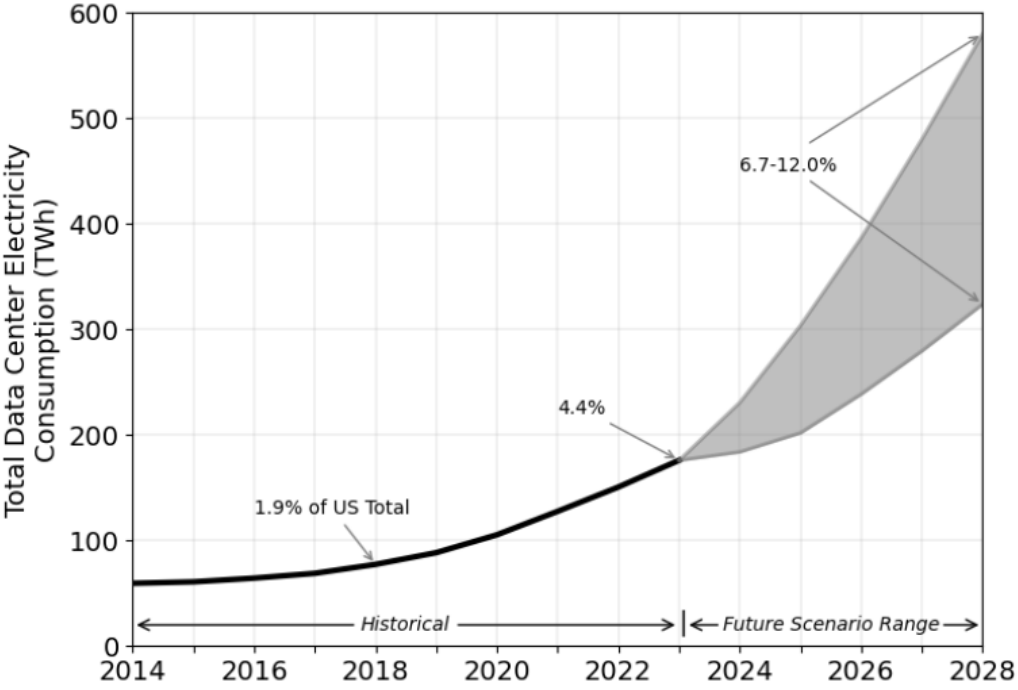


Figure 2: Total U.S. data center electricity use from 2014 through 2028 (Shehabi et al., 2024).

Over a similar time horizon (2023-2030) The U.S. Department of Energy (DOE) conservatively estimates total electricity system peak demand to grow by approximately 200GW (Downing et al., 2023; Razdan et al., 2025). For energy system planners, this presents significant challenges to meeting and managing the anticipated loads, and especially the peak loads of projected data center deployments.

A typical data center's primary electricity consuming hardware includes the IT equipment (e.g., servers, storage systems, and network infrastructure), auxiliary components (e.g., uninterrupted power supply, security systems, and lighting), and the data center cooling system which is critical for maintaining an optimal temperature that prevents hardware malfunction and ensures equipment longevity. Figure 3 shows that as much as 40% of a data center's energy consumption is attributable to its cooling load (Aljbour et al., 2024).

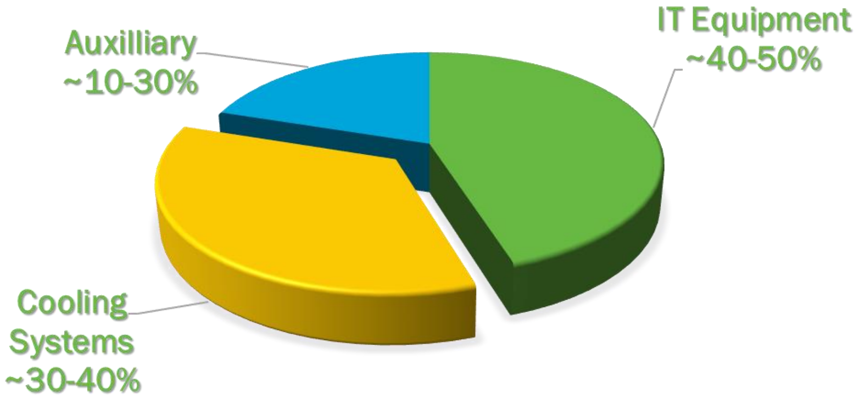


Figure 3: Data center primary electricity-consuming hardware and equipment (Aljbour et al., 2024).

Technologies which can improve data center cooling energy efficiency while also taking advantage of seasonal or diurnal pricing and energy arbitrage, present a unique opportunity to reduce overall data center energy consumption and manage peak loads. Cold Underground Thermal Energy Storage ("Cold UTES") technologies are a long duration energy storage (LDES) solution, offering an important opportunity to reduce and shift data center peak cooling loads while also significantly improving cooling efficiency (high coefficient of performance, or COP), as compared to alternative cooling methods. By reducing cooling loads at peak hours, Cold UTES can increase the effective load carrying capacity of the transmission system, serving more customers without the need for new transmission construction. Transmission planners commonly refer to this as a "non-wires alternative." Cold UTES can enable both a time-shift in when the cooling loads occur and a net reduction in the amount of power required for cooling. Since data centers can have large cooling loads, the reduction of peak demand can provide significant value as megawatts of new power plant capacity that do not have to be built or operated. These avoided megawatts are colloquially referred to as "negawatts," after Lovins (1990). The US Department of Energy's Geothermal Technologies Office (GTO) is supporting work to understand the grid and system-wide value, costs, and impacts of large-scale and widespread deployment of this emergent cooling solution.

2. COLD UTES EXPLAINED

Underground thermal energy storage (UTES) is described by the U.S. Geological Survey as a subset of a larger suite of geologic energy storage technologies that are capable of storing excess energy for later use to meet seasonal demand or provide baseload power when renewable energy sources are variable (Buursink et al., 2023). UTES can be further subdivided into three main thermal energy storage configurations: Aquifer Thermal Energy Storage (ATES), Borehole Thermal Energy Storage (BTES), and Reservoir Thermal Energy Storage (RTES; also referred to as GeoTES after McLing et al., 2022 and Akindipe et al., 2024). The DOE's Cold UTES initiative described herein, intentionally focuses on BTES and RTES opportunities and does not consider ATES (Figure 4).

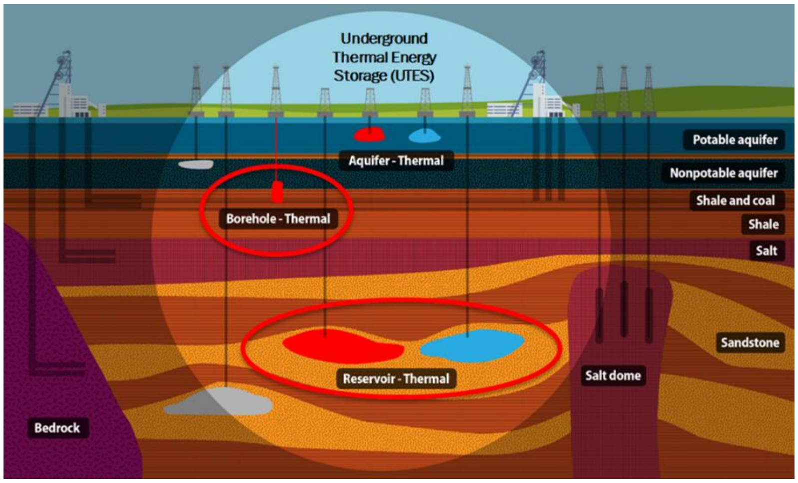


Figure 4: Schematic cross section, modified after Buursink et al. (2023), showing examples of Underground Thermal Energy Storage (UTES) as Aquifer Thermal Energy Storage (ATES); Borehole Thermal Energy Storage (BTES); and Reservoir Thermal Energy Storage (RTES). The DOE's Cold UTES project will focus on BTES and RTES technologies.

An RTES system is characterized by deep, laterally extensive, porous and permeable strata, saturated with brackish or saline (non-potable) water. An RTES system is an open loop system which flows water or brine as the heat transfer fluid through the formation. A BTES system on the other hand may or may not have the same permeability and porosity needed to function as a volumetrically significant RTES system, and thus it relies instead on conductive heat transfer with the near-wellbore rock volume. A BTES system would typically be accessing shallower depths, and thus from a capital cost development perspective, and once the drilling cost differentials are considered, may be cheaper, or comparable to the capital costs required to develop an RTES system. Ginosar et al. (2023) and references therein provide more complete descriptions of the technical details of UTES systems.

While there are technical challenges in developing these systems, and the applicability and techno-economic viability of a given RTES or BTES system will be site-specific, UTES system technology is generally location-agnostic. That is to say that a particular site which lacks the necessary geology to host a suitable RTES system at depth, may still be developable as a shallower BTES system instead – or vice versa. While more research is needed to identify the national scale potential of UTES systems, in principle these factors make the technology applicable to a wide range of potential end-uses and across very broad geographic regions of the U.S.

The key property of a UTES system enabling its function as LDES for cooling (or heating) is the heat capacity of the rock, or its ability to store and release heat. To illustrate this point, consider the following generalized parameters of a rock mass of varying porosity comprising a hypothetical UTES system shown in Table 1.

Table 1: Volumetric heat capacity of a cubic meter of saturated rock mass comprising a hypothetical UTES and normalized for porosity.

	Volumetric Heat Capacity		Volumetric Heat Capacity of a m3 of saturated rock mass normalized to porosity	
Porosity	Water	Rock	Water	Rock
	KJ/m3-ºK	KJ/m3-ºK	KJ/ºK	KJ/ºK
5%	4,184	2,650	209	2,518
10%			418	2,385
20%			837	2,120

Volumetric heat capacities for water and rock were calculated using densities of 1g/cc and 2.65g/cc and specific heat capacities of 4,184 J/kg-ºK and 1,000 J/kg-ºK for pure water and typical rock respectively. The specific heat capacity of rock was taken as an average of about 1,000 J/kg-ºK in the temperature ranges of interest (Table 12; Robertson, 1988). Volumetric heat capacities are normalized across a range of porosities (5-20%) that might be encountered for sedimentary strata. When normalizing for porosity, it is evident that the volumetric heat capacity of a UTES system lies predominantly in the rock and not the pore fluid. The fluid is mostly the medium by which heat can be physically conveyed from a surface-based application, down a well, and to- or from-the rock mass.

Often, preconceived biases conveyed by the term "geothermal" lead people to assume that a technology like UTES is only focused on storing something "hot." It is noteworthy that UTES technologies can store "cold," or the absence of heat, just as easily as they can store "hot," or the presence of heat. For an RTES system, the limitation is a lower temperature bound defined by the freezing point of water, below which the RTES functionally becomes unworkable as there is no longer a practical means to physically move heat through the system. A BTES system, however, can be cooled to temperatures below the freezing point of water because antifreeze can be added to the BTES closed loop.

2.1 UTES as a Grid-Scale Battery

Few, if any renewable energy technologies can match the energy storage capacities or durations of UTES. Figure 5 shows a graph of typical energy storage methods and their respective storage capacities and durations (after Buursink et al., 2023). UTES is capable of storing GWh to TWh-scale capacities of energy and doing so for durations of weeks or even years. In fact, a UTES can be designed to provide 4 to 12-hour diurnal energy storage as well, making it a versatile LDES technology with attributes that can mimic a large, grid-scale battery.

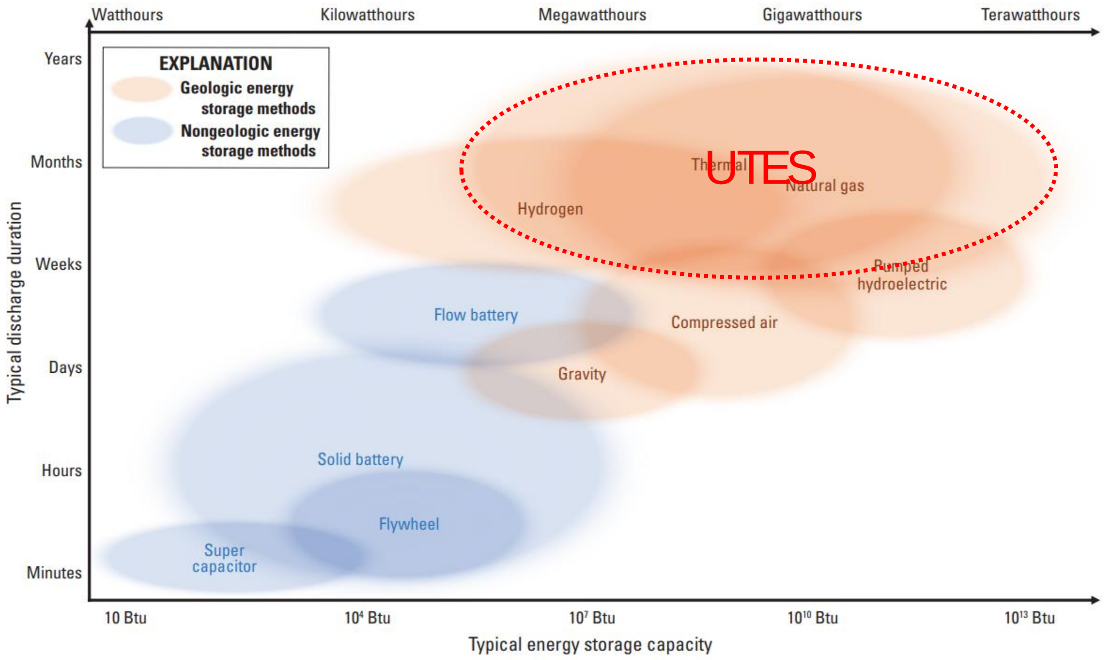


Figure 5: A graph of typical energy storage capacity compared to typical discharge duration for various geologic and nongeological energy storage methods (modified after Buursink et al., 2023).

A UTES system is an adaptable geothermal energy technology that can be used both for direct thermal energy storage and for conversion of thermal energy to electrical power. Although the UTES system itself stores energy in the form of thermal energy, depending on the system design and configuration, a UTES system can be charged and discharged either directly as thermal energy or indirectly as electrical energy by using an energy conversion device (Figure 6).

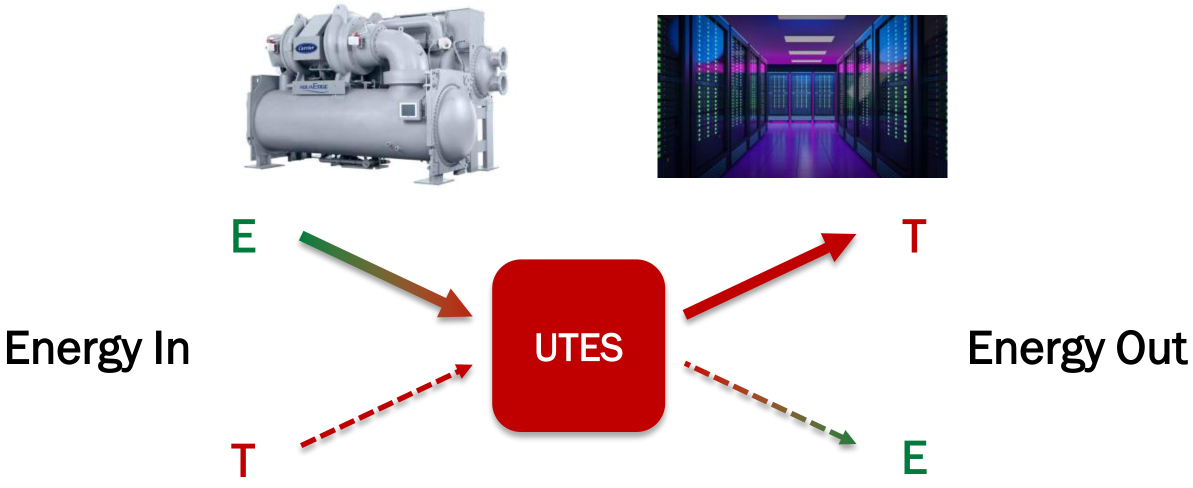


Figure 6: Conceptualization of a UTES, modified after McLaughlin et al. (2023). The energy pathway considered in this study is indicated by the thick, solid arrows (electrical energy converted to thermal energy with a heat pump | UTES storage | discharged as direct thermal energy out to cool a data center).

Figure 6 conceptualizes that a UTES system can be configured in the four main ways listed below, where (d) indicates *direct* thermal energy input and output and (c) indicates *conversion* of thermal energy to electrical energy (or vice versa) using a heat pump or a heat engine:

1. Electrical energy in (c) | UTES | Thermal energy out (d)
2. Electrical energy in (c) | UTES | Electrical energy out (c)
3. Thermal energy in (d) | UTES | Thermal energy out (d)
4. Thermal energy in (d) | UTES | Electrical energy out (c)

The round-trip efficiency of a UTES system configuration will be dependent on the number of thermal to electrical conversions, where each one implies a loss of thermodynamic efficiency. As indicated by the thick, solid arrows in Figure 6, the focus of the DOE's Cold UTES project is on the use of off-peak, low-cost grid electricity to remove thermal energy using a heat pump, such as a water chiller (shown) or a dry cooler (not shown) in order to pre-cool the subsurface. This energy is stored in a UTES system (either a BTES system or RTES system) as a thermal reservoir of cooled rock. By rejecting heat to this reservoir, either diurnally or seasonally, the stored cold thermal energy can then be dispatched directly to provide cooling to a data center and to offset the consumption of what would otherwise be expensive peak-cooling demand. For this reason, the total annual energy consumption of a Cold UTES system may in fact not be "minimized," however the real value of this implementation is its ability to leverage time-of-use pricing from a grid or facility cost perspective.

2.2 UTES as a Virtual Power Plant

The DOE's 2023 *Pathways to Commercial Liftoff: Virtual Power Plants*, and its recent 2025 update describe the ways in which energy resources can shape demand on the grid (Downing et al., 2023; Razdan et al., 2025). These reports refer to the ability of an energy resource to "shift," "shape," "shed," and "shimmy" grid loads, and the importance of these load modifying capabilities are described. When a Cold UTES system, or any geothermal LDES system is deployed at scale, it can have the effect of shifting, shaping, shedding, and shimmying grid loads (Figure 7).

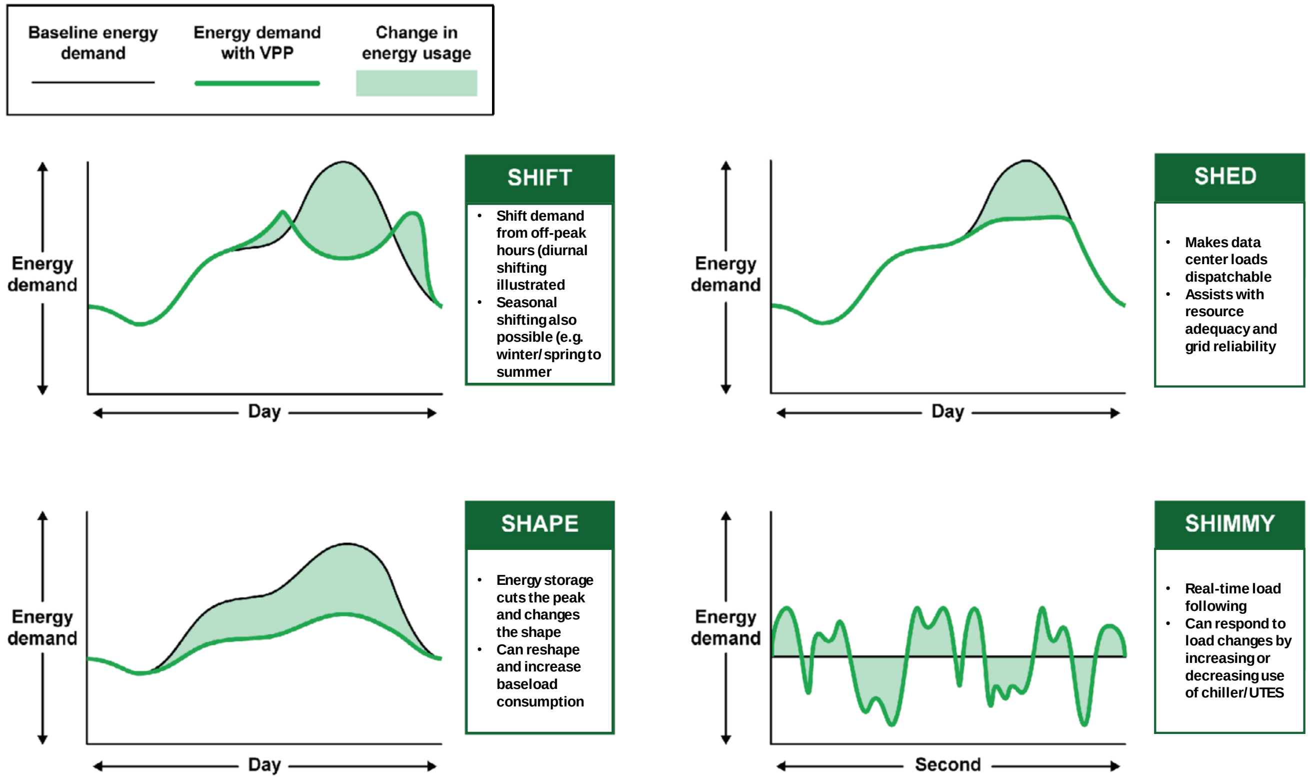


Figure 7: Ways in which a Cold UTES system can affect demand on the grid. Modified after Downing et al. (2023).

A Cold UTES system at a data center, can shift demand from peak to off-peak hours – either diurnally or seasonally. It also has the effect of cutting the peak and reshaping the load curve to broaden and flatten it, resulting in a variety of positive grid benefits and value, especially in terms of reduced resource adequacy requirements and reduced planning reserve margin. A Cold UTES system at a data center adds flexibility and dispatchability to the operation. In the event of a grid emergency, such as a generator tripping offline, a data center with a Cold UTES has the optionality to quickly switch its cooling configuration and shed load. This allows data center operators in certain grids to participate in demand response programs and receive compensation for doing so. Finally, a Cold UTES system at a data center can potentially shimmy grid loads, facilitating real-time control of load fluctuations and assisting with frequency regulation.

The DOE's 2023 *Liftoff* report on virtual power plants (VPPs) explicitly defines VPPs as:

> *…aggregations of distributed energy resources (DERs) such as rooftop solar with behind the meter batteries, electric vehicles (EVs) and electric water heaters, smart buildings and their controls, and flexible commercial and industrial loads that can balance electricity demand and supply, as well as provide utility-scale and utility-grade grid services.*

For the reasons elaborated above, when applied to data centers or to any industrial cooling (or heating) load, a UTES system can function as a VPP.

2.3 The Cold UTES Value Proposition to Data Centers

A central hypothesis of the DOE's research on Cold UTES for data centers, is that it can be cost-effectively integrated with standard data center cooling approaches, but with an optimized time-of-use, such that both the system integration (grid) benefits and benefits to the data center facility are maximized as compared to conventional data center cooling approaches. When integrated within a data center heat rejection loop, a Cold UTES system should be a commercially attractive and technically viable solution for large data center cooling loads. A Cold UTES system is expected to be commercially attractive by having a financeable return on investment and positive net-present value as compared to a conventional, large air-cooled system. A Cold UTES system will have a higher capital cost, owing to the subsurface development requirements, i.e. drilling wells to support a Reservoir Thermal Energy Storage (RTES) or Borehole Thermal Energy Storage (BTES) system. However, it is expected that this additional capital cost will be offset "behind the meter" and at the facility level by:

1. Reduced cooling system capital costs resulting from reduced peak hour loads and reduced or eliminated derating of cooling equipment capacity in hot weather conditions
2. Reduced electric utility power costs for both demand ($/MW-month) and energy ($/MWh)
3. The value derived from faster computing clock speeds
4. The value of avoided water consumption resulting from the use of a Cold UTES

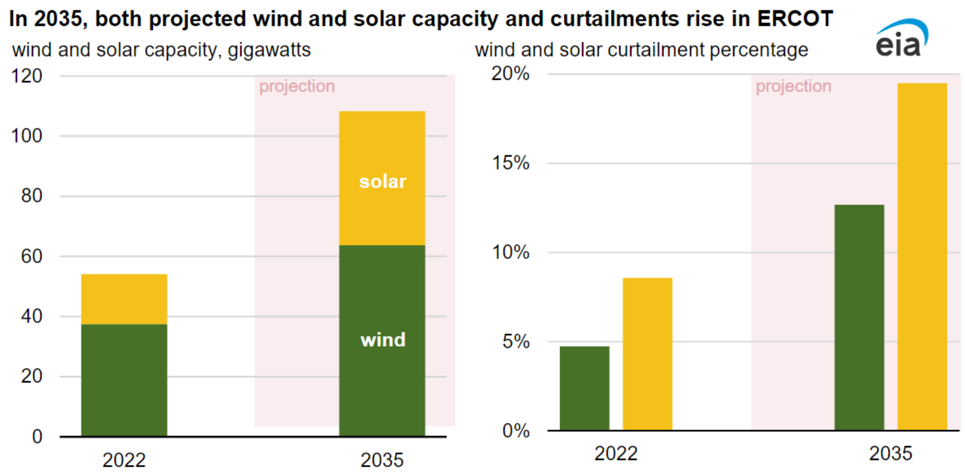
On the grid, or in "front of meter," the additional capital cost of drilling and construction of the UTES can be offset by grid level capital and operating cost reductions such as:

1. The avoided capital and operational costs of new generation (e.g. new peaking power plants)
2. The avoided cost of new grid energy storage batteries (both short and long duration)
3. The value of avoided transmission construction resulting from optimized utilization of existing transmission assets
4. Reduction of curtailed power, (monetization of low or negatively priced power) by using that power to cool the subsurface and charge the Cold UTES system

Utilities and grid system operators may ultimately choose to capture this value by offering incentives that could help developers to finance the deployment of the Cold UTES technology. Ultimately the DOE anticipates that a Cold UTES system can provide an LDES and industrial-scale cooling solution that is commercially attractive and technically viable for data centers. GTO's R&D initiative will evaluate the potential of these systems to provide significant savings and value to data center operators, utilities, and grid system operators.

3. A DATA CENTER COLD UTES EXAMPLE

To illustrate how a Cold UTES can be deployed as an economic cooling solution for a data center, it is useful to consider a hypothetical installation within ERCOT. ERCOT provides a good example of the value opportunity for a Cold UTES system because there are significant amounts of curtailed renewable energy (wind and solar) on the ERCOT system, which is projected to increase over time. At around 37GW of installed capacity, ERCOT has the most deployed wind energy in the U.S. As shown in Figure 8, in 2022 there were periods throughout the year where as much as 2GW of installed wind capacity were curtailed. This is projected to grow to as much as 7 or 8GW of wind energy curtailments by 2035.



Data source: U.S. Energy Information Administration, UPLAN model simulation of ERCOT power market, and Potomac Economics, *State of the Market Report for the ERCOT Electricity Market*, May 2023

Figure 8: ERCOT wind and solar capacity and curtailments (Warady et al., 2023).

Ambient weather conditions are an important consideration when evaluating the arbitrage opportunity of a Cold UTES system. As an example, Figure 9 shows average high and low temperatures for major cities across Texas (Weatherspark, 2024). As the graph shows, from roughly October to May, ambient temperatures are frequently cool, presenting an opportunity to charge a UTES with "free" cooling (blue ellipse) and/or with a chiller (green ellipse) which would operate at a high COPs in those weather conditions. When this cold energy is seasonally stored, it can be made available to offset peak cooling demands (red ellipse) during the subsequent hot summer months from June to September, when a chiller would otherwise be running on expensive peak power and at a low COP to meet the cooling demands. The shift in generation technology would be substituting otherwise curtailed wind power (stranded capital assets with no emissions) for expensive simple cycle gas turbine power (a resource with regulated emissions). In fact, as seen in Figure 9, even during the hot summer months, the potential exists to deploy Cold UTES systems to leverage off-peak charging during cooler parts of the day (green dashed ellipse) to achieve diurnal storage.

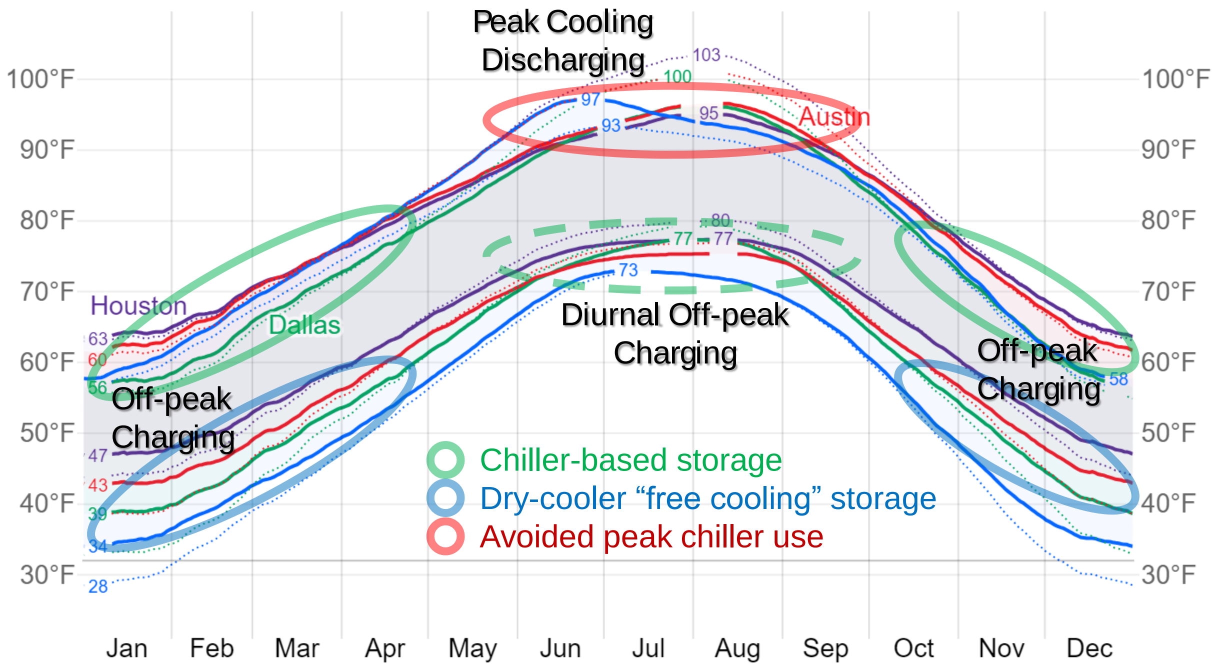


Figure 9: Average high and low temperature lines for Houston (purple), Dallas (green), Austin (red), and El Paso (blue) over the period from 1980 to 2016 (Weatherspark, 2024). Solid green and blue ellipses indicate many hours available to charge a UTES using some combination of "free" cooling (blue ellipse) and/or a water chiller (green ellipse) operating at a high COP. Cold energy can be stored and seasonally dispatched to avoid running a chiller at the summer peak (red ellipse). The dashed green ellipse indicates additional opportunities for diurnal energy storage, even during summer months.

Most of the wind energy curtailments in ERCOT occur in the spring when ambient temperatures are relatively cool. And yet, only a few months later during summer, ERCOT net-loads routinely break records (Grid Status, 2024). This represents a significant amount of stranded, off-peak, low or negatively-priced energy that could be monetized, given the appropriate technology implementation – such as a Cold UTES system. The reason why this does not currently occur is because capturing the pricing arbitrage requires technologies with seasonal storage durations and very large capacities. To date, a suitable LDES technology has not been successfully commercialized to capture this value.

3.1 Configuration of a Cold UTES at a Data Center

What is described above, and increasingly replicated across the U.S. with some variation, is an energy arbitrage opportunity where otherwise off-peak, low-cost, or curtailed grid electricity can be converted to cold thermal energy and stored in a BTES or RTES system by pre-cooling the rock mass. The BTES or RTES system can later be discharged by accepting the partial or full heat rejection load of the data center during peak cooling demands. The following section describes an example configuration at a data center that could capture this arbitrage value.

Consider a hypothetical data center with a 100MW cooling load operating within the ERCOT grid. This data center will have some implementation of IT equipment cooling which could consist of any number of conventional or emerging chip cooling methods, such as direct air-cooling, liquid immersion cooling, two-phase immersion cooling, etc. The specific means by which the IT equipment is cooled is of secondary importance, because the heat produced by the IT equipment is ultimately transferred to and rejected from the facility through a water loop. Shehabi et al. (2024) describes a variety of major cooling systems that can achieve the overall facility heat rejection. For the example described herein, we consider a dry cooler without adiabatic assist (no evaporative cooling) and with the implementation of a water-cooled chiller and a Cold UTES that can leverage time-of-use and arbitrage the grid energy.

Dry coolers reject heat to the ambient air and enable "free" cooling during favorable weather conditions. When the outdoor dry bulb temperatures are low, dry coolers efficiently reject heat using ambient air alone. When the outdoor dry bulb temperatures are close to, or higher than the water temperatures exiting the data center, a water-cooled chiller provides the additional needed cooling. It achieves this by raising the heat rejection temperatures above that of ambient outdoor temperatures, so that dry cooling can still operate without any need for water-consumptive evaporative cooling (e.g. a cooling tower). Figure 10 shows this basic configuration as described.

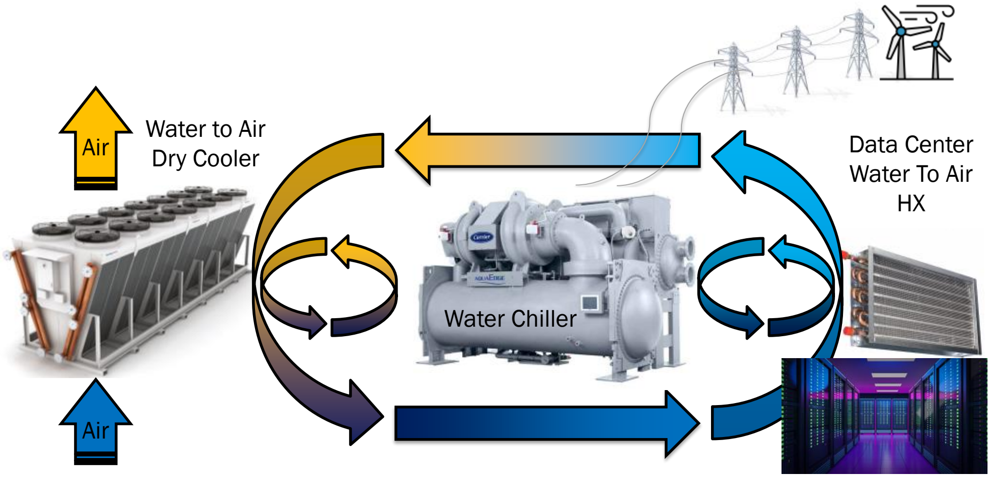


Figure 10: A hypothetical 100MW data center conventional cooling configuration consisting of a water chiller working in tandem with a dry cooler. This configuration does not include a UTES, leaving the data center and grid operators exposed to peak power demands and pricing.

The system as shown in Figure 10 exposes both the data center and the grid operators to the technical and economic challenges of peak power demands and pricing. Since the data center cooling equipment is sized for peak demand and derated in hot weather conditions, the data center developer must install oversized equipment at considerable capital cost to address the extreme design temperatures. This equipment may end up operating at its rated condition for only a few hours of the year, and thus becomes a large and expensive stranded asset. If instead a Cold UTES is added to the system as shown in Figure 11, low priced or otherwise curtailed grid energy can be consumed during off-peak periods (spring months) to run the water chiller at a high COP (~5) and charge the UTES system.

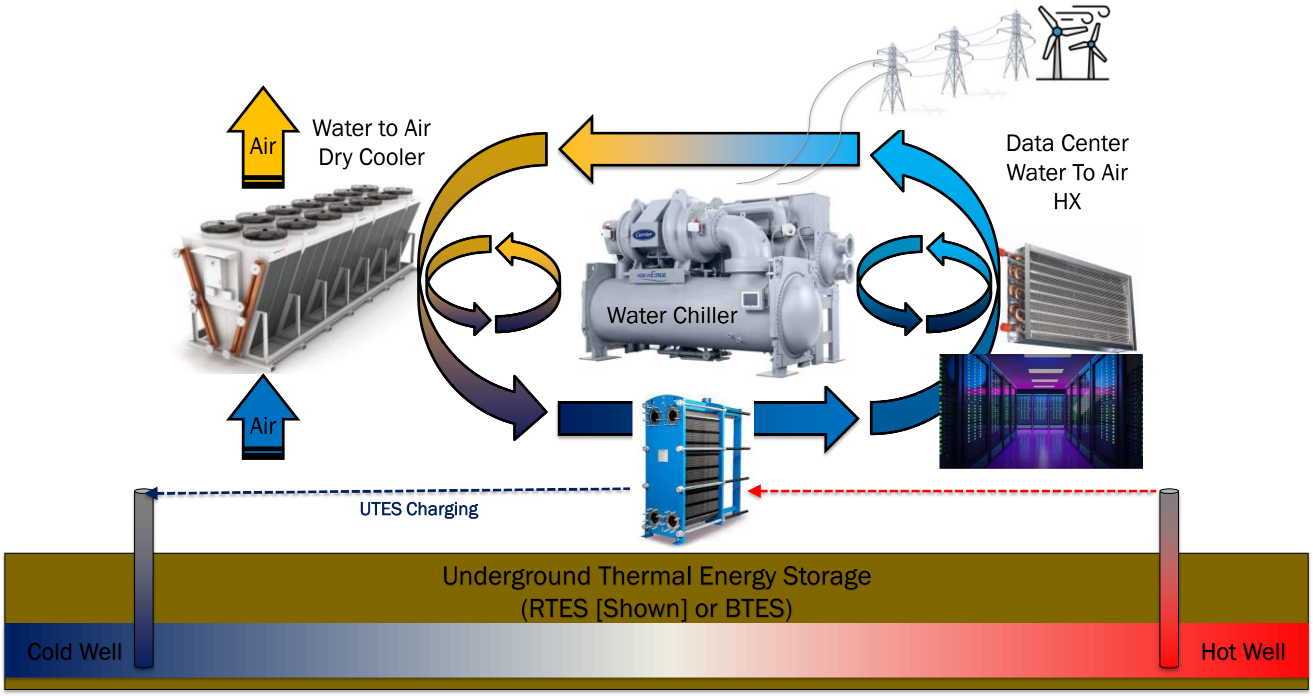


Figure 11: A hypothetical 100MW data center cooling configuration consisting of a water chiller working in tandem with a dry cooler and a UTES. The system shown is in CHARGING mode during spring when grid power is off-peak, ambient temperatures are cool, "free" cooling is available, and the chiller can operate at high COP (~5) to provide all data center cooling loads in addition to charging the UTES system.

With the addition of the dry cooler, this combination can super-charge a Cold UTES system, effectively pre-cooling the rock with GWh to TWh-scale cold energy capacity and storing that capacity for months until it is needed to offset expensive grid power during the summer peak. Figure 12 shows the system operating in the summer at peak, where the cold energy which was stored from the prior season (Figure 11) can now be discharged to meet 100% of the data center cooling load. In this configuration no grid energy is required to run either the water chiller or even the dry cooler. Not only is running the chiller at peak power prices made unnecessary by the implementation of a Cold UTES, but that chiller would have otherwise had to operate at a relatively low COP (~2-3) given the hot ambient air temperatures during the ERCOT summer peak.

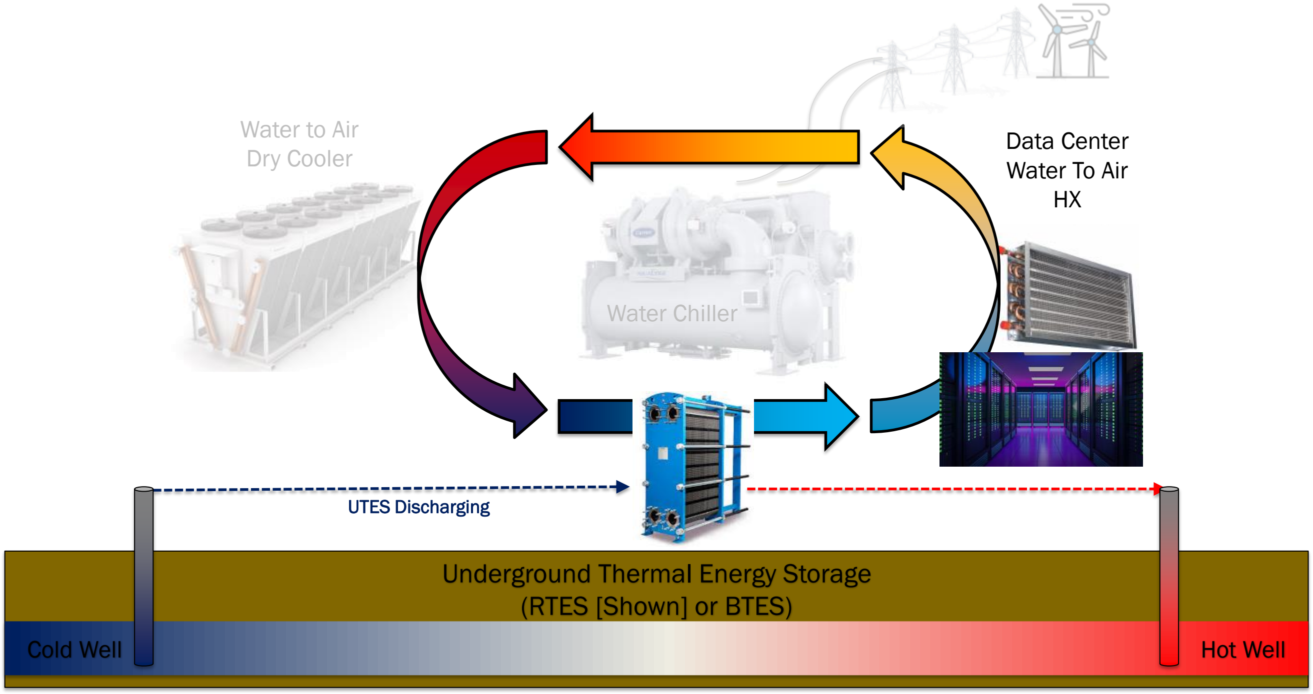


Figure 12: A hypothetical 100MW data center cooling configuration consisting of a water chiller, dry cooler, and a UTES. The system shown is in DISCHARGING mode during the summer where the Cold UTES is providing 100% of the data center cooling load and no chiller or dry cooler is required. Expensive peak grid power for cooling is largely avoided at the peak when ambient temperatures are hot.

When a chiller is operating in hot weather, the chiller capacity is derated, and therefore additional equipment must be purchased to meet the peak load. For a Cold UTES system, since the cooling system is not as significantly derated with respect to the maximum summer air temperature, less cooling equipment capacity is required, presenting a capital savings opportunity to the data center.

3.2 An Example of the Potential Economic Value

To further illustrate how this operational configuration and strategy can produce a significant financial arbitrage opportunity, it is instructive to consider the ERCOT real-time pricing and grid fuel mix during a typical spring day and during the summer peak. April 18, 2024, was chosen as the typical spring day, and as can be seen on Figure 13, the grid fuel mix was predominantly natural gas (most likely combined cycle), wind, and solar, with some baseload coal and nuclear. April 18, 2024, was an average and typically windy spring day, and while a detailed nodal analysis was not performed, it is likely that some nodes on the ERCOT grid were close to full capacity and wind energy was close to zero cost or possibly even curtailed in some locations. On average the real-time locational marginal pricing on ERCOT was around $10/MWh. It is noteworthy that there are many days like this during the spring months on ERCOT.

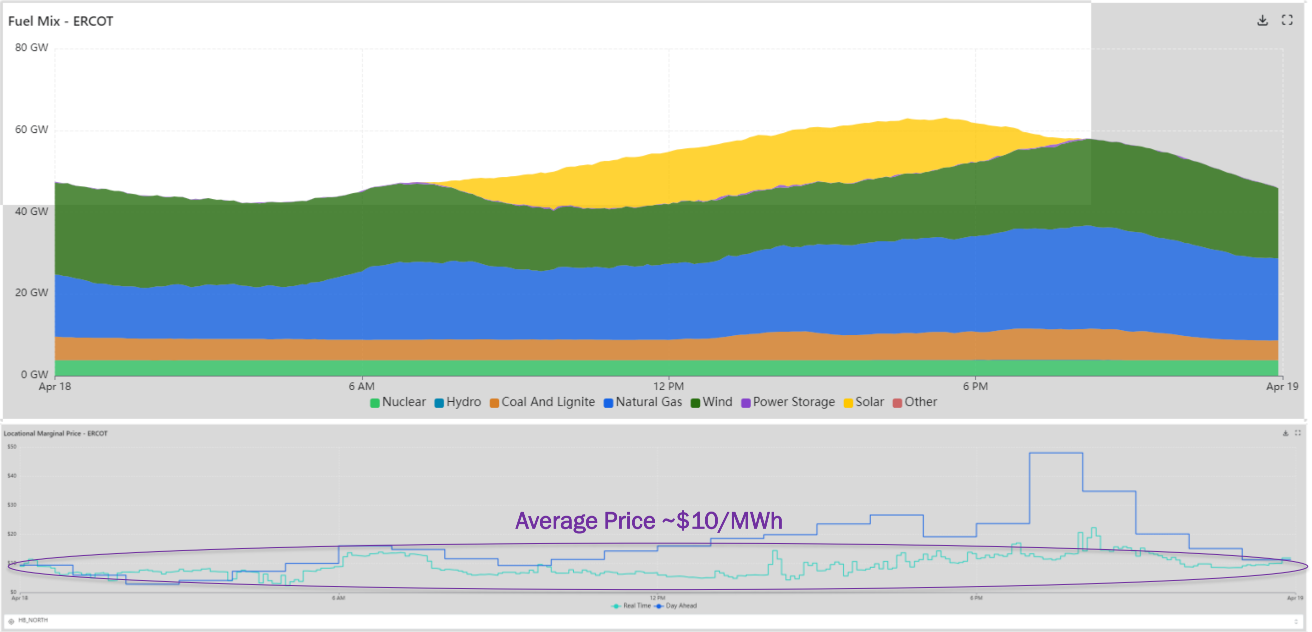


Figure 13: ERCOT fuel mix and real-time pricing on April 18, 2024, a typical spring day (Grid Status, 2024)

In contrast, the August 20, 2024, summer peak was an exceptional day on ERCOT from many perspectives (Grid Status, 2024). August 20th broke records as the day with the highest net-load in ERCOT history (70.9GW at 7:50PM). A review of weather data demonstrates ambient temperatures across much of Texas (Houston, Dallas, Abeline, Austin, and El Paso) reaching above 105ºF (40.5ºC) and at the time of peak net-load staying above 95ºF (38ºC) with high humidity and relatively low wind speeds throughout the day (Weatherspark, 2024). As shown in Figure 14, the ERCOT fuel mix reflects these weather conditions, with significant amounts of natural gas generation, much of which was likely from single-cycle gas turbines, especially approaching the net-load peak from around 6pm and into the evening. Despite much of the day seeing ~$15/MWh real-time locational marginal pricing, several hours at peak net-load hit the $5,000/MWh cap on ERCOT.

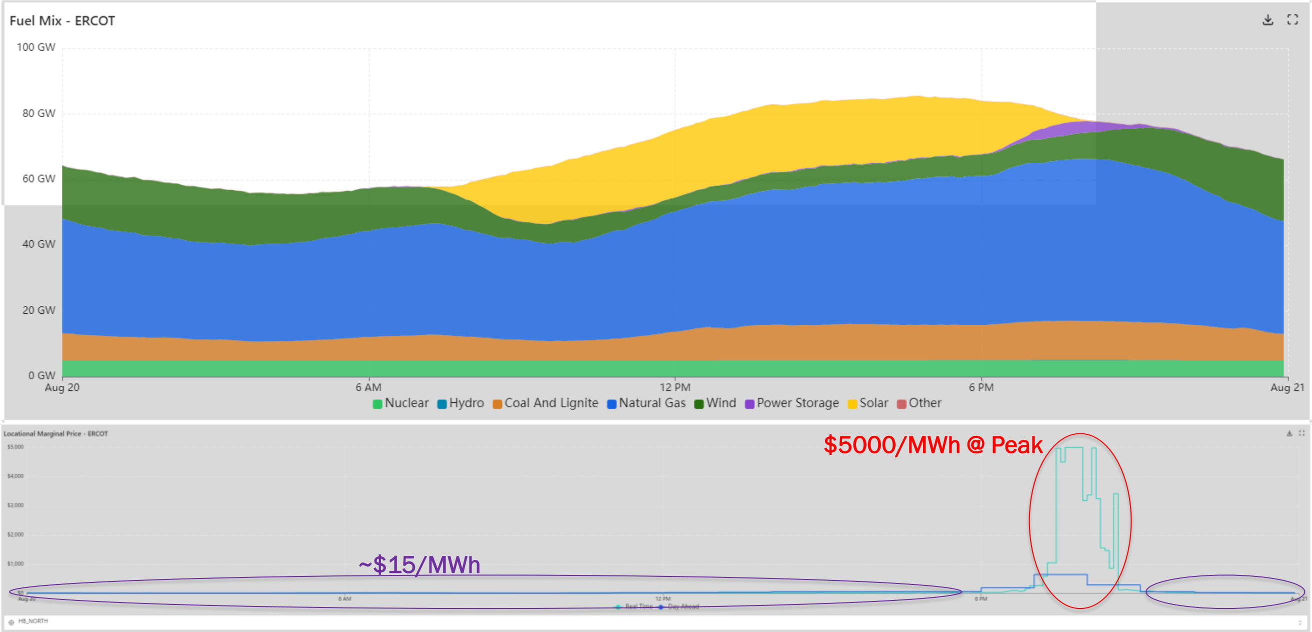


Figure 14: ERCOT fuel mix and real-time pricing on August 20, 2024, at the summer peak (Grid Status, 2024)

Cooling a large data center installation in these conditions and without a Cold UTES system would have exposed the facility operator to significant pricing extremes in the summer and significant opportunity costs during the spring months. Conversely, a data center with a Cold UTES system, could have charged the UTES system with low-cost (~$10/MWh) grid energy with high proportions renewable energy in the fuel mix. This seasonally stored energy could have been made available to offset the expensive ($5,000/MWh) summer peak hours. This is illustrated in the calculations shown below in Table 2 for a hypothetical data center with a 100MW cooling load.

Table 2: Simplified example energy pricing arbitrage enabled by cooling a data center with a Cold UTES system.

Grid condition		Average Spring Day	Peak Summer Hour
Date		4/18/2024	8/20/2024 @7:50PM
Ambient Temperature	°C	Cool (20°C)	Hot (38°C)
Data Center Cooling Load	MW	100	100
Chiller Efficiency	COP	5	2
Grid Energy for 100MW Cooling	MW	20	50
Grid Real-Time Price	$/MWh	$ 10	$ 5,000
Hours Procured	hours	1	1
Energy Procured	MWh	20	50
Energy Cost	$	200	250,000
Energy Cost Savings	$	$	(249,800)

As seen in Table 2, charging a Cold UTES to meet one hour of a 100MW cooling load on an average spring day at a relatively high COP (~5) would require procuring 20MWh (20MW for 1 hour) at a relatively low cost of $200 (20MWh at $10/MWh off-peak power price). That same amount of energy seasonally stored in the Cold UTES would avoid a data center operators' exposure to $250,000 for only one hour at the summer peak. Those avoided costs derive from avoiding the need to buy 50MWh at a peak power price of $5,000/MWh. Because of the hot ambient temperatures during the peak, and low chiller COP (~2), 50MW of peak grid energy would be required to meet that same 100MW data center cooling load.

From this analysis it is evident that the differential seasonal COPs amplify the economic arbitrage, resulting in a total energy cost savings of $249,800 for only that one hour. Considering that the Cold UTES could in fact be charged continuously for days, weeks, or even months under the right off-peak weather and pricing conditions, the scale of the energy arbitrage opportunity could be on the order of GWh to TWh and many millions of dollars in peak energy cost savings. For a prospective data center developer, these potential savings represent a financeable return on investment for building the Cold UTES. None of these benefits consider the significant "front of meter" grid values that have been previously mentioned, such as the value of avoided generation and transmission that likely result from deploying Cold UTES systems for cooling at scale.

5. U.S. DEPARTMENT OF ENERGY'S R&D INITIATIVE

The Geothermal Technologies Office (GTO) at the DOE is investing significant resources to research and develop Cold UTES technologies as applied to data center cooling loads. GTO have stood-up a joint effort to be led by the National Renewable Energy Laboratory in partnership with Lawrence Berkeley National Laboratory, Princeton University, and the University of Chicago to evaluate how Cold UTES can be a commercially attractive and technically viable option for meeting large data center cooling loads (Bersine, 2025). The project officially kicked-off in January, 2025, and is on an aggressive timeline (1 to 1.5 year period of performance) to produce results that will inform data center developers and operators, utilities, grid and system operators, and policymakers on the transformative opportunities Cold UTES presents to data center markets. The project is not currently a national-scale analysis, but is intended to provide insights that are applicable nationally. To quickly execute the project, conduct analyses, and demonstrate value, the project will instead focus on two key data center markets. The project work scope comprises five main tasks that are briefly described below:

Task 1 - System analysis of Cold UTES cooling technology for data centers. This task focuses on system model development, sizing, optimization, and sensitivity analysis for a data center cooling system with and without Cold UTES. The models will consider system configurations, performance, and costs at the data center installation scale.

Task 2 - Data center load projections. This task will include a comprehensive review of relevant data center load forecasts, estimates of hourly load shapes and a determination of how these loads will change with the implementation of a Cold UTES system. These load projections will be essential prerequisites to the grid impacts modeling in Task 3.

Task 3 - Grid Impacts Modeling. Using the analysis compiled by the teams from Tasks 1 and 2, load projection changes resulting from a Cold UTES will be input to grid capacity expansion and production costs models to assess a variety of impacts including, but not limited to, system costs, peak and net peak load reductions, changes to grid fuel mix, impacts on resource adequacy, and impacts to transmission.

Task 4 - Technical Advisory Groups. Three Technical Advisory Groups (TAGs) will be established: 1) utility engagement, 2) subsurface technologies, and 3) data center developers. Task leads will meet regularly with the TAGs to share project progress updates and to obtain feedback and guidance on project activities.

Task 5 - Reporting & Communications. No matter how compelling the results, unless they are disseminated, the technology adoption of Cold UTES will stagnate. This task aims to leverage the project results, which are expected to quantify and demonstrate the grid and data center facility value of Cold UTES, and disseminate this information amongst key stakeholders including utilities, data center developers, grid operators, policymakers, and regulators. This task includes the development of a Final Technical Report on the project findings.

6. CONCLUSIONS

The DOE conservatively estimates total electricity system peak demand to grow by approximately 200GW. Within this environment, data center energy loads are growing rapidly and contributing to a new era of rising U.S. electricity demands. This presents significant challenges to meeting and managing the anticipated loads, and especially the peak loads of projected data center deployments. As much as 40% of a data center's total energy consumption is for the cooling system, presenting an important opportunity for new energy efficiency and LDES technologies to play a key role in managing data center loads.

Cold UTES implemented as part of a data center cooling system gives the ability to shape, shed, shift, and shimmy grid energy loads. It meets the definition of a VPP and can provide significant grid value and benefits in "front of the meter." A Cold UTES system is a LDES system capable of delivering significant cost savings and value not only to the grid (utilities and system operators) but also to data center developers and operators. It is a unique and scalable method to avoid exposure to peak pricing over seasonal or even diurnal timescales. These savings can be integrated within the business case of a development project, making the return on investment financeable and providing a positive net present value.

The DOE and its Geothermal Technologies Office (GTO) anticipates that Cold UTES systems could play a crucial role in addressing the challenges associated with the explosive growth in data center energy demands, and it is therefore a key technology solution that deserves attention and investment in research and development. For these reasons, GTO is investing significant resources to research and develop Cold UTES technologies as applied to data center cooling loads. GTO have stood-up a joint effort to be led by the National Renewable Energy Laboratory in partnership with Lawrence Berkeley National Laboratory, Princeton University, and the University of Chicago to evaluate how Cold UTES can be a commercially attractive and technically viable option for meeting large data center cooling loads. Results of the project will be made available in a forthcoming report, anticipated in FY26.

REFERENCES

Akindipe, Dayo, McTigue, Joshua, Dobson, Patrick, Atkinson, Trevor, Witter, Erik, Kumar, Ram, Sonnenthal, Eric, Umbro, Mike, Lederhos, Jim, Adams, Derek, & Zhu, Guangdong (2024). Techno-Economic Analysis and Market Potential of Geological Thermal Energy Storage (GeoTES) Charged With Solar Thermal and Heat Pumps. https://doi.org/10.2172/2474842

Aljbour, J., Wilson, T., & Patel, P. (2024). Powering Intelligence: Analyzing Artificial Intelligence and Data Center Energy Consumption. EPRI White Paper no. 3002028905.

Bersine, A., (2025). *Reducing Data Center Peak Cooling Demand and Energy Costs With Underground Thermal Energy Storage*. National Renewable Energy Laboratory, News Blog., January 17, 2025.

Buursink, M.L., Anderson, S.T., Brennan, S.T., Burns, E.R., Freeman, P.A., Gallotti, J.S., Lohr, C.D., Merrill, M. D., Morrissey, E.A., Plampin, M.R., and Warwick, P.D., 2023, Geologic energy storage: U.S. Geological Survey Fact Sheet 2022–3082, 4 p., https://doi.org/10.3133/fs20223082.

DOE (2025). Clean Energy Resources to Meet Data Center Electricity Demand., US Department of Energy website blog accessed Jan 29, 2025.

Downing, J., Johnson, N., McNicholas, M., Nemtzow, D., Oueid, R., Paladino, J., & Wolfe, E. B. (2023). Pathways to commercial liftoff: Virtual power plants. US Department of Energy Report.

Ginosar, Daniel M., Atkinson, Trevor A., Adhikari, Birendra, Toman, Jakub N. M. N., & Podgorney, John (2023). Reservoir Thermal Energy Storage Benchmarking (Rev. 3). https://doi.org/10.2172/1997222

Grid Status (2024). *Breaking Down a Record-Setting Day in ERCOT: Market highlights and explanations from a day of records in ERCOT.* August 21, 2024Lovins, Amory B. "The negawatt revolution." Across the board 27.9 (1990): 18-23.

McLaughlin, L., Gluesenkamp, K., Balliet, H., Ma, Z. Technology Strategy Assessment: Thermal Energy Storage (TES) (2023). 3rd Annual Energy Storage Grand Challenge Summit. U.S. Department of Energy

McLing, Travis L., Dobson, Patrick, Jin, Wencheng, Spycher, Nic, Doughty, Christine, Neupane, Ghanashyam Hari, Smith, Robert W., & Atkinson, Trevor A. (2022). Dynamic Earth Energy Storage: Terawatt-year, Grid-scale Energy Storage Using Planet Earth as a Thermal Battery (GeoTES): Phase I Project (Final Report). https://doi.org/10.2172/1885826

Razdan, S., Downing, J., White, L. (2025). Pathways to Commercial Liftoff: Virtual Power Plants 2025 Update. US Department of Energy Report

Robertson, E. C. (1988). Thermal properties of rocks (No. 88-441). US Geological Survey.

Shehabi, A., Smith, S.J., Hubbard, A., Newkirk, A., Lei, N., Siddik, M.A.B., Holecek, B., Koomey, J., Masanet, E., Sartor, D. (2024). 2024 United States Data Center Energy Usage Report. Lawrence Berkeley National Laboratory, Berkeley, California. LBNL-2001637.

Warady, D., Hodge, T., Aramayo, L. (2023). *As Texas wind and solar capacity increase, energy curtailments are also likely to rise.*, U.S. Energy Information Administration., Today in Energy: July 13, 2023.

Weatherspark (2024). *Climate and Average Weather Year Round in Texas*. Website accessed December 9, 2024.