WHAT'S NEXT FOR DATA CENTERS

As AI and other technologies surge in popularity, so has the demand for data centers to power global digital transformation. However, a greater demand for data centers also introduces demand for greater sustainability measures and the need to educate communities on the positive change data centers bring. Here is what you can expect from the data center industry in 2024 and beyond.

INCREASED ADOPTION OF EFFICIENT COOLING SYSTEMS

The advancement in AI technologies comes with a potential consequence for data centers – increased heat generation and density. As climate concerns increase, there will be continued pressure on data center companies to utilize more efficient cooling methods to reduce their carbon footprint.

The sustainability challenge for providers is to maximize operational efficiencies and reduce overall energy use, all while maintaining cooling system resiliency and providing a flexible environment for a wide range of rack densities and dynamic load profiles from AI driven High Performance Computer (HPC).

In 2024 and beyond, we will see a steady, if not exponential rise in liquid-to-chip server deployments as more manufacturer's produce these options and end-user adoption increases.

Liquid cooling is a far more efficient means for heat removal than air, so in theory the more it is deployed, the lower a data centers PUE (ratio of IT energy to overall facility energy) will be as less energy is required for heat removal than would be used with air alone. The higher the percentage of direct liquid cooling, the more efficiently a data center can operate and obtain a significant reduction in PUE. **Overall, this shift from liquid to air is beneficial.**

Not only do these servers utilize a high degree of efficient liquid cooling, but they produce more computing per square foot than any systems that have come before. The more the kW/rack densities increase on average, the less floor space and ultimately building (i.e., "carbon") is required to accommodate the loads.

To be able to accommodate the rising demand, facilities will require the necessary baseline cooling infrastructure and a higher level of operational finesse than we have seen before.

Existing facilities may require a significant Capex investment if not already equipped with a fluid-based cooling system. Having a glycol or chilled water base cooling system is a necessity, as refrigerant based systems alone cannot accommodate this requirement. An air-cooled, closed-loop system is typically the baseline choice regarding sustainability concerns, as it has a negligible demand on water once in operation, and provides the most flexibility in supply water temperatures to accommodate servers requiring low inlet fluid temperatures. (Evaporative cooling strategies may use less energy on average, and still be a strong consideration for dryer-cooler climates, but require special consideration with respect to the continued water demand they will require as an offset to yearly energy savings).

Strategically sized and placed Cooling Distribution Units (CDUs) provide the flexibility to distribute liquid direct to in-rack manifolds for seamless distribution to liquid-cooled servers. While many providers have migrated away from raised floors and gone with overhead only distribution, raised floors may come back in style for new builds to mitigate the inherent risks of overhead liquid and cross-congestion with dense power distribution and other overhead systems. Manufacturers like Schneider Electric and Vertiv, are launching CDU product lines. CDU sizes, currently available up to 1 MW, may increase much like available busway ampacities did, to accommodate the larger deployments of high-density clusters.

While a high percentage of heat from liquid-cooled servers can be removed via water, they still require a percentage of conditioned air, and supporting network, storage, and other potentially high-density rack deployments still rely solely on directed air. With the use of a raised floor and proper containment, mid-to-high-density air-cooled IT deployments can often be managed, but the lack of precision makes maintaining high efficiencies in conjunction with system resiliency challenging due to the fan speed and volume requirements to cover large air-throw distances. Close-coupled cooling such as rear-door heat exchangers and in-row cooling provide a way to leverage a base chilled water system as a supplemental solution for high-density, non-liquid cooled installations. **Supplemental refrigerant based closed-coupled cooling is a viable alternative for providers that have already maxed out their cooling plants.**

With the dynamic nature of all High-Performance Compute (HPC), a robust monitoring system with precision control and automation will be a necessity. **A data center populated with traditional server technologies is more likely to "hum along" at a more constant and predictable rate, allowing operators to monitor trends, cycle equipment, and manually optimize energy usage of cooling equipment while maintaining necessary levels of redundancy.** Power dense AI deployments have a more dynamic and unpredictable load profile, with server clusters that can go from idle to warp speed in the blink of an eye. Loads can shift around the geography of a data center, creating instant hot spots and exceeding the speed at which a traditional cooling system can respond to the demand, risking system level shutdowns from equipment overload conditions. Operators can find themselves playing whack-a-mole trying to re-direct air flow and adjust system performance parameters on the fly.

This creates a new challenge for energy management. **Systems need to be designed to accommodate the peak demands but are traditionally automated to adjust pump and fan speeds to maintain the target supply and return air temperatures that provide the right balance of redundancy and operational efficiency.** Without proper precision monitoring and controls, operators might have to keep systems running as in peak demand mode, sacrificing efficiency, just to ensure reliability.

There is a high learning curve for operators as they fine tune methods to deal with the demand response while continuing to preserve energy. Fine-tuned air flow management via containment and blanking panels to keep hot and cold air streams separate is even more important.

In time, the AI systems themselves will likely be the master of their own domains, controlling the environment they live in and learning as they go how to ensure uptime AND preserve energy.

DATA CENTERS AS LEARNING CENTERS

NIMBYism (as in "not in my backyard") is unavoidable when it comes to data center development. Data centers are typically depicted as loud, messy, and unattractive, and the industry has not done much to change this narrative. In 2024, data center leaders will prioritize community education around data centers.

With more education around data centers, communities will learn about the technological contributions data centers bring, as well as the possible career paths one could have in the information economy. Although the development of a new data center does not bring many jobs to the local economy in the long-term, they are still an essential part of our tech-driven world. Data center leaders must make community outreach a priority as new buildings are developed. One way to do this is by offering the data center as a learning destination for local school districts so students can take field trips to the facility and learn the essential functions of a data center. Additionally, data center developers can consult architects and designers to make sure the building fits the aesthetic and character of the local community to ensure it is not an eyesore.

The size of modern data centers has caused communities to raise concerns about the noticeability to residents. These concerns show a need for more education around how these new and highly efficient data centers are effectively replacing hundreds of thousands of smaller inefficient legacy facilities that are being decommissioned and re-purposed as companies move to the cloud. As an industry, data centers need to pay more attention to political outreach and explain how they are saving energy and reducing carbon. Data centers are a multibillion-dollar industry and can facilitate energy efficiency throughout the economy.

NUCLEAR-POWERED DATA CENTERS ON THE RISE

There is a growing appetite among hyperscale and colocation providers for green, dependable, close-proximity power generation. There is not only a shortage of steady, baseline power generation available in marquee datacenter cities like Ashburn, Dallas, and Phoenix, but also the ability to DELIVER power as existing transmission lines are maxed out and new easements and high-voltage lines can take years to plan, approve, and build. The scale of developments is also at an unprecedented level, with hyperscalers and other major colocation providers planning campus deployments in the 300 MW to 1 Giga-Watt range, taxing the power grid that airports, homeowners, and commercial businesses alike depend on. Hyperscalers already lead the way in the procurement of green power purchase agreements (PPE's) which fund additional solar and wind generation projects. It makes sense that large developers would want to secure their own sources of onsite power, or locate their facilities near steady, affordable, green power sources. Proximity to power generation cuts down on transmission line losses from power traveling over long distances, maximizing the reliability and the amount of power delivered relative to power produced.

Once data centers are built and fit-out, they provide a steady base load on power grids, a good pairing to a nuclear source which operates most effectively with a steady base load for continuous operation.

While continued growth in all forms of green energy are important to combatting climate change, generation from wind and solar by nature are more volatile, and not as well suited to serve base loads, which are still heavily supported by fossil fuels. Nuclear power stands as a green alternative to elevate steady-state power production on the grid, making us less reliant on non-renewable energy sources. Once a nuclear power plant is built, like hydropower, it produces virtually zero emissions, making it the most power dense, green energy source option available.

Were it not for the cost and risks, nuclear power might still be the defacto standard for power generation today.

Small Modular (Nuclear)Reactors (SMRs): While the affordability of SMR's remains to be determined, there is much to be optimistic about regarding their future as a staple green energy source for our power demands. Small Modular (Nuclear)Reactors (SMR's) are factory built and modular, allowing the delivery of pre-manufactured components for onsite assembly, significantly reducing project timelines and (hopefully) cost. They are much smaller in size than a legacy nuclear plant and can safely be deployed in closer proximity to population centers, with enhanced safety systems requiring no human intervention for shut down, effectively eliminating the risk for any radioactive propagation.

While technically SMR's can range from 5 to 1000 MW's, the practical size for SMR's purpose built for data centers could range from as small as 20 MW from companies like Last Energy, to upwards of 470 MW each as are being developed in the UK by Rolls Royce, though SMR's in the 250-300MW range tend to be the most talked about in the US.

Advocates of SMR's have a long way to go to overcome public perception on safe operation and proximity to population centers. At the mention of "nuclear power" most people still associate with the Chernobyl meltdown, or images of the three-eyed fish from the Simpsons sitcom.

While we likely will not see SMR's being deployed at scale for the better part of a decade, what we will see more of in 2024 and beyond, is increase in investor financing for the technology, more hyperscale and large developer ambassadors, and regulatory approvals in various geographies starting to pave the way. Micro-Nuclear manufacturers like Last Energy will continue to raise capital and look for opportunities to implement modular SMRs.

Microsoft itself has announced that nuclear power will be a key component to achieving its net zero commitments, appointing Archana ("Archie") Manoharan as director of nuclear technologies to "oversee a program to develop small-scale atomic reactors to power datacenters as an alternative to fossil fuels." This year already, rezoning was approved to pave the way for Green Energy Partner's planned one giga-watt data center campus adjacent to Surry Nuclear Power Plant which is operated by Dominion Energy. Day one power for the "Surry Green Energy Center (SGEC)" will be from legacy nuclear, with a 10–15-year plan to install up to six 250MW SMR's.

Looking forward, we can expect to see more savvy developers and hyperscalers announcing plans to partner with existing power companies for long-term SMR backed data center developments, particularly in locations believed to have favorable market conditions for nuclear advancement.

DATA CENTERS IN 2024 AND BEYOND

As the information economy continues its rapid growth, data center leaders are constantly looking for ways to evolve data centers. Of the many development considerations, sustainability, community perception and alternative power sources will dominate the industry in 2024 and beyond. Keeping these trends in mind can help data center planners make the right choices.

HED

Scott Hays, PE | *Mission Critical Market Sector Leader, Sustainable Design Leader*
Explore more market and design insights at *www.hed.design/insights*