Starline
A brand of legrand

Powering the Future:
AI's Impact on Data Center Design

INTRODUCTION

The data center landscape is undergoing a seismic shift due to the growth of artificial intelligence. Gartner notes that 70% of executives report exploring generative AI for their organization; IDC forecasts that the AI software market is expected to be worth nearly $792 billion by 2025 and is projected to grow at a rate of 18.4% through that year.

This growing adoption leads to profound changes in data center power requirements and infrastructure planning. As AI technologies advance, data centers' power demands and infrastructure requirements must evolve to accommodate increasing power densities, changing hardware configurations, and dynamic workloads. Data centers must be agile enough to scale up or down quickly to meet fluctuating demands without disrupting operations. This adaptability ensures efficient resource utilization, future-proofs the facility, and creates a scalable data center for seamless expansion to meet growing computational needs. Together, adaptability and scalability are essential cornerstones in designing data centers that can effectively harness the potential of AI while staying responsive to evolving business and technology trends.

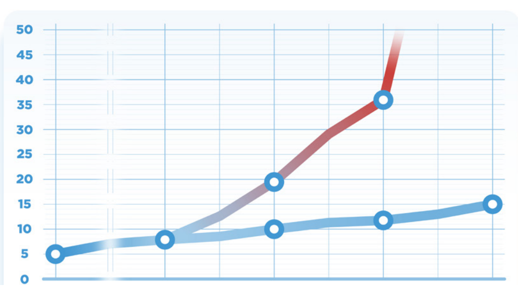
DESIGN CRITERIA

The average data center design RFP in the late 2010's required approximately 5kW per cabinet and grew to 10kW per cabinet from approximately 2012 to 2018. As an industry we quickly moved on from single phase 120-volt inputs, spent several years at single phase 208 volts, moved into three phase 208 volts, and on to other voltages, like 415 volts. The voltage was coming up, and the current, protective devices, and conductor sizes were all going down. Select applications in the traditional data center environment started ticking up, and per rack power would start to increase to 10-20% of the overall environment. Designs began to bifurcate into density and availability zones where we could leverage power distribution and heat rejection technologies based on the various power densities required in an overall space. All the while utilization was lagging design, presenting an inefficient balance on space utilization, and mechanical and electrical systems.

POWER DESIGN

The integration of AI into data centers has significantly shifted power requirements—projections show a 50% increase in the power footprint of a traditional data center by 2025. AI servers demand power levels surpassing traditional servers to support advanced AI algorithms, data processing, and GPU-intensive workloads at a rate 3-7X higher. Design criteria has transcended the above whereby entire facilities are starting with a basis of design cabinet 17kW and higher. Now the voltage and rack density in kW have increased so too does the current. Current RFPs are calling for 17.3kW to more than 50kW per cabinet. Gone are the days of 1U pizza box servers; this environment boasts multiple 8U+ systems, each approaching 10kW! With a 42U+ cabinet, five systems and 50kW every two linear feet has become a reality.

POWER DENSITY PER CABINET



In response to the rapidly increasing power needs, data centers must scale in different ways, and typical strategies like remote power panels and (RPPs), power distribution units (PDUs), and even uninterruptible power supplies (UPS) are being bypassed in lieu of higher ampacity busway direct from switchboards and switchgear, directly to the cabinet.

In many cases, stalwart design elements like dual power path and 100% UPS and generator backup, are being reimagined with unitary power paths, batteries at the rack, or limiting UPS and generator only to networking equipment and limited local storage. The solutions are more industrial and utilitarian. Power resiliency is giving way to a robust and fast network, geographic diversity and on-ramps to various cloud services. However, a fully integrated, reliable and effective real-time monitoring and alerts system designed specifically for these ampacities and busway systems can help bridge the gap where redundancy was reduced and reliability is still required.

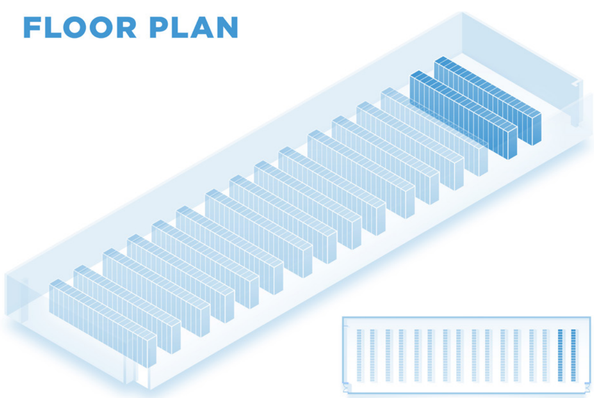
HEAT REJECTION

While power density increases, so too does the need to reject heat. 120CFM per kW has long been the minimum basis of design for air at the face of a server. Considering the legacy 1U pizza box server and 120CFM, every 8U would require 960CFM, minimum. Current AI systems require nearly 1,400CFM every 8U, a >50% increase, vertically. Moving this much air starts to challenge the typical 4′ aisle delivery approach where there is a 4′ deep cabinet with equal hot and cold aisles on either side of it.

There are myriad ways to approach this, from raised access floors with perimeter computer room air conditioners, to in-row cooling, and now immersion. For the purposes of 2′ wide and 4′ deep spacing, 4-5 AI systems will require 5,600 – 7,000CFM per cabinet. Considering a “cold aisle” where the fronts of two such systems face each other, the scenario presents 11,200 to 14,000CFM every two linear feet. Selecting one of the highest performing perforated raised access floor tiles, with a 56% unobstructed air flow, tops out at 2,340CFM and would push the cold aisle to a minimum of 8′ wide! This is double the typical basis of design, and could be as much as a 12′ cold aisle in an air only delivery.

This simplified scenario is not accounting for raised floor height, ceiling height, static pressure, pitch and roll. Like the various voltages changing the approach to ampacity, and ultimately power density, air delivery and heat rejection present numerous variables in design necessary for balanced air handling.

If the cabinet format is constant, and we hold the hot aisle constant too, a new environment, with traditional approaches, would be a minimum of 30-60% larger! Retrofitting an existing environment, depending on the floor plate, and the ability to move air, could result in an IT system population utilizing only 10% of the floor space! Because of these concerns, many are taking a hybrid approach to compensate for the additional forced air cooling. While effective in heat dissipation, the use of any form of liquid cooling (direct-to-chip, rear door heat exchangers and even submersion cooling) can have consequences for power delivey systems. Depending on how/if these systems are powered, harmonics can be introduced into IT loads, and if retrofitting a slab on grade installation, water ingress ratings should be considered.



SPATIAL CONSIDERATIONS

As the power density promotes augmentation to power delivery, the design has to account for additional space, at least vertically, potentially pushing ceilings slightly higher. To hedge that, as the aisles increase in ways described above, the additional space may be able to absorb the changes to larger electrical distribution to the cabinet level.

Buildings are expensive, and wider aisles and taller ceilings shifts initial capital to very inefficient cost centers. Certainly, few would want multiple 8-12' aisles that are effectively, purposely empty spaces.

In some cases, immersion cooling could be the answer, but introducing pumps and other powered appliances, and the weight on systems in more than a single floor deployment will yield structural augmentation. Air to air, direct evaporative systems, water to door or chip, are other possibilities, each with their own challenges and limitations. Exterior equipment can help reduce floor plates, but may introduce new plenums and pressurized spaces that start to bring it back in scope, mitigating the benefit.

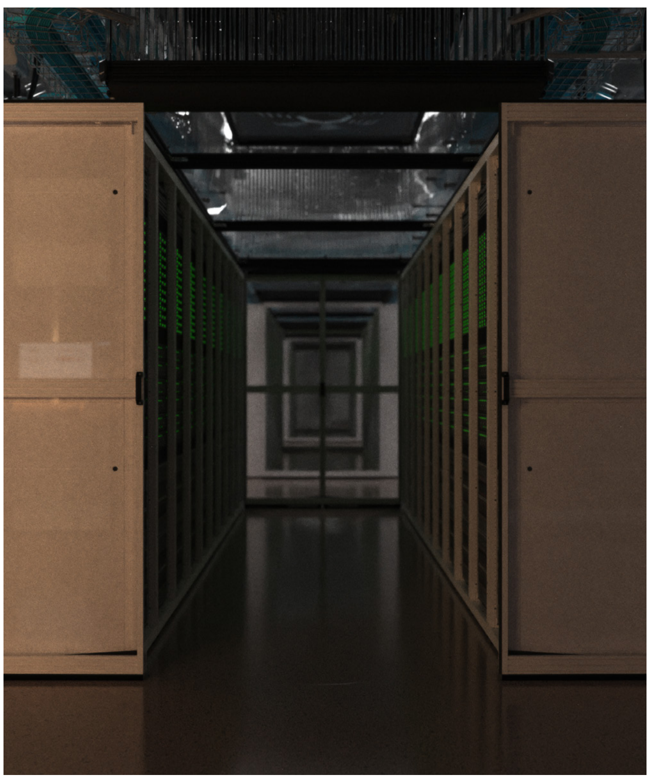
Exterior electrical equipment, in purpose built structures, offer a balance to the size of the core and shell, and promote maximizing the use of interior space for IT system use and circulation, but exasperates operation logistics, like the time for an operator to get from the white space to an exterior system that requires isolation.

SPECIAL CHALLENGES

Operationally, electrical busway is already a preferred method of data center power delivery. Industrial systems with the ability to connect every 2' is a requirement. Scheduling periodic maintenance, including visual inspections, and torquing and tightening are becoming part of the annual facility preventative maintenance plan. These have always been required and recommended, but on manufacturer and/or NETA intervals.

Cold aisles are a misnomer, and becoming hot, hot aisle are becoming hostile! Front connections on IT systems have been evolving and promoted by organizations like the Open Compute Project. A tech working in the hot aisle of an AI facility would require entrance between processes, where the systems are idle, commanding the system to operate at a lower percentage utilization, driving IT system heat rejection down temporarily, or in an extremely limited window, where the tech would be exposed to significant temperatures, like a dry sauna!

Shifting electrical distribution equipment to the outside can expand underground works and their associated calculations, building penetrations, and design coordination. It also shifts individual building design into a campus planning exercise, where the doors, paths and access between buildings and systems require careful consideration.



CONCLUSION

AI is driving the size of data center campuses from MW to GW, and driving up the scale of power distribution, air flow and physical space requirements. The aggressive growth in power density is challenging the entire industry from utility service to stand-by power topology to the cabinet itself. We are stretching and contracting physical real estate to move air, reject heat, and create pathways for the same. Design criteria are evolving with technology. We must embrace new possibilities and approaches to progress. Keeping an open mind and collaborating across fields can help us make the most of available technologies and create a better future for all.

As AI and other high-performance compute (HPC) applications grow, modern data centers face increasing demand for power distribution systems. Server Technology's PRO4X and Raritan's PX4 intelligent rack PDUs provide robust and granular revenue-grade power monitoring for high-density deployments up to 415V Inputs. Starline Track Busway's flexibility, adaptability, and IP54 rating make it perfect for these applications, especially when paired with the Starline CPMs revenue-grade metering. These products help data center facility managers adapt or outfit their facilities to meet today's and tomorrow's demands.

Discover further information on Server Technology's and Raritan's next-generation, Rack PDU Solutions. To learn more about Track Busway and CPMs, visit starlinepower.com.

Starline
A brand of legrand

Starline Holdings, LLC
724.597.7800
starlinepower.com