Is Your Data Center Infrastructure Equipped for Artificial Intelligence?

Learn how Artificial Intelligence is transforming the future of data center operations, particularly in the areas of power management, fiber connectivity, and cooling solutions

PANDUIT™

13 Most Frequently Asked Questions
About Artificial Intelligence Driven Data Center Environments

The landscape of digital transformation is rapidly evolving, and integrating artificial intelligence (AI) into the planning for futureproofing your data center environment has become increasingly critical. As AI continues to evolve, many data center operators are left with lingering questions.

To assist with strategic planning around the power, fiber, and cooling components of an AI data center build, our team has compiled an eBook featuring the 13 most frequently asked questions and definitions. This resource aims to provide data center operators with valuable insights and practical guidance on designing and optimizing their infrastructure to meet the evolving demands of AI.

Table of Contents

Preparing for the Now

1 How long will the AI revolution continue, and do we expect the bubble to pop like it did for the internet in 2000?

It is always hard to accurately predict the future, but this boom is backed by some of the largest financially strong companies (Microsoft, Google, Amazon, Meta, Oracle, NVIDIA) as compared to the small startup-heavy internet craze of the late 1990s. The answer will be whether companies can achieve a good return on investment on their AI investments. Currently, there is more infrastructure spending than revenue, but there are several areas such as customer service, cybersecurity, and predictive analysis that look to offer cost reductions and efficiency improvements that can more than pay for the initial costs.

2 Do I really need to invest in having AI?

Another way to look at it is “Can you compete if you do not utilize AI and your competitors do?” The classic example is the race that Microsoft and Google are having to see which company can offer the best AI platform for search results. A significant portion of Google’s revenues come from advertising and placement on their search results and if they do not offer the best search engine, they could potentially lose billions of dollars. The same is true for many other smaller companies – if their competitors can offer better service or can be more efficient (think of Walmart dominating the retail sector by being better at logistics and customer service) they stand to lose.

3 What do I need to do differently to prepare for having an AI enabled network?

Once you have decided the scope of what you want to accomplish and where you want to have it housed (Cloud, Colocation, or On-Premises) you will need to determine how many GPUs you need.

If you are not using a Cloud provider and want to own your equipment, you will need to select suppliers, architects, systems integrators, and contractors similar to the process of creating a traditional data center. However, for AI you will need to pay close attention to the amount of power needed for the building and for each rack. AI requires much higher power which may eliminate some locations and is driving many data centers to build their own solar and wind farms as well as small modular nuclear reactors to augment the power supply from the grid.

If the power per rack is above 40kW, a more efficient system such as Rear Door Heat Exchangers (RDHX) or Direct to Chip (DtC) liquid cooling needs to be considered. Most current data centers do not offer liquid cooling, so this is an important aspect to consider.

You also need to give added attention to your network cabling and pathways. AI requires 4-6 times more fibers so you will want to work with an expert in structured cabling like Panduit who can make sure the cables are managed so that your racks and pathways do not get overly congested which can lead to damaged fibers.

View the full product roadmap for AI

Fiber

Panduit Base-8 fiber cabling is designed to save capital expenses,
rack space, and power requirements

4 What type of cabling infrastructure is needed for an AI network?

Higher-speed networks use optical fiber today and AI is no different. AI servers need more communication lanes than traditional servers so you will see 4-6 times the number of fibers. You will use similar fiber cable types, but AI servers are connected via 8 fiber MPO connectors instead of 2 fiber Duplex LC connectors. Due to this higher density, structured cabling which organizes the cabling and manages excess slack is even more important for AI data centers.

5 Do you expect to see copper cabling in an AI network?

Twisted pair copper cables will continue to be used for connecting the <10G out-of-band management as well as all low-speed sensor signals such as thermal and power measurements. For higher speed short reach (<3m) internal to a rack, passive Direct Attach Copper (DAC), Active Copper Cable (ACC) or Active Electrical Cable (AEC) can be used. While using less power than fiber cabling due to not needing a transceiver, DAC and ACC / AEC cables have a larger outside diameter and a larger bend radius which takes up more room in the often-congested pathways.

As data rates increase, the reach of copper-based cables will continue to shrink, and eventually, optical fiber will likely be used for all >10G connectivity. Optical fiber cabling is currently used for all longer reach rack-to-rack cabling including server-to-switch and switch-to-switch.

6 Does AI need Multimode (MMF) or Single-mode (SMF) cabling?

Most pods or rows of AI are between eight to eighteen cabinets so cabling lengths from server-to-switch are generally less than 30m and rarely more than 50m, which is ideal for MMF. MMF transceivers are less expensive than SMF transceivers so most customers try to use MMF if they can. Spine switches are often located in a different row than the servers and Leaf switches so SMF would be commonly used for these longer runs.

Note that Panduit MMF cables work well with both InfiniBand transceivers and ethernet transceivers. Our SignatureCore™ OM4+ MMF cables offer the lowest loss, longest length, and best bandwidth available.

7 Are new connector standards going to be used for AI?

Today's AI systems all use traditional MPOs with 8 fibers, but new smaller-sized connectors such as the MMC and SN-MT are just starting to be used to increase patch panel density in some of the largest AI data centers. Most transceivers are still using the MPO form factor, but that may change especially as the data rates move to 1.6TB.

Learn how Base-8 fiber cabling helps AI improve data center efficiency

Power

Panduit G6 Power Distribution Units are designed for the modern data center, integrating seamlessly with facility systems

8 What is the industry doing to reduce the power needed by AI?

The GPU and server manufacturers are trying to develop materials and methods to reduce power consumption. Even though overall power per rack is increasing, new chips with higher petaflops per watt are making the servers capable of handling more tasks with fewer servers which is more efficient.

Optical industry experts are also working on ways to lower the power needed by optical transceivers. They have been generally successful in coming up with materials and processes to reduce the power of each new transceiver within a few years of initial launch, but more recently they have made strides in more out-of-the-box solutions like Linear Drive Pluggable Optics (LPO), Linear Receive Optics (LRO), and Co-Package Optics (CPO) which show promise in reducing transceiver power by up to 50%.

9 How many power distribution units (PDUs) will each AI server rack need?

That is dependent on how many servers are in each rack, how much power each GPU needs, how many power supplier each server has and what power redundancy is required by the servers. It is becoming more common to see 4 or 6 PDUs per rack.

10 What power rating will my PDUs need to be for an AI network?

Most new data centers are being built with 415v 3-phase power to handle the power needs. For these data centers, you can use a 60 amp or 100 amp PDU which should handle most applications. When looking for a site for your AI network you should give preference to sites that have 415v over those with 208v/210v.

11 Can I use a basic PDU for AI or do I need to use ones with intelligence?

All data centers are seeing an uptick in the number of devices and sensors that collect information such as temperature, safety warnings, and security. To best manage all this information, it is highly recommended that AI data centers use intelligent PDUs with the preference being the monitored switch per outlet (MSPO) which offers power monitoring and outlet level switching.

Discover how the G6 PDU provides comprehensive, accurate, energy measurement data to efficiently use power resources, improve uptime, measure power usage effectiveness, and drive green data center initiatives to save energy and money

Cooling

Panduit FlexFusion™ cabinets are compatible with active rear door heat exchangers (RDHX)

12 How does physical media work with liquid immersion cooling?

Very few customers are planning on using immersion cooling as it is much more expensive and generally not needed unless you are doing a supercomputer or cryptocurrency mining. You need to use a PUR (polyurethane) jacketed DAC or Active Optical Cables (AOC) to connect to immersed GPUs as the fluid will interfere with optical signaling unless it is hermetically sealed as is the case with AOC.

13 What is the recommended cooling method?

Rear Door Heat Exchangers (RDHX) and direct-to-chip cooling (cold plates) will be much more popular than immersion cooling. Panduit is developing offerings for our FlexFusion cabinets to come with a Rear Door Heat Exchanger and/or fluid manifolds for direct-to-chip cooling.

Discover the history of cooling and what the future holds for AI-driven data centers

Definitions

Artificial Intelligence Accelerators: Chips (GPU, TPU, FPGA, and ASICs) are very efficient at parallel processing tasks (doing multiple computations) at the same time.

Artificial Intelligence (AI): Is the term used for high-level computational systems that can make complex decisions and predict outcomes based on analyzing large amounts of data.

ASICs (Application Specific Integrated Circuits): These chips are customized for particular use as compared to general-purpose chips like Intel's CPUs. Most network switches and routers use ASICs.

CPU (Central Processing Unit): Traditionally used as the main chip in personal computers and servers, these chips are generally not as powerful as GPUs and TPUs for handling larger amounts of data. However, most AI networks use them for management and networking as they are significantly lower cost and link well to existing networks that are CPU-based.

Deep Learning (DL): This is the most advanced AI system and will pull in a data set much larger than either AI or ML. It uses massive neural networks that pass through many more layers to more deeply interpret potential outcomes and is designed to be self-directed to learn on its own.

FPGAs (Field Programmable Gate Arrays): These chips are similar to ASICs as they can be made for a specific application, but like CPUs, they are not customized by the manufacturer. They are programmable, meaning the end user makes the adjustments so the chip becomes specialized to their exact requirements.

Generative AI (GI): Refers to the generation of new content such as written reports, music, art, new medicine, or new designs. The most popular are the Large Language Models such as ChatGPT.

GPU (Graphic Processing Unit): Designed to crunch high amounts of data these chips migrated from originally being focused on handling the large amount of 3D graphics data in video games to being the most commonly used chips in AI. NVIDIA and AMD dominate the graphics card business for personal computers, but NVIDIA has attained an 80% share in AI due to its high-performance chips and full ecosystem.

Inferencing: An AI system where the model uses the database from the training system to create new output or predictions to produce actionable results/content. Inferencing networks, unlike training, will usually need to be located close to the end users to offer the lowest latency. It is expected to be a mainstay of Edge data centers.

Large Language Models (LLM): Refers to AI software that can create written or verbal responses to queries, ChatGPT is the most popular example.

Machine Learning (ML): A higher functioning AI system that will use a larger data set than standard AI and have a more sophisticated ability to create new designs, recognize images, and improve with experience.

Predictive AI: Uses data to develop trends that can predict future outcomes such as when to do preventative maintenance, what content a user would like based on previous choices, when to sell a stock, or diagnose a disease.

TPU (Tensor Processing Unit): Developed by Google for their most powerful systems, these chips are like GPUs in computational power but are focused on handling high volumes of low-precision computing instead of fewer more complex tasks.

Training: An AI system that has a model that is fed a curated dataset so that it can "learn" everything it needs to know about the type of data it will analyze. It then will supply that data to an inferencing model which will respond to queries from end users. The training system is never directly accessed by end users and can be designed to access just a closed set of data or the entire internet. The servers used for training systems are separate from the servers used for inferencing and do not need to be in the same building. Training is generally more computationally intensive and requires more servers and higher bandwidth which means it also uses more power than inferencing. This can lead to training data centers being located where power is plentiful and at low cost. However, the opposite is true for popular models, such as ChatGPT, which can have millions of queries per day and are likely to use more servers for inferencing than training.

We have the knowledge and experience to help you make the most of your infrastructure investment.

panduit.com

Let’s connect
panduit.com/contact-us

PANDUIT®
infrastructure for a connected world

©2025 Panduit Corp. ALL RIGHTS RESERVED. CPEB16-WW-ENG 1/2025