

AHEAD



Simplifying and Accelerating
AI Infrastructure Deployments
with AHEAD and Cisco



How AHEAD helps enterprises deploy AI-ready infrastructure from Cisco

Enterprises are under increased pressure to incorporate AI into their business strategies and operations. In fact, 97% of organizations reported an urgency to deploy AI-powered technologies, yet only 13% of IT organizations have infrastructure prepared for AI today.

Deploying AI at scale comes with significant challenges and some demanding workloads requiring optimized and scalable infrastructure. However, many existing data centers can no longer handle the physical size, power, and cooling requirements of modern AI infrastructure.

In addition, many organizations lack the in-house AI expertise necessary to build, deploy, and maintain models and infrastructure effectively. Partnering with AI experts, investing in training, and adopting turnkey technology solutions will be crucial for long-term success.

Cisco AI Pods are ideal data center solutions to simplify the deployment of AI-ready infrastructure. By leveraging this modular and standardized solution, organizations can safely and effectively operationalize AI now while ensuring their infrastructure can easily adapt to the future.

In this whitepaper, we'll discuss the benefits of Cisco AI Pods, how AHEAD helps plan and deploy this AI infrastructure, and why leading enterprises choose to partner with AHEAD and Cisco together.



CISCO AI PODS:

Purpose-Built AI Infrastructure Solutions

Cisco AI Pods are purpose-built AI infrastructure solutions with integrated compute, storage, networking, and NVIDIA software. This lowers the total cost of ownership of deploying and operating modern infrastructure while providing greater control over resources as AI requirements evolve.



Reliability

Cisco AI Pods are validated for edge inferencing, retrieval augmented generation (RAG), and scale-out inferencing use cases, showcasing consistent performance and dynamic scalability.

Scalability

AI Pods offer a scalable path from initial deployment to advanced applications. All pre-designed AI Pods integrate UCS compute from Cisco and accelerated GPUs from NVIDIA to support high performance workloads and are modular so they can be replaced with the latest chipsets in the future.

Comprehensive Observability and Management

Centralized management with Cisco Intersight reduces operational complexity and ensures reliability even during intermittent connectivity outages. Intersight also reduces latency for real-time processing and responses. Cisco ND Insights provides enterprise level network management, observability, and reporting of AI network infrastructure metrics to ensure maximum cluster-wide performance.

Open Source Compatibility

This solution supports a wide range of leading AI, virtualization, and container platforms, so that organizations can develop and train models with the latest open source technologies. This includes the OpenShift Container Platform by Red Hat for Kubernetes-based deployments.

Storage Add-on Options

FlexPod from NetApp and FlashStack from Pure Storage are pre-sized appliances and software that can be added to an AI Pod to handle data-heavy workloads like generative AI and large language models.

How AHEAD Accelerated the Implementation of a Large RAG Cisco AI Pod

[TaxSlayer](#), a leading tax preparation and tax software company in the U.S., wanted to adopt AI for fraud detection and prevention along with advanced Python use cases related to deeper reporting and marketing automation. However, TaxSlayer needed to modernize their aging infrastructure to meet the growing demands of AI.

Facing compute limitations, security concerns, and no scalable path to AI, TaxSlayer partnered with AHEAD and Cisco to execute a dual-site data center refresh and critical security architecture uplift. This engagement also included an implementation of the first large RAG AI Pod in the U.S. commercial space, enabling TaxSlayer to move existing AI models and workloads to on-premise infrastructure.

AHEAD worked closely with Cisco to navigate the complexities of bringing together ecosystem partners to re-engineer TaxSlayer's entire compute environment. Once the design was finalized, AHEAD accelerated the deployment by pre-integrating the third party ecosystem that's part of the AI POD within our AHEAD Foundry facility.

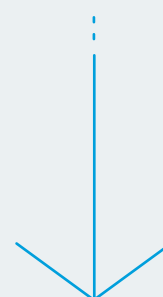
By shifting from cloud-first to a hybrid-ready strategy, TaxSlayer improved performance, reclaimed capacity, and moved in a direction that addresses both immediate operational needs and long-term scalability. If cloud costs or complexity become a barrier, TaxSlayer now has a proven path for AI expansion using its Cisco AI Pod deployments.



Simplifying AI Infrastructure Deployment

Although Cisco AI Pods are pre-integrated, these solutions can still be complicated for organizations to deploy in their data centers. Many organizations also want to customize their AI Pods with additional third-party integrations.

Having an AI infrastructure partner you can trust allows you to concentrate on what applications and innovative business capabilities you want to build. AHEAD is a leader in AI with deep technical expertise in designing, building, and operationalizing AI infrastructure.



Here's how AHEAD approaches an AI infrastructure engagement:



DEFINING REQUIREMENTS

Our AI experts work with Cisco and your team to determine the elements of your Cisco AI Pod. Then we consider other customizations like security, data protection, and HPC compute that may be needed, as well as the internal and upstream network designs.



RACK INTEGRATION

AHEAD has a team of experts in our [Foundry™ rack integration](#) facility that will work with your data center team or co-location provider to design how every piece of equipment will be racked and cabled in the real-world space.



FLOOR PLAN DESIGN

Even though our racks are plug-and-play, it's still important to make sure the space is a perfect fit for a Cisco AI Pod configuration. This means AHEAD will review your environment to ensure it meets all the requirements for deployment.



POWER & COOLING DESIGN

Although AI racks typically use at least 12kW of power and often upwards of 30-40kW of power and cooling load per rack, many existing data centers are not equipped to handle this density. AHEAD will work with your data center team to ensure the design accounts for all power and cooling requirements so the solution works once it's built at Foundry and installed onsite.



NETWORKING DESIGN

AI Infrastructures typically have multiple networks. There is an east-west GPU-to-GPU (backend) network, a north-south data network (frontend), a management network, and often a dedicated storage network. Cisco has top offerings for each of these networks, and AHEAD will work with your networking team to design the optimal networks for your Cisco AI POD.



RACK ELEVATION & CABLING PLAN

All key components are designed at the same time to avoid project delays later on. This detailed design extends to all components adjacent to the Cisco AI Pod, including storage, HPC servers, security devices, and more.



DEPLOYMENT

Once the final design is agreed on, AHEAD Foundry will build and deliver the Cisco AI Pod solution, with our [AHEAD's Hatch® platform](#) providing full visibility into the supply chain for all Cisco AI Pod procurements.



ONBOARDING & MANAGEMENT

After physical implementation and deployment, AHEAD can also provide oversight and accountability with ongoing program management for your Cisco AI Pod infrastructure.

Accelerating AI Adoption with AHEAD

AHEAD is a long-standing Cisco partner with the scale to deploy and manage Cisco AI Pod pre-integrated solutions through AHEAD Foundry, and simplify asset management for clients with our AHEAD Hatch software. We also offer professional and managed services to support full platform operations.

AHEAD Foundry is our comprehensive services for building, configuring, testing, and validating deployment capabilities at our integration facility using a plug and play approach. In addition, our full-stack integration facilities with liquid-cooling capacity are coming online this year to support future chip generations requiring increased density.

The [AHEAD Hatch IT Lifecycle Management platform](#) can track all components, design documentation, and the full asset lifecycle. Hatch also provides consolidated data at the site level for your infrastructure, and can easily integrate with ServiceNow and other leading enterprise platforms to share data in real time. This accelerates every aspect of planning, executing, and supporting your AI infrastructure initiatives.

Along with building and deploying Cisco AI Pods, AHEAD offers additional professional and managed services to reduce the burden on your IT team. Our experienced team of data scientists, architects, and engineers can help you implement and manage world-class computing solutions across on-premises and cloud environments, accelerating the impact of full stack infrastructure investments.

[Contact AHEAD](#) to learn more about implementing a Cisco AI Pod and adopting a hybrid infrastructure strategy for AI.





Combining cloud-native capabilities in software and data engineering with an unparalleled track record of modernizing infrastructure, we're uniquely positioned to help accelerate the promise of digital transformation.

Visit us at ahead.com.

National Hubs

CHICAGO

444 W. Lake Street
Suite 3000
Chicago, IL 60606

NEW YORK

500 5th Avenue
Floor 17
New York NY 10010

ATLANTA

1117 Perimeter Center
W406
Atlanta, GA 30338

SAN FRANCISCO

2000 Crow Canyon Place
Suite 250
San Ramon, CA 94583