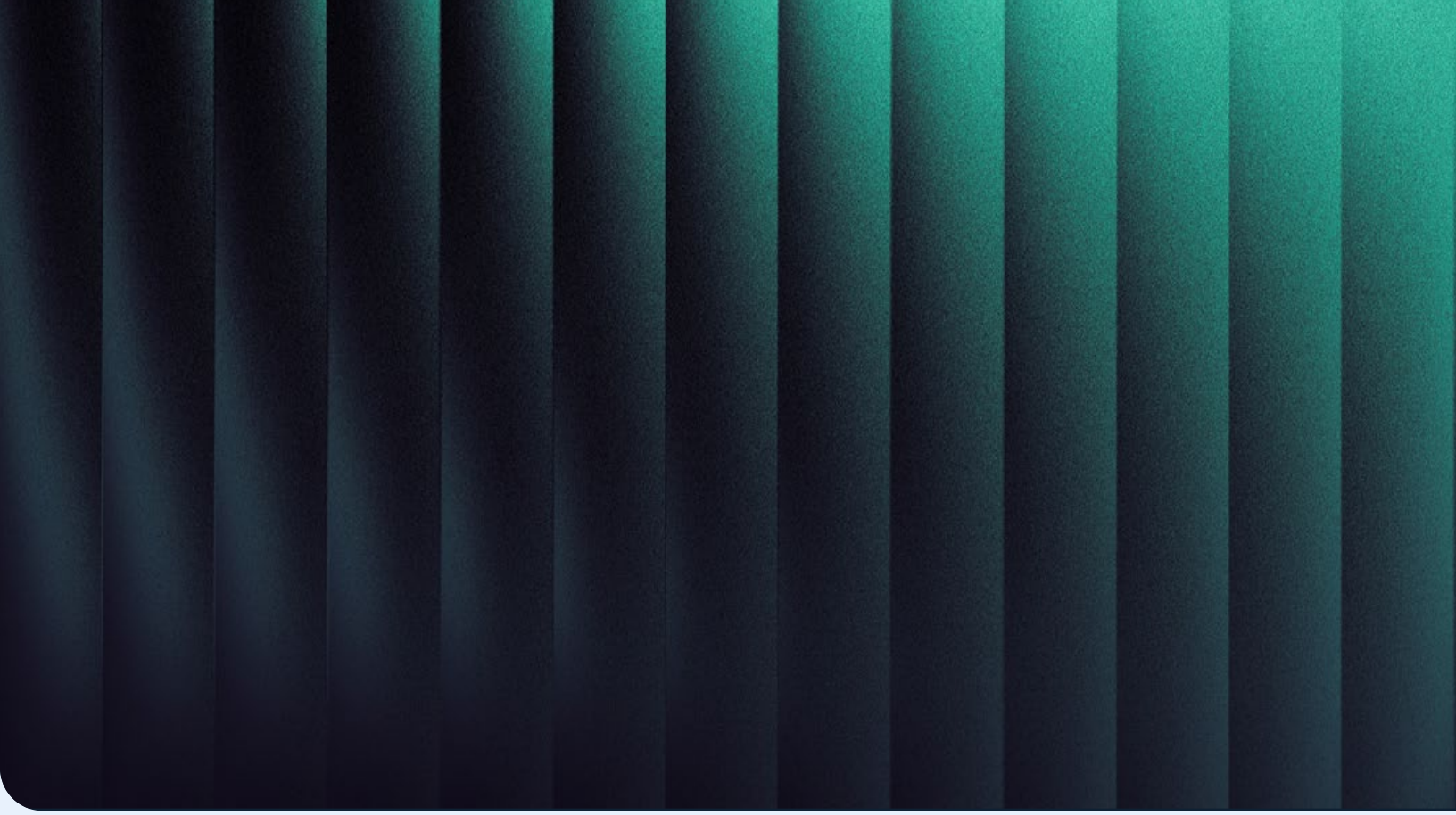
AHEAD

Adapt & Scale

A PRACTICAL PATH TO TRANSFORM EXISTING FACILITIES INTO HPC-READY DATA CENTERS

How AHEAD, Dell, and NVIDIA enable organizations to adapt their existing infrastructure for HPC workloads



Although AI has taken the spotlight recently, the demand for high performance computing (HPC) workloads like simulations, modeling, and analytics is also growing faster than many existing environments can accommodate. In fact, HPC workloads and dense enterprise workload consolidation architectures have similar power densities as AI workloads, which means they can strain existing power, cooling, and space capacity in much the same way.

The trend toward increasing power density per rack is even impacting traditional compute. For example, a fully populated 42U dual-socket enterprise server rack now draws 40kW or more, up sharply from the sub-15kW racks that were typical for enterprise computing in the past. This means data centers need new approaches to handle modern workloads regardless of whether AI or HPC is the primary driver.

The good news is HPC readiness does not require an all-at-once infrastructure replacement or an entirely new data center. Organizations can create an adaptable foundation by understanding their current constraints, separating workloads by need, adding capacity in modular waves, modernizing data and networking, and embedding lifecycle and operational practices that make future expansion predictable.

According to a recent report from IDC, AI and HPC workloads may also share data center space with other types of lower-power density workloads to make the best use of total available facility power capacity. This means existing facilities can be adapted to more easily run various types of infrastructure as business needs evolve to optimize the overall efficiency of enterprise infrastructure.

The Case for Adaptability

HPC workloads are becoming more diverse, from CPU-intensive simulation and modeling use cases to GPU-accelerated analytics and interactive visualizations. Besides varying server requirements, data centers need to support data-intensive pipelines that stress storage and network traffic along with hybrid deployment environments across cloud and on-premises infrastructure. That's why it's important to shift infrastructure investment planning toward adaptable architectures that can handle the demands of various workloads.

Fixed infrastructure can cause avoidable friction because capacity is often procured and provisioned in large, infrequent increments that slow HPC workload rollouts. Compute may also be available while storage or networking becomes the bottleneck. Additional friction is introduced when data is duplicated or moved unnecessarily between different teams and platforms. Finally, limited telemetry makes it difficult to quantify performance and connect infrastructure consumption to outcomes.

With 600kW per rack on the horizon and 1MW per rack being planned, data center power supply and distribution also has to be reconsidered. For example, a data center that's capable of handling a wide variety of modern workloads will likely need a power distribution of over 100kW per rack. In turn, cooling and heat dissipation systems will likely need to be upgraded to accommodate these more demanding HPC workloads. However, it is possible to scale aisle by aisle using modular power and cooling solutions in 10MW increments to more efficiently utilize space and facility-level energy supplies.

For these reasons, we recommend viewing data center adaptability as an operating model rather than a single initiative. We define infrastructure adaptability across six dimensions:

CAPACITY

Adding compute, storage, and network resources in manageable waves

PERFORMANCE

Matching architectures to workload characteristics and service levels

DATA

Moving, protecting, and governing data without forcing a single storage pattern

FACILITY

Upgrading power and cooling systems incrementally to maximize efficiency

OPERATIONS

Using automation, observability, and repeatable runbooks

ECONOMICS

Connecting utilization and outcomes to investment decisions

Assessing Current Capabilities and Constraints

Before considering any infrastructure upgrades, it's crucial to assess the current capabilities of your existing facilities. The available rack space along with electrical distribution, redundancy, and operational constraints will often dictate the types of hardware you can realistically deploy and where. You'll also want to consider any existing server, storage, and network assets that can be repurposed to increase current utilization.

After getting a baseline, you may discover it's not compute capacity that's the real constraint for HPC workloads. Data storage capacity and throughput could be the actual bottleneck preventing adequate performance. In addition, network contention and east-west traffic limitations can slow workloads even when compute and storage are sufficient.

Next, you can evaluate your HPC workloads and map them to your existing architecture to discover any gaps and potential areas for infrastructure investments.

WE RECOMMEND ORGANIZING WORKLOADS BY:

- CPU versus GPU intensity
- Batch versus interactive access
- Data volume and data-movement profile
- Latency and throughput requirements
- Security, compliance, and data-sovereignty needs
- Growth rate and business or research criticality

Scaling Modular Infrastructure for HPC

The key to preparing your data center for HPC workloads is scaling in a way that's adaptable and extensible. By using repeatable building blocks or modules, new resources can be integrated on demand without redesigning the whole environment. This ensures near-term capacity additions are made quickly and with minimal disruption, while longer-term platform decisions are implemented more deliberately.

The first design principle we recommend is to keep compute, storage, and networking decisions connected while still being independently extensible. This means using architectures that allow capacity and performance to grow at different rates, and planning for storage and network growth that can keep pace with new compute.

Next, we recommend making the data layer a shared accelerator by supporting the access patterns of both HPC and AI workloads. By establishing policy-based data placement, protection, retention, and lifecycle management, organizations can reduce unnecessary copies and manual processes across teams.

Finally, it's crucial to design new infrastructure for repeatable deployments by standardizing configurations and validation criteria. This makes it easier to pre-integrate and test components before they arrive onsite, reducing risk and time to deployment. We also recommend defining repeatable processes for provisioning, configuration, monitoring, and documentation that can be easily automated.

Foundry

Streamlining Deployments with AHEAD Foundry

AHEAD Foundry™ is our comprehensive services for building and integrating all hardware and networking infrastructure at scale in our facility rather than at your sites. This enables organizations to quickly transform infrastructure investments into operational capacity.

Our AHEAD Foundry offering includes:

PRE-DEPLOYMENT CONFIGURATION

Pre-configured products leave AHEAD facilities in a consistent foundational state, ensuring reliability and performance, and economies of scale.

RACK INTEGRATION

Integrated solutions decrease DOAs (dead on arrival), reduce missing parts, and improve time to value.

WAREHOUSE AND KITTING

Kitting allows organizations to ship all items to our facility, regardless of who they are purchased from. Items are consolidated and palletized to simplify

GLOBAL LOGISTICS SERVICES

Our dedicated global logistics team ensures compliance and efficiency across global shipments.

AHEAD HATCH IT LIFECYCLE MANAGEMENT

Our proprietary platform consolidates and normalizes all data from various software tools to provide a single-pane view of hardware engineering and logistics. This includes direct integration into Dell production facilities for real-time inventory tracking.

AHEAD can pre-configure, integrate, and kit racks in AHEAD Foundry before shipping them to data centers globally. This turnkey approach accelerates the rollout of new hardware infrastructure and reduces the time to value for customers. And for especially complex deployments like AI and HPC infrastructure, Foundry offers fewer surprises and a faster path to production-ready AI at scale.

Operating the Adaptable Platform

After deploying new capacity, how your teams manage it determines whether it will deliver long-term business value. Operating an adaptable HPC environment means combining observability, workload orchestration, and lifecycle management into a single framework for efficiently adding capacity and refreshing hardware.

Start by implementing observability to measure infrastructure health, capacity, utilization, and job-level performance. We recommend instrumenting compute, memory, GPU, network, and storage utilization and tracking capacity consumed, capacity available, and adoption to inform expansion decisions.

In addition, lifecycle management involves tracking assets, configurations, warranties, support coverage, and refresh windows across all environments. This data will enable your operations team to proactively plan incremental refreshes so the infrastructure platform can evolve without destabilizing production workloads.

If your organization needs to enhance its hardware inventory, asset, and contract management capabilities, AHEAD Hatch® is our proprietary IT lifecycle management platform that simplifies infrastructure operations. The platform centralizes in-depth asset data so operations teams can make more informed infrastructure decisions that maximize the long-term value of hardware.

Designing a Phased Roadmap

Now that we've covered how to assess, scale, and operate HPC-ready infrastructure, we'll walk through how to design a phased roadmap from scratch.

PHASE 1

Assess and prioritize

First, we recommend assessing your current capabilities and constraints by creating an inventory of your current environment and facility. Then you can identify the highest-value workloads and the most material bottlenecks to prioritize your HPC workloads based on realistic conditions. After organizing your target workloads and doing a gap analysis, you can define success metrics and a decision framework to guide your infrastructure investments.

PHASE 2

Prove the pattern

Before a large scale rollout of additional HPC capacity, we recommend running a pilot program with one or two workloads representative of your target architecture. This allows you to validate the compute, storage, network, scheduling, and data-access design under real-world conditions. From there, you can establish performance baselines and operational runbooks for scaling further.

PHASE 3

Integrate and deploy a production wave

After a successful pilot program, you can begin implementing modular capacity additions in incremental phases. Each wave should include the security, privacy and data protection controls necessary for production use. Then you can onboard initial user groups and capture adoption feedback that can be incorporated into the next production wave.

PHASE 4

Scale with evidence

In the final phase, we recommend tracking key operations metrics and revisiting your initial roadmap as workloads, facility constraints, and economics change. This allows you to expand capacity strategically based on utilization, queue demand, workload results, and business or research outcomes.

Building an HPC-Ready Data Center with **AHEAD and Dell**

Now that we've covered how to adapt and scale a facility for HPC workloads from start to finish, we'll look at a validated architecture designed by AHEAD and Dell.

You need both sides to make any optimization decision worthwhile.

Dell Hardware for HPC & AI

Dell's infrastructure portfolio is designed to support a variety of HPC and AI use cases without locking enterprises into a single architecture. This allows organizations to select appropriate hardware solutions based on workload requirements and facility constraints.

Here are a few suitable Dell solutions for the most demanding modern workloads:

Dell PowerEdge is a compute server product line that's adaptable for different types of workloads. You can use GPU-optimized Dell PowerEdge XE platforms for AI training, inference, and GPU-accelerated HPC where justified by workload requirements. Meanwhile, Dell PowerEdge R-series platforms is ideal for CPU-intensive HPC, orchestration, management, and supporting services.

Dell PowerScale provides a foundation for consolidating and simplifying the data layer of AI and HPC environments, while Dell's newer Exascale offering brings multiple AI and HPC storage capabilities together in a single, highly scalable solution. Paired with Dell Automation Platform, which automates compute provisioning across Dell PowerEdge servers, this approach helps organizations deploy and manage infrastructure more consistently as workloads evolve. The combination supports flexible deployment models and enables compute and storage resources to scale in a controlled manner, complementing broader data center modernization efforts while accommodating diverse AI and HPC requirements.

Dell PowerStore is an intelligent and adaptable storage solution that can support high-density workloads with high performance and cost efficiency. This solution offers flexible storage options for shared HPC and AI data, whether it's block, file, or object access patterns. You can choose high-throughput shared file access for active research and AI data sets, or capacity-oriented storage for growing unstructured data.

AHEAD ADVISORY & MANAGED SERVICES

AHEAD is the leader in data, developer, and infrastructure platforms for enterprise organizations. We can partner with your organization for any phase across the entire lifecycle, from an initial assessment to deployment and ongoing operations. Our advisory team can help evaluate your baseline infrastructure and offer expert advice that aligns workloads, constraints, goals, and investment priorities. From there, we can create a modular architecture design and expansion roadmap to guide your infrastructure investment decisions.

Along with advisory services, our engineers can configure, test, document, and validate your Dell hardware before deployment. AHEAD has the infrastructure expertise, Dell certifications, and deep experience to ensure clients' success when adapting their data centers to handle HPC workloads. In fact, we are a Dell Technologies Titanium Black Partner focused on consultative solutions and one of the top 3% of Dell partners globally.

In addition, AHEAD's state-of-the-art lab outside of Chicago has a comprehensive Dell portfolio available for testing configurations, proof of concepts, and exploratory solution development. Our Automation Hub also has 400+ custom IP solutions to accelerate digital strategies, including accelerators for Dell PowerStore, Dell PowerFlex, and more. These can help organizations determine whether Dell solutions are right for their use cases before committing to a large-scale rollout while also automating many aspects of deploying and operating Dell infrastructure.

Adapt Your Data Center for HPC Workloads

Enterprises face considerable challenges when adapting data centers for HPC workloads, which often includes significant changes in infrastructure and operations. It's important to remember that HPC readiness is a continuous capability, not a single procurement event. The most resilient path starts with assessing your current environment and improving it in measured waves.

If you're looking for experienced partners to guide you, AHEAD and Dell can help your organization assess constraints, validate architectures, integrate infrastructure, and operate an adaptable platform for AI and HPC. Contact AHEAD now to start with an HPC and AI readiness assessment.

AHEAD

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.

Global Hub Offices

CHICAGO
444 W. Lake Street
Suite 3000
Chicago, IL 60606
United States

NEW YORK
500 5th Avenue
17th Floor
New York, NY 10010
United States

UNITED KINGDOM
310 Wharfedale Rd
Winnersh, Wokingham
RG41 5TP
United Kingdom

NETHERLANDS
Kerkenbos 10-123
6546 BJ Nijmegen
The Netherlands

INDIA
L-21, One Horizon Center
Golf Course Road, DLF Phase V
Sector 43, Gurugram 122002
India