Futurum

NetApp®

Enterprise AI Evolution: Maximizing Data Value in a Hybrid World

JULY 2025

Executive Summary

As artificial intelligence (AI) transitions from experimental projects to mission-critical enterprise applications, organizations face a fundamental shift in how they approach AI implementation, data management, and infrastructure design.

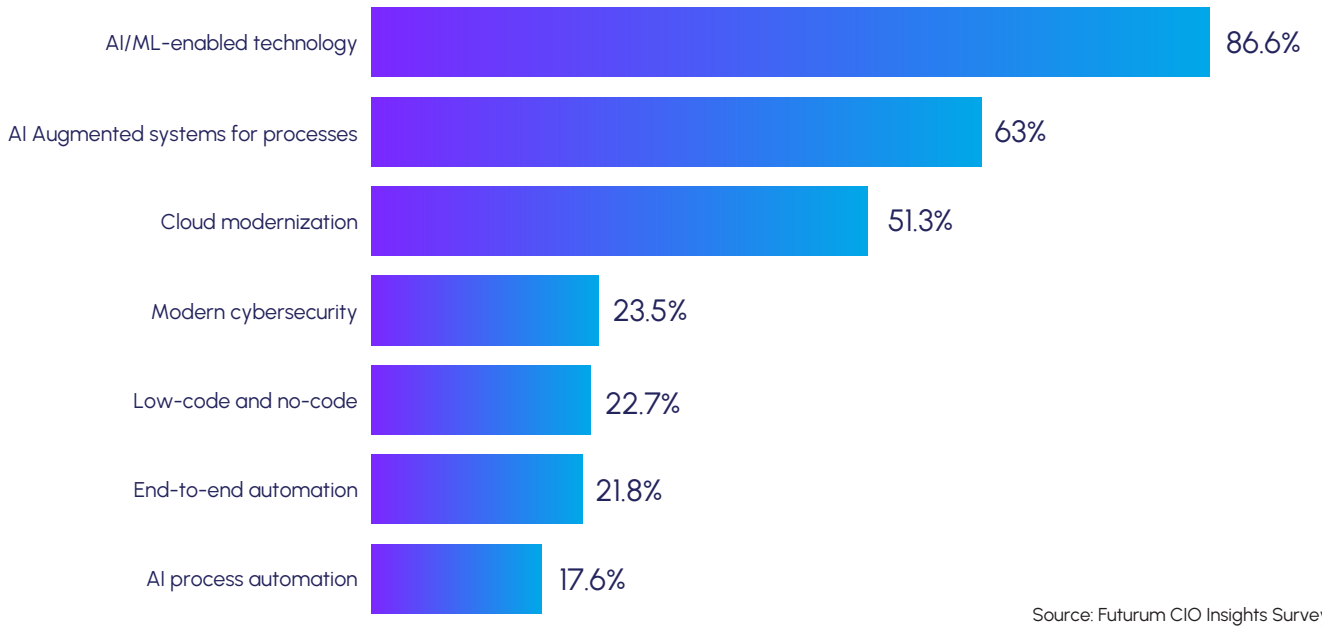
Organizations are increasingly moving beyond AI experimentation to production deployments, creating new demands for manageability, cost efficiency, and collaboration across teams. To maximize the value of their data for AI initiatives, businesses are focusing on data integrity, governance, and accessibility within their current systems – making intelligent data infrastructure a foundational consideration for AI-ready enterprise architecture. Furthermore, most enterprises are finding that a hybrid approach, combining cloud services with on-premises infrastructure, offers the most effective and adaptable path for AI implementation.

Enterprise AI Shift: Meeting Evolving Customer Expectations

The Mainstreaming of Enterprise AI

AI has decisively moved beyond proof-of-concept to become a foundational element of enterprise technology strategies. According to the Futurum CIO Insights Survey (Q1 2025), AI/ML-enabled technologies dominate emerging IT investments, with 87% of CIOs identifying them as a primary focus area.

CIO Technology Investment Priorities (Q1 2025)



Source: Futurum CIO Insights Survey, Q1 2025

CIOs are particularly focused on applying AI-augmented systems for process efficiencies (63%) and integrating AI with cloud modernization initiatives (51%). This represents a significant shift from experimental AI to operational transformation at scale.

This mainstreaming creates new challenges as organizations scale from isolated experiments to enterprise-wide deployment. The stakeholder ecosystem has expanded beyond data scientists to include IT operations, security teams, compliance officers, and business users, each bringing distinct requirements and concerns. As AI becomes embedded in critical business processes, the expectations around reliability, governance, and cost management have risen significantly.

Build vs. Buy: The Changing AI Implementation Paradigm

As AI adoption accelerates, organizations are reevaluating their infrastructure approach. The emergence of turnkey AI appliances provides ready-to-deploy alternatives to custom-built infrastructure, particularly appealing to enterprises seeking to reduce complexity.

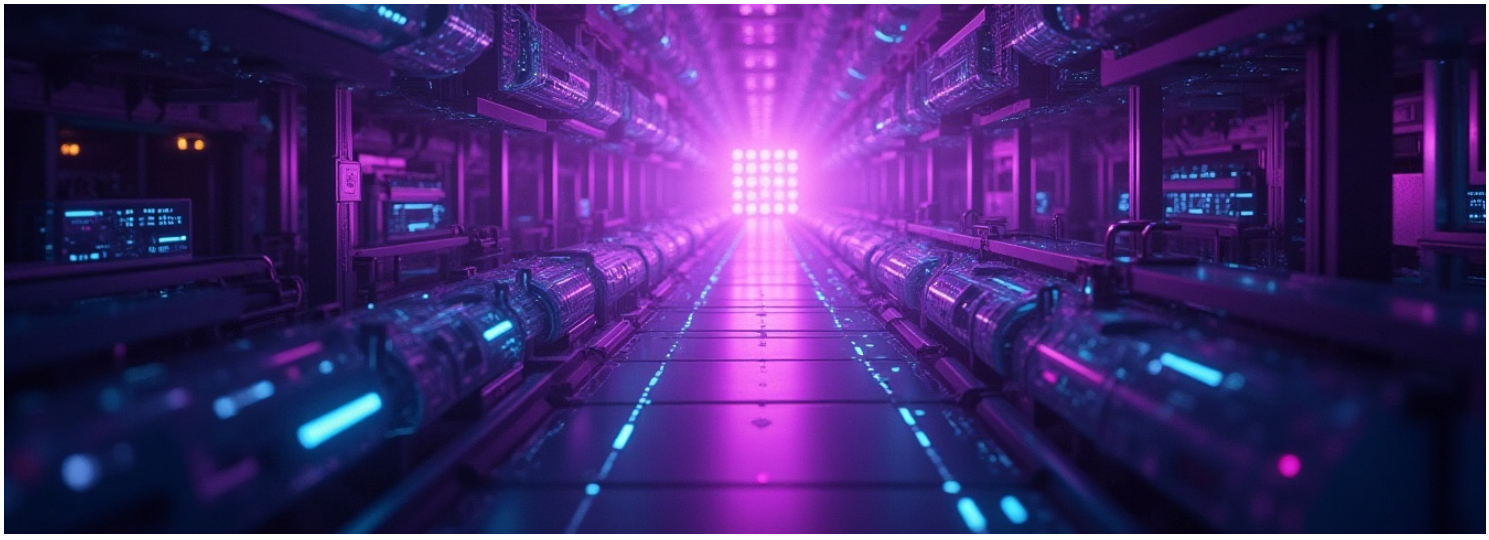
Price sensitivity varies significantly across organization sizes, but a common trend is the accelerated depreciation cycles for AI infrastructure. What was once a 3- to 5-year investment horizon has contracted to approximately 18 months as AI technologies evolve at unprecedented speed. This rapid evolution creates pressure to select infrastructure that can adapt to changing requirements without complete replacement.

Organizations therefore increasingly find themselves balancing the flexibility of custom infrastructure against the simplicity and speed-to-value of packaged solutions. The optimal approach often combines standardized components for common workloads with specialized elements for unique requirements.

IT Integration and Management Challenges

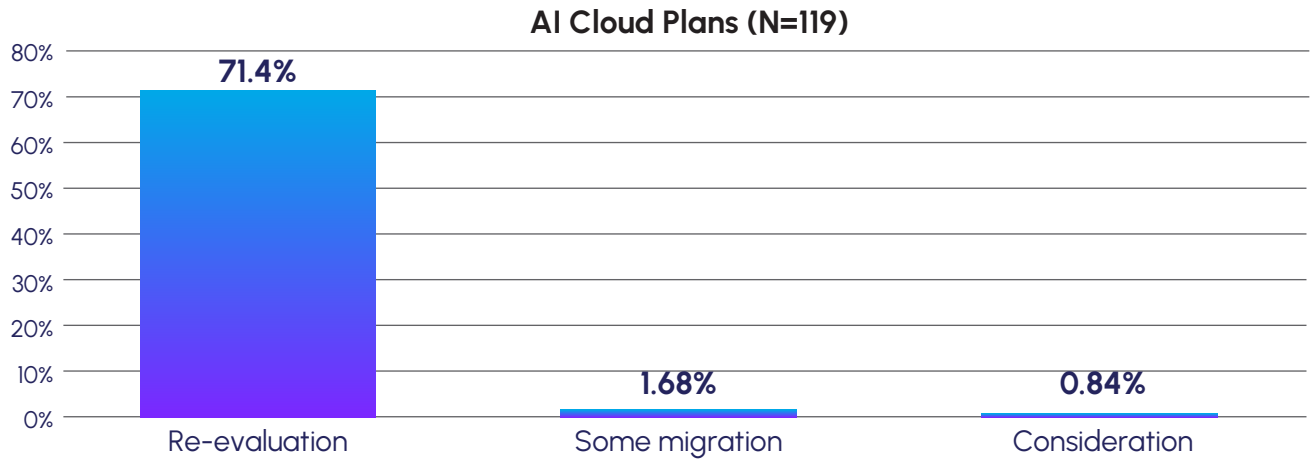
The shift to production AI is driving earlier involvement of IT teams in the AI development lifecycle. Where data science teams previously drove much of the initial AI investment, IT leaders are now engaged from the outset to ensure enterprise-grade infrastructure readiness, integration with existing systems and security frameworks, comprehensive cost management and ROI calculations, and operational resilience with high availability.

This evolution highlights a critical need for AI infrastructure that satisfies both data science innovation requirements and IT operational standards. This is precisely the role of Intelligent Data Infrastructure – a unified framework that brings together AI-readiness, cyber resilience, and operational scalability into a single, integrated foundation. The most successful organizations establish collaborative processes where data science, business stakeholders, and IT operations develop shared frameworks for evaluating technology investments. These frameworks balance innovation speed with enterprise stability, creating sustainable approaches to AI scaling.



The Need for Multi-Workload Infrastructure

Organizations are increasingly recognizing the limitations of niche storage implementations focused on single workload types. The business case for flexible infrastructure that serves multiple applications has become compelling as AI workloads diversify beyond training to include inference, vector databases, and data preparation. Budget constraints push for consolidated infrastructure that can serve both AI and traditional workloads—further elevating the importance of adopting an Intelligent Data Infrastructure approach that flexibly supports multiple workload types without compromise.



Source: Futurum CIO Insights Survey, Q1 2025

Figure 2: Cloud workload reevaluation plans

As Figure 2 shows, 71% of organizations plan to comprehensively reevaluate their cloud workloads in 2025 to optimize placement between private and public environments. This signals a pivotal moment in enterprise infrastructure strategy, with AI requirements driving much of this reassessment. The overwhelming preference for comprehensive workload evaluation rather than incremental changes suggests organizations recognize the strategic importance of optimizing their entire infrastructure portfolio for AI readiness.

Security and Governance in Distributed AI Environments

The shift to production AI amplifies security and governance challenges across distributed environments. Organizations must maintain consistent security policies whether data resides on-premises, in public clouds, or hybrid configurations. Traditional security approaches that rely on perimeter-based controls prove inadequate for AI workloads that frequently access data across multiple locations and require dynamic resource allocation.

Regulatory compliance adds another layer of complexity, particularly for organizations operating across multiple jurisdictions. Data residency requirements, such as those mandated by GDPR in Europe or industry-specific regulations like HIPAA in healthcare, create constraints on where AI processing can occur. Organizations must implement governance frameworks that can adapt to varying compliance requirements while maintaining operational efficiency.

NetApp's unified data fabric approach addresses these challenges by providing consistent security and governance controls regardless of where data physically resides. Through integrated identity management, encryption in transit and at rest, and comprehensive audit trails, organizations can maintain security posture across hybrid environments. The platform's policy-driven approach ensures that governance rules automatically follow data, reducing the risk of compliance violations during AI development and deployment.

Maximizing Data Value: Breaking Down Data Silos

The Data Challenge in AI Development

AI development has triggered exponential growth in enterprise data requirements, pushing organizations to reconsider their data management approaches. Highly regulated industries such as healthcare, financial services, and pharmaceuticals have pioneered integrated approaches to balance innovation with strict compliance mandates.

Key to their success has been robust metadata management - tracking not just the data itself but its lineage, quality metrics, access patterns, and governance requirements. This "data about data" has become as strategically important as the primary data assets. Organizations with mature metadata practices can accelerate AI development while maintaining governance, creating a sustainable foundation for innovation.

The volume, variety, and velocity of data needed for effective AI development create significant operational challenges. Organizations must balance the need for comprehensive data access against performance and cost constraints. Many early AI initiatives struggled with this balance, creating expensive data lakes that failed to deliver expected value due to governance gaps or performance limitations.

Connecting Models to Enterprise Data Assets

Organizations are adopting sophisticated techniques to bridge the gap between open-source models and proprietary enterprise data. Fine-tuning existing models on domain-specific data improves relevance and accuracy, while implementing Retrieval-Augmented Generation (RAG) architectures combines LLMs with enterprise knowledge bases. Developing secure integration patterns maintains data sovereignty while leveraging model capabilities.

RAG implementations represent a critical first step in most enterprise AI journeys, serving as a foundational milestone toward more advanced AI capabilities. Organizations must first develop comfort with prompts and queries against their own data and establish trust in AI-generated results before considering more sophisticated implementations like autonomous AI agents.

Importantly, this initial RAG step requires enterprises to develop comprehensive security and governance policies that can later be extended and scaled for agentic AI deployments.

While technology vendors rapidly develop agentic AI frameworks, enterprise adoption of these advanced capabilities remains constrained by risk mitigation concerns. Successfully implementing RAG architectures builds the operational confidence and security frameworks necessary for eventual deployment of more autonomous AI systems.

These approaches address a fundamental disconnection while introducing critical security considerations. Most cutting-edge AI models originate outside the enterprise, creating potential intellectual property and data sovereignty risks when integrating with internal data assets. Organizations must implement zero-trust architectures that verify every access request, encrypt data at multiple layers, and maintain detailed access logs for compliance auditing.

The challenge intensifies when dealing with fine-tuning or RAG implementations that require persistent access to enterprise data. NetApp's approach enables secure model-data integration through encrypted data channels, role-based access controls, and data lineage tracking that maintains compliance visibility throughout the AI lifecycle. This ensures that organizations can leverage external AI capabilities without compromising their data security posture.

The emergence of vector databases as a specialized storage layer for AI applications exemplifies these security and integration challenges. Organizations must balance the benefits of purpose-built vector databases against the complexities of managing another data silo that requires its own security controls, access policies, and governance frameworks. Integration strategies that connect vector representations with source data management systems prove critical not only for maintaining data consistency but also for ensuring that security policies and audit trails remain intact throughout the AI pipeline. NetApp's unified approach enables organizations to implement vector databases while maintaining centralized governance, ensuring that specialized AI storage doesn't create security gaps or compliance blind spots.

Data movement and Gravity Challenge

One of the most underestimated challenges in enterprise AI deployment is the physics of data movement at scale. Organizations frequently underestimate the time, cost, and complexity of moving large datasets between environments. A 100TB training dataset that seems manageable in theory becomes a months-long project when accounting for network bandwidth limitations, data validation requirements, and compliance checkpoints.

Data gravity effects become pronounced as datasets grow. Rather than moving petabyte-scale datasets to computational resources, organizations find it more practical to bring computation to the data. This shift requires rethinking traditional IT architectures that assume data mobility and instead designing for data locality. The most successful AI implementations develop tiered strategies that leverage high-performance local storage for active training data while maintaining broader datasets in cost-effective remote storage.

Geographic data distribution adds another dimension of complexity. Organizations with global operations often discover that their most valuable AI training data spans multiple continents, each with distinct regulatory requirements. The European Union's data residency requirements, China's data localization laws, and various industry-specific regulations create a complex web of constraints that must be navigated without sacrificing AI model quality.

NetApp's approach addresses these challenges through intelligent data tiering and caching strategies that optimize for both performance and compliance. By maintaining data locality where required while enabling secure, high-performance access to distributed datasets, organizations can develop AI models that leverage their full data estate without violating geographic or regulatory constraints.

Solving Advanced AI Data Challenges

As AI matures within enterprises, organizations face increasingly sophisticated data challenges. Data scientists spend up to 70% of their time locating and preparing relevant data. Streamlining this process represents a major opportunity for productivity gains. Vector database management becomes more complex as vector embeddings proliferate, requiring organizations to manage storage growth and optimization, particularly for frequently updated knowledge bases. Ensuring data currency becomes critical as organizations must develop sophisticated synchronization and change detection mechanisms so AI systems operate on current rather than outdated information.

Leading organizations are tackling these challenges through comprehensive data cataloging, automated lineage tracking, and integration of governance controls into data pipelines. These capabilities enable data scientists to discover and utilize relevant data assets efficiently while maintaining compliance with internal and external regulations.

The most sophisticated enterprises are developing unified data access layers that abstract the complexities of underlying storage systems, providing consistent access patterns regardless of whether data resides in cloud object stores, on-premises file systems, or specialized databases.

The Hybrid AI Future: Integrating Cloud Services with Enterprise AI

The Reality of Hybrid AI Deployments

Practical enterprise AI deployments rarely exist solely in one environment. Instead, organizations are developing sophisticated hybrid patterns. Cloud bursting enables training models on-premises for routine workloads while leveraging cloud resources for periodic intensive training. Mixed-location processing maintains sensitive data on-premises while utilizing cloud-based model services for inference. First-party integration connects enterprise data systems with cloud provider AI PaaS/SaaS offerings.

A persistent challenge remains the disconnect between cloud AI services and on-premises data repositories, often requiring complex extract-transform-load (ETL) processes that introduce delays and compliance risks. Forward-thinking organizations are developing architectures that minimize data movement while enabling seamless integration between cloud services and on-premises data assets.

The complexity of these hybrid patterns creates significant operational challenges, particularly in governance and cost management. Organizations must develop comprehensive frameworks that maintain visibility and control across environments while optimizing resource utilization. Those that succeed in developing these frameworks gain significant competitive advantages through greater agility and lower operational overhead. These outcomes are exactly what Intelligent Data Infrastructure is designed to deliver—ensuring governance, efficiency, and adaptability across hybrid AI deployments.

AI Workload Diversity and Infrastructure Requirements

The AI workload spectrum has expanded dramatically, encompassing high-intensity training for foundation models, continuous fine-tuning of domain-specific models, vector search and retrieval, real-time inference for customer-facing applications, and batch inference for internal analytics. Each workload presents distinct infrastructure requirements, further complicated in multi-cloud environments where organizations must navigate different storage paradigms, networking configurations, and compliance frameworks such as the EU's Digital Operational Resilience Act (DORA).

This diversity creates tension between standardization for operational efficiency and specialization for workload optimization. Organizations must balance these competing priorities, developing infrastructure strategies that provide consistency where possible while accommodating specialized requirements where necessary.

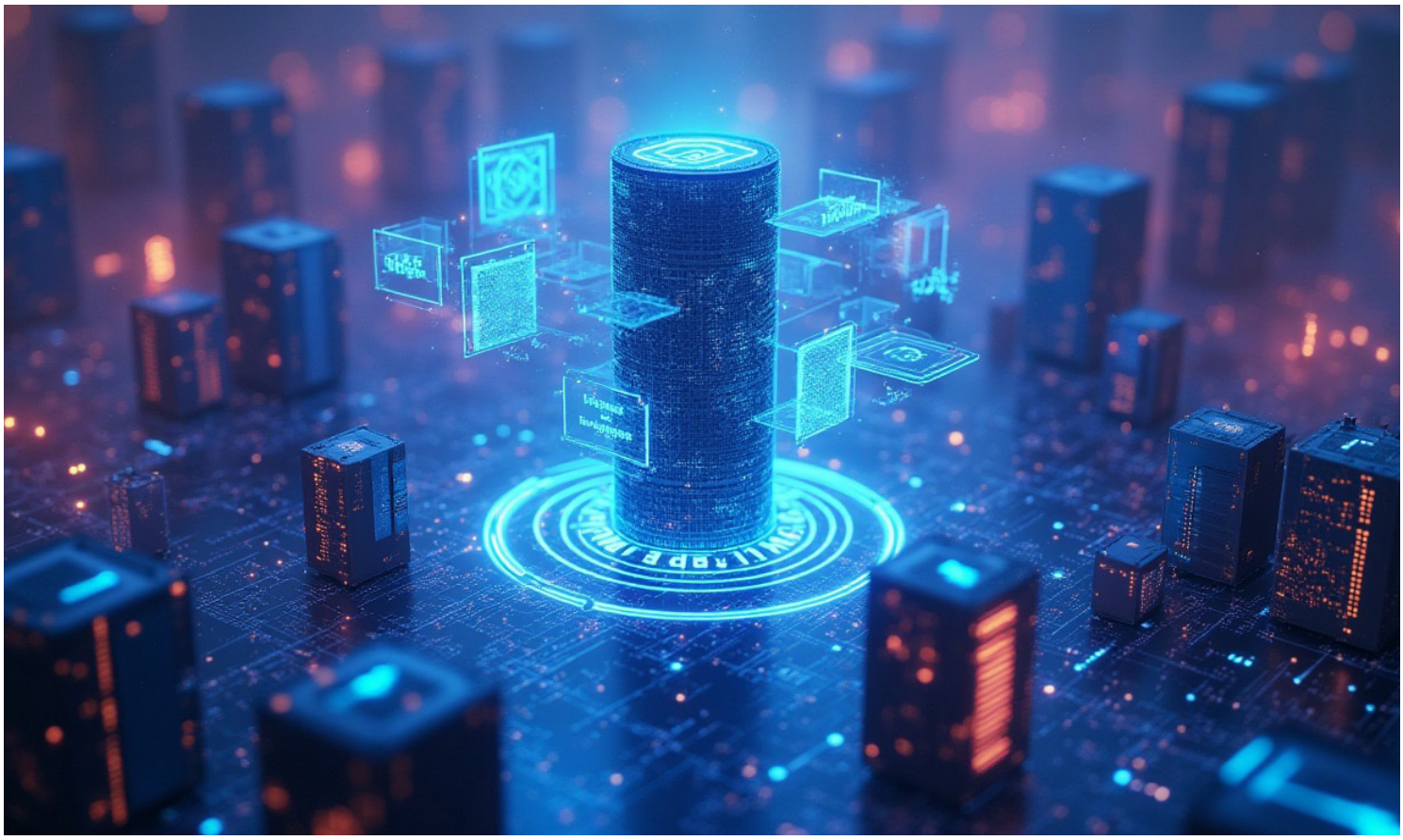
The rapid evolution of AI technologies compounds these challenges, as infrastructure decisions made today must accommodate workloads that may not exist yet. This uncertainty drives interest in flexible, scalable approaches that can adapt to emerging requirements without wholesale replacement.

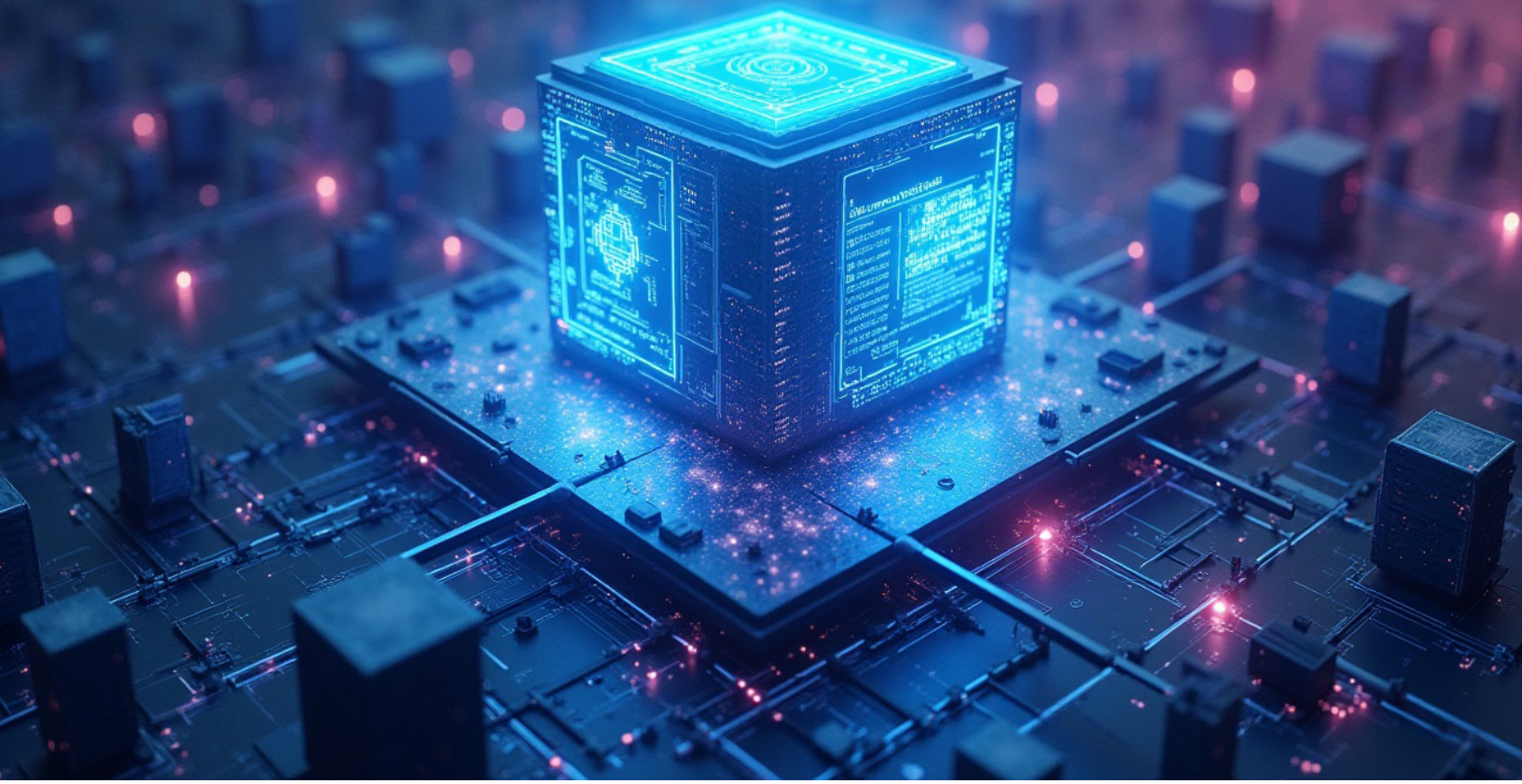
Unified Control Plane for Distributed AI

Forward-thinking enterprises are implementing unified control planes to manage distributed AI environments. These systems provide centralized policy management across environments, consistent data access patterns regardless of physical location, and the ability to consume data from disparate sources without infrastructure refresh.

This unified approach is a core principle of Intelligent Data Infrastructure, enabling organizations to evolve their strategy incrementally rather than requiring wholesale replacement of existing systems. It recognizes that most enterprises have significant investments in on-premises infrastructure that must be leveraged alongside cloud capabilities to maximize ROI.

The most effective control planes extend beyond infrastructure management to encompass data governance, model lifecycle management, and deployment automation. This comprehensive approach enables organizations to accelerate AI development while maintaining appropriate controls, significantly reducing the operational friction that often slows enterprise AI adoption.





Conclusion

As AI becomes central to enterprise operations, organizations must rethink traditional approaches to infrastructure, data management, and cloud strategy. The challenges outlined throughout this brief require sophisticated, integrated solutions rather than point products.

NetApp's unified data fabric directly addresses these enterprise AI challenges through several key capabilities. The platform's consistent security and governance framework ensures that data protection policies automatically follow data across hybrid environments, maintaining compliance visibility whether processing occurs on-premises or in cloud environments. For data movement and gravity challenges, NetApp's intelligent data tiering and caching capabilities optimize data placement based on usage patterns, regulatory requirements, and performance needs while respecting local data sovereignty requirements.

The integration challenges between AI models and enterprise data assets are addressed through NetApp's support for modern AI architectures including vector databases, object storage, and high-performance file systems. This unified approach eliminates data silos while maintaining enterprise governance controls. Critically, NetApp's platform integrates with existing storage environments, enabling organizations to evolve incrementally rather than requiring disruptive infrastructure replacement.

These outcomes are enabled through Intelligent Data Infrastructure aimed to unify data services, simplify operations, and embed intelligence from the start. Organizations that succeed in enterprise AI will be those that solve the fundamental challenges of data access, security, and governance at scale.

Important Information About This Report

AUTHORS

Nick Patience
Vice President & Practice Lead,
AI Platforms | The Futurum Group

PUBLISHER

Futurum Research

INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in context, displaying author's name, author's title, and "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations.

LICENSING

DISCLOSURES

The Futurum Group provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

NetApp®

ABOUT NETAPP

NetApp is the intelligent data infrastructure company combining unified data storage, integrated data services, and CloudOps solutions to turn a world of disruption into opportunity for every customer. NetApp creates silo-free infrastructure, then harnesses observability and AI, to enable the best data management. As the only enterprise-grade storage service natively embedded in the world's biggest clouds, our data storage delivers seamless flexibility and our data services create a data advantage through superior cyber-resilience, governance, and applications agility. Our CloudOps solutions provide continuous optimization of performance and efficiency through observability and AI. No matter the data type, workload, or environment, transform your data infrastructure to realize your business possibilities with NetApp. Learn more at www.netapp.com or follow us on X, LinkedIn, Facebook, and Instagram.

Futurum

ABOUT THE FUTURUM GROUP

The Futurum Group is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

Futurum

CONTACT INFORMATION: The Futurum Group LLC | futurumgroup.com | (833) 722-5337