



Four must-haves to fully leverage your data for AI workloads

The era of AI experimentation has officially given way to the AI factory. For any organization looking to transform its data from a static archive into a high performance engine, there are four non-negotiable “must haves”.



AI has raced from experiment to expectation almost overnight. Boardrooms are asking where copilots, assistants, and predictive models can move the needle, while teams in the trenches are wiring proof-of-concepts into every corner of the business. The common thread across all of it isn't just models or GPUs, but data.

Every new use case needs more of it, in more shapes, at higher speeds, and under tighter scrutiny from security, compliance, and the business. If your enterprise is like most, all that AI ambition sits on top of a data estate that was never designed for this level of volume, variety, and velocity.

Pilots work in the lab but stall in production because the underlying data is fragmented, poorly governed, or stored on infrastructure that can't keep up. In fact, according to a recent IDC survey, 70% of enterprise AI projects stall before making it into durable, repeatable production.¹ Retrieval-Augmented Generation (RAG) projects can't see the most important content. Models drift because feedback and telemetry aren't captured in a way that can be reused. Risk teams slow everything down because they don't have clear answers on where sensitive data lives or how it's being used. In short, your organization likely doesn't have an AI problem—it has a data problem.

To succeed, you must move away from the DIY debt of unvalidated systems and tightly coupled architectures that stifle scaling and waste expensive compute power. Your focus must shift toward unified, disaggregated, and governed environments that can feed the relentless demands of today's AI workloads.

Must-have #1: Data unification

A fragmented data estate spread across legacy storage, public clouds, databases, edge locations, and application-specific repositories creates bottlenecks

in preparing data for AI, slowing ingestion, limiting access, and driving inconsistent or biased model outputs. Fragmented estates are particularly bad for AI because they block the two things AI needs most: complete context and consistent truth.

By moving to a data fabric, you transform your fragmented estate from a liability to an asset. Instead of locked data, you have a seamless, intelligent data plane that allows AI to access information regardless of where it physically resides. It doesn't mean moving every byte into one giant database or repository, but rather, creating a virtualized logical layer that provides a single-pane-of-glass view. The result is an ecosystem where context is king. With data that is unified, the AI can draw connections that were previously invisible, moving from "knowing a fact" to "understanding a relationship." By providing a consistent, governed, and high-fidelity stream of information, your unified estate makes sure the AI can make complex, multi-step decisions with the full weight of your organization's intelligence behind it.

Must-have #2: Data governance

Lacking proper governance is risky because AI is not a traditional software program, but more like a "black box." Without oversight, you are essentially deploying a high-speed system that can make billions of unmonitored decisions every day. It destroys trust in the data, in the outputs, and in your ability to use AI safely, fairly, and repeatedly at scale.

Proper data governance starts with automatic metadata enrichment. By tagging data with context like provenance and sensitivity, you essentially create a "digital passport," so AI agents only interact with secure, verified, and compliant information. Proper governance looks like a continuous loop where every AI model is tracked in a centralized inventory, its data sources are

¹ "IDC Executive CIO QuickPoll Series: Operationalizing AI," IDC, November, 2024.

verified through automated lineage, and its decisions are monitored in real time for drift, or bias. Instead of manual checklists, you have automated guardrails in place—software that automatically masks sensitive data before the AI sees it and blocks outputs that violate legal or safety protocols.

Must-have #3: Performant, scalable, intelligent storage

Legacy storage that lacks performance, scalability, and intelligence essentially functions as a “digital warehouse” that throttles your AI potential. Data is trapped in rigid, tightly coupled hardware that leads to GPU starvation. Its scale-up design makes growth disruptive and expensive. And without internal intelligence, it provides no insight into the data it holds, forcing manual and error-prone preparation.

On the other hand, there’s storage built for the modern AI factory. This architecture is typically disaggregated, allowing compute and capacity to scale independently, so you never pay for resources you don’t use. It also leverages ultra-high-speed all-flash performance bolstered by cutting-edge innovations like remote data memory access (RDMA). Most importantly, this storage is data-aware, meaning it uses an intelligent software layer to automate governance, enrich metadata, and tier data across the edge and cloud simultaneously. Instead of a bottleneck, your storage becomes a strategic asset that keeps your AI models grounded in a real-time, high-fidelity, and fully governed single source of truth.

Must-have #4: Ecosystem validated for AI

Without an AI-validated ecosystem, your team must act as a hardware integrator instead of an AI innovator. Every driver update or hardware addition carries the risk of a cascading failure. Plus, hidden bottlenecks result in chronic GPU starvation and a significantly lower return on investment. And an unvalidated stack creates a governance vacuum that can lead to severe legal and operational consequences.

An ecosystem validated for AI is essentially a pre-certified, stress-tested blueprint where every component—from the GPUs and high-speed networking to the storage and software drivers—is guaranteed to work together seamlessly. Not only does this save your organization months of integration troubleshooting and GPU-starvation bottlenecks, but it also prevents the finger pointing that often occurs between hardware

and software vendors when something breaks. On the governance side, a validated stack typically comes with patterns for identity, access, data residency, backup/recovery, and observability that others have already used in production. This gives your risk and compliance teams more comfort because you’re not inventing controls from scratch around an experimental mix of tools.

How HPE solutions drive your success

HPE’s modern technology addresses the challenges of AI data preparation by replacing rigid, siloed systems with a flexible and intelligent architectural layer. To solve the issue of data fragmentation, our unified data software creates a logical fabric across your entire organization. Rather than undergoing the risky and expensive process of moving petabytes of information into a single location, HPE’s software connects disparate sources—from edge devices to cloud repositories—into a single source of truth. By automating identity resolution and metadata management, it ensures that AI agents have the full, real-time context they need to make accurate decisions without the delay of manual data cleaning.

By integrating an intelligent data layer and delivering real-time computing capabilities directly where data resides, HPE storage empowers you to enrich information seamlessly for AI pipelines, dramatically accelerating time to first inference. To eliminate performance bottlenecks like processor starvation, HPE’s disaggregated infrastructure decouples storage from compute, allowing each to scale independently. This approach uses ultra-fast, direct-access networking to bypass traditional hardware middlemen, ensuring that high performance GPUs are constantly fed with a high-velocity stream of data.

In addition, automated governance frameworks are embedded directly into the data pipeline. These tools provide always-on monitoring that tracks data lineage, masks sensitive information in real time, and enforces compliance guardrails. Lastly, by using HPE’s validated ecosystem solutions, your enterprise can bypass the technical debt of custom builds and move from pilot to full-scale production in a fraction of the time.

A smarter path forward

To operationalize AI at scale, your organization must break free from data silos, rigid infrastructure, and fragmented pipelines. HPE technology is engineered to remove bottlenecks, accelerate AI development, and give you confidence that your models are fueled by high-quality, well-governed data.

Learn more at

[HPE.com/data](https://hpe.com/data)



Visit [HPE.com](https://hpe.com)

[Chat now](#)

© Copyright 2026 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

a00155950ENW

HEWLETT PACKARD ENTERPRISE

hpe.com

