How to get your data AI-ready

AI has the potential to turn your business data into intelligence. But opening opportunities from such insight demands a polished, single source of truth. You need to build a solid data foundation to deliver truly effective AI. Here's how.

At a glance

1. Properly preparing your data foundation is a critical first step to AI success that businesses recognize, with 61% listing data management as their top area for AI investment.
2. However, low data maturity levels highlight a concerning reality. Businesses aren't as ready as they think they are, leading to potentially lackluster AI deployments with delayed delivery of business benefits or unusable results.
3. The secret to good data isn't a secret at all. If businesses first tackle data accessibility and storage, they'll be primed to support AI excellence.

Outdated. Incomplete. Hidden. Siloed. Mislabeled. Insecure. If any of these describe your company's data estate, your chances of AI success are slim. This is because AI output is only as good as your data input. You need a great stream of rich, relevant data. But you also need to ensure that this data is consistently and effectively organized, cleansed, stored, and accessed—if it's not, it won't matter how artfully coded your AI algorithm is. Building a strong data foundation is a critical first step to AI success.

In early 2024, Hewlett Packard Enterprise commissioned Sapio Research to survey businesses' AI journey. Findings from this research, **Architect an AI Advantage**,[1] reveal that many organizations believe they have moved past the initial AI-preparation stage. In fact, 72% consider their data AI-ready, with a similar number confident that it is also accessible and clean. As a result, 43% of IT leaders are already pushing AI out across their enterprises.

> "Unless you have a very diverse and historically rich set of data, your AI models will not be as credible or realistic."
>
> – **Healthcare manufacturing CISO, on AI limitations in the absence of good data**

[1] "Architect an AI Advantage," Sapio Research, January 2024

With AI implementation already in motion and 94% of businesses planning to increase AI investment over the next 12 months, leadership teams are eager to see quick results—unrealistically so, according to half of our surveyed IT leaders. And yet, these same respondents say they have already recorded some wins—for example, 42% have boosted operational efficiency and 29% have seen a reduction in errors.

The concern around expectations may stem from IT's understanding of AI's potential beyond operational streamlining. Implemented well, AI can transform business data into intelligence that strengthens business-led projects. Slow or elusive results on this front reflect the limitations of current AI deployments, and this signals that businesses need to step back and look at their data processes. Ultimately, if the data isn't properly prepared, AI rollouts are destined to be lackluster, causing delayed delivery of benefits or unusable results.

Despite organizations' high confidence in their data preparedness, a deeper look at our recent research reveals that the problem may be rooted in their **data maturity levels, which are surprisingly low:**

— A **mere 7%** of surveyed IT leaders can run real-time data pushes or pulls to drive innovation or external data monetization.

— **Only a quarter** have the means to run advanced analytics.

— **Fewer than 6 in 10** can handle all the key stages of data preparation—including accessing and storing.

The picture around data preparedness is crystallizing—businesses aren't as caught up as they need to be, and their AI projects will suffer as a result. So, what should organizations do to prepare?

Assess data accessibility

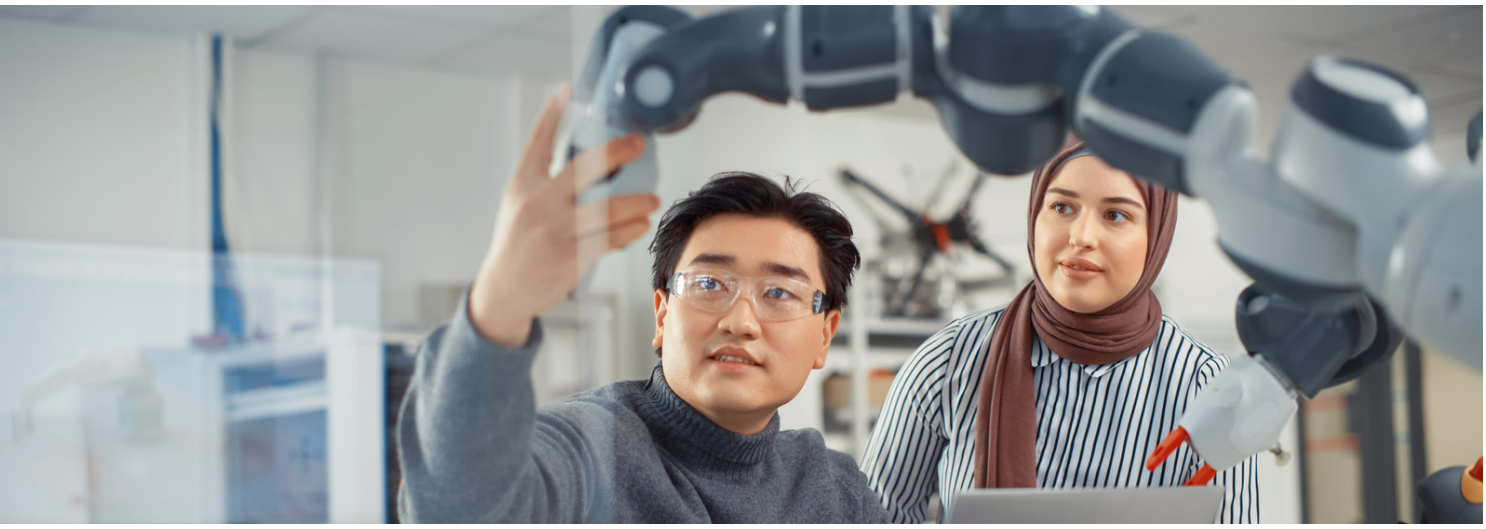
Organizations clearly understand the importance of data, and this is reflected in high investment levels. The question is, can enterprise data scientists find and access all sources of relevant data across the entire business to put it to good use? Better yet, can they do so in real time?

— Our latest HPE survey shows that **fewer than 4 in 10** (37%) IT leaders have set up shared data models with centralized business intelligence, meaning access to data will be unequal across the business.

— This mirrors findings from HPE's 2022 survey[2], where **34% of respondents** shared that their company's data was isolated in individual applications or locations, meaning data requests are at risk of returning incomplete results.

Right now, it appears that **organizational data is scattered across locations and applications**, between on-premises and cloud-based infrastructure. Businesses need to know where their data lives and how to pull it together. **Providing global data visibility across fragmented data sources** will help provide equal access to all data and effectively link related datasets for accurate push and pull requests. **Best practice requires an overarching data-first and AI by design architecture** that brings all data together across applications and locations. The goal is to provide unified access to the most up-to-date data across the organization, no matter where it resides.

One of the key learnings from the 2022 survey was that **removing data silos across hybrid architectures is critical to success**. This requires setting up a single source of truth for your data, but what does this mean? This is not about having all your data stored on a single server or database. Rather, it is about putting a supportive system and governance rules in place to ensure all data types and sources act in unison wherever they happen to be.



[2] "Lack of data capabilities impedes organizations' success, global survey finds," HPE, December 2022

This supportive data environment is **made possible through a hybrid cloud architecture**, where data can reside anywhere but still be treated as a unified, cohesive source. For this reason, it is imperative that data management decisions are made in parallel with decisions around enabling architecture.

Once your data is accessible as a consolidated source, the next step involves organizing data that your business has classified as relevant. In the beginning stages of organization, companies are encouraged to take a **people-led approach to data governance**. Software that automates this process will be essential down the line, but in the initial stages, it won't be capable of discerning what data is relevant to your company's unique business use cases. Here, the team behind data preparation will therefore be of utmost importance, with members like your chief technology officer, chief data officer, or chief data and AI officer, leading the charge on the relevance of data.

"Historically, every department of the bank had their own solutions, their own data warehouses, and databases. The power of AI really comes when we can pull that data together and look at insights for a customer across all of our products."

– Financial services LOB executive, on unifying data to fully realize AI value

Consider AI-specific data storage

A more deliberate approach to data storage must be implemented with AI in mind. This is because legacy data-storage systems don't support modern organizations' end goal of running AI workloads.

The **capacity and performance of any storage solution are key metrics, and both must be scalable**. Data consolidation and organization are capacity-focused tasks while activities such as training AI with data are performance reliant. These metrics are intertwined, meaning any AI data storage solution should be an investment that not only considers capacity but also supports eventual performance-centric activities using the data that is stored.

A cap on scalability is a serious challenge in today's storage landscape, with many legacy solutions able to scale storage capacity just fine, but unable to scale performance in parallel as data volumes skyrocket. Unfortunately, when it comes to AI data storage, performance-at-scale limitations are the norm for legacy scale-out NAS solutions built on **shared-nothing** architectures. **What's needed is a highly resilient storage solution built on a disaggregated, "shared-everything," modular architecture** that allows you to scale performance and capacity independently.

The necessary solution should also be **designed to handle data at exabyte scale**, and this is where **hybrid again stands out as the most flexible and scalable of architectures**. Hybrid cloud can easily spread out your asset storage wherever data is generated or ingested—whether it's from an on-premises private cloud or the edge. This exemplifies scalability.

For storage needs, the architecture you choose makes a difference. Hybrid and private cloud architectures are better suited to storage requirements than the **public cloud, where businesses would need to contend with latency**. IT leaders generally recognize this, with a third choosing to store their AI data in a private cloud and a similar number (30%) opting for hybrid architecture to meet this need. The remaining 36% chose other data housing options, such as on-prem, public cloud, and edge.

Ensure continued momentum

Once data is consolidated and organized, and scalable storage of it is accessible, businesses can focus on **data protection** as well as **automation to keep their data in good order**.

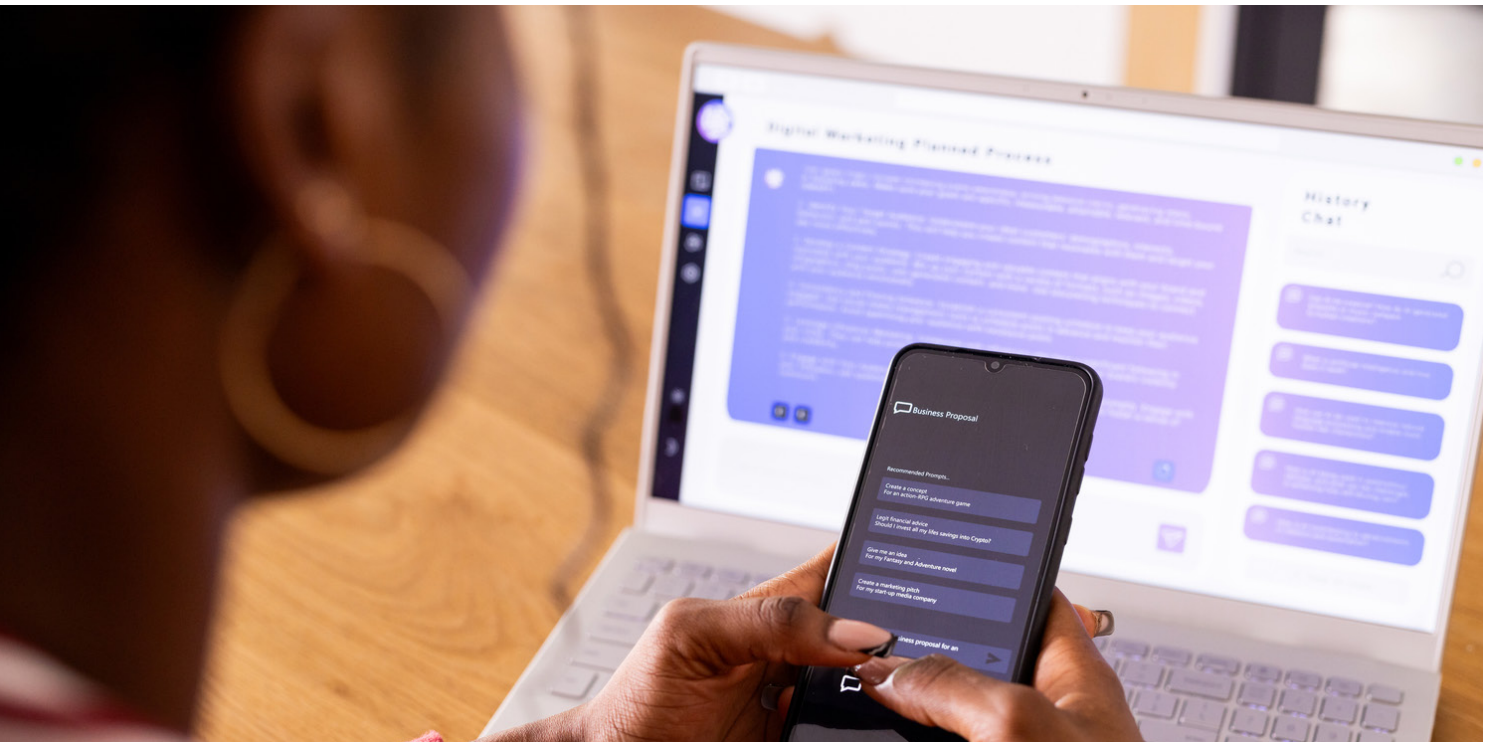
The ability to secure AI datasets is a big concern for 73% of IT leaders. The question here is whether effective data management can help mitigate risk, and learnings from research[3] commissioned by HPE in 2023 show it can. Findings revealed that data-first businesses are **better able to mitigate risk** and that they are 3.5x more likely to entirely automate IT tasks like protection.

To continue tapping into the benefits of automation, businesses might also consider employing AI to help with data cleansing, where algorithms make quick work of finding and addressing inconsistencies and duplication. Establishing smart, automated data-governance policies can continue the task of ensuring data is properly classified and sorted as it's collected—and, perhaps more crucially, before it is used to train AI models to harness intelligence insights.

With the starting steps outlined, and the path ahead illuminated, businesses are set to make the most of an AI architecture that is supported by a **data-first approach**. Reiterating the importance of data, the 2023 research also found that data-first leaders are 13.5x more likely to beat their competitors to market. The takeaway is clear: **data excellence is pivotal to the success or failure of a business in the age of AI**.

"There are laws that patient data from this country cannot be stored in any other country or in certain countries. So, then there will be layers of roles for who can access patient data and who cannot. You don't want to do it one-off, so processes are put in place to determine access going forward."

– **Manufacturing medical devices data scientist, on automating processes for secure data access**



About HPE

HPE is the edge-to-cloud company that helps organizations accelerate outcomes by unlocking value from all of their data, everywhere. Built on decades of reimagining the future and innovating to advance the way people live and work, HPE delivers unique, open, and intelligent technology solutions, with a consistent experience across all clouds and edges, to help customers develop new business models, engage in new ways, and increase operational performance.

Learn more at

HPE.com/data

Visit HPE.com

Chat now

a50010470ENW, Rev. 1

HEWLETT PACKARD ENTERPRISE

hpe.com