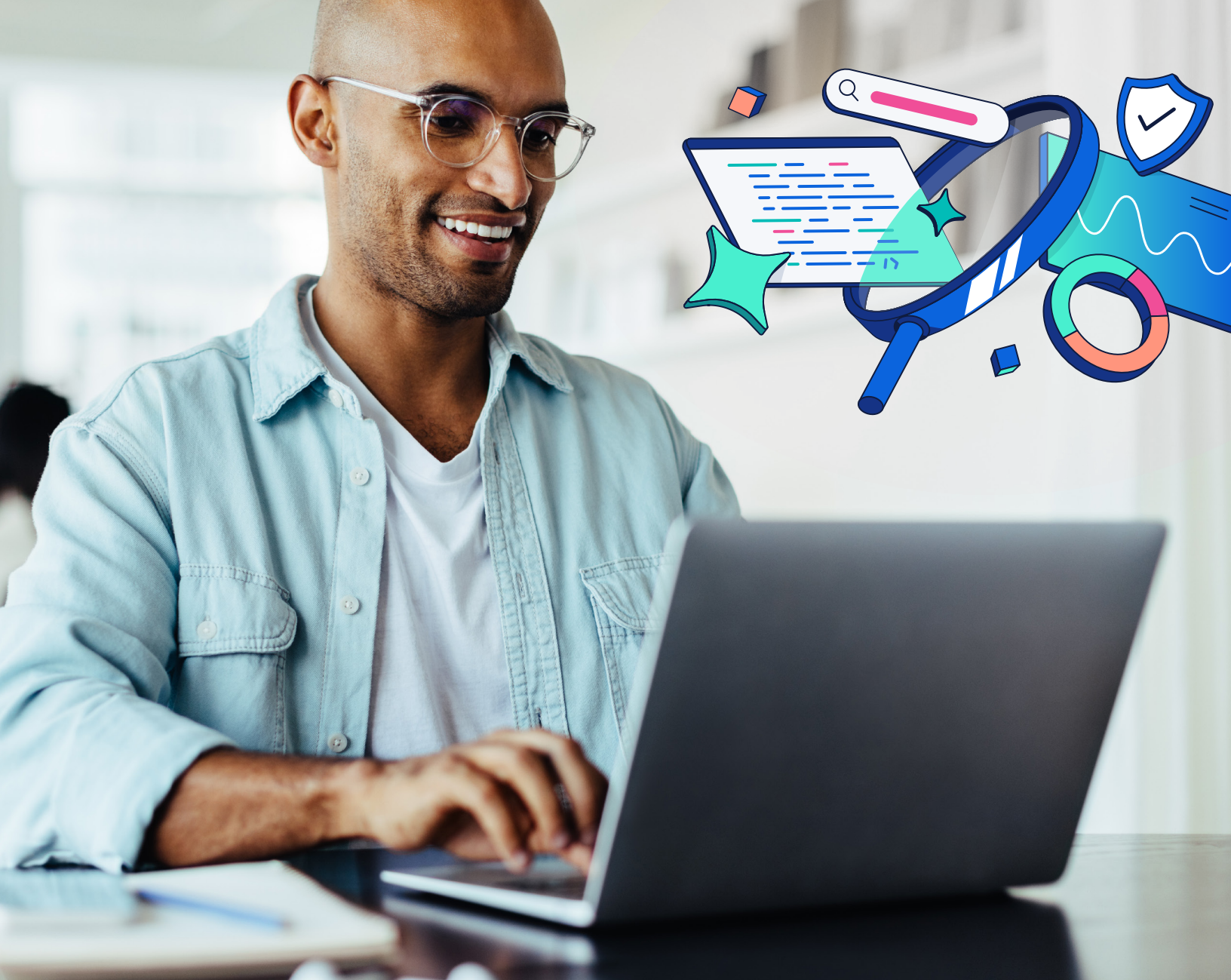




Accelerating AI at scale: Elastic and Dell AI Data Platform with NVIDIA

Table of contents

Executive summary	3
The road to AI readiness	4
The hidden value of unstructured data	5
Enabling enterprise AI adoption	
Closing the context gap	
From unstructured data to the DELL AI Data Platform	
AI search at scale	6
How to apply AI in a business context	
Sovereign AI: The secure enterprise solution	
How to use AI for your business	
The Dell AI Data Platform	8

Executive summary

Enterprise AI is full of promise — and full of stalled pilots. However, the gap between ambition and production often lies with data, not models.

Most organizations are sitting on vast reserves of unstructured data — documents, contracts, logs, and call transcripts — that their AI can't search or reason from. This results in models that miss context and produce unreliable outputs. Over time, AI initiatives can lose the trust of the teams they're meant to support.

The solution

The Dell AI Data Platform, with search powered by Elastic on NVIDIA accelerated computing, gives enterprises the foundation to turn that data into a strategic asset:



Turn data into actionable insights

One engine handles vector, keyword, and hybrid search across all your data, enabling your teams and AI agents to get the right answer in the right business context every time.



Deploy and scale in days, not quarters

A validated, petabyte-ready infrastructure stack grows from one use case to enterprise-wide without rebuilding.



Own where your data lives

The Dell AI Data Platform runs in your environment, under your controls, keeping regulated and proprietary data on infrastructure you govern.

The road to AI readiness

Enterprise organizations are investing heavily in AI. In fact, after finding success from chatbots and other simple AI activations, many organizations feel ready to scale. AI ambition is clear: Enterprises are seeking efficiency, automation, and a boost to their bottom line. However, across industries, organizations are discovering that AI investment doesn't necessarily guarantee meaningful outcomes. AI pilots may look promising, and industry-specific use cases make sense in theory, but moving from proof of concept to production often exposes underlying gaps.

Why do so many AI readiness strategies fail? The answer is: a weak data foundation.

Too many enterprise organizations treat data as a passive resource. Many organizations assume they can simply plug existing data into AI systems and see immediate results. In the real world, too much unactionable data leads to what's known as [context rot](#).

AI operations that scale successfully understand the importance of a strong data foundation.

90%

of enterprise data is unstructured, including documents, emails, contracts, call transcripts, logs, and more, according to IDC.¹

Unstructured doesn't feed cleanly into a model. Without a dedicated engine to find, surface, and contextualize it, that data goes unused — or worse, actively undermines AI performance.

Most AI readiness plans skip a foundational step: building the infrastructure that turns raw, unorganized data into something a large language model (LLM) can actually work with. In the AI gold rush, organizations often feel pressure to deploy new models and agents immediately.

However, AI tools are only as effective as the data environment powering them.

Skip that step, and AI may:



**Confidently produce
wrong answers**



**Miss critical
context**



**Degrade trust among the
people it's meant to help**

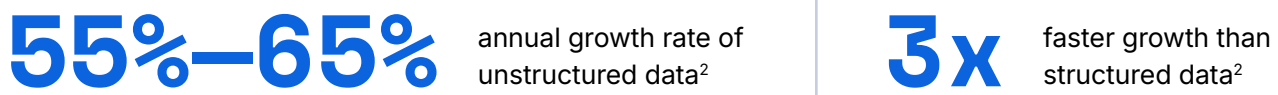
The Dell AI Data Platform — powered by Elastic's data search engine and NVIDIA accelerated computing — is designed to transform raw, unstructured data into AI-ready context.

The hidden value of unstructured data

Your AI is only as good as the data it can search and retrieve.

While this conclusion may seem obvious, the implications run deep — and it's a necessary shift in framing for how most IT leaders think about AI. Even the most advanced LLM will deliver mediocre results if it sits on top of a weak data foundation. In effect, this means that **your competitive advantage in AI depends less on the model itself and more on the infrastructure supporting it**. It also means that the gap between enterprise AI success and failure is less about what AI you buy or how quickly you deploy it. Rather, AI success depends on whether your organization can operationalize its data.

This is no small challenge as enterprise data volumes compound. According to research from Gartner:



Organizations that proactively manage unstructured data will create long-term value, while those that continue to ignore it will fall behind due to model inadequacies and a flood of unfindable information.

Enabling enterprise AI adoption

How do enterprise organizations actually [operationalize all this data](#)? Currently, most are only using a fraction. Structured data, which lives in databases and data warehouses, accounts for only a small minority of total enterprise data. The insights that could differentiate an enterprise AI deployment are largely locked in the other 90% of unstructured data that doesn't fit neatly into existing systems.

As a consequence, AI only sees part of the picture. It answers based on what it can access instead of what actually exists. That creates performance bottlenecks, heightened operational costs, and AI deployments that miss context altogether.

As organizations transition from pilot to production, the problem only becomes more difficult to address. An AI agent that misses critical business context in a live workflow is a liability. These stakes are driving IT leaders to reevaluate their data infrastructure from the ground up.

Closing the context gap

What really happens when an LLM has access to data that is outdated, redundant, or irrelevant to the query at hand? This is where context rot enters the picture in the real world.

Context rot occurs when there's too much noise and the signal disappears, bogging the model down. It can't distinguish what matters from what doesn't. Accuracy drops, yet confidence remains high. Teams end up with answers that sound authoritative but miss the mark.

Solving context rot requires a discipline called [context engineering](#), or the deliberate process of designing, structuring, and maintaining the data pipelines that feed AI systems. Context engineering ensures that what reaches the LLM is the right information, correctly formatted, at the right time — not the full archive of everything an organization has ever stored.



Context engineering

Designing, structuring, and maintaining the data pipelines that feed AI systems

[Learn more](#)

From unstructured data to the Dell AI Data Platform

Unstructured data is usually treated as a problem to be managed down, whether that's archiving, compressing, or culling. However, significant business value lives within this unstructured data. Data sources like customer conversations, internal research, institutional knowledge, and operational logs provide the context AI needs to reason well. Overlooking this data leads to AI deployments that are technically functional but lack essential strategic context, risking context rot.

To scale effectively, organizations need infrastructure that can tame unstructured data and a context engineering strategy that delivers the right information to LLMs in the right format. With a proper data foundation and AI readiness strategy, the information that ultimately reaches the model is relevant and properly structured.

AI search at scale

As a business scales, the volume of data it generates grows exponentially. And with it, so do storage costs. To see a return on investment, all this data must be operationalizable — and AI working from stale context delivers very little business value.



The biggest challenge

Making that data searchable for an intelligent system to find the right context across petabytes of distributed, siloed, heterogeneous data in real time.



The underlying problem?

Most businesses' processes are not built to index new data, unify access, and deliver consistent performance as workloads scale.

To elevate their operations, teams are seeking solutions to analyze data, derive insights, and scale without constant manual intervention. They need instantaneous intelligence that doesn't break the bank.

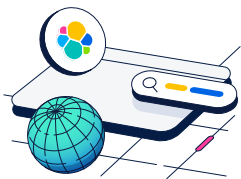
How to apply AI in a business context

Moving past the pilot stage means preventing tool sprawl with random, disconnected AI deployments. Without a sophisticated data foundation and a cohesive AI strategy, businesses are likely playing in the shallow end of their data rather than using their data repositories to their full potential.

Many organizations are seeking solutions to take advantage of business-specific context in their AI deployments. But using business context for AI tools goes beyond just plugging AI into existing protocols. It requires specific foundations that allow for scale and data sovereignty.

To operationalize unstructured data for AI, organizations must invest in continuous indexing and a vector retrieval engine to automate data preparation and simplify access to enterprise data.

Key concepts:



Vector database

A vector database stores data as embeddings that ascertain meaning rather than exact text. A [vector database](#) is the foundation of [semantic search](#), helping AI systems surface content conceptually related to the query, even if it doesn't share the exact same keywords. For instance, a query for "revenue impact of supply chain delays" returns relevant results even if no document uses those exact words.



Semantic search

Whereas traditional keyword search matches exact terms, [semantic search](#) understands intent. It finds the result that answers the question, not just the result that contains the same words. Alongside keyword precision and metadata filtering, AI-driven semantic search is how enterprise AI moves from approximate to accurate retrieval.

Sovereign AI: The secure enterprise solution

Sovereign AI is emerging alongside a strategic emphasis on data governance for regulated industries and any organization that handles sensitive data.

The compliance pressure is becoming more intense for cybersecurity, regulations, and customer obligations. The EU AI Act entered into force in August 2024 and began phasing in critical obligations as of February 2025.³ This measure joins a widening field of regulations with direct implications for AI deployment, including GDPR, India's Digital Personal Data Protection Act, and China's PIPL.⁴ Emerging regulations further increase complexity for enterprises operating across jurisdictions.



Data sovereignty concerns are no longer exclusively the domain of financial services or healthcare organizations. Customer trust, IP protection, and long-term risk management are driving organizations to keep proprietary data in-house regardless of any specific mandate. Every AI query is now a potential data liability.

This risk increases as data volumes scale. A secure sovereign AI system maintains data integrity even as businesses grow. An AI tool's ability to secure business-specific context is a must in the new data landscape.

An on-premises AI stack can meet the demands of data sovereignty.

The Dell AI Data Platform gives organizations full control over their data and models, meeting strict regulatory requirements without sacrificing performance.



How to use AI for your business

For IT leaders ready to scale, deployment must be strategic and data-first. The practical path to using AI in a business context happens in three steps:

- STEP**  Audit your current unstructured data landscape and ask how much business value is invisible to your AI.
- STEP**  Identify the gaps between the data your models can access and the data that actually drives decisions.
- STEP**  Evaluate whether your current stack can scale without a linear increase in cost and complexity. If the answer is no, focus on what an integrated alternative would look like.

The Dell AI Data Platform

Enterprise AI accelerated by NVIDIA and Elastic

For IT leaders, the Dell AI Data Platform offers a new way to stand up a production-ready AI stack in days, all without stitching together components from multiple vendors or rebuilding your infrastructure every time a use case changes.

The Dell AI Data Platform decouples storage from intelligence. Dell PowerScale and ObjectScale handle data at petabyte scale, while Elastic transforms that data into searchable, AI-ready context.

NVIDIA cuVS COMPARED
TO CPU-ONLY APPROACHES

12x faster indexing throughput
than CPU-only approaches

NVIDIA cuVS COMPARED
TO CPU-ONLY APPROACHES
ON A COST-ADJUSTED BASIS

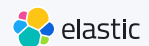
5x higher indexing
throughput

6x faster force
merging

NVIDIA cuVS accelerates the whole stack, delivering indexing throughput nearly 12 times faster than CPU-only approaches. On a cost-adjusted basis, that translates to approximately five times higher indexing throughput and six times faster force merging, meaning faster time to production without proportionally adding to infrastructure spend.

Elastic's hybrid search capability sits at the heart of the stack. Rather than choosing between keyword matching and semantic understanding, [hybrid search](#) combines both to deliver precision and contextual depth.

Search at scale



Elastic is the data search engine at the core of the Dell AI Data Platform. It handles vector, keyword, and hybrid search across all your enterprise data in a single index, giving AI agents and business teams the power to surface accurate, context-rich answers without jumping between tools or systems.

GPU-accelerated performance



NVIDIA cuVS powers the compute layer, accelerating indexing and retrieval so your infrastructure keeps pace with your data. The result is faster time to production, more throughput per node, and an AI stack that scales without a proportional increase in hardware spend.

The unified foundation



Dell PowerScale and ObjectScale provide the storage backbone — purpose-built for AI workloads at petabyte scale. The AI Data Platform bundles infrastructure, compute, and search into one validated, Dell-sold stack so your team can deploy faster and buy from a single vendor.



Unlocking the intelligence layer with Elastic

As businesses look beyond AI pilots and prepare to scale, Elastic is the layer that elevates the stack from a data repository into an agile AI system.

Elastic makes sense of unstructured data so that LLMs stay grounded in relevant context. It uses NVIDIA accelerated computing to execute hybrid search at a scale that traditional databases simply can't reach. And it does this continuously, indexing new data in real time to ensure that AI reflects the current state of business.

As AI evolves to take on autonomous action, the organizations that lead will be those whose data is searchable, context-rich, and secured within their own environments. The Dell AI Data Platform integrates Elastic as the data search engine and runs with NVIDIA accelerated computing to ensure it is built to compound that advantage over time.



Ready to get started?

Schedule a [data-readiness call](#) and explore our offerings with [NVIDIA](#) and [Dell](#).

[Schedule now](#)

¹ IDC, "Worldwide Global DataSphere Forecast, 2024–2028: AI Everywhere, But Upsurge in Data Will Take Time," May 2024; IDC, "The Untapped Value of Unstructured Data," August 2023.

² Athento, "Unstructured information: the great challenge (and opportunity) for companies in 2025," June 2025.

³ Future of Life Institute, "The EU Artificial Intelligence Act."

⁴ Protecto, "Data Sovereignty in the Age of AI: Why It Matters and How to Get It Right," October 2025.