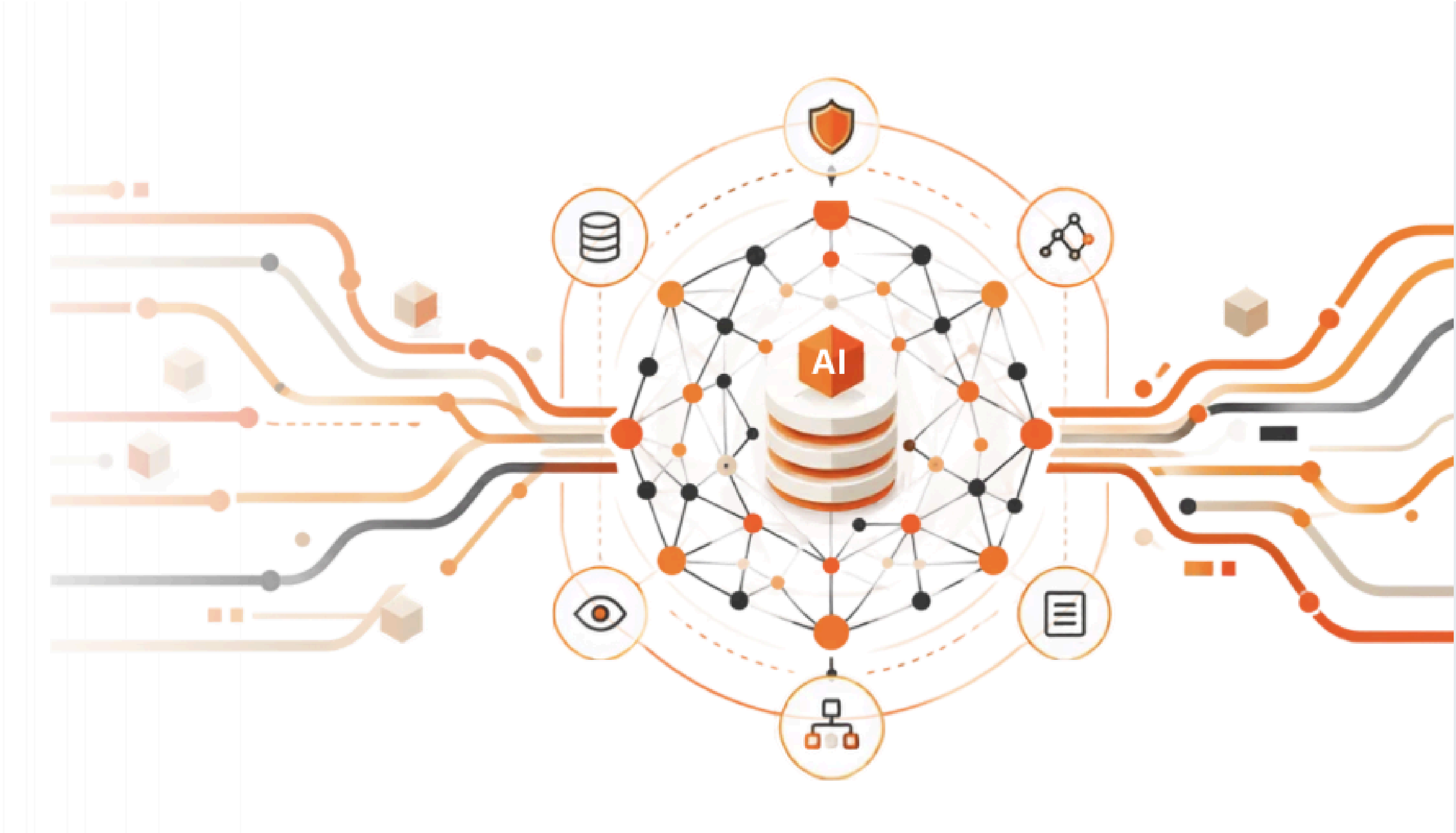
Modern Data 101 Community

September 2026

The State of Data Architecture for Enterprise AI



The data platform decisions for converting AI experimentation into sustained, enterprise-wide value.

Modern Data 101 Community

Table Of Contents

1

Executive Summary

From The Modern Data 101 Community

This issue is packed with findings from industry leaders like Deloitte, MIT Sloan, MIT Sloan Executive Education, Gartner, Snowflake, CIO.com, The Modern Data Company and Datamatics.

AI has moved out of the pilot stage, but most initiatives that succeed in a controlled pilot fail to sustain value at scale. Research from Deloitte, MIT, Gartner, Snowflake, and a new 540-practitioner survey from the Modern Data 101 Community points to one consistent cause: not model quality, but data architecture, the platform, governance, and design choices that determine whether data can reach the model at all.

This brief distils that research into what actually separates organisations converting AI into durable advantage from those stuck in perpetual pilot mode, and what the best data platform architecture looks like in practice: fixing data discovery and context before adding capability, engineering data liquidity deliberately, building governance into pipelines rather than retrofitting it, and pairing all of it with literate leadership.

2

The AI Paradox

Enterprise appetite for AI is no longer in question. CIO.com's 2025 State of the CIO survey found data-enabled AI is now the top CEO priority for IT organisations. But Deloitte's research frames the resulting execution gap as an "AI paradox": most AI pilots succeed in controlled testing, yet only a small fraction go on to create sustainable value at scale. A joint MIT Technology Review Insights and Snowflake survey of 275+ leaders quantifies why: most businesses don't yet have a data foundation genuinely ready for generative AI.

Few of those businesses have identified, let alone built, the best data platform architecture the moment demands, which is exactly the gap the rest of this brief works through.

78% of businesses lack a data foundation that is "very ready" to support generative AI. Source: MIT Technology Review Insights & Snowflake

3

The Data Activation Gap

The Modern Data 101 Community's 2026 survey of 540+ practitioners across 64 countries, The Modern Data Report 2026: The Data Activation Gap, confirms the industry pattern from the practitioner's chair. Most respondents already work with leading platforms and AI features; the constraint is activation.

Discovery, context, and trust

Finding the right data is now the single biggest time sink: 89% cite it as a top-three time drain, and 34% rank it their number-one activity, driven by too many data versions (71%), missing or outdated documentation (67%), and no discovery tooling (60%). Even when data is found, 57% struggle to interpret it because context and definitions don't travel with it, and 46% say they cannot fully trust their data for business decisions, while 84% hit conflicting versions of the same metric regularly. For AI specifically, 68% say their data isn't clean or trustworthy enough for AI operations, a paradox, since only 30% currently use AI itself to help close that gap.

The
Data Activation Gap

The cost, and the fix practitioners want

The cost is measurable: 66% report productivity loss from constant reconciliation, 54% cite inflated costs, and 47% link poor data directly to lost revenue. What practitioners want in response isn't more tools; it's convergence: 80% rank a semantic layer with standardised definitions as the top AI enabler, and 77% say a converged, well-governed data platform would be valuable, with 40% calling it "extremely valuable". Read together, those preferences are practitioners describing the best data platform they don't yet have.

4

Data Architecture and the Best Data Platform

Awareness of the data-AI relationship isn't the issue; most organisations already invest in data, but often toward goals that predate AI, like compliance and reporting. What's changed is the speed and autonomy with which AI now consumes that data, raising the bar on timeliness, traceability, and access control. Gartner research cited by Datamatics estimates enterprises will abandon up to 60% of AI projects through 2026 due to poor data quality alone.

Up to 60% of AI projects will be abandoned by enterprises through 2026 due to poor data quality. Source: Gartner, cited by Datamatics

This plays out consistently across leading digital organisations, per CIO.com: HPE built its internal AI hub on a dedicated private cloud so sensitive use cases never leave a controlled environment, while Virgin Atlantic consolidated scattered databases specifically to cut the time analysts spent reconciling data before analysis could even begin. In both cases, AI capability followed data architecture.

Data Architecture and the Best Data Platform

So what does that architecture look like? Datamatics' research breaks the best data platform's architecture into five layers that must each work before the next adds value:

- Ingestion, batch, streaming, API, and change-data-capture combined; Gartner forecasts streaming adoption for agentic AI will exceed 60% of organisations by 2028, up from under 15% in 2025.
- Refinement, a tiered "medallion" structure moving raw data through validation into curated, AI-ready datasets.
- Orchestration, pipelines modelled as dependency graphs, so a failure upstream halts bad data before it spreads.
- A semantic layer, standardised definitions so every dashboard, model, and AI agent means the same thing by a metric, plus reverse ETL to push insights back into daily tools.
- Governance and observability, built continuously into the pipeline, instead of being audited after the fact.
- This is what separates the best data platform from a merely capable one instead of feature count, but whether these five layers are engineered together.

5

Data Liquidity as the Foundation

MIT's Center for Information Systems Research (MIT CISR) introduces data liquidity, how easily data can be reused and recombined across the business. Organisations with high liquidity outperform peers on customer experience, speed to market, and decision-making, but it doesn't happen automatically. A multiyear study of Caterpillar identifies three deliberate levers:

1. Architecture designed for reuse: a modular platform with clearly owned stages from raw ingestion to stable, reusable master data.
2. Prepared, quality-checked master data, with an explicit data-quality function catching and routing problem records.
3. Least-privilege permissioning, access scoped to the task, with entitlements transparent to the people who need them.

Data liquidity is a managerial discipline; organisations that treat data as a reusable asset compound the value of every subsequent AI investment. It's also what turns a capable platform into the best data platform for AI: it is seldom the number of features it ships, but how freely value moves through it.

6

Trust, Governance, and the Economics of Scale

Adoption still depends on trust and cost. The MIT Technology Review Insights and Snowflake survey found 59% of leaders rank governance, security, or privacy above every other AI consideration, ahead of accuracy or cost, while 72% still want AI to improve efficiency, reflecting a dual mandate to be both trustworthy and transformative.

59% of leaders prioritise data governance, security, or privacy above all other AI considerations. Source: MIT Technology Review Insights & Snowflake

On cost, Gartner projects that by 2030, inference on a trillion-parameter model will cost over 90% less than in 2025. But agentic AI needs 5–30x more tokens per task than a standard chatbot, so falling per-token prices won't offset rising total spend, organisations that architect efficiently, routing routine tasks to smaller models, will still outspend inefficient rivals that treat every task as a frontier-model problem. That routing itself depends on the same governance and observability layer covered in Section 3, precisely why the best data platform treats governance as infrastructure, built in from the start, rather than bolted on after an incident.

7

AI Literacy as Infrastructure

Technology and governance aren't sufficient alone. MIT Sloan Executive Education's research argues AI literacy- judging where AI creates value, where it introduces risk, and when human judgment should override it, is now a leadership competency. McKinsey's State of AI research, cited in that work, finds 88% of organisations already use AI regularly in at least one business function, so AI-informed decisions now touch hiring, pricing, and customer engagement broadly.
As AI takes on more of the data work itself, human judgment doesn't disappear; it moves up the stack, into defining intent, setting guardrails, and staying accountable for outcomes. Even the best data platform still needs literate leadership to interpret what it produces.

8

Strategic Imperatives for the Best Data Platforms

1. Fix discovery before adding capability. With 89% of practitioners naming discovery their top time drain, documentation and naming conventions do more for AI readiness than another analytics feature.
2. Treat data architecture as reusable infrastructure. Modular, service-oriented platforms are the discipline behind building the best data platform for AI, letting new use cases draw on existing, governed data instead of bespoke integration each time.
3. Engineer data liquidity deliberately, architecture, preparation, and permissioning are management disciplines, and not a byproduct of a platform purchase.
4. Build governance into pipelines continuously. Embedded quality checks and observability cost less than retrofitting after an incident, and are what practitioners say would help AI enablement fastest.
5. Pair the technology stack with AI literacy at the leadership level, so judgment and evaluation criteria scale as fast as the architecture does.

9

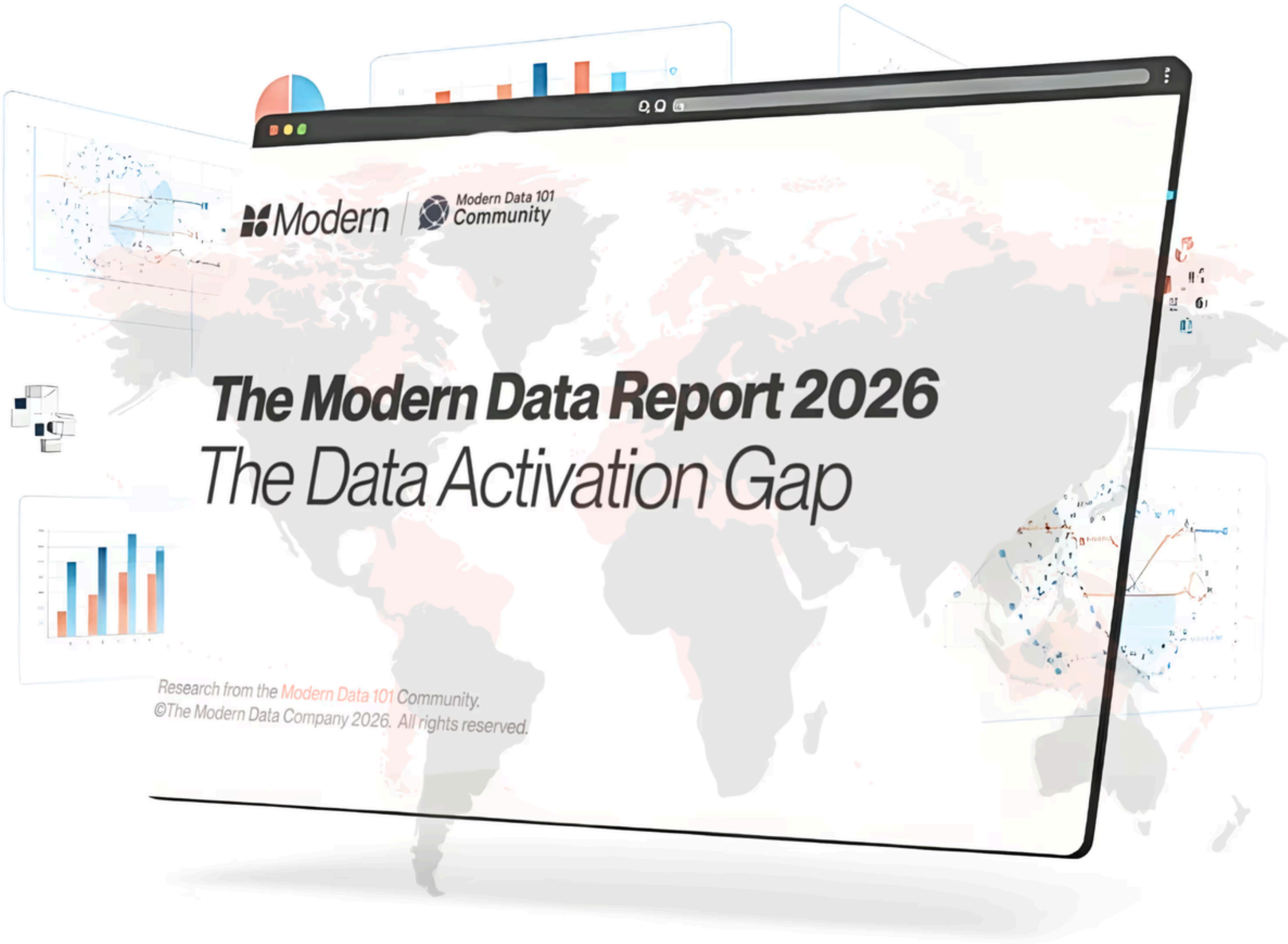
Conclusion

The organisations that lead the next phase of enterprise AI won't be the ones running the most experiments, across Deloitte, MIT, Gartner, Snowflake, and the Modern Data 101 Community's report, the same conclusion holds: they'll be the ones whose data architecture, platform, liquidity, governance, and literate leadership can operate at the speed of the AI systems built on it. Scaling AI means scaling the best data platform architecture beneath it first.

References

1. The AI Paradox: Why Your Data Foundation Matters Most. Deloitte Luxembourg, April 2026. https://www.deloitte.com/lu/en/our-thinking/future-of-advice/the-ai-paradox.html
2. The Modern Data Report 2026: The Data Activation Gap. Modern Data 101 Community, survey of 540+ practitioners across 64 countries, 2026.
3. How to Build Data Foundations for AI Exploration. CIO.com, August 2025.
4. How Data Moves Through a Modern Enterprise Platform: From Raw Data to AI-Driven Decisions. Datamatics Global Services, July 2026.
5. Building a Data Foundation for AI Excellence. Harvard Business Review, sponsored content from Avalara, March 2026.
6. Data Liquidity Leads to AI Success. MIT Sloan School of Management, June 2026.
7. How Solid Data Strategies Are Fueling Generative AI Innovation. Snowflake, in partnership with MIT Technology Review Insights
8. AI Literacy: A Key Piece of an Executive's Skill Set. MIT Sloan Executive Education, February 2026.
9. Gartner Predicts That by 2030, Performing Inference on an LLM With 1 Trillion Parameters Will Cost GenAI Providers Over 90% Less Than in 2025. Gartner, Inc., March 2026.

Modern Data 101 Community



540+ Respondents, 64 Countries, 29 Industries

Access the report now

Modern Data 101 Community

Subscribe

moderndata101.com