Modern™ WhitePaper

Context for Enterprise AI:

A DataOS® Approach

What it takes for AI to understand, trust, and act on the data it relies on

Table of Contents

Executive Summary

In 2023, researchers at UC Berkeley published *MemGPT: Towards LLMs as Operating Systems*, proposing a useful way to think about how AI manages information. A model uses a small workspace to work on immediate requests but has the capacity to draw on more content as needed, much like a few files on a desk, backed by a much larger filing cabinet.

The analogy points to a problem enterprise AI still faces. Giving AI access to more information is not enough. It needs a reliable way to find the right information, understand what it means, and bring the right context into each task.

Context is what tells an AI system what data means, how it relates to other data, where it came from, and how it may be used. In the enterprise, that includes business meaning, relationships, quality, provenance and lineage, governance, and operational state.

Research increasingly shows the consequences when that context is missing or difficult to use. In a Stanford-led study, researchers found that performance was generally strongest when relevant information appeared at the beginning or end of a long context and declined when it was placed in the middle. In multi-document tests, GPT-3.5-Turbo's performance dropped by more than 20% and, in some cases, fell below its performance with no supporting documents at all.

Our own research shows the same gap inside the enterprise. In *5 Emerging Trends in Enterprise Data & AI*, based on early findings from our third Modern Data Survey, 61% of more than 500 data practitioners say reliable business context is critical or very important for their AI agents, yet only 16% say their organization has engineered that context as a deliberate, managed product. A Gartner projection points to the consequences: by 2028, 60% of agentic analytics projects relying solely on the Model Context Protocol will fail because they lack a consistent semantic layer.

This paper presents a DataOS approach to making context operational at the data layer. DataOS works across existing enterprise data infrastructure and platforms to connect context to data in reusable data products, where it can be maintained and used consistently across AI, applications, and analytics.

We'll show how this works in practice, including how one Fortune 500 manufacturer used DataOS to get a GenAI use case into production 80% faster.

Making Context Operational

Enterprise context rarely lives in one place. Business definitions may sit in catalogs and semantic layers. Relationships may be represented elsewhere, while quality rules, governance policies, lineage and operational state live in still other systems. AI may have access to the data without receiving these different parts of its meaning together.

Making context usable requires that meaning to be represented in a consistent, machine-readable form. Catalogs document it. Semantic layers standardize it. Knowledge graphs connect it. Ontologies define the concepts and relationships that give the data its business-specific meaning.

The question is how much of that meaning must be formally defined. Modern's co-founder and CTO, Animesh Kumar, addresses this in his discussion of the minimal ontology principle in *The Ontology Engineering Challenge is Not Building It, but Mirroring the Delta*. His argument is that **organizations do not need to define every business concept from scratch. They need to identify and encode the places where their meaning differs from what a model would otherwise assume.**

For data leaders, the larger challenge is connecting these components so context remains consistent, stays attached to the underlying data and reaches AI at the point of use. This becomes more demanding as agents move from interpreting information to acting on it. An agent must understand which actions are permitted, under what conditions they may occur and how changes to operational systems should be governed.

The enterprise challenge is therefore not simply to create context, but to maintain and deliver it consistently across models, applications and use cases.

Context Belongs with the Data 01

Keeping context with the data means preserving more than its current definitions and metadata. The meaning, rules and operating conditions that applied when the data was used must also remain connected to it.

Consider a financial institution reviewing a lending decision several months after it was made. The decision log may still exist, along with the applicant's credit file, pricing tables, dataset documentation and lending policies. Yet reconstructing the decision can require significant investigation:

- Which pricing table was active?
- Which source supplied the debt-to-income ratio?
- Which version of the lending policy was applied?
- Did the underlying data meet its quality requirements?
- What information was available to the model or agent at the time?

The individual records may all be available without providing a reliable account of how they came together in that decision.

This distinction matters whenever AI participates in consequential business processes. **The context needed to understand a decision includes more than what the data means today. It also includes what it meant, what rules applied, and what condition the data was in at the time it was used.**

The context connected to enterprise data has two broad components:

Business context establishes what the data means and how it should be interpreted. It includes definitions, calculations, relationships among entities, applicable policies and permitted uses.

Operational context establishes the condition and history of the data when it was used. It includes provenance, lineage, quality, freshness, pipeline status, SLA performance and other evidence that the data was valid for its intended purpose.

Time cuts across both. Business meaning evolves, relationships change, policies are revised and operational conditions shift. A metric may be calculated differently in 2027 than it was in 2026. Preserving only its current definition would erase part of the context needed to interpret earlier results.

These components become useful when they remain connected to one another and to the data itself. A fulfillment agent should be able to determine what "available inventory" means, whether the inventory data is current and whether it is permitted to commit that inventory without reconstructing those answers across separate systems.

Context therefore needs to remain attached to the data throughout its use, including the meaning, policies and operational conditions that applied at a specific point in time.

The Data Product Brings Data and Context Together

02

If context belongs with the data, the next question is how to keep them connected.

DataOS uses the data product as the authoritative unit for bringing data and context together. In DataOS, a data product connects enterprise data with the context required to understand, govern, trust and consume it. Business meaning, relationships, quality, provenance and lineage, governance, and operational state become part of the product rather than separate information that every consumer has to assemble independently.

The meaning, quality and governance are not bolted on afterward by whoever happens to be using the data. They are already part of the product.

Take customer profitability. The calculation may combine transaction data, product costs, discounts, returns, service costs and account information from several systems. Finance determines what profitability means. Without that definition, an AI system may apply a generic understanding of profitability rather than Finance's approved calculation. Data teams determine which sources and transformations produce it. Governance determines who can access customer-level information. Quality rules establish acceptable freshness and completeness.

Building a customer profitability data product brings that context together with the data.

Now consider what happens when the organization introduces a new AI application that needs profitability data. The team does not have to start by determining which revenue field to use, rediscovering the calculation, mapping customers to accounts, establishing source lineage and rebuilding access logic specifically for the AI application. The product already contains that understanding.

The same product can serve a customer-facing application, an analytics team, an AI agent and an executive dashboard. Context improvements made to the product can be reused across those consumers rather than implemented independently for each one.

The data product changes the economics of context. Without a reusable architecture, every new AI use case creates another context-engineering exercise. Teams retrieve definitions, assemble metadata, map relationships, determine which sources can be trusted, implement policies and put that information into the application's prompts, retrieval system, agent tools or code. Then another use case arrives and much of the work begins again.

DataOS moves that work upstream. Context is connected to the data product when it is built rather than reconstructed every time an AI system needs the data. The organization can invest in the meaning and reliability of customer, inventory, order, product, financial or operational data once and reuse that investment across multiple applications.

This also changes where context is maintained. If Finance changes the definition of customer profitability, that change can be reflected in the data product. If an upstream source changes, lineage can reflect it. If a quality threshold fails, the product can expose its current status. If a governance policy changes, the policy remains connected to the data as it is consumed.

The data product also provides the reference point for reconstructing the context used in a specific decision. When its definitions, policies, lineage and quality status are versioned, an observer can identify what was in effect when the data was consumed rather than relying on the product's current state.

Context becomes part of the lifecycle of the data rather than documentation surrounding it. That distinction becomes increasingly important as AI systems begin to act. A person looking at a dashboard can notice that a number looks unusual, check another system, ask the data team a question or decide not to use the information.

An agent operating inside a workflow may have seconds to make the same determination. The data product needs to provide enough context for that decision to be made reliably at the point of consumption.

Connecting the Dots Across the Business 03

A data product maintains context about the data it represents. It does not, by itself, provide an enterprise-wide model of how data products, entities, and domains relate to one another. Yet a great deal of the reasoning enterprises need from AI runs across exactly those boundaries: which campaigns are associated with which revenue, which suppliers affect which product lines, which customers connect to which accounts.

A knowledge graph connects that context across the enterprise. It makes relationships across data products, entities, and domains explicit, so an AI agent can follow the connection from a campaign to the revenue associated with it, or from a supplier to the product lines it affects, without a person mapping that path by hand.

None of this replaces the semantic layer or the data product. A semantic layer still defines shared business meaning and metric logic. A data product brings data and its context together. A knowledge graph connects entities and context across products and domains.

DataOS brings these pieces together so AI can use a more complete understanding of the enterprise than any one of them can provide alone.

How DataOS Brings Context into the Data Product

04

Enterprises already have substantial investments in warehouses, lakehouses, databases, streaming infrastructure, SaaS applications, catalogs, governance systems, and analytics platforms. Much of the context required for AI is already being created across that environment.

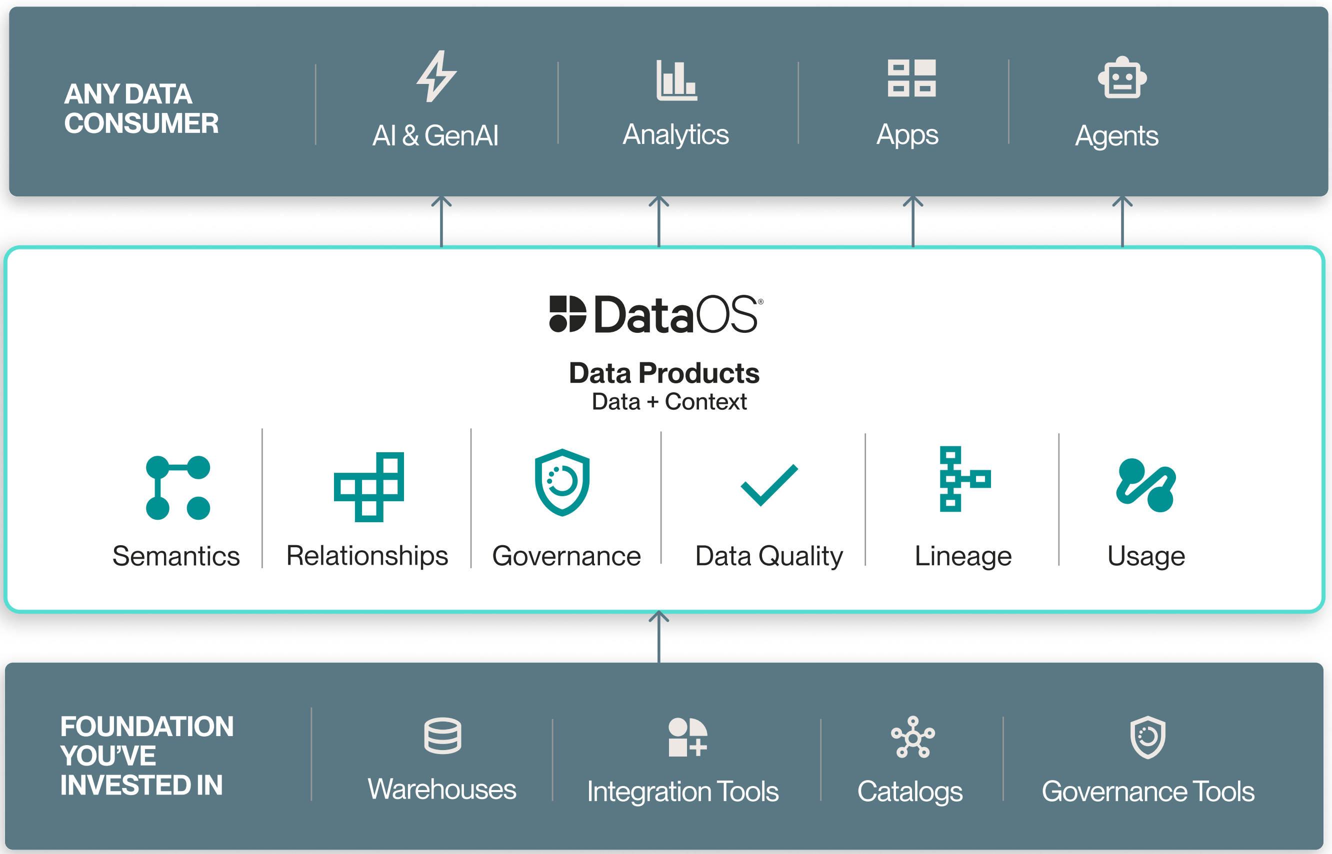
The data architecture needs a way to bring those pieces together and operationalize them.

DataOS works across the existing data environment to create governed data products from data already distributed across it. Those products provide a point where data and context can be managed together. At the bottom of the architecture are the systems enterprises already use to store and process data: cloud warehouses, lakehouses, databases, operational applications, streams, files, and other sources.

DataOS works across that foundation. Within DataOS, the data is modeled with its business semantics and relationships. Quality expectations are defined and monitored. Governance policies are applied. Provenance and lineage are maintained. Operational metadata captures the state of the product and the processes supporting it.

Those capabilities come together in the data product. At the consumption layer, AI systems, agents, applications, analytics, and teams access the product with the context already associated with it. **The diagram on the following page shows how these layers come together in DataOS.**

DataOS Connects Enterprise Data and Context



DataOS works across existing enterprise data infrastructure to create governed data products where data and context are managed together and made available to AI, applications, analytics, and teams.

DataOS also keeps the data product consistent across the existing data environment. Some platforms define and manage data products within their own platform environment. DataOS builds data products across the existing data environment, so the same contract, including semantics, quality rules, and governance policies, can remain consistent whether the underlying data sits in Snowflake, BigQuery, Databricks, or Postgres. **For an enterprise running more than one engine, which many do, that distinction affects how consistently data products and their context can be used across the enterprise.**

This also provides a practical answer to a problem that can otherwise become overwhelming: how much enterprise context needs to be modeled? An organization does not need to encode everything it knows before it can deploy useful AI.

Start with the Use Case
For a specific business process or AI use case, identify the data and context required to support it. For an order fulfillment process, that might include orders, customers, products, inventory, warehouses, shipping commitments, and the relationships, metrics, and policies required to interpret and use them. For lending, for example, it may include applicants, accounts, credit information, pricing, risk rules, decision history, and regulatory policies.

The goal is to associate enough context with those data products for the systems consuming them to interpret and use them reliably.

With DataOS, enterprise context grows through use. As additional data products and domains are added, their definitions, relationships, governance, quality rules, and operational signals become part of a growing base of enterprise context that can be governed, maintained, and reused.

That changes what happens when the next AI use case arrives. Instead of beginning with another effort to find the right data and reconstruct what it means, teams can begin with data products where much of that work has already been done.

Proof in Practice: An 80% Faster Path to GenAI

05

A Fortune 500 technology manufacturer with more than 160 million connected devices worldwide was generating continuous telemetry it struggled to put to use at scale. The data could support predictive maintenance, quality optimization, and new revenue streams, but several platform initiatives with multiple vendors had produced two-to-three-year timelines, fragmented systems, and growing leadership skepticism about whether device data could ever be connected.

Rather than begin another lengthy platform replacement, the company implemented DataOS across its existing infrastructure. The first test began with roughly fifty sample telemetry files delivered to the DataOS team on a Friday. By Tuesday, the team had turned those files into a working Device360 data product with a unified view of device-performance data, natural-language queries, and real-time predictive risk monitoring. Within 12 weeks, a test of the approach became a decision to adopt DataOS across the enterprise.

Compared with the company's previous delivery approach, **DataOS reduced the time required to move from enterprise data to a working GenAI use case by 80%.** The results also included a 50% reduction in repetitive data engineering tasks, sub-second query latency across more than 160 million devices, and production-ready deployments in four to six weeks compared with a typical cycle of twelve to eighteen months. As the company's Head of Engineering for its Global Innovation Center put it, developers shifted "from 'just producing tables' to building reliable, discoverable data products with clear ownership and governance, and a direct line to driving business results."

The telemetry had always existed. DataOS connected it with the context needed to create a governed, reusable data product that other teams and systems could build on.

Start With What the Business Means 06

The context that gives enterprise data meaning already exists across the business. The challenge is keeping it connected to the data so every new AI initiative does not have to find and reconstruct it again.

At the data layer, the DataOS platform keeps data and context connected through governed data products across the systems and platforms an enterprise already has. That context can be maintained and reused as the data is consumed across the enterprise.

As more data products are built, that shared understanding grows. New AI use cases can build on what the organization has already defined and governed rather than starting over. The goal is for each new AI use case to start with more enterprise understanding than the one before it..

References

Challapally, A., Pease, C., Raskar, R., & Chari, P. (2025). The GenAI Divide: State of AI in Business 2025. MIT Project NANDA.

Gartner. (2026). Market Guide for Agentic Analytics. Deepak Seth, Georgia O'Callaghan, Fay Fei, and Jeroen Cornelissen. 9 February 2026.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157–173. https://doi.org/10.1162/tacl_a_00638

Packer, C., Wooders, S., Lin, K., Fang, V., Patil, S. G., Stoica, I., & Gonzalez, J. E. (2023). MemGPT: Towards LLMs as operating systems. arXiv. https://doi.org/10.48550/arXiv.2310.08560

The Modern Data Company. (2026). *5 emerging trends in enterprise data & AI: Interim findings from the third Modern Data Survey.*

Contact Us

About The Modern Data Company

The Modern Data Company is the creator of DataOS®, a data platform that works across an enterprise's existing data systems and platforms to make data ready for AI, applications, and analytics. DataOS brings business context, governance, quality, and operational intelligence to enterprise data, giving teams and AI a consistent way to understand and use it without starting from scratch for every new use case.

Trusted by Fortune 500 companies, public sector organizations, and enterprises around the world, DataOS puts complex enterprise data to work faster and at a lower cost, without requiring organizations to replace or consolidate the infrastructure they have already built.

To schedule a demo or learn more about DataOS: **info@tmdc.io**

www.themoderndatacompany.com

Modern™