Modern WhitePaper

AI Starts with Data: Are you Ready?

A **Scorecard** for Enterprise AI Readiness

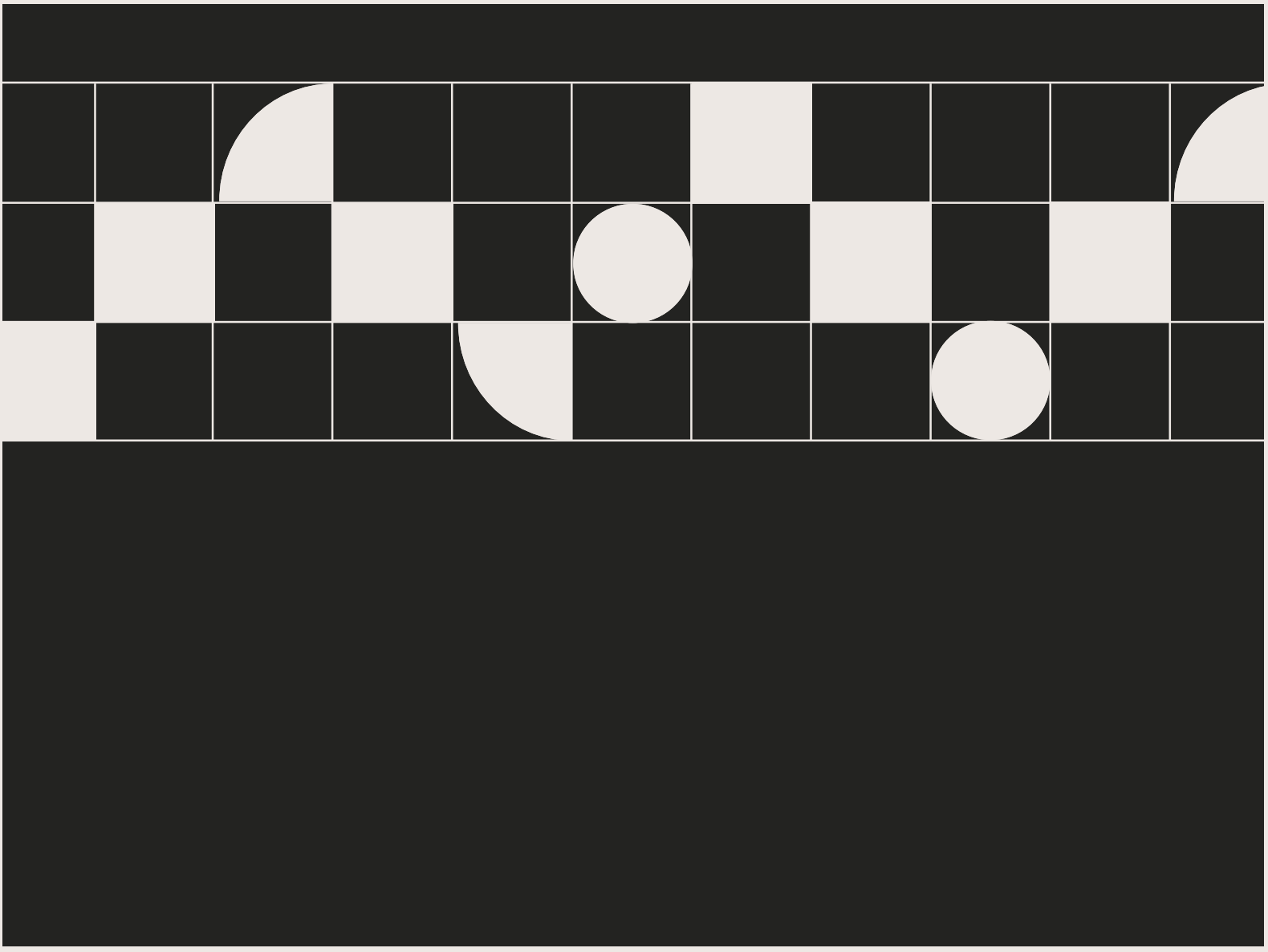


TABLE OF CONTENTS

Executive Summary

AI Success Starts with AI-Ready Data

Despite years of investment in robust data platforms, enterprises are finding that these systems—though effective for traditional analytics—struggle to meet the demands of transformative AI initiatives. According to S&P Market Intelligence, the share of companies abandoning most of their AI initiatives before production jumped from 17% in 2024 to 42% in 2025. **AI projects can show initial promise but often fail to scale or deliver consistent results across different business units. The problem isn't necessarily the AI models—it usually lies in how data is managed, harmonized, and delivered to those models.** Most data architectures were built for human reporting and analytics rather than for powering intelligent, adaptive systems. As a result, data environments often lack the governance, context, composability, and activation capabilities required to fuel AI at scale, creating friction that hinders adoption and limits impact.

For business leaders, Chief Data Officers (CDOs), Chief Information Officers (CIOs), and more, the imperative is not to replace what has already been built, but to evolve it. Preparing an AI-native data foundation efficiently means building on existing investments while addressing the critical gaps that limit scalability, such as fragmented governance, lack of reusability, and inconsistent semantic structures. This deliberate evolution toward an AI-native data ecosystem allows organizations to unlock AI's full potential without disruption, ensuring data becomes a trusted, reusable, and business-aligned asset that accelerates outcomes across the enterprise.

The Four Elements Framework for Efficient AI Readiness

This white paper introduces the **Four Elements Framework**, a first-principles approach to building an AI-native data ecosystem. These four capabilities represent the essential characteristics of AI-ready data: **Trust, Context, Productized** (into data products), and **Actionable.** When built into a data foundation, these capabilities ensure that data is Trusted through built-in governance, enriched with business Context through semantic meaning, Productized and Accessible into reusable assets, and Actionable on demand. By applying this framework, organizations can move beyond BI- and dashboard-oriented architectures toward a unified foundation built for AI. The result is an enterprise data layer that accelerates deployment, strengthens trust and consistency, and scales seamlessly and cost effectively across business and technical domains.

Organizations that implement the Four Elements Framework achieve measurable impact: the speed of data activation increases by up to 90%, model deployment accelerates by a factor of five, and AI initiatives are significantly more likely to succeed in production.

AI Begins with Data: A First-Principles Approach

The gap between AI's promise and scalable business outcomes isn't about the models—it's often about the data. Enterprises invest billions in AI agents, yet there's a fundamental mismatch between what these agents need and what today's data infrastructure provides. **Data warehouses, lakes, and analytics platforms were built assuming humans would do the thinking. That assumption is now obsolete.**

AI agents don't consume data like people do. They don't wait for queries, interpret dashboards, or pause for clarification. Most existing architectures were designed for business intelligence, not for the trust, context, composability, and AI-native execution that intelligent systems demand. The result is fragmented workflows that leave even advanced AI initiatives struggling to deliver consistent business value.

Unlocking AI's full potential requires evolving from systems built for business intelligence to data foundations designed for intelligent, adaptive workloads. It demands rethinking how data is structured, governed, and made instantly available. **This evolution calls for four foundational capabilities—trust, context, productized, and actionable**—which transform data infrastructure into an AI-native ecosystem that builds, deploys, and scales AI solutions with measurable outcomes and lasting competitive edge.

We call this the Four Elements Framework for AI-Ready Data. This is a first-principles approach to building data infrastructure designed not only to support AI but to accelerate it.

The Four Elements Framework for AI-Ready Data

1. Trust: Built-in Governance for AI Confidence

Building trust in data is more than accuracy—it's about making every asset transparent, validated, and traceable. Instead of relying on scattered governance tools, a converged approach embeds compliance, quality, and lineage directly into the data flow. Each dataset carries a clear record of origin, transformation, and use, creating a real-time map of its journey. With policies applied automatically, organizations gain consistent oversight and confidence that every dataset, whether used by analysts, data scientists, or AI models, is reliable, compliant, and ready for action.

Why Trust Matters for AI

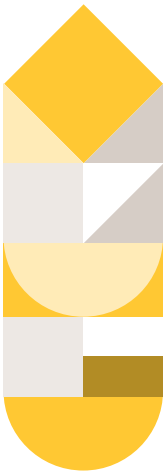
AI doesn't fix untrusted data—it magnifies its weaknesses. Even small gaps in accuracy, completeness, or compliance can escalate quickly when scaled through AI. Built-in governance ensures AI operates on data that is accurate, explainable, and policy aligned. End-to-end lineage enables organizations to see exactly how data flows, allowing for the rapid identification of issues, validation of model inputs, and proof of compliance. This transforms AI from a high-risk, unpredictable effort into a dependable driver of measurable, scalable business value.

Best-in-Class Capabilities

End-to-end lineage and ownership – Complete traceability from source to consumption, with clear accountability for data assets.
Automated policy enforcement – Embedded governance and access controls that proactively maintain compliance and data security without slowing delivery.
Real-time monitoring and quality checks – Continuous visibility into pipeline health, performance, and anomalies, ensuring reliable data with rapid resolution.
Self-service discovery – Empowered business and technical teams with governed, secure access to trusted data on demand.

2. Context: Semantic Consistency for AI and Analytics

Semantic consistency ensures data carries the same meaning everywhere, always grounded in the business context. Raw numbers, codes, and IDs are translated into a shared business language with standardized metrics, dimensions, and entities. Relationships are modeled to mirror real-world operations, so concepts like 'customer,' 'revenue,' or 'churn' are always consistent. This bridges the gap between technical data structures and business concepts, embedding institutional knowledge into the data foundation. As a result, analysts, applications, and AI models draw from trusted, context-aware definitions, eliminating ambiguity and accelerating insights to drive confident decisions at scale.



Why Context Matters for AI

AI systems perform best when they understand not just what happened, but why it matters in a business context. Without semantic consistency, models must infer meaning from raw inputs, risking misinterpretations that undermine trust, relevance, and adoption. By embedding context into standardized definitions of metrics, dimensions, and entities, the semantic layer ensures that every piece of data carries shared meaning across the organization.

This context-rich consistency enables AI to connect related data points, recognize patterns, and generate insights that mirror domain expertise. It also allows models to scale across diverse use cases, transferring learning from one domain to another without constant retraining. Most importantly, it ensures outputs are aligned with business priorities, making AI more explainable, actionable, and impactful at every level of the enterprise.

Best-in-Class Capabilities

Unified semantic layer – Applies metrics, dimensions, and entities consistently across all use cases, so both AI and analytics draw from the same business-aligned model.
Cross-domain knowledge graph – Provides a unified business view by linking entities, relationships, and attributes across systems, without technology constraints.
Semantic alignment with lineage – Ensures business metrics are modeled consistently and backed by end-to-end traceability, making them auditable.
Impact analysis and dependency mapping – Surfaces downstream implications of changes, ensuring that data and metadata activation stay consistent across systems and domains.

3. Productized: Reusable Data Assets to Scale AI

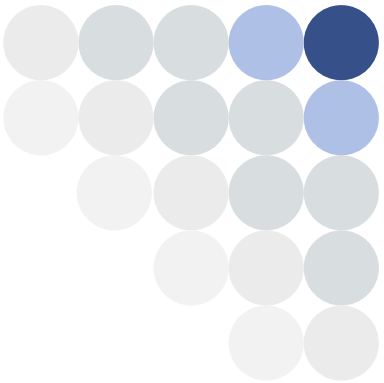
Productizing data means treating it as reusable products rather than one-off outputs. This approach transforms data into a business asset—a data product—that can be easily discovered, trusted, and applied consistently across multiple initiatives. Concepts like “active customer” or “monthly recurring revenue” are embedded into reusable, composable assets that maintain consistency across all use cases. Managed through a full product lifecycle, these assets evolve into a living library of trusted, high-value resources that adapt with changing business priorities.

Not all “data products” are created equal. The term is often applied loosely today, but true data products—as defined by ODPS and Gartner—go much further. They embed business context, discoverable through semantic search, guarantee SLOs, and enforce governance for compliance. They provide out-of-the-box consumption capabilities for BI and Agents, which are observable from both operational and business perspectives, and can be deployed like containers on any cloud infrastructure. This is what separates ad hoc outputs from business-ready, reusable data products that enable AI to scale with confidence.

Why Productization Matters for AI

One-off data preparation for each AI project hinders progress, introduces inconsistencies, and increases costs. Treating data as a business asset—engineered, productized, and managed through its full lifecycle—makes it reusable, reliable, and ready to fuel AI at scale. Data products embody this shift: packaging context, lineage, quality, and standardized metrics into assets that can be easily discovered, trusted, and applied across business initiatives.

By embedding business meaning directly into the data and eliminating repetitive preparation, productized assets reduce AI deployment time from months to weeks, while ensuring consistent results across domains such as marketing, finance, and operations. They also provide reusability at scale, giving AI agents and applications business-ready inputs without rework. Enterprises that adopt this approach accelerate time-to-value, lower operational costs, and gain the agility to adapt AI initiatives as priorities evolve, without having to start over for every new project.



Best-in-Class Capabilities

- **Delivery with performance optimization** – Data is delivered quickly and reliably at scale, ensuring responsiveness for both analytics and AI workloads.
- **Standardized and interoperable formats** – Data is structured in consistent formats that enable seamless integration across applications and AI consumption patterns.
- **Consistent global metrics and definitions** – Single source of truth for business metrics, ensuring data is understandable across all use cases.
- **Cataloged and versioned assets** – Data assets are enriched with descriptions and ownership, and version-controlled to ensure they remain trusted, reusable, and aligned to business needs.

4. Actionable: Data Delivered at the Speed of AI

AI agents, applications, and business users can only move as fast as the data they consume. Traditional approaches force teams to spend time on manual preparation, integration, and translation before use—slowing insights and limiting the impact of AI. What's required is actionable data: high-throughput access, flexible query patterns, and seamless integration with AI and BI workflows.

Out-of-the-box consumption endpoints make this possible by exposing data in ready-to-use formats through APIs, MCP, and BI integrations that users and agents already know. Context-rich datasets are pulled on demand and arrive governed, traceable, and consistent, with semantics preserved end to end. The result: humans and agents get the right data when they need it, enabling faster execution, more thoughtful decisions, and reliably better outcomes.

Why Actionability Matters for AI

Traditional data infrastructure, built for static reports and dashboards, creates bottlenecks for AI agents and applications that require fast and context-aware data access with flexible connectivity. Without an AI-native execution capability grounded in productized data, access remains fragmented, lacks context, and is misaligned with business priorities. Productized data is instantly actionable, removing these constraints and enabling effortless connectivity through multiple access endpoints—APIs, REST, SQL, GraphQL, and MCP—so agents and applications can quickly pull exactly what they need at the required speed. This ensures AI systems always operate on trusted, context-rich data with the precision and agility intelligent workloads demand, empowering them to act immediately on business-ready inputs. For leaders, this is not a backend detail—it is a strategic enabler that turns data from passive storage systems into an active driver of reliable, business-aligned AI at scale.

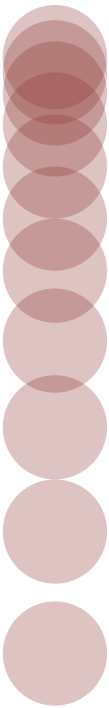
Best-in-Class Capabilities

Simplified data access gateway – Declaratively abstract physical data silos to simplify integration and apply consistent access controls across all activations.

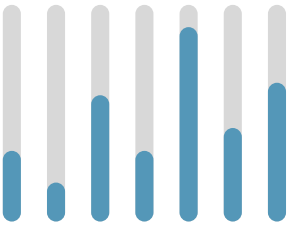
High-performance APIs – Built-in APIs for AI workloads, delivering low-latency, high-throughput access for model training, inference, and data apps.

MCP, API – REST, SQL, and GraphQL support – Empowers LLMs, agents, and applications with context-rich data when it's needed.

Notebook and Dev environment integration – Direct connections to Jupyter notebooks, IDEs, and engineering tools streamline experimentation and model development.



Take your AI Readiness Assessment



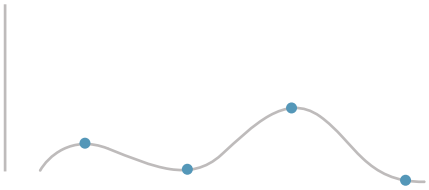
The path to AI Readiness begins with an honest assessment of your current data ecosystem measured against the Four Elements Framework. This evaluation reveals not only capability gaps but also strategic opportunities to create lasting competitive advantage through data and AI.

This scorecard is a practical tool to evaluate how prepared your data ecosystem is for the AI era. Based on the Four Elements Framework—Trust, Context, Productized, and Actionable—it highlights strengths, identifies gaps, and provides a benchmark to guide next steps.

How to Use this Scorecard

This scorecard presents a series of questions, each rated on a scale of 1 to 10. For each question, select the rating that best reflects your current capability:

1 (Ad Hoc): No or minimal capability in place.
5 (Reactive): Moderate capability, but with gaps in coverage or manual processes.
10 (Purpose-Driven): High capability with comprehensive, automated, or systematic implementation.



Use these anchors to calibrate your score. A “1” reflects significant gaps and ad-hoc approaches, while a “10” reflects a fully AI-native data foundation. Most organizations will fall somewhere in between—the scorecard helps reveal your current readiness and highlights where the most significant opportunities for improvement lie.

Once completed, your responses generate a total score out of 150, providing a clear snapshot of your current state and a starting point for planning improvements.

0–80 points: Early stage, with significant gaps to address.
81–130 points: Progressing, with strong capabilities and room to strengthen further.
131–150 points: Strong foundation, purpose-driven, and AI-ready.

Need Help?

We’re here to support you. If you’d like help filling out the questions or reviewing your results, our team can guide you through a quick but comprehensive AI readiness assessment—turning your score into a strategic roadmap for activating your data foundation and driving AI excellence.

Contact us at info@tmdc.io
or schedule a meeting.

Questions

Lineage

1. How quickly and accurately can you trace end-to-end data lineage from source systems to AI outputs, and use that lineage to assess the impact of any change?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Raw data with no lineage visibility. Impact analysis is manual, updates take weeks with a high level of rework.
5 (Reactive): Partial lineage documented with basic tagging and categorization, but coverage is incomplete and often manual. Impact analysis can take days, and at least 30% of changes require manual rework.
10 (Purpose-Driven): Comprehensive lineage tracking across all systems, connecting entities and relationships through knowledge graphs. Impact analysis and changes propagate the same day, almost automatically, with less than 5% rework.

Parallel, Governed Workflows

2. How integrated and governed are your analytical workflows, and how rapidly and reliably can they run in parallel to deliver business-ready outcomes?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Workflows are fragmented across tools; governance is manual or absent; runs are sequential and slow; outcomes take weeks or months.
5 (Reactive): Some workflows are connected with partial governance; parallelism is limited; outcomes are delivered in days to weeks with frequent rework.
10 (Purpose-Driven): Workflows are fully integrated, end-to-end, with automated governance; parallel execution is standard, and outcomes are delivered in hours to days with minimal rework.

Self-service Access

3. How well is data access democratized while maintaining security/compliance, and how long does it typically take for a new user to gain access to data?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Data access is siloed, controlled manually, and often inconsistent. Days to weeks for granting access; duplicated extracts or shadow copies of data are common.
5 (Reactive): Broader access is available, but security/compliance controls are applied reactively and unevenly. Most approvals are received within 1–3 days; some may require duplicate extracts.
10 (Purpose-Driven): Self-service access available to all authorized users/agents with fine-grained controls ensuring security and compliance. Access grants in minutes or less; no redundant copies.

IT dependence

4. Can business and AI teams quickly access enterprise data independently, without relying on IT?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Access is through IT, requiring tickets and manual interventions. IT-managed handoffs; multi-day cycles per request.
5 (Reactive): Some self-service access exists, but teams often still depend on IT for delivery.
10 (Purpose-Driven): Teams can directly access and activate trusted data products through self-service tools. Zero-ticket onboarding; governed endpoints available same day.

Context Aware

5. Can your systems recognize and differentiate the contextual meaning of the same data across use cases, with definitions that are co-owned and aligned between business and technical teams?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No recognition of context; data is interpreted differently in each use case, with no business and technical alignment.
5 (Reactive): Some contextual differences are documented, and a few shared definitions exist, but ownership is fragmented and inconsistent.
10 (Purpose-Driven): Context is consistently recognized across use cases, with business and technical teams co-owning definitions.

Semantic Layer

6. Do you have a semantic layer that connects business concepts to data with traceability for compliance, and how quickly can new definitions be implemented?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No semantic layer; business concepts and technical data are disconnected, with no traceability. Changes roll out per tool, over weeks.
5 (Reactive): A partial semantic layer exists, with some business concepts linked to data and limited lineage documentation. New metrics reach priority tools in days.
10 (Purpose-Driven): A comprehensive semantic layer links all business concepts to technical data, with automated lineage ensuring full compliance and auditability. Publish once goes live across BI/APIs in hours.

Data Preparation Effort

7. How much of your data preparation effort is spent on repetitive, manual tasks compared to automated reuse?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Nearly all data preparation is manual and custom for each project. 70–90% manual per project; heavy one-off pipelines.
5 (Reactive): Some reusable data products reduce repetitive work, but most preparation is still case-by-case, more than 30% manual work
10 (Purpose-Driven): Preparation is largely automated using production-grade, reusable data products across projects, less than 10% manual.

Versioning and Cataloging

8. Is your data versioned, cataloged, and delivered as organized assets that AI systems can consume without extra preparation, and how fast can a new use case be spun up?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Data is raw and unorganized, requiring extensive manual cleaning or transformation before AI can use it. Each use case needs weeks of prep.
5 (Reactive): Some data is catalogued and reused for specific use cases, but versioning and consistency are limited. New use cases take days to weeks.
10 (Purpose-Driven): Data is versioned, catalogued, and delivered as AI-ready assets that are composable and immediately usable across domains. Fully versioned/cataloged, AI-ready assets, new use cases in hours to days.

Global Metrics

9. Do you maintain consistent global metrics across teams and use cases, and how quickly can you correct metric drift?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Each team defines and calculates metrics independently, leading to discrepancies. Drift persists for weeks.
5 (Reactive): Some shared metrics exist for common use cases, but gaps and inconsistencies remain. Drift detected/corrected in days.
10 (Purpose-Driven): A unified catalog of standardized, global metrics is consistently applied across all teams and systems. Drift is auto-detected, corrections propagate in hours.

Business-Aligned Data Assets

10. Are your data assets designed for specific business outcomes rather than generic consumption?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): Data is stored generically, with no alignment to business outcomes.
5 (Reactive): Some data assets are tailored to use cases, but most remain generic.
10 (Purpose-Driven): Data assets are engineered as outcome-driven products, reusable and composable for multiple business needs.

AI Explainability

11. How effectively can you audit AI decision paths—such as model inputs, outputs, and interpretability—to ensure regulatory compliance?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No visibility into AI decision paths, model inputs/outputs, or interpretability.
5 (Reactive): Some audit logs and partial traceability of model inputs/outputs exist, but interpretability is limited and not standardized.
10 (Purpose-Driven): Full transparency with traceable model inputs/outputs, clear interpretability of decisions, clear documentation, and audit-ready compliance reporting.

AI Agent Data Access

12. Can AI systems reliably access data through readily available APIs, such as REST, SQL, GraphQL endpoints, and MCP, and how quickly can they be exposed?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): AI systems require custom integration for each data source, with no standard APIs. Weeks to expose new endpoints.
5 (Reactive): Some APIs are available, but coverage is limited and not optimized for AI. Days to expose.
10 (Purpose-Driven): A comprehensive AI-native execution layer exposes consistent, well-documented APIs (REST, SQL, GraphQL, MCP) across all data domains. Endpoints can be exposed in hours.

Dev Tools Integration

13. Can both technical teams and business applications access production data securely and directly through notebooks (such as Jupyter), development tools, or APIs for rapid delivery of AI use cases, and how long does environment setup take?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No direct access; teams must rely on exports or IT-managed pipelines. The setup takes weeks.
5 (Reactive): Some notebook integrations exist, but coverage is inconsistent and limited to technical users. Set up in days.
10 (Purpose-Driven): Full support for direct, secure access to production data for both technical and business teams via notebooks, APIs, and modern development tools. Environment Setup in hours

Real-time Access

14. Does your infrastructure support real-time data access for AI inference with minimal latency?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No support for real-time access; data must be batch-processed before use. Seconds-to-minutes latency is unacceptable for inference.
5 (Reactive): Some workloads support near real-time queries, but latency is inconsistent. 95th percentile latency seconds.
10 (Purpose-Driven): Infrastructure is optimized for low-latency, real-time AI inference across all workloads. 95 / 99th percentile latency seconds across critical workloads.

Data Quality Monitoring

15. Do you have real-time monitoring and alerting for data quality issues to proactively prevent cascading model failures in AI systems?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

1 (Ad Hoc): No monitoring; issues found manually or only after failures.
5 (Reactive): Some monitoring exists, but alerts are inconsistent, delayed, or need manual fixes.
10 (Purpose-Driven): Automated, real-time monitoring with proactive alerts, anomaly detection, and self-healing to prevent cascading failures.

Scoring Sheet

Question #	Score (1-10)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
Total Score	

Add an AI-Ready Foundation to Your Data Stack with DataOS

The path to AI-ready data doesn't require replacing your existing infrastructure. You can augment it with an intelligent layer that delivers the four essential elements at scale. **DataOS serves as this AI readiness layer, seamlessly connecting to your current data sources while embedding trust, context, productization, and actionability directly into your data foundation, quickly and efficiently.**

As an operating system for your enterprise data, DataOS creates a unified experience that preserves the specialized strengths of your existing systems while eliminating their limitations. This approach protects your technology investments, minimizes operational disruption, and establishes the foundation for sustained competitive advantage in the AI era.

DataOS maximizes your entire data ecosystem

For IT teams, DataOS delivers seamless integration without the rip-and-replace complexity. It simplifies infrastructure management by breaking down data silos and reducing maintenance overhead, while maintaining full compatibility with your existing stack.

For business teams, DataOS provides intuitive access to context-rich, governed data products that accelerate the journey from insight to action. Teams can discover, trust, and activate data assets that drive innovation across applications and AI initiatives.

The result is an AI-ready data foundation that scales with your ambitions and adapts to business needs.

Ready to activate your data?

Need help knowing where you stand with AI readiness?

Contact us at info@tmdc.io or schedule a meeting.

About the Modern Data Company

The Modern Data Company is redefining data management for the AI era, transforming data from a technical challenge into a company's most powerful business asset. Modern's DataOS platform is the world's first operating system for data, integrating with any stack and transforming it into an AI-native data ecosystem, no rip and replace.

Fortune 1000+ enterprises use DataOS to scale AI and solve mission-critical data challenges. With DataOS, enterprises are accelerating AI adoption by up to 90% while reducing data consumption and platform costs by 50%. Modern's rapidly expanding customer base includes global category leaders across a wide range of global industries. They trust DataOS to power their AI and business transformation.

www.themoderndatacompany.com
Info@tmdc.io

Modern