WHITE PAPER

# AI-Ready Data Foundations

Helping Mid-Market Companies Reinvent for an AI World

By Jim Perry, Bill Raper, Uttam Channegowda, and James Palavos

**G2O Data & Analytics**

POINT OF VIEW • Q3 2026

CONTENTS

# In this document

01

# A Note to the Reader

You already know the gap. Your platform is modern. Your governance is documented. Your team is capable. And yet when leadership asks whether the organization is ready to scale AI, the honest answer is more complicated than the investment history suggests.

The problem is not what you bought. It is what was left unbuilt between the platform and the business. The semantic enforcement, the operational controls, the governed definitions that turn documented intent into consistent, trusted output at runtime. That gap is where AI pilots stall, where agents produce confidently wrong answers, and where human experts become the quality control system no one budgeted for.

This paper is about that gap: what it looks like in mid-market manufacturing and distribution environments, why it persists despite meaningful investment, and what it takes to close it in a way that holds in production. G2O has spent years working in this space, not just at the architecture level but at the operational level where definitions either hold or they don't. If you are at Level 2 on the maturity curve and wondering why Level 4 feels further away than it should, this paper was written for you.

— **Uttam Channegowda, VP Data & Analytics at G2O**

## AI readiness is not a platform problem. It is an operationalization problem.

Platform modernization, governance, metadata, master data, semantic enforcement, data quality, security, stewardship, and operating model all have to work together before AI can scale reliably. The semantic layer matters as one component of a larger system, but not as a standalone solution. The organizations that succeed with AI are the organizations that operationalize trust before AI arrives. The ones that don't find out the hard way and the AI program quietly stalls.

The problems AI exposes were already there. Long before AI became the priority, organizations struggled with familiar challenges: multiple definitions of customer, conflicting revenue calculations, fragmented product hierarchies, duplicate supplier records, and reporting environments dependent on tribal knowledge to explain discrepancies. AI did not create those problems. It exposed them. The question is no longer “is governance documented?” The question is “does governance survive at runtime?”

Mid-market companies are under pressure to reinvent themselves for an AI world. Leadership teams are asking for faster decisions, better customer experiences, improved productivity, more automation, and measurable returns from AI investment. Still, many organizations are still operating from data foundations designed for dashboards, reports, and manual analysis but not for AI agents, retrieval-based systems, predictive models, and automated decisioning.

That creates a dangerous gap. When the data foundation is not ready, people lean in to compensate. Analysts reconcile definitions. Data engineers patch pipelines. Business experts explain exceptions. Leaders ask for manual validation before trusting the output. What gets described as “human-in-the-loop” becomes something heavier: human-IS-the-loop. Trust depends on heroic human effort rather than built-in controls. That tax is manageable for pilots. It does not scale in production.

The scale of this gap is now documented at industry level. Accenture Research’s 2026 study of 2,000 companies across 15 countries and 9 industries found that only 7% of what Accenture calls “data reinventors” have reached the level of data-readiness required to scale advanced AI at enterprise speed. The remaining 93% have invested in cloud platforms and governance tooling but have not yet built the operational foundation required to make AI outputs reliable.

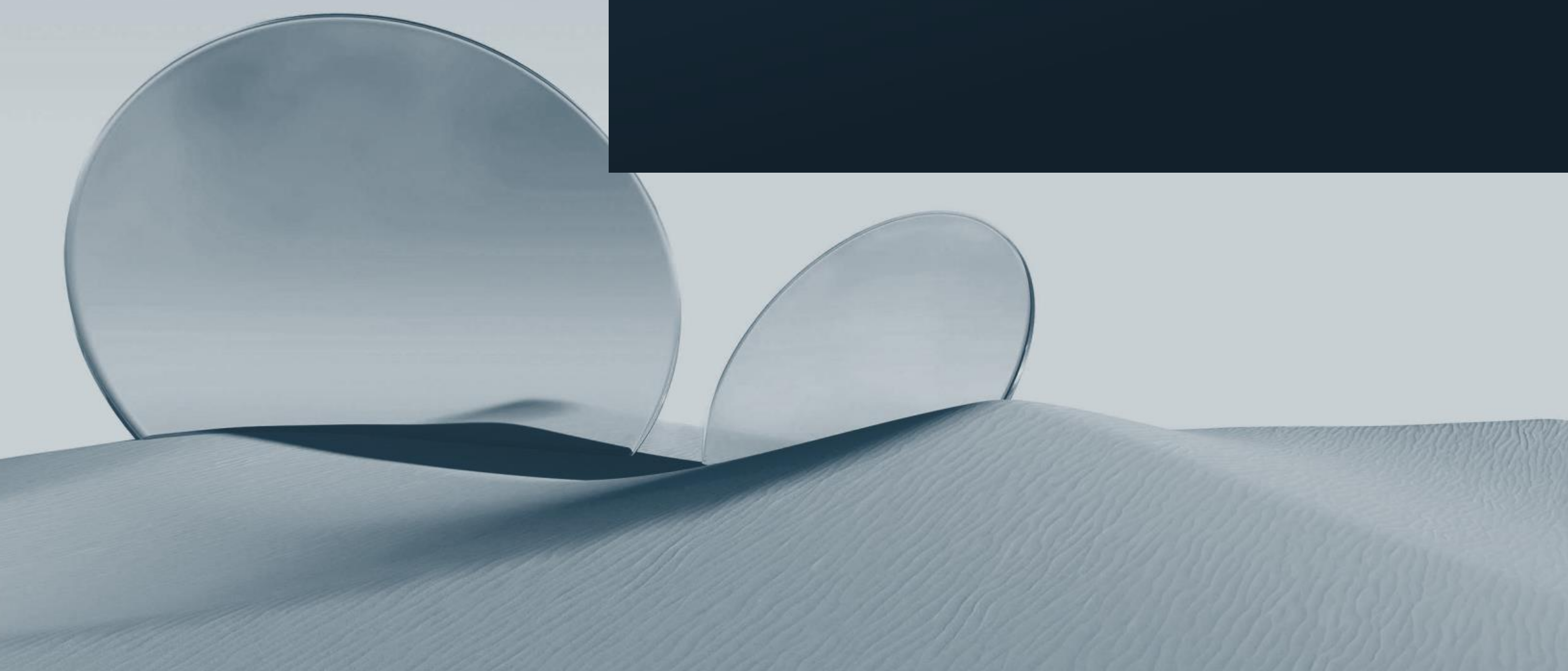
G2O helps mid-market companies move beyond early AI maturity by building the data foundation, operating model, and roadmap required to scale AI with confidence. The goal is not another technology stack. The goal is to help organizations build the production system they need for trusted AI.

PART ONE

01

# The case

Why the platform investments landed but the trust didn't — and why the window is now.



02

# The Investment Paradox

Over the last five years, many mid-market companies have invested heavily in making data teams more effective and efficient.

They modernized platforms so teams could move faster, scale compute, reduce infrastructure friction, and support more advanced analytics. Snowflake, Databricks, and Microsoft Fabric are tools we see often that can help accomplish this. In factory terms, they upgraded the plant: more capacity, better machinery, faster throughput, and more room for advanced production.

They invested in metadata and governance so teams could find data, understand lineage, assign ownership, document definitions, apply policy, and improve trust. Purview, Collibra, Alation, Atlan, Unity Catalog, and Snowflake Horizon are examples of effective tools. This in factory terms would be a better operating manual: parts catalogs, supplier records, safety policies, approval processes, and traceability.

These investments were necessary. They made the ‘factory’ faster, better organized, and easier to govern. But they did not fully solve the problem AI now exposes: whether every output is built from the same business meaning at runtime.

The dashboard era was forgiving. A Power BI report that defined “revenue” slightly differently from a Looker dashboard was an annoyance. A business team could reconcile the difference in a meeting, document the exception, or escalate the disagreement. That was like discovering that two production teams used slightly different inspection checklists and correcting the issue after the fact.

The AI era is not forgiving.

A RAG model cites contradictory revenue figures from internal documents. A forecasting model silently underperforms because its feature pipeline defines “active customer” differently than the validation dashboard. An AI agent given a tool that queries raw tables produces confidently wrong outputs depending on which join path it picks.

In factory terms, this is not a reporting inconvenience. It is an assembly-line defect. The product may look complete, but it was built using inconsistent specifications.

The gap is not in the platform, which may be computing efficiently. It is not in the catalog, which may contain accurate documentation. The gap is between team enablement and operational consistency.

Metadata describes what data means. Governance defines who owns it and how it should be used. But the semantic layer enforces those meanings at the point of consumption. It turns documented definitions into operational behavior.

In the factory analogy, the investments made the factory floor faster and the operating manual clearer. But without standardized work instructions and calibrated measurement at each production station, different teams can still assemble the product differently.

Most enterprises have invested in a better factory and a better operating manual. Fewer have implemented the production controls that ensure every output is built the same way.

That is the semantic gap.

03

# Why Now: The Mid-Market AI Inflection

Five forces have moved AI-ready data foundations from technical aspiration to executive requirement.

The factory is no longer producing only dashboards and reports. It is being asked to produce AI-generated answers, model-driven decisions, automated recommendations, and agentic actions. Those outputs require tighter production controls than traditional analytics did.

For mid-market companies, this inflection is especially important. Many have enough data, scale, process complexity, and leadership urgency to benefit meaningfully from AI. But they often do not have the depth of data operating model, governance maturity, semantic enforcement, or specialized teams found in the largest enterprises. That makes the foundation gap more visible and the human-is-the-loop tax more expensive.

## 1. Leadership Expectations Have Moved Faster Than Data Foundations

Mid-market leadership teams are under pressure to show AI progress quickly. They want productivity gains, improved customer engagement, better forecasting, smarter operations, and faster decision cycles. Many have already funded pilots, copilots, dashboards, automation experiments, or RAG prototypes.

But AI value depends on whether the organization can trust the data behind the output. If definitions are inconsistent, source systems conflict, documents are outdated, permissions are unclear, or data quality is not monitored, AI creates more work instead of less.

## 2. Agentic AI Requires a Semantic Contract

When a language model is given tools that call SQL against raw tables, the model is not operating from business meaning. It is operating from available structure. That creates risk.

An agent may choose the wrong table. It may follow the wrong join path. It may apply a local definition of a metric that conflicts with the enterprise definition. It may produce a confident answer without understanding that the same business term has multiple interpretations across systems.

The Model Context Protocol and the broader agentic ecosystem make this problem acute. Agents need a trusted interface to enterprise data. Without a semantic contract, they cannot reliably determine which interpretation of a metric is correct. The semantic layer becomes the controlled production interface for AI agents.

## 3. Native Platform Features Changed the Build-Versus-Buy Math

Snowflake Semantic Views and Databricks Unity Catalog metric views are now part of the native platform conversation. Microsoft Fabric is also pushing a more integrated data and analytics experience. CDOs and data leaders are being asked why they have not adopted the semantic capabilities now available to them.

This changes the decision. The question is no longer whether a semantic layer is an exotic architecture choice. The question is which semantic pattern fits the enterprise's consumption model, operating model, and platform strategy.

## 4. Regulation and Risk Made Definitional Consistency a Business Issue

AI governance is no longer only a technology concern. It is becoming a regulatory, risk, and disclosure concern. For higher-risk AI systems, organizations increasingly need to demonstrate lineage, explainability, control, and appropriate governance over the data used by models.

Metadata documents lineage and ownership. Governance defines policy. But semantic layers enforce consistent meaning at the point where data becomes an output. A documented definition in a catalog does not guarantee that an AI system, dashboard, or model actually used that definition.

## 5. AI Has Expanded the Definition of "Data Foundation"

The foundation must now support structured data, unstructured data, governed retrieval, semantic definitions, feature reuse, vector search, real-time signals, agent tools, access controls, lineage, and observability. It also has to support people: the domain experts, stewards, engineers, and business owners who understand the context behind the data.

04

## Start Simple

Think of the mid-market AI journey as moving from a workshop to a high-volume smart factory.

Your data platform (Snowflake, Databricks, or Microsoft Fabric) is the factory floor. It provides the space, machinery, storage, and compute power to move raw materials through production. It is modern, scalable, and essential. But a factory floor by itself does not guarantee consistent output.

Your metadata and governance layer is the factory's operating system: the parts catalog, supplier records, safety standards, inspection rules, ownership model, and traceability logs. It tells you what each part is, where it came from, who approved it, how it can be used, and what rules apply. This is the role of Purview, Collibra, Alation, Atlan, Unity Catalog, Snowflake Horizon, and similar tools.

**But there is still a missing piece.**

The semantic layer is the set of standardized work instructions and calibrated measuring tools used at every production station. It ensures that every team builds from the same definition of "customer," "revenue," "active account," or "gross margin." Without it, one station may install a "wheel" one way, another station may interpret "wheel" differently, and the finished cars may look similar but perform inconsistently.

Again, that is the risk many mid-market companies face as they push into AI. The platform can process the data. The catalog can describe the data. But unless definitions are enforced when data is actually queried and used, every dashboard, model, pipeline, and AI agent can still produce a different answer.

Specialized capabilities are like specialized factory equipment. MDM is needed when parts arrive from multiple suppliers with different identifiers. A feature store is useful when the same components must be reused across many ML production runs. Vector infrastructure supports AI-specific retrieval and search. Streaming is required when the assembly line must respond in real time. A knowledge graph or ontology serves as the plant map when use cases require complex relationships, explainable reasoning, or policy-aware action. These tools matter, but they are not universal; they are added when the production model requires them.

Without the full foundation, the factory may run, but only because people apply herculean effort to inspect, reconcile, correct, and explain every car before it leaves the line. That is not true human-in-the-loop AI. It is human-is-the-loop AI: a model where trust depends on manual intervention rather than built-in controls. With the right foundation, the enterprise can produce consistent, governed, AI-ready outputs at scale.

The foundation does not replace people. It makes their expertise scalable. Every effective factory still depends on experienced operators, engineers, quality leaders, and production managers who understand the product in context. Enterprise AI is the same. The people who understand the data, the business process, the exceptions, the customer context, and the operational tradeoffs are essential to making AI trustworthy. The goal is not to remove those experts from the loop; it is to stop making them the entire loop.

PART TWO

# 02

# The architecture

Three foundation layers, two cross-cutting controls, and the capabilities you only add when a use case demands them.

05

# The Architecture: Three Foundation Layers Plus Cross-Cutting Controls

The model that fits the post-2026 environment has three foundation layers that are always required, a set of cross-cutting controls that must operate across those layers, and conditional capabilities added only where the AI use case mix demands them.

This is the architecture of an AI-ready factory. Every factory needs a floor, an operating system, and standardized production controls. It also needs sensors, inspections, safety systems, and access controls that run across the plant. Some factories also need specialized equipment, but only when the product line requires it.

## Foundation Layers: Always Required

### Platform

The cloud data platform provides separated storage and compute. It is usually chosen before the AI conversation begins. Snowflake, Databricks, and Microsoft Fabric are common enterprise patterns. This is the factory floor: the machinery, storage, throughput, and production capacity.

### Metadata and Governance

The metadata and governance layer includes catalog, lineage, business glossary, classifications, ownership, stewardship, and access policies. It describes data and the rules that govern it. Purview, Collibra, Alation, Atlan, Unity Catalog, and Snowflake Horizon are common examples. This is the factory operating system.

A strong operating system tells the enterprise what materials exist, where they came from, who owns them, what rules apply, and how they should be handled. It creates visibility and accountability across the plant. But visibility is not the same as control. A catalog entry does not guarantee that every production station follows the same instruction.

### Semantic Layer

The semantic layer provides executable metric and dimension definitions to queries at runtime. It enforces the meanings the metadata layer describes. It is the activation layer that turns documentation into operational governance. This includes things like dbt Semantic Layer, Cube, AtScale, LookML, Snowflake Semantic Views, and Databricks Unity Catalog metric views.

This is the standardized work instruction layer: the common measurement system used across every production station. It ensures that “revenue,” “customer,” “active account,” “gross margin,” and other key business concepts are not reinterpreted every time a dashboard, model, pipeline, or agent needs them.

## Semantic Layer Versus the Broader Context Layer

The semantic layer is not the entire context layer, but it is a critical part of it. Metadata and governance describe assets, ownership, lineage, classifications, and policy. The semantic layer enforces business definitions for metrics, dimensions, and governed consumption. Vector infrastructure supports retrieval over unstructured content. Knowledge graphs or ontology-based approaches may be needed when relationships, policies, entities, events, and actions become more complex.

## Cross-Cutting Control: Data Quality, Observability, and Trust

Across all three foundation layers, enterprises need active controls for data quality, freshness, completeness, anomaly detection, and operational monitoring. In manufacturing terms, this is the sensor network and inspection system built into the production line.

For AI, these controls matter because models and agents often consume data continuously and at scale. A stale table, broken pipeline, missing field, inconsistent document, or silent schema change can quickly become a bad recommendation, incorrect answer, or automated decision.

Quality controls also need ownership. When a data product fails a test, someone must know who owns the issue, what business process is affected, and whether downstream AI systems should continue using that data.

## Cross-Cutting Control: Security, Privacy, and Access Enforcement

AI-ready data foundations must also enforce security, privacy, and access controls at every layer. AI increases the risk surface because models and agents can combine data, summarize sensitive content, and expose information through new interfaces.

The enterprise needs clear policies for who can access which data, which fields are sensitive, which documents are restricted, which prompts or tools can call governed data, and which outputs require auditability. This includes role-based access, row- and column-level controls, PII classification, masking, permissions inheritance, policy enforcement, and usage logging.

Security and privacy cannot be bolted on after the model is built. They must be part of the production line.

## Conditional Capabilities: Added by Use Case

The next set of capabilities should not be treated as universal requirements. They should be added only when the AI use cases demand them. In manufacturing terms, these are specialized production capabilities. Not every factory needs every machine.

A vendor that recommends every layer regardless of use case is selling a stack. The right plan starts with the AI use cases and works backward to the capabilities required.

| CONDITIONAL CAPABILITY | REQUIRED WHEN | NOT REQUIRED WHEN |
|---|---|---|
| **Master Data Management (MDM)** | Use cases cross entity boundaries, such as Customer 360, supplier consolidation, householding, or cross-channel attribution. | Single-system analytics, transactional ML, or document-based RAG. |
| **Feature Store** | Multiple production ML models share features at scale and training-serving consistency is required. | Early-stage ML or one or two production models with limited reuse. |
| **Vector Infrastructure** | RAG, semantic search, embedding-based recommendation, or retrieval over unstructured content is required. | Structured analytical workloads dominate. |
| **Streaming** | Real-time agentic decisioning, in-flight fraud detection, sensor-based response, or | Decision windows are measured in minutes or |

| | | |
|---|---|---|
| | second-level decision windows are required. | hours and micro-batch is sufficient. |
| **Knowledge Graph or Ontology** | AI use cases require complex relationships, explainable reasoning, policy-aware actions, or connected entity context. | Metrics, reporting, and simpler retrieval use cases can be handled through semantic, catalog, and vector patterns. |

The discipline is that conditional means conditional. A vendor that recommends every layer regardless of use case is selling a stack. The right plan starts with the AI use cases and works backward to the capabilities required.

### A Note on MDM in the AI Era

Master data has been treated for years as a back-office discipline: important, expensive, slow, and often deferred. The AI era changes the math. A language model cannot reason about customer context when systems disagree on who the customer is. A forecasting model cannot predict demand when product definitions vary across fulfillment, inventory, and planning systems. An agent that retrieves from multiple repositories cannot resolve which version of “supplier” is canonical.

MDM stays conditional, meaning not every AI use case requires entity resolution. But where the use case crosses entity boundaries, MDM is foundational, not optional. The teams selling AI without addressing the entity problem are selling a model that will produce confident answers grounded in inconsistent identifiers.

06

# Unstructured Data Readiness

AI-ready foundations must also account for unstructured data including but not limited to: documents, policies, emails, contracts, transcripts, knowledge articles, PDFs, images, and operational notes. For many generative AI use cases, these are the raw materials the factory consumes.

Making unstructured data AI-ready requires more than loading documents into a vector database. The enterprise needs source ownership, permissions inheritance, document quality controls, metadata tagging, version management, chunking strategy, retrieval testing, and citation back to authoritative sources.

In factory terms, the organization needs to know which materials are certified, which are obsolete, which are restricted, and which can safely be used in production. A model that retrieves from an outdated policy document or an unauthorized contract repository is not just making a retrieval mistake. It is using the wrong material on the line.

Unstructured data also needs governance in context. A document may be useful for one audience but restricted for another. It may contain embedded definitions that conflict with certified enterprise metrics. It may be authoritative for one region, product, or business unit but not another.

For RAG, semantic search, and agentic workflows, the quality of retrieval becomes a production control. The question is not only whether the model can find a document. The question is whether it can find the right document, respect permissions, understand its context, cite the source, and avoid using obsolete or conflicting material.

07

# What We See in the Field

The three patterns below are not derived from research reports or industry surveys. They come from the first conversations G2O has with mid-market manufacturing and distribution clients. Those conversations that happen before a scope of work is written, before a platform is evaluated, before a single line of code is touched.

Our first question will be “Which of your business partners are we bringing into this conversation?”

The semantic gap, or the distance between how data is defined in a system and how it is understood in the business, cannot be closed by a technology team working alone. It can only be closed by people who understand how data is actually used, debated, and trusted across the organization. That knowledge lives in the business, not the platform.

What follows is what G2O consistently finds when that conversation happens.

FIELD PATTERN

## The Governance Program That Stopped at Documentation

### What leaders asked for

Organizations had already invested in governance platforms such as Collibra, Purview, or similar catalog tools. They had documented definitions, assigned stewards, and built business glossaries. The expectation was that reporting consistency would follow.

### What the real issue was

The definitions existed, but they were not operational. Teams continued to recreate business logic inside reports, SQL scripts, dashboards, and downstream applications.

In one financial services engagement, the organization maintained hundreds of metrics across multiple reporting environments. The governance platform described the approved definitions, but nothing required teams to use them. As a result, executives could still compare reports across departments and find numbers that did not match.

### What had to change

The organization needed to move from documented governance to runtime governance. Metric definitions had to become executable standards, not reference material. Business logic needed to be consumed from a certified source rather than rebuilt independently by each reporting team.

### What G2O built

G2O established a governed semantic layer that consolidated key metric definitions into a single certified model. Reports, dashboards, and analytics tools could then consume business logic from the same source of truth instead of recreating it in separate environments.

### The Results

Definitions stopped being recommendations in a catalog and became the runtime standard for reporting, analytics, and future AI consumption.

FIELD PATTERN

## When MDM Is Really a Definitions Problem

### What leaders asked for

A regional distributor wanted to implement vendor master data management using a lakehouse-native MDM capability.

### What the real issue was

The business had never resolved what counted as a “vendor.” A logistics provider was treated as a vendor by finance, a partner by operations, and a customer in one business line.

The same legal entity appeared three different ways across systems, and each treatment was operationally valid.

### What had to change

Finance, procurement, operations, and the commercial team needed to agree on an entity hierarchy before match rules were written. The governance model had to respect legitimate business differences instead of flattening them into one oversimplified definition.

### What G2O built

G2O facilitated joint working sessions to resolve the entity definitions, ownership model, and stewardship responsibilities. The MDM platform then encoded the agreed hierarchy rather than forcing a technical definition onto the business.

### The Results

The MDM capability became sustainable because the business had agreed to the model before implementation. Stewardship lived inside finance, procurement, and operations, so the hierarchy still held when new vendors were onboarded a year later.

FIELD PATTERN

## AI Exposed a Metric Layer Nobody Owned

### What leaders asked for

A consumer-facing services firm wanted to evaluate the fastest path for a generative AI capability that could analyze high-volume operational interactions.

### What the real issue was

The hard question was not which AI platform to use. It was which metric definitions the AI would inherit. Terms like “successful outcome,” “high-value account,” and “first-touch resolution” varied across dashboards. None of the dashboards were wrong, but none of them agreed.

### What had to change

The business needed to stop treating metric drift as a reporting issue and start treating it as a foundation issue. Metric ownership had to move into the operating function, with certified definitions that could be enforced at the query layer.

### What G2O built

G2O helped define and certify the core metrics before the AI capability was built. The work established business ownership, semantic enforcement, and a governed path for AI consumption.

### The Results

The AI program shipped against governed definitions instead of inherited dashboard logic. Both the internal center-of-excellence path and the vendor platform option became viable because the foundation underneath them was certified.

In each case the pattern repeats. The technology was buildable from day one. The business definitions were the bottleneck. The engagements that succeeded were the ones where the business showed up early, owned the definitions, and stayed involved after the platform was in production. When key business partners are left out of the conversation, builds like these fail -- we know, because that's when we get called in to fix it.

That is the semantic gap. Closing it is not a tooling problem. It is a communication problem with business consequences.

"Every engagement starts the same way. The technology team reaches out. We tell them we need the business in the room. That’s not a request — it’s a condition. You cannot build definitions that hold in production if the people who live by those definitions aren’t part of building them." — Jim Perry, G2O Data & Analytics

PART THREE

# 03

# The path

Where mid-market companies actually sit, and the fixed-price route from here to trusted AI.

08

# The Competitive Reality: Before the Agentic Factory Comes the AI-Ready Data Factory

The market conversation is moving quickly toward AI-enabled factories, digital twins, physical AI, agentic operations, and software-defined manufacturing. Large global consultancies and platform alliances are positioning this future at industrial scale: AI agents working with humans and machines, digital twins optimizing operations, and manufacturing data platforms supporting repeatable performance across factories.

That vision can only be realized with the right foundation.

For many mid-market companies, the immediate problem is not whether AI can transform the enterprise. It is whether the organization has the trusted data, governed definitions, quality controls, access model, and operating discipline required to make AI outputs reliable in the first place.

G2O's position is different by design. We help mid-market companies build the AI-ready data factory first. That means establishing the production foundation: platform readiness, metadata and governance, semantic enforcement, data quality, observability, security, unstructured data readiness, and the operating model required to sustain trust.

This is not a smaller version of a global transformation program. It is a more practical path for companies that need to move from early AI maturity to scalable AI readiness.

The destination may be an agentic enterprise. The starting point is a data foundation strong enough to support it.

# Where Most Mid-Market Companies Sit

Progression to an AI-ready foundation is not binary. This maturity curve is a factory maturity curve. Organizations move from craft production, where every team builds its own way, to governed production, where common definitions, controls, and automated interfaces make outputs repeatable.

| LEVEL | NAME | DEFINING CHARACTERISTIC | TYPICAL COHORT |
|---|---|---|---|
| 1 | **Metric Chaos** | Definitions live in BI code, spreadsheets, extracts, and ad hoc SQL. No working glossary. Disagreements resolved by escalation. | Most mid-market today |
| 2 | **Documented** | Glossary exists in catalog. Ownership assigned. Definitions documented but not enforced at query time. | Mid-market with mature catalog investments |
| 3 | **Certified** | Gold metrics governed. Executive reporting constrained to a certified set. Two-tier reality persists. | Leading domains in mid-market |
| 4 | **Semantically Operational** | Semantic layer is the default access path for BI, ML, and AI. Definitions executable, governed, observable. | Few mid-market today |
| 5 | **Agent-Ready** | Semantic layer exposed via MCP. Agents query certified definitions. Definitional changes propagate. | Emerging |

In G2O's field experience, most mid-market companies sit at Level 1 or Level 2, with leading domains beginning to reach Level 3. Companies at Level 1 have invested in cloud platforms and may have a catalog license, but lack a working business glossary, consistent metric definitions, or operational governance. Companies at Level 2 have a documented glossary but no runtime enforcement; every dashboard, model, and pipeline still re-implements the definitions. Few mid-market companies have crossed into Level 4. Level 5 is emerging, especially as organizations begin to explore governed agent access through interfaces such as MCP.

The published research is consistent with what we see in the field. The EDM Association's 2026 Global Data Management Benchmark, surveying more than 435 organizations across 50+ countries and structured around the DCAM Data Management Capability Assessment Model, found that only approximately 31% of organizations have advanced data strategy capability, leaving the majority without the operational foundation required to scale AI. Notably, while more than 70% report formal governance structures, most lack the funding models, measurement frameworks, and enterprise alignment required to operationalize them. Governance has been documented but not yet enforced.

The move from Level 2 to Level 4 is not primarily a tooling migration. It is an operating model change. It requires domain ownership, metric certification, stewardship workflow, change control, testing, quality monitoring, access enforcement, and consumption patterns that make the semantic layer the default path rather than an optional layer.

In factory terms, the goal is to stop relying on heroic inspectors at the end of the line and start building quality into the line itself.

# Why G2O

Most data engagements start and end with the technology team. The brief comes from a data engineer or an IT leader. The solution is scoped, built, and delivered within that same boundary. The business is informed at milestones. The output is technically correct and operationally disconnected.

This is how G2O works differently.

When an inquiry arrives, regardless of who sends it, G2O's first conversation is about who else needs to be in the room. Not in a steering committee. Not in a sign-off meeting. In the working sessions where definitions are debated, business logic is documented, and the meaning of “customer,” “revenue,” “active account,” and “gross margin” gets resolved into something the organization can actually enforce.

That is not a methodology preference. It is a prerequisite for the work we do.

The semantic gap is not a technical problem. It is a business problem that lives in technology. Closing it requires understanding how data is used in practice: which definitions are trusted, which are contested, which are quietly ignored, and which ones different parts of the organization have been quietly maintaining in parallel for years.

Larger firms bring frameworks. G2O brings immersion. The difference shows up not in the architecture diagram but in whether the governance model holds six months after the engagement closes. This is because it was built on how the business actually works, not assumptions.

G2O's data and analytics practice serves mid-market clients across manufacturing, financial services, distribution, and consumer products. Our engagements are structured around production outcomes. Our measure of success is not whether the solution was delivered, but whether the business trusts it.

"We don't build definitions for your data team. We build definitions your business team will actually use. Those are two very different deliverables." — Jim Perry, G2O Data & Analytics

11

# Tiered Path

What an engagement with G2O looks like in practice.

| TIER | DURATION | INVESTMENT | BUSINESS OUTCOME |
|---|---|---|---|
| **Assess** | 3–4 weeks | $65–85K fixed | Maturity score, capability map, prioritized 18-month roadmap, and business case input. |
| **Foundation Sprint** | 6–8 weeks | $200–275K fixed | One domain live: metadata, semantic layer, governance, data quality controls, and 2–3 AI use cases validated. |
| **Activate** | 4–8 months | Custom | Multi-domain rollout, conditional capabilities deployed, MCP or agent enablement, observability, access enforcement, and stewardship live. |

12

# Next Step

A 30-minute conversation. No slides. Just three key questions.

1. How many places is “active customer” defined in your environment today, and how do those definitions differ?
2. Which of your AI use cases have a single, certified, traceable metric definition behind them?
3. If an AI agent queried your data tomorrow, would you trust the answer?

These three questions usually surface more in 30 minutes than a six-month assessment from a larger firm. If the answers indicate Level 1, we’ll talk about what comes next. If they indicate Level 2 or 3, we’ll explore the bridge to Level 4. If you are already at Level 4, you’re part of the 7%. You probably don’t need us, but we would like to learn from you.

> **The destination may be an agentic enterprise.**
> **The starting point is a data foundation strong enough to support it.**

*Sources  Accenture Research, “AI-Ready Data for Advanced AI” (2026); EDM Association, “2026 Global Data Management Benchmark Report” (May 2026).*

G2O Data & Analytics

G2O.COM