O'REILLY®
Report

Foundations for Better Enterprise Data Management

A Practical Guide for Business and Technology Leaders

Emma McGrattan

ACTIAN™
a division of HCLSoftware

Put Your Data Foundations to Work

Turn complexity into clarity and data into business value.

Architecture
Modernize infrastructure with hybrid cloud, high-speed pipelines, and seamless integration — no need to start from scratch.

Governance
Turn ownership, policies, and access rules into action with embedded controls that enable — not slow down — your teams.

Quality & Observability
Catch issues early, track data health continuously, and build trust into your delivery and consumption layers.

Metadata
Accelerate discovery and reuse with knowledge graph–powered cataloging, automated lineage, and rich business context.

Access & Security
Enable governed, role-aware data access across your data landscape at scale, with compliance built in.

Actian helps make enterprise data scalable, trusted, and AI-ready. **Explore our solutions at actian.com**

Foundations for Better Enterprise Data Management

A Practical Guide for Business and Technology Leaders

Emma McGrattan

O'REILLY®

Foundations for Better Enterprise Data Management
by Emma McGrattan

Published by O'Reilly Media, Inc., 141 Stony Circle, Suite 195, Santa Rosa, CA 95401.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Aaron Black
Development Editor: Jeff Bleiel
Production Editor: Gregory Hyman
Copyeditor: Charles Roumeliotis
Cover Designer: Susan Brown
Cover Illustrator: Ellie Volckhausen
Interior Designer: David Futato
Interior Illustrator: Kate Dullea

February 2026: First Edition

Revision History for the First Edition
2026-02-06: First Release

979-8-341-67232-1

[LSI]

Table of Contents

Preface

Over the years, I've seen too many organizations pour time and money into data platforms, pipelines, and catalogs only to end up with the same old problems: data that is siloed, inconsistent, and impossible to trust. I've heard from business analysts who can't answer basic questions without a week of reconciliation, and from engineers whose "transformational" projects are killed before they deliver anything of value.

Since the advent of AI, the pressure has only gotten more intense. AI and digital initiatives don't just expose weak data practices; they put them under a microscope and blow them up. Cracks in the foundation show up fast: models that spit out garbage, and dashboards that leadership can't trust.

I wrote this report for the people who carry that burden every day: the data teams. The ones who get the late-night calls when pipelines break, who are asked to explain numbers that don't add up, and who are expected to deliver trust in the middle of chaos. This isn't a sales pitch and it's not theory. It's a practical playbook you can put to work right now to get your data under control, earn confidence, and build foundations that last.

Who This Report Is For

If your job involves designing, building, maintaining, or governing data systems, I wrote this report for you. The contents here are meant for:

- Data engineers who wrestle with pipelines that never seem to run cleanly
- Database administrators and architects who balance scale, performance, and compliance under constant pressure
- Data stewards and governance leads trying to make sense of messy, inconsistent definitions that erode trust
- Analysts who know that "good enough" data rarely is

Above all, I wrote this for the practitioners, the ones asked to deliver reliable, trustworthy data when the stakes are high and the spotlight is on. This is not meant as a primer for business executives; it's a hands-on guide for the people in the trenches who make data work.

How to Use This Report

This report is organized around five foundational pillars of modern data management: Architecture, Governance, Quality and Observability, Metadata, and Access and Security. I explain why the pillars matter, what challenges to anticipate, and what practices can help strengthen the foundations.

You may choose to read the report from cover to cover or focus on the areas most relevant to your current challenges. The intent is not to prescribe a rigid methodology, but to provide a practical reference that supports prioritization, discussion, and planning. By the end, you should be equipped with a structured starting point for improving your organization's data foundations and sustaining them as business needs evolve.

O'Reilly Online Learning

For more than 40 years, *O'Reilly Media* has provided technology and business training, knowledge, and insight to help companies succeed.

Our unique network of experts and innovators share their knowledge and expertise through books, articles, and our online learning platform. O'Reilly's online learning platform gives you on-demand access to live training courses, in-depth learning paths, interactive coding environments, and a vast collection of text and video from O'Reilly and 200+ other publishers. For more information, visit *https://oreilly.com*.

How to Contact Us

Please address comments and questions concerning this report to the publisher:

O'Reilly Media, Inc.
141 Stony Circle, Suite 195
Santa Rosa, CA 95401
800-889-8969 (in the United States or Canada)
707-827-7019 (international or local)
707-829-0104 (fax)
support@oreilly.com
https://oreilly.com/about/contact.html

We have a web page for this report at *https://oreil.ly/better_data_mgmt*.

For news and information about our books and courses, visit *https://oreilly.com*.

Find us on LinkedIn: *https://linkedin.com/company/oreilly-media*.

Watch us on YouTube: *https://youtube.com/oreillymedia*.

Acknowledgments

This report was made possible through the collaboration and support of many individuals. Special thanks to my colleagues at Actian and HCLSoftware who contributed their insights, challenged assumptions, and provided real-world perspectives that shaped the content.

I am truly grateful to Anthony Corbeaux and Jennifer Jackson for their sponsorship and guidance, and to the extended thought leadership and marketing teams for helping refine the vision for this work.

CHAPTER 1

Breaking Out of the Current Data Management Mess

Data management has never been more strategic, or more misunderstood.

In this report, *data management* refers to the practical work of turning raw data into something the organization can reliably use across its lifecycle from creation to consumption. That includes architecture and integration, metadata and definitions, quality and observability, and access and security. *Data governance* is the mechanism that makes those practices consistent and accountable: it assigns ownership, sets policies and standards, and ensures the right controls are embedded into day-to-day workflows.

Most organizations don't fail because they lack data, talent, or technology; they fail because their data initiatives are fragmented, reactive, and disconnected from business context. This chapter explores why so many well-funded data projects stall or collapse under their own weight, and what foundational principles can help leaders turn data chaos into progress.

The Fragile Foundations of Enterprise Data

Enterprise leaders have long been told that data is the essential fuel for digital transformation—popularized by outlets such as *The Economist*, which famously argued that the world's most valuable resource is no longer oil, but data. Yet, despite massive investments in platforms, catalogs, pipelines, and talent, most data programs continue to underperform. The reasons are both structural and cultural.

On the structural side, enterprises often operate with fragmented architectures: multiple warehouses, lakes, lakehouses, and integration tools that don't work well together. Data lives in silos owned by different business units, each with its own rules, standards, and definitions. This fragmentation makes even simple questions (such as "What is our revenue by product line?" or "What is our customer churn rate?") difficult to answer consistently or with confidence, often resulting in different teams arriving at different answers to the same question.

Culturally, the absence of clear ownership leads to confusion and finger-pointing when issues arise. Governance, the ownership and control layer of data management, is either too heavy-handed, slowing progress with red tape, or too light, failing to enforce quality or accountability. Project teams frequently underestimate the effort required to ensure data quality, focusing instead on flashy tools or analytics layers. As a result, dashboards look impressive but hide unreliable data underneath.

These dynamics explain why many data projects stall or fail. Business stakeholders lose trust, adoption declines, and technical teams are asked to "fix" issues that should have been addressed at the foundation. Too often, the outcome is frustration: more spend, more tools, and little to show in measurable business value.

The current wave of AI adoption has amplified these long-standing challenges. Advanced models can automate decisions, generate insights, and personalize experiences, but only if the data they rely on is accurate and unbiased. Without strong data foundations, AI amplifies existing flaws. A model trained on incomplete customer records may misclassify segments, while one exposed to biased historical data may reinforce inequities at scale (as happened with the Apple Card in 2019).

Enterprises also face rising scrutiny from customers, regulators, and their boards. It is not enough to deploy AI applications; organizations must be able to explain how models were trained, what data was used, and how outcomes are monitored. Transparency and accountability are now as critical as accuracy. The result is that data, once viewed primarily as an IT concern, has become a boardroom topic.

This shift has created urgency but also risk. Some organizations are rushing AI pilots into production, without adequate oversight, only to face reputational damage when models behave unpredictably. Others are paralyzed, delaying initiatives out of fear of getting it wrong, or because of a lack of trust in their data. In both cases, weak data foundations are the root cause. AI has not created new problems; what it has done is exposed and magnified existing ones.

The Illusion of Progress

Companies often confuse activity with outcomes. Standing up a new lakehouse, rolling out a data catalog, or hiring a data science team can look like transformation on the surface. There's budget spent, impressive slide decks built, and a sense of momentum. But if the basics are still broken, like pipelines that fail regularly or numbers that change depending on who you talk to, then what was the point?

This illusion is powerful because it buys time. Executives can point to investments and claim progress, while the underlying problems remain unsolved. Dashboards may look slick, but if sales and finance are still debating revenue, or operations can't reconcile inventory, the window dressing doesn't matter. The danger is that this illusion of progress masks serious underlying issues. Leaders believe the foundations are solid enough to layer on AI or advanced analytics, only to watch those projects stall or backfire. A model doesn't care how much you spent on a lakehouse; it cares whether the training data is consistent and complete. When the foundations aren't right, the cracks show up quickly, and the organization ends up distrusting the new and expensive tools.

Real progress looks different. It's measured in fewer reconciliations, faster decision cycles, and business conversations that move past arguing over numbers to debating actual choices. It's visible when trust in the data grows, when business teams stop building shadow systems because they no longer need to. Those are the signs that the foundation is finally working for you, not against you. Once that shift begins, every improvement compounds; trust deepens, decisions accelerate, and the organization starts moving with clarity and confidence. That's what real progress feels like.

The Trust Deficit

When trust in the data doesn't exist, people create their own workarounds. Finance teams reconcile numbers on the fly, sales managers keep private spreadsheets, analysts pull their own extracts rather than rely on shared dashboards. Over time, the official system becomes a reference point rather than the source of truth. The organization ends up with multiple versions of the same metrics, none of them fully trusted, and a culture of second-guessing every number.

Trust and confidence are related but not identical. Data trust is the belief that numbers are correct; data confidence is the willingness to act on them without hesitation. A dashboard may technically be accurate, but if leaders don't feel confident in how the numbers were produced, they won't make decisions based on them. Instead, they will delay, debate, or demand manual validation.

The trust deficit rarely begins with bad intent. It starts with small disconnects between people, processes, and technology. When business users aren't part of defining metrics, when data owners can't clearly explain lineage, when there is no consistent approach to managing core entities such as customers, products, or suppliers, or when quality issues surface without accountability, confidence starts to slip. Trust doesn't collapse overnight; it erodes one inconsistency at a time. Every unexplained discrepancy chips away at credibility until skepticism becomes the default. Rebuilding starts by acknowledging that erosion and making transparency the norm.

Restoring trust and building confidence takes commitment. It requires transparency about how data is sourced, how it flows, and what checks are in place to make sure that it's complete and accurate. Business users appreciate tangible signals that the data is current and reliable, like freshness indicators, quality scores, and access to lineage graphs. Each time those signals hold up under scrutiny, confidence grows. Each time they fail, confidence erodes even further.

Over time, confidence becomes a force multiplier. When teams believe the data is accurate and trustworthy, they stop hedging with shadow systems and start focusing on decisions. Meetings shift from "Is this number right?" to "What should we do about it?" That shift is the real payoff of addressing the trust deficit and building data confidence.

Talent Burnout and the Cost of Delay

Weak foundations drain teams. Engineers spend more time repairing pipelines than building new ones. Analysts clean data instead of answering questions. Governance leads argue over definitions without the authority to settle them. Skilled people end up maintaining brittle systems rather than doing meaningful work.

The problem isn't just overwork; it is the loss of progress. Constant firefighting replaces creativity. Short-term fixes take the place of lasting solutions and the system becomes a tangle of one-off workarounds that no one fully understands. What looked like a shortcut becomes a debt that's impossible to repay.

Eventually, people disengage. Some leave for roles where they can actually build and innovate. Others stay but withdraw, delivering the minimum because they no longer see impact. Collaboration erodes and tensions rise as business teams demand answers and data teams struggle to keep up. The business pays for this too. Bad data delays decisions, damages trust, and stalls AI efforts before they reach production. Opportunities slip because teams are busy reconciling instead of innovating.

Fixing the burnout cycle starts with fixing the foundations. When pipelines run reliably, engineers can focus on designing better systems. When definitions are clear and enforced, governance leads can shift from refereeing arguments to enabling trust. When data is clean by default, analysts can spend their energy driving insights instead of cleaning up messes. Strong foundations protect both the business and the people who keep it running.

A New Kind of Thinking

Breaking out of the cycle means changing how organizations see data. For too long, it's been treated as a byproduct of business systems, something collected and stored without intent. That mindset is what created the current challenges. Strong data foundations do not come from technology alone; they come from leaders who make data a business priority. When executives own data outcomes and tie them directly to business results, the organization starts to shift from aspiration to accountability.

Data must be treated as a product, with clear ownership, quality standards, and guardrails built into day-to-day workflows. This doesn't mean adding layers of bureaucracy. It means putting in place patterns and practices that are repeatable and lightweight enough to keep pace with the business. Strong foundations aren't about slowing people down; they're about creating the conditions where teams can move fast with confidence.

Leaders set the tone by measuring progress based on outcomes such as trust in the data, decision speed, and the impact of analytics and AI. When data discipline becomes business discipline, teams align, and data becomes part of how success is defined and rewarded.

Business and technology leaders must work together. Data teams cannot deliver value without agreement on definitions, and business leaders cannot expect reliable analytics if they do not understand the trade-offs behind data quality and governance. When both sides align on outcomes, data shifts from cost center to competitive advantage.

Success should be measured by improved trust, faster decision cycles, and clearer outcomes. The goal is not perfect data but predictable, reliable data that people can act on with confidence. That is what separates organizations that move forward from those that remain stuck.

The current state is not inevitable. With the right mindset and building blocks, data management can become an engine for strategy, transformation, and AI.

CHAPTER 2

Traits of a High-Functioning, Data-Driven Organization

Many enterprises make bold claims about being data-driven, yet only a fraction actually turn those investments into a consistent, measurable impact. What sets high-functioning organizations apart isn't the volume of data at their disposal or the sophistication of their platforms, but their ability to make data a trusted, dependable engine for decisions and innovation.

This chapter explores six traits that distinguish these organizations: treating data as a strategic asset, establishing accountability, building agility, improving data literacy for everyone, managing data as a product rather than a byproduct, and aligning data strategy tightly with business goals. Each trait reinforces the others, creating the cultural and technical foundation for data that truly drives outcomes.

These six reinforcing traits are not independent capabilities. Together, they form a reinforcing system that defines what a high-functioning data-driven organization looks like in practice, as shown in Figure 2-1.

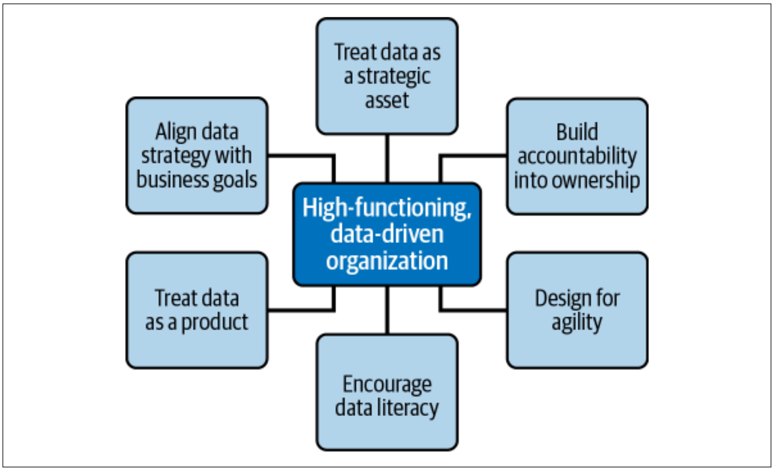

Figure 2-1. The six reinforcing traits of a high-functioning, data-driven organization, spanning culture, accountability, agility, literacy, product thinking, and business alignment

Trait 1: Treat Data as a Strategic Asset

In high-functioning organizations, data is treated as a strategic asset. Leaders across functions (finance, marketing, supply chain, operations) speak a shared language about metrics and performance. Core concepts such as "customer," "revenue," or "churn" are defined consistently across the enterprise, establishing shared semantics that eliminate the confusion that arises when different teams calculate the same measure in different ways. Decisions are made based on shared truths rather than conflicting reports.

When data is positioned as a business asset, its value compounds. Teams make faster decisions, leaders debate choices instead of numbers, and the organization builds confidence in how information is generated, maintained, and used. The result is not just better analytics. It is a better business rhythm.

Trait 2: Build Accountability into Data Ownership

Another defining trait of a truly data-driven organization is accountability. Ownership for key data assets is clearly assigned, and responsibilities are tied to business outcomes, not just technical operations. Governance is embedded into processes so that compliance, quality checks, and lineage tracking happen automatically.

This approach prevents governance from becoming a bottleneck or an afterthought. Instead, it becomes an enabler of innovation. It ensures that teams can move quickly without sacrificing trust or control. Accountability makes data everyone's business, yet no one's burden.

Trait 3: Design for Agility

High-functioning organizations also exhibit agility. When markets shift, customer behaviors change, or new technologies emerge, these organizations can adapt quickly because they trust the data guiding their actions. Instead of spending weeks reconciling spreadsheets or debating whose numbers are "right," they can move swiftly, confident in the integrity of the information they are using.

Agility in this context is not about speed alone. It is about responsiveness. Reliable data allows leaders to test, learn, and iterate without fear of compounding errors. When data foundations are strong, agility becomes a competitive advantage rather than a risk.

Trait 4: Encourage Data Literacy

Even the best systems won't deliver value if people don't know how to make sense of what they see. Data literacy—the ability to read, interpret, and communicate using data—is one of the most powerful enablers of a data-driven culture.

The challenge isn't about turning everyone into a data scientist. Most employees don't need to build models or design pipelines. What they do need is enough understanding to ask the right questions, put results into context, and recognize where data has blind spots. A marketing leader, for example, should be able to distinguish correlation from causation when looking at campaign results. A product manager should spot the risks of using incomplete or biased datasets when evaluating a new feature.

Organizations that get this right don't leave literacy to chance. They create baseline training for all roles, not just the technical ones. They make documentation easy to find and use plain business language in it. Most importantly, they build a culture where questioning the numbers is healthy—where "These figures don't feel right" sparks an investigation instead of an argument. In this environment, employees at every level are not just passive consumers of reports, but active participants in turning data into insight and action.

Trait 5: Treat Data as a Product

Historically, enterprises have treated data as an incidental output of business processes. Systems were designed to manage operations, with data captured along the way and stored for compliance or reporting purposes. This approach has led to fragmented ownership, inconsistent quality, and difficulty scaling data use across new initiatives.

Modern organizations are shifting to a data product mindset. When data is treated as a product, it is intentionally designed, documented, and maintained with the needs of its consumers in mind. This includes not only technical specifications such as schemas, APIs, or refresh intervals, but also guarantees around usability, quality, and support. A data product is expected to meet the same standards of clarity and service as any customer-facing product.

For example, consider a retail company that maintains a "customer 360" dataset. As a byproduct, it might be an inconsistent mix of loyalty data, ecommerce interactions, and call center logs, assembled in an ad hoc manner. As a product, it would be curated with defined ownership, consistent identifiers, clear lineage, and documentation so that marketing, sales, and analytics teams can confidently build campaigns or models without revalidating every field.

This mindset shift is crucial for scalability. When data is managed as a product, it becomes reusable across projects, reducing duplication of effort and enabling innovation at speed. Teams no longer need to rebuild pipelines from scratch or reconcile conflicting datasets; they can consume trusted data products, accelerating delivery while maintaining consistency.

Trait 6: Align Data Strategy with Business Goals

Ultimately, strong data management is not an end in itself. It is valuable only to the extent that it supports business outcomes.

Organizations that excel in data management understand this connection deeply. Their data strategies are not built around abstract notions of "modernization" or "platform upgrades," but around the specific outcomes the business is seeking to achieve.

This alignment requires clarity on priorities. For some enterprises, the priority may be AI readiness, ensuring data is clean, well-governed, and documented so models can be trained responsibly and retrieval-based AI systems can reliably access trusted, up-to-date information. Solutions here include both training pipelines, where data is used to build or fine-tune models, and retrieval-augmented generation (RAG) systems, where models dynamically retrieve enterprise data at query time rather than learning from it. For other enterprises, the priority may be agility: creating the ability to pivot quickly when markets shift or supply chains are disrupted. For still others, it may be compliance: building governance processes that satisfy regulators without slowing innovation.

In practice, this alignment means that technical metrics such as system uptime, data latency, and pipeline throughput are paired with business metrics such as reduced time to market, improved customer retention, and faster compliance reporting. Investments in data foundations are justified not by the number of tools deployed, but by their impact on real-world business performance.

High-functioning organizations also recognize that alignment is not static. Business priorities evolve and data strategies must evolve with them. This requires continuous dialogue between business and technology leaders, as well as mechanisms for monitoring, measuring, and adjusting the value delivered by data initiatives. When done well, this approach transforms data from a compliance necessity into a source of competitive advantage, fueling innovation, improving customer experience, and creating resilience in the face of change.

Taken together, these six traits define what "good" looks like in modern data management. They are not isolated best practices but interlocking foundations that allow organizations to move with confidence. The next step is turning these traits into repeatable patterns that scale. That is the essence of *governance by design*: building trust, accountability, and agility directly into the systems and workflows that power the business.

CHAPTER 3

The Five Foundational Data Management Pillars

High-performing data organizations don't succeed by luck or by buying the right tools. They succeed because they build on strong foundations. Behind every trusted report or reliable model is a set of disciplines that keep data connected, governed, and ready for use. These disciplines form five foundational pillars that shape how data is created, managed, and shared across the business.

These pillars aren't buzzwords or abstract frameworks. They're the practical foundation that makes data work in the real world. Together, they turn vision into discipline and ensure that data keeps up with the speed and precision the business expects:

1. Architecture
: Sets the blueprint for how data flows and connects across systems

2. Governance
: Defines ownership, accountability, and the rules of responsible use

3. Quality and Observability
: Makes reliability visible and measurable

4. Metadata
: Provides the context that turns data into knowledge

5. Access and Security

Controls who can use which data, under what conditions, and with what protections

Together, these five pillars form the structural foundation that enables data to be trusted, discoverable, and usable across the enterprise, as illustrated in Figure 3-1.

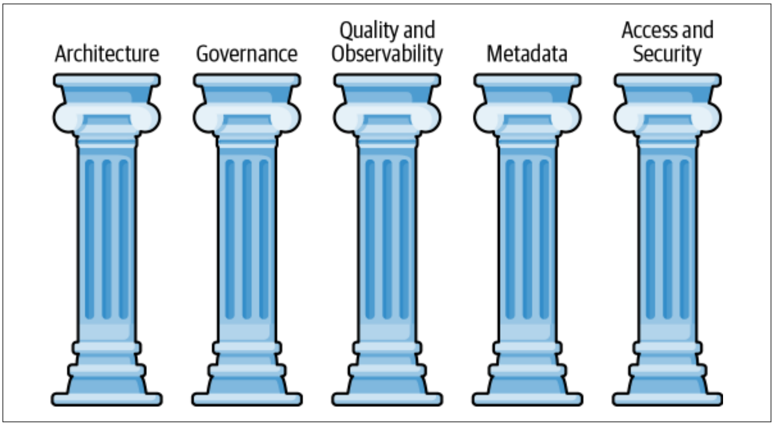

Figure 3-1. The five foundational pillars of data management working together to support trusted, scalable data use

We explore the pillars in this order because they follow the natural journey of how data gains value. Architecture lays the groundwork for movement and connection. Governance defines how that flow is controlled. Quality and Observability ensures that what moves through those systems can be trusted. Metadata gives that data meaning and discoverability. And Access and Security ensures that it can be safely used by the right people at the right time.

When these pillars work together, they create a self-reinforcing system—one that balances flexibility with control and makes trust the default rather than the exception. In the sections that follow, we'll look at each pillar in turn: what it is, why it matters, the common pitfalls that signal trouble, and what good looks like in practice.

Pillar 1: Architecture

What it is and why it matters: Architecture shapes how data is collected, stored, integrated, and made available across the enterprise. Without the right architectural foundation, data becomes fragmented, inconsistent, and difficult to scale. Architecture determines not only performance and cost but also whether teams can collaborate effectively. A well-designed architecture enables agility, supporting new business initiatives, AI workloads, and analytics use cases without requiring constant rework.

Just as importantly, architecture provides the guardrails for long-term resilience. A poorly designed system might deliver quick wins but will buckle under regulatory pressure, security demands, or sudden shifts in business strategy. When architecture is treated as a strategic asset rather than a one-time project, it gives organizations confidence that they can innovate responsibly, knowing the foundation will flex as data volumes grow, compliance rules tighten, or new technologies emerge.

At its core, data architecture is the blueprint for how data flows across the organization. Examples range from centralized warehouses to federated approaches, from data lakes to lakehouses, to hybrid designs that combine elements of each. Beyond structure, architecture establishes the principles for interoperability: how data moves between systems, how metadata is captured and shared, and how governance and security are applied consistently. A well-designed architecture is not about choosing tools; it is about how those tools work together to serve business outcomes.

Modern data architectures also increasingly rely on abstraction layers such as APIs, data fabrics, and data virtualization to hide underlying complexity from users. This shift reflects the reality that organizations rarely operate on a single platform. Instead, architecture must knit together legacy systems, SaaS applications, and cloud-native services, while still presenting a coherent view of data to business and AI systems. The measure of success is not how "modern" the components look in isolation, but whether the overall architecture reduces friction, accelerates access, and adapts as new use cases arise.

What to watch for: In the context of architecture, beware of "tool sprawl," where every team adopts its own platform without regard for enterprise integration. Watch for brittle pipelines that break under growth, or rigid architectures that lock the organization into outdated patterns. The goal is not to chase the latest trend, but to design architectures that fit business needs, balance centralization with flexibility, and evolve over time.

Example in practice: A European retailer routed all sales, inventory, and customer data through a centralized warehouse. While this kept management reports clean and consistent, innovation suffered. Regional teams waited three to four weeks for headquarters to onboard each new data source, which was long enough to miss promotional windows and lose revenue opportunities. The fix was a hybrid architecture: centralized systems for financial reporting, federated domains for regional analytics. The shift cut onboarding time to less than a day, gave regional teams the freedom to experiment, and preserved enterprise-wide consistency for leadership.

Pillar 2: Governance

What it is and why it matters: Governance establishes the rules, roles, and processes that ensure data is used responsibly and consistently. It defines ownership, sets policies, and embeds mechanisms such as governance by design and data contracts. When implemented well, governance creates accountability and confidence rather than bureaucracy.

Modern governance practices emphasize embedding controls directly into data systems and workflows instead of layering them on afterward. This approach, often called *governance by design*, ensures that compliance, security, and quality checks are part of everyday operations. A related concept, the *data contract*, is a formal agreement between data producers and consumers that defines expectations for quality, freshness, and schema stability. Together, governance by design and data contracts turn policies from static documents into active safeguards that operate automatically.

Governance is not only a compliance requirement; it is the foundation of trust. When ownership is clear and policies are enforced automatically, teams can innovate without hesitation. Governance also underpins ethical responsibility. As AI adoption accelerates, organizations must ensure that data is used fairly, transparently, and with respect for privacy and consent.

What to watch for: Governance can fail in two ways—by being too heavy-handed, creating bureaucracy that slows down teams, or by being too lax, leading to confusion and risk. Look for signs of unclear accountability (such as datasets with no known owner, or policies that are documented but not enforced). The goal is governance that is practical, lightweight, and embedded in day-to-day operations.

Example in practice: A financial services company in the Nordics introduced data contracts between data and business teams. Each contract specified expectations for freshness, schema stability, and quality thresholds. By automating enforcement in the data pipeline, governance became invisible but effective. The result was fewer surprises, fewer broken reports, and more trust between business and technology teams.

Pillar 3: Quality and Observability

What it is and why it matters: The Quality and Observability pillar determines whether data can be trusted when it is used. Quality refers to the accuracy, completeness, and reliability of data. Observability is the ability to monitor and measure that quality in real time through indicators such as freshness, lineage, and anomaly detection. Together, they turn data management from a reactive process into a proactive discipline that strengthens every stage of the data lifecycle.

Quality goes beyond simple correctness or formatting. It also includes context and meaning. For example, a hospital's patient record may show the correct patient ID and demographics but still be dangerously incomplete if allergy information or recent lab results are missing. In finance, a transaction feed may be technically valid but misleading if settlement dates or counterparties are absent. Observability provides the visibility to catch these gaps before they affect patient safety, regulatory compliance, or business performance.

Modern observability practices draw from software engineering, where monitoring and logging are built into systems by default. Applied to data, this means embedding quality checks, automated alerts, and lineage tracking directly into pipelines so teams can see where data came from and how it has changed. By making data quality transparent and measurable, organizations build confidence in their information and reduce the cost of downstream fixes.

Observability also creates transparency between technical and business teams. When users can see indicators such as freshness, completeness, and anomalies alongside the data itself, they can make informed decisions, even when minor imperfections exist. This shared visibility builds a culture where quality is everyone's responsibility, not just the data team's.

What to watch for: If users cannot tell whether data is current or complete, they will hesitate to use it or make decisions based on it. Watch for manual, reactive processes where teams discover problems only after business stakeholders complain. Mature organizations design observability into their systems so quality issues are detected and resolved before they cause downstream impact.

Example in practice: An airline deployed data observability dashboards to track flight operations in real time. When anomalies appeared, such as missing updates from airports or irregular delay codes, the system flagged them automatically. Analysts could respond immediately, preventing errors causing schedule delays or incorrect customer notifications. Over time, confidence in the dashboards grew because teams trusted that issues would be detected and addressed before they impacted passengers.

Pillar 4: Metadata

What it is and why it matters: Metadata is data about data—the context that describes what a dataset is, where it came from, who owns it, and how it has been used. It includes catalogs, lineage tracking, and business glossaries that make data discoverable and understandable. Metadata serves as the connective tissue between raw data and human understanding. It translates tables, files, and streams into meaningful assets by attaching business definitions, usage notes, and ownership details.

At the heart of this context is semantics: the shared meaning of business terms, metrics, and relationships. Without agreed semantics, two teams can look at the same dataset and reach different conclusions. Many organizations operationalize semantics through semantic layers and knowledge graphs, which provide a consistent, machine-readable understanding of concepts such as customers, products, and revenue across analytical, operational, and AI use cases. These structures help ensure that data is not only discoverable but interpreted consistently wherever it is used.

Modern metadata systems go beyond static documentation. They capture relationships, usage patterns, and quality signals in real time, creating a "living blueprint" of the data landscape. This allows organizations not only to answer, "What is this dataset?" but also "Who has used it, how trustworthy is it, and how does it connect to the rest of the ecosystem?" In this way, metadata becomes the foundation for governance by design, enabling AI-driven discovery and surfacing recommendations that reduce friction for business users.

Without metadata, even high-quality data remains underutilized. Teams spend time searching for the right dataset, duplicating work, or questioning definitions. Metadata provides the context necessary for collaboration and reuse by clarifying ownership, lineage, and purpose. It also underpins compliance by making it possible to trace how data was created, transformed, and used.

Equally important, metadata turns data from a technical artifact into a business asset. When datasets are treated as data products, metadata acts as their labeling: definitions, quality indicators, and usage notes that make them consumable and reusable. Paired with data contracts that set clear expectations for delivery and quality, metadata shifts conversations from "Can we trust this?" to "How can we use this to drive additional value?" This confidence and accessibility unlocks innovation, shortens time to insight, and reduces reliance on experts who otherwise serve as institutional gatekeepers.

What to watch for: Beware of catalogs that are populated but not maintained, or lineage diagrams so complex that they are ignored. Metadata initiatives often fail because they are treated as one-off projects rather than living systems. The real test of success is whether business and technical users can actually find and understand the data they need when they need it.

Example in practice: At a pharmaceutical company, scientists once spent weeks piecing together trial data from disparate systems, relying on informal knowledge of where to find it. To address the problem, the company rolled out an enterprise data marketplace. Instead of raw tables, the marketplace presented curated data products such as "clinical trial outcomes" and "adverse event reports." Each product came with rich metadata packaging—business definitions, lineage, quality scores, and ownership details.

To ensure reliability, the marketplace also employed data contracts between producers and consumers, specifying expectations for update frequency, schema stability, and quality thresholds. These contracts operationalized governance, giving scientists confidence that the data products they used would remain consistent and trustworthy.

The result was transformative. Analysts and researchers could browse for data products as easily as shopping in an online store, request access with a click, and trust that what they received was accurate, current, and compliant. The marketplace turned metadata from a back-office function into a shop window for discovery and compliance, accelerating research timelines while reducing reliance on tribal knowledge and IT bottlenecks.

Pillar 5: Access and Security

What it is and why it matters: The Access and Security pillar defines who can use data, under what conditions, and with what safeguards. This includes balancing democratization, i.e., making data broadly available, with the need to protect sensitive information. It establishes the frameworks that determine how data is accessed, shared, and monitored across the enterprise.

At its core, this pillar is about trust. Employees need confidence that the data they're using is both safe and appropriate for their role, while organizations need assurance that sensitive assets are not misused. Effective access and security practices combine technical controls such as authentication, encryption, data masking, and audit trails with clear governance policies that align with regulatory and ethical standards.

Modern practices emphasize "just enough access" rather than "all or nothing." This means applying fine-grained permissions, dynamic masking, and real-time monitoring to ensure that the right people can work with the right data, in the right context, without friction. In effect, security is embedded into the flow of work, allowing democratization and protection to coexist rather than compete.

Data that is locked away provides no value, while data that is too open creates risk. Striking the right balance is critical for innovation and compliance alike. When access processes are clear, automated, and aligned to business needs, projects move faster and employees are less likely to resort to risky workarounds. Conversely, if security is viewed as an obstacle, it undermines both compliance and culture.

Done right, this pillar turns security into an enabler of trust, speed, and collaboration.

What to watch for: Rigid access controls that prevent teams from doing their work are a red flag, as are lax policies that expose sensitive data. Watch for shadow practices, such as employees exporting datasets to spreadsheets or sharing them informally, which often indicate that official access is too restrictive. The goal is controlled democratization, i.e., policies and platforms that make data accessible to those who need it, with safeguards proportionate to the risk.

Example in practice: A healthcare provider implemented role-based access with fine-grained controls. Doctors and nurses could access patient records relevant to their treatment roles, while researchers saw only anonymized datasets. Security was enforced automatically, but access was seamless, reducing the need for workarounds and ensuring compliance with the Health Insurance Portability and Accountability Act (HIPAA) of 1996 and Europe's General Data Protection Regulation (GDPR).

Bringing the Pillars Together

Each pillar is essential, but none stands alone. Architecture without governance creates technical debt. Governance without quality and observability erodes trust. Metadata without access is just documentation, not discovery. Access without security introduces risk. And without a culture that embraces these principles, even the best-designed systems fail to deliver.

When the five pillars work together, they create the foundation for treating data as a product. Architecture ensures those products are built on scalable platforms. Governance, expressed through data contracts, makes expectations explicit and enforceable. The Quality and Observability pillar provides the assurance that data products are reliable. Metadata gives them their packaging through definitions, lineage, and usage notes that make them discoverable and understandable. The Access and Security pillar defines how those products are shared responsibly across the organization.

The strength of an organization's data management lies not in the existence of any single pillar, but in how they interlock to form a stable, evolving foundation. When designed and maintained together, the pillars shift enterprises from fragmented, reactive practices to cohesive, proactive strategies. They enable transparency for trust, scalability for growth, and adaptability for change. The outcome is not just compliance or efficiency; it is the ability to innovate responsibly, empower employees with confidence, and unlock the full value of AI and analytics.

Many of the challenges described in this report (silos, inconsistent metrics, and the need for lineage) are rooted in a long-standing architectural pattern: the separation of operational systems (OLTP) from analytical systems (OLAP). For decades, organizations have isolated production workloads from large analytical reads to protect performance and reliability. While the technologies have evolved from traditional extract, transform, load (ETL) pipelines, to extract, load, transform (ELT), to streaming platforms, and now to vector databases supporting retrieval-augmented generation (RAG), the underlying pattern has remained consistent. Data is extracted from source systems, transformed, and prepared for consumption elsewhere.

This separation is precisely what makes lineage, governance, and quality controls essential. Every transformation between source and consumption introduces the risk of semantic drift, data loss, or misinterpretation. Lineage provides the visibility needed to understand how data has moved and changed, while governance and observability ensure those transformations remain intentional, accountable, and trustworthy.

When these disciplines connect, they create the conditions for trusted, reusable data products that can be easily discovered, governed, and consumed. Many organizations express this maturity through an enterprise data marketplace: an environment where data products are cataloged, quality is visible, and access is managed through clear rules and automation. The marketplace is not a sixth pillar, but a reflection of what happens when the five pillars operate together effectively, turning good data management into a living system that accelerates innovation.

CHAPTER 4

Putting the Foundations into Action

Strong data foundations are built gradually. The most successful organizations do not try to fix everything at once. They start with focused initiatives that prove the pillars can work in practice, show value quickly, and create patterns that scale. This chapter outlines how to apply the foundational pillars through pilot projects, smart prioritization, clear roles, and repeatable templates. Over time, these efforts accumulate into an enterprise-wide foundation that supports innovation, agility, and trust.

One practical way to decide where to begin is to balance effort against business impact using a simple prioritization model like the one shown in Figure 4-1.

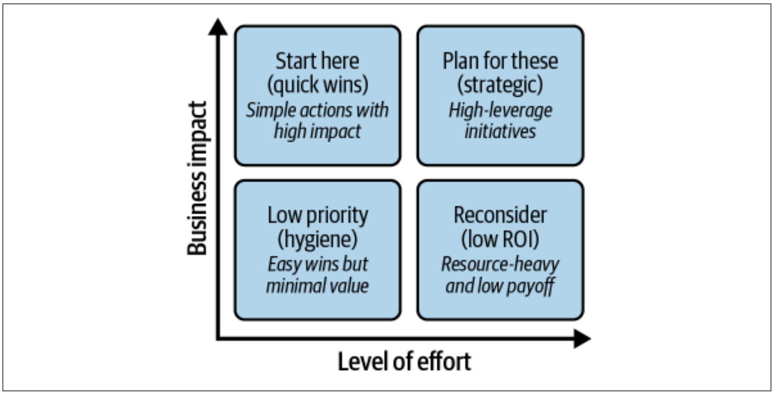

Figure 4-1. Prioritizing foundational data initiatives by business impact and implementation effort to distinguish quick wins, strategic investments, hygiene tasks, and low-return work

Start Small with Pilot Projects

The temptation in enterprise data programs is to attempt a complete overhaul. Clean every dataset. Replace every platform. Rewrite every governance policy. In reality, this approach delays results and drains teams.

Organizations that succeed pick one well-defined pilot that exercises one or two pillars meaningfully. A good pilot does not attempt to solve the entire foundation. It proves that the foundation works.

A good pilot shares three qualities:

Characteristic	What it means	Pillars demonstrated
Clear scope	Limited enough to complete in weeks or months, not years	Architecture, Quality and Observability
Visible impact	Solves a real problem for a defined group of stakeholders	Governance, Access and Security
Reusable patterns	Produces templates that can be extended to other domains	Metadata, data-as-product mindset across all pillars

Example in practice (Quality and Observability pillar): A consumer bank piloted data quality improvements in its mortgage application workflow. By improving a small number of critical fields, manual rework dropped and approvals accelerated. Once the monitoring patterns proved reliable, the same approach was extended to credit cards and personal loans. The pilot demonstrated the Quality and Observability pillar in action and produced reusable standards for other domains.

Prioritize by Pain Point or Business Value

Choosing where to start affects how the foundations are perceived. When trust is low, fixing visible pain builds credibility. When alignment is high, investing to support strategic initiatives accelerates value.

There are two prioritization strategies:

- *Start with pain points* when reports are disputed, dashboards are ignored, or business users lack confidence in the data. This supports rebuilding trust across the Quality and Observability and Governance pillars.
- *Start with business value* when a strategic initiative exists, such as AI personalization or real-time customer experiences. This requires strengthening Architecture, Metadata, and Access and Security to support scale.

Example in practice (Metadata and Governance pillars): A global manufacturer faced repeated delays in regulatory reporting. Instead of reworking every financial dataset, the company focused on adding clear metadata and governance controls only to regulatory reporting-critical datasets. Reconciliation time dropped, audit readiness improved, and trust increased. This was a pain-point priority that proved the Metadata and Governance pillars in a highly visible environment.

Establishing Essential Roles

Technology alone cannot sustain strong foundations. Clear ownership and collaboration across roles prevent backsliding.

Let's look at four essential roles:

Role	Responsibility	Pillars influenced
Data owners	Provide accountability for critical datasets. Ensure data remains accurate and aligned with business needs.	Governance, Quality and Observability
Data stewards	Maintain definitions, apply quality rules, manage documentation. Translate policy into something teams can actually use in practice.	Metadata, Governance
Data engineers and architects	Design and build the pipelines, maintain the platforms, and figure out how all the pieces fit together for scale.	Architecture, Quality and Observability
Business stakeholders	Define meaning and ensure metrics align to business outcomes.	Data-as-product mindset across all pillars

When these roles work in isolation, gaps appear. When they share accountability, the pillars strengthen one another.

Example in practice (Governance and Metadata pillars): A healthcare provider assigned data owners within clinical teams and paired them with IT data stewards. Ownership drove accountability, while stewardship ensured data quality and documentation. The result was cleaner patient records and higher trust in clinical decision systems.

Scaling What Works

Pilots prove the concept, but scaling is where the payoff comes. The challenge is to avoid reinventing the wheel with every new domain. Successful organizations capture what worked in the pilot and create templates from these approaches. The next team shouldn't have to start from scratch; they should be able to reuse proven patterns.

Scaling doesn't mean copy-paste. Each domain has its quirks. But with a shared foundation, 80% of the work is standardized and only the last 20% is domain-specific. That balance creates both speed and consistency.

Example in practice (Quality and Observability, Metadata, and Architecture pillars): An insurance company improved data quality in the claims process, then packaged the rules and observability dashboards into templates. When the policy administration team adopted them, rollout time was reduced and common mistakes were avoided. Templates operationalized the Quality, Metadata, and Architecture pillars across departments.

Avoiding the "Big Bang" Trap

It's worth stating plainly: I have never seen the big-bang approach work. Ambitious programs that try to fix everything at once run out of steam. Stakeholders lose interest, budgets get cut, and the organization is left with half-finished platforms and no measurable value.

The better path is incremental and agile, focused on small, testable improvements that deliver value, generate feedback, and build momentum over time. Build credibility with small wins. Celebrate each success and use it to fund and justify the next. Keep connecting every initiative back to business impact. Over time, the small wins accumulate into enterprise-wide progress. It looks slower at the start, but it moves faster in the long run because adoption sticks and trust grows with every step.

Five Signs Your Foundations Are Taking Hold

Organizations with strong foundations show recognizable patterns:

- *Data is trusted* and used without secondary validation.
- *Ownership is clear* and accountability is visible.
- *Issues are detected early* through observability rather than support tickets.
- *Access is balanced*, enabling use without exposing risk.
- *Data strategy drives outcomes* tied to revenue, cost reduction, experience, or AI readiness.

Example in practice: A global logistics company introduced observability dashboards for shipment data. Teams shifted from reconciling errors to resolving operational issues. The change was visible on the warehouse floor, not just in reporting decks. Confidence grew because teams trusted that data issues would be surfaced before decisions were made. This demonstrated the Quality and Observability pillar, working in harmony with Architecture and Governance.

Bringing It Together

Pilot projects prove that foundations work in practice. Prioritization ensures effort impacts what matters most. Well-defined roles sustain progress. Templates enable scale. Together, these practices convert the pillars from theory into an operating model that the organization can rely on.

The result is not just fewer data problems. It is faster decisions, clearer accountability, and a foundation strong enough to support AI, analytics, and innovation at scale.

CHAPTER 5

Your Action Plan for the First 90 Days

Strong data foundations are built through steady, visible progress. The goal in the first 90 days is not to redesign architecture, rewrite governance, or clean every dataset. It is to demonstrate momentum, build confidence, and create patterns that can scale. This chapter provides a practical action plan for the first three months, focused on setting ownership, choosing the right pilot, establishing trust signals, and aligning teams around shared definitions and outcomes. These actions are small by design. Each one contributes to a foundation that becomes stronger, more resilient, and easier to expand over time.

Rather than treating the first 90 days as a checklist, it can be more useful to view this period as a progression of focus areas and outcomes, as shown in Figure 5-1.

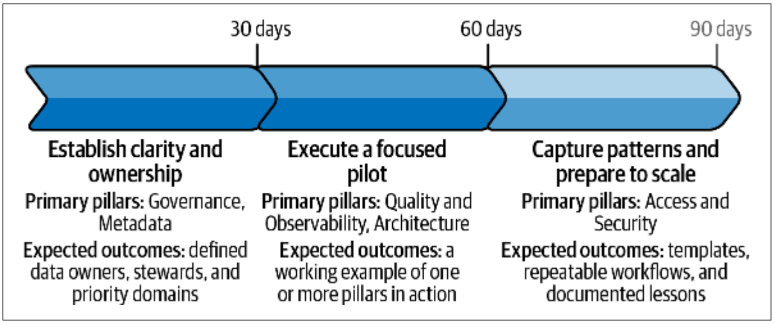

Figure 5-1. A 90-day phased approach to building data foundations, progressing from ownership and clarity to pilot execution and scalable practices

By the end of the first 90 days, the organization should have a clear source of truth for at least one business-critical dataset, improved trust in its quality, observable signals that show how data is performing, and shared ownership across business and technical stakeholders. The foundation does not need to be complete to be effective. It only needs to be real, visible, and repeatable.

Days 1–30: Establish Clarity and Ownership

Pillars in focus: Governance, Metadata

Objective: Create visibility into critical data assets and assign accountability.

Roles involved: Data owners, business leaders, data architects

Actions:

- Identify the 10 to 15 datasets most critical to decision making or compliance.
- Assign a named data owner for each dataset, responsible for quality and usability.
- Baseline current quality, lineage, and access issues.
- Create shared definitions for 5 to 10 core metrics such as customer satisfaction, revenue, and churn.

Measurable outcomes:

- A published list of critical datasets with assigned owners
- Initial quality scores (freshness, completeness, accuracy) for those datasets
- A business glossary for the most-used metrics

Days 31–60: Build Early Controls and Monitoring

Pillars in focus: Quality and Observability, Architecture

Objective: Put lightweight structures in place to improve reliability and transparency.

Roles involved: Data stewards, engineers, governance leads, analytics teams

Actions:

- Implement basic data contracts for two or three critical datasets.
- Address the most critical data quality issues in the pilot datasets, focusing on fixes that reduce manual rework or prevent downstream breakage.
- Pilot a data observability dashboard for one or two high-value pipelines.
- Document lineage for key datasets to clarify how data flows and transforms.
- Train teams to interpret and act on data quality and lineage signals.

Measurable outcomes:

- Early warning alerts for data anomalies
- At least one schema change caught before breaking downstream reports
- Lineage diagrams visible to both business and technical users
- Positive feedback from teams using the pilot systems

Days 61–90: Embed Practices and Align with Business Goals

Pillar in focus: Access and Security

Objective: Connect foundational improvements to business value and demonstrate momentum.

Roles involved: Executives, data product managers, security and compliance, business stakeholders

Actions:

- Review and refine access policies to balance democratization with security.
- Extend observability practices to additional pipelines or domains.
- Demonstrate how foundational improvements supported one real initiative, such as faster reporting or more accurate AI outputs.
- Share progress with executive sponsors and reaffirm commitment to the next phase.

Measurable outcomes:

- Reduced manual reconciliation or data preparation effort
- One business initiative showing measurable improvement due to stronger data foundations
- Updated access policies reviewed and approved
- Leadership endorsement to expand the program

How to Use This 90-Day Action Plan

This plan is a structured starting point. It builds trust and momentum by delivering visible improvements early. After 90 days, the next phase focuses on scaling: extending contracts, observability, metadata, and governance practices across more domains and embedding literacy across teams.

Looking Ahead

The first 90 days are about building credibility: assigning ownership, piloting controls, and demonstrating tangible business value. But strengthening data foundations is not a sprint; it is a long-term capability that must evolve with the organization.

Beyond the initial playbook, leaders should shift focus to scaling and sustaining:

- *Scale practices* across domains by extending contracts, observability, and metadata to a broader set of datasets and pipelines.
- *Embed governance by design* so that quality checks, lineage tracking, and access controls are built into workflows rather than bolted on later.
- *Invest in literacy and culture*, ensuring that employees at all levels have the skills and confidence to use data responsibly.
- *Measure business outcomes,* not just technical outputs, to keep the program aligned with strategic priorities like AI readiness, regulatory compliance, and customer experience.

Organizations that treat this plan as a launchpad, rather than an endpoint, build a foundation that supports innovation and resilience. Over time, data management shifts from a reactive cost to a proactive advantage.

Bringing It All Together

Improving data foundations is not about perfection. It is about progress that compounds. Every clarified definition, every pipeline that becomes reliable, every business team that gains confidence in the numbers contributes to a shift in how the organization thinks and operates. Over time, that shift becomes culture.

What we have outlined in this report is a path that any enterprise can follow. It begins with recognizing the patterns that hold organizations back, moves toward defining what "good" looks like in practice, and then introduces the pillars and behaviors that support it. The first 90 days provide a manageable, concrete place to begin. They show that change is possible, that trust can be rebuilt, and that teams can work differently with the right structures in place.

As momentum builds, the work moves from being a project to becoming a habit. Conversations change. Meetings focus on decisions instead of reconciliations. Teams share the same language for data and outcomes. And the organization gains confidence in its ability to respond to new opportunities, market shifts, and emerging technologies.

This is the real value of strong data foundations: they make the organization more resilient and more adaptable. They enable innovation not by adding complexity, but by removing friction. They create the conditions where analytics and AI can deliver meaningful impact because the underlying data is reliable, understood, and governed with intent.

The work continues beyond the first 90 days, but the payoff begins almost immediately. Trust grows. Decision cycles shorten. People have the confidence to act. And the organization moves forward with clarity rather than hesitation.

The opportunity is within reach. The foundations are the starting point. The next step is simply to begin.

About the Author

Emma McGrattan is the chief technology officer at Actian, where she leads the strategy behind some of the industry's most innovative data solutions. With over two decades of experience delivering mission-critical data technologies, she has helped organizations across industries design and modernize their data architectures, embed governance by design, and become AI-ready. A recognized thought leader and sought-after speaker, Emma is known for turning complex data challenges into clear, actionable blueprints. Her work bridges the gap between technology and business, enabling teams to move faster with trusted, scalable data. This report draws on her hands-on experience and no-nonsense approach to understanding and building data systems that actually work.