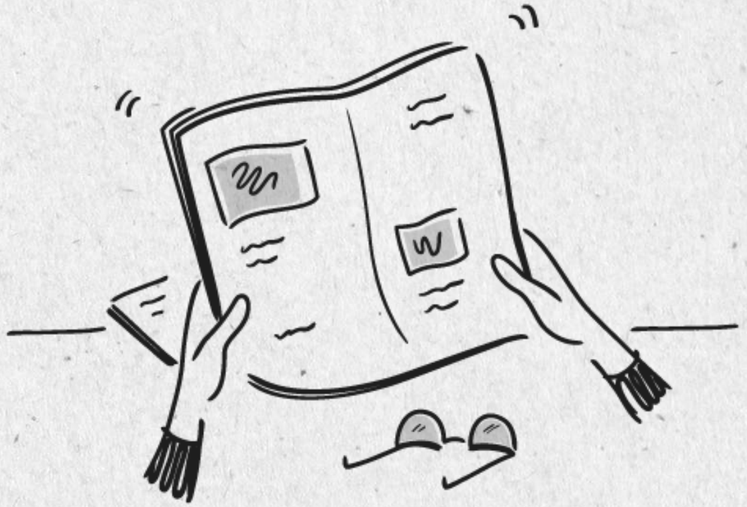
STATE OF DATA PRODUCTS

QUARTERLY RELEASE: COMMUNITY CURATION

Stepping Stones to Become AI-Natives

Bridging the Gaps to AI-Readiness

Closing the AI readiness gap—industries race to redesign data architecture with Lakehouse 2.0 and modular data developer platforms..



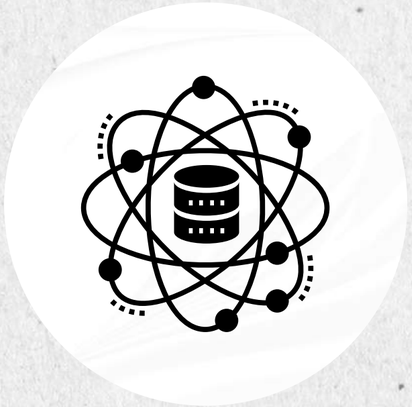
A TECH-REPORTING INITIATIVE
RITWIKA CHOWDHURY X MODERN DATA 101

WHAT'S IN STORE

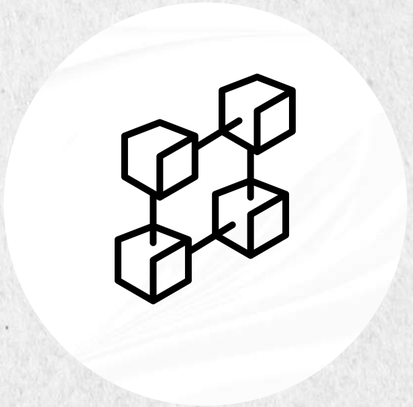
The Lack of AI Readiness in Organisations

Adopting Modern Design Principles for Data Platforms

Driving the Value of Metadata



The Data Lakehouse 2.0 Vision



Architectural Changes through Data Developer Platforms

...and so much more!

Editor's Note: 2025 Q1

A Boardroom Breakthrough: Will Lakehouses Change the AI & Data Trajectory?

With generative AI sparking a revolution, business leaders are forced to rethink their data infrastructures. Traditional systems, bogged down by complex ETL pipelines and stale data, no longer cut it in a world that demands real-time insights and agile decision-making.

Despite heavy investments in generative AI, many companies find themselves mired in a maze of data siloes and an overwhelming array of tools and vendors. These fragmented infrastructures create bottlenecks that delay insights and stifle agile decision-making.

The promise of AI falls flat when critical data isn't readily accessible or actionable.

The answer to this conundrum?
A new vision of the Lakehouses—delivering the agility and real-time analytics needed to transform data chaos into a strategic asset.

For the modern CDO, this isn't just tech talk—it's about winning in the marketplace. With Lakehouses, companies can harness AI-driven insights to streamline operations, reduce costs, and stay ahead of the competition. Yet, like any bold new strategy, this approach comes with its own set of challenges, requiring a keen eye on innovation and meticulous execution.

From The Editor's Desk,
Modern Data 101

2025 Kicks Off with Soaring Agents

As 2025 unfolds, organizations are diving headfirst into the generative AI (Gen AI) revolution. No longer content with mere experimentation!

Organizations are proactively establishing structures and processes to harness their potential effectively. Companies are redesigning workflows, enhancing governance frameworks, and addressing emerging risks associated with AI implementation. This proactive approach is crucial for integrating AI into business operations while ensuring compliance and ethical standards.

Who's the catalyst for these changes?

"Strategic Leadership Is Non-Negotiable."

- Dion Hinchcliffe, the Vice President of CIO Practice at The Futurum Group

Today, nearly 57% of CEOs personally lead their organization's AI strategy, underscoring the critical importance of direct executive involvement in driving innovation and ensuring enterprise-wide alignment. (Source)

More numbers on the case:

- 28% of organizations utilizing AI report that their CEO oversees AI governance.
- 17% of these organizations indicate that their board of directors holds this responsibility.
- Many organizations report that AI governance is jointly managed by multiple leaders.

(Source)

The tangible value of AI emerges from fundamentally reengineering business operations. We refer to this as AI-Readiness.
However, despite significant breakthroughs in AI across various sectors, many organizations struggle to truly identify themselves as AI-Ready.

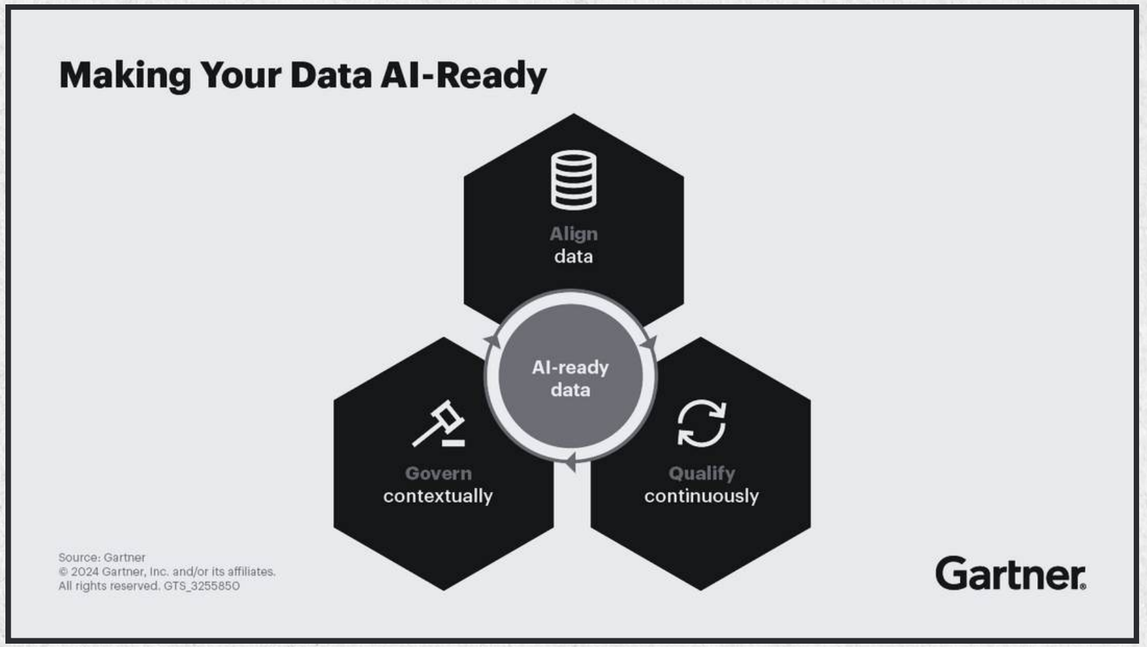
Reality Check: Not All Can Keep Up With the Agents

The year 2025 has begun with AI at full throttle, dominating boardroom discussions and technology roadmaps. Yet, as AI investments soar, a stark reality is settling in: Enthusiasm alone isn't enough.

Despite 90% of companies investing in AI, fewer than 2 in 5 report meaningful business gains. The disconnect?

AI can't generate value without AI-ready data.

In fact, Gartner forecasts that through 2026, organizations will actually abandon 60% of AI projects unsupported by AI-ready data.



The pillars of AI readiness | Source: Gartner

"There is no AI strategy without a Data Strategy."

Sridhar Ramaswamy, in one of his articles, mentions how (Source).

ln most cases, CEOs and data leaders we talk to realise how crucial data readiness is for their AI initiatives. The truth? AI won't scale without a solid data strategy—yet many CEOs see data as a blocker.

The AI Gold Rush: Why Data Readiness is a Missing Link?

AI doesn't just require data — it demands AI-ready data. While many organizations have solid data management foundations, most lack the agility, structure, and metadata practices necessary to support AI's evolving needs. Gartner emphasizes that AI-specific data innovations, such as vector data stores, chunking, sampling, embedding, and retrieval-augmented generation (RAG), must be iteratively integrated into existing data management frameworks.

"It is a process and a practice based on the availability of metadata to align, qualify and govern the data," ~ mentions Rital Sallam, Chief of Research, Data, Analytics and AI.

Minor data errors can cause major fallout. Without robust quality safeguards, AI outputs become unreliable—limiting real-world adoption as businesses lose trust. But to make this shift, they need a structured, scalable approach—one that aligns with the three pillars of AI readiness:

The Pillars of AI Readiness

- Align: AI must learn from real-world patterns, errors, and outliers, not just clean, idealized datasets.
- Govern: Context is king; metadata-driven governance ensures AI models evolve dynamically.
- Qualify: Data readiness is a continuous process, requiring refinement and adaptability.

Many organizations still rely on static data management, leaving AI pipelines built but starved of the context-rich, dynamic, and governed data needed for real impact.

The result? AI pipelines are built, but AI remains starved of the high-context, structured, and dynamically governed data it needs to be truly operational.

A Community of Data Leaders Resonate on Needing a Trusted Data Foundation!

Jonathan Eide
VP, Business Operations + Data Science/Engineering + TPM, Coinbase

2. The Foundation: Data Quality is Non-Negotiable

For AI: **garbage in → garbage out**.

Most analytics questions are simple and well structured. What makes it hard for AI to deliver correct answers is the fact that data models can be messy, hence getting the right answer is not a statistical or analytical challenge, but a data organization problem. We can't have models guessing ("hallucinating"), misinterpreting basic queries, or producing outputs that are directionally right, but specifically wrong. Therefore our priority is ensuring a **trusted data foundation**:

Bernard Marr
Internationally Best-selling #Author #KeynoteSpeaker #Futurist #Business, #Tech & #Strategy Advisor

The Hidden Obstacles To AI Success

While investment enthusiasm runs high, implementation reality is more sobering. The **CDO Insights 2025 report** revealed that a staggering 97% of organizations struggle to demonstrate business value from their generative AI investments—a significant roadblock to securing continued funding and executive buy-in.

Technical challenges are equally prevalent, with 92% reporting that issues like data quality, responsible AI use, and compliance concerns are delaying the progression from proof-of-concept to production. As Roscoe emphasized, companies are grappling with questions like: "How will this AI model behave? What risks are associated with putting this into production?"

In my experience working with organizations across sectors, I've observed additional barriers: executive teams not thinking ambitiously enough, scattered pilot projects without strategic cohesion, leadership knowledge gaps, skills shortages, cultural resistance, and data silos that prevent AI from reaching its potential.

George Firican • 1st
Award Winning Data Governance Leader | DataVenger | Founder of Lig...
View my services
4mo •

The bottom line?
Invest in data governance before you invest in AI. What do you think?

Most relevant

T. Scott Clendaniel • Following 4mo
100K LinkedIn Followers | UPenn Wharton #AI | Gartner Director | On a ...

George Firican "Truer words were never spoken." - Batman

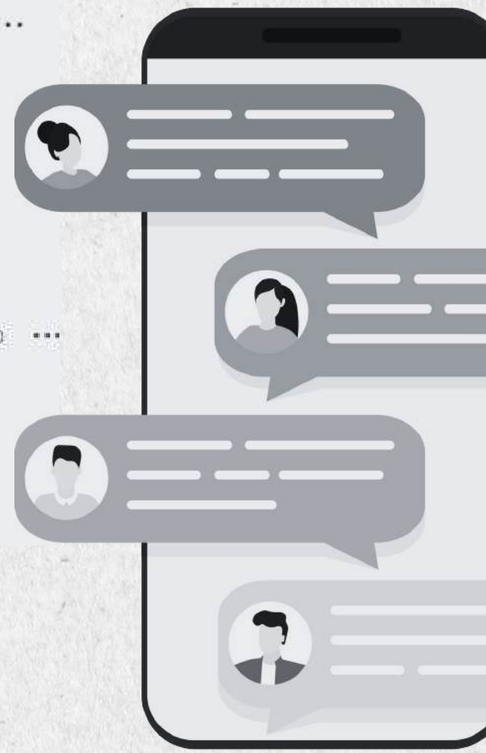
AI without Data Governance is like building a skycraper on a swamp.

Eric Mack • 2nd 3mo
Helping Businesses Successfully Share Their Data using Maxxphase Dat...

Great insight Data gov quality trusted data, but also when leveraged correctly becomes the back bone of data sharing through out the enterprise. Data Standards, driven by data gov, are going to be key not only for AI, but any data driven initiatives. Thanks for the vid enjoyed it!

Fred Lardaro • 2nd
I sell SkyBell... the best Wi-Fi video doort

Data Governance is good Finance



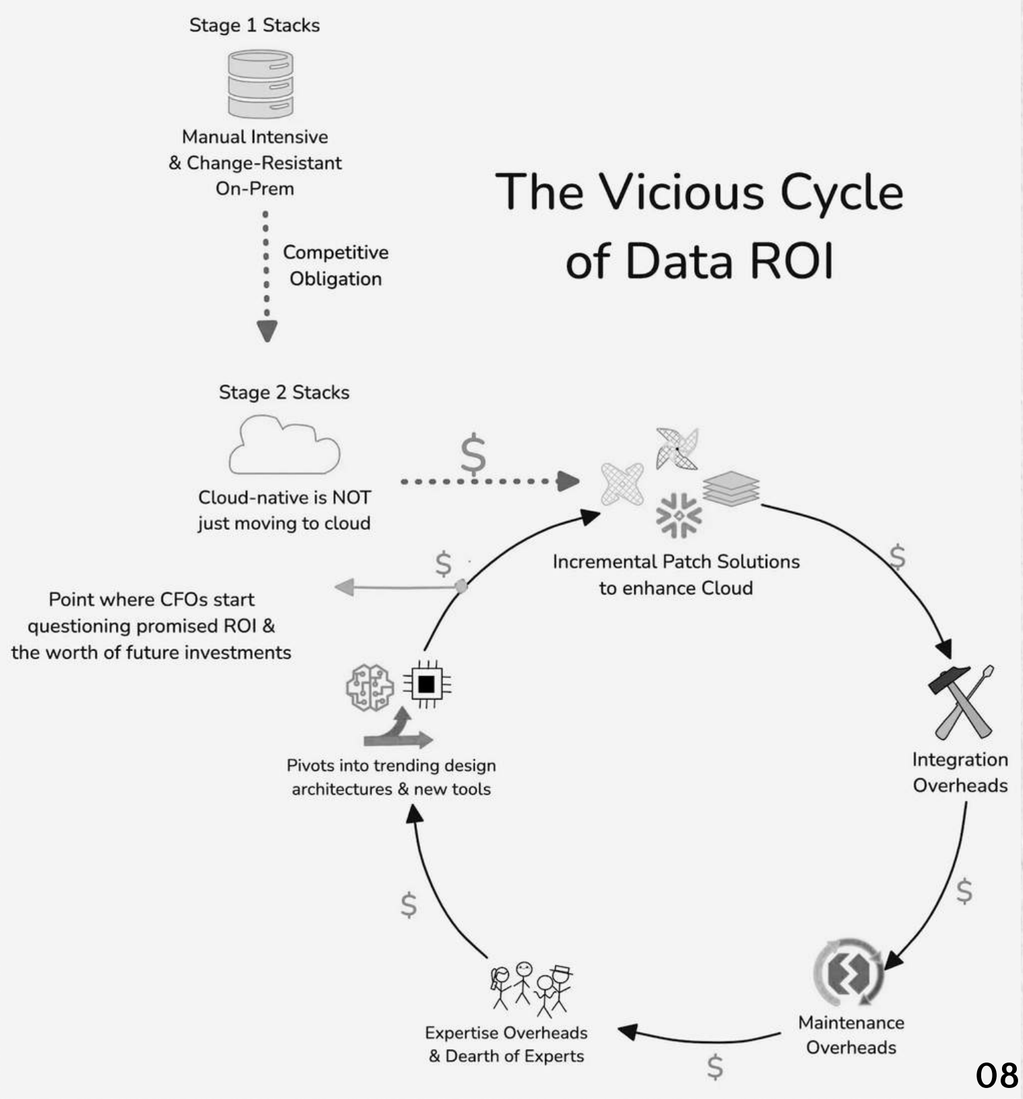
Is the Conventional Data Architecture to Blame?

A patchworked data stack is not benefiting the aspiring data & AI roadmap. It's important to go agile, modular, and composable. And what's critical is to find and address the root cause of the data problem.

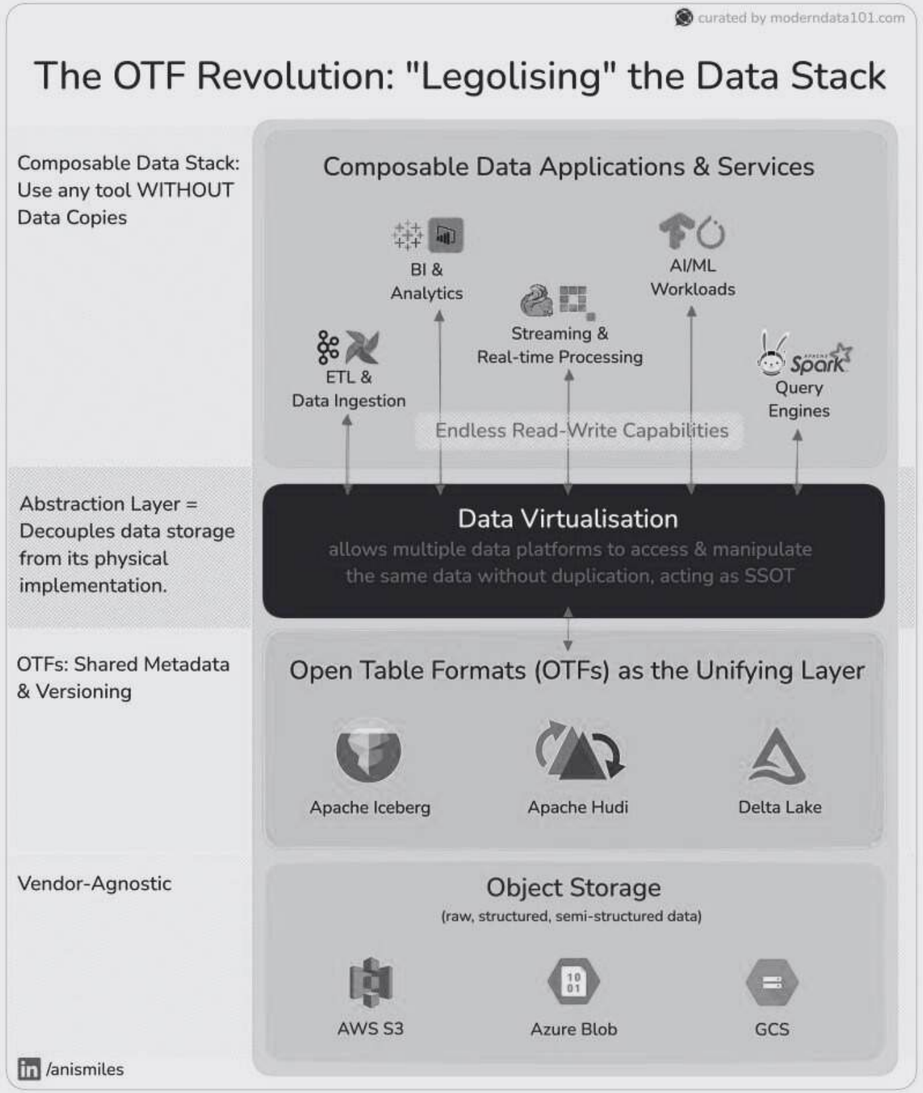
A leading problem identified is the fragmented tooling landscape.

"Over 70% of data leaders & practitioners are driven to use more than 5-7 different tools or work with 3-5 vendors for different tasks, such as data quality and dashboarding. About 10% use more than 10 tools, showing the increasing complexity of the Data Landscape," says Saurabh Gupta, CSO at Modern Data.

The image shows the vicious cycle of diminishing data ROI, a consequence of fragmented or siloed tolling and data architecture | Source: Srujan Akula, the Co-Founder & CTO at The Modern Data Company.



Rethinking Conventional Data Architectures: Applying New Design Principles



By adopting modern design principles, companies are moving away from static, siloed architectures. The modern data platform is agile, self-service-friendly, composable, and purpose-driven—which significantly reduces the data-to-insights time. For most organisations, one of the clearest requirements is to address tooling overwhelm, where they carry the operational overhead of integrating multiple products, tools, or stacks, ensuring they work well together without corrupting data.

Take for instance, design principles that focus on composing data formats. In traditional designs, each domain/team/client would have had to comply with specific data storage formats to benefit from the stack. However, the OTF revolution unifies the storage experience despite not disrupting existing format choices like Apache Iceberg, Apache Hudi, or Delta Lake (see image).

"Instead of forcing all workloads into one stack, OTFs unlock true composability—one shared data layer with infinite stack possibilities. **The roadmap of the data stack isn't monolithic. It's modular.**" (Source)

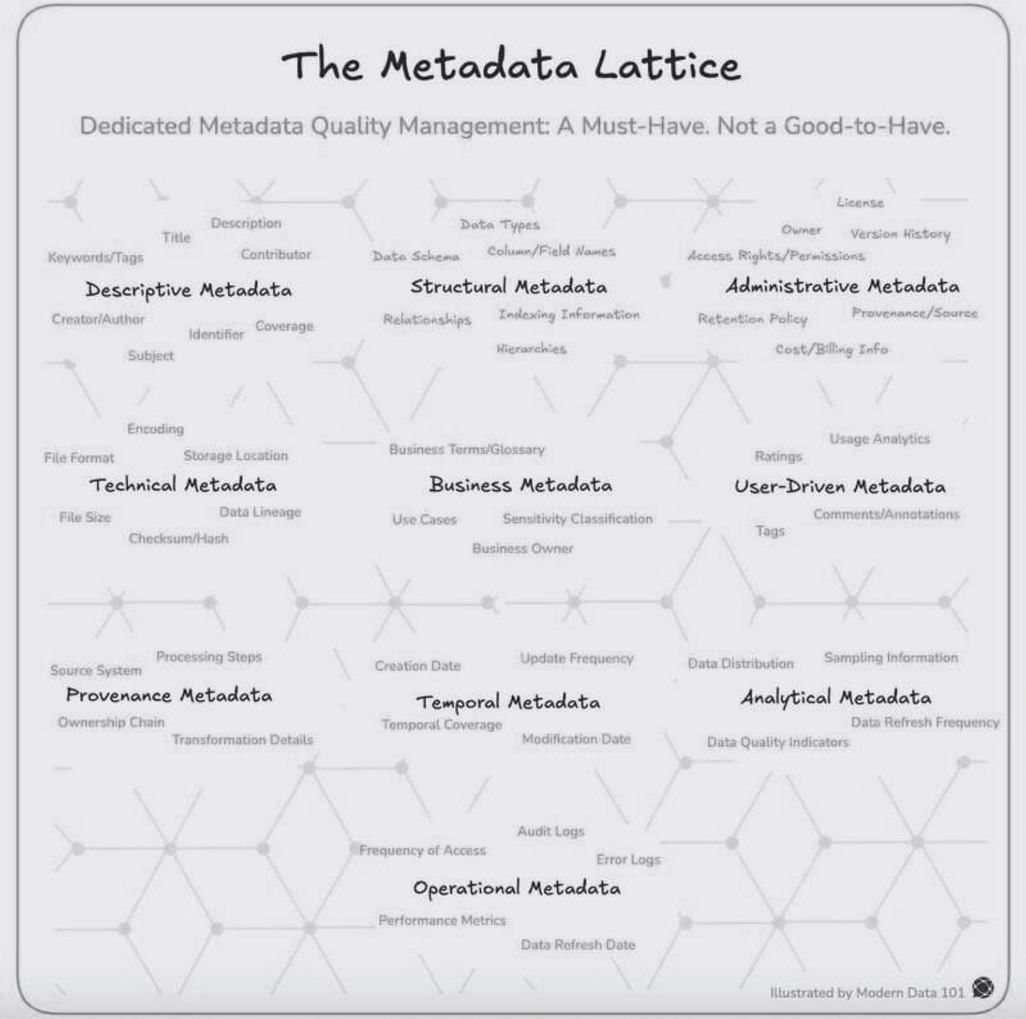
Rethinking Conventional Data Architectures: Applying New Design Principles

And this isn't just about composing table formats or data storage, it's about the modular design approach that helps deal with overwhelming choices layered on the legacy data stack over the years and decades. Organisations are adopting new design principles to holistically enhance the people-infra landscape. And the moment we talk "people", two distinct aspects surface: culture & context.

Nailing the Context Game by Pushing Metadata to the Forefront

Without metadata, data remains an untapped resource, lacking the clarity needed for effective utilization. This underscores why metadata stands as one of the foundational pillars of data products. It's noteworthy that the volume of metadata can sometimes surpass that of the actual data. This extensive metadata includes elements such as tags, descriptions, and freshness indicators, all aimed at making data more accessible and understandable. However, this abundance also raises a critical question: If the pathways to data usage are compromised, isn't there a risk of negatively impacting business outcomes?

Image: Diagnosing the aspects of data culture at your organisations | Source: Mirko Peters



The need to have metadata management to avoid corrupted data outcomes | Source: Modern Data 101

Nailing the Context Game by Pushing Metadata to the Forefront

Managing the Metadata Lattice

Employing advanced technologies that ensure precision becomes imperative, minimizing areas where quality might be compromised.

Animesh Kumar, the CTO & Co-founder of Modern, mentioned that while metadata quality measures share similarities with general data quality metrics—such as completeness, timeliness, and accuracy—certain parameters are uniquely vital to metadata, including:

- Granularity: Detailed metadata enhances data understanding.
- Interoperability: Ensures metadata's usability across diverse systems.
- Semantic Clarity: Clear metadata conveys precise data context.

Context is King: Elevating Metadata for Enhanced Data Utilization

Laurent LETOURMY
Head of Data @ Devoteam France | Data & AI fan

Universal Data Access Through Semantic Understanding

The semantic layer transcends its traditional role in BI to become a **universal knowledge graph** that connects data across domains, a central repository of business logic and metric definitions, an interpreter between natural language and technical implementations, and a guarantor of consistent data interpretation across all tools and platforms.

This universality means that whether data is being accessed by any business user using natural language (and **it works**!), an AI agent answering to queries, an automated process making decisions, or a data scientist building models, the same semantic understanding, business rules, and data relationships are applied consistently.

JESSICA TALISMAN
MAR 20, 2025

When the business understands the composite of a semantic system, leaders and stakeholders have more visibility into the requirements of a semantic knowledge management system. With clear goals and outcomes in sight, **organizations can build and maintain semantic knowledge management systems with confidence, knowing that these investments are essential to support data infrastructure, data transformation, and AI initiatives.**

Pietro La Torre · 1st
Data strategy🎯, management 🚣 and tales 🔏
7h •

Ontologies bring order to knowledge, making it explicit and machine-readable instead of tribal wisdom passed around in slack threads. The wiser have also rules to model changes.

Transitioning to Lakehouse 2.0 for Enabling the New Design

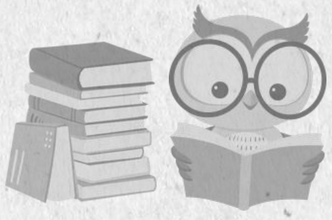
As the emphasis on AI readiness is growing, attention is shifting towards promoting infrastructures and platforms that are inherently aligned with AI objectives. So, along with actively cultivating AI-ready data through robust metadata management & dynamic governance, organisations felt the need to have adaptable infrastructures, and that's where a new vision for lakehouses emerged.

Inherent Challenges of Traditional Data Lakehouses

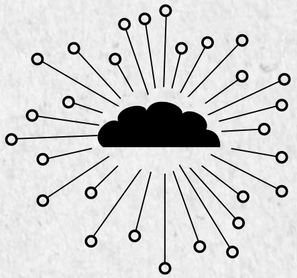
Despite the promise of combining data lakes (flexibility) with data warehouses (structure & reliability) to offer simpler, cheaper and unified architecture, the reality of conventional lakehouses was different.

The Reality:

- Vendor lock-in, where users had to use one vendor's tools (compute, storage format, catalog), and they couldn't easily swap out engines like Trino or DuckDB
- Lack of modularity
- Storage-compute coupling, where query engines had to be built around the storage layer's logic
- Table format lock-in that drove "open" formats. Like Delta was tightly coupled to specific engines, and users couldn't plug-and-play different engines or tools easily



Data Lakehouses 2.0? What Changed & Why it Matters?

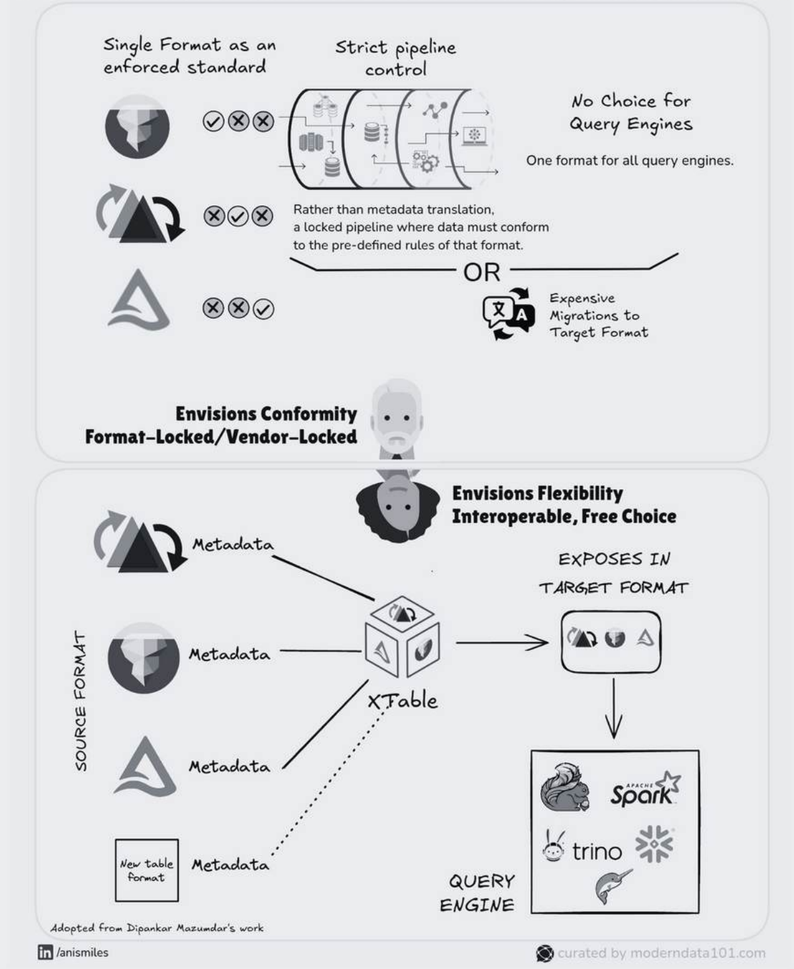


Lakehouse 2.0 is a new, more modular, open, and flexible architecture for data platforms. Unlike earlier "monolithic" lakehouse setups, Lakehouse 2.0 emphasizes freedom of choice, allowing users to mix and match best-in-class components (like query engines, catalogs, governance layers, etc.).

This new generation architecture embraces open table formats like Apache Iceberg, Hudi, and Delta (as a spec, not just a product). By decoupling the storage and compute, users can now run, for instance, Spark, Trino, DuckDB, Snowflake, etc., on the same data without rewriting pipelines.
Open table formats (OTF) allow metadata handling, ACID transactions, and schema evolution, thereby enabling any engine to work with the data — true plug-and-play.

The shift to Lakehouse 2.0 promotes composable Lego-like stacks that allow teams to build custom data stacks using best-of-breed tools. Hence, more flexibility, cost efficiency, and future-proofing!
Aligning with the vision of Lakehouse 2.0, the CTO of The Modern Data Company talks about its interoperability capabilities and highlights the revolution that OTFs and cloud adoption (for instance, S3 + Iceberg) have brought forth, which introduces the element of choice and interoperability. With OTFs, teams can seamlessly trigger interactions between diverse table formats without the need for data duplication or extensive reconfiguration. This strategy allows organizations to select the most suitable tools, formats, & literal plug-&-play stacks for their specific requirements without being constrained by a singular system with enforced one-vendor ecosystems.

Are we Transitioning to Data Lakehouses 2.0?



Tools like Apache XTable facilitate this interoperability by acting as translators, enabling seamless interaction across different data formats.

This shift towards adaptable data architectures signifies a move away from monolithic systems, offering businesses the agility to respond to evolving data demands effectively.

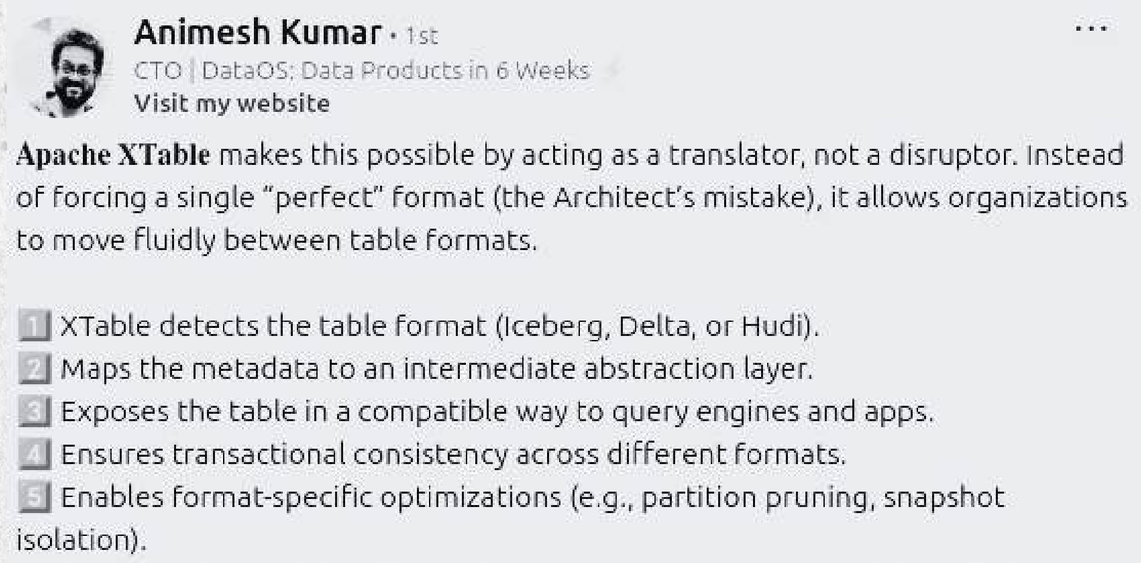
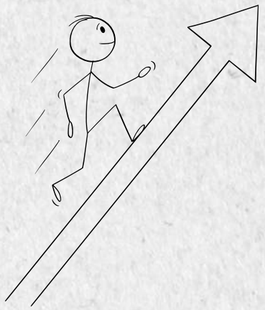


Image on Left: Lakehouse Vision: The Oracle Vs. The Architect | Source: MD101 Archives

Data Developer Platforms as the New Enabler



While Lakehouse 2.0 seems attractive, the question of how to achieve this vision is the primary concern. And that's how we land at approaches like the Data Developer Platform that borrows its principles from IDP.

It's about leveraging the principles of software engineering into data engineering. This approach is now gradually seen to be resonating with the industry's needs. And that's how we land at approaches like the Data Developer Platform that borrows its principles from IDP.

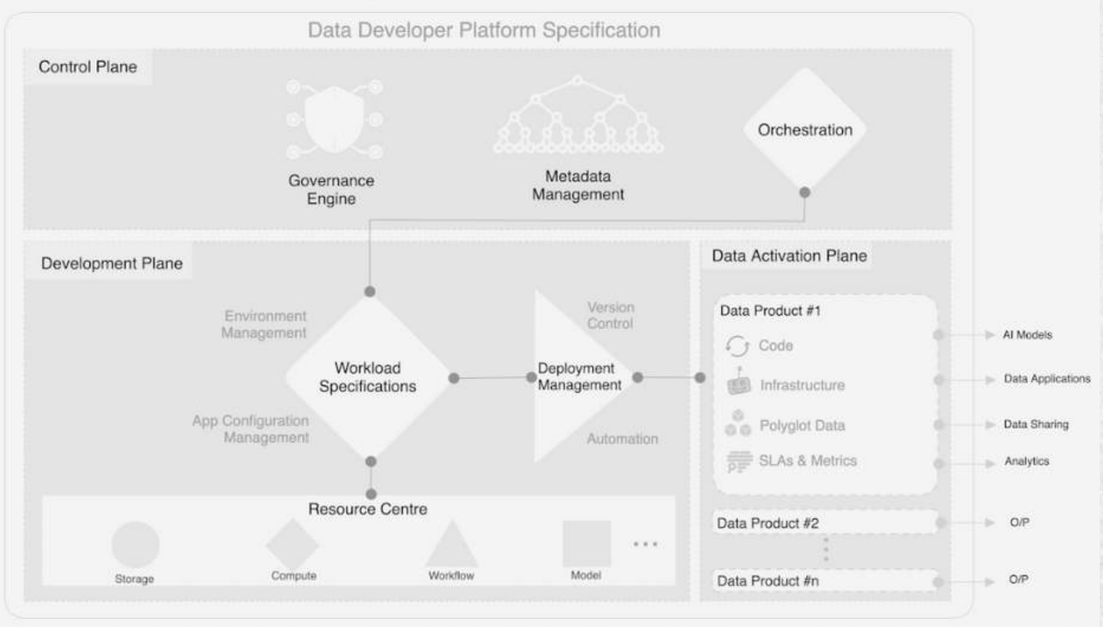
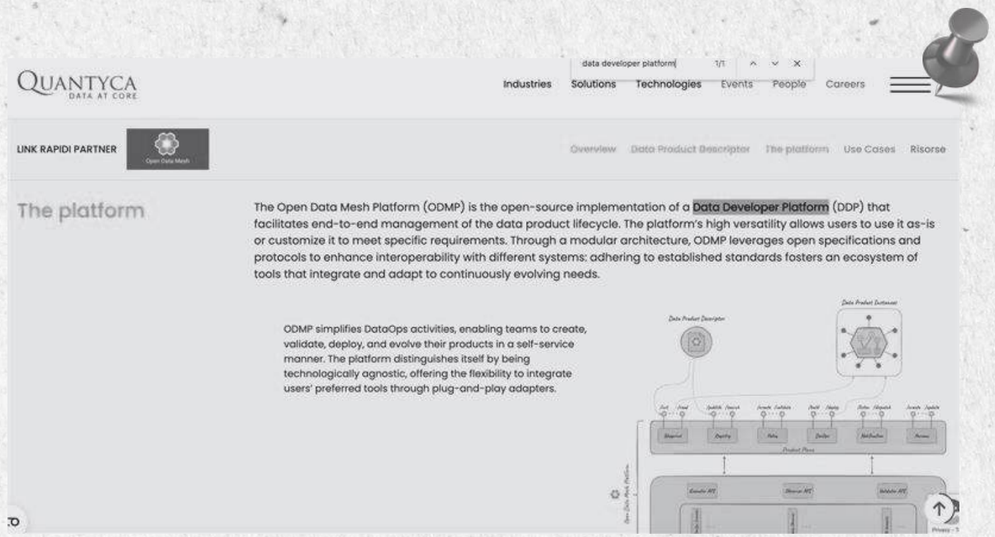


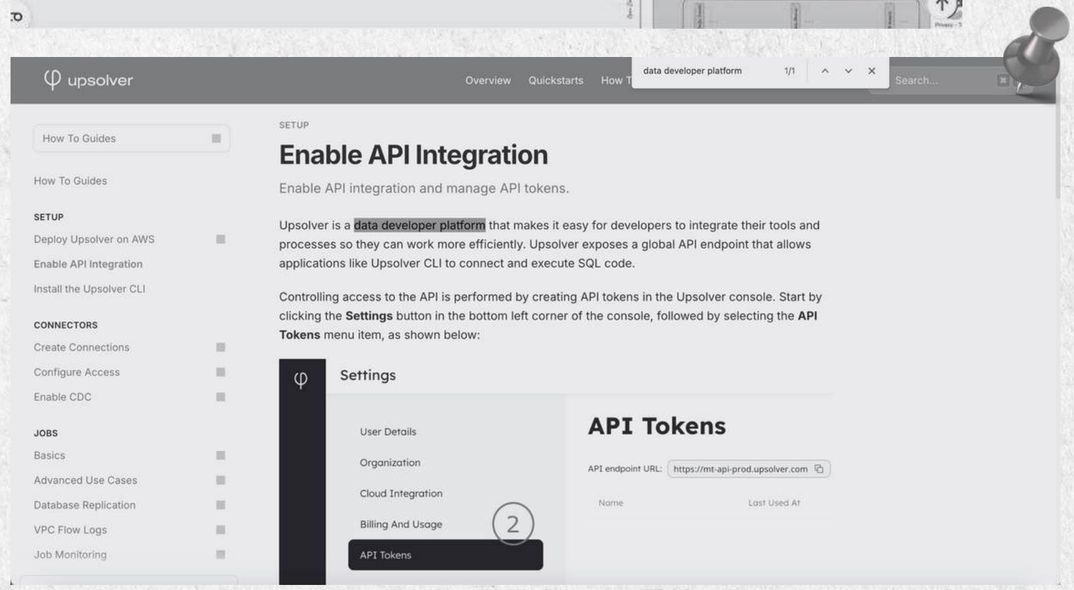
Image: The Architecture of Data Developer Platform | Source: data developerplatform.org

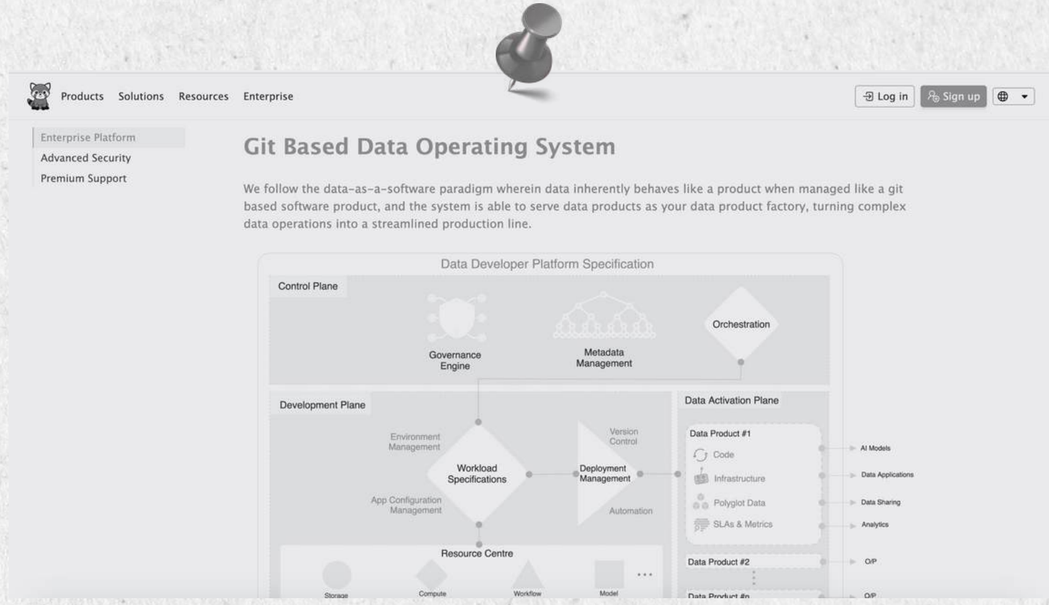
Enabling a unified data management with a modular composable architecture

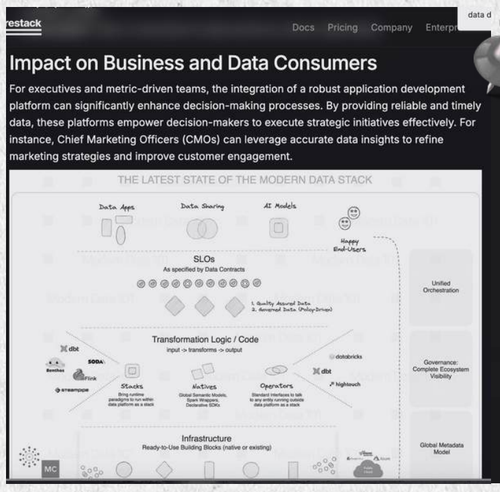
At its core, a DDP offers a standardized framework that abstracts away the complications of distributed systems. This means data engineers can focus on extracting insights rather than wrestling with integration challenges. By harmonizing various tools and services, the platform creates a seamless environment that's in perfect sync with the goals of Lakehouse 2.0.

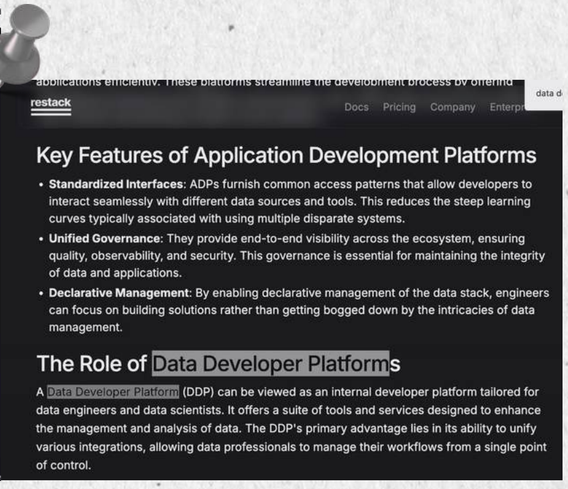
How Multiple Industry Experts Find DDPs Beneficial









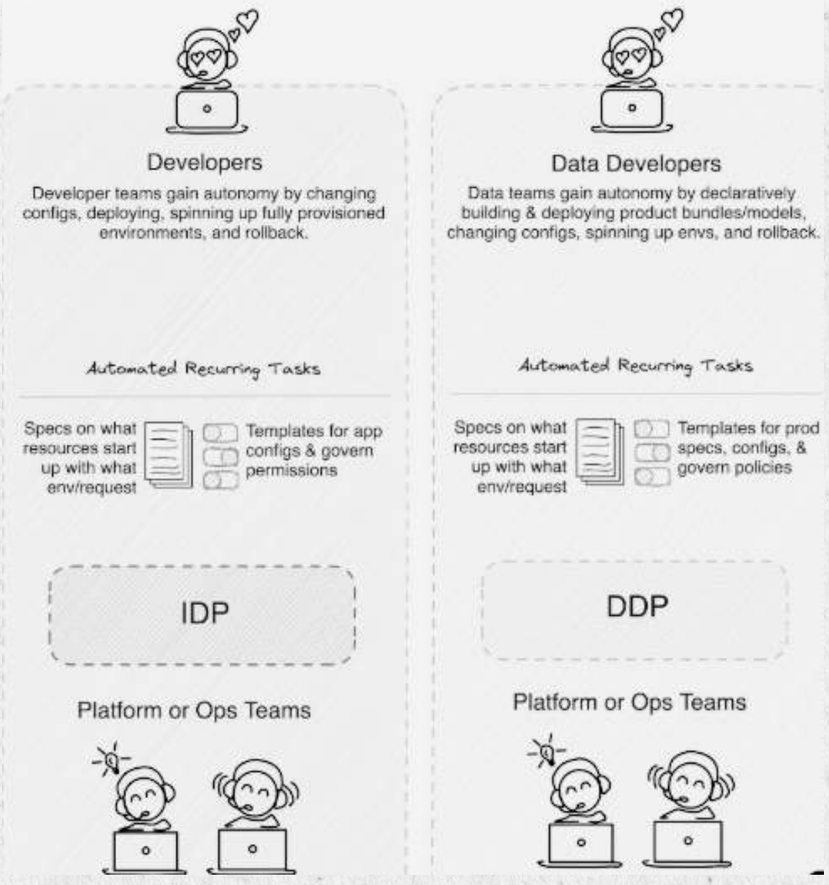


Enabling Unified Data Management with a Modular Composable Architecture

It 's not just about simplicity. The DDP is built for scale. Whether you're provisioning resources or automatically adjusting computing power and storage, the platform grows with your needs. It effortlessly handles large-scale data processing and analytics, ensuring that performance remains top-notch even as data volumes explode.

The DDP also smooths out the bumps in the development lifecycle. With streamlined processes for packaging, deployment, versioning, and release management, it provides consistent, reliable environments for development, testing, and production. This consistency is key to efficiently rolling out Lakehouse 2.0 architectures across your organization.

And in today's data-sensitive world, security can't be an afterthought. With robust, attribute-based access controls built in, the DDP enforces strict data governance policies, ensuring data integrity and security across every layer of your unified architecture.



Applying software principles of an IDP for data to ensure seamless DataOps and usage | Source: Modern Data 101 archives

Join The Modern Data 101 Crusade!

If this playbook was helpful for your implementation or even as a means to organise your thoughts, we'd highly recommend you join us and help us build the Data Product Expertise in the data industry.

We are ALL THINGS DATA PRODUCTS!

Find a truly community-driven voice with experts from all walks of the data industry voicing their insights in the Data Product space on our channels. We'd love for you to come forward and share your insights with us so we can improve our delivery and truly create for YOU.

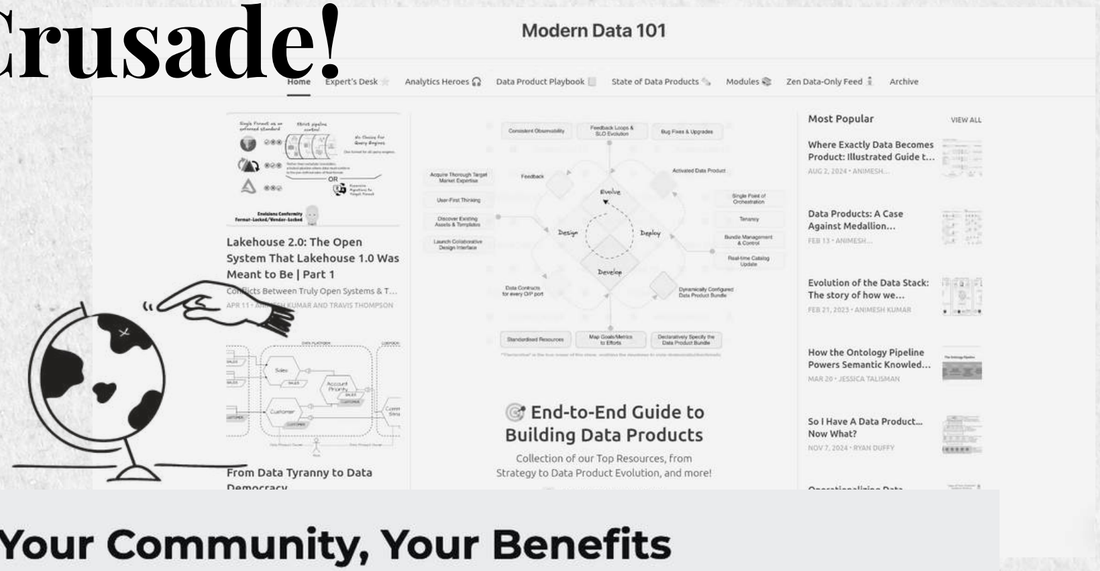
Subscribe

9000+ Data Peers

100+ Resources

30+ Contributors

120+ Countries Reached



Your Community, Your Benefits

Explore a world of reliable information, connect with experienced peers, share insights, boost visibility, receive valuable feedback, and propel your career forward in the ever-evolving data industry.

Access To Resources

In an age of infinite data perspectives, have trustworthy information to make a judgment.

Opportunity to Network

When learning about data, surround yourself with people more experienced than you.

Data Product Expertise

Find all things data products, be it strategy, implementation, or a directory of top data product experts & their insights to learn from.

Visibility & Recognition

Showcase your expertise, contribute to discussions, and build your reputation within the community.

Career Growth

Stay updated on the latest trends and job opportunities in the data industry, and gain valuable career advice.