

The Top Data Trends for 2025

REPORT

Industry experts share their
predictions for the year ahead



Table of Contents

3	Contributors
4	Introduction: The Year of Practical Innovation
5	From data pipelines to massive knowledge pipelines
7	Multiple engines, unified storage
9	Stronger data solutions through consolidation
11	A healthy industry reset
12	AI agents making up for industry amnesia
13	Balancing AI ambitions with energy grid demands
15	Widespread adoption of AI assistants
16	Data lakes as the foundation for AI
18	Beyond the hype: AI deployments at scale
21	Closing the gap between IT and business
22	Problem-solving pragmatism over data dogmas
23	AI-powered, next-level customer engagement
25	Automation of Automation
27	From data gravity to platform gravity
29	New insights from unstructured data
32	Iceberg and open table formats surge
33	AI-enhanced data engineering efficiency and workflows
35	Rapid value from query-enabled LLMs
37	Data management disruption with semi-structured data

Contributors



FRANK BELL

Founder, Data Thought Leader
ITS – Snowflake Solutions



KENT GRAZIANO

The Data Warrior



TEJAS MANOHAR

Co-CEO and Co-Founder
Hightouch



TAYLOR BROWN

COO and Co-Founder
Fivetran



JUSTIN GRIMME

Manager of Business Intelligence
Shane Co.



BARR MOSES

CEO and Co-Founder
Monte Carlo



PATRICK BUELL

Co-Founder and VP of Consulting
Hakkoda



CINDI HOWSON

Chief Data Strategy Officer
ThoughtSpot



ARMON PETROSSIAN

CEO and Co-Founder
Coalesce



JOHN COSGROVE

Partner
Cloudwerx



SATISH JAYANTHI

CTO and Co-Founder
Coalesce



CHRIS TABB

COO and Co-Founder
LEIT DATA



LEE DERKS

CEO
DigBI Consulting



ARTYOM KEYDUNOV

CEO and Co-Founder
Cube



PAVEL TIUNOV

CTO and Co-Founder
Cube



MATT FLORIAN

Practice Director, SAP Delivery
Hakkoda



HUGO LU

CEO and Founder
Orchestra



GU XIE

Head of Data Engineering
Group 1001



ADAM FOKKEN

CTO and Co-Founder
phData

2025

The Year of Practical Innovation

2024 was momentous. Bold generative AI experiments captured imaginations and investment dollars, and the cultural conversation turned to the data that underpins them. In the new year, AI will continue to shape society, but in increasingly practical ways tied to tangible business value. As IT teams and business leaders reckon with ROI and persistently high interest rates, many are refocusing on practical applications rather than aspirational ones. Simultaneously, AI is reshaping infrastructure, creating a demand for data literacy while challenging business culture and governance practices to keep up. We see a lot of reasons to get excited about the year ahead, in which smart data practices will continue to shape the world around us.

The start of a new year isn't just a time to look ahead and anticipate the innovations to come in 2025, it's also a period of reflection. At Coalesce, we're grateful to have spent 2024 forging new partnerships, introducing new integrations, and innovating in an ecosystem that continues to challenge and excite.

As many of the experts we've interviewed here have pointed out, human beings are always at the core of meaningful change. Coalesce's ability to continually transform, improve, and pioneer in the data space is informed by our partnerships with the brightest minds in the field. A number of these experts have shared their observations with us on where the industry is headed in 2025. With their intellect and insight, we anticipate an inspired start to the new year.

Welcome to 2025!

From data pipelines to massive knowledge pipelines



JOHN COSGROVE

Partner
Cludwerx

Massive knowledge pipelines will go mainstream

Hands down, my biggest prediction—and I admit it's a bit of a cheat because I'm already working on it—is the monumental mainstreaming of what I'm calling massive knowledge pipelines. These pipelines refer to the industrialization of RAG (retrieval-augmented generation) into AI. We're at this incredible intersection, a "perfect storm" if you will. If we were still talking about conveyor-belted unstructured data like we have been over the past five years with traditional machine learning, it wouldn't be that remarkable. But that's not the case anymore.

What we're seeing now is a new type of pipeline, not concerned with traditional relational structures like tuples, but with semantic structures instead. This pipeline needs to be comfortable with a polyglot approach to meaning, which is a far more comprehensive way of thinking about data that's more in line with the older concepts of ontology and semantic knowledge. It's no longer just about tables—there are many different structures at play here. This follows from predictions made last year about the rise of vector and graph databases. Now, we have clear patterns emerging for piping trusted knowledge into the context window of your preferred LLM, and doing so at scale.

I emphasize "massive" for two reasons: First, we're talking about powering LLMs that have been completely unleashed. By the end of 2025, I would be surprised if we're not seeing at least 10% or even 20% of all customer inquiries involve LLM interaction. That's a huge leap from where we started: essentially zero. The forward consumption potential is massive.



What surprises me is the stubbornness of people who remain skeptical about generative AI's future potential, even though to me it's so **clearly accelerating faster than we thought.**

JOHN COSGROVE

Partner
Cludwerx



I think 2025 is going to be a deeply destabilizing year for everyone's sense of comfort with technology. All of us in this industry are going to have to grapple with the fact that **our jobs are changing**. By 2026, I think everyone else will have to as well.

JOHN COSGROVE

Partner
Cloudwerx

But the real reason I call it a massive knowledge pipeline is because we've spent the last several years setting up these data lakes filled with documents, videos, images, and petabytes of tables. This polyglot landscape is already primed for action. So when we talk about converting data into vectors, for example, it's ready to go. Many companies have already made the investment, and for those that haven't, now's the time. This is a real wake-up call. You can't afford to lag behind, because if you do, you'll be at a disadvantage in the world of gen AI.

The majority of the industry, I believe, has some form of this up and running. What we're going to see now is these data lakes emptying into new types of pipelines. They'll still function like pipelines in the sense that they convey information, knowledge, and material from wherever it resides to a specific endpoint, an augmented retrieval point for generation. But the nature of these pipelines is fundamentally different. They're more comprehensive, more nuanced.

As someone who has thrown himself wholeheartedly into gen AI over the last two or three years, working with it daily at scale, I can say with confidence that we are only beginning to unpack the vast richness of the latent semantic space. There are so many ways to anchor your knowledge space, your ontology, to achieve a particular RAG outcome—and you can tailor it to your needs. That calls for a level of configuration, tooling, and iteration that sounds like modern data pipeline development, like working with SQL or Python. But now, instead of code, we're dealing with something more akin to a library index.

Multiple engines, unified storage



SATISH JAYANTHI
CTO and Co-Founder
Coalesce

More companies will rely on multi-engine compute

Organizations will increasingly use more than one compute engine to process data. We saw this before with legacy on-prem systems, but now we're going to start seeing it with cloud systems, too. The tendency is always to consolidate things, but this increasingly doesn't matter—it's OK to have two or more systems, especially because there are many developers out there who are using different systems. On top of that, not all compute engines address the same problems, and sometimes you have to use a certain compute engine for a specific use case. In fact, some database vendors (such as Snowflake) are encouraging organizations to use a more open source format, such as Iceberg tables, that makes the data easier to access even by their competitors' systems.

We'll see a big move to unified storage

But even if you have multiple engines, all these engines need to talk to a single storage layer. This is why one big trend we will start to see is the move to unified storage, likely in these open source Iceberg tables. And so, rather than consolidating systems,

it will become more valuable to figure out how to have a unified view of everything instead. People will want to interact with all these compute engines in a way that's uniform, using a single interface that's approachable for a wide variety of skill sets. So while it's difficult to predict what the big compute engine companies will do in the short term, the future data architecture that makes the most sense will be one that has unified storage of data, multiple engines (each chosen for whatever use case it best solves for), and one unified view that offers built-in governance and ease of both development and operations from that same system.



People are coming to the realization that building an AI solution is very easy, but **building an AI solution that actually adds value** is much more difficult.

SATISH JAYANTHI
CTO and Co-Founder
Coalesce



Data quality and observability will grow in importance

Even though the issues of data quality and observability have been discussed for a while now, their importance will only continue to grow in the year ahead, given the key role they play in AI. This means having a good data quality and observability system—one that is well integrated with your data transformation process—will be crucial going forward.

Data mesh will finally take off

While data mesh has been a hot topic in the industry for a little while now, the challenge has been transforming it from an interesting theory into a useful, real-world model. But as this new data architecture emerges, it will bring us to the point where data mesh finally seems realistic to implement.

We shouldn't assume that some new technology is going to show up one day that will solve all our problems. There's a lot of attention paid to technology evolution, but let's not lose sight of what we're using that technology for—to process data.

And those data volumes are only going to keep growing. **So even though the technology continues to improve, the problem also keeps getting bigger.**

SATISH JAYANTHI

CTO and Co-Founder

Coalesce

Stronger data solutions through consolidation



ARMON PETROSSIAN

CEO and Co-Founder
Coalesce

We'll see vendor consolidation emerge as a trend

People have been talking about how a lot of recently venture capitalist-backed companies will run out of money soon, but I think it's already happening. When most VC-backed businesses get funded, there are usually strategies they can employ when they start running out of money—for example, using bridge funding or drawing from their line of credit to get about another six months of runway. But for many startups in the data space, these two last “safety valves” have already been pulled.

If your startup raised a round of funding in 2021 or 2022, you're probably going to be running out of cash by early next year. And so I anticipate that in the coming year, a lot of vendors in the data space will get folded into larger companies. The good news is, this will actually benefit the larger industry. Too much VC funding has led to too many options in the market, which can be overwhelming to buyers. Consolidation in each specific category will provide customers with fewer solutions that do more things. It will also offer increased stability as customers don't have to worry that the solution they're using might go out of business.

Vendors themselves will benefit from consolidation trend

This consolidation will also be to the benefit of the vendors themselves. As the big cloud data warehouses become more and more feature-rich, they are starting to step on the toes of a lot of the peripheral point solutions. Unique differentiation becomes harder and harder unless you're building a much broader platform. So this consolidation trend will result in many companies coming together to form stronger, larger organizations.

Q

&

A



ARMON PETROSSIAN

CEO and Co-Founder
Coalesce



SATISH JAYANTHI

CTO and Co-Founder
Coalesce

What were some 2024 trends that surprised you?

We were both struck by how quickly the open table format, such as Iceberg tables, was embraced this past year. The popularity of this format is one trend we expect to see continuing into the future.

What's a trend you hope to see less of in 2025?

No more unthoughtful AI use cases. Of course AI is important, but AI initiatives should be focused and well thought through. Unfortunately, this past year we saw a lot that weren't.

What's the biggest misconception you hope gets put to rest in 2025?

That SQL is waning in importance. The truth is, SQL is only becoming more commonly used and has become very mature. There is no better language to deal with large data sets.

A healthy industry reset



ARTYOM KEYDUNOV

CEO and Co-Founder
Cube

There will be a healthy industry reset and market consolidation

The analytics industry is like a roller coaster: everything is going great and then it goes down, as if it's resetting. It happened with Hadoop and Big Data, then the Modern Data Stack got hot in 2021 during COVID, while last year and this year it wasn't entirely doom and gloom, but it felt close to it. And I believe it was actually healthy for the industry to do this reset. Every vendor, every company, and even every data team should think about the value they create and how to deliver on that value.

This has led some to wonder if maybe we have too many vendors in the space, too many tools. My feeling is that this sentiment is going to continue at least into the next year, and we'll reach the point where some consolidations will happen. We'll go from the chaos of multiple vendors in certain categories to a more structured market.

Practical, production-grade AI applications

On the other hand, I think we'll start to see more interesting production applications of AI in some tools. Production-grade applications may not end up as cool as initially predicted two years ago—we're probably going to start small. More tools will get their Copilot, for example. Copilot is great. It's useful, yet it doesn't feel entirely like magic; rather, it feels like a better auto-complete feature. I believe many tools will get AI-powered improvements or better workflows, which won't seem groundbreaking right away, but would be a very good production-grade use of AI and maybe unlock some improvements that were not possible before.

AI agents make up for industry amnesia



PAVEL TIUNOV
CTO and Co-Founder
Cube

AI will make up for industry amnesia

The terms we are now discussing as big trends in our industry, such as the semantic layer, existed thirty, even forty years ago—they're not new. We're experiencing some industry amnesia right now, and we might see the rise of AI-assisted data engineering pretty soon, with agents coming in and filling the gap of this memory because it's basically a literacy question: Do you have time to fill the gap of forty years in your head? Usually, the answer is "no" and that's why we keep reinventing the wheel around foundational concepts in analytics.

What business leader do you most admire and why?



Bob Muglia, investor, advisor, and former CEO of Snowflake:

"His combination of leadership skills and technical acumen is impressive. He can talk business or he can talk tech."

—**Kent Graziano**, The Data Warrior



Judith Faulkner, Founder and CEO of Epic Systems:

"She's put so much strategy in place to make sure the business doesn't get acquired, go public, or lose its mission—the goal is to turn it over to employees in the future, because it's their work that built the company."

—**Matt Florian**, Hakkoda



Armon Petrossian, CEO and Co-Founder of Coalesce:

"Watching what Armon has done to build Coalesce into the company and product it is today with Satish has been incredible. He's down-to-earth and approachable, while being hugely successful."

—**Lee Derks**, DigBI Consulting



Erik Duffield, CEO of Hakkoda:

"I learn a lot from him. He's pragmatic, he's got lots of experience, and tons of energy."

—**Patrick Buell**, Hakkoda



Richard Branson, Founder of Virgin Group:

"From a brand and ethics perspective, I admire him a lot."

—**Chris Tabb**, LEIT DATA

Balancing AI ambitions with energy grid demands



KENT GRAZIANO
The Data Warrior

The growing AI power dilemma will finally demand action

The electrical energy (kWh) that it's going to take to power the AI activities that organizations are envisioning will be tremendous. We're reaching the point where problem-solving for this is pressing. Right now AI is on a collision course with green energy initiatives. Can we power everything we want to by using alternative energy? If we can't, where is the compromise coming from? How are we going to navigate this split? This discussion is already happening, but it's going to become a focal point in 2025 as more organizations develop AI-based solutions and bring AI products on board. It comes down to a standard business question: What's the return on investment? With technology, just because we can doesn't mean we should. Use technology appropriately—that's a lesson we'll take away from the next few years.

The convergence of data quality and culture will become essential to AI success

Due to the expansion and democratization of AI in various forms, we're seeing more good as well as bad examples of it. This is making everybody sit up and pay closer attention to the intersection of data quality and data culture. If you don't have the fundamentals in place, you can't slap AI on your solution. If you don't have the right knowledge base, you can't use the tools properly. If you don't have the basics down, AI will only help you build something that's not very good faster.

If the point of AI is to empower more people with data, a solid data culture has to exist at the business level, not just the technical one. A more robust level of understanding of data across the organization is a growing need because a lot of quality issues go back to the source of the data. If you want to build AI and feed new data into your data products, the way to do that efficiently is to make sure there's quality on the input side, and that's not just IT's problem anymore (and it never really was).

If people keep believing that, it becomes an organizational problem—these issues can no longer be solved by IT alone. Good quality data is everyone’s responsibility, and there has to be an organizational-level understanding of that. In other words, data culture has to be as much top-down as bottom-up.

Centralization of data will become essential

The only way to achieve data governance effectively is not with an assortment of open source tools and with data scattered around the cloud somewhere, but by having your data live inside an ecosystem that you can control and govern. That is a development I see coming a long way in 2025. More product companies are able to embed AI capabilities into their tools to support this. Copilot and Cortex in Snowflake are examples of this consolidation of AI features and functionality inside a data platform. This helps achieve one of the promises of AI: simplifying the use of data and information for nontechnical people—and doing that requires centralization and governance of the tools and the platforms.

This level of oversight includes building and managing private data sets. If you want to control and minimize the potential for hallucinations in your AI products, you have to control the input. You must curate the data to ensure quality.



AI is costly. And just because you build it doesn’t mean they’ll come. There are lots of capabilities out there—are you using them judiciously? When you look at all the compute power and what it costs in terms of energy, where and how much is the return?

Where is the balance between the benefit versus the cost? **This is foundational, and we’re just now arriving at these questions.**

KENT GRAZIANO

The Data Warrior

Widespread adoption of AI assistants



LEE DERKS
CEO
DigBI Consulting

There will be widespread adoption of AI assistants in both work and life

The implementation and adoption of AI as an assistant is one of the most important trends for 2025. Right now AI is in an era similar to what the internet was in the early '90s—on the upswing and gaining popularity. If AI is implemented and adopted well, it will complement professional and personal lives alike. I see these assistants becoming more accepted in the workplace this year, too.

Integration of Microsoft Copilot into the Office platform, Gemini into Google search, Snowflake Copilot, and Apple Intelligence are all game changers.



Data warehouses are due for some changes. The adoption of unstructured data, AI to classify it, and zero ETL integrations **are all gaining momentum.**

LEE DERKS
DigBI Consulting

While not yet 100% accurate, they provide a great starting point for when questions or issues arise. Further down the line, this technology paves the way for personal robotics. Tesla Optimus, Boston Dynamics, and other pioneers have already developed incredible designs for these.

On a visit this year to China, I witnessed robots at hotels automatically request an elevator, deliver items, and display cute emoticons—it's fascinating.

The move toward simplification will continue

I see more leaders wanting to mesh data from a variety of sources into a single platform so they can gain a holistic view of their business. Gone are the days of allowing multiple sprints to set up a data pipeline. Solutions like Coalesce demonstrate that within a few minutes, raw data can be easily transformed, machine learning forecasts generated, and all of it executed in the current data warehouse.

Additionally, analytics, pipelines, and data transformation are now easily accomplished without expensive resources—and that will continue. This includes rethinking how we work in data warehouses. Companies are realizing that using intuitive, cutting-edge technologies can save them thousands of dollars in development and maintenance costs.

Data lakes as the foundation for AI



TAYLOR BROWN
COO and Co-Founder
Fivetran

Data lakes will increasingly serve as the foundation for AI

By 2025, the explosion of unstructured and semi-structured data—from IoT devices, social media, and other digital touchpoints—will force businesses to rethink their data architectures. Traditional databases will struggle to handle the volume and variety of this data. Companies will need fast, reliable, and secure access to all their data to build accurate AI models, and data lakes will be the most efficient way to provide this flexibility and scale, especially with open file formats like Iceberg or Delta Lake ensuring compatibility across systems.

Fivetran's Managed Data Lake Service will play a key role in this shift. As businesses increasingly rely on data lakes to store diverse data sets, the need for automation in data ingestion and management will grow. Fivetran's solution addresses this by enabling companies to automatically ingest, store, and govern massive amounts of data, ensuring they can scale their AI initiatives without the burden of manually managing complex data pipelines.



AI models require large amounts of clean, high-quality data to function effectively and produce accurate results. **Enterprises will increasingly leverage user-friendly data integration tools** to centralize data from various operational data stores to create a corpus for AI training.

TAYLOR BROWN
Fivetran

Data governance will become a strategic business priority

Ensuring data quality, compliance, and security will remain a hot topic next year. Organizations will continue on the journey of establishing robust governance frameworks, and I expect data lake governance practices to be on the rise. Not only are data lakes essential to AI workloads, the data volumes for each organization are growing considerably, making the cost advantages in compute and storage offered by data lakes more meaningful.

The rise of open file formats

In 2025, we'll see open file formats become the standard across industries. These formats allow businesses to future-proof their data architecture, avoiding vendor lock-in while maximizing flexibility and interoperability across cloud environments. Fivetran's automated data integration capabilities will be crucial in supporting the transition to open file formats. As companies increasingly adopt these formats, Fivetran provides the infrastructure to efficiently move data from diverse sources into environments where these formats can be leveraged for AI and analytics projects. By automating the ingestion of data into systems that support open formats, Fivetran enables organizations to scale their data operations more rapidly and adapt to innovations in the evolving data landscape.

Beyond the hype: AI deployments at scale



GU XIE

Head of Data Engineering
Group 1001

AI at scale for all—with plenty of growing pains along the way

I think we're still going to see a lot of trends centered around AI. It took center stage in 2024, but now we'll see its evolution. Whereas it was experimental before, now we're going beyond the question "what can we do with this technology?" and are thinking through how it can be deployed—specifically, deployed at scale.

I'm seeing more companies strategize around this, and I don't mean just big tech companies—I mean mid-cap, large-cap, even small-cap—leveraging generative AI technology to either automate existing workflow processes or enhance customer experience and servicing. This is the moment when we move beyond the hype. I firmly believe there's a tremendous amount of value in gen AI and that's not going to change. But we're done with experimentation and ready to leverage AI day-to-day to improve operational processes. Naturally, this will introduce lots of challenges. It will mean strengthening MLOps to even include AIOps so everything is done in a way that is governed and deployed correctly.

We talk a lot about data governance, but I think pretty soon we're going to be talking about AI governance as its own thing as we emerge into this new world. Good data is the ultimate prerequisite to

using AI, so governance was a huge trend last year. Now we're more laser-focused on AI governance. All the concerns that exist with data analytics are replicated and intensified in these projects. With impending AI regulation, how do we know we can build these in a way that doesn't create lots of regulatory risk over the long term?

Economic conditions will trigger vendor consolidation

We're past the zero-rate era now, so getting funding is much harder. Cutting costs is important for everyone, whether they're startups or enterprises, and we're seeing more companies faltering. Whereas before we had a proliferation of new products and services, now we're going to see increased consolidation across vendors, even among major players, which will force the decision—what side are you on? Which kingdom do you want to work in? That's going to happen in 2025 and in the years beyond.

A silver lining is there will be a greater emphasis on identifying what companies really need, with less overlap of tools and confusion in the marketplace. In the past that confusion has created major challenges for organizations that want to achieve meaningful outcomes, especially those trying to leverage AI. Smaller, more prescriptive vendors tied

to particular ecosystems will be purchased and integrated into bigger systems. This will reduce some decision-making, but at the same time could reduce innovation and increase the cost of switching platforms.

Cross-functional teams become more common

Organizations need to move away from putting teams into functional silos. Historically, you have had teams of data analysts, data engineers, et cetera, but when you organize by function in this siloed way, you don't achieve positive outcomes. Instead, an increase in cross-functional teams—which include a mix of technical and nontechnical folks, including people from the business side—will lead to better outcomes.

What business leader do you most admire and why?



Simon Sinek, business leadership

author and speaker: “I’ve been Team

Simon and Team Brené Brown for a long time. There are many other business leaders I admire in the industry, such as Eric Schmidt, whose recent commentary out of Stanford has been mind-blowing. I also admire a lot of people in philanthropy. I’m not a fan of getting rich just for the sake of it; I’m more interested in how we deploy capital to make meaningful change in the world.”

—**John Cosgrove**, Cloudwerx



Ali Ghodsi, CEO of Databricks: “There’s something about the approach of

building the technology in a bit of an organic way, which I admire. It wasn’t about ‘Open source is so hot right now. What can we do to acquire loads and loads of customers in four years by selling something for free and then really start making money?’ It was more like ‘We’ve got this problem. We need to do something about it.’ And now that they’re scaling, seeing how much they’ve leaned into being a bit of a commercial machine—it’s really impressive.”

—**Hugo Lu**, Orchestra



Satya Nadella, CEO of Microsoft: “What

Satya Nadella has done at Microsoft

is insanely impressive. He inherited the company in a weak position and made some really massive bets, and they paid off so incredibly well.”

—**Armon Petrossian**, Coalesce



People still think AI outcomes can happen without data teams, but these intelligent systems cannot drive higher-level reasoning without data analysts and engineers.

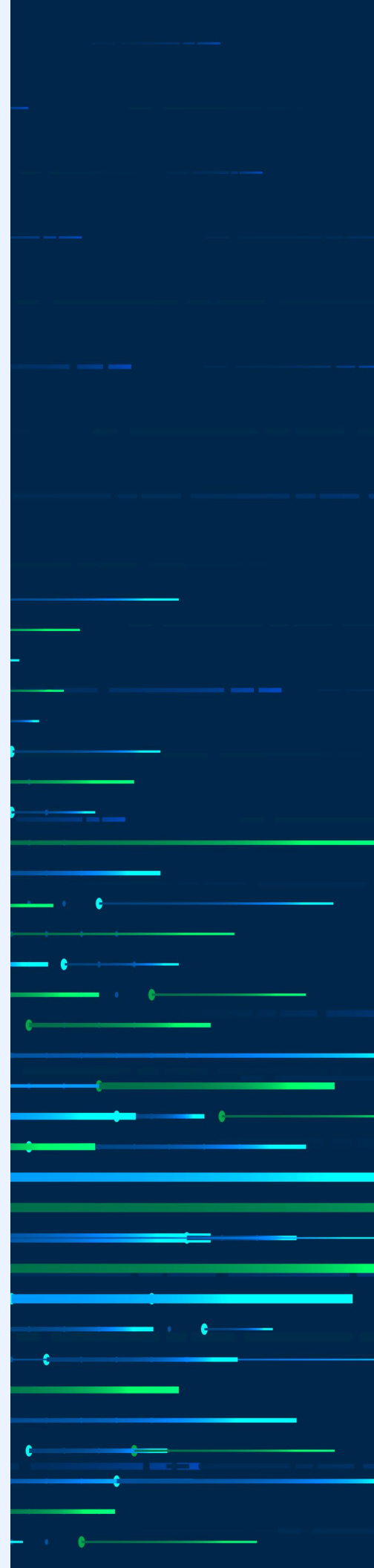
I don't think we're headed toward a future where data teams are replaced.

Instead, we're upscaling them, giving them greater efficiency.

To drive the next outcome, businesses shouldn't reduce data teams, they should spread them across the entire business to create a more decentralized, data-focused approach.

GU XIE

Group 1001



Closing the gap between IT and business



PATRICK BUELL

Co-Founder and VP of Consulting
Hakkoda

As AI accelerates, the need to lessen the divide between business and IT will grow

We're in the midst of an industrial revolution. There are so many unknowns right now regarding how AI will impact our lives. The reality is that we're going to keep doing a lot in this space, and it's moving really fast. There's enough use cases and tangible proof points that we can't ignore the value of AI and the potential it has to impact everyone in the world.

Given that, the people aspect of AI is where my mind goes most often. I've been doing IT my whole career, and technology is never the reason why a business can't succeed or transform—it always comes back to people, processes, and culture. In order to modernize business operations, there needs to be less separation between business and IT. This is why I think data mesh is still relevant, because without it, shadow IT can prevail. With business users accelerating AI use, there are dangers to approaches that keep organizations divided. Without something like data mesh in place as AI is integrated, companies will end up further deepening the divide between the business and IT sides. Modernizing means working in a different way.



It's never been a sexy domain, but **is this the year that data governance finally becomes cool?**

There's such a need for data trust at this moment that it might be.

PATRICK BUELL

Hakkoda

Organizations will require value engineering frameworks to prioritize AI use cases

We must start process mapping our AI decisions and pinning down the most pragmatic use cases. As use cases proliferate, how do you prioritize what your company does with AI? How do you democratize access to new functionality? How you make those decisions requires truly measuring business operations and value engineering your transformations. This underscores the importance of reducing the separation between IT and business in order to gain better intelligence on products, services, and operations. Conceptually, we can do a lot with AI, but how expensive will that be? Is the ROI truly there? I've witnessed even large enterprises hesitating because of the investment required. Value engineering frameworks help us understand how best to deploy new technologies. That matters a whole lot in this moment because it's how businesses will be able to move the needle forward.



What's the one gadget that you can't live without?



Smartphone: "I disconnect from work on vacation, but I still use my phone for directions, reservations, and most importantly, pictures."

—Lee Derks, DigBI Consulting



Airpods: "I can't live without these. I need to tune out when I'm working—I have to have music or a podcast on in the background."

—Gu Xie, Group 1001



Meater Smart Thermometer: "It's new to me but I'm excited about it. It integrates with your phone. It's like pressing the easy button for cooking food."

—Justin Grimme, Shane Co.



Fly fishing rod: "It helps me detox from the other gadgets. It's just between me, the river, the fish, and the present moment."

—Matt Florian, Hakkoda



Old-school fountain pens: "I have a million gadgets, but the ones I truly can't live without are my smartphone and my Apple Watch. Oh, and my fountain pens. I love them—I need that tactile connection."

—John Cosgrove, Cloudwerx

Problem-solving pragmatism over data dogmas



MATT FLORIAN
Practice Director, SAP Delivery
Hakkoda

Problem-solving pragmatism will overcome data dogma

It's so common in this industry, the idea that you must choose a side. Hard lines are drawn around what's right and what's wrong. Instead of focusing on a business problem and the solution best suited to the organization, the problem is often approached with a sense of dogma from the data side. That's not really approaching the problem, it's just applying your worldview to it. You can't apply the same solution to everything you see. All businesses have nuance to what they do, and as professionals we have to embrace that. We need technology that is flexible and IT that doesn't adhere to a rote view of problem-solving so the solution can have greater longevity.

In the last year we've had lofty ideas about AI dominating the conversation, but they weren't tied to real meaning, real value. What problems do you want to solve? We need to get back to the most pragmatic pieces of how we're solving problems, with tangible numbers and value, not lofty goals and narrow-minded approaches. No hype, just real problem-solving—that's how you get actual wins.

Great progress will be made on data interoperability

We won't get there in 2025, but seeing the progress around making data interoperable is really exciting. The ability to securely access data that is meaningful and informs the decision you're trying to make, helping you produce better outcomes—that can improve people's lives. Right now we're all looking for opportunities to do this securely without compromising privacy and trust with customers and vendors. There's potential here for it to go wrong, but there's also so much potential for it to go right and make life better as a result. There's such a drive to see this happen and make data function in the world more meaningfully. We can build the future we were promised.

AI-powered, next-level customer engagement



TEJAS MANOHAR
Co-CEO and Co-Founder
Hightouch

AI will take customer engagement to the next level

In 2025 and beyond, the role of AI will extend far beyond generative AI capabilities, with businesses harnessing its power to drive automated decision-making across the entire organization. As companies build robust data foundations within their data warehouses, AI and machine learning-powered decision layers sitting on top of these warehouses will become crucial for optimizing operations, enhancing customer engagement, and automating marketing at scale.

In marketing specifically, AI will transform how brands personalize and automate every step of the customer journey. Marketers will move past manual A/B testing and static targeting, embracing ML-driven experiences that continuously learn and adapt for each user. This shift will not only deliver more impactful, personalized engagement but also allow marketers to focus on higher-level strategy and oversight as AI takes on the complexity of real-time decision-making and campaign optimization.

Data warehouses will become ROI-driving engines

We're seeing companies embrace the data warehouse not just as a destination for data, but as the most valuable source of data that powers business applications and production workflows. If your data warehouse is just someplace your store data without acting on it, it's just another silo. It's far more valuable as a single source of truth if you can make sure that all of your other company systems stay updated with the data that's in it. Companies that use the data warehouse as an ROI-driving engine have the right incentives in place to solve upstream issues such as data quality, data staffing, and analytics engineering workflows.

Growing adoption of open table formats, unified streaming and batch workflows

There are two interesting trends that we're closely following: First, we're seeing far more adoption of open table formats, such as Iceberg.

These allow you to use various compute engines (whether proprietary or open source) on top of your data lake. Second, we're seeing a unification of streaming and batch workflows in the data warehouse. Products such as Snowflake's Dynamic Tables, Databricks Delta Live Tables, and BigQuery continuous queries are all great examples of this trend. The data warehouse increasingly can support low-latency use cases and data streaming.

More companies will start thinking of their data assets as products

Companies are thinking about data as a product. We've worked with enterprises that have even codified rules for what makes their data assets "products," such as reusability and uniqueness. This makes data investments both more valuable and sustainable.

The best way to get started here is to sit with your customers (business stakeholders) and understand what they're actually trying to solve. They aren't trying to solve pipelines, reports, or SQL—those are just mechanisms to solve the actual business job to be done. Your data products deliver them the right solution in a workflow that makes sense for them—whether it's a report, an automated notification, or a data sync into a downstream tool.



One listening recommendation from the past 6 months?

Lo-fi beats: "For putting pen to paper and getting stuff done, I always listen to chill lo-fi beats. Having it in the background helps me get in the zone. I also have light systems in my home that react to music—it really creates an ambiance."

—**Gu Xie**, Group 1001

Charley Crockett, \$10 Cowboy: "I'm a huge music fan and I listen to a lot. This one stays with me, it's one of my favorites of the year."

—**Justin Grimme**, Shane Co.

Stick Figure: "Personally, professionally, this band keeps me grounded. They're better than any self-help book out there."

—**Matt Florian**, Hakkoda

Evaluate, Interpret, Plan, Act: "Andrew Ng's incredible, succinct, and elegantly presented talk at Stanford in April, where he outlined on a whiteboard his simple framework for how agentic behavior might start to emerge. It was absolutely incredible for its scope and elegance, and how it captures the ways in which this is changing the world."

—**John Cosgrove**, Cloudwerx

Automation of Automation



FRANK BELL

Founder, Data Thought Leader
ITS - Snowflake Solutions

“Automation of Automation”, human + AI trends accelerate

The [Automation of Automation concept](#) explores the idea that all processes have inputs and outputs, and that more and more data and AI processes and automations will be linked together to create massive improvements and efficiencies in productivity and knowledge. The management of these automations will increase in importance, with data automations requiring more tools that manage data operations and data AI outputs and inputs. We'll also see the continued progress and usage of human + AI solutions that will combine human and AI subtasks within tasks to achieve high-quality outcomes, using what generative AI does best combined with what a human does best. Automated code generation and testing at speeds and iterations we have never seen before is already happening.

The cloud democratizes AI and centralizes knowledge

AI democratization through advanced cloud platforms is also another exciting development that's around the corner. We're approaching a future where

businesses of all sizes will be able to leverage highly sophisticated AI tools, thanks to platforms like Snowflake that simplify data management and AI deployment. While it might not fully mature in 2025, this is a game changer for companies that traditionally lacked access to cutting-edge data technologies.

Organizations will push for more cross-functional, data-literate teams

Leaders are increasingly focused on fostering a data-driven culture where data literacy is as important as technical skill. While AI is transforming how data is processed, the human element remains crucial for now. Data teams are being built with a mix of technical talent and domain expertise to ensure AI models are not just powerful but also practical and aligned with business needs. There has been a shift toward cross-functional teams, with data professionals collaborating closely with business units to drive insights and value. Data literacy programs are becoming more common to enable non-technical stakeholders to better leverage both AI and data in decision-making.



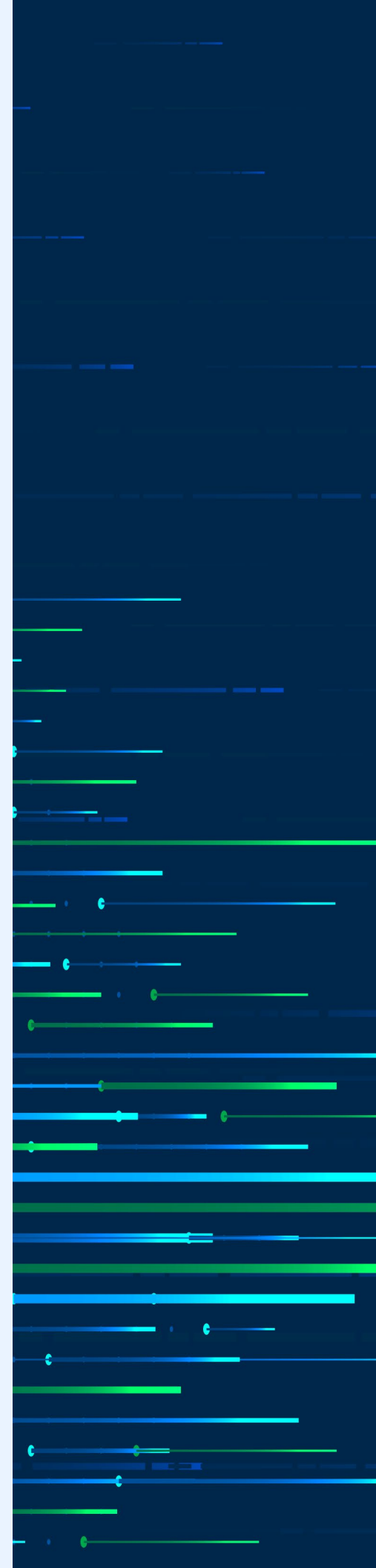
AI and ML models along with their data inputs and outputs require ongoing governance, monitoring, and retraining as data and business conditions evolve.

In 2025, I hope businesses start to **treat the combination of data and AI as an evolving tool that needs continuous oversight around quality to be effective.**

While some of the hype will dissipate as the market matures, it will likely remain a challenge in 2025 due to the limited understanding of AI and data capabilities as well as the fast pace of AI innovation.

FRANK BELL

ITS – Snowflake Solutions



From data gravity to platform gravity



CHRIS TABB
COO and Co-Founder
LEIT DATA

Data gravity will evolve into “platform gravity”

The concept of data gravity refers to the idea that the more data we have co-located in a single location, the more data, applications, and data services it attracts, thereby multiplying the value in that data store or location.

It’s gained popularity in recent years, but I believe we’ve now come to the next evolution of data gravity: platform gravity. Whereas before we were just bringing data into one place, now we’re bringing decisioning, applications, AI, and more to be run in the Data Cloud. This triggers fears of vendor lock-in for some, but it also allows you to solve so many technical problems along the way, such as avoiding the risk and cost of having multiple copies of data moving around. You can do so much in a single platform like Snowflake, such as run AI algorithms, containers, and connect to APIs. You have this entire accessible ecosystem and you can bring the compute load there, too. This trend will continue in the future, with the eventual combining of analytics and operational workloads into a single platform.

Data quality and modeling become more crucial than ever

It seems everyone has a budget for AI these days, but the only AI projects that will be successful are those that have ownership, data quality checks, and modeling in place in a way that’s extendable and easy to change. I don’t mean just a general data modeling process, modeling within your platform, semantic modeling, modeling for business value. I mean modeling expansively—the whole project modeled from conception to business application, and linked to business processes.

Without quality data you get bad results, and this comes into play with modeling. Generative AI has definitively proved to everyone—especially those in the C-suite—that models trained on bad data get bad results. The people making decisions understand that they need to be able to trust that data if they want the benefit that comes from making data-driven decisions. You can’t justify allocating budget for AI if you aren’t doing these things.

Templates eliminate friction

Templates, sometimes called “pattern catalogs,” help to decentralize architectural patterns and make them self-serviceable for business users, but without sacrificing control, security, automation performance, and the ability to maintain functionality at a later date. This is increasingly important as the gap between business and technology closes. Technology is easy—it’s the people and processes that are hard.

Leveraging a template approach—that is, using a roadmap that defines the right guardrails at the right time and templatizes patterns that are repeatable—makes everything easier. Where is there friction in delivery? By “friction” I mean time, complexity, and effort to perform an activity. The frequency of that dictates whether it is worth optimizing. We want self-service, but we also want to standardize data modeling and data security to promote reusability while minimizing security risks. Using a pattern catalog to do that can make a huge difference in creating data products that are supportable, extendable, scalable, and delivered at pace.



Architecture is a mindset, not a role. It’s something that should be shared.

The gap between business and technology is closing, so product teams are getting closer to data programs.

As a result, **data is more closely aligned to business strategy than ever.**

CHRIS TABB
LEIT DATA

New insights from unstructured data



BARR MOSES
CEO and Co-Founder
Monte Carlo

Adding structure to unstructured data will become a money-making endeavor for data teams

Before the rise of generative AI, most of the insights companies extracted from their data were from structured formats—data neatly arranged in columns, rows, and tables, making it easier to organize, clean, analyze, and utilize for various purposes.

Now, we're seeing many of our customers build successful enterprise AI products that leverage previously untapped unstructured data. A great example is a Fortune 500 insurance company that built an LLM to analyze customer service calls (audio files, or unstructured data), generate transcripts, and apply a simple rating scale (structured data in columns of single digits). By structuring the unstructured, this new data set of customer satisfaction scores unlocked further opportunities for machine learning modeling, uncovering new insights like predictive values or churn rates, which would have previously been impossible—or, at the very least, expensive.



What I'd love to see more in 2025 is **organizations embrace process just as much as they do technology.**

Simple operational changes, such as determining who owns data quality for what domain, or establishing strong SLAs for data reliability that map back to business objectives, could reap huge benefits for the overall health of an organization's data and AI estate.

BARR MOSES
Monte Carlo

However, teams shouldn't simply slap an LLM onto an unstructured data set and hope for the best. Strong data reliability principles must be applied to your model, newly structured data, and any subsequent analysis to ensure the insights are trustworthy and actionable. These include setting SLIs for key data metrics like freshness, distribution, and volume, and measuring against those SLIs over time to assure that its reliability isn't just a flash in the pan. And, of course, exploring ways to automate these processes as much as possible to account for the large volume and complex nature of the data.

Big players will continue to jump headfirst into the AI stack

While we saw Snowflake, Databricks, Fivetran, Monte Carlo, and others come together to form the "Modern Data Stack," the post-modern data + AI stack is starting to crystallize, and that will only accelerate in 2025. Earlier this year, Databricks and Snowflake announced their own LLMs with Cortex AI and DBRX respectively, for customers to query data with natural language, build gen AI pipelines within their respective platforms, and generally make AI easier for their customers.

And in tandem they launched open source versions of their catalogs (Snowflake Polaris and Databricks Unity Catalog) to enable greater interoperability, scale, and perhaps most importantly, governance. Both these moves are clear indicators that the two major players in the cloud database space are not only going all in on AI, but putting a premium on helping their customers ensure their data is AI-ready.

Data downtime will be a growing problem for the industry

Ever since we founded Monte Carlo back in 2019, we've been fighting an ongoing battle against what we refer to as "data downtime," the periods when data is unavailable, inaccurate, incomplete, or otherwise unusable.

Unfortunately, I don't see that problem going away any time soon, and in fact with the increase in production-grade AI, this problem will only grow. Monte Carlo conducted a survey a few months ago with Wakefield Research that revealed two-thirds of data leaders had doubts about their data's AI readiness, despite 100% of them feeling major pressure from their leadership to build gen AI products as soon as possible. If left unchecked, this glaring dichotomy could mean data downtime wreaking havoc on even the best-intended data and AI strategies.

I see 3 major trends in data management that I think will only continue to gain steam in 2025:

01. AI-POWERED EVERYTHING

When it comes to maintaining the quality of your data, whether it's monitor creation, anomaly detection, root-cause analysis, or any number of other critical data reliability workflows, AI can take it to the next level.

As I've mentioned, whether it's for real-time data streaming, RAG architectures, or pretty much any other AI use case, the quality of the data is everything. I expect a massive spike in interest for how data teams can unlock the power of AI to take data quality management to the next level.

02. BRING ON THE OPEN TABLES

Over the past few months, we've seen open source table formats such as Apache Iceberg soar in popularity as a way to more quickly and easily process large data sets at scale. While modern data warehouses and lakehouses offer both compute and storage, Iceberg focuses on providing cost-effective, structured storage that can be accessed by the many different engines that may be leveraged across your organization at the same time. A word of caution to data leaders exploring this option for their business: because of the engine-agnostic nature of open tables, challenges such as data inconsistency, incompleteness, and inaccuracy are likely to be hurdles you'll need to jump. Nevertheless, as the lakehouse becomes a de facto solution for many organizations, open-table formats like Iceberg are sure to continue to grow in popularity.

03. MORE DATA, MORE AI... BUT LOWER CLOUD COSTS

Today's data leaders are faced with an impossible task: use more data, create more impact, leverage more AI—but lower those cloud costs. The good news? There are a variety of ways you can reduce cloud storage and compute costs using the tools already in your data stack. For example, your data warehouses and lakes allow you to configure resources like memory, storage, compute resources (CPU/GPU capacity), and concurrency settings (determining how many queries or users can access the warehouse). And you can leverage tools such as data observability platforms to monitor the performance of your Snowflake and Databricks queries, making sure you're not eating up precious compute with poor-performing workloads.

Companies will double down on data products

I think we'll see a major trend in companies not only thinking about their data like a product (which is important!), but structuring their data teams and operational workflows around the data products themselves.

Thinking about data as a product can do a few very useful things for data leaders: First and foremost, it provides a helpful framework with which to approach AI, which technically is a data product in and of itself. It also helps align resources and effort around the data that actually matters to the business. It's impractical to build high-powered monitors and dedicate precious engineering and analyst resources to every single data asset in your data estate. Instead, focus on those data products that actually move the needle for the business, and apply governance to them.

From there, you can start making the decisions that move your data quality efforts from a case study in firefighting to a well-oiled, operationalized machine. For example, is my team ready to move toward a domain-driven, decentralized structure, grounded by a centralized platform team to codify tools and workflows? Is the data product fundamental (a centralized asset that's being used by hundreds of users across several domains), or derived (an asset built for a specific use case and owned by one or a handful of users, usually in one business domain)? This line of questioning will help your organization gain a clearer picture of how to tackle and ultimately operationalize quality.

Iceberg and open table formats surge



HUGO LU
CEO and Founder
Orchestra

Iceberg and open table formats will surge in popularity

Why can't data engineers just build microservices like software engineers? One of the main objections is the need for duplication of data as well as the high costs of moving data between microservices. As a result, there's a tendency to try to centralize data in a single place to minimize these costs. Historically, this has meant centralizing data in an on-premises warehouse or cloud warehouse. Now with open table formats, data teams can leverage object storage as the central layer. This has many benefits: lower data movement costs, cheaper and more flexible storage, interoperability with different query engines, and a minimized security footprint (all enterprises use cloud object storage anyway). But the greatest benefit of an open table format is that software teams can keep their copies of the data there as well. Open table formats and Iceberg finally offer a chance for "unification," and in conjunction with these other benefits should become increasingly popular as more success stories emerge.

Data teams will focus more on data foundation basics than AI

Some data teams are still viewed as cost centers, and although the hype around AI is real in some areas, it remains to be seen how this will affect data teams. Indeed, employees leveraging AI are mainly using tools such as ChatGPT or GitHub Copilot. Very few real, widely adopted use cases for AI exist right now. As economic conditions become more uncertain, companies without a built-out map of how data teams can provide value will practice caution, both in budget allocation and project prioritization. It's unlikely that AI will take precedence, and instead consolidation, data quality, and simplification will be the chief focus for many enterprise data teams over the coming year.

AI-enhanced data engineering efficiency and workflows



ADAM FOKKEN
CTO and Co-Founder
phData

Generative AI will greatly streamline data engineering operations to improve efficiency and workflows

The most important trend we're seeing is the increasing integration of generative AI models into data engineering. Data warehousing ecosystems have started to employ their own gen AI assistants, allowing a dialogue-based interface for everything, including querying data, generating code, creating test data, and implementing business rules. We've found this vastly improves the efficiency of DE workflows and allows them more time to focus on creating and managing pipelines.

The potential for fully autonomous data engineering platforms using AI is exciting. Even if it's just automating simpler "cookie cutter" pipelines, it could open up so much time for data scientists and data engineers to focus on more strategic initiatives.

Growing demand for domain-specific data expertise will give rise to the popularity of fusion teams

The demand for domain-specific expertise in data engineering and machine learning is growing. Enterprises are seeing more business unit-level demand for data and AI applications. This results in a need for engineers who can bridge that gap between technical knowledge and industry-specific understanding. These engineers move faster, are more prescriptive, and can better determine the best features to include for that specific vertical.

Related to that, we see organizations increasingly move toward data fusion teams to co-develop data solutions. Fusion teams are a blend of data engineers, machine learning engineers, analytic engineers, and business domain experts. These types of teams are really helpful in building the foundation for AI applications. By bringing business domain experts and data experts together, the business is able to have more ownership and awareness of these AI apps, leading to increased buy-in and utilization of them.

A reckoning that without a solid data foundation, AI projects will fail

Everyone wants to reap the benefits of using AI, but without a solid foundation of clean data, organizations won't get the results they're looking for. Garbage in, garbage out, as they say. I think this will change as these businesses start to realize they're not seeing the output they expected from their models, and they'll learn that having clean data is the most important aspect of getting AI right.



One misconception I hope gets put to rest in 2025 is the belief that data engineering and machine learning are purely technical disciplines. **Effective solutions come from engineers with technical skills, domain knowledge, and business acumen.**

ADAM FOKKEN
phData



What's the wildest client request you've ever had?

A 2-hour flight for a 30-minute meeting—the next day: “From Seattle to San Francisco. This was before Zoom, of course.”

—**Lee Derks**, DigBI Consulting

A mainframe to a new architecture model in less than two years. “It took 10. And we were on a death march for about five of them.”

—**Chris Tabb**, LEIT DATA

Brewery supply line report: “When I worked for a brewing company, they had all these sensors that integrated into their supply line. We used these different data sources to build a report that would alert brewers when something hit a certain trigger and the beer would have to be purged from the line. SAP said it wouldn't be possible, but we built it.”

—**Justin Grimme**, Shane Co.

A magical pipeline fix: “To magically fix 25 or 26 different database pipelines that were supposedly the source of truth for a client's accounts, inventory, and till takings—without anyone knowing what the correct answer was. They said, ‘We'll know it when we see it.’ That was also when I learned exactly what the worst possible statement of work (SOW) looks like, and why I'll never sign one like that again.”

—**John Cosgrove**, Cloudwerx

Rapid value from query-enabled LLMs



JUSTIN GRIMME
Manager of Business
Intelligence
Shane Co.

Query-enabled LLMs for business users will ease pressure on data teams

Something that I saw at the Snowflake World Tour this year that got me excited was the company's Cortex AI, which offers a way for data teams to build out a model and make it accessible to business users. It takes something similar to a ChatGPT and integrates it with your data so that users can ask questions. For example, you could ask it what your sales were in Q3 of this year versus last year for a specific product and region. Building data models that support this kind of functionality means users don't have to go to the IT team and request a specific report. Instead, the IT team builds an answer portal where users don't have to log in to a system and they don't need to know any code—they can simply ask a straightforward question in everyday language and get a correct answer back. This is an area where AI can provide very rapid value against your actual data set, and I'm excited to see it continue to develop in 2025.

Consumer attitudes toward data privacy will harden

For a long time, we've been living in the Wild West when it comes to data. It goes out into the world and it's difficult to rein in. But consumers now have more control over the data they share, and I think we'll start to see more people opting out of data clauses and even taking back their data from large companies. The United States generally seems to be behind Europe when it comes to data protection and privacy, but this movement is already trending in California. I think we're going to see similar momentum across the country in 2025.

People are increasingly aware of their data and where it goes, especially because conversations around AI are more common, and AI specifically can feel invasive. People want to control their own data—it's going to become a priority.



Proprietary data gets augmented

Data is an asset. And the Internet of Things is so large that trying to find the data that is actually actionable and incorporating it into your data sets can be difficult, but it is a point of focus. Similarly, mining proprietary data to find “dark data” is also increasingly important. Dark data is the data that is already in your platform but you’re not gaining any insights from. Finding more insights in the data that you already have and augmenting your existing data with related data to make it more valuable is a big, big thing.

Purpose-built platforms proliferate

I think more small and mid-size businesses are going to move away from big systems like SAP, and instead use their own purpose-built data and analytics platforms. Some bigger companies are building their own platforms, too, but more as sidecars. Big providers that have been in the space for decades now have a price tag associated with them that’s far too large. It’s hard to stick with them, and replacements for systems like SAP can be built at a fraction of the cost. These are modern, sustainable, and scalable for each company’s specific industry and allow for integration of new, different data sets and sources—thus unlocking the possibility of answering more complex questions. It’s complicated to replace these kinds of systems. It takes a lot of work, so smaller organizations are better able to make the move.

People want trust.

Building something reliable and actionable that you have confidence in means providing transparency for all users.

Visibility into the model allows you to integrate AI more easily down the road because the trust is there. **Users have confidence in what you’re delivering.**

JUSTIN GRIMME

Shane Co.

Semi-structured data upends existing data management practices



CINDI HOWSON

Chief Data Strategy Officer
ThoughtSpot

Semi-structured data will upend archaic data management practices

The data and analytics industry spends \$103B annually in storing structured data. Spending for semi-structured data in terms of text, voice, images, slides, and documents is unknown and often separate from relational data. IDC estimates that semi-structured data accounts for [90% of the world's data](#). Generative AI promises to make this data more accessible and useful. For example, Versalytics and VoiceBase use call data to understand sales and technician performance to optimize customer service. Unilever is using satellite imagery to understand the impact of sustainability efforts in the many consumer products they produce.

AI startup A79.ai allows life sciences clinicians to analyze comments of adverse reactions in clinical trials.

Historically, proprietary storage and processing systems have made semi-structured data difficult to access, hence why it is often called dark data. As we wrote in [ThoughtSpot's 2024 data and analytics trends and predictions](#), gen AI has made dark data more accessible. And yet our data management practices remain primarily focused on structured data.

In order to more effectively analyze all types of data, we must evolve our data management practices to include a variety of storage formats and the requisite security and governance policies. Analyzing this data distinctly offers some value, but analyzing these data sets together is the game changer. **Consider a few examples:**

Use case	Semi-structured data	Structured data	Game-changing questions
Customer loyalty	Call center recordings	Customer sales	How does empathy of agent correlate with customer retention and NPS?
Insurance	Satellite images of structures	Insurance policies and claims	Can signs of structural weakness pre-empt claims?
Supply chain	Satellite images of deforestation	Units produced	How many units of oil did I produce compared to number of palm trees used?
Life sciences	Clinician notes about adverse reactions	Patient demographics	What characteristics and health of a patient are more likely to correlate with an adverse reaction?

The evolution in data management practices is both organizational and technical. Knowledge management teams have often looked after semi-structured data through document management platforms, whereas data warehouse teams have looked after structured data. These systems also were often isolated by different storage formats. Security policies rarely spanned both.

Vendors are taking different approaches to bring the different data types together. For example, Snowflake's new Document AI essentially puts structure on top of PDF documents. A number of analytics vendors have added support for Glean. Data governance vendor Securiti recently launched Gencore, which offers governance across data sets.

In order to tap into the transformative combination of semi-structured and structured data, business teams and data teams must collaborate to:

- Imagine the art of the possible in formulating analytics use cases. Instead of asking, "What transactional data do we have?", flip the thinking to "If we really want to know our customers, patients, and employees, we need to see this type of data."
- Extend data enablement (governance) policies to semi-structured data sources.
- Evaluate solutions from existing data and analytics vendors as well as new solutions to mine this previously dark data.

RECOMMENDED RESOURCES:

[What is unstructured and semi-structured data?](#)

[An example LLM that summarizes product reviews](#)

[How Airbnb uses unstructured data](#)

[Google Gemini video for analyzing structured and unstructured data](#)



Coalesce revolutionizes data transformations to accelerate the delivery of data projects. Recognizing data transformation's critical role in the analytics lifecycle, we've created an inclusive developer platform that automates most SQL coding without sacrificing flexibility. Our platform boosts data team efficiency tenfold, allowing faster data pipeline development while empowering organizations to concentrate on extracting maximum value from their data.

Discover more at [Coalesce.io](https://coalesce.io).

