



If AI Is So Smart, Why Can't It Understand Your Enterprise Data?



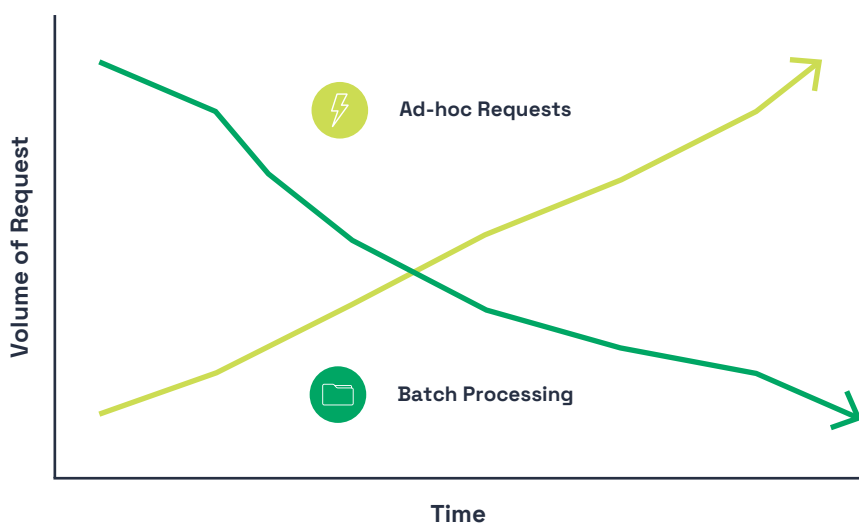
Table of Contents

The Era of Instant Gratification: The Promise of Generative AI	3
Ad-Hoc Analysis is a Problem for Data and Business Teams Alike	4
3 Challenges Standing in the Way of Instant Enterprise Data Analysis	5
Challenge 1: Data Silos	5
The Limits of Lightweight ETL/ELT in Complex Environments	6
Quality Assurance (QA) in an Unstable Foundation	7
Challenge 2: Data Pipelines Can't Keep Up with Real-Time, AI-Driven Analytics	7
The High Tax on Business Questions	8
The Contextual Bottleneck of Pipelines	8
The Velocity Trap	9
Challenge 3: AI Lacks Enterprise Data Context	9
3 Real-World Examples of Enterprise Data Context	10
When Definitions Drift, AI Fails Silently	11
Context is Localized and Fragmented	11
Generative AI's Opportunity for the Instant Gratification Era	12
A New Approach is Emerging	12
Key takeaways	13

The Era of Instant Gratification:

The Promise of Generative AI

The enterprise analytics mix is changing quickly from batch processing to ad-hoc requests. Today, **76% of data teams** spend half of their time preparing ad-hoc reports for business teams. In industries like retail, finance, or tech, data analysts might spend up to **70% of their time** on ad-hoc requests.



That time is only expected to increase, as generative AI tools like ChatGPT, Gemini, and Claude have reset executives' expectations for quick results. In fact, **83% of executives** report increased pressure to make faster strategic choices over the past three years. More than half of these leaders admit they've faced decision regret due to a lack of timely data access.

However, data teams know that getting quick and reliable answers from enterprise data is a massive challenge: **61% of companies** say the average turnaround time for a data request is more than 1 week. That's because the process of carefully constructing pipelines takes time. **Data teams find themselves struggling to scale with the demand for faster, iterative answers.**

By the time the data team builds a pipeline, the business user's original question may have changed. For example, if a marketer is trying to figure out which customers have the highest propensity to churn in the next 8 weeks, and it takes up to 4 weeks to get an answer, will they have enough time to run an effective campaign? Is the question still the right one, given the amount of elapsed time?

Even when data teams do arrive at an answer, more than half say they get up to four follow-up questions from the business team for each of their fulfilled data requests. This back-and-forth motion shows that business users either may not know exactly what hypothesis they're trying to test, or whether their original hypothesis will evolve with further data. Since getting to the answer takes data teams so much effort, the process of proving a hypothesis is anything but fast — or easy.

Ad-Hoc Analysis is a Problem for Data and Business Teams Alike

There's a double-edged sword to ad-hoc data requests. Business users might not know if the hypothesis they want to test is worth pursuing until they look at the data. At the same time, undefined questions are difficult for data teams to answer.

As one data analyst on [Reddit](#) put it,

“20% of every ad hoc I’ve done has been asking people what question they’re looking to answer.”

This is a very different challenge than that of “self service BI,” a promise that has been discussed for decades. For business users within a single platform, domain-specific AI agents can remove technical barriers for data analysis challenges. However, data analysts still are very much an essential part of extracting consistent and reliable answers to business questions from complex, distributed enterprise data. Generative AI can help accelerate this process, but only if the right data architecture is in place first.

These questions remain:

- Can you use the power of (generative) AI to get instant access to all enterprise data?
- Can data analysts use generative AI to test hypotheses and create faster, more accurate data products for the business?
- And can you bring in all of the enterprise context you need to make these answers reliable and scalable?

That would be the real nirvana of using generative AI to get answers from enterprise data. AI could transform the ad-hoc workflow for data teams and business stakeholders alike. The good news is, this ultimate goal is achievable today.

Let's first explore some of the challenges of the traditional, pipeline-based approach to ad-hoc analysis. From there, we'll preview what data teams need to do to use AI to solve these challenges.

3 Challenges Standing in the Way of Instant Enterprise Data Analysis

As ad-hoc requests continue to consume more of data teams' time, three big architectural challenges stand in the way of their success. In the age of AI-powered instant gratification, the status quo of moving data to analyze it no longer works. The three barriers below can make it impossible for most data teams to fulfill ad-hoc requests quickly and accurately.

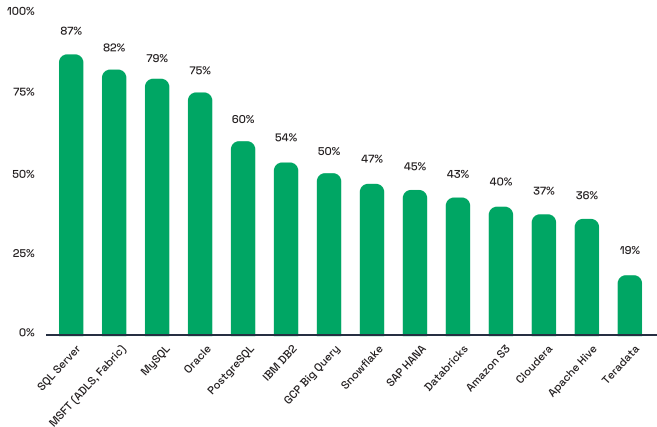
Challenge 1: Data Silos

Today's enterprises deploy an average of **976 applications, but only 28% are integrated** — a statistic that reveals a troubling disconnect between tool adoption and data unification. This fragmentation isn't just inconvenient, it's compounding. As of late 2024, 68% of organizations cited data silos as their top data architecture concern, a 7-point increase year-over-year. This signals that despite investments in integration tooling, the problem is getting worse, not better.

Each platform, application, or SaaS tool introduces its own logic for structuring, storing, and formatting data. This heterogeneity creates semantic mismatches and integration brittleness — challenges CDOs and data architects know all too well. Every new point solution added to the stack creates exponential complexity, especially when internal teams need to map relationships across schemas that were never designed to talk to each other.

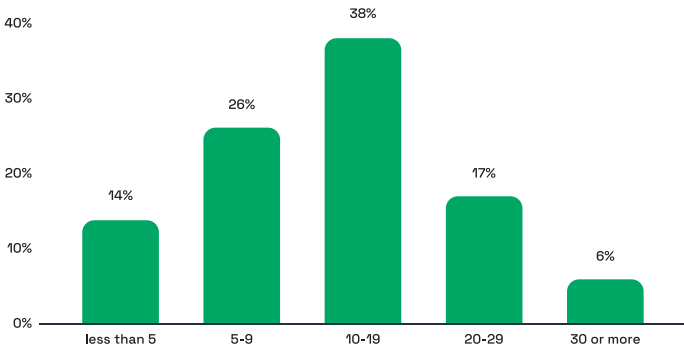
Proliferation of Data Sources

% of G2000 companies with HQ in NA & Europe



Number of Data Sources

% of G2000 companies with HQ in NA & Europe



Source: ZoomInfo, Company Reporting

Note: Based on a sample of 50 data sources based on the Gartner Magic Quadrant for Cloud Database Management Systems and their open source alternatives. Excludes companies with no available data.



The Limits of Lightweight ETL/ELT in Complex Environments

To cope, organizations have embraced more accessible and lightweight ETL and ELT frameworks, which help data teams to stand up pipelines faster. However, faster doesn't always mean more reliable. **These tools often assume that underlying source data is stable and that transformations are repeatable — assumptions that rarely hold in the real world.**

In practice, source data changes constantly: Field names are altered, API contracts shift, new values emerge without notice, and upstream systems introduce silent breaking changes. Despite strong data modeling or carefully tuned SQL pipelines, inaccurate or incomplete data frequently makes its way into dashboards, forecasts, and executive reports. This fragility forces data teams into a reactive posture — spending more time triaging anomalies than generating insights. When it comes to ad-hoc analysis, this process makes it impossible to get instant answers for business stakeholders.

This is a problem [@AnalyticsMentor](#) describes well in this [Medium article](#).

“When I was at eBay, one of my partner analytics teams owned a metric for calculating losses. The calculation wasn't technically complex, but there were 9 different fields that had to be pulled into the calculation. Unfortunately, the code and data wasn't properly architected, which meant that when other teams, such as my team, wanted to calculate losses, we didn't get it by querying a single value from a single table. We had to copy their code and logic. Again, not difficult, but a major issue nevertheless.

This is because any time the partners would change their formula, I'd have no way of knowing that it changed. The Losses team could change the logic in their code, but what about me and everyone else? Again, my code would likely run without error, but the results would be wrong. And they were. But even with proper engineering practices, issues arise. This is frequently experienced when production tables are changed or deprecated.”



Quality Assurance (QA) in an Unstable Foundation

Although QA checks on production reports are considered best practice, they're not a true safeguard in modern data environments. When ad-hoc analysis becomes the norm — which it often does in response to business questions — there's little time to run complete QA passes. Worse still, self-service business users are frequently querying datasets they don't fully understand, increasing the risk of flawed decisions based on misunderstood (or inaccurate) data.

As data environments grow in complexity, the illusion of end-to-end control breaks down. **Even organizations with mature data practices struggle to get accurate answers to business questions when the data they need is stored across multiple domains.**

Challenge 2: Data Pipelines Can't Keep Up with Real-Time, AI-Driven Analytics

For years, data teams have relied on creating pipelines and moving data to deliver clean, trusted answers at scale. Pipeline tools are highly effective for repeatable, predictable workloads, but are insufficient for ad-hoc queries. Why? Many times, data teams are moving data before understanding what is required to effectively answer the question.

However, with the volume of ad-hoc requests on the rise, pipelines simply weren't built to keep up. AI has re-engineered expectations toward high-volume and high-velocity decision making.

The problem isn't with the pipeline tools themselves. It's with the architecture. Pipelines prefer more static, definitive requirements, and are meant to transform data from Point A to Point B in a controlled way. To contrast, ad-hoc workflows are fast, iterative, and unpredictable — all data teams have is an idea they want to explore. **Simply put, it's hard to accelerate the pipeline to the speed of human curiosity.**



The High Tax on Business Questions

In theory, asking more business questions should be cheap. But, in reality, each new business question comes with a high tax. Data teams must locate the right source, map fields, clean or modify the dataset, and ultimately build or modify their pipelines. This effort does not scale. The cost of building a single pipeline is high, in terms of time and human resources. **As a result, ad-hoc questions often fall through the cracks, or worse, teams make quick decisions without data at all.**

The industry has spent billions building pipelines, but that value chain collapses under pressure as business needs shift from scheduled, batch reports to iterative, ad-hoc requests.

The Contextual Bottleneck of Pipelines

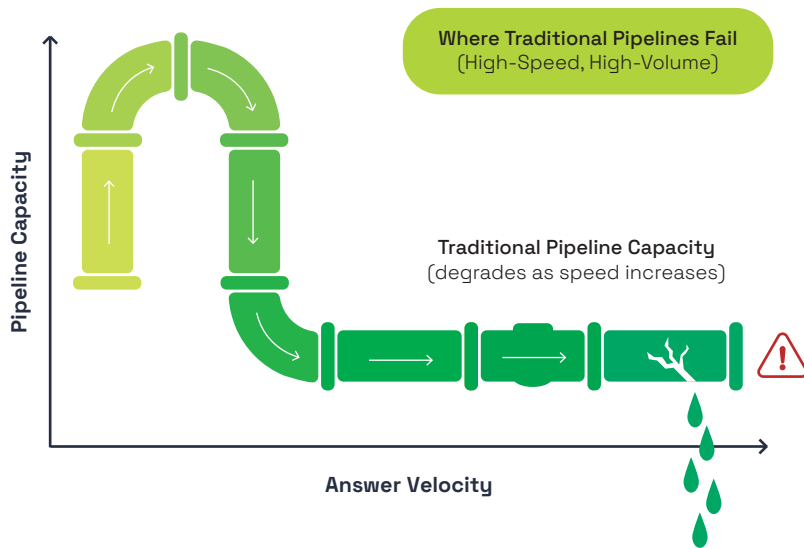
Even with automation, pipelines struggle with the contextual complexity that modern AI analytics demands. There's a three-way context problem:

1. Understanding the user's intent
2. Interpreting the available data
3. Knowing the semantic relationship between them

Most pipelines aren't context-aware. They move data, but they don't reason about it. This makes them poorly suited for AI workflows where semantic understanding is the core challenge, not just schema alignment.

The Velocity Trap

The faster ad-hoc questions come, the worse pipelines perform. They were built to optimize for volume, not velocity. Teams end up with too many bottlenecks: Think handoffs and dependencies between data engineering, analytics, and business users.



Despite modern tooling, **82% of companies** admit they are making decisions using stale data. This can lead to devastating consequences: 85% say relying on out-of-date data has led to incorrect decisions and lost revenue opportunities. The root cause? By the time the pipeline delivers a usable dataset, the window of business relevance has already closed.

Challenge 3: AI Lacks Enterprise Data Context

Generative AI excels when it has clear and consistent context. The more background data teams can give AI, the better results it will return. But, enterprise data systems aren't set up to deliver that context. At least not in the way generative AI can understand.

All too often, context lives in people's heads — not the data. Even when core metrics live in the same system, the meaning of those core metrics can change depending on who is using them. Take for example, a company whose base revenue figure lies in a central data warehouse or data lake. The CFO calculates revenue according to accounting rules, while the CRO uses the same number with a different formula aligned to compensation and sales goals.

Those definitions are critical — but they don't exist in the data or the data platform. Without context, a generative AI system would see "revenue," without knowing which revenue definition matters most. At best, you'll get an educated guess. At worst, you'll get inaccurate data back. Over **85% of AI projects** fail because the data is messy or incomplete.

3 Real-Word Examples of Enterprise Data Context

A seemingly simple business question could yield a variety of results, depending on the context available in enterprise data systems. Here are three examples of how contextual definitions — and the resulting answers to business questions — change depending on who’s asking and where the data is sourced.

“What’s the ROI of our last marketing campaign?”

- **Marketing Domain:** May define “revenue” as total contract value or number of leads converted. Sourcing out of CRM (Salesforce) and marketing system (Adobe)
- **Finance Domain:** May define “revenue” as recognized revenue in a certain accounting period. Sourcing out of financial system (Netsuite)
- **Sales Domain:** May track contract bookings that include future revenue — whether recognized yet or not. Sourcing out of CRM (Salesforce)

“Which products were most profitable last quarter, and what contributed to that profitability?”

- **Operations Domain:** ‘Profit’ might be seen in terms of direct costs (e.g., raw materials, labor). Sourcing out of financial system (Netsuite) and logistical operations (SAP)
- **Finance Domain:** May consider overhead allocations, depreciation, or other accounting adjustments. Sourcing out of financial system (Netsuite)
- **Marketing Domain:** Might associate ‘profitability’ with uplift from campaigns or brand contribution (not always included in standard cost accounting). Sourcing out of financial system (Netsuite) and marketing system (Salesforce Marketing Cloud)

“How many employees will we need next quarter to meet operational demands?”

- **HR Domain:** May consider staff availability, training needs, or attrition rates. Sourcing out of HR system (Workday)
- **Operations Domain:** Forecasts are based on production schedules or service demands. Sourcing out of logistical operations (SAP) and field service management (Service Now)
- **Finance Domain:** Is concerned with budget constraints, payroll forecasting, and approvals. Sourcing out of financial system (Netsuite) and field service management (Service Now)

When Definitions Drift, AI Fails Silently

Across the enterprise, the same term can exist in Snowflake, Databricks, Tableau, and many other systems — yet, be filtered, labeled, or calculated differently in each. “Revenue by product,” for example, might be straightforward in one system but totally ambiguous in another. The problem is not just duplication, but misalignment.

In other words, there’s a growing gap between the semantic model used on the frontend (where business users explore dashboards), and the data structures on the backend (which were built for scale, rather than clarity). As a result, AI systems trained on backend data alone won’t understand how frontend users interpret it. And because enterprise data doesn’t use a shared ontology, there’s no consistent signal to guide AI.

Even tagging isn’t enough. Tags are rarely complete, often outdated, and typically are applied differently across systems.

Context is Localized and Fragmented

Today’s data teams are expected to stitch together meaning across a wide variety of systems. As we’ve demonstrated above, terminology and definitions can vary greatly depending on the eye of the beholder. As data context continues to sprawl across data warehouses, data lakes, and BI platforms, the issue isn’t just with integration. It’s with the orchestration of meaning. **Without context, the same terms in two systems contain different truths.**

Some platforms have developed their own domain-specific AI agents to help interpret and answer business questions. While powerful, these agents’ context and understanding is limited to their single domain. The true promise lies in generative AI agents that can reason effectively — with the proper context — across domains.



Domain-Specific Agents

vs



Universal AI Agent

Generative AI's Opportunity for the Instant Gratification Era

The shift toward ad-hoc, AI-driven analysis is not a passing trend. It's the new normal. Business users expect answers instantly. Data teams are under pressure to deliver them. Yet, the modern data stack, designed for repeatable batch processing workflows, is buckling under the weight of fast-moving, iterative business questions.

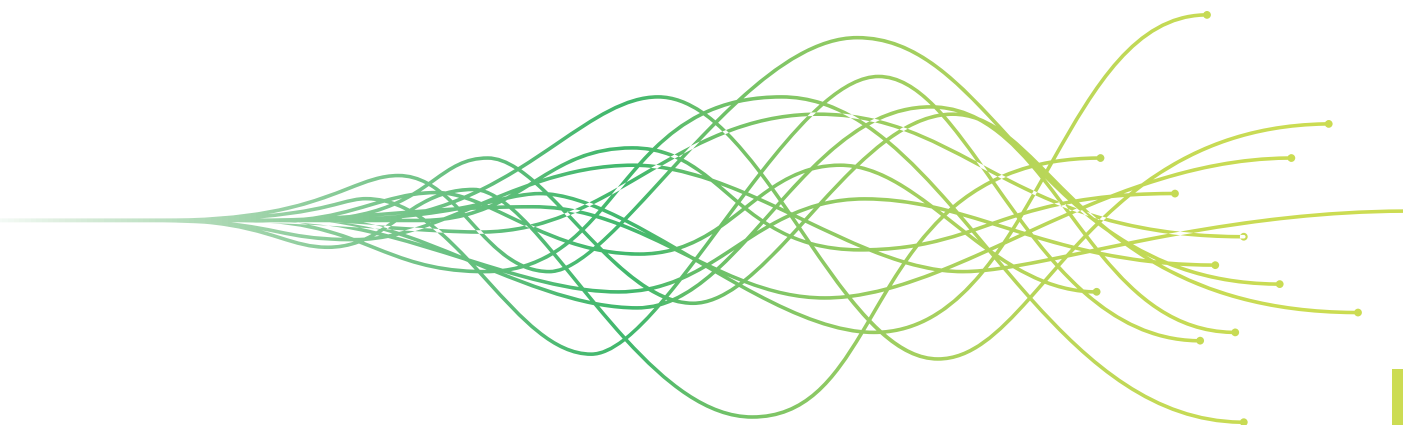
This is an architectural problem. Data teams need more than just pipeline automation to scale. Pipelines simply weren't built for instant answers.

A New Approach is Emerging

To meet the promise for instant, generative AI-powered analytics in the enterprise, **we need a shift in how context, collaboration, and computation come together.** Agentic architectures, designed to reason, act, and adapt, offer a path forward. When agents can interpret semantics, infer intent, and translate business questions into valid, executable SQL queries, we can start to close the gap between business curiosity and technical capability.

These agents are not replacements for analysts. They're force multipliers for data teams. For expert SQL users, they accelerate hypothesis testing for ad-hoc business questions. For less technical users, they offer guided exploration — similarly to how Data Robot has democratized data science. In both cases, the outcome is the same: Higher quality answers, delivered faster, with fewer dependencies.

The future of ad-hoc analysis won't be built on faster pipelines alone. It will require architectural changes that embrace AI, preserve enterprise data context, and empower data teams to scale. With the right foundation, AI can finally deliver on its promise: Instant, reliable, enterprise-grade insights.



Key takeaways

- **Ad-hoc demand is growing fast.** It's outpacing the capabilities of today's pipeline-based data stacks.
- **Traditional tools aren't enough.** They lack the context and semantic understanding to answer real business questions.
- **Generative AI offers a way forward.** But only when it's paired with agentic systems that can reason across messy, distributed, and evolving enterprise data.
- **Agentic architectures can scale data teams.** They can make technical users more productive and enable less-technical users to explore safely.
- **The opportunity is real — and available now.** Organizations that are ready to rethink how they design, structure, and scale their data workflows can explore our resources to learn more.

Want more detail about using AI to deliver instant insights and analytics?
Read: [Unlocking AI for Enterprise Data: Overcoming the Roadblocks to Trusted, Real-Time AI Insights](#)

