blue margin

WHITE PAPER • 2026

AI Needs a Foundation Too

Why “just point it at your data” doesn’t work, and what does

AI Needs a Foundation Too

Why "just point it at your data" doesn't work, and what does

A Blue Margin white paper. June 2026.

The short version

By now you have probably heard it, or said it yourself. With AI, you no longer need a data or BI team or partner, because you can point it at your business and start asking questions. The software vendors are encouraging this. Salesforce, Oracle, Microsoft, Snowflake, Databricks, HubSpot, and a dozen smaller players are all selling some version of the same promise. Ask your data a question in plain English, get a trusted answer.

That promise is absolutely achievable. The demos are not lying about what is possible. But the demos do not show what it takes to get there, and that gap is where mid-market companies run into trouble. Point a general-purpose AI, or a single vendor's built-in AI, at raw business data and the answers come back fast, confidently stated, and sometimes wrong, in ways the person asking cannot always detect. This is not a criticism of AI. It says less about the models than about what they have been pointed at.

The work that makes AI answers trustworthy is the same work that has always made business intelligence trustworthy: Integrating data across your siloed systems, modeling it, governing it, and encoding your business definitions. AI has not eliminated that work. If anything it raises the stakes. A governed dashboard can still be wrong, but the modeling, testing, and agreed definitions behind the dashboard make it right far more often. An AI pointed at unprepared data has none of that backing, and because it answers fluently and rarely the same way twice, its mistakes are both more likely and easier to miss.

The skip-the-data-foundation shortcut takes several forms, and each one runs into the same wall when it meets a real business. Your data is scattered across systems that have to be integrated before any question that matters can be answered. Your definitions, the ones your team argues about, have to be encoded somewhere the AI can read them. The numbers that go in front of a board have to come back the same way every time. And at scale, regenerating routine reports with AI can get expensive fast. Some companies try to sidestep all of this by consolidating everything inside one application vendor's cloud and letting that vendor's AI analyze it, which solves nothing and adds a lock-in problem that is hard to reverse. None of these are reasons to avoid AI. They are reasons to do it right with a real data foundation and governed reporting underneath it. The companies that get value from AI analytics run it alongside their BI, not instead of it. You need both to run the business well.

This paper walks through those problems one at a time, using the vendors' own documentation as the evidence, and lays out what it actually takes to get trustworthy answers from AI.

The promise you're being sold

If you are a mid-market CEO or CFO, you have heard some combination of these four pitches in the last year. They take different shapes, but they feed one belief, that you can skip the foundation work, that you no longer need dedicated BI, or that a single tool will simply handle all of it.

1. **Export to a general-purpose LLM.** Send your data to ChatGPT or Claude, ask questions, get answers. No software to buy, no vendor to hire.
2. **Switch on the embedded copilot.** Your ERP or CRM already has AI built in, whether NetSuite, Salesforce Einstein, or Microsoft Copilot. Just turn it on and ask.
3. **Consolidate on a single vendor's cloud.** The most sophisticated of the four. Land all of your data in one vendor's cloud, the Salesforce cloud or the Oracle cloud, and run all of your modeling and analytics there with that vendor's AI. It is persuasive because it concedes something true, that you do need to bring your data together, then makes a leap that does not follow, that the place to do it is inside one application vendor's walls.
4. **Wait and see.** This one usually comes from inside the company. Microsoft will build this into Power BI for free soon, or we'll build our own with AI when we get around to it, so why work on it now. While the official answer is to hold off, the reality is someone is often already pasting exports into ChatGPT to get by, which just runs the company straight into the first pitch's problems with no foundation underneath.

The vendor language is nearly interchangeable. Databricks invites you to "Talk with your data."[1] ThoughtSpot promises to bring "the analytical and reasoning skills of a data analyst to the fingertips of every user."[2] Snowflake says that "any Snowflake user can analyze structured data and get reliable answers to their questions just by asking natural-language questions."[3] Microsoft claims that "even people without technical expertise in AI or a deep understanding of the data structure can receive precise and context-rich answers."[4] HubSpot offers to let you "Ask questions and get answers grounded in your CRM data and

---

[1] Databricks, "AI/BI Genie product page: 'Talk with your data.'"

[2] ThoughtSpot, "ThoughtSpot Launches Spotter, the Autonomous Agent for Analytics," GlobeNewswire, November 13, 2024.

[3] Snowflake, "Getting Started with Cortex Analyst: Augment BI with AI."

[4] Microsoft Learn, "Fabric data agent creation" (ms.date May 11, 2026). Cited for the natural-language-access claim.

role."[5] The implied takeaway for a busy executive is that AI has collapsed the entire analytics stack into an LLM chat box.

It has not. And the clearest evidence comes from the vendors themselves.

What the shortcut quietly assumes

Every one of those pitches works beautifully under three conditions. The data lives in one clean system, everyone agrees on what the metric definitions are, and the person asking already knows the right way to ask their questions. A polished vendor demo satisfies all three by construction. It runs against one curated dataset, with definitions pre-loaded, answering a question the presenter chose in advance.

A real mid-market company often satisfies none of these in the wild.

Consider a composite picture drawn from companies we talk to every day, a $90 million IT services firm running NetSuite for financials, Salesforce for CRM, a separate ticketing system for service delivery, a remote-monitoring tool for operations, and a pile of spreadsheets holding the logic that ties it all together. Customer records do not share IDs (or even the exact same name) across those systems. "Revenue" means one thing to the finance team and another to sales. Margin by customer cannot be calculated at all today, because costs are allocated across five business lines through cross-charges that live in one analyst's head. The board wants a profitability view in 90 days. The controller builds it by hand, exporting from the ERP into pivot tables, the way she has for years.

Now point an AI at that. Which system does it read? It cannot see across all of them, because nothing has logically joined them. When it answers a question about a customer, which of the conflicting definitions for a metric does it use? When it computes margin, where does it get the cost allocation logic that exists only in a spreadsheet and a person's memory? It does not, so at worst it makes something up that looks plausible. None of this is the AI's fault. The questions a business cares about live in the overlap between systems, and the AI has been handed a single piece of the picture. The work of assembling the full picture, integrating those systems, reconciling the definitions, and encoding the allocation logic, is exactly what the shortcut assumes has already been done. If it hasn't, no tool can reliably do it for you on-the-fly in the background.

---

[5] HubSpot, "Breeze AI: 'Ask questions and get answers grounded in your CRM data and role.'" Marketing-page copy can drift; verified verbatim on May 31, 2026.

Where it breaks

Cross-silo blindness, and why the application vendors concede it

The most valuable analytics questions span systems. Profitability joins revenue from the ERP to cost from operations. Customer health joins CRM activity to support tickets to billing. Pipeline reliability joins sales forecasts to historical delivery. None of these can be answered from inside one application, because the data resides in several of them.

This is where the consolidation pitch tries to win. Put everything in the Salesforce cloud or the Oracle cloud, the argument goes, and the cross-silo problem disappears. But watch what the vendors actually do, as opposed to what their marketing says. Salesforce's flagship Data Cloud capability is called Zero Copy, and its entire design is to query data that remains in your external warehouse or lakehouse rather than absorbing it into Salesforce.[6] When the application vendor's own best architecture is to reach out to your neutral data platform and read it in place, the vendor has conceded where your analytical data actually lives. Databricks, describing the common case of integrating SAP and Salesforce data, is blunt about the alternative. Connecting applications directly "create[s] multiple data copies, introduce[s] latency, and make[s] governance difficult," and the answer is "one governed data & AI platform."[7] **Even Marc Benioff, Salesforce's CEO, who sells AI agents as hard as anyone alive, concedes the point. "Data has to be managed a certain way. That doesn't go away with agents. It doesn't go away with AI."[8]**

The single vendor consolidation pitch collapses on its own terms. To make a single application vendor's cloud the home for analytics, you first have to ingest, map, and model all the data that originated everywhere else. That work is building a data platform. You can do it on a neutral foundation that you control, or you can do it inside one vendor's walls and become locked-in.

Missing business context

An AI model knows what "revenue" means in general. It cannot know that your company books revenue as ARR only, excludes professional services, booking them to a separate entry, and recognizes it differently in the fourth quarter. Multiply that by every metric, every department's private vocabulary, every business rule that an experienced analyst applies without thinking, and you have the real gap. As Harsha Chintalapani, who built data platforms at Uber and co-created OpenMetadata, put it in a recent interview, "your LLM

---

[6] M. Orun, "Salesforce Data Cloud Zero Copy: When (and When Not) to Use It," Salesforce Ben, March 21, 2025.

[7] K. Satyavarapu, "SAP and Salesforce Data Integration for Supplier Analytics on Databricks," Databricks Blog, February 7, 2026.

[8] Marc Benioff, quoted in "Salesforce Agentforce and Data Cloud: A path to the software-only hyperscaler," SiliconANGLE, May 12, 2025.

issues are really data issues."[9] The engineers at Tinybird, who build analytical infrastructure for a living, found the same thing. Large language models, in their words, "are surprisingly bad at working with analytical data," and "struggle even to figure out what they're being asked."[10] The gap is easy to see in practice. Ask the AI for last quarter's revenue and it returns a clean, confident figure, except it counted the professional-services line that finance books separately and leaves out of the number the board sees. The AI did nothing wrong. No one told it which definition you use, so it picked one, and its answer quietly addresses a different question than the one you asked.

The solution is not a better model. It is capturing your definitions once, in a place the AI is required to read before it answers. That kind of context layer, where your company's vocabulary, exceptions, and business rules are written down in a form the AI can understand, is what separates a tool you can trust from one that flatters you with fast answers. Once it exists, the AI stops guessing. It reads what you mean, the same way every time, and answers accordingly.

Non-determinism, or ask three times and get three answers

A Power BI dashboard built on a properly modeled data layer is deterministic. The same question returns the same number every time, until the underlying data changes. An AI is probabilistic by design, meaning it generates each answer fresh from patterns in language rather than looking up a fixed result, so the same question can come back worded, formatted, or even calculated differently. A 2026 study that ran identical data-analysis prompts ten times each across six models found that the runs produced different conclusions about the same question, and that this held even at the most conservative "creativity" settings, where you would have expected identical output.[11] The researchers' recommendation was to run an AI analysis multiple times and consider the distribution of answers.

In everyday business use, the inconsistency shows up as much in the form of the answer as in the figure itself. Ask for last quarter's revenue by region and you may get a clean table one time, a paragraph with the number buried in it the next, and a chart the time after, and even specifying the format you want does not guarantee you'll get the same output each time.

That is often fine for a researcher. It is a disaster for a Monday morning operating review, where the regional VP needs the same revenue number calculated the same way and

---

[9] H. Chintalapani, interviewed in "Your LLM issues are really data issues," Stack Overflow Blog, April 28, 2026.

[10] J. Sancha and R. Moreno, "Why LLMs struggle with analytics (and how we fixed that)," Tinybird, January 15, 2026.

[11] J. Cui and R. Alexander, "Same Prompt, Different Outcomes: Evaluating the Reproducibility of Data Analysis by LLMs," arXiv:2602.14349, 2026.

showing up in the same place every week, and where two managers reading the same metric must see the same figure or the meeting derails. Variability that is acceptable for exploration is unacceptable for the repeated, governed reporting that runs a company.

Consistency is a property of the modeled layer underneath, not of the model itself. Define the metric once, where the calculation is fixed, and the same question stops coming back as a different number.

Confident, invisible errors

The dangerous failures are the ones that look like successes, where the AI returns a wrong answer that reads as a right one. Jamie Davidson, who runs the BI vendor Omni, puts it well. Text-to-SQL "fails silently and confidently. The output may look right, even if it's wrong."[12] The benchmarks bear this out. On academic test databases with a handful of tables, modern models translate questions to SQL with roughly 90 percent accuracy. Point those same models at realistic enterprise databases, the kind with a thousand columns or more, and accuracy collapses. On Spider 2.0, a benchmark built from real enterprise data warehouses, the best system at release solved 21.3 percent of the tasks, against more than 90 percent on the older, simpler benchmark.[13] On BIRD, another large-database benchmark, the strongest published results reach the low 80s in percentage terms, while a human expert scores nearly 93.[14] That gap is what separates a tool you can trust from one that embarrasses you in a board meeting.

This is also why even "90 percent accurate" is the wrong thing to celebrate. As one practitioner with two decades in analytics wrote, "[a]n accuracy of 80% or even 90% is, unfortunately, not enough,"[15] because the one query in ten that is confidently wrong, undetected, destroys trust in the other nine. Picture telling your board the Northeast is your strongest region and shifting next year's budget to match, then learning a quarter later that the AI had double-counted one large account and the Northeast was never the strongest.

---

[12] J. Davidson, "Why text-to-SQL fails," Omni, April 8, 2026.

[13] J. Lei et al., "Spider 2.0: Evaluating Language Models on Real-World Enterprise Text-to-SQL Workflows," arXiv:2411.07763, 2024 (rev. 2025); Spider 2.0 leaderboard. The 21.3 percent figure is the best system reported at the benchmark's release; engineered agentic systems have since scored higher on its sub-tracks, which is precisely the curated-layer approach this paper argues for.

[14] J. Li et al., "Can LLM Already Serve as a Database Interface? A BIg Bench for Large-Scale Database Grounded Text-to-SQLs" (BIRD), NeurIPS 2023; BIRD leaderboard (accessed May 2026; human baseline 92.96 percent).

[15] G. Zavaleta, "Why 90% Accuracy in Text-to-SQL is 100% Useless," Towards Data Science, January 12, 2026.

The fix is not a more careful prompt. It is the modeled, tested foundation underneath, where the joins are defined and each metric is computed one agreed way. AI pointed at a well-prepared data layer produces answers that are not just faster but are reliable. The same foundation that makes a dashboard trustworthy makes the AI trustworthy, because they are drawing from the same source.

The single-vendor cloud trap

Of the four pitches, the consolidation play is the one to push back on hardest, because its consequences last the longest. Set aside accuracy for a moment and assume it worked technically. You would still have made a strategic mistake. Concentrating all of your enterprise data inside one application vendor's proprietary cloud is the textbook definition of lock-in, and the lock-in is economic, not technical. Once the bulk of your data lives in one provider, the cost and complexity of ever moving it is what traps you. The major cloud providers understood this so well that they charged steep fees to extract your own data, and only waived those exit fees in 2024 under regulatory pressure from the European Union, and even then only for customers leaving.[16] When the industry waives a fee solely for departing customers, the fee was a retention mechanism.

There is also an incentive problem. An application vendor's analytics layer is built to keep you in the application vendor's ecosystem. It is optimized for retention, not for giving you the most honest, complete view across all of your data, including the data that lives in their competitors' systems. The defensible position is the one a veteran data architect summarized cleanly, "the data belongs to the company, not the engine."[17] Keep your data in open formats on infrastructure you control, and you can point any tool at it, including next year's better tool. Lock it in one vendor's cloud, and you are betting your entire data strategy on a single company's roadmap and pricing. Every other failure mode in this paper can be fixed with better engineering. This one can only be fixed by moving, and moving is exactly what the vendor has priced you out of.

The economics are not what they look like

AI inference is cheap right now. It will not stay this cheap. The frontier providers are running deep losses to capture the market. OpenAI's own investor documents, as reported in late 2025, project operating losses approaching three-quarters of revenue in 2028 and roughly $115 billion in cumulative cash burn through 2029.[18] Today's token prices reflect a

---

[16] R. Losio, "AWS Waives Egress Fees for Customers Exiting the Cloud," InfoQ, March 16, 2024 (following comparable waivers by Google and Microsoft, prompted by the EU Data Act).

[17] A. Buckenhofer, "Open Table Formats: From Vendor Lock-In to Data Sovereignty," data.KISS, April 12, 2026.

[18] D. Smith, "OpenAI plans to report stunning annual losses through 2028," Fortune, November 12, 2025, citing reporting on OpenAI investor documents by The Wall Street

land grab, not steady-state economics. The early movers are already hitting the wall. Uber, after rolling agentic AI out to its engineers, burned through its entire 2026 AI budget in four months, and its operating chief, asked whether that spend was translating into shipped value, answered plainly, "That link is not there yet."[19] The lesson is not that AI is too expensive to use. It is that pushing high-volume, repetitive work through it, with no discipline about where it earns its spend, is how the bill runs away from you. The practical consequence is that high-volume, repeated reporting, the hundred managers pulling the same dashboard every morning, is comparatively free in a governed BI tool and becomes genuinely expensive if every one of those views is regenerated by an AI on demand. The economics reinforce the same division of labor the accuracy argument points to. Use governed BI for the high-volume, repeated questions. Use AI for the long tail of exploration.

The part nobody is selling you

Step back and notice what every vendor's fine print has in common. Snowflake's natural-language analytics requires a hand-authored semantic model, which its own documentation calls "the key that unlocks" the product.[20] Databricks requires analysts to package metric definitions, sample queries, and join relationships, and concedes that "curating semantics is still a chore."[21] ThoughtSpot requires you to build models that establish "a domain-specific glossary for the LLM that reflects your unique business context."[22] Microsoft tells you to configure each AI data agent with instructions and example queries, a process it compares to "building a Power BI report,"[23] and states plainly that its Copilot "can't, and doesn't, aim to replace the people who today create and manage

---

Journal and The Information. Figures are company-wide projections, which include training and R&D, not a per-query inference margin.

[19] J. Angelo, "Uber burned through its entire 2026 AI budget in four months," Fortune, May 26, 2026, reporting remarks by COO Andrew Macdonald (originally reported by The Information). The example is enterprise AI coding tools, not BI, but the economics are the same: unbounded, high-volume AI usage scales cost faster than the value it returns.

[20] Snowflake, "Getting Started with Cortex Analyst" (the semantic model as "the key that unlocks Cortex Analyst's power").

[21] Databricks, "AI/BI Genie is now Generally Available," June 12, 2025; see also the Genie documentation on defining join relationships.

[22] ThoughtSpot, "Introducing Spotter: Your AI Analyst."

[23] Microsoft Learn, "Fabric data agent creation" (ms.date May 11, 2026). Cited for the comparison of configuring a data agent to "building a Power BI report."

reports."[24] Microsoft even caps its Fabric data agent responses at 25 rows by 25 columns,[25] which tells you it is built for conversational lookups, not for running a business.

Read those caveats together and the real picture emerges. The natural-language chat box is the last mile. Everything that makes it work, the integration across silos, the modeling, the encoded definitions, the governance, is the road, and the road still has to be built. To be fair, the application vendors are not hiding this. They are just selling the last mile and assuming you will build, or buy, the road.

There is a well-understood way to build that road. Modern data platforms use what is called a medallion architecture, a series of layers that progressively refine raw data into something trustworthy. Raw data lands in a bronze layer exactly as it arrives from each source system. A silver layer cleans, deduplicates, and joins it across systems, which is where the cross-silo problem actually gets solved. A gold layer models it into analytics-ready form, with business rules applied and metrics defined once, so that the same question produces the same answer. This is the foundation that has always made BI trustworthy, and it is the same foundation that makes AI trustworthy. Even renaming tables and columns so their meanings are easily inferred matters more than it sounds. Research presented at a leading database conference in 2025 confirmed that simply giving tables and columns clear, human-readable names measurably improves an AI's ability to query them correctly.[26] An AI handed a column called `ARH_TXN_AMT` guesses. An AI handed a column called `invoice_amount_usd` does not have to.

How you actually get there

Blue Margin builds this foundation for mid-market companies as an ongoing service, and over the past year we have extended it to make data genuinely consumable by AI. **We do it by adding a fourth layer on top of the gold layer, which we call the Platinum Layer.** The point of the approach is analytics you can trust and own, running on a deterministic base in your own environment, with no lock-in to any single vendor.

The gold layer remains the deterministic foundation. It is what your Power BI dashboards run against, and it is the source of the numbers you defend in a board meeting, *the same answer every time*. Then the Platinum Layer is engineered specifically for an AI to read. It takes the modeled gold data and prepares it three ways.

---

[24] Microsoft Learn, "Overview of Copilot in Fabric" ("Copilot can't, and doesn't, aim to replace the people who today create and manage reports").

[25] Microsoft Learn, "Fabric data agent creation," Limitations (responses "capped at a maximum of 25 rows and 25 columns").

[26] K. Luoma and A. Kumar, "SNAILS: Schema Naming Assessments for Improved LLM-Based SQL Inference," Proceedings of the ACM on Management of Data (SIGMOD), 2025. Open-access copy: UCSD ADA Lab technical report (PDF).

First, it flattens the data into wide, denormalized tables, because AI reasons far more reliably over data when it does not have to perform complex joins, and it pre-computes the hard aggregation logic so the AI is not left to invent it.

Second, it materializes those tables so answers come back in seconds rather than after a query crawls through layers of views.

Third, it adds plain-language context files that teach the AI about the business. They spell out what each metric means, where the same word means different things to different departments, which questions it can answer reliably and which tend to mislead, and what the company's top priorities are this year.

Fourth, connecting the AI to all of this is a piece of software we call the Platinum Connector, a custom MCP server we build bespoke for each client, which mediates between the AI and the data, enforces that the AI reads the business context before it answers, and adds capabilities that off-the-shelf connectors lack.

None of this infrastructure replaces your governed reporting. It exists to deliver what the AI shortcut promises but cannot straight out of the box. **An executive or an analyst asks a question in plain English and gets an accurate answer back fast, because the integration, modeling, definitions, and context the AI needs have all been built underneath it.**

Here is what that looks like from the executive's chair. Take the $90 million IT services firm from earlier, the one whose controller rebuilt the board's profitability view by hand every quarter. A few months after the foundation is in place, the CFO opens a chat window and asks, in plain English, for trailing-twelve-month margin by service line. The answer comes back in seconds, already reconciled to the same numbers the board sees, because the cost-allocation logic that used to live in one analyst's head is now encoded in the layer the AI is required to read. Margin by customer, which could not be calculated at all before, is now a question anyone with access can ask and trust. The controller is no longer assembling the quarterly view from pivot tables; it is live, and her time goes to interpreting the numbers instead of stitching them together. The weekly operating review stops waiting on whoever owns the spreadsheet, because the figures are already there and they hold up to scrutiny. That is the payoff, and it only exists because the work underneath it got done.

This is why we are emphatic that AI-driven analytics and traditional BI are complements, and that you need both. Power BI off the gold layer is fixed, fast, inexpensive to use, governed, and consumed by hundreds of people without anyone having to craft a prompt. It answers the questions you already know you need to answer, the same way every week. AI off the Platinum Layer is dynamic and exploratory. It answers the long tail of questions you did not know to build a dashboard for, and it puts that capability directly in the hands of the people who know the business best. A CEO or CFO who is given access can interrogate the data without waiting on anyone, and answer a one-off question or test a hunch in the moment.

We have a shorthand for this internally. The dashboard is your operating system, the AI is your analyst. A human analyst is invaluable for exploration and fast turnarounds, and

you still validate the analyst's number before you put it in front of your board. The same discipline applies here. Use AI for exploration and speed. Validate against your governed reporting before you defend a number publicly. Done this way, AI makes research and analysis dramatically faster, because the analyst can take the query the AI wrote, confirm it, and answer leadership in minutes instead of days. This is the answer to the belief in all its forms, that you can do it yourself, that you no longer need BI, or that one tool will cover everything. The foundational work is what makes any of it trustworthy, and that work does not disappear.

A buyer's checklist

If a vendor or an internal champion tells you that you can do your business intelligence with AI and skip the data foundation, ask these six questions. The answers will tell you quickly whether you are looking at a real solution or a demo.

1. Can it answer a question that spans my ERP, my CRM, and a spreadsheet at the same time, using my actual customer mapping? If the data is not integrated first, the honest answer is no.
2. Where do my business definitions live, and who maintains them? If the answer is "the AI figures it out," it does not, and it will pick a definition of its own, silently.
3. If I ask the same question three times, do I get the same answer in the same form three times? For anything that goes in front of your board, you need that to be a definite yes.
4. When it is wrong, how would I know? If the tool cannot tell you when it is unsure, it will be confidently wrong without warning.
5. Who owns and controls the underlying data, and what does it cost to leave? If the answer concentrates everything inside one application vendor's cloud, you are buying lock-in along with the analytics.
6. What does this cost at full scale, when a hundred people use it every day, at next year's token prices rather than today's subsidized ones?

A real solution has good answers to all six, because someone did the foundational work. The shortcut has good answers to none of them, because its entire premise is that the foundational work is no longer necessary. It still is. What has changed is the cost of skipping it. A governed dashboard is not immune to error; a bad data load or a mis-edited measure can throw it off. But the architecture, governance, and QA behind a well-built report keep it right most of the time and make its mistakes findable. An AI is harder to police, because its answers vary and each one reads as fluent and sure. The same foundational work is what moves it from confidently wrong to dependably right. Skip it, and the AI hands you a wrong number in clean, confident prose, and no one in the room knows.

What getting there looks like

If this argues for anything, it is doing the foundational work deliberately, either with an internal team or with a partner, rather than hoping a tool will paper over its absence. Here is what that involves in practice, so you know what you would actually be taking on.

Blue Margin delivers this as an ongoing managed service, not a one-time build. A Blue Margin team of data engineers, analytics consultants, and a customer success lead works with you in two phases. The first is the foundation: integrating your source systems, building the modeled gold layer, and standing up the governed reporting your team needs to run the business. At the end of this phase (shorter than you might expect), you have dashboards that hold up to scrutiny and a data layer you can defend in a board meeting. The second phase adds the AI layer on top, which is faster to build because the foundation is already there. Most of the work at this stage is writing the context files, encoding your definitions and business rules, and connecting the AI to the data in a way that enforces those rules before it answers. The result is that both your BI and your AI run from the same trusted source.

Everything lives in your environment and your accounts, in open formats you own, so nothing about this creates the lock-in this paper warns against. The team stays on to maintain and extend it as your business and your definitions change, the same way reporting gets maintained today. When your source systems change, when a definition shifts, or when leadership starts asking a new class of questions, the foundation evolves with you rather than drifting out of date.

You do not need a data team of your own to start, and you are not signing up to build and manage one. What you need is someone on your side who can speak for how the business actually works, what the metrics mean and where the hard parts of the data are, so the foundation reflects your reality rather than a generic template.

The fastest way to find out whether this fits is a short conversation about your systems, your reporting today, and the questions leadership keeps asking that no one can answer quickly. That will tell both of us whether there is a foundation worth building.

Notes and sources

Current as of May 2026. Vendor product claims are quoted from the vendors' own documentation and marketing; independent evidence is drawn from peer-reviewed benchmarks and named practitioners. Each source title links to the original.

Source Archive Appendix

Dated Wayback Machine snapshots for every source cited in Confident, Fast, and Wrong: What "Just Use AI" Leaves Out of Your BI

Vendor marketing and news pages change without notice. These are permanent, timestamped snapshots captured via the Internet Archive on May 31, 2026, cross-referenced to the white paper's footnote numbers. If an original page is altered or removed, cite the archived copy. Two pages block the Wayback crawler and are listed at the end for one-click manual capture.

---

[1] AI/BI Genie product page: 'Talk with your data.' - Original: https://www.databricks.com/product/business-intelligence/genie - Archived 2026-05-03: http://web.archive.org/web/20260503192921/https://www.databricks.com/product/business-intelligence/genie

[2] ThoughtSpot Launches Spotter, the Autonomous Agent for Analytics - Original: https://www.globenewswire.com/news-release/2024/11/13/2980332/0/en/ThoughtSpot-Launches-Spotter-the-Autonomous-Agent-for-Analytics.html - Archived 2025-05-29: http://web.archive.org/web/20250529001110/https://www.globenewswire.com/news-release/2024/11/13/2980332/0/en/ThoughtSpot-Launches-Spotter-the-Autonomous-Agent-for-Analytics.html

[3] Getting Started with Cortex Analyst: Augment BI with AI - Original: https://www.snowflake.com/en/developers/guides/getting-started-with-cortex-analyst/ - Archived 2026-05-31: http://web.archive.org/web/20260531182943/https://www.snowflake.com/en/developers/guides/getting-started-with-cortex-analyst/

[4] Fabric data agent creation - Original: https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent - Archived 2026-05-06: http://web.archive.org/web/20260506180017/https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent

[5] Breeze AI: 'Ask questions and get answers grounded in your CRM data and role.' - Original: https://www.hubspot.com/products/artificial-intelligence - Archived 2026-05-31: http://web.archive.org/web/20260531183229/https://www.hubspot.com/products/artificial-intelligence

[6] Salesforce Data Cloud Zero Copy: When (and When Not) to Use It - Original: https://www.salesforceben.com/salesforce-data-cloud-zero-copy-when-and-when-not-to-use-it/ - Archived 2026-02-15: http://web.archive.org/web/20260215205910/https://www.salesforceben.com/salesforce-data-cloud-zero-copy-when-and-when-not-to-use-it/

[7] SAP and Salesforce Data Integration for Supplier Analytics on Databricks - Original: https://www.databricks.com/blog/sap-and-salesforce-data-integration-supplier-analytics-databricks - Archived 2026-05-31: http://web.archive.org/web/20260531224742/https://www.databricks.com/blog/sap-and-salesforce-data-integration-supplier-analytics-databricks

[8] Salesforce Agentforce and Data Cloud: A path to the software-only hyperscaler - Original: https://siliconangle.com/2025/05/12/salesforce-agentforce-data-cloud-path-software-hyperscaler/ - **Not yet archived**, capture at: https://web.archive.org/save/https://siliconangle.com/2025/05/12/salesforce-agentforce-data-cloud-path-software-hyperscaler/

[9] Your LLM issues are really data issues - Original: https://stackoverflow.blog/2026/04/28/your-llm-issues-are-really-data-issues/ - Archived 2026-05-31: http://web.archive.org/web/20260531190453/https://stackoverflow.blog/2026/04/28/your-llm-issues-are-really-data-issues/

[10] Why LLMs struggle with analytics (and how we fixed that) - Original: https://www.tinybird.co/blog/why-llms-struggle-with-analytics-and-how-we-fixed-that - Archived 2026-05-31: http://web.archive.org/web/20260531190534/https://www.tinybird.co/blog/why-llms-struggle-with-analytics-and-how-we-fixed-that

[11] Same Prompt, Different Outcomes: Evaluating the Reproducibility of Data Analysis by LLMs - Original: https://arxiv.org/html/2602.14349 - Archived 2026-05-31: http://web.archive.org/web/20260531190556/https://arxiv.org/html/2602.14349

[12] Why text-to-SQL fails - Original: https://omni.co/blog/why-text-to-sql-fails - Archived 2026-05-31: http://web.archive.org/web/20260531192343/https://omni.co/blog/why-text-to-sql-fails

[13] Spider 2.0: Evaluating Language Models on Real-World Enterprise Text-to-SQL Workflows - Original: https://arxiv.org/abs/2411.07763 - Archived 2026-05-23: http://web.archive.org/web/20260523225001/https://arxiv.org/abs/2411.07763 - Original: https://spider2-sql.github.io/ - Archived 2026-05-31: http://web.archive.org/web/20260531204246/https://spider2-sql.github.io/

[14] Can LLM Already Serve as a Database Interface? A BIg Bench for Large-Scale Database Grounded Text-to-SQLs - Original: https://arxiv.org/abs/2305.03111 - Archived 2026-05-24: http://web.archive.org/web/20260524075129/https://arxiv.org/abs/2305.03111 - Original: https://bird-bench.github.io/ - Archived 2026-05-31: http://web.archive.org/web/20260531204421/https://bird-bench.github.io/

[15] Why 90% Accuracy in Text-to-SQL is 100% Useless - Original: https://towardsdatascience.com/why-90-accuracy-in-text-to-sql-is-100-useless/ - Archived 2026-02-09:

http://web.archive.org/web/20260209232138/https://towardsdatascience.com/why-90-accuracy-in-text-to-sql-is-100-useless/

[16] AWS Waives Egress Fees for Customers Exiting the Cloud - Original: https://www.infoq.com/news/2024/03/aws-egress-fees/ - Archived 2026-05-31: http://web.archive.org/web/20260531210445/https://www.infoq.com/news/2024/03/aws-egress-fees/

[17] Open Table Formats: From Vendor Lock-In to Data Sovereignty - Original: https://buckenhofer.com/2026/04/open-table-formats-parquet-delta-lake-iceberg/ - **Not yet archived**, capture at: https://web.archive.org/save/https://buckenhofer.com/2026/04/open-table-formats-parquet-delta-lake-iceberg/

[18] OpenAI plans to report stunning annual losses through 2028 - Original: https://fortune.com/2025/11/12/openai-cash-burn-rate-annual-losses-2028-profitable-2030-financial-documents/ - Archived 2026-05-31: http://web.archive.org/web/20260531214456/https://fortune.com/2025/11/12/openai-cash-burn-rate-annual-losses-2028-profitable-2030-financial-documents/

[19] Uber burned through its entire 2026 AI budget in four months - Original: https://fortune.com/2026/05/26/uber-coo-ai-spending-tokens-claude-code/ - **Not yet archived**, capture at: https://web.archive.org/save/https://fortune.com/2026/05/26/uber-coo-ai-spending-tokens-claude-code/

[20] Getting Started with Cortex Analyst - Original: https://www.snowflake.com/en/developers/guides/getting-started-with-cortex-analyst/ - Archived 2026-05-31: http://web.archive.org/web/20260531182943/https://www.snowflake.com/en/developers/guides/getting-started-with-cortex-analyst/

[21] AI/BI Genie is now Generally Available - Original: https://www.databricks.com/blog/aibi-genie-now-generally-available - Archived 2026-05-31: http://web.archive.org/web/20260531220629/https://www.databricks.com/blog/aibi-genie-now-generally-available - Original: https://docs.databricks.com/aws/en/genie/ - Archived 2026-05-31: http://web.archive.org/web/20260531222042/https://docs.databricks.com/aws/en/genie/

[22] Introducing Spotter: Your AI Analyst - Original: https://www.thoughtspot.com/blog/introducing-spotter-ai-analyst - Archived 2026-05-31: http://web.archive.org/web/20260531222109/https://www.thoughtspot.com/blog/introducing-spotter-ai-analyst

[23] Fabric data agent creation - Original: https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent - Archived 2026-05-06: http://web.archive.org/web/20260506180017/https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent

[24] Overview of Copilot in Fabric - Original: https://learn.microsoft.com/en-us/fabric/fundamentals/copilot-fabric-overview - Archived 2026-05-31: http://web.archive.org/web/20260531222829/https://learn.microsoft.com/en-us/fabric/fundamentals/copilot-fabric-overview

[25] Fabric data agent creation - Original: https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent - Archived 2026-05-06: http://web.archive.org/web/20260506180017/https://learn.microsoft.com/en-us/fabric/data-science/concept-data-agent

[26] SNAILS: Schema Naming Assessments for Improved LLM-Based SQL Inference - Original: https://dl.acm.org/doi/10.1145/3709727 - Archived 2025-02-16: http://web.archive.org/web/20250216214540/https://dl.acm.org/doi/10.1145/3709727 - Original: https://adalabucsd.github.io/papers/TR_2025_SNAILS.pdf - Archived 2025-06-05: http://web.archive.org/web/20250605015230/https://adalabucsd.github.io/papers/TR_2025_SNAILS.pdf

---

Pending manual capture

These pages refuse the automated crawler. To capture each, paste the save link into a browser once (it takes ~20 seconds), then record the resulting `web.archive.org/web/...` URL here:

- Footnote [8]: https://siliconangle.com/2025/05/12/salesforce-agentforce-data-cloud-path-software-hyperscaler/
  - Save link: https://web.archive.org/save/https://siliconangle.com/2025/05/12/salesforce-agentforce-data-cloud-path-software-hyperscaler/
- Footnote [17]: https://buckenhofer.com/2026/04/open-table-formats-parquet-delta-lake-iceberg/
  - Save link: https://web.archive.org/save/https://buckenhofer.com/2026/04/open-table-formats-parquet-delta-lake-iceberg/
- Footnote [19]: https://fortune.com/2026/05/26/uber-coo-ai-spending-tokens-claude-code/
  - Save link: https://web.archive.org/save/https://fortune.com/2026/05/26/uber-coo-ai-spending-tokens-claude-code/

blue margin

Turn Your Data Into a BI and AI-Ready Growth Engine

BLUEMARGIN.COM