

The data leader's guide to → AI-ready data



Contents

AI is leading the conversation, be it in the boardroom or in the break room. And the one thing that can make AI go from a buzzword to real business value is good data. High-quality, well-governed and accessible data can fuel accurate, trustworthy and scalable AI. And as conversations focus on data for AI, organizations lean on data leaders to drive a strategic shift. In the age of generative AI (gen AI), building a strong data foundation and accelerating innovation must go hand-in-hand. But managing the data and making it ready for AI has become more complex than ever. That's why you need a two-pronged approach—data for AI and AI for data.

Let's explore how this approach can help you overcome today's data challenges and outline actionable steps to ready your data for AI—while using AI to make data management smarter.

01 →

AI has a data problem. Data needs an AI solution.

02 →

Turn data sprawl into data readiness

03 →

The AI action plan for data leaders

04 →

Get your data AI ready with IBM

AI has a data problem. Data needs an AI solution.



If data volume guaranteed scalable gen AI, businesses would be enjoying peak operational efficiency. But explosive data growth has left organizations with more data than they can effectively manage. Unmanaged sprawl is more than an inconvenience—it's a risk, leading to repeated errors that can cost you time and effort, if not more. A governed data foundation, on the other hand, reduces incidents, shortens recovery time and helps you accelerate delivery.

Creating a plan to organize data and another to make it AI ready isn't the way to go. What you need is a unified strategy. According to a recent study by IBM Institute for Business Value, 72% of surveyed leaders view proprietary data as the key to unlocking the true value of gen AI, and 68% say an integrated enterprise-wide data architecture is critical for collaboration and innovation.¹ Combining

data and AI in your strategy prevents duplicated effort, reduces tool sprawl and builds trust in AI outputs. When your data is AI ready, it's also ready for analytics and insights downstream.

For an AI-ready business, unstructured data is a treasure trove, yet only 26% of surveyed businesses say they can turn unstructured data into business value.² Gen AI thrives on documents, web pages, social posts and surveys, especially when your architecture supports retrieval augmented generation (RAG) and source attribution for traceable, auditable outputs, making unstructured data an opportunity despite the complexity. Data quality also relies on a solid governance strategy. Your approach to governance can bring structure, visibility and accountability to your data and data products. Bear in mind, governance must be always on, not a one-time check.

In many organizations, engineers spend roughly 50% of their time searching for information in an ocean of data, much of it unstructured and not easily retrievable. This not only diminishes productivity but also inflates costs, funding more data retrieval rather than problem-solving.³

Brian Mullaney
Principal and Chief Revenue Officer
CrushBank

The challenge every data leader faces is to unify, govern and scale across silos, platforms and clouds, while overcoming operational hurdles such as workforce readiness. The scale of this challenge is itself a great use case for introducing AI in your data management practices.

No matter where you are in your journey, there are sustainable strategies and solutions that can help you build a successful data foundation for your present and future AI strategy.



Gen AI has amplified
the challenges data
leaders face today

79%

of CDOs say they're still early in the process
of defining how to scale and govern them.²

26%

of CDOs are confident their data capabilities
can support new AI-enabled revenue streams.²

Turn data sprawl into data readiness

In the early 1960s, IBM programmer George Fuechsel brought an important phrase into the public lexicon—*garbage in, garbage out*. What it means is, low-quality data can't yield high-quality results, even with the help of powerful technology. If garbage data dominates your data landscape, AI adoption will only accelerate and emphasize garbage outcomes.

AI that's ready for business is powered by high-quality data that can support decision-making, analysis and performance across the organization.

That's why it's crucial for data leaders to fully invest in data quality.

So, how do you keep the garbage out and harness the power of high-quality data? Here are five questions data leaders must answer to lay the groundwork for a trusted, protected data foundation.

01



Can you effectively discover, access and use data wherever it resides to support AI?

To effectively support your AI workloads, your data needs to be accessible, reliable and ready to consume. However, most organizations struggle to manage the vast amounts of structured and unstructured data stored in different formats, siloed across multiple clouds and on-premises storage.

Data sources are often varied and they may address specific—but unrelated—use cases. Existing systems can present a particular challenge if they lack the APIs to connect with AI environments.

Many of the elements of traditional data management architecture are relevant and important to harness the potential of AI. It is equally important to understand the data

that exists and see how it can be used. In a survey, 81% of responding leaders say they bring AI to data rather than centralizing data for AI.² You must also evaluate what access and governance policies are in place and what needs to be added to support your use case.

You need to connect the right data and build pipelines for future AI requirements. This means supporting multiple integration types to accommodate structured and unstructured data. Beyond pipelines, you must also have semantic layers to provide governed metrics, reduce rework and misinterpretation, and add retrieval layers to ground AI outputs in authoritative enterprise content. A recent IDC report says over 80% enterprises have already adjusted their data strategies in response to AI.⁵ Effective integration can improve accuracy and usability and help you make the most of unstructured data.

02

Do you trust the quality and security of the data you are using for AI?

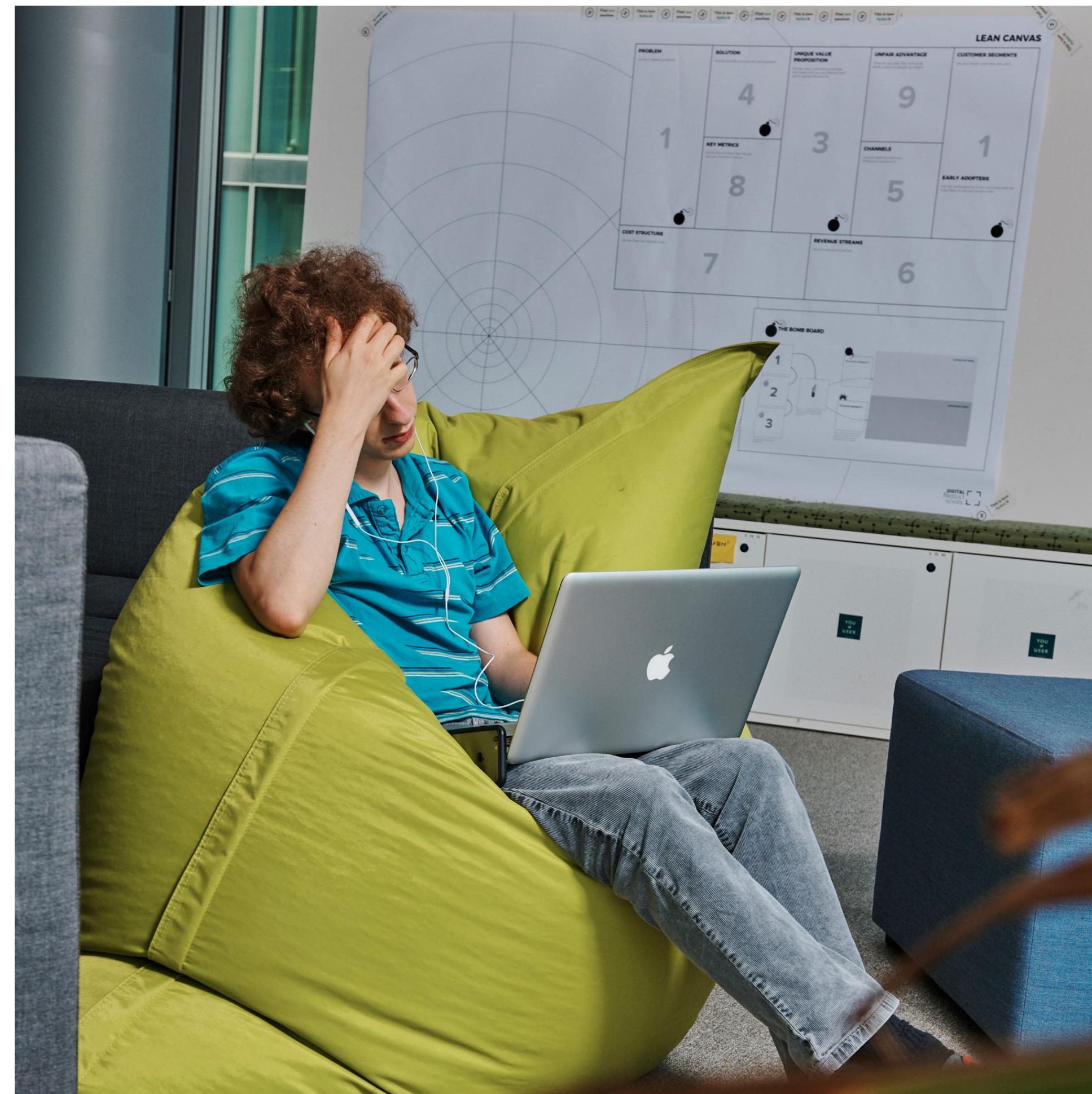
One of the biggest barriers to AI adoption is the absence of data you can trust. If you can't find your data, trace its lineage, control access to it and keep it secure, you don't have a base for reliable models. Trust also depends on operational quality such as accuracy, consistency and readiness. According to IBM Cost of a Data Breach Report 2025, 63% of the breached organizations lacked AI governance policies and 97% of the AI-related breaches involved AI systems that lacked proper access controls.⁶

Data quality depends on a solid foundation that includes understanding lineage, implementing governance and enabling observability to identify issues before

Turn data sprawl into data readiness

AI outputs are impacted, but it does not end there. You must adopt an always-on approach by setting clear standards to measure data freshness, routinely classify and label sensitive data, apply proper access rules, and maintain an audit trail. By taking these measures, you can turn quality from a promise into a practice.

You must also protect your data from the risks introduced by AI, especially gen AI. Of the data breaches covered in IBM Cost of a Data Breach Report 2025, 16% involved attackers using AI, often in phishing or deepfake impersonation.⁶ Setting up gen AI controls such as answers grounded in sources that are cited and keeping humans in the loop can serve as guardrails without slowing down everyday work.





03

Is your data fabric ready for the workloads of today and tomorrow?

Your data fabric is ready when teams can reach and use quality data across all systems without breaking rules, duplicating data or rewriting pipelines every quarter. You must look at data integration as more than an IT detail. It helps you avoid tool sprawl, duplicated data and uneven rules across departments.

Even organizations with mature data infrastructure struggle when integration is manual, brittle or siloed. As AI workloads become more dynamic, your data has to be accessible in real time.

Your data has a shelf life and what's relevant today need not be tomorrow, especially in the era of gen AI. It is important to gain access to critical data when you need it, where you need it. The most efficient way to move data is through a single logical integration layer that adapts to context—whether extract, transform, load (ETL), streaming or replication—eliminating constant pipeline rewrites. A modern data fabric connects producers and consumers seamlessly across environments. It can automate routing using metadata-driven integration, apply consistent governance and offer policy-aware data connectivity wherever data resides. This can free up your team from manual busywork to focus on strategic initiatives.

04

Have you considered the people, processes and technology needed to ensure your data is AI ready?

Your data strategy and architecture should reflect new considerations—around people, processes and technology—for AI. The pace at which organizations are adopting AI is creating a skill gap that is proving difficult to fill. In a recent study, 82% of surveyed CDOs report they're hiring for data roles related to gen AI that didn't exist last year.² One effective way to address the gap is to rethink your hiring strategy and look inwards. Internal recruitment and upskilling programs can help you find talent who are already well-versed in the domain and can bring a fresh perspective to the challenges.

Further, the right tools and processes can help you avoid tool sprawl and increase productivity with little to no extra resources. Modern platforms embed intelligent agents to automate repetitive tasks while keeping humans in control for critical approvals. Strengthening your data operations is critical to AI readiness. It includes identifying which data to access, evaluating cost of ownership and selecting the right migration approach.

Identify the right people and equip them with the right tools and training to successfully adopt and scale AI, optimize data value and meet business objectives.

05

Can your data storage infrastructure meet the intense performance and capacity needs of AI workloads?

Large language model-driven AI workloads often benefit from a cloud-first approach, particularly for training and rapid experimentation, while on-premises storage remains essential for regulated and latency-sensitive use cases. Hybrid multicloud adoption continues to increase but hybrid storage alone doesn't make data AI ready.

Operationalizing and monetizing AI requires a foundation that can unify access across disparate storage environments. Beyond storage, add a semantic layer for consistent business definitions and a retrieval tier for unstructured content so data is not only stored but is also governed and made usable.

Successful AI initiatives require data that is organized, enriched and governed across environments.



Now that you've assessed your current state, let's explore steps to get your data AI ready.

The AI action plan for data leaders

As a data leader, it's up to you to position data as the fuel for better AI outcomes. Building a strong culture of data democratization, where every employee understands data as a strategic asset, should be a guiding principle of data strategy.

Introduce agentic AI into your data practices to improve productivity

When it comes to ensuring data readiness, agentic data management can yield significant results. By adopting agentic AI capabilities within your data practices, you can improve productivity and empower your team to focus on strategic decisions. Gartner predicts that by 2027, the use of AI assistants and AI-enhanced workflows within data integration tools will reduce manual effort by 60%.⁷ Agentic AI can reduce the burden of the mundane by augmenting data operations such as alerts, triaging pipeline breaks, proposing routine fixes and running playbooks. It can also help maintain governance standards and reduce sprawl—all while keeping a human in the loop.

But before you can fully reap the benefits of agentic AI for data management, you need to make sure you're implementing it in a data estate that's already optimized.

CDOs from organizations that deliver higher ROI on both AI and data investments are 20% more likely to say they understand the increased data access requirements of agentic AI. They are also likely to say they are developing policies for secure and controlled data sharing within their organization.²

Ensure you can store, discover and access the right data

To ensure data readiness for AI, your underlying data stores need to be highly reliable, protected and sufficiently flexible to respond to new deployment options. With databases, consider a fit-for-purpose approach. This method is about using the right database for the right workload, data types and price-performance. The right data strategy lets you build on your organization's existing data architecture and optimize it to fit your AI strategy.

Some types of databases well-suited to manage structured data aren't designed to store, access and analyze unstructured data. A lakehouse is a strong

foundation, but a lakehouse-only approach is no longer sufficient in a gen AI-driven landscape. Give it a boost with a robust data fabric and integration framework—along with capabilities such as semantic and retrieval layers—to help ensure that unstructured data doesn't go undiscovered or underutilized.

Unstructured data can introduce new file formats into your data landscape. Your organization needs a solution that can accommodate these new formats. Otherwise, you run the risk of your teams using AI models and tools outside your governance parameters, without approvals or guardrails, which gives rise to shadow AI.





Build trust in your data with effective governance

Training your AI models with low-quality data is like filling your brain with misinformation. It's a recipe for ill-informed decision-making, cognitive bias and risky behavior.

Advanced AI models touch every workflow, so governance must be built in and always on. The last thing you want to do is amplify substandard results. An always-on approach ensures sensitive fields are autoclassified, policies are

checked at ingestion and at each retrieval, and lineage is tracked across domains. It also helps ensure audit logs are protected and are tamper-resistant, access is role-based and consent and retention rules are enforced. Set up a strong governance framework that gives you control and visibility over data product ownership, operations, controls, policies and standards. So, whether you choose to run it on autopilot or take full control, you can manage your data efficiently and effectively.

Don't let data security and compliance cost you

There's a steep price to pay for insecure data. The average cost of a data breach was USD 4.44 million according to the IBM Cost of a Data Breach Report 2025. Further, shadow AI added USD 670,000 to the average among companies with high levels of shadow AI.⁶ It's important to protect your data, no matter where it resides. Monitoring it across the pipeline—whether it's at rest or in motion or in use—presents its own challenges. Invest in the right security solutions to ensure your data is protected no matter where it resides.



Data change management

- Provide control and oversight of the changes made to data-related processes.

Data responsibility and accountability

- Align stewardship roles to responsible individuals or teams and clearly define the duties required for data ownership.

Data quality and consistency

- Ensure data is accurate and complete, and that processes are uniform.

Data integration

- Determine how data is pulled in from various sources, weeding out duplicate and low-quality data and transforming it into a unified format.

Data standards

- Help your organization establish the minimum accepted data quality levels and the criteria for data access.

Chart the course for the future

Every data leader wants their data landscape to be highly organized and accessible like a library, where books are easily located as they're carefully cataloged and neatly displayed. But realistically, a lot of data environments probably look like a secondhand bookstore, a disorganized space overcrowded with books randomly placed all over the store. And such disorderliness needs to be managed. As you become more adept at managing all your data assets and embedding agentic data management capabilities, your landscape will begin to look increasingly like a well-managed library.

Time to build your data library now. Start with a unified data strategy, bringing data for AI (governed foundations) and AI for data (agentic workflows) together. Support it with agentic AI, keeping humans in the loop. Design a data architecture that is kitted for any curveballs gen AI may throw at it. This will set you up for the demands of today and the decade to come. Harness your data to meet constantly evolving business strategies and technological advancements with enhanced clarity.

The data leader's checklist for data readiness

- Discoverability**
Ensure teams can effectively discover, access and use data wherever it resides.
- Quality basics**
Make ensuring data quality an always-on practice, with lineage, governance and observability measures.
- Guardrails**
Embed governance and gen AI controls from the get-go.

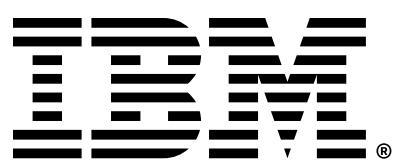
- Integration**
Adopt a modern data fabric with metadata-driven integration.
- Unstructured retrieval**
Operationalize unstructured data with a retrieval tier (RAG) and source logging.
- Ownership**
Define data stewardship and product ownership.

Get your data AI ready with IBM

Discover the data and AI solutions empowering modern data leaders. No matter where you are in your journey, IBM has the technology, expertise and experience to help you overcome your data challenges and achieve AI readiness.

[Explore IBM data and AI solutions →](#)





1. [5 mindshifts to supercharge business growth](#), IBM Institute for Business Value, 5 May 2025
2. [The 2025 CDO Study: The AI multiplier effect](#), IBM Institute for Business Value, 12 November 2025
3. [Fast, accurate answers to business-critical questions](#), IBM
4. [Real-Time Data Integration for Business in Real Time](#), IDC, June 2025
5. [Worldwide Data Integration Software Platforms 2025 Vendor Assessment](#), IDC Marketscape, October 2025
6. [Cost of a Data Breach Report 2025](#), IBM, July 2025
7. [Magic Quadrant for Data Integration Tools](#), Gartner, 8 December 2025

© Copyright IBM Corporation 2026

IBM and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation, in the United States and/or other countries. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on ibm.com/legal/copytrade.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

No IT system or product should be considered completely secure, and no single product, service or security measure can be completely effective in preventing improper use or access. IBM does not warrant that any systems, products or services are immune from, or will make your enterprise immune from, the malicious or illegal conduct of any party.

The client is responsible for ensuring compliance with all applicable laws and regulations. IBM does not provide legal advice nor represent or warrant that its services or products will ensure that the client is compliant with any law or regulation.