

Point of View

Why Good Data Is the Secret Ingredient for Scalable AI

A Look at Informatica's IDMC CDI

In the past, data integration tools (like ETL platforms) were built for structured data—think spreadsheets and databases. But today, data comes in all shapes and sizes: XML and JSON provide semi-structured data, and emails, images, and social media posts are categorized as unstructured data. As data processing needs evolve, organizations must quickly, securely, and accurately handle massive volumes of complex, unstructured data.

The problem? Data is often siloed across systems, is of low quality, or non-compliant, resulting in companies struggling to scale AI as data isn't ready. According to Informatica's CDO Insights 2025 survey, 67% of organizations could not operationalize even half of their GenAI pilots due to issues with data readiness and the maturity required to support AI at scale.¹ So, the message is loud and clear: AI success depends heavily on a strong data foundation. Even the most promising AI models can't deliver real-world impact without a solid foundation.

The solution? Informatica, a market leader in data processing and data integration, is focused on aligning itself with the fast-paced changes in data demands. Informatica is evolving from a traditional ETL platform to a cloud-native, AI-powered data integration platform. In today's fast-paced, data-driven world, AI is no longer a futuristic idea—it's a strategic imperative. AI requires high-quality, trusted, and mature data to generate true insights and meaningful outcomes.

Informatica's Intelligent Data Management Cloud (IDMC) is changing how enterprises manage data. Its advanced Cloud Data Integration (CDI) features help organizations build a secure, scalable, and reliable data foundation for AI, using different data types and smart automation across data lakes, warehouses, and marts.

The AI Challenge: Data Complexity and Fragmentation

Effective and tangible outcomes of AI models depend not only on massive datasets but also on high-quality, diverse, and secure data. Many enterprises find themselves stuck in a series of challenges that create roadblocks to achieving their AI ambitions. Some of the key challenges are:



Siloed systems: Many organizations suffer from fragmented data systems, leading to incorrect analysis and poor decision-making.



Poor data quality: According to Gartner's estimate, organizations are paying a heavy price (~\$12.9 million annually) due to poor data quality.²



Measurement difficulties: 30% of CDAOs say their top challenge is the inability to measure the impact of data, analytics, and AI on business outcomes.³



Limited scalability: Only 22% of organizations could define and track the business impact metrics for their data and analytics use cases.³



Compliance risks: 35% of data leaders realize governance as the most critical success criteria for data-driven transformation.⁴

Challenges in Data Management

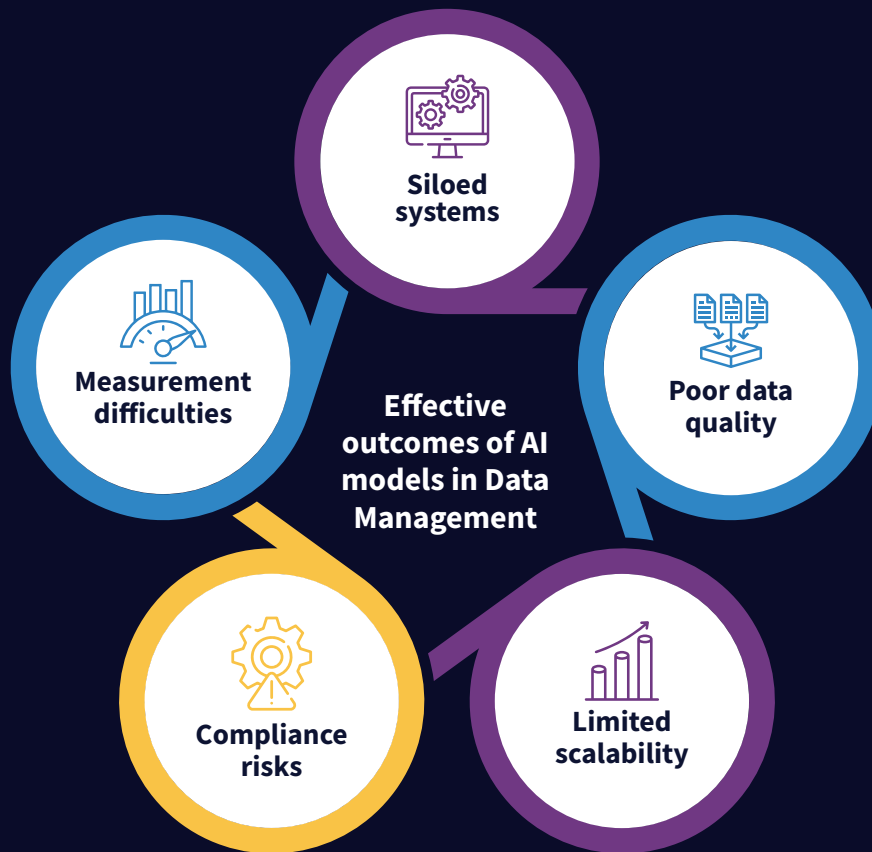


Figure 1: Challenges in data management

The above challenges are not only technical hiccups but also critical roadblocks to bringing data insights and informed decision-making.

IDMC: A Unified Platform for AI-Ready Data

Informatica IDMC is a cloud-native, AI-powered, advanced data management platform that combines data engineering, governance, privacy, and analytics into a unified experience. With holistic data management at its core, IDMC empowers organizations to fuel their AI initiatives with high-quality, compliant, and ready-to-action data. The result? Smarter decisions, faster innovation, and measurable outcomes.

Enabling AI with the key capabilities of IDMC CDI



Cloud-native scalability

IDMC CDI is a cloud-native platform that scales elastically across multi-cloud environments. This is crucial for AI workloads that require dynamic compute resources for training and inference.



Low-code/no-code data integration

With intuitive interfaces and automation, CDI empowers data engineers and citizen integrators to build complex data pipelines without deep coding expertise. This accelerates AI project timelines and democratizes access to data.



Real-time data processing

AI models need fresh data to remain relevant. CDI supports real-time ingestion and transformation, ensuring that AI systems operate on the most current information available.



Data quality and governance

IDMC CDI seamlessly integrates with Informatica's data quality and governance tools, ensuring that AI models are trained on trusted, compliant data. This is important in regulated industries like healthcare and finance.



Support for modern architectures

IDMC supports modern data architectures, such as data mesh and data fabric. This allows decentralized teams to manage data autonomously while maintaining global governance. This flexibility is key to scaling AI across large enterprises.



AI-Powered metadata management

Informatica's CLAIRE engine uses AI to automate metadata discovery, lineage tracking, and impact analysis. This enhances transparency and trust in AI models by providing clear data provenance.

Real-World Applications of IDMC CDI with AI

IDMC CDI helps organizations manage their data more efficiently with features like data ingestion, transformation, and orchestration. This makes it easier to build AI solutions that address specific business needs. Let's look at some of the real-world applications.

Healthcare

Automating Patient Admissions and Clinical Documentation

Hospitals and medical centres utilize AI to process documents containing unstructured data, such as Electronic Health Records (EHRs), referrals, and medical forms. This information is then stored in vector databases for easy searching and retrieval. IDMC CDI can support this process in many ways, such as:

- **Document ingestion and chunking:** CDI splits large medical documents into manageable and smaller chunks for processing.
- **Vector embedding:** These chunks are transformed into semantic vectors using GenAI models. This process would enable the meaningful interpretation and efficient retrieval of unstructured medical data to improve the accuracy and relevance of search results.
- **Integration with Pinecone:** IDMC CDI pipelines push data into Pinecone vector databases for fast semantic search.
- **Querying and retrieval:** IDMC CAI (Cloud Application Integration) orchestrates queries and retrieves insights using OpenAI chat completion. The response time for querying and retrieval should ideally be ~2-5 seconds for well-optimized queries.

Takeda Pharmaceuticals moved 96% of data to the cloud using Informatica IDMC. As a result, the company now has the productivity improved by 40% with better access to its data, greater trust in its information, and faster progress on AI projects that improve patient outcomes.⁵

Insurance

Claims Processing with GenAI

GenAI is transforming claims workflows by automating the extraction, validation, and routing of data from multiple sources. Informatica IDMC CDI plays the following roles in insurance processes:

- **Data aggregation:** IDMC CDI integrates policy, customer relationship management (CRM), and claims data from disparate systems.
- **Orchestration:** IDMC CAI triggers large language models (LLMs) to summarize applications and assess risks.
- **Governed access:** Informatica Catalogue ensures secure and compliant data access.

Informatica's GenAI-powered IDMC platform serves many insurance customers globally. Below are the few examples of such use cases

- Claims automation for health insurance clients⁶
- Risk assessment for property and casualty insurance customers
- Policy and claim management for life insurance companies⁶

Retail

Personalized Customer Experience

Retail organizations are leveraging GenAI to deliver personalized experiences by integrating their transactional and consumer behavioural data with predictive analytics. Informatica IDMC CDI plays the following crucial roles:

- **Real-time data integration:** CDI connects customer touchpoints across POS, e-commerce, and CRM systems.
- **GenAI blueprint:** IDMC CDI's low-code/no-code service enables rapid deployment of GenAI apps.
- **Metadata intelligence:** IDMC CDI helps in enhancing personalization by enabling the semantic understanding of consumer behaviour.

Informatica's partnership with Databricks allows retailers to build scalable GenAI apps using IDMC and Unity Catalog, democratizing access to AI-powered personalization.⁷

The above use cases prove CDI's ability to support advanced AI architectures like retrieval-augmented generation (RAG), which blends structured and unstructured data to deliver mature and more contextualized outputs.

Here's the kicker: IDMC CDI's low-code/no-code platform allows users to seamlessly integrate with large language models (LLMs) and other AI services without getting into technical depth, enabling faster time-to-value and AI democratization in action.

Accelerating AI Readiness with Informatica IDMC

The multiple IDMC releases over the last couple of years introduced a series of enhancements focusing on simplified and accelerated enterprise AI adoption. These releases ensure Informatica's commitment to enabling organizations to become truly AI-ready by addressing the fundamental challenges of data engineering and management.⁸

Key advancements in IDMC include:

- **Expanded multi-cloud support:** Continuous focus on interoperability across major cloud platforms ensures seamless data movement, processing, and integration, enabling organizations to operate in hybrid and multi-cloud environments with greater flexibility and control.
- **Tools and interfaces for enhanced productivity:** With the new low-code/no-code platform, intelligent automation features, and continuous improvements to user interfaces, IDMC CDI enables data teams to design and manage complex data pipelines, significantly reducing the development life cycle.
- **Streamlined data preparation for GenAI:** Purpose-built features for preparing structured and unstructured data for generative AI use cases, including support for RAG pipelines, vector database integration, and LLM connectivity.⁶
- **AI-powered data management using CLAIRE GPT:** CLAIRE GPT enables natural language-driven, AI-powered data management within Informatica's Intelligent Data Cloud.



- **Streamlined data pipeline creation and documentation using CLAIRE Copilot:** CLAIRE Copilot streamlines data pipeline creation via natural language, offers intelligent transformation suggestions, automates documentation, processes unstructured data, and empowers non-technical users to build integrations through intuitive, prompt-based workflows. The Data Ingestion Agent automates the process of collecting data from various sources.
- **Form a cohesive framework within IDMC using IDMC AI agents:** The newly introduced data ingestion, ELT, and modernization agents would enable organizations to automate and scale their data integration and AI initiatives with minimal manual intervention.
 - **The data ingestion agent** automates the process of collecting data from diverse sources. It simplifies the setup of ingestion pipelines using natural language prompts and smart recommendations.
 - **The ELT agent** streamlines the transformation of raw data into structured formats directly within target systems. It uses AI to generate and optimize ELT pipelines, reducing manual coding and accelerating data preparation.
 - **The modernization agent** supports automated pipeline generation, metadata mapping, and integration with modern platforms. This agent is key to scaling AI initiatives by enabling seamless access to cloud-based, AI-ready data environments.

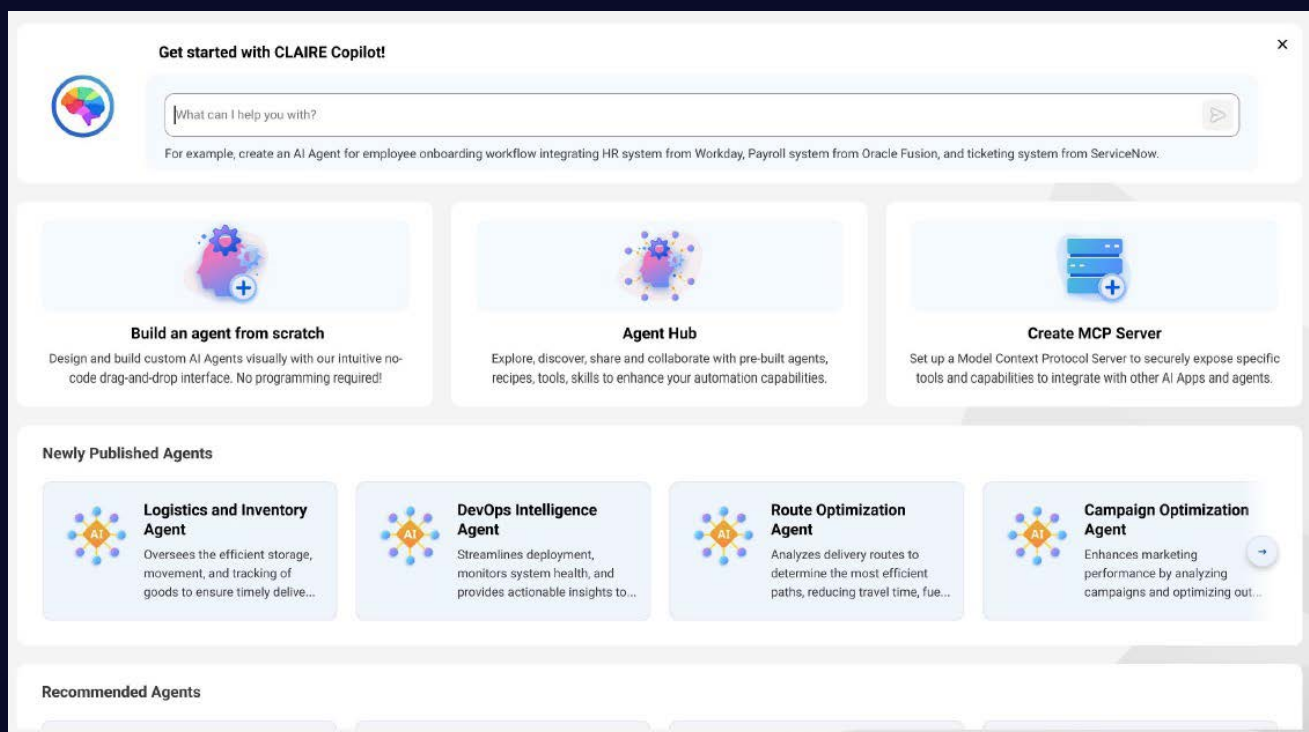


Figure 2: IDMC Claire Copilot and IDMC AI Agent Services

IDMC significantly reduces architectural complexity by consolidating critical data engineering and management functions such as data and application integration, governance, quality, and privacy into a unified platform. This simplification enhances operational efficiency and ensures that data is consistently available, trusted, compliant, and mature for deploying scalable and effective AI solutions.

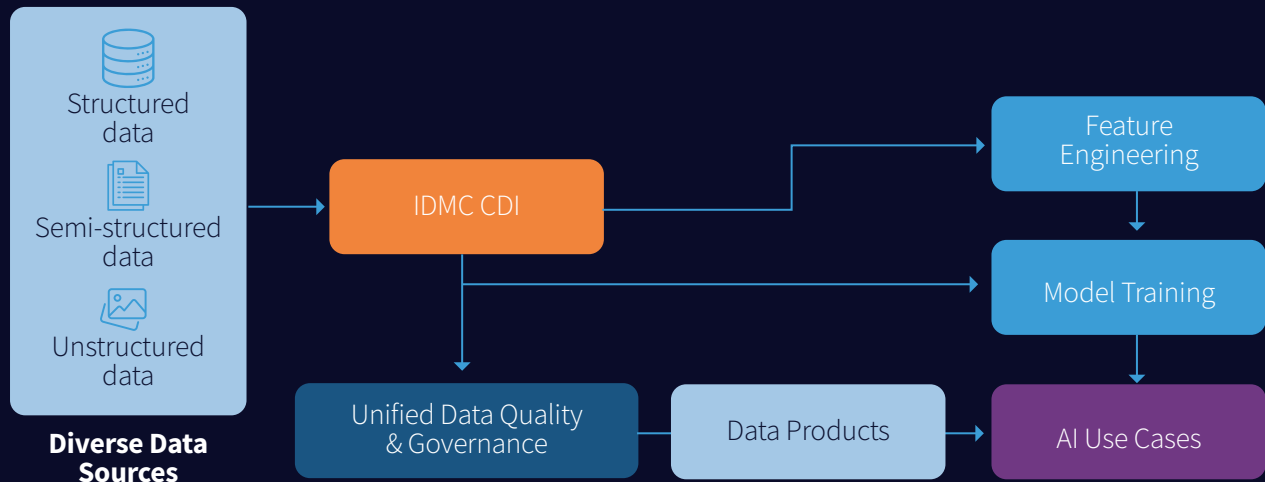
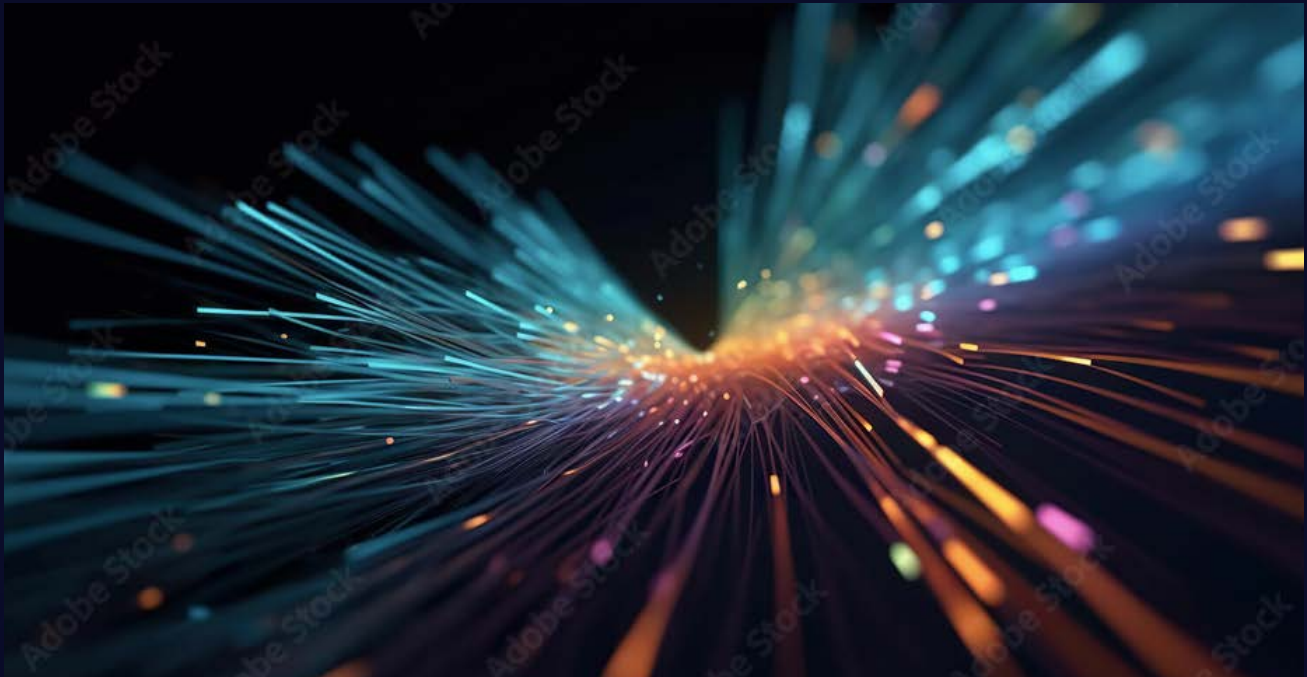


Figure 3: Process Flow Diagram for Data Engineering using CDI integrated with AI



Conclusion

Building a Robust and Scalable AI Foundation

Building scalable, effective, and efficient AI solutions not only relies on cutting-edge algorithms but also on a robust, intelligent data infrastructure. Without a high-quality, trusted, and easily explorable data foundation, even the most advanced AI models won't be able to deliver real value.

That's where Informatica's IDMC comes in. IDMC's robust Cloud Data Integration (CDI) capabilities (ingestion, transformation, and orchestration) help organizations with the required tools and architecture to manage data at enterprise scale across complex environments while ensuring compliance, integrity, and efficiency.

With AI moving toward generative models, real-time analytics, and autonomous decision-making, it's more important than ever to have a data foundation that is ready for AI, automated by metadata, and easy to explore. IDMC helps meet these needs by connecting disparate data sources, managing metadata effectively, and supporting robust data governance. By bringing together separate data sources into reliable, AI-ready resources, organizations gain the tools to innovate quickly, automate tasks, and keep pace with the rapid pace of digital change. LTIMindtree, one of the top SI partners of Informatica, is here to help customers drive adoption of the Informatica full stack. LTIMindtree has best-in-class Informatica capability and a dedicated centre of excellence that focuses on modern AI-powered Informatica cloud data management stacks.

Are you looking to get more value from your data and drive innovation with AI?

Learn how Informatica's IDMC can help you create a reliable and flexible data foundation that is ready for the future. Get started today! For more information, write to us at **data.analytics@ltimindtree.com**.

Citations

- ¹ *Accelerating AI Readiness with Informatica IDMC — Connected data management in the cloud*; Vaibhav Suresh, March 13, 2025:
<https://www.informatica.com/blogs/accelerating-ai-readiness-with-informatica-idmc-connected-data-management-in-the-cloud.html>
- ² *The Urgent Need for Data Transformation: Bridging the Gap Between Data and Action*; Jamelle Brown, Forbes Councils Member, April 07, 2025:
<https://www.forbes.com/councils/forbestechcouncil/2025/04/07/the-urgent-need-for-data-transformation-bridging-the-gap-between-data-and-action/>
- ³ *Gartner Survey Finds One-Third of CDAOs Cite Measuring Data, Analytics and AI Impact as Top Challenge*; STAMFORD, Conn., February 20, 2025:
<https://www.gartner.com/en/newsroom/press-releases/2025-02-20-gartner-survey-finds-one-third-of-cdao-cite-measuring-data-analytics-and-ai-impact-as-top-challenge>
- ⁴ *Top data and AI challenges to master for an AI-first business transformation in 2025*; April 22, 2025:
<https://www.softwebsolutions.com/resources/top-data-and-ai-challenges-in-2025.html>
- ⁵ *Informatica: Takeda Fuels AI-Ready Data to Propel Patient Outcomes with Informatica*; Gartner Data & Analytics Summit, March 03, 2025:
<https://www.gartner.com/en/conferences/na/data-analytics-us/sessions/detail/3863413-Informatica-Fuel-GenAI-with-AI-ready-data>
- ⁶ *Gen AI-Powered Industry Use-Cases with IDMC*; Masaki Nakasima; Anita Ayyagari; Sreehari Desai; Ravi Tubati, March 20, 2025:
<https://success.informatica.com/explore/tt-webinars/gen-ai-powered-industry-use-cases-with-idmc.html>
- ⁷ *Informatica and Databricks Expand Partnership to Boost GenAI Capabilities*; Slice of Healthcare Team, June 11, 2024:
<https://www.sliceofhealthcare.com/p/informatica-and-databricks-expand-partnership-to-boost-genai-capabilities>
- ⁸ *IDMC July 2024 Release: Elevating Cloud Data Management*; Nathan Turajski, July 15, 2024:
<https://www.informatica.com/blogs/idmc-july-2024-release-elevating-cloud-data-management.html>

About the Author



Debarshi Ray

Associate Principal, Data Engineering, LTIMindtree

Debarshi brings over 25 years of experience in the software industry, specializing in on-premises and cloud-enabled projects, data migration and integration, platform modernization, data warehousing, advanced analytics, and optimization.

LTIMindtree is a global technology consulting and digital solutions company that partners with enterprises across industries to reimagine business models, accelerate innovation, and drive AI-centric growth. Trusted by more than 700 clients worldwide, we use advanced technologies to enable operational excellence, elevated customer experiences, and long-term value creation. With a workforce of more than 86,000 talented and entrepreneurial professionals across over 40 countries, LTIMindtree — a Larsen & Toubro Group company — is dedicated to solving complex business challenges and delivering transformation at scale. For more information, please visit ltimindtree.com.