HIMSS®

ACCELERATING THE PHARMACEUTICAL VALUE CHAIN WITH DATA AND AI

As competitive pressure grows and R&D becomes more expensive and complex, life sciences firms must leverage AI to compete and win

Pharmaceutical companies must reduce time to market for new therapies

Life sciences companies have always been under pressure to get life-saving drugs to providers and patients, but the evolving healthcare landscape has increased the pressure to accelerate the pharmaceutical value chain and reduce the 10–15-year development timeline.

- **Research and development costs are rising.** In the past 12 years, R&D costs have increased by almost $1 billion per year, according to a Deloitte report.[1]
- **Regulatory requirements are more complex and costly**. Evolving global guidelines now demand more extensive and longer-term safety and efficacy data, particularly post-market, creating significant overhead for compliance and clinical maintenance.
- **Clinical trials are increasingly complex and lengthy.** Trials for novel therapies now require more specialized designs, larger and more diverse patient populations and vast amounts of data collection, with the clinical phase of development now exceeding eight years on average.
- **The shift to precision medicine increases per-patient costs.** The focus on targeted specialty therapies for smaller patient populations inherently drives up the R&D cost per patient, creating pressure to accelerate time-to-market to secure a return on investment.

Despite these trends, life sciences firms must:

- Develop a greater volume of highly specific drugs designed for smaller patient populations as the adoption of precision medicine becomes more widespread
- Rely on the use of genomic and biomarker data to develop targeted therapies
- Manage more complex and decentralized clinical trials

Fortunately for pharmaceutical companies, data and artificial intelligence (AI) have the power to significantly reduce time to market and optimize operations.

> **Data and AI have the power to significantly reduce time to market and optimize operations.**

Dive deeper

Watch "Accelerating the Pharmaceutical Value Chain with Agentic AI," featuring a fireside chat with Rameez Chatni and Nick Calavancia on informa techtarget.

A unified view of data is the key to realizing value from AI initiatives

AI can accelerate R&D by speeding up target identification, optimizing clinical trial design and predicting failures earlier. However, the path to AI implementation is blocked by significant technical and business barriers that limit the building and training of AI models and the ability to trust their results. Ultimately, pharmaceutical companies must build a unified, contextual view of enterprise data to drive successful AI transformation.

Business and organizational barriers

- **Data silos:** Data is historically isolated by functional area (discovery, preclinical, clinical, manufacturing, commercial) and stored in disconnected systems that lack interoperability, and in different cloud and on-premises environments. Siloed data prevents researchers from correlating results across the entire value chain.
- **Cultural resistance:** Teams often maintain a preference for traditional tools and workflows, leading to resistance to sharing data openly or adopting new, data-centric practices. Data is frequently viewed as a departmental asset rather than an enterprise resource.
- **Lack of domain-specific AI talent:** A shortage of AI professionals who can bridge the gap between advanced data science and deep pharmaceutical domain expertise is often a significant bottleneck for AI implementation.

Technical and data barriers

- **Unstructured data:** An estimated 80% of life sciences data is unstructured (e.g., scientific literature, research notes, patient free-text records).[2] This information is rich in context but cannot be processed by traditional databases, forcing manual extraction and creating massive knowledge gaps.
- **Lack of context and semantics:** Even when data is successfully extracted, it lacks context. For example, an AI model might see a single identifier used to represent a protein target in a lab, a clinical trial participant in a safety database and a patent number in a legal repository, but without a model to define the relationships between these entities, the data is meaningless to AI.
- **Data quality and governance:** Datasets often suffer from inconsistency, missing values and varying quality standards across different systems, reducing the reliability and auditability of AI-generated insights, which is a critical concern in a highly regulated industry.

To overcome these barriers, life sciences companies must adopt a data architecture that unifies data across distributed systems and silos, ensures consistent security and governance, and provides the necessary context to heterogeneous data to turn isolated information into actionable, intelligent insights.

Pharmaceutical companies must build a unified, contextual view of enterprise data to drive successful AI transformation.

An open data lakehouse provides a foundation of trusted data for AI

AI is only as valuable as the data it's trained on. It requires authoritative, trusted and governed data that meets a high-quality bar and is accessible to every role across the pharmaceutical value chain. An open data lakehouse architecture that combines the flexibility and scalability of data lake storage with the management, governance and analytic capabilities of relational databases provides the intelligence layer that ensures AI success.

AI needs the lakehouse

The lakehouse is the architecture that is best suited for storing and processing all data, from structured clinical records to semi-structured genomic files to unstructured scientific text — all in a single place. By collapsing siloed data lakes and data warehouses into a single architecture, lakehouses reduce data movement and ensure that AI models are trained on the richest data available.

End-to-end AI/ML support

The open data lakehouse provides support for every step of the AI workflow:

- Easy, governed, secure and federated access to a consistent and unified view of enterprise data
- Elastic scalability to support the varying compute requirements for model training and experimentation
- Model explainability with governance capabilities that enable teams to recreate a view of the data at specific points in time

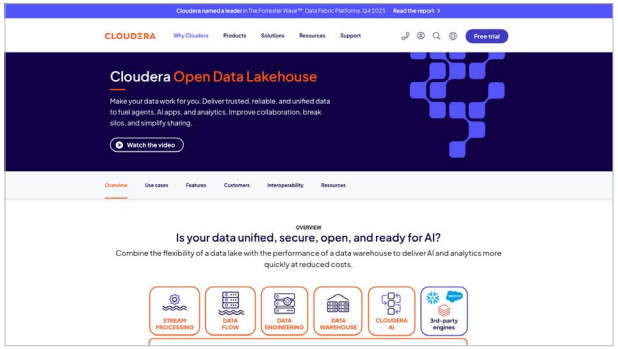
Openness and future-proofing

One of the biggest challenges in pharmaceutical companies is the proliferation of point solutions and heterogeneous systems that often rely on proprietary data formats and query engines. An open data lakehouse built on open standards, formats and technologies ensures the data is always accessible
by a variety of tools and engines.

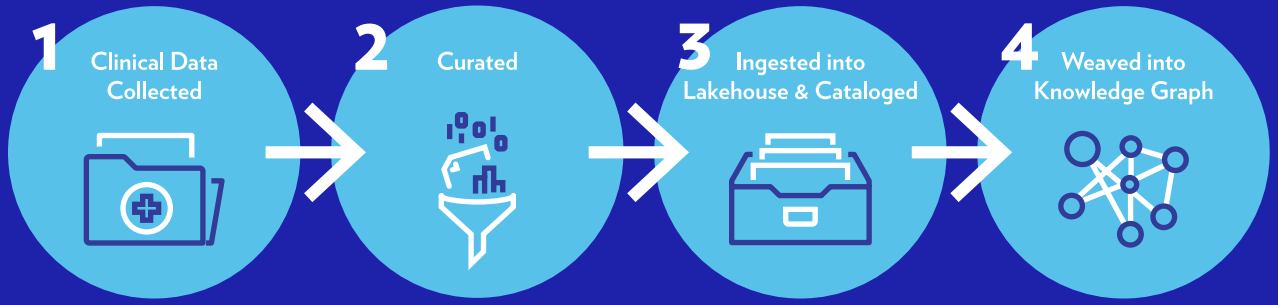
By providing a single, governed and highly scalable environment for all data and AI workloads, the open data lakehouse establishes the intelligence layer that is essential for AI to move from experimental pilot projects to high-value, production-ready solutions
that accelerate the pharmaceutical value chain.

Dive deeper

Learn more about Cloudera Open Data Lakehouse



Building on the AI intelligence layer with context



The knowledge graph builds on the AI intelligence layer and transforms a collection of siloed records into a dynamic, interconnected knowledge network, making complex scientific and operational data accessible, traceable and intelligent.

A knowledge graph is a framework that models the relationships between data entities (nodes) with semantic labels (edges). Instead of simply storing data in rows and columns, knowledge graphs model the real-world complexities of the pharmaceutical value chain. This map enables AI models to provide reliable insights and even find relationships and correlations between datasets that might take researchers weeks to months to find on their own.

- The biggest technical hurdle to building an effective knowledge graph is the complexity of life sciences data estates. Data teams must:
  - Leverage an open, hybrid architecture that can unify data and metadata wherever it resides, including on-premises data centers and cloud environments.
  - Store, process and integrate both structured and unstructured data to ensure a holistic view of organizational data.
  - Centralize security and governance and ensure consistent application of policies for end-to-end good practices (GxP) compliance, explainability and auditability of data and AI.

A knowledge graph built on a secure and flexible unified data architecture is the intelligence layer that converts an organization's vast, fragmented data estate into a reliable foundation for AI that accelerates the pharmaceutical value chain.

Dive deeper

Read more about knowledge graphs and more here: Cloudera's Top Takeaways from AWS re:Invent 2025.

Managing distributed, heterogenous environments

In the life sciences industry, data is an asset, but it also carries a regulatory responsibility. The success of AI depends entirely on the organization's ability to provide trustworthy data with security and compliance built into every step of the data lifecycle. This challenge is even more difficult in distributed environments.

Life sciences companies must build a security and governance framework that ensures policies are applied consistently regardless of where the data resides.

Universal security and access controls: Security and access control policies should be written once and applied and enforced consistently across lakehouse and AI workloads, even as data moves, to ensure compliance.

End-to-end lineage and traceability: Life sciences organizations must be able to prove that the data used for decisions like drug target validation or clinical submissions has not been tampered with. This requires tracking the lineage of every data point from its source through each transformation and consumption. They must be able to use tools like time travel to recreate a snapshot of the data at any point in time to address GxP compliance audits and other regulatory requirements.

Compliance built on openness: By standardizing on open formats and technologies, life sciences companies can ensure security and governance is comprehensive, transparent and simplified. Organizations maintain control over their data and all associated policies. Compliance and security are universal and interoperable. And GxP validation is easier and less expensive.

By providing a common security and governance framework across environments, life sciences companies can ensure that compliance is built into every step of the AI lifecycle and innovate more quickly and reliably.

Life sciences companies must build a security and governance framework that ensures policies are applied consistently regardless of where the data resides.

Bring AI to your data, wherever it lives

To fully realize the potential of generative AI and agentic AI, life sciences companies must bring AI to the data.

The pharmaceutical value chain requires an intelligence layer that spans from remote clinical trial sites to the central data platform. The ability to serve AI models at the point of ingestion enables real-time insights and agentic workflows that accelerate decision making and automate operations.

Intelligence anywhere encompasses three core principles:

- **Edge intelligence:** AI models deployed closest to the source enable proactive safety monitoring and real-time quality control, preventing costly failures and accelerating response times.
- **Deep learning in the core:** Data is processed and stored in the core, where AI models are trained and retrained on a complete, trusted view of enterprise data and external knowledge.
- **AI for BI:** Empower knowledge workers with generative AI augmented by a knowledge graph, enabling employees to be conversational with the data and to uncover trusted insights more rapidly.

An edge-to-AI strategy ensures that life sciences firms operate with continuous intelligence, moving beyond slow, batch-based decision making to a real-time, AI-driven enterprise.

> **The pharmaceutical value chain requires an intelligence layer that spans from remote clinical trial sites to the central data platform.**

Realizing AI's potential in life sciences

The urgency to accelerate R&D and reduce time to market is no longer just a strategic goal — it is a necessity. The compounding pressures of rising costs, regulatory complexity and the shift to precision medicine demand a fundamental change in how pharmaceutical organizations operate.

AI is the key to unlocking the potential of massive volumes of enterprise data, but to be successful, life sciences companies must:

- Build an AI intelligence layer by ensuring consistent, unified access to data with an open data lakehouse.
- Centralize security and governance with a framework that applies consistent policies across heterogeneous environments.
- Deliver context with a knowledge graph that models semantic relationships within the data, improving accuracy, insights and results from AI.
- Bring AI to the data anywhere it resides, enabling agentic workflows for real-time decision making and insights and empowering knowledge workers to interact with data using natural language.

This unified, intelligent data architecture is the only way for life sciences to overcome the business and technical barriers to AI adoption and accelerate the pharmaceutical value chain to get life-saving therapies to providers and patients faster than ever.

To learn more about Cloudera, go to www.cloudera.com.

References

1. Deloitte. 2024. Perspective: Be brave, be bold. Measuring the return from pharmaceutical innovation. https://www.deloitte.com/us/en/Industries/life-sciences-health-care/articles/measuring-return-from-pharmaceutical-innovation.html
2. Wood, T. November 2, 2024. Unstructured data. *Fast Data Science*. https://fastdatascience.com/natural-language-processing/unstructured-data/

CLOUDERA

About Cloudera

Cloudera is the only platform for data and AI - anywhere. With 100x more data under management than cloud-only vendors, Cloudera empowers global enterprises to transform data of all types, on any public or private cloud, into valuable and trusted insights. Our open data lakehouse delivers scalable and secure data management with portable cloud-native analytics, enabling customers to bring GenAI models to their data while maintaining privacy and ensuring responsible, reliable AI deployments. The world's largest brands in financial services, insurance, media, manufacturing, and government rely on Cloudera to use their data to solve what seemed impossible—today and in the future.

Produced by HIMSS