



UX | PRODUCT BUILD | QA | APPLICATION MODERNIZATION

Building Agentic Intelligence for Real-Time Supply Chain Insight

gravity9 partnered with one of the world's largest pharmaceutical distribution companies; a Fortune-25 enterprise operating across every stage of the healthcare supply chain. Their business encompasses everything from manufacturing and warehousing to national distribution and pharmacy-level delivery. With operations on this scale, access to real-time data is mission critical. However, the company's data landscape was highly fragmented. Key operational data was spread across multiple platforms, including Warehouse Management Systems, Enterprise Resource Planning tools, and EPCIS platforms. Each system served its purpose, but they didn't communicate well with each other.

As a result, retrieving key insights such as where a product was last shipped, what lot numbers were affected by a delay, or how much of an order was fulfilled required hours of manual effort and technical know-how. Non-technical teams like customer service and logistics coordinators were particularly affected, unable to access the data they needed without relying on technical colleagues.

With a clear need: they wanted to empower their people with fast, intuitive access to critical supply chain data, without requiring deep technical expertise or custom queries. The solution needed to be intelligent, accurate, scalable—and delivered quickly.



Utilized Technology Stack

Data Storage: MongoDB Atlas

Agentic AI Workflows: LlamalIndex

Review of Challenges

The challenge began with the sheer complexity of the data itself. Supply chain events such as aggregation, shipping, disaggregation, and decommissioning generated deeply nested, interrelated records. Questions like "Where was this serial number last

seen?" or "What's the current status of lot number X?" couldn't be answered by any single data source. They required correlation across multiple systems and documents, each with its own structure and logic.

On top of that, the company faced a usability gap. Most of their existing systems required advanced querying knowledge. This created bottlenecks where only a handful of technical users could retrieve the answers the broader team needed. Even when those answers came, they often arrived too late to support real-time decision-making or customer communication.

Speed was another pain point. Operational decisions often had to be made in the moment; tracking an urgent delivery, responding to a product recall, or checking fulfillment accuracy. The existing tools couldn't keep up.

Finally, there was a question of scalability. The data volume was only growing, and any solution would need to support increasingly complex use cases without reducing performance or requiring constant manual intervention

Our Solution

We proposed and delivered a powerful yet elegant solution: a conversational Logistic AI Assistant capable of answering complex supply chain questions in natural language. Built using LlamalIndex and layered over MongoDB, the Logistic AI Assistant enabled users to ask questions like "What percentage of customer X's shipment was fulfilled today?" or "Which warehouse last scanned this serial number?" and receive accurate, contextual answers in real time.

The core of the solution was a hybrid architecture that combined an intelligent agentic framework with a robust fallback retrieval-augmented generation (RAG) pipeline. When a user asked a question, the system first attempted to resolve it through a purpose-built ReAct agent equipped with specialized tools. If the agent couldn't resolve the query, the system automatically transitioned to a RAG-based approach. This second path translated the question into a structured MongoDB query, retrieved the relevant data, and synthesized a clear answer using a large language model.

This dual-layered approach offered the best of both worlds: high performance for common questions, and depth and flexibility for more complex or novel ones. Everything was containerized for seamless deployment in either cloud or on-premise environments. An essential requirement for a client working under strict regulatory and data security standards.

Our Approach

The team began by immersing itself in the client's supply chain workflows and technical landscape. Understanding the underlying business problems helped us design a solution that wasn't just technically capable, but genuinely helpful for the people who would use it every day.

We then translated these workflows into a modular architecture using LlamaIndex. Our team created a set of custom tools tailored to MongoDB's structure, allowing the agent to retrieve precise data quickly and consistently. We also developed a conversational interface that could refine user questions, resolve ambiguity, and maintain context across exchanges.

To ensure the solution remained scalable and robust, we engineered a fallback mechanism that could handle large, complex queries without losing performance. By using LLMs to translate natural language into structured queries and then integrate the results back into natural language responses, we created a seamless user experience that masked the complexity beneath.

Throughout the project, we maintained close collaboration with the client's internal teams. We prioritized early wins and iterative feedback, ensuring the solution was intuitive, practical, and aligned with operational needs. The entire system was designed for rapid deployment and future extension, with all components built to be reusable, modular, and secure.

Subsequent Outcomes

The Logistic AI Assistant launched smoothly and began delivering value immediately. Non-technical users, many of whom had never run a database query could now ask detailed supply chain questions and receive answers in seconds. What once took a team of analyst's hours now happened in near real-time.

The operational impact was significant. Customer service teams responded more confidently and quickly to order inquiries. Logistics coordinators tracked inventory movement without escalating to data teams. Managers made faster, more informed decisions. The overall responsiveness of the organization improved across the board.

The solution also reduced manual effort dramatically. By automating the process of data correlation across systems, the Logistic AI Assistant freed up hours of work each week, time that teams could now spend on higher-value tasks. Internally, word of the tool's success spread, generating interest from other departments and laying the foundation

for future enhancements.

Client Feedback

Assistant quickly became a trusted part of daily operations, not just for logistics and customer support, but for anyone who needed fast answers from complex data. Leaders within the organization highlighted the project as a model for how AI can be deployed responsibly and effectively at scale.

What stood out most was the shift in mindset. Teams who once viewed their data systems as opaque and intimidating now saw them as accessible and empowering. The conversational interface and intelligent responses bridged a longstanding gap between business needs and technical infrastructure.

By building a solution that spoke the language of its users, both literally and metaphorically gravity9 helped unlock real operational intelligence and set the stage for broader digital transformation.

**WE'RE
BUILDING
BETTER
DIGITAL
PRODUCTS**

