

Case Study:

Data Ingestion Using AI

What the Company Does / Industry

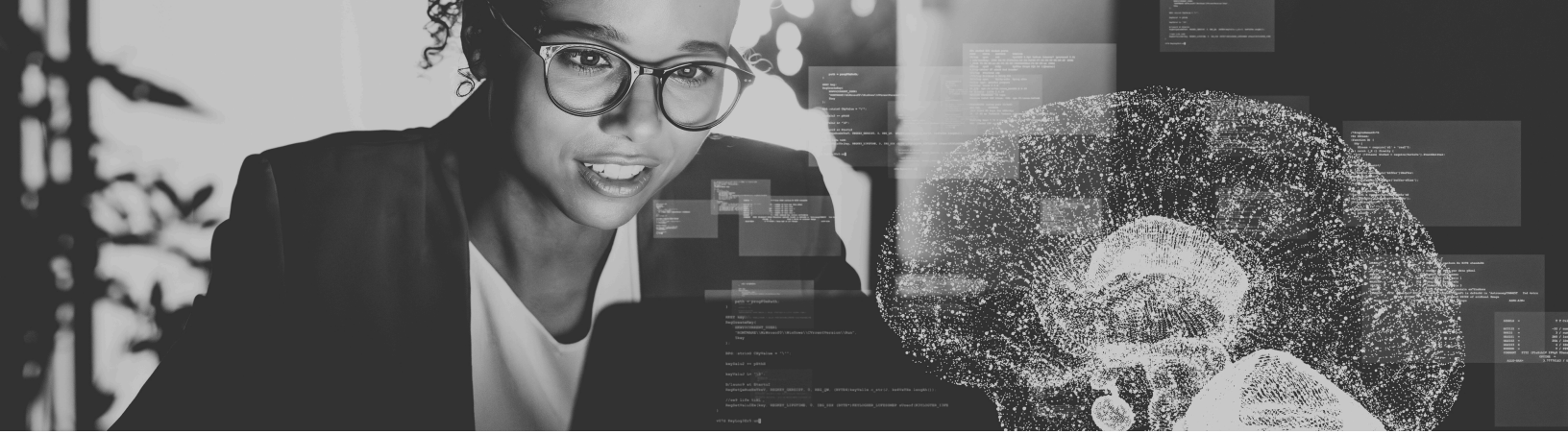
The company operates in the omnichannel commerce and retail data space, helping brands understand product availability, pricing, and catalog accuracy across major retailers. Their platform ingests data from retailer APIs, websites, and scrapers at massive scale.

The Problem That Led to the Project

The company relied on a rules-based and human-driven product classification process:

- Millions of product records daily
- Inconsistent retailer data (text, images, formats)
- High manual effort to match products to a canonical catalog
- Limited ability to scale without ballooning headcount

They needed automation without sacrificing accuracy.



How We're Solving It Today and How It Evolves Over Time

Current State (Pragmatic, Production-Focused)

Today's solution prioritizes reliability, scale, and cost control over full autonomy:

- ✓ A centralized, cloud-native scraping and orchestration pipeline
- ✓ Retailer integrations that are still explicitly configured, but far more standardized than before.
- ✓ Automated ingestion and delivery of product, price, and inventory data into the company systems.
- ✓ Removal of much of the manual human effort previously required to operationalize scraped data.

This approach is intentionally more rigid than the long-term vision – not because AI isn't desirable, but because fully autonomous scraping is not yet cost-effective or reliable at the scale the company requires.

In short: the current architecture stabilizes and automates what was previously fragile and manual, while laying groundwork for what comes next.

Long-Term Vision (Adaptive, AI-Driven Ingestion)

The long-term goal is to move away from retailer-specific scraping logic entirely.

The future state envisions a system where the company can:

- 📍 Point the platform at a retailer's website

- ⬡ Automatically discover products, pricing, and inventory
- ⬡ Adapt dynamically as site layouts change

This will be enabled through a combination of:

- ⬡ Vision AI and OCR to interpret product pages visually
- ⬡ Vector search and LLM-based analysis to match scraped content to canonical products
- ⬡ Confidence scoring and selective human-in-the-loop review where needed

That vision promises resiliency over rigidity – but achieving it efficiently and affordably requires careful evolution, not a premature leap.

If your business depends on ingesting data from external systems you don't control, automation alone isn't enough. Long-term success requires architectures that balance pragmatism today with adaptability tomorrow. The right path modernizes fragile processes first – and introduces AI where it truly adds resilience, not just novelty.



Thinking through a similar challenge?

This project is one example of how the right cloud strategy creates real business impact. Connect with Serverless Solutions to see what's possible for your team.

