databricks

EXECUTIVE BRIEF

How Retail and CPG Leaders Are Putting AI to Work

Based on Hundreds of Databricks Deployments

Introduction

Retail and CPG companies sit on one of the richest datasets in the global economy: every transaction, every click, every shipment, every shelf scan, every loyalty signal. Working together, this information holds the power to sharpen every operational decision, anticipate every shift in consumer demand, and deepen every customer relationship. Today, though, most of it cannot work together. It remains fragmented across e-commerce platforms, POS systems, supply chain tools, loyalty programs and third-party data sources, each with its own version of the truth.

The promise of AI shines bright especially in consumer industries. The companies pulling ahead are not simply deploying AI. They are deploying AI on a unified, governed data platform where the connections between signals become visible and actionable. This is how a single pilot becomes a scaled system, where every new use case compounds on the last and AI stops being an experiment and starts driving measurable growth, margin, and customer lifetime value.

Key takeaways

Multi-agent architectures, not single-purpose chatbots, are the production standard for leading companies.

Mondelez manages over 20,000 models in both batch and real time using MLflow, powering use cases across sales execution, marketing, supply chain and revenue growth management from a single, governed platform.

Data governance is the architectural prerequisite for AI at scale.

Unity Catalog's fine-grained access controls, column-level security and full lineage through agent workflows allow retailers and brands to safely activate customer and supplier data across use cases like customer 360 personalization, dynamic pricing and inventory management. Crucially, these capabilities must operate without compromising consumer privacy or adherence to regulations such as GDPR.

The most effective AI works across every data type — structured and unstructured — in a single coordinated system.

SQL and Genie agents handle database queries while RAG and Vector Search agents analyze product catalogs, customer reviews, trade promotion plans and supplier documents. A merchant or brand manager can ask a question and get an answer that pulls from both live sales data and detailed product or campaign materials, without needing to know where the data lives.

Time to value is measured in weeks, not quarters.

Across Databricks deployments, companies are reaching production faster than expected: Carhartt achieved a two-week time-to-market for new features; Alpura accelerated deployment times by over 85%, reducing app delivery from 12 weeks to just 2–5 days for data products and applications.

The efficiency gains are real and attributable.

Grupo Casas Bahia achieved a 14x productivity gain, analyzing 1,400 comments in the time manual processes handle 100. Cafe24 dramatically reduced their global expansion timelines, launching a new mall in just 20 minutes and achieving first sales within 10 days.

At Burberry, reducing clickstream latency by 99% enabled real-time AI-ready Customer 360 profiles, powering 40+ personalization models for product recommendations, propensity scoring and lifetime value. Client advisors can access a customer's latest online behavior in-store, turning browsing signals into immediate, high-conversion interactions.

Align your stakeholders before you build.

The most successful deployments involve CMOs, CDOs, CTOs/CIOs, data science leaders, heads of supply chain and merchandising, sales leaders, IT security and finance from day one. This cross-functional collaboration is what separates pilots that scale from pilots that stall.

14x
Productivity Gain

90%
Reduced Processing Costs

99%
Latency Reduction

Use cases

Multi-Agent Supervisor Architectures

Most retailers and consumer goods companies rely on dozens of disconnected systems — from POS and e-commerce platforms to trade promotion tools, syndicated data feeds and supply chain systems. Multi-agent architectures solve this by assigning specialized agents to handle different tasks, all coordinated by a central supervisor that routes each question or workflow to the right system in real time. This enables workflows such as dynamically coordinating pricing, promotions and inventory decisions, optimizing trade promotion performance across retailers, powering real-time personalization and recommendation engines, and enabling supply chain planners to respond to demand signals and disruptions as they happen.

Mondelez manages over 20,000 models in both batch and real time using MLflow across sales execution, marketing, supply chain and revenue growth management. Cafe24 built a GenAI data platform powered by Agent Bricks that orchestrates specialized agents across reasoning, analysis and execution. Together, these approaches unify data exploration, analytics and operational action into a single end-to-end workflow.

Built on: *Unity Catalog (access controls and data lineage), Genie agents (structured data/SQL), RAG agents (unstructured documents), function-calling agents (vector search), Databricks Model Serving*

Natural Language Data Access and Analytics

What if merchants, planners, brand managers and revenue growth teams could get answers from their data just by asking a question — no SQL, no analyst queue, no waiting? At Grupo Casas Bahia, Genie reduced time to analyze data from 5–6 hours to minutes, enabling teams to move from delayed reporting to real-time decision-making across merchandising and operations.

Users can ask questions in their native language, Portuguese, and get answers to questions like "Which top 5 categories have moved the most products today?" Instead of waiting days for reports, decisions happen in the moment, closer to the customer and the shelf.

At Danone, Databricks unifies data across global business units and powers GenAI-driven analytics through tools like Unity Catalog and Genie, breaking down silos and delivering faster, more accurate insights. Initiatives like Danone's WeFi project and internal GenAI chatbots accelerate time to actionable insight across marketing, supply chain and commercial teams.

Built on: *Databricks Model Serving, open-source LLM support (Llama, Mistral), Genie text-to-SQL, Mosaic AI critique agent framework*

Predictive Demand and Customer Intelligence

The biggest risks are missed demand signals, stockouts and delayed customer insights. Predictive models use real-time data across sales, inventory and customer behavior to identify risks and opportunities before they impact revenue.

Alpura monitors OEE (overall equipment effectiveness) dashboards and receives automated alerts to anticipate supply chain disruptions earlier. Flaconi's prediction times dropped from 20 minutes to 300 milliseconds, enabling real-time recommendations that improve conversion and speed purchasing decisions. They expect their net order income to increase by 5%.

Built on: *Databricks streaming pipelines, Delta Lake, real-time ML model serving, Mosaic AI*

Data Governance as Enabler

Governance is what makes enterprise-wide AI possible in retail and consumer goods, where sensitive customer, transaction and supplier data must meet strict requirements like GDPR and evolving consumer privacy standards. Unity Catalog's access controls, data masking and lineage ensure AI can scale without introducing compliance risk.

At Adidas, a GenAI chatbot analyzes millions of customer reviews with secure, governed access to data and models. Pilot Company sits on terabytes of data. Unity Catalog enforces access controls and maintains data and model integrity for all their GenAI applications, enabling faster innovation without sacrificing trust.

Built on: *Unity Catalog (column-level security, dynamic data masking, data lineage tracking through agent workflows, audit logging), fine-grained access controls for multi-agent systems*

Strategic Recommendations

Build Your Foundation

Before your first agent deployment, implement Unity Catalog. Its built-in access controls, data lineage and audit logging are what allow you to expand AI beyond your first team without creating governance debt. Start with high-friction workflows like promotion performance analysis, demand forecasting or inventory visibility, where teams rely on manual reporting and delayed insights. Then run a focused two-week pilot on a real use case, such as enabling merchants or brand managers to query performance data in natural language and act on it immediately.

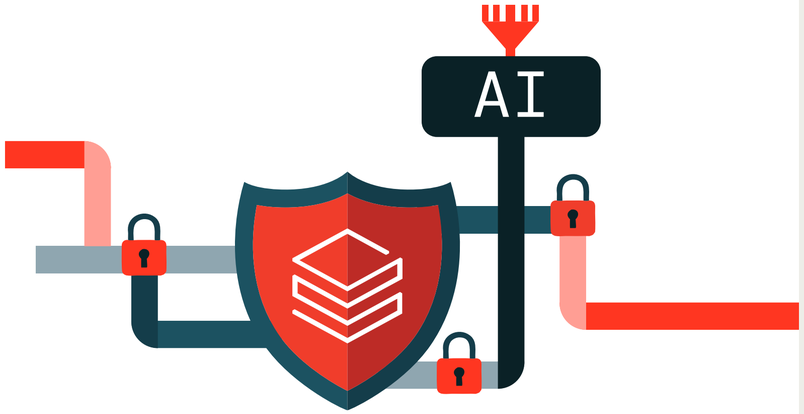
Scale What Works

Once a workflow is proven in one business unit, treat it as a template for the next. Databricks enables reusable pipelines, shared notebooks and governed data assets that let you take a text-to-SQL workflow or multi-agent architecture and expand it across an organization without rearchitecting. Use MLflow to track model performance and Unity Catalog to manage access as new teams and data domains come online. The goal is to build a library of reusable agent components instead of a collection of isolated deployments.

Skip What Doesn't

Not every investment is worth making. Three to avoid:

- **DIY MLOps infrastructure.** Databricks and major cloud providers have already solved this; building your own means spending engineering time on maintaining infrastructure instead of solving business problems.
- **Single-purpose AI point tools.** Every additional vendor increases the total cost of ownership and adds integration debt.
- **Premature model optimization.** LLMs improve every quarter. Build on an open architecture that lets you swap in better models as they become available without rebuilding your workflows.



Results from the Field

Three patterns emerge across these deployments:

Efficiency gains are concentrated in data operations: processing, querying and enrichment. This repetitive, high-volume work is exactly what AI handles best.

Speed improvements are transformational, not incremental. Cafe24's ability to launch a global mall in just 20 minutes and Alpura's 85% faster deployment cycles fundamentally change how quickly teams can move from insight to execution.

Scale compounds — each AI workflow absorbs more throughput without adding operational cost. Grupo Casas Bahia analyzes 1,400 comments in the time manual review would get through 100, while automatically classifying 33,500 customer reviews each month. This level of scale would be impossible to manage manually.

In every case, the underlying enabler is the same: Databricks Platform that connects raw operational data to the AI layer without requiring teams to stitch together point solutions to bridge the gap.

adidas	~92%	Cost Savings More efficient LLMs
BURBERRY	99%	Reduction in latency for customer clickstream data
PETSMART	13%	Lift in autoship transactions driven by adaptive targetings
Walgreens	20%	Increase in Productivity Demand Planning

Realistic timeline

When it comes to AI, business decision-makers often wonder:
How long before this technology delivers something real?

Based on Databricks deployments across retailers and brands:

DAYS	**Working MVP** TomTom accelerated prototype-to-production from weeks to days
2–5 DAYS	**Production-Ready Apps** Alpura delivers data products in days, reducing deployment time from 12 weeks
15 MINUTES	**Infrastructure Setup** Flaconi's infrastructure setup reduced from two days to minutes, without dependencies from other departments

Real, measurable value can be delivered in weeks, not the multi-year transformation cycles most enterprises are used to. The companies that move fastest have two things in place before they start: a governed data foundation, so agents have clean, accessible data to work with, and a clear business owner who defines what 'working' looks like. Both are achievable before your first sprint begins.

Conclusion

AI in Retail and CPG firms is no longer an experiment; it's an operational capability that leading companies are deploying at scale right now. The companies pulling ahead aren't doing it with a collection of point solutions or a single chatbot. They're doing it with a unified platform that brings data engineering, business teams, governed access, multi-agent orchestration and model serving together in one environment.

Databricks is the defensible choice for retail and CPG AI at scale, and it comes down to three advantages:

Open architecture:
Native support for ChatGPT, Claude, Llama and other open-source models means no model lock-in, and you never have to rebuild when a better model emerges.

Governance at the platform layer:
Unity Catalog makes it possible to deploy AI broadly across an organization without compromising data protection or compliance.

The compounding advantage of a unified platform:
Every new use case built on Databricks shares the same data assets, pipelines and governance framework, so the tenth deployment takes a fraction of the effort the first one did.

The question for retail and CPG brand leaders isn't whether to deploy AI. It's whether to spend the next two years stitching together tools that don't scale, or build on a unified foundation where every deployment makes the next one faster, smarter and more efficient.

databricks