

# Databricks Enabled Data and AI Use Cases for Retail Chain

Authored By: Aman Balwan

## **1. Executive summary**

Retail is rapidly shifting toward AI native operating models in which every customer interaction, merchandising decision and supply chain move is informed by real time data and machine learning. Leading brands such as adidas, Burberry, Walgreens and Hershey already use the Databricks Data Intelligence Platform to transform customer engagement, demand forecasting and decision making at scale.

For a multi brand, multi product retail chain, this shift is an opportunity to create a single AI ready data foundation that powers omnichannel personalization, optimized inventory, high ROI marketing and self-service analytics for executives and field teams. ***This document outlines that vision, describes concrete Databricks use cases and proposes an execution roadmap with success metrics for leadership.***

## **2. Strategic context and rationale**

Consumer expectations for personalized, seamless, omnichannel experiences continue to rise while cost and margin pressures force retailers to run leaner inventories and more efficient operations. At the same time, data remains fragmented across ecommerce, stores, loyalty programs, media platforms and suppliers, which limits the ability to deploy AI where it can have the most impact.

The **Databricks Data Intelligence Platform** addresses this fragmentation by **unifying data warehousing, data engineering and machine learning on a single lakehouse architecture that is built on Delta Lake.**

### **3. Target vision: an AI ready retail operating system**

The target state is an **AI ready operating system for retail, which means a single governed data and intelligence layer that powers all major retail functions**. In this model, all behavioral, transactional and operational data lands in **Delta Lake on Databricks in near real time and is governed through Unity Catalog as one source of truth**.

Customer, product, inventory, marketing and operational datasets are enriched and linked into an AI ready graph that supports both traditional BI and advanced ML or GenAI workloads. AI models and agents, including demand forecasting, price optimization and GenAI shopping assistants, are trained and deployed on the same platform and their outputs are pushed back into ecommerce, store, supply chain and marketing systems to drive measurable business outcomes.

### **4. Current pain points in multi brand retail**

Multi brand, multi category retailers typically experience recurring structural challenges that limit growth and efficiency.

- Customer data is scattered across ecommerce, apps, stores, loyalty systems, call centers and external partners, which makes it hard to calculate customer lifetime value and limits personalization.
- Operational data on inventory, sales, promotions and suppliers is often delayed and inconsistent, which contributes to out of stocks on winning products and excess stock on slow movers.
- Webrooming journeys, where customers browse online and buy in store, are not stitched together so store associates lack insight into what customers researched before arrival.
- Business teams rely heavily on spreadsheet based reporting and email attachments, which creates long turnaround times for new analysis and makes it difficult to maintain a single version of the truth.
- AI pilots often run on separate tools and infrastructure, which makes it difficult to move from proofs of concept to enterprise grade governed AI products.

**Addressing these pain points requires a unified data and AI foundation that can support many functions without creating new silos.**

## **5. Platform foundation on Databricks**

### **5.1 Unified Lakehouse on Delta Lake**

All priority datasets, including point of sale transactions, ecommerce events, mobile app telemetry, loyalty and CRM data, product catalog and hierarchy, supplier data, store operations and marketing signals, are landed in Delta Lake tables on Databricks. Delta Lake provides ACID transactions, schema evolution, time travel and scalable performance for streaming and batch workloads, which ensures reliability for downstream analytics and AI.

### **5.2 Behavioral data with Snowplow and Mosaic AI**

As demonstrated by Burberry, clickstream and behavioral events from web and app can be captured with Snowplow and delivered into Databricks as AI ready and well-structured events. This creates a high-quality feed for customer journey analysis, real time personalization and marketing attribution while maintaining full ownership of the behavioral data.

As an alternative or complement, Mosaic AI on Databricks can be used to build machine learning and GenAI workloads directly on behavioral data stored in the lakehouse, enabling capabilities such as user journey analysis, churn prediction and real time sentiment decoding without relying on a specific behavioral collection tool. In this model, clickstream data can be ingested through any preferred data collection solution, and Mosaic AI provides the vector search, modeling and agent framework needed to power downstream personalization and customer experience use cases.

## **5.3 Governance and security with Unity Catalog**

Unity Catalog serves as the central governance layer for data, machine learning models and AI agents, and it provides fine grained access control, lineage and audit capabilities. This is essential for handling customer and supplier data responsibly, meeting privacy regulations and ensuring that GenAI applications only access approved and

## **6. Priority use cases**

### **6.1 Voice of customer and product feedback intelligence**

**Objective:** *Turn unstructured customer feedback into a continuous and actionable signal for product, pricing and experience decisions.*

Adidas uses Databricks and Mosaic AI to analyze more than two million customer reviews with a retrieval augmented generation pattern, which improves analyst productivity and reduces infrastructure costs. Reviews are embedded into vectors, stored in Mosaic AI Vector Search and exposed through a GenAI assistant that internal teams query in natural language to surface sentiment, themes and product issues.

A similar solution can ingest product reviews, survey responses, call center transcripts and social media content into Databricks, and then apply LLMs to extract topics and sentiment across dimensions such as price, quality, packaging, service and delivery. Databricks customer feedback patterns show how these models can automatically classify feedback, detect emerging issues and summarize insights by brand, category and geography, which replaces manual, spreadsheet-based analysis.

### **6.2 Real time Customer 360 and omnichannel personalization**

**Objective:** *Personalize experiences across web, app and store by unifying behavioral and transactional data into an AI ready Customer 360.*

Burberry implemented Snowplow with Databricks to build real time Customer 360 profiles that power around forty personalization models, including product recommendations,

propensity scores and lifetime value. As clickstream data flows in from [Burberry.com](#), Snowplow structures events and Databricks updates model scores, which allows in store client advisors to see opted in customers digital behavior on mobile devices.

For this retailer, unifying loyalty, ecommerce and store transaction data with behavioral events in Databricks would enable several high value capabilities.

- Personalized product recommendations across all brands and categories, driven by inferred preferences, sizes and recent browsing.
- Propensity models for purchase, churn, cross sell and upsell, which allow finely targeted offers instead of broad discounts.
- Webrowsing visibility, where store associates can review a customers recent browsing and saved items to tailor in store recommendations.

Burberry also shows how to extend cookie lifetimes through server side tracking while honoring GDPR and consent management, which is important for long purchase cycles.

### **6.3 Store level demand forecasting and inventory optimization**

**Objective:** *Predict demand and optimize stock by store, channel and product in order to reduce out of stocks and markdowns.*

Walgreens built an Information, Data and Insights platform on Databricks that processes roughly forty thousand events per second and supports machine learning based demand forecasting and inventory optimization across its pharmacies. These models determine store level stocking needs, improve productivity by about twenty percent and enhance product availability.

For this retailer, Databricks can aggregate historical sales, promotions, pricing, supplier lead times, seasonality and local signals and then use these data to train forecasting models at store and SKU and day level. These forecasts can then drive replenishment recommendations, dynamic safety stock rules, assortment decisions and alerts when live sales deviate from forecasts, and this enables proactive corrective actions.

Hershey created a Commercial Data Store on Databricks, and this shows the value of a single, timely source of truth for commercial performance that supports better daily decisions by sales, revenue management and supply chain teams.

#### **6.4 Marketing attribution and media optimization**

**Objective:** *Maximize return on marketing spend using data driven attribution, audience building and budget planning.*

Burberry used Snowplow and Databricks to refine marketing taxonomies, referrals and consent signals and to build a rich and fully owned attribution dataset plus a tailored last click attribution model. This work provided a clearer view of which channels and campaigns are most effective and enabled smarter budget allocation.

On Databricks, the retailer can centralize impressions, clicks, onsite behavior, conversions and offline sales and then use time series and causal models to measure incremental lift by channel and campaign. Scenario models can then test alternative budget distributions and forecast net sales and margin outcomes by brand and category, which improves alignment between marketing spend and commercial goals.

#### **6.5 Executive and field self service intelligence**

**Objective:** *Replace static spreadsheets with governed, real time dashboards and natural language analytics for leadership and field teams.*

Hershey built a Commercial Data Store on Databricks and now centralizes commercial data and serves it through Databricks SQL into tools such as Power BI and Tableau, which consolidates multiple reports into one executive summary dashboard with flexible filters. This approach reduced reliance on complex spreadsheet macros and manual reporting and enabled business users to explore data by retailer, category, promotion and region without IT involvement.

For this retailer, curated semantic layers in Databricks SQL can expose measures like sales, margin, sell through, stock positions and customer metrics to BI tools, with access

governed by Unity Catalog roles. Executives and regional leaders can track KPIs such as like for like sales, inventory turns, forecast accuracy and NPS on mobile dashboards and can use natural language query features built on GenAI to answer diagnostic questions.

## 6.6 GenAI assistants for shoppers and associates

**Objective:** Use GenAI to enhance customer experiences and associate productivity through conversational interfaces.

Adidas built a GenAI solution on Databricks that shows how a retrieval augmented assistant can surface complex internal insights in an accessible way for non-technical users. Databricks Mosaic AI and agentic workflows extend this pattern to assistants that can reason over product catalogs, policies and knowledge bases and can take actions through APIs.

For this retailer, potential assistants include several applications.

- A digital shopping assistant on web and app that helps customers find the right product across brands, answers questions about fit, compatibility, availability and alternatives and uses real time data.
- An associate copilot that provides store staff with instant access to Customer 360 profiles if consent is given, inventory information, cross sell suggestions and troubleshooting guides.
- An internal knowledge assistant that consolidates SOPs, HR information, planograms and training content to reduce the time employees spend searching for information.

Running these assistants on the governed Databricks platform helps ensure that they respect access controls and data use policies.

## 6.7 Data governance, privacy and compliance by design

**Objective:** Maintain trust and regulatory compliance while expanding data and AI usage.

**Unity Catalog provides centralized governance for data and AI assets**, including permissions, lineage and audit across the Databricks environment. **This central control is**

**critical as more teams and use cases rely on sensitive customer, employee and supplier data.**

**Burberrys deployment with Snowplow and Databricks illustrates how to maintain GDPR compliance while extending anonymous cookie lifetimes using server side tracking, supported by robust consent tracking and selective data collection.** Together, Snowplow and Databricks create a privacy preserving foundation for Customer 360 and personalization that keeps full control over what is collected, processed and used.

## **7. Execution roadmap**

A phased roadmap can reduce risk while delivering measurable value at each stage.

### **Phase 0 (0 to 6 months): Establish lakehouse and governance**

- Land priority datasets such as POS, ecommerce and app events, product catalog and core inventory in Delta Lake with a clear and minimal data model and Unity Catalog based access controls.
- Integrate Snowplow or an equivalent tool for high quality behavioral data capture into Databricks in near real time.
- Stand up Databricks SQL and initial executive dashboards for core financial and operational KPIs.

### **Phase 1 (6 to 12 months): Customer intelligence and feedback**

- Build an AI ready Customer 360 that combines transactions, loyalty data and initial behavioral data for selected brands and regions.
- Implement the voice of customer GenAI use case to analyze reviews and service feedback and deliver recurring insight reports to product and customer experience teams.

### **Phase 2 (12 to 24 months): Personalization and supply chain**

- Deploy personalization models such as recommendations, propensity and churn into ecommerce, app and CRM channels with controlled A or B tests to quantify uplift.
- Roll out store level demand forecasting and inventory optimization and integrate models with replenishment and allocation processes.

## **Phase 3 (24 months and beyond): GenAI assistants and scaled automation**

- Launch GenAI assistants for customers and associates that use Mosaic AI and governed enterprise data.
- Expand automation to decisions such as promotion planning, markdown optimization and supplier negotiations in areas where models prove reliable and trusted.

Across all phases, the organization should invest in data literacy, governance practices and a cross functional operating model in which business, data and technology teams jointly own outcomes rather than separate initiatives.

## **8. Executive metrics and success criteria**

To ensure that Databricks investments translate into enterprise value, leadership should track a concise set of metrics that align with the use cases described above.

### **Customer and revenue metrics**

- Conversion rate by channel.
- Average order value and items per order.
- Customer lifetime value and retention rates.
- Personalization uplift in terms of incremental revenue and engagement versus control groups.
- NPS, CSAT and key complaint categories.

### **Supply chain and operations metrics**

- Forecast accuracy at store and SKU level.
- On shelf availability and stock out rates.
- Inventory turns and days of supply by category.

- Markdown and waste rates.
- Productivity gains for planners and store staff, for example time saved from automation.

## **Marketing efficiency metrics**

- Incremental sales generated per marketing dollar.
- ROAS by channel and campaign.
- Cost per acquisition by audience.
- Time required to produce attribution and performance insights for stakeholders.

## **Data and AI maturity metrics**

- Percentage of critical data domains managed under Unity Catalog.
- Number of ML and GenAI models in production that support key processes.
- Coverage of top processes, including forecasting, replenishment, recommendation and feedback analysis, by AI enabled solutions.
- Adoption of self service analytics and GenAI assistants by business users.

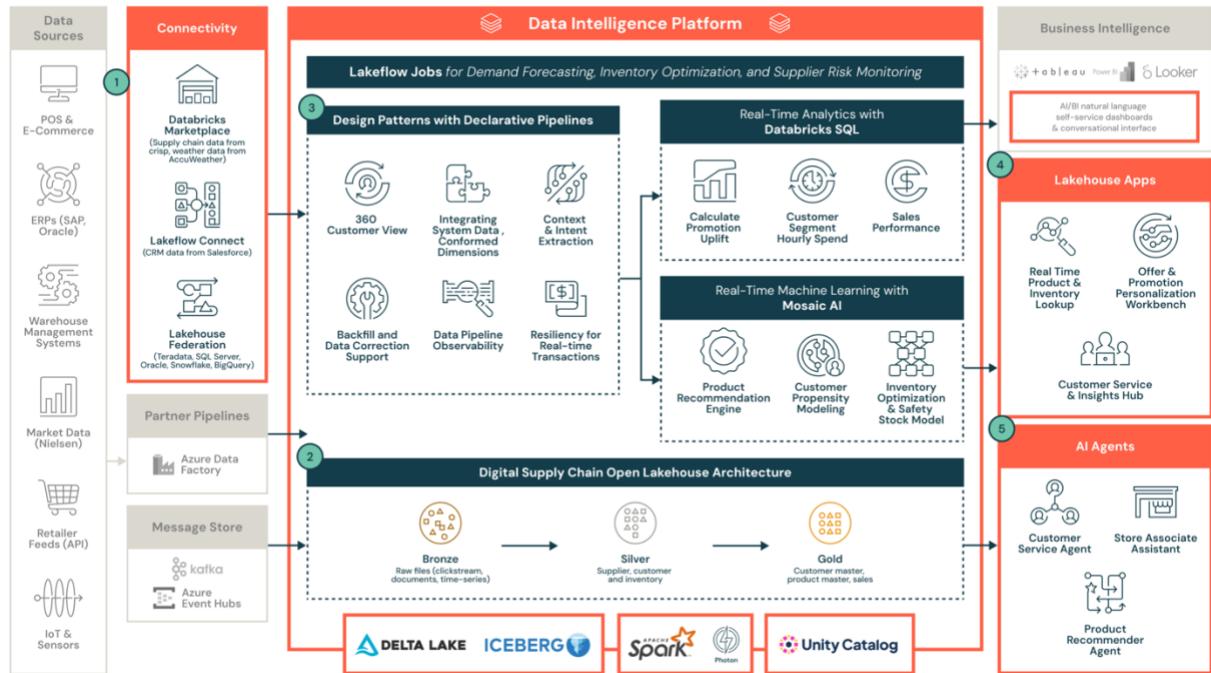
## **9. Conclusion**

Databricks provides an integrated foundation for building an AI ready operating model in retail that unifies data, governance, analytics and AI on a single platform. **By prioritizing use cases such as customer feedback intelligence, real time Customer 360, demand forecasting, marketing attribution, self-service analytics and GenAI assistants**, the retailer can unlock substantial improvements in customer satisfaction, revenue, margin and productivity.

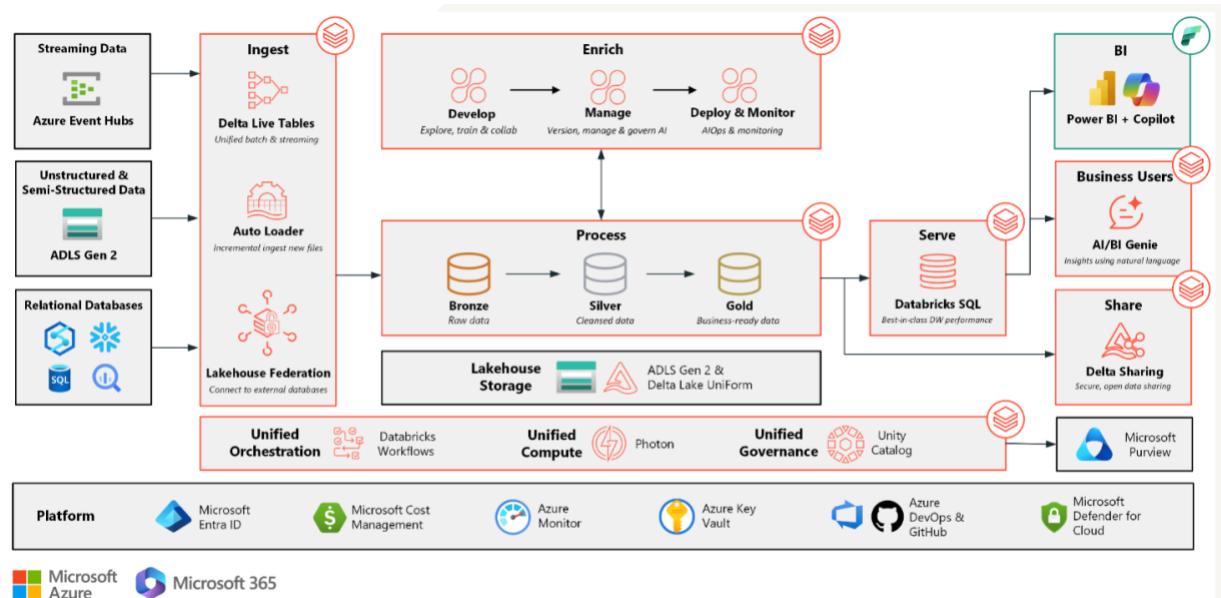
With a clear roadmap and disciplined measurement against the outlined metrics, Databricks can move from a technology choice to a core enabler of long term competitive advantage in an increasingly AI driven retail landscape.

## Additional: Pipeline examples

Generalized:



Using Azure Services:



References:

<https://www.databricks.com/solutions/industries/retail-industry-solutions>

<https://www.databricks.com/product/unity-catalog>

<https://www.databricks.com/customers/burberry-snowplow>

<https://www.databricks.com/customers/adidas>

<https://www.databricks.com/customers/walgreens>

<https://www.databricks.com/customers/hershey>

<https://www.databricks.com>