EBOOK

Collaborating Across the Retail Value Chain With Data and AI

By leveraging Lakehouse for Retail

databricks

Contents

databricks

INTRODUCTION

Retail needs real-time collaboration

Retail's digital transformation over the past 24 months has been remarkable, as e-commerce market penetration in the first 90 days of 2021 surpassed the aggregate 10 prior years, retailers transformed their businesses from brick-and-mortar locations to local stocking and delivery sites to reflect social distancing requirements, in addition demand forecasting or customer personalization algorithms had to be rewritten to reflect the new reality in supply chain and consumer preferences. Clearly digital transformation is reshaping retail.

When you consider that retailers today seek to improve customer experiences and optimize supply chains to build customer loyalty, most do that through use cases, such as fine-grained demand forecasting, safety stock analysis or foot traffic forecasting — based upon enterprise data. But this current capability really only considers what data can be accessed, at best, one arm's length away and many times it's not even real-time data. Databricks stepped back and realized that there are still fundamental issues that are not being addressed in this transformation.

The true promise of digital transformation is one of collaboration throughout the value chain — collaboration among manufacturers, retailers and the distribution channels — so that data, insights and innovation can be shared up and down the value chain, producing predictive insights and monetized returns for the ecosystem.

Collaboration allows use cases to be unshackled from current enterprise data and integrated across the value chain. Examples being:

- Fine-grained forecasting would include insight from manufacturers into their suppliers' supply chain constraint schedules or distribution centers' stocking and inventory levels
- Safety stock analysis for direct to consumer would be supported with not just the demand signals from their customers but also the upstream supply chain feeding this channel

Lakehouse for Retail is doing just this. It's an ecosystem and business solution approach enabling retailers, suppliers and partners to collaborate and innovate around data and AI. It eliminates the technical limitations that have constrained collaboration across the value chain and enables businesses to operate their business in real time, deliver more accurate analytics that leverage all their data, and drive collaboration and innovation across their value chain.

- For **retailers, manufacturers and channel partners,** it's an ecosystem solution that enables use cases addressing retail's most vexing challenges in customer insight, supply chain optimization, customer-centric merchandising or reimagining the role of stores by providing access to all data across the value chain with prebuilt accelerators that jump-start POC to profit and an ecosystem of partners that support your business transformation initiatives
- For **ISVs and SIs,** it's an open platform that allows the building solutions supporting use case deployment or intercommunication of data between companies

Welcome to Lakehouse for Retail

"One of the biggest challenges in retail today is the volatility in the supply chain. Acosta is invested in AI to help us improve our analytics in the supply chain, but we need to match that with better coordination with our partners. The Retail Lakehouse excites us because the open data sharing will enable us to collaborate with partners of all sizes to drive better execution."

Chad Pester, VP, Data and Analytics,
Sales and Marketing, Acosta

databricks

Retail Transformation Trends

Retailers are not sitting idly by in this environment. Today, successful retailers are addressing these challenges by leveraging fast and connected data from all corners of the enterprise. Four trends powering winners:

Embracing real-time retail

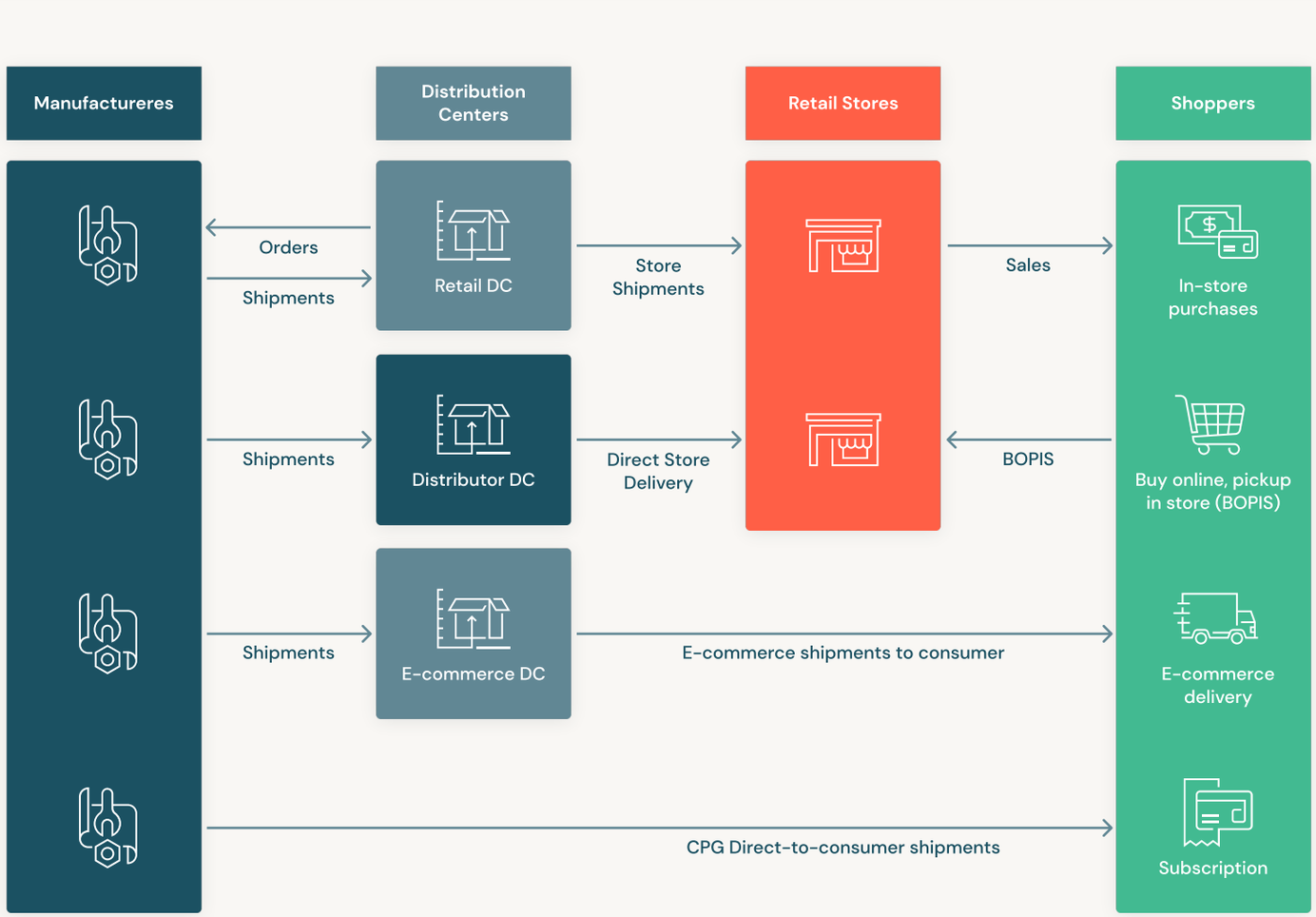
The meteoric rise of e-commerce has put pressure on brick-and-mortar stores to eliminate stock-outs, provide alternatives in order fulfillment pathways or next-best offers, and provide blended digital/live buying paths. Retailers are responding by leveraging real-time data input from social channels, customer reward programs, point of sale, enterprise resource planning, and inventory and logistics to develop a holistic real-time picture of the business, delivering higher granularity demand forecasting, stock availability and in-store merchandising.

Reimagining the importance of the consumer relationship

Consumers want to be connected — on their own terms. Data privacy regulations (GDPR, CCPA) and the elimination of third-party cookies have diminished retailers' insight, while the rise of e-commerce is allowing for consumers to be far more insightful in comparison shopping and, hence, less brand loyal. Re-establishing a direct and primary relationship with consumers has been a major priority for retailers and consumer goods companies for several years.

Retail Value Chain and Route to Market

Current State Insights



Addressing volatility in the supply chain

The biggest factor facing the retail environment at present is volatility and the risk it presents to the business. Inflation, a lack of employees to operate stores, warehouse capacity constraints, and delays in shipping are all pressing concerns in retail, but what makes this especially challenging is the uncertainty in understanding what is going to happen next.

Volatility in our economy has rendered most time-series models ineffective. Companies are exposed to significant risks and are making decisions based on instinct or are avoiding major decisions altogether.

This trend is expected to last for at least the next two years.

Addressing the need to unlock collaboration to foster innovation and growth

Retail has always operated a value chain built around collaboration. Distributors directly stocking stores, brokers resetting categories or partners performing audits on the store for price and display compliance. To address the environment's current volatility, there is further room for collaboration up and down the value chain:

- Manufacturers benefit from early knowledge of social media–driven demand forecasting signals concerning customer preferences
- Distribution centers benefit from early knowledge of external factors, such as weather, road construction, fleet maintenance and how it will affect product flow to, within and from warehouses
- The entire value chain becomes more agile with the knowledge of shifting customer preferences or buying channels

"84.51° continues to advance our data science capabilities by working with thousands of consumer-packaged-goods companies every day that are looking for new ways to interact with our data and science. The Retail Lakehouse will help bring our products and data closer to our customers and partners, driving better collaboration than ever. We see this as a competitive advantage for 84.51°."

Nick Hamilton
VP Engineering, 84.51°

Retail Data Challenges

Operating in real-time. Companies get into trouble when they make decisions ahead of the information. E-commerce, warehouse operations, plant production and more are turning traditional businesses into real-time businesses. Traditional data platforms, such as data warehouses, were not designed to operate in real time. Data silos within the business and between partners in the value chain prevent data access and holistic insight. Companies need to be able to both **rapidly ingest this at scale** and **make insights available across the value chain in real time.**

Delivering accurate analysis. The volatility in the market has eliminated any margin of error in demand forecasting, inventory, customer preference or product pricing. Poor decisions now lead to unprofitability. Legacy analytics tools perform analysis in aggregate or only analyze a portion of the problem. These legacy tools do not have the ability to scale out their compute and advanced analytic methods, resulting in an analysis that takes much longer than is allowed and is less accurate. These compromises on time and accuracy cost businesses money. Companies need the ability to eliminate the technical roadblocks so they can **perform fine-grained analysis** for **all products** and **within their tight service level agreements.**

Leveraging all your data. The majority of challenges that retail businesses have and are facing are externalities. The ability to monitor and test these externalities for their potential impact on business is essential to minimizing the risk through the business. The challenge? Legacy platforms are ill-suited or incapable of dealing with the types of data that manage these inputs. Companies need the ability to **ingest and analyze all types of data** — structured, unstructured and semi-structured — from any data source.

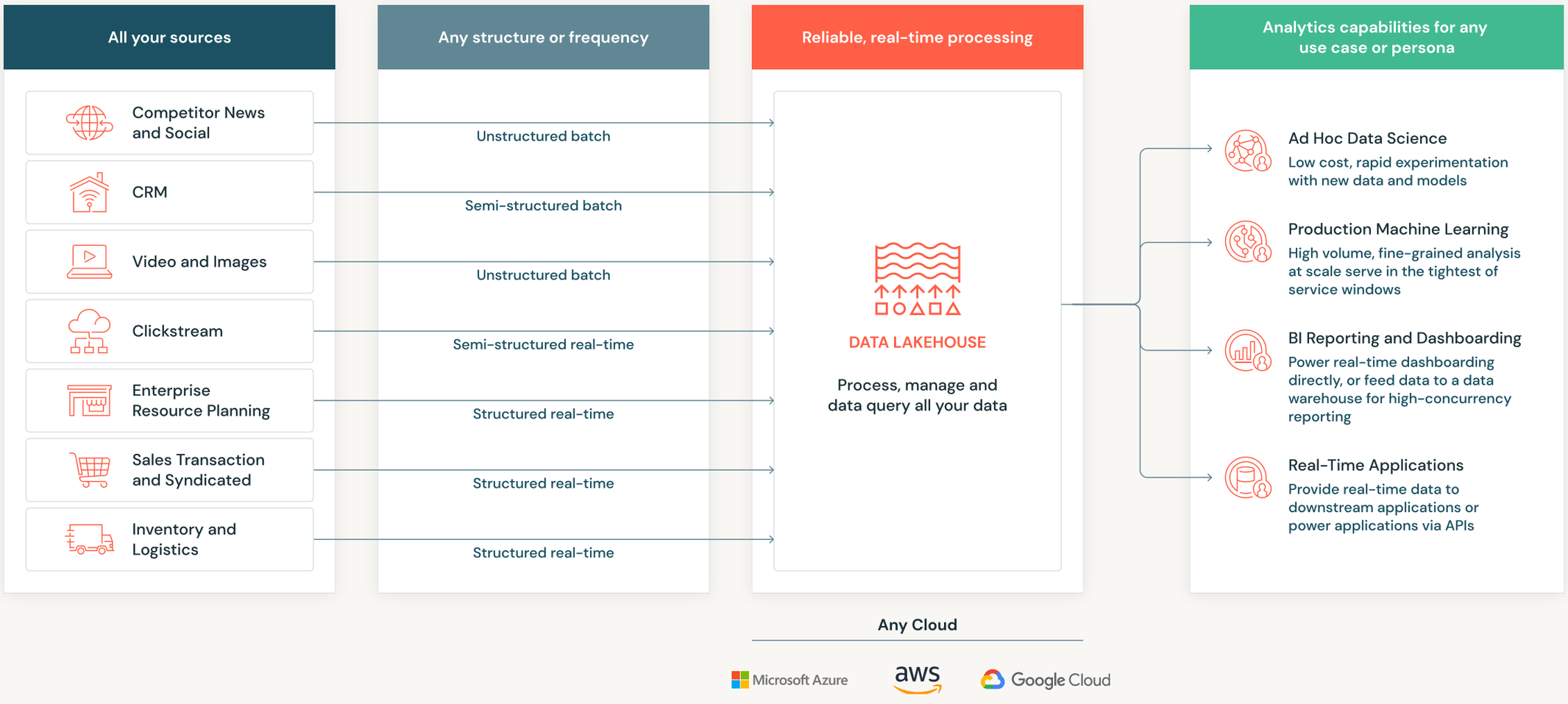
Unlocking data driven collaboration. The retail and consumer goods value chain has always been collaborative — with programs such as category captains, vendor-managed inventory, and vendor-managed replenishment bringing together manufacturers, distributors and retailers — but it has been limited to companies that can afford expensive systems for integration. Those existing systems are also limited, with most interactions happening in nightly or weekly batch processes. Out of the thousands of suppliers that call on a retailer, a few dozen can afford the expensive systems. This limitation leads to lower service levels and lower revenues.

Retail needs an **inexpensive and open method of collaboration** around data and analysis that opens interaction and innovation to all partners in the value chain. Leading retailers are looking at driving collaboration and innovation in three primary models:

- **White label.** Companies, often retailers and service providers, create analytics built around their view of the business. With white-label offerings, these companies would provide an environment in which customers could bring their own data to enrich the analysis and customize these analytics.
- **Clean room.** Retailers, suppliers, media agencies and service providers need the ability to securely and anonymously integrate data about consumers to drive greater consumer experiences
- **Data syndication.** Companies want easier and faster access to partner data. They look at data syndication as a way to streamline syndication without slow and complicated ETL, FTP, SSH and EDI systems.

The Lakehouse for Retail

Deliver better customer insights and more efficient supply chain in retail with Databricks



The Lakehouse for Retail is a platform that enables retailers, suppliers and partners to collaborate and innovate around data and AI. It eliminates the technical limitations that have constrained collaboration across the value chain and enables businesses to operate their business in real time, deliver more accurate analytics that leverage all their data, and drive collaboration and innovation across their value chain.

databricks

SOLUTIONS

Real Time for Retail

Operating in Real Time

Rapid data ingestion at scale makes advanced insights available across the value chain in real time, reducing costs and minimizing errors.

Retailers make mistakes when they make decisions without information. These mistakes can manifest themselves in many ways, including some of the following examples:

- Underestimating demand, leading to expedited shipping costs to rush delivery
- Incorrectly predicting how much of an item to produce, leading to excess carrying costs, missed sales and higher waste/scrap
- Reacting to breakdowns, leading to unplanned outages that disrupt entire production cycles
- Order fulfillment with incomplete or inaccurate data, leading to additional shipping costs or higher rates of return
- Missing an opportunity to engage the consumer based on current data, leading to missed sales opportunities

Processing data in real time enables all parts of the value chain to see the status of operations without delay and make better-informed decisions that help them avoid these problems.

Differentiated Capabilities

The Lakehouse uses technologies that include Delta, Delta Live Tables, Autoloader and Photon to enable customers to make data available for real-time decisions.

- The Lakehouse for Retail supports the largest of data jobs at near real-time intervals

  Example: Customers are bringing nearly 400 million events per day from transactional log systems at 15-second intervals

  Because of the disruption to reporting and analysis that happens during data processing, most retail customers load data to their EDW during a nightly batch. Some companies are even loading data weekly or later.

- The Lakehouse event-driven architecture provides a simpler method of ingesting and processing batch and streaming data than legacy approaches, such as lambda architectures. This architecture handles the change data capture and provides ACID compliance to transactions.
- Delta Live Tables simplifies the creation of data pipelines and automatically builds in lineage to assist with ongoing management
- The Lakehouse allows for true real-time stream ingestion of data, and even analytics on streaming data.

  Data warehouses require the extraction, transformation, loading and then additional extraction from the data warehouse to perform any analytics.

- Photon provides record-setting query performance, enabling users to query even the largest of data sets to power real-time decisions in BI tools

Common Use Cases

- Real-time dashboarding
- Personalization
- On-shelf availability
- Perpetual inventory
- Arrival time prediction
- Order picking and consolidation

SOLUTIONS

Fine-Grained Analysis

Companies need the ability to eliminate the technical roadblocks so **they can perform fine-grained analysis** for **all products** and **within their tight service level agreements**.

Data warehouses and legacy analytic platforms face two primary limitations when it comes to performing advanced analytics on large analytic jobs:

- They operate in serialized manners, meaning they calculate in a series. This limits how many calculations they can complete within a time period to the number of available threads.
- They lack support for advanced analytic techniques. They offer support for SQL. Some offer R or SAS embeddings, but they lack the basic functionality to perform robust analysis.

Because of these limitations, users of these systems make compromises on their analysis. They perform analysis at aggregate levels, they limit the number of items being analyzed, and they are limited on the robustness of the analytics being performed.

Improved accuracy of decisions, leading to higher revenue growth, greater incrementality and lower costs.

- By performing forecasting at finer grains of analysis and avoiding the assumptions implicit in aggregate forecasts, Databricks customers average a double-digit improvement in forecast accuracy over legacy approaches
- Companies that use Databricks distributed computing to run personalization models are able to more precisely tailor recommendations. In some categories, variance (error) has been reduced from 29% to 3% (71% accuracy to 97% accuracy).
- In time-sensitive analyses, the ability to distribute compute means that companies can analyze all items and not just high-priority items

Differentiated Capabilities

The Lakehouse uses technologies that include Delta, Databricks Managed Clusters, Hyperparameter optimization and MLflow, among others, to enable customers to perform fine-grained analysis.

- Delta provides optimized storage and querying that reduce the retrieval time of all types of data without the costly extraction that is required of data warehouse systems. This additional cost is reflected in the additional time required and incremental processing charges.
- Databricks provides users with the ability to fully distribute model calculations. The results of these analyses are captured and persisted in Delta for fast retrieval.
- Hyper-parameter optimization enables users to automatically test thousands of features for their contribution to model accuracy, leading to more accurate analysis
- Databricks clusters offer pre-loaded and tested libraries that enable companies to tap into a wide range of analytics
- MLflow provides a streamlined repository for tracking the results of experiments and managing the deployment of models

Common Use Cases

- Demand forecasting
- On-shelf availability
- Customer segmentation
- Personalization
- Pricing optimization

SOLUTIONS

360-Degree View of the Business

Companies need to **quickly and inexpensively use multimodal data** in their analysis.

Only 5%–10% of a company's data is structured. Tapping into the other 90% of data helps businesses better understand the environment around them and make better decisions.

Examples:

- Companies that leverage call center transcripts can better understand problems with the fulfillment or product quality
- Images and videos of items on shelf can be converted into useful information that enables companies to immediately identify out-of-stock, pricing variances, display compliance and other problems
- Images of production items can be used to identify high rates of errors or waste that can be used to optimize production processes
- Satellite imagery can be used to identify traffic patterns, drought, shipping delays and more. These can help in identifying potential disruptions to supply chains well in advance of other data.

Differentiated Capabilities

The Lakehouse supports the use of all types of structured, unstructured and semi-structured data types with Delta, Spark and partnerships with companies, including Labelbox and John Snow Labs to enable customers to take advantage of all data types.

- Unlike EDWs that store unstructured data as blobs, Delta enables companies to store data of all structures and immediately access that unstructured data to transform it into useful information
- Databricks partners with leading companies such as Labelbox, which enable companies to rapidly label images for use in computer vision projects
- Databricks partners with John Snow Labs, which provides enhanced NLP libraries to assist in the classification of text data

Common Use Cases

- ESG analysis
- Call center analytics
- Vendor compliance analysis (NLP)
- Human resources job screening
- Product quality
- Item availability
- Spill detection
- People tracking
- Planogram compliance
- Display compliance

SOLUTIONS

Stronger Collaboration Across the Value Chain

Retail needs an **inexpensive and open method of collaboration** around data and analysis that opens interaction and innovation to all partners in the value chain.

Retailers need to improve the speed of operations, build richer analytics and reduce the cost of collaboration across the value chain.

- Direct sharing via the Lakehouse eliminates the need for timely and costly ETL processes and batch-oriented data exchange processes making it easier to collaborate with partners around data will enable companies to spend less on data engineering and invest more resources toward innovation

Differentiated Capabilities

The Lakehouse provides capabilities, including Unity Catalog and Delta Sharing, which enable companies to manage and govern their data, assign rights and securely share data with partners.

Unity Catalog provides a holistic view of all data types. It integrates with existing data catalogs and allows companies to manage and govern their data.

Delta Sharing provides open sharing. Unlike legacy data warehouses which require all parties to be on the warehouse, Delta Sharing is an open standard and allows companies to securely share data with partners regardless of the platform they use.

Databricks white label provides companies with the potential to create shared analytics within their own environment that are accessed by partners and customers

Common Use Cases

Transformational use cases with partners within the value chain that rely upon collaborative data sharing. Such as:

- Demand forecasting
- Fulfillment optimization
- Single view of inventory
- Customer segment (profiles)
- In-stocks
- Price and markdown optimization
- Predictive maintenance
- Fleet optimization
- Network (logistics and DC) optimization

Customers That Choose Us

Retailers today seek to manage the disruptions across their supply chain and drive stronger customer experiences through the use of data and AI.

The Retail Lakehouse is a platform that enables retailers, suppliers and partners to collaborate and innovate around data and AI. It eliminates the technical limitations that have constrained collaboration across the value chain and enables businesses to operate their business in real time, deliver more accurate analytics that leverage all their data, and drive collaboration and innovation across their value chain.

databricks

About Databricks

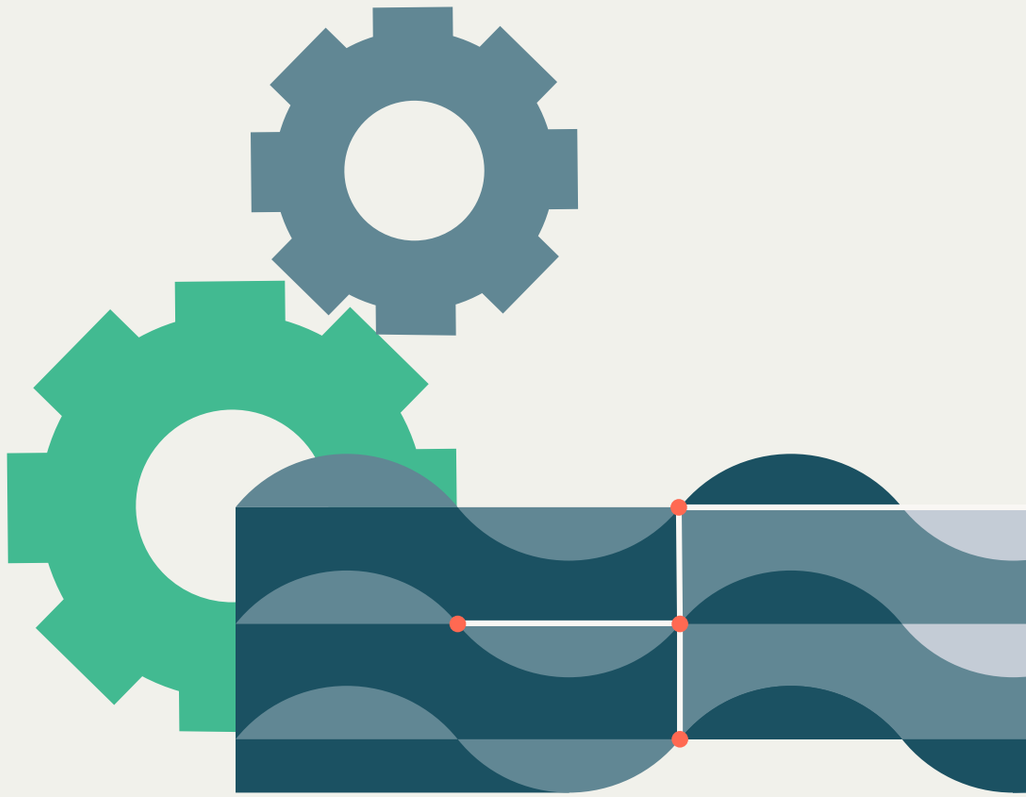
Databricks is the data and AI company. More than 5,000 organizations worldwide — including Comcast, Condé Nast, Acosta and over 40% of the Fortune 500 — rely on the Databricks Lakehouse Platform to unify their data, analytics and AI. Databricks is headquartered in San Francisco, with offices around the globe. Founded by the original creators of Apache Spark,™ Delta Lake and MLflow, Databricks is on a mission to help data teams solve the world's toughest problems. To learn more, follow Databricks on Twitter, LinkedIn and Facebook.

Get started with a free trial of Databricks and start building data applications today

START YOUR FREE TRIAL

To learn more, visit us at:

Retail Industry Solutions



databricks