eBook

Big Book of Retail & Consumer Goods Use Cases

Driving real-time decisions with the Lakehouse

databricks

Contents (1/2)

databricks

Contents (2/2)

databricks

Introduction

Retailers are increasingly being challenged to make time-sensitive decisions in their operations. Consolidating e-commerce orders. Optimizing distribution to ensure item availability. Routing delivery vehicles. These decisions happen thousands of times daily and have a significant financial impact. Retailers need real-time data to support these decisions, but legacy systems are limited to data that's hours or days old.

When seconds matter, only the Lakehouse delivers better decisions

Retail is a 24/7 business where customers expect accurate information and immediate relevant feedback. The integration of physical and e-commerce customer experiences into an omnichannel journey has been happening for the past 20 years, but the pandemic provided a jolt to consumer trends that dramatically shifted purchasing patterns.

In reaction to these industry changes, retailers have responded with significant, rapid investments — including stronger personalization, order fulfillment, and delivery and loyalty systems. While these new targeted capabilities have addressed the immediate need — and created expectations of making decisions in real time — most retailers still rely on legacy data systems, which impedes their ability to scale these innovations. Unfortunately, most legacy systems are only able to process information in hours or days.

The delays caused by waiting for data are leading to significant risks and costs for the industry.

- **Grocers** need to consolidate order picking to achieve profitability in e-commerce, but this requires up-to-the-minute order data. Not having this information causes them to spend more resources on having people pick orders separately, at a higher operating cost.

- **Apparel retailers** must be able to present the correct available inventory on their website. This requires that in-store sales be immediately reflected in their online systems. Inaccurate information can lead to lost sales, or worse, the customer becoming unsatisfied and moving to different retailers.

databricks

Lakehouse

- **Convenience fuel retailers** must collaborate with distribution centers, direct-to-store delivery distributors and other partners. Having delayed data can lead to out-of-stocks, costing stores thousands of dollars per week.

The margin of error in retail has always been razor thin, but with a pandemic and inflationary pressures, it's at zero. Reducing the error rate requires better predictions and real-time data.

Use Case Guide

In this use case guide, we show how the Databricks Lakehouse for Retail is helping leading organizations take **all of their data in a single lakehouse architecture, streamline their data engineering and management, make it ready for SQL and ML/AI**, and **do so very fast within their own cloud infrastructure environment based on open source and open standards**. These capabilities are all delivered at world-record-setting performance, while achieving a market-leading total cost of ownership.

Databricks Lakehouse for Retail has become the industry standard for enabling retailers to drive decisions in real time. This use case guide also highlights common use cases across the industry, and offers additional resources in the form of Solution Accelerators and reference architectures to help as you embark on your own journey to drive better customer experiences with data and AI.

databricks

Modern Data Platform for Real-Time Retail

Retailers continue to adapt to rapidly shifting dynamics across the omnichannel. In navigating these changes, retailers are increasingly focused on improving the real-time availability of data and insights, and performing advanced analytics delivered within tight business service windows.

Common challenges

In response to the surge in e-commerce and volatility in their supply chains, retailers are investing millions in modernizing distribution centers, partnering with delivery companies, and investing in customer engagement systems.

Warehouse automation is expected to become a $41B market according to Bloomberg. Increasingly, distribution centers are being automated with robotics to power dynamic routing and delivery. Shoppers that became accustomed to having fast, same-day, and sometimes even overnight delivery options during the pandemic now expect them as the norm. Retailers understand that the shipping and delivery experience is now one of many touchpoints that merchants can use to develop customer brand loyalty.

$41B Market | Retail Warehouse Automation

Bloomberg

Yet while retailers modernize different areas of their operations, they're constrained by a single point of weakness, as they are reliant on legacy data platforms to bring together all of this data.

Powering real-time decisions in modern retail requires real-time ingestion of data, transformation, governance of information, and powering business intelligence and predictive analytics all within the time required by retail operations.

databricks

- **Ingesting large volumes of transactional data in real time.** The biggest blocker to crucial insights is the ability to ingest data from transaction systems in real time. Transaction logs from point-of-sale systems, clickstreams, mobile applications, advertising and promotions, as well as inventory, logistics and other systems, are constantly streaming data. Big data sets need to be ingested, cleansed and aggregated and integrated with each other before they can be used. The problem? Retailers have used legacy data warehouses that are built around batch processing. And worse, increasing the frequency of how often data is processed leads to a "hockey stick" in costs. As a result of these limitations, merchants resort to ingesting data nightly to deal with the large volumes of data and integration with other data sets. The result? Accurate data to drive decisions can be delayed by days.

- **Performing fine-grained analysis at scale within tight time windows.** Retailers have accepted a trade-off when performing analysis. Predictions can be detailed and accurate, or they can be fast. Running forecasts or price models at a day, store and SKU level can improve accuracy by 10% or more, but doing so requires tens of millions of model calculations that need to be performed in narrow service windows. This is well beyond the capability of legacy data platforms. As a result, companies have been forced to accept the trade-off and live with less accurate predictions.
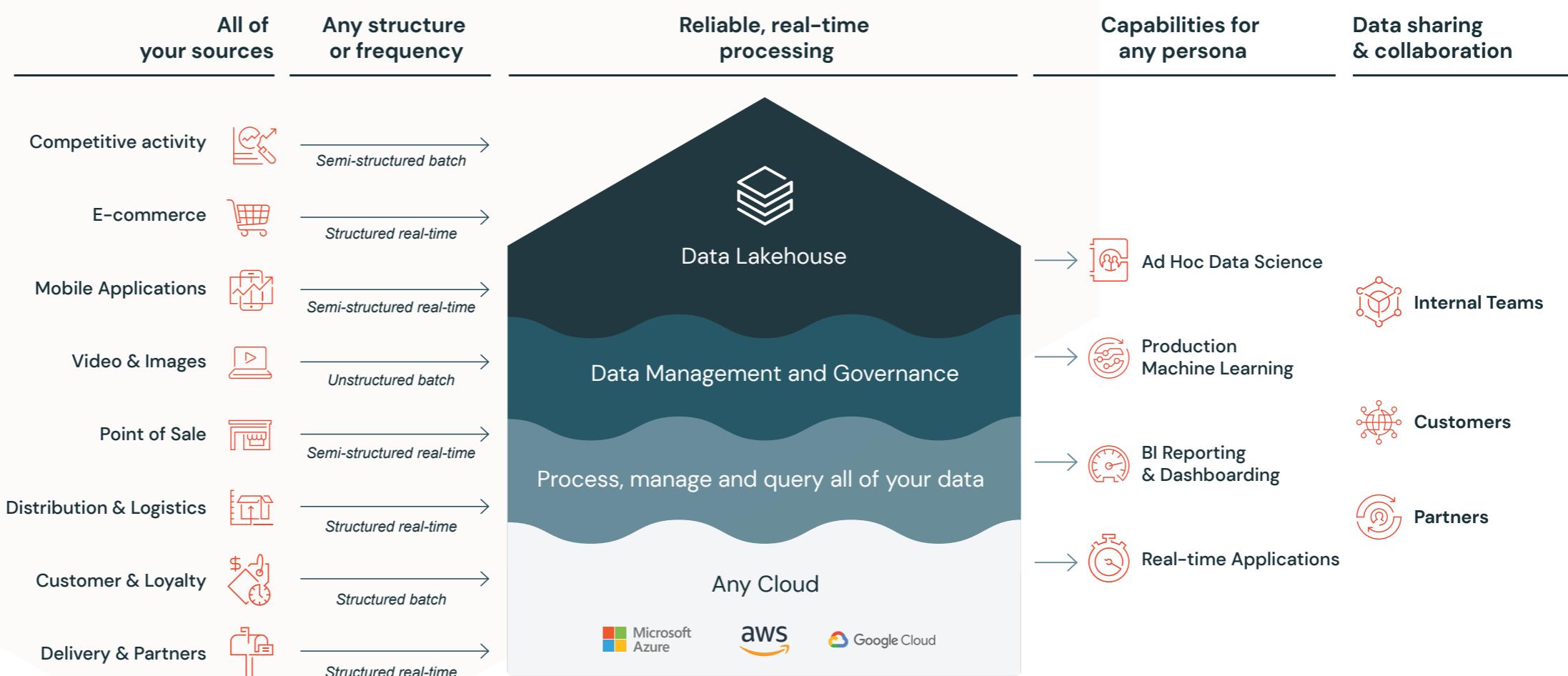
- **Powering real-time decisions on the front line.** Data is only useful if it drives decisions, but serving real-time data to thousands of employees is a daunting task. While data warehouses are capable of serving reports to large groups of users, they're still limited to stale data. Most retailers limit the frequency of reports to daily or weekly updates and depend on the staff to use their best judgment for decisions that are more frequent.

- **Delivering a hyper-personalized omnichannel experience.** The storefront of the 21st century is focused on delivering personalized experiences throughout the omnichannel. Retailers have access to a trove of customer data, and yet off-the-shelf tools for personalization and customer segmentation struggle to deal with high volumes, and the analytics have high rates of inaccuracy. Retailers need to deliver personalized experiences at scale to win in retail.
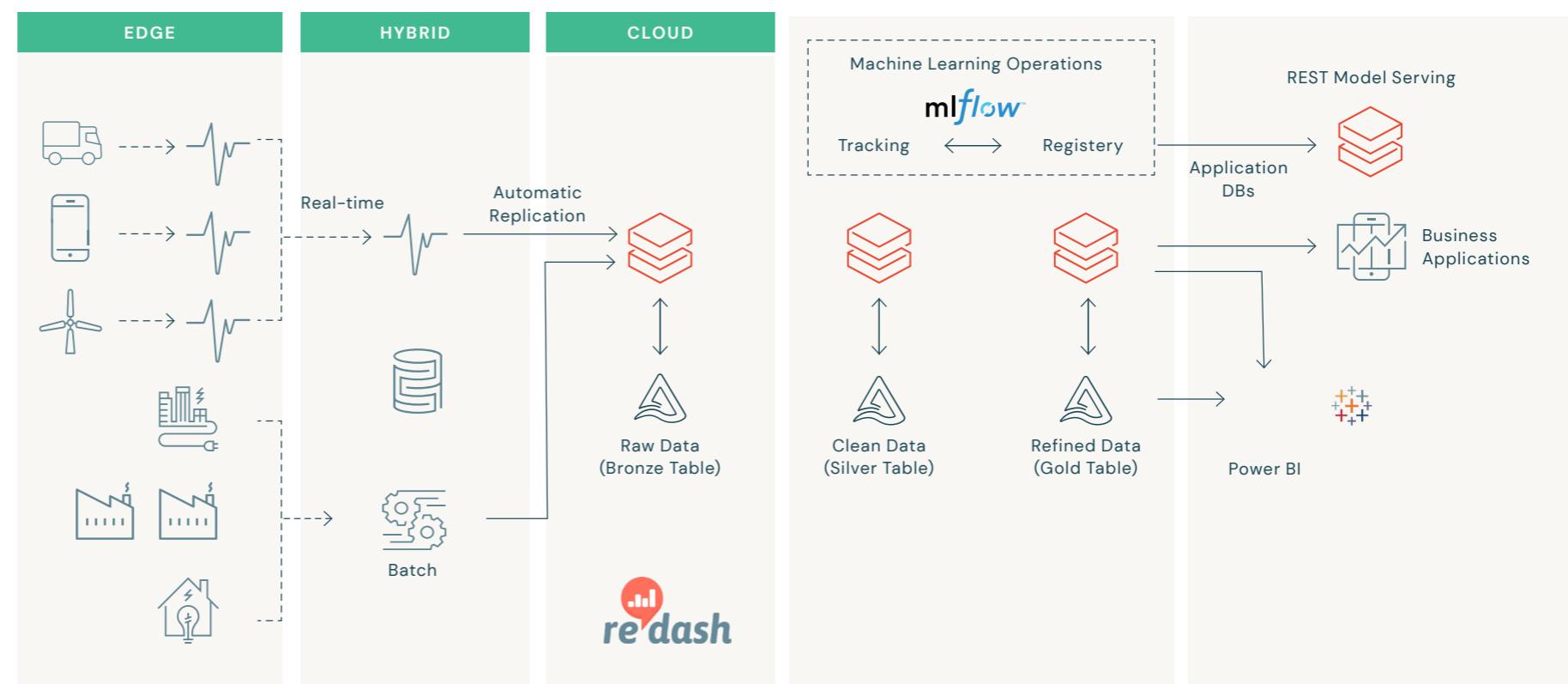
databricks

The Lakehouse for Retail

Databricks Lakehouse for Retail solves these core challenges. The Lakehouse unlocks the ability to unify all types of data — from images to structured data — in real time, provide enterprise-class management and governance, and then immediately turn that data into actionable insights with real-time reporting and predictive analytics. It does this with record-setting speed and industry-leading total cost of ownership (TCO) in a platform-as-a-service (PaaS) that allows customers to solve these pressing problems.

All of your sources	Any structure or frequency	Reliable, real-time processing	Capabilities for any persona	Data sharing & collaboration
Competitive activity	Semi-structured batch	**Data Lakehouse**	Ad Hoc Data Science	Internal Teams
E-commerce	Structured real-time	**Data Management and Governance**	Production Machine Learning	Customers
Mobile Applications	Semi-structured real-time	Process, manage and query all of your data	BI Reporting & Dashboarding	Partners
Video & Images	Unstructured batch	Any Cloud	Real-time Applications	
Point of Sale	Semi-structured real-time	Microsoft Azure aws Google Cloud		
Distribution & Logistics	Structured real-time			
Customer & Loyalty	Structured batch			
Delivery & Partners	Structured real-time			

databricks

Reference Architecture

At the core of the Databricks Lakehouse for Retail is technology that enables retailers to avoid the trade-offs between speed and accuracy. Technology such as Delta Lake enables the Lakehouse — a new paradigm that combines the best elements of data warehouses and data lakes — to directly address these factors by enabling you to unify all of your data — structured and unstructured, batch and real-time — in one centrally managed and governed location. Once in the Lakehouse, e-commerce systems, reporting users, analysts, data scientists and data engineers can all leverage this information to serve models for applications and power real-time reporting, advanced analytics, large-scale forecasting models and more.

How it works

The Lakehouse for Retail was built from the ground up to solve the needs of modern retail. It blends simplicity, flexibility and lower cost of ownership with best–in–industry performance. The result is differentiated capabilities that help retailers win.

	Data in real time	Use all of your data	Robust data management	Real–time reporting	Time-sensitive machine learning
LEGACY DATA WAREHOUSE	**No.** Data warehouses are batch oriented, restricting data updates to hours or days.	**No.** Data warehouses have very limited support for unstructured data.	**Limited.** EDWs support the management of structured data.	**No.** EDWs offer quick access to reports on old data.	**No.** EDWs must extract data and send it to a third party for machine learning.
DATA LAKES (HADOOP)	**No.** Data lakes are batch oriented.	**Yes.** Data lakes offer support for all types of data.	**No.** Data lakes lack enterprise–class data management tools.	**No.** Data lakes were not designed for reporting, let alone real–time reporting.	**Yes.** Data views can be materialized, enabling front-line employees with real-time data.
ROBUST DATA	**Yes.** Support for real–time streaming data.	**Yes.** Supports all types of data in a centrally managed platform.	**Yes.** Delta and Unity Catalog offer native data management and governance of all data types.	**No.** Data lakes are able to support large analytics, but lack the ability to meet business SLAs.	**Yes.** The Lakehouse can scale to process the most demanding predictions within business SLAs.

databricks

- **Data in real time.** Retail operates in real time and so should your data. The Lakehouse offers support for streaming data from clickstream, mobile applications, IoT sensors and even real-time e-commerce and point-of-sale data. And Delta Lake enables this world-record-leading performance while maintaining support for ACID transactions.

- **Use all of your data.** Retailers are increasingly capturing data from mobile devices, video, images and a growing variety of other data sources. This data is extremely powerful in helping to improve our understanding of consumer behavior and operations. The Lakehouse for Retail enables companies to take full advantage of all types of data in a cost-efficient way, in a single unified lakehouse architecture.

- **Robust data management and governance** that companies need to protect sensitive data, but was lacking from earlier big data systems. The Lakehouse offers transactional integrity with ACID compliance, detailed data security, schema enforcement, time travel, data lineage and more. Moving to a modern data architecture does not require sacrificing enterprise maturity.

- **High-performance predictive analytics.** Machine learning models, such as demand forecasting or recommendation engines, can be run in hours without compromising accuracy. The Lakehouse can scale to support tens of millions of predictions in tight windows, unlocking critical and time-sensitive analytics such as allocating inventory, optimizing load tenders and logistics, calculating item availability and out-of-stocks, and delivering highly personalized predictions.

Value with Databricks

By using Databricks to build and support your lakehouse, you can empower your business with even more speed, agility and cost savings. The flexibility of the Databricks Lakehouse Platform means that you can start with the use case that will have the most impact on your business. As you implement the pattern, you will find that you're able to tackle use cases quicker and more easily than before. To get you started, this guidebook contains the use cases we most commonly see across the Retail and Consumer Goods industry.

databricks
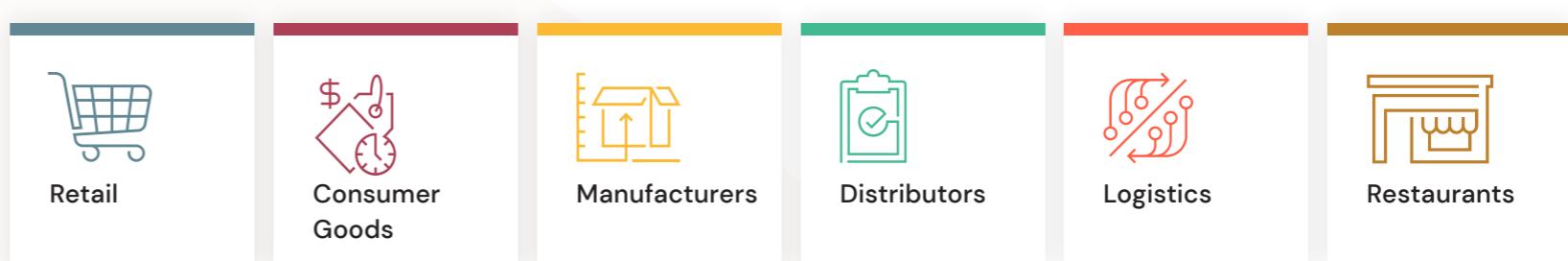
CHAPTER 3

Use Case:
Real–Time Supply
Chain Data

Overview

As companies see a surge in demand from e-commerce and delivery services, and seek increasing efficiencies with plant or distribution centers, real-time data is becoming a key part of the technical roadmap. Real-time supply chain data allows customers to deal with problems as they happen and before items are sent downstream or shipped to consumers, which is the first step in enabling a supply chain control tower.

RELEVANT FOR

Retail	Consumer Goods	Manufacturers	Distributors	Logistics	Restaurants

Challenges

- **Batch data** — existing data warehouses bring data in batch, creating a lag between when something is happening and when a customer can act on it

- **Complex analysis in real time** — if ingesting data in real time wasn't a big enough challenge, companies have the added pressure to take immediate action on it

- **Complex maintenance** — ETL tools to bring data in batch are often complex and costly to maintain

databricks

Value with the Databricks Lakehouse

Databricks has enabled real-time streaming of supply chain data across a variety of customers for specific plant operations or as part of a supply chain control tower.

- **Near real-time ingestion and visibility of data** — one customer experienced a 48,000% improvement in speed to data, with greater reliability

- **Cost-neutral** — because Delta's efficient engine requires smaller instances, many customers report that they were able to move from batch to real-time at neutral costs

- **Simplified architecture and maintenance** — leveraging Delta for ingestion streamlines the pattern for real-time data ingestions. Customers frequently report that the amount of code required to support streaming ingestion is 50% less than previous solutions.

- **Immediate enablement of additional use cases** — customers can now prevent problems as they're happening, predict and prevent issues, and even gain days on major changes such as production schedules between shifts

Solution overview

Databricks allows for both streaming and batch data sets to be ingested and made available to enable real-time supply chain use cases. Delta Lake simplifies the change data capture process while providing ACID transactions and scalable metadata handling, and unifying streaming and batch data processing. And Delta Lake supports versioning and enables rollbacks, full historical audit trails, and reproducible machine learning experiments.

Typical use case data sources include:

Supply planning, procurement, manufacturing execution, warehousing, order fulfillment, shop floor/historian data, IoT sensor, transportation management

databricks

HOW TO GET STARTED
Contact your Databricks account team to have them perform a free proof-of-concept with your real-time data.

CASE STUDY gousto

With Databricks, Gousto was able to implement real-time visibility in their supply chain. Gousto moved from daily batch updates to near real-time streaming data, utilizing Auto Loader and Delta Lake. The platform provided by Databricks has allowed Gousto to respond to increased demand during the coronavirus outbreak by providing real-time insight into performance on the factory picking lines.

CASE STUDY BUTCHERBOX™
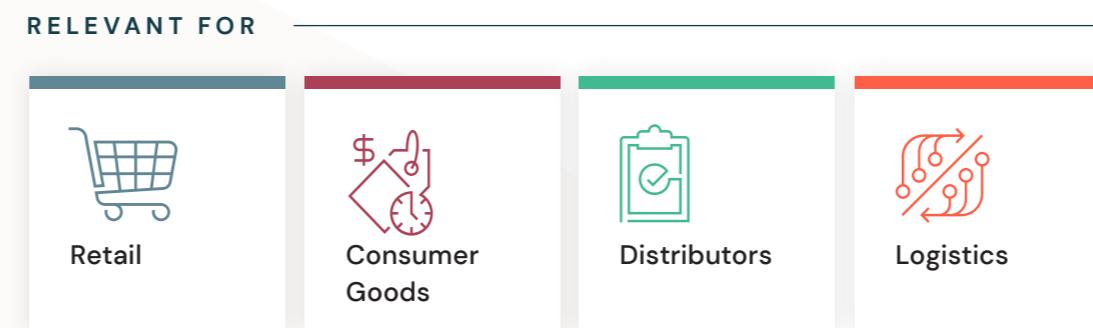
As a young e-commerce company, ButcherBox needed to act nimbly to make the most of the data from its hundreds of thousands of subscribers. With Databricks Lakehouse, the company could pull 18 billion rows of data in under three minutes.

Now, ButcherBox has a near real-time understanding of its customers, and can also act proactively to address any logistical and delivery issues.

databricks

CHAPTER 4
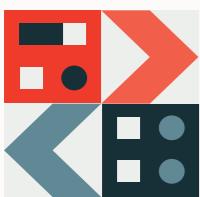
Use Case:
Truck Monitoring

With many industries still feeling the effects of supply chain issues, being able to increase the efficiency of trucks on the road can make all the difference in getting goods into the hands of customers in a timely manner. Real-time data is making it easier for companies to get immediate insights into truck manufacturing delays, maintenance issues, supply chain issues, delivery schedules and driver safety.

RELEVANT FOR

Retail	Consumer Goods	Distributors	Logistics

Challenges

- Siloed data makes it difficult to get a comprehensive understanding of fleet performance

- A lack of real-time insights can delay responses to manufacturing or supply chain issues

- Not having effective automation and AI increases the risk of human error, which can result in vehicular accidents or shipment delays

databricks

Value with the Databricks Lakehouse

Databricks empowers companies to get real-time insights into their fleet performance, from manufacturing to delivery.

- **Near real-time insights** — the greater speed to data means a quicker response to issues and the ability to monitor driver safety more immediately

- **Ability to scale** — although consumer demands are constantly evolving, Databricks can handle fleet expansion without sacrificing data quality and speed

- **Optimizing with AI/ML** — implementing AI and ML models can lead to more effective route monitoring, proactive maintenance and reduced risk of accidents

Solution overview

Databricks enables better truck monitoring, quickly ingesting data on everything from vehicle manufacturing to route optimization. This results in a more complete and real-time view of a company's fleet, and these analytics provide companies with the tools they need to scale and improve their operations.

Typical use case data sources include:

Supply planning, transportation management, manufacturing, predictive maintenance

HOW TO GET STARTED

Contact your Databricks account team to have them perform a free proof-of-concept with your real-time data.

CASE STUDY   **EMBARK**

With 94% of vehicular accidents attributed to human error, Embark used the Databricks Lakehouse Platform to unlock thousands of hours of recorded data from its trucks and then collaboratively analyze that data via dashboards. This has resulted in more efficient ML model training as Embark speeds toward fully autonomous trucks.

databricks
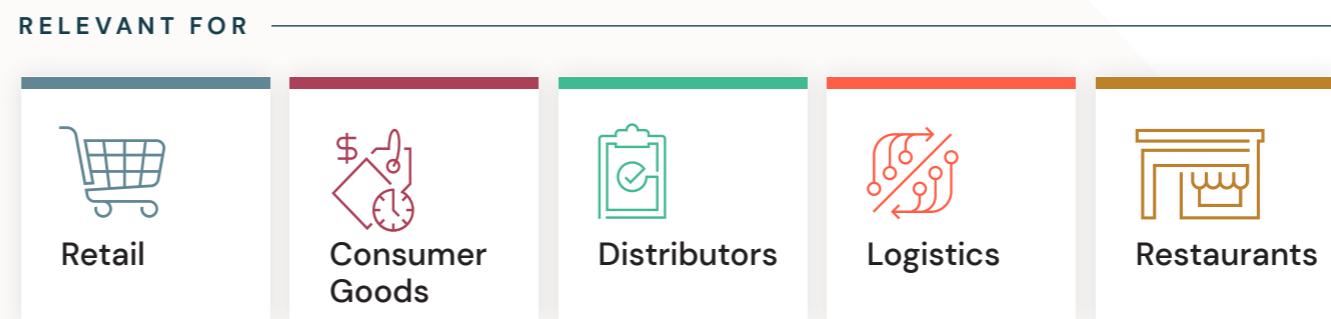
Use Case:
Inventory Allocation

Overview

Replenishment planning is the process of determining what needs to go where. It is used by replenishment planning, distributors and consumer goods companies performing vendor-managed replenishment (VMR) or vendor-managed inventory (VMI) to make daily decisions on which product needs to be sent to which store and on what day.

Replenishment is challenging for companies because it deals with rapidly changing data and the need to make complex decisions on that data in narrow service windows. Retailers need to stream in real-time sales data to signal how much of a product has been sold in order. Inaccurate sales data leads to an insufficient number of products being sent to stores. This results in lost sales and low customer satisfaction.

Inventory allocation is a process that might be performed multiple times a day during peak seasons, or daily during slower seasons. Companies need the ability to scale to perform tens of millions of predictions multiple times a day — on demand and dynamically — during peak season without paying a premium for this capability throughout the year.

RELEVANT FOR

Retail	Consumer Goods	Distributors	Logistics	Restaurants

databricks

Challenges

- Customers must complete tens of millions of inventory allocation predictions within tight time windows. This information is used to determine which products get put on trucks and go to specific stores.

- Traditional inventory allocation rules cause trade-offs in accuracy in order to calculate all possibilities in the service windows

- Legacy tools have rudimentary capabilities and have limited ability to consider flavors, sizes and other attributes that may be more or less popular by store

Value with Databricks
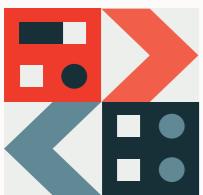
Customers are able to complete inventory allocation models within SLAs with no trade-off for accuracy.

- **Speed —** on average, customers moving to Databricks for demand forecasting report a double-digit improvement in forecast accuracy

- **Ability to scale** and perform fine-grained (day, store, item) level allocations

- **Provide more robust allocations** by incorporating causal factors that may increase demand, or include information on flavors or apparel sizes for specific stores

Solution overview

The objective of inventory allocation is to quickly determine when to distribute items and where — from warehouses and distribution centers to stores. Inventory allocation begins by looking at the consumption rate of products, the available inventory and the shipping schedules, and then using this information to create an optimized manifest of what items should be carried on which trucks, at what point, and at what time. This becomes the plan for route accounting systems that arrange deliveries.

Inventory allocation also deals with trade-offs related to scarcity of items. If an item has not been available in a store for a long time, that store may receive heightened priority for the item in the allocation.

databricks

Typical use case data sources include: point of sale, digital sales, replenishment data, modeled safety stock, promotions data, weather

View our webinar covering demand forecasting with Starbucks and then read our blog about demand forecasting.

Demand forecasting with causal factors.

Our most popular notebook at Databricks. This blog walks you through the business and technical challenges of performing demand forecasting and explains how we approached solving it.

On-demand webinar for demand forecasting.

Video and Q&A from our webinar with Starbucks.

HOW TO GET STARTED

Contact your Databricks account team to have them perform a free proof-of-concept with your real-time data.

CASE STUDY *H&M*

H&M turned to the Databricks Lakehouse Platform to simplify its infrastructure management, enable performant data pipelines at scale, and simplify the machine learning lifecycle. The result was a more data-driven organization that could better forecast operations to streamline costs and boost revenue.

CASE STUDY edmunds

Edmunds is on a mission to make car shopping an easy experience for all. With the Databricks Lakehouse Platform, they are able to simplify access to their disparate data sources and build ML models that make predictions off data streams. With real-time insights, they can ensure that the inventory of vehicle listings on their website is accurate and up to date, improving overall customer satisfaction.

databricks

CHAPTER 6

Use Case:
Point of Sale
and Clickstream

Overview
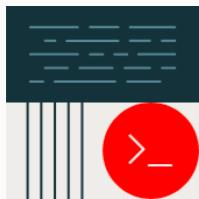
Disruptions in the supply chain — from reduced product supply and diminished warehouse capacity — coupled with rapidly shifting consumer expectations for seamless omnichannel experiences are driving retailers to rethink how they use data to manage their operations. Historically, point-of-sale (POS) systems recorded all in-store transactions, but were traditionally kept in a system that was physically in the store. This would result in a delay in actionable insights. And now with consumers increasingly shopping online, it's crucial to not only collect and analyze that clickstream data quickly, but also unify it with POS data to get a complete and real-time snapshot of each customer's shopping behavior.

Near real-time availability of information means that retailers can continuously update their estimates of item availability. No longer is the business managing operations based on their knowledge of inventory states as they were a day prior, but instead is taking actions based on their knowledge of inventory states as they are now.

RELEVANT FOR →

Retail	E-commerce

Challenges

- Retailers with legacy POS systems in their brick-and-mortar stores are working with siloed and incomplete sales data

- Both POS and clickstream data need to be unified and ingested in real time

databricks

Value with Databricks

Databricks brings POS and clickstream data together for a unified data source that leads to real-time insights and a clearer understanding of customer behavior.

- **Single source of truth** — a centralized, cloud-based POS system means it can be merged with clickstream data

- **Near real-time insights** — the greater speed to data means businesses get the latest insights into customer purchasing behaviors and trends

- **Scalability** — companies can scale with Databricks to handle data from countless transactions

databricks

Use Case:
On–Shelf Availability

Overview

Ensuring the availability of a product on shelf is the single largest problem in retail. Retailers globally are missing out on nearly $1 trillion in sales because they don't have on hand what customers want to buy in their stores. Shoppers encounter out–of–stock scenarios as often as one in three shopping trips. All told, worldwide, shoppers experience $984 billion worth of out–of–stocks, $144.9 billion in North America alone, according to industry research firm IHL.
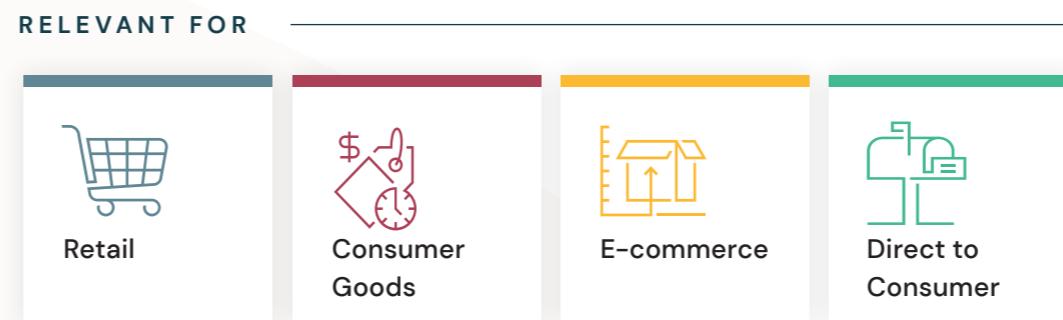
In the past, if a customer faced an out–of–stock, they would most likely select a substitute item. The cost of going to another store prevented switching. Today, e–commerce loyalty members, such as those who belong to Walmart+ and Amazon Prime, are 52% more likely than other consumers to purchase out–of–stock items online. It is believed that a quarter of Amazon's retail revenue comes from customers who first tried to buy a product in–store. In all, an estimated $36 billion is lost to brick–and–mortar competition, and another $34.8 billion is lost to Amazon or another e–retailer, according to IHL.

On–shelf availability takes on a different meaning in pure e–commerce applications. An item can be considered in stock when it is actually in a current customer's basket. If another customer places the same item in their basket, there is the possibility that the first customer will purchase the last available item before the second customer. This problem is exacerbated by retailers who use stores to keep inventory. In these situations, customers may order an item that is picked for delivery at a much later time. The window between ordering and picking creates the probability of out–of–stocks.

On–shelf availability predicts the depletion of inventory by item, factors in safety stock levels and replenishment points, and generates a signal that suggests an item may be out of stock. This information is used to generate alerts to retail staff, distributors, brokers and consumer goods companies. Every day, tens of thousands of people around the world do work that is generated by these algorithms.

The sheer volume of data used to calculate on–shelf availability prevents most companies from analyzing all of their products. Companies have between midnight and 4 AM to collect all of the needed information and run these models, which is beyond the capability of legacy data systems. Instead, companies choose the priority categories or products to analyze, which means a significant percentage of their unavailable products will not be proactively addressed.

databricks

One of the biggest challenges with on-shelf availability is determining when an item is actually out of stock. While some retailers are investing in computer vision and robots, and others employ the use of people to manually survey item availability, most retailers default to a signal of determining when an item has not been scanned in an acceptable time.

RELEVANT FOR

| Retail	Consumer Goods	E-commerce	Direct to Consumer
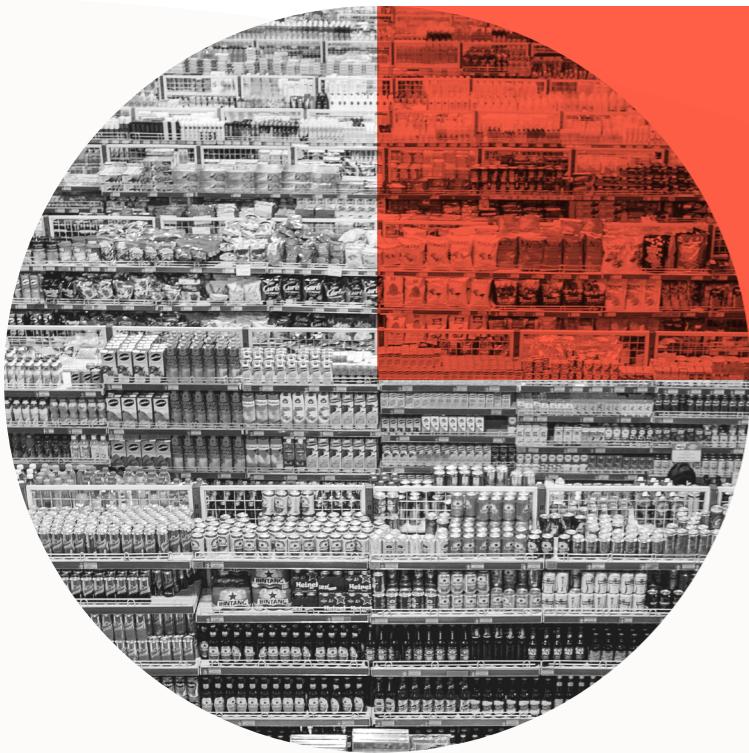
Challenges

The biggest challenge to generating on-shelf availability alerts is time. Companies may receive their final sales data from the preceding day shortly after midnight. They have less than 4 hours from that point to ingest large volumes of t-log data and calculate probabilities of item availability. Most firms are encumbered by a data warehouse process that only releases data after it has been ingested and aggregates have been calculated, a process that can require multiple hours per night.

For this reason, most firms make sacrifices in their analysis. They may alternate categories they analyze by different days, prioritize only high-impact SKUs, or run analysis at higher-level and less-accurate aggregate levels. Among the challenges:

- Processing large volumes of highly detailed data and running millions of models in a narrow time window

- Companies are spending hundreds of thousands of dollars annually to generate these daily alerts for a few categories

- Dealing with false positives and negatives in predictions

- Distributing information quickly and efficiently to internal systems and external partners

databricks

Value with Databricks

Databricks enables customers to generate on-shelf availability (OSA) predictions at scale with no compromises.
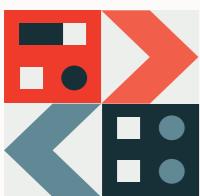
- Delta removes the data processing bottleneck. Delta enables retailers to stream in real time or to batch process large volumes of highly detailed and frequently changing point-of-sale transaction data.

- Easily scale to process all OSA predictions within tight service windows using Apache Spark™

- Manage features and localize models with additional causal data to improve accuracy with MLflow

- Easily deploy information via streams, through API for mobile applications or partners, or to Delta for reporting

- Enable retailers to monetize their data by directly licensing OSA alerts

Solution overview

Databricks enables companies to perform on-shelf availability analysis without making compromises to the breadth or quality of predictions.

It begins with Delta Lake — a nearly perfect platform for ingesting and managing t-log data. One of the biggest challenges in t-log data is the frequent number of changes to a transaction that can occur within a data. Delta Lake simplifies this with transaction awareness using a transaction log, and creates additional metadata for easier retrieval. Data is made available in a fraction of the time needed in data warehouse-based systems. This is why the largest retailers in the world are using Delta Lake for processing t-log data.

Once data is available, users need to generate predictions about item availability on the shelf. With its extremely performant engine and the ability to distribute computation across countless nodes, Spark provides the perfect platform for calculating out-of-stocks. Customers no longer need to run in aggregate or against a subset of data.

databricks

And lastly, data is only useful if it drives better outcomes. Databricks can write the resulting data into Delta Lake for further reporting, or to any downstream application via APIs, feeds or other integrations. Users can feed their predictive alerts to downstream retail operations systems or even to external partners within the tightest service windows, and in enough time to drive actions on that day.

Typical use case data sources include: point-of-sale data, replenishment data, safety stock calculations, manual inventory data (optional), robotic or computer vision inventory data (optional)

<div style="color:gray">CASE STUDY</div> reckitt

Reckitt distributes its products to millions of consumers in over 60 countries, which was causing the organization to struggle with the complexity of forecast demand, especially with large volumes of different types of data across many disjointed pipelines. Thanks to the Databricks Lakehouse Platform, Reckitt now uses predictive analytics, product placement and business forecasting to better support neighborhood grocery stores.

HOW TO GET STARTED
Solution Accelerator: On-Shelf Availability
In this solution, we show how the Databricks Lakehouse Platform enables real-time insights to rapidly respond to demand, drive more sales by ensuring stock is available on shelf, and scale out your forecasting models to accommodate any size operation.

databricks

Use Case:
Customer and Vehicle Identification

Overview

COVID-19 led to increased consumer demand for curbside pickup, drive-through and touchless payment options. Retailers that were able to implement these new services have been able to differentiate overall customer experiences and mitigate catastrophic hits on revenue levels.
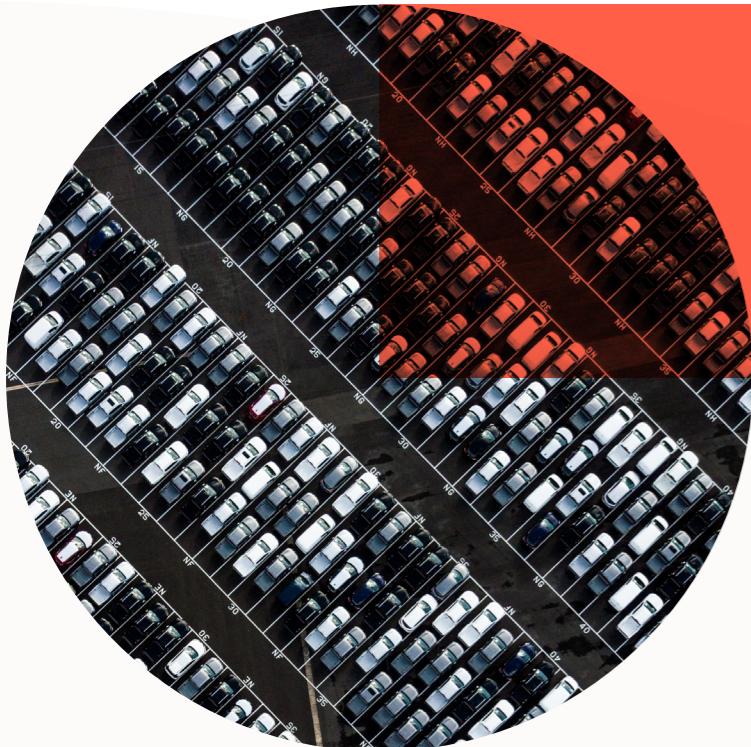
For retailers to create a seamless contactless experience for customers, they need real-time data to know when a customer has arrived and where they're located, as well as provide updates throughout the pickup journey. And through the use of computer vision, they can capture that data by employing optical recognition on images to read vehicle license plates.

Retailers can also use information captured from license plates to make recommendations on buying patterns. Looking ahead, facial recognition also has the potential to provide retailers with valuable information to better serve their customers in real time.

RELEVANT FOR →

Retail	Consumer Goods	Drive-Through Food Retailers

Challenges

- Ineffective data processing can lead to suboptimal order preparation timing

- Without real-time data, it can be difficult to provide customers with live updates on their order status

databricks

Value with Databricks

Databricks makes it possible to not only identify customers and vehicles in real time but also provide real-time communications throughout the entire shopping and curbside or drive-through experience.

- **Near real-time insights** — the greater speed to data means retailers can get the right order preparation timing

- **Recommendations** — being able to quickly access and refer to data from previous visits will ensure each subsequent visit is equally as or more seamless than the last

- **Optimizing with AI/ML** — implementing AI and ML models can lead to more effective geofencing, vehicle identification and order prediction

CASE STUDY *H&M*

CASE STUDY reckitt

databricks

CHAPTER 9

Use Case:
# Recommendation Engines
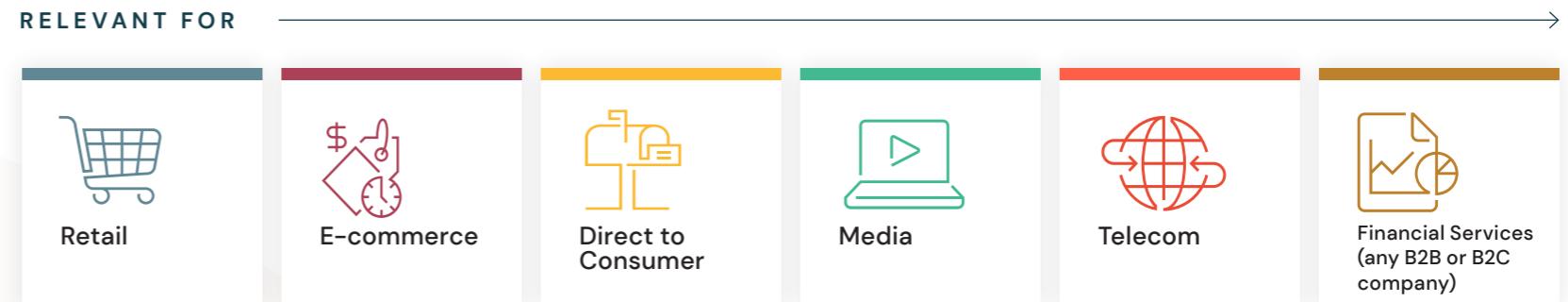
Overview

Customers that feel understood by a retailer are more likely to spend more per purchase, purchase more frequently with that retailer, and deliver higher profitability per customer. The way that retailers achieve this is by recommending products and services that align with customer needs.

Providing an experience that makes customers feel understood helps retailers stand out from the crowd of mass merchants and build loyalty. This was true before COVID, but shifting consumer preferences make this more critical than ever for retail organizations. With research showing the cost of customer acquisition is as much as five times as retaining existing ones, organizations looking to succeed in the new normal must continue to build deeper connections with existing customers in order to retain a solid consumer base. There is no shortage of options and incentives for today's consumers to rethink long–established patterns of spending.

Recommendation engines are used to create personalized experiences for users across retail channels. These recommendations are generated based on the data collected from purchases, items interacted with, users' behavior across physical and digital channels, and other data such as from customer service interactions and reviews. Leveraging a Customer 360 architecture that collects all user clickstream and behavioral data, marketers are able to create recommendations that are integrated with other business objectives such as highlighting items that are on promotion or product availability.

Creating recommendations is not a monolithic activity. Recommendation engines are used to personalize the customer experience in every possible area of consumer engagement, from proactive notifications and offers, to landing page optimization, suggested products, automated shipment recommendations, cross–sell and upsell, and even suggestions for complementary items after the purchase.

databricks

RELEVANT FOR

Retail	E-commerce	Direct to Consumer	Media	Telecom	Financial Services (any B2B or B2C company)

Challenges

Recommendation engines are very difficult to do well. Many companies use off-the-shelf recommenders, but traditional off-the-shelf systems suffer from high rates of inaccuracy. In our analysis, we found general recommenders with 29% variance, meaning that of every 10 recommendations delivered, 3 would be irrelevant.

- **Massive volumes of highly detailed and frequently changing data.** Recommendation accuracy is improved by having recent data, and yet most systems struggle to handle the large volumes of information involved.

- **Creating a 360 view of the customer.** Identity and being able to stitch together all customer touchpoints in one place are critical to enabling this use case. More data, including transaction and clickstream data, is critical for driving accuracy and precision in messaging.

- **Processing speed.** Retailers need to be able to frequently refresh models based on constantly changing dynamics, and deliver real-time recommendations via APIs.

- **Automation.** This is an "always-on" use case where automation is essential for scalability and responsiveness based on frequent model updates.

databricks

Many firms choose to use recommender systems from Amazon or Google. Using these systems trains the general recommendation engine in a way that helps competitors improve the accuracy of their own recommendations.

Value with Databricks

Recommendations are one of the most critical capabilities that a retailer maintains. This is a capability that retailers must own, and Databricks provides a solid platform for enabling this.

Using Databricks as the foundation for their Customer 360 architecture to deliver omnichannel personalization, sample value metrics from a media agency include:

- **200% ROI for 70% of retailers** engaging in advanced personalization

- **10% improvement** in conversions

- **35% improvement** in purchase frequency

- **37% improvement** in customer lifetime value

Solution overview

Recommendations are only as good as the data that powers them. Delta Lake provides the best platform for capturing and managing huge volumes of highly atomic and frequently changing data. It allows organizations to combine various sources of data in a timely and efficient manner, from transactions, demographics and preference information across products, to clickstream, digital journey and marketing analytics data to bring a 360 view of customer interactions to enable omnichannel personalization.

By identifying changes in user behavior or engagement, retailers are able to detect early signals that indicate a propensity to buy or a change in preferences, and recommend products and services that will keep consumers engaged.

databricks

Typical use case data sources include: Customer 360 data, CRM, loyalty data, transaction data, clickstream data, mobile data:

- **Engagement data** — transaction log data, clickstream data, promotion interaction

- **Identity** — loyalty data, person ID, device ID, email, IP address, name, gender, income, presence of children, location

- **User lifecycle** — subscription status, payment history, cost of acquisition, lifetime value, propensity to churn

HOW TO GET STARTED

Databricks has created four Recommendation Engine accelerators, with content-based and collaborative filter methods, and both item- and user-based analysis. These accelerators have been further refined to be highly performant to enable frequent retraining of models.

To begin working on recommendation engines, contact your Databricks account team.

CASE STUDY wehkamp

For Wehkamp to provide the best shopping experience for their customers, they turned to Databricks for help with their data analytics and machine learning needs, resulting in a highly engaging web shop personalized to each of their customers.

CASE STUDY Columbia

Columbia's legacy ETL was unable to support batch and real-time use cases at scale. After migrating to Databricks, the company is now able to more efficiently and reliably work with its data, resulting in smarter business decisions.

CASE STUDY PANDORA

Pandora wanted to drive stronger online engagement with their customers, so they used the Databricks Lakehouse Platform to create more personalized experiences and boost both click-to-open rates and quarterly revenue.

databricks

CHAPTER 10
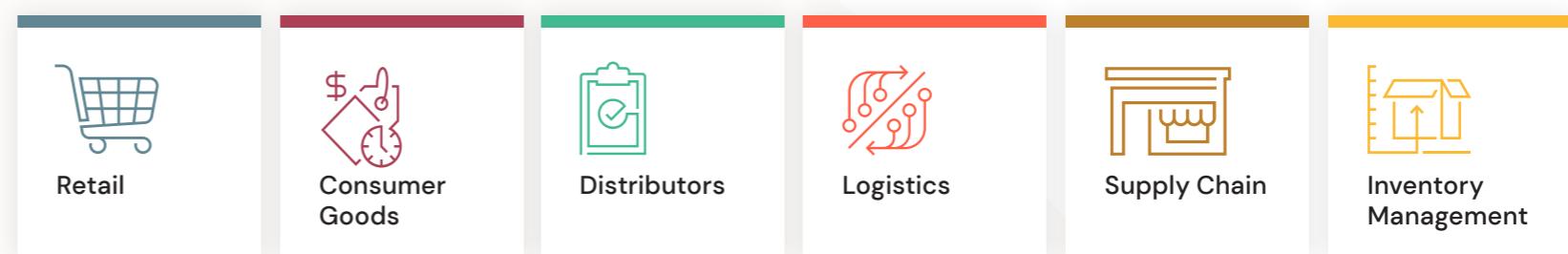
Use Case:
Perpetual Inventory

Overview

With the rapid adoption of digital channels for retail, staying on top of your inventory is crucial to meeting customer demand. As a result, the periodic inventory system is now outdated — instead, using a perpetual inventory model allows businesses to perform immediate and real-time tracking of sales and inventory levels. This has the added benefit of reducing labor costs and human error, ensuring that you always have an accurate overview of your inventory and can better forecast demand to avoid costly stockouts.
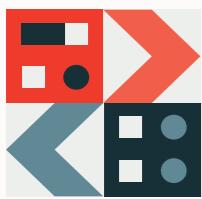
The key to building a perpetual inventory system is real-time data. By capturing real-time transaction records related to sold inventory, retailers can make smarter inventory decisions that streamline operations and lower overall costs.

RELEVANT FOR

Retail	Consumer Goods	Distributors	Logistics	Supply Chain	Inventory Management

Challenges

- Companies need to scale to handle ever-increasing inventory and the data associated with the products

- Data needs to be ingested and then processed in real time (or near real-time) to provide a truly accurate view of inventory

databricks

Value with Databricks

Databricks enables real-time inventory updates, giving businesses the insights they need to properly manage inventory and to forecast more accurately.

- **Near real-time insights** — the greater speed to data means inventory is automatically updated with the latest sales data

- **Detailed records** — with all inventory updates and movements being tracked as they happen, companies know they're getting the most accurate information at any point

- **Optimizing with AI/ML** — using AI and ML can help with forecasting demand and reducing inventory management costs

databricks
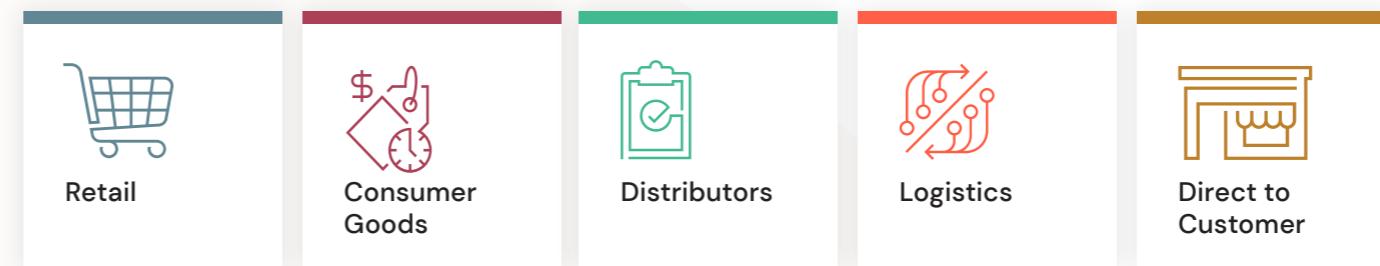
Use Case: Automated Replenishments

Overview

Customers favor convenience more than ever when it comes to their goods, and automated replenishments help meet that need. Whether it's through a connected device or smartphone app, real-time data plays a key role in ensuring consumers get a refill automatically delivered at the right time.
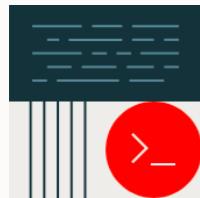
On the manufacturing side, this real-time data can also help with vendor-managed replenishment (VMR), reducing the time needed to forecast, order and receive thousands of items.

RELEVANT FOR

Retail	Consumer Goods	Distributors	Logistics	Direct to Customer

Challenges

- Being able to ingest large amounts of data quickly is crucial to actually fulfilling the replenishment orders

- With VMR, there may be a disconnect between the vendor and customer, resulting in a forecast for replenishment even when the customer can't fulfill that order

databricks

Value with Databricks

Databricks enables real-time inventory updates, giving businesses the insights they need to properly manage inventory and to forecast more accurately.

- **Near real-time insights** — the greater speed to data means businesses can stay on top of customer needs

- **Scalability** — companies can scale with Databricks to handle thousands of SKUs, each with its own unique properties and expiry dates

- **Optimizing with AI/ML** — using AI and ML can lead to better forecasting and predictions

databricks

CHAPTER 12
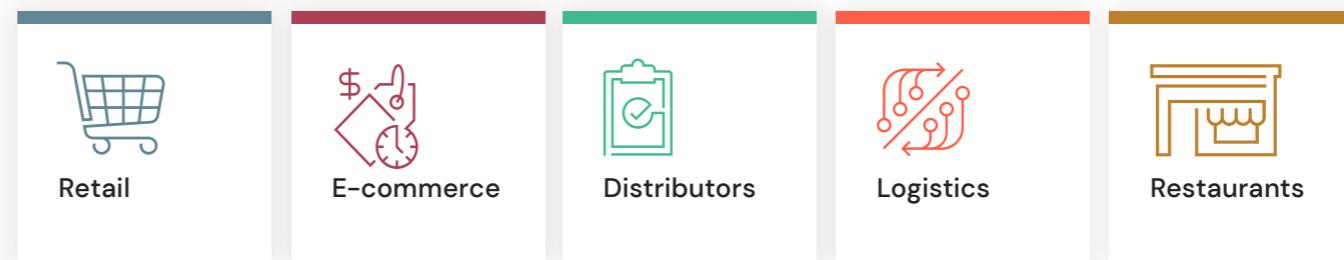
Use Case:
Fresh Food Forecasting

Overview

Fresh food typically accounts for up to 40% of revenue for grocers, and plays an important role in driving store traffic. But fresh food is also incredibly complex to manage — prices can be volatile, there is a wide range of suppliers to work with and the products expire, which creates significant amounts of waste.
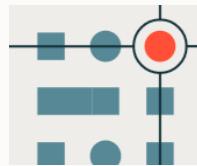
In order to avoid losing significant revenue, businesses need to properly forecast when food is nearing its sell–by date, the current levels of customer demand (also taking into account seasonality), and the proper timing for replenishing food stock. Being able to tap into real–time data is key to staying on top of the ever–changing needs around fresh food.

RELEVANT FOR

| Retail | E-commerce | Distributors | Logistics | Restaurants |

Challenges

- Because of the perishable nature of fresh food, customers need to be able to ingest data quickly enough to conduct daily forecasting and daily replenishment

- Customers are running aggregate-level forecasts, which are less accurate than fine–grained forecasting

- Customers are forced to compromise on what they can analyze

databricks

HOW TO GET STARTED

Contact your Databricks account team to get started with inventory allocation. Databricks does not have a Solution Accelerator.

View our webinar covering demand forecasting with Starbucks and then read our blog about demand forecasting.

Fine-grained time series forecasting at scale. This blog details the importance of time series forecasting, walks through building a simple model to show the use of Facebook Prophet, and then shows off the combination of Facebook Prophet and Adobe Spark to scale to hundreds of models.

On-demand webinar for demand forecasting. Video and Q&A from our webinar with Starbucks on demand forecasting.

Value with Databricks

Customers average double-digit improvement in forecast accuracy, leading to a reduction in lost sales and in spoiled products, as well as lower inventory and handling costs.

- **Improved accuracy** — on average, customers moving to Databricks for demand forecasting report a double-digit improvement in forecast accuracy

- **Ability to scale and perform fine-grained (day, store, item) level forecasts** — rapidly scale to tens of millions of model iterations in narrow service windows. Companies need accurate demand forecasts in a few hours.

- **Eliminate compromises on what to analyze** — customers do not need to select winners or losers among the products they forecast. They can predict demand for all products as frequently as required.

Solution overview:

Databricks is well suited to handling forecasting for fresh food at scale. Forecasting begins with the Databricks Solution Accelerator. It enables companies to rapidly build fine-grained forecasting of items — forecasting that can be efficiently scaled to tens of millions of predictions in tight service windows.

Typical use case data sources include: historic point-of-sale data, shipment data, promotions, pricing, expiration dates and weather.

CASE STUDY **BUTCHERBOX**

ButcherBox faced the complex challenges of securing inventory with enough lead time, meeting highly variable customer order preferences and unpredictable customer sign-ups, and managing delivery logistics. With Databricks, the company was able to create a predictive solution to adapt quickly and integrate tightly with the rest of its data estate.

CASE STUDY **sam's club**

Sam's Club needed to build out an enterprise-scale data platform to handle the billions of transactions and trillions of events going through the company. Find out how Databricks became a key component in the shift from on-premises Hadoop clusters to a cloud-based platform.

databricks

CHAPTER 13

# Use Case:
Propensity-to-Buy
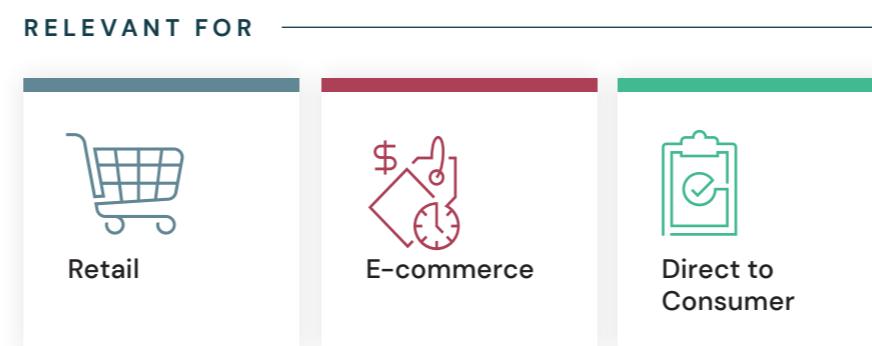
Overview

Customers often have repeatable purchase patterns that may not be noticed upon initial observation. While we know that commuting office workers are likely to purchase coffee at a coffee shop on weekday mornings, do we understand why they visit on Thursday afternoons? And more importantly, how do we predict these buying moments when customers are not in our stores?

The purpose of a propensity-to-buy model is to predict when a customer is predisposed to make a purchase and subsequently act on that information by engaging customers. Traditional propensity-to-buy models leveraged internal sales and loyalty data to identify patterns of consumption. These models are useful, but are limited in understanding the full behavior of customers. More advanced propensity-to-buy models are now incorporating alternative data sets to identify trips to competing retailers, competitive scan data from receipts, and causal data that helps to explain when and why customers make purchases.

Propensity-to-buy models create a signal that is sent to downstream systems such as those for promotion management, email and mobile alerts, recommendations and others.

RELEVANT FOR →

Retail	E-commerce	Direct to Consumer

databricks

Challenges
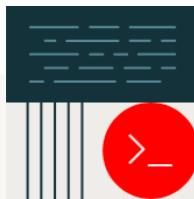
- Customers do not want to be inundated with messages from retailers. Companies need to limit their outreach to customers to avoid angering them.

- Companies need to traverse and process vast sums of customer data and generate probabilities of purchase frequently

- Companies need to look at external data that helps build a propensity-to-buy model that captures the full share of the customer wallet. They need to quickly test and incorporate additional data that improves the accuracy of their models.

Value with Databricks

- Databricks allows companies to efficiently traverse huge volumes of customer data over time, and efficiently synthesize this into data for analysis

- Companies need to traverse and process vast sums of customer data and generate probabilities of purchase frequency

- Companies need to look at external data that helps build a propensity-to-buy model that captures the full share of the customer wallet. They need to quickly test and incorporate additional data that improves the accuracy of their models.

Solution overview:

Propensity-to-buy analytics determine the signals that indicate the probability a customer is in a buying moment. Historic propensity models relied on sales data to identify buying patterns, but newer approaches are incorporating behavioral data. Proximity to a coffee shop might push a consumer over the threshold of a buying moment. Traditional, batch-oriented operations are insufficient to solve this problem. If you wait until that night, or even later in the day, you have lost the opportunity to act.

databricks

HOW TO GET STARTED

To begin working on propensity-to-buy, leverage our Propensity Scoring Solution Accelerator

With the propensity to buy, speed becomes a critical force in determining key inflection points. Databricks enables marketers to ingest data in real time and update probabilities. Lightweight queries can be automated to refresh models, and the resulting data can be fed automatically to downstream promotions, web or mobile systems, where the consumer can be engaged.

As this data is streamed into Delta Lake, data teams can quickly capture the data for broader analysis. Calculating a propensity to buy requires traversing interactions that are episodic in nature, and span broad periods of time. Delta Lake helps simplify this with scalable metadata handling, ACID transactions and data skipping. Delta Lake even manages schema evolution to provide users with flexibility as their needs evolve.
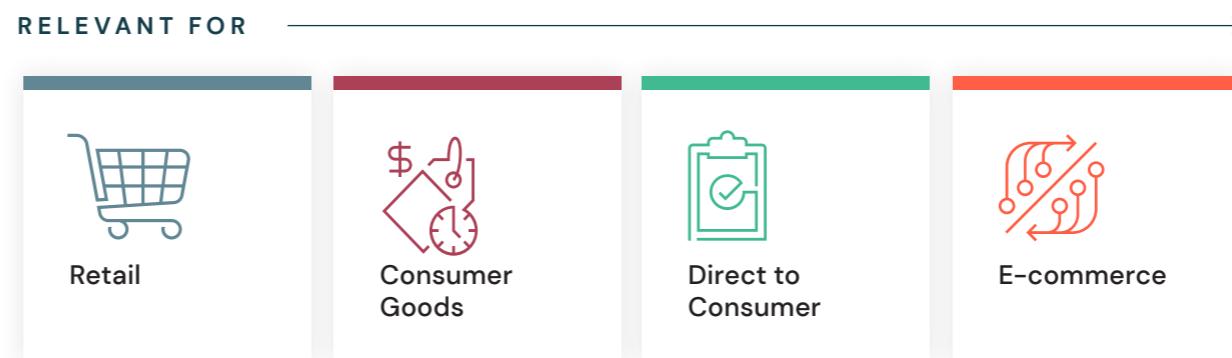
Typical use case data sources include: point-of-sale data with tokens, loyalty data, e-commerce sales data, mobile application data, competitive scan or receipt data (optional), place of interest data (optional)

databricks

CHAPTER 14

# Use Case:
Next Best Action
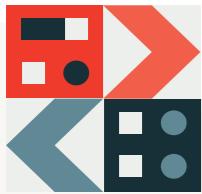
Overview

The e-commerce boom over the last couple of years has given consumers ample choice for digital shopping options. If your business isn't engaging customers at every point in their purchasing journey, you risk losing them to a competitor. By applying AI/ML to automatically determine — in real time — the next best action for customers, you can greatly increase your conversion rates.

RELEVANT FOR

Retail	Consumer Goods	Direct to Consumer	E-commerce

Challenges

- Siloed data makes it difficult to create an accurate and comprehensive profile of each customer, resulting in suboptimal recommendations for the next best action

- Companies need to ingest large amounts of data in real time and then take action on it immediately

- Many businesses still struggle with training their ML models to properly determine the next best action (and self-optimize based on the results)

databricks

HOW TO GET STARTED

To begin working on propensity-to-buy, leverage our Propensity Scoring Solution Accelerator

Value with Databricks:

- Databricks provides all the tools needed to **process large volumes of data and find the next best action** at any given point in the customer journey

- **Near real-time insights** — the greater speed to data means businesses can react immediately to customer actions

- **Single source of truth** — break down data silos by unifying all of a company's customer data (including basic information, transactional data, online behavior/purchase history, and more) to get a complete customer profile

- **Optimizing with AI/ML** — use AI to create self-optimizing ML models that are trained to find the best next step for customers

databricks

CHAPTER 15

Customers That Innovate With Databricks Lakehouse for Retail

Some of the top retail and consumer packaged goods companies in the world turn to Databricks Lakehouse for Retail to deliver real-time experiences to their customers.

Today, data is at the core of every innovation in the retail and consumer packaged goods industry. Databricks Lakehouse for Retail enables companies across every sector of retail and consumer goods to harness the power of real-time data and analytics to solve strategic challenges and deliver more engaging experiences to customers.

Get started with a free trial of Lakehouse for Retail and start building better data applications today.

Start your free trial

Contact us for a personalized demo at: databricks.com/contact

databricks

About Databricks

Databricks is the data and AI company. More than 7,000 organizations worldwide — including Comcast, Condé Nast, H&M 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.

Sign up for a free trial

databricks