

RETAIL TRANSFORMATION, A BUSINESS IMPERATIVE

How digital transformation readiness defines the gap
between leaders and laggards

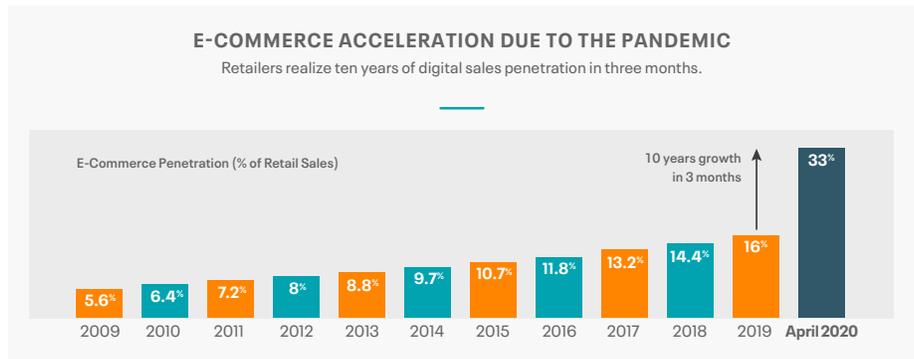


“The rules of engagement of retail have changed, but many are still playing by the rules of 60 years ago.”

Source: IHL Group

The once emerging retail and consumer goods renaissance, driven by data and analytics, seems like a distant memory. The global pandemic arrived, and it not only justified retail's earlier digital transformation strategic plans and investments, but catapulted some prepared retail leaders into record setting revenues and profits, while forcing rapid bankruptcies for those who had not sufficiently prioritized data as a strategic asset.

COVID-19 quickly tested all retail and consumer goods companies' readiness in their individual digital transformation journeys in a matter of weeks. While retail leaders may have been well positioned, laggards suddenly accelerated strategic plans, many still in the midst of their three-year execution horizons. Regardless, those successfully managing the pandemic's stress test of e-commerce, in-store insights, and supply-chain visibility and fulfillment capabilities started with a foundation built on consumer experience and high-stakes competitive insights.



Source: ShawSpring Research, Bank of America, U.S. Department of Commerce, Forrester Analytics, McKinsey analysis

Until recently, retail and consumer goods technology primarily served to ensure transactions with customers were quick and efficient. Now, retailers and brands have an unprecedented opportunity to enact change by leveraging data, analytics, technology, shared collaboration frameworks and streaming data to elevate the customer experience and improve operational efficiencies in much more meaningful ways.

Connected Consumers (Remain) in Control

Consumers are more in control than ever. Mobile and socially connected, they are armed with more knowledge than ever before. They're curious about where products come from, how they are made and the social causes that companies embrace. They are harder to influence, and they enjoy controlling their own shopping experience—redefining the meaning of "loyalty."

As a result, leading retailers and brands are integrating more tech (data, analytics and devices) into every step of the consumer purchase path to create a consumer-first approach. The entire retail experience—starting with production, replenishment and continuing into merchandising, marketing and on to fulfillment and returns—has an opportunity (and challenge) to shape customer perception. Loyalty is becoming more important now than ever.

In fact, every aspect of the retail value chain has the potential to make or break the customer experience today. From developing new products to delivering personalized campaigns—analytics, artificial intelligence (AI) the internet of things (IoT) and enterprise data management capabilities are empowering brands and retailers to better understand their shoppers and deliver differentiated products, services and experiences. As customers' expectations for best-in-class shopping continues to increase—thanks to the immediate and personalized relationships established by tech giants and retail upstarts—data-led intelligence will remain essential for retailers and their trading partners to thrive in this competitive landscape.

Leaders Focus on Four Key Strategies

Leading retailers and brands today are laser-focused on optimizing four key strategies crucial to delivering a next-level shopper experience built on the ability to better understand customers:

91%

of consumers are more likely to shop with brands that recognize, remember and provide relevant offers and recommendations.

Source: Personalization Pulse Check Report, Accenture Interactive

\$1.1T

annual cost to retail due to out of stocks and overstocks

Source: IHL Group

infusing customer insights into product, upgrading marketing and merchandising; creating an agile supply chain that both optimizes in-stock and fulfillment capabilities; and reimagining brick and mortar stores as a competitive differentiator. More specifically:

- **Personalized Experiences:** Retailers are leveraging customer profiles, producing higher customer engagement results and reduced marketing costs by delivering targeted, relevant, contextual content and recommendations. Leaders are moving to “segments of one,” defined as tracking and understanding individual behaviors and using the data to customize offers, products or services to the individual customer.
- **Customer-Centric Merchandising:** Moving from a product to a customer-centric merchandising strategy requires a deep understanding of customer tastes and preferences. Granular product attribution and retail execution data shared among trading partners is critically important for both to better respond through localized assortments, tailored promotions, dynamic pricing and product development for an ever-changing, diverse consumer base.
- **Supply-Chain Agility:** In today’s retail environment, retailers realize that building demand forecasts simply based upon historical transaction, promotional and pricing data alone is not good enough. Including new data sources like demand signals—such as weather, social commentary, competitor pricing, local event calendars, shipping and returns policies and demand transfer dynamics— not only boosts forecast accuracy, but it greatly enhances inventory visibility, reduces out-of-stocks and improves fulfillment optimization capabilities.
- **Reimagining Stores:** With more than [78% of global retail sales projected to still occur in-store by 2023](#), traditional retailers are increasingly realizing that the brick-and-mortar footprint is a competitive differentiator. Enabling better pricing and convenience to consumers than digital pure-plays, physical stores now serve as micro-fulfillment centers—through BOPIS (buy online, pick-up in store) or curbside delivery—and drive down overall cost-to-serve. In-store customer insights and engagement opportunities are now possible using data captured from sensors, video and beacons. This technology allows retailers to measure and respond in real-time to shopper behavior, measure geo-location, traffic, dwell times and conversion metrics. For merchants, the ability to capture shelf, rack, table and bin inventory levels allows them to prevent out-of-stocks (lost sales), monitor merchandising (display, pricing, promo, POG), meet compliance initiatives and share these new insights with trading partners they may have. Traditional retailers now have access to an entirely new data monetization opportunity that many have been missing out on for years.

25%

Digital supply chains respond 25% faster due to real-time information

Source: BCG

A Meaningful Response

Retailers moving faster than their competition know that it isn’t enough to merely collect all types of data—internal, external and multi-structured—something many retailers already do well. They are now making advances in analyzing streaming data, embedding meaningful analytics and responding in near real time to “moment of truth” business opportunities. Customer interactions, digital media, inventory status, supply chain network and trading partners can all create insights, events and engagement.

These opportunities move in data streams that demand rapid opportunistic response times and agile problem solving. Retailers must have capabilities to analyze customer browsing behavior and gain inventory visibility across the supply chain, fulfillment optimization, delivery routes and out-of-stocks. In-store customer journey insights and social commentary can provide heads up intelligence for customer service representatives. Successful response is not about having all the data, it’s about using it in a meaningful and “right time” response. It’s about sorting, filtering and identifying the events that matter in real time in order to respond effectively.

To compete in today’s environment, retailers have embraced a new approach. They are using solutions that enable ingestion of streaming data from a seemingly endless variety of sources, storing and processing across a hybrid infrastructure and running analytics and improving precision in machine learning (ML) algorithms. At the same time, leading retailers are maintaining strict enterprise data security, governance and control across all environments with no data silos.

Successful response is not about having all the data, it’s about using it in a meaningful and “right time” response.

91%

of organizations struggle to reach data maturity

Source: Gartner

Challenges to Retail Digital Transformation

Technology is dramatically reshaping the face of shopping and giving forward-thinking retailers and brands a significant competitive edge. However, many still face hurdles on the road to modernization:

- Culture:** Retailers have a profound cultural resistance to the notion of “vendor lock in.” Proprietary solutions that leave retailers bound to a single solutions provider are distrusted. Retailers are eager to own their IP (intellectual property) and leverage it to be masters of their own destiny. Put another way: Retail is inherently distrustful of black-box solutions. Leaders would rather own the competencies that support digital transformation than sign on with a vendor featuring a variety of point solutions (and data silos), leaving retailers with no understanding of the underlying math, logic or results.
- Cost:** Historically, retailers also have been conservative with their IT spend because they live with razor-thin margins, and that cultural predisposition remains. While the top 50 or so global retailers are investing aggressively in new technologies, the rest—partly by habit and partly by necessity—are wary of pursuing solutions that come with too big a price tag. Along these same lines, many feel they can’t invest in top-dollar data scientists and software developers, leaving them unable to make the most of their already available data or integrate new and emerging data streams as they arise.
- Infrastructure:** In terms of the data itself, many retailers are still lacking an enterprise data strategy and comprehensive infrastructure solution, one that can handle multi-structured and multi-location (on premise or multi-cloud) data. Making this more challenging is the perpetual growth of disparate data sources, from IoT devices to external partners to enterprise data. Easily accessible data must be leveraged across on-premise and multi-cloud environments to drive solutions that deliver answers to the most complex business questions.
- Real-Time Complexity:** Data is both retail’s friend and enemy. Retail has never had a problem acquiring data, it does, however, have a problem with ingesting, analyzing and acting on data in real time. To manage the complexity of real-time information, retail data management platforms must enable real-time analytics on streaming data, effectively ingesting, storing and processing data to instantly deliver insights and action.
- Handling the Volume and Variety of IoT Data:** To enable use cases like supply-chain optimization or marketing to “segments of one,” retailers demand a platform that can handle diverse data structures and schemas—everything from intermittent data originating from click-stream or point of sale, to unstructured data (e.g., images, video, text and barcodes) from the edge, delivered through a variety of supported drivers and protocols.
- Diverse Analytical Capabilities:** Existing platforms offer limited ability to provide insights and analytics into platform usage and performance. For retail solutions, a data platform must provide a wide range of analytical options—including everything from SQL analytics and search capabilities, to tools to support advanced analytics, ML modeling and AI development. To be most effective, it must also allow for tight integration with leading business intelligence (BI) solutions that offer reporting dashboards and and specialized business analytics capabilities.

AI and IoT are about to reinvent how industries utilize data—and could drive an era of productivity growth in our cities, farms, factories, and consumer venues.

Source: Morgan Stanley Research



of the top 20 global retailers rely on Cloudera, including over 250 retailers, food service, hospitality and consumer products companies worldwide.

Open Source Technology Drives Innovation

Retail savvy tech giants have invested heavily in data analytics as a competitive differentiator. Data-driven exemplars like Amazon, Alibaba, Facebook, and Google are continuing to use data analytics as a competitive wedge. They recognize that new data types and advanced analytics are not only changing customer demand and containing operational costs, they are driving consumer expectations.

Building and retaining their own data analytics competency and owning their intellectual property is fundamental to who these technology titans are. To advance their competency, they continue to add people skills and implement change management initiatives based upon data and analytics. To leverage their intellectual property, they're organized and equipped to infuse their own software development with advanced ML and AI capabilities.

While the largest traditional retailers and brands can organize internal resources and keep the development of systems in-house, many others have neither the capacity nor the executive commitment to invest on such a grand scale. With an open source data platform, however, retailers can own and control both their data and their intellectual property. Rather than depend on a single software vendor timeline for updates and new releases and face punitive data egress charges from cloud-based point solutions, retailers relying on open source solutions are enmeshed in a broad ecosystem of developers, all generating new ideas and new capabilities. Retailers and brands need no longer rely on "black boxes" that give them no visibility to the underlying math, algorithms or result sets. Open source puts the retailer and brand closer to the source of innovation, thanks to some of the biggest companies in the world actively contributing to that dynamic ecosystem. In this environment, retailers and brands are freed from their biggest cultural concern around digital transformation. In an open-source environment, they own the intellectual property, the enterprise data and the analytic competencies. They own the data lifecycle in its entirety.

Cloudera is uniquely positioned to help retailers and brands make the best use of the vast amounts of data already on hand, operationalizing insights derived from a wide array of emerging data sources. Key to this advantage is its open source approach that meets retailers' and brands' practical and cultural approaches to corporate digital transformation.

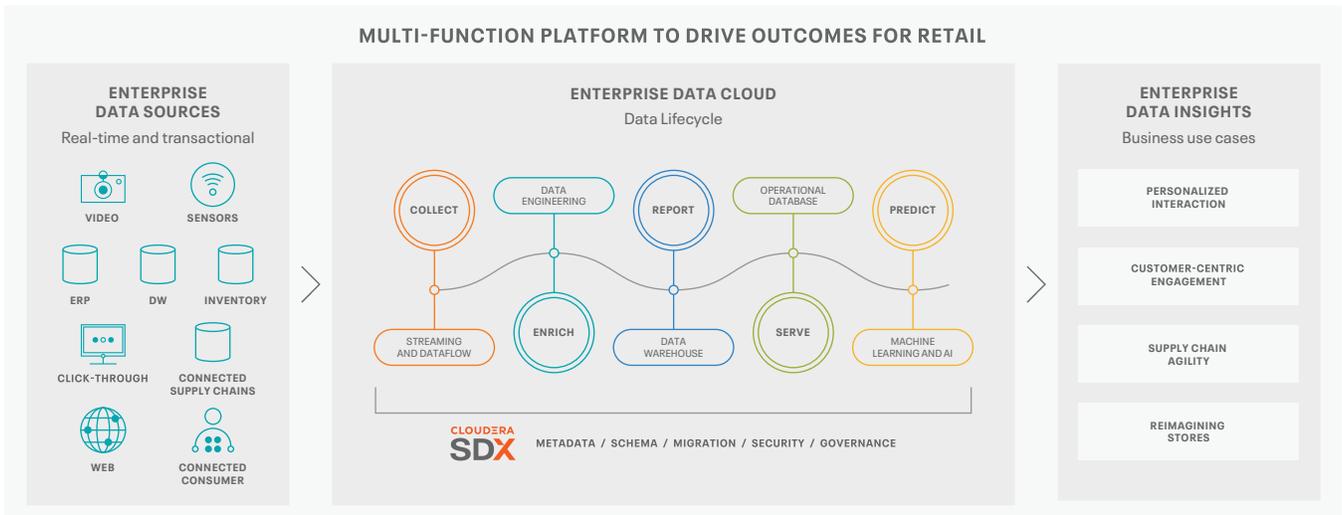
COMMON CONSUMER GOODS BUSINESS INITIATIVES

More effectively connect with shoppers and convert them into engaged, loyal customers		Differentiate the brand from the competition to increase sales, volume, and margin
Deliver timely, personalized customer experiences and informed decisions—at the moment of truth		Orchestrate conversations with consumers to drive visits, improve retention and up/cross-sell opportunities
Deliver relevant content, knowledge, and communications to in-store employees and consumers anytime, anywhere		Decrease cost to serve, empower front line employees/field sales reps to improve customer experiences
Deliver on the brand promise with responsive fulfillment, supply chain, customer service, and field sales visibility		Improve consumer experience, satisfaction, whilst reducing defects, out of stocks, and fulfillment costs

End-to-End Data Lifecycle Management with Cloudera Data Platform

Today, leading retail organizations worldwide are adopting an enterprise data cloud strategy using the open source-based [Cloudera Data Platform](#) to manage the end-to-end data lifecycle. From collecting data from multiple sources, to storing, processing, analyzing, serving and predicting to drive actionable insights and use cases, the platform handles both data in motion and data at rest.

With [Cloudera Data Platform](#), data can be ingested from a variety of sources (including both streaming and enterprise data sources), enriched and processed across a hybrid infrastructure. Analytics or ML algorithms can be applied to all data, all while maintaining strict enterprise data security, governance and control across all environments.



Here is an overview of the end-to-end data lifecycle:

1. Collect

With [Cloudera Data Platform](#), retail organizations can easily ingest data from multiple sources, including both traditional as well as new and streaming data sources such as IoT, beacons, RFID or associate-worn connected devices. Any type of data can be ingested and loaded without altering its format or fidelity, preserving data integrity and enabling security and governance.

[Cloudera DataFlow](#) provides a scalable, real-time streaming analytics engine that ingests, curates and analyzes data for key insights and immediate actionable intelligence. It can ingest and process real-time data in motion from streaming data sources (such as clickstreams, log streams, social feeds, IoT sensors, smart devices, etc.) as well as from traditional enterprise data sources, such as databases, data warehouses as well as enterprise systems that manage CRM, customer care, inventory and supply chains. It addresses the key challenges enterprises face with data in motion including:

- Ingesting and enriching real-time data streaming at high volume and high scale.
- Driving data processing and predictive analytics on high-speed data in motion.
- Tracking provenance and lineage of streaming data.
- Managing and monitoring edge applications and streaming sources.
- Enabling edge intelligence with ML models and edge processing.

2. Enrich

Preparing data for analysis and insights is the foundation of any data-driven exercise. [Cloudera Data Engineering](#) helps enrich, transform and cleanse a wide variety of data and makes it easier than ever to create and execute end-to-end data pipelines. For IT operations, this means delivering real-time business information blended with additional datasets that provide context for decision-makers. The underlying platform also delivers the ability to execute a wide range of



Why Cludera

HYBRID AND MULTI-CLOUD

Run analytics on any cloud, easily and securely moving data and metadata between on-premises file systems and cloud object stores.

ANALYTICS FROM EDGE TO AI

Apply real-time stream processing, data warehousing, data science and iterative machine learning across shared data, securely, at scale on data anywhere.

SECURITY AND GOVERNANCE

Use a common security model, role and attribute-based access policies and sophisticated schema, lineage and provenance controls on any cloud.

100% OPEN

Open source, open compute, open storage, open architecture and open clouds. Open for developers, partners and open for business. No lock-in. Ever.



high-performance data processing workloads, including batch and real-time stream processing using Apache Spark and Spark Streaming. It is also supported by multiple storage data stores including Apache HBase, Apache Kudu and cloud object storage.

3. Report

With Cludera, retailers have the flexibility to run multiple analytical options to drive insights, intelligence and action from all data. Depending on the business needs, organizations can analyze data in a variety of ways—including interactive SQL, text search, integration with leading BI and visualization tools—or run advanced analytics and ML models.

Traditional data warehouses are inadequate to meet the increased scale, economics and analytics demands that today’s retailers are experiencing. [Cludera Data Warehouse](#) is an auto-scaling, highly concurrent and cost-effective analytics service that ingests high scale data, whether structured, unstructured or from edge sources. It supports hybrid and multi-cloud infrastructure models by seamlessly moving workloads between on-premises and any cloud for reports, dashboards, ad-hoc and advanced analytics, including those leveraging AI. Cludera Data Warehouse offers zero query wait times, reduced IT costs and agile delivery, while maintaining consistent security and governance.

4. Serve

The [Cludera Operational Database](#) serves traditional structured data alongside new unstructured data within a unified end-to-end open source platform. It enables processing of streaming data and real-time analytics on that continuously changing data, ensuring the latest data and analysis can be injected into decision-making. Users can not only serve real-time data at scale, with high concurrency and low latency, but they can also use data science at scale to easily build, score and deploy ML models into production.

5. Predict

Retailers can close the loop on the data lifecycle by using [Cludera Machine Learning](#) to make predictions that will drive key business outcomes. Cludera Machine Learning can help organizations accelerate enterprise data science at scale—from research to production—with self-service, collaborative workflows for building and operationalizing ML models and interactive, visual applications. Without interrupting business workloads, retailers can rapidly onboard new data science teams, giving them on-demand access to governed business data, open-source tools and auto-scaling computing resources for end-to-end ML, all without wait time. Using Python, R and Scala directly in the web browser, Cludera Machine Learning delivers a powerful self-service experience for data science teams developing, prototyping and deploying new ML capabilities to production.

Use Cases in Retail and Consumer Goods

Analytics, ML, AI and IoT are rapidly reshaping the retail competitive landscape. The most innovative retailers and brands are applying an entire host of data acquisition concepts not only to data at rest, but also to data in motion and streaming data sources, in real time. By adding in ML and analytics, these retailers are transforming their online and in-store customer experiences, supply chains, merchandising and marketing activities. However, in order to capitalize on this fundamental shift, retailers and brands need the ability to ingest, process, store and analyze all types of data (structured, unstructured and semi-structured data) regardless of where it originates—from the consumer, from the retail shelves and aisles or from consumer click-through—to the data center, in any public or hybrid cloud. With these abilities at hand, retailers can harness petabytes of data to identify consumer and supply-chain patterns, detect stock or supply-chain delivery anomalies, identify emerging consumer trends, optimize fulfillment costs and interact with consumers in more personalized, contextual and relevant ways.

With greater visibility and insights at the point of decision, retailers and brands can combine external data with the data they generate every day to gain real-time visibility into their operations and supply chains. These insights allow them to improve operational performance and customer experiences to help reduce costs and grow revenues—and grab market share from the competition.

Today, many leading retailers and brands across the globe use the [Cloudera Data Platform](#) to drive connected customer processes that enable personalized interactions and improved experiences. They also depend on [Cloudera Data Platform](#) to optimize their supply chain and inventories across all fulfillment and delivery channels. See how four major retailers used Cloudera Data Platform to deliver value:



1. Personalized Experiences—Require ‘Segments of One’

A leading global drug store chain specializing in filling prescriptions, health and wellness products, health information and photo services sought to improve its digital personalization capabilities with customers. Considering that the chain fills 900 million prescriptions annually, interacts with over 87 million loyalty customers and sees more than 10 million sales transactions daily, this was truly a big data opportunity.

The business set out to improve the precision and timeliness of personalized, targeted email offers by moving the email offer closer to the customer’s last interaction.

The drug store chain was experiencing a **lag of three to five days between its last interaction** with a customer and delivering personalized offers due to sluggish analytic workload performance and high costs associated with its legacy data platform. To improve its business capabilities and customer relevance, the company turned to Cloudera to establish a modern data architecture capable of handling massive data workloads. A new system based on [Cloudera Data Platform](#) ingested and analyzed 365 days of customer purchase history, including in-store point-of-sale and e-commerce data across all loyalty customers to produce category-based affinity scores showing individual customer preferences. **More important, the platform reduced the time of insight from three to five days to less than three hours.**

The [Cloudera Data Platform](#) allowed secure integration of all data across the enterprise from a variety of data inputs, rapid analytic model iteration and development.

The result: The company delivered increased value to customers with more relevant and timely content while gaining a scalable and cost-effective data platform for further customer insights and analytics across all lines of business. In addition, the modernization program (anchored by Cloudera) attracted new IT talent, refreshing and upskilling the firm for future growth.



62%

of retailers report that the use of information (including big data) and analytics is creating a competitive advantage for their organizations.

Source: Analytics: The Real-world Use of Big Data in Retail, IBM

2. Customer Centric Merchandising—Location, Location, Location

Product placement in apparel and department stores is well documented as positively affecting retail sales. One leading U.S. department store retailer known for clothing, accessories and jewelry was receiving sales data that was not specific enough to suggest changes to the layout of its brick-and-mortar stores to maximize product sales. The company's online channel had an advantage as it could compare what shoppers viewed with what they bought, an insight lacking from brick-and-mortar stores' data.

To address this intelligence gap, the retailer built a [Cloudera Data Platform](#)-based enterprise solution that provided microdata on in-store shopper location, enabling analysis similar to the insights gleaned from the online business. The retailer leveraged IoT data that captured the in-store location of shoppers using the company's mobile app on their smartphones. Data then streamed from [Cloudera DataFlow](#) to [Cloudera Data Platform](#) providing:

- Intra-day insight on the flow of customers and traffic throughout stores.
- Identification of merchandising hot and cold spots.
- Insights leading to real-time, proximity-based marketing capabilities.

Real-time delivery of personalized offers quickly demonstrated a 40% conversion lift, and the data was further leveraged to redesign and optimize product placements within stores.

The key to success was the ability to stream real-time data using [Cloudera DataFlow](#) from multiple sensors and seamlessly integrate this new data into the retailer's [Cloudera Data Platform](#), where cleansing, further analysis and advanced analytics could be run to deliver further value.



3. Supply Chain Agility—A ‘Last Mile’ Commitment

When enterprise and IoT data are leveraged to improve forecast accuracy and reduce out-of-stocks, retail supply chains can demonstrate a significant return on investment. Gartner, IDC and ISM have reported that incorporating big data helps improve demand forecasts, building supply chain agility that can provide a 2% average revenue increase and a 15% average inventory reduction simultaneously.

A leading European retailer focused on groceries and general merchandise—with over 6,500 store locations and 6,000 delivery vans—**saw costs spiraling with its delivery services as it tried to keep its customer fulfillment promise.** Existing route optimization tools were not reflecting true delivery drive times, and prebuilt demand forecast models for intra-day orders were leading to further inefficiencies. **Additionally, fleet maintenance costs were growing due to unplanned downtime and the necessity for replacement vehicles.**

This retailer leveraged Cloudera to build an analytics solution for fulfillment delivery that improved data access to omni-channel orders, logistics and delivery capacity and provided advanced analytic modeling, A/B testing and optimization. Working with the incumbent point-solution provider, the retailer brought the analytic modeling and intellectual property in-house, improving the accuracy of delivery order demand forecasts and further optimizing of delivery routes. By partnering with Cloudera, the retailer reduced point-solution costs, improved analytic agility and established an approach to use with other legacy business applications. A follow-on use case of predictive analytics for vehicle maintenance improved uptime of the delivery fleet.

Illustrative, measurable business impacts were realized:

- Improved intra-day online order fulfillment demand forecast accuracy by 3%.
- Improved customer delivery capacity and speed.
- Cut 140 vehicles from its fleet, saving \$21 million in maintenance and labor costs.



4. Reimagining Stores—Real-Time Pulse on Freshness, Waste

Supermarkets discard an estimated 43 billion pounds of food every year, according to a [recent study](#).

Retail grocers are committed to doing better, but food waste is still such a pervasive problem that only one supermarket chain earned a B on the [food waste “report card”](#) recently issued by the Center for Biological Diversity. Just a handful of chains earned Cs, while the rest of the country’s most recognizable grocery stores scored Ds or Fs.

The leading global mass merchant—that scored highest in rankings—recognized a need to improve cold storage temperature fluctuations on grocery products, understanding that both high and low temperature variations could lead to excessive shrink (waste).

This retailer deployed [Cloudera DataFlow](#) to tap real-time streaming data from thousands of cold storage sensors across its vast network of brick and mortar stores. The solution ingested and aggregated data from these temperature sensors with location and on-hand inventory data to predict, monitor and respond to possible changes in perishable food products such as produce, dairy and meat.

Predictive analytics allowed the retailer to proactively respond not only to product life cycle impacts, but also the potential risk of cold storage equipment down-time. Automating the closed-loop process using pre-built business rules and alerts gave individual maintenance teams and store department managers actionable instructions to ensure product freshness and reduce waste.

The resulting application of streaming data and advanced analytics is expected to be a major contributor to **improving freshness, reducing food waste and cutting cold storage maintenance costs resulting in over \$500 million in annual business impact.**

About Cloudera

At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Data Analytics is Transforming Retail

Retailers live on thin margins. Within the current business environment, retailers must now more than ever accelerate investment into people, process and technology by leveraging data as a strategic asset. Data and analytic capabilities are rapidly becoming the key differentiators in the retail and consumer goods industry, often defining winners and losers.

Digital transformation is often used to describe a two-to-three-year strategic endeavor. Those who have already laid the foundation to more effectively leverage data, analytics and new digital technologies as a competitive advantage are seemingly thriving in this uncertain environment, even as the pandemic may force latecomers to make difficult choices.

The Cloudera Data Platform enables retailers and consumer goods companies to maintain their momentum and accelerate digital transformation by leveraging data from any source whether on-premise, cloud or hybrid platforms—powered by open-source technology. Cloudera delivers the data lifecycle solution through Cloudera Data Platform, from edge to AI.

Visit our website and learn more at cloudera.com