The Retail Sector Boosts Sales with Hadoop®

A Modern Data Architecture for Retailers

Contents

Introduction

The combined forces of technological innovation, changing consumer preferences and a prolonged period of slow economic growth are driving rapid changes in the retail industry. For retailers, Apache Hadoop can be a powerful tool for responding to that change.

In many countries, "new normal" post-recession consumer spending growth rates are lower than before the crisis. At the same time, technology has empowered consumers to be smarter, more social shoppers. Online ratings, reviews, and price comparisons influence their shopping and when it comes time to buy, they can purchase on a mobile device, online or in a store. The new shopping hours are 24x7.

Retail e-commerce continues to grow quickly, at 15.5 percent in the US in 2014 and now accounts for six percent of total retail sales.[1] At the same time, brick-and-mortar retailers are seeing market growth slow, bringing overall retail sales growth to below four percent a year.[2] These trends have driven some physical retail formats like travel agents and bookstores into oblivion. Surviving retailers face margin pressure from new online entrants and consumer expectations of greater price transparency.

In this climate, retailers need data-driven ways to segment customers, interact with them, and manage supply chains to match products with consumer demand.

Hadoop improves customer segmentation. Retailers mine a rich trail of consumer data from sources previously unavailable at a large scale: online reviews, blogs, and social network messages. The new data sources in Hadoop deliver retailers unprecedented insights into consumer behavior and preferences. Hadoop data also improves customer interactions with confident next-product-to-buy (NPTB) recommendations. It can also recognize customers across multiple channels and touch points to produce in-store or online personalized recommendations, along with real-time coupons to whet their shoppers' appetites. According to a McKinsey study, these retail programs from big data analytics can enhance retailer margins by 60%.[2]

Finally, retailers use Hadoop for better forecasting, purchasing and supply chain optimization. They can optimize distribution routes and inventory levels to appeal to the local and rapidly changing tastes of individual consumers.

Hortonworks' founding architects pioneered Apache Hadoop YARN, which moved the platform beyond its batch-only roots. Hortonworks invested heavily in YARN and spearheaded its development in the Apache community.

RESULTS WITH HADOOP

Luminar's data ingest captures

15 terabytes from 2,000 raw sources

Up from 2 terabytes from 300 sources

TrueCar stores and processes data at

23¢ per gigabyte

Instead of a traditional platform's $19 per gigabyte

A major department store promotes with

Real-time delivery

Across all media channels

Clickstream analysis for retailers now

Stores both click files and ad impressions

In the same data lake

---

1) Total US Retail Sales Top $4.5 Trillion in 2013, Outpace GDP Growth, eMarketer, April 10, 2014
http://www.emarketer.com/Article/Total-US-Retail-Sales-Top-3645-Trillion-2013-Outpace-GDP-Growth/1010756

2) Big data: The next frontier for innovation, competition and productivity, McKinsey Global Institute, June 2011

Now Hadoop 2 (with YARN) includes the following enhanced capabilities, which nobody understands better than Hortonworks.

Multi-use, Multi-workload Data Processing: Hadoop supports multiple access methods (batch, interactive, and real-time) to a common data set. Analysts can view and transform data in multiple ways at once. This speeds time-to-insight and strengthens confidence in their findings.

New Opportunities for Analytics: Hadoop's schema-on-read architecture lets users store data in its raw format. Analysts then define unique schemas for the data they need for a particular research question or application.

New Efficiencies for Data Architecture: Hadoop runs on low-cost commodity servers and reduces overall cost of storage. This makes it affordable to retain all source data for much longer periods, which provides applications with far deeper historical context.

Data Warehouse Optimization: ETL workloads also benefit from Hadoop's favorable economics. Data with low per-unit value can be extracted and transformed in Hadoop. This data's value grows in the aggregate, when it is joined with other data and stored for longer.

These advantages explain why retailers throughout the world adopt Hadoop. In fact, Hortonworks Data Platform (HDP) is currently in use at many of the largest retail companies in the US and Europe.

This white paper illustrates real-life improvements in business performance achieved by Hadoop in twelve common use cases. Retailers derive the most value from Hadoop when they begin by focusing on specific line-of-business challenges, such as those outlined here. This complements the value that Hadoop creates in horizontal functions such as IT, which are not covered specifically in this white paper.

Hadoop for Retail

ANALYSIS

Recommendation engine	Product mix
Brand health	Web path optimization
Price sensitivity	A/B testing
Top-down clustering	

X (Excel) Datameer splunk> MicroStrategy R

X (Excel) tableau SOFTWARE mahout SAS platfora Microsoft

DATA REPOSITORIES

Hortonworks

hadoop

EDW

ERP

CRM

RDBMS

Governance & Integration

Batch	Script	SQL	NoSQL	Stream	Search	Others

YARN : Data Operating System

HDFS
(Hadoop Distributed File System)

Security

Operations

TRADITIONAL SOURCES

CRM	WEB TRANSACTIONS	PRODUCT CATALOG	STAFFING
ERP	POS TRANSACTIONS	INVENTORY	STORES

EMERGING & NON-TRADITIONAL SOURCES

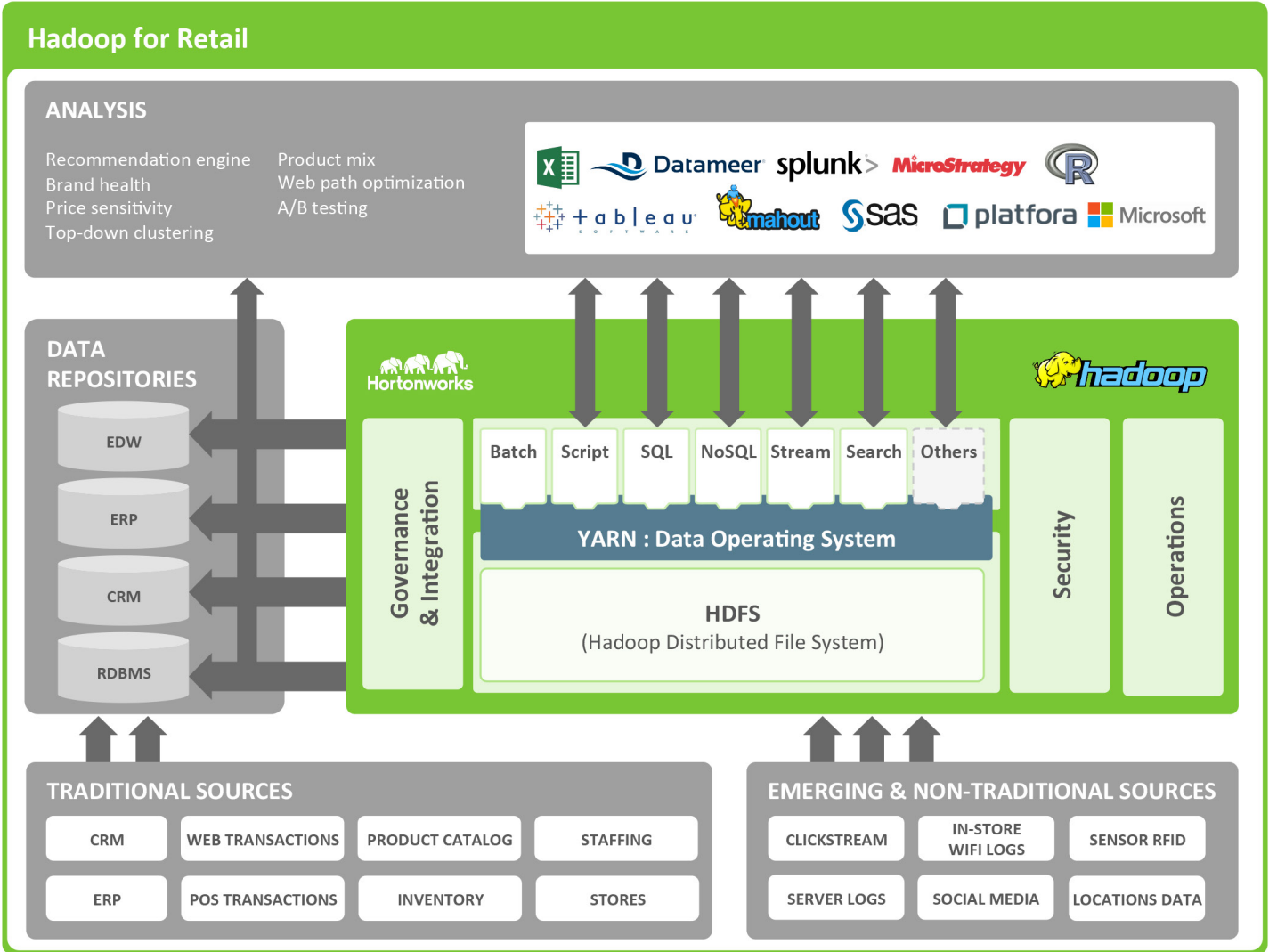| CLICKSTREAM	IN-STORE WIFI LOGS	SENSOR RFID
SERVER LOGS	SOCIAL MEDIA	LOCATIONS DATA

Figure 1: Ingest all the data, store it at scale and analyze it with the tools you already use.

Marketing

Hadoop drives better business outcomes in marketing through its ability to improve segmentation, conversion, share-of-wallet and customer lifetime value.

360° VIEW OF LIFETIME CUSTOMER VALUE

Business challenge

Retailers interact with customers across multiple channels, so customer interaction and transaction data is often stored in different data siloes. This means that few retailers can accurately correlate eventual customer purchases with marketing campaigns and online browsing behavior to predict the lifetime value (LVC) of its customers.

For example, a major home improvement retailer sells online and across thousands of stores worldwide. But its data on website traffic, POS transactions and in-home services existed in fragmented silos. Lack of a unified "golden record" of customer behavior limited the retailer's efforts to quantify the lifetime value of its relationships. Without this granular segmentation information, the company couldn't prioritize marketing efforts according to expected LCV.

Solution

The retailer installed Hadoop at its head office to consolidate all customer information into one data lake. Now the golden record ties together structured ERP and CRM data with clickstream, social media, geo-location and POS transactional data. The company uses this data lake for precise, highly targeted segmentation and LCV calculations. That comprehensive information drives specific outbound actions such as customized coupons, promotions and emails.

Impact

Now Apache Hadoop gives this retailer a 360° view of its customers, segmented on multiple dimensions including: shopping basket analysis, preferences expressed on social media, and observed feedback from marketing campaigns. Better analytics increase sales, reduce inventory expenses and improve customer lifetime value. Every year, Hadoop's economies of scale save the company millions in recurring data warehouse expense.

CUSTOMER SEGMENTATION

Business challenge

Luminar, a unit of Entravision Communications Corporation, is the leading big data analytics provider focused on delivering actionable insights on U.S. Latino consumers. Adult U.S. Latinos represent $1.5 trillion in annual purchasing power and advertisers use Luminar's analysis to reach that audience.

Luminar was challenged to store enough empirical data to provide confident recommendations to its retail clients. It wanted to move away from the status quo: the same incomplete and out-of-date sample data as used by other market research businesses. Luminar's data came from 300 sources on nearly 15 million adult U.S. Latinos, including credit card transactions, television set-top data, voter records, and social media. It needed a stable platform to store and process that complex data.

Solution

Luminar chose Apache Hadoop as a solution to meet its requirements for cost, efficiency, and flexibility. Hadoop stores and processes the data that feeds Luminar's targeting models and the company knows that its Hadoop environment can scale with the business. Now Luminar can tackle a greater number of projects and process data for more clients in parallel. Moreover, they can now isolate the unique, actionable marketing signals that their clients demand.

Impact

Luminar increased its data ingest by a factor of eight. It was capturing 2 terabytes of new transactional data per month from 300 raw sources. Now with Hortonworks Data Platform, it ingests 15 terabytes from 2,000 raw sources. Before Hadoop, it could take Luminar several days to ingest and join data to refresh a model. Now it takes only hours. Luminar offers its high-value clients a self-service data exploration portal through a Tableau web interface.

PRICING TRANSPARENCY

Business challenge

Most car buyers can research makes and models online before they visit a nearby dealership to work out the last piece of critical information: the price. Lack of pricing transparency breeds customer mistrust and slows the sales process. This harms both the buyer and the seller. When surveyed, consumers expressed the belief that dealers make an average of 19.7% profit, and that 13.2% would be a fair profit.[3] In reality, the average dealer profit margin is closer to 3-4%.[4]

TrueCar's mission is to provide truth and transparency to increase trust and improve the car buying experience. The company ingests and analyzes immense amounts of data in order to estimate what other buyers paid for a given vehicle. As the data and complexity of analysis increased, TrueCar's existing SQL Server data warehouse became increasingly expensive and difficult to scale.

Solution

TrueCar moved all of its mission-critical price data to a Hadoop data lake. Through Apache Flume, server log data from more than one thousand feeds flows into Hadoop. Data covers approximately 8,400 dealers, 8 million vehicles and 250 million car images.

TrueCar enriches this data via hundreds of processes and offers local price estimates with a high level of confidence through its mobile apps, over 400 branded partner sites, and its flagship site TrueCar.com. The company has helped sell more than 1.4 million cars so far.

Impact

TrueCar realized two profound benefits. First, they can better serve their customers (both buyers and dealers) by storing and processing enough data to make trustworthy price estimates–and the solution scales easily as more cars and dealers come into the system. Just as importantly, the company now stores and processes its data at a cost of 23 cents per gigabyte. A traditional platform would have cost 19 dollars per gigabyte, without any processing capability.

3) TrueCar version of the 2013 Strategic Vision New Vehicle Experience Study

4) National Automobile Dealers Association

BRAND SENTIMENT ANALYSIS

Business challenge

Enterprises lack a fast, reliable way to track their brand health. They may find it difficult to analyze how advertising, competitor moves, product launches or news stories affect how customers perceive their brand. Traditional brand studies or focus groups are slow, expensive and subject to sample bias.

Now sentiment analysis of social media can provide a large volume of current, targeted information on customer feelings. This requires a flexible and cost-effective platform to capture, store and process social streams. Traditional schema-on-load databases are not suited to the highly variable content in social posts, and their relatively high cost limits the number of feasible use cases. Retailers must pick a limited number of hypotheses before exploring the data.

Solution

Hadoop enables quick, unbiased snapshots of brand opinions expressed in social media. Analysts can interactively explore sentiment from Twitter, Facebook, LinkedIn or industry-specific social media streams. This yields daily (or hourly) snapshots on changes to customer opinion. Analysts can also look for words or phrases commonly used by the firm's champions and detractors—for use in future campaigns.

Impact

With better understanding of customer perceptions, retailers align their communications, products and promotions with those perceptions. They can conduct micro-assessments on the ROI on their marketing efforts as measured by positive brand sentiment expressed by specific customer segments, in particular locations, at different times of the year. With Hadoop's schema-on-read architecture, the retailer can explore multiple hypotheses at once (and add others that come up through exploration.)

eCommerce and Customer Service

Hadoop drives better business outcomes in marketing through its ability to improve segmentation, conversion, share-of-wallet and customer lifetime value.

PRODUCT RECOMMENDATION ENGINE

Business challenge

A major specialty department store wanted to improve its product marketing precision. The marketing team wanted to roll out personalized promotions, coupons and product recommendations over multiple customer touch points: in-store, kiosk, web and mobile apps. The company was particularly interested in enabling in-store, real-time product promotion among its shoppers.

But the company's customer data was fragmented, and this prevented it from developing those data-driven marketing promotions. For example, the website or a kiosk should not recommend a product that the same shopper had already purchased in the store. Financial and political obstacles blocked proposals for an IT project to modernize the data architecture.

Solution

Now a Hadoop data lake integrates all the raw data from customers across different product lines. The company ingests and integrates data in real-time and batch, in both structured and unstructured formats. An ETL process uses Apache Pig and Apache Hive to transform the raw data, which is then consumed by Mahout and R machine learning algorithms. The retailer can now deliver real-time recommendations and promotions through all channels, including its website, store kiosks and mobile apps.

Impact

This retailer built an omni-channel recommendation engine similar to what Amazon does online. Thirty-five percent of what consumers purchase on Amazon and seventy-five percent of what they watch on Netflix comes from such product recommendations based on that type of analysis.[5] This retailer can vary recommendations based on weather, loyalty, purchase history, abandoned carts or life stage triggers—and deliver those to shoppers in its stores.

---

5) Big Data and the Creative Destruction of Today's Business Models, A.T. Kearney, 2013, http://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Todays+Business+Models.pdf

WEB PATH OPTIMIZATION

Business challenge

Online shoppers leave billions of clickstream data trails. Clickstream data can tell retailers the web pages customers visit and what they buy. More importantly, the clickstream holds clues as to why online shoppers fail to complete a transaction. This can only be done at scale (ten clickstreams do not a pattern make).

But the huge volume of unstructured weblogs is difficult to ingest, store, refine and analyze for insight. Storing web log data in relational databases requires challenging transformations to convert it to an appropriate schema. It's also expensive. The marginal value of
each new clickstream may not be worth the cost to store, so online marketers make hypotheses and then hand-pick data sets that they believe will either confirm or deny those guesses.

Solution

Web retailers use clickstream data to understand user paths, do basket analysis, run A/B tests and prioritize site updates. Apache Hadoop can store all web logs, for years, at a low cost. Hadoop's flexibility and efficiency at scale mean that online marketers can plan far more tests to run simultaneously on the same raw data. Data access tools like Hive, HBase, Storm and Solr allow data analysts to do both recurring analysis and impromptu data exploration.

Impact

A Hortonworks customer conducts clickstream analysis for one of the world's largest retailers. Clickstream data flowed in at the rate of 100s of megabytes per hour and billions of rows per month. Now the agency stores ad impression files and click files in the same data lake, and then easily joins them for customer insight. With better targeting, their client places fewer, more effective ads, which improves both the overall web experience and its online sales.

In-Store Experience

Retailers use Hadoop to store and process data on how customers interact with their store environments. They use that insight to improve space allocation, display products more effectively, and deliver real-time offers in stores.

PRODUCT PLACEMENT AND STORE LAYOUT

Business challenge

A major omni-channel retailer knew that in-store layout, merchandising and product placement affected sales. Yet the company's brick-and-mortar stores lacked "pre-cash register" visibility into how its customers shopped before they made decisions. The company wanted the same level of customer path visibility and analysis that its clickstream data gave for customers visiting its website.

In-store sensors, RFID tags and QR codes could fill that data gap, but those technologies generate data in formats and volumes that the company's legacy systems were ill-equipped to handle. The retailer became a relatively early adopter of Hadoop because the platform did not enforce a schema-on-load paradigm that would have hampered ingestion and storage of the location data needed for the program.

Solution

The company began testing Apple's iBeacon technology in its flagship stores. iBeacons capture in-store location data from the shoppers' iPhones and Android devices. The data then streams into Hadoop, revealing how customers move through the retail stores (which can be compared to the location of particular product categories). As the iBeacon program grows, Apache Hadoop can store and process that huge volume of sensor and micro-location data.

Impact

Though the results of this specific pilot are not public, the retailer's big data analytics program boosted store sales by 10 percent. As data in Hadoop helps the company optimize its in-store experience, it sees the potential for additional programs that reduce unnecessary inventory and improve customer satisfaction through smarter product placement and updates to store layouts.

IN-STORE PERSONALIZED OFFERS

Business challenge

Another specialty department store wanted to personalize offers and present them in real-time (rather than stay with the traditional, one-size-fits-all approach to promotions).

Data fragmentation was the main obstacle to building those rapid, personalized offers. Both transactional and non-transactional data were stored in its EDW. The legacy architecture's relatively high cost of storage and schema-on-load architecture meant that it was neither economically nor technically feasible to store all the data for as long as the retailer would have liked.

Data needed to flow in from all channels to a single repository, and then the offers needed to flow back out across all channels: to the web, in-store kiosks and mobile apps.

Solution

The company piloted this new Hadoop-enabled capability in five of its stores. Shoppers who walked into one of those locations could opt in for offers via their smartphones. If a shopper lingered in the shoe department she might receive a coupon based on shoes that she looked at online but never purchased. She can redeem this coupon in the store or online. The retailer is testing its hypothesis that customers are more likely to respond to offers while they're already in the store shopping.

Impact

Grocers have long improved sales with product-specific coupons printed on the back of POS receipts, but those coupons are limited, static offers delivered after the purchase. Now this retailer stores data in Hadoop and formulates product recommendations with tools like Mahout and R. Then they deliver those offers across all channels with tools like Apache HBase. That system can deliver multiple offers in real-time and those offers can change dynamically in response to variations in the customer and product mix.

Procurement and Supply Chain

Hadoop drives better business outcomes in procurement and supply chain management by increasing efficiency across the entire value chain. This improves tactical optimization decisions and strategic realignment of the supplier landscape.

INVENTORY MANAGEMENT

Business challenge

Retail margins are always under pressure, and better inventory management is one tool to maintain profitable prices. For example, hype around newly launched products can support a higher price point—but those hot products need to be on the shelves. Retailers that underestimate a huge wave of demand quickly stock-out of those lucrative products and lose sales.

Overstocking excess inventory is similarly costly, leading to mark downs that reduce margins and the customers' perception of value. Some retailers respond by shipping excess inventory to another location, but this adds another cost that erodes margins further. Retailers need to manage inventory with real-time, end-to-end visibility. Just like the story of Goldilocks, the prosperous retailer needs not too little, not too much, but just the right amount of goods on hand.

Solution

Hadoop helps overcome these challenges by providing a real-time view of product sales, leading indicators of product demand, and also a longer, deeper amount of supply chain data. This helps optimize inventory levels and can trigger automatic inventory replenishment. Text, click stream and sensor data complements more structured data on purchase history and product delivery times. This combined data spans warehouses, drop ship vendors, logistics partners, social media tools, and stores.

Impact

One A.T. Kearney study[6] cites a retail chain that "quickly moved hundreds of millions of dollars in store overstocks to various other stores and has since built a predictive model of distribution to limit overstocking altogether. The chain is now capturing pricing, promotion, and loyalty-card data to create even deeper insights into what, when, and why their customers buy." The company now tracks the entire dataset, instead of only the top 100 overstocked SKUs.

6) *Big Data and the Creative Destruction of Today's Business Models, A.T. Kearney, 2013, http://www.atkearney.com/documents/10192/698536/Big+Data+and+the+Creative+Destruction+of+Todays+Business+Models.pdf*

EFFICIENT ORDER PICKING

Business challenge

Order picking in a logistics warehouse is a labor-intensive process. A fulfillment center may do piece picking, zone picking, wave picking, pick-to-box, or some combination of these methods. But this is a repeatable, high-volume process similar to a manufacturing assembly line. Both real-time and historical analysis of order picking data is useful for continuous improvement of those operations. Faster order picking means faster shipments, which improves order fulfillment and customer satisfaction.

Large online retailers like Amazon use a range of automated mechanisms to pick orders. Now Hadoop's storage and processing efficiencies can bring some of that power to smaller retailers who want to improve their order-picking processes (without robotics or millions in up-front investment).

Solution

Hadoop stores data from different sources like orders, product inventory, warehouse configurations, and workflow sensors. This can then be analyzed according to the company's priorities within the overall picking process, for trouble-shooting and input to staffing and equipment decisions. With enough data, machine-learning algorithms can simulate the order picking process and recommend optimizations.

Impact

WD, a Western Digital company, is one Hortonworks customer[7] that optimizes its manufacturing process with HDP. Fulfillment warehouses can achieve the same type of results: faster response to changing conditions, lower levels of breakage and returns, and cost reductions. Changes can be simulated with BI tools running with Hadoop data, before rolling out the final process to the warehouses and stores.

7) http://hortonworks.com/customer/western-digital

VENDOR MANAGEMENT

Business challenge

Most retailers work with multiple vendors in their supply chains. These include drop ship vendors, third party logistics vendors, transportation vendors, and packaging vendors. For busy retailers, it can be difficult to analyze and rank the relative quality of their vendors. Hadoop-driven analysis provides an objective view of vendor performance against a set of key performance indicators. These KPIs include vendor profitability, on-time service, customer feedback and complaints.

In addition, product vendors are generating ever more digital data on their offerings and inventory levels, which they make available to retailers. Retailers that ingest that data and integrate it with their own operations can manage the relationships more effectively. For example, they might establish alerts to request on-demand shipment of hot products as sales velocity data predicts stock-outs.

Solution

Retailers integrate Hadoop with vendor data feeds and then add their own ERP data, social network feeds related to vendor deliveries, POS data and inventory numbers. Like many other Hadoop data lake solutions, this combined data is useful for a variety of applications. Historical batch processing can make purchase recommendations across multiple vendors and products. Interactive SQL data exploration yields insights helpful in vendor negotiations. Real-time streaming analysis can generate alerts if vendor KPIs do not stay within pre-defined ranges.

Impact

Unified analysis of vendor KPIs ensures that retailers optimize their vendor relationships without additional effort or changes to existing operations. Vendors also benefit from improved data transparency, since they receive clear signals from the retailer about what they must do to keep her business. In fact, greater data availability means that both vendors and retailers spend less time on the phone with each other—leaving more time to focus on delivering and selling product.

STRATEGIC SOURCING

Business challenge

Sophisticated retailers take a holistic approach to optimizing the performance of their supplier base. They are looking for ways to exploit their commercial buying power through strategies such as volume concentration, component price transparency, and variable or index pricing. These best-in-class retailers also aim to create technical advantages through product specification changes, process improvements, or by restructuring value chain relationships.

Retailers typically use auctions for strategic sourcing, but auction data is complex. For example, one retailer launched an auction for outbound and reverse logistics services. More than 100 carriers bid for some or all of 10,000 discrete routes, each with more than 20 characteristics like "frequency" and "temperature control". Traditional systems fail once the analysis is extended to the unstructured text, image and GPS data contained in bids. And when the same auction is repeated, data on the nature and outcomes of previous auctions is rarely available.

Solution

Hadoop provides the ideal platform to combine, correlate and interrogate these diverse data sources. This enables the retailer to run multiple, simultaneous simulations to weigh the outcome of various award combinations. Suppliers are encouraged to provide complex inputs, with as many bundles, conditional offers and creative deal terms as possible. In some cases, these bundles cover 3 to 5 years, several sourcing groups, many countries and a few hundred million dollars of expense annually. Without a data system that can scale with the amount and complexity of the data, more vendor information is a liability that overwhelms decision-makers. With Hadoop, more data means better decisions that increase profits.

Impact

Retailers who implement these types of data-rich, rigorous strategic sourcing initiatives typically generate 10–15% savings in cost of goods sold. Moreover, they build an internal data asset for year-over-year analysis that helps improve the procurement process. The holiday shopping season comes the same time every year, but purchasing teams don't need to start every new year from scratch. Now they can "remember" years of detailed purchasing decisions before they make the next one. Hadoop enables that.

With Hadoop, retailers can predict customers' purchasing and buying behaviors to develop tailored pricing, space, and assortment at stores. They can improve their contact with customers across all channels: online, phone or in the store. They can optimize global sourcing to drive efficiencies and reduce the cost of goods sold.

These retail company examples show what enterprises are learning in other industries: Hadoop presents superior economics compared to legacy data warehousing and storage technologies and it also uncovers exciting new capabilities for growing the business.

Build a Modern Retail Data Architecture with Enterprise Hadoop

To realize the value of your investment in big data, use the blueprint for Enterprise Hadoop to integrate with your EDW and related data systems. Building a modern data architecture enables your organization to store and analyze the data most important to your business at massive scale, extract critical business insights from all types of data from any source, and ultimately improve your competitive position in the market and maximize customer loyalty and revenues. Read more at: http://hortonworks.com/hdp

Hortonworks Data Platform provides Enterprise Hadoop

Hortonworks Data Platform (HDP) is powered by 100% open source Apache Hadoop. HDP provides all of the Apache Hadoop related projects necessary to integrate Hadoop alongside an EDW as part of a Modern Data Architecture. It ships with efficient Data Management and versatile Data Access capabilities, along with three capabilities enterprises require for widespread adoption: Data Governance & Integration, Security and Operations.
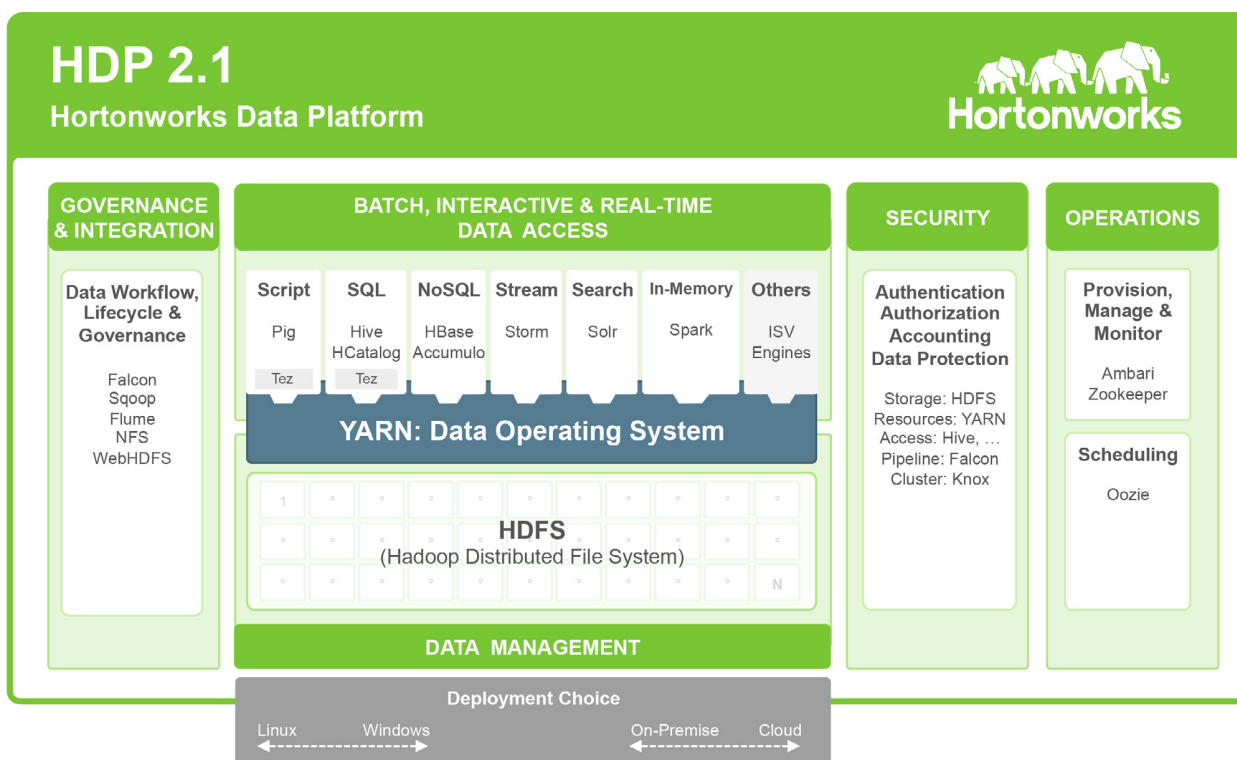

Figure 2: Five core capabilities—data governance & integration, data management, data access, security, and operations.

Why Hortonworks for Hadoop?

Founded in 2011 by 24 engineers from the original Yahoo! Hadoop development and operations team, Hortonworks has amassed more Hadoop experience under one roof than any other organization. Our team members are active participants and leaders in Hadoop development, designing, building and testing the core of the Hadoop platform. We have years of experience in Hadoop operations and are best suited to support your mission-critical Hadoop project.

For an independent analysis of Hortonworks Data Platform and its leadership among Apache Hadoop vendors, you can download the Forrester Wave™: Big Data Hadoop Solutions, Q1 2014 report from Forrester Research.

About Hortonworks

Hortonworks develops, distributes and supports the only 100% open source Apache Hadoop data platform. Our team comprises the largest contingent of builders and architects within the Hadoop ecosystem who represent and lead the broader enterprise requirements within these communities. Hortonworks Data Platform deeply integrates with existing IT investments upon which enterprises can build and deploy Hadoop-based applications. Hortonworks has deep relationships with the key strategic data center partners that enable our customers to unlock the broadest opportunities from Hadoop. For more information, visit www.hortonworks.com.