

E-Commerce and NoSQL in Retail



Donald Soares is chief technology officer of the retail and consumer industry at MarkLogic. Donald.Soares@marklogic.com

Donald Soares

Abstract

Few industries have access to more data regarding consumers, products, and channels than the retail and consumer industries. Data-derived insights should be at the heart of what drives this business. Yet beyond the hype, most attempts at using big data to build a competitive advantage have been dismal failures. This is especially true for e-commerce initiatives. This article examines the reasons for the failures, and what can be done to fix the problem.

Introduction

In retail today, an estimated 80 percent of new data sources involving in-store transactions, e-commerce, and consumer and product information are not considered for analysis. That's not terribly surprising considering that retailers face immense challenges arising from the sheer volume, velocity, and variability of the data they must manage.

Consider a few examples of complex big data projects and the problems they attempt to solve:

- Data volume makes it difficult for A-Z Retail to gain insights from the one billion consumers who buy its products each week so it can engage with customers through targeted, personalized messages.
- The complexities caused by data velocity prevent Retail 123 from effectively managing its global supply chain and being able to analyze real-time data from sensors, RFID chips, shipment notifications, and receipts.
- Retailer X wants to manage a customer loyalty database of 70 million consumers—combining demographics, purchase history, preferences, and location information—and leverage it for live consumer purchase transactions and promotions both online and in-store. How will the company manage the variability of this data?

These considerations are everyday challenges to retailers that typically rely on mainframes or relational database management systems (RDBMSs)—neither of which were designed to handle the complexities of modern data management. We are increasingly seeing that mainframes and RDBMSs that are traditional to retail lack the flexibility and scalability to handle the volume, velocity, and variability issues that are inherent to big data.

Why Do Big Data Initiatives Fail in Retail?

The major reasons for the failure of big-data initiatives in retail include the following:

Challenges with Structure and Schema

Retail systems largely utilize RDBMSs that are driven by Oracle, IBM DB2, Microsoft, and SQL as the programming language for managing the data stored within them. Ironically, in the case of retailers, structure has become one of the primary challenges of using RDBMSs to store and manage content. For RDBMSs to perform well, data flowing into them must first be mapped with a predefined schema or a set of constraints that define how it is structured and organized for analysis. RDBMSs assume a rigid tabular structure in which all entries in a column are of the same data type. For instance, addresses in the United States typically have a street number, city, state, and ZIP code. Because this data format is always consistent, address information will fit neatly into a relational table that has columns defined as “street,” “city,” “state,” etc.

Managing Unstructured Data Sources

As you can see, the schema-driven rows and columns of the RDBMS approach work well enough for the internal structured data of enterprise resource planning (ERP) or point of sale (POS). The problem is that it doesn't work so well for managing relatively unstructured data, including:

- External unstructured data from consumer activity. This type of data provides rich information, but unfortunately it doesn't fit well into rows and columns for analysis. Examples include blogs, tweets, digital images, video, social media commentary, clickstream data, XML, and HTML Web pages.

- Internal unstructured data. This type of data is vital to your organization's internal capabilities and expertise, but right now it's stored all over the place in multiple legacy systems and databases and therefore not used. Examples include PowerPoint presentations, PDF files, recipe sites, SharePoint site posts, and online forums that capture expertise.
- External structured data. This type of data is currently used only marginally and includes loyalty, location, credit, competitive, demographic, and census data as well as machine-sensor and mobile data that would make for a logical extension of consumer analytics and product development.

Getting Data into the Database

Data ingestion, or getting content into a relational database, is daunting. To make it fit within a predefined row and column structure, users first need to analyze the details of the content to identify the schema and then map it into rows and columns. This is a costly, time-consuming first step that many can find insurmountable. Understandably, many a big data project has met with a premature death at just this stage. Imagine trying to develop the perfect schema for a data model that manages data on a billion consumers across multiple countries, brands, and channels. You get the picture.

Enabling Analytics and Deriving Insight

Then there is the question of integrating and analyzing the data to derive insights. In retail, having a real-time operational and transactional system is critical because you want to make live promotional decisions online or in-store, check product availability by channel, and respond to consumer queries immediately and accurately. This is not so easy when data is stored across different legacy systems that were never designed to “talk” to each other. This is illustrated by the fact that loyalty data is almost never linked to social-media data, and online-purchase data is rarely linked to store-location or loyalty-card data.

Data Challenges to Retail E-Commerce

The case for e-commerce is solid. It's the fastest-growing segment of the retail industry today. According to the Census Bureau, in Q2 2015, online sales (which

accounted for 7.2 percent of total retail sales) increased 14.1 percent over the same quarter in 2014 (Department of Commerce, 2015) whereas retail sales grew just 1.0 percent in that one-year period. To put this in perspective, Walmart's global e-commerce sales rose by 22 percent to \$12.2 billion while total retail sales grew by just 1.8 percent. This trend shows no signs of slowing down.

In fact, when announcing a recent \$1 billion investment in e-commerce at its shareholder meeting in June 2015, Walmart CEO Doug McMillon was quoted as saying, "One customer can shop with us in so many different ways—in stores, on their phones, at home—we'll win one customer at a time." (Tabuchi, 2015)

With numbers like \$12.2 billion on the line, it's easy to understand a \$1 billion investment in e-commerce. Still, the sobering reality is that e-commerce continues to face significant challenges and missed opportunities because success will boil down to converting online visitors into buyers. According to an e-commerce report from Q4 2014, a mere 2.84 percent of visitors to e-commerce websites actually bought anything. Worse, analysis of traffic patterns showed that fewer than 1 percent of shoppers on smartphones made a purchase. (Monetate, 2015)

Even when online shoppers added items to their carts, two out of three (66 percent) did not end up completing the transaction at all. (Monetate, 2015) This is actually more disastrous than it might seem. After all, for the giant e-commerce retailers (such as Walmart, Best Buy, and Target) the stakes are incredibly high. A mere increase from 3 to 5 percent in online sales conversions would result in sales growth of \$2 billion. That's difficult to ignore.

Why are retailers missing out on potentially lucrative e-commerce opportunities? At a high level, most retailers are unable to effectively manage and extract value from big data with the database technologies that are common in the industry.

1. Product data complexity and management. A typical retailer's product data is extremely complex, consisting of structured and unstructured data that needs to be ingested and integrated into a database.
2. Issues with search and updates. Can your consumers find what they're looking for, or is your search functionality stuck in the Dark Ages? Regrettably, online retail suffers from a classic case of the "Goldilocks Syndrome" in that product search results—much like the three bears' household goods—are rarely just right for consumers. Following are a few search scenarios that you can consider testing for yourself:
 - You want to buy a wireless sound system for your home. An online search for "wireless sound systems" on an electronics retailer's website might yield 1,271 results. If you tried to narrow the search by clicking "wireless," you'd be down to 271 options. For comparison, a similar search on a mass merchandiser's site would yield 13,232 results, and filtering on "audio speaker systems" would reduce the results to 1,180. The overall online shopping experience would be quite different from walking into a store where an associate would point you to the wireless sound section that holds just two major brands and talk you through the differences between them.
 - You could try to buy a 17-inch laptop on an office supply website by performing a search for "17-inch laptops." Your top 10 recommendations would include five privacy filters followed by a series of laptop bags, which is certainly not likely to induce anyone to buy a laptop computer!
 - Alternatively, a search for a "summer male jacket" on a general merchandiser's site may yield a woman's denim vest, a Barbie doll summer dress, and two sets of earrings. For contrast, a similar search on a different retailer's site might yield a

man's cologne, four CDs, and an extension cord. Online search clearly delivers a very different experience from what you can find walking into a store and asking about the latest summer styles in jackets.

Complicating matters are the issues caused by product updates. Product-related data changes with new models, product innovations, and the addition of new options (such as color, size, and packaging). Unfortunately for most e-commerce search technologies that are bolted onto relational databases, the pre-existing schema that determines how the data is to be sorted and searched restricts the number of attributes that can be associated with the data. This prevents consumers from filtering down intelligently through the options available.

As a result, the consumer search is constrained and often fruitless! In addition, many searches lack context and use a "bag of words" approach without benefit of intelligence. This results in consumers being steered to the wrong product options. Finally, from a development perspective, updating predefined schemas is difficult and requires significant additional coding and time to add new product features and attributes.

3. Relationships and context between product data and information. Remember consumers buy solutions to meet their needs, not products. Consider these examples:

- Recipes for dinner (e.g., wine and cheese pairings)
- Entertainment solutions (e.g., TV + DVR + cables + delivery/installation + service)
- Products and accessories (e.g., printer + correct toner cartridge)

In the absence of context or links between product data, the retailer loses its ability to cross-sell and upsell (both in-store and online) via intelligently linked product recommendations.

The NoSQL Value Proposition for E-Commerce

These data disconnects have created a need for a more flexible and scalable database that can easily operate in today's modern infrastructure. Traditional mainframes and RDBMSs lack the flexibility and scalability to handle the volume, velocity, and variability inherent in big data. NoSQL technology represents a transformation in perspective. Instead of getting the schema just right before doing anything else, NoSQL advocates loading up the data first and then seeing where the problems are. This problem-oriented approach focuses on how the data will be used (queried) rather than how the data must be structured to fit within a traditional RDBMS.

Essential Attributes of an Enterprise NoSQL Platform

An Enterprise NoSQL database platform provides a flexible answer to the challenges we've outlined. What differentiating capabilities should retailers look for when considering NoSQL for e-commerce? The following are essential.

1. Product data management. Is the NoSQL database the right fit for complex data management? For retailers, the scale and complexity of product data that needs to be ingested is a major barrier that must be addressed up front. The database must be optimized to store JSON, XML, RDF, and geospatial data. It must also ingest other types of data—from RDF relationships to text, geospatial data, binary video files, and PDFs—without the need for conversion. Simply put, your users should expect better answers because they start with better data.
2. Ease of development and changes to product data. One of the major problems with schema-based relational databases is that they limit attributes based upon which searches can be conducted, and they are difficult to update. Also, changes take considerable development time. Your NoSQL database should allow you to change the data without mapping it to a fixed schema or hiding data in opaque objects. You can still store all the information you would find in the row of a relational table, but because it is stored and indexed as documents, you don't have to nor-

malize the data and you don't have to worry about how the shape of the data changes over time. This leaves you enormous amounts of time and energy that would ordinarily be invested in ETL processes, and you also gain agility with future development.

3. Superior search capabilities. A NoSQL database must also provide a superior search experience without limits on attribute search and with the flexibility to drill down into the data to research alternative products and features. The database must have the search features that users expect from an enterprise search application, such as type-ahead suggestions, relevance ranking, facets, snippeting, highlighted search terms, proximity boosting, relevance ranking, and language support. To reiterate, search functionality should be integrated into the database so you don't have to bolt on capabilities from another solution. This simplifies your architecture and makes management incredibly easy for DBAs and developers. Having integrated search means one less platform to worry about.
4. Semantic capabilities to provide context and relationships. The best NoSQL database solutions use semantics to store billions of relationships between associated and linked product types. For example, if you're looking to buy a particular brand and model of HDTV, semantics determine what cables, sound box, or service and installations are linked with that model. With semantics, you can store and query these billions of facts and relationships and infer new facts. These facts and relationships provide context for a better search that enables users to:
 - Find more relevant information by expanding the terms entered
 - Present more and better information about whatever the user is searching for
 - Publish information dynamically to Web, print, or mobile

A Transformational Opportunity for Retail E-Commerce

E-commerce represents the most significant growth opportunity for retail, but it's also the channel that causes consumers frustration in terms of search and fulfillment. An enterprise NoSQL data platform will drive e-commerce sales growth while empowering retailers to provide better value and convenience to digital consumers. Although considered critical, simplified product data management and intelligent search and semantic linkages are merely the tip of the iceberg of enterprise NoSQL opportunities.

The shift to NoSQL could save you from spending a year figuring out the perfect data model and schema to analyze and store data on 70,000 products and a billion consumers. Instead, you can load the data, index it automatically, and enable consumers to intelligently query it to find what they need quickly. An enterprise NoSQL database is technology for mere mortals (no prognostication needed for perfect data modeling) and presents a transformational opportunity for retailers in terms of e-commerce sales growth and customer satisfaction and loyalty.

References

- U.S. Department of Commerce, Bureau of the Census [2015]. Quarterly Retail E-Commerce Sales, 2nd Quarter 2015, August 17. https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf
- Tabuchi, Hiroko [2015]. "Wal-Mart, Lagging in Online Sales, Is Strengthening E-Commerce," *New York Times*, June 5.
- Monetate [2015]. Ecommerce Quarterly Q4 2014, Feb. 17.