Introduction

The volume, variety, and velocity of data being produced in all areas of the retail industry is growing exponentially, creating both challenges and opportunities for those diligently analyzing this data to gain a competitive advantage. Although retailers have been using data analytics to generate business intelligence for years, the extreme composition of today's data necessitates new approaches and tools. This is because the retail industry has entered the big data era, having access to more information that can be used to create amazing shopping experiences and forge tighter connections between customers, brands, and retailers.

A trail of data follows products as they are manufactured, shipped, stocked, advertised, purchased, consumed, and talked about by consumers – all of which can help forward-thinking retailers increase sales and operations performance. This requires an end-to-end retail analytics solution capable of analyzing large datasets populated by retail systems and sensors, enterprise resource planning (ERP), inventory control, social media, and other sources.

How does one start a big data project? In an attempt to demystify retail data analytics, this paper chronicles a real-world implementation that is producing tangible benefits, such as allowing retailers to:

- Increase sales per visit with a deeper understanding of customers' purchase patterns.
- Learn about new sales opportunities by identifying unexpected trends from social media.
- Improve inventory management with greater visibility into the product pipeline.

A set of simple analytics experiments was performed to create capabilities and a framework for conducting large-scale, distributed data analytics. These experiments facilitated an understanding of the edge-to-cloud business analytics value proposition and at the same time, provided insight into the technical architecture and integration needed for implementation.
Big Data Basics

Data has been getting bigger for a while now. The volume of data generated or processed in 2014 alone is expected to exceed six zettabytes; that is 1,200 times more than all the data ever generated prior to 2003.

Unstructured Data

One of the reasons data is getting bigger is it is continuously being generated from more sources and more devices. Making matters more difficult, much of the data is unstructured, coming from videos, photos, comments on social media forums, reviews on web sites, and so on. As a result, this data is often made up of volumes of text, dates, numbers, and facts that are typically free form by nature and cannot be stored in structured, predefined tables.

Certain data sources are arriving so fast, there may not be enough time to store them before applying analytics to them. And that is why conventional data analytics and tools alone do not enable retail IT to store, manage, process, and analyze all the data they may need to utilize.

Competitive Advantage

So what if retailers just ignore big data; after all, is it worth all the effort? It turns out that it is. According to McKinsey Global Institute, big data has the potential to increase net retailer margins by 60 percent. Likewise, companies in the top third of their industry in the use of data-driven decision making were, on average, five percent more productive and six percent more profitable than their competitors, wrote Andrew McAfee and Erik Brynjolfsson in a Harvard Business Review article.

In order to generate the insights needed to reap substantial business benefits, new innovative approaches and technologies are required. This is because big data in retail is like a mountain, and retailers must uncover those tiny, but game-changing golden nuggets of insights and knowledge that can be used to create a competitive advantage.

Big Data Technologies

In order for retailers to realize the full potential of big data, they must find a new approach for handling large amounts of data. Traditional tools and infrastructure are struggling to keep up with larger and more varied data sets coming in at high velocity. New technologies are emerging to make big data analytics scalable and cost-effective, such as a distributed grid of computing resources. Processing is pushed out to the nodes where the data resides, which is in contrast to long-established approaches that retrieve data for processing from a central point.

Hadoop is a popular, open-source software framework that enables distributed processing of large data sets on clusters of computers. The framework easily scales on servers based on Intel Xeon processors, as demonstrated by Cloudera, a provider of Apache Hadoop-based software, support, services, and training. Although Hadoop has captured a lot of attention, other tools and technologies are also available for working on different types of big data analytics problems.

Real-World Use Cases

Intel teamed up with Living Naturally, a trusted name in retail technology, in order to demonstrate big data use cases for 3,000 natural food stores in the United States. Living Naturally develops and markets a suite of online and mobile programs designed to enhance the productivity and marketing capabilities of retailers and suppliers. Since 1999, Living Naturally has worked with thousands of retail customers and 20,000 major industry brands.

In a period of three months, a team of four Intel analysts worked with Living Naturally on a retail analytics project spanning four key phases: problem definition, data preparation, algorithm definition, and big data solution implementation. During the project, the following use cases were investigated in detail:
• **Product Pipeline Tracking**: When inventory levels are out of sync with demand, make recommendations to retail buyers to remedy the situation and maximize return for the store.

• **Market Basket Analysis**: When an item goes on sale, let retailers know about adjacent products that benefit from a sales increase as well.

• **Social Media Analysis**: Prior to products going viral on social media, suggest retail buyers increase order size to be more responsive to shifting consumer demand and avoid out-of-stocks.

A key objective of the team was to ensure the big data solution complemented how people do their jobs. This meant the data had to be presented to retail buyers in an actionable and easy-to-use manner. This is best illustrated by the solution's intuitive user interface, which is shown in Figure 1 and described in the following sections:

**Product Pipeline Tracking**

Figure 2 defines the fields of the retail buyer’s user interface (UI) used to order products. In this case, the product is cashews, and the upper left corner shows the orders (30, 55, 30, 25) the retail buyer made over the last four weeks. Next to the graph is a shopping cart, indicating the buyer plans to order 20 more. Moving right along the UI, the next shipment will contain a partial order of 18 more products arriving from Houston, Texas, as seen in the lower left corner. Currently, there are 120 units in inventory.

The retail buyer has a good track record ordering this product, getting an A grade and a ‘thumbs up’ from the tool. The tool makes a suggestion to increase the reorder from 18 currently in the shopping cart to 38 units. The buyer can override the suggestion by moving the ‘smiley face’ left or right to decrease or increase the amount of the next order.
Social Media Analysis

In the previous example, why is the tool suggesting a greater order size when 120 units are already in inventory, and the prior orders have kept pace of demand and then some? Because the health benefits of cashews were discussed in a recent social media post, which can be read by clicking the ‘Twitter’ bird in the UI shot in Figure 3. The tool estimated the demand increase created by this tweet and other related social media postings (also reflected in the retailer’s own increased sales numbers), and recommended a suitable reorder quantity for cashews.

Market Basket Analysis

The tool also lets the retail buyer know about associated products for cashews, recommending the buyer increase orders of papayas, which customers are more likely to purchase along with cashews due to another social media-driven microtrend (Figure 4). In other words, if cashew sales are about to spike due to social media postings, papaya sales are likely to increase as well.

Problem Definition

Before starting a big data project, it is important to have clarity of purpose, which can be accomplished with the following steps:

Identify the Problem

Retailers need to have a clear idea of the problems they want to solve and understand the value of solving them. For instance, a clothing store may find that four of ten customers looking for blue jeans cannot find their size on the shelf, which results in lost sales. Using big data, the retailer’s goal could be to reduce the frequency of out-of-stocks to less than one in ten customers, thus increasing the sales potential for jeans by over 50 percent.
Big data can be used to solve a lot of problems, such as reduce spoilage, increase margin, increase transaction size or sales volume, improve new product launch success, or increase customer dwell time in stores.

Working with a retail solution provider to 3000 natural food stores, the Intel team identified an out-of-stocks problem associated with products going viral thanks to social media and Internet postings of a celebrity or expert endorsing them. For example, a popular medical doctor with a health-oriented television show recommended raspberry ketone pills as a weight reducer to his very large following on Twitter*. This caused a run on the product at health stores and ultimately led to empty shelves, which took a long time to restock since there was little inventory in the supply chain.

**Identify All Data sources**

Retailers collect a variety of data from point-of-sale (POS) systems, including the typical sales by product and sales by customer via loyalty cards. However, there could be less obvious sources of useful data that retailers can find by “walking the store,” a process intended to provide a better understanding of a product’s end-to-end life cycle. This exercise allows retailers to think about problems with a fresh set of eyes, avoid thinking of their data in silo’d terms, and consider new opportunities that present themselves when big data is used to find correlations between existing and new data sources, such as:

**Video:** Surveillance cameras and anonymous video analytics on digital signs

**Social Media:** Twitter, Facebook*, blogs, and product reviews

**Supply Chain:** Orders, shipments, invoices, inventory, sales receipts, and payments

**Advertising:** Coupons in flyers and newspapers, and advertisements on TV and in-store signs

**Environment:** Weather, community events, seasons, and holidays

**Product Movement:** RFID tags and GPS

The Intel team focused on the data sources and business solutions listed in Table 1. Using social media chatter, the team sent useful product information and desired promotions to customers based on their interests. In addition, social media and supply chain data were analyzed together to minimize out-of-stocks and overstocks, which will be described in more detail in a following section.

**Conduct ‘Cause and Effect’ Analysis**

After listing all the possible data sources, the next step is to explore the data through queries that help determine which combinations of data could be used to solve specific retail problems. For the Intel team, this meant using social media data to inform health stores up to three weeks before a product is likely to go viral, providing enough time for retail buyers to increase their orders. The team queried social media feeds and compared them to historic store inventory levels and found it was possible to give retail buyers an early warning of potentially rising product demand. In this case, it was important to find a reliable leading indicator that could be used to predict how much product should be ordered and when.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Social Media</th>
<th>Supply Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Sources</td>
<td>• Twitter*</td>
<td>• Orders</td>
</tr>
<tr>
<td></td>
<td>• Facebook*</td>
<td>• Invoices</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Shipments</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Receipts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Transactions</td>
</tr>
<tr>
<td>Business Solutions</td>
<td>• Implement 1:1 digital marketing</td>
<td>• Increase new product sales</td>
</tr>
<tr>
<td></td>
<td>• Send personalized promotions</td>
<td>• Maximize profit by understanding retail buyer behavior</td>
</tr>
<tr>
<td></td>
<td>• Influence customer sentiment</td>
<td>• Make product recommendations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Measure marketing campaign effectiveness</td>
</tr>
</tbody>
</table>

Table 1. Data Sources and Business Solutions
Determine Metrics That Need Improvement
After clearly defining a problem, it is critical to determine metrics that can accurately measure the effectiveness of the future big data solution. A recommendation is to make sure the metrics can be translated into a monetary value (e.g., income from increased sales, savings from reduced spoilage) and incorporated into a return on investment (ROI) calculation.

The Intel team focused on metrics related to reducing shrinkage and spoilage, and increasing profitability by suggesting which products a store should pick for promotion.

Verify the Solution Is Workable
A workable big data solution hinges on presenting the findings in a clear and acceptable way to employees. As described previously, the Intel team made sure the solution provided retail buyers with timely product ordering recommendations displayed with an intuitive UI running on devices used to enter orders.

Consider Business Process Re-engineering
By definition, the outcomes from a big data project can change the way people do their jobs. Will the data simplify or complicate things for employees? Will employees need to be trained on a new device? Will the POS system need to be modified to deliver a different data set? It is important to consider the costs and time of the business process re-engineering that may be needed to fully implement the big data solution.

Calculate an ROI
Taking all the inputs previously discussed, retailers should calculate an ROI to ensure their big data project makes financial sense. At this point, the ROI is likely to have some assumptions and unknowns, but hopefully these will have only a second order effect on the financials. If the return is deemed too low, it may make sense to repeat the prior steps and look for other problems and opportunities to pursue. Moving forward, the following steps will add clarity to what can be accomplished and raise the confidence level of the ROI.

Data Preparation
Retail data used for big data analysis typically comes from multiple sources and different operational systems, and may have inconsistencies. Particularly for the Intel team, transaction data was supplied by several systems, which truncated UPC codes or stored them differently due to special characters in the data field. More often than not, one of the first steps is to cleanse, enhance, and normalize the data.

The Intel team performed a number of data clean tasks:
- Populate missing data elements
- Scrub data for discrepancies
- Make data categories uniform
- Reorganize data fields
- Normalize UPC data

Algorithm Definition Process
After identifying the retail data sources, a set of algorithms can be selected for data exploration. This is a trial-and-error process because it is very difficult to assess the effectiveness of an algorithm before applying it to a particular data set. Algorithms are like atomic pieces, and figuring out which ones are needed and work together is like assembling a puzzle.

Table 2 lists commonly used algorithms and the types of business problems they help to solve.
<table>
<thead>
<tr>
<th>Analysis Technique</th>
<th>Algorithms</th>
<th>Business Problems</th>
</tr>
</thead>
</table>
| **Classification** | ❘ Decision trees, neural network, and Naïve Bayesian Networks.  
❘ A Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional independencies. | ❘ Churn Analysis: Learn which customers are most likely to switch to a competitor.  
❘ Risk Management: Decide whether a loan be approved for a particular customer.  
❘ Directed Advertising: Personalize advertisements (in-store, online) to better attract customer attention. |
| **Business Solutions** | ❘ Similar to classification, logistic and linear regression techniques look for patterns that determine a numerical value. | ❘ Predict coupon redemption rate based on face value, distribution method, distribution volume, and season. |
| **Clustering/Segmentation** | ❘ These algorithms assign a set of observations into subsets (clusters), which are similar according to pre-designated criteria.  
❘ They allow for different assumptions on the data structure and evaluations based on internal compactness within a cluster and/or separation between different clusters. | ❘ Maximize profit by creating a profile of perfect buyer behaviors.  
❘ Customer Segmentation: Determine behavioral and descriptive profiles of customers. |
| **Association Rules** | ❘ Association rule learning is a method for discovering interesting relations between variables in large databases.  
❘ The method helps identify items that frequently appear together and determine rules about the associations. | ❘ Identify product adjacency for cross-selling purposes.  
❘ Determine frequently shopped products.  
❘ Perform market basket analysis. |
| **Forecasting/Prediction** | ❘ This approach estimates the value of a variable of interest at some specified future date.  
❘ It uses formal statistical methods, employing time series, cross-sectional or longitudinal data, and techniques that deal with seasonality, trending, and noisiness of the data. | ❘ This approach estimates the value of a variable of interest at some specified future date.  
❘ It uses formal statistical methods, employing time series, cross-sectional or longitudinal data, and techniques that deal with seasonality, trending, and noisiness of the data. |
| **Sequence Analysis** | ❘ This method finds patterns in a series of events.  
❘ Both sequence and time-series data are similar; they contain adjacent observations that are order-dependent. The difference is that where a time series contains numerical data, a sequence series contains discrete states. | ❘ Model customer purchases as a sequence, (e.g., a customer first buys a computer, and then buys speakers, and finally buys a webcam). |
| **Anomaly Detection** | ❘ These are distance based techniques: k-nearest neighbor, cluster analysis-based outlier detection, pointing at records that deviate from learned association rules.  
❘ They detect patterns (i.e., anomalies) in a given data set that do not conform to an established normal behavior and can be used to validate data entry and identify outliers from expected norm. | ❘ Detect credit card fraud and network intrusion.  
❘ Conduct manufacturing error analysis. |
| **Computer Vision** | ❘ This type of algorithm translates images/high-dimensional data from the real world to provide numerical or symbolic information for decision-making.  
❘ Use models are constructed with the aid of geometry, physics, statistics, and learning theory for vision perception.  
❘ Automated image analysis is performed on video sequences to multiple camera views with many applications. | ❘ Detect events (surveillance), model objects interaction, reconstruct scenes, recognize objects, etc. |

Table 2. Algorithm Examples
Social Media Analysis

The goal of this data exploration was to determine if social media could be used to identify changes to customer demand for products. The initial premise was that social media analysis could identify emerging social topics and viral events, and correlate them with products and sales. This capability would enable retailers to detect micro trends and events early enough to make informed ordering, pricing, and product promotion decisions. The Intel team focused on known and emerging influencers in the social media world.

Figure 5 shows the data exploration approach, which basically looked for correlations between social media events and changes in the sales of related products. The input data included Google* Trends, tweets from a popular doctor with a TV show, product attributes, and business-to-consumer (B2C) point-of-sale data.

An example of the output is presented in Figure 6. The blue line shows the trend for social media activity related to raspberry ketone, a potential weight-reducing compound found in red raspberries. The trend line peaks twice (weeks two and four), which is followed by an increase in sales about a week later, as shown by the green line.

Big Data Analytics Solution Implementation

The Intel team implemented an analytics solution architecture based on a Cloudera Distribution of the Apache Hadoop platform, which could handle the large volume, velocity, and variety of data that has to be processed. Supply chain and sales data were analyzed along with social media data relevant to the natural foods domain; however, the framework was designed to scale for additional data like weather, sensor data from the Internet of Things (IoT), or other sources that are of value to retailers.

Figure 7 shows a conceptual solution stack of the big data analytics solution, containing layers of features required to analyze large volumes of data and gain actionable insights from the chosen algorithms. The left side of the figure shows the data inputs, and the right side lists elements of the framework spanning the computing infrastructure through to the retailer buyer UI.
Capabilities and Software Components
Various solution components and associated technologies were used, which are highlighted in Figure 8. At a high level, Twitter data is streamed live using the fire hose API exposed by Twitter, and supply-chain data is normalized, processed, prepared, and stored in relational databases. These databases are loaded in Hadoop Distributed File System (HDFS) and HBase, which is a NoSQL database. The data is processed using MapReduce programs, machine learning algorithms, and statistical tools, and the output is recommendations for retail buyers as well as actionable interactive visualizations.

Social Media Analysis
To enable the social media analysis capabilities, a base framework was created to access and store live Twitter streams, depicted by the process flow in Figure 9. The known social media influencers were identified as previously discussed. Public chat trends were tracked by extracting content from tweets, messages were tokenized and tagged using natural language processing techniques, and sentiment scores calculated using open source libraries. The social media analysis was tied to product sales trends to make recommendations.
Product Pipeline Tracking

A retailer’s profitability is closely tied to supply chain management, a critical function that helps ensure the right amount of inventory is on hand. Without proper product management, even small variations in customer demand can lead to major inventory imbalances as business partners in the supply chain try to make adjustments with imperfect information. Inventory can swing wildly, creating an unstable situation known as the bullwhip effect, which results in stock-outs or excess inventory.

It is possible to avoid these inventory issues by providing effective communication and coordination of the supply chain after connecting the dots between orders, inventory, and sales. For example, retailers can improve information sharing by providing sales insight to the supply chain so everyone’s demand predictions are based on the same information, thereby reducing the bullwhip effect. Essentially, the supply chain should be viewed as a glass pipeline that provides stakeholders with information transparency for data types such as customer demand, available capacity, and inventory levels. Efficiency should increase with better information flow, assuming incentive structures and the necessary technologies are also in place.

The Intel team developed a glass pipeline analytics capability by first evaluating the availability and integrity of product-oriented data structures shown in Figure 10. The data was bucketed into four categories: unavailable, limited and suspect, available but dirty, and mostly clean data. After merging the various product-oriented data structures, the team was able to generate useful reports such as the inventory-turns graph in Figure 11. This was the first step in creating product-based actionable insights to ensure the right inventory is in the right place at the right time. The glass pipeline also allows the retail buyer to select individual product volumes, pricing, and promotions while ensuring maximum store returns.

---

**Figure 10. Glass Pipeline Data Structures**

- **Not data available**
- **Limited data / highly suspect**
- **Moderate / larger amount of data requiring scrubbing**
- **Very strong data requiring limited to no scrubbing**
The details with respect to this specific analytics solution can be found in a companion solution blueprint. This solution blueprint provides a deep dive discussion on the business and technical aspects of social media analytics.

**Market Basket Analysis Implementation**

Several techniques were applied in order to understand consumer buying patterns and determine product mixes that best suited consumers and maximized overall return. The basis of the analytics begins with association rule learning; and in particular, the machine learning technique was applied in on-line buying and extended to address the likelihood of shopping baskets of much larger size and diversity.

**Market Basket Analysis**

To determine product associations, the Intel team used association rules to mine data and discover relationships between products in large-scale transaction data recorded by point-of-sale systems. Brick and mortar retailers, especially grocery, have much more complicated associations between products than on-line retailers due to the greater average number of items in a consumer’s market basket. There are also other types of associations and relationships that may occur in the store (based on co-location, recipe ingredients, etc.). This called for a fresh approach to providing prioritized market basket association signals to the buyer.

Market Basket Association Rules analysis employs a well-defined algorithmic method. The following provides a specific example of simple grocery association between milk, bread, and butter to explain how the method works. In this case, the following rule was tested to determine the likelihood that customers who buy either milk or bread will also buy butter:

\[
\{\text{milk, bread}\} \Rightarrow \{\text{butter}\} \quad [\text{support, confidence}]
\]

Three measures were used to place constraints on the significance of rules:

- **Support**: The proportion of transactions in the data set which contain items: milk or bread.
- **Confidence**: The probability of finding butter under the condition that these transactions also contain milk or bread.
- **Lift**: A measure of the improvement in the occurrence of the butter given the presence of milk or bread.

The conclusion is that the purchase of milk or bread is accompanied by the purchase of butter 25 percent of the time based on this data set.

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Milk</th>
<th>Bread</th>
<th>Butter</th>
<th>Beer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

With association rule techniques as the basis for creating the patterns of consumer buying and products with natural affinities, the solution took into account the diversity and size of shopping baskets and provided a multi-tiered analysis approach. The idea is to identify emergent basket rules and patterns from transaction datasets to detect product affinities. These product affinities were then combined with social media analysis, pricing strategy, promotions, and customer product recommendations. Figure 13 shows a high level process flow of the solution.

**Basket Analysis Using Transposition and Encoding**

For every product set, the solution merged the “sets” for every possible pair of products and counted the results. The merge need not be restricted to a pair, so multiple products can be considered at a time. The result was then merged with other products to find basket rules with the cost of only counting the transactions. The solution calculated the number of times the products occurred together, divided by the total number of transactions. The solution processed the following steps:

1. For each product, run through the rest of the list.
2. Merge the transaction lists together using AND (only keep a transaction if in both lists).
3. If the number of transactions in common is greater than the support value, put the new combination in the result at a level equal to the number of keys in the product list, and keep the list of transactions in common.
4. When a level is finished, and there are results, repeat the process, but use the result from that level as the input.

The details with respect to this specific analytics solution can be found in a companion solution blueprint, whereas this solution blueprint focuses on the business and technical aspects of social media analytics.

![Figure 13. Market Basket Analysis High-Level Process Flow](image)

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Product Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>100012345</td>
<td>Vegetables</td>
</tr>
<tr>
<td>100012346</td>
<td>Canned Goods</td>
</tr>
<tr>
<td>100012347</td>
<td>Vegetables</td>
</tr>
<tr>
<td>100012348</td>
<td>Bread</td>
</tr>
<tr>
<td>100012349</td>
<td>Chips &amp; Pretzels</td>
</tr>
<tr>
<td>100012350</td>
<td>Canned Goods</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Association Rules Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Support</strong> = ( \frac{(X \cup Y)\text{.count}}{n} )</td>
</tr>
<tr>
<td><strong>Confidence</strong> = ( \frac{(X \cup Y)\text{.count}}{X\text{.count}} )</td>
</tr>
<tr>
<td><strong>Lift</strong> = ( \frac{\text{Support}}{\text{Supp}(X) \cdot \text{Supp}(Y)} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (DRY GROCERY\SHELF STABLE FOODS\CANNED GOODS, FRESH\PRODUCT\VEGETABLES)</td>
<td>0.005387144</td>
</tr>
<tr>
<td>2 (DESSERTS, DRY GROCERY\SHELF STABLE FOODS\COOKIES CRACKERS AND CRISPBREADS)</td>
<td>0.003611720</td>
</tr>
<tr>
<td>3 (DRY GROCERY\SHELF STABLE FOODS\COOKIES CRACKERS AND CRISPBREADS, HERBS BULK)</td>
<td>0.007000243</td>
</tr>
<tr>
<td>4 (DRY GROCERY\SHELF STABLE FOODS\CANNED GOODS, DRY GROCERY\SNACKS\CHIPS AND PRETZELS)</td>
<td>0.003363160</td>
</tr>
<tr>
<td>5 (DRY GROCERY\SHELF STABLE FOODS\SOUP AND BROTH, DRY GROCERY\SNACKS\CHIPS AND PRETZELS)</td>
<td>0.0024992898</td>
</tr>
<tr>
<td>6 (DRY GROCERY\SNACKS\CHIPS AND PRETZELS, FRESH BAKERY\BREAD AND BAKED GOODS\BREAD)</td>
<td>0.003525485</td>
</tr>
<tr>
<td>7 (DRY GROCERY\SHELF STABLE MEAL SOLUTIONS\MEAT POULTRY SEAFOOD, FRESH BAKERY\BREAD AND BAKED GOODS\BREAD)</td>
<td>0.004159565</td>
</tr>
<tr>
<td>8 (DRY GROCERY\SHELF STABLE MEAL SOLUTIONS\MEAT POULTRY SEAFOOD, DRY GROCERY\SNACKS\CHIPS AND PRETZELS)</td>
<td>0.003276926</td>
</tr>
</tbody>
</table>
Physical Infrastructure

The physical implementation architecture of the analytics framework is shown in Figure 14, including the hardware configuration, operating systems, and important software components used in the solution.

Accelerate Big Data Implementations

Big data presents retailers with game-changing opportunities to solve age-old business problems as well as create a competitive advantage through unique insights. This paper provided an overview of an actual big data analytics implementation, thereby providing a primer for those just getting started in this field.

Working to make implementing big data easier, Intel has a team of experts who create data analytics applications and analytics toolkits for retailers so they can focus on solving business problems instead of on system development. Intel has been in the big data business for over 30 years, managing state-of-the-art manufacturing sites and complex supply chain networks. Applying this knowledge to in-store analytics, Intel data analytics consulting services yield actionable data that allows retailers and brands to respond to customers’ desires in a more responsive and predictive manner.

For more information about Intel solutions for the retail industry, visit [www.intel.com/retail](http://www.intel.com/retail).

Visit [http://www.cloudera.com](http://www.cloudera.com) for information on Cloudera Big Data solutions.

---


