insideBIGDATA

insideBIGDATA Special Report

Reinventing the Retail Industry Through Machine and Deep Learning

By Daniel D. Gutierrez, Managing Editor and Resident Data Scientist, insideBIGDATA

BROUGHT TO YOU BY DELL EMC | NVIDIA

Introduction

The use of data for competitive advantage is not new to the retail industry — having taken a pioneering role for cultivating customer analytics both online and in-store. As a result, modern retailers leverage a growing collection of retail data to understand everything from customer buying behavior, product trends and product pricing optimization, to what to stock, how much to buy, what products to suggest to repeat customers, precise ad targeting, and so much more.

Today, deep learning techniques are poised to disrupt the retail industry. As artificial neural networks become more and more efficient, and as graphics processing units (GPUs) get more and more powerful, so does their influence on retail. Imagine a world in which deep learning-based systems know precisely what a customer wants. In such a scenario, retailers would cater to their customers more quickly and more efficiently. Using deep learning, sophisticated image classification and recognition algorithms would instantly locate the product with the lowest price and the best quality, saving consumers both time and money.

Imagine further a world in which deep learning techniques predict future preferences and needs. Retailers would no longer buy excessive stock in items in order to make up for an eventual rise in the demand, because they would know ahead of time the demand for certain products. It would be a situation of mutual gain for both retailers and customers. The possibilities are endless.

A recent study by McKinsey found that U.S. retailer supply chain operations who have adopted data and analytics have seen up to a 19% increase in operating margin over the last five years. Data is clearly effective for retailers, but it's all about putting it to work in the right areas and adding in predictive capabilities.

In this white paper...

We will explore how AI, machine learning and deep learning are transforming the retail industry in many positive directions including:

- Inventory and supply chain management
- Analysis of customer buying patterns
- Analyzing traffic patterns
- Providing assortment planning
- Performing retail analytics at scale
- Fighting cyberfraud

In order to sort out important nomenclature, a recent insideHPC Special Report — *Riding the Wave of Machine Learning & Deep Learning* — offers a closer look at the differentiation between AI, machine learning and deep learning, and the impact they're having on the global business landscape.

CONTENTS

Intelligent Retail Inventory Management

One prominent area of the retail industry where machine learning is serving to embody the concept of smart automation is supply chain and inventory management. Companies are now able to leverage an abundance of data in new ways,

Inventory management, picking, packing and shipping are all time and resource-intensive processes that tend to have a dramatic impact on a company's bottom line.

e.g. preventing costly facility malfunctions, meeting customer expectations in terms of product demand and service, and advancing ROI over the long term.

Here are a few compelling use case examples:

- Using autonomous retail robots on the front-end by helping customers find products, and also on the back-end for inventory audits;
- Utilizing Internet-of-things (IoT) to provide connected monitors of the supply chain and equipment that can send alerts about potential problems before they become an issue that would disrupt the movement of merchandise;
- Employing predictive analytics technology that can be used to help determine supply chain availability and demand based on weather.

In the retail world, supply chain efficiency is essential for competitive success. Inventory management, picking, packing and shipping are all time and resource-intensive processes that tend to have a dramatic impact on a company's bottom line.

Here is a short list of ways that predictive analytics provides benefit to retailers in terms of intelligent inventory management:

- Filling customer needs more quickly through running detailed simulations that allow the effects of lateness or missed deadlines to be assessed before they become an issue;
- Reducing downtime due to faults and breakage. In supply chain logistics, predictive maintenance is starting to be used in data-driven "picking and packing" operations, as well as across transport channels;
- Cutting "shrinkage" and maximizing stock. In retail, a certain number of units will be lost due to damage, inventory mismanagement, errors of stocktaking, as well as fraud and theft. Using predictive analytics, there are a multitude of opportunities to reduce — and perhaps in some areas eliminate — this "shrinkage."
- Classification techniques can be used to make predictions such as the likelihood that an order will be late and by how many days.

AI is changing inventory management and supply chains. Specifically, AI adoption, by way of the use of optimization algorithms, is revolutionizing inventory agility by reducing stock depletions and maximizing stock levels. The use of AI in supply chains is helping companies take large strides in innovation by reducing the time to market and to evolve by establishing an agile supply chain capable of foreseeing and dealing with uncertainties.

Analyzing Buying Patterns

Another important area where the retail industry is taking advantage of machine learning is in analyzing buying trends and patterns to identify and prepare for personalized digital cross-sell and upsell opportunities — showing customers the most relevant products based on their behavior at that moment. Further, technology can help prevent customers from abandoning a retailer's website by sensing the first signs they might drop off and causing live chat assistance windows to pop-up.

Given the capabilities of AI and machine learning, it's easy to see how they can be powerful tools for retailers— reading and listening to data, understanding and learning from it, and instantly and accurately recommending the next best action *without needing to be explicitly programmed.* This is a boon for retailers seeking to accurately predict demand, anticipate customer behavior, and optimize and personalize customer experiences. For example, it can be used to automate:

- Personalized product recommendations based on data about each customer's unique interests and buying propensity;
- The selection of additional upsell and cross-sell options that drive greater customer value;
- Chat bots that can drive intelligent and meaningful engagement with customers;
- Recommendations on additional services and offerings based on past and current buying data and customer data;
- Planogram analyses, which support in-store merchandizing by telling people what's missing, comparing sales to shelf space, and accelerating shelf replenishment by automating reorders;
- Pricing engines used to make tailored, situational pricing decisions.

Retailers are able to collect large volumes of transaction-based and behavioral data from their customers. And, as data volumes grow and processing power improves, machine learning becomes increasingly applicable in a wider range of retail areas to further optimize business processes and drive more impactful personalized and contextual consumer experiences and products.

How Retail Analytics is shaping Retails Future, courtesy of SAP

Analyzing Traffic Patterns

Yet another important area of technology being cultivated by retailers is the use of deep learning for predicting retail foot traffic trends. Using sensors can detect in-store foot traffic by combining video and audio from real-time traffic along with mobile fingerprinting from shoppers' smartphones.

Consumer behavior coupled with location data offers unique insights that are built on an understanding of billions of anonymous location-enabled daily device movements, and used by brands for sophisticated consumer analytics and insights in a retail setting. These insights are foundational to a wide range of business decisions, including advertising strategy and planning, customer segmentation, merchandising, product development, and investment strategy.

A typical deep learning scenario for predicting retail foot traffic is as follows: use a series of neural networks along with retailer traffic data; make a prediction and check accuracy; if predicted vs. actual is divergent use a process called "back propagation" to go back through all the steps, make adjustments, and adjust weights to achieve a more accurate result the next time around; repeat the process many times, and each time the system gets a bit smarter about how to predict traffic.

Thanks to technologies like deep learning, retail foot traffic counting and analysis has taken significant strides. Some vendor solutions perform deep learning that enables us to better estimate traffic. In doing so, the industry is leveraging two specific components of modern deep learning that are critical to success: access to extremely large training data sets, and newly available compute resources using GPUs. We're now able to crunch data sets at a rate that was impossible just a few years ago using techniques that go far beyond the basic statistics and linear regressions of the past. These statistical techniques are nodes within a hierarchy of neural networks, and they allow us to quickly test and tweak our programs by using cutting-edge deep learning maneuvers.

Assortment Planning

Retailers frequently struggle with assortment planning and allocation optimization to ensure the right product is delivered to the right store in time for expected consumer demand. *Assortment planning* addresses the fundamental condition of what items will be available for sale. Effective assortment planning ensures that the right mix and range of retail goods are in stock for customers in every channel and category.

With deep learning, a statistical model can predict consumer demand to ensure inventory movement is directed to the right distribution center or store. Many organizations solve this problem with a batch-oriented allocation process that delays real-time analysis and takes time to resolve for impatient consumers, or they have relied strictly on traditional CPU's that have low turnaround times.

With a trained model built on item, store, and consumer demand, machine learning becomes an engine that can optimize assortments, supply chain routes, as well as feed an engine for targeted promotions and consumer-specific recommendations.

Scale in Retail: The Power of the GPU

NVIDIA has a retail customer with a high volume of data. Analytics processes have calculations requiring a blend of items and stores that resulted in 800,000,000 line items for processing. Algorithms are used to tune forecasts and route goods to stores. A past process took many hours on a server farm of CPUs, and ran only once per two-week cycle. The company's current process on 2 GPUs takes 40 minutes, which yields an opportunity to run much more frequently to make more insightful decisions.

Retail Analytics at Scale

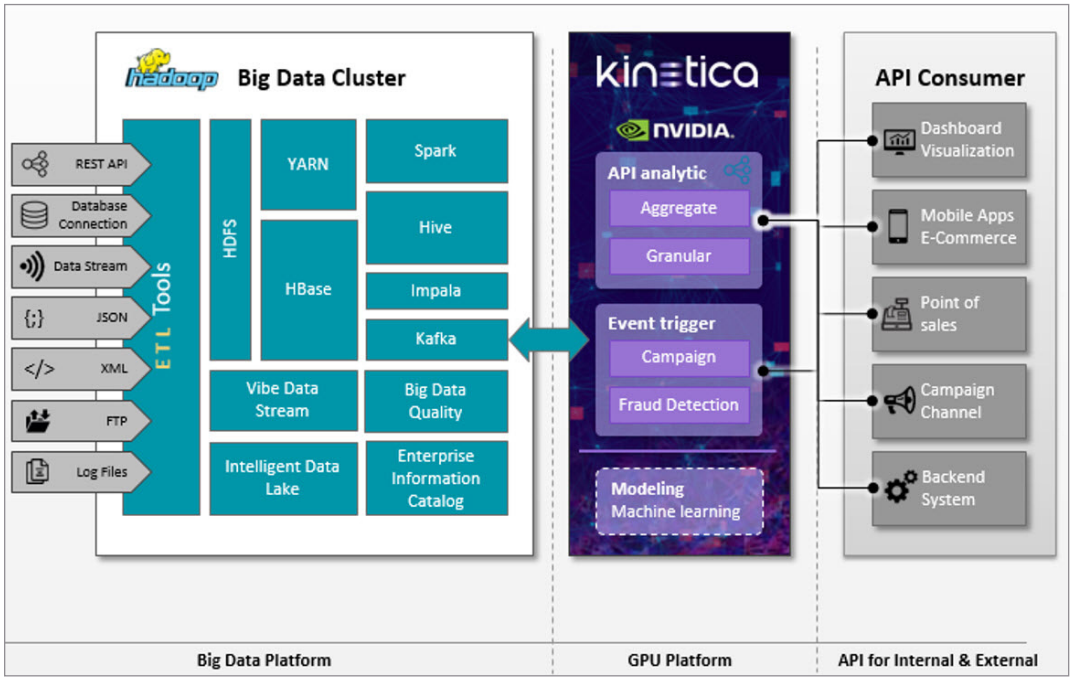
The growing challenge confronted by the retail industry is that, with ever larger data sets, the database and visualization tools that once served them well are no longer up to the task. Insights needed in milliseconds often take minutes, or even hours. At that slow pace, the tendency is to accept what's available versus what's needed. This capitulation is a prescription for failure. The modern retailer needs to process large amounts of retail data in real time, in order to capitalize on insights discovered in their retail analytics.

Many forward-leaning retailers are now working with GPU-accelerated analytics databases like those from Kinetica and MapD to seamlessly harness the power of massively parallel GPUs to deliver real-time interactive analytics at scale. In general, these technology solutions are in-memory databases with parallel processing using GPUs. A lot of the emergent machine learning algorithms combined with big data require a greater density of computing. GPUs have as many as 4000 cores as opposed to traditional CPUs having 16 to 32 per socket.

In the retail space, some players use recommendation engines that apply deep learning at scale. For example, say a customer is moving about in a store with a basket of goods and the individual items incorporate RFIDs. With use of an application the retailer might make additional purchase recommendations based on the basket composition as the customer is heading toward checkout. This is based on a deep model that is neural network based, combined with training using a framework like TensorFlow. In order to run this level of processing, there's a very large data set consisting of millions of transactions.

Accelerated analytics using GPU databases enables a 360-degree customer profile, offering a multi-dimensional set of customer attributes to describe their behavior: facial recognition, customer profile (age, marital status, income range, etc.), purchase behavior, visit behavior, merchants visited at physical locations, and lifestyle. These attributes can be achieved using tools like TensorFlow or Keras for facial recognition, clustering for basket fingerprint, and clustering for socio-economic status. Such actionable insights allow retailers to engage customers with a new and better experience.

A GPU-tuned database and visualization layer takes advantage of the most powerful compute engines in the world to deliver lightning fast analytics to some of the most powerful brands and retailers on the planet. Take the example of one of the world's largest sporting goods brands.



GPU-accelerated analytics database architecture courtesy of Kinetica

A GPU database is used to determine, in real-time, how to adjust inventory levels to better manage micro-segments within their customer population that had previously gone undetected until days or weeks after the opportunity had passed. Now, armed with analytics at scale, this brand can easily outpace their competition, extract more from their channel partners, analyze demographic visualizations, and add incremental growth to a competitive industry.

Deep Learning for Maintaining Privacy and Combating Cyberfraud

We're also seeing deep learning being used by the retail ecosystem to protect the privacy and security of consumer data while generating insights. Dell EMC is working with MasterCard and other organizations to combat cyberfraud.

With such sensitive material as customers' social security numbers and credit cards flowing throughout the cloud, innovative means of achieving protection and deep threat analysis are essential components for any well-run retail organization.

The company also is working closely with partners like Cloudera and NVIDIA to integrate Dell EMC's platform and make an organization like MasterCard able to capitalize on the convergence between tools like Hadoop and the process of machine learning to provide the benefits of data protection to organizations of all types.

Security is arguably one of the most important aspects of big data. With such sensitive material as customers' social security numbers and credit cards flowing throughout the cloud, innovative means of achieving protection and deep threat analysis are essential components for any well-run retail organization.

MasterCard utilizes Dell EMC machine learning technologies

MasterCard leverages AI to help protect consumers against credit card fraud. With approximately two million applied rules to automate spend tracking; the company handles 160 million transactions per hour and 52 billion per year. Utilizing Dell EMC machine learning technologies, the company has accelerated the speed with which transaction data can be retrieved and validated, as well as applied new rules to prevent unauthorized card usage. Of equal importance to stopping unauthorized charges is ensuring that genuine charges are not falsely flagged as fraudulent and prohibited. As their algorithms become more intelligent, MasterCard is moving closer to a model of complete and proactive oversight, where inaccuracies are prevented before they occur, and customer disruption is minimized.

Dell EMC Machine Learning and Deep Learning Ready Bundles

At the forefront of AI, Dell EMC is providing the technology needed by the retail industry today and uniquely provides an extensive portfolio of technologies — spanning workstations, servers, networking, storage, software and services — to create the high performance computing (HPC) and data analytics solutions that underpin successful machine and deep learning implementations.

The Dell EMC Ready Bundles combine pre-tested and validated servers, storage, networking and services optimized for machine and deep learning applications.

Dell EMC offers a portfolio of *Ready Solutions* designed to provide faster, better and deeper insights, delivered with efficiency and security. What's more, you can rely Dell EMC's team of experts to help you adapt as machine and deep learning evolve over time.

The Dell EMC Machine and Deep Learning Ready Bundles are part of a new portfolio that delivers on the commitment of democratizing HPC and helping customers achieve faster, better and deeper data insights. These Ready Bundles combine pre-tested and validated servers, storage, networking and services optimized for machine and deep learning applications. Customers will benefit from the introduction of the new Dell EMC PowerEdge C4140 server, supporting latest generation NVIDIA Tesla V100 GPU accelerators with PCIe and NVLink high-speed interconnect technology.

Dell EMC Deep Learning Ready Bundle with NVIDIA

The *Dell EMC Deep Learning Ready Bundle with NVIDIA* provides a GPU-optimized solution stack that can shave valuable time from deep learning projects. Dell EMC engineers can help you configure, test and tune GPU-enabled hardware and software, with included services to help data scientists load and discover insight from data more quickly.

NVIDIA NVLink™ connects multiple V100 GPUs at up to 300 GB/s to deliver 30X higher inference performance than CPU-based servers.

The Dell EMC Deep Learning Ready Bundle with NVIDIA is built around NVIDIA® Tesla® V100 GPUs. With 640 tensor cores, this powerful accelerator was the first to break the 100 teraFLOPS barrier for deep learning performance. NVIDIA NVLink™ connects multiple V100 GPUs at up to 300 GB/s to deliver 30X higher inference performance than CPU-based servers. This level of throughput and efficiency makes scaling-out machine and deep learning services much more feasible. Train AI models that used to require weeks of computing resources in just a few days. Scale AI resources more easily and bridge the gap between the data science, IT and lines of business.

Dell EMC Machine and Deep Learning Reference Configurations

Dell EMC has partnered with Bright Computing® to offer the software stack on Dell EMC hardware in a portfolio of reference configurations for multiple use cases. The Dell EMC HPC team has in-depth experience working closely with Bright Cluster Manager® to create solutions from a portfolio of servers, storage, networking, software and services. That expertise is then multiplied through collaboration with Customer Solution Centers, worldwide HPC Innovation Centers, the HPC Innovation Lab and the broader data analytics community. These experts can work with you to create a solution with the right features, at the right price.

Summary

As we've seen in this white paper, the landscape of traditional retail is undergoing a seismic shift. A rapidly evolving competitive environment, a global move towards digital shopping, and the ever-changing sentiments of highly informed buyers are forcing a new perspective in the industry. From this new perspective, we're seeing the adoption of accelerated analytics, robotics, and deep learning.

The use of AI in retail spans every aspect of the industry. Whether your goal is to optimize your supply chain, use existing data to increase sales conversion, or customize the consumer shopping experience with predictive modeling, deep learning can help you meet your challenge.

Retailers have either adapted to the new retail technology trends in retail data analytics or they have unceremoniously exited the business — beset by an inability to adapt to changing customer preferences and expectations. Those that have adapted have done so by navigating as deftly as possible, by gathering as much retail data as possible and by carefully sifting through it with new analytical techniques designed to not only spot trends but to predict them.

Whether your goal is to optimize your supply chain, use existing data to increase sales conversion, or customize the consumer shopping experience with predictive modeling, deep learning can help you meet your challenge.

Where do you stand? It definitely is time to take action to engage these technologies for competitive advantage. Here are some useful resources designed to help you take the next step with Dell EMC and NVIDIA:

- Dell EMC HPC Community, an online destination where you can interact with Dell Rockstars, independent experts and technology enthusiasts who solve technical issues, assist community members, and influence new products and services.
- NVIDA Deep Learning Institute (DLI), offering hands-on training for developers, data scientists, and researchers looking to solve the world's most challenging problems with deep learning using the latest GPU-accelerated deep learning platforms.

Further Resources

Dell EMC/AI

Dell EMC Machine Learning Knowledge Center

McKinsey Report: *The age of analytics: Competing in a data-driven world*

insideHPC Special Report: *Riding the Wave of Machine Learning & Deep Learning*

