Use Case — Retail CPG Vertical
Intel Big Data Enablement — Sales & Marketing
Predictive Analysis

Intel and Cloudera Improve Company's Predictive Analysis Abilities and Boost Sales

Improved predictive analysis boosts the company's holiday sales, while reducing inventory, by accurately targeting product to stores in each state.

Why Intel and Cloudera

Intel and Cloudera take the guesswork out of Hadoop. Using a unique collaborative approach, we delivery the best performance, security, and quality distribution, built on open standards. Working with more vendors across the ecosystem, only a solution built on CDH can ensure freedom from lock-in, enabling you to build a robust big data solution to meet the needs of your business today and into the future.

- Uniquely aligned product roadmaps for software and hardware to drive innovation faster, providing more industry firsts than any other Hadoop alternative.
- Deep partnerships with virtually every provider in the data center, streamlining the process for building Big Data solutions.
- Proven track records of identifying the driving industry standards, so you don't run the risk of stranding yourself on an island.

By accurately predicting consumer purchasing, a large international confectioner increased year-to-year comparable store sales for the Easter holiday by 15 percent. The Company also reduced inventory during this period by 15 to 40 percent and improved inventory turnover by a factor of 3.5, using predictive analysis tools from partners like SAP and Cognilytics on an operational data store built on Cloudera.

Results

- Company increases store sales 15 percent for Easter 2014 over the previous year.
- Company reduces inventory by 15 to 40 percent and increases inventory turnover by 3.5 times.
- Company's sales and marketing teams use secure, historical in-house data to create cost-effective, geographically divided E2E campaigns.
- Decision-makers and line managers receive real-time insights, which enables them to react quickly to changing consumer trends.
- Company saves over $3 million on software licensing by balancing workloads between Cloudera Enterprise and SAP* Business Warehouse (SAP BW) environments.

Business drivers

Candy manufacturing worldwide generates about $200 billion annually, with North America accounting for 18 percent, or about $36 billion.[1] Because seasonal sales of chocolates and candies constitute about a quarter of the US annual total,[2] many retailers see holidays such as Easter, Halloween, Christmas, and Valentine's Day[3] as critical opportunities to increase sales.

With social media and the digital economy gaining traction, the Company's business leaders and line managers needed to react faster to changing consumer trends. As another Easter season drew near, the Company decided to plan inventory, sales, and marketing targets for its various products at different stores.

The Company's sales and marketing teams asked:

- How do we predict the right inventory level for a product at the store level?
- What packaging should we use at different stores?
- Which markets will support upsell?

1. First Research. "Candy Manufacturing Industry Profile." April 2015. *http://www.firstresearch.com/Industry-Research/Candy-Manufacturing.html.*
2. Packaged Facts. "Chocolate in the US." September 2010. *http://www.candyindustry.com/articles/83201-u-s-chocolate-sales-up-volume-down.*
3. The Gale Group. "Encyclopedia of American Industries: Candy, Nut, and Confectionery Stores." *http://business.highbeam.com/industry-reports/retail/candy-nut-confectionery-stores.*

Accurately predicting sales for a key holiday such as this would enable the Company to allocate the right quantities of various products to the appropriate stores. This would help increase revenue while optimizing store level inventories. Better managed inventory reduces overall cost by reducing price markdowns and the transfer of goods between stores.

Solution details

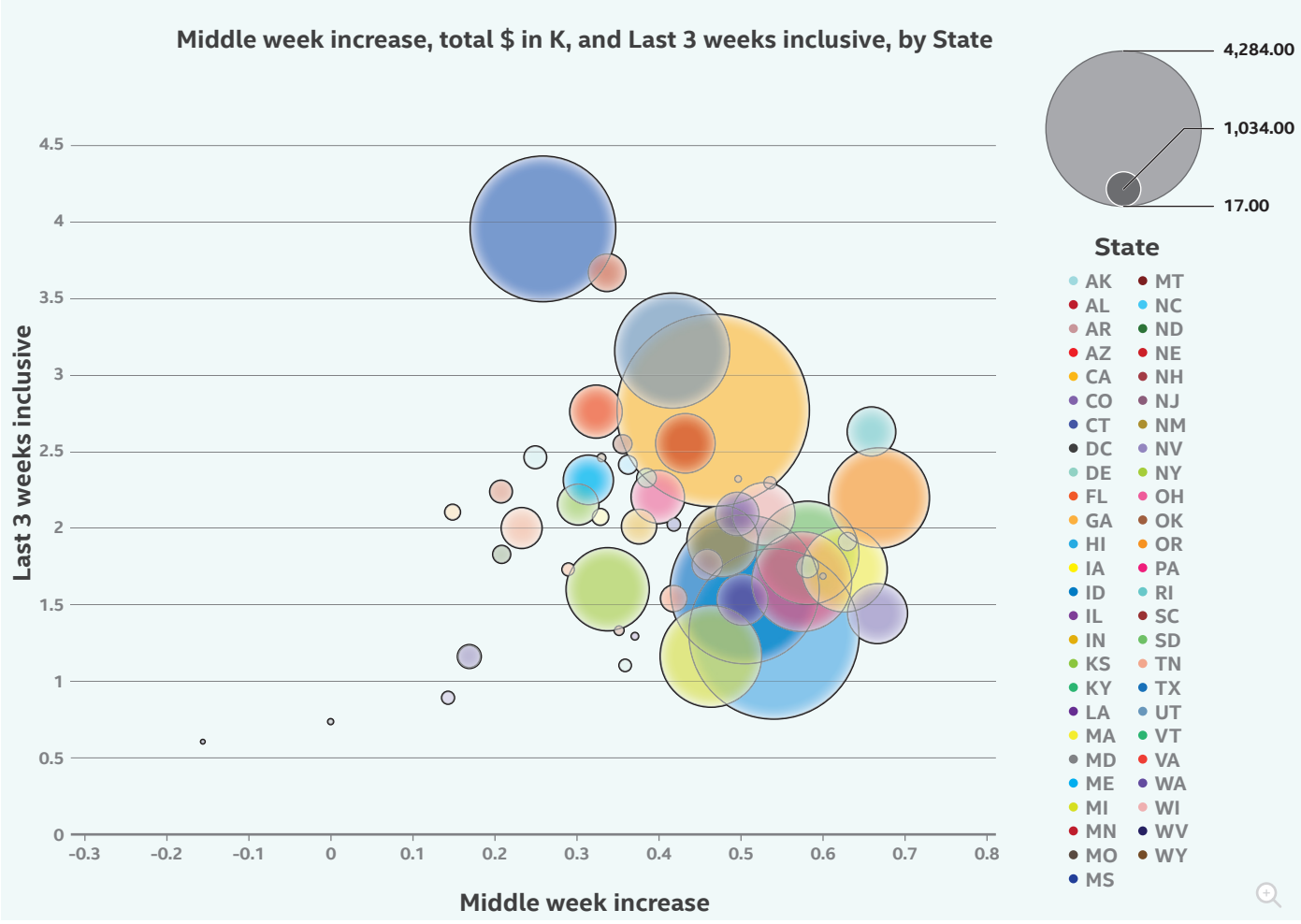
The Company deployed Cloudera Enterprise, integrating it with existing tools and platforms, including SAP* Business Warehouse (SAP BW). For agility and cost, the Company hosted the solution on cloud infrastructure provided by Virtustream*.

Because the extract, load, transform (ETL) process for legacy data on disparate systems is so compute-intensive, the Company wanted an architecture that freed their analytical workloads from ETL processing. They designed a system that performed ETL on Cloudera, where they could optimize the parallel processing required, and reserved SAP BW for analytical workloads.

For their modeling and analytics, the Company continues to run SAP Lumira* and Cognilytics predictive analyses directly on SAP BW. Data collection and ingestion now links historical data from Cloudera to weekly sales data for each store in geographical markets using the SAP HANA* Smart Data Access (SDA) feature.

Because of the close integration of these two environments, SAP and Cognilytics business intelligence tools can now provide business leaders and line managers with up-to-date reports and information, allowing them to react to changing consumer trends through various interactive dashboards. *Figure 1* shows a sample screen from one of the Company's dashboards, showing a state-by-state graphic representation of sales leading up to Easter.

Figure 1 **Predicting Easter sales.** Data analysis tools show known/unknown consumer trends by state for Easter candy purchases.



Cloudera Enterprise

The Company selected Cloudera for several reasons:

- Workload variety.
- Volume, scalability, and cost.
- Processing power.
- Enterprisewide features and manageability.
- Security and protection.

The Company needed a robust platform like Hadoop that could support the diverse applications they run. For example, an interactive dashboard that supports root cause analysis and an iterative "fail-early" process requires a platform that supports low-latency queries against historical data items linked to data in the Company's SAP HANA* environment. Furthermore, many of the Company's applications and machine learning algorithms (like MapReduce) run on historical data in batch mode. Hadoop supports batch processing, parallel processing, and other workloads. Hadoop can ingest data in any format and convert data downstream to more sophisticated formats like Parquet or event HBase tables, for optimization purposes.

The Company collects weekly sales information for PoS transactions from various stores. The Company's line managers must transform this data and use it in their predictive analyses. They needed a platform like Hadoop that could perform ETL processing quickly on both current and historical data.

Another attractive feature of the Hadoop platform is its inherent scalability. It is easy and cost-effective to add nodes to a Hadoop cluster as the workload or the number of customers increases. The Company's portfolio of products is very large, and predicting sales for each of these products requires a large volume of data. As the Company adds geographical regions and increases the historical data size, they needed an affordable solution that would meet projected growth numbers. Compared to traditional data warehouse solutions, storing and processing historical data in Hadoop is very cost-effective.

Security and data protection played a key role in the Company's decision to use Cloudera. Cloudera features like encryption, secure access, snapshots, backup, auditing and lineage provide backup and disaster recovery (BDR) and regulatory compliance for protecting customer information.

Cloudera's comprehensive security package includes complete governance—data protection, integrated authentication, authorization, encryption of data at rest and in motion, key management, SNMP support, AD/Kerberos integration, quota management, automated backup/disaster recovery (BDR), audit, lineage, and rolling updates. These features allow IT departments to track data, manage user interactions, and rest assured that their data is protected.

Summary

Historical analysis of geographical sales data at the store level—linked to weekly store sales for individual products—increases the accuracy of future sale predictions. By deploying this Big Data architecture before Q1, the Company was able to use the previous two years' sales data for Easter to predict 2014 Easter market requirements. Adding current sales from PoS (Point of Sales) terminals to the historical analysis helped them identify changing consumer trends and ultimately grow their Easter sales.

Let us help your business too.

Spotlight on Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data, an enterprise data hub built on Apache Hadoop™. Cloudera offers enterprises one place to store, access, process, secure, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data.

Cloudera's open source Big Data platform is the most widely adopted in the world, and Cloudera is the most prolific contributor to the open source Hadoop ecosystem. As the leading educator of Hadoop professionals, Cloudera has trained over 40,000 individuals worldwide. Over 1,600 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry plus top public sector organizations globally run Cloudera in production.

For more information, visit **www.cloudera.com**.

Meeting your needs

We look forward to meeting with you to define your requirements and meet your objectives.

- **Accelerate time to value:** Achieve real-time cost savings, respond to market trends, and drive innovation.
- **Secure Big Data:** Deploy a sustainable Big Data program that doesn't put your organization, or you, at risk.
- **Maintain control:** Work with a partner who educates your team so you become self-sufficient.
- **Increase business potential**: Create and execute a plan that helps you adapt now, and in the future.

Hadoop sizing guide

		Cluster size		
		Small	**Medium**	**Large**
CPU		**Intel® Xeon® Processor E5 v3**		
Storage (TB)		<72 TB	72 to 570 TB	>570 TB
Node count	**Master**	2 to 3	4 to 7	≥8
	Slaves	<12	12 to 95	≥ 96
Memory (GB)	**Master**	64 GB	128 GB	≥256 GB
	Slaves	48 GB	96 GB	≥128 GB
Network		1 Gbps	10 Gbps	10 Gbps

Hardware configuration is highly dependent on workload. A high storage density cluster may be configured with a 4 TB JBOD hard disk, while a compute intensive cluster may be configured with a higher memory configuration.

Contact us

Contact your sales rep or e-mail us.

Intel.com/bigdata/services