

1

WALMART***How Big Data Is Used To Drive Supermarket
Performance*****Background**

Walmart are the largest retailer in the world and the world's largest company by revenue, with over two million employees and 20,000 stores in 28 countries.

With operations on this scale it's no surprise that they have long seen the value in data analytics. In 2004, when Hurricane Sandy hit the US, they found that unexpected insights could come to light when data was studied as a whole, rather than as isolated individual sets. Attempting to forecast demand for emergency supplies in the face of the approaching Hurricane Sandy, CIO Linda Dillman turned up some surprising statistics. As well as flashlights and emergency equipment, expected bad weather had led to an upsurge in sales of strawberry Pop Tarts in several other locations. Extra supplies of these were dispatched to stores in Hurricane Frances's path in 2012, and sold extremely well.

Walmart have grown their Big Data and analytics department considerably since then, continuously staying on the cutting edge. In 2015, the company announced they were in the process of creating

BIG DATA IN PRACTICE

the world's largest private data cloud, to enable the processing of 2.5 petabytes of information every hour.

What Problem Is Big Data Helping To Solve?

Supermarkets sell millions of products to millions of people every day. It's a fiercely competitive industry which a large proportion of people living in the developed world count on to provide them with day-to-day essentials. Supermarkets compete not just on price but also on customer service and, vitally, convenience. Having the right products in the right place at the right time, so the right people can buy them, presents huge logistical problems. Products have to be efficiently priced to the cent, to stay competitive. And if customers find they can't get everything they need under one roof, they will look elsewhere for somewhere to shop that is a better fit for their busy schedule.

How Is Big Data Used In Practice?

In 2011, with a growing awareness of how data could be used to understand their customers' needs and provide them with the products they wanted to buy, Walmart established @WalmartLabs and their Fast Big Data Team to research and deploy new data-led initiatives across the business.

The culmination of this strategy was referred to as the Data Café – a state-of-the-art analytics hub at their Bentonville, Arkansas headquarters. At the Café, the analytics team can monitor 200 streams of internal and external data in real time, including a 40-petabyte database of all the sales transactions in the previous weeks.

Timely analysis of real-time data is seen as key to driving business performance – as Walmart Senior Statistical Analyst Naveen Peddamail tells me: “If you can't get insights until you've analysed your sales for a week or a month, then you've lost sales within that time.

WALMART

“Our goal is always to get information to our business partners as fast as we can, so they can take action and cut down the turnaround time. It is proactive and reactive analytics.”

Teams from any part of the business are invited to visit the Café with their data problems, and work with the analysts to devise a solution. There is also a system which monitors performance indicators across the company and triggers automated alerts when they hit a certain level – inviting the teams responsible for them to talk to the data team about possible solutions.

Peddemail gives an example of a grocery team struggling to understand why sales of a particular produce were unexpectedly declining. Once their data was in the hands of the Café analysts, it was established very quickly that the decline was directly attributable to a pricing error. The error was immediately rectified and sales recovered within days.

Sales across different stores in different geographical areas can also be monitored in real-time. One Halloween, Peddemail recalls, sales figures of novelty cookies were being monitored, when analysts saw that there were several locations where they weren’t selling at all. This enabled them to trigger an alert to the merchandizing teams responsible for those stores, who quickly realized that the products hadn’t even been put on the shelves. Not exactly a complex algorithm, but it wouldn’t have been possible without real-time analytics.

Another initiative is Walmart’s Social Genome Project, which monitors public social media conversations and attempts to predict what products people will buy based on their conversations. They also have the Shopycat service, which predicts how people’s shopping habits are influenced by their friends (using social media data again) and have developed their own search engine, named Polaris, to allow them to analyse search terms entered by customers on their websites.

BIG DATA IN PRACTICE

What Were The Results?

Walmart tell me that the Data Café system has led to a reduction in the time it takes from a problem being spotted in the numbers to a solution being proposed from an average of two to three weeks down to around 20 minutes.

What Data Was Used?

The Data Café uses a constantly refreshed database consisting of 200 billion rows of transactional data – and that only represents the most recent few weeks of business!

On top of that it pulls in data from 200 other sources, including meteorological data, economic data, telecoms data, social media data, gas prices and a database of events taking place in the vicinity of Walmart stores.

What Are The Technical Details?

Walmart's real-time transactional database consists of 40 petabytes of data. Huge though this volume of transactional data is, it only includes from the most recent weeks' data, as this is where the value, as far as real-time analysis goes, is to be found. Data from across the chain's stores, online divisions and corporate units are stored centrally on Hadoop (a distributed data storage and data management system).

CTO Jeremy King has described the approach as "data democracy" as the aim is to make it available to anyone in the business who can make use of it. At some point after the adoption of distributed Hadoop framework in 2011, analysts became concerned that the volume was growing at a rate that could hamper their ability to analyse it. As a result, a policy of "intelligently managing" data collection was adopted which involved setting up several systems designed to refine and categorize the data before it was stored. Other technologies in use

WALMART

include Spark and Cassandra, and languages including R and SAS are used to develop analytical applications.

Any Challenges That Had To Be Overcome?

With an analytics operation as ambitious as the one planned by Walmart, the rapid expansion required a large intake of new staff, and finding the right people with the right skills proved difficult. This problem is far from restricted to Walmart: a recent survey by researchers Gartner found that more than half of businesses feel their ability to carry out Big Data analytics is hampered by difficulty in hiring the appropriate talent.

One of the approaches Walmart took to solving this was to turn to crowdsourced data science competition website Kaggle – which I profile in Chapter 44.¹

Kaggle set users of the website a challenge involving predicting how promotional and seasonal events such as stock-clearance sales and holidays would influence sales of a number of different products. Those who came up with models that most closely matched the real-life data gathered by Walmart were invited to apply for positions on the data science team. In fact, one of those who found himself working for Walmart after taking part in the competition was Naveen Peddamail, whose thoughts I have included in this chapter.

Once a new analyst starts at Walmart, they are put through their Analytics Rotation Program. This sees them moved through each different team with responsibility for analytical work, to allow them to gain a broad overview of how analytics is used across the business.

Walmart's senior recruiter for its Information Systems Operation, Mandar Thakur, told me: "The Kaggle competition created a buzz about Walmart and our analytics organization. People always knew

BIG DATA IN PRACTICE

that Walmart generates and has a lot of data, but the best part was that this let people see how we are using it strategically.”

What Are The Key Learning Points And Takeaways?

Supermarkets are big, fast, constantly changing businesses that are complex organisms consisting of many individual subsystems. This makes them an ideal business in which to apply Big Data analytics.

Success in business is driven by competition. Walmart have always taken a lead in data-driven initiatives, such as loyalty and reward programmes, and by wholeheartedly committing themselves to the latest advances in real-time, responsive analytics they have shown they plan to remain competitive.

Bricks ‘n’ mortar retail may be seen as “low tech” – almost Stone Age, in fact – compared to their flashy, online rivals but Walmart have shown that cutting-edge Big Data is just as relevant to them as it is to Amazon or Alibaba.² Despite the seemingly more convenient options on offer, it appears that customers, whether through habit or preference, are still willing to get in their cars and travel to shops to buy things in person. This means there is still a huge market out there for the taking, and businesses that make best use of analytics in order to drive efficiency and improve their customers’ experience are set to prosper.

REFERENCES AND FURTHER READING

1. Kaggle (2015) Predict how sales of weather-sensitive products are affected by snow and rain, <https://www.kaggle.com/c/walmart-recruiting-sales-in-stormy-weather>, accessed 5 January 2016.
2. Walmart (2015) When data met retail: A #lovedata story, <http://careersblog.walmart.com/when-data-met-retail-a-lovedata-story/>, accessed 5 January 2016.