_____

HADOOPING THE MARKETING GAME

IVA SORIĆ
Poslovna Inteligencija d.o.o.
Stubička 50b, 10000 Zagreb, Croatia
iva.soric@inteligencija.com

DRAŽEN OREŠČANIN
Poslovna Inteligencija d.o.o.
Stubička 50b, 10000 Zagreb, Croatia
drazen.orescanin@inteligencija.com

MARKO ŠTAJCER
Poslovna Inteligencija d.o.o.
Stubička 50b, 10000 Zagreb, Croatia
marko.stajcer@inteligencija.com

ABSTRACT

Big data allows marketers to define and describe their customers at a more detailed and precise level than it was even considered possible just a couple of years ago. With the speed this field is evolving, big data is becoming increasingly important in the marketing domain.

While it is not a problem to collect the data, the problem remains to leverage these vast amounts of data, coming from different sources, in an efficient manner. Most marketing departments use many marketing applications, causing data fragmentation problems. This paper describes a different, data-driven approach that refers to this problem by using the Data Lake concept allowing the marketers to leverage data as their biggest and most valuable resource in a more agile and flexible manner. The goal is to collect the data, across all applications and channels, consolidate it and analyze all at once, regardless of source and type, which will enable a more complete picture of customer behavior than ever before, answer questions that were previously unanswerable, give rise to many other possibilities, and by extension – lead to more profit.

We describe in detail the strategy and its benefits, give an overview of technologies in terms of functionalities inside the Data Lake architecture, and finally, we elaborate the advantages of this concept, in comparison to traditional marketing techniques, through some of the many use cases that show how this approach can respond to the challenges of marketing today and in the future.

KEYWORDS: Data Lake, big data, data-driven marketing, customer interactions, big data analytics

_____

1 INTRODUCTION

The digital age has brought extensive changes to consumers' buying process. In this hyperconnected world we live in, customers have easier access to offers, and more options to interact with brands and each other, than ever before. Depending on the industry, their journeys with a brand or a company include physically visiting a shop, visiting their website, blogs and social media, email communication, credit card payments, etc. Massive Internet expansion and digitalization also brings many more features, like mobile apps, sensor data, GPS, cameras, call recordings, etc. All these channels of interactions generate data about customers, and that data contains information about their requirements, preferences, taste, and their lifestyle. Customers are, empowered by technological advances, becoming more demanding than ever, and the data they generate on the way to purchase, is marketers' best asset in trying to respond to their needs.

All that is driving marketing's aspiration to become more data-driven. It is all about understanding their customers better, and the way to do that is by using all the information about them they can get. Data is the greatest resource marketers have, but leveraging information hidden in it, i.e. turning vast amounts of data into valuable insights remains a challenge. Marketing now has to deal with data of various types – structured and unstructured, transactions and social media. This is where big data methods and tools come in. Traditional marketing technologies cannot provide the agility and flexibility they badly need.

Most marketing departments have multiple marketing applications like CRM, CMS, email, mobile apps, campaign management, etc. There is an overabundance of tools and platforms, and every time a new demand arises, marketers will first look for a solution in the form of a new tool. However, dramatical changes in customer behavior are driving changes in this composition. When asked about their priorities, 78% of marketers said that the 'ability to integrate' was their most important goal (Source: Forrester's Q2 2014 Global Enterprise Marketing Software Suite Customer Reference Online Survey). Therefore, traditional marketing systems are being replaced with comprehensive platforms based on a centralized repository, which will solve data fragmentation problems and address the critical need for fast insights. Basic marketing objectives are keeping their place, but more sophisticated big data methods will gain companies real competitive advantage.

2 DATA-DRIVEN MARKETING

All the changes prompted mostly by Internet expansion, caused dramatical changes in the marketing domain too, pushing customers and their wishes to the first place. Marketing is falling behind, and cannot keep pace with their customers' individual activities. Big data tools and technologies have the power to transform marketing, and the right ingredients to build a better marketing machine, one that is able to address inbound demands. Therefore, marketers

_____

should take advantage of that, and embrace the innovations. This chapter describes some of the biggest problems marketers are facing and how big data technologies can address them. [2] [4]

Problems:

- Dealing with customer interactions across channels. On their decision journeys, customers use websites, email, blogs and social platforms, mobile applications, etc. They produce large amounts of data while searching, exploring, buying and discussing products. All these touchpoints are capturing relevant data about an individual, but collecting and managing data coming from various sources is not a simple task. Having many different data sources also means various data types and formats – structured and unstructured. There is social networks data, emails, web logs, mobile data, M2M (machine-to-machine) data, etc. The ability to efficiently use external sources, along with traditional enterprise data such as product and historical customer data, is one of the requests the digital world imposes on marketing experts.

- Data fragmentation. The data fragmentation problem is often labelled as the biggest obstacle in conducting modern marketing. Marketing departments mostly use specialized systems and applications for different purposes. Each of those applications, and its data, lives in its own silo, unaware of other processes the same customer participates in. When it comes to the ability to integrate, marketers are falling short. After acquiring several technologies, they now need a way to seamlessly connect data across systems, to create a complete view of who the customer is. For example, integrating CRM, email marketing campaigns and clickstream analysis can provide extremely valuable insights. In addition, they often have redundant or conflicting data sitting in separate silos, which makes linking and integrating that data expensive, time consuming and bothersome.

- Speed and relevance. Marketers know how important timely and relevant reaction is in their field. As the speed and volume of customer interactions grows, as well as their desires and expectations, companies are struggling to respond adequately and in a timely manner. Businesses are failing to keep the pace, and realizing that traditional databases do not scale. The need to deliver faster decision making based on customer insights is becoming critically important.

How big data approach can address those problems:

- Centralized data repository inside a platform. Considering the problems of data fragmentation, and the rising number of various channels and touchpoints that participate in customers purchase journeys, a centralized data repository makes sense. Instead of using multiple applications, marketers are turning to more comprehensive solutions, hubs or platforms, which will give them the agility and performances they need. Big data technologies allow them to store all kinds of data, both structured and unstructured, and join them to get a better, exhaustive view of customer behavior across

all channels and applications. That simplifies data integration, reduces time, and provides possibilities that were before hard or impossible to accomplish.

- Improved analytics and real-time technologies. Big data technologies provide more than a distributed storage system. There are a lot of processing tools, including sophisticated analytics and visualization tools. The ability to develop more accurate predictions, powerful analytics and data mining are some of the keys to effective data-driven marketing. Moreover, big data technologies can respond to ever-growing demands for faster insights, which are proving to be so important in marketing. Managing real-time interactions at scale is something that really brings a shift to the process.

3 THE DATA LAKE CONCEPT

3.1. BIG DATA AND HADOOP

There is no unique and precise definition of big data. Everybody is talking about the volume, velocity and variety, some add more 'V-s', but what it is really about is finding a way to use that data efficiently. Although there is a lot of talk around this topic, in most cases really mastering your data remains unfulfilled, at least partially.

In the marketing data domain, the first and most obvious 'V' - volume is definitely a factor, especially in larger companies. But more importantly, it is about the variety of data types and formats. That part is especially prominent in marketing, because it has the most diverse data sources, and new ones will most probably appear in the future. There is also the speed at which data comes in and has to be dealt with, e.g. sensor data. Big data technologies make it possible to gain real-time insights, which is especially important in marketing where timely and relevant response is essential.

Since marketers always want to know more about their customers, all the data they can gather, from all the sources they have, and can acquire, is a valuable resource to them. It is where customers offer information about themselves. But it takes certain knowledge and skills to get them. The holy grail of customer experience – the '360 degree customer view' is an illusion without mastering the data game. And big data technologies are a step in the right direction. The next chapter explains some basic big data terms and gives an overview of the Data Lake concept.

A word that most commonly goes hand in hand with the term 'big data' is Hadoop. Hadoop is not a synonym for big data software, but it has established itself as a great prospect in the field. It is not one tool; it is a platform of tools, which makes it suitable for a variety of use cases. The main purpose of Hadoop is storage and fast distributed processing of large amounts of data

_____

across clusters of computers, but there are also a lot of functionalities and tools built on top of it.

Some of the advantages of Hadoop over traditional data warehouses (although it can be used with DWH, it is not necessarily a replacement) are:
- Agility. It is suitable for all kinds of data – structured and unstructured.
- It is built out of commodity hardware, so it is far more affordable.
- As the need arises, the system can be expanded through the addition of nodes in the cluster, i.e. scale-out.
- It enables advanced real-time data processing and analytics.

The reason Hadoop is convenient for storing most disparate data types, is the 'schema on read' concept. Basically, it means you do not have to pre-process data or structure it before saving. You can save anything and decide what to do with it later – when needed. That flexibility is the essential difference; it makes adding data from various sources easy, which is especially suitable for unpredictable marketing demands.

3.2. DATA LAKE

In short, Data Lake is a platform for managing data in a more flexible manner. It allows marketers to bring all their data together in one place (Hadoop), covering all marketing channels. It brings the agility and accessibility marketing needs, and the main benefit is – no more data silos. That is where this concept scores big points.

The technology stack does not have explicitly defined components; they can vary depending on the environment and preferences [3]. Here is a schema:
- Marketing applications (CRM, web analytics, marketing automation)
- DWH
- Hadoop
- Data management tools
- Third party applications as additional data sources
- Predictive analytics tools
- Visualization tools

_____

Figure 1. Technology stack
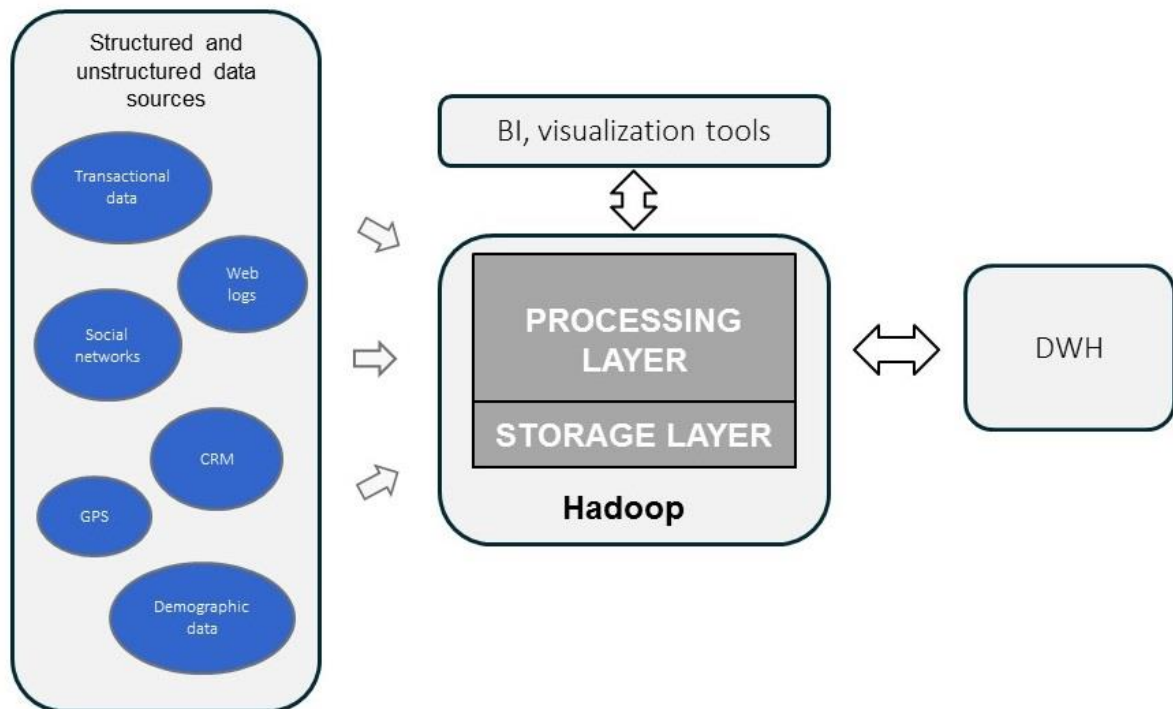
The concept assumes eliminating and/or integrating capabilities regarding marketing applications. Therefore, we suggested three core marketing applications – CRM, web analytics and marketing automation. Enterprise data warehouse, if you have it, can be used as a source for feeding the Data Lake. There is of course Hadoop – the Data Lake itself, data management tools – to automate the process of combining data from different sources and ensure data quality. Third party applications can be a great resource for enriching the data and bring new insights. Predictive analytics tools can be used to e.g. score leads, and visualization tools to communicate all the great findings with broader audience.

As customers use more and more channels and devices, the number of different data sources is continuing to rise. A customer may interact with a business by email, catalogues, visiting their website, linking to social network accounts, visiting physical e.g. retail locations or speaking to an agent at the call center. Also, each marketing application or program generates new customer data – from transactional data, social media, to GPS data. A centralized data repository turned out to be a logical and good choice. With this solution, adding new sources and new functionalities is not a problem. Marketers can continue to use traditional DWH, but also make the most of the digital world information. It is a comprehensive solution – a platform that gives marketers the flexibility they need.

_____

Key benefits of the concept are:
- Ability to seamlessly transmit content between systems
- Reduce complexity in data management
- Agility
- Speed and relevance

The Data Lake concept can address exactly the problems identified by marketers as their most common challenges. The biggest gain this approach brings is the view across channels, silos and platforms. It enables marketers to run through their data to match specific marketing activities to specific customers. They can drill down to individual transaction-level data. This complete visibility is what makes a difference. The biggest challenge of the concept is how to properly link data from all systems, clean it and validate, but there are APIs that make it work, and the final result is worth the initial trouble.

4 USE CASES

The Data Lake concept can have a positive impact on marketing strategies in different industries and it is suitable for many different use cases, particularly where data fragmentation is a common problem. For the purpose of this article, we describe two use cases for two big industries – telco and retail.

4.1. CASE 1 – TELCO

In comparison with other industries telecoms generate larger amounts of data. They are also more prone to accepting new technologies and innovations. That is why they are a logical beneficiary for the presented concept.

We are assuming a telecom operator wishes to improve customer experience, make more personalized offers to their clients, which will lead to higher revenue per customer, and better customer retention rates. The main problem in creating personalized offers is, as analyzed before, that all the data relevant for a single customer is split across systems. That complicates the process of collecting all necessary data, linking it by a common ID and analyzing it, let alone doing it fast enough to make relevant real-time offers.

For example, customer data in telecommunications can include his past purchases and used tariffs, usage patterns and billing data, customer support requests, clickstream analysis, social networks activity, location and demographic data, etc. Using the Data Lake concept we described, all data regardless of the source or type, is in one place – Hadoop storage, ready to be analyzed. Marketers can use visualization tools to create dashboards that contain detailed customer data arranged in a way that neatly comprehends his preferences. There are some great

_____

visualization tools that support connections to Hadoop in the market. The dashboard captures customer's taste and preferences, and allows marketers or sales agent to drill down by specific attributes for different perspectives and insights. They are also pretty and easy to use, suitable for nontechnical staff members.

To demonstrate one of the use cases, consider a customer who contacts the call center. The agent who takes the call can automatically get detailed information about the customer on the line at the moment of interaction. Assume the agent gets a recommendation dashboard, which will generate the best offers considering the data he has, in real-time. There is a wide range of data that needs to be utilized to create this view, for example behavioral data, calls to customer support, contract details, previous campaigns responses, sentiment analysis applied to social media data, customer interest rating derived from his website and other activities, scores expressing how valuable or how prone to converting the customer in question is. The key is to bring different sources together, so that predictive models under the hood can create recommendations based on all relevant data. The dashboard can be updated in real-time, and the agent is presented with all potential products and tariffs to offer to the customer on the line. That approach is expected to bring many benefits such as upsell opportunities, higher retention rates, improved customer satisfaction and ultimately, turning consumers in loyal customers.

4.2. CASE 2 – RETAIL

Retailers can use Data Lake to combine historical purchase data with real-time streams, in order to personalize their promotions and boost sales. Previous buying habits and patterns can be used in predictive analytics to create recommendations. Together with real-time data such as geolocation, marketers can achieve the ultimate goal – target the right customer with the right offer, at the right time and in the right place. Again the Data Lake serves as a central repository for data from different sources, and big data technologies enable real-time processing and powerful analytics capabilities.

Retailers are turning to mobile apps for new possibilities to improve customer experience, and that is a great resource for data-driven marketing. By using GPS data from their mobile subscribers, detailed profile and product data, marketers have the ability to develop customized offers, suited to the customer but also to the company's demands. For example, a retail company can notify their customers about special offers and sales that match their preferences and are happening in their vicinity at that exact moment.

5 CONCLUSION

Technological advances brought by the digital age, especially massive Internet expansion, brought changes to the marketing domain too. Customers today have more options than ever,

they use more and more devices and channels, have easier access to information, and that makes them more demanding. As their expectations grow, their interactions with companies and each other multiply, as a side effect they generate a lot of data. It is a challenge for marketers to use that data efficiently and respond to their growing demands. Since traditional marketing technologies and methods have proven to be inadequate for the task, they are being replaced with big data technologies.

In a field that is so unpredictable and ever-changing, as marketing is, the main goal is to provide flexibility and speed. The Data Lake solution is suited for marketing problems because it allows all data from various sources to be collected together and analyzed at once, thus providing the agility. It also brings advanced analytics and real-time technologies, thus providing performances. Therefore, big data technologies bring the tools and methods marketers need to develop an effective data-driven marketing strategy.

REFERENCES

[1] Forbes Insights, 2015, *Data Driven and Digitally savvy: The Rise of the New Marketing Organization.*

[2] Forrester Research, Inc., January 2014, *Marketing's Big Leap Forward*.

[3] Franz Aman, Anish Jariwala (2016), *The Marketing Data Lake*, USA, ISBN 978-0-9976012-0-6.

[4] Rusty Warner, Forrester Research, Inc., October 19, 2015, *Combine Systems Of Insight And Engagement For Contextual Marketing.*

[5] Stephan Kudyba (2014), *Big Data, Mining, and Analytics: Components of Strategic Decision Making,* CRC Press.