International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-5, Issue-5, September-2017
DOI:10.21276/ijircst.2017.5.5.3

Analysis of Customer Behavior in Online Retail Marketplace Using Hadoop

Garima Shrivastava, Shailesh Shrivastava

***Abstract*— To provide a modernized technology foundation for transforming the retail experience, retail CIOs and IT leaders need to support the "seven Cs" of customer interaction: connected, continuous, convenient, contiguous, consistent, collaborative and customized [1]. The biggest differentiator of any online marketplace is how well the store understands its customer and provides the personalized experiences. Retailers interact with customers across multiple channels, yet customer interaction and purchase data is often isolated in data siloes. This paper attempts to identify the use cases to accurately correlate eventual customer purchases with marketing campaigns and online browsing behavior. The data used in this analysis was collected from a start- up online retailer in India.**

Index Terms— **Adaptive Retail, Big data, Hadoop, RDBMS.**

I. INTRODUCTION

Today, customers purchase not just limited to stores by span across multiple channels such as web and mobile. This opens new use cases using real time information through social media customer forums and blogs and make purchase decisions through ratings, reviews, and price comparison and product recommendations. Across all these transactions, customers leave some information about their inclinations and activities that retailers can follow to adopt customer-centric strategies that help them involve customers. There are millions of people online any time and they all are a potential consumer in the online market. Since there are so many providers, the most important thing for organisations is to understand what are consumer wants and needs in this competitive business environment [3].

Manuscript received September 19, 2017.

Garima Shrivastava, Microsoft data scientist, MS Computer Science Former Specialist Data Science and Mining JSS Step, Noida, India

Shailesh Shrivastava, PMP, MS Software Engineering, Technical Program Manager, Ericsson Inc, Noida, India

.

Customer behaviors are influenced by different factors such as culture, social class, references group relation, family, salary level and salary independency, age, gender etc. and so they show different customer behaviors. These differences are seen more specific when it is considered between two different consumer groups from different countries [8].

Big Data helps to improve decision making, fast transformation, policy making, providing solution and mechanism for the development of society in many eras [1].

In this paper, we will go over how to load and query data for a web retail store in what has become an established use case for Hadoop: deriving insights from large data sources such as web logs. By combining web logs with more traditional customer data, we can better understand customers and understand how to optimize future promotions and advertising. We have used the trial version of Hadoop sandbox for this work.

II. TECHNOLOGY USED

The traditional RDBMS cannot handle big data therefore non-relational database Hadoop is being used.

Apache Hadoop is a layered structure to process and store massive amounts of data. Apache Hadoop is an open source framework for distributed storage and processing of large sets of data on commodity hardware. We have used Hortonworks Sandbox for a single node implementation of HDP. It comes packaged as a virtual machine to make evaluation and experimentation with HDP fast and easy.

This work involved below four key aspect of Hadoop that we needed to consider for analysis.

A. Data Management: Used Hadoop Distributed File System (HDFS) that provides scalable and reliable data storage that is designed to span large clusters of commodity servers.

B. Data Access: Data access helps to interact with data in wide variety of ways. In out project, we have used Map Reduce which process large amounts of structured and unstructured data in parallel across cluster of machines, in a reliable and fault-tolerant manner.

Microsoft SQL replaced by Map Reduce which is the distributed querying and data processing engine used to extract data from big datasets hosted on clusters in Hadoop implementation.

C. Data integration: Quickly and easily load data, and manage according to policy. Workflow Manager allows to easily create and schedule workflows and monitor workflow jobs. It is based on the Apache Oozie workflow engine that allows users to connect and automate the execution of big data processing tasks into a defined workflow.

D. Operations: To Provision, manage, monitor and operate Hadoop clusters at scale. We have used Apache Ambari 1.4.4 for administration and monitoring system for Apache Hadoop clusters.

III. BIG DATA REQUIRES HIGH-PERFORMANCE ANALYSIS

After we gathered Retail data (user data, product data and Omniture data) below are the key steps which we have performed.

- Load data into HDFS using Ambari User Views.
- Stored our Files in HDFS as tsv files.
- Created tables and load data into a Hive warehouse, from where it can be queried.

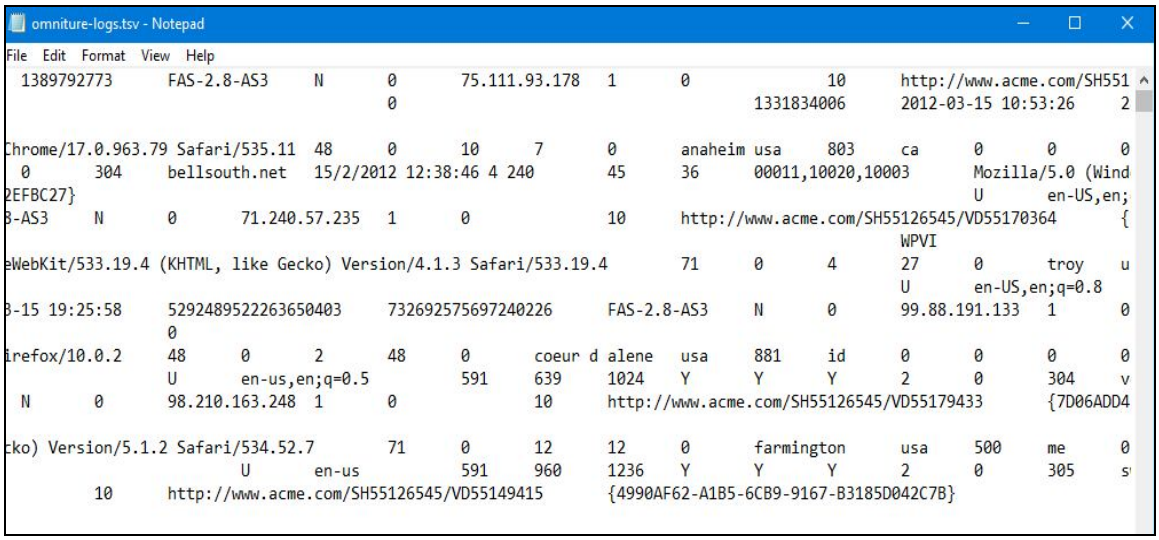

Fig. 1: Omniture.tsv file

- Analyzed this data using SQL queries in Hive User Views and store it as ORC.
- Hive can query data from Crile format, text files, ORC, JSON, parquet, sequence files and many of other formats in a tabular view. Using SQL, we could view our data as a table and created queries like we would do in an RDBMS.

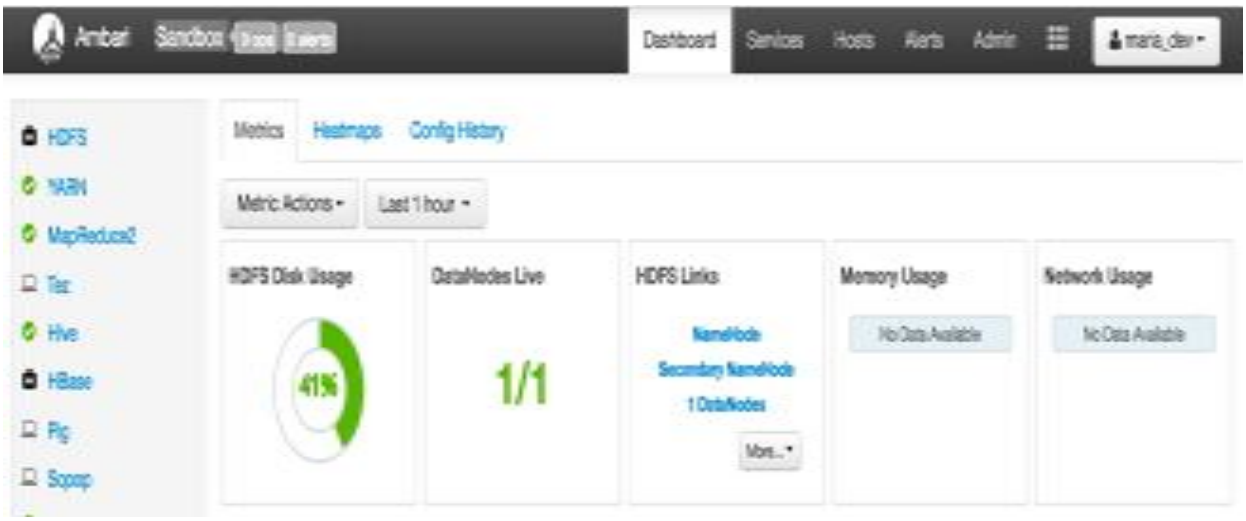

Fig. 2: Ambari user view

- Lastly, we have used Zeppelin to analyze derived data from the Retail Data tables. The business objective was to better understand the Consumer Expense behavior. To accomplish this, we have applied a series of transformations to the source data, mostly though SQL. For Data Visualization, we have used

ISSN: 2347-5552, Volume-5, Issue-5, September-2017
DOI:10.21276/ijircst.2017.5.5.3

Zeppelin to generate a series of charts to better understand the consumer behavior.

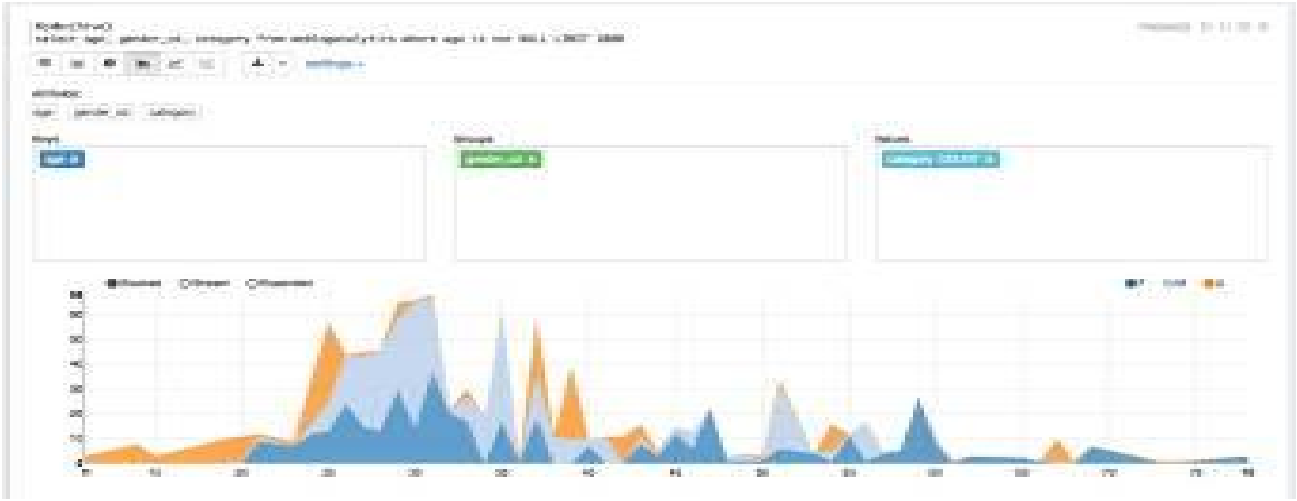

Fig. 3: Consumer analysis: Understand Customer demographics.

Example of this big data might be petabytes (1,024 terabytes) or Exabyte (1,024 petabytes) from all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured data that is often incomplete and inaccessible.

IV. CHALENGES FOR ORGANIZATIONS

One of the major challenge, which current retailers are facing is with order fulfilment and Order returns. Retail customer are becoming increasingly demanding in not just better prices and value for money but also overall order fulfilment. As online retailers focus on raising the bar, they need to focus on tackling increased volumes of returned merchandise besides fulfilling customer orders.
The solution to this problem lies in using Big data however when dealing with larger datasets, organizations face difficulties in being able to create, manipulate, and manage big data. Big data is particularly a problem in business analytics because standard tools and procedures are not designed to search and analyse massive datasets.
For most organizations, big data analysis is a challenge. Consider the utter volume of data and the many different formats of the data (both structured and unstructured data) collected across the entire organization and the many ways different.

The first challenge is in breaking down data silos to access all data an organization stores in different spaces and often in different systems. A second big data challenge is in creating platforms that can pull in unstructured data as easily as structured data. This massive bulk of data is typically so large that it's difficult to process using traditional database and software methods.

V. CUSTOMER CENTRIC MARKETING

Big Data analytics allows retailers to take on new techniques to absorb customer with their brands. They provide most promotional offers through mail, vouchers and coupons. The traditional product-centric approach to merchandising and marketing, which for years worked well to provide sales results, is rapidly being revolutionized. Below are top three questions related to marketing and insights on how big data could help.
Merchandising: The key question here is: "What merchandising application providers should a retailer consider when seeking customer centricity?"
Analytics: How can analytics be used to optimize merchandising and marketing decisions? We will look at multichannel customer analytics, algorithms, advanced analytic providers and analytic use cases for retail.
Marketing: How can retailers use customer data effectively during merchandising and marketing processes? Topics researched include customer behavior, social media and real-time offer engines. In view for the Internet to spread out as a retail channel, it is imperative to realize the consumer's mind-set, intention and conduct in light of the online buying practice: i.e., why they employ or falter to use it for purchasing? Consumer attitudes seem to have a significant influence on this decision [5].

Big Data allows retailers to make recommendations and advertisements to the customers, thereby enlightening sales and allied services [3]. Big Data promotes real-time transparency, enable predictive alerts within the supply chain, and exposes supply demand trends. This optimizes inventory and helps evading stock out incidents.

VI. BIGDATA USECASES FOR TRANSFORMING BUSINESS

There are several use cases for big data analytics in retail, because of opportunities for analysis, and will affect how the information is technically processed and organizationally accomplished.

- Build a 360-degree view of customer.
- Advanced analytics functions such as clustering and outlier detection make data-driven decisions.

ISSN: 2347-5552, Volume-5, Issue-5, September-2017
DOI:10.21276/ijircst.2017.5.5.3

- Retailers can analyse brand sentiment from Twitter, Facebook, LinkedIn or industry-specific social media streams for better understanding of customer perceptions, and align their communications, products and promotions with those perceptions to improve the product appeal to customer.
- Retailers that can geo-locate their mobile subscribers can deliver localized and personalized promotions. This requires connections with both historical and real-time streaming data.
- Online shoppers leave billions of clickstream data trails. Clickstream data can tell retailers the web pages customers visit and what they buy (or what they don't buy) on their site. But at scale, the huge volume of unstructured weblogs is difficult to ingest, store, refine and analyze for insight.
- Apache Hadoop can store that huge volume of unstructured sensor and location data. The intelligence allows retailers to reduce costs and simultaneously improve customer in-store satisfaction. This improves same-store sales and customer loyalty.
- Feedback and reviews.

The Use of big data is to optimize the company inventory based on regional and seasonal preferences of customers in each geographic area they serve. Furthermore, the use of big data is to implement an in-store, mobile navigation system that signals customers to sales based on their preferences and location in store.

Fig. 4: Interest category distribution

VII. FUTURE TRENDS IN RETAIL ANALYTICS(MOBILE ANALYTICS AND IOT)

Instead of interfacing via legacy business intelligence systems, modern mobile analytics lives at the core of decision making for major brick and mortar stores and their distribution centers. More than ever, retailers are leveraging them in-store WIFI investments to empower cashiers and even distribution associates with analytics in hand. For example, if a customer wants a product that isn't in stock, an employee with a mobile analytics app will have far more actionable insights and be able to provide the customer with a product or service much faster.

products, merchandising displays and even foot traffic pathways now have sophisticated sensors that collect and relay information for analysis. This year, an influx of beacons, Wi-Fi based sensors, and radio frequency identification (RFID) tags will be utilized to track items throughout the supply chain, and improve accuracy for in-store inventory levels. With connectivity everywhere, and data from in-store mobile devices growing in volume, so too will the potential for actionable insights.

companies are exposing live IoT data about product counts both in-store, and online—with the exact location of the product, down to the isle and bin at a specific store.

VIII. CONCLUSION

Big Data solutions permit retailers to identify and assess patterns that affect product demand, thereby boosting supply chain. Big Data solutions capitalize on campaign effectiveness by providing customer view and analyzing customer actions.

We categorized customer expense behavior based on consume interest category, Consumer age and geographical location (state) of the consumers. Below is the final analysis where we have analyzed clothing was the most popular category for customers to visit the website.

REFERENCES

[1] Sayali Gayakwad, Pranali Nale, Ravindr Bachate, "Survey on Big Data Analytics for digital world", Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016, published in IEEE International Conference.

[2] Gartner study, "Retail Industry Research Focuses on Digital Business Transformation", ID: G00310308, Published 28 Jun 2016.

[3] Garima Shrivastava, Ujwala Thakur, "Big data in retail management", International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-2, 2014, Issue-4.

[4] Gu, L.; Zeng, D.; Li, P.; Guo, S., "Cost Minimization for Big Data Processing in Geo-Distributed Data Centers", IEEE, 2014, pp. 1.

[5] Dr.Gagandeep Nagra , Dr.R Gopal, "An study of Factors Affecting on Online Shopping Behavior of Consumers", International Journal of Scientific and Research Publications, Volume 3, 2013, Issue 6, ISSN 2250-3153.

[6] Huang, Wenliang; Chen, Zhen; Dong, Wenyu; Li, Hang; Cao, Bin; Cao, Junwei, "Mobile Internet and Big Data Platform in China Unicom", IEEE, Vol. 19, 2014, pp. 95-101.

[7] Joseph, R.C.; Johnson, N.A.," Big Data and Transformational Government", IEEE, Vol. 15, 2013, pp. 43-48.

[8] Turan, A., H., "Internet Shopping Behaviour of Turkish customers: comparison of two competing models", Journal of Theoretical and Applied Electronic Commerce Research, Vol.7(1), 2012, pp.77-93.

[9] Nazir, S., Tayyab, A., Sajid, A., Rashid, H., Javed, I., "How online shopping is affecting consumers buying behaviour in Pakistan?", International Journal of Computer Science Issues, Vol.9(3), 2012, pp.486-495.

[10] Uygun, M., Ozciftci, V., Divanoglu, S., "Factors affecting online shoping behaviour of consumers", Organizasyon ve Yonetim Bilimleri Dergisi, Vol.3(2), 2011, pp.373-385.

Garima Shrivastava is a Microsoft Data Scientist, MS Computer Science,Former Specialist Data Science and Mining JSS Step, Noida, India

Shailesh Shrivastava, PMP MS Software Engineering, Technical Program Manager, Ericsson Inc

