Big Data - an opportunity and challenge for E-commerce

Sachin S. Deshmukh

Abstract — In today's world internet is one of the necessities of the human life. Almost all the systems in the world are online now. E-commerce is also one the hot buzz in today's world. As the size of E-commerce, trade is increasing so is the opportunity and problem associated with it. One of the major problems associated with the E-commerce is the data, which is produced and need to be process for effective business intelligence. This paper discusses the Big Data's opportunity and problem for the E-commerce.

Index Terms— Big Data, Business Intelligence, Business processing, Database, Data explosion, E-commerce, Unstructured data.

— — — — — — — — — ◆ — — — — — — — — —

1 INTRODUCTION

The data in the world has been exploding tremendously through internet and its forms like E-commerce. This ever-growing data is of various forms and very hard to process.

Billions of E-commerce customers using mobile device and internet are flooding a million bits of data. The increasing volume of enterprise information, customer information the catalog of goods and its prices, fuels the exponential growth of data. Analysis of this large dataset, popularly known as Big Data will be the biggest challenge and opportunity for the E-commerce and business processing.

E-commerce is a type of trade where the buying and selling of products or services is conducted over electronic systems such as the Internet and other computer networks. It uses the World Wide Web at least at one point in the transaction's life cycle, although it may include a wider range of technologies such as e-mail, mobile devices, social media, and telephones. Hence it is obvious that lot of bits of data will be generated.

2 WHAT IS BIG DATA?

While there is more and more data will be stored over time, the question is how much data does it contain? The amount of data in our world is ballooning in size everyday. The amount of information now is measured in zettabyte and petabyte. Over the next decade the number of electronically stored pieces of data or files that encapsulate the information in the digital universe and the number of servers managing the world's data stores will grow by ten times. The data that is growing too big moves very fast and does not fit in the structure of database architectures. Such large, complex and unstructured data difficult to handle process and store is called as the "Big Data".

———————————————

• *Sachin S. Deshmukh is currently pursuing masters degree program in Computer Science and Engineering from Prof Ram Meghe College of Engineering And Technology, SGBA University, India, PH-07620090812.*
E-mail: sachindeshmukh24@gmail.com

3 HOW BIG IS BIG DATA?

Trends associated with the electronic customer such as- social media, E-commerce, mobile-commerce – have created millions data residing in form of e-mails, photo, blogs, videos, social network message etc. Processing such voluminous and diverse data sets to better understand customer behavior is tremendous challenge and opportunity for E-commerce companies. In coming years this challenge will continue to grow.

Every day, around 2.5 billion gigabytes of data is being created in a variety of forms, such as social media posts, information gathered in sensors and medical devices, videos and transaction records. The volume of data in future is set to grow 800% over next five years and 80% of it will reside as unstructured data. A 30-fold increase in the internet content over the past 10 years to 35 zeta bytes with 60% annual growth in the data volume is made.

This expansion of data is the biggest challenge for the E-commerce industry. As this data is very huge to process and very hard to capture because this data is created at very high velocity. E-commerce industry is realizing that the problem of the Big Data cannot be solved by using traditional database management tools, as this data is unstructured and big in volume.

There are two types of data present on the E-commerce site – one, transactional data from ERP, CRM, SCM stored in relational databases and second, the un- structured data from disruptive technologies including Web, emails, photos etc. All this data both structured and un-structured is covered in Big Data's span which need to be captured and processed.

Facebook has 500 million users who actively generate data of various forms such as text, pictures and video. Similarly, E-commerce giant Walmart handles one million transactions in one hour. These transactions are imported into databases of size more than 2.5 peta bytes (2560 terabytes) of data. 10,000 payment card transactions are made every second around the world. Each transaction requires quick and robust processing, through various third party merchants around the globe. Amazon.com the E-commerce giant handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps

Amazon running is Linux-based and as of 2005, they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB. From these figures, the estimation about the size and structure of Big Data can be made.

4 HOW E – COMMERCE COMPANIES USE BIG DATA?

E-commerce company reaps many big benefits from Big Data collaboration. An E-commerce store takes a snapshot of traffic behaviors. This traffic can be from search engines or online ads visited most frequently. Collecting data about which links a user clicks more frequently, Facebook like for a particular products, tweets made by people. Also how much time a user spends on a specific page of E-commerce site, information from cookies stored on the customers computer and which products are most viewed by the user helps marketing, and produce better campaigns, and it give managers better insight into products and customers. In this way customer, behavior is studied even if the customer is not buying anything.

A team of data analyst process the data which digital customer leave on the internet in form of text, image, video. This data extracted contains the actual behavior of the customers on the internet and current trend. This valuable data helps the E-commerce sites to plan micro-strategies for future. In addition, companies take data such as competitor pricing, product sales, regional preferences and customer actions to determine the right price to close the sale and suitable for customer.

E-commerce companies use Big Data to quickly gather information from multiple parties on multiple products in order to accurately convey expected delivery time to their customers. Big Data helps E-commerce companies identify events in business before they occur. For example, from last year's most sold price range and the most sold color of sweaters, an E-commerce company can automatically order its inventory to get sweaters of particular color and price in advance.

5 BIG DATA AN OPPORTUNITY FOR E-COMMERCE

Big Data provide big opportunities for E-commerce as it helps the virtual E-commerce shop to interact and study their digital customer closely. If the Big Data analysis is correctly applied to E-commerce, it could make a difference with "a before" and "an after" in their online shop experience. The European E-commerce Conference held at October 2012 in Madrid, it was claimed that the Big Data, used as a tool to analyze macro-data, would be a trend to increase business that will move 34,000 million dollars in 2013. The predictions are also made in 2015 it will give 4.4 million direct jobs to satisfy the needs from analyzing and managing the Big Data.

In March 2012, The White House announced a national "Big Data Initiative" of more than $200 million for Big Data research projects. The European Commission is funding a 2-year-long Big Data Public Private Forum to guide supporting actions from the European Commission in the successful implementation of the Big Data economy.

Following are the few points, how Big Data create opportunities for E-commerce:

5.1 Personalization of products for customers

Online shopper's demand personalized customer service. Increasingly, they also require personalized product price. With Big Data, retailers can analyze customer interactions across all channels like social, mobile, and web to determine how the customer is using the products they bought or wish to buy. For example, E-commerce site can give directives to their supply chain to offer some customers a configurable product, where they can select features like color or wired versus wireless. A different segment of customers could get products that are environmentally friendly and another segment could be offered value-added features like gift-wrapping. All this can be done in real-time by matching the available products with the target customer segments by using Big Data.

With more data collected from customers through various touch points such as loyalty programs, visitor-browsing patterns, and past purchase behavior, E-commerce companies can process this information to carry out customer segmentation and thus push out personalized content and promotions. This personalized content can be promoted to customers through social networking sites, e-mails and online advertising. Personal purchase behavior, visitors interested products can be captured by the clicks that customers make while visiting an advertising, Facebook likes made by the customer for a particular product.

5.2 Dynamic pricing for customer

A relatively accurate profile can be created for every customer after collecting and processing data provided by Big Data. This customer profile will thus offer insights as to what price would re-engage the customer and persuade him to buy another product from the site. This has proven the most effective for customer retention strategies. For example, E-commerce site would decide whether a Rs. 10 off or 20% discount would work the best on any particular customer.

E-commerce company can use dynamic pricing to compete on price with other E-commerce companies, which requires incorporating data from multiple sources, including competitor pricing, regional preferences, product sales, and customer actions to determine the best price to close the sale. E-commerce giant like Amazon already has this functionality in place, which gives its business a huge competitive advantage.

5.3 Customer service

68% of online customers leave because of poor customer service so improving this mistake can help an E-commerce company to boost its sales. Outstanding customer service is a critical contributor to the success of an ecommerce site. E-commerce site like Zappos and Netflix, are excellent examples of good customer service. Big Data can help in coordinating the various communication channels including phone calls, emails, or live chat features.

Understanding of the customers expectations from the customer service desk can help to allocate better resources while

helping the customer. Any customer conflict can thus be resolved more effectively within a shorter period. For example, if a customer complains through the online chat on the website and tweets about it, it will be good to be aware of the complains when he calls customer service. This will help the customer feel listened to and valued.

5.4 Supply chain management

For an E-commerce company transparency and careful management of its supply chain process are important. Big Data allows E-commerce company to determine patterns, which can be useful to forecast any potential problems and disruptions to the process, and thus quickly act on preventative measures. For instance, any changes in warehousing or shipping updates can be captured real-time and communicated to the E-commerce company immediately. This will help the E-commerce to make back up plan for the upcoming problem.

5.5 Predictive analytics of business

Analytics is crucial for all E-commerce company, regardless of size. Without analytics, it is difficult to sustain online business. Big Data can help businesses to identify events before they occur.

Big Data allows to derive a deeper picture of the different channels in business, including sales, and inventory. Knowing how to forecast gives you the added flexibility in deciding the next steps of your business operations. The effectiveness of the forecast depends on how the data analysis is done and data available. A good example of this is predicting the revenue from a certain product in the next quarter. Knowing this, an E-commerce company can better manage its inventory costs and avoid key out-of-stock products.

5.6 Managing fraud

Fraud is a biggest problem of an online business. It affects the customer service team. Common types of fraud include denial of product delivery and purchasing of product with stolen credit cards. Both of these results in credit card chargeback. All this reverse transaction affects the E-commerce companies. Big Data can help to increase fraud detection. However, it requires the right infrastructure, to detect fraud in real-time. This will lead to a safer environment to run online transaction safely.

For detection of fraud, most E-commerce site need to process their sales transactions against defined fraud patterns. If it has not done in near real-time, it could be too late to catch the fraudsters. For new customers, they have to analyzing system log files to identify a customer's location and his IP address, browser, and operating system. Create alerts if anything looks different from the norm. All this can be done using Big Data solutions.

5.7 Real-time delivery management of products

Tracking of the shipped goods has improved. Big Data can helps to improve it even further by enabling real-time delivery management that analyzes weather, traffic, and truck location

feeds to determine the exact time of delivery to customer. To track high-priced items in the shipment, E-commerce companies can use sensors. Services from several logistics vendors, such as FedEx Sense Aware, can help an E-commerce company to control its supply chain by providing real-time information on the shipment's environmental conditions as it travels, its location, whether it has been opened or not.

5.8 Better vendor management

Most E-commerce companies have to work with multiple vendors in their supply chain. These include drop ship vendors, third party packaging vendors, transportation vendors, and logistics vendors. Big Data analytics solutions can enable real-time management by reviewing vendor performance against a set of key performance indicators. These key performance indicators include complaints, customer feedback, and on-time service. These can be tracked in real-time by integrating with vendor systems, social network feeds related to vendor deliveries, and product packaging. Companies also benefit, as they know exactly what is required from them to continue to keep your business.

5.9 Automated sourcing of product

Losing customer due to out-of-stock product is big loss for E-commerce site. Big Data solutions can help to overcome this challenge by having a real-time view of the product sales, product demand and sourcing process. With Big Data solutions, E-commerce companies can stop marking products as "out of order" as they will know in advance the time when the product needs to be sourced. This also helps to reduce the number of customers returning from the site due to out of order of the product.

6 BIG DATA AN CHALLENGE

Today an online customer makes purchase decision using the rating, reviews, price comparisons and product recommendation. For this decision, they take the help of real time information like social media, customer forum and blogs. At all these touch points and across all transactions, the online customers leave trail of information about their preferences and behavior that E-commerce site can follow to find their way through the market. E-commerce companies need to gather, aggregate, process, analysis this structured and unstructured information to adopt customer intensive strategies that help them to engage customers.

A crucial problem is that E-commerce companies don't know how to make use of such enormous data. Even an E-commerce company invests huge figure sums to derive insight from information streaming in from suppliers and customers, less than 40% of employees have sufficiently mature processes and skills to do so. Big Data needs to be complemented by "big judgment". Though the Big Data gives good decision-making scenario, but the ultimate decision making process is in the hand of human. This makes it vulnerable to risk or misjudgment.

Challenges of Big Data can be summarized as:

6.1 The Volume

Biggest challenge of Big Data is the size of Big Data. Most E-commerce sites generate more data than what their systems are able to handle. This is challenge to structure of the databases currently E-commerce sites are using. Now the Big Data calls for scalable storage and a distributed approach to querying. As the data is generating and growing exponentially it is very much difficult to bind a Big Data in a particular structure. Many E-commerce companies have huge amounts of archived data, perhaps in the form of logs, but not the capacity to process it.

6.2 The Velocity

The speed of data going in and out on average E-commerce site is around 1 petabyte/ second. The rate of change in the data and how quickly it must be used to create real value is what velocity refers to. This challenge can get worse if an E-commerce company's data analysis or data storage operates much slower than the speed of actual data generation. The velocity problem could happen if millions of customers click on the E-commerce's website at the same time or thousands of sales transactions take place every second.

6.3 The Variety

Big Data consist of variety of data, both structured and unstructured. Data can be in any form text, video, image etc. Hence, it is challenge to accommodate various form of data structure. Very few times the data represent itself in structured form and ready for processing for business intelligence and to produce the desired insights.

6.4 The Value

Another challenge of Big Data for an E-commerce site is to deriving valuable insights from data, which is consider as the most important use of the Big Data analysis. It is difficult to ask the right questions to get the most value out of Big Data than to collect all the data. For example, a company may analyze data from social networks, databases and customer service call records at the same time. However, it is much more important that on which data should an E-commerce company must rely and should consider for further analysis.

Technical and application challenges
6.5 High-speed networking

Now a day's one terabyte can be stored on disk for just $100 but transferring that on network requires an hour within a cluster system and roughly a day over a typical high-speed Internet connection. These bandwidth limitations of the internet increase the challenge of making efficient use of the computing and storage resources in a cluster used in Big Data. Link between geographically dispersed clusters and to transfer data between a cluster and an end user increases the challenge further. For example for E-commerce site like Amazon.com, which serves customer all over the globe, this network problem is certainly a big issue. This slow bandwidth makes it difficult to make transfer the information of transaction between the server and client spread across the globe.

6.6 Cluster computer programming

Big Data requires data to be distributed and processed across various nodes, grouped in form of distributed computer. Programming of distributed computer systems, which are required to process very large data sets in reasonable amounts of time, is much more difficult. The software use to analyze the Big Data has to distribute the data across the cluster and have to perform computation across the nodes in a cluster, and detect and correct hardware and software errors that occur in systems of this scale. Innovations have been made in methods to organize and program such systems, like the MapReduce programming framework introduced by Google. However, much more powerful and general techniques must be developed to fully realize the power of Big Data computing across multiple domains.

6.7 Data analysis techniques

Today data analysis is still in its early stages of development. Many data analysis algorithm are present but they work on restricted data structure. However, Big Data requires variable data structure. Hence, it is challenge to develop data analysis techniques, which can work efficiently on variety and huge amount of data. Many techniques do not scale beyond data sets of a few million elements or cannot tolerate the statistical noise and gaps found in real world data. Further research is required to develop techniques that apply in real world situations and on data sets of trillions of elements. The automated or semi-automated analysis of data is the heart of big-data computation for E-commerce. Big Data requires new tools to query and report on the data. The bright future for use Big Data of E-commerce will require new data integration tools, new querying tools, new reporting tools and new dashboards.

6.8 Maintaining security and privacy

As E-commerce organizations are capturing user data, it has become very important to make sure that this data need to be in kept safely and privacy of the user must be maintained.

6.9 Data source

Big Data cannot solve the problem of data integration. Generally, E-commerce companies try to get "good" data, "OK" data, and sometimes even "some" data about their customers. Companies cannot have much data to analyze, without an ability to extract data from systems. The dynamic nature of business today, with mergers and acquisitions, makes the problem worse.

7 CONCLUSION

Although many E-commerce companies are collecting huge amounts of data, only a few are making proper utilization of the insights that are provided by Big Data analysis. Use of Big Data in E-commerce will definitely affect the way we are shopping today and will affect the E-commerce companies

working. Big Data combined with cloud computation will affect the future of the E-commerce, than the previous generation computational revolution. E-commerce will make profit in their trade by making proper analysis and processing of the Big Data in future. In addition, data extracted from the Big Data analysis can be used by Expert Systems in future to make more business decisions, which would be far better than human decision-making capacity. This would definitely make E-commerce trade more profitable and more successful. However many technical challenges mentioned in this paper must be resolved before it can be utilized fully. Advanced research should be conducted extract the benefits of Big Data in E-commerce fully, by overcoming the existing technological and application challenges. This survey paper enlightens opportunities and the challenges residing in the use of Big Data in E-commerce.

8 REFERENCES

[1] "A Practical Guide to Big Data Opportunities, Challenges & Tools" Laura Wilber
[2] "Big Data for Development: Challenges & Opportunities" by Emmanuel Letouzé, May 2012.
[3] Margaret Rouse, "Big Data", 2011
[4] The Times of India, article- "Big Data to create a boom in job market"
[5] "IBM What is Big Data? — Bringing Big Data to the enterprise". www.ibm.com. Retrieved 2013-08-26
[6] Layton, Julia. "Amazon Technology". Money.howstuffworks.com. Retrieved 2013-03-05
[7] "Scaling Facebook to 500 Million Users and Beyond". Facebook.com. Retrieved 2013-07-21.