

BIG DATA DRIVEN E-COMMERCE ARCHITECTURE

Ahmad Ghandour

Department of MIS, College of Business Administration

Al Ain University of Science and Technology, UAE

ahmad.ghandour@aau.ac.ae

Abstract

Online businesses are increasingly investing in Big Data Analytics. Not only does it allow them to gain deeper insights into customer behavior and industry trends, but it also lets them make more accurate decisions to improve merchandising, marketing, customer retention, and every other aspect of their business. However, various problems may emerge particularly when the infrastructure of Big Data fails to perform optimally, hence leading to important intelligence and information to become unavailable or delayed. The essence of this study is not only to understand the benefits and challenges of applying Big Data Analytics but also to introduce a reference architecture as an optimum solutions to online businesses. The reference architecture represents the building blocks from which for every customer-specific situation an optimum solution can be designed by combining various technologies.

Keywords: Big Data, e-Commerce, Hadoop, Analytics, Reference Architecture

INTRODUCTION

Big Data as such is an innovative topic which changes the way companies make decisions and do business requiring new mindset with new infrastructure concepts. The term itself has been an internet buzzword since 2012. Since then, big data has promised to be more utilized in future, as organization both small and large employs big data analytics in creating a competitive advantage. Big Data is defined as data that exceeds the processing capacity of conventional database management system because of its volume, velocity, and variability. Within this data lie valuable patterns and information that previously require amount of work and cost to extract them ([Dumbill, 2012](#)). Today, *Big Data is not just about the data itself; it is about analytics when affordable technologies are used to quickly store, discover and analyze massive datasets.* Big

Data Analytics is becoming feasible at affordable cost. Many ecommerce organizations depend on Big Data Analytics in gleaning real time; and gaining valuable insights that drive profitability and smarter business decisions. However, problems may emerge particularly when the infrastructure of Big Data fails to perform optimally, hence leading to important intelligence to become unavailable or delayed ([Graham, 2012](#)). It is true that the application of Big Data analytics comes with performance challenges, as well as bottlenecks, which plague the presently distributed application, subsequently putting Big Data projects at risk. This paper is set to examine technological challenges that are associated with the application of Big Data analytics in ecommerce and suggest a building blocks as a reference architecture. The study is intended to assist online business managers to effectively understand the issues behind using Big Data analytics and how to deal with it.

This paper is structured as follows. In section 2, a review of literature on benefits and challenges is presented. Section 3 present a reference architecture of applying Big Data Followed by a discussion. Last, summary and conclusion conclude the paper.

LITERATURE REVIEW

A range of recently published research literature on Big Data is reviewed to explore their current state, issues and challenges learned from their practice. Research by the firm in conjunction with the Massachusetts Institute of Technology (MIT) in 2012 showed that companies that used big data and analytics outperformed their peers by 5% in productivity and 6% in profitability. Meanwhile, the Big Success with Big Data, 2014 report by Accenture based on a survey of senior executives from companies across 19 countries and seven industries, revealed that 92% of executives from those that were applying big data were satisfied with the results. They expected big data to have the biggest impact on their organization over the following five years in customer relationships (63%), product development (58%) and operations (56%). ([Jobidon, 2014](#)).

Big Data is not only a term that refers to data volume, but also it refers to data velocity, and variety which rarely is structured, but rather unstructured. Its volume is in the range of many Terabytes to Petabytes. The volume presents the most immediate challenge to conventional IT structures, it call for scalable storage, and a distributed approach to querying. Data velocity refers to the increasing rate at which data flows to the organization and the speed of the feedback loop, taking data from input through to decision ([Dumbill, 2012](#)).

The term Big Data is also refers to the analytical processing capabilities. A Big Data system process a high volume data coming from versatile internal and external data sources that doesn't fall into a neat relational structures. Input data to Big Data systems could be emails,

IT logs, audio and video files, chatter from social networks, web server logs, traffic flow sensors, satellite imagery, broadcast audio streams, banking transactions, the content of web pages, scans of government documents, GPS trails, telemetry from automobiles, financial market data, to name a few. While it is a daunting task to analyze data, the challenge for most eCommerce businesses is to collect, store and organize data from multiple sources. Big data paves the way for more organized data and enables businesses to better understand information from many different resources and consequently make more informed decision, which in turn enhances a company's performance. For example pricing based on historical sales data can be set effectively according to variables, such as seasonality and competitor prices. Tracking user's behavior can also alter pricing based on traffic and conversions to maximize sales and profit. Big Data provides a detailed insight in relation to the behavior of customers. Consequently, this could help an eCommerce business in tailoring their products and services to the client's preferences and needs. In addition, each customer profile relates to specific products or services alongside helping in streamlining and targeting the promotional and marketing efforts. This endeavor will help businesses cut their expenses across the board, as well as increasing their revenue through high value customers. ([Coudry & Turow, 2014](#)).

Another area where Big Data are beneficial to the company is in regard to communication. Companies with customer profiles are better positioned to understand what to communicate, what strategy to employ in these communication, and the better time of communicating.

Big Data is also beneficial to clients since they are able to receive information, and or services that are specifically customized to them. In other words, customers with profiles can receive information based on their circumstances and requirements. In most cases, customers would need a customized segmentation systems that can be able to support a massive data centric initiative that could be able to leverage and organize underutilized databases. Furthermore, the communication between customers and the company is enhanced where customers find it easy to express their concerns and needs at their convenience without having to undergo lengthy or excruciating process ([Hughes, 1994](#)).

Rud ([2001](#)) explains that, Big Data analytics provides a more effective, and faster way of collecting, managing, dispensing and controlling structured and unstructured information across ecommerce channels. The author notes that online content that is accurate creates trust and assist in improving brand image. In the event when information about a product is not adequate, difficult to find or inconsistent, there is a tendency for customers to find other alternatives. Consequently, this could negatively affect company performance and sales alongside the brand perception with long lasting impact.

By use of statistical analysis, distributed processing, operations research algorithms, and real time capabilities, ecommerce entities may be able to establish mechanisms of practical pricing that are tailored on each customer profile. Consequently, this can help sales personnel by making better sale since they now have factual data that is also real-time. Through data analytics, it becomes possible for firms to perform estimation of price elasticity. For instance, in the event when prices are hiked for a specific product or service, or when the demand declines or any other effects. Data analytics have greatly leveraged the challenges of segmentation and price elasticity in the B2B environment owing to the sets and subsets, as well as the drill downs in the product sets([Unilog, 2015](#)).

According to Unilog, ([2015](#)) data analytics have facilitated real-time demand deliveries and forecasts, predicting the present capacity and optimizing supply chain levels. In addition, it facilitates identification of the most important customers, strategies of expansion, and sources of competitive supply, lead times and the capabilities of raw materials. The author also notes that for customers, big data analytics have facilitated visibility with collaboration of real time supplier, as well as traceability of the real status of their orders or queries.

Jagadish et al ([2014](#)) explains that data obtained from the RFID sensor, the GPS, bins and pallets, social media sources, scanning devices, mobile, and ERP systems could be unstructured. Big data analytics can be used in breaking data from these sources into meaningful chunks in providing significant information, checking bottlenecks in the supply chain and making it more effective. Performing analysis on real time data from freight such as air, sea or land may require longer time in processing owing to the larger scale. Data analytics could be employed in synchronizing inventory data, supplier data, and POS in effectively providing information regarding process improvement. Generally, the author identifies big data analytics as providing a major impetus in the visibility of the supply chain, product traceability, and geo-location mapping.

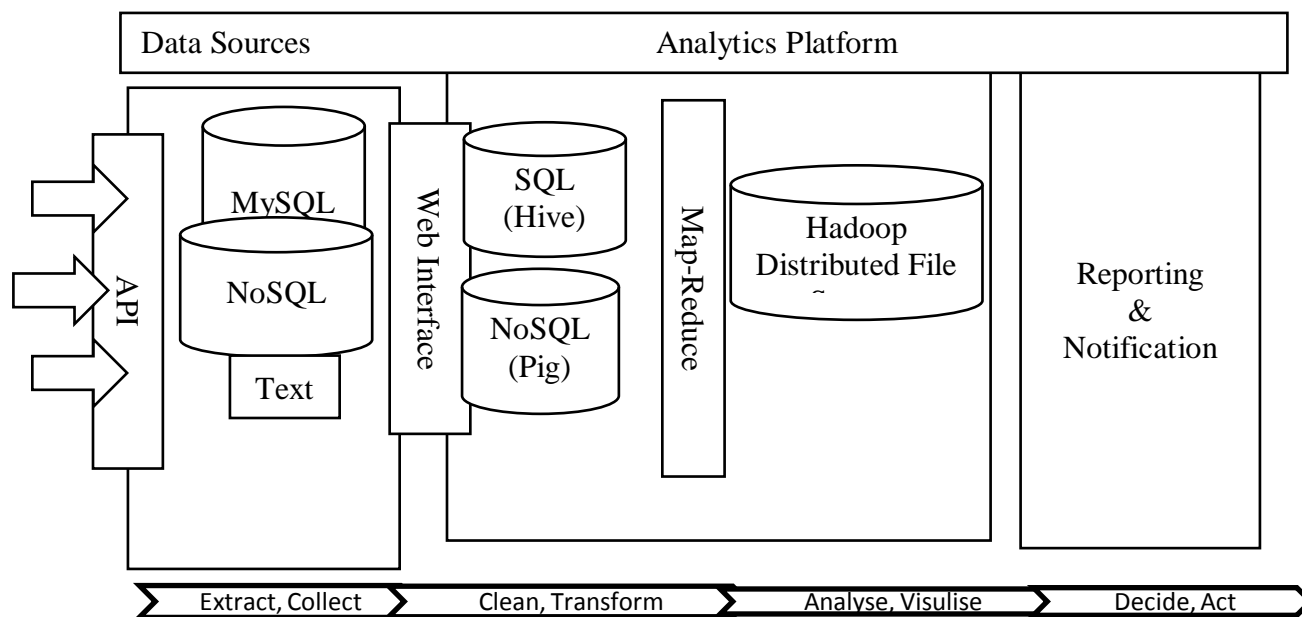
As McGuire ([2013](#)) observes, social networking has presented business entities with avenues of customer engagement and helps in providing market opportunities as well as insights for lead generation. The development of social media in marketing platforms has been increasingly used by organizations in building social signals that are very crucial in many SEO digital marketing campaigns. Apparently, the emergence of various media platforms has offered internet marketers a broader range of marketing opportunities. The availability of large quantity of data from online customer interactions could be helpful for online sales personnel in making sales that are more effective. The author identifies a good example as being a shopping cart abandonment. This provides information concerning the products intended for the purpose of purchasing and the sales personnel can conduct a follow up by using a cell phone in getting the

correct information. In addition, sales representatives in identifying the patterns of customers could use real-time updates. This will subsequently help them in selling products at the most appropriate times with the most optimized pricing options. Additionally, it offers opportunities in cross selling and up selling. Streamlined data provided options in market segmentation for peak sales forecasting sales accurately, deploying sales resources. Pipelines and sales forecasts had been based on historical trends. Nonetheless, Big Data analytics could be helpful in precise and relevant information and therefore, the high hypothetical numbers no longer drives sales personnel.

REFERENCE ARCHITECTURE

Based on the traditional method of the business intelligence process and taking into account what Big Data is, its benefits and challenges, a high-level building blocks focuses on eCommerce businesses and how to alleviate them with Big Data services is portrayed in Fig.1.

Figure 1. Reference Architecture



1. Extract and collect: Traditionally, data is structured and typically stored in a static data warehouse which will be queried periodically. For Big Data, there are versatile data sources from which Data is extracted and it is unstructured. An eCommerce business has numerous data sources to consider such as competitor offers, pricing data, sales data, stock level, advertising data and customer sentiment. Usually data is processed through API and become available from SQL, NoSQL and text files. Data from all the data sources is

imported into a distributed management system such as Hadoop. The integration of the different databases with Hadoop varies but a simple web form is possible

2. Transform and Clean: The next step is transforming and loading data from different sources into Hive table on Hadoop's file system and/or data could be stored in scalable NoSQL stores that integrate with Hadoop such as MongoDB. By means of tools such as Map-Reduce the data is then cleansed and pre-processed.
3. Analyze and visualize: The analysis and visualization of the analytics results may be applied directly to the distributed data management system. Alternatively, data transformed by distributed parallel processing may be exported as distilled essence into another data management system. This will usually be a SQL or NoSQL database. Analytics tasks will then be applied to this database.
4. Decide and act: At this stage business gets reporting and notifications depending on the challenges were addressed at the input. Such as product pricing, advertisement campaigns, product adoption or discontinuation and stock level.

The reference architecture represents the building blocks from which for every customer-specific situation an optimum solution can be designed by combining various technologies. The more important it is not to face the end user with this high complexity. This is achieved by corresponding connectors between the individual subsystems and by data adapters enabling a smooth transition.

DISCUSSION

This evaluation finds that in a wide range of areas of application, there is a collection of data in an unparallel scale. Decisions that were done previously on the basis of guesswork or curved from models of reality could now be easily undertaken using data driven models. In essence, Big Data analytics drives many aspects of the society including financial services, mobile service, physical science, and online businesses. Until recently, most ecommerce businesses have realized the opportunities in utilizing analytics and data to improve productivity, gaining a competitive edge and improving decision-making process. In essence, analytics defines the variation between winners and losers ([Graham, 2012](#)).

Despite its related advantages, Big Data analytics has presented new problems, which are yet to be solved. For instance, they cannot be used as standalone processes; rather, they have to be accompanied by human judgment. Other problems include issues related to multiple comparisons, and false results that may not be noticed easily ([Couldry & Turow, 2014](#)).

Studies including that of Lee et al (2013) established that 8 percent of all servers in a data center encounter a hardware issue on an annual basis. The most common frequent of these causes is the hardware failure. Additionally, Big Data analytics have to deal with unpredictable communication latencies, as well as network partition owing to link failures. These requirements necessitate a scalable application in treating failures as regular events, which have to be handled carefully in ensuring that the application operation may not be interrupted.

New sources of data, ranging from various business transactions, the internet, high resolution sensors and social media have created a tidal wave of high quantity data. This data has to be captured, integrated, processed, archived and analyzed. The application of big data systems has become increasingly common not only for organizations that conduct businesses online, but also in many aspects of the society. These systems have represented a long term and major investment, which necessitate considerable financial commitments alongside deployment of the massive software scale. Many aspects of data in such sectors as health, intelligence analysis, military operations that represents big data applications which grows at an exponential rate and need for scalable software solutions in sustaining future operations.

For ecommerce businesses, the Big Data challenges are intimidating. The complex and swiftly evolving context of non-standardized technologies established on a radically divergent data models. Using new technologies in designing and constructing a massive scale data generates challenges related to software architecture. It therefore, becomes important for companies to select a database technology, which best supports the mission needs of cost effective, timely and efficient delivery, as well as future growth.

SUMMARY AND CONCLUSION

Big Data is vital if online businesses are to stay abreast of consumer's demand and preferences. From this review, it is apparent that Data Analytics could assist business organizations in enhancing their performance and customer experience. It has facilitated personalization with interactive and rich product displays as well as aspects of dynamic pricing. This could assist with customer retention and acquisition alongside improving brand image and loyalty. In addition to increased sales opportunities for companies, it also provides useful insights regarding process improvement across the seller and buyer networks. Data analytics is also helpful for organizations in capitalizing on their data since costs are saved through self-service options. Finally, it highlights where improvements need to be done in terms of sales performance improvement.

With the growth of data storage estimated to be 30 to 60 percent, businesses should establish a long-term strategy for addressing the various challenges of managing projects that

require growing data sets with linear and predictable costs. In addressing the challenges of big data, there should be a new generation of scalable data management system to replace the outdated ones. These systems have to offer a strong, consistent guarantee on the basis of a vertical scaling of storage hardware and scaling of computation. The new databases should also have a high scalability and performance in employing simpler data mechanisms of data consistency, low-cost hardware in enhancing availability, and performance.

However, while the architecture introduced is not solid as technology changes rapidly, an operational architecture embedding new tools that can seamlessly grow with demands is what online businesses need to acquire to improve their performance. In essence such architecture enable the business to explore scenarios to run different analysis and make a better decisions.

WAY FORWARD

This paper proposes an architecture for online businesses to analyze their Big Data. The usefulness of this architecture should be tested in theory and practice. Further work should derive such an architecture based on empirical results and how online business are analyzing their Big Data.

REFERENCES

- Couldry, N., & Turow, J. (2014). Advertising, big data and the clearance of the public realm: marketers' new approaches to the content subsidy. *International Journal of Communication*, 8, 1710-1726.
- Dumbill, E. (2012). What is big data? An introduction to the big data landscape. Retrieved from <http://radar.oreilly.com/2012/01/what-is-big-data.html> on 17th April 2015
- Graham, M. (2012). Big data and the end of theory? *The Guardian*.
- Hughes, A. M. (1994). *Strategic Database Mining*. Chicago: Probus Publishing.
- Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.
- Jobidon, P. (2014). Using big data to enhance e-commerce. Retrieved from <http://www.bpostinternational.com/en/content/using-big-data-enhance-e-commerce> on the 17th April 2015
- McGuire, T. (2013). Making data analytics work: Three key challenges. Retrieved from http://www.mckinsey.com/insights/business_technology/making_data_analytics_work on the 17th April 2015
- Rud, O. P. (2001). *Data mining cookbook: modeling data for marketing, risk, and customer relationship management*. New York: John Wiley & Sons.
- Unilog. (2015). Big Data Analytics in B2B ecommerce- Making Big Decisions. <http://www.unilogcorp.com/pdf/Big-Data-Analytics-in-B2B-ecommerce-v2.pdf> on the 17th April 2015