International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Analyzing E-Commerce Trends Through Big Data

Karamjeet Kaur[1], Yatika Hasija[2]

[1]Student, CT University
[2]Professor, CT University

Abstract
Big data is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale.

Keywords: Analytics, Algorithms, Data Storage, Cloud Computing, IOT, Big Data Processing

Introduction:
Imagine a world without data storage; a place where every detail about a person or organization, every transaction performed, or every aspect which can be documented is lost directly after use. Organizations would thus lose the ability to extract valuable information and knowledge, perform detailed analyses, as well as provide new opportunities and advantages. Anything ranging from customer names and addresses, to products available, to purchases made, to employees hired, etc. has become essential for day-to-day continuity. Data is the building block upon which any organization thrives. Now think of the extent of details and the surge of data and information provided nowadays through the advancements in technologies and the internet. With the increase in storage capabilities and methods of data collection, huge amounts of data have become easily available. Every second, more and more data is being created and needs to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations need to get as much value as possible from the huge amounts of stored data. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted. The contribution of this paper is to provide an analysis of the available literature on big data analytics. Accordingly, some of the various big data tools, methods, and technologies which can be applied are discussed, and their applications and opportunities provided in several decision domains are portrayed Tools For Big data Processing: Large numbers of tools are available to process big data. In this section, we discuss some current techniques for analyzing big data with emphasis on three important emerging tools namely Map Reduce, Apache Spark, and Storm. Most of the available tools concentrate on batch processing, stream processing and interactive analysis. Most batch processing tools are based on the Apache Hadoop infrastructure such as Mahout and Dryad.

1. **Apache Hadoop and Map Reduce:** The most established software platform for big data analysis is Apache Hadoop and Map reduces. It consists of hadoop kernel, map reduces, hadoop distributed file system (HDFS) and apache hive etc.
2. **Apache Mahout:** Apache mahout aims to provide scalable and commercial machine learning techniques for large scale and intelligent data analysis applications. Core algorithms of mahout including clustering, classification, pattern mining, regression, dimensionalty reduction, evolutionary algorithms, and batch based collaborative filtering run on top of Hadoop platform through map reduce framework. The goal of mahout is to build a vibrant, responsive, diverse community to facilitate discussions on the project and potential use cases.
3. **Apache Spark:** Apache spark is an open source big data processing framework built for speed processing, and sophisticated analytics t is easy to use and was originally developed in 2009 in UC Berkeleys AMP Lab. It was open sourced in 2010 as an Apache project. Spark lets you quickly write applications in java, scala, or python. In addition to map reduce operations, it supports SQL queries, streaming data, machine learning, and graph data processing. Spark runs on top of existing hadoop distributed file system (HDFS) infrastructure to provide enhanced and additional functionality. Spark consists of components namely driver program, cluster manager and worker nodes.
4. **Dryad:** It is another popular programming model for implementing parallel and distributed programs for handling large context bases on dataflow graph. It consists of a cluster of computing nodes, and an user use the resources of a computer cluster to run their program in a distributed way. Indeed, a dryad user uses thousands of machines, each of them with multiple processors or cores.
5. **Storm:** Storm is a distributed and fault tolerant real time computation system for processing large streaming data. It is specially designed for real time processing in contrasts with hadoop which is for batch processing.
6. **Apache Drill:** Apache drill is another distributed system for interactive analysis of big data. It has more flexibility to support many types of query languages, data formats, and data sources. It is also specially designed to exploit nested data
7. **Jasper soft:** The Jasper soft package is an open source software that produce reports from database columns. It is a scalable big data analytical platform and has a capability of fast data visualization on popular storage platforms, including Mango DB, Cassandra, Redis etc. One important property of Jasper soft is that it can quickly explore big data without extraction, transformation, and loading (ETL).
8. **Splunk:** In recent years a lot of data are generated through machine from business industries. Splunk is a real-time and intelligent platform developed for exploiting machine generated big data. It combines the up-to-the-moment cloud technologies and big data. In turn it helps user to search, monitor, and analyze their machine generated data through web interface.

IOT for big data analytics: Data though collected by the devices need to be filtered to make it relevant and useful. The redundancy in the data being collected is predominant due to the sheer nature of the framework of IoT. The data is continuous hence the extraction of valuable information is not simple. This requires a good mechanism of protocols and software to ensure that the data is secured and also significant.

Cloud Computing For Big Data: Cloud Computing is the delivery of computing services such as servers, storage, databases, networking, software, analytics etc., over the Internet (“the cloud”) with the aim of

providing flexible resources, faster innovation and economies of scale. Cloud computing has revolutionized the way computing infrastructure is abstracted and used.

Issue and Challenges Big data and anaylsis: Challenges in big data can be broadly alienated into three types the first type is data challenges, the second type is data process challenges, and the third type are data management. Data challenges are the challenges that are associated with the characteristics of big data. Process challenges are the challenged that faced during the processing of data whereas management challenges pertaining to tackling the data such as providing security. The characteristics of big data bring many challenges to it such as its high volume, variety, etc. Process challenges are related to data acquisition, pre-processing, data analysis, and data visualization whereas management challenges are related to privacy and security.
Some of the prominent challenges are discussed as follow. 3.1.1. Volume Challenges. The unprecedented increase in data through internal and external sources has resulted in a massive amount of data. This high volume of data brings the challenges to the data itself such as the storage of the data for processing is not possible through traditional tools and thus more innovative methods should be developed to handle this data deluge. 3.1.2. Variety Challenges. The challenge associated with variety is related to its different forms. The massive data can be present in the form of structured, semi-structured, and unstructured. Research studies show that 95% of the data is present in unstructured form. Therefore, converting it into a form so that the analysis can be performed is a big challenge. 3.1.3. Velocity Challenges. Velocity indicates the speed of the data generated through the devices. Data can be processed in two ways batch processing and real-time processing. In batch processing, the data is stored and then processed whereas real-time processing is continuous. In online shopping, real-time processing is required to generate value for customers.

Suggestions For Future work: While many large companies are already edging closer to, if not already fully embracing, all of these trends, giving them an edge over their competitors, the future of big data analytics is no longer locked behind a wall of price barriers. Data engineers and scientists are developing innovative ways to uncover insights hidden beneath the heap of data without requiring the budget of a Fortune 500.We're going to see a lot more small and mid-size companies incorporating big data analytics into their business strategies embrace it.

Method For big data: It is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information. In a research it supports the researcher to reach to a conclusion the user purchasing the goods from e-commerce websites.

Conclusion:
The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and cost-effectively for the first time in history. These capabilities are neither theoretical nor trivial. They represent a genuine leap forward and a clear opportunity to realize enormous gains in terms of efficiency, productivity, revenue, and profitability. The Age of Big Data is here, and these are truly revolutionary times if both business and technology professionals continue to work together and deliver on the promise.

Reference:

1. Demirkan, H., Delen, D.: Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud, Decision Support Systems 55, 412–421 (2013)
2. Barton, D., Court, D.: Making Advanced Analytics Work For You, Harvard business review, 90 (10), pp. 79–83 (2012).
3. McAfee, A., Brynjolfsson, E.: Big data: the management revolution, Harvard business review, 90 (10), pp. 60–68 (2012)
4. X. Wu, G. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 28 (2014) 97–106.
5. Paakkonen, P., Pakkaka, D.: "Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems", Big DataResearch2, 166–186 (2015)
6. Thusoo, Z. Shao, S. Anthony, D. Borthakur, N. Jain, J. S. Sarma, R. Murthy, H. Liu, Data warehousing and analytics infrastructure at Facebook, in: 2010 ACM SIGMOD International Conference on Management of Data, Indi-anapolis, Indiana, USA, 6–11 June (2010) 8 Go Muan Sang, Lai Xu, Paul de Vrieze
7. Kreps, J., Narkhede, N., Rao, J.: Kafka: a distributed messaging system for log pro-cessing, in: The 6th International Workshop on Networking Meets Databases, Athens, Greece, 12 June (2011)
8. Wu, L., Sumbaly, R., Riccomini, C., Koo, G., Kim, H.J., Kreps, J., Shah, S.:
1. Avatara: OLAP for web-scale analytics products, in: 38th Inter-national Conference on Very Large Databases, Istanbul, Turkey, 27–31 August (2012)
9. Mishne, G.: Fast data in the era of big data: Twitter's real-time related query suggestion architecture, in: The 2013 ACM SIGMOD International Conference on Management of Data, New York, New York, USA, 22–27 June (2013)
10. Lin, J., Ryaboy, D.: Scaling big data mining infrastructure: the Twitter experience, ACM SIGKDD Explor. Newsl. 14, 6–19 (2013)
11. Lee, G.L., Lin, J., Liu, C., Lorek, A., Ryaboy, D.: The unified logging infrastructure for data analytics at Twitter, in: The 38th International Conference on Very Large Databases, Istanbul, Turkey, 27–31 August (2012)
12. Busch, M., Gade, K., Larson, B., Lok, P., Luckenbill, S., Lin, J.: EarlyBird: real-time search at Twitter, in: 2012 IEEE 28th In-ternational Conference on Data Engineering, Washington, DC, USA, 1–5 April (2012)
13. Amatriain, X.: Big & Personal: data and models behind Netflix recommendations, in: The 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Chicago, Illinois, USA, 11 August (2013)
14. Amatriain, X., Basilico, J.: System architectures for personalized recommendations, Available via Netflix, http://techblog.netflix.com/2013/03/system-architectures-for.html, accessed 05 August, 2016.
15. Boulon, J., Konwinski, A., Qi, R.: Chukwa: a large-scale monitoring system, in: Cloud Computing and its Applications, Chicago, Illinois, USA, 22–23 October (2008)