Big Data and Cloud Computing in Consumer Behaviour Analysis

M.S.R.P. Dias
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka

K.A. Dilini T. Kulawansa
Department of Computational Mathematics
Faculty of Information Technology
University of Moratuwa
Moratuwa, Sri Lanka

Abstract— **Big Data and Cloud Computing are two major concepts which have been constantly growing since their emergence. Big data is generated in enormous scales daily in online activities, which are used in analytics related to many fields of study such as banking, healthcare, chemistry, data mining, finance, marketing etc. Consumer behavior analysis can be done through analyzing the big data which is generated through transactions. Effective application and management of resources is important in IT implementations related to this subject, for which cloud computing can effectively contribute. This paper presents technology used in big data and cloud computing, how they are used in consumer behavior analysis, existing barriers and possible future directions in which these technologies will be used in consumer behavior analysis.**

Keywords—Cloud Computing, Big Data, Consumer behaviour analysis, marketing

I. Introduction

The datasets which is hard to handle using traditional data handling methods, due to the volume, velocity and variety are called "Big Data" [3].Unlike in the past decades when data collecting was difficult and data was scarce, data is generated in massive amounts due to various activities which are high in volume, velocity and variety in different fields of study. When these data sets are stored in computerized systems, it is essential that they should be properly fetched, managed and processed using non-conventional software tools [15], which ultimately leads to generation of high-value results. Big data analytics presents various technologies and advanced techniques related to analytics, on massive sets of data[3].

Cloud Computing refers to a vast collection of integrated networked hardware, software and platform provisions, which has the ability to minimize grid computing and rigid hardware structures [2]. Cloud computing can be deployed in several models: mainly public cloud ,private cloud, hybrid cloud and community cloud[9], which can be selected according to user's requirements. Big Data analytics require two major factors to succeed: scalable storage that can accommodate growing data and high processing power to complete analytical tasks in the allowed time [18]. Integrating cloud computing in to big data solutions contributes in satisfying above needs. Consumer behaviour can be defined as the study of individuals, groups or organizations in a bid to understand the process of their selection, securing, using and disposing the products, services, experiences or ideas [1]. Analyzing consumer behavior is an important aspect, since it contributes in better customer relationship management, understanding customer psychology and attracting customers [1]. Using big data which are generated constantly through consumer activities to analyze customer behavior is highly effective in identifying consumer behavior patterns. The information which are retrieved can be used to create market strategies which increase the probabilities of buying behaviour [16].

Sophisticated computational resource requirements for big data analytics can be achieved via cloud computing technologies.

The paper is organized as follows. In the second section of the paper, an overview of big data, cloud computing and consumer behavior analysis is provided. Major researches and findings in big data and cloud computing in consumer behavior analysis is described in section three. Section four and five presents the current problems and issues of big data and cloud computing in consumer behavior analysis. Section six presents the discussion. Finally, the future directions and possibilities are presented in section seven.

II. Overview of Big Data and Cloud Computing IN Consumer Behaviour Analysis

Big data has become an important integral part of many fields. Big data analytics are applied and used in healthcare, banking, marketing etc. Big data analytics are used to visualize and analyze huge datasets which are produced in such fields. Consumer behavior depends on various factors. This includes various demographic factors as well as consumer psychology. Hence, pre-determining and identifying patterns in consumer behaviour is extremely important and profitable for business organizations, since the business organizations can make key decisions in production and marketing, towards success. At present, Net-shopping and online buying and selling has become essential features for the society. The participants of these online activities always leave "digital footprints", which can be used to trace back to him/herself. Huge pools of datasets are daily created due to online transactions, and BDA can be used on these datasets, to recognize consumer behavioural patterns including the preferred item or service

categories of customers, most generally consumed items or services within a time period etc . Big Data Analytics require costly computational requirements, since parallel computing must be performed on huge datasets which could be unlikely to be afforded by medium and small scale businesses. To overcome such barriers, cloud services can be used. Platform as a Service(PaaS) and Software as a Service (SaaS) can be used most of the time according to the organization's consent, which allows the organizational IT infrastructures to wield a massive computational power which they don't 'physically' posses. When implementing such cloud solutions, security must be properly maintained.

III. Major researches in Big Data and Cloud Computing in Consumer Behaviour Analysis

Over the past years the researchers have involved in studies and researches regarding applications of big data and cloud computing and how they can be related to consumer behavior analysis.

A research done by Qiuchen Li, Jin kun Xing, On Liu and Woonkian Chang [15] on impact of big data analytics (BDA) on customers' online behavior, shows strong correlation between the browsing and the purchasing of sales items which were presented to the customer through predicting via BDA.

Saavi Stubseid and Ognjen Arandjelovic have compared two data classification methodologies [14] which use readily measurable features, for implementing machine learning for determining customers' purchasing decision. The methodologies have been tested using a dataset of 642,709 entries gathered through customer transactions.

Daniel Pop states about machine learning via cloud computing. The softwares such as BigML, BitYota provides machine learning as SaaS [22] as well as AWS.

Bernie M. Purcell's article titled "Big Data using cloud computing" [7] presents (with references) that the businesses can use cloud computing services to substantially lower the cost of parallel computing requirements for Map Reduce process, rather than using clustered network assisted storage (NAS) devices.

T. Thiraviyam and V. O. Chidambaram stresses that Artificial Intelligence (AI) applications can be used for propensity modelling, which helps to identify correlate customer characteristics with anticipated customer behaviours [24]. Cloud related platforms such as IBM Watson, Microsoft Azure, Amazon Cloud, Google Cloud AI can be used to achieve these task.

According to Mr. Bhupesh Kumar and Mr. Praveen Shende [11] , special data-collecting tools such as Cisco's NetFlow Monitor, provides nearly real-time network traffic monitoring and data flow within multiple criteria selection, by using IP addresses and protocols and enabling effective evaluation of user security and user performance trust. This enhances security of the cloud systems, opening more secure cloud service options to businesses.

IV. Current issues of Big Data and Cloud Computing in   Consumer Behaviour Analysis

ALTHOUGH THERE ARE IMPROVEMENTS IN USING BIG DATA AND CLOUD COMPUTING IN CONSUME BEHAVIOUR ANALYSIS, THERE EXIST CERTAIN LIMITATIONS AND CHALLENGES WHICH ARE OPEN TO SOLUTIONS.

WHEN CONSIDERING THE USE OF BIG DATA ANALYTICS, THE PERFORMING ORGANIZATIONS SHOULD ALWAYS KEEP THE PROPER COMPUTATIONAL POWER NEEDED BY THE PROCESS. ALSO DIFFICULTIES ARISE IN THE PROCESSING OF DATA WITH HIGH VARIETY [13]. THE ORGANIZATION SHOULD BE RESPONSIBLE FOR THE COLLECTED DATA AND DATA MINING TECHNIQUES.

USING CLOUD COMPUTING HAS ITS OWN CHALLENGES. ORGANIZATIONS ARE SLOW IN ACCEPTING CLOUD COMPUTING TO SUPPORT THEIR BUSINESSES DUE TO RISKS OF DATA LOSS, DATA CLEANING, ACCOUNT HIJACKING AND SECURITY ISSUES. [9] ALSO CONCERNS ABOUT DATA HONESTY AND PROTECTION EXISTS IN CLOUD NETWORKS[10].

When an organization adopts cloud services, confidential datasets including cloud users' and consumers' private data can be possibly used and included for computations [23]. Organization will have to uncover these data to clouds as well as facing the risks of User-level, CSP level and Network level security threats. [8] CSP level attacks include guest-hopping attacks which will be used to infiltrate information using a virtual machine, from a virtual machine hosted by the same hardware, SQL injection attacks which crashes databases, malicious insiders who are employees of data centres who

accesses the sensitive data regarding an organization's customers for tgeir personal gains etc. Cloud users are rendered with an initial vulnerability to these attacks mostly, since they occur in CSP level. [8]

In application of cloud and big data combined solutions, extremely qualified staff with java programming experience is needed to combine Hadoop with the existing ERP systems [23]. At present, there is a short-coming of cloud framework applications which process big data interactively [19].

V. Applications of Big Data and Cloud Computing In Consumer Behaviour Analysis

Big Data and Cloud Computing concepts have been implemented in different applications to make them more user friendly and support consumer behaviour analysis.

To process big data effectively, there should be methods which consume minimal computational power and data bandwidth to achieve maximum amount of processing. Apache Hadoop framework has been developed to perform big data processing over huge sets of data using simple programming models. It has HDFS (Hadoop Distributed File System) to achieve effective data storage and Map Reduce for processing purposes. Map reduce is able to implement parallel processing and achieve data locality, which reduces a massive amount of data bandwidth while processing. Hence the framework can be easily used to develop applications to analyze consumers' data through online transactions.

The recommendation systems are a key component of online marketing, which suggest items to potential consumers via considering the results of BDA performed on online transaction data. The Amazon Web Services (AWS) recommendation systems is a prime example. It is based on events, ratings and filtering. The activities on the AWS website are captured and stored,to the extent of every valid click of a customer. It takes into account, the ratings given by the users, user's browsing history and filters the potential recommendations and presents the best ones to the customer. The system is capable of considering the customer feedbacks also. The recommendation systems of E-bay, Alibaba share some of these features too.

The supply chain management system in Walmart is another application of effective BDA. It is able to gather and consumer data from all the shops and able determine the under-performed and over-performed products, and also able to identify consumer behaviour patterns.

For the big data analytics to be adopted by the small scale and medium scale businesses, considerably large hardware and computational requirements should be fulfilled. Cloud computing can be adopted for this purpose. AWS cloud computing services is a great application of cloud computing on big data processing. AWS services include Elastic Computing Cloud (EC2), Virtual Private Cloud(VPC), S3, Relational Database Systems (RDS), Elastic Load Balance(ELB) and Auto -Scaling . EC2 allows users to acquire raw servers for hosting purposes, VPC allows users to acquire parts of Amazon Cloud which can be used to create networks, S3 supports storage, RDS supports hosting and managing various kinds of databases, ELB and Auto Scaling ensures the continuous functioning of the users' servers or applications. Data mining in cloud computing contributes to the centralization of organizational management software, while providing effective and reliable services to the users [17]. All the above services combined, organizations are easily able to gather and mine the consumers' data and host their own customer behaviour analysis within cloud servers. Big Data can be stored directly in cloud storages and processing can be done using the cloud servers' computational power.

VI. Conclusion

Vast development of technology has extended BDA in to various fields of study and made cloud solutions which are more effective. Businesses can adopt these solutions to reduce their cost of computational power, compensate their lack of officers with high IT knowledge, and build flexible storage infrastructures. To analyze consumer behavior, business organizations can use cloud platforms and services to process and analyze big data. But as of now, these benefits come in a background with both positive and negative aspects. Following are summarized details of some related researches, used in preparation of this paper.

TABLE 1: SUMMARY OF SOME RELEVANT RESEARCHES

Title of the Research Paper	Aim and Problems Addressed	Limitations/ Challenges	Future Directions
"A Systematic review of Consumer Behaviour Prediction"	Providing a systematic review of Customer behaviour prediction studies with a focus on components of customer relationship management, methods and data sets.		
"Big Data Analytics: A Literature Review Paper"	Analyzing some of the different analyzing methods and tools which can be applied	Capturing, storing, searching, sharing and visualizing big data can	Improving the level of customer intelligence by using big data can be done.

	to big data. Discussing different opportunities provided by application of big data analytics in different domains of decisions.	become difficult without considerable computational power availability and other issues.	Improving Risk Management and Fraud Detection can be done to minimize security issues.
"Big Data, the perfect instrument to study today's consumer behaviour"	Discussing the evolution of consumer behaviour. Presenting advantages and disadvantages of using big data in studying consumers' behaviour Discussing how big data is analyzed.	More the data is registered, larger the problems will be that the analysts need to solve. People don't have a personal life anymore, since data of their activities are recorded. It is hard to ensure who use the data and how they are used. This data can lead to an avalanche of errors and incorrect results, which affect the entire business	Machine Learning can be used together with big data to enhance customer experience by suggestions and predicted results directed towards the users/ customers. Innovative business models can be bought to the organizations which results in more customer-responsive products.
"Review Paper on Big Data Analytics in Cloud Computing	Discussing the technologies and reviewing cloud computing usage in big data implementations	So far, researchers have failed to collect and present the features that are essential to big data as a whole, so that many has the impression that big data is something that we can not easily process. IT industry has faced many challenges since big data is gathered in massive amounts, daily, worldwide.	The vast amount of varied data produced on social media platforms can be used for identifying patterns via BDA in much better and improved ways.
"Big Data Using Cloud Computing"	Discussing the importance of cloud computing for big data	The computational requirements for Map Reducing	Different Cloud deployment methods and services are available and
	applications.	process are expensive, which is not affordable to most of the business organizations.	improving, at much more affordable costs which can be used by medium scale and small scale businesses.
"Cloud Computing Review: Concepts, technology, challenges and security"	1.Reviewing cloud computing concept :its historical evolution, concepts, tools etc. 2.Discussing the security threats in cloud computing	Security threats: 1. SQL injection attacks 2. Cross site scripting attacks 3. Cookie poisoning. 4. Sniffer attack. 5. DDoS attacks 6. Abuse of nefarious use of coud computing 7. Insecure API s. 8. Malicious insider actions 9. Vulnerability of shared technology. 10. Data loss/leakage. etc. Data loss. Limited migration ability from one CSP to another CSP.	There is a need of designing solutions of higher efficiency rate to address security issues : 1.data transfer across gateway 2.long-term viability, compromised services 3.regulatory compliance , virtualization Etc.
"Big Data Challenge: A Data Management Perspective"	Discusses big data concept in the view of data management perspective.		1. Data integration 2. Data reduction 3. Data querying and indexing 4. Data analyzing and mining Improvements for above aspects can be done and researched.
"Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions"	Investigate how cloud computing has impacted the field of ML. Surveying ML as a Software as a Service (SaaS)		Improvements in SaaS for Machine Learning (ML) Improvements for Text Mining (to improve extraction of data from social media)
"Performing Customer Behaviour	Presenting a proposed Map Reduce	1. Processing a huge set of data which have	The proposed system can be applied for

112

Analysis using Big Data Analytics"	implementation of a well-known statistical classifier.	a high variety is difficult. 2.Efficient use of storage should be maintained.	customer data visualization, using Map Reduce and HDFS.
."Big Data Computing using Cloud Based Technologies : Challenges and Future Perspectives"	Surveys big data concepts, discusses analytics techniques which can be used on big data. Gives a basic idea about existing tools, frameworks and platforms which are for big data computing.	1. A solid and scientific base must be developed to select the method which needs to be chosen. 2. There is a need for more efficient and scalable algorithms. 3. The developed algorithms can't be implemented unless appropriate technology platform has been selected.	1. Energy-effeicient and cost-effective solutions 2. Introducing proper standards for the solutions for the end-users 3. The data characteristics which are used to classify data as big data should be generated. 4. Feasibility and commercial viability should be improved for analitical applications.
"The Impact of Big Data Analytics on Customers' Online Behaviour"	Exploring effects of analyzing big data, brought on online behaviour of the customers ,to examine whether big data analytics is successful .		Firms could develop more person-specific promotion strategies to increase purchasing rate,in ways which increase more impulsive behaviour, using big data analytics

VII. Future directions of Big Data And Cloud Computing In Consumer Behaviour Analysis

Technologies related to the networking, cloud computing and big data analytics are constantly improving. The educated parties approve cloud computing as a solution which supports the businesses [6]. Also the marketing aspects and trends are prone to timely changes, imposing timely changes in consumer psychology, hence varying the consumer behaviour. Nonetheless, the future researches can focus on developing personalized marketing techniques for the customers, to maximize the profit and the number of long-term customers.[12]. Technologies should be developed to easily utilize structured and unstructured data to analyze consumer behaviour.[12]. Creative designs of cloud models can be designed to minimize the security risks and delegate the authorities properly to relevant parties[7]. Machine learning should be combined with big data analytics in creative methods which creates personalized customer experience[4].

The vast amount of data which can be gathered through social media can be incorporated in these methods [5].AI applications for real time big data analytics is a major upcoming trend [20]. Two major purposes of data mining and analytics are classification and prediction [6]. Incorporating machine learning and cloud based AI in big data analytics will definitely improve those aspects and can be employed by organizations to identify consumer behaviour patterns. Efficient algorithms for data reduction, integration and querying are open for researching, to improve the data management tasks of big data and cloud implementations [25].Also the security aspects in cloud environments have much space for improvements.

REFERENCES

[1]   Afolabi Ibukun, Olufunke Oladipupo, Rowland E. Worlu, Akinyemi I.O. "A Systematic review of Consumer Behaviour Prediction", Covenant Journal of Business and Social Sciences Volume: 7 2016

[2]   Rizwan Khan "A Review Paper on Cloud Computing", www.researchgate.net 2018

[3]   Nada Elgendy, Ahmed Elragal "Big Data Analytics: A Literature Review Paper", www.researchgate.net August,2014

[4]   Cristina Stoicescu "Big Data, the perfect instrument to study today's consumer behaviour",www.dbjournal.ro

[5]   Saneh Lata Yadav, Asha Sohal "Review Paper on Big Data Analytics in Cloud Computing" www.researchgate.net July, 2017

[6]   Fitri Purnamasari, Imam Subekti, Aulia Fuad Rahman "Consumer Behaviour of Interest and Use of Cloud Computing Technology as support Business Activities", Journal of Applied Management Volume 16 Number 1, March 2018

[7]   Bernice M. Purcell(Holy Family University) "Big Data Using Cloud Computing", www.researchgate.net

[8]   Nidal M. Turab, Anas Abu Taleb, Shadi R. Masadeh "Cloud Computing Challenges and Solutions",International Journal of Computer Networks & Communications(IJCNC) Vol. 5, No. 5, September 2013

[9]   Mahantesh Birje, Praveen Challagidad, R.H. Goudar, Manisha T. Tapale "Cloud Computing Review: Concepts, technology, challenges and security",www.researchgate.net January,2017

[10] D. Geethika, J. Naga Spandana, P.M.S.S. Chandu "Service Check: Effective User Behaviour Analysis& Trust Management-Identify Best Cloud",International Journal of Computer Engineering &Technology Volume 16, Issue 1, January-February 2017

[11] Mr. Bhupesh Kumar Deweangan, Mr. Praveen Shende "Survey on User

Behaviour Trust Evaluation in Cloud Computing", International Journal of Science, Engineering and Technology Research Volume 1, Issue 5, November 2012

[12] Qianling Chen, Min Zhang, Xiande Zhao "Analysing customer behaviour in moblie app usage", Industrial Management & Data Systems, Vol. 117 Issue: 2, pp.425-438

[13] Anindita A Khade "Performing Customer Behaviour Analysis using Big Data Analytics", Procedia Computer Science 79 (2016) 986-992

[14] Saavi Stubseid, Ognjen Aranjelovic "Machine Learning Based Prediction of Consumer Purchasing Decisions: The Evidence and Its Significance", www.researchgate.net

[15] Qiuchen Li, Jinkun Xing, Ou Liu, Woonkian Chong "The Impact of Big Data Analytics on Customers' Online Behaviour",Proceedings of the International MultiConference of Engineers and Computer Scientists 2017, Volume 2 March 15-17,2017,Hong Kong

[16] Abhijit Raorane, Raja Kulkarni ""Data Mining Techniques: A Source for Consume behaviour Analysis", 2011,International Journal of Database Management Systems. 03. 10.5121/ijdms.2011.3304.

[17] Miss Rohini A. Dhote, Dr. S.P. Deshpande "Data Mining with Cloud Computing", International Jpournal of Advanced Research in Computer Engineering &Technology, Voluume 5 Issue 1, January 2016

[18] Samiya Khan, Kashish Ara Shakil, Mansaf Alam "Big Data Computing using Cloud Based Technologies : Challenges and Future Perspectives", Book "Networks of the future : Architectures, Technologies amd Implementations, Publiusher: Chapman and Hall/ CRC

[19] Smaranika Mohapatra, Jharana Paikaray, Neelamani Samal "Future Trends in Cloud Computing and Big Data", Journal of Computer

Sciences and Applications, 2015, Vol. 3, No. 6, 137-142

[20] Affreen Ara, Aftab Ara "Cloud for Big Data Analytics Trends", IOSR Journal of Computer Engineering (IOSR-JCE), Volume 18, Issue 5, Ver. IV (Sep.-Oct.2016).PP 01-06

[21] Bhumika Pahwa, Dr. S. Taruna, Dr. Neeti Kasliwal "Role of Data Mining in analyzing consumer's online buying behaviour", International Journal of Business and Management Invention ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X www.ijbmi.org || Volume 6 Issue 11 || November. 2017 || PP—45-51

[22] Daniel Pop (Phd. ,West University of Timisoara) "Machine Learning and Cloud Computing: Survey of Distributed and SaaS Solutions", www.researchgate.net

[23] Emad Al-Shawakfa, Hiba Alsghaier "An empirical study of Cloud Computing and Big Data Analytics", International Journal of Innovative Computing and Applications

[24] T. Thiraviyam(V.O. Chidambaram College) , "Artificial Intelligence Marketing" , www.researchgate.net

[25] Jinchuan Chen (Renmin University of China), Yueguo Chen, Xiaoyong Du, Cuiping Li "Big Data Challenge: A Data Management Perspective", Frontiers Computer Science , 2013, 7(2): pg 157-164

About author :-

I am Ravindu Dias, an undergraduate of University of Moratuwa, Sri Lanka, following the degree Bachelor of Science in IT.

K. A. Dilini T. Kulawansa
Senior Lecturer
Department of Computational Mathematics, Faculty of Information Technology,
University of Moratuwa., Sri Lanka.