Challenges against Big Data as a Service: A Survey

Santosh Kumar Sharma[1*], Ajay Pratap[2], Harsh Dev[3]

[1,2]Amity Institute of Information Technology, Amity University, Lucknow
[3]Dept. of Computer Science & Engineering, Pranveer Singh Institute of Technology, Kanpur

*Corresponding Author: santosh.sharma.ddn@gmail.com, Tel.: +08791302181

Abstract— Big Data refers to the huge amount of data in terabytes and petabytes. These kinds of data is collected from different heterogeneous sources like defence, government data, health industry, banking sector, social media, sensor data, public data, transaction data, etc. And these data is yet growing exponentially with time. Big Data can be structured, semi-structured and unstructured by nature and handling such kind of data through traditional data management system has become very complex. Also there are a number of applications available in today's era which requires high speed data transmission capacity for the storing and computing large amount of data. This requirement can be fulfilled using Big Data as a Service. To handle the large amount of unstructured data the organizations use Big data as a service to free up the organizational resources by taking the advantages of predictive analytics skills of an outside service provider including storage and computing services, rather than hiring in-house staff. In this paper we have given the overview of Big Data as a service, its advantages, application areas and challenges faced by industry.

Keywords—Component Big Data, Big Data as a service, cloud computing.

I. INTRODUCTION

In today's scenario each and every person is using internet facility through mobile phone, personal computers, etc. And everyday huge amount of data is generated from different sources such as social networking sites (facebook, twitter), from the organization, from the transportation system, banking system, from government sectors, from banking systems, etc. These data are heterogeneous and unstructured in nature which includes text, audio, video, images, and so on. If we are talking about the banking system, the number of customers has increased a lot. Today people used banking facilities through Internet such as mobile banking, e-banking, net banking, credit/debit card, etc and because of it different kind of diverse data is generated every day. To handle such kind of data the concept of Big Data was proposed in 2005 by O'Reilly Media. Big Data is characterized by 5Vs as volume, velocity, variety, value and veracity [1]. Here volume refers to the quantity of data, velocity refers to the time taken in the processing of the high volume of data, variety refers to the different types of data, value refers to the worth of the data which is to be extracted and veracity refers to the trust and data reliability. But there were some problems like storage and processing of huge unstructured data growing very fast involved with Big Data since traditional data management technologies were unable to manage and process the large amount of data due to complexity involved in data. Also a lot of demanding mobile network services is available today such as cloud computing, grid computing, mobile cloud computing, etc which requires high data transmission rates for the specific operations like storing, managing, analyzing, processing the data on the cloud computing servers along with streaming of high definition videos with high speed. Cloud Computing is defined as a internet or network solutions for providing easy an d reliable services in the terms of software, infrastructure and platform to information technology resources [2]. Three categories of Cloud services are Platform as a Service (PaaS), Software as a Service (Saas), Infrastructure as a Service (IaaS) and Network as a Service (NaaS) [3][4][5]. To extract the meaningful information from the huge amount of data, investigation or data analysis is also required. So now data storage along with proper computing has become an important utility in this computing world. Big Data paradigm refers to a very large, complex and unstructured data set that traditional data management systems are inefficient to capture, store and process to extract the suitable information for the competitive advantages. As a result Big Data required more computing power and storage capacities as provided by cloud computing.
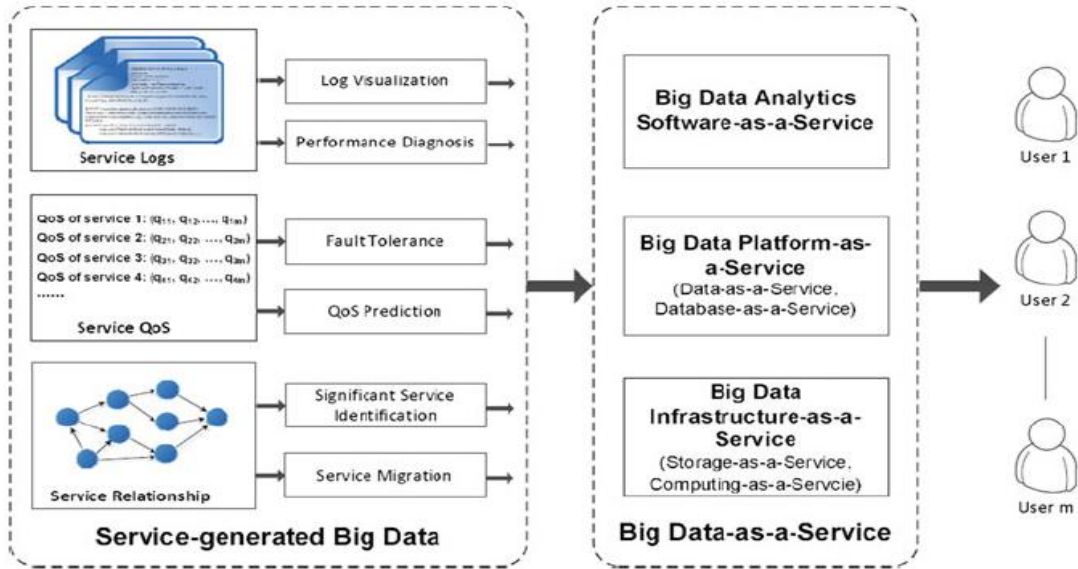
Fig 1: Service-generated big data and big data-as-a-service as presented by Zheng et al. [6]

Rest of the paper is organized as follows: in section II related works are given, in section III we have defined some of the benefits of BDaaS, in section IV application areas are defined, in section V some of the challenges against BDaaS are given and at the last conclusion is given.

II. RELATED WORK

Zibin Zheng et. al proposed the concept of service generated Big Data and Big Data as-a-Service platform as in fig 1 which consist of three layers as Big Data Infrastructure as a service (BDIaaS), Big Data Platform as a Service (BDPaaS), and Big Data Analytics Software as a Service (BDAaaS) [6]. It provides the functionality for managing, analyzing and processing different types of service generated Big Data along with the framework of Big Data as a Service. It has been employed to provide the Big Data services and analytics results to users for improving the competence and reducing the cost of computations.

Kim Yunkon et al surveyed enterprises and their products (outcomes) after that deduct criteria to classify and evaluate cloud based Big data services [7].

Big data platform provides the necessary recourses for accessing analyzing and building analytics applications on top of large data sets [8]. M. Rouse defined Big Data Analytics is the process of analyzing a huge amount of data of heterogeneous types to discover hidden patterns, hidden correlations and meaningful information [9]. Therefore for taking the service benefits, business organizations turn to Big Data Analytics as a Service [10]. Siham Yousfi et. al introduced a solution based on Hadoop ecosystem that collects and extracts data from social media and database tables [11]. A framework for Big Data as a Service was proposed by Xiaokang Wang et. al [12]. This framework is divided in three plane as sensing plane, cloud plane and application plane. The function of this plane is to organize data generated by various sources in CPSS (Cyber-Physical-Social Systems). Cloud plane is used for integration and processing, where Big Data including structure, unstructured and semi structured data are decomposed into smaller groups according to their applications. The purpose of application plane is to provide needed proactive services in various application domains. And then Big Data Analytics methods such as learning, mining and recommendation should be redesigned.

III. BENEFITS OF USING BDAAS

There are so many advantages or benefits of using Big Data as a Service like cost cutting, cloud infrastructure, Data Analytics, Data Storage, etc. as It provides storage as a service and computing as a service. In fig 2 major benefits are shown:
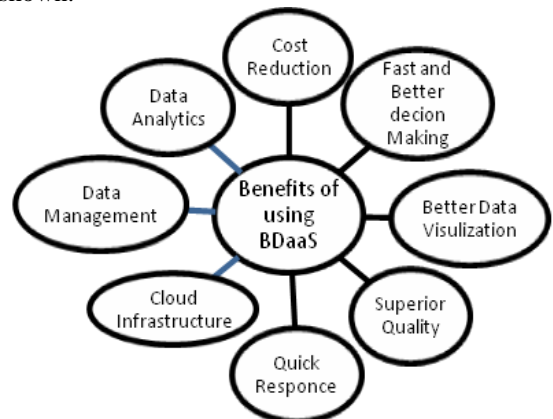

Fig 2: Benefits of Big Data as a Service

1. Cost Reduction: It could be very costly to manage an in-house data delivery mechanism, especially for those companies that don't use data as a core business process. BDaaS reduces these costs as it's a cloud-based functionality and delivering data applications to the business enterprises in an on-demand and cost-effective manner.

2. Fast and better decision making: Big Data Analytics as a Service gives the ability to analyze new sources of data in a proper and fast manner therefore businesses can analyze information immediately and make decisions based on what they've learned.

3. Better Data Visualization and Simplicity: Big data analytics tools allow users to query and manipulate data in a simplified form. Data can be accessible to any of the enterprise's divisions or locations quickly as the architecture of data delivery is very simplistic. It means that there is an option to incorporate data structure changes, or if any kind of new changes are required in the business environment, then alterations are fairly easy to implement.

4. Better Quality: Since the majority of data is controlled by the BDaaS provider, this results in an added layer of security and a high level of control over data quality. Security and privacy of the data is fairly managed by the Big Data as service provider, therefore business organizations need not to worry about the data theft.

5. Quick response: As most of the BDaaS providers today are based on Service Oriented Architecture (SOA), therefore offering a very high degree of flexibility while accessing mission-critical data from a cloud-based DaaS provider.

6. Data Management: Big Data as a Service provides direct access to the data within less amount of time and also provides the much possibilities for complex data analysis and modification.

7. Cloud Infrastructure: Big Data as a Service uses the infrastructure as a service of cloud including computing as a service and storage as a service for providing the better services to the business organizations also enables instantiation of IT infrastructure and determine capabilities of overlying infrastructure.

8. Data Analytics: It also provides the facilities where users can access Analytics as a service without having to deal with programming domains of Big data as a Service.

IV. APPLICATION OF BIG DATA AS A SERVICE

Multiple business organizations use the services of Big Data as a service to enhance the productivity of their organization without investing the cost for developing the infrastructure and other resources. There could be several application areas where Big Data Analytics is applied such as [13]: Education, Healthcare industry, Government Sector, Media & Entertainment Industry, Weather Patterns, Transportation Industry, Banking Sector, Manufacturing and Natural Resources, Internet of Things, Insurance, Retail and whole sale trade, Energy and utilities as shown in Fig 3.

V. CHALLENGES AND OPEN RESEARCH ISSUES

Although a lot of research work has been done in the field of Big Data as a Service but still more insight research work is needed. In the following section some of the challenges and open issues are described as identified in [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30]:

A. Diversity with increasing volume of data
As data is coming from different sources and these are mostly unstructured in nature. Also they belong to different formats and types. Although a lot of work has been done for handling such kind of diverse data set but still more work is needed.

B. Technologies used for information extraction
It refers to the selection of algorithms and methodologies for extracting the valuable information from the huge amount of heterogeneous data. Since data is heterogeneous and unpredictable therefore we should be very careful during selection of proper algorithm for cleaning and analyzing the Big Data sets.

C. Security
As data, which is coming into enterprises is made available from a wide range of sources, some of which cannot be trusted and might be harmful for the business organizations. This leads to the another challenge in front of the business to how to secure their data from various security breaches. Therefore it's necessary to work more on introducing Data Security best practices to secure data collection, storage and retrieval.

D. Privacy of Data
The gathering of data and applying different analytical tools for extracting information raises several privacy challenges. As we know that data is spread and replicate around the world therefore protecting privacy has become very difficult in today's scenario. A solution to this problem was to treat all data as personally identifiable and subject to a regulatory framework. Although, doing so might discourage business organizations from using de-identification methods and, therefore, increase privacy and security risks of accessing data.
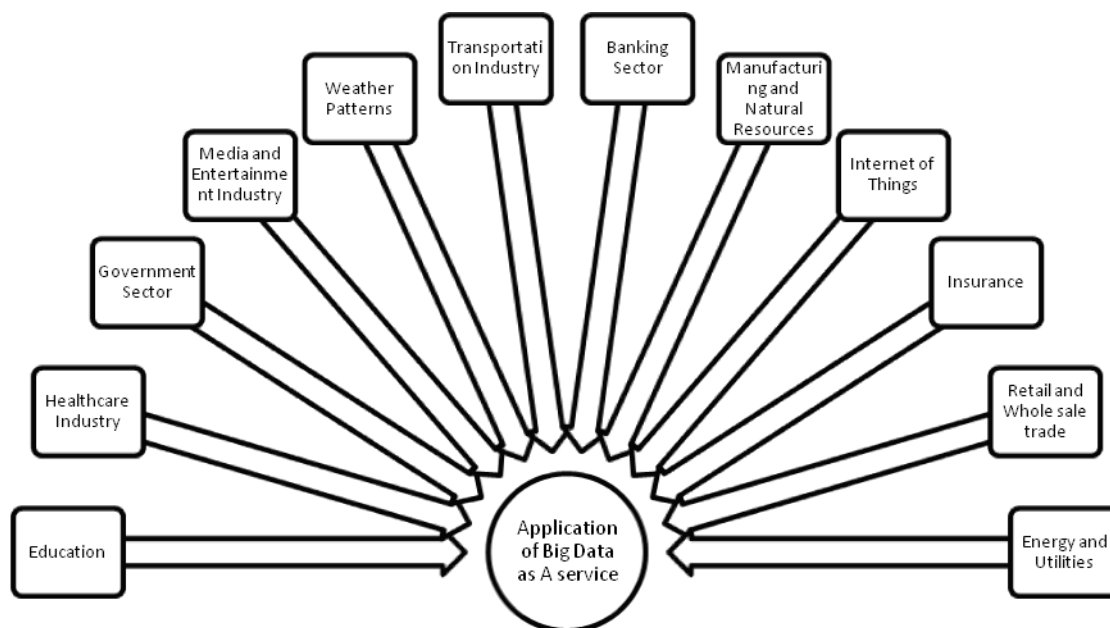

Fig. 3 Application areas of Big Data as a Service

E. Big Data Modelling

It refers to the applicable framework model for Big Data as a Service. As very less attentions is devoted in this area therefore more work is needed to design the accurate and perfect framework for modelling the Big Data nicely.

F. Storage Problem

Big Data means the huge amount of data coming from different sources in unstructured way. Therefore

storing such kind of data is a problem. The solution is to use Storage as a Service and Computing as a Service to store and process massive data. Although these technologies provide Infrastructure for automation of data collection, storing and visualization, big data enforce significant challenges to the conventional infrastructure, due to the characteristics of volume, velocity and variety.

G. Clear understanding of data and talent shortages

Clear understanding of the data to be processed is very important aspect when we are deploying the correct Big Data Analytics. If we unable to clearly characterize the data according to their types and formats it is very tough to implement the right analytics technique for extracting information. Also finding the right talent for the Big Data Analytics is another challenge in front of the business organizations as hiring of the right kind of people with the right experience and the right set of skills is more tedious and time consuming process.

H. Location of heterogeneous data

There is another challenge about the location of the data as the data can be in different locations. It includes the type of

processing which should be applied to the data, and where the processing should take place. It means that either the data is moved to a processing environment or the processing is performed where the data is localized. These challenges should also be taken into consideration while deploying the Big Data to a cloud system environment.

VI. CONCLUSION

Big Data Analytics technologies are evolving with the exponential rise in data availability. Big Data as Service is a service generated platform where cloud computing services are utilized with Big Data for the betterment of the organizations. In this paper we have given the overview of Big Data as a Service along with service generated framework. Also we have given the application areas and benefits of Big Data as a Service. And at the last we have identified some of the challenges against Big Data as a Service which are very important and crucial and need to be explored more for the betterment of the organizations.

REFERENCES

[1] ISO/IEC JTC 1. Information technology Big Data, Preliminary Report 2014. ISO/IEC 2015.
[2] J. Srinivas, K. Reddy, and A. Qyser, "Cloud Computing Basics," *Build. Infrastruct. Cloud Secur.*, vol. 1, no. September 2011, pp. 3–22, 2014.
[3] Eli Colins, "Bringing Big Data System to the Cloud", IEEE Cloud Computing, 2014.
[4] ISO/IEC 17789 , " Information Technology- Cloud Computing-Reference Architecture" ISO/IEC 2014.10.
[5] Mohammad Sajid, Zahid Raja, "Cloud Computing: Issues and Challenges", International Conference on Cloud, Big Data and Trust, 2013.11.

[6] Z. Zheng, J. Zhu, and M. R. Lyu, "Service-generated big data and big data-as-A-service: An overview," pp. 403-410, 2013.

[7] Yunkon Kim,Yong –Hyun Kim,Ga-Won Lee, Eui-Nam Huh, "Survey of Big Data as a Service Type", 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015.

[8] "Big data-as-a-service: A market and technology perspective," EMC Solution Group, Tech. Rep., 2012.

[9] M. Rouse, "Definition of big data Analytics", http://searchbusinessanalytics.techtarget.com/definition/big-dataanalytics, January 2012.

[10] "Why big data analytics as a service?" http://www.analyticsasaservice.org/why-big-data-analytics-as-aservice/ ,August 2012.

[11] Siham Yousfi, Dalila Chiadmi, "Big Data-as-a-Service Solution For Building Graph Social Networks", IEEE, 2015.

[12] Xiaokang Wang, Laurence T. Yang, Huazhong Liu and M. Jamal Deen, "A Big Data-as-a-Service Framework: State-of-the-Art and Perspectives", IEEE TRANSACTIONS ON BIG DATA, VOL. 4, NO. 3, JULY-SEPTEMBER 2018.

[13] Konstantinos Vassakis, Emmanuel Petrakis and Ioannis Kopanakis, "Big Data Analytics: Applications, Prospects and Challenges", Springer International Publishing, 2018.

[14] Ye, F., Wang, Z., Zhou, F., Wang, Y., Zhou, Y.: Cloud-based big data mining and analyzing services platform integrating R. In: Proceedings of the 2013 International Conference on Advanced Cloud and Big Data (CBD), Nanjing, pp. 147–151 (2013)

[15] http://www.bigdatavalue.eu/

[16] Georgios Skourletopoulos, Constandinos X. Mavromoustakis, George Mastorakis, Jordi Mongay Batalla, Ciprian Dobre, Spyros Panagiotakis and Evangelos Pallis, "Big Data and Cloud Computing: A Survey of the State-of-the-Art and Research Challenges", Springer International Publishing Switzerland 2017.

[17] McAfee, A., Brynjolfsson, E.: Big data: the management revolution. Harv. Bus. Rev. 90(10) 60–66, 68, 128 (2012)

[18] Lavalle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. 52(2), 3–22 (2010)

[19] The 2 types of data strategies every company needs. In: Harvard Business Review, 01 May 2017. https://hbr.org/2017/05/whats-your-data-strategy. Accessed 18 Jun 2017

[20] Waller, M.A., Fawcett, S.E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J. Bus. Logist. 34(2), 77–84 (2013)

[21] Janssen, M., van der Voort, H., Wahyudi, A.: Factors influencing big data decision-making quality. J. Bus. Res. 70, 338–345 (2017)

[22] V. Markl, "Breaking the chains: On declarative data analysis and data independence in the big data era," Proc. of VLDB Endowment, vol. 7, no. 13, pp. 1730–1733, August 2014.

[23] P. Russom, Big Data Analytics, TDWI best practices report, TDWI Research, 2014, http://www.iso.org/iso/home/news index/ news archive/news.htm?refid=Ref1821.

[24] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," IEEE Network, vol. 28, no. 4, pp. 46–50, July 2014.

[25] D. Wu, M. J. Greer, D. W. Rosen, and D. Schaefer, "Cloud manufacturing: Strategic vision and state-of-the-art," Journal of Manufacturing Systems, vol. 32, no. 4, pp. 564–579, 2013.

[26] K. E. Martin, "Ethical issues in the big data industry," MIS Quarterly Executive, vol. 14, p. 2, 2015.

[27] K. A. Salleh and L. Janczewski, "Adoption of big data solutions: A study on its security determinants using sec-toe framework," in CONF-IRM 2016 Proceedings, 2016. [Online]. Available: http://aisel.aisnet.org/conrm2016/66

[28] N. Rahman, "Factors affecting big data technology adoption," http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1099, 2016.

[30] IDC, "Six patterns of big data and analytics adoption," 3 2016, http://www.oracle.com/us/technologies/big-data/six-patterns-bigdata-infographic-2956541.pdf