e-ISSN 2320 –7876 www.ijfans.org
Vol.10, Iss.03, 2021
Research Paper **UGC CARE Listed (Group -I) Journal**

Big Data Environment AndApplications

Raju Barskar[1*]
[1*]Department of Computer Science and Engineering, UIT-RGPV Bhopal (M.P), India

Gulfishan Firdose Ahmed[2]
[2]Department of Computer Science, JNKVV, College of Agriculture, Powarkheda (M.P), India

***Corresponding Author:**Raju Barskar
*Department of Computer Science and Engineering, UIT-RGPV Bhopal (M.P), India

***Abstract*:**
Big data is data sets that are beyond the power of storage and processing capacity. A huge amount of data is generated from applications like the Internet, social networks, Bio-informatics, sensors, Weather forecasting, etc. Processing this huge dataset using Traditional database systems is impractical. To process these data sets, two processing techniques are used, one is the Batch processing technique and the other is the non-batch processing technique. In Batch processing, huge data are collected over a particular period, processed, and produced output. Non-batch processing includes continual input of data and processing in small time. As per thekinds of literature and papers in the field of big data processing techniques, it was found difficult for the new scholars to analyze the big data generated by various application areas with various processing techniques. In this paper, the suggestions of the most efficient and suitable processing techniques technique are given to process different types of datasets generated by various application areas, along with the challenges in storing and processing and the advantages of analyzing them. Suggestions of the most efficient techniques will help new scholars to opt for better techniques to handle big data generated by various applications and to give optimum results.

***Keywords*:**Big Data, routing protocols, mobility model, MapReduce, Wireless Sensor Network, Apache SPARK,TOSSIM.

1. Introduction

Big Data is a huge dataset, whose huge size is beyond the power of storage and processing capacity. Data increases exponentially from various application areas. The definition of Big Data depends on the three properties, of dimensions: Volume, Variety, and Velocity. Figure 1 explains the three V's of big data. Volume focuson the size or magnitude of data, variety explains the number of types of data and velocity explains the speed of processing data[1].

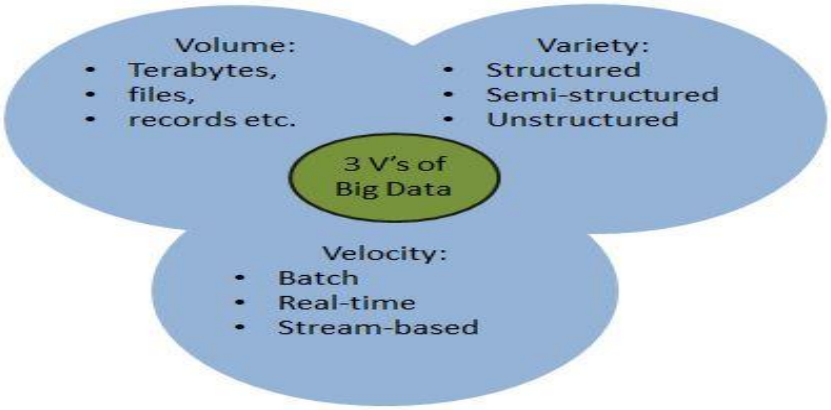

Figure 1: Traveling pattern of a Mobile node using theRandom Waypoint Model

Big Data is a set of Structured (Meta Data, Tabular Databases, etc.), Unstructured (Texts, videos, images, audio, etc.), and Semi-Structured (XML, JSON, Log Files, etc.) datasets. It is impractical to analyze this type of dataset using traditional database systems because, the challenges with the traditional database systems are: storage, processing, search, analysis, transfer, privacy etc. So, we needed some new techniques to manage it[2].

Big Data is being generated by almost every technology around us all the time. It is produced by social and digital processes, and transmitted by mobile devices and system sensors.

Data increased by 90% in the last 2 to 3 years. Figure 2 is brief about the generation of big data over twelve years. Data is generated from various sources: Web, Social networking sites, Sensors, GIS, Whether Forecasting, Bio-Informatics, Medical Science, etc. A survey shows, in 2014, Data generated every minute was: Facebook users share 2.5 million pieces of content, 3,00,000 tweets on Twitter, 2,20,000 new photos on Instagram, 72 hours of video content on youtube, 200 million messages over the Internet, 50,000 apps downloads on Apple, $80,000 online sales on Amazon, etc[3]. These all data are unstructured.

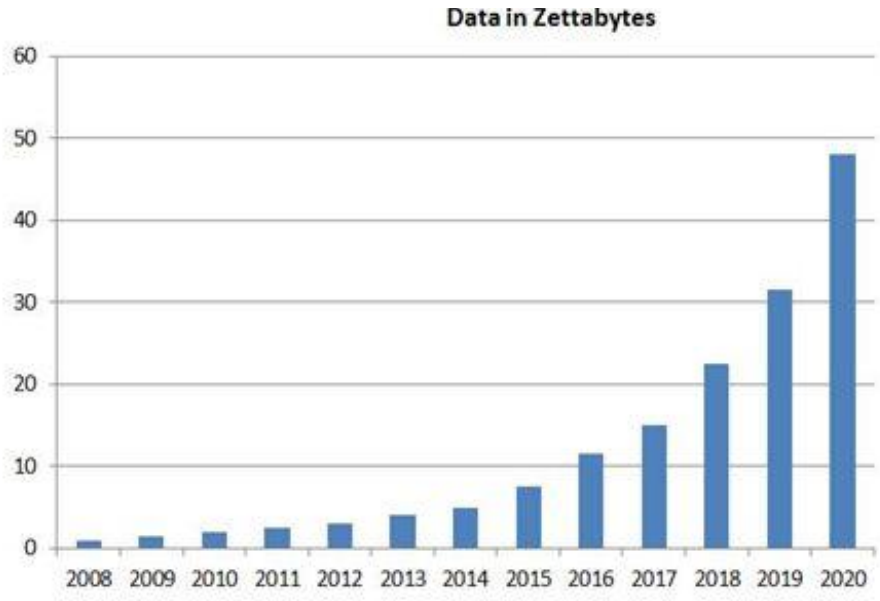

Figure 2: Traveling pattern of a Mobile node using theRandom Waypoint Model

Big Data has various applications in the field of Natural Language Processing, Human Behaviour Monitoring, GIS, Bio-informatics, Medical Science, Traffic Data Monitoring, Forecasting, Cloud

control systems, Multimedia, Body Sensor Networks, and many more. In all these areas, big data analysis plays an important role. Big Data can be processed by two processes: Batch processing and Non-Batch processing. The two Non-Batch processing techniques used to process big data are Real-time processing and Stream processing. In-Memory Computing and Real-time queries over Big Data are the two popular frameworks used for processing Real-time Big data processing. Storm and S4 are the two popular frameworks used for processing Stream processing.

In section two of this paper; various Big Data application areas and big data generated by these application areas are discussed. In section three, the various big data processing techniques are discussed. Then in section four, there is a comparative analysis between various application areas, their challenges, and the appropriate technologies used to handle big data generated by various sources[4].

2. Big Data Application Areas

Big data is generated from various application areas. Figure 3 shows the various big data application areas. In this section, the discussion about the application, how big data is generated from that particular application, the benefit to handle big data, and which processing technique is used to handle the big data, is done[5].

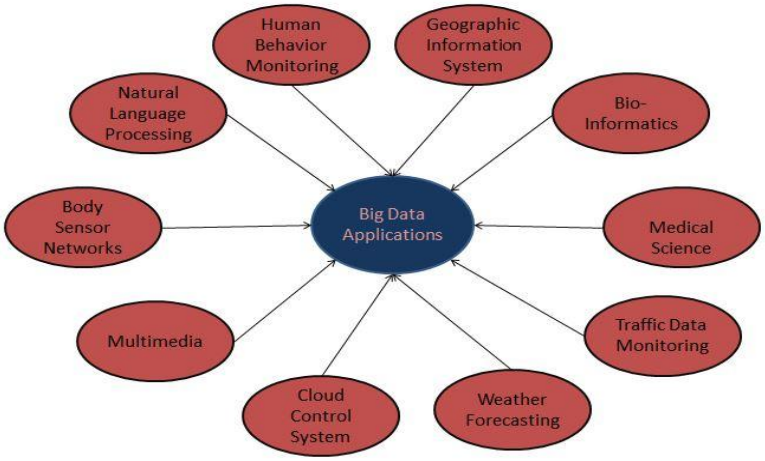

Figure 3: The Various Big Data Application Areas

2.1 Natural Language Processing (NLP)

Natural Language Processing is the study of designing machines or programs that can understand verbal and written communications. Extracting meaningful information from a large volume of unstructured human language is a Big Data problem. The unstructured data like voice calls, emails, text messages, etc. is increasing exponentially and need to be able to be accurately analyzed, which would lead to more accurate insights and better predictive model into human behavior. Figure 4 explains the big data generated by Natural Language Processing[6].

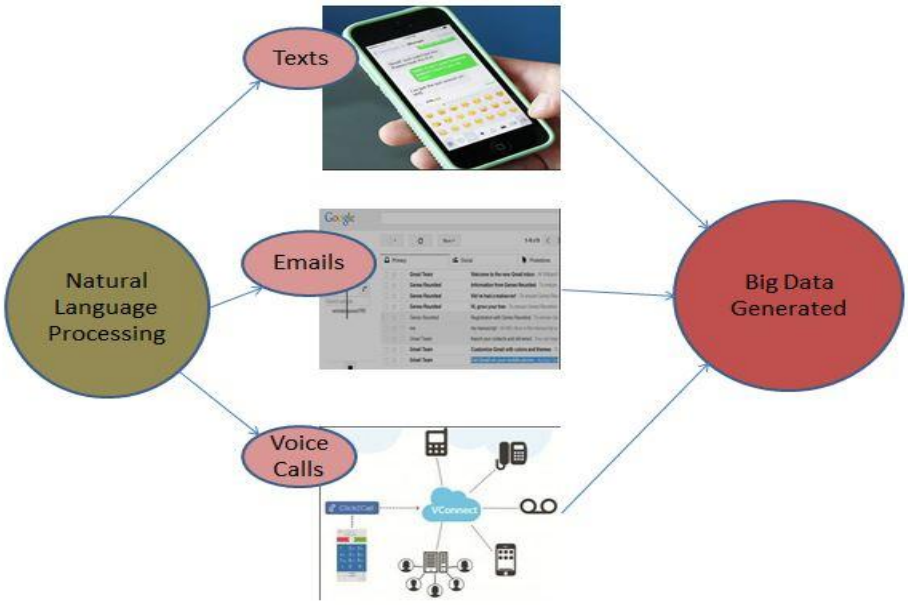

Figure 4: The Big Data Generated by Natural Language Processing.

Hadoop frameworkcan solve the NLP problems but Hadoop is a batch processing technique, so it requires a complete input set of each NLP module. It is very difficult or impractical to get a complete input set of each NLP module at the starting stage of execution. Apache SPARK[5] which is built on top of Hadoop and is also an extension of Hadoop is used to overcome this problem by executing interactive and streamed queries[7-8].

2.2 Human Behavior Monitoring

Everything we do, it generates data. Human life is becoming increasingly datafied. Emotions and sentiments, Relationships and interactions, Speech, Offline, and back-office activities, culture etc. generate a huge set of data. Figure 5 shows the big data generation by Human Behavior Monitoring. Analyzing big data on human behavior can lead to a detailed insight and better predictive model[9].

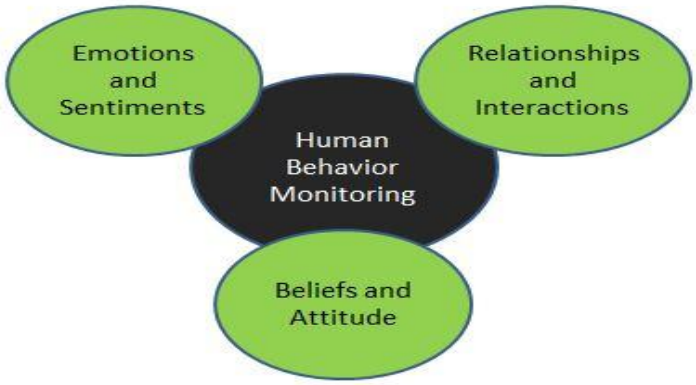

Figure 5: The Big Data Generation by Human Behavior Monitoring

Big Data technology Apache Hadoop can process this huge set of data. For better results, Stream-based processing Storm[7] can be used[10].

2.3. Geographic Information System (GIS)

It is a powerful system, which is designed for making better location decisions. It includes storing, manipulating, managing, collecting, sorting, select of geographical data. Figure 6 shows the big data generation by Geographic Information System. Controlling GIS[11], Big Data will enable several transformative societal applications. Societal applications in the context of understanding climate change, next-generation routing services, and emergency and disaster response.

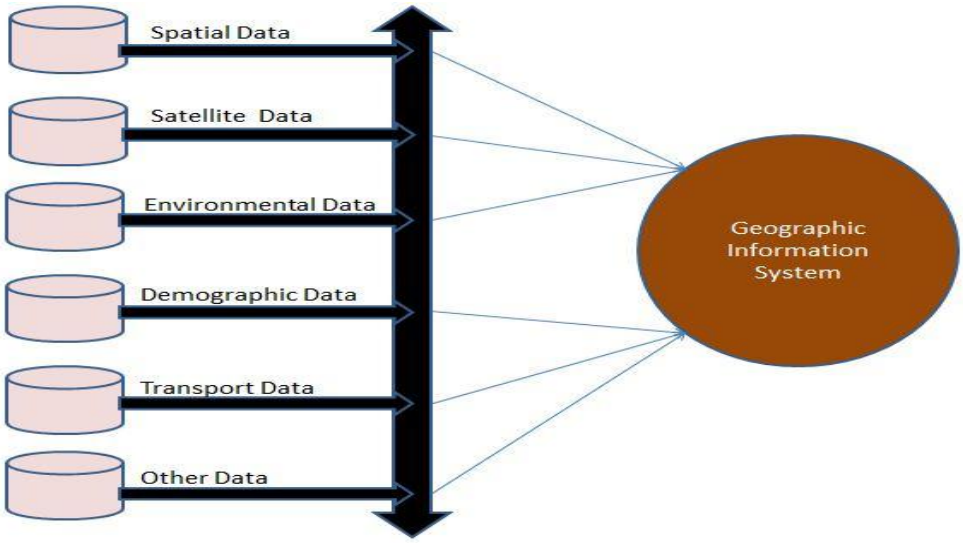

Figure 6: The Big Data Generation by Geographic Information System.

Apache Hadoop can process this large amount of data using MapReduce and HDFS. Apache Spark, which has a cheaper "Reduce" step, will need to be evaluated with iterative GIS workloads[12].

2.4. Bio-Informatics

Bio-informatics is the study of understanding the molecular mechanism of life on earth by analyzing Genomic information. Genomic information includes genomic sequencing and expressed gene sequencing. The sequencing mechanism has improved these days, which leads the volume of a sequence of data being produced to exceed the capabilities of the traditional method of a database model. Figure 7 explains the big data generated by Bio-Informatics. Big data possess a great impact on the bioinformatics field and a researcher in this field facesmany difficulties in using biological big data to extract valuable information from the data easily thereby enhancing further advancement in the decision-making process related to diverse biological processes, diseases, and disorders.

A tool Hadoop MapReduce platform, such as BioPig[12] and Crossbow[13] has been developed for sequence analysis[13].

2.5. Medical Science

Data is growing faster than medical Science can consume it. The unstructured data generated by medical science is huge (around 80% of the total relevant medical data). From genetic to genomic, internal imaging to the motion picture, treatment to life course assessment, etc. are all Big Data. Figure 8 tells about the big data generated by Medical Science. There is a need for Big Data technology to capture all the information of every patient. Big data techniques can be used to evaluate data generated from routine care of entire patients[14].

By statistically analyzing big data effectively, it will be beneficial to single Doctor's clinics as well as large hospital networks[15]. The easy and efficient analysis of big data benefits such as (i) detection of diseases can be done earlier; (ii) identification of health care fraud can be done more

quickly. Open source platform Hadoop MapReduce is used for big data analytics in Medical Science[16].

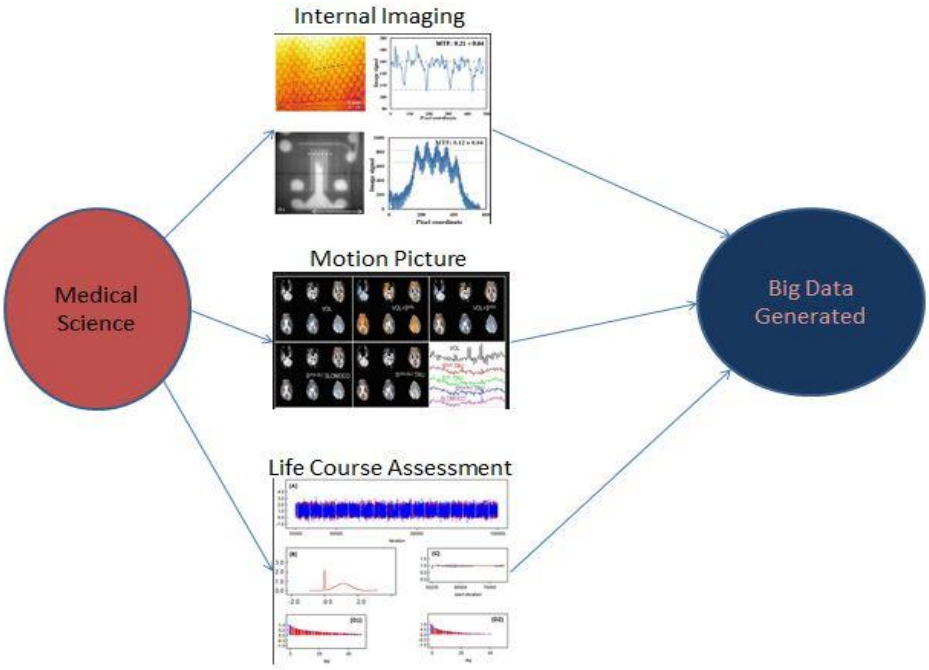

Figure 8: The Big Data Generation by Medical Science.

2.6. Traffic Data Monitoring

Traffic Data Monitoring analysis is significant for improving network resources and user experience. Detecting Diagnosing, fixing network problems, route profiling, capacity planning, congestion management, etc. are the applications that generate huge amounts of data**.** Figure 9 shows the big data generation by Traffic Data Monitoring[17].

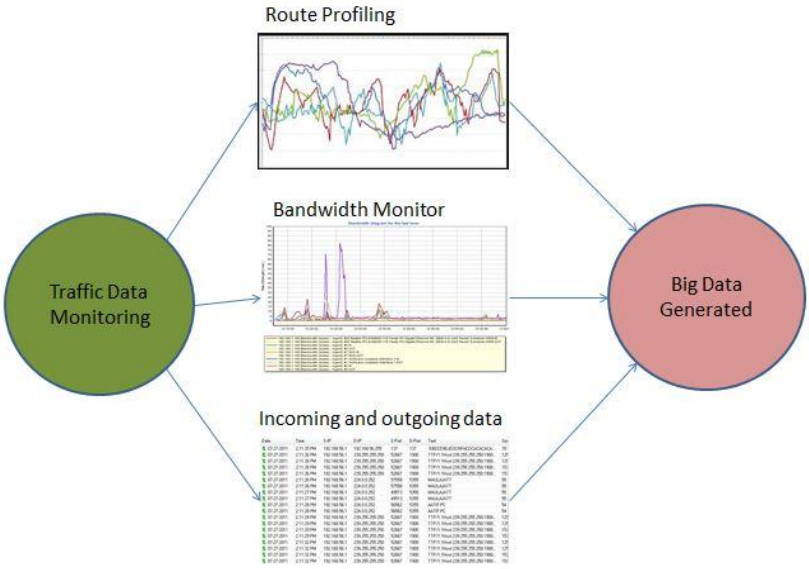

Figure 9: The Big Data Generation by Medical Science.

The analysis of this huge data will help in enhancing extensibility, easing-out programming, optimizing opportunities and efficiently processing the data generated from routers, switches, and from the website access logs at fixed interval[18].

Network traffic analysis is done by Apache Hadoop MapReduce framework and Hadoop Pig which is a platform for analyzing large data sets[19].

2.7. Weather Forecasting

Human has tried to get a better understanding of weather and forecast. Nowadays, satellite sensors and other resources are used by whether forecasting to fascinate everyone.

The volume of environmental data is increasing exponentially. So, we need a big data technique to manage, store, and process Big Data. Figure 10 shows the big data generated by Weather Forecasting.

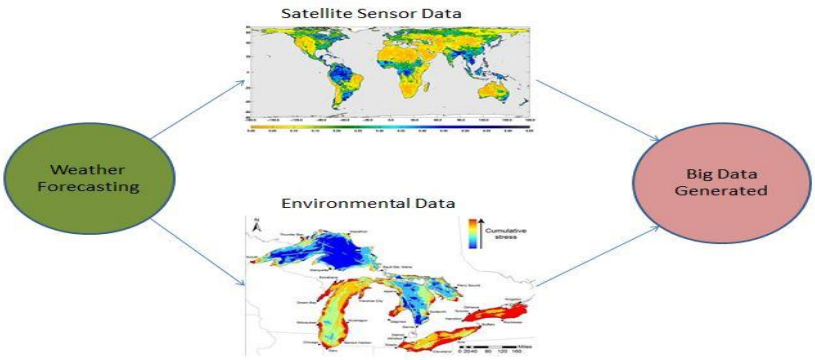

Figure 10: The Big Data Generation by Weather Forecasting.

one mobile entity and the remaining sensors are static. Mobile entities can communicate with neighboring sensors. Following the role played, mobile entities can be either mobile base station, which acts as network data collector, or mobile sensors that detect changes in the environment or serve as data relay nodes. The mobility of the mobile SB patterns or mobile sensors could be used to improve network performance, such as the network lifetime. These mobile units can be introduced naturally or artificially placed. The mobility model of each mobile entity is generally determined according to the specific application and the size of the WSN.

In reality, the mobility and deployment design of an MWSN is a complex problem that involves design requirements, mobility capability of mobile sensors, network environment, and application constraints such as time requirements. According to these design constraints, mobility strategy, collaborative model, data packets, and the routing protocol should be approached with caution in terms of network performance[20].

3. Conclusion and Future Work

This paper discusses big data applications in diversefieldsof computer science. This paper focuses on research areas and suggests the most efficient and suitable processing techniques technique is given to process different types of datasets generated by various application areas, along with the challenges in storing and processing and the advantages of analyzing them. Suggestions of the most efficient techniques will help new scholars to opt for better techniques to handle big data generated by various applications and to give optimum results. In this papaer basically used for various researcher to study the basics of the big data environment and its application. By using this basics analysis will make by applying various algorithms which concern with big data analytics.

References

[1]. Yick, J., Mukherjee, B., & Ghosal, D., "Wireless sensor network survey", *Computer networks*, *52*(12), pp.2292-2330, 2008.
[2]. Tan, L., Zhu, Z., Ge, F., & Xiong, N., "Utility maximization resource allocation in wireless networks: Methods and algorithms", *IEEE Transactions on systems, man, and Cybernetics: systems*, *45*(7), pp.1018-1034, 2015.
[3]. Azeem, M., Haleem, A., Bahl, S., Javaid, M., Suman, R., & Nandan, D., "Big data applications to take up major challenges across manufacturing industries: A brief review",*Materials Today: Proceedings*, *49*, pp.339-348, 2022.
[4]. Osinga, S. A., Paudel, D., Mouzakitis, S. A., & Athanasiadis, I. N., "Big data in agriculture: Between opportunity and solution", *Agricultural Systems*, *195*, 103298,2022.
[5]. Oulanouar, I., Rachedi, A., Lohier, S., & Roussel, G., "Energy-aware object tracking algorithm using heterogeneous wireless sensor networks", In *2011 IFIP Wireless Days (WD)* (pp. 1-6). IEEE, 2011.
[6]. Prasanna, S., & Rao, S., "An overview of wireless sensor networks applications and security", *International Journal of Soft Computing and Engineering*, *2*(2), pp.2231-2307, 2012.
[7]. Nikoletseas, S., & Rolim, J. D., "*Theoretical aspects of distributed computing in sensor networks",*(pp. 439-477). Berlin: Springer, 2011.
[8]. Addo-Tenkorang, R., & Helo, P. T., "Big data applications in operations/supply-chain management: A literature review",*Computers & Industrial Engineering*, *101*, pp.528-543, 2016.
[9]. Furht, B., &Villanustre, F., "Big data technologies and applications", 2016.
[10]. Guo, T., & Wang, Y., "Big data application issues in the agricultural modernization of China", *Ekoloji*, *28*(107), pp.3677-3688, 2019.
[11]. Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J., "Big data service architecture: a survey", *Journal of Internet Technology*, *21*(2), pp.393-405, 2020.
[12]. Fan, J., Han, F., & Liu, H., "Challenges of big data analysis", *National science review*, *1*(2), pp.293-314, 2014.
[13]. Wang, G., Ng, T. E., & Shaikh, A., "Programming your network at run-time for big data applications", In*Proceedings of the first workshop on Hot topics in software defined networks* (pp. 103-108), 2012.
[14]. Oussous, A., Benjelloun, F. Z., Lahcen, A. A., &Belfkih, S., "Big Data technologies: A survey",*Journal of King Saud University-Computer and Information Sciences*, *30*(4), pp.431-448, 2018.
[15]. Ge, X., & Jackson, J., "The big data application strategy for cost reduction in automotive industry", *SAE International Journal of Commercial Vehicles*, *7*(2014-01-2410), pp.588-598, 2014.
[16]. Sarawagi, A., Pandey, R., Barskar, R., & Pandey, S. P.,"A Real Time Stream Data Processing and Analysis Model and Catchments over Twitter Stream Data", *International Journal of Computer Applications*, *975*, 8887.
[17]. Arputhamary, B., and L. Arockiam. , "Data integration in Big Data environment." , *Bonfring International Journal of Data Mining* 5, no. 1, pp.1-5, 2015.
[18]. Sun, Y., Shi, Y., & Zhang, Z., "Finance big data: management, analysis, and applications", *International Journal of Electronic Commerce*, *23*(1), pp.9-11, 2019.
[19]. Ali, A., Qadir, J., Sathiaseelan, A., Zwitter, A., & Crowcroft, J., "Big data for development: applications and techniques", *Big Data Analytics*, *1*(1), pp.1-24, 2016.
[20]. Pyne, S., Rao, B. P., & Rao, S. B. (Eds.)., "*Big data analytics: Methods and applications*", Springer, 2016.