International Journal of Modern Electronics and Communication Engineering (IJMECE) ISSN: 2321-2152
Volume No.-6, Issue No.-6, November, 2018

A Literature Survey on Big Data Analytics in Context of Smart Cities

Ms. Aditi Tulchhia
Research Scholar
RTU, Kota
E-mail: aditi.tulchhia@gmail.com

Dr. Monika Rathore
Associate Professor
ISIM, Jaipur
E-mail: monikadrathore@gmail.com

Abstract: "Smart City" is a city which is advanced in use of technologies. A Smart City is able to understand its surroundings by examining methodically its data so that it can instantly make updating to resolve the issues and improving the quality of life of people. The extremely huge volume, broad variety and high velocity of city's data require making use of technologies of "Big Data" to obtain valuable understandings from it. Smart City concept is widely preferred because it builds up the quality of urban people's life, comprising various dimensions like, smart healthcare, smart parking, smart community and many more. Big data is offering the possibility for the cities to gain valuable understandings from a huge volume of data which is collected from different sources. This work starts from analysing the city's visibility that means data collection from various networks, sensors and devices that were embedded in its infrastructure. In context of improving system application of a city, the data should be conceptualized in useful shape, and hence the smartness of data driven city can be attained. This paper is a literature review on Big Data Analytics in the context of Smart Cities which addresses Big Data applications in Smart Cities, also discussing the 5 V's of big data, technologies of big data, smart city concept and applications. This article reviews the applications where Big Data technology can make a city smart. This review also demonstrates the practical applications of Big Data in context of a Smart City in smart health, smart energy, smart traffic systems and smart public safety domains

Keywords: Literature Survey, Big Data Analytics, Smart Cities, 5 V's of Big data

I. INTRODUCTION

Various serious challenges are faced by smart cities before global undertaking, but the combined use of Big Data and several other technologies to solve contemporary urban issues should ultimately results in their adoption.

Due to urbanization, there occurred many extraordinary challenges. Therefore, the concept of Smart Cities appeared as an approach to solve these challenges and however citizens are provided with better quality of life. By using ICT (Information and Communication technology), smartness of a city can be achieved. Thus, as a result, the smart cities are making use of various digital technologies in different types of disciplines of the city.

This paper is a literature review on Big Data Analytics and Smart Cities whose main focus is on providing answer to two basic research questions. The very first question is: What are the Big Data Analytics frameworks that can be applied to smart cities? The second question is: What are the main foundations based on which frameworks for big data analytics can be designed for smart city context? This literature review analyzes 14 articles which addresses Big Data applications in Smart Cities. The article also discusses about what is Big Data 5 V's of big data, technologies of big data, smart city concept and applications.

Several contributions can be made by Big Data for the development of Smart Cities. However, many different challenges are faced by Smart Cities. Some of them can be social issues, confidentiality of data, lack of funds etc. These challenges can be overcome by integration of Big Data and several other technologies.

II. WHAT IS BIG DATA ?

The data that is generated at an extraordinary rate across the world is referred to as Big Data. This data could be either structured, semi structured or unstructured. Big Data is defined as extremely huge data sets which can be computationally analyzed to disclose trends, relations and pattern, which are especially related to behavior of human and interactions.

To work with such a huge amount of data, a platform named Hadoop is preferred. In future only those enterprises will be successful, those who can make use of all the data. In order to occupy recently developed markets and customer base, enterprises should gain valuable understandings from the data at huge volumes and high velocity.

5 V'S OF BIG DATA:

The characteristics of Big Data can be defined by following 5 V's:

Volume: The organizations or enterprises collect the data from different sources, including business transactions, social media, and machine-to-machine data or sensor and hence this

huge volume of the data becomes a complex issue in Big Data analytics.

Velocity: Data streams in at an extraordinary speed. Velocity refers to the rate of generation of new data. This high velocity of data is due to our dependence on the sensors, internet and machine-to-machine.

Variety: The data to be collected comes in various types of formats:

1) Structured- numeric data in traditional databases.
2) Unstructured- text documents, audio, video, email, financial transaction and stock ticker data.

The generated data is totally heterogeneous and so the understanding of the type or format of Big Data is a very important factor to unlock its value.

Veracity: Veracity refers to the trustworthiness or messiness of the data. Before decrypting and implementing Big Data for business requisite, it is very important to know that whether the available data is coming from a reliable source or not**.**

Value: Last but not least, big data must have value. Value refers to the worth of the data being obtained. The huge endless volume of data is one of the important characteristics of big data but unless this data is useless, unless it can be turned into value.

BIG DATA TECHNOLOGIES:

Big data can be defined as any immense amount of structured, semi structured and unstructured data that has the potential to be extracted for information.

The importance of Big Data technologies is that they provide more accurate analysis which leads to more definite decision-making. To make use of the Big Data power, an infrastructure is required to manage and process extremely huge volumes of unstructured data and structured data in real-time and can protect data security and privacy.

Some technologies used for big data analytics:-

1. **Hadoop: -** Hadoop is an Apache open source framework written in java that makes use of simple programming models to enable processing of huge datasets in distributed fashion over collection of computers. The application which is based on Hadoop framework, will work in a surrounding which provides storage in distributed manner and computation is done over collection of computers.
2. **MongoDB: -** MongoDB is an open-source document database that provides high availability, high performance, and automatic scaling. To facilitate development, MongoDB hinders the need for an Object Relational Mapping (ORM).
3. **MapReduce: -** MapReduce refers to a programming model which allows writing applications to process Big Data in parallel on multiple nodes. MapReduce perform its work by dividing a task into small parts and then assign those smaller tasks to many computers. At last, the results from all the computers are collected at one place and then integrated to form the result dataset.
4. **Hive:-** Hive refers to a data warehouse infrastructure tool to process structured data in Hadoop. It is located on top of Hadoop to sum up Big Data, and makes querying and analyzing easy. Originally Facebook developed it but later it was taken by the Apache Software Foundation who developed it further as an open source under the name Apache Hive.
5. **Apache Pig:-** Apache Pig is an abstraction over MapReduce. Apache Pig refers to a tool or platform which is used to examine huge datasets representing them as data flows. It is usually used with Hadoop, all the data manipulation related operations in Hadoop can be performed using Pig.

III. WHAT IS SMART CITY

An urban area which makes use of different types of sensors to collect electronic data can be referred to as Smart City. This collected electronic data is used to provide information for effective management of resources and valuables. This comprises of data collected from devices, citizens, and assets that is operated and examined to observe and manage transportation and traffic systems, water supply networks, power plants, waste management, information systems, law enforcement, schools, hospitals, libraries, and other community services.

The concept of smart city combines information and communication technology (ICT), and different types of physical devices and sensors that are connected to the network, to improve the effectiveness of city operations and services and connect to residents. Smart city technology enables city officials to communicate directly with both city infrastructure and community and to monitor what is happening in the city and how the city is evolving.

Applications of Smart Cities:-

- **Smart parking**: Refers to monitor the city for availability of parking spaces.
- **Structural Health**: Monitoring of material conditions and throbbing in Bridges, buildings, and historical monuments.
- **Noise Urban maps:** *Real time monitoring of sound in centric zones and bar areas.*
- **Electromagnetic field levels**: The energy radiated by WiFi routers and cell stations is measured.
- **Traffic Congestion**: To optimize walking and driving routes, vehicles and pedestrian levels are monitored.

- **Waste management**: Identification of rubbish levels in containers to improve the trash collection routes.
- **Smart roads:** On the basis of climatic conditions and unpredicted events like traffic jams or accidents, warning messages and diversions are given by intelligent highways.

IV. LITERATURE REVIEW

Md Ileas Pramanik, Raymond Y.K. Lau, Haluk Demirkan, Md Abul Kalam Azad The author proposed a big data enabled smart healthcare system framework (BSHSF), providing a business model in the context of healthcare which represents the intra and inter organizational associations. In this research, for maximizing the capacity of big data analytics in the context of healthcare, used applications of 3 technologies: Intelligent Agent, Machine learning and Text Mining. They examined the change of technologies and applications in the context of data, city and healthcare [1].

Veda C.Storey, Il-YeolSong reviewed the five Vs of big data, volume, variety, velocity, value and veracity. The author identified the challenges that exist in context of big data, analyzes the importance of conceptual modeling in big data and suggested effective conceptual modeling efforts with respect to big data. The author concluded that integrating database management and the researcher's efforts in the discipline of conceptual modeling is needed. Also development is required to address challenges that arise as big data, the Internet of Things, and many other [2].

Xiufeng Liu, Alfred Heller, Per Sieverts Nielsen The author proposed a framework which is capable of managing smart city data. The data management includes collection of data, cleaning of data, anonymization and publishing of data. To verify the effectiveness and efficiency of a smart city, the author used a real world smart city data set for the evaluation of the framework. To manage smart city data, this framework can be considered a generic solution. A powerful and flexible ETL tool, BigETL, for big data processing was developed by the author for advancing smart city data management. The author also implemented an algorithm for linear regression and a data quality checking and anonymization library was also developed [3].

Nikolaos Panagiotou, Nikolas Zygouras, Ioannis Katakis, Dimitrios Gunopulos, Nikos Zacheilas, Ioannis Boutsis, Vana Kalogeraki, Stephen Lynch, and Brendan O'Brien A set of frameworks and techniques was developed by the author. The goal of this development is to provide urban data management. They examined a city named Dublin on the basis of real problems and data. To detect traffic congestion along with elasticity technique, they combined Apache's Spark and Hadoop so that varying volume of nodes can be adjusted automatically. The data sources at real time that were used from Dublin city are: BUS, SCATS and Twitter. This framework along with Complex Event Processing (CEP) helps to generate or modify the event detection rules. Social media used here is Twitter, which gives real time information. The proposed framework provides accurate, scalable real time event detection and the challenges that arise in urban data analysis were resolved by the techniques used [4].

Ibrahim Abaker Targio Hashem, Victor Chang, Nor Badrul Anuar, Kayode Adewole, Ibrar Yaqoob, Abdullah Gani, Ejaz Ahmed, Haruna Chiroma proposed a future business model of big data for smart cities and identified the business and technological research challenges. They described application based on smartness and the state-of-the-art communication technologies which were used in the smart city context. The state-of-the-art technologies include IoT, embedding sensors and equipment, Hadoop Distributed File System (HDFS), Cloud Computing, MapReduce. This article offered a comprehensive view of the contribution of big data in smart city. They concluded that to gain valuable information and also for decision making, big data plays an important role [5].

Ashwani Gupta, Nitin Katiyar, Brijendra Singh, Sweta Bhattacharya This article compares those big data architectures whose main focus is on the way of collecting data, data processing and enable fast decision making in a smart city context. This paper includes the "literature review of smart sensing using optical sensors to give users a basic understanding of how data in a smart city is collected and then monitored in real time". Also it provides review of two architectures BASIS and CIDAP focusing on its functionalities and usage criteria. Finally the author concluded that BASIS presented an architecture which is tested using different verification cases and CIDAP presented an architecture which is analyzed from an on-going European project in Santander [6].

Lixia Bao, Kai Yang,YiBo Wang, Junfeng Zhao proposed a novel pavement performance framework based on the exploration of big data in the context of smart city by using tensor decomposition. The proposed framework combines miscellaneous data from various sources and divided into 3 dimensions: spatial mode, temporal mode and feature mode. The count data is the element of tensor that describes the correlation of 3 modes. Through decomposition of tensor, some rank-one sub-tensors are generated which are used to predict the future pavement performance. CP-ARP (CANDECOMP/PARAFAC- Alternating Poisson Regression) algorithm is used for tensor decomposition with count data. They concluded that the collection and integration the data which is related to pavement performance from different sources can be performed by the proposed framework. By the potentially low-rank group, accurate pavement performance can be predicted easily [7].

Zaheer Khan, Ashiq Anjum, Kamran Soomro and Muhammad Atif Tahir A prototype is developed using MapReduce. This prototype is capable of determining how cloud infrastructure can be used to examine a selective set of Bristol Open Data. The analysis of Bristol Open Data was done by determining associations between selected indicators in an urban environment. Hadoop and Spark are used to perform various experiments and results are presented. The results concluded that when jobs are submitted to the cluster, Hadoop experiences significant overhead because of expensive operations for accessing data. On the other hand, Spark is faster and experiences significantly less overhead. Therefore, Spark is more convenient for the chosen Bristol open dataset [8].

Gao, S., et al designed and implemented novel Geo processing platform based on Hadoop for mining, storing, analyzing crowd sourced gazetteer entries from various web sources. This architecture comprises:- Web Crawler, Hadoop cluster, GIS client. They used Hadoop clusters and connected GIS to cloud computing environment for Big Geo data analysis, to offer new vision on enriching future gazetteers. MapReduce algorithm is used to count co-occurrence words is which makes it possible to quickly obtain place semantics parts and popular tags to characterize a place [9].

Majdi Beseiso, Abdulkareem Eid Al-Alwani proposed a framework that can be used to resolve challenges that are associated to data interoperability and related smart city environment challenges. They used various data related to smart city through different sources and stored in a NoSQL database. This data was scattered data, which was then transformed to machine-process able data by this framework. This proposed framework was compared with District Information and Management for Energy Reduction (DIMMER). The author found out that this model could be applied in multiple servers but it is not easy in DIMMER. They concluded that data was in resource description framework (RDF), so data interoperability issues were resolved and the use of semantic framework addressed the challenges related to integration of dynamic data and retrieval of information [10].

Anu Peisker, Soumya Dalai The author proposed a conceptual framework for data analytics applications to enhance development in rural areas. This study is based on supporting various sectors: banking, healthcare, education, governance and agriculture. To create a conceptual framework for both rural and urban development, important industry insights and opinion of the experts were considered as the fundamental data, which was collected from interviews of industry experts. The proposed framework comprises of 5 levels: - Data integration, Data normalization, Data Storage, Metrics or Correlation and Analytics. The analysis of huge amount of data was discusses for all the mentioned sectors of rural and urban areas. The author concluded that the knowledge of industry as well as analytics can make use of valuable understandings and insights. Thus skilled solution architecture can be constructed.[11]

V. Bassoo, V. Ramnarain-Seetohul, V. Hurbungs, T.P. Fowdur, and Y. Beeharry The author presented 4 state-of-the-art techniques for big data analytics. In this paper, the author also discussed the applications of big data analytics for five different smart city sectors. Finally, the challenges on security of big data are discussed and analytics used for smart cities is described. The big data analytics and techniques used for smart cities in this paper are: Edge Analytics, Semantic Analytics, Security Analytics and Predictive Analytics. The five smart city sectors mentioned above are: Transport, Weather Forecasting, Agriculture, Healthcare and Education. The application of Big Data analytics were discussed for all these 5 sectors. The security challenges comprises of privacy, integrity and confidentiality of smart city data. The author finally concluded that in future different types of IoT devices can be installed to collect huge amount of data and Big Data analytics are very important to gain values from such a growing volume of data [12].

Hisham Elhoseny, Mohamed Elhoseny, A. M. Riad, and Aboul Ella Hassanien The author proposed a framework that enables the adaption of traditional e-learning systems, which are to be made suitable for applications of smart cities. A Smart Learning framework is proposed which determines a conceptual model for platform of Big Data learning analytics. The author discussed the Related Work that is of applying smart learning in the applications of IoT and Big Data. The systems of smart learning which already exist were compared and discussed. The paper described the smart learning framework which comprises of 3 layers: Smart Information Transmission layer, Smart Devices layer and Smart Application Layer. These all 3 layers together enable transmission of data between a learner and framework. The author concluded that the framework is convenient to work with huge size of data in smart surroundings. Also different types of smart learning paradigms were investigated [13].

M. Victoria Moreno, Fernando Terroso-Sáenz, Aurora González-Vidal, Mercedes Valdés-Vela Antonio F. Skarmeta, Miguel A. Zamora and Victor Chang This paper mainly focuses on the interest or main foundations of Big Data for Smart Cities. The author also proposed a general architecture which is based on IoT. This architecture was divided into 4 layers. The 3rd layer in this architecture is the management layer where implementation of Big Data techniques was done. A case study of the smart campus of University of Murcia was discussed for some smart services implemented in that campus. On the other hand the analysis of information associated to Public Tram Service of the Murcia City was discussed. Finally, the benefits of Big Data applications for smart buildings and pattern recognition in urban areas were elaborated. Basically the two applications of big data were described for smart city services. These services

were:- public transport service of Murcia city and efficiency of energy and managing comfort the smart buildings. They concluded that they are able to provide mean energy savings of 23% per month [14].

V. CONCLUSION

This article is a literature review on Big Data Analytics frameworks in Smart Cities. The main focus of this paper is to answer the two basic questions. The very first question is: What are the Big Data Analytics frameworks that can be applied to smart cities? The second question is: What are the main foundations based on which frameworks for big data analytics can be designed for smart city context? In order to answer these questions, 14 articles have been reviewed and analyzed. In this review, several smart city frameworks that use big data analytics were discussed.

Now, some shortcomings have to be observed, which addresses future research; even though a fairly selected research articles were analyzed, many of them were not included in this paper because they did not meet the criteria of this study. The process of analysis denoted that the proposed frameworks which are available, lack many important features, like model determination and ensemble, privacy and security. Since the primary goal of Smart Cities is improving quality of life of citizens, advancement of the ethical and cultural aspects of life of citizens indicates a significant challenge. It is observed that this aspect is not given the necessary attention. Further more research is to be done for studying human behaviors, privacy, innovations, education, sentiments and emotion is required.

REFERENCES

[1].Md Ileas Pramanik , Raymond Y.K. Lau , Haluk Demirkan , Md Abul Kalam Azad ,"Smart Health: Big Data Enabled Health Paradigm within Smart Cities", Expert Systems With Applications (2017).

[2].Veda C.Storey, Il-YeolSong, "Big data technologies and management: What conceptual modeling can do", Data &Knowledge Engineering 0169-023X/& Elsevier B.V. 2017.

[3].Xiufeng Liu, Alfred Heller, Per Sieverts Nielsen, "CITIESData: a smart city data management framework", Springer-Verlag London 2017.

[4].Nikolaos Panagiotou, Nikolas Zygouras, Ioannis Katakis, Dimitrios Gunopulos, Nikos Zacheilas, Ioannis Boutsis, Vana Kalogeraki, Stephen Lynch, and Brendan O'Brien, "Intelligent Urban Data Monitoring for Smart Cities", B. Berendt et al. (Eds.): ECML PKDD 2016, Part III, LNAI 9853, pp. 177–192, 2016.

[5].Ibrahim Abaker Targio Hashem, Victor Chang, Nor Badrul Anuar, Kayode Adewole, Ibrar Yaqoob, Abdullah Gani, Ejaz Ahmed, Haruna Chiroma, "The role of big data in smart city", International Journal of Information Management 36 (2016) 748–758.

[6].Ashwani Gupta, Nitin Katiyar, Brijendra Singh, Sweta Bhattacharya, "Comparitive Study on Big Data Architectures Proposed For Smart Cities Context", IJPT| Dec-2016 | Vol. 8 | Issue No.4 | 25440-25451.

[7].Lixia Bao, Kai Yang,YiBo Wang, Junfeng Zhao, "A Novel Pavement Performance Prediction Framework in Smart City Based on Tensor Decomposition", " Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress, 2016".

[8].Zaheer Khan, Ashiq Anjum, Kamran Soomro and Muhammad Atif Tahir, "Towards cloud based big data analytics for smart future cities", Journal of Cloud Computing: Advances, Systems and Applications (2015) 4:2.

[9].Gao, S., et al, "Constructing gazetteers from volunteered Big Geo-Data based on Hadoop", Computers, Environment and Urban Systems (2014).

[10]. Majdi Beseiso, Abdulkareem Eid Al-Alwani ," An Interoperable Data Framework to Manipulate the Smart City Data using Semantic Technologies", "International Journal of Advanced Computer Science and Applications" January 2017.

[11]. Anu Peisker and Soumya Dalai, "Data Analytics for Rural Development", "Indian Journal of Science and Technology, Vol 8(S4), 50-60", February 2015.

[12]. Bassoo, V. Ramnarain-Seetohul, V. Hurbungs, T.P. Fowdur and Y. Beeharry, "Big Data Analytics for Smart Cities", "Springer International Publishing AG ", 2018.

[13]. Hisham Elhoseny, Mohamed Elhoseny, A. M. Riad, and Aboul Ella Hassanien, "A Framework for Big Data Analysis in Smart Cities", "Springer International Publishing AG, part of Springer Nature ", 2018.

[14]. M. Victoria Moreno, Fernando Terroso-Sáenz, Aurora González-Vidal, Mercedes Valdés-Vela Antonio F. Skarmeta, Miguel A. Zamora and Victor Chang, "Applicability of Big Data Techniques toSmart Cities Deployments", "1551-3203 (c) IEEE", 2016.