Big Data Analytics on IOT: Challenges, Open Research Issues and Tools

Mantripatjit Kaur[1*], Anjum Mohd Aslam[2]

[1] Dept. of CSE-IBM, Chandigarh University, Mohali, India
[2] Dept. of CSE, Chandigarh University, Mohali, India

Corresponding Author: mantripat63@gmail.com, Tel.: 9781790900

Abstract— A tremendous repository of terabytes of data is generated every day from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of such huge data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The main objective of this paper is to explore the potential impact of big data challenges, open research issues, and various tools associated with it. As a result, this article provides a platform to explore big data at numerous stages. Additionally, it opens a new horizon for researchers to develop the solution, based on the challenges and open research issues.

I. INTRODUCTION

Smart devices, Smartphones, Smart cars, Smart homes, Smart cities a smart world. These notions have come into the for many years. Achieving these goals has been investigated, to date, by various diverse research communities. Few such common research fields are: Internet of Things (IoT), Mobile Computing (MC), Wireless Sensor Networks (WSN) etc. However, as technology grows in each of such fields there is an elevated overlap of principles and research questions. Brief and narrower definitions of each of these fields are no longer suitable. Further, research in these areas usually depends upon numerous technologies including real-time computing, machine learning, security, privacy, big data, and others. All in all, the smart vision of the world includes much of computer science, computer engineering, and electrical engineering. Greater interactions among these communities will speed up the progress.

A. Internet of Things (IOT)

The technology of communication between human beings and things, or among things has developed rapidly [1]. Recently, machines are getting into the act to control countless autonomous gadgets using internet and create Internet of Things (IoT). As a result, these appliances are using the internet, just like human uses the web browsers. Internet of Things is a spotlight for researchers because of its most promising opportunities and challenges. In the upcoming era, everything will be connected and intelligently controlled. The concept of IoT is drawing attention of realistic world due to the development of mobile devices, ubiquitous communication technologies, cloud computing, and data analytics. Also, IoT offers collective challenges of volume, velocity and variety. Broadly, just like the internet, Internet of Things allows the devices to exist in plethora of places and provides facility to applications ranging from trivial to the crucial. On the contrary, it is still unable to understand IoT well, including its definitions, content and differences from other similar concepts. Numerous technologies such as computational intelligence, and big-data can be used collectively to improve the data management a knowledge discovery of large scale automation applications.

B. Big Data

Big data is a term mostly used to refer to the increase in the volume of data that are difficult to store, process, and analyze through traditional database technologies. We are in the age of big data [4],[5]. Big data has three typical characteristics which are volume, variety, and velocity [6]. That is, the size of big data is extremely large, the data types of big data are diverse such as number, text, and figure. In addition, the data processing of big data is so rapid. So, we should consider these characteristics when we analyses the big data. The nature of big data is indistinct and it involves considerable processes to identify and translate the data into new insights. The term "big data" is relatively new in IT and business. However, several researchers and practitioners have utilized

the term in previous literature. For instance, [14] referred to big data as a data which is in zetta bytes.

According to a study [8] by Gartner, the revenue that is generated from IoT-enabled services and products will exceed $300 billion by 2020. This, however, is only a hint of much larger issues soon to arrive. There is going to be a huge amount of data that IoT will generate, and in present world, well-analyzed data is of extreme value. It will force companies to quickly upgrade their ongoing processes, tools, and technology to accommodate vast data volumes and take full advantage of services that will be delivered by Big Data.

C. Relationship between IOT and Big Data

Processing all the data from IoT is a job in big data. Various types and formats (structured, unstructured, semi-structured) of data generated from various devices and sensors in IoT are handled and transformed by the "variety" of Big Data collection. The massive amount data originated from IoT are ingested and stored by the "volume" of Big Data loading [10]. The near real-time processing and analysis of tremendous data from IoT are performed in a timely manner by the "velocity" of Big Data analytics. In addition to this, the intelligence and insights of smart IoT objects are realized and augmented by the "value" of Big Data. The same is described in Fig 1 [9] below.
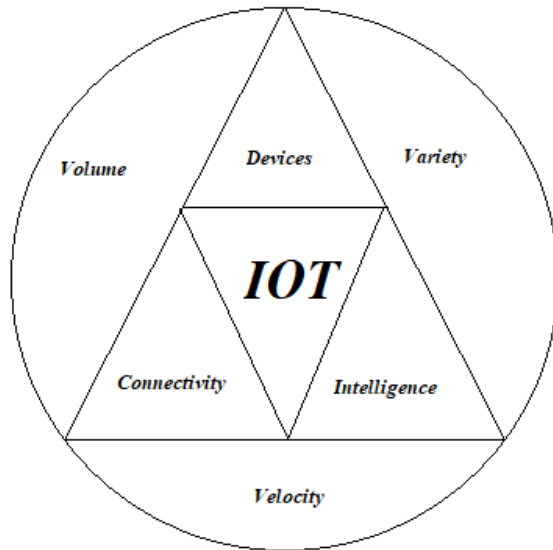

Fig1: Relationship between Big Data and IOT

II. RELATED WORK

Zhang et al. [13] devised Firework, a new computing paradigm that provides distributed data processing and sharing in an IoT-based, collaborative edge environment. Firework combines physically distributed data by allowing virtual data views to end users using predefined interfaces.

These interfaces are in the form of a set of functions and a set of datasets. Firework focuses on minimizing data access latency by moving the processing closer to the data producers in the edge network. Firework instance has numerous stakeholders who must register their datasets and corresponding functions that are abstracted as data views. These data views are available to all participants of that framework instance such that they can merge multiple data views into a single job to perform data analytics in detail.

Rathore et al. [14] describes a smart city management system based on IoT that exploits big data and analytics. The data are collected by deploying various sensors, including weather and water sensors, vehicular networking sensors, surveillance objects, smart home sensors, and smart parking sensors. An architecture for the system is also designed for its model and implementation. The proposed system is implemented using the MapReduce Hadoop ecosystem in a real environment. The implementation process has many steps, including data generation, data gathering, data combining, data categorization, data pre-processing, and decision making. Spark, an engine for big data processing is used over Hadoop for the efficient processing of big data. Smart systems are utilized as sources of city data to create a smart city as an implemented system. However, the proper developed smart system is yet to be deployed and its accuracy is still untested.

Ahlgren et al. [15] explains the significance of IoT usage to deliver services for improving the lives of citizens, like transportation, air quality, and energy efficiency. The authors emphasize that IoT-based systems must be based on open data and standards, involving interfaces and protocols, to enable third-party innovations by mitigating manufacturer lock-ins. Based on this idea, a Green IoT platform is designed in Sweden to deter- mine the advantage of open platforms and open data for the development of smart cities. However, few guidelines related the procurement of an open IoT infrastructure, including common data formats and open application programming interfaces (APIs), must be framed.

Sezer et al. [16] propose an augmented framework integrating semantic web technologies, big data, and IoT. The key requirements for the proposed framework are analysed, and the conceptual design of the envisioned IoT system is given based on the analysis results. The conceptual framework includes five layers, namely, data acquisition, extract-transform-load (ETL), semantic-rule reasoning, learning, and action. The data acquisition layer, which collects data from different sources, can be considered as an input layer for the framework. The ETL layer provides sensor drivers to transform the data received from various types of sensors. The semantic-rule reasoning supports a reasoning engine to make inferences from the resource description framework (RDF) data received from the ETL layer. The learning layer extracts many features from the data and forms machine-learning-based models.

The action layer provides predetermined actions for the   output of the learning layer.

Table 1: A Literature Survey of various IOT big data analytics use cases.

Analytic Types	Requirements	Tools	Pros and Cons
Collaborative Edge Environment [13]	Needs distributed data processing and sharing	Firework	+ provides virtual views to end users + minimize data access latency
Smart City Management System [14]	Needs an architecture that can collect data from sensors including weather sensors and vehicle sensors	Map-Reduce Ecosystem	+ efficient processing of big data using Spark over Apache Hadoop.
Network Edge [17]	Real Time Data Processing	GeeLytics	+address low latency analytics resulting from large amount of IOT data.
Health Care [22]	Real time data processing of massive data generated by body sensors	Apache Kafka and Spark	+ able to process large amount of data in real time -lacks security solution
Business Intelligence [28]	the size of data is larger than the memory level and traditional database is unable to handle such vast data	Map Reduce	+ mostly belong to offline

Cheng et al. [17] proposes GeeLytics, an edge analytics platform that performs real-time data processing at the network edges and in the cloud. This platform addresses the issues like geo-distributed and low-latency analytics resulting from the large amounts of IoT data. GeeLytics is designed to support dynamic stream processing topologies by considering the system characteristics of heterogeneous edge/cloud nodes, and the current system workload.

Wang et al. [18] discuss the challenges and opportunities resulting from IoT and big data for the maritime cluster. They also develop a new framework for integrating industrial IoT with big data and analytics technologies. Implementing such framework can help increase output and productivity as well as allow whole clusters to continue acting as leaders in the global maritime industry.

Vuppalapati et al. [22] explains the role of big data in healthcare and find that body sensors generate huge amounts of health-related data. Basically, two challenges are analyzed in this con- text, namely, integrating these massive data points with Electronic Health Records (EHR) and presenting these data to doctors in real time. Based on these observations, they propose a sensor integration framework that suggests a scalable cloud architecture that can provide a comprehensive approach to the EHR sensor system. Apache Kafka and Spark are used to process large amounts of data in a real-time. Although visualizing the health of patients in real time can help detect urgent situations, this model lacks a security solution.

III. ISSUES AND CHALLENGES

IoT systems are much capable to provide solution to many problems, but myriad of challenges remains unaddressed. The solutions to some of these challenges are yet to be provided by big data and analytics, while others require more efforts from the IoT community, hardware and platform vendors, governments, and policy makers.

Table 2: Various issues and challenges in IOT.

Issues in IOT	Challenges
Architecture	Many researchers proposed various architecture still not yet standardized the Architecture
Interoperability	Machine to Machine (M2M) communication
Security	-The entire communication between the physical device is wireless communication ii. -low capability devices participate.
Privacy	Profile access operations between devices without interferences are difficult
Availability	Ability to provide the anytime, anywhere, anything services are challenge.
Scalability	Ability to add new devices does not affect the Quality of Service is a challenge in.
Mobility	The connecting user or devices can be getting the services while on the move is a challenge.

Although, research on big data in IOT in its early stages, IOT has been broadly accepted by many organizations. Several existing issues have not been fully addressed yet as shown in table 2. New challenges continue to emerge from applications by organization. Due to convergence of big data and IOT, most issues of big data and IOT are resolved. Some remain unresolved.

IV. CONCLUSION

In recent years data are generated at a dramatic pace. Analyzing these data is challenging for a general man. To this end in this paper, we survey the various research issues, challenges, and tools used to analyze these big data. From this survey, it is understood that every big data platform has its individual focus. Some of them are designed for batch processing whereas some are good at real-time analytic. Each big data platform also has specific functionality. Different techniques used for the analysis include statistical analysis, machine learning, data mining, intelligent analysis, cloud computing, quantum computing, and data stream processing. We believe that in future researchers will pay more attention to these techniques to solve problems of big data effectively and efficiently

REFERENCES

[1]. D. Bandyopadhyay, and J. Sen, (2011) "*Internet of Things: Applications and Challengers in Technology and Standardization,*" Wireless Pers Commum Springer, Vol. 58, pp. 49-69.

[2]. L. Atzori, A. Iera, and G. Morabito, (2010) "*The Internet of Things: A survey,*" Computer Networks, Vol. 54, Iss. 15, pp. 2787.

[3]. L. Atzori, A. Iera, and G. Morabito, (2010) "*The Internet of Things: A survey,*" Computer Networks, Vol. 54, Iss. 15, pp. 2787-2805.

[4]. S. Lohr, (2012) *The Age of Big Data,* The New York Times, February 11, 2012.

[5]. N. R. Shikalgar, and D. Badgujar, (2013) "*Online Review Mining for Forecasting Sales,*" International Journal of Research in Engineering and Technology, Vol. 2, Iss. 12, pp. 53-55.

[6]. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, (2011) *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute.

[7]. S. Jun, and J. Choi, (2014) *"Big Data and Statistics, What's the Connection*?*"* Proceedings on KIIS Spring Conference, Vol. 24, No. 1, pp. 183-184.

[8] IEEE (2014) IEEE xplore, http://ieeexplore.ieee.org. [10]. USPTO (2014) The United States Patent and Trademark Office, http://www.uspto.gov.
[9]http://cloudonomic.blogspot.com/2015/03/iot-and-big-data

[10] M.R. Bashir , A.Q. Gill , *Towards an iot big data analytics framework: Smart buildings systems, in: High Performance Computing and Communications;* IEEE 14th International Conference on Smart City; IEEE 2nd International Con-ference on Data Science and Systems (HPCC/SmartCity/DSS), 2016

IEEE 18th International Conference on, IEEE, 2016, pp. 1325–1332 . [11] C. Lee , C. Yeung , M. Cheng , *Research on iot based cyber physical system for industrial big data analytics, in: Industrial Engineering and Engineer-ing Management* (IEEM), 2015 IEEE International Conference on, IEEE, 2015, pp. 1855–1859 .

[12] P. Rizwan , K. Suresh , M.R. Babu , Real-time smart traffic management system for smart cities by using internet of things and big data, in: Emerging Techno- logical Trends (ICETT), International Conference on, IEEE, 2016, pp. 1–7 .
[13] Q. Zhang , X. Zhang , Q. Zhang , W. Shi , H. Zhong , Firework: Big data sharing and processing in collaborative edge environment, in: Hot Topics in Web Sys- tems and Technologies (HotWeb), 2016 Fourth IEEE Workshop on, IEEE, 2016, pp. 20–25 .

[14] M.M. Rathore , A. Ahmad , A. Paul , Iot-based smart city development using big data analytical approach, in: Automatica (ICA-ACCA), IEEE International Con- ference on, IEEE, 2016, pp. 1–8 .

[15] B. Ahlgren , M. Hidell , E.C.-H. Ngai , Internet of things for smart cities: interoperability and open data, IEEE Internet Comput. 20 (6) (2016) 52–56 .

[16] O.B. Sezer , E. Dogdu , M. Ozbayoglu , A. Onal , An extended iot framework with semantics, big data, and analytics, in: Big Data (Big Data), 2016 IEEE Interna- tional Conference on, IEEE, 2016, pp. 1849–1856 .

[17]B. Cheng , A. Papageorgiou , F. Cirillo , E. Kovacs , Geelytics: Geo-distributed edge analytics for large scale iot systems based on dynamic topology, in: Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on, IEEE, 2015, pp. 565–570 .

[18] H. Wang , O.L. Osen , G. Li , W. Li , H.-N. Dai , W. Zeng , *Big data and industrial internet of things for the maritime industry in northwestern norway,* in: TEN- CON 2015-2015 IEEE Region 10 Conference, IEEE, 2015, pp. 1–5 .

[19] J.L. Pérez , D. Carrera , *Performance characterization of the servioticy api: an iot-as-a-service data management platform, in: Big Data Computing Service and Applications (BigDataService), 2015* IEEE First International Conference on, IEEE, 2015, pp. 62–71 .

[20] M. Villari , A. Celesti , M. Fazio , A. Puliafito , Alljoyn lambda: an architecture for the management of smart environments in iot, in: Smart Computing Work- shops (SMARTCOMP Workshops), 2014 International Conference on, IEEE, 2014, pp. 9–14 .

[21] A.J. Jara , D. Genoud , Y. Bocchi , *Big data for cyber physical systems: an analysis of challenges, solutions and opportunities, in: Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS),* 2014 Eighth International Conference on, IEEE, 2014, pp. 376–380 .

[22]C. Vuppalapati , A. Ilapakurti , S. Kedari , *The role of big data in creating sense ehr, an integrated approach to create next generation mobile sensor and wear- able data driven electronic health record (ehr), in: Big Data Computing Service and Applications* (BigDataService), 2016 IEEE Second International Conference on, IEEE, 2016, pp. 293–296 .

[23] Z. Ding , X. Gao , J. Xu , H. Wu , Iot-statisticdb: a general statistical database clus- ter mechanism for big data analysis in the internet of things, in: Green Com- puting and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, IEEE, 2013, pp. 535–543

[24] A. Ahmad , M.M. Rathore , A. Paul , S. Rho , Defining human behaviors using big data analytics in social internet of things, in: Advanced Information Network- ing and Applications (AINA), 2016 IEEE 30th International Conference on, IEEE, 2016, pp. 1101–1107 .

[25] E. Ahmed, M.H. Rehmani, Introduction to the special section on social collab- orative internet of things, 2017.

[26] I.-L. Yen , G. Zhou , W. Zhu , F. Bastani , S.-Y. Hwang , A smart physical world based on service technologies, big data, and game-based crowd sourcing, in: Web Services (ICWS), 2015 IEEE International Conference on, IEEE, 2015, pp. 765–772 . [27] R.P. Minch , Location privacy in the era of the internet of things and big data analytics, in: System Sciences (HICSS), 2015 48th Hawaii International Confer- ence on, IEEE, 2015, pp. 1521–1530 .

[28] Jourdan and R. K. T. E. Rainer Marshall, ''Business intelligence: An analysis of the literature 1,'' Inf. Syst. Manage., vol. 25, no. 2, pp. 121–131, 2008.

Authors Profile

Er.Mantripatjit Kaur pursed Bachelor of Computer Science and Engineering in 2013 and Master of Computer Science and Engineering in 2016 from Punjab Technical University. She is currently working as an Assistant Professor at Chandigarh University. She has published 6 research papers in reputed international conferences and journals. Her main research work focuses on IoT, Cloud, Big data-Hadoop.

Er. Anjum Mohd Aslam pursed Bachelor of Computer Science and Engineering from Punjab Technical University in 2014 and Master of Computer Science and Engineering from National Institute of Technical Teachers Training and Research, Chandigarh in year 2017. She is currently working as an Assistant Professor at Chandigarh University. She has published 5 research papers in reputed international conferences including Springer. Her main research work focuses on Cloud Computing, IoT, Big data and Analytics, Fog Computing.