A Literature study of Big Data Analytics and IoT

K.Sathiesh Kumar [1], M.Athigopal[2], M.Arun[1], S.Malathi[3]

[1]*Assistant Professor, Department of Computer Science and Applications,*
Sri Krishna Adithya College of Arts and Science College, Coimbatore, India
[2]*Assistant Professor, Department of Networking,*
Subbalakshmi Lakshmipathy College of Science, Madurai, India
[3]*Assistant Professor, Department of Information and Computer Technology,*
Sri Krishna Adithya College of Arts and Science College, Coimbatore, India

Abstract

IoT devices generate continuous streams of data in a scalable manner. Users must be able to handle this data and make it actionable. The actions performed on these high volumes of stream data may include analytics, statistics operation, metric calculation, or event correlation. The actions may vary depending on the big data scenario, and the data may not always be stream data. Put these factors into consideration before you build an analytic solution for your IoT data. Proper IoT data analytics should be growth, right-size infrastructure, and performance focused. A hybrid approach is best for performance and future growth. A hybrid deployment consists of dedicated hosting, colocation, managed hosting, and the cloud. It provides a single optimal environment by combining the best features from multiple platforms. For performance, the best option would be a single customer dedicated server. When setting up proper analytics infrastructure for IoT data, companies should remember that not all IoT data is important. This paper is a complete literature review about the IoT and Big data analytics.

Keywords: Smart objects, Big data, IoT.

Introduction

The Internet of Things (IoT) has been gaining momentum in both the industry and research communities due to an explosion in the number of smart mobile devices and sensors and the potential applications of the data produced from a wide spectrum of domains.

The collection of large and complex data sets which are difficult to processusing common database management tools or traditional dataprocessing applications.Big data is not just about size. Finds insights from complex, noisy,heterogeneous, longitudinal, andvoluminous data. With the rapid development of IoT, big data, and cloud computing, the most fundamental challenge is to explore the large volumes of data and extract useful information or knowledge for future actions. The key characteristics of the data in IoT era can be considered as big data; they are as follows.

(i) Large volumes of data to read and write: the amount of data can be TB (terabytes), even PB (petabytes) and ZB (zetta byte), so we need to explore fast and effective mechanisms.

(ii) Heterogeneous data sources and data types to integrate: in big data era, the data sources are diverse; for example, we need to integrate sensors data, cameras data, social media data, and so on and all these data are different in format, byte, binary, string, number, and so forth. We need to communicate with different types of devices and different systems and also need to extract data from web pages.

(iii) Complex knowledge to extract: the knowledge is deeply hidden in large volumes of data and the knowledge is not straightforward, so we need to analyze the properties of data and find the association of different data.

Big data and Analytics

Big data analytics enables organizationsto analyze a mix of structured, semi structured and unstructured data in search of valuable business information. Makinsey's internal Think-Tank, the Mckinsey Global Institute, published a major study in June 2011 on Big Data [1]. Its overloading conclusion: Big Data is "a key basis of competition and growth". The term Analytics (including its Big Data form) is often used broadly to cover any data-driven decision making [2]. The term analytics divided into two groups: Corporate analytics and Academic research Analytics. In Corporate Analytics, Team uses their expertise in statistics and Data mining. In Academic Analytics, Researchers analyze data to test Hypotheses and form theories [2].

L. Xu et al., [3] reviewed the privacy issues related to data mining by using a user-role based methodology. They differentiated four different user roles that are commonly involved in data mining applications, i.e. data provider, data collector, data miner and decision maker. A. Belle et al., [4] reviewed that the Big Data focused on three areas of interest: medical image analysis,physiological signal processing, and genomic data processing. V. Sujatha et al., [5] analyzed that the data sets from statistical models or complex pattern recognition models may be fused into predictive models that combines data set of patients' treatment information and prognostic outcome results. S. Vennila and J. Priyadarshini., [6] promoted that the security in Big data is a challenging research issue. If Integration of MapReduce, a machine for privacy preserving, is designed for the analyzing of data would provide better privacy.

Kovalchuk et al., [7] represented an early stage of the work aimed to the development of a general-purpose concept of the P4 CDSS rising from a treatment-level scope to a hospital-level scope. J. Cunha, C. Silvaa and M. Antunes[8] proposed a generic

 ISBN: 978-93-87793-72-9

functional architecture with Apache Hadoop framework and Mahout for handling, storing and analyzing big data that can be used in different scenarios. Z. Liu et al., [9] presented an agent-based model of emergency department that was implemented inNetlogo simulation environment. Case studies have been carried out for proving two of the possible uses of the simulator, one to meet the increasing patient arrival overcrowding problem, and the second a quantitative analysis of the influence of ambulance response time (for departure) over the ED behavior.

Srivathsan and Y. Arjun[10] proposed that PrognotiveComputing recognize patterns and formulates its own structure to provide a solution or gives a predicted alert so as to find a solution by ourselves .The System provides a handle of Health care and life span of numerous life forms. A. Abbaset al., [11]stated that they propose a cloud based framework that effectively manages the health related Big-data and benefits from the ubiquity of the Internet and social media. The framework facilitates the mobile and desktop users by offering: (a) disease risk assessment service and (b) consultation service with the health experts on Twitter. F. Zhanget al., [12] proposed a task-level adaptiveMapReduce framework. This framework extends the generic MapReduce architecture by designing each Map and Reduce task as a consistent running loop daemon. The beauty of this new framework is the scaling capability being designed at the Map and Task level, rather than being scaled from the compute-node level.

Analysis of IoT

In 2012, Miorandi et al. [13] survey the IoT mainly from the perspective of the key issues and research challenges and some initiatives going on to address them. Barnaghi et al. [14] look at developments in the semantic web community, analyzing the advantages of semantics but also highlighting the challenges they face and review work on applying semantics to the IoT. Chen et al. [15] study, using bibliometrics; some of the key research areas in business intelligence and analytics, some application areas and propose a framework to classify them. In 2013, Sagiroglu et al. [16] give an overview of the big data problem, methods to handle the big data, analysis techniques and challenges. Vermesan et al. [17] look at the vision, applications, governance and challenges of the IoT and some proposed solutions like semantics.

In 2014, Perera et al. [18] present a study of context-aware computing and discuss how it can be applied to the IoT. Zanella et al. [19] survey the enabling infrastructure and architecture for the Internet of Things in an urban, connected, smart city scenario while Xu et al. [20] review the development of IoT technologies for industries. Zhou et al. [21] discuss the challenges brought to data analytics by big data from the perspective of various applications while Kambatla et al. [22] discuss trends with a focus on hardware and software platforms,

 ISBN: 978-93-87793-72-9

virtualisation and application scopes for analytics. Another big data survey is done by Chen et al. [23] who look at challenges and work done from each stage of “data generation, data acquisition, data storage, and data analysis”. They also look at applications of big data briefly, where one such area is the IoT. Finally, Stankovic[24] proposes a set of research directions and considerations for future research on the IoT.

In 2015, Granjal et al. [25] survey existing protocols for protecting communications on the IoT, comparing against a set of fundamental security requirements, and highlight the open challenges and strategies for future research work. Al-Fuqaha et al. [26] focus on giving a thorough summary of protocols for the IoT and how they work together for applications in big data scenarios.

In 2016, Ray [27] surveys domain-specific architectures for the IoT providing a brief summary of whether cloud platforms in the IoT support data analytics. Razzaque et al. [28] survey middleware platforms for the IoT against a set of comprehensive service and architectural requirements. In 2017, Akoka et al. [29] perform a systematic mapping study, a method for structuring a research field, to classify big data academic research and identify trends in the research. Both analytics and the IoT were identified as popular topics. Reviews by Lin et al. [30] and Farahzadia et al. [31] focus on specific IoT research areas of fog computing architectures and middleware for cloudcomputing platforms.Sethi et al. [32] take the approach of surveying IoT architectures, protocols and applications which help them organizetaxonomy of IoT research.

Conclusion

The Internet of Things (IoT) has huge potential to provide advanced services and applications across many domains and the momentum that it has generated, together with its broad visions, make it an ideal frontier for pushing technological innovation. We have shown that analytics plays a role in many applications, across many domains, designed for the IoT and will be even more important in the future as the enabling infrastructure develops and scales and the deployment of devices becomes truly ubiquitous. This paper provides a systematic review of analytics applications in the IoT to the task of understanding analytics as it develops. This results in a layered taxonomy that defines and categorizes analytics by their capabilities and application potential for research and application roadmaps.

References

[1]James Manyika, Michael Chui, Brad Brown, Jacques Bhuhin, Richard Dobbs, Charles Roxburgh, Angela Hungh Byers, "Big Data: The next frontier for innovation, competition and productivity", June 2011.

[2] D.Fisher, R.Deline, M.Czerwinski and S. Drucker,"Interaction with big data analytics", Volume 19, No.3, May 2012

[3]L. Xu et al., "Information Security in Big Data: Privacy and Data Mining", IEEE Access, vol. 2, pp. 1149-1176, 2014.

[4]A. Belle, R. Thiagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard and K. Najarian "Big Data Analytics in Healthcare", Hindawi Publishing Corporation, BioMed Research International, vol. 2015, article ID 370194, 16 pages.

[5]V. Sujatha et al., "Bigdata analytics on Diabetic Retinopathy Study (DRS) on real-time data set identifying survival time and length of stay", Procedia Computer Science, vol, 87, pp. 227-232, 2016.

[6]S. Vennila and J. Priyadarshini., "Scalable Privacy Preservation in Big Data A Survey", Procedia Computer Science, vol. 50, pp. 369 –373, 2015.

[7]Kovalchuk et al., "Personalized Clinical Decision Support with Complex Hospital-Level Modelling", Procedia Computer Science, vol.66, pp. 392–401, 2015.

[8]J. Cunha, C. Silvaa and M. Antunes "Health Twitter Big Bata Management with Hadoop Framework", Procedia Computer Science, vol. 64, pp. 425 –431, 2015.

[9]Z. Liu et al., "Quantitative Evaluation of Decision Effects in the Management ofEmergency Department Problems", Procedia Computer Science, vol. 51, pp. 433 –442, 2015.

[10]S. Ma and Y. Arjun, "Health Monitoring System by Prognotive Computing using Big Data Analytics", Procedia Computer Science, vol. 50, pp. 602 –609, 2015.

[11]A. Abbasetal.,"Personalized healthcare cloud services for disease risk assessment and wellness management using social media", PervasiveandMobileComputing,vol. 28, pp. 81–99, Jun. 2016.

[12]F. Zhanget al., "A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications", FutureGenerationComputerSystems, vol. 43–44, pp.149–160, 2015.

[13] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and ImrichChlamtac. 2012. Internet of Things: Vision, Applications and Research Challenges. Ad Hoc Networks 10, 7 (2012), 1497–1516.

[14] PayamBarnaghi, Wei Wang, Cory Henson, and Kerry Taylor. 2012. Semantics for the Internet of Things: Early Progress and Back to the Future. International Journal on Semantic Web and Information Systems 8, 1 (2012), 1–21.

[15] Hsinchun Chen, Roger H L Chiang, and Veda Storey. 2012. Business Intelligence and Analytics: From Big Data To Big Impact. MIS Quarterly 36, 4 (2012), 1165–1188.

[16] SerefSagiroglu and DuyguSinanc. 2013. Big Data: A Review. In International Conference on Collaboration Technologies and Systems.

[17] OvidiuVermesan and Peter Friess. 2013. Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems. River Publishers.

[18] CharithPerera, ArkadyZaslavsky, Peter Christen, and DimitriosGeorgakopoulos. 2014. Context Aware Computing for the Internet of Things: A Survey. IEEE Communications Surveys and Tutorials 16, 1 (2014), 414–454.

[20] LidaXu, Wu He, and Shancang Li. 2014. Internet of Things in Industries: A Survey. IEEE Transactions on Industrial Informatics PP, 4 (2014), 1–11.

[21] Zhihua Zhou, Nitesh V Chawla, Yaochu Jin, and Graham J Williams. 2014. Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives. IEEE Computational Intelligence Magazine 9, 4 (2014), 62–74.

[22] KarthikKambatla, GiorgosKollias, Vipin Kumar, and AnanthGrama. 2014. Trends in Big Data Analytics. J. Parallel and Distrib. Comput. 74, 7 (2014), 2561–2573.

[23] Min Chen, Shiwen Mao, and Yunhao Liu. 2014. Big Data: A Survey. Mobile Networks and Applications 19 (2014), 171–209.

[24] John A Stankovic. 2014. Research Directions for the Internet of Things. IEEE Internet of Things Journal 1, 1 (2014), 3–9.

[25] Jorge Granjal, EdmundoMonteiro, and Jorge Sa Silva. 2015. Security for the Internet of Things: A Survey Of Existing Protocols and Open Research Issues. IEEE Communications Surveys and Tutorials 17, 3 (2015), 1294–1312.

[26] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and MoussaAyyash. 2015. Internet of Things: A Survey on Enabling Technologies, Protocols and Applications. IEEE Communications Surveys and Tutorials 17, 4 (2015), 2347–2376.

[27] ParthaPratim Ray. 2016. A Survey on Internet of Things Architectures. Journal of King Saud University – Computer and Information Sciences (2016).

[28] Mohammad AbdurRazzaque, MarijaMilojevic-Jevric, Andrei Palade, and SiobhánCla. 2016. Middleware for Internet of Things: A Survey. IEEE Internet of Things Journal 3, 1 (2016), 70–95.

[29] Jacky Akoka, Isabelle Comyn-Wattiau, and Nabil Laoufi. 2017. Research on Big Data - A Systematic Mapping Study. Computer Standards & Interfaces 54, 2 (2017), 105–115. https://doi.org/10.1016/j.csi.2017.01.004

[30]Jie Lin, Wei Yu, Nan Zhang, Xinyu Yang, Hanlin Zhang, and Wei Zhao. 2017. A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications. IEEE Internet of Things Journal (2017).

[31] AmirhosseinFarahzadia, Pooyan Shams, JavadRezazadeh, and Reza Farahbakhsh.2017. Middleware Technologies for Cloud of Things - A Survey. Digital Communications and Networks 3, 4 (2017), 1–13.

[32] PallaviSethi and Smruti R Sarangi. 2017. Internet of Things: Architectures, Protocols, and Applications. Journal of Electrical and Computer Engineering 2017 (2017).