

A survey and analysis of big data management based on computational methodologies

Shweta Kumari^{1*}, Kailash Patidar², Rishi Kushwah² and Gaurav Saxena²

M.Tech Scholar, Department of Computer Science Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India¹

Assistant Professor, Department of Computer Science Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India²

Received: 1-July-2020; Revised: 24-August-2020; Accepted: 28-August-2020

©2020 Shweta Kumari et al. This is an open access article distributed under the Creative Commons Attribution (CC BY) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In a business enterprise there is an enormous amount of data generated or processed daily through different data points. It is increasing day by day. It is tough to handle it through traditional applications like excel or any other tools. So, big data analytics and environment may be helpful in the current scenario and the situation discussed above. This paper discussed the big data management ways with the impact of computational methodologies. It also covers the applicability domains and areas. It explores the computational methods applicability scenario and their conceptual design based on the previous literature. Machine learning, artificial intelligence and data mining techniques have been discussed for the same environment based on the related study.

Keywords

Big data, Machine learning, Artificial intelligence, Data mining.

1.Introduction

The term big data means the handling of high volume of data in terms of specific data requirement and arrangement [1]. The major aim is the data arrangement task which can be handled in such a way to provide velocity, variety and high volume [2–5]. In today's world the amount of data generated is very high so by the use of big data framework, different computational aspects can be applied successfully [4–6]. The main capabilities of this environment are data processing with high volume, time saving and the better efficiency [8, 9].

Big data environment can be better understood by the following attributes [8–12]:

- Volume: The amount of data processed or generated comes under the volume category.
- Variety: It is the type of data under structured and unstructured category.
- Velocity: data or information processed or generated speed comes under the velocity category.

- Value: knowledge processing under the process of the above generation and processing in different levels.

Big data processing covers data, analytics and decision (Figure 1). Figure 2 shows the big data applications.

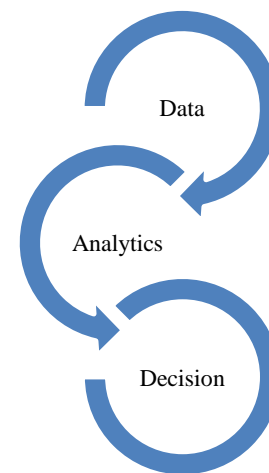


Figure 1 Big data processing

*Author for correspondence

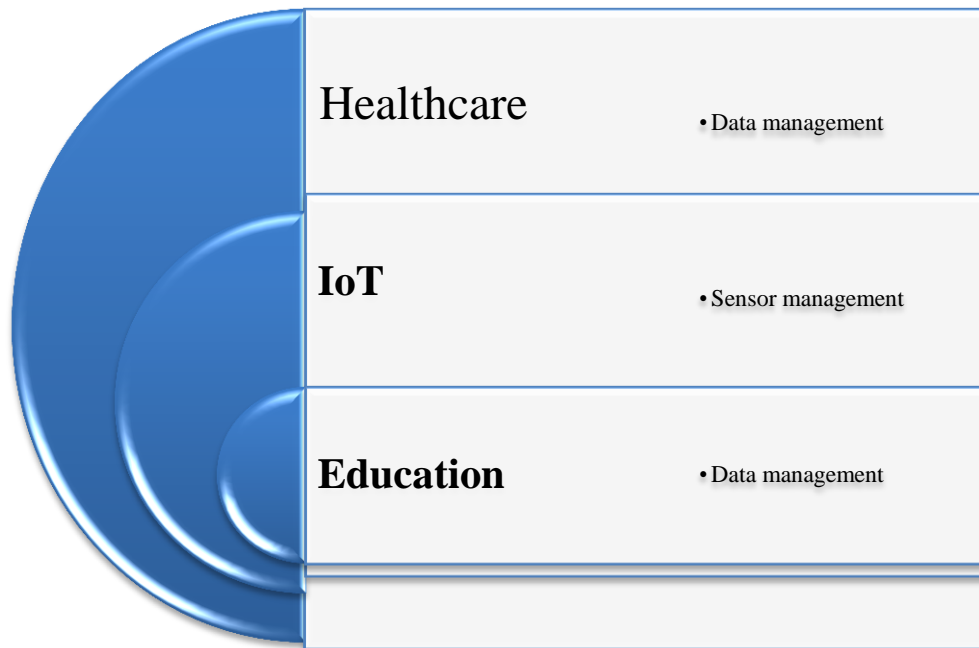


Figure 2 Big data applications

The main objective of this paper are as follows:

1. To explore the computational aspects applicability and the domain.
2. To explore the applicability areas and domains where the efficient use of this framework is possible.
3. To explore the ways to hybridize the decision-making scenario based on big data for the future framework.
4. To explore the applicability of computational methods like machine learning and data mining for the better data management.

2.Literature survey

In 2020, Li et al. [13] discussed about big data management in case of mineral resources. They have discussed about the system design process. It has been discussed in terms of resource management in terms of big data. The GIS technology has been discussed based on big data. It covers the resource management aspect along with the geological exploration. Their approach has improved the ability of production management.

In 2020, Li [14] discussed about the applicability of big data for the improvement of postgraduate management. Their approach is based on big data management. Their results show that the applicability of the approach in terms of management of postgraduates in case of colleges and universitie.

In 2020, Chaves et al. [15] discussed about data mining technologies in terms of exploration of large volume of data. They have discussed different aspects of big data along with the applicability in business analytics. They have also discussed regarding the subject which may be capable in achieving positive ideal for the used and the companies.

In 2020, Chen et al. [16] discussed about the big data in terms of strategic management. They have discussed a novel strategic method based on big data for the cultural industry. Their results support the approach.

In 2020, Chunlei et al. [17] discussed the big data in terms of workload evaluation. They have suggested the major drawback is the lack of relevant models in the same domain. They have proposed natural language processing with machine learning algorithms. They have obtained the reference value and then the workload evaluation has been performed. Their result is found to be efficient in the identification and the prediction of attribute class impact factors.

In 2020, Deng et al. [18] discussed about the enterprise management in case of big data. They have discussed the complete characteristics along with the

reengineering process. It is beneficial for the enterprise in case of big data.

In 2020, Du [19] discussed about the management of book materials. It has been discussed in terms of degree of information management. The management of materials and the book data has been analyzed in terms of big data visualization and knowledge discovery. It has been initiated in terms of data management, technology and electronic libraries.

In 2020, In [20] discussed about the improvement in terms of entrepreneurial practice ability. It has been discussed in terms of college students' Huizhou merchants. The environment considered here are big data. They have considered network sharing platform. They have discussed the cross regional exchange of the entrepreneurship.

In 2020, Li et al. [21] discussed about the complex macroeconomic environment. It has been discussed in terms of trust product allocation. The environment suggested is the big data. They have suggested that for financial innovation big data may found to be useful. They have extended their impact and suggested that China's trust industry needs to actively adopt big data technology.

In 2020, Ha and Back [22] discussed about the use of social network services (SNS). It has been discussed in terms of data collection, data processing and data analysis. They have suggested that it takes material resources as well as time. They have proposed data filtering algorithm. It has been used for the garbage data filtering purpose. Their filtering accuracy found to be improved due to recursive learning approach. They have achieved 70% accuracy.

In 2020, Kesheng et al. [23] discussed about the big data methods. It has been discussed for the behavior patterns understanding. This can be done in terms scientific quantitate analysis. They have considered data mining techniques. It has been used for the study and analysis of student campus network behavior analysis. They have used data support operation for the student affairs based on the scientific development.

In 2020, Neves and Cruvinel [24] discussed regarding the semantic model. It has been discussed for the digital database's structuring. The data has been originated from big data. They have examined different structure. These structures belong to structured, semi-structured and unstructured data. It

has been used for agricultural risk management. Their approach is beneficial for Data Mart and Data Warehouse (NoSQL). Their configuration covers agriculture data which involves multiple modes operating environment.

In 2020, Youzhuo et al. [25] discussed about the big data. It has been discussed in terms of big data and IT industry. They have suggested a retrieval system. It is based on Lucene architecture. It has been used based on the integrated architecture. It has been used in big data framework. It has been used in terms of distribution management system. Their proposed application has the capability of information retrieval in the fastest way and it is also convenient.

In 2020, Zhang [26] discussed about the big data prevention technology. They have used the combination of big data technology along with the financial market. It is useful in monitoring the financial market. They have suggested the combination of big data along with the financial market analysis.

In 2020, Zeng [27] suggested the application of big data. It has been used for military logistic support. It has been used for logistics support. It has been based on big data. It has been used in military logistics support.

In 2018, Pillmann et al. [28] discussed about the mobile Internet of Things for the sensor utilization and vehicle enabling. It has been addressed with the big data marketplace which has been leverage which exploits the crowds-sourced sensor data. It has been used for the harmonized data model. They have developed common vehicle information model. It has been developed based on generic representation for the whole data value and processing chain. It consists of sensor data though different components and data value process chain in case of big data environment.

In 2020, Yang [29] proposed a big data-based application for the financial field. It has been processed for the analysis the possible risks. Their approach is based on big data. It has been used for the application platform. They have suggested that their application is useful in the personal privacy data and promote the standardization of big data.

3.Discussion and comparative analysis

The previous literature suggest that big data framework has the capability to fin in different domain and area. It has been found that

computational methods can be used to enhance the processing performance and helpful in better data

management. *Table 1* depicts the approaches applied and applicability.

Table 1 Different approaches and its applicability

S. No	Reference	Approach used	Applicability	Results
1	[30]	Secure authentication and data sharing	They have proposed a novel security system architecture for the big data environment. It includes secure authentication and data sharing in cloud. It covers big data outsourcing, big data sharing and management.	They have covered user privacy and data security aspects in big data environment.
2	[31]	Internet of things and big data	Their main purpose is to check the applicability of Internet of Things and big data for the business and digital transformation. They have discussed the outbreak for the same combination in terms of mass of disorganized knowledge.	It shows the multidisciplinary aspects of Internet of Things and Big data.
3	[32]	Network security strategies	They have focused on network security and protection strategies of big data. They have highlighted different factors. These factors allow the affecting network for different security platforms. It has been performed in the big data framework. They have also discussed security and network issues.	They have discussed strategies for reliable platform for big data.
4	[33]	Big data for smart manufacturing	They have proposed a reference architecture for the industrial big data with a quantitative model It is based on decision making trial and evaluation laboratory.	They have provided a top-level planning for the reference architecture.
5	[34]	Big data frameworks for big service composition	They have proposed a big service composition model. They have used fuzzy relational concept analysis.	Their approach is found to be secure and efficient.

4.Problem identification

The following problems have been identified based on the literature discussed and analyzed.

1. There is the need of checking the applicability of big data in IoT, big medical services and data visualization aspect.
2. There is the need of a proper methodological analysis for the network capability and different pathways.
3. There is the need of computational hybridization along with the applicability check with different domains.
4. There is the need of proper data management using clustering and classification technique.
5. There is the need of discussing stream processing strategies and it applicability.

5.Conclusion

In this paper a meta- analysis has been presented for the big data environment. It covers the analytical and conceptual aspect along with the methodological applicability in different areas. It also covers the study and discussion of big data environment considering machine learning and data mining algorithms. Several previous literatures have been discussed and analyzed. Based on the discussion, the need of proper data management with the

applicability testing in different areas is the major finding.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Tsai CW, Lai CF, Chao HC, Vasilakos AV. Big data analytics: a survey. *Journal of Big data*. 2015; 2(1):1-32.
- [2] Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. *Journal of Business Research*. 2017; 70:263-86.
- [3] Hussin SK, Omar YM, Abdelmageid SM, Marie MI. Traditional machine learning and big data analytics in virtual screening: a comparative study. *International Journal of Advanced Computer Research*. 2020; 10(47):72-88.
- [4] Kaur S, Baghla S. Harnessing supremacy of big data using hadoop for healthy human survival making use of bioinformatics. *International Journal of Advanced Technology and Engineering Exploration*. 2018; 5(48):460-8.
- [5] Mtey MM, Dida MA. Towards interoperable e-Health system in Tanzania: analysis and evaluation of the current security trends and big data sharing dynamics.

- International Journal of Advanced Technology and Engineering Exploration. 2019; 6(59):225-40.
- [6] Bertot JC, Gorham U, Jaeger PT, Sarin LC, Choi H. Big data, open government and e-government: issues, policies and recommendations. *Information Polity*. 2014 ; 19(1, 2):5-16.
- [7] Pouyanfar S, Yang Y, Chen SC, Shyu ML, Iyengar SS. Multimedia big data analytics: a survey. *ACM Computing Surveys (CSUR)*. 2018; 51(1):1-34.
- [8] Dhas JJ, Vigila SM, Star CE. Forecasting of stock market by combining machine learning and big data analytics. In *international conference on soft computing systems 2018* (pp. 385-95). Springer, Singapore.
- [9] Omollo R, Alago S. Data modeling techniques used for big data in enterprise networks. *International Journal of Advanced Technology and Engineering Exploration*. 2020; 7(65):79-92.
- [10] Atat R, Liu L, Wu J, Li G, Ye C, Yang Y. Big data meet cyber-physical systems: a panoramic survey. *IEEE Access*. 2018; 6:73603-36.
- [11] Shobha K, Nickolas S. Time domain attribute based encryption for big data access control in cloud environment. *ACCENTS Transactions on Information Security*. 2017; 2(7):73-7.
- [12] Liebowitz J, editor. *Big data and business analytics*. CRC Press; 2013.
- [13] Li D, Gong Y, Tang G, Huang Q. Research and design of mineral resource management system based on big data and GIS technology. In *5th IEEE international conference on big data analytics (ICBDA) 2020* (pp. 52-6). IEEE.
- [14] Li J. Research on the management mode of graduate students in colleges and universities based on big data. In *2020 international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 617-20). IEEE.
- [15] Chaves A, Moura Í, Bernardino J, Pedrosa I. The privacy paradigm: an overview of privacy in business analytics and big data. In *15th iberian conference on information systems and technologies (CISTI) 2020* (pp. 1-6). IEEE.
- [16] Chen C, Zuo R, Ni H. Research on the application of big data in the field of strategic management of cultural industry. In *international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 795-8). IEEE.
- [17] Chunlei Z, Yin J, Qianli X. The workload assessment of national grid big data projects based on content recommendations and text classification. In *international conference on cloud computing and big data analytics (ICCCBDA) 2020* (pp. 482-90). IEEE.
- [18] Deng L, Ye S, Zhou X. Research on two-way integration business process reengineering under big data. In *information technology and mechatronics engineering conference (ITOEC) 2020* (pp. 1699-703). IEEE.
- [19] Du Q. Research on the application of big data in book management. In *international conference on measuring technology and mechatronics automation (ICMTMA) 2020* (pp. 773-5). IEEE.
- [20] In Gao X. Research on the entrepreneurship practice of college huizhou merchants based on big data. In *international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 604-7). IEEE.
- [21] Li D, Gong Y, Ren M, Li D. The research and design of trust business management and analysis system based on big data technology. In *5th IEEE international conference on big data analytics (ICBDA) 2020* (pp. 68-72). IEEE.
- [22] Ha IK, Back BH. Effective garbage data filtering algorithm for SNS big data processing by machine learning. In *international conference on artificial intelligence in information and communication (ICAIC) 2020* (pp. 520-4). IEEE.
- [23] Kesheng L, Yikun N, Zihan L, Bin D. Data mining and feature analysis of college students' campus network behavior. In *IEEE international conference on big data analytics (ICBDA) 2020* (pp. 231-7). IEEE.
- [24] Neves RA, Cruvinel PE. Model for semantic base structuring of digital data to support agricultural management. In *14th international conference on semantic computing (ICSC) 2020*(pp. 337-40). IEEE.
- [25] Youzhuo Z, Yu F, Ruifeng Z, Shuqing H, Yi W. Research on lucene based full-text query search service for smart distribution system. In *3rd international conference on artificial intelligence and big data (ICAIBD) 2020* (pp. 338-41). IEEE.
- [26] Zhang T. Research on big data's prevention technology of financial systemic risk. In *international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 663-6). IEEE.
- [27] Zeng Y. Analysis on the influence and countermeasures of big data in military logistics support. In *international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 648-51). IEEE.
- [28] Pillmann J, Sliwa B, Wietfeld C. The automat CVIM- a scalable data model for automotive big data marketplaces. In *19th IEEE international conference on mobile data management (MDM) 2018* (pp. 284-5). IEEE.
- [29] Yang R. Research on the risk and supervision method of big data application in financial field. In *international conference on intelligent transportation, big data & smart city (ICITBS) 2020* (pp. 695-8). IEEE.
- [30] Narayanan U, Paul V, Joseph S. A novel system architecture for secure authentication and data sharing in cloud enabled big data environment. *Journal of King Saud University-Computer and Information Sciences*. 2020.
- [31] Sestino A, Prete MI, Piper L, Guido G. Internet of things and big data as enablers for business digitalization strategies. *Technovation*. 2020.
- [32] El Alaoui I, Gahi Y. Network security strategies in big data context. *Procedia Computer Science*. 2020; 175:730-6.

- [33] Zhang X, Ming X, Yin D. Application of industrial big data for smart manufacturing in product service system based on system engineering using fuzzy DEMATEL. Journal of Cleaner Production. 2020.
- [34] Sellami M, Mezni H, Hacid MS. On the use of big data frameworks for big service composition. Journal of Network and Computer Applications. 2020.



Shweta Kumari received the B.Tech degree in Computer Science Engineering from DCE Darbhanga, Bihar in 2014. She is Currently pursuing M.Tech from SSSUTMS, Sehore, MP. Her area of research is Data Mining and Big Data.

Email: shweta57dce@gmail.com