

Big data-analysis, map reduced framework, security & privacy challenges and techniques in health sector

R. Sarkar^{1,*}, M. Telugu², N. Kuntla³

¹Assistant Professor, VIT-AP University, Amaravati, Andhra Pradesh, India

^{2,3}Undergraduate Student, VIT-AP University, Amaravati, Andhra Pradesh, India

Abstract

INTRODUCTION: Data is increasing exponentially. Data processing is an essential component in all industries, including health care. Even though a lot of progress has been made, it has been noted that in the recent decade, the health industry is capable of efficiently utilizing data and providing perfect Advancements in therapies.

OBJECTIVES: the main objectives include of finding the right problems in the security systems and to review the methods of present data processing methods.

METHODS: Methods involved are Quantitive analysis, Descriptive analysis, Data cleaning and Extraction.

RESULTS: The outputs of the reduce function are combined across all reducer nodes to produce the final output.

CONCLUSION: Big data analytics has enormous potential to accelerate the health care industry and that can only be done with some innovative methods and security plays a crucial role and can be a good catalyst in the user experience Elements.

Keywords: Big Data, Data Analytics, Map Reduced Framework, Privacy, Security

Received on 16 October 2023, accepted on 30 October 2023, published on 02 November 2023

Copyright ©2023 R. Sarkar *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetpht.9.4292

*Corresponding author Email: rajarshi.sarkar@vitap.ac.in

1. Introduction

As we are moving forward with technological advancements, Data is increasing exponentially. Data processing is an essential component in all industries, including health care. Even though a lot of progress has been made, it has been noted that in the recent decade, the health industry is capable of efficiently utilizing data and providing perfect Advancements in therapies. Our study describes how the health industry is improving its services through the significant application of data analytics. It covers the methodologies for using data analytics in the health business as well as the instrument utilized for analysis. Despite the latest technology tools, there are still numerous concerns and challenges in using data. This paper also discusses the difficulties encountered in the implementation of analytics, as well as the privacy and security problems that pose a

significant danger to data, and techniques for maintaining data security and privacy.

2. Methods

In the current era, there has been a rapid increase in viral and chronic diseases with an increase in population ¹. This is leading to the generating of various huge data sets from the health sector as Electronic Health Records and Data a key element in the healthcare industry since it enables healthcare professionals to make better decisions and enhance patient outcomes for ideal medical treatment.

Big Data Analytics now has an essential role in the health sector as it is helping to draw deeper insights from several health records through various methods and in providing conclusions drawn from the records for giving ideal treatment to patients with the same kind of diseases. It also can identify people who have a high chance of being effect by illnesses for a particular disease and symptoms based on previous

records, and this enables doctors to take quick action and prevent the occurrence in various ways. Figure 1 describes the applications of data analytics in the health sector in various ways.

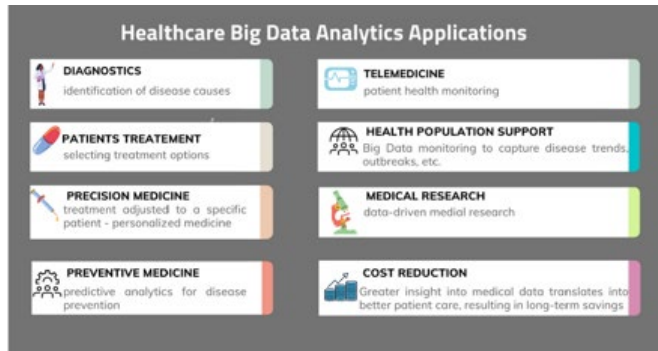


Figure 1. These are the health care Big-data applications.

2.1. Data Analysis on Health Records

Now, let us see how we apply phases of Data analytics in the health sector to provide conclusions^{2,3}.

2.1.1 Data Collection

Initially, diverse and huge data sets are collected from different health centers and stored in data warehouses or databases. Data is obtained from different sources and in different forms, it may include graphs, images, x-rays, and samples.

2.1.2 Data Cleaning, Extraction

We Extract the required and relevant records from the huge health records and eliminate inaccurate patient records. We segregate data into images, text, and graphs as it is essential to categorize data in a well-structured way for efficient analysis.

2.1.3 Data Analysis

Now we start analyzing the patient’s data by finding out meaningful insights from the organized health records by finding out the similarities in any symptoms, analyzing its effect on different age groups, the severity of the disease, etc. We use different types of data analysis^{4,5} which includes following:

Descriptive Analysis: Elaborate on the patient details, health conditions of patient and medical history using digital tools for a brief analysis.

Diagnostics Analysis: It draws the reasons for the occurrence of a particular disease based on other patient records and factors triggering the disease.

Predictive Analysis: It predicts the future threats to the patients and the further impact of disease in various age

groups. Mainly focus on predicting the future states by using conclusions from records.

Prescriptive Analysis: It provides a brief prescription including optimal treatment for patients, furthermore new inventions required to tackle the diseases.

3. Tools Used in Analysis of Health Records

Analyzing health records is very crucial and challenging as it contains sensitive data of the patients. So, we use very promising technologies for analyzing the data. There are several tools utilized in the health industry for efficient usage of data and data analysis, but the following few^{1,7,8} stand out.

3.1 Hadoop

Hadoop is an open-source big data processing technology that can store and manage massive amounts of both organized and unstructured data. It stores data in a parallel and distributed manner. In healthcare, it is often used to process and analyze claims data, clinical trial data, electronic health records (EHRs), and other types of healthcare data. It consists of many data nodes and one Master Node. All the records obtained from different sources are stored in Data nodes and the locations of each record are stored in Master Node. A key feature of this technique which makes it used more often is, Replicates of a record present in one data node will be available in all other data nodes to avoid data loss if there is any failure in that particular node.

3.2 MapReduce Framework

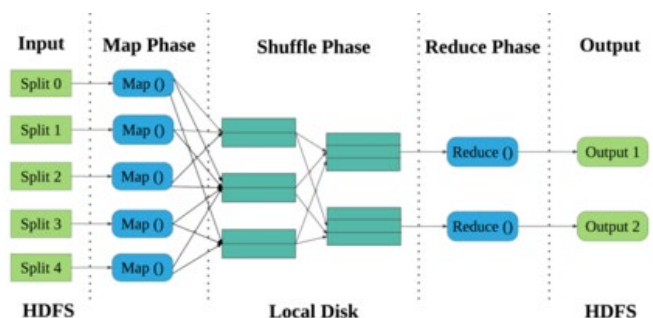


Figure 2: MapReduce Framework

MapReduce framework works by performing a sequence of tasks across several nodes in a distributed system. An overview of the MapReduce framework's functionality is provided below:

Data splitting: The input data is split up into smaller units. The processing of each block is subsequently allocated to a mapper node.

Map: A collection of intermediate key-value pairs is generated by the mapper nodes after they individually analyze the input data chunks. Each data point is processed by the map function simultaneously.

Shuffle: The intermediate key-value pairs generated by the mappers are shuffled by first grouping them according to their keys, then collecting the values connected to each key individually. The data is redistributed during this step, which is referred to as the shuffle phase, so that each reduction node may work with a small and easy data set.

Reducing: The reducer nodes compute the intermediate key-value pairs that were created during the shuffle phase. The corresponding values for each key are processed concurrently by the reduce function.

Output: The outputs of the reducer function are combined across all reducer nodes to produce the final output.

Conclusion

Big data analytics has enormous potential to speed up the healthcare business. As we progress towards the modern world powered by AI, the possibilities are endless. Proper infrastructure, as well as clean and ethical practices, permit a proper system that can yield benefits for society. With the availability of the most recent trends and Frameworks, we may utilize current developments and tackle current problems. Healthcare practitioners can make intelligent choices, which has a significant influence on the broader healthcare system.

However, the advancement comes at a cost. The innovations that assist us have negative side effects. Users' concerns about privacy and security remain unanswered. Even the most advanced approaches fall short of ensuring security and privacy. The cost element also plays a part, which opens the possibility of working on resources and developing betting systems that can balance the cost factor. Even though we are currently dealing with challenges, organizations, and the government are working tirelessly to resolve them and advance the healthcare system. Technology is undoubtedly a

Future Scope

There is a lot of scope to research and develop in the Sector of Health care. Every method or practice comes with some limitations. There are a lot of challenges to be addressed and these are the main challenges that create opportunities to work in the future.

Cost Reduction- The storage systems and Software are demanding a high level of financial investment. With the modern technological ways, we have scope to Redesign Nd modify them for better efficiency and with a better cost.

Prescriptive Analysis- Prescriptive analysis is not so strong enough to recommend and prescribe accurate output. The AI Algorithms and datasets need to be more efficient along with

the frameworks that all these are operating with. This creates a scope to make it better and to research more efficient ways to make the systems and analysis much more reliable.

References

- [1] Munusamy, Ambigavathi & Sridharan, D.. (2018). Big Data Analytics in Healthcare. 269-276. 10.1109/ICoAC44903.2018.8939061.
- [2] Wang, Y., Kung, L., Wang, W. Y. C., & Cegielski, C. G. (2018). An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, 55(1), 64-79.
- [3] Ambigavathi, M., & Sridharan, D. (2018, December). Big data analytics in healthcare. In 2018 tenth international conference on advanced computing (ICoAC) (pp. 269-276). IEEE.
- [4] Sangjukta Das, Suyel Namasudra, A Novel Hybrid Encryption Method to Secure Healthcare Data in IoT-enabled Healthcare Infrastructure, *Computers, and Electrical Engineering*, Volume 101,2022,107991,ISSN 0045-7906,https://doi.org/10.1016/j.compeleceng.2022.107991.(https://www.sciencedirect.com/science/article/pii/S0045790622002609).
- [5] Wang, Yichuan & Byrd, Terry. (2017). Business Analytics-Enabled Decision-Making Effectiveness through Knowledge Absorptive Capacity in Health Care. *Journal of Knowledge Management*. 21. 517-539. 10.1108/JKM-08-2015-0301.
- [6] Zeng, Xuezhi & Garg, Saurabh & Wen, Zhenyu & Strazdins, Peter & Zomaya, Albert & Ranjan, R.. (2017). Cost-Efficient Scheduling of MapReduce Applications on Public Clouds. *Journal of Computational Science*. 26. 10.1016/j.jocs.2017.07.017.
- [7] A.S. Thanuja Nishadi (2019); Healthcare Big Data Analysis using Hadoop MapReduce; *International Journal of Scientific and Research Publications (IJSRP)* 9(3) (ISSN: 2250-3153), DOI: http://dx.doi.org/10.29322/IJSRP.9.03.2019.p87104
- [8] Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform*. 2018 May 10;15(3):20170030. doi: 10.1515/jib-2017-0030. PMID: 29746254; PMCID: PMC6340124.
- [9] Batko, K., Ślęzak, A. The use of Big Data Analytics in healthcare. *J Big Data* 9, 3 (2022). https://doi.org/10.1186/s40537-021-00553-4
- [10] Dash, S., Shakyawar, S.K., Sharma, M. et al. Big data in healthcare: management, analysis, and prospects. *J Big Data* 6, 54 (2019). https://doi.org/10.1186/s40537-019-0217-0
- [11] L. Wang and R. Jones, "Big Data, Cybersecurity, and Challenges in Healthcare," 2019 SoutheastCon, Huntsville, AL, USA, 2019, pp. 1-6, doi: 10.1109/SoutheastCon42311.2019.9020632.
- [12] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2, 1-10.

- [13] Nambiar, R., Bhardwaj, R., Sethi, A., & Vargheese, R. (2013, October). A look at challenges and opportunities of big data analytics in healthcare. In 2013 IEEE international conference on Big Data (pp. 17-22). IEEE.
- [14] Sun, J., & Reddy, C. K. (2013, August). Big data analytics for healthcare. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1525-1525).
- [15] Archana, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, 50, 408-413.
- [16] Raj, P., Raman, A., Nagaraj, D., Duggirala, S., Raj, P., Raman, A., & Duggirala, S. (2015). Big data analytics for healthcare. *High-Performance Big-Data Analytics: Computing Systems and Approaches*, 391-424.
- [17] Manogaran, G., Lopez, D., Thota, C., Abbas, K. M., Pyne, S., & Sundarasekar, R. (2017). Big data analytics in healthcare Internet of Things. *Innovative healthcare systems for the 21st century*, 263-284.
- [18] Khan, Z. F., & Alotaibi, S. R. (2020). Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. *Journal of healthcare engineering*, 2020, 1-15.
- [19] Wong, Z. S., Zhou, J., & Zhang, Q. (2019). Artificial intelligence for infectious disease big data analytics. *Infection, disease & health*, 24(1), 44-48.
- [20] Abidi, S. S. R., & Abidi, S. R. (2019, July). Intelligent health data analytics: a convergence of artificial intelligence and big data. In *Healthcare management forum* (Vol. 32, No. 4, pp. 178-182). Sage CA: Los Angeles, CA: SAGE Publications.
- [21] Singh, R. K., Agrawal, S., Sahu, A., & Kazancoglu, Y. (2023). Strategic issues of big data analytics applications for managing health-care sector: a systematic literature review and future research agenda. *The TQM Journal*, 35(1), 262-291.
- [22] Khanna, D., Jindal, N., Singh, H., & Rana, P. S. (2023). Applications and Challenges in Healthcare Big Data: A Strategic Review. *Current Medical Imaging*, 19(1), 27-36.
- [23] Bag, S., Dhamija, P., Singh, R. K., Rahman, M. S., & Sreedharan, V. R. (2023). Big data analytics and artificial intelligence technologies based collaborative platform empowering absorptive capacity in health care supply chain: An empirical study. *Journal of Business Research*, 154, 113315.