Sustainable Computing: Informatics and Systems 35 (2022) 100719

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom

Smart teledentistry healthcare architecture for medical big data analysis using IoT-enabled environment

Muhammad Babar [a], Muhammad Usman Tariq [b], Mohammad Dahman Alshehri [c], Fasee Ullah [d,*], M. Irfan Uddin [e]

[a] Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan
[b] Abu Dhabi School of Management, Abu Dhabi, United Arab Emirates
[c] Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[d] Department of Computer Science and IT, Sarhad University of Science and Information Technology, Peshawar, Pakistan
[e] Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

ARTICLE INFO

Keywords:
Big Data analytics
Medical data
Smart health
Internet of Things

ABSTRACT

The current spread out in Big Data analytics and the medical Internet of Things (IoT) originated the recognition of smart health. Smart health is the integration of devices, sensors, cameras, and objects or things embedded with sensors that generate an enormous amount of data known as big data. IoT-enabled smart health management assimilates the digital traces generated by things and humans including teledentistry. The key purpose of teledentistry is to analyze the data and produce valuable insights and hidden patterns to offer services and overcome the existing challenges. This article proposes a data management design using Big Data analytics for smart teledentistry planning. The proposed scheme is the customization of the present parallel platforms to accomplish effective processing. The design offers three different modules which are the data pre-processing, data ingestion, and data processing module. The pre-processing is performed to speed up the real data processing. The recent schemes lack well-organized and effective parallel data loading and proficient communication. The Big Data storage and ingestion are achieved using optimized utility and parallel mechanisms. Optimized RDD-enabled (resilient distribution) Yet Another Resource Negotiator (YARN) based proposal is designed for resourceful cluster administration and computation of medical Big Data. The proposed data design is experimentally simulated with authentic datasets using Apache Spark 3.0 framework. The proposed architecture is compared with existing state-of-the-art proposals. The simulation results reveal the efficacy of the proposed system.

1. Introduction

Smart health is one of the modern conceptions in the modern age [1]. Smart health has fascinated consideration by providing an image of the all-inclusive set-up of physical objects' network [2]. Smart health involves communication between different IoT objects in an intelligent fashion. With the rise of connected devices, the idea of smart health and particularly teledentistry is also gaining prominence and has been broadly acknowledged [3]. The interactions among all these components create a new type of smart application and service. The internet allowed us to connect through these devices in multiple ways. There are main elements to organize the future Internet as Intelligent medical objects (devices) including extended Networks, and data analytics [4]. In health organizations, computerized health information systems have become a major issue of concern to all managers. In the practical lives of health workers, health information systems play a major role. Health information systems and technologies have moved distances far beyond human reach, allowing digital, text, audio, and image data to be stored [5].

Smart health aims to connect the physical organization, the IT necessities, and the social necessities to affect the conjoint intellect of the hospitals [5]. The creation of smart health is an approach to alleviate the problems which emerge from rapid urbanization and its population growth. Apart from cost association, once the smart hospitals are implemented, can deal with the ability to provide efficient services to the citizens. IoT is helping build smart health to improve the lives of

* Corresponding author.
E-mail addresses: muhammad.babar@aiou.edu.pk (M. Babar), m.tariq@adsm.ac.ae (M.U. Tariq), alshehri@tu.edu.sa (M.D. Alshehri), fasee.csit@suited.pk (F. Ullah), irfanuddin@kust.edu.pk (M.I. Uddin).

https://doi.org/10.1016/j.suscom.2022.100719
Received 21 September 2021; Received in revised form 24 December 2021; Accepted 5 March 2022
Available online 12 March 2022
2210-5379/

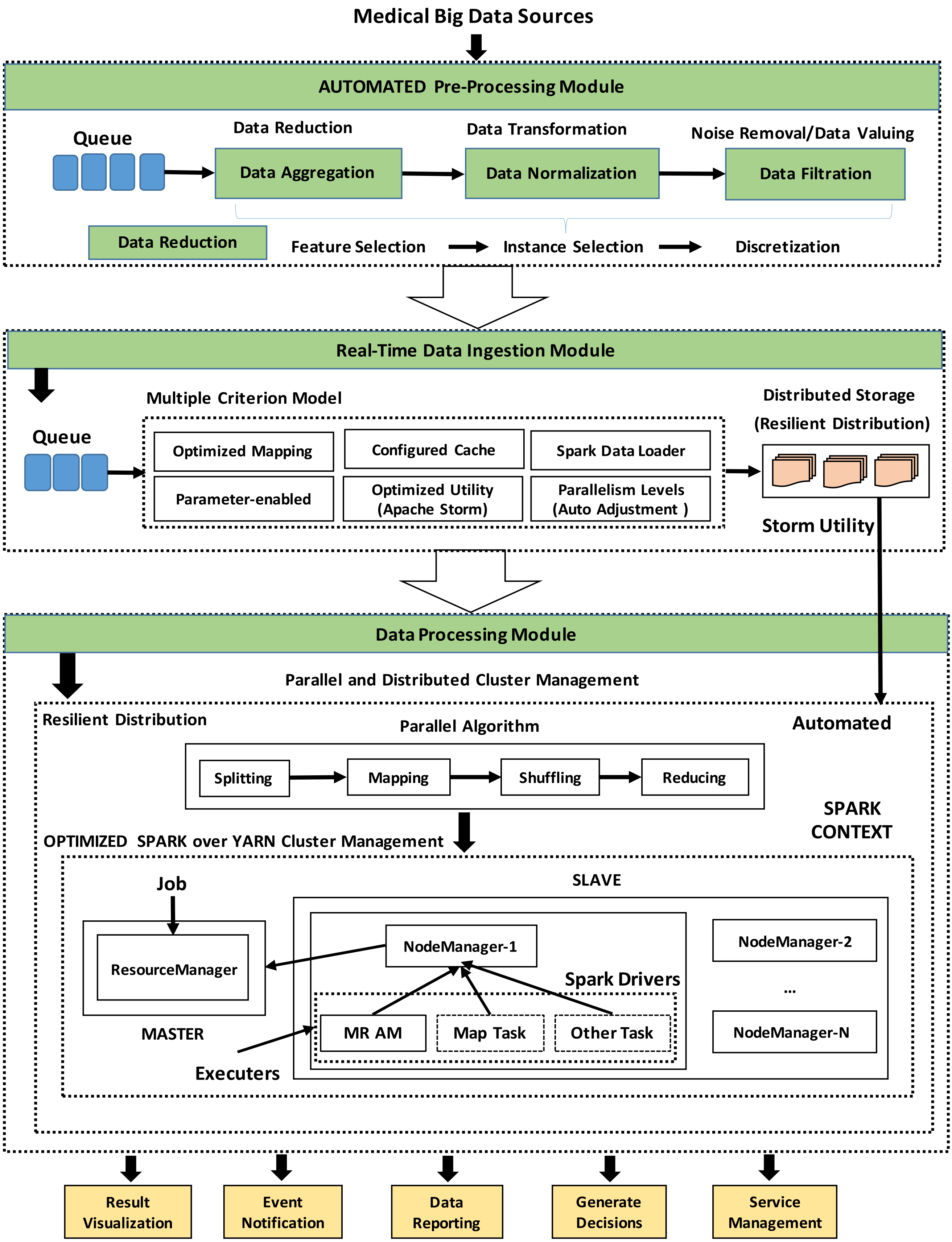

Fig. 1. Proposed system architecture.

people, making it more productive and functional [6]. The whole world is now talking about smart dental services. Smart health services and teledentistry change the life of citizens. Smart health has been advanced and it would be varied between user experience and the acceptance of the consumer by technology [7]. Health policy experts, clinicians, and healthcare institutions may use big data to make data-driven decisions that will improve patient treatment, disease management, and healthcare decisions. An accessible substructure is essential to manage the enormous incursion of devices.

Big Data analytics is integrated with healthcare to a truthful smart community [8]. The secure artificial intelligence of things is also preferred for big data analytics as privacy is the key issue in healthcare [9,10]. Apache Hadoop is used for parallel processing and distributed storage. Traditional approaches face bottlenecks to process the heterogeneous information. Besides, there are some research have been done to comprehend the belief of smart dental health that covers health management, controlling, and monitoring [11]. The diverse environment resolves the communication practice of the IoT up to large extent. It is due to the integration of the existing expertize and the mobility of the IoT nodes that can move in the networks in healthcare [12]. Therefore, any node can easily set up a communication and can access anything. Many healthcare ideas focus on big data and IoT when there are more vital problems. Smart health is about people, not technology. In general, big data systems are concerned with the provision of data for use in the enterprise, and today it has taken on a new organized level where it has been viewed as a platform for organizational resources.

This article proposes a data management design using Big Data analytics for smart teledentistry planning. Optimized RDD-enabled (resilient distribution) Yet Another Resource Negotiator (YARN) based proposal is designed for resourceful cluster administration and computation of medical Big Data. The proposed scheme is the customization of the present parallel platforms to accomplish effective processing. The Big Data storage and ingestion are achieved using optimized utility and parallel mechanisms. A specific dataset (dental health dataset) is used to comprehend the planned smart health scheme to enhance big data analytics.

2. Literature review

Cities population is rising hastily as people move to the urban area. This increase in population demands an efficient smart health environment. Smart health is one of the major stakes for the excellence of life of citizens. It is certain to enrich health services containing dental health services. Moreover, it helps improve people's quality of life. Economic success leads to city growth that, in turn, breeds more success. Recently, researchers and academia both have comprehensively considered healthcare and dentistry [13,14]. The IoT-enabled smart health generated huge and gigantic data that required to be processed [15]. Several devices are associated with everyday objects and workplace structures, which gather run-time information in a correct and incorrect format. Numerous societies use various systematic methods i.e. genetic algorithms, dynamic programming, distributed platforms, neural networks, and sentiment analysis. The imprint of connected objects via the internet is prolonged with the incredible growth of smart devices [16,17]. With the growth of hospitals and the population rising at an alarming pace, the potential of smart devices is also poised for incredible growth. The customary smart health community can have huge benefits to the industry and researchers.

Likewise, from a data point of view, smart city planning proposals are also given [18,19]. The investigative research relies on novel skills and today societies are entirely reliant on these skills [20]. It is important to attentively scrutinize the collected Big Data to propose an efficient smart scheme [21,22]. Big data refers to the exponential development and widespread availability of discrete, continuous, categorical, or hybrid data that is difficult to manage and analyze using traditional software tools and technology. Though offline (e.g., batch communication) processing can support to develop the smart health, it lacks real-time decision-making. Many techniques are developed to investigate big data based on the traditional computation approach (e.g., Hadoop) [23]. There is an architecture for medical record management in a secure environment [24]. Its designers describe Spark as a quick and general engine for processing massive amounts of data. It is quicker than prior techniques to working with Big Data [25], such as traditional MapReduce. Spark's speed is because it operates on memory (RAM), which allows it to process data considerably quicker than it would on disc.

To avoid health complications at the very early stages of infections, many researchers are exploring novel methods and strategies for using Big Data machine learning algorithms. Spark handles vast volumes of data very quickly since the parallel processing structure is used to manage data feeds from major data sources [26]. Apache Spark streaming from wearable devices that have sensor and system data to process medical wide data on runtime [27]. A web-based platform is proposed for healthcare big data analysis that was developed using data mining methods. The proposed architecture is divided into three layers including the Apache Spark Layer, the Workflow Layer, and the Web Service Layer. Regular Resilient Distributed Datasets (RDD) operations are supported by the Apache Spark Layer, which provides the fundamental Apache Spark functionality [28]. Meanwhile, this layer offers a caching mechanism to allow for the most efficient use of the findings that have already been computed. This layer contains a range of nodes for big data mining, each of which performs a distinct function, such as data source, algorithm model, or assessment tool.

Unfortunately, there is no specific model to efficiently utilize the parallel and distribute platforms. Likewise, a comprehensive method is necessary to process big medical data efficiently. The literature extracts noteworthy challenges that need to be addressed. To advance the processing and storage challenges of Big Data produced in the IoT-based smart health environment, this article proposes a specific scheme. The big data analytics-based solution chokes the encounters in IoT-enabled healthcare.

3. Proposed framework

The proposed scheme is the optimization of the current schemes to accomplish proficient processing. The design offers three different modules which are the data pre-processing module, data ingestion module, and data processing module depicted in Fig. 1. The pre-processing is performed to speed up the real data processing. The existing models lack effective data loading and effective communication. Normalization and filtration are performed to achieve pre-processing. The Big Data storage and ingestion are achieved using optimized utility and parallel mechanisms. Optimized RDD-enabled YARN-based proposal is designed for efficient cluster management and processing of Big Data.

3.1. Automated pre-processing module

The data gathered from the IoT medical devices in the healthcare domain may include many anomalies such as missing values, out-of-range output, noise, and so forth. This module is utilized to pre-process the data to remove the anomalies. Besides, the pre-processing process speeds up the processing mechanism as it saves the time being utilized on anomalies. Moreover, the pre-processing improves the processing too. The proposed pre-processing module is an improved automated process that performs the pre-processing automatically. The proposed pre-processing tool integrates with system architecture includes two major phases. The initial phase includes data aggregation, data normalization, and data filtration, while the latter includes the ease the convolution by feature selection. The normalization is performed using Eq. (1). The data aggregation is shown in Algorithm 1. The data splitting and divide-and-conquer mechanisms are utilized in the pre-processing module of the proposed system architecture. The data filtration is carried out using Kalman Filter. The filter is used to remove the noise from the available data. The proposed architecture utilizes the M/M/1 queuing model. Message Queue is a method that places a data item (message) on the peak of a queue which is not demanding an immediate reaction to take on the computation. The M/M/1 model is achieved using Eq. (2) where C is a chunk of data.

$$Y = \frac{X - Xmin}{Xmax - Xmin} * (n - m) + m \quad (1)$$

$$M = \frac{C}{1 - C} = S + \frac{C^2}{1 - C} \quad (2)$$

Furthermore, every chunk including m nodes can be presented as Eqs. (3) and (4), where the C is a chunk of the dataset and S is a set.

$$C = \sum_{k=1}^{M} Set_k \quad (3)$$

$$S = \sum_{l=1}^{X} MN_l \quad (4)$$

The data aggregation presents the assembled information in a précis form. It is beneficial for huge data that shortages the valued information. It is favored to acquire a grip of more data about groups.

Algorithm 1. Aggregation.

```
{
  IF  Dataset is equal to Required_Dataset
        Aggregate ()
  Else
      Fragmented Dataset (BLeft block, Bright block)
      BLeft Aggregation
      Bright Aggregation
      BLeft and BRight Collection
}
```

3.2. Real-time data ingestion module using multiple criterion model

Data ingestion is a significant stair to process and analyze the data, and then gain business values out of it. Besides, forcing the map-reduce to access data from external sources is error-prone, repetitive, and costly. The competent data ingestion is realized to decrease the ingestion time and increase the processing time. Realizing quality data ingestion considerably enrich the speed. In the proposed scheme, an improved integrated solution is proposed using the multiple criteria model to conquer the difficulty of additional time taken by loading large files to HDFS. It is an enhanced solution that improves the performance of data ingestion by avoiding copying data. Therefore, the customized configuration is preferred to be done before loading. Besides, the block size of HDFS and replication factor are also customized according to the proposed architecture. Moreover, the level of parallelism has also been configured accordingly. Therefore, the entire configuration regarding the in-memory cache is optimized to improve the efficiency of the data loading.

Data loading is a very significant part of the proposed system because unless we do not load data, the data cannot be administered. Apache Storm is integrated and optimized for data loading in real-times. It offers a program launch ability. This modification will occur and be applied to those files that are going to be loaded. This modification does not apply to the files already existing in the HDFS. Only the files copied after the change will have the new setting. The proposed architecture uses the customized block size and replication of the HDFS. The Hadoop components can be configured as Hadoop is an open-source framework. Therefore, the HDFS component of the Hadoop is configured concerning block size and replication. The parallelism is balanced by block size optimization too through splitting the data into a fixed block size. The traditional size of the blocks takes more time and offers less parallelism. The HDFS is optimized using the Hadoop documentation by configuring the configuration file called hdfs.site.xml. the optimization is performed using Eq. (5). The philosophy managing the IoT network is to assign and adjust the right values to the produced data units and the threshold value of the network shall be done along the negative gradient direction.

$$x_{k+1} = x_k - \eta_k g_k \quad (5)$$

Where the x_k denotes the weight values, g_k denotes the gradient of the current function, and the η_k denotes the rate.

3.3. Resilient distribution-enabled big data processing

The data processing module performs the actuals Big Data processing. Big data processing is carried out based on Resilient Distributed Datasets (RDDs). This multipurpose scheme permits controlling the perseverance and supervision of the partitioning of information related to big data, among other features. It includes parallel data processing, distributed storage, and sophisticated cluster management. The cluster management is realized using optimized YARN integrated with RDD. It is exploited for effective processing and cluster management distinctly. In RDD-enabled YARN, the different tasks belong to its corresponding application and the ApplicationMaster is running in a container controlled by specific NodeManagers. The NodeManager has several containers that are formed dynamically. All the elements of the YARN (e.g., NodeManagers, ResourceManager, containers, and ApplicationMaster) cooperate with one other as shown in Fig. 2. The application is submitted using the Hadoop jar command in CLI or using Java IDE to RM in a similar way as in traditional MR.

Several limitations are considered while taking this verdict such as the capacity algorithm. The RM employs a particular scheduler whose focal point is only on scheduling activities; it compacts when resources of the cluster are accessed and who access it. When a new application submission is recognized by RM, initially the scheduler chooses a container to launch ApplicationMaster. The Application Master will be in charge for the entire life cycle of the application where it starts.

4. Implementation and results

The proposed design is experimentally simulated with an authentic dataset using Hadoop 3.0 and Apache Spark 3.0 frameworks. The traditional Spark is customized using an optimized YARN. The Apache

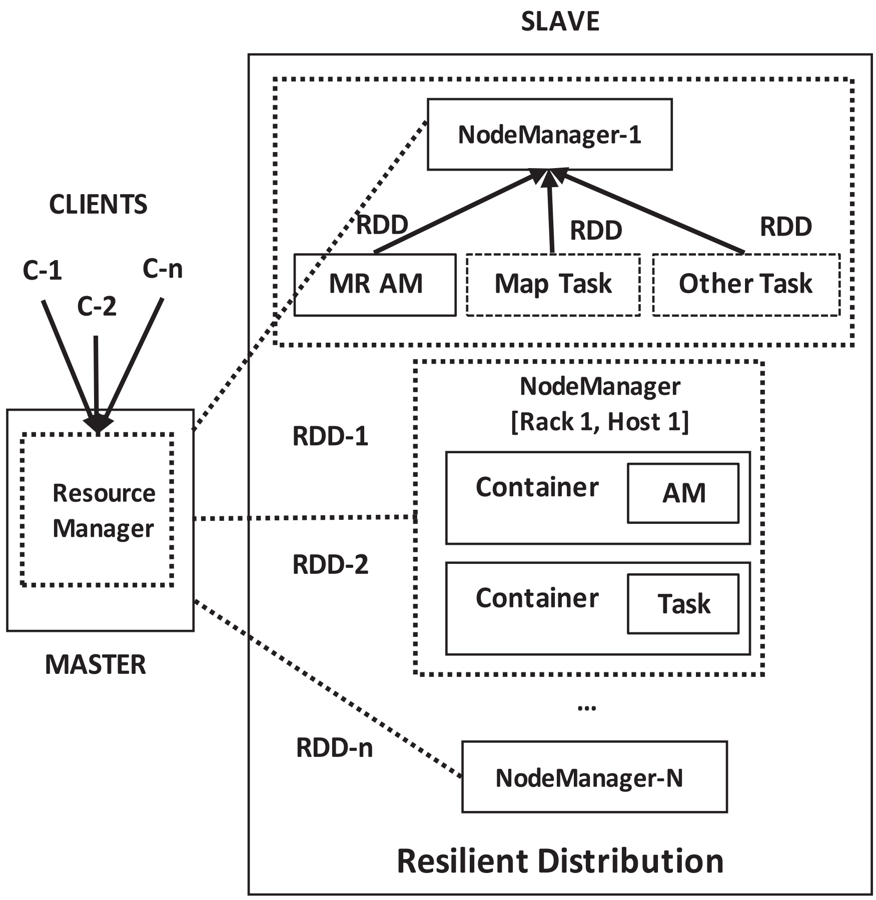

Fig. 2. Optimized YARN (RDD-enabled).

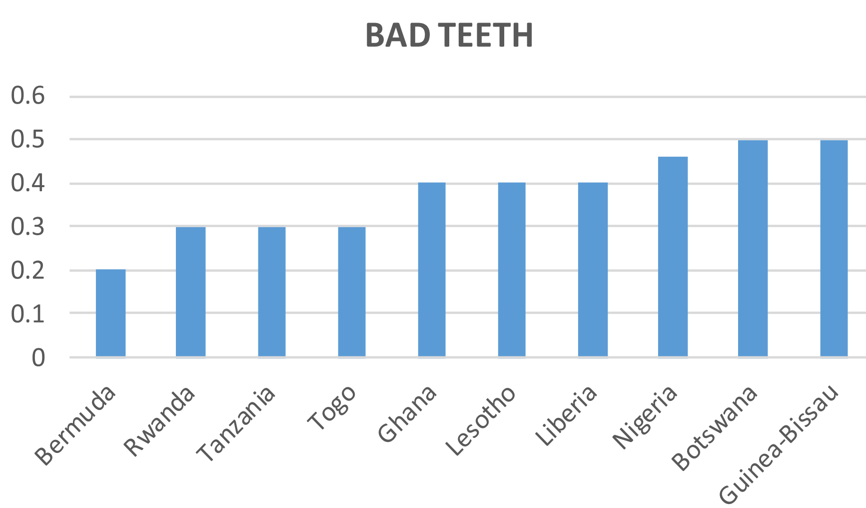

Fig. 3. Countries with lowest number of bad teeth.

Strom provided by Spark is used for stream data ingestion. The data loading outcomes are derived using the classical and proposed approaches both and the results are compared default setting. The execution time and the throughput of the overall architecture are compared with existing works. The experiments outcomes and comparison reveal the efficacy of the proposed scheme. In the context of import syntax, the Hadoop general arguments (parameters) ought to precede some import parameters. The connect string is used to illustrate the connection to the database to import data from a table of databases into HDFS. The connect string is used to communicate with Sqoop with the –connect argument and it is like a URL. The –password or -P and –username arguments are used to provide authentication. Connect strings beginning with jdbc:mysql:// are automatically handled in Sqoop. Similarly, the –driver arguments are used to specify the required JDBC driver. After proper connection, authentication, and driver specification, the required table is chosen which data is required to be loaded to HDFS. The data is usually imported in a table-centric fashion.

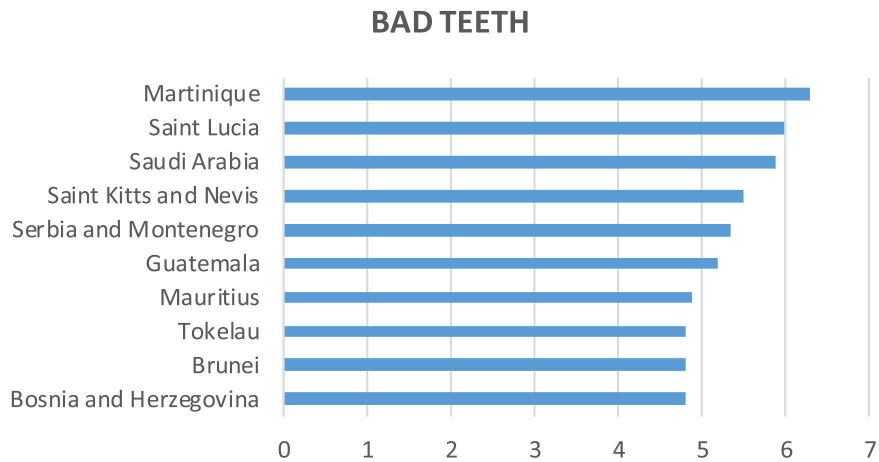

Fig. 4. Countries with highest number of bad teeth.

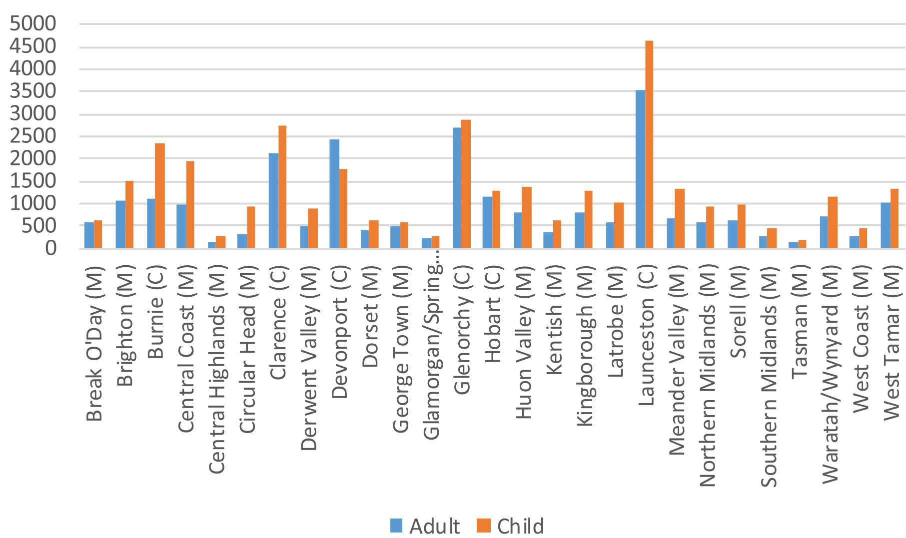

Fig. 5. Customers per LGA.

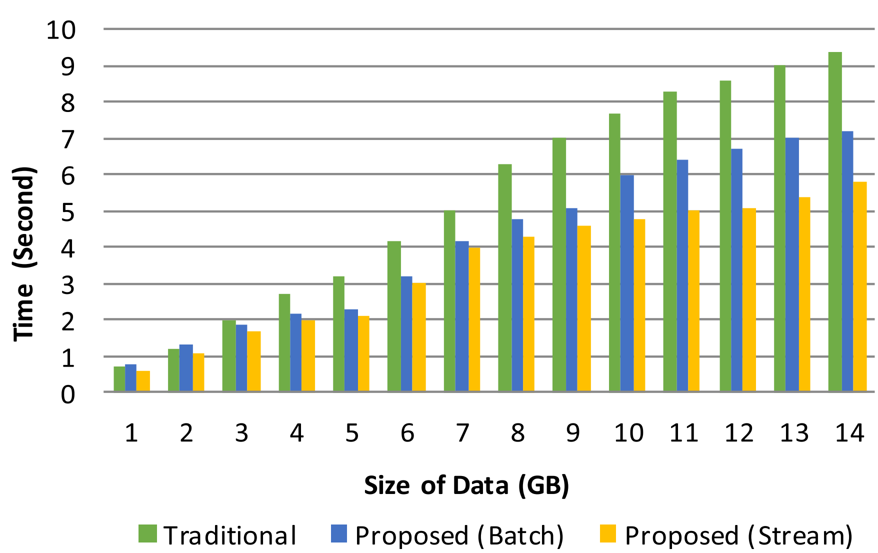

Fig. 6. Overall data loading efficiency.

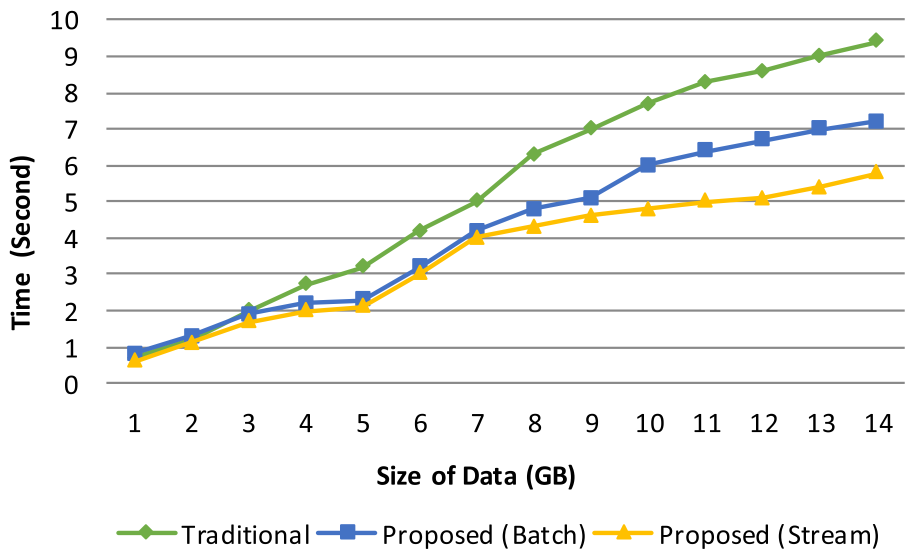

Fig. 7. Dataset threshold.

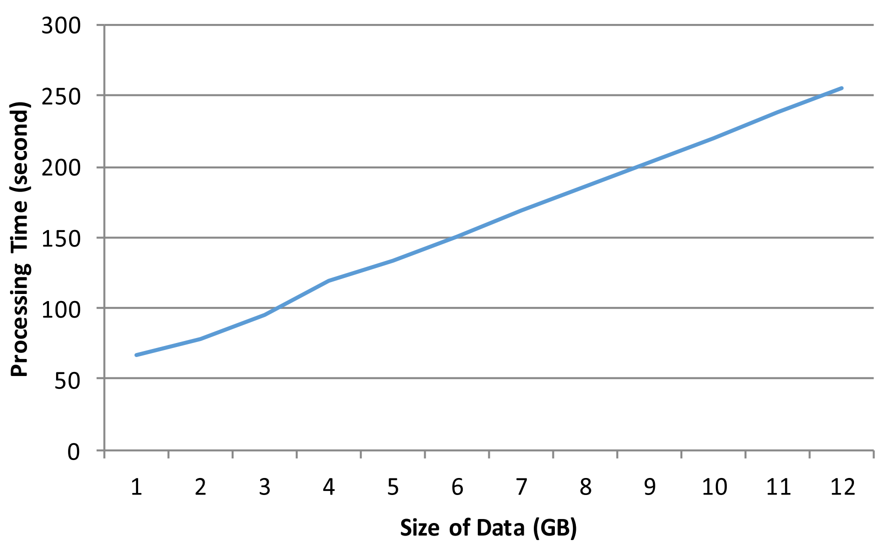

Fig. 8. Processing time of proposed scheme.

4.1. Data analysis and visualization

The data analysis of dental data is provided in this section. The dataset includes the number of bad teeth of the children provided by WHO [29]. The 10 countries having the best dental performance in the work are depicted in Fig. 3. Besides, the 10 countries having more bad teeth of children are also depicted in Fig. 4. The average of bad teeth per child in the world is 2.1. The dataset of the health department of Tasmania is also taken into consideration. The data includes the counts of customers per local government area (LGA) that were treated. The data information is provided in Fig. 5.

4.2. Real-time data ingestion efficiency

The data loading module is equipped with the real-time streaming utility using the multi-criteria model. It is observed that it takes no time when the loaded data is up to 2.5–3 GB. The results do not reveal the significant difference of data ingestion either loaded automatically or manually. It is proved that the proposed offline data ingestion module improves the data loading efficacy than the traditional default setting. In addition, the stream-enabled data ingestion produces better results than the offline data loading format. The data loading efficiency of the proposed scheme is depicted in Fig. 6.

Fig. 7 validates the threshold value for data loading using the proposed RDD-enabled data ingestion with the integration of a multiple criterion model. The threshold is the point where the data loading efficiency of the proposed system is increased. The threshold value is 2 GB that is the considered for using the proposed architecture.

4.3. Processing time efficiency

Fig. 8 reveals the processing (execution) time of jobs using the proposed solution. The processing time is given about the small, medium, and large sets of data. It is perceived that the execution time efficiency increases when the size of data is enlarged. It is due to the efficiency of data loading and enhancement. The processing time of the proposed optimized RDD-enabled YARN solution is also highlighted in Fig. 9 regarding different categories of the dataset (e.g. small, medium, large). The small dataset size is up to 2.5 GB, the medium dataset size is 6–10 GB, and the large dataset size is around 16 GB.

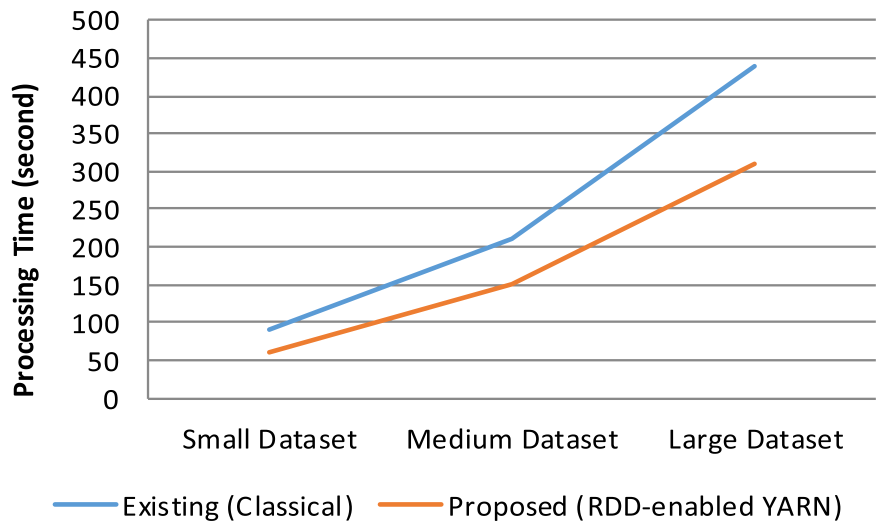

Fig. 9. Processing time comparison with classical approach.

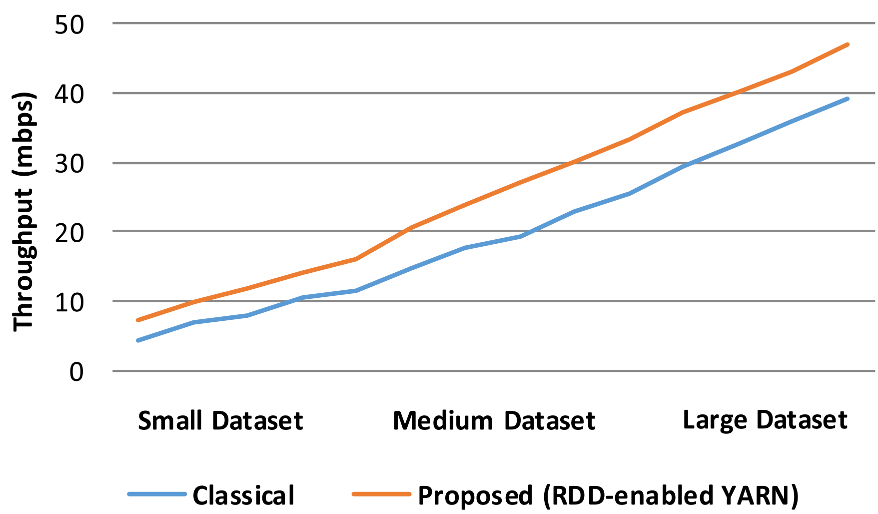

Fig. 10. Throughput comparison with classical approach.

4.4. System evaluation about throughput

Moreover, the throughput efficiency of the proposed architecture is also measured. The throughput of the proposed customized YARN-based solution is also highlighted in Fig. 10 along with the traditional approach. The throughput of the YARN-based proposed architecture is higher than the traditional solution.

5. Conclusion

This article proposes a data management design using Big Data analytics for smart dentistry planning. The proposed scheme offers three different modules which are the data pre-processing module, data ingestion module, and data processing module. The pre-processing is performed to speed up the real data processing. The recent schemes lack well-organized and effective parallel data loading and proficient communication. Normalization and filtration are performed to achieve pre-processing. The Big Data storage and ingestion are achieved using optimized utility and parallel mechanisms. Optimized RDD-enabled YARN proposal is designed for effective cluster management with RDDs. The proposed data design is experimentally simulated with an authentic dataset using Spark 3.0 framework. The comparison of the proposed architecture is provided with a traditional approach. The results reveal the efficiency of the proposed scheme.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Taif University Researchers Supporting Project number (TURSP-2020/126), Taif University, Taif, Saudi Arabia.

References

[1] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, T. Sato, Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health, IEEE Wirel. Commun. 28 (3) (2021) 54–61, https://doi.org/10.1109/MWC.001.2000374.

[2] Muhammad Babar, Ataur Rahman, Fahim Arif, Gwanggil Jeon, Energy-harvesting based on internet of things and big data analytics for smart health monitoring, Sustain. Comput. Inform. Syst. 20 (2018) 155–164.

[3] J. Virdee, R. Sharma, S. Ponduri, Spotlight on teledentistry, Br. Dent. J. 228 (11) (2020), 815-815.

[4] Rahat Iqbal, Faiyaz Doctor, Brian More, Shahid Mahmud, Usman Yousuf, Big data analytics: computational intelligence techniques and application areas, Technol. Forecast. Soc. Chang. 153 (2020), 119253.

[5] Md. Milon Islam, Ashikur Rahaman, Md. Rashedul Islam, Development of smart healthcare monitoring system in IoT environment, SN Comput. Sci. 1 (2020) 1–11.

[6] Smita Ambarkar, Sanjay, Narendra Shekokar, Toward smart and secure IoT based healthcare system. Internet of Things, Smart Computing and Technology: A Roadmap Ahead, Springer, Cham, 2020, pp. 283–303.

[7] Hugo Bergier, Loïc Duron, Christelle Sordet, Lou Kawka, Aurélien Schlencker, François Chasset, Laurent Arnaud, Digital health, big data and smart technologies for the care of patients with systemic autoimmune diseases: where do we stand? Autoimmun. Rev. (2021), 102864.

[8] Peng-Ting Chen, Chia-Li Lin, Wan-Ning Wu, Big data management in healthcare: adoption challenges and implications, Int. J. Inf. Manag. 53 (2020), 102078.

[9] K. Yu, Z. Guo, Y. Shen, W. Wang, J.C. Lin, T. Sato, Secure artificial intelligence of things for implicit group recommendations, IEEE Internet Things J. (2021), https://doi.org/10.1109/JIOT.2021.3079574.

[10] K. Yu, L. Tan, S. Mumtaz, S. Al-Rubaye, A. Al-Dulaimi, A.K. Bashir, F.A. Khan, Securing critical infrastructures: deep learning-based threat detection in the IIoT, IEEE Commun. Mag. (2021), https://doi.org/10.1109/MCOM.101.2001126.

[11] Delin Zeng, Yenni Tim, Jiaxin Yu, Wenyuan Liu, Actualizing big data analytics for smart cities: a cascading affordance study, Int. J. Inf. Manag. 54 (2020), 102156.

[12] S. Sonune, D. Kalbande, A. Yeole, S. Oak, Issues in IoT healthcare platforms: a critical study and review, in: 2017 International Conference on Intelligent Computing and Control (I2C2), IEEE, 2017, pp. 1–5.

[13] A. Sharma, G. Singh, S. Rehman, A review of big data challenges and preserving privacy in big data. Advances in Data and Information Sciences, Springer, Singapore, 2020, pp. 57–65.

[14] S. Boubiche, D.E. Boubiche, A. Bilami, H. Toral-Cruz, Big data challenges and data aggregation strategies in wireless sensor networks, IEEE Access 6 (2018) 20558–20571.

[15] S.E. Bibri, The IoT for smart sustainable cities of the future: an analytical framework for sensor-based big data applications for environmental sustainability, Sustain. Cities Soc. 38 (2018) 230–253.

[16] M. Ge, H. Bangui, B. Buhnova, Big data for internet of things: a survey, Future Gener. Comput. Syst. 87 (2018) 601–614.

[17] Yan Su, Xianping Wang, Innovation of agricultural economic management in the process of constructing smart agriculture by big data, Sustain. Comput. Inform. Syst. (2021), 100579.

[18] M. MazharRathore, A. Paul, A. Ahmad, M. Anisetti, G. Jeon, Hadoop-based intelligent care system (HICS): analytical approach for big data in IoT. ACM Transactions on Internet Technology (TOIT), December 2017. ISSN 1533-5399 (print), 1557-6051 (electronic).

[19] H. Sun, A.J. Song, Jara, R. Bie, Internet of Things and big data analytics for smart and connected communities, IEEE Access 4 (2016) 766–773, https://doi.org/10.1109/ACCESS.2016.2529723.

[20] Rudyar Cortés, Xavier Bonnaire, Olivier Marin, Pierre Sens, Stream processing of healthcare sensor data: studying user traces to identify challenges from a big data perspective, Procedia Comput. Sci. 52 (2015) 1004–1009.

[21] M.M. Rathore, A. Paul, A. Ahmad, G. Jeon, IoT-based big data: from smart city towards next generation super city planning, Int. J. Semant. Web Inf. Syst. 13 (1) (2017) 28–47.

[22] Md. Ileas Pramanik, Raymond Y.K. Lau, Haluk Demirkan, Md. Abul Kalam Azad, Smart health: big data enabled health paradigm within smart cities, Expert Syst. Appl. 87 (2017) 370–383.

[23] M. Rathore, A. Ahmad, A. Paul, G. Jeon, Efficient graph-oriented smart transportation using internet of things generated big data, in: 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, 2015, pp. 512–519.

[24] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, H. Lu, Towards secure and privacy-preserving data sharing for COVID-19 medical records: a blockchain-empowered approach, IEEE Transactions on Network Science and Engineering, doi: 10.1109/TNSE.2021.3101842.

[25] Lekha R. Nair, D.Shetty Sujala, D.Shetty Siddhanth, Applying spark based machine learning model on streaming big data for health status prediction, Comput. Electr. Eng. 65 (2018) 393–399.

[26] Dong-Her Shih, Thi Hien To, Ly. Sy. Phu Nguyen, Ting-Wei Wu, Wen-Ting You, Design of a spark big data framework for PM2. 5 air pollution forecasting, Int. J. Environ. Res. Public Health 18 (13) (2021) 7087.

[27] Lekha R. Nair, D.Shetty Sujala, D.Shetty Siddhanth, Applying spark based machine learning model on streaming big data for health status prediction, Comput. Electr. Eng. 65 (2018) 393–399.

[28] Md Rezaul Karim, R. Sahay, D. Rebholz-Schuhmann, A scalable, secure and realtime healthcare analytics framework with Apache Spark, in: Proceedings of the 2nd INSIGHT Student Conference on Data Analytics, The Insight Centre for Data Analytics, 2015, pp. 83–83.

[29] ⟨https://www.kaggle.com/angelmm/healthteethsugar⟩.