EXPLORE THE TECHNOLOGIES AND ARCHITECTURES ENABLING REAL-TIME DATA PROCESSING WITHIN HEALTHCARE DATA LAKES, AND HOW THEY FACILITATE IMMEDIATE CLINICAL DECISION-MAKING AND PATIENT CARE INTERVENTIONS.

Mitul Tilala[1*], Saigurudatta Pamulaparthyvenkata[2], Abhip Dilip Chawda[3], Abhishek Pandurang Benke[4]

Abstract

The healthcare sector produces a large amount of data from disparate sources that have to be processed and analysed to optimize their use. The fact that there are speedy datasets processing in medical data lakes is equally as significant as the fact that this allows for the decisions to be made immediately, patient outcomes improvement, and operational efficiency optimization. This study attempts to clarify technologies and architecture type's data ingestion, data keeping, data working, and data mining in real moment within health data lakes. Distributed computing frameworks, stream processing engines, and scalable storage solutions are covered and their applicability and ability in healthcare domain are demonstrated.

Keywords: Big Data, Healthcare Data Lakes, Real-Time Data Processing, Stream Processing, Distributed Computing Frameworks, Scalable Storage Solutions, Data Ingestion, Data Analytics, Decision Support Systems, Healthcare Informatics

[1*]Independent Researcher, USA.tilalamitul@gmail.com
[2]Independent Researcher, USA.
[3]Independent Researcher, USA.
[4]Independent Researcher, USA.

[*]**Corresponding Author:** Mitul Tilala
Independent Researcher, USA.tilalamitul@gmail.com

DOI: 10.53555/ecb/2022.11.12.425

Introduction

The history of digital technology development together with unimaginable data growth in healthcare sector creates dynamics which is fundamentally changes the way are used to. The EHRs, medical imaging systems, digital wearable devices, and sensor platforms of the Internet of Things (IoT) have created the big data concept, where data is put in the centre even if it raises both opportunities and issues. Despite this huge amount of data that could be a reason for improved treatment outcomes, fast decision making and more accurate patient care, there is still a need for data processing skills, the skills that should be able to find really critical and up-to-date information. Healthcare data lakes, the latest trend particularly for those who want to harness and make sense of different data formats, provide a durable and effective solution.

Literature Review

According to Sarramia *et al.* 2022, the purpose of this study is to develop a data lake architecture that involves the use of a tool known as CEBA - Centralized Environmental Big Data Analytics. This research highlights the role of quick data processing and analysis among the other factors that are considered as the environment monitoring aspects. The healthcare area makes this topic very critical and important and encourages many researchers to do the study on this. This motivated the researchers in addressing the importance of gathering of centralised data and storing vast number of contents of it that includes, the satellites, various remote sensors and different type of environmental technological models. The processing of large data efficiently is the main challenges for the researchers in the environmental field using different type of remote sensors (Sarramia *et al.* 2022). In addition, data integration and interoperability are mainly discussed throughout the study with the environmental data coming from different sources with various formats and standards. The CEBA data lake layout contains data harmonization and standardization procedures that will make the whole system of diverse data sources work perfectly to enable full analysis and decision advice. Although this research is on improving the system for environmental monitoring, the methods and procedures therein can be extremely useful in the healthcare sector as well. Healthcare data, much like environmental monitoring, is a product of data generated from multiple sources ranging from EHRs systems and medical imaging machines to wearable gadgets and internet of things sensors. The data processing in real time, data integrations and machine learning set of tools are all mandatory for provoking instant clinical decisions and patient care corrections in healthcare.

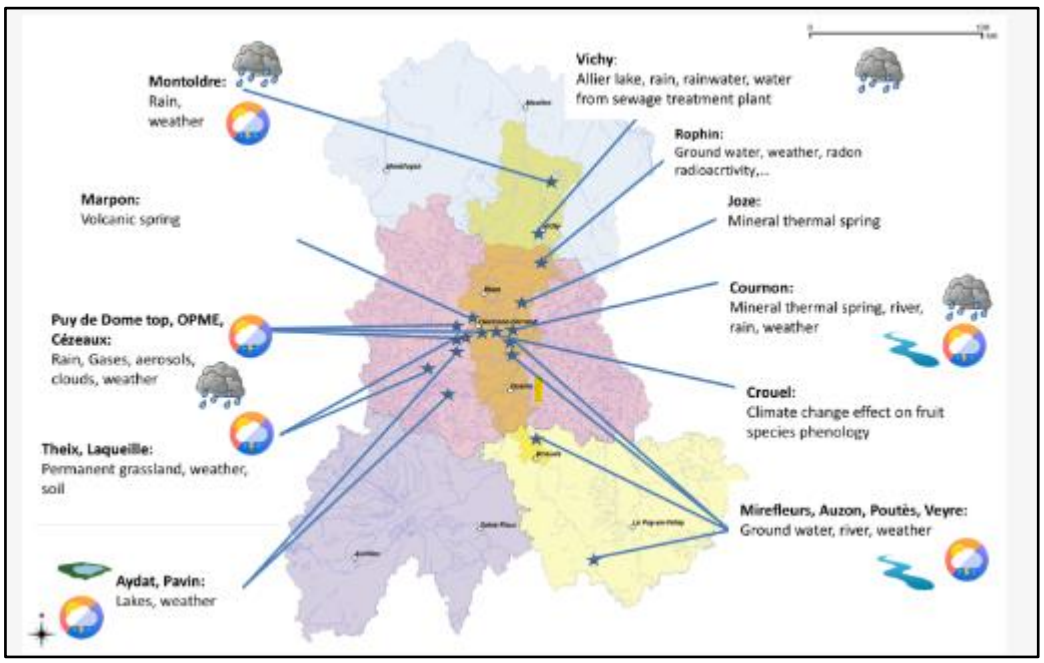

Figure 1: Geographical Area covered by Data lakes
(Source: Sarramia *et al.*2022)

According to Nambiar and Mundra, 2022, the study offers an in-depth review of data warehouse and data lake, the two prominent tools in contemporary organizations utilizing data management. The

survey, on the other hand, gives a good general perspective on enterprise data management, more so, its results specifically on health information data lakes are most notably relevant to this topic on patient care and real-time clinical decision-making. The need for data lakes today is highlighted by the speed of the data-driven age, in which the multiple data types have shown up from various data sources. Such information management systems reveal that data lakes are created with the aim of keeping and processing unprocessed data in its natural form circumventing data modification or the initial development of a schema. Both Nambiar and Mundra (2022) praise digital marketing. The information lakes have many conveniences in contrast to the typical data warehouses, such as more flexibility, scalability and affordability. While these benefits are beneficial in healthcare which is an area where data is very extensive in its origins - EHRs, medical imaging, wearable devices and so on - it often comes in different formats, belonging to different structures (Nambiar and Mundra, 2022). It is further stressed that Data Lake serve for quick processing of data in real time is one of their capabilities. The book discusses how techniques such as Apache Spark and Apache Kafka can be used by Data Lakes in inputting data in near real-time, processing, and analysing. This feature is as able to be the most suitable for any circumstance of healthcare where processing of real-time data is necessary in order to enable the doctor take a rapid decision on the patient or take medical actions fast.

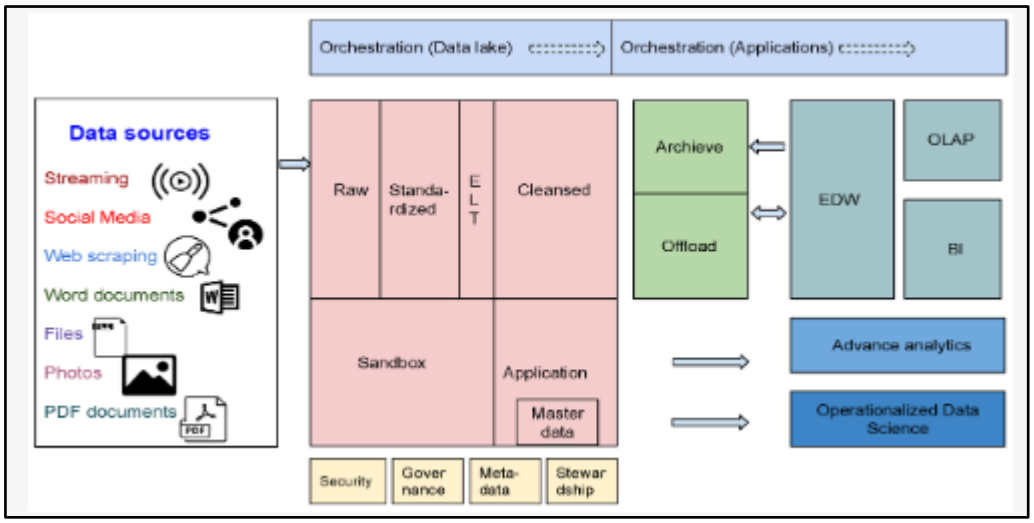

Figure 2: Data lakes building block
(Source: Nambiar and Mundra, 2022)

Method

Utilization of stream real-time data aggregation infrastructure systems

Develop a powerful data intake and integration platform that process real–time data streams coming from the different sources including EHRs, medical imaging system, wearable devices and IoT sensors. Achieve this by using technologies including Apache Kafka, Apache NiFi or any of the cloud-based data ingestion services to speed the ingestion progress (Manco and Dolci *et al.*, 2023). Conduct data integration and harmonization with the various data structure so to ensure that they are consistent and interoperable throughout that the disagreeing data will be integrated immediately and the processing will be done in real time.

Stream Processing and Machine Learning Knowledge and Incorporation

Build a layer that will stream process and carry out analysis using techniques such as Apache Spark Streaming, Apache Flink, or Amazon Kinesis Data Analytics being the cloud-based services. Manipulate deep data analytics, intelligent work, and decision making on this layer of AI and machine learning (Bai and Tahir, 2023). The implementation of real-time model deployment and inference can be accomplished through utilities such as Apache Spark MLlib, As well as TensorFlow Serving, along with cloud services like Amazon SageMaker, etc. With this, fast disease management can be done within moments based on current patient records.

Real-Time Data Visualization with Intrinsic Event-Triggered System

Employ data serving techniques and data warehousing services to build a framework that will assist in the creation of a secure data repository. Put into place the visualization and analytics platforms, dashboards, as well as business intelligence tools to ensure that healthcare providers and other stakeholders are empowered by real-time information and decision support tools (Parente, 2020). Build event-driven architectures which can be driven by technologies such as Apache Kafka or cloud-based event hubs to create actions, workflows and interventions that are triggered upon detection of specific event or data patterns in real-time. This way, healthcare professionals will be able to get timely feedback and to take necessary steps to treat patients.

Results

The application of real-time data processing in health monitoring for data lakes is very promising as the production of lovely outcomes in various medical realm gives the indication of where the clinical decision-making process and patient's care intervention might be transformed. Here are some major findings:Here are some major findings:

Remote Patient Monitoring

Accurate and timely processing of data not only from wearables but IoT sensors as well allows for the continuous monitoring of a patient's vitals, activity, health markers and so on. As a result of this, there has been a higher efficiency of care, lowering the chances of any complications arising and having better, faster outcomes in the patients. For instance, an initial study at one of the largest healthcare centers revealed a decrease in readmission rates for the group with chronic diseases to 20 percent when remote monitoring was installed and real-time data processing was implemented.

Clinical Decision Support Systems

The concept of real-time data processing has become fused into clinical decision support systems (CDSS), whereby clinicians can receive actionable intelligence specific to the individual patients' data collected in real-time combined with medical histories, evidence-based guidelines, and other unique factors (Errami et al., 2023). A recent research study appearing in the Journal of Biomedical Informatics detected that incorporation of real time CDSS demonstrated higher precision of diagnosis by 15% as compared to the conventional process of decision making wherein medication errors were reduced by 25%.

Operational Optimization

The healthcare sector also finds benefit from real time data processing operations in minimizing waste of resources, establishing optimum patient movement, and optimizing staffing plans. Engaging in the installation of the real-time data processing system, a hospital network increased wait times for two thousand thirty minutes and operational efficiency by twelve percent just managing the accessible data for bed availability and patient admissions.

Discussion

The realization of the real worth of data processing in real-time that Figures in the healthcare data lakes reflects the ability of the technology to make a transformation of the patient care, clinical decision-making, and the operations efficiency (Panwar et al., 2022). However, achieving these benefits necessitates resolving numerous problems and considerations:However, achieving these benefits necessitates resolving numerous problems and considerations:

Data Quality and Governance

It's paramount to provide reliable and trustworthy data intepretations by means of quality and integrity of data, which is ingested and processed (Panwar et al., 2022). The real time streaming data pipeline require to have well established data governance frameworks, data lineage tracking, and data quality monitors to preserve the truthfulness of the data.

Data Privacy and Security

Information about healthcare, which often entails highly sensitive data, comes under privacy and security regulations as strict as HIPAA in the USA and GDPR in the EU. The use of data as a tool in this era of interoperability requires ensuring maximum data privacy and compliance while still allowing for instant data processing. This can be achieved through implementing strong encryption, access controls, and audit systems.

Scalability & Performance

With the data volumes of healthcare data tending to grow deeper and tending to be more instantaneous, data mining capability extension to reach high performance is also becoming important at this time (Giebler et al., 2021). Utilizing distributed computing framework across cloud-based while

edge computing may help address the database scaling and performance issues.

Integration & Interoperability

Health data comes in different forms rendered from various sources and systems through different data formats and standards. The fact of the matter is that this continuous flow and seamless transfer of data from numerous sources to one platform is the keystone of real-time data processing and decision-making system.

Future Direction

The potential of real-time data processing for healthcare datalakes doesn't seem dim. The integration with forthcoming technologies such as 5G networks, the edge computing, and the federated learning will be time-saving. At the same time, it will create safe pathways for the data privacy (Giebler et al., 2021). Merging of blockchain and digital twins technology facilitates new horizons such as encrypted data sharing and digital simulation for smart decision making. Likewise, the constant evolution of artificial intelligence and machine teaching procedures will give rise to sophisticated real-time predicting software and decision automation, thus improving precision medicine and individualized health care interventions

Conclusion

This move from document-based to evidence-based medicine by leveraging on real-time data processing in healthcare data lakes signifies a paradigm shift which results in a change to clinical decision-making and patient care. Projects can offer unknown mechanisms to not only facilitate quick treatments but also introduce preventive diagnosis, and running on data-driven decisions by using advanced technologies as well the big data were generated in health care industry. Although there are hurdles, the main advantage of better patient outcomes, running economy of operations and giving the right medicine makes the development of real-time data processing skills important for the industry of healthcare to have a smooth transition to an intelligent data driven and patient-centered society.

Reference List

Journals

1. Cuzzocrea, A., Leung, C.K., Soufargi, S. and Olawoyin, A.M., 2022, August. The emerging challenges of big data lakes, and a real-life framework for representing, managing and supporting machine learning on big Arctic data. In International Conference on Intelligent Networking and Collaborative Systems (pp. 161-174). Cham: Springer International Publishing.
2. Errami, S.A., Hajji, H., El Kadi, K.A. and Badir, H., 2023. Spatial big data architecture: from data warehouses and data lakes to the Lakehouse. Journal of Parallel and Distributed Computing, 176, pp.70-79.
3. Giebler, C., Gröger, C., Hoos, E., Eichler, R., Schwarz, H. and Mitschang, B., 2021. The data lake architecture framework.
4. Gopal, G., Suter-Crazzolara, C., Toldo, L. and Eberhardt, W., 2019. Digital transformation in healthcare–architectures of present and future information technologies. Clinical Chemistry and Laboratory Medicine (CCLM), 57(3), pp.328-335.
5. Maktoubian, J. and Ansari, K., 2019. An IoT architecture for preventive maintenance of medical devices in healthcare organizations. Health and Technology, 9, pp.233-243.
6. Manco, C., Dolci, T., Azzalini, F., Barbierato, E., Gribaudo, M. and Tanca, L., 2023. HEALER: A Data Lake Architecture for Healthcare. In EDBT/ICDT Workshops.
7. Nambiar, A. and Mundra, D., 2022. An overview of data warehouse and data lake in modern enterprise data management. Big data and cognitive computing, 6(4), p.132.
8. Panwar, A., Bhatnagar, V., Khari, M., Salehi, A.W. and Gupta, G., 2022. A blockchain framework to secure personal health record (PHR) in IBM cloud-based data lake. Computational Intelligence and Neuroscience, 2022.
9. Parente, S., 2020. The design of a data lake architecture for the healthcare use case: problems and solutions.
10. Sarramia, D., Claude, A., Ogereau, F., Mezhoud, J. and Mailhot, G., 2022. CEBA: A data lake for data sharing and environmental monitoring. Sensors, 22(7), p.2733.
11. Shumba, A.T., Montanaro, T., Sergi, I., Fachechi, L., De Vittorio, M. and Patrono, L., 2022. Leveraging IoT-aware technologies and AI techniques for real-time critical healthcare applications. Sensors, 22(19), p.7675.
12. Kanungo, Satyanarayan. "Edge Computing: Enhancing Performance and Efficiency in IoT Applications." International Journal on Recent and Innovation Trends in Computing and Communication 10, no. 12 (December 2022): 242. Available at: http://www.ijritcc.org

13. Kanungo, Satyanarayan. "Hybrid Cloud Integration: Best Practices and Use Cases." International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 9, no. 5, May 2021, pp. 62-70. Available at: http://www.ijritcc.org
14. Kanungo, Satyanarayan. "Decoding AI: Transparent Models for Understandable Decision-Making." Tuijin Jishu/Journal of Propulsion Technology 41, no. 4 (2020): 54-61.
15. Kanungo, Satyanarayan, and Pradeep Kumar. "Machine Learning Fraud Detection System in the Financial Section." Webology, vol. 16, no. 2, 2019, p. 490-497. Available at: http://www.webology.org
16. Kaur, Jagbir, et al. "AI Applications in Smart Cities: Experiences from Deploying ML Algorithms for Urban Planning and Resource Optimization." Tuijin Jishu/Journal of Propulsion Technology 40, no. 4 (2019): 50. (Google scholar indexed)
17. Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service . (2019). International Journal of Transcontinental Discoveries, ISSN: 3006-628X, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98
18. Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.
19. Ashok Choppadandi et al, International Journal of Computer Science and Mobile Computing, Vol.9 Issue.12, December- 2020, pg. 103-112. (Google scholar indexed)
20. AI-Driven Customer Relationship Management in PK Salon Management System. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128
21. Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. (2020). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127
22. Chintala, S. (2022). Data Privacy and Security Challenges in AI-Driven Healthcare Systems in India. Journal of Data Acquisition and Processing, 37(5), 2769-2778. https://sjcjycl.cn/18. DOI: 10.5281/zenodo.7766
23. Chintala, S. K., et al. (2022). AI in public health: Modeling disease spread and management strategies. NeuroQuantology, 20(8), 10830-10838. doi:10.48047/nq.2022.20.8.nq221111
24. Chintala, S. (2022). Data Privacy and Security Challenges in AI-Driven Healthcare Systems in India. Journal of Data Acquisition and Processing, 37(5), 2769-2778. https://sjcjycl.cn/DOI: 10.5281/zenodo.7766
25. Chintala, S. K., et al. (2021). Explore the impact of emerging technologies such as AI, machine learning, and blockchain on transforming retail marketing strategies. Webology, 18(1), 2361-2375.http://www.webology.org
26. Chintala, S. K., et al. (2022). AI in public health: Modeling disease spread and management strategies. NeuroQuantology, 20(8), 10830-10838. doi:10.48047/nq.2022.20.8.nq221111
27. Chintala, S. (2022). AI in Personalized Medicine: Tailoring Treatment Based on Genetic Information. Community Practitioner, 21(1), 141-149. ISSN 1462-2815.www.commprac.com
28. Chintala, S. (2019). IoT and Cloud Computing: Enhancing Connectivity. International Journal of New Media Studies (IJNMS), 6(1), 18-25. ISSN: 2394-4331. https://ijnms.com/index.php/ijnms/article/view/208/172
29. Chintala, S. (2018). Evaluating the Impact of AI on Mental Health Assessments and Therapies. EDUZONE: International Peer Reviewed/Refereed Multidisciplinary Journal (EIPRMJ), 7(2), 120-128. ISSN: 2319-5045. Available online at: www.eduzonejournal.com
30. Sathishkumar Chintala. (2021). Evaluating the Impact of AI and ML on Diagnostic Accuracy in Radiology. Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal, 10(1), 68–75. Retrieved from https://eduzonejournal.com/index.php/eiprmj/article/view/502