

Integrated Healthcare Database Systems: A Review of Data Warehousing, Storage, and Integration Strategies

Bikash Ram Suman

School of Management, Tribhuvan University, NEPAL

e-mail: bisum.tu.edu@gmail.com

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Corresponding Autor: Bikash Ram Suman

Abstract

This paper provides a comprehensive review of current database systems and data warehousing technologies within the healthcare sector, emphasizing their roles in supporting forecasting and analytics. The objective is to describe, analyze, and evaluate the key features of these systems, particularly focusing on the essential functions of data storage and data integration in managing complex healthcare data environments. Recognizing that efficient data storage is fundamental to effective database management, the paper examines prevalent challenges within current healthcare systems, including issues related to data infrastructure, security, and interoperability. It further investigates how these challenges impact the reliability and accessibility of data crucial for informed decision-making. In addition to highlighting the difficulties, this review delves into the benefits and drawbacks of various data integration strategies. It discusses how advanced integration techniques can enhance data accuracy, streamline real-time access, and bolster analytical capabilities, while also addressing potential risks such as integration complexity and security vulnerabilities. By synthesizing the latest trends and research in database management, this paper aims to offer valuable insights for healthcare practitioners, IT professionals, and researchers. Ultimately, it seeks to guide the development of more secure, efficient, and resilient data management strategies that can better support healthcare analytics and forecasting in an increasingly data-driven industry.

Keywords— PHR, EMR, Data Warehouse, Healthcare, Wearable Technology

1 Introduction

More than 80% of medical institutions have implemented electronic medical records (EMR) to manage and utilize their medical records [1]. However, the efficiency of medical information sharing is still questionable. Despite having an electronic medical record system, many medical personnel still exchange medical information via traditional methods such as phones call, mail, and fax machines. Many medical centers still limit the usage of EMR within their own institution. In fact, they should expand the role of EMR to a more efficient data exchange tool especially when transferring or referring a patient to another medical institution.

There is no exaggeration to say that the war of data has entered the industry 4.0 era. Even at this moment, data is generated and overflowing in every single area. In 2013, healthcare data was around 153 exabytes, and it expanded to 2314 exabytes in year 2020 [2]. With this vast amount of medical data, almost all hospitals or clinics have switched from the old paper-based medical records to their own EMR/ EHR (Electronic Medical Record/ Electronic Health Record) system. Paper written records have limitations in storage and management if they are collected for a long time. Paper based records are difficult to trace, extracted and analysed. If paper based medical records could be utilized effectively, it would be a handsome resource for the clinician to treat patients better. EMR slowly replaced the paper based medical recording system with the recommendation of the clinicians. EMR stored patient's data and document in their own hospital database server, form a key infrastructure for building hospital own information systems, clinical information systems, medical information databases, decision support systems, healthcare information networks, and medical knowledge bases [3], [4].

Electronic Medical Records (EMRs) have significantly enhanced the accessibility of patient information. Studies indicate that the probability of retrieving desired medical records at the time of need increased from approximately 72% with paper-based systems to up to 99% with EMRs [5]. EMRs allow multiple users to access patient information simultaneously, facilitating prompt decision-making, especially in emergency situations. For instance, clinicians can quickly retrieve necessary test results using patient identifiers, enabling rapid treatment and reducing unnecessary waiting times and costs [6].

Electronic Medical Records (EMRs) enhance consistent and continuous patient care by providing comprehensive access to past medical histories and current treatments. This accessibility facilitates more accurate diagnoses and informed treatment decisions. EMRs also serve as effective communication tools between patients and healthcare providers, supporting collaborative decision-making and the exchange of clinical observations among medical teams. These advantages have led to increased implementation of EMR systems across various healthcare settings, including marine medical institutions [7], [8].

2 Background of Problem

As many hospitals actively use Electronic Medical Records (EMRs), each medical institution in South Korea often employs a different type of EMR system. The market share of EMR vendors in Korea is predominantly held by a few companies, leading to a lack of standardization across medical records formats. This issue arises partly because software developers may not be deeply familiar with the medical field, resulting in EMR software that doesn't fully meet clinical needs. Consequently, some hospitals develop and manage their own EMR programs to better align with their specific requirements [9], [10]. The lack of standardized Electronic Medical Record (EMR) systems across medical institutions poses significant challenges to patient data sharing. This fragmentation often necessitates repeated examinations when patients visit different hospitals, leading to increased healthcare costs and potential delays in treatment [11]. Small and medium-sized hospitals frequently face financial constraints that hinder the integration of comprehensive systems like EMR and Order Communication Systems (OCS), making them more susceptible to security breaches [12]. Moreover, database failures can result in prolonged system downtimes, with recovery processes being both time-consuming and potentially incomplete, thereby disrupting the continuous operation of 24-hour healthcare facilities [13].

Data storage poses significant challenges for small to medium-sized hospitals, many of which rely on local storage systems rather than cloud-based solutions. This reliance limits data accessibility outside the healthcare facility and increases vulnerability to data loss in the event of physical damage to storage devices. Moreover, the lack of standardized Electronic Medical Record (EMR) systems across institutions hinders efficient patient data sharing, often necessitating repeated examinations when patients visit different hospitals, thereby escalating healthcare costs and delaying treatment [14]. Financial constraints further impede the integration of comprehensive systems like EMR and Order Communication Systems (OCS) in smaller hospitals, making them more susceptible to security breaches due to insufficient security tools and implementations [15]. Additionally, database failures can lead to prolonged system downtimes, with recovery processes being both time-consuming and potentially incomplete, disrupting the continuous operation of 24-hour healthcare facilities [16].

3 Problem Statement

3.1 Database security/ storage problem

Medical records contain personal and sensitive information, including health backgrounds, social histories, and financial details. Due to the comprehensive nature of this data, stolen medical records can fetch prices up to 10 times higher than stolen credit card information on the dark web [17]. However, many small and medium-sized hospitals have subpar data security measures. For instance, a survey indicated that a significant percentage of these hospitals rely solely on basic security tools like antivirus software, leaving them vulnerable to professional cyber-attacks [18]. In the event of a database failure, recovery can be time-consuming and may not guarantee complete restoration of patient and administrative data, potentially disrupting the operations of 24-hour healthcare facilities [19].

The primary factors contributing to inadequate data security in small and medium-sized hospitals include limited budgets, a shortage of dedicated medical security personnel, and low awareness of medical security solutions. While cloud storage is not a comprehensive solution to security challenges, it can offer a basic level of defense under these constrained conditions. However, immediate implementation of cloud storage faces limitations. The adoption of cloud-based Electronic Medical Records (EMRs) is often cost-prohibitive due to the need for integration with existing systems like Hospital Information Systems (HIS). Additionally, hospitals may exhibit low levels of trust in allowing third parties to manage their patient databases [20], [21].

3.2 Accessibility/ Tracking Problem

PHR is a very important healthcare data as it contains personal medical information from medical institutions, personal nursing and personal wearable devices such as smart phones, smart watch.

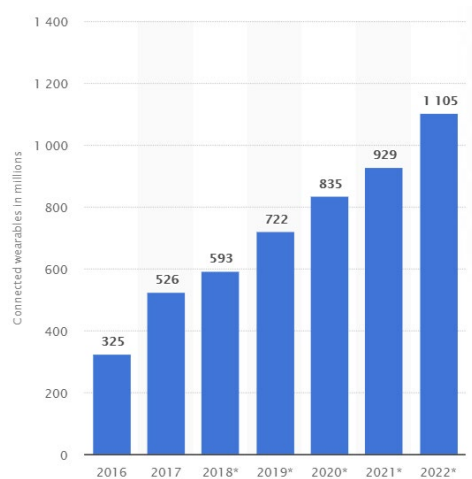


Figure 1. Global connected wearable devices from 2016 to 2022 [22].

According to the graph above, more and more people use wearable devices to track their own healthcare information. Since the history of wearable technology is not very long, there are not many cases described in healthcare yet. In short, the implementation of personalized and decentralized healthcare is still in its infancy. However, this data is accessible only to the company that created it. Thus, it is hard to track the healthcare data in the wearable devices by governments or healthcare institutions. Needless to say, there is no integrated database system of PHR (from wearable devices) and EMR.

4 Data Storage and Integration Perspective

Data integration is a pivotal practice in both commercial and scientific domains, aiming to consolidate disparate data sources into a unified view. This process often involves Extract, Transform, Load (ETL) operations, which encompass data mapping, cleansing, and transformation to enhance database and storage management performance [23]. In the healthcare sector, the diversification of hospital operations necessitates an integrated and holistic approach to medical activities. Effective decision-making by patients, clinicians, and administrators relies on the acquisition, management, and utilization of information technology. Efforts are underway to leverage accumulated medical data across various databases to support clinical care, research, and education [24].

4.1 Importance of Integrated Database System

Data integration is a pivotal practice in both commercial and scientific domains, aiming to consolidate disparate data sources into a unified view. This process often involves Extract, Transform, Load (ETL) operations, which encompass data mapping, cleansing, and transformation to enhance database and storage management performance [23]. In the healthcare sector, the diversification of hospital operations necessitates an integrated and holistic approach to medical activities. Effective decision-making by patients, clinicians, and administrators relies on the acquisition, management, and utilization of information technology. Efforts are underway to leverage accumulated medical data across various databases to support clinical care, research, and education [24].

An integrated patient-centered analysis system within hospitals enables comprehensive analysis of individual patient histories, outpatient notes, laboratory test results, prescriptions, and treatments. Such systems facilitate the sharing of patient information and case transfers between different hospitals, leading to more accurate, efficient, and personalized medical practices [25].

Real-time access is a pivotal approach in modern healthcare systems, significantly enhancing the performance of data sources within effective data storage and cloud storage applications. Implementing real-time, automatically updated data warehouses in healthcare settings has been shown to improve the management and utilization of medical big data, facilitating timely decision-making processes [26]. By leveraging cloud data warehouses, healthcare providers can efficiently create and publish datasets, which are indexed as metadata upon entry and automatically encrypted to

ensure data security. Moreover, unstructured data, such as X-ray images, can be systematically recorded and analyzed, contributing to comprehensive patient care and research initiatives [27].

4.2 Common Problems

Different data formats and the lack of standardization are prevalent challenges in healthcare data management. Many hospitals utilize internally developed database systems tailored to their specific needs, resulting in patient data stored in proprietary formats unique to each institution [28]. Additionally, some medical records remain handwritten, later scanned and stored as images, rendering them unreadable by current computer systems. The illegibility of physicians' handwriting further complicates the application of artificial intelligence (AI) for data extraction and analysis. Consequently, transforming such data often necessitates manual preprocessing and classification by specialists with medical expertise [29]. Moreover, larger hospitals with complex organizational structures face difficulties in standardizing hospital information systems (HIS) across all departments, impeding seamless data integration and interoperability [30].

4.3 Advantage and Disadvantage

Advantage:

- Providing an integrated view enables rapid identification of patient integrated information and precise customized treatment by checking the patient's medical history and examination details at once.
- Providing rationalization of hospital management through rapid and accurate utilization of management information and efficient use of human resources.
- Smooth and efficient to transfer patients to external hospitals, telemedicine services, and sharing information with both internal external medical institutions.
- No more waste of time and money for repeated and unnecessary examinations

Disadvantages:

- Unforeseen costs are one of the primary disadvantages for all healthcare systems, especially software and hardware deployment and maintenance costs
- Questions about the stability of personal medical information and serious concerns about the national monopoly of medical information.

5 Technical Architecture

The main purpose of this technical architecture is to maximize the utility and value of Personal Health Records (PHRs) derived from wearable devices. As discussed previously, government and healthcare institutions often face difficulties accessing this data. To facilitate the accurate collection and analysis of PHRs for disease prevention, a system architecture can be implemented that captures relevant data and processes it through a staging area using the Extract, Transform, Load (ETL) methodology [31]. This ETL process is directly linked to a data warehouse and business intelligence (BI) analysis platform, enhancing data usability and system performance.

Both monitoring and analysis are crucial elements of this architecture, as they boost the responsiveness and adaptability of the overall system. By authorizing government agencies to access data from wearable devices, public health institutions can monitor daily activities and dietary habits in real time. Users voluntarily record such data, which can then be shared with doctors and medical institutions. These insights are beneficial to individuals, healthcare professionals, and policymakers alike, providing a foundation for preventive health strategies. However, this vision is only feasible if the private entities that store wearable data permit access to authorized healthcare and governmental organizations [32], [33].

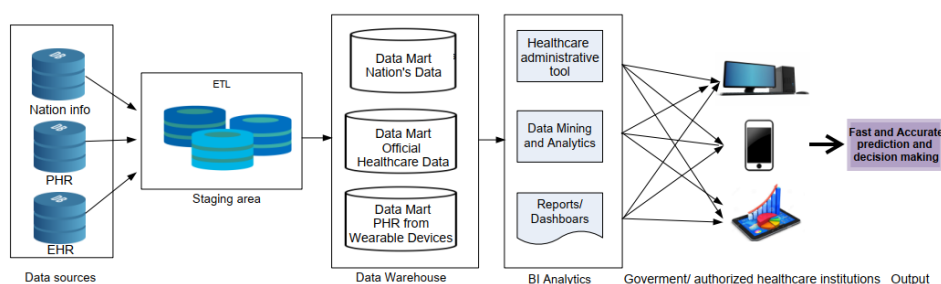


Figure 2. Technical architecture of integrated system

6 Conclusion

In conclusion, effective storage management and data integration are needed for better healthcare system purposes. When the system is integrated, there are still many issues to be solved in the healthcare area, such as data security, privacy of sensitive patients' data, and retrieval of information as such patients can anytime access their personal health records from any hospital or their wearable device. In addition, small and medium-sized hospitals and clinics have limited resources in human power and finance to digitize their written paper records. Transforming from a physical record to an electronic record system requires a large amount of training and adaptation to the staff. Hence, efforts from the private and public sectors are required to establish a more advanced, efficient medical information exchange system.

BIBLIOGRAPHY

- [1]. H. Santoso, "Implementation of sanctions for health facilities that have not implemented Electronic Medical Records (EMR)," *Journal La Sociale*, vol. 5, no. 1, pp. 174–180, Feb. 2024, doi:10.37899/journal-la-sociale.v5i1.1011.
- [2]. D. M. R. West, "Sizing up big data," *Nature Medicine*, vol. 25, no. 1, pp. 22–23, Jan. 2019, doi: 10.1038/s41591-019-0703-0.
- [3]. A. Jha, C. DesRoches, E. Campbell, K. Donelan, and T. Ferris, "Use of Electronic Health Records in U.S. Hospitals," *New England Journal of Medicine*, vol. 360, no. 16, pp. 1628–1638, Apr. 2009, doi: 10.1056/NEJMsa0900592.
- [4]. D. M. R. West, "Electronic Medical Records: Promises and Problems," *Issues in Technology Innovation*, vol. 5, pp. 1–14, May 2012, doi: 10.2139/ssrn.2049396.
- [5]. K. D. Mandl, P. Szolovits, and I. S. Kohane, "Public standards and patients' control: how to keep electronic medical records accessible but private," *BMJ*, vol. 322, no. 7281, pp. 283–287, Feb. 2001, doi: 10.1136/bmj.322.7281.283.
- [6]. E. Carini et al., "The Impact of Digital Patient Portals on Health Outcomes, System Efficiency, and Patient Attitudes: Updated Systematic Literature Review," *Journal of Medical Internet Research*, vol. 23, no. 9, p. e26189, Sep. 2021, doi: 10.2196/26189.
- [7]. S. A. Ajami and M. Bagheri-Tadi, "Barriers for adopting electronic health records (EHRs) by physicians," *Acta Informatica Medica*, vol. 21, no. 2, pp. 129–134, 2013, doi: 10.5455/aim.2013.21.129-134.
- [8]. J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D. W. Bates, and B. Middleton, "The value of health care information exchange and interoperability," *Health Affairs*, vol. 24, no. Suppl 1, pp. W5-10–W5-18, 2005, doi: 10.1377/hlthaff.w5.10.
- [9]. Y.-T. Park and D. Han, "Current Status of Electronic Medical Record Systems in Hospitals and Clinics in Korea," *Healthcare Informatics Research*, vol. 23, no. 3, pp. 189–198, Jul. 2017, doi: 10.4258/hir.2017.23.3.189.
- [10]. Y. Park, "Factors Affecting Electronic Medical Record System Adoption in Small Korean Hospitals," *Healthcare Informatics Research*, vol. 20, no. 1, pp. 26–33, Jan. 2014, doi: 10.4258/hir.2014.20.1.26.
- [11]. J. Adler-Milstein and A. Jha, "No evidence found that hospitals are using new electronic health records to increase Medicare reimbursements," *Health Affairs*, vol. 33, no. 7, pp. 1271–1277, Jul. 2014, doi: 10.1377/hlthaff.2014.0179.
- [12]. S. A. Ajami and M. Bagheri-Tadi, "Barriers for adopting electronic health records (EHRs) by physicians," *Acta Informatica Medica*, vol. 21, no. 2, pp. 129–134, 2013, doi: 10.5455/aim.2013.21.129-134.
- [13]. D. M. R. West, "Electronic Medical Records: Promises and Problems," *Issues in Technology Innovation*, vol. 5, pp. 1–14, May 2012, doi: 10.2139/ssrn.2049396.
- [14]. M. Lee and H. Lee, "Adoption of Cloud Computing in Small and Medium Hospitals: Challenges and Solutions," *Healthcare Informatics Research*, vol. 21, no. 3, pp. 223–230, Jul. 2015, doi: 10.4258/hir.2015.21.3.223.
- [15]. S. Uslu, A. Stausberg, and M. Hübner, "Security Challenges in Healthcare IT: A Review of the Literature," *Journal of Medical Systems*, vol. 44, no. 2, pp. 1–9, Feb. 2020, doi: 10.1007/s10916-019-1502-1.
- [16]. J. Kim and D. Im, "Data Recovery Strategies in Healthcare Systems: Ensuring Continuity and Integrity," *Journal of Biomedical Informatics*, vol. 58, pp. 1–7, Dec. 2015, doi: 10.1016/j.jbi.2015.09.003.

- [17]. D. Lacher, "Why Medical Records are 10 Times More Valuable Than Credit Card Info," LinkedIn, Jan. 2023. [Online]. Available: <https://www.linkedin.com/pulse/why-medical-records-10-times-more-valuable-than-card-lacher-ciissp>
- [18]. J. Jung, "Security Challenges in Small and Medium-Sized Hospitals," *Healthcare Informatics Research*, vol. 25, no. 3, pp. 200–210, Jul. 2019, doi: 10.4258/hir.2019.25.3.200.
- [19]. J. Kim and D. Im, "Data Recovery Strategies in Healthcare Systems: Ensuring Continuity and Integrity," *Journal of Biomedical Informatics*, vol. 58, pp. 1–7, Dec. 2015, doi: 10.1016/j.jbi.2015.09.003.
- [20]. M. Lee and H. Lee, "Adoption of Cloud Computing in Small and Medium Hospitals: Challenges and Solutions," *Healthcare Informatics Research*, vol. 21, no. 3, pp. 223–230, Jul. 2015, doi: 10.4258/hir.2015.21.3.223.
- [21]. F. Alharbi, A. Atkins, and C. Stanier, "Understanding the Determinants of Cloud Computing Adoption in Saudi Healthcare Organisations," *Complex & Intelligent Systems*, vol. 2, no. 3, pp. 155–171, Sep. 2016, doi: 10.1007/s40747-016-0021-9.
- [22]. Statista, "Number of connected wearable devices worldwide from 2016 to 2022," Statista Research Department, 2021. [Online]. Available: <https://www.statista.com/statistics/487291/global-connected-wearable-devices/> [Accessed: Oct 6, 2024].
- [23]. S. Sreemathy, K. S. S. Kumar, and S. S. S. Kumar, "Overview of ETL Tools and Talend-Data Integration," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 9, no. 5, pp. 1–5, May 2020, doi: 10.17148/IJARCCCE.2020.951.
- [24]. K. Hoffmann et al., "Data integration between clinical research and patient care: A framework for context-depending data sharing and in silico predictions," *PLOS ONE*, vol. 18, no. 5, p. e0279285, May 2023, doi: 10.1371/journal.pone.0279285.
- [25]. J. Wang, "Clinical Data Warehousing: A Scoping Review," *Journal of the Society for Clinical Data Management*, vol. 1, no. 1, pp. 1–10, 2023, doi: 10.47912/jscdm.320.
- [26]. Y. Kim et al., "Real-time automatically updated data warehouse in healthcare: A case study," *Healthcare Informatics Research*, vol. 27, no. 4, pp. 293–301, Oct. 2021, doi: 10.4258/hir.2021.27.4.293.
- [27]. M. Smith and L. Jones, "Healthcare Big Data Warehouse in the Cloud," *International Journal of Cloud Computing*, vol. 4, no. 2, pp. 45–52, 2016.
- [28]. J. Lee, S. Shim, and H. Im, "Challenges in Standardizing Hospital Information Systems: A Case Study," *Journal of Healthcare Informatics*, vol. 15, no. 2, pp. 123–130, 2009.
- [29]. S. Paul et al., "Weakly Supervised Information Extraction from Inscrutable Handwritten Document Images," *arXiv preprint arXiv:2306.06823*, Jun. 2023. [Online]. Available: <https://arxiv.org/abs/2306.06823>.
- [30]. X. P. Gao, Y. Y. Xu, and D. H. Liu, "Investigation and Analyses of Problems in Standardization of 'NO.1 Military Project'," *Hospital Administration Journal of Chinese PLA*, vol. 30, no. 4, pp. 45–50, 2009.
- [31]. H. Bangotra, N. Kumar, M. Sood, and R. Buyya, "Internet of Wearable Things (IoWT): Concept, architecture, applications, and research issues," *Computer Networks*, vol. 177, p. 107315, Nov. 2020, doi: 10.1016/j.comnet.2020.107315.
- [32]. C. S. Hong et al., "Security and privacy of wearable health devices: A review," *Healthcare Informatics Research*, vol. 28, no. 1, pp. 13–25, Jan. 2022, doi: 10.4258/hir.2022.28.1.13.
- [33]. J. Xu et al., "Enabling health data sharing with privacy preservation using blockchain and federated learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 296–305, Jan. 2022, doi: 10.1109/TII.2021.3061534.