International Journal of Information Technology and Computer Science Applications (IJITCSA)
p-ISSN: 2964-3139 e-ISSN: 2985-5330
Vol. 03, No. 03, page 126 - 131
Submitted 12/10/2025; Accepted 17/12/2025; Published 20/12/2025

Complications in healthcare integration models and correlated data infrastructure proposition

* **Ohmar Shiraz Arfeen, Rauf Shahzad Shahrin, and Ashraf Zeeshan Ahmad**
Department of Computer Science, DHA Suffa University, Karachi 75500, PAKISTAN
e-mail: ohm67arfeen@gmail.com, RaufShahzadShahrin@gmail.com, AshrafZeeshanAhmad@gmail.com

Corresponding Autor: Ohmar Shiraz Arfeen

Abstract
In healthcare systems, proper data integration models are necessary in order to provide swift treatment for patients. Without integrity and proper management of patient data, it can result in losing many lives due to unwanted delays in getting the necessary data. This study aims to solve this problem by looking at different data-related technological perspectives and discussing which is best suited for the healthcare sector. Multiple papers on different technological perspectives are reviewed to identify the underlying problems and how they can be tackled individually without getting drawbacks in return. Most impactful problems are highlighted and discussed extensively. The findings show that a data warehouse is the most viable option for tackling the highlighted problems due to its highly centralized infrastructure and data consistency. Elaborations are made on the viability of a data warehouse and how it can help healthcare systems in terms of effective data management.

Keywords— Data Warehouse, Data Lake, Big Data, Database, Interoperability, Data Storage, Snowflake

1 Introduction
We now live in a time where quality healthcare provision is of utmost importance in the light of the Covid-19 pandemic that has left the world economy and physical accessibility in a substandard state. To ensure quality healthcare, the way medical-related data is stored and managed needs to be integrated system so that patients can clear out their confusion and in turn get their treatment effectively and promptly. We can think of establishing an integration that is well-balanced and systematic by looking at data-related technological perspectives like a data warehouse, data lake, big data, and database models.

1.1 *Discussion on different technological perspectives*
If we were to look from the perspective of traditional database models, there are certain problems regarding how outdated records are managed as it causes inefficiencies in terms of the operability of the data system. There is also concern regarding how network communication becomes a problem in data management, as remote access requires frequent information updates, as described by Wilkes et al. [1]. Another point discussed is the importance of interoperability in healthcare systems, as high-quality and low-cost healthcare services can be delivered across different countries.

Furthermore, interoperability standards must be established to ensure the privacy and security of healthcare systems. The more secure healthcare systems are, the easier they are to manage properly and effectively, as highlighted by Gavrilov, Vlahu-Gjorgievska, and Trajkovik [2]. Their work explains how a data warehouse can be used to develop an interoperable data storage and management system. During the pandemic, it became necessary to attend to patients swiftly due to the surge of infected cases worldwide. To accomplish this, it is imperative to filter query-relevant data from enormous pools of information through schemas designed for specific medical purposes. This process enables personalized medication whereby precise treatments and therapeutics are prescribed to individual patients after filtering all relevant data that may influence therapeutic response, as explained by Sarathkumar, Liu, Wang, and Wang [3].

Additionally, this results in less data being stored for patients, reducing overall storage costs since larger data volumes require greater storage capacity. The paper further noted that many systems still require pre-processing of medical records before storage, contributing to unwanted delays when accessing multiple data sources. In the era of rapidly emerging data types, digitized health information is expanding dramatically, with data coming from internal and external sources—mobile devices, wearable sensors, Electronic Health Records (EHR), radiology images, videos, clinical notes, social media, blogs, remote monitoring systems, and newer forms such as imaging and sensor readings. These sources collectively increase the need for Big Data solutions capable of managing the vast silos of information within the healthcare industry.

The presence of numerous forms of data—structured, semi-structured, and unstructured—makes healthcare data inherently challenging, as noted by Prabha and Anitha [4]. In the genomics era, the volume of data captured from biological experiments and routine healthcare processes is expanding at an unprecedented pace. This wealth of information promises major advances in healthcare research and breakthrough treatments, but it also introduces new technological, administrative, and knowledge-management challenges. Adibuzzaman et al. [5] emphasized that healthcare systems are already overwhelmed by massive data volumes, increasing storage costs and creating management issues due to the rising complexity of data organization. This paper revisits these ongoing problems and proposes an integration model to address them.

2 Problem Formulation

Despite the fact that there are various backgrounds described regarding the problems above, only two of the most concerning aforementioned will be discussed. These two problems directly impact the integrity of the data systems in the healthcare aspect where proper attendance to patients for their wellbeing is concerned.

2.1 *Old data record management*

As mentioned in the introduction, an overwhelming amount of data is accessed and stored daily within healthcare systems. Over time, this accumulation can lead to a substantial buildup of data in hospital storage systems. The efficiency of the database system is detrimentally affected due to the presence of such large volumes of outdated data, resulting in inconsistencies within the database as the informative value of old records diminishes. Conversely, setting a deletion threshold and removing all data older than a specified date may also create systemic issues, as some historical information remains essential for workflow continuity, medical decision-making, and future reference.

Blind deletion of data can lead to the destruction of valuable information and cause operational disruptions within the source system. Therefore, it is necessary to establish a mechanism that maintains a balance between retaining and removing outdated data, as emphasized by Wilkes, Paul, and A. [1]. Relevant data must be filtered from irrelevant records within the pool of older data through a systematic procedure, ensuring that only non-essential information is removed. Additionally, historical data that remains stored may influence trend analysis outcomes, as illustrated in Figure 1.

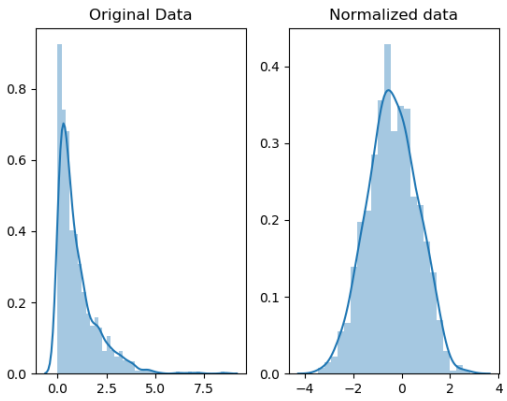


Figure 1. Comparison of original and normalized data

There can be conditions when compared to the recent data, some of the old data were not affected by certain variables that are brought about by immediate occurrences which result in miscalculations during the analysis process. Additionally, due to the lack of a systemic data integration system to selectively differentiate relevant data

from irrelevant ones, more storage space is used resulting in higher costs. Maintenance of that vast pool of data also takes a share in high costs due to having to hire more people in order to withhold a fluid workflow.

2.2 *Less efficient management due to problems in network communication and security*

In hospital facilities, multiple computers often do share common information throughout their databases spread across as a communication network. Online database systems are necessary not only because of real-time data synchronization but also due to limited functionality available for offline systems. Remote access is also of importance, which needs to be irrespective of the network communication link, but the problem comes when modifications are made to the root database system as it may require access to that particular data by another database within the same network.

Moreover, shared database systems require frequent information updates in their local memory in order to function properly. Without effective data integration, frequent data updates and modifications require continuous surveillance, intensive labor, and substantial time investment, as noted by Wilkes, Paul, and A. [1]. Undoubtedly, this also results in higher operational costs for the data management sector. With network communication comes an increased risk of network security issues, as data exchange between multiple databases can lead to data leaks or data loss. Therefore, a centralized control system is necessary—one capable of frequently communicating across the entire database network to assess which information should be transmitted and to which database components. As long as online services are required, network-related challenges will persist, particularly when handling large-scale systems.

3 Problem Solution

3.1 *Essentiality of data integration*

As discussed previously in the introduction, the data usage of healthcare is rapidly growing at an alarming rate due to the fact that various methods of data attainment are emerging accordingly with technological advancements. It is important to get in control of those data as it can significantly improve clinical results, patient care, and also financial statements for the organizations related to healthcare. In order to effectively monitor and attend to patients with rare conditions, it is important for there to be old data records for referencing purposes so that the follow-up effects can be analyzed better.

At the same time, deletion of unnecessary data in hospital records allows more valuable information about new patients in critical condition to be stored. Improved data management also enhances optimization in patient scheduling, patient intake, and verification of insurance and billing forms. Furthermore, physicians' work-related information—such as privileges, organizational affiliations, faculty appointments, and work locations—can be tracked more efficiently. Enhanced network communication within the data system also improves call center productivity and enables rapid verification of applications and appointment information, as real-time data becomes more accessible, as discussed by Islam, Hasan, Wang, Germack, and Noor-E-Alam [6].

3.2 *Problems in implementing data integration*

- Wrong data formatting
  Actions like data analysis and visualizations cannot be performed on anomalous data that is ambiguous or improperly formatted. Manual data formatting, validation, and correction are tedious tasks that consume a significant portion of IT professionals' time, as noted by Kamble, Angappa, Milind, and Jaswant [7].
- Data duplication
  Making sure that data duplication is not present, while also establishing a foundation for a well-integrated data system to prevent future duplication, is one of the key challenges faced by IT professionals. The presence of multiple duplicated records can result in missed sales opportunities, as significant time is wasted attempting to reach contacts who are no longer associated with their organizations, as highlighted by Boskova and Stadler [8].
- Low-quality data
  Managing data quality is also one of the challenges, as poor-quality data can lead to lost revenue, reputational damage, and missed insights. Accurate business decisions can only be made when proper data quality management practices are implemented, as emphasized by Kamble, Angappa, Milind, and Jaswant [7].
- Data misplacement

IT professionals must manually curate data from contrasting sources and combine them, a process that consumes significant time. This creates inefficiency, as such time should instead be spent on data insight analysis and reassessing business practices of value, as noted by Kamble, Angappa, Milind, and Jaswant [7].

- Lack of understanding of data
  When sharing data between technical and business teams, some data definitions are not thoroughly understood, causing miscommunication that disrupts workflow. Implementing data governance and data stewardship within the data integration process to resolve these issues is one of the prominent challenges faced by IT professionals, as highlighted by Kamble, Angappa, Milind, and Jaswant [7].

3.3 *Advantages and Disadvantages of data integration in healthcare systems*

3.3.1 Advantages

- Higher efficiency
  As a lot of important information is collected from different patients and the time of doctor and the patient meeting is limited, specific data filtering services provided by data integration can allow doctors to stay focused on each individual patient [9].
- Error reduction
  Data integration can result in computerized orders from doctors which reduces the errors caused by misunderstanding doctor's handwriting as well as mistakes in transcriptions.
- Early Detection
  A well-integrated data system can be helpful in the early detection of medical issues due to predictive assessments and algorithmic processes [9].
- Providing lifesaving information
  Data integration can provide doctors with relevant information from past records to help them determine what is the best possible treatment for patients having similar symptoms.

3.3.2 Disadvantages

- Increased workload
  Data integration brings about advancements that are beneficial but at the same time increases the workload of the medical organizations. Frequent data updates need to be managed properly to avoid orders being stalled due to lack of items in stock [9].
- High cost
  Establishing a well-integrated data system is costly as organizations not only have to buy the necessary software but also train the staff in how to use it [9].
- Technical problems
  As inevitable it is for data integration computer systems to have a breakdown, it is very concerning for the healthcare sector as patients' lives are at stake and the provision of medication will get problematic.

3.4 *Proposition of Technical Architecture*

The technical architecture to be proposed is based on the Snowflake schema of the data warehouse. The business scenario to be exemplified is keeping data records for patients that are discharged from a hospital. Each row of the fact table represents a patient discharged from a certain branch on a certain date by a certain specialist in the hospital. The dimensions for this model were created by going through business questions that are shown in Figure 2 below.

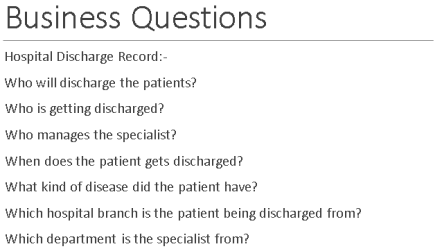


Figure 2. Business questions for dimension tables

The fact measures were also determined through the questions shown in Figure 3 below.

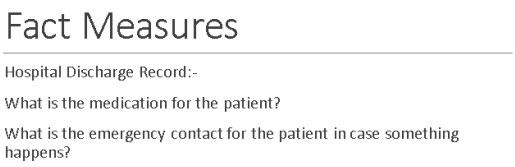


Figure 3. Fact measures for the fact table

These sequential processes led to the creation of the snowflake schema shown in Fig.4 below that is based on a typical hospital-related business scenario which is keeping records of discharged patients.

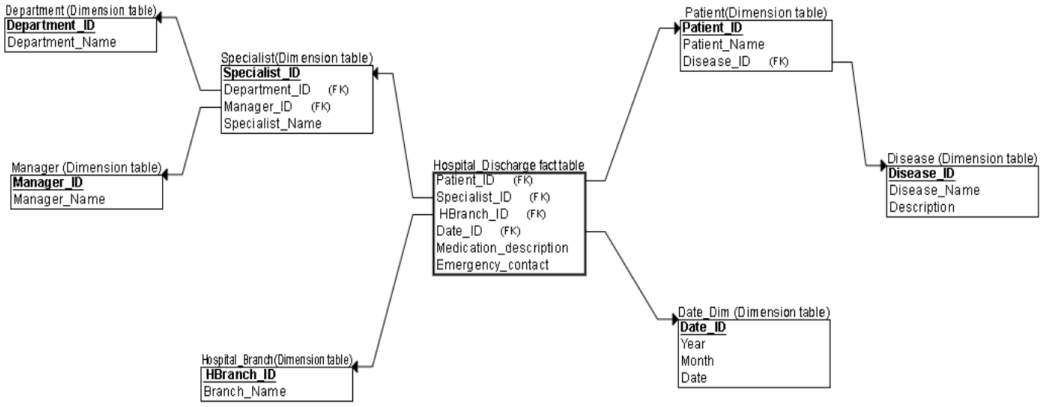
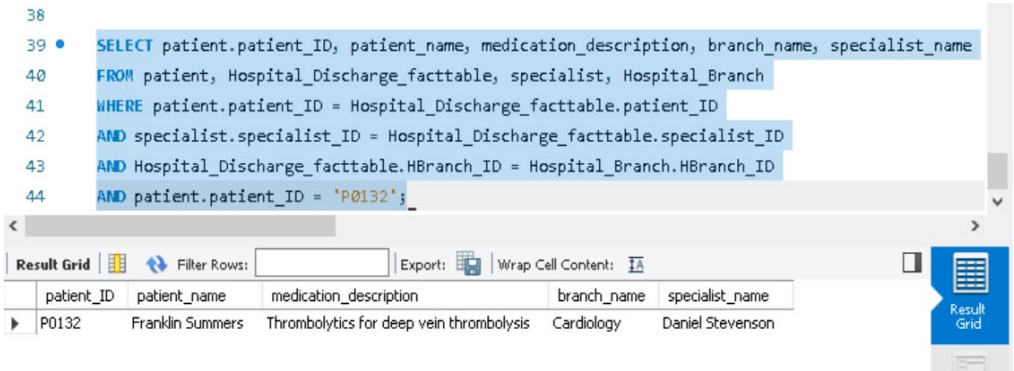


Figure 4. Snowflake schema

The relationships are also displayed denoting the interconnectivity of the dimensions and the fact table in the schema. There are sub-dimensions that split off from patient and specialist dimensions which increase their data quality due to complementary background information provided and in turn make the fact table more sensible. The scheme also provides a lot of options when it comes to joining tables for a filtered view to getting data based on varying business requirements like an example shown below in Figure 5 where selective information about a specific patient discharged is shown by executing a complex multi-table join query.



```
38
39 • SELECT patient.patient_ID, patient_name, medication_description, branch_name, specialist_name
40   FROM patient, Hospital_Discharge_facttable, specialist, Hospital_Branch
41   WHERE patient.patient_ID = Hospital_Discharge_facttable.patient_ID
42   AND specialist.specialist_ID = Hospital_Discharge_facttable.specialist_ID
43   AND Hospital_Discharge_facttable.HBranch_ID = Hospital_Branch.HBranch_ID
44   AND patient.patient_ID = 'P0132';
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

patient_ID	patient_name	medication_description	branch_name	specialist_name
P0132	Franklin Summers	Thrombolytics for deep vein thrombolysis	Cardiology	Daniel Stevenson

Result Grid

Figure 5. Data of a specific discharged patient

This alone shows how connected the dimensions and the fact table are in terms of relationship and perfectly fulfill the purpose of the business scenario mentioned. Finally, the snowflake scheme is incorporated into the full technical architecture shown in Figure 6.

4 Conclusion

In conclusion, for the long term, potential data warehouse storage problems can be addressed by incorporating a cloud-based system from the beginning. The cost saved from not having to purchase physical hardware can instead be allocated toward acquiring scalable cloud storage solutions, as cloud platforms provide virtually unlimited capacity and elasticity [10], [11]. Furthermore, to justify the technical architecture being proposed, it effectively resolves the issues described in the problem statement. First, old data record management is no longer problematic, as historical records can be retained efficiently due to the snowflake scheme requiring less storage through normalized table structures [12]. Additionally, normalization significantly reduces data redundancy, improving both performance and maintainability.

Network security can also be strengthened by implementing fine-grained access policies and permissions for authorized hospital users. User privileges can be customized and granted in the form of defined roles by the database administrator, improving accountability and compliance with healthcare data regulations [13]. By restricting sensitive data transfers, the risks of data loss or leakage during transmission are greatly reduced. The centralization inherent in the snowflake schema further contributes to enhanced connectivity and interoperability across the interconnected hospital information systems, strengthening the reliability of the overall network.

BIBLIOGRAPHY

[1]. G. J. Wilkes, E. S. Paul, and A. P., Healthcare Database Management Offline Backup and Synchronization System and Method, U.S. Patent US20030204420A1, 2003. [Online]. Available: https://patentimages.storage.googleapis.com/1a/b0/8d/a9a29f5bd62c45/US20030204420A1.pdf

[2]. G. Gavrilov, E. Vlahu-Gjorgievska, and V. Trajkovik, "Healthcare data warehouse system supporting cross-border interoperability," Health Informatics Journal, 2019, doi: 10.1177/1460458219876793.

[3]. S. Rangarajan, H. Liu, H. Wang, and C.-L. Wang, "Scalable Architecture for Personalized Healthcare Service Recommendation Using Big Data Lake," in Service Research and Innovation, 2018, doi: 10.1007/978-3-319-76587-7_5.

[4]. P. S. Mathew and A. S. Pillai, "Big Data solutions in Healthcare: Problems and perspectives," in 2015 Int. Conf. Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015, doi: 10.1109/ICIIECS.2015.7193211.

[5]. M. Adibuzzaman, P. DeLaurentis, J. Hill, and B. D. Benneyworth, "Big data in healthcare – the promises, challenges and opportunities from a research perspective: a case study with a model database," AMIA Annual Symposium Proceedings, vol. 2017, pp. 384–392, 2018.

[6]. M. Islam, M. Hasan, X. Wang, H. Germack, and M. Noor-E-Alam, "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining," Healthcare, vol. 6, no. 2, p. 54, 2018, doi: 10.3390/healthcare6020054.

[7]. S. S. Kamble, A. Gunasekaran, M. Goswami, and J. Manda, "A systematic perspective on the applications of big data analytics in healthcare management," International Journal of Healthcare Management, vol. 12, no. 3, pp. 226–240, 2019, doi: 10.1080/20479700.2018.1531606.

[8]. V. Boskova and T. Stadler, "PIQMEE: Bayesian Phylodynamic Method for Analysis of Large Data Sets with Duplicate Sequences," Molecular Biology and Evolution, vol. 37, no. 10, pp. 3061–3075, 2020, doi: 10.1093/molbev/msaa136.

[9]. M. Sony, "Pros and cons of implementing Industry 4.0 for the organizations: a review and synthesis of evidence," Production & Manufacturing Research, vol. 8, no. 1, pp. 244–272, 2020, doi: 10.1080/21693277.2020.1781705.

[10]. A. Sultan, "Cloud computing for education: A new dawn?," International Journal of Information Management, vol. 30, no. 2, pp. 109–116, 2010.

[11]. M. Armbrust et al., "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[12]. R. Kimball and M. Ross, The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd ed. Wiley, 2013.

[13]. S. Zeng, F. Luo, and M. Sadiq, "A survey of access control models in cloud computing," IEEE Access, vol. 8, pp. 210747–210764, 2020.