Preprints.org

Article

Leveraging Data Warehousing for Healthcare Delivery – A case of Outpatient Department

Aggrey Obbo [*] , William Wasswa , Annabella Basaza-Ejiri

Article

Leveraging Data Warehousing for Healthcare Delivery—A Case of Outpatient Department

Aggrey Obbo [1,†,*], **William Wasswa** [1] **and Annabella Basaza-Ejiri** [2]

1   Department of Software Engineering, Mbarara University of Science and Technology; awwasswa@must.ac.ug (W.W.)
2   Department of Information Technology, Makerere University Kampala; annabella.habinka@cit.ac.ug (A.B-E)
*   Correspondence: aggrey@must.ac.ug or agrymike@yahoo.com.
†   Faculty of Computing and Informatics, Mbarara University of Science and Technology, P. Box 1410, Mbarara, Uganda

Abstract: Mortality rates in sub-Saharan Africa remains unacceptably high, and disproportionately affecting economically disadvantaged communities. And healthcare delivery which is considered a major contributor to this is challenged by the limited resources and high service costs, prompting many patients to opt for outpatient department services. This study addresses the need for a data-driven approach to healthcare policy formulation by proposing a model for the implementation of a data warehousing system that integrates outpatients into the national healthcare framework. By leveraging data warehousing, healthcare delivery can be made more efficient through a robust repository that supports decision-making and policy formulation. The research focuses on the outpatient department in resource-scarce settings, proposing a novel multidimensional data model and practical framework implementation. The benefits include improved patient experiences, cost-effective service delivery, and enhanced patient management through electronic health records. Additionally, mobile apps are explored for patient navigation and reminders to enhance treatment adherence. The study's results indicate that data warehousing can significantly improve healthcare delivery by consolidating critical data and facilitating better planning, resource allocation, and policy formulation. Ultimately, this approach offers valuable insights for optimizing healthcare systems in resource-constrained environments, enabling data-driven policies, and improving health outcomes across sub-Saharan Africa.

Keywords: data warehousing; healthcare; electronic health records; outpatients

---

1. Introduction

Mortality rates in sub-Saharan Africa remain unacceptably high, and this has disproportionately affected the economically disadvantaged communities. Although factors like household dynamics [1] and disease outbreaks like cancer and the covid pandemic [2,3] have accelerated the problem, the limited access to quality healthcare services resulting from inadequate infrastructure, disproportionate resource allocation and policies that are not evidence based have had a significant contribution to this critical issue [4–7]. While the insufficient pays and high costs of living has caused a number of medical staff to "*juggle*" between medical facilities and thus report for duty only during their scheduled shifts [8]. These imbalances arising from the high patient-to-doctor ratio has led to countries using surgical task-shifting from physicians to non-physician clinicians (NPCs) [9] and other subsidized services offered by community health workers, mobile clinics, and traditional healing practices that may not appropriately meet patient needs. Additionally, their places of residence may be far from health facilities, making it hard to reach them in times of emergencies. This situations exacerbates the challenge of providing timely medical care in many communities. In addition, a vast majority of healthcare providers are privately owned and therefore charge fees that may not be affordable to low income patients especially in rural communities. With this background, healthcare access and delivery in low-income regions remain limited and call for special attention. Despite these obstacles, however, resilience and resourcefulness characterize efforts to improve healthcare accessibility and outcomes in the region. This work proposes an innovative approach that leverages data warehousing techniques to identify and address disparities, thereby enhancing healthcare delivery for the poor and socially disadvantaged. It then goes ahead to propose tools to enhance better health outcomes for outpatients.

Data warehousing is the process of collecting, storing, and managing data from different sources in a centralized repository for effective data-driven decision-making. The purpose of a data warehouse is to provide a unified and integrated view of data for analysis and reporting. The concept of data warehousing originated in the 1980s with the traditional ETL (Extract, Transform, Load) architecture. It involved extracting data from multiple heterogeneous sources, transforming it, and loading it into data warehouses. In data warehouses, data was organized and structured for efficient querying and analysis[10] however, due to technological advancements, these have evolved into data lakes. A data warehouse stores clean, processed, and structured data for analysis and reporting, while a data lake stores unstructured, semi-structured, and unstructured data for flexible exploratory data analysis. This work maintains the term data warehousing since these two technologies complement each other and are not replacements for each other. Data warehouses also support complex analytical queries and business intelligence applications compared to data lakes. Data warehousing facilitates business intelligence (BI) activities by providing a solid foundation for reporting, querying, and data analysis. It allows organizations to make informed decisions based on a comprehensive and consistent view of their data. Popular data warehousing technologies include Informatica, Oracle Data Integrator, and AB Initio [11].

In the sphere of healthcare, data warehousing can enhance patient care by consolidating and organizing vast healthcare data. It enables comprehensive analysis, aiding in treatment decisions, resource optimization, and trend identification. This centralized repository ensures quick access to patient records, facilitates interoperability among different systems, and supports evidence-based practices. Ultimately, data warehousing promotes efficiency, accuracy, and informed decision-making, contributing to improved healthcare outcomes and delivery. Several studies have been conducted with different objectives in the areas of data warehousing and healthcare. Among others, they include: The role of data warehousing in the detection, control, and prevention of adverse events in the health of an economy was highlighted by Willie and Davies in their work, "Role of data warehousing in healthcare epidemiology" [12]. A study was also conducted that described a portal framework that offers a comprehensive interoperability services to enable national e-Health platforms to set up cross-border health information networks compliant with European Patients Smart Open Services [13]. The huge increases in medical devices and applications today have generated large data volumes that only exacerbated the need for a more efficient means of managing, processing, and mining data. Most of the sub-Saharan countries have, however, not taken up data warehousing because of inadequate knowledge of data warehousing, challenges of limited financial resources, inadequate infrastructure, and a lack of skilled personnel. Additionally, governance issues contribute to slow adoption, hindering the implementation of advanced health data management systems in the region.

This paper underscores the significance of data warehousing in healthcare delivery. Focused on highlighting data warehousing concepts to facilitate streamlined tracking of outpatients and central databases either at health facilities or regional health centers to enable an overall view of the health sector. The study emphasizes data-driven insights to inform evidence-based policy formulation, demonstrating the pivotal role of robust data management in healthcare decision-making. The paper is organized as follows: Section 2 provides an introduction to data warehousing in and healthcare; Section 3 outlines the materials and methods and finally, the conclusion and suggestions for future research are presented in section 4.

2. Related Work

A data warehouse as a system stores and processes large amounts of data from various sources within a business. It is considered an integral component of Business Intelligence (BI), and enables businesses make better, more informed decisions by applying data analytics to large volumes of Information. And a data warehouse is characterized by four fundamental characteristics that include: Subject oriented-ness, Time variant, Non-volatility and integrated [14,15]. In the context of healthcare service delivery, data warehousing is characterized by the following key attributes:

- *Subject Oriented*. Subject oriented-ness in the context of a data warehouse refers to the design principle where the data warehouse is organized and optimized around business subjects or topics, rather than the operational processes or functions of an organization. This approach allows for a more user-friendly and business-focused structure, making it easier for analysts and decision-makers to retrieve relevant information. Subject oriented-ness ensures the data focuses on specific healthcare aspects, aiding in comprehensive analysis and decision-making. This approach enhances the effectiveness of healthcare services by providing valuable insights into patient care, operations, and outcomes.
- *Integrated*. The integrated-ness of a data warehouse refers to the degree to which data from various sources and formats are combined and made accessible in a unified manner within the warehouse. A highly integrated data warehouse ensures that data is consistent, accurate, and can be easily analyzed for decision-making purposes. Data warehousing in healthcare involves consolidating and organizing diverse healthcare data sources into a unified repository. Integrated-ness ensures seamless access and analysis, fostering data-driven decision-making in healthcare services. This approach enhances efficiency, patient care, and enables comprehensive insights for informed healthcare management.
- *Volatile*. Volatility refers to the frequency and extent of changes in the data. If data within the warehouse frequently changes or is updated, it is considered volatile. This could be due to regular updates, inserts, or deletions of data. A data warehouse unlike an operational database, should be non-volatile. In healthcare, data warehousing volatility refers to the fluctuation in the volume and types of data collected over time. It impacts the ability to maintain stable and consistent data for analysis, making it challenging to derive meaningful insights. This volatility can stem from evolving healthcare protocols, changing patient demographics, and technological advancements. This heightens the demand for data ware housing.
- *Time-variant*. And a data warehouse is considered time-variant because it stores historical data and allows users to analyze changes in data over time. Data warehousing in healthcare involves storing and managing vast amounts of patient-related information. Time variance in this context refers to tracking changes in health data over time, allowing for historical analysis. The time variance property also aids healthcare providers in making informed decisions, improving patient outcomes, and enhancing overall efficiency in delivering quality care.

Healthcare data relevant to policy formulation exemplify the four key attributes of data warehousing in the following ways: They are non-volatile since, unlike operational data stores, where new patient information, treatment records, and medical research are updated frequently, data at central repositories are only updated after a relatively longer duration of time. In addition data is organized around specific healthcare subjects, such as patient demographics, diagnoses, and treatments; hence, they are subject-oriented. Health data also has to be integrated, given that data from various sources, including hospitals, health centers, and insurance providers, is combined for comprehensive analysis. And finally, health data is time-variant because it is historical, capturing changes over time to track patient progress, treatment outcomes, and epidemiological trends. The above-mentioned characteristics only point to the fact that healthcare delivery can leverage the strengths of data warehousing.

2.1. Data Warehouse Architecture

An effective data repository needs to both unify and regulate huge amount of data to enhance the analytics. There are a number of data repositories, but the most popular ones are: enterprise data warehouses and data lakes [16]. A data warehouse architecture is an intentional design comprising data services and subsystems that enable the consolidation of disparate data sources into a single repository for business intelligence and analysis. Since its inception in the late 80's and early 90's data warehousing has transformed due to the increase in data Volumes, Velocity and Variety (VVV) together with the need for efficiency and flexibility of emerging data management solutions. The warehousing architecture has as a result progressed through the following:

Traditional data warehousing architecture. The traditional data warehousing architecture was made up of a data source, data store, data marts, a staging area, and a presentation/data access area [17]. And within the data warehouse had extraction, transformation, and loading processes. A data mart is a subset of a data warehouse focused on a particular department, or subject area. The staging area is an intermediate storage area where data is temporarily held before it is processed and loaded into a data warehouse or another target system. It helps in ensuring data quality, consistency, and integrity before it is moved to the final destination. The presentation area refers to part of the system where data is presented, visualized, and analyzed. It could involve dashboards, reports, charts, and other visualization tools that make complex data more understandable and actionable. A traditional data warehousing architecture consisted of a central repository with a schema-based structure, optimized for query performance and reporting as highlighted in Figure 1 below.
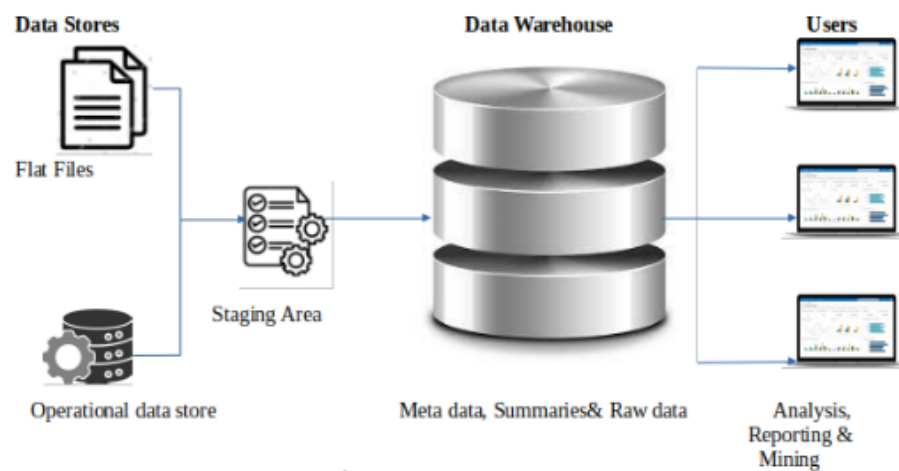
Figure 1. Data warehousing architecture.

Enhanced data warehousing architecture. In addition to the traditional ETL architecture, the enhanced data warehousing architecture incorporated sophisticated ETL tools and enhanced reporting techniques into the data warehousing processes. The introduction of Business Intelligence (BI) tools were to enable in-depth analytics and reporting for both structured and unstructured data.

Cloud based data warehousing architecture. With the rapid increase in data accompanied by data related applications, the cost of maintaining data locally has proportionately increased. The costs relate to data storage mediums, power, System Administrators etc. Cloud computing aims at eliminating or at least reducing this burden. Cloud computing further facilitates users by provisioning Software-as-a-Service (SaaS), Platform-as-aService (Paas) and Infrastructure-as-a-Service (Iaas) for organizations that may have limitations at inhousing fully fledged data warehouses locally. Similarly, with the advent of cloud based computing, data warehousing moved to the cloud [18]. This brought about a shift from the traditional ETL architecture to the Extraction, Loading and Transformation (ELT) architecture. Data was extracted, loaded and leveraging on the processing power of the modern data ware housing solutions and tools, the data is thereafter transformed. Examples of cloud based data warehousing solutions include: Google BigQuery, Amazon Redshift. This architecture offered flexibility, scalability and cost efficiency. the cloud based data warehousing architecture enabled real-time analytics, allowing for businesses to make faster data-driven decisions.

Data lakes. In response to the need for handling vast amounts of unstructured and semi-structured data, there arose the need for a more flexible and scalable approach to data storage. This led to the emergence of data lakes [19,20]. In the data lakes architecture, data is stored in its raw format and structured only when read, an architecture referred to as schema-on-road. This introduced greater flexibility in data ware housing. Data lakes store structured, semi-structured, and unstructured data, making them ideal for big data analytics and machine learning applications. Despite their advantages

however, data lakes face challenges of data governance, quality control, and ensuring efficient data retrieval.

Each stage of the evolution of data warehousing brought improvements in handling and analyzing data, addressing the growing complexities and demands of modern data environments. While data warehouses offer structured, high-quality data with strong governance and security, data lakes provide flexibility and scalability for diverse data types, making them suitable for big data and advanced analytics. In the healthcare sector, leveraging the strengths of both data warehousing and data lakes can lead to more comprehensive and insightful data analysis, ultimately enhancing patient care and operational efficiency. In this work however, the researcher maintains the use of the traditional data warehousing architecture because of the strengths highlighted in Table 1 below.

Table 1. Comparing traditional data warehousing architecture to data lakes architecture.

Attribute	Data warehousing	Data lakes
Data structure	Employs a defined schema	Stores structured, semi-structured, and unstructured data
Data quality	High, due to rigorous ETL processes	Variable, raw data storage can lead to quality issues
Query Performance	Optimized for fast query performance and reporting	Can be slower for complex queries
Security	Robust security measures and access controls	Security can be complex due to diverse data formats
Analytics	Advanced analytics and BI tools for structured data	Suitable for big data analytics and machine learning
Integration	Comprises of well-integrated, consistent data from multiple sources	Can integrate diverse data but requires significant effort
Governance	Strong data governance and compliance capabilities	More challenging to enforce governance and compliance

2.2. Data Modeling

The data warehouse modeling process works on designing and organizing data models within a data warehouse. It involves setting up appropriate databases and schema that can then be transformed and stored in a way that supports decision making. There are four prominent data warehousing data models namely: dimensional models, the relational/normalized models, anchor model and data vault model [21]. Two major data modeling techniques stand out - multidimensional modeling and normalized modeling [22]. Multidimensional modeling focuses on organizing data into a structure that reflects the business's measures, dimensions, and context. It arranges data into a star schema or snowflake schema, emphasizing easy querying and analysis. This approach is well-suited for decision support systems like data warehouses, as it simplifies complex data relationships and allows for efficient retrieval of information. On the other hand, normalized modeling follows traditional database normalization principles, aiming to eliminate data redundancy and ensure data integrity. While normalized modeling is effective for transactional systems, it can introduce complexity in querying and analyzing data for decision-making purposes, making it less ideal for data warehousing environments where performance and ease of analysis are crucial.

Data warehouse design patterns are solutions to recurring challenges in building and managing data warehouses. They aid in organizing, storing and accessing data in a way that supports analytical needs and goals. In addition to other generic models [23]. According to Kimball's, these patterns are

based only on the structure of the data model for the data warehouse and not on all business needs. Kimball suggested generic patterns for creating data models of data warehouses [24,25]. The patterns were categorized as generic star schema and and data warehouse bus matrix patterns. And the five groups of patterns illustrated in the different case studies included: 1) fact table related patterns; 2) dimension table related patterns; 3) dimension's hierarchy related patterns; 4) fact changing related patterns; and finally 5) product heterogeneity related patterns.

The data warehouse bus matrix patterns is concerned with ensuring consistency of information throughout the enterprise. This is done by mapping each subject with each common dimension. The generic star schema modeling on the other hand uses two main types of tables - dimension and fact tables. Dimension tables contain detailed information about entities and events. The dimensions take the form of: conformed dimension (shared across multiple data marts), slowly changing dimensions (capture historical changes in data), junk dimension (stores unrelated attributes), degenerate dimension (derived from fact table), role-playing dimension (same dimension used in multiple roles), outrigger dimension (extends existing dimension), shrunken dimension (subset of a larger dimension), and calendar date dimension (specifically for time-related analysis). These dimensions serve different purposes in organizing and analyzing data within the data warehouse environment. The dimension-related patterns can play the following crucial roles in the following ways:

- *Generic dimension*: It helps provides a consistent framework for data collection, enabling comparison across different parameters such as age, community or gender. This is possible because generic dimension pattern involves standardized dimensions, such as patient demographics like age, gender or even address. This then improves data integrity.
- *Date dimension*: Date dimension track events with respect to time. For example, they can be used to track patient visits, treatments, or even follow-ups. With data dimensions, it is possible to carry out trend analysis, such as identifying peak times for outpatient visits or evaluating the effectiveness of treatment schedules over time.
- *Degenerate dimension*: Because degenerate dimensions deal with dimensions that don't have associated tables, for example outpatient visit IDs. In healthcare delivery, they can help in uniquely identifying each patient visit without necessarily requiring a separate dimension table, streamlining data storage and retrieval processes.
- *Role-Playing dimension*: This pattern allows a single dimension to play multiple roles. For instance, the date dimension can represent appointment date, follow-up date, and treatment start date. It reduces redundancy and simplifies the data model, aiding in efficient reporting and analysis.
- *Junk dimension*: Junk dimensions combines several low-cardinality attributes, such as patient consent types or visit reasons, into a single dimension. In doing this, clutter in the data model is reduced making it easier to manage and query the attributes without overwhelming the schema.
- *Outrigger Dimension*: In working with outrigger dimensions, secondary dimensions are linked to primary dimensions such as linking patient data to socio-economic status. This enriches the primary data as it provides deeper insights into how socio-economic factors affect healthcare delivery and outcomes in low-resource communities.

Fact tables, contain data about different facts and measures. There are at least four fact table related dimensions that can be adopted.

- *Transactional Fact Table*: Records individual patient visits, capturing details like date, time, and services provided during each outpatient encounter.
- *Periodic Snapshot Fact Table*: Stores aggregated data over a specific time period, useful for tracking trends and changes in outpatient activities, such as monthly or quarterly summaries of patient visits.
- *Accumulating Snapshot Fact Table*: Monitors the evolution of a process, such as the stages a patient undergoes during treatment, providing a comprehensive view of the patient's journey over time.

7 of 19

As exemplified by the above highlighted points, fact and dimension related patterns ensure organized, scalable and insightful data management, which is very crucial for improving outpatient care in resource-constrained settings. Their implementation at health facility databases will facilitate detailed analysis of outpatient activities. For example, the periodic and accumulating snapshot fact tables would link to time dimensions, aiding in trend analysis and patient journey tracking. This structured approach to fact and dimension tables implementation ensures efficient management and analysis of outpatient medical data in the health center, supporting informed decision-making and improving healthcare delivery to the community.

2.3. Enabling Tools

Over the past few decades, mobile devices have revolutionized people's lives. This has included both the way people receive and extend medical care. On the other hand, the penetration rate of smartphones into the populace has led to a proportionate increase in mobile health applications, including mobile health (mHealth) applications. Their potential to provide low-cost and around-the-clock access to their users on a global scale makes them good candidates for resource-constrained communities. Mobile apps have been used for wellness management, disease management, self-diagnosis, medication reminders, electronic patient portals, and in physical medicine and rehabilitation [26]. This work proposes two apps: 1) a navigation tool, and 2) a treatment adherence reminder app. The treatment reminder does not recommend personal medication but rather provides reminders about already prescribed dosages and scheduled appointments at the health facility. The navigation tool aids outpatients who may not yet be familiar with health facility procedures and navigation at the facility.

Multidimensional data modeling offers several advantages with regard to developing user apps: By structuring data into dimensions such as patient demographics, appointment schedules, medical conditions, and treatment plans, healthcare providers can easily analyze patient trends and behaviors. This allows a mobile app to provide personalized directions to outpatients based on their specific needs and preferences, enhancing the overall patient experience. Additionally, using snapshot fact tables within the multidimensional model enables healthcare professionals to track treatment adherence over time accurately. By capturing data at regular intervals, snapshot fact tables provide insights into patient compliance with prescribed medications or therapies, allowing for timely interventions and adjustments to treatment plans as needed to improve patient outcomes. Overall, multidimensional data modeling facilitates efficient data analysis and decision-making, contributing to better patient care by the health facility professionals.

3. Materials and Methods

The main objective of this research was to highlight the application of data warehousing practices in healthcare delivery and come up with a framework that incorporates patients with limited resources and from relatively low-income communities for effective healthcare delivery. Related data warehousing fundamentals are discussed. The work goes ahead to provide a comparison between two main data warehousing architectures with regard to key requirements of data structure, quality, query performance, security, analytics, integration, and governance. Effective extension of healthcare services to rural communities is only possible with accurate data from all the key stakeholders. This includes poor patients who may not be able to afford treatment beyond outpatient services. In addition, effective data-driven policy formulation is highly dependent on efficient data aggregation from multiple stakeholders.

The study was organized around the following subject areas: data warehousing, healthcare delivery, data lakes, outpatients and electronic health records. And on the basis of these, phrases were created including cloud computing, data lakes, data warehousing architecture and electronic health records. It involved a literature review of peer-reviewed papers published over the past two decades. although a few older material were also looked at to understand the evolution of data warehousing.

These topics were researched to gather literature in the field. Efforts were directed towards major digital libraries that included: MDPI, IEEE Explore, ScienceDirect, Springer Link, ACM, and Google Scholar. This review was aimed at understanding the application of data warehousing concepts and technologies in healthcare delivery, highlighting the strengths of data warehousing in healthcare.

To carry out the needs assessment, selected health center IVs(HCIVs) in both urban and rural areas were visited. HCIVs were selected because they posses basic the basic healthcare infrastructure compared to Health center IIs (HCIIs) and Health Center Is (HCIs). They are also more easily accessible by the the local communities compared to the regional, referral and national hospitals that are more scattered. The survey was purposive in nature to ensure that needs of both the rural and urban facilities were assessed.

To improve patient experience and enhance service delivery, particularly for outpatients who may not be familiar with health facility navigation and procedures, enabling tools were proposed. These tools are intended to enhance patient treatment adherence and simplify patient navigation at the facility. The navigation tools reduce long queues and related delays at the health facilities, while the reminder apps improve treatment adherence. The combined insights from literature and survey informed the design and development of healthcare delivery framework and the enabling tools.

4. Results

4.1. Electronic Health Records (EHR)

An Electronic Health Record (EHR) provides an electronic version of a patient's medical history, maintained by the health provider over time. This data includes key administrative clinical data relevant to both the patient and care giver. Electronic health records automate access to information and can streamline health facility workflow. In addition, EHR support other care-related activities including evidence-based decision support, policy formulation and health service quality management.

The Outpatient Department (OPD) in the context of healthcare, refers to the department, where patients receive medical treatment or consultation without being admitted to the hospital overnight. Because of the extra bills associated with admission at medical facilities of Sub-Saharan Africa, the OPD is common place. And like other medical records, data from the OPD is crucial for maintaining a comprehensive and organized healthcare data necessary for both regional and national policy formulation. This goes a long way in facilitating efficient patient care and guidance on the overall delivery of health services.

Table 2 below presents a list of data commonly collected by the OPDs of the various visited healthcare facilities. The list may not necessarily be exhaustive, but presents a generality of the facilities.

Table 2. Sample categorical and specific health facility data.

Serial No.	Categorical data	Details
1	Patient Registration Form	Personal details (name, age, gender) Contact information (address, phone number) Identification details (national ID, passport) Next of kin information
2	Medical History Form	Previous illnesses and surgeries Allergies and medication history Family medical history Immunization records
3	Outpatient Visit Record	Date and time of visit Chief complaints or reason for the visit Vital signs (blood pressure, temperature) Preliminary diagnosis and treatment provided
4	Prescription and Medication Record	Medication details (name, dosage, frequency) Prescribing healthcare provider Instructions for use Duration of the prescription
5	Laboratory and Diagnostic Test Results	Date and type of tests conducted Results of tests, imaging, or other diagnostics Interpretation of results Follow-up recommendations
6	Follow-up Appointment Record	Date and time of scheduled follow-up Purpose of the follow-up Recommended actions or treatments Any additional instructions for the patient
7	Billing and Payment Records	Details of services rendered Cost of services Payment method and amount paid Outstanding balances if any
8	Consent Forms	Signed consent for treatments or procedures Acknowledgment of privacy practices Consent for release of medical information Other relevant consent forms
9	Referral and Consultation Forms	Details of referred specialists Reason for referral Consultation notes and recommendations Follow-up plans post-consultation
10	Quality Improvement and Feedback Forms	Patient feedback on services received Suggestions for improvement Incidents or complaints Staff performance assessments

4.2. The Healthcare Service Delivery Framework

Health facilities in sub-Saharan Africa are structured into regional, and national units all under the supervision of a statutory body. For example, Uganda's health system comprises a tiered structure. At the top is the National Referral Hospital, providing specialized care and training. It's supported by 14 Regional Referral Hospitals across the country. Health centers serve primary care needs, offering basic services. The National Medical Stores ensures efficient procurement and distribution of medical supplies nationwide.

Figure 2 above is a proposed robust data warehousing framework for health data integration from multiple disparate sources including Ministry of Health (MoH), National Medical Stores (NMS), and Regional Referral Hospitals (RRH). The framework incorporates the outpatients into the healthcare realm. Data is obtained from multiple heterogeneous sources where they have been stored as flat files or operational systems. At the staging area, data is transformed to prepare them for loading into the data warehouse.
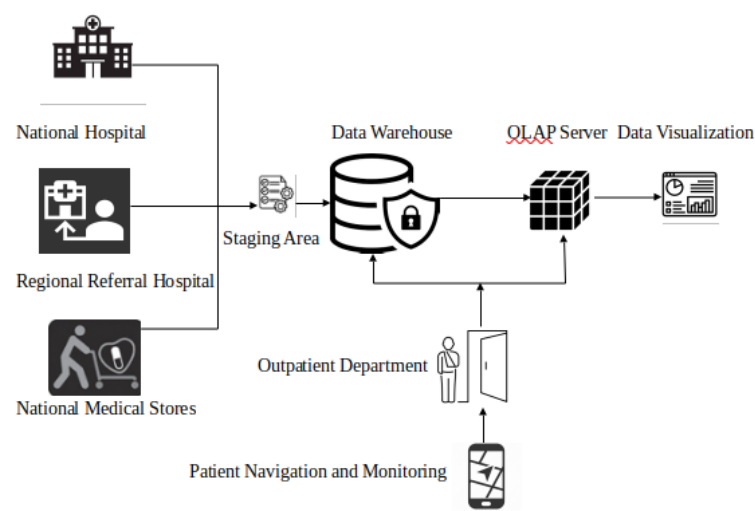

Figure 2. Healthcare Delivery Framework

The warehouse comprises metadata, raw data, and summaries. Metadata provides context and structure to raw data, describing its characteristics and origin. Metadata is stored in an information directory or metadata registry. Raw data contains unprocessed information. Summaries distill raw data into concise representations, offering insights and facilitating understanding without needing to analyze the entire dataset.

Online Analytical Processing (OLAP) is a technology that organizes large business databases and supports complex analytical queries without negatively affecting transactional systems. It employs both predefined and customized queries to enable efficient access to data warehouses. In healthcare delivery, it facilitates dynamic analysis and reporting for informed decision-making and improved patient care.

The operational data stores within the data warehouse infrastructure at local medical facilities are designed to facilitate seamless authorized user access, encompassing both local and remote stakeholders. Remote accessibility extends to regional or national healthcare institutions, inclusive of entities vested with statutory oversight in medical data management or regulation of medical inventory. Also, within the facility, patients may necessitate aid in navigating the healthcare environment. On the other hand, medical staff encounter challenges in ensuring treatment adherence by patients. A mobile application is devised to furnish navigation support and facilitate monitoring of treatment adherence. The mobile devices are enabled with both navigation and notification features. leading to enhanced engagement and management.

4.3. The Healthcare Data Model

The multi-dimensional data model is a star schema for monitoring patient data. MySQL implements this schema by creating fact tables for patient encounters and dimension tables for attributes like date, patient, and diagnosis. The MySQL Workbench was used to facilitate schema design, while MySQL query browser enabled querying and analysis of the data.

Below are the fact table specifics of the designed schema: VisitID (Primary Key) - Unique identifier for each visit, PatientID (Foreign Key referencing Patient dimension) - Foreign key referencing patient dimension, linking the fact to patient details, FacilityID (Foreign Key referencing HealthFacility

dimension) - Foreign key referencing a HealthFacility dimension, linking the fact to facility details, VisitDate - Date and time of the outpatient visit, RegistrationStatus - Indicates whether the patient completed the registration process, VisitRecordStatus - Indicates the status of the visit record, PrescriptionRecordStatus - Indicates the status of the prescription record, MedicationRecordStatus - Indicates the status of the medication record, LabTestResultsStatus - Indicates the status of the laboratory test results, DiagnosticTestResultsStatus - Indicates the status of diagnostic test results, FollowUpAppointmentStatus - Indicates the status of the follow-up appointment record, BillingRecordStatus - Indicates the status of the billing record, AppointmentRecordStatus - Indicates the status of the appointment record. The developed star schema can be seen in Appendix A1.

4.4. Enabling Tools

In the context of mobile apps for outpatient navigation at the medical facility and monitoring of treatment adherence, multidimensional data modeling offers several advantages: First is that by structuring data into dimensions such as patient demographics, appointment schedules, medical conditions, and treatment plans, healthcare providers can easily analyze patient trends and behaviors. This allows a mobile app to provide personalized directions to outpatients based on their specific needs and preferences, enhancing the overall patient experience. Additionally, using snapshot fact tables within the multidimensional model enables healthcare professionals to track treatment adherence over time accurately. By capturing data at regular intervals, snapshot fact tables provide insights into patient compliance with prescribed medications or therapies, allowing for timely interventions and adjustments to treatment plans as needed to improve patient outcomes. Overall, multidimensional data modeling facilitates efficient data analysis and decision-making, contributing to better patient care by the health facility.

4.4.1. OPD Messaging App

A messaging app is designed to enhance treatment adherence by sending personalized treatment plans and timely reminders to patients or their caretakers. It delivers regular medication schedules, dosage instructions, and lifestyle recommendations directly to patients' mobile devices. Additionally, the app sends automated reminders about upcoming hospital visits and due dates for medical check-ups, ensuring patients stay informed and engaged with their treatment regimens. By providing clear and consistent communication, the app aims to improve health outcomes, reduce missed appointments, and support patients in managing their health effectively. To send reminders to outpatients regarding their due dates at a medical facility and medication adherence, a python script that leverages a messaging service such as Twilio was used to send short messages to patients or their caretakers. For functionality of the python script, the following were done:

- Setting up a list of patients including their specific details that include but may not be limited to: phone number, appointment date, and medication schedule.
- Using the schedule library, reminders for appointments, and medication adherence are scheduled.
- To send reminders to patients, the twilio messaging service is used.
- A function was written to handle the sending of scheduled reminders to patients.

The reminder app can be found in Appendix A2.

4.4.2. The OPD Navigation App

In modern healthcare, efficient navigation within multi-unit health facilities can greatly enhance the outpatient experience. Data warehousing plays a crucial role in enabling this seamless navigation by centralizing and organizing diverse datasets from various departments such as the laboratory, pharmacy, cash office, security, etc.

Outpatients are empowered to navigate these facilities effortlessly through a mobile app that leverages geo-fencing technology. As soon as a patient specifies the next point of service, the app analyzes the available data, including unit locations and operational statuses, to calculate the optimal

route. The app then provides real-time directions to guide the patient via the shortest path to their chosen destination, minimizing wait times and ensuring a smooth flow through the facility.

This advanced navigation system is supported by databases that contain comprehensive information about each unit, including current capacity, services offered, and geographical data. By integrating these databases, the app can dynamically adjust the patient's route based on factors such as traffic, availability, and urgency of services.

Once the patient completes all necessary services, the facility's security database is automatically updated to reflect the patient's journey and completion status. This data-driven approach ensures a secure and organized process, from the patient's arrival to their departure.

Moreover, data warehousing allows the health facility to continuously improve its navigation system by analyzing usage patterns and patient feedback. This ongoing refinement leads to better patient outcomes, increased satisfaction, and streamlined operations. In summary, leveraging data warehousing for outpatient navigation within multi-unit health facilities offers a holistic approach to healthcare delivery. By providing real-time, personalized guidance and continuously improving services, health facilities can ensure a safer, more efficient, and more pleasant experience for patients. This innovative use of data is a step forward in revolutionizing healthcare delivery.

Geofencing technology was used in the medical facility to enhance patient navigation by creating virtual boundaries around the various departments and providing real-time notifications to guide patients. The system works as follows:

- *Geofencing Setup*. Virtual geofences (geographic boundaries) are created around each of the departments in the medical facility.
- *Patient Tracking*. GPS or Bluetooth-based tracking is used to monitor patients locations within the facility. The patient's location can be tracked through a mobile app installed on their smartphone.
- *Notifications*. When the patient gets to a specific geofence, they receive a notification on their mobile app with information relevant to that department, such as *"This is the pharmacy. Move to the reception"*. Notifications can also include details such as waiting times, instructions, or other important information.
- *Shortest Route Guidance*. Using indoor mapping and routing algorithms, the app provides the patient with the shortest route to a specified department. The mobile app also offers turn-by-turn directions and updates based on the patient's current position.
- *System Integration*. The geofencing system is integrated into the medical facility's database to keep track of appointments and patient schedules. It can also be configured to provide personalized notifications based on a patient's schedule.
- *Privacy and Security*. To ensure that patient location data is collected and stored securely, data anonymization and encryption techniques can be used.

The code for the navigation app can be seen in Appendix A3.

4.5. Discussion

The implementation of data warehousing in healthcare delivery, particularly in low-resource communities, presents a great opportunity to improve patient outcomes and positively streamline the overall healthcare process. This study proposes a data warehousing framework, along with a navigation and reminder app especially tailored for patients in low-resource communities.

In sub-Saharan Africa and many developing economies, healthcare facilities often struggle with fragmented and inaccessible patient data [27,28]. This is attributed to inefficient health structures and poor infrastructure. This has led to sub-optimal care and, hence, high mortality rates. This is where data warehousing, which involves the aggregation, storage, and management of large volumes of data from diverse sources, can be particularly important.

A well-structured data framework provides a comprehensive view of the data. In this study, a healthcare data framework based on data warehousing is proposed. We believe the proposed framework will facilitate comprehensive and real-time access to healthcare data. This will enable the

formulation data-driven policies, which will in turn improve clinical decision-making and resource allocation.

The results clearly show that the framework will significantly enhance data accessibility and interoperability among different healthcare stakeholders at different levels of service delivery. This is especially important when patients cannot keep their records and may find themselves visiting multiple facilities that lack integrated systems. A unified data repository will allow healthcare providers to have access to patient histories, leading to more accurate diagnoses and treatment plans. This is also expected to aid health authorities in planning targeted interventions.

Another critical factor in healthcare is non-adherence to treatment and follow-up on appointments. This can be attributed to patients lacking adequate education and the logistics required to travel long distances to reach health facilities. The proposed navigation and reminder apps developed as part of this study, to a good extent, address these concerns. The app mitigates these issues by reminding patients with clear, accessible information about their appointments and treatment schedules. The reminder feature thus helps reduce missed appointments. The navigation app on the other hand provides practical assistance in reaching and navigating through healthcare facilities. And because the developed apps are integrated into the databases, current and accurate data is used.

Leveraging data warehousing in healthcare delivery offers significant benefits, especially in low-income communities where data accessibility and patient adherence are not commonplace. The proposed framework and apps can significantly enhance the efficiency and effectiveness of healthcare delivery, ultimately leading to better health outcomes.

5. Conclusions

The research on data warehousing in healthcare significantly advances understanding across academic, community, and government sectors. Academically, it elucidates key data warehousing concepts tailored to healthcare needs, introduces a novel multidimensional data model, and outlines a practical implementation framework. These contributions enrich existing literature and pave the way for further inquiries into efficient healthcare data management. Within the community, the research yields substantial benefits by enhancing patient experiences, enabling cost-effective service delivery, and improving patient management, particularly in resource-constrained environments such as sub-Saharan Africa. The integration of electronic health records supports evidence-based decision-making and informed policy formulation, ultimately fostering improved healthcare outcomes. And to governments, the research offers critical insights into optimizing healthcare systems, especially in low-resource regions. Utilizing mobile applications for patient navigation and treatment reminders, the study suggests strategies to enhance healthcare accessibility and adherence, thereby improving public health outcomes.

The successful deployment of data warehousing technologies in low-resource settings are however influenced by several other factors that call for more attention. These include:

- Data warehousing necessitate the availability of fairly efficient infrastructure, availability of healthcare personnel, and patient acceptance and engagement with the technology. Future research focusing on longitudinal studies to assess the long-term impact of data warehousing and patient engagement apps on health outcomes.
- In addition, notable challenges includes obtaining informed consent for integrated database development, particularly when demonstrating their potential is complex. Additionally, despite ongoing research into patient data security within data warehousing, significant gaps still remain [29,30].
- Additionally, exploring partnerships with local governments and non-governmental organizations could facilitate the scalability and sustainability of these initiatives.

Addressing these challenges is essential for realizing the full potential of data warehousing in transforming healthcare delivery and policy. This study lays a robust foundation, highlighting areas for future research and practical implementation to drive improvements in global healthcare systems.

Overall, this research bridges academia, community, and government efforts by offering practical solutions for enhancing healthcare delivery, particularly in resource-constrained environments. Its findings have implications for policy-making, resource allocation, and technological innovation in healthcare systems globally.

Abbreviations

The following abbreviations are used in this manuscript:

|	
BI	Business Intelligence
EHR	Electronic Health Records
ETL	Extraction, Transformation and Loading
HCI	Health Center I
HCII	Health Center 1I
HCIII	Health Center 1II
HCIV	Health Center 1V
MoH	Ministry of Health
NRH	National Referral Hospital
NMS	National Medical Stores
NPCs	Non-Physician Clinicians
OPD	Out Patient Department
OLAP	Online Analytical Processing
RRH	Regional Referral Hospital

Appendix A. Data Schema
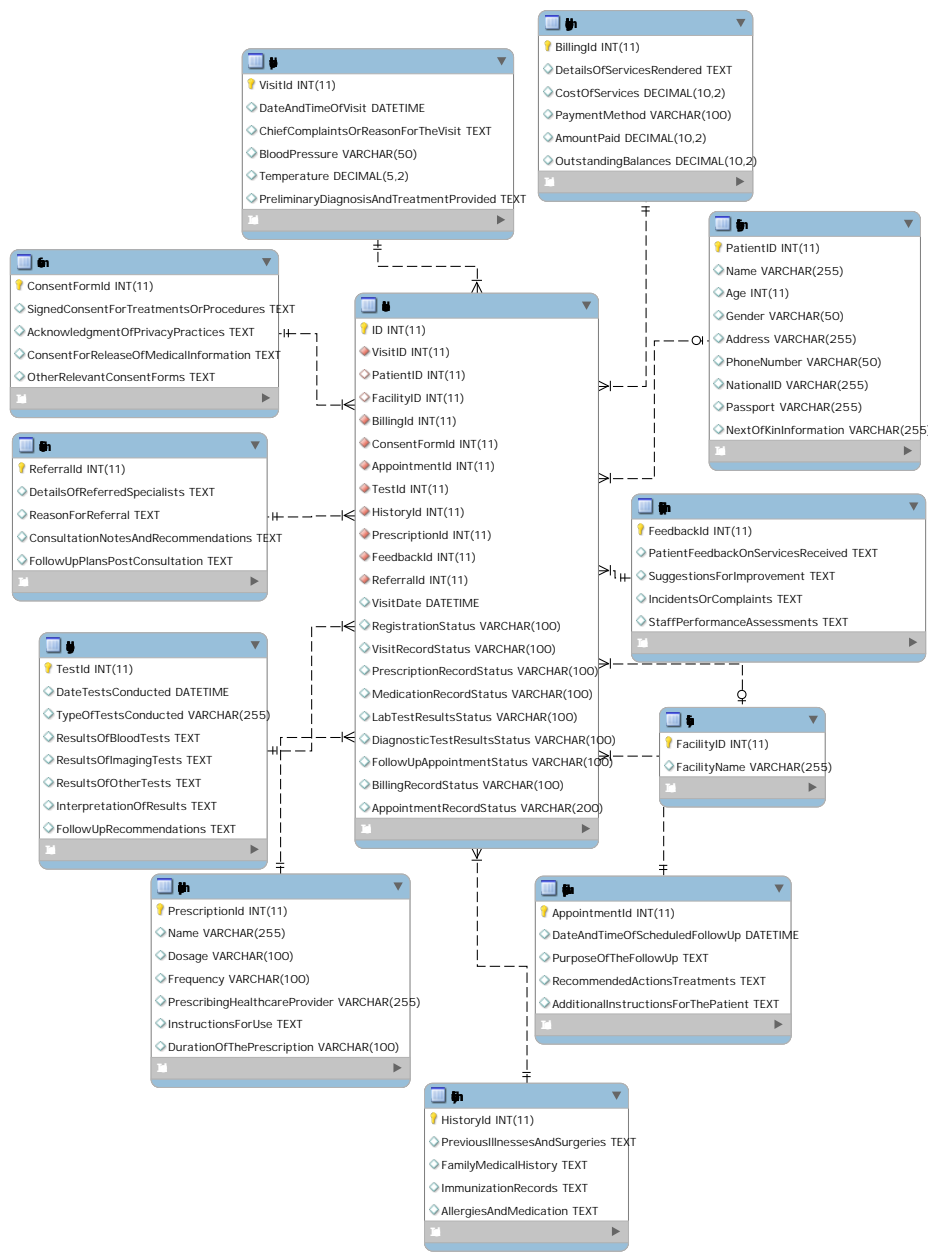
Figure A1. Healthcare delivery data schema

Appendix B. Reminder App.

File: /media/aggrey/981E91861E915DD...ery/Reminder App - Aggrey Okori Page 1 of 2

```
#Install the required libraries:

pip install schedule twilio
#Reminder App:
#Importing necessary libraries
import schedule
import time
from twilio.rest import Client

# Configuration of twilio
account_sid = 'your_account_sid'
auth_token = 'your_auth_token'
from_phone_number = 'your_twilio_phone_number'

# Initializing twilio client
client = Client(account_sid, auth_token)

# Specifying list of patients together with their respective phone numbers, appointment
dates, and medication schedules
patients = [
    {
        'name': 'John Doe',
        'phone_number': '+256700700',
        'appointment_date': '2024-04-30',
        'medication_schedule': [
            {'name': 'Aspirin', 'time': '09:00'},
            {'name': 'Ibuprofen', 'time': '14:00'},
            {'name': 'Paracetamol', 'time': '20:00'},
        ]
    },
    # More patients are added here
]

# Defining the function to send SMS reminders
def send_sms_reminder(phone_number, message):
    message = client.messages.create(
        body=message,
        from_=from_phone_number,
        to=phone_number
    )
    print(f"Sent reminder to {phone_number}: {message.body}")

# Function that sends appointment reminders
def send_appointment_reminder():
    today = time.strftime('%Y-%m-%d')
    for patient in patients:
        if patient['appointment_date'] == today:
            message = f"Hello {patient['name']}, you have an appointment today at the
medical facility."
            send_sms_reminder(patient['phone_number'], message)

# Function to send medication reminders
def send_medication_reminder():
    current_time = time.strftime('%H:%M')
    for patient in patients:
        for medication in patient['medication_schedule']:
            if medication['time'] == current_time:
                message = f"Hello {patient['name']}, please take your medication:
{medication['name']}."
                send_sms_reminder(patient['phone_number'], message)

# Schedule reminders
schedule.every().day.at("08:00").do(send_appointment_reminder)  # Appointment reminders
schedule.every().day.at("09:00").do(send_medication_reminder)  # Medication reminders
schedule.every().day.at("14:00").do(send_medication_reminder)
schedule.every().day.at("20:00").do(send_medication_reminder)
```

Figure A2. Healthcare Reminder App.

Appendix C. Navigation App

File: /media/aggrey/981E91861E915DD…y/Navigation App - Aggrey Ob… Page 1 of 2

```python
#Implementinge geofencing system using a Python-based mobile application:

python
#Importing libraries
from geopy.distance import geodesic
from datetime import datetime
import geopy
import json

# Defining the geofences for each department/Unit
geofences = {
    'X-ray facility': {
        'latitude': 40.1234,
        'longitude': -74.1234,
        'radius': 50,  # in meters
    },
    'laboratory': {
        'latitude': 40.1235,
        'longitude': -74.1235,
        'radius': 50,  # in meters
    },
    'pharmacy': {
        'latitude': 40.1236,
        'longitude': -74.1236,
        'radius': 50,  # in meters
    },
    'payment office': {
        'latitude': 40.1237,
        'longitude': -74.1237,
        'radius': 50,  # in meters
    },
    'security office': {
        'latitude': 40.1238,
        'longitude': -74.1238,
        'radius': 50,  # in meters
    }
}

# Checking if a patient is within a geofence
def is_within_geofence(patient_location, geofence):
    distance = geodesic(patient_location, (geofence['latitude'],
geofence['longitude'])).meters
    return distance <= geofence['radius']

# Managing patient navigation
def handle_patient_navigation(patient_location, destination_department):
    # Determine the geofence for the specified department
    geofence = geofences[destination_department]

    # Determine if the patient is within the geofence
    if is_within_geofence(patient_location, geofence):
        # Notify patient with relevant information
        notification = f"You are now at the {destination_department}. Please proceed as
instructed."
        print(notification)
        # Perform any alternative actions such as update the system or provide directions
    else:
        # Calculate the shortest route to the destination department
        # Implemented using the indoor routing algorithm
        shortest_route = calculate_shortest_route(patient_location, geofence)
        print(f"Proceed to the {destination_department}. Follow this route:
{shortest_route}")

# Function to calculate the shortest route (this function should be implemented)
def calculate_shortest_route(patient_location, geofence):
    # Implement indoor routing algorithm here
```

Figure A3. Healthcare Navigation App.

References

1. Ekholuenetale M, Wegbom AI, Tudeme G, Onikan A. Household factors associated with infant and under-five mortality in sub-Saharan Africa countries. *IJCCEP* **2020** *14*, 1–5.
2. Bray F, Parkin DM, Gnangnon F, Tshisimogo G, Peko JF, Adoubi I, Assefa M, Bojang L, Awuah B, Koulibaly M, Buziba N. Cancer in sub-Saharan Africa in 2020: a review of current estimates of the national burden, data gaps, and future needs. *Lancet* **2022** *23(6)*, 719–28.

3. Agwanda B, Dagba G, Opoku P, Amankwa MO, Nyadera IN. Sub-Sahara Africa and the COVID-19 pandemic: reflecting on challenges and recovery opportunities. *JDS*. **2021** *37(4)*, 502–24.

4. Kiross GT, Chojenta C, Barker D, Loxton D. The effects of health expenditure on infant mortality in sub-Saharan Africa: evidence from panel data analysis. *Health Econ. Rev.* **2020** *10* 1–9.

5. Chewe M, Hangoma P. Drivers of health in sub-Saharan Africa: a dynamic panel analysis. *Health Policy OPEN*. **2020** *1* 100013.

6. Anyangwe SC, Mtonga C. Inequities in the global health workforce: the greatest impediment to health in sub-Saharan Africa. *IJERPH*. **2007** *4(2)*, 93–100.

7. Langnel Z, Buracom P. Governance, health expenditure and infant mortality in sub-Saharan Africa. *Afr. Dev. Rev.* **2020** *32(4)*, 673–85.

8. van Heemskerken P, Broekhuizen H, Gajewski J, Brugha R, Bijlmakers L. Barriers to surgery performed by non-physician clinicians in sub-Saharan Africa—a scoping review. *Hum resour health* **2020** *18*, 1–2.

9. Willcox ML, Peersman W, Daou P, Diakité C, Bajunirwe F, Mubangizi V, Mahmoud EH, Moosa S, Phaladze N, Nkomazana O, Khogali M. Human resources for primary health care in sub-Saharan Africa: progress or stagnation?. *Hum resour health* bf 2015 *13*, 1–1.

10. van Heemskerken P, Broekhuizen H, Gajewski J, Brugha R, Bijlmakers L. Barriers to surgery performed by non-physician clinicians in sub-Saharan Africa—a scoping review. Hum resour health. 2020 Dec;18:1-2.

11. Muñoz L, Mazon JN, Trujillo J. ETL process modeling conceptual for data warehouses: a systematic mapping study. IEEE Lat. Am. Trans. 2011 Jun 16;9(3):358-63.

12. Wyllie D, Davies J. Role of data warehousing in healthcare epidemiology. *JHI)* **2015**; *89(4)*, 267–70.

13. Gavrilov G, Vlahu-Gjorgievska E, Trajkovik V. Healthcare data warehouse system supporting cross-border interoperability. *J. Health Inform*. **2020** *26(2)*, 1321–32.

14. Tang C, Wang H, Liu J, Liu C. A design for AIS data warehouse of traffic flow. In Proceedings of the 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8 Aug 2017; 895–898.

15. Ponniah P. *Data warehousing fundamentals: a comprehensive guide for IT professionals* John Wiley & Sons; New York, USA, 2004; pp. 20–23.

16. El Aissi ME, Benjelloun S, Loukili Y, Lakhrissi Y, Boushaki AE, Chougrad H, Elhaj Ben Ali S. Data lake versus data warehouse architecture: A comparative study. In Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, Singapore, India, 22 July 2021; 201–210.

17. Kimball, R. and Ross, M. *The data warehouse toolkit: the complete guide to dimensional modeling*, 2nd ed., John Wiley & Sons: New York, USA; 6–14.

18. Oliva SZ, Felipe JC. Optimizing public healthcare management through a data warehousing analytical framework. *IFAC-PapersOnLine* **2018**; *51(27)*, 407–12.

19. Soliman MA, Antova L, Sugiyama M, Duller M, Aleyasen A, Mitra G, Abdelhamid E, Morcos M, Gage M, Korablev D, Waas FM. *A framework for emulating database operations in cloud data warehouses*. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, 11 June 2020, 1447-–1461.

20. Hai R, Koutras C, Quix C, Jarke M. Data lakes: A survey of functions and systems. *IEEE Trans. Knowl. Data Eng*. **2023** *35(12)*, 12571–12590

21. Khine PP, Wang ZS. *Data lake: a new ideology in big data era.* In Proceeding of the ITM web of conferences, Beijing, China, 02 February 2018.

22. Kimball, R. and Ross, M. *The data warehouse toolkit: the complete guide to dimensional modeling*, 2nd ed., John Wiley & Sons: New York, USA; 16–26.

23. Bojičić I, Marjanović Z, Turajlić N, Petrović M, Vučković M, Jovanović V. A comparative analysis of data warehouse data models. In Proceeding of the International Conference on Computers Communications and Control (ICCCC), Oradea, Romania 10th May 2016, 151–159.

24. Kimball, R. and Ross, M. *The data warehouse toolkit: the complete guide to dimensional modeling*, 3rd ed., John Wiley & Sons: New Indianapolis, Indiana.

25. Viqarunnisa P, Laksmiwati H, Azizah FN. Generic data model pattern for data warehouse. In Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, Bandung, Indonesia, 17 Jul 2011, 1–8.

26. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. *PM&R*. **2017** *9(5)*, 106–15.

27. Murphy SN, Avillach P, Bellazzi R, Phillips L, Gabetta M, Eran A, McDuffie MT, Kohane IS. Combining clinical and genomics queries using i2b2–Three methods. *PloS one.* **2017** *12(4)*, 5317760.
28. Naeem M, Jamal T, Diaz-Martinez J, Butt SA, Montesano N, Tariq MI, De-la-Hoz-Franco E, De-La-Hoz-Valdiris E. Trends and future perspective challenges in big data. In Advances in Intelligent Data Analysis and Applications: In Proceedings of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, Arad, Romania, 15–18 October 2019, 309–325.
29. Shahid A, Nguyen TA, Kechadi MT. Big data warehouse for healthcare-sensitive data applications. *Sensors* **2021** *21(7)*, 2353.
30. Berndt, D.J., Fisher, J.W., Hevner, A.R. and Studnicki, J. Healthcare data warehousing and quality assurance. *Computer*, **2001** *34(12)*, 56–65.