Homework / Lab 1

Proposing a Data Warehouse

MGS 6577LEC F1S: Cloud Data Warehousing & Data Engn (19774 Fall24)

Submitted By

Anjali Pandey

Contents

Driving Business Growth Through Data: Business Data Warehouse Proposal for ABC Corp.

In today's world healthcare and its data are evolving rapidly, data became a powerful tool for driving better decision making and improving member services. ABC Corp., a health insurance company in United States is facing various challenges in managing and analysing data. There are certain areas of the business that are using spreadsheets, manual queries for the day-to-day operations, leading to inefficiencies in reporting and difficult in accessing historical data and obstacles in adopting advanced technologies like AI and ML.

This proposal outlines the need for centralized data warehouse that will streamline data management, improve reporting capabilities, reduce latency and provide a unified data source for the business. Data warehousing will enable ABC Corp. to address the issue of health insurance database system. A healthcare data warehouse (HDW) is a main centralized warehouse that collects and stores organized, useful healthcare information from various sources such as ERP, HER, radiology, labs etc. The main component of any healthcare platform includes connectivity layer, data analytics, enrichment layer and reporting features. The refined data is stored in the Data Warehouse, where Business Intelligence reporting is accessed. This helps insurers to make informed decisions and optimize operations, with specialized applications to meet the organizational needs.

Background-

Every year the healthcare industry is increasing its focus on data analytics. The main idea behind this data analysis is about combining different types of information like patient record, claims data, claims provider, demographic from various sources. This can be significantly challenging and complex because the data comes from different sources, which leads to problems in making sure that the data is accurate, data quality, standardization and security. And there are strict rules in protecting patients' privacy such as HIPAA, making it more challenging to manage and share health care data properly. Currently departments are often using manual methods such as spreadsheets, manual querying separate databases to extract and compile data which led to face various challenges -

1. **Inconsistent data format and Data silos**: Different healthcare systems and organizations use varying coding schemes and data structures making it difficult to combine data seamlessly. This ends up with fragmented information and complex data information and mapping. This leads to incomplete or missing information on data analysis.
2. **Data Cleaning, Standardization and Validation** – These are the essential steps which involves data cleaning before analysis, standardizing medical codes to consistent formats for accurate comparisons and ensuring data accuracy by cross checking.
3. **Privacy and Security Challenges** – protecting sensitive information by converting data into a code that can be accessed only by a specific code. This means that even if patient data is accessed, it will remain inaccessible to unauthorized individuals.
4. **Scalability Issue** – Health care data is huge and require efficient storage and processing capabilities to handle large queries while maintaining optimal query performances.

5. **Queries run manually** – When dealing with large volume of data, it strains resources and introduces risk related to consistency because the reports are queried from operational system. This increases the chances of errors.
6. **Problem in getting report**- When the company wants to make a decision, they often ask different system for different piece of information. And this can take long time, and it's tough to understand whether the information is helpful or not. Because of this the company might not be able to make decisions quickly.
7. **Challenges with Old data** – Historical datasets are archived in cold storage, which is hard to access sometimes. As a result, for a good decision making the companies are not able to predict about the data pattern based on past experiences.
8. **AI and ML Limitation –** The ABC Corp data is messy, and they need their data to be organized to use these technologies for further analysis.

Therefore, from the past people tried to manage their data by using quick fixes like manual reporting and spreadsheet to create custom reports. But these methods took lot of time, efforts and the expected results were not visible. And this lead to show a stronger, more centralized database system, the one that can handle reports more efficiently and support advanced data analysis.

Requirements-

A health care data warehouse should meet several requirements to implement a successful data warehouse including-

1. **Data Integration**- This feature specifically involves the collection of healthcare data from various sources such as EHRs, ERP software, public health databases, sales etc., and organize it in unified format providing a complete information about patient and effective analytics.
2. **Analytics and Reporting** – Healthcare Data warehouses are equipped with powerful analytical tools, such as implementation of ETL processes to extract data from transactional system and convert it into meaningful data and load to our databases and machine learning in healthcare. For reports and dashboards data visualization tools are included like Tableau or Power BI.
3. **Data Security and Compliances** – Data warehouse for healthcare provides security and data protection, encryption, authentication and authorization are required to protect data.
4. **Decision Support –** Decision Support team is required to simplify decision making processes. This includes analytical reports, BI tools, dashboards for visualization. The data scientist and AI specialist are also required to design and implement ML AI models that can help in the prediction of centralized data.
5. **Scalability and Flexibility –** The data warehouse is capable of handling large amounts of data, and it must be scalable. This also includes the estimated timeline of project duration which will include data migration, system integration and staff training.
6. **AI and ML Integration-** For predictive analysis and model training, a strong foundation of AI and ML is to be initiated by providing clean, organized, structured and historical data. This helps in gaining the patters of the data. Additionally, strict adherence to security standards such as HIPAA, DICOM, FHIR plays a key role in data security.

Solutions –

1. **Centralized Data Management/ Data Integrity Complexity –** The data warehouse will gather and consolidate all data from all major sources into one centralized repository. By implementing data exchange methods such as HL7, FHIR and ETL technologies and provides a unified healthcare information.
2. **Data Security and Privacy –** To ensure data security, implementing data encryption techniques, security monitoring systems and some control rules will help us in protecting the patient's data. Some of the health care data security tools are like- Protected Health Information, Industry-standard data encryption, Data residency and retention etc.
3. **Streamline Reporting & Foster Data Driven Culture-** The integration process and smooth implementation of health care data warehouse into your existing healthcare warehouse is important. The data team can analyse the systems to be integrated and if required create necessary API and interface for the effective communication between healthcare data warehouse and other systems. These dashboards will serve as invaluable decision-support tools, empowering stakeholders to make informed choices based on up-to-date information.
4. **Predictive analytics by AI-ML Driven -** With all the relevant data, for further predictive analytics is AI ML model is implemented. These models can then predict the further analysis, customer behaviour, detect fraudulent activity and even recommend personalized insurance product. The data needed for AI is high-quality, well structed and consistent

   AI intersecting data warehousing in several ways, enhancing traditional way of databases and reporting system.

   - AI can use historical data and predict the disease progression and risk factors, helping insurers to take proactive measures. The technology can detect insufficient reimbursement codes and suggest matching fields to it . Such tools are so helpful that hospital prepare insurance claims faster while minimizing errors.
   - Predictive analytics uses various statistical modelling, data mining and machine learning to analyse data from various sources such as electronic health record, patient demographics and previous insurance records.
   - Natural Language Processing (NLP) a form of AI can work with unstructured data very effectively such as unstructured data like doctors, hospital records etc. By feeding all these records into the warehouse the AI can quickly extract information and automatically process claims. This whole helps in reducing the manual work, speeding up the process and ensure faster payouts for customers.
   - AI can recommend personalized plans or services with the help of data stored in warehouse, AI can help in tailoring individual needs by ensuring customer satisfaction and improving retention. It can evaluate patient's previous medical expense, lifestyle habit, and health status to formulate a plan for them in the basis of their need at a price that reflects their personal health profile.
   - AI can also help in freeing up human resources and speeding decision maker by automating the creation of reports, analysing the large set of data and automating data extraction, loading and transformation in ETL pipeline.

- Machine learning algorithms and modelling can actually help us in flagging errors and marks duplicate records, missing information, ensuring data integrity for better data analysis.
- Effectively training AI systems will enhance the protection and security of healthcare data, while also enabling the responsible and efficient use of AI in healthcare. This ensures both data integrity and the delivery of more accurate, fair, and informed healthcare decisions.

Authorization-

A data warehouse will then empower different department in the ABC Corp. to access and analyse their own data, without waiting for the **IT team** to get the information for them by manual querying. In a data driven decision making environment people tend to choose to believe on data and then navigate their action towards decision making. By having access to data warehouse, departments like claims, sales etc can quickly generate their own data, reports, analyse trend and understand pattern will help them to take informed decisions. There are several decision makers for the ABC Corp. **Chief Information Officer (CIO)** – responsible for managing all technical infrastructure of data warehouse and supports the company IT business objectives. **Chief Financial Officer (CFO)** – are responsible for finances of the project, approving the budget of the project. **Chief Executive Officer (CEO)** – responsible for final project approval, in achieving the business goals and alignment of ABC Corp. success.

Stakeholder Engagement-

ABC Corp, after successfully implementing data warehouse require active involvement and collaboration across multiple departments to ensure the alignment and smooth adoption and understanding of new system to everyone in the department.

For this **regular communication** is required, for which bi-weekly status updates or monthly progress meetings should be conducted among stakeholders from each department to ensure transparency and regular updates about the project. **Trainings** will be provided to the department who are heavily involved in data analysis such as claims, sales and finance, interpret report and how to leverage data for decision-making. Before full scale deployment of the project, **testing** of the product is necessary. A **dedicated support team** is required assist users with any technical or operational issues. This team is responsible for the entire product after the deployment.

Conclusion-

ABC Corp. on implementing data warehouse will enhance its ability to generate timely and accurate report by pulling out all the data together at one place generating reports quickly with less prone to errors. A data warehouse will not only collect the data, but also automate the process, saving time and reduces the errors. AI ML provides well organized data, allowing for predictive analytics and automation. This will lead to better customer satisfaction, foster a data driven decision-making and help company quickly adapt to future market challenges.

References –

1. https://arcadia.io/resources/healthcare-data-warehouse
2. What Is a Data Warehouse?
3. Research data warehouse: using electronic health records to conduct population-based observational studies.
4. Big data in healthcare – the promises, challenges and opportunities from a research perspective: A case study with a model database.
5. Using a Data Warehouse in Healthcare: Architecture, Benefits, and Use Cases
6. Data & Analytics Expertise/ Center for Health Research - Kaiser Permanente(kpchr.org)
7. Big Data & You: A Guide to Data Analytics in the Health Insurance Industry
8. Healthcare Data Warehousing [Full Guide]
9. Medical Insights: Predictive Analytics in Healthcare.
10. Exploring the Intersection of Artificial Intelligence and Clinical Healthcare: A Multidisciplinary Review
11. A Beginner's Guide to Leveraging a Healthcare Data Warehouse
12. BHI data warehouse offers many benefits
13. ChatGPT
14. Gemini
15. Quillbot: Your complete writing solution