STUDY OF MULTIMODAL DATA WAREHOUSE ASSURING PRIVACY AND SECURITY ON CARDIOLOGY DATASETS AND THEIR SOURCES

Thantilage, Ranul.[1], **Kechadi, Tahar.**[1], **Le-Khac, Nhien-An.**[1], **Owen, Kathryn.**[2]

[1]University College Dublin, Ireland
[2]Ulster University and Southern Health and Social Care, Northern Ireland
ranul.thantilage@ucdconnect.ie

INTRODUCTION

The healthcare industry has been growing each year. In Europe the healthcare analytics market valued at 2.68 Billion USD in 2020. It is now estimated that by 2025 it will reach 8.68 Billion USD, accounting for a growth of 26.52% during the forecasted period. [1]

At current clinical data is not only used for clinical reasons. The healthcare industry is looking for relationships within records. These can differ from a patients current condition to previous treatment records. [2] On the other hand there are high levels of cost associated with medical treatments. Since the early years integrating data warehousing in healthcare has aided in reducing these costs. [3]

Privacy, it is a major topic discussed today. In the health industry each record of data will be personal information of a patient. Therefore, the architecture of clinical DW should assure privacy in the design phase itself. Any data leaked to unauthorized persons could be catastrophic, for both the patient as well as the clinical organization.

Different regions and countries would have their unique ways of testing and treatment for a medical condition. This would also mean slight differences in the tests carried out. Therefore, the DW architecture should be welcoming to varying sets of data for a similar medical condition.

MATERIALS AND METHODS

Data Warehousing has evolved over the years. They not only focus on analysis and decision making, but also keeps a close focus on "ever-accelerated" updating of science and technology. [4]

The main objective of the study is to evaluate the available data warehouse architectures and choose best suited architecture for clinical data warehousing. Benefits and draw backs of each architecture would be investigated considering requirements such as data privacy, scalability, traceability, and distribution.

Identified Research Gaps:

- Researched as individual components

- Streamlined data transfer, sharing and rapid approvals

- Data distribution

- Privacy-in-design

RESULTS

Current Proposed Approach:

- Distributed big data architecture
- Privacy-in-design: User authentication, Data anonymization, encryption, filtration
- Cross platform standardization
- Security in all stages: data input, processing, transport, storage and analysis
- Cloud based
- Patient-based data ownership management
- Criticality based automated data access control
- Anonymous data only available for research and analysis purposes

Data Categorization:

- Demographic Data
- Clinical History
- Fitness Data
- Treatment and Diagnostic Data
- Medical Test Data
- Pharmaceuticals

DISCUSSION

Clinical data warehousing has key requirements. Generally clinical data comes in form of both structured and unstructured data which is mostly distributed. Therefore, it is best suggested to adopt a big data-oriented architecture for a clinical data warehouse.

A key factor noticed during the research is the fact that privacy and security is not much talked about when discussing about DW architectures. This concludes that privacy and security features are added later on, to most DWs. As clinical DWs deals with heaps of confidential healthcare data it will be further researched in the future to propose a data warehouse architecture with 'security in design' as a major component of the DW architecture.

REFERENCES

[1] "Europe Healthcare Analytics Market Size, Growth | 2020 to 2025 | UK, Germany, France, Span, Italy." [Online]. Available: https://www.marketdataforecast.com/market-reports/europe-healthcare-analytics-market. [Accessed: 10-Oct-2020].

[2] A. Khalaf Hamoud, A. Salah Hashim, and W. Akeel Awadh, "Clinical Data Warehouse a

Review," *Iraqi J. Comput. Informatics*, vol. 44, no. 2, 2018, doi: 10.25195/2017/4424.

[3] E. F. Ewen, C. E. Medsker, L. E. Dusterhoft, K. Levan-Shultz, J. L. Smith, and M. A. Gottschall, "Data warehousing in an integrated health system; building the business case," *Dol. Proc. ACM Int. Work. Data Warehous. Ol.*, vol. Part F1292, no. October, pp. 47–53, 1998, doi: 10.1145/294260.294271.

[4] Q. Yang, M. Ge, and M. Helfert, "Analysis of data warehouse architectures: Modeling and classification," *ICEIS 2019 - Proc. 21st Int. Conf. Enterp. Inf. Syst.*, vol. 2, pp. 604–611, 2019, doi: 10.5220/0007728006040611.