

Postgraduate Seminar Series

Venue: Graduate Seminar Room

Date & Time: September 12, 2026 at 2:30 PM

Speaker Information

Sultan Mahmud (Std. No. 0421052097) is a full-time M.Sc. student in the Department of CSE at BUET. He completed his undergraduate studies from Noakhali Science and Technology University in 2020. His research interests lie in the field(s) of Machine Learning, Deep Learning, Natural Language Processing, and Database Systems. He is currently doing his postgraduate thesis under the supervision of Dr. Abu Sayed Md. Latiful Hoque. In this talk, he will present his ongoing research and discuss the key aspects and findings of his current work.



A Comprehensive Schema Analysis for the Design of Privacy-Preserving Clinical Data Warehouse

Clinical Data Warehouses (CDWs) are increasingly being used in healthcare to bring together and analyze large amounts of clinical information collected from different sources, such as hospitals, diagnostic centres, and laboratories. By using Online Analytical Processing (OLAP), these systems allow healthcare professionals and researchers to examine clinical data from different perspectives and at different levels of detail. This can be useful for identifying disease patterns, studying treatment outcomes, and supporting clinical decision-making. However, the usefulness and efficiency of a CDW largely depend on how its underlying schema is designed. Clinical data are often complex, diverse, and continuously changing, which makes schema design particularly challenging. Issues such as organizing dimensional hierarchies, deciding the appropriate level of normalization, maintaining good query performance, integrating data from heterogeneous sources, and preserving accurate temporal information for longitudinal analysis all need to be carefully considered. At the same time, clinical data contain highly sensitive patient information that must be protected. Therefore, a CDW must strike a careful balance between providing efficient analytical capabilities and maintaining patient privacy while complying with regulatory requirements such as HIPAA.

This study presents a comprehensive schema analysis and optimization approach for the design of a privacy-preserving Clinical Data Warehouse. Clinical data were collected from multiple healthcare sources, including hospitals, diagnostic centres, and laboratories, and integrated into a centralized warehouse environment using a structured Extraction, Transformation, and Loading (ETL) process. Three dimensional schema models—star schema, snowflake schema, and fact-constellation schema—were designed and implemented to support multidimensional clinical analysis. A privacy-preserving mechanism based on the KSRL algorithm was applied to protect sensitive patient information during data integration. The performance of the proposed schemas was evaluated using real datasets by measuring storage utilization of fact and dimension tables and the execution efficiency of OLAP operations such as roll-up, drill-down, slice, and dice. The experimental results demonstrate that optimized schema design and appropriate OLAP optimization techniques can significantly improve query performance, reduce storage overhead, and enhance the overall efficiency of clinical data warehouse systems.