****

**Emerging Non-Traditional Database System**

**PROJECT REPORT ON**

**Library Datawarehouse System**

**SUBMITTED BY:**

Parneet Kaur (110124612)

Pooja Vishwakarma (110124478)

**PROFESSOR:**

Dr. C.I. Ezeife

School of Computer Science,

University of Windsor

**Abstract**

Data warehousing encompasses methodologies and techniques aimed at extracting, transforming, and loading data from various sources to facilitate in-depth analysis. A prevalent approach in data warehousing involves multidimensional modeling, utilizing fact and dimension tables for OLAP (Online Analytical Processing). Facts represent specific values of interest, while dimensions impart meaning to these facts, organized into hierarchies with multiple levels.

Libraries grapple with the challenge of managing vast amount of data as count of books in libraries increases tremendously, managing membership details, borrowers’ data, maintain the history which books are issued more, when member issued the book, rearranging them based on order, checking count of books available of an author requires technology to handle. This project report concentrates on developing a system that integrates two branches that is two source databases, constructed on MySQL workbench focusing on three modules borrowers, books and members. The application constructs a data warehouse to store historical data, employing fact and dimensions table to enable the retrieval of essential information through database views and a user-friendly GUI.

The database schema is generated using a web page provided by the mining tool, with data types chosen based on attribute value analysis. Input to the data warehouse comprises source databases of books, members and borrowers’ modules, along with their relationship tables. The Extract, Transform, Load (ETL) process converts these inputs into facts and dimension tables for the data warehouse. Ultimately, query results are obtained based on views presented on the web page.

**Introduction**

**Database**

A database comprises structured collections of detailed records or data, typically stored within a computer system. The management of a database is typically overseen by a Database Management System (DBMS). Together, the data and the DBMS form a database system, often simply referred to as a database, which includes associated applications. In commonly used databases today, data is organized into tables with rows and columns, facilitating efficient processing and querying. This structure enables easy access, management, modification, updating, monitoring, and organization of data. The Structured Query Language (SQL) is commonly employed in most databases for writing and querying data.

**Data Warehouse**

A Data Warehouse serves as a centralized repository of data that facilitates informed decision-making. Data is sourced from transactional networks, relational databases, and other channels, flowing into the data center typically on a daily basis. Business analysts, data scientists, and decision-makers leverage Business Intelligence (BI) software, SQL clients, and various analytics applications to access and analyze this data.

The architecture of a data warehouse typically comprises three tiers. The bottom tier encompasses the database server, responsible for loading and storing the data. The middle tier houses the analytics engine, utilized for accessing and analyzing the stored data. Finally, the top tier consists of the front-end client, which presents results through reporting, analysis, and data mining tools.

A data warehouse employs Online Analytical Processing (OLAP) techniques for supporting decision support systems, organizing and accessing data in a multidimensional structure known as a data cube. Within this structure, a fact table holds the measurements or facts, along with foreign keys that link to dimension tables. Dimension tables, in turn, are denormalized tables containing descriptive information associated with the facts.

**ETL (Extract, Transform, Load)**

It is a critical process in data management that involves extracting data from diverse source systems, transforming it through operations like calculations and concatenations, and ultimately loading it into a Data Warehouse system. This process facilitates the movement of data from multiple sources into a centralized data warehouse.

As data sources evolve or change, ETL processes ensure that the Data Warehouse is automatically updated to reflect these modifications. By migrating data into a Data Warehouse and converting it into consistent formats and types, ETL plays a crucial role in ensuring data integrity and coherence within the overall system.

This project endeavors to consolidate various concepts into the creation of a Library Data Warehouse, encompassing modules such as books, members, and borrowers’ information. It amalgamates operational databases from two different branches, each utilizing relational database systems like MySQL. These databases house data pertaining to as books, members, and borrowers within their respective branches.

The Data Warehouse aims to address queries that may not be efficiently answered by individual source databases alone. ETL processes play a crucial role in cleaning and amalgamating data from these two different source databases. Additionally, a user-friendly GUI is provided to facilitate easy access to different views and queries.

Evaluations have been conducted, demonstrating satisfactory performance. This report offers an overview of the tools utilized and extends to discuss structured automatic data extraction techniques. It delves into the concept of modeling the data warehouse and explores various techniques aimed at achieving superior results.

**Problem Statement**

In the contemporary landscape, where education stands as a pivotal factor, libraries worldwide require robust systems for efficient information management and data extraction. As we are booming towards for technical advanced era, requirement of maintaining system which can handle large amount of data efficiently, along with the details what number of books are present in a library, whether a book of specific author is present or not, details about the borrower, their history of issuing a book can help in making decision which can boom the businesses of library.

A data warehouse emerges as a solution to address these challenges by providing a centralized repository for storing data from diverse sources in a uniform format. The schema of the data warehouse facilitates the retrieval of historical and real-time data related to books, members, borrowers, and more. This consolidated approach streamlines decision-making processes, enabling quick access to updated information.

Given the varied hardware and software platforms across two different branches and operational databases, an ETL (Extract, Transform, Load) tool plays a crucial role. It standardizes data from disparate sources into a uniform platform compatible with the data warehouse or operational databases.

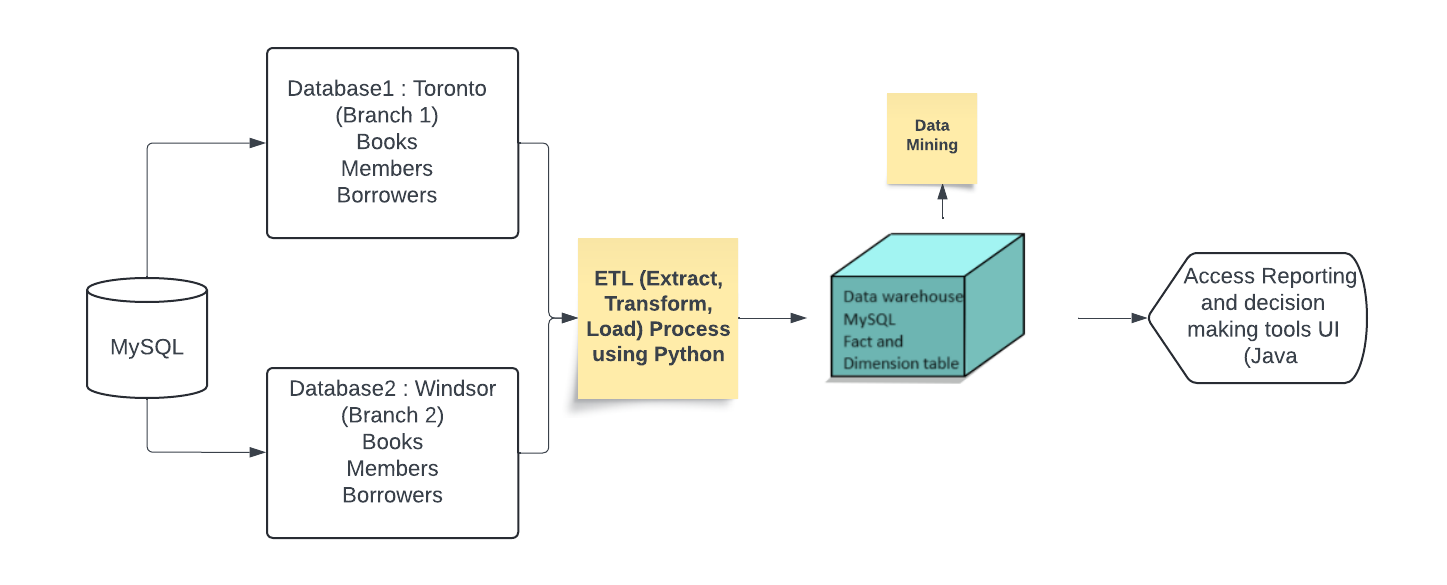
Therefore, there is a compelling need to design and implement a Library Data Warehouse. Such a system provides a structured framework for organizing vast amounts of data, storing it centrally, and offering a user-friendly interface for querying and accessing information from all operational source databases.

**Purpose of the Application**

The application is designed to establish a data warehouse accommodating data from two distinct branches (Toronto and Windsor). It incorporates database principles, data warehousing techniques, and an ETL (Extract, Transform, Load) process to amalgamate information from various source databases. Additionally, a graphical user interface (GUI) is integrated to provide an intuitive environment for querying and accessing query results.

The books, members, and borrowers’ modules are constructed within MySQL databases for two distinct branches. These modules are established as operational source databases. Subsequently, a data warehouse is constructed in MySQL workbench, consolidating data from both sources. The data warehouse is designed with a Star comprising one two tables while the dimension tables hold relevant information pertaining to the fact tables.

This application constructs a schema by scrutinizing the structure and content pertinent to two branches at present. It is developed subsequent to an analysis of different branches including the details of books, members and borrowers, which are utilized for disseminating information for user access through a systematic process, involving data collection, cleansing, processing, storage, and displaying the results of executed queries in a suitable user-friendly interface.



*Figure 1: Library Data Warehouse System Architecture*

**Implemented Module**

**Branch Module**

The branch module that is Toronto and Windsor contains the information related to book i.e. Id for uniquely identify the book, book name, author and number of copies present of that book currently in the library. Also, the detail of members including is, name, address, phone number email and also the borrower details, book issue and return details of the borrower along with issue date and return date. The attributes in both the tables of the databases differ in the schema, data types and syntax having null and not null values. The books table is connected to members tables through borrower’s module by a relationship table which will link the constraint in both modules of the two databases. These two source databases will be combined into the warehouse (MySQL) and fact table and its corresponding dimension tables are prepared for extracting information and running multiple queries in the views which can’t be answered with single source databases.

**Relationship between the modules**

**Books – Members via Borrower/Takes Relationship**

Books table is linked to members table through the borrowers relationship table in both the database with all the foreign key constraints that can be the primary keys or the table attributes in books and members tables. The fact table in the data warehouse contains the attributes from all tables including borrowers, members, books, dates and their relationship tables along with certain other attributes (e.g. historical information of borrower history in past 6 month or a year) required to extract information that source databases can’t answer single handedly.

**Database and Data Warehouse Schema**

The schema diagrams of source databases and data warehouse for two branches (Toronto and Windsor) are completely implemented in MySQL for visualization purpose.

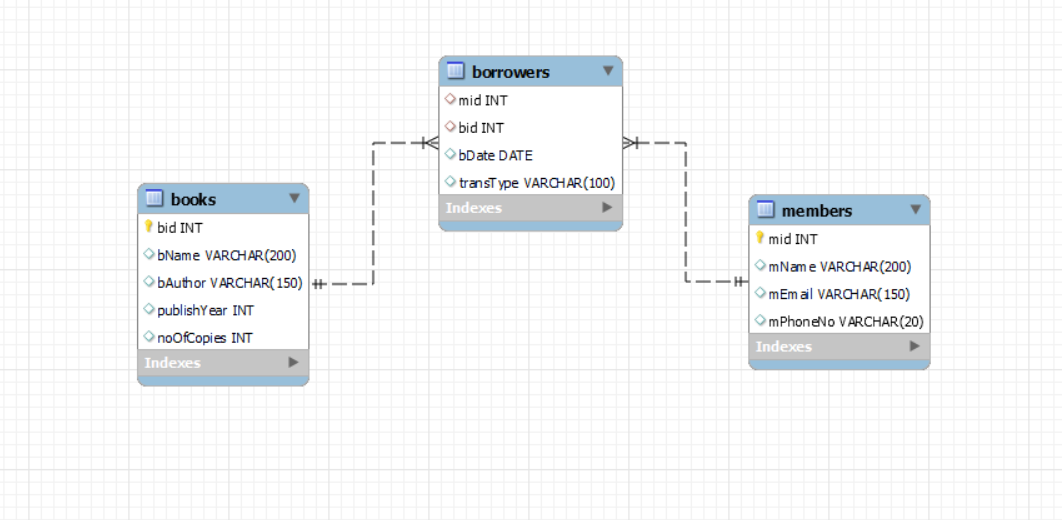
1. **Windsor schema for databases**

**Source database 1: Windsor**

Books (Bid, BName, BAuthor, BPublishYear, BNoOfCopies)

Members (Mid, MName, MEmail, MPhoneNo)

Borrowers (Bid, Mid, BDate, BTransactionType)



*Figure 2: Source database for Windsor branch*

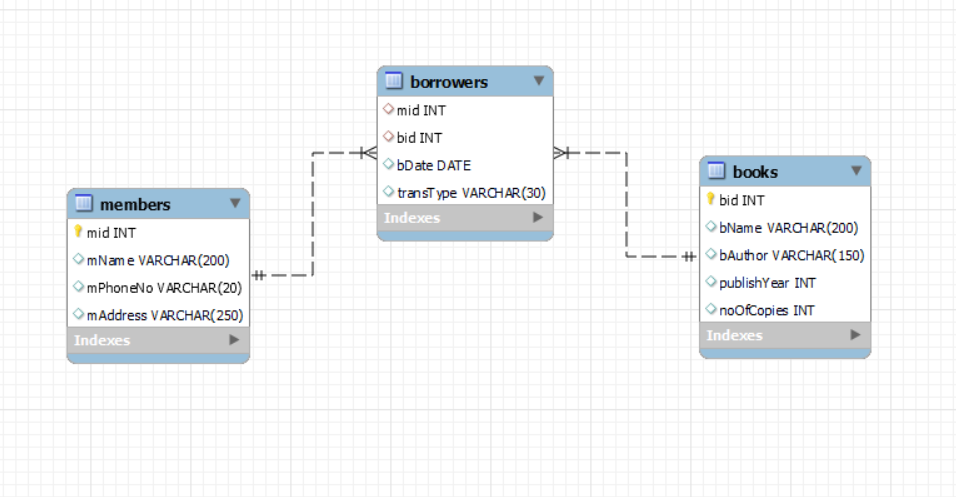
1. **Toronto schema for databases**

**Source database 2: Toronto**

Books (Bid, BName, BAuthor, BPublishYear, BNoOfCopies)

Members (Mid, MName, MPhoneNo, MAddress)

Borrowers (Bid, Mid, BDate, BTransactionType)



*Figure 3: Source database for Toronto branch*

1. **Datawarehouse Star Schema for library**

**Fact table**

fact (bid, mid, transType, dateId, branchId)

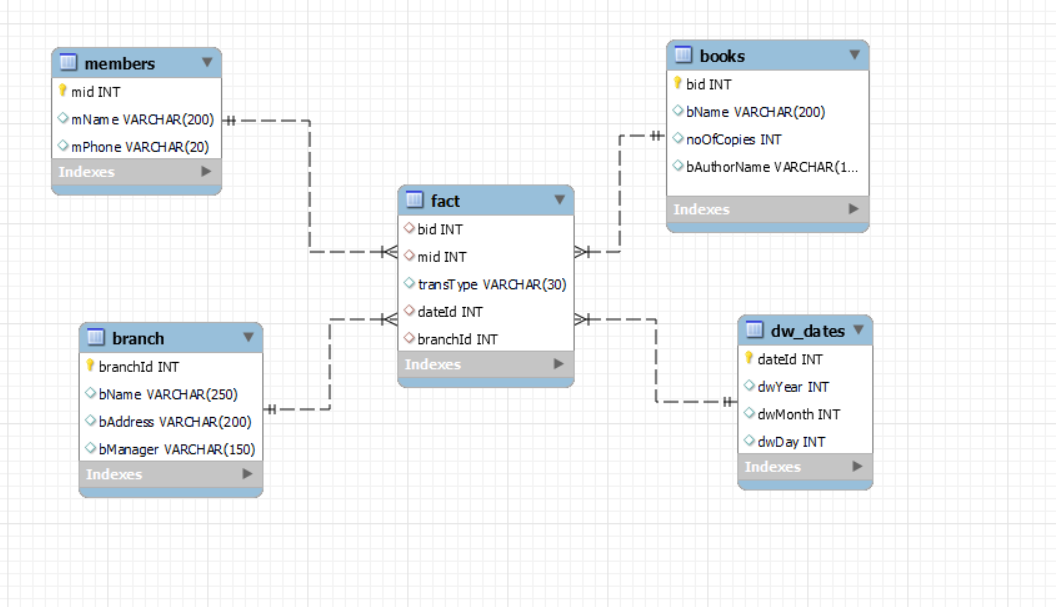
**Dimension tables**

Books (bid, bname, noOfCopies, authorName)

Branch (branchId, branchName, branchAddress, branchManager)

Dw\_dates (dateId, dwYear, dwMonth, dwDay)

Members (mid, mName, mPhoneNo)



**Tools and Software**

The entire project is built on Windows 11 operating system. We have used MySQL workbench version 8.0.34 for creating both databases Windsor and Toronto and also for creating the data warehouse. The GUI of the system is built on AWT using Java version 20. For ETL (Extract, Transform, Load) process, we have used python 3.12 and also imported “pymysql” for database connectivity.

**Screenshots of the Implemented System**

**Conclusion**

This project implements Library data warehouse system for two different branches Windsor and Toronto and integrates them to the warehouse. It is considering two source databases and warehouse from MySQL workbench database. Since the data used here is structures, we are using ETL for transferring data from csv to source databases and from source databases to warehouse. The warehouse helps in keeping the historical and current data maintained in a central repository and answers multiple queries related to books and issue details of the members. GUI has been built for user friendly environment and for searching the data warehouse while firing the queries or calling the views in the back end. The project aims to suffice fair requirements of a library warehouse and is being used for two different branches.

**Future Work**

The project can be extended to combine the databases from more branches or with multiple other databases or combination of both. Furthermore, this project can be extended to further recommendation system, based on members history can recommend books from same author or same subject books from other authors.

**Contributors**

**Source Database creation**

* SDB1 – Windsor - Parneet Kaur
* SDB2 – Toronto – Pooja Vishwakarma

**Data Warehouse Creation**

* Dimension and Fact tables Creation with Referential Integrity Constraints – Parneet Kaur
* Indexing – Pooja Vishwakarma

**ETL Design**

* ETL for transferring data from csv to source databases – Pooja Vishwakarma
* ETL for transferring data from source databases to data warehouse – Parneet Kaur

**Application Design**

**SQL Queries**

Worked by both