MODULE NAME: DATA MINING AND WAREHOUSING.

DEPARTMENT : ICT

OPTION : IT

LEVEL : 8 YEAR 4

CLASS : IT Btech

LIBRARY USAGE DATA WAREHOUSE SYSTEM

A Comprehensive Data Integration Solution for Modern Library Management

GROUP 1: MEMBERS

NAMES	REG NO	MARKS
HIMBAZWA Rebbeca	25RP21086	
NISINGIZWE Jean Pierre	25RP20888	
UMUGWANEZA Aimee Providence	25RP18890	

Date: 4th Feb 2026

DESIGN DOCUMENT

1. Introduction

This document presents the complete design and technical specification of the **Library Data Warehouse System**, developed to support data-driven decision-making in library operations. The system integrates heterogeneous data sources into a centralized analytical repository.
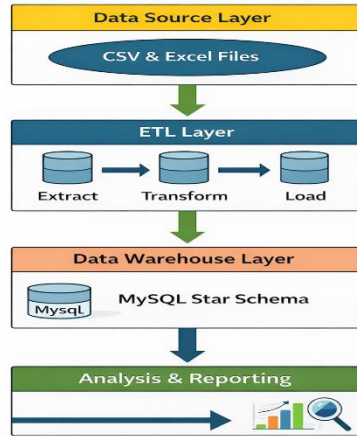
2. Project Overview

Modern libraries generate data from multiple systems such as book lending, digital resource usage, and room booking platforms. These systems operate independently, making reporting time-consuming and inconsistent.

This project consolidates these data sources into a single **data warehouse optimized for analytics**.

3. System Architecture

Figure 3.1: System Architecture of the Library Data Warehouse

The system follows a **three-tier architecture**:

- ➤ **Data Source Layer**: CSV and Excel files from operational systems

- ➤ **ETL Layer**: Python-based data extraction, transformation, and loading

- ➤ **Data Warehouse Layer**: MySQL star schema optimized for analytics
  Data Sources → Staging Layer → ETL Process → Data Warehouse
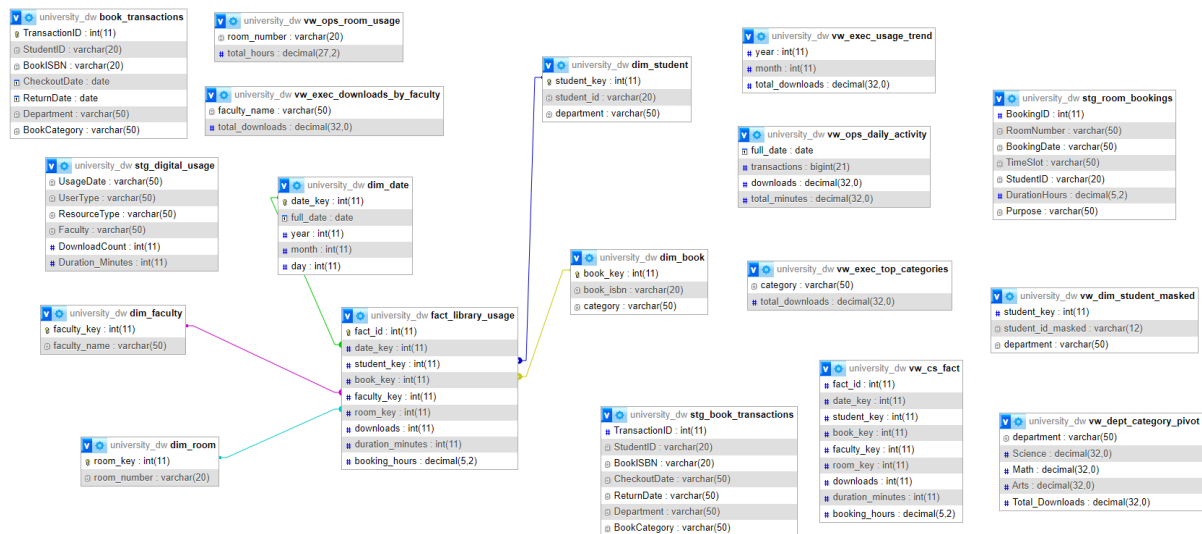
4. Data Warehouse Schema

The data warehouse uses a **star schema** design with one central fact table and multiple dimension tables to support fast analytical queries.

5. Entity Relationship Diagram (ERD)

ERD Description

- Central table: fact_library_usage

- Dimension tables:

  - ➢ dim_date

  - ➢ dim_student

  - ➢ dim_book

  - ➢ dim_date

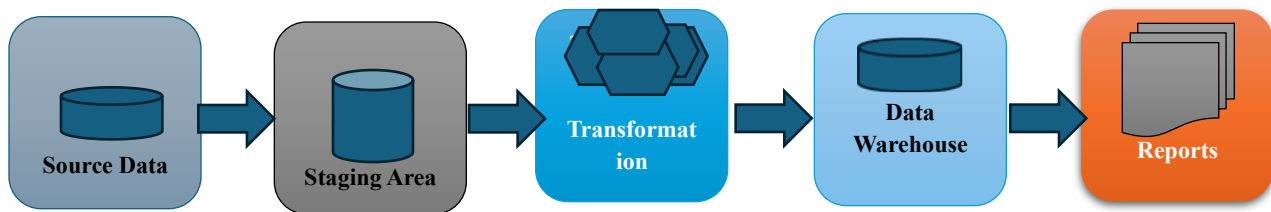**Figure 5.1:** ERD of the Library Data Warehouse (Star Schema)

6. ETL Process Flow

ETL Description

The ETL pipeline performs:

1. Data extraction from source files

2. Validation and data quality checks

3. NULL handling and transformation

4. Loading into dimension and fact tables

Figure 6.1: ETL Process Flow Diagram

Data Flow:
Sources → Staging → Cleaning → Transformation → Dimensions → Fact → Reporting

7. NULL Handling Strategy

NULL values are handled contextually:

*   Meaningful NULLs (e.g. book not returned) are preserved

*   Missing values are imputed or flagged

*   All NULL decisions are logged

8. Data Quality Framework

The system enforces:

- Completeness

- Accuracy

- Consistency

- Validity

- Uniqueness

Automated checks run during every ETL cycle.

9. Dimensional Model Justification

The star schema:

- Improves query performance

- Simplifies reporting

- Matches business understanding of library usage

10. Technology Stack

- **Database**: MySQL

- **ETL**: Python

- **Data Processing**: Pandas

- **Documentation**: PDF

11. Implementation Details

The project is organized into:

- Design documents

- Database scripts

- ETL scripts

- Source data files

12. Performance Considerations

- Indexing on foreign keys

- Optimized joins

- Scalable schema design

13. Security & Compliance

- Role-based access control

- Audit logging

- Secure credentials handling

14. Future Enhancements

- Dashboards

- Predictive analytics

- Additional data sources

15. Conclusion

The Library Data Warehouse provides a robust, scalable, and analytics-ready solution for library management, enabling informed decision-making and improved service delivery.