Analysis and Design of Data Warehouse for E-Library Using Snowflake Schema

Made Dwi Mulyawan[1*], Ida Bagus Alit Swamardika[2], and Komang Oka Saputra[3]

[1]Department of Electrical and Computer Engineering, Post Graduate Program Udayana University
[2,3]Department of Electrical and Computer Engineering, Udayana University
*dwi.mulyawan@student.unud.ac.id

Abstract Information systems have become an important requirement in helping human work, one of which is e-library. This system is used by librarians in carrying out daily business processes in the library, such as borrowing and returning books, managing book collections, and procuring the latest book collections so that the library has updated books. Over time, the data stored in the database will increase a lot, so when analyzing information that requires data with a certain period will take quite a long time. So, in this case, the use of a data warehouse is needed to help the librarian in providing the information needed. This research will present a data warehouse design using snowflake schema by following the design stages in Fourth Steps Dimensional Modeling, as one of the solutions in providing the required data warehouse design in the e-library.

Index Terms—**Data Warehouse, Snowflake Schema, Fourth Steps Dimensional Modelling, OLTP, Information System**

I. INTRODUCTION

CURRENT developments in information technology have developed rapidly, making it possible to develop a system that can simplify human work. Information systems can do storage of a set of data that can later produce information. This information can later be used as a support in business processes in a company. Also, the information system is one of the innovations to change the whole system or work process that is done or run conventionally to be easier and integrated with the information system.

One information system that is utilized in facilitating the management process, is the e-library information system. E-Library is a system used in performing performance management or transaction processes related to borrowing and returning books, distributing books, monitoring books being borrowed, and making an inventory of books in the library. So that the information system becomes an important part of the library in processing the information needed.

Over time the data stored in the e-library will increase, so when analyzing information that requires data with a certain period will take quite a long time. So that the use of a data warehouse is needed to help the librarian in analyzing the information needed. The data warehouse is a database that is designed to be able to do queries quickly [1]. By using a data warehouse, you can see patterns or see trends, rolling up to be summarized, drilling down to get detailed information and making reports [2] such as library material collection reports, book borrowing reports, and library visit data collection to facilitate librarians in making decisions quickly and accurately.

Based on these problems, it is necessary to design a data warehouse in the e-library that can be used to analyze in a structured and detailed process of processing data transactions that exist in the e-library. The data warehouse design in e-library is done using the Snowflake Schema method. Snowflake Schema is a development of Star Schema which applies the normalization method for dimension tables to simplify and speed up the query process that is performed [3].

II. LITERATURE REVIEW

A. Related Research

Some of the literature used as a reference or basis in designing this data warehouse research as follows:

1) Research conducted by [4] conducted a data warehouse design to respond to the Bali Province expenditure report and evaluate grant recipients. The design was carried out using Kimball's fourth step method to design a data warehouse that is compact, dynamic, and supports the creation of grant funding reports and grant fund recapitulation.

2) In research [5] a data warehouse design was carried out to facilitate obtaining information to support accreditation standards and as a reference in obtaining results. The data warehouse design is done using a star schema in changing the dimensions of the business into a logical data model.

3) Research conducted by [6] is a data warehouse that can be used as a source of information about batik company

management. The design was carried out using a snowflake scheme and MOLAP (Multidimensional Online Analytical Processing). By using a snowflake scheme, normalized tables can reduce data redundancy and update the data analysis as needed.

4) Research [7] implements a data warehouse in the Online Community Complaints System used by the Denpasar City Government. data warehouse design is done using a snowflake schema that aims to minimize the excess data. The data warehouse in this study is used to simplify the process of analyzing the number of complaints by category, platform, and follow-up.

5) In research [8] a data warehouse design is done using the snowflake schema to optimize e-travel business. The use of snowflake schemes is a solution for companies to predict business trends, maintain quality, improve competitiveness, and exist in the long run.

Based on research that has been done in implementing and designing a data warehouse, this research will do a data warehouse design in the e-library by using snowflake schema through the fourth steps of dimensional modeling.

B. Data Warehouse

The data warehouse is a collection of data that has a subject-oriented, integrated, time-variant, and fixed nature of data collection to support the management decision-making process [9]. The database is designed more for querying and analyzing transaction processes, usually containing historical transaction data and possibly data from other sources [10],[18]. Data warehouses separate analysis workloads from transaction workloads and allow organizations to combine or consolidate data from various sources [14],[16].

Data warehouses typically store historical data that functions as a comprehensive database to support all decision analysis needed by an organization to provide detailed information and summaries [11]. There are 4 characteristics in a data warehouse, namely:

1) Subject Oriented
   The data warehouse is designed to meet the needs of data analysts based on a particular subject.

2) Integrated
   The data warehouse must be able to integrate different data sources as diverse operating systems, so before data from various sources are stored in a data warehouse, it must eliminate inconsistent data.

3) Non-Volatile
   Transaction data that has been entered into the data warehouse will never be updated or deleted. Data stored in a data warehouse is not intended for daily activity data, but rather the data is entered into the data warehouse at certain time intervals that are tailored to business needs.

4) Time-Variant
   Historical data storage (historical data) is a type or piece of data that is very important in a data warehouse, where data is often stored several times, such as when a transaction occurs, changes in transaction time, or a transaction occurs, changes in transaction time, or

when a transaction is canceled. A data warehouse is said to be good if the data warehouse is valuable [20].

C. Snowflake Schema

Snowflake Schema is a logical table structure in a multi-dimensional database that is designed so that the ER diagram resembles the shape of a snowflake [1]. Snowflake Schema is a development of Star Schema by adding additional dimensions. The dimension table is normalized which divides data into additional tables. The structure of the Snowflake Schema can be seen in Figure 1.
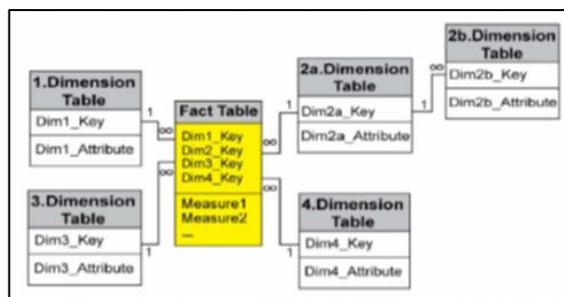

Fig. 1. Snowflake Schema Structure

Figure 1 shows the structure of the Snowflake Schema, where the structure contains a fact table surrounded by dimension and sub-dimensional tables. Dimension table which carried out the normalization process aims to be able to simplify and speed up the query process in making a report that is designed in the e-library data warehouse.

D. Fourth Step Dimensional Modelling

In the data warehouse, Kimball recommends a very unique development methodology, including a bottom-up that develops a data mart in every business process [12],[13]. The development methodology consists of four processes::

1) Select the Bussiness Process
   This stage identifies the business processes needed to determine the dimensions and facts at the end of the data warehouse modeling.

2) Declare the Grain
   The contents of the fact table can be identified through grain statements. The grain will map and match information needed in the data warehouse [15].

3) Identity the Dimensions
   This stage is carried out the process of developing dimension tables to meet the needs provided in the fact table [19].

4) Identity the Fact
   The fact table is the embodiment of the data warehouse and is the main target of business needs. The fact table will display the results of business processes that can meet business needs.

III. RESEARCH METHODOLOGY

This research was carried out through several processes needed to achieve the expected results. The research methodology is shown in Figure 2.

Fig. 2. Research Methodology

Figure 2 shows the structure of the research methodology that was conducted. The problem of managing book collections in libraries is finding as to the basis for designing a data warehouse. Data collection is carried out through the process of observing the history of transitions that occur in the distribution of books, such as incoming books, books that are still in a loan, books that are damaged, and so on. And some related literature in the development of data warehouses. Furthermore, the business needs that have been collected will be the basis for designing a data warehouse that is developed through four steps of dimensional modeling, namely:
1. Select the Business Process
2. Select the Grain
3. Identify the Dimensions
4. Identify the Fact.

Through these stages, the results of the data warehouse design are expected to be a solution to the business needs of the e-library to assist in analyzing the information needed in book collection management as a material consideration in the procurement of the latest book collection. The results of the design will produce conclusions as an evaluation topic so that the data warehouse design can run effectively and be applied to the e-library.

IV. RESULT AND ANALYSIS

The results of this study are reviewed in several explanations following the design methodology described in the previous section. The process of designing a library data warehouse is carried out in several stages as follows:

A. Analysis of Business Needs

This stage is the main thing to determine what is needed from the data warehouse design to be developed. The needs analysis is carried out based on the e-library database shown in Figure 3.

Fig. 3. OLTP E-Library

Figure 3 is the OLTP design as a database that functions as a daily requirement in the transaction process in the e-library. Based on the OLTP design, the business processes in the library are very diverse, namely:
1) Member Data is a variety of data stored for member information to be able to access the library information system.
2) Book Data is a variety of data stored for book information that is owned by the library.
3) Data Publisher is a collection of data relating to a book publisher that becomes a collection in a library.
4) Author data is a collection of data relating to the author of a book that is a collection in the library.
5) Category data is a variety of data used to categorize the types of books that are collected in a library.
6) Data Shelf is data that is used to store information on the location of the storage of books placed on a shelf in the library.
7) Fine data is data used to store member information that is late returning books, damaging books, and eliminating library collection books.

Various data is a business requirement in the e-library, librarians will use some data that can be needed for the information needed. The business needs of the e-library data warehouse design are determined on the management side of the collection of books owned by the library, such as recapitulation of incoming and outgoing books, reports of damaged and missing books, favorite book reports, and reports on the number of collection books by category. Where the information is needed as a consideration in the procurement of the latest book collection. Book collection management is analyzed as a business need that requires some data following the target, among others:
1) Member Data is needed for member data information needed in the transaction process.
2) Book Data is needed to find out what books are collections in the e-library and borrowing history from the members of the books.
3) Publisher Data is needed to know the history of incoming book transactions and simplify the process of categorizing existing books.
4) Author Data is needed to categorize the collection of books based on the author.

5) Category Data is needed to group books according to the type of collections available in the e-library.
6) Data Rack is used to determining the location of book storage in the library to facilitate librarians in managing existing book collections.

The five data are estimates for writers that can be used in designing e-library data warehouse as an analysis needs in the process of library book collection management as a material consideration in making the latest book collection procurement.

B. Select the Business Process

The initial stage of the Fourth Steps Dimensional Modeling is the selection of business processes. Business processes are determined based on the analysis of information needs that have been carried out in the previous stages. The business processes chosen in the data warehouse design are:
1) Book Loan Transactions
2) Favorite Book
3) Lost and Damaged Book
4) Total of Book Collection

The selected business process is needed in book management in the library as a consideration in the procurement of the latest book collection.

C. Declare the Grain

Data recorded in the fact table will be determined via grain. Grains in the data warehouse are used to record the data processing of book lending transactions in e-libraries by members. Determination of grain at this stage can be described in the following Table 1.

TABLE I
DECLARE THE GRAIN

GRAIN / DIM	BOOK TRASACTION	FAVORITE BOOK	LOST AND DAMAGED BOOK	TOTAL BOOK COLLECTION
BOOK	√	√	√	√
MEMBER	√	√	√	
PUBLISHER		√	√	√
AUTHOR		√	√	√
CATEGORY		√	√	√
SHELF			√	√
TIME	√	√	√	√

Table 1 is the process of determining grain with various dimensions needed to realize a fact table based on the selected business process.

D. Identify the Dimensions

Dimensions are identified to meet the needs of the fact table. Each dimension table chosen is a data table that has important data records for output from the data warehouse. Tables 2 through Table 11 are a collection of dimensions identified as supporting data tables for the realization of fact tables. In designing a data warehouse for analyzing the information needs needed by librarians as a material consideration in procuring the latest book collections, it requires a dimension table, such as:

TABLE II
MEMBER DIMENSION

dim_member
id_member
identity
name
address
gender
phone_number
email

TABLE III
BOOK DIMENSION

dim_book
id_book
book_title
id_author
id_publisher
id_shelf
id_category
publication_year
ISBN
id_source
id_status

TABLE IV
SOURCE LIBRARY DIMENSION

dim_source_library
id_source
source

TABLE V
BOOK STATUS DIMENSION

dim_book_status
id_status
status

TABLE VI
PUBLISHER DIMENSION

dim_publisher
id_publisher
publisher_name
publisher_address
phone_number
email

TABLE VII
AUTHOR DIMENSION

dim_author
id_author
author_name
email

TABLE VIII
CATEGORY DIMENSION

dim_category
id_category
category

TABLE IX
SHELF DIMENSION

dim_shelf
id_shelf
shelf_name

TABLE X
TIME DIMENSION

dim_time
id_time
date
month
year

TABLE XI
TRANSACTION STATUS DIMENSION

dim_transaction_status
id_trans_status
status

E. Identify the Fact

At this stage, the fact table is selected based on the identified dimension table. The fact table in the data warehouse interprets the business processes in the e-library and makes it easy for librarians to analyze the information needs that are used as a consideration in procuring the latest book collections. The fact table identified for the e-library data warehouse design is shown in Table 12.

TABLE XII
TRANSACTION FACT

fact_transaction
id_transaction
id_member
id_book
id_time
id_trans_status

Table 12 is some of the data used in the fact table that is formed from the relationships between the dimension tables. In the case study, the authors designed a fact table called fact_transaction. This fact table is the center of various dimensions to meet information needs in managing book collections in e-libraries.

F. Snowflake Schema Design

The data warehouse design to meet the information needs of book collection management in e-libraries using the snowflake schema is shown in Figure 4 as follows:
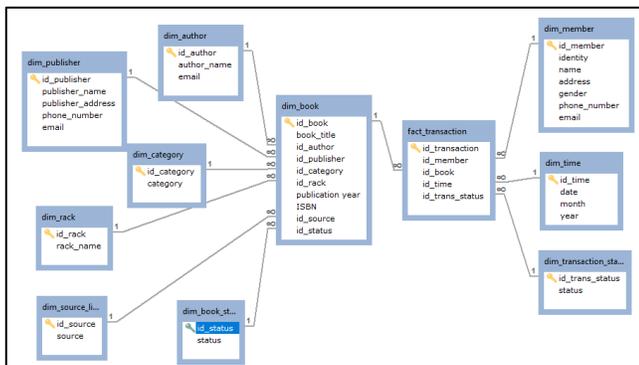
Fig. 4. Snowflake Schema Design

Figure 4 is a snowflake schema used to design an e-library data warehouse. With this data warehouse design scheme and model, the management of book collections in e-libraries

will be enhanced through the ability to analyze information from various available data. Business processes that have been determined in the previous process can help writers in designing an ideal and effective e-library.

G. Analysis of Snowflake Schema

There are weaknesses and strengths in implementing a data warehouse using the snowflake schema. According to [6] by using the snowflake schema, normalized tables can reduce data redundancy and speed up the processing of data analysis needed. Some of the underlying considerations in using this scheme, because it can ease the burden of memory used, as well as easier maintenance and updating process. Weakness will have an impact on the level of query complexity, higher data structures, and difficulty in finding data. However, the design of this scheme is still feasible to be used in certain case studies that involve a lot of dimension to dimension connectivity to support the needs of fact tables. As the design of data warehouse in e-library that has been done in this study.

V. CONCLUSION

The conclusion of this research is the design of a data warehouse in e-library will be very possible in helping librarians in providing the information needed in managing their book collections. By analyzing the expected business needs in this data warehouse design, the use of snowflake schema is very suitable in this case study, where the e-library there are many dimensions needed as supporting data in the fact table. The fourth steps dimensional modeling method is also very helpful in the design process of the e-library data warehouse. Where each stage of the method is very simple to implement.

REFERENCES

[1] Connolly, T. M., & Begg, C. E. (2010). Database Systems: A Practical Approach to Design, Implementation, And Management. Boston: Pearson Education.
[2] Foster, E. (2016). Database systems: a pragmatic approach. New York: Apress.
[3] Poniah, P. (2001). Data Warehouse Fundamentals: a Comprehensive Guide for IT Professional. New York: John Wiley & Sons
[4] Narendra, A.. N., Murpratiwi, S. I. and Sudarma, M. (2017) 'Design of E-Grant Application Data Warehouse', IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 11.
[5] Budiarta, K., Ananta Wijaya, P. A. and Partha, C. G. I. (2017) 'Analysis and Design of Data Warehouse on Academic STMIK STIKOM Bali', IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 35.
[6] Gede, L., Suardani, P. and Sudarma, M. (2018) 'Data Warehouse Implemantation To Support Batik Sales Information Using MOLAP', IJEET (International Journal of Engineering and Emerging Technology), 3(1), pp. 45–51.
[7] Putra Wijaya, I. P. A., Wahyudin, W. and Mataram, M. (2017) 'Data warehouse Implementation on Denpasar City Online Community Complaints System', IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 67.
[8] Prabawa, I. N. A. et al. (2019) 'Analysis and Design Data Warehouse For E-Travel Business Optimization', IJEET (International Journal of Engineering and Emerging Technology), 4(1).
[9] Inmon, W. H., & Hackathorn, R. D. (1994). Using the Data Warehouse. New York: John Wiley & Sons.

[10] Lane, Paul. (2002). Oracle91 Data Warehousing Guide. Oracle Corporation.

[11] Kimball, R., & Margy, R. (2002). The Data Warehouse toolkit: the complete guide to dimensional modelling. New York: John Wiley & Sons.

[12] Kimball, R., & Margy, R. (2010). The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. First Edition. Indianapolis: John Wiley & Sons

[13] K. A. B. Permana, G. B. Subiksa, and M. Sudarma (2017), "Design Data Warehouse For Centralized Medical Record," IJEET (International Journal of Engineering and Emerging Technology), vol. 2, no. 2, pp. 47–51

[14] Ambara, M. P., Sudarma, M. and Kumara, I. N. S. (2016) 'Desain Sistem Semantic Data Warehouse dengan Metode Ontology dan Rule Based untuk Mengolah Data Akademik Universitas XYZ di Bali', MITE (Majalah Ilmiah Teknologi Elektro), 15(1), p. 8. doi: 10.24843/mite.2016.v15i01p02.

[15] Iswardani, P. R., Pramana, I. W. S. and Saputra, K. O. (no date) 'Design of Data Warehouse for Monitoring Hotel's Food and Beverage Cost', IJEET (International Journal of Engineering and Emerging Technology), vol. 4, no. 1, pp. 1–4.

[16] Shobirin, K. A., Iskandar, A. P. S. and Swamardika, I. B. A. (2017) 'Data Warehouse Schemas using Multidimensional Data Model for Retail', IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 84. doi: 10.24843/ijeet.2017.v02.i01.p17.

[17] Sudarmojo, Y. P. (2018) 'Design of Library Data Warehouse Using OLTP Result of Services Analysis', IJEET (International Journal of Engineering and Emerging Technology) 3(1), pp. 62–65.

[18] Pradhana, I. G. N. A., Giriantari, I. A. . and Sudarma, M. (2018) 'Analisis dan Perancangan Sistem Pengelola Data Menuju Implementasi Data Warehouse Untuk Mendukung Administrasi E-Procurement', MITE (Majalah Ilmiah Teknologi Elektro), 17(2), p. 245. doi: 10.24843/mite.2018.v17i02.p12.

[19] Sulastra, I. M. D. J., Sudarma, M. and Kumara, I. N. S. (2015) 'Pemodelan Integrasi Nearly Real Time Data Warehouse Dengan Service Oriented Architecture Untuk Menunjang Sistem Informasi Retail', MITE (Majalah Ilmiah Teknologi Elektro), 14(2), p. 11. doi: 10.24843/mite.2015.v14i02p03.

[20] Anshari, M., Suryadharma, I. P. and Rinas, I. W. (2017) 'Implementation of Data Warehouse in Human Resource Information System Using SEM-GeSCA', IJEET (International Journal of Engineering and Emerging Technology), 2(1), p. 16.