Design of Data Warehouse in Library Circulation using Pentaho

Totok Suprawoto[1]

[1]*Department of Information System, STMIK AKAKOM, Jl. Janti No. 143, Yogyakarta*

Keywords: library collection, library circulation, data warehouse, pentaho.

Abstract: Library is a source of knowledge that can be accessed by students, lecturers and employees both from inside and outside campus. Library as a tool that can be used for the development and improvement of knowledge for its members, plays an important role in disseminating information and providing references for the academic community. Current and previous circulation of lending and returning data books can be stored in a data warehouse. Data warehouse is a database that is designed to work on the query process, create reports and further analysis for the decision making process. In this research, 5 dimensions have been designed, namely: the time dimension, the member dimension, the publisher dimension, and the author dimension. Furthermore, it is implemented using 2 fact tables, namely: borrowing tables and book returns tables. Data Warehouse that has been built can be used by management to use certain software to produce information that can be used for decision making.

1 INTRODUCTION

Library is one source of knowledge that can be accessed by students and employees. The library collection includes textbooks, research journals, theses, scientific magazines, lecture modules, ebooks, and e-journals. The library is a tool used for the development and improvement of knowledge, plays a very large role in the series of information dissemination because the library provides collections that can be used as a reference for the academic community. Circulation of library materials, is one of the main activities of the library in providing services to its members.

There are two activities in the circulation of library collections, namely lending activities and library material return activities. Borrowing activities carried out by members are divided into two types, namely: 1) lending to lecturers with a maximum of 4 titles in 1 semester, and 2) lending to students with a maximum of 2 titles with a maximum borrowing of 1 week and can be extended again once. Book return transactions are carried out in stages and / or at the same time as long as they are still within the deadline for borrowing. If a member is late returning books, a late fee will be imposed. Many books are borrowed by members, are not returned on time and it is difficult to get information about who the borrower is and when it should be returned. So that it can provide opportunities for other members to borrow certain books to support their learning process. In addition, library leaders still find it difficult to obtain information re-

lated to books most borrowed by members, making it difficult for leaders to make plans for procurement of books.

Data Warehouse is a database that designed to work on the query process, create reports and analysis. Data stored in a data warehouse is historical data from an organization / company that has not been stored in detail. Data Warehouse is different from OLTP (Online Transactional Processing) data whose data is stored until the process is complete. The traditional approach to the ETL process takes data from the data source, places it in the staging area, and then transforms and loads it to the datawarehouse shown in Figure-1 below.

Figure 1: Data Warehouse Architecture

Research (Kurniawati, 2006) on the design and manufacture of cluster analysis applications on book transaction data in Petra Christian University Library. This study aims to determine the relationship between lending patterns and student GPA using the cluster analysis method. Being research from (Supriyatna and Wahyudi, 2012) about data warehouse design in the Bina Sarana Informatika library, aims to build a database that can be used by leaders to analyse

and make decisions appropriately, and make applications that can analyze deeply on the subjects desired. Researchers (Andri and Tujni, 2015) designed a Data Warehouse as a library repository that was implemented using Pentaho Kettle assistive software. Research (Suprawoto et al., 2017) on integration of morbidity data of puskesmas patients into the data warehouse at the district health office in the Bantul, emphasized on how to integrate outpatient morbidity data from puskesmas in the Bantul district health office. Furthermore, the data is further processed to produce information according to the needs of leaders to support decision making.

2 METHOD

2.1 Data Collection

The first step in this research is data collection. For this purpose researchers conducted data collection by taking data from the library information system (Sipusta) which was the object of research and conducted a literature review to explore information related to library data management systems and business processes.

In this study, researchers used a top-down approach, this approach begins with defining organizational goals and policies and then analyzing information needs and then down to transaction processing. Before starting to create a data model for a data warehouse, identified information and data requirements specifications available in the Library. The data obtained at this stage are operational data from the circulation of books in the library. Furthermore, a deeper analysis of the information needs of the leadership is carried out. Furthermore, the data obtained will be used as input to the system analysis process.

2.2 System Analysis

The data source used in this study came from a book circulation database at library. The database taken consists of book data, library members, loan transactions and library book return transactions. The Data Warehouse design methodology used in this study uses the nine-step methodology. The process of integrating data is carried out with the concept of ETL (Extracts, Transformation, Loading). Data that has been integrated and stored in the same format is further grouped into the form of dimension tables and fact tables.

Pentaho Kettle software can be used as a device to integrate data. Pentaho Kettle provides facilities for ETL (Extraction, Transformation and Loading) processes. The main elements of Pentaho Kettle are Transformation and Job. Transformation is a set of instructions to change the input into the desired output. Whereas Job is a collection of instructions for carrying out transformation. There are three main components in Pentaho Kettle namely, Spoon, Pan and Kitchen. Spoon is the user interface for creating Job and Transformation. Pan is a tool that functions to read, change and write data, and Kitchen is a program that executes jobs.

2.3 System Design

Before designing the system, an analysis of business processes from the circulation of library collections in library includes: 1) Member registration (students and lecturers), 2) Procurement of library collections (books, journals, magazines), 3) Borrowing books, 4) Returning books, and 5) Free submission of student libraries.

After knowing the business processes that occur in the library, then do the grain selection process. The selected grain will be used as a fact table in the Data Warehouse. Based on the business processes that have been defined, the resulting grains include: the number of books, the number of members (number of students and number of lecturers), and the number of books borrowed.

Next identify and adjust the dimensions associated with the fact table. From the results of the identification of fact tables, 6 dimensions can be determined in this design, namely: 1) Dimensions of members, 2) Dimensions of books, 3) Dimensions of time, 4) Dimensions of categories, 5) Dimensions of publishers, and 6) Dimensions of authors.

The next step is to choose a fact table based on the selection of grains in the previous stage. The fact table obtained from the analysis consists of borrowing and returning library collections in book form. The design of the fact table produced in this study is the borrowing fact table and the book return fact table. The loan fact table and the book return fact table can be seen in Figure-1 and Figure-2.

2.4 Detailed Design Stage

2.4.1 Save Initial Calculations and Fact Tables

Aggregation in the loan fact table is total borrowing of books based on time (days, weeks, months and years) and aggregate factual returns of books are total returns based on time (days, weeks, months and years).
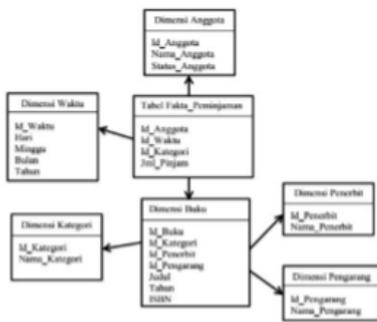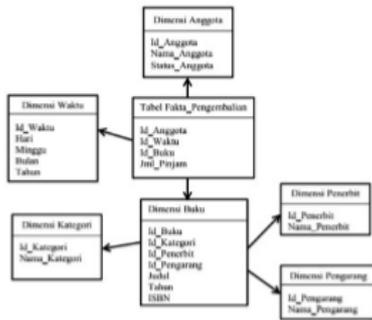
Figure 2: Borrowing Facts Table

Figure 3: Return Fact Table

2.4.2 Complete Dimension Tables

The book dimensions table and its attributes in this study include: Id_Book, Id_Category, Id_Publisher, Id_Author, Title, Year, and ISBN.

Field	Type (Length)	Information
Id_Buku	Varchar(8)	Book Identity Number
Id_Katego ri	Varchar(3)	Category Identity Number
Id_Pengar ang	Varchar(6)	Author Identification Number
Id_Penerb it	Varchar(6)	Issuer Identity Number
Judul	Varchar(10 0)	Book title
Tahun	Varchar(4)	Publication Year
ISBN	Varchar(20)	ISBN number

Figure 4: Book Dimensions.

The book dimensions in Figure 4 function to hold the book data, there are 7 attributes, with the primary key Id_Book, and 3 guest keys, namely: Id_Category, Id_Author, and Id_Penerbit.

Field	Type (Length)	Information
Id_Kateg ori	Varchar(3)	Category Identity Number
Nama_K ategori	Varchar(20)	Copy Categories (Programming, Information Systems, Management, Expert Systems, DSS, Economy, Social, Culture)

Figure 5: Book Category Dimensions.

The Category dimension in Figure 5 serves to hold the book category data, there are two attributes, with

the primary key Id_Kategori and is equipped with the Nama_Kategori attribute.

Field	Type (Length)	Information
Id_Penga rang	Varchar(6)	Author Identification Number
NamaPen garang	Varchar(30)	Author's Name

Figure 6: Author Dimensions.

Author dimension in Figure 6 has the function to accommodate the data of the author's name, there are 2 attributes, with the primary key Id_Pengarang and equipped with the name_Pengarang.

Field	Type (Length)	Information
Id_Pener bit	Varchar(6)	Publisher Identity Number
Nama_P enerbit	Varchar(30)	Name of Book Publisher

Figure 7: Publisher Dimensions.

The Publisher Dimension in Figure 7 functions to hold the data of the book publisher's name, there are 2 attributes, with the primary key Id_Penerbit and is equipped with the Nama_Penerbit attribute.

Field	Type (Length)	Information
Id_Anggota	Varchar(9)	Member Identity Number
Nama_Anggot a	Varchar(30)	Library Member Name
Status	Enum()	Member Status (Lecturer, Student)

Figure 8: Members Table.

The Member Dimension in Figure 8 serves to hold the data of the name of the author of the book, there are 3 attributes, with the primary key Id_Anggota, which is equipped with the attribute Nama_Anggota and Status.

Field	Type (Length)	Information
Id_Waktu	Varchar(5)	Time Identity Number
Hari	Date	Borrowed Date
Minggu	Varchar(10)	Weekly
Bulan	Varchar(2)	Monthly (1, 2, 3,….. 12)
Tahun	Varchar(4)	Annual (2017, 2018,….)

Figure 9: Time Dimension.

The Time Dimension in Figure 9 serves to hold the data of borrowing and returning books, there are 5 attributes, with the primary key Id_Waktu and equipped with attributes of hari, minggu, bulan and tahun.

Field	Type (Length)	Information
Id_Anggota	Varchar(5)	Member Identity Number
Id_Waktu	Varchar(9)	Time Identity Number
Id_Kategori	Varchar(3)	Category Identity Number
Jml_Pinjam	Integer ()	Number of books borrowed

Figure 10: Borrowing Fact.

The Borrowing Facts in Figure 10 have the function to store the data of borrowing books from mem-

bers, there are 4 attributes, with a combined key (composite key) Id_Anggota, and Id_Waktu, Id_Kategori, which is equipped with attributes Jml_Pinjam.

Field	Type (Length)	Information
Id_Anggota	Varchar(5)	Member Identity Number
Id_Waktu	Varchar(9)	Time Identity Number
Id_Kategori	Varchar(3)	Category Identity Number
Jml_Pinjam	Integer ()	Number of books borrowed

Figure 11: Return Fact Table.

The Borrowing Facts in Figure 11 functions to collect the book return data from members, there are 4 attributes, with a combined key (composite key) Id_Anggota, and Id_Waktu, Id_Kategori, which is equipped with attributes Jml_Pinjam.

2.5 Database Duration Selection

In designing this library Data Warehouse, the duration of time used is 3 years. The data stored in this Data Warehouse will be the last 3 years of data.

3 RESULTS AND DISCUSSION

3.1 Making Facts on Borrowing Books

The results of the design of the fact of borrowing books, after being implemented into Pentaho can be seen in Figure 12. In the loan table consists of five attributes, namely: anggota_pinjam_buku associated with dimensi_anggota, waktu_peminjaman associated with dimensi_waktu, jum_buku_pinjaman associated with dimensi_buku, and the end node is tot-peminjaman.
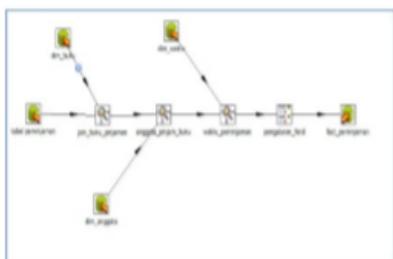

Figure 12: Formation of Book Borrowing Facts

3.2 Making Fact Returns Books

While the results of the design of the return of books after being implemented using Pentaho can be seen in Figure 13. In the return table consists of six attributes, namely: anggota_pinjam_buku which is connected

with dimensi_anggota, waktu_pengembalian associated with dimensi_waktu, jum_buku_pengembalian associated with dimensi_buku, associated with pengaturan_field and the end node is tot_pengembalian.

Figure 13: Making of Fact Returning Books

Anggota_kembali_buku associated with dimensi_anggota, waktu_pengembalian associated with dimensi_waktu, jum_buku_pengembalian associated with dimensi_buku, associated too with pengaturan_field, and the end node is tot_pengembalian.

4 CONCLUSIONS

From the results of the discussion in the previous chapter it can be concluded that the design of the Library circulation Data Warehouse with case studies in the Library can be designed, in 5 dimensions, namely: the time dimension, the member dimension, the publisher dimension, the author dimension, and the category dimension. Furthermore, it is implemented using 2 fact tables, namely: borrowing tables and book returns tables. Data warehouse that has been built can be utilized by interested parties by using certain software to produce information that can be used to support decision making.

The advice given for this research is so that it can be expanded even further, to produce a more detailed output in helping the decision making process that is more specific to the leadership.

REFERENCES

Andri, A. and Tujni, B. (2015). Analisis dan perancangan data warehouse perpustakaan (studi kasus: Perpustakaan universitas binadarma palembang). In *Seminar Nasional Informatika (SEMNASIF)*, volume 1.

Kurniawati, A. S. (2006). *Perancangan dan pembuatan aplikasi cluster analysis terhadap data sirkulasi buku di Perpustakaan UK. Petra*. PhD thesis, Petra Christian University.

Suprawoto, T., Sela, E. I., and Windarti, S. (2017). Prototipe integrasi data morbiditas pasien puskesmas kedalam data warehouse di dinas kesehatan kabuten bantul. *Jurnal TAM (Technology Acceptance Model)*, 7:1–10.

Supriyatna, A. and Wahyudi, M. (2012). Perancangan data warehouse pada perpustakaan bina sarana informatika. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.