Design of Data Warehouse for University Library using Kimball and Ross 9 Steps Methodology

Case Study: Udayana University Postgraduate Library

I Nyoman Darma Kotama[1*], Anak Agung Gede Oka Kessawa Adnyana[2], Komang Oka Saputra[3]

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

Abstract Udayana University has a library that purposed to serve all the student and lecturer needs in Post graduate program. To help achieve University goals to be a world class university, improvement of library management and assessment of its value and function is needed. Data Warehouse concept is a long proven solution that exist to help the upper management know the state of the institution through its own data. Effective design of Data Warehouse can help the management of institution to decide critical evaluation for its institution. On past research Kimball & Ross propose a technique to build an effective design of Data Warehouse named "9 Steps Methodology". This paper is aiming to build an effective design of Udayana Postgraduate Library Data Warehouse using Kimbal & Ross Methodology. We hope by this blueprint we can contribute to the future of the library by providing additional choice of Data Warehouse Design. After we follow through all the 9 steps and using past 3 years data of the library, we successfully designed model of Data Warehouse that consist of 8 dimension tables and 9 facts tables.

***Index Terms*— Data Warehouse, University Library Management, Library Data Warehouse, Kimball 9 Steps Methods.**

I. INTRODUCTION

Udayana University has a vision to become a world class university in 2025. In spirit to achieve that goals, some departments within itself should be efficiently designed and should be enabled to easily integrate with new technology. One of the departments is Udayana Postgraduate Library (UPL). UPL is one of department in Udayana University that provides the students and lecturers access to the books and journal about sciences.

With current state of technology that moves very fast, a library should state its effectiveness and values within the institution [1]. UPL is department that exist from 2001 [2] and until now it has to keep the technology and data up to date, especially in book inventory data.

Usage of data warehouse concept, is used almost in every big organization to help the management to manipulate or share its data. A Library that integrate data warehouse will have a blueprint and clear plan about how the future data and old data will exist within each other's[3]. Data Warehouse concept core ideas is not just for easy integration amongst data, but to help the upper managements to decide what business decision to do when the data is clearly visible and easily readable[4].

Kimball & Ross explain about how the effective Data Warehouse design is made using Kimball 9 Step Methodologies[5]. Many research has successfully implemented this method into its data warehouse design, such as designing data warehouse for a shop [6], designing a biological data mart [7], and designing an enterprise data warehouse system for helping decision support system [8].

With these facts we decided to propose a design of Data Warehouse from Udayana Postgraduate Library department using Kimball 9 Steps Methodologies.

II. Purpose Of Paper

We hope from this design, upper management can easily integrate new data and easily know what the data means, so they can decide what is better for the busines. Second, we hope from this blueprint, we can contribute to decision support system that in the future can be built to help the department catch up with current trends of enterprise technologies.

III. Literature Review

A. Data Warehouse

Data Warehouse is a collection of data that has several characteristics such as; separate, available, integrated, time-stamped, subject-oriented, non-volatile and accessible [9]. Under the same view, another research stated that Data Warehouse is a collection of data that built based on subject-oriented approach, integrated, non-updateable, time variant for data that has historical properties and to support the decision process [10].

Not only a collection of data, but Data Warehouse can be stated as a system that collects and consolidate data periodically from data source into the dimension table or database that already normalized. Data Warehouse is built to solve the technical and business problem that involves data creation and collection in order to make precise business decisions [11].

B. Kimball 9 Steps Methodologies

Kimball and Ross proposed a several steps of method to designing an effective Data Warehouse. They named it "9 Steps Methodologies", each steps consist with separate task that must be fulfilled in order to make Data Warehouse, and it was :

1. *Choosing the Process*
   In this part we choose the main subject that we are going to converts into data structure. This main subject is refer to a business process that has the main impact on whole business process.
2. *Choosing the grains*
   Here grains is refer to a data that can be represent from the fact table. Fact table is a collection of data that written into a table that come from choosing the main process and separating it from the noise.
3. *Identify and conforms the dimension*
   After we choose the grain, then we continue to make the dimension tables that connected with the fact tables. In this step, grain that chosen from the fact table, also represent grain that chosen from the dimension table.
4. *Choosing the facts*
   In this step we choose which fact table that we going to focus
5. *Store Pre-calculation in the Fact Table*
   Here we calculate and measure about the parameters and attributes that we going to use in the chosen facts table. By choosing and measuring attributes, we hope that every time the data logic happened, the attributes is well calculated and have minimum error margin.
6. *Round out the dimension tables*
   From all of the identified dimension, then we make a structured description about the chosen attributes and what it supposed to do.
7. *Choosing the Database Durations*
   After we determine about the data warehouse contents, in this steps we conclude how long the data from the business we are going to record and process, as example 3 years from 2016-2019.
8. *Determine the Need to Track Slowly Changing Dimension (SCD)*
   Chosen dimension can changed and causing problems in the futures, so to prevent that we need to define which variables, attributes and facts that has possibility to change overtime in this step.
9. *Design the Physical Database*
   On this last part, all the gathered facts, dimension and attributes is drawn into a Physical Database Model that represent the final product of the Data Warehouse Blueprint.

IV. Research Method

A. Research Methods

To get the best result, we separated our research methods into 5 tasks:

1. Literature Researches
   In this part we gather all the required information about Data Warehouse designs, tools and technologies that can help us to research this blueprint.
2. Interview with the Library Management
   We interviewed the library management and issuing a database request with range time 3 years.
3. Design and Implementation
   After we gather enough data about data warehouse and Kimball methods, we draw our blueprints into Physical Database Models and then analyze it
4. Result Analysis
   After we draw the Physical Database Model, then we analyze it to match the core problems of the Library
5. Conclusion
   In this part we write the conclusion of the research into a research paper.

V. Analysis and Results

A. Designing Library Data Warehouse Database

1 *Choosing the Process*

Choosing the process means determining what are the main subjects. Main subjects refer to business processes that will able to answer every important business question with specific characteristics to each process. In this research we choose 3 processes according to the library transactions, they are book lending, book returns, and book procurements. book lending and returning is the most transaction done by number. And the main business process in library environment.

2 *Choosing the grains*

In this step, we determine what will be represented by the fact tables. Defining grain on fact tables will affect grain on dimension tables.

We will represent book lending, book procurement, and book status to keep track about book collections in the library.

3 *Identify and conforms the dimension*

after we choose the fact tables, we also need to determine the dimension tables that are related to the fact tables. Dimensions are viewpoints to describe facts inside the fact tables. there are several dimensions we identified,

- Student dimension
- Study program dimension
- Book dimension
- Study field dimension
- Time dimension
- Writer dimension
- Author dimension
- Librarian dimension

4 *Choosing the facts*

Grains from fact tables define useable facts. In this step we measure what kind of information that will be required inside the fact tables. we select 3 facts, there are lending facts, book status facts, and procurement facts.

5 *Store Pre-calculation in the Fact Table*

In this step, every calculation result from an attribute will be considered whether saved or not inside the database. It is important to reduce errors every time we process a calculation on these attributes. We do some pre-calculations on every fact tables and their relations to every dimension. We done the pre-calculations all 3 facts table that we have been choose before based on every dimension related. This step will ease the ETL process because all dimension have been listed well.

6 *Round out the dimension tables*

Here we describing every dimension tables that related to our fact tables. according to our pre-calculations before, we will have 8 different dimension tables as our point of views.

7 *Choosing the Database Durations*

In this step we choose to store data from the last 3 years. This step is chosen based on library policy to track and analyze their performance based on years before.

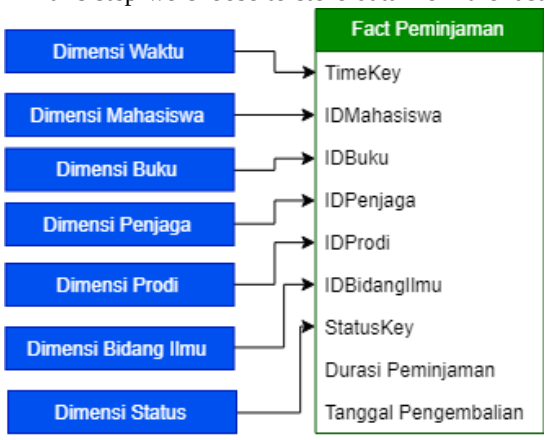

Fig. 1. Lending Fact Table Pre-calculation

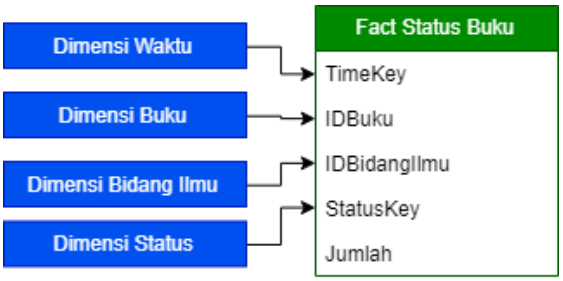

Fig. 2. Book status fact pre-calculation

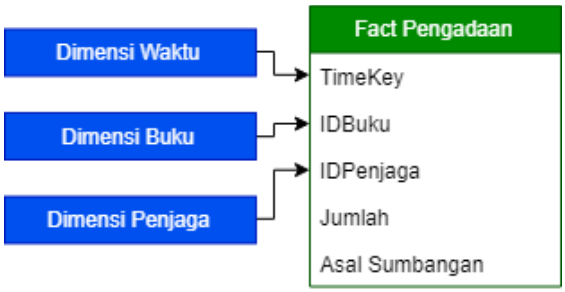

Fig. 3. Procurement fact pre-calculation

8 *Determine the Need to Track Slowly Changing Dimension (SCD)*

In this data warehouse, we chose to use 2 methods. First method is every time there is a change, we add a new record to easily track them. We use this on field study dimension and author dimension to easily track the change that happened. On the other hand, we chose to update existing record to a new record every time there is a change on program study dimension. That's because a program study change their program name sometimes, but still studying on the same field (possibly updated with new field of study) but we need to keep on track

Fig. 4. Library's dimension tables

from what program the book transaction happened.

9 *Design the Physical Database*

The final step is creating database based on all 8 steps before, after defining the fact tables, dimension tables, and relations between fact and dimension tables, this is the final design of data warehouse database of Udayana University Library described on fig. 5.

VI. Conclusion

Based on the design process that has been done, using 9 steps design methodology by Kimball and Ross in designing a data warehouse for Udayana University Library, there are several conclusions that can be made.

First, for data warehouse designing, besides the technical side and data analyzing, we also required to pay attention to what business targets from the library. So in the future, the data that has been gathered will be able to help on future decision making.

We found 9 steps design methodology will able to create a data warehouse that can fulfill the business requirements for a company that can be oriented on the present and the future.

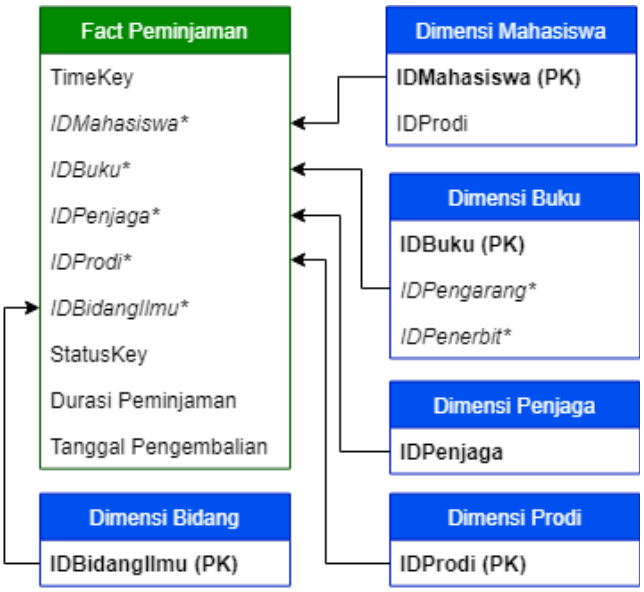

Fig. 5. Database design on Lending Fact

VII. Suggestion

From the steps that we follow through we successfully made a blueprint about Udayana University Post Graduate Library Data Warehouse. We suggest, future study around this topic is about implementing the Data Warehouse design into a real case, as example Data Mining, or integration into Decision Support System.

References

[1] J. R. Matthews, *The evaluation and measurement of library services*. ABC-CLIO, 2017.

[2] "UNUD | Pascasarjana - Sejarah Program S2 Magister Ilmu Ekonomi." [Online]. Available: https://fe.unud.ac.id/pascafeb/pages/view/sejarah-program-s2-magister-ilmu-ekonomi. [Accessed: 21-May-2019].

[3] A. Dahlan, E. Utami, and E. T. Luthfi, "Perancangan Data Warehouse Perpustakaan Perguruan Tinggi Xyz Menggunakan Metode Snowflake Schema," *J. Teknol. Inf. Respati*, vol. 8, no. 24, 2017.

[4] A. Amborowati, "Perancangan dan pembuatan data warehouse pada perpustakaan stmik amikom yogyakarta," 2009.

[5] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, and B. Becker, *The data warehouse lifecycle toolkit*. John Wiley & Sons, 2008.

[6] D. Subuh and W. Yasman, "Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho," *Pros. SENIATI*, vol. 5, no. 2, pp. 29–36, 2019.

[7] A. Kasprzyk, "BioMart: driving a paradigm change in biological data management," *Database*, vol. 2011, 2011.

[8] J. M. Firestone, *Enterprise information portals and knowledge management*. Routledge, 2007.

[9] S. Kelly, "Mining data to discover customer segments," *Interact. Mark.*, vol. 4, no. 3, pp. 235–242, Jan. 2003.

[10] W. H. Inmon, *Building the Data Warehouse*. New York, NY, USA: John Wiley & Sons, Inc., 1992.

[11] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.