Analysis And Design Of Data Warehouse And Data Mart Budget

University as higher education institutions must be able to manage budgets properly. The budget is a future financial plan which includes the expectations of university management. This research will design a data warehouse, which is a place where data can be stored on a large scale. In this research, a data warehouse will be designed as a place to store budget data. The method applied in this study is the Kimball method with a nine-step methodology. The result of this research is a data warehouse design and budget data mart.

KeyWords: Budget, Design, Data warehouse, Data mart, Nine-step methodology

1 Introduction

University is an institution of higher education and research which provides academic degrees in various fields[1]. The university as the highest institution providing the utmost education should be able to manage finances the greatest way possible. However, nowadays there are still many obstacles in the reality[2]. In realizing qualified education, it is necessary to have a thorough and professional management of accessible resources in university. One of the resources which need to be managed properly in higher education is financial problems[3].

Budget is a management tool in controlling, communicating, evaluating work, coordinating and motivating. An effective budgeting system facilitates the value creation process[4]. Budget management is a common problem in various fields of science. Problems occur due to the absence of excellent budget management, planned, measurable and the realization can be monitored. Almost at all levels of education, institutions, schools or colleges these problems occur[5]. The budget is also a reference for company leaders in running their company in the direction that has been planned and as a control tool to find out within a certain period of time until the implementation is appropriate to plan based on a predetermined budget[6]. Annual budget and expenditure planning activities are routine activities that must be prepared for each unit[7].

Information technology makes it easy for users to carry out all activities, one of which is the dissemination of information[8]. An information system is a series of activities carried out to obtain information which will support the formation of a decision and provide reports to other parties as needed[9].

Recently, the management of the budget plan for Setia Budi University, Surakarta, still applies the old system, that is doing bookkeeping in physical form and Microsoft Excel for each work unit. This can cause problems, such as differences in budget data from each work unit, financial bureau and foundation in one fiscal year. One solution to overcome this problem is the implementation of a data warehouse. The data warehouse is important in terms of strategic decision making with its ability to integrate heterogeneous data from several sources of information in a common storage space, for querying and analysis[10]. Therefore it is necessary to design a data architecture in the form of a data warehouse that can integrate transaction data from various storage sources into a set of historical data which can be processed into qualified information which is prepared to apply and certainly does not take long[11].

This study applies the Kimball method with a nine-step methodology. This method has the advantages of making and designing data warehouses faster, star schemas which are easy to follow, easy management of data warehouse systems and effective database operations.

This Article was:
submitted: 19-06-23
accepted: 30-06-23
publish on: 14-07-23

How to Cite:
2 Background

2.1 Budget. Budget is the amount of money spent in a certain period to implement a program, while spending is the acquisition of goods or services from sellers with the aim of buying at that time[9]. The budget is a future financial plan that includes management’s expectations of income, expenses and other financial transactions within one year. The budget preparation stage is a very significant stage because budgets which are ineffective and not performance oriented can in fact fail programs which have been prepared before[6]. Therefore it can be defined that the budget is an amount of money or funds used in a certain period to obtain goods and services carried out by individuals or an agency. Procedure is a series of related tasks which are applied to ensure identical execution of work. The budget preparation procedure is divided into the following stages: the planning guideline determination stage, the budget preparation stage, the budget determination stage, and the budget implementation stage[12].

2.2 Data Warehouse. Data Warehouse is a system which is used to do extraction, cleansing, adjustment, and provide source data into dimensional data storage and then employ it to support and implement questions and analysis for decision making purposes [13]. The data warehouse is designed for analysis rather than transactional processing, and usually contains historical data derived from transaction data, but can contain data from other sources[10]. The data warehouse has its own different characteristics, including[14]:

(1) Subject Oriented which means that the data warehouse is designed to analyze data based on certain subjects in the organization, not on application processes or functions.
(2) Integrated explicitly being able to store data that approach from separate sources into a format which is consistent and integrated each other.
(3) Time – variant or time range, specifically the time span of a data related to the accuracy of the data.
(4) Non – volatile to be exact data in the data warehouse cannot be updated in real time but is refreshed from the operational system on a regular basis.

In designing a data warehouse, an appropriate architecture must be determined. The data applied for the data warehouse must be extracted from one or more data sources, then converted into a form which is easy to analyze and consistent with the existing data in the data warehouse, and finally loaded into the data warehouse. This is called the ETL process[15]. The processes in data warehouse development are ETL (Extract – Transform – Loading). Extract is the entire process related to data collection and collection (including data from outside which is needed). Transform is the process of preparing data to ensemble the needs such as cleaning, decoding and integration. Loading is the process of storing and organizing data in an existing data warehouse structure[16].

2.3 Data Warehouse Design Methods. The data warehouse design method, according to Kimball, includes 9 stages known as the nine-step methodology[17]. The nine stages as follow:

(1) Choosing the process
(2) Choosing the grain
(3) Identifying and conforming the dimensions
(4) Choosing the facts
(5) Storing Pre-Calculation in The Fact Table
(6) Rounding Out The Dimension Tables
(7) Choosing The Duration of Database
(8) Tracking Slowly Changing Dimension
(9) Deciding The Query Priorities and The Query Models

3 Research Methodology

3.1 Literature Review. In previous research by (Khotimah and Sriyanto, 2016)[11] entitled Design and Implementation of Data Warehouses to Support Academic Systems (Case Study on Muhammadiyah Kotabumi Stkip) published in 2016, created a data warehouse to support the academic system by using STKIP Muhammadiyah Kotabumi Lampung academic data. Then research by (Andri and Baibul, 2015)[18] entitled Analysis and Design of Library Data Warehouses (Case Study: Binardarma University Palembang Library) published in 2015, created a library data warehouse using library data from Binardarma University Palembang.

From several previous literature studies, no one has discussed data warehouses for budgets. Therefore, it becomes the basis for research on Data Warehouse Analysis and Design and Budget Data Marts (Case Study: Setia Budi University, Surakarta).

3.2 Tools and Materials. The tools used in this study are:

(1) Laptop with AMD Ryzen 3 3200U processor specifications with Radeon Vega Mobile Gfx 2.60 GHz RAM 8 GB, Windows 11 64-bit operating system
(2) Microsoft Excel 2016 as an application to process data
(3) Apache Hive application for data warehouses
(4) Python and Google Colab as applications for converting data

The material used is data on budget plans for work units or faculties in Setia Budi University, Surakarta and applying data from several hotels which provide meeting facilities as an example of data from outside Setia Budi University.

3.3 Data Collection. Data collection was carried out by looking at work unit or faculty budget plan data in Setia Budi University Surakarta for the 2017/2018 fiscal year and taking data from the websites of several hotels that provide meeting facilities as an example of data from outside Setia Budi University.

3.4 Research Procedure.

• System analysis and design

The data warehouse designs applied in this study makes use of the Nine-Step Methodology [17]. In this method there are nine steps in designing a Data Warehouse, as follows:

(1) Choosing The Process
(2) Choosing The Grain
(3) Identifying and Conforming The Dimensions
(4) Choosing The Fact
(5) Storing Pre-Calculation in The Fact Table
(6) Rounding Out The Dimension Tables
(7) Choosing The Duration of Database
(8) Tracking Slowly Changing Dimension
(9) Deciding The Query Priorities and The Query Models

• Implementation

At the implementation stage, it is carried out by identifying parties who have an interest in the budget. From the identification results, there are several parties who have an interest in the budget, as follows:

(1) Work units or faculties to propose budgets and request budgets that have been set
(2) Finance bureau to verify budget proposals and budget requests
(3) Foundation to determine budget proposals and conduct disbursement of budget requests

From the parties which have these interests require the similar information regarding the budget proposal data, budget determination data and budget request data. By using the ETL process, the budget data pipeline is still separated in each work unit or faculty, finance bureaus and foundations can be moved into one data warehouse.
4 Result and Discussion

4.1 Data Warehouse Preparation. The data warehouse design process will be carried out using a nine-step methodology. The process involved in creating a data warehouse design is as follows:

1. Choosing The Process
   Business processes which occur in the budget includes:
   (a) Submission of budget proposals
   (b) Determination of budget proposals
   (c) Budget request

2. Choosing The Grain
   Grains are data from prospective facts which can be analyzed. Based on the existing business processes, a grain is determined which describes the fact table as follows: includes the number of work units, the number of activity descriptions, the amount of budget time, and the number of budget requests.

3. Identifying and Conforming The Dimensions
   The third step in designing the Data Warehouse is identifying the dimensions associated with the fact table. From the identification results, it can be determined that the dimensions involved include:
   (a) Work unit dimensions
   (b) Description of activities dimensions
   (c) Budget time dimension
   (d) Budget request dimensions

4. Choosing The Fact
   Choose the facts used in the data mart. Each fact has data which can be calculated, and can later be displayed in the form of reports, graphs or charts. Here are the facts that users will utilize:
   - The number of work units includes unit_code, unit_name.
   - The number of activity descriptions includes activity_code, unit_code, activity_name, activity_type, nominal, time_code.
   - Total budget time includes time_code, month, year.
   - The number of budget requests includes application_code, unit_code, activity_code, activity_name, nominal, time_code.
   - The number of meeting_places includes place_name, room_name, capacity, package, price.

5. Storing Pre-Calculation in the Fact Table
   The aggregation in the fact table for submitting budget proposals is the total number of budget proposals based on activity descriptions. The fact aggregation of budget determination is the total budget determination based on activity descriptions. Moreover, the fact aggregation of budget requests is the number of budget requests based on activity descriptions.

6. Rounding Out The Dimension Tables
   Adding a text description on the dimension table. The description is conveyed so that it can be understood by the user.

7. Choosing The Duration of Database
   The data to be managed in the data warehouse is the budget plan for the 2017/2018 fiscal year.

8. Tracking slowly changing dimension
   The attributes in the dimension table do not always have a fixed/static value. The value changes in the dimension table may alter in quite a long time. Therefore it is necessary to update the dimension table to maintain the accuracy and consistency of the data.

9. Deciding the query priorities and the query modes
   In this process, what is done is to consider the influence on the physical design, such as the existence of summaries and aggregates as well as ETL (Extract, Transformation Loading), backup, and security processes that provide limits to the user are factors which must be considered.

4.2 Star Schema. The following star schema is created for data warehouse and budget data mart designs.

4.3 Data Preparation. The process in the data warehouse development is ETL (Extract – Transform – Loading).

4.4 Apache Hive Implementation. The next step is to prepare the processed data in .csv format. The data is entered into the
data warehouse that will be created. Next, a data warehouse table structure is created based on the dimensions and facts in accordance with those that have been finished at the system analysis and design stage. Table of dimensions and facts and their attributes in this study include:

1. **Work_Unit Table**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit_Code</td>
<td>String</td>
<td>Work Unit Code</td>
</tr>
<tr>
<td>Unit_Name</td>
<td>String</td>
<td>Work Unit Name</td>
</tr>
</tbody>
</table>

2. **Activity Description Table**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity_Code</td>
<td>String</td>
<td>Activity Code</td>
</tr>
<tr>
<td>Unit_Code</td>
<td>String</td>
<td>Work Unit Code</td>
</tr>
<tr>
<td>Activity_Name</td>
<td>String</td>
<td>Activity Name</td>
</tr>
<tr>
<td>Activity_Type</td>
<td>String</td>
<td>Activity Type</td>
</tr>
<tr>
<td>Nominal</td>
<td>Integer</td>
<td>Activity Nominal</td>
</tr>
<tr>
<td>Time_Code</td>
<td>Integer</td>
<td>Budget Time Code</td>
</tr>
</tbody>
</table>

3. **Budget Timetable**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time_Code</td>
<td>Integer</td>
<td>Budget time code</td>
</tr>
<tr>
<td>Month</td>
<td>String</td>
<td>Month Budget</td>
</tr>
<tr>
<td>Year</td>
<td>Integer</td>
<td>Year Budget</td>
</tr>
</tbody>
</table>

4. **Budget requests Table**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request_Code</td>
<td>String</td>
<td>Request code</td>
</tr>
<tr>
<td>Unit_Code</td>
<td>String</td>
<td>Work unit code</td>
</tr>
<tr>
<td>Activity_Code</td>
<td>String</td>
<td>Activity code</td>
</tr>
<tr>
<td>Activity_Name</td>
<td>String</td>
<td>Activity name</td>
</tr>
<tr>
<td>Nominal</td>
<td>Integer</td>
<td>Activity nominal</td>
</tr>
<tr>
<td>Time_Code</td>
<td>Integer</td>
<td>Budget time code</td>
</tr>
</tbody>
</table>

5. **Emeeting_Place Table**

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place_Name</td>
<td>String</td>
<td>Place Name</td>
</tr>
<tr>
<td>Room_Name</td>
<td>String</td>
<td>Room Name</td>
</tr>
<tr>
<td>Capacity</td>
<td>Integer</td>
<td>Room Capacity</td>
</tr>
<tr>
<td>Package</td>
<td>String</td>
<td>Meeting Package</td>
</tr>
<tr>
<td>Price</td>
<td>Integer</td>
<td>Price</td>
</tr>
</tbody>
</table>

After the data warehouse table structure is formed, the next step is to start the Apache Hive application.

1. The first step is to start running the Apache Hadoop application

(2) The second step is to run the Apache Derby application

(3) The third step is to initialize the metastore data on Apache Hive

(4) The fourth step is to run Hive Server 2

(5) The fifth step is to run the Apache Hive application

After the Apache Hive application is ready to employ, the design of a financial data warehouse can be executed.

1. Create a financial database
2. Create a Work_Unit table
3. Create an Activity_Description Table
4. Make a Request_Budget Table
5. Create the Emeeting_Place Table
6. Load data on each table which has been created

From the data which has been entered into each table, it can then be used to obtain budget information as needed. For example, if you need budget information for one work unit in one month. Then the information will be displayed as below.

![Example of Work Unit Budget Information](image)

### 4.5 Prototype Data Warehouse Test.

(1) **Data Completeness**: Ensuring that all data from various sources is loaded into the data warehouse. The total number of records uploaded from the source system must match the total number of records loaded into the data warehouse[19]. This test aims to ensure the entirety of data from sources which go through the data warehouse.

<table>
<thead>
<tr>
<th>No</th>
<th>Data Source Transformation Name</th>
<th>Data Warehouse Table Name</th>
<th>Number of Source Records</th>
<th>Number of Data Warehouse Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>work_unit_data.csv</td>
<td>work_unit</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>activity_description_data.csv</td>
<td>Activity Description</td>
<td>2764</td>
<td>2764</td>
</tr>
<tr>
<td>3</td>
<td>budget_time_data.csv</td>
<td>Budget Time</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>get_request_data.csv</td>
<td>Get Request</td>
<td>1120</td>
<td>1120</td>
</tr>
<tr>
<td>5</td>
<td>emeeting_place_data.csv</td>
<td>Emeeting Place</td>
<td>81</td>
<td>81</td>
</tr>
</tbody>
</table>

It can be seen that the number of records in the data source transformation is identical the number of records in the data warehouse.

(2) **Data Transformation**: The process of preparing data to suit the needs such as cleaning, decoding and integration[19]. This test aims to ensure the data transformation process which will be applied in the data warehouse.
<table>
<thead>
<tr>
<th>No</th>
<th>Data Source Name</th>
<th>Data Source Transformation Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>work_unit_data.xlsx</td>
<td>work_unit_data.csv</td>
</tr>
<tr>
<td>2</td>
<td>activity_description_data.xlsx</td>
<td>activity_description_data.csv</td>
</tr>
<tr>
<td>3</td>
<td>budget_time_data.xlsx</td>
<td>budget_time_data.csv</td>
</tr>
<tr>
<td>4</td>
<td>budget_request_data.xlsx</td>
<td>budget_request_data.csv</td>
</tr>
<tr>
<td>5</td>
<td>emeeting_place_data.xlsx</td>
<td>emeeting_place_data.csv</td>
</tr>
</tbody>
</table>

Table 7 Data Transformation

Data Transformation can be seen from the transformation of data from data sources of type .xlsx to .csv

(3) Data Quality: The data warehouse (ETL) system must ensure the quality of the data loaded into it by rejecting (or correcting the data)[19][20]. This test aims to test the quality of the data that will be entered into the data warehouse.

<table>
<thead>
<tr>
<th>No</th>
<th>Data Source Name</th>
<th>Transformation Name</th>
<th>Loading Process to Data</th>
<th>Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>work_unit_data.csv</td>
<td>activity_description_data.csv</td>
<td>Succeeded</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>activity_description_data.csv</td>
<td>Succeeded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>budget_time_data.csv</td>
<td>budget_request_data.csv</td>
<td>Succeeded</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>budget_request_data.csv</td>
<td>emeeting_place_data.csv</td>
<td>Succeeded</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 Data Quality

In table 8 Data Quality can be seen that the process of loading data into the data warehouse was successful or received from the data source.

(4) Scalability and Performance: The data warehouse must ensure system scalability with increasing loads. Hence, there should be no performance degradation when executing queries, with results anticipated within a certain time frame[19][21]. This test aims to test the scalability and performance when running queries.

<table>
<thead>
<tr>
<th>No</th>
<th>Query</th>
<th>Proses Query (detik)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SELECT * FROM work_unit;</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>SELECT * FROM activity;</td>
<td>0.575</td>
</tr>
<tr>
<td>3</td>
<td>SELECT * FROM budget_time;</td>
<td>0.48</td>
</tr>
<tr>
<td>4</td>
<td>SELECT * FROM budget_request;</td>
<td>0.788</td>
</tr>
<tr>
<td>5</td>
<td>SELECT * FROM emeeting_place;</td>
<td>0.578</td>
</tr>
</tbody>
</table>

Table 9 Query Performance

Query Performance table shown that the average queries processing time in the data warehouse is less than 1 second. This test illustrates that the average time needed for query execution is 1 second[22].

5 Conclusion

Based on the results of the analysis performed, it can be concluded that:

(1) Creating a data warehouse and budget data mart using a nine-step methodology. The fundamental to this method is the selection of business processes, the determination of the dimension table and the determination of the fact table.

(2) Data integration is carried out by extracting from several data sources by taking the required data items. Data sources were obtained from work unit or faculty budget data in Setia Budi University and the sample data from external Setia Budi University.

(3) The data integration process begins with extracting the data obtained from the data which has been collected on Google Drive and followed by transforming it based on system analysis and design. Furthermore, the data in the appropriate format is entered into the data warehouse through the load process.

(4) Data warehouse testing applies four parameters, as follows: Data Completeness, Data Transformation, Data Quality, Scalability and Performance.

References


[13] D. Sugarto, H. Leslie Hendri Spits Warnars, and Winarno, “Perancangan Data Warehouse Penujian (Studi Kasus PT. Subafood


