The Development of Data Warehouse for Contractor System

Wimora Sarwindo, Abba Suganda Girsang

Abstract — In the middle of the crowded infrastructure of facility and infrastructure while the data of project budget has been increasing every day, PT XYZ requires new technology of data warehouse in order to respond the needs of recording and extracting project budget data quickly and accurately because Ms Excel, recent software used by managing-data staff to store data manually, is fairly time-consuming to record and extract daily realization of project budget. Therefore, in order to help the company to learn the result of the project budgets that have already been spent for the various needs of project construction, the technology of data warehouse should be present in order to extract and record data more consistently and time-saving. Data warehouse technology to support the development of Data Warehouse is Pentaho because Pentaho is database application that is supported with database MYSQL and OLAP feature as a means to analyze and process data contained in multidimensional tables once the process of ETL is executed. ETL means the process of selecting and extracting data from one source into the designated source in order to produce better-quality data result. OLAP is developed by using pivot 4j view integrated by pentaho server to map the structure of fact and dimension table properly so that, the tables can get a proper connection to display desired information in term of graphs for the need of dashboard display. The dashboard that is coded through mxd queries in CDE Dashboard shows the final result of either monthly or yearly project budget obviously and reliably so that the company can learn the report and make decision that needs to be taken to control the budget costs and maintain the company profit. Both CDE Dashboard and pivot 4j view are integrated with pentaho server. Finally, report is presented by using pentaho report designer as a means to extract needed data more rapidly and consistently because data are called through id stored in different table, but the same field.

Keywords : Data Warehouse, Pentaho, MYSQL, Realization of Project Budget CDE Dashboard.

I. INTRODUCTION

Construction service is a part of important activity in Indonesia's development in order to increase participants not only from outside the city, but also from overseas. The activities of construction projects consist of energy, mining, transportation and airports, manufacturing, tourism, electrical and mechanical installations, stations, defense, etc. Project construction activities have been bringing positive impacts in Indonesia which these activities not only increase Indonesia's revenue, but also provide opportunities for project teams to develop so that the project team can become part of Indonesia's infrastructure development.

However, the budget that needs to be prepared both through government and private sector for each year is very costly due to many reasons such as quality of materials, quantity of materials needed, etc.

PT. XYZ is the company that runs the project field. The recent technology that the company applies to input data containing daily budget from every contractors is Microsoft Excel. However, The main problem for PT XYZ is that Microsoft Excel, which is still used by users for the recording of project budget data, doesn’t help the user to record the project budget data optimally, especially when recording and searching the budget needs for various areas that carry out different project. As a result, it tends to take much longer time, afflict lack of accuracy in recording data based on time history of the project budget while the budget data keeps coming every time.

Therefore, the design and implementation of data warehouse is provided to answer the needs of project report data from the beginning to the end.

The purpose of this case study is to help user record the project budget report more accurately. The benefit of this study case is to help the company can learn more and review the result of the report on how much the project budget has been issued. After the report is reviewed, the head of project management will then make a decision in order to control the cost of project budget needed for the future. The method of developing data warehouse is by using Kimball method as a solution to analyze and process data starting from:

a. ETL (Extract, Transform, Load) to convert database properly into data warehouse.
b. Mapping of schema by using the method of OLAP (Online Analytical Processing) through pentaho server integrated with pivot4j view.
c. CDE Dashboard that integrated with pentaho server as well to display the total of monthly and yearly project budget in graphical view.
d. The process of presenting data as information through report and dashboard.

II. ACTIVITY STEPS

Framework of thought is a step-by-step explanation of each step in conducting research based on literature reviews and relevant or related research results ranging from the topic of the related problem to the results of the solution of the research conducted.
The Development of Data Warehouse for Contractor System

The purpose of the literature review is to support logical arguments from the result of research made by individual researcher. Here are several step-by-step phases to make data warehouse designed properly.

A. Studying Introduction and Literature

The introduction is a kind of adaptation to learn initial business process. Meanwhile, literature is a clear and specific learning method in order to encourage the curiosity about the boundaries and problems of research and sources for conducting research.

B. Learning Company Profile

This step is the introduction of company profile which the field of business process has been running.

C. Identifying Beginning Condition

This step is to identify the technology equipped to record daily data such as Microsoft Excel.

D. Interviewing User Needs

This step is to get the response to reach when new technology of data warehouse is developed. This step also describes some requirements to solve the problem of recent condition.

For example, when the car owner goes to a car service, the owner’s car will meet the repairman to fix some car’s problems such as the fog lamp doesn’t light, car’s battery runs out, parking sensor sound terribly, and so forth.

The example of the illustration above shows that the author plays the role as a repairman. Meanwhile, the company plays the role as car owner.

E. Collecting Data

In this step, data are collected and processed through the proposed method. Therefore, Entity Relationship Diagram is designed to map the field of data needed for each table.

F. Designing Star Schema

This step is a bridge in order to develop the advanced data warehouse to make data into certain information.

G. Executing Advanced Development of Data Warehouse

This step is to convert data with ETL (Extract, Transform, Load) method, structure the table dimensions by using pentaho schema workbench, OLAP Cube to provide facilities for interactive and complex query requests for users by using pentaho server collaborating with pivot4j, and finally, the report result is presented by using pentaho report designer 7.

H. Evaluating Result of Developing Data Warehouse

This step is to get the response by data-managing staff once new technology of data warehouse is developed. Related to point D which the fourth step is interviewing the user needs, the purpose of this step is to describe how positive the impact of data warehouse usage once the company experiences to use it.

This step is parable of the result of recent condition once the car is already repaired thoroughly. The result of the test once the car is already repaired can be varying such as well-repaired, repaired with condition, or some parts of car aren’t still repaired completely.

I. Making Conclusion

The final step is to make the conclusion once new technology of data warehouse is developed in order to help the company get desired data. Meanwhile, suggestion is written as an evaluation to identify the weakness in research. The weakness in research will be revised in the future so that the weakness in the future research would be improved while the good research will be maintained.

III. PROPOSED METHOD

In this chapter, the process of developing data warehouse to make data become better report to help company learn the result of data. Processes are needed as below:

A. ETL Process

In this chapter, the first step of developing data warehouse is ETL (Extract, Transform Load) process. ETL process will be executed for each dimension by filtering and transforming previous data into data that has been transformed. The software to process the ETL is *pentaho data integration 6.1*. Therefore, this data needs two databases to support ETL process. Here are two databases provided below:

a. *dbbeforeetl*. This database contains all of data that have not been processed through ETL. After that, all of data are stored in Microsoft Excel, this file must be stored in type of *CSV (CommaDelimited)*. Finally, this file can be imported in *phpmyadmin MYSQL* to be translated into database. This step follows the flow of Entity Relationship Diagram (as shown in *Fig. 1*) as the initial step before data is processed through ETL.

b. *dbafteretl*. This database contains all of data that have been processed through ETL. This step follows the flow of Star Schema (as shown in *Fig. 2*) as the next process to develop data warehouse.
Here are five dimension tables stored in dbaferetl database once conversion of data from excel to MYSQL is already executed:

a. This is the process of ETL for dim_kontraktor as shown in Fig. 3.

Fig. 3 ETL dim_kontraktor

b. This is the process of ETL for dim_supplier as shown in Fig. 4.

Fig. 4 ETL dim_supplier
c. This is the process of ETL for dim_kota as shown in Fig. 5.

Fig. 5 ETL dim_kota
d. This is the process of ETL for dim_time as shown in Fig. 6.

Fig. 6 ETL dim_time
e. This is the process of ETL for dim_proyek as shown in Fig. 7.

Fig. 7 ETL dim_proyek
f. This is the process of ETL for fact_anggaran as shown in Fig. 8.

Fig. 8 ETL fact_anggaran

to map every structure of table as dimension by using pentaho schema workbench. In this mapping process, fact table will have to connect to every dimension by defining foreign key.

2) Hence, the foreign key in fact table will connect to primary key in dimension table in order to adjust the content of data in accordance with id from each dimension, so that the result of data will be presented properly.

3) Here are five schema structures executed for each table from hierarchy below the dimension to level below the hierarchy:

a. This is the process of structuring fact for fact_anggaran as shown in Fig. 9.

Fig. 9 The Structure of fact_anggaran

b. This is the process of structuring dimension for dim_kontraktor as shown in Fig. 10.

Fig. 10 The Structure of dim_kontraktor
c. This is the process of structuring dimension for dim_supplier as shown in Fig. 11.

d. Fig. 11 The Structure of dim_supplier
e. This is the process of structuring dimension for dim_time as shown in Fig. 12.

B. Structuring Schemas

1) When ETL process has been executed properly for each dimensions and 1 fact through transformation, the next step is...
The Development of Data Warehouse for Contractor System

f. This is the process of structuring dimension for dim_kota as shown in Fig. 13.

g. This is the process of structuring dimension for dim_proyek as shown in Fig. 14.

Fig. 12 The Structure of dim_time

Fig. 13 The Structure of dim_kota

Fig. 14 The Structure of dim_proyek

Special for OLAP process for fact_anggaran, It needs a cube as the parent while the child of the cube is dimension usage. The function of dimension usage is as a relation to call the dimension whose source is located outside the cube based on id as primary key according to the respective dimension table inside the cube.

After that, pivot 4j view supported by pentaho server 7 makes the process of OLAP Cube more easily, especially to let the user get the access to capture data. Finally, the dashboard is presented to display graphical information about monthly and yearly project budget.

C. OLAP Cube

Here are the executions of OLAP Cube through pivot 4j view integrated with pentaho server:

1. OLAP Cube for Supplier as shown in Fig. 15.

2. OLAP Cube for Time as shown in Fig. 16.

3. OLAP Cube for Contractor as shown in Fig. 17.

4. OLAP Cube for City as shown in Fig. 18.
5. OLAP Cube for Project as shown in Fig. 19.

IV. THE RESULT OF ANALYSIS

A. Sales Report

After the cube and dimensions are well-structured, Pentaho Report Design 7 is used to display more interactive report about trends of project. The SQL command will provide the desired data to be displayed budget in term of supplier, contractor, project, time, and city.

Here are several results to present desired data for the company to present data by using SQL syntax in order to select desired data:

a. Data for Showing Top 10 of Contractor’s Project Budget as shown in Fig. 20.

And here is data report of budget for Contractor as shown in Fig. 21.

### Top 10 Anggaran Proyek Untuk Kontraktor

<table>
<thead>
<tr>
<th>ID Kontraktor</th>
<th>Nama Kontraktor</th>
<th>Total Anggaran</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB0144</td>
<td>PT. ADHI KARYA (Persero), Tbk.</td>
<td>13,033,815,000</td>
</tr>
<tr>
<td>DB0143</td>
<td>PT. FORZA PROPERTI SERPONG</td>
<td>11,523,840,000</td>
</tr>
<tr>
<td>DB0146</td>
<td>PT. NUSA RAYA CIPTA, Tbk.</td>
<td>10,484,395,000</td>
</tr>
<tr>
<td>DB0135</td>
<td>PT. TRIKON ABADI SEHTERA</td>
<td>7,106,470,000</td>
</tr>
<tr>
<td>DB0147</td>
<td>PT. HIRO MINDO PERKASA</td>
<td>6,206,226,000</td>
</tr>
<tr>
<td>DB0131</td>
<td>PT. DELSON MANDIRI SEHTERA</td>
<td>5,240,205,000</td>
</tr>
<tr>
<td>DB0141</td>
<td>PT. WASKITA KARYA (Persero)</td>
<td>4,964,735,000</td>
</tr>
<tr>
<td>DB0148</td>
<td>PT. CSCEC - SAR IJO</td>
<td>4,741,235,000</td>
</tr>
<tr>
<td>DB0127</td>
<td>PT. GEOHAS UTAMA INDONESIA</td>
<td>4,738,822,500</td>
</tr>
<tr>
<td>DB0145</td>
<td>PT. PINTANG SEWU SEHTERA</td>
<td>4,246,330,000</td>
</tr>
</tbody>
</table>

b. Data for Showing Top 5 of Supplier’s Project Budget as shown in Fig. 22.

### Chart Top 5 Anggaran Proyek Untuk Supplier

And here is data report of budget for Supplier as shown in Fig. 23.

### Fig. 18 OLAP Cube For Sum of Project Budget For City Since 2010

### Fig. 19 OLAP Cube For Sum of Project Budget For Project Since 2010

### Fig. 20 Top 10 of Contractors Who Have Most Project Budget

### Fig. 21 Data of Top 10 of Contractors Who Have Most Project Budget

### Fig. 22 Top 5 of Suppliers Which Have Most Project Budget

### Fig. 23 Data of Top 5 of Suppliers Which Have Most Project Budget
The Development of Data Warehouse for Contractor System

Top 5 Anggaran Proyek Untuk Supplier

<table>
<thead>
<tr>
<th>ID Supplier</th>
<th>Nama Supplier</th>
<th>Total Anggaran</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR005</td>
<td>Building Materials Store Kuda Mas PD</td>
<td>75,752,185,000</td>
</tr>
<tr>
<td>BR006</td>
<td>Toko Talang Patinda Adv</td>
<td>17,765,460,100</td>
</tr>
<tr>
<td>BR007</td>
<td>Toko Jaya aluminium &amp; kaca</td>
<td>16,888,512,500</td>
</tr>
<tr>
<td>BR002</td>
<td>Toko Besi Permata</td>
<td>11,586,465,000</td>
</tr>
<tr>
<td>BR011</td>
<td>PT. Semco Baturaja Persero</td>
<td>10,226,005,000</td>
</tr>
</tbody>
</table>

Fig. 23 Data of Top 5 of Suppliers Which Have Most Project Budget

c. Data for Showing Top 10 of City’s Project Budget as shown in Fig. 24.

Top 10 Anggaran Proyek Untuk Kota

<table>
<thead>
<tr>
<th>ID Kota</th>
<th>Nama Kota</th>
<th>Total Anggaran</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG0001</td>
<td>Jakarta</td>
<td>22,815,240,000</td>
</tr>
<tr>
<td>AG0002</td>
<td>Tangerang</td>
<td>19,035,905,000</td>
</tr>
<tr>
<td>AG0022</td>
<td>Mataram</td>
<td>17,970,115,000</td>
</tr>
<tr>
<td>AG0021</td>
<td>Denpasar</td>
<td>10,865,535,000</td>
</tr>
<tr>
<td>AG0003</td>
<td>Bali</td>
<td>9,808,133,000</td>
</tr>
<tr>
<td>AG0008</td>
<td>Bandung</td>
<td>9,662,507,500</td>
</tr>
<tr>
<td>AG0006</td>
<td>Bogor</td>
<td>9,258,031,000</td>
</tr>
<tr>
<td>AG0004</td>
<td>Karawang</td>
<td>6,558,238,500</td>
</tr>
<tr>
<td>AG0011</td>
<td>Sukabumi</td>
<td>5,418,390,000</td>
</tr>
<tr>
<td>AG0007</td>
<td>Palembang</td>
<td>4,862,620,000</td>
</tr>
</tbody>
</table>

Fig. 25 Data of Top 10 of Cities Which Have Most Project Budget
d. Data report for showing top 10 of Project’s Project Budget as shown in Fig. 26.

Fig. 24 Top 10 of Cities Which Have Most Project Budget

Chart Top 10 Anggaran Proyek Untuk Kota

Fig. 26 Top 10 of Projects Which Have Most Project Budget

And here is data report of budget for project as shown in Fig. 27.

Fig. 25 Data of Top 10 of Cities Which Have Most Project Budget

Fig. 26 Top 10 of Projects Which Have Most Project Budget

And here is data report of budget for project as shown in Fig. 27.
Fig. 27 The Extraction of Contractor’s Project Budget On March, 2012

e. Chart for showing yearly budget since 2010-2019 as shown in Fig. 28.

Fig. 28 Graph of Accumulated Yearly Project Budget Since 2010 – 2019

And here is data report of yearly project budget as shown in Fig. 29.

Fig. 29 Data of Accumulated Yearly Project Budget Since 2010 – 2019

f. Chart for showing monthly project budget in 2013 as shown in Fig. 30.

Fig. 30 Graph of Accumulated Monthly Project Budget For Project In 2013

And here is data report of monthly project budget for projects that have been accumulated for every month since January to December, 2013 as shown in Fig. 31.
In order to make the graphical information show the desired output, its methods can be used by using:

1) **MDX Queries.** This method is used by copying and pasting the code showed at **pivot 4j view** once the structure of fact and dimension tables have already been structured by using **pentaho schema workbench**.

2) **SQL Queries.** This method is used by writing code manually based on formulas for joining tables, summing total, grouping tables, and so forth.

In this journal, the method of sql query to display dashboard is **SQL Queries.** Here are some dashboards to display:

a. The display of dashboard above on January, 2014 as shown in **Fig. 34**.

![Fig. 34 The Dashboard Above To Show Line Chart On January, 2014](image)

The line chart in **Fig.34** shows the total of monthly project budget starting from January until December, 2014. The position of line in this **line chart** will only change when the **year** is selected. The line in this **line chart** won’t change when the **month** is selected because **month** will react to a couple of **pie charts** (in **Fig. 35**) located below the **line chart** to change their results.

b. The display of dashboard above on January, 2014 for project and city as shown in **Fig. 35**.

![Fig. 35 The Dashboard Below To 2 Pie Charts On January, 2014 For Project And City](image)

A couple of pie charts shown in **Fig. 35** consist of:

1) **Top 10 Proyek.** This pie chart shows the total of project budget based on the name of projects which hold the most project budget on January 2014.

2) **Trend Lokasi Proyek.** This pie chart shows all of locations where the total of project budget hold the most project budget on January 2014 due to the activity of projects in certain location.

C. The display of dashboard above on February, 2014 as shown in **Fig. 36**.
The line chart in Fig. 36 shows the total of monthly project budget starting from January until December, 2014. Similar to the line chart in Fig. 34, the position of line in this line chart will only change when the year is selected. The line in this line chart won’t change when the month is selected because month will react to a couple of pie charts (in Fig. 37) located below the line chart to change their results.

The display of dashboard above on February, 2014 for project and city as shown in Fig. 37 below.

A couple of pie charts shown in Fig. 37 consist of:

1) **Top 10 Proyek**. This pie chart shows the total of project budgets based on the name of projects which hold the most project budget on February 2014.

2) **Trend Lokasi Proyek**. This pie chart shows all of locations where the total of project budget hold the most project budget on February 2014 due to the activity of projects in certain location.

V. CONCLUSION

With the support of technology of data warehouse, it can help the company extract data of daily project budget much more time-saving and accurately because data are selected through **id** from every tables such as the extraction of contractor who is in charge of certain project.

Furthermore, the interactive and creative graphics in dashboard help the company monitor both monthly and yearly total of project budget spent for some requirements of materials to operate the project, so that the company can learn and evaluate the report much better.

This technology has been set properly, so that it has become easy to access this tool to open the dashboard anytime. Hence, the company will be able to make the best decision about the ways to control the project budget even though the required materials must be bought for the needs of project activities.

Since this journal only contains project budget realization, it’d be the best if the project budget plan is present as well in order to show the comparison between project budget plan and project budget realization.

REFERENCES


The Development of Data Warehouse for Contractor System


AUTHORS PROFILE

Wimora Sarwindo achieved his Bachelor's degree for Information System, Bina Nusantara University in 2014. He was also an accounting staff in PT. Sekawan Manajemen Konsultan in 2015-2017. In April 2017, he continued to study Master of Informatics Engineering at Bina Nusantara University. He also has been studying web course at Dumet School Grogol since March 2018. He also worked for junior web developer in June 2018 – August 2018. His two major interests in studying are web development and advanced database system.

Abba Suganda Girsang is currently a lecturer at Master in Computer Science, Bina Nusantara University. He got Ph.D. in the Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan. He graduated bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta, Indonesia, in 2000. He then continued his masters degree in the Department of Computer Science in the same university in 2006–2008. He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and also worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003–2015. His research interests include swarm, intelligence combinatorial optimization, and decision support system.