DEVELOPMENT OF INDEPENDENT DATA MART ON OPTIK YUDA

Dinar Priskawati[1], Dian Dharmayanti[2]

Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
Email : narpris@gmail.com[1], diandmdr@yahoo.com[2]

ABSTRACT

Optik Yuda is one of the stores that sell visual aid tools that located in Kompleks Ruko Sukajadi number 18 D and has been established since 1983. With official permission by Dinas Kesehatan Kota Bandung Optik Yuda can accept customers using BPJS Kesehatan (Badan Penyelenggara Jaminan Sosial Kesehatan). The main operational activities undertaken by the Optik Yuda is selling frames and lenses to customers. Customer orders are recorded in the invoice as a sales transaction data. Data form that available is still invoices and books led to the reporting process and analysis of sales data to be slow. As a result of the data is still in invoices and books are difficulties in conducting an analysis of customer interest and also the difficulty in obtaining a desired strategic information.

Independent data mart is a collection of information from various operational databases for needs of data analysis using multidimensional concept models. Based on the problems, then it need to be develop an independent data mart to manage sales data and represent strategic information from sales both detail and global to be easily analyzed. Development of data mart is using Snowflake schema and OLAP (On-Line Analytical Processing) use a Roll-Up and Drill-Down and Slice-and-Dice technique and represented through the pivot table.

Development of independent data mart is to assist and simplify the administration in the reporting making process and data analysis. Moreover simplify the process of calculating the sales reports and help in the data management and data storage.

Keyword : Data mart, Snowflake schema, Roll-Up and Drill-Down, Slice-and-Dice, OLAP

1. INTRODUCTION

Optik Yuda is one of the stores that sell visual aid tools that located in Kompleks Ruko Sukajadi number 18 D and has been established since 1983. With official permission by Dinas Kesehatan Kota Bandung Optik Yuda can accept customers using BPJS Kesehatan (Badan Penyelanggara Jaminan Sosial Kesehatan), therefore the main customers of the shop is among the lower middle class. The main operational activities undertaken by the Optik Yuda is selling frames and lenses to customers. Besides, Optik Yuda receive frames, eyeglass lenses and contact lenses orders according to customer wishes and doctor's prescription. If the customers does not have a doctor's prescription, the optics can do eye examinations on the spot. Customer orders are recorded in the invoices as sales transaction data

Based on interviews with administration, the problem encountered by Optik Yuda is data that related to sales transaction still in the form of notes and books. Because the data is still in the form of notes and books there is a problem in the reporting process and sales data analysis requires considerable time even reporting process become late. Data analysis was performed by the management for example is what brand are most popular among general and BPJS customer, how much general and BPJS customer transaction, products from which supplier are more sold, and the sales related issues. Another problem encountered is when calculating of total sales experience a miscalculation or sales transaction is not counted or missing. As a result of miscalculation total sales is the information in report become invalid. Data analysis and sales report is really important to define business strategy for next month and determine the sales target. Besides, the other problem that occur are in data management is not good that will result the risk of data loss and the difficult in searching a specific data without knowing time period.

Based on problem that found in Optik Yuda, then need a system that can manage sales data and can provide strategic information from sales both detail and global to be easily analyzed. Therefore, the solution are provided to Optik Yuda is to develop a independent data mart [2] software that

can assist employee to obtain information about sales well and quick. Moreover, it can simplify employee to make a sales report. And it can also display the report in chart in order to see sales fluctuation every month.

1.1. *Data Mart*

A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization. Given their single-subject focus, data marts usually draw data from only a few sources. The sources could be internal operational systems, a central data warehouse, or external data.

Data mart or data warehouse characteristic according to Inmon [3]:

1. Subject Oriented

Data warehouse is subject oriented means data warehouse designed to analyze data based on certain subjects within the organization, rather than on the particular application or function. The Data warehouse is organized around the main subject of the company (*customers*, *products* and *sales*) and not organized in the main application areas (*customer invoicing*, *stock control and product sales*).

2. Integrated

Data warehouse can store data that come from separate sources into a format that is consistent integrated with each other. Thus, the data can not be broken down because the data is an entity that supports the overall concept of the data warehouse itself. Terms of the integration of data sources can be met in various ways such as consistent in naming variables, consistent variable in size, consistent in coding structure, and consistent in the physical attributes of the data.

3. Time Variant

All data in the data warehouse can be said to be accurate or valid at any given time. To view the time interval used to measure accuracy of a data warehouse, we can use the way include: The simplest way is to represent the data warehouse at a certain time range, such as between 5 to 10 years into the future.

4. Non Volatile

The fourth characteristic of the data warehouse is non-volatile, mean that the data in data warehouse is not updated in real time but refresh of the operational system on a regular basis.

There are two basic types of data marts: dependent and independent. The categorization is based primarily on the data source that feeds the data mart. Dependent data marts draw data from a central data warehouse that has already been created. Independent data marts, in contrast, are standalone systems built by drawing data directly from operational or external sources of data, or both. [2].

The main difference between independent and dependent data marts is how you populate the data mart; that is, how you get data out of the sources and into the data mart. This step, called the Extraction-Transformation-and Loading (ETL) process, involves moving data from operational systems, filtering it, and loading it into the data mart. [1].

1.2. ETL (Extract, Transform, Load)

ETL Process happened when data is enter until exit the Staging Area. At this stage, the data is extracted from OLTP system (Extract) and sent to Staging Area. Staging Area is a temporary location for the placement of the data sources that have been altered (Transform), to be prepared for the next process. In the Staging Area, data form can be set of raw text files (flat files), databases tables, or both. Data integrated with other data and converted into a form that is more homogeneous. Furthermore, the data is loaded (Load) to the data mart table are specified.

1. Extract

The first part of an ETL process is to extract from the data source is called extract, as data retrieval process is not taking up the existing data in the operational database.

2. Transform

The second process is the transformation of data that has been extracted into the required format. It is necessary to remember that captured data came from different sources which may have different standardization.

3. Load

Load stage is a load data into the final target, which generally is a data warehouse or data mart. Depending on the needs of organization, this process varies widely.

1.3. Multidimensional Modeling

Multidimensional modeling is a technique for visualizing the data model as a set of the size described by the business aspect [2]. This is particularly useful for summarizing and collating data and show data to support data analyst.

There are three concepts that concept that can model more clearly in determining the fact table, measure, and dimensions, that is:

1. Star Schema

Star schema is the basic model of multidimensional modeling that can be seen in figure 1 star model has a single parent table called the fact table and a collection of small tables called dimension tables, arranged in circular patterns around the fact tables.

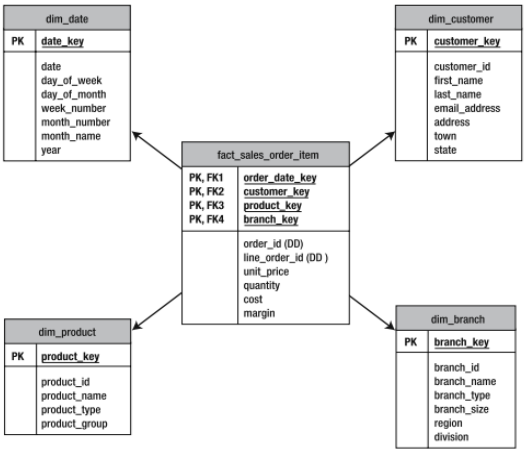


Figure 1 Star Schema [4]

2. Snowflake Schema

Besides the star schema to show the design of data mart can also use snowflake models as an example can be seen in figure 2. In a snowflake models, there is a fact table surrounded by several dimension table. But the dimension table itself can be as if in the form of a fact table that also has its own dimensions.

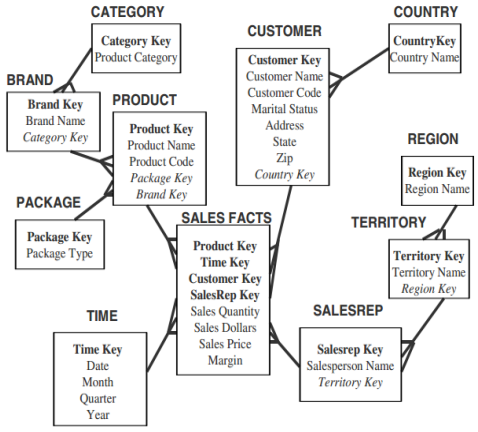


Figure 2 Snowflake Schema [5]

2. RESEARCH CONTENT

2.1. Analysis

2.1.1. Problem Analysis

Optik Yuda is one of the stores that sell the visual aid tools such as frames and lenses. Based on the results of interviews and observation in a place of research, as for some of the problems that appear are as follows:

1. Administration had difficulty in conducting an analysis of customer's because the information is insufficient information.
2. The data is still separated in the form of invoices and books make it difficult to obtain a desired strategic information.

2.1.2. Data Analysis

Based on the processes that are currently running and there are 8 operational from the results of normalization, customer table, supplier frame table, supplier lens table, frames table, lens table, transaction table, lens detail table, and districts table. The scheme of OLTP relation can be viewed in figure number 3 below:

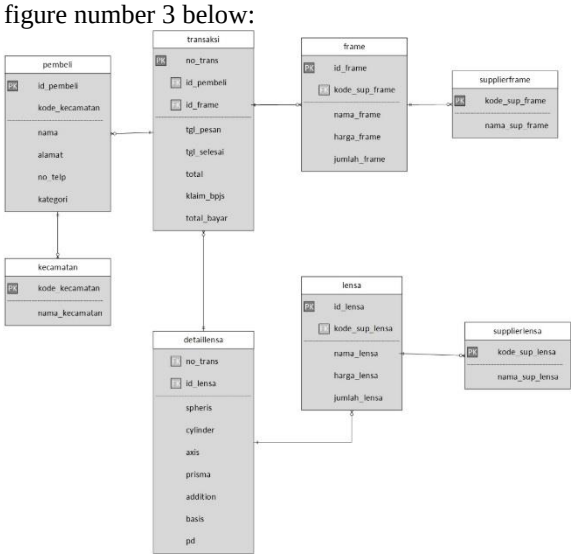


Figure 3 OLTP Schema Relation

2.1.3. Information Needs Analysis

In this stage of the analysis of the information needs will be conducted the establishment facts table candidate and dimensions table candidate. There are four stages that are conducted, selecting business process, determining the grain and strategic information, identifying dimension, and identifying facts. For more details are as follows:

1. Selecting Business Process

Based on the results of the analysis and interviews then the selected business processes are as follow:

a. Sales of frames and lenses toward customer.
b. General and BPJS customer reporting.

2. Determining Grain and Strategic Information

Grain is the stage to determining what records will be used in a facts table. In this case the determination of the grain will be adjusted with business processes. Grain and strategic information that have been selected can be viewed in table 1 below:

Table 1 Formed Information Strategic

No	Grain	Strategic Information
1	Sales transactions total of frames and lenses.	1. Information of the most sold frame brand information in each month and year 2. Information of the most sold lens brand information in each month and year 3. Information of sales transaction amounts in general and by category of customers in each month and year

No	Grain	Strategic Information
2	Sales transactions total toward general customer.	1. Information of the most sold frame brands to the general customer in each month and year 2. Information of the most sold lenses brands to the general customer in each month and year.
3	Sales transactions total toward BPJS customer.	1. Information of the most sold frame brands to BPJS customer in each month and year.
		2. Information of the most sold lenses brands to BPJS customer in each month and year.
		3. Information of BPJS total claims in each month and year.
4	Customer and district of customer.	1. Information of districts of customer that make the most transaction in each month and year.
5	Sales of items total from frames supplier and lenses.	1. Information of sales of items total from each frame supplier and lenses in each month and year.

3. Identifying Dimensions

In this stage, there will be identified and adjusted the dimension of the grain that has been selected. The dimensions are is the district, customer, time, frames, lenses, frame supplier, and lens supplier.

4. Identifying Facts

In this stage, there will be the selection of facts that will be used. This fact tables adjusted to the grain and strategic information that has been formed. In this case, the fact that is used is a table of sales transactions. Because of there are two customer categories in the sales process, that is the general and BPJS, then there will be two tables of facts that represent a common sales transactions and the claims fact BPJS table.

2.1.4. Data Mart Architecture

Architecture that was used in the data mart development will be refer to two-layer architecture on a data warehouse [1]. The Analysis of this architecture is divided into four layer, that is source layer, data staging, data mart layer, and analysis. Here is the architecture of the data mart than can be viewed in Figure 4:

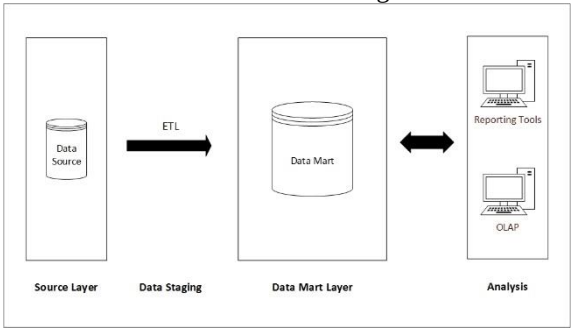


Figure 4 Data Mart Architecture

2.1.5. Source Layer Analysis and Data Staging

In this layer, the data is still in the form of operational data. Data source that was used in the data-marts development form the data logic in operational databases. In this stage, there will be happen ETL process, that is the process of Extract, Transform, and Load to the data source.

1. Extraction Process

At the time of data extraction process will form a new field to show timestamp, in this case, formed timestamp is a timestamp to the customer. This timestamp will consist the time and the date of the default of MySQL Server. Table 2 shows the table and extracted field. .

Table 2 Extract Table

No	Table Name	Field
1	pembeli	id_pembeli
		nama
		alamat
		kecamatan
		no_telp
2	supplierframe	kode_sup_frame
		nama_sup_frame
3	supplierlensa	kode_sup_lensa
		nama_sup_lensa
4	frame	id_frame
		kode_sup_frame
		nama_frame
		merek_frame
		harga_frame
5	lensa	id_lensa
		kode_sup_lensa
		nama_lensa
		harga_lensa
6	transaksi	no_trans
		id_pembeli
		id_frame
		tgl_pesan
		tgl_selesai
		total
		klaim_bpjs
		total_bayar
		no_trans

No	Table Name	Field
7	detaillensa	no_trans
		id_lensa
		spheris
		cylinder
		axis
		prisma
		addition
		basis
		pd
8	kecamatan	kode_kecamatan
		nama_kecamatan

2. Transformation Process

In the process of transformation being conducted is the selection process and conditioning process.

a. Selection

In this selection process was conducted selection field from the tables that have been extracted. In this, election field means removes the field that will not be used in the analysis process. Here is a field name that have been removed in the selection process:

1) In the customer table (tabel pembeli) does not require field names, addresses, and no_telp.
2) In the frame table (tabel frame) does not require jumlah_frame.
3) In the lenses table (tabel lensa) does not require field jumlah_lensa.
4) In the transaction table does not require field tgl_selesai and total_bayar..
5) In the lens detail table (tabel detaillensa) does not require field spheris, cylinder, axis, prisma, addition, basis, dan pd.

The example of selection process can be viewed in Table 3 below:

Table 3 Detail Lensa Table Selection

detail_lensa	
No	**Field**
1	no_trans
2	id_lensa
3	spheris
4	cylinder
5	axis
6	prisma
7	addition
8	basis
9	pd

→

detail_lensa	
No	**Field**
1	no_trans
2	id_lensa

b. Merging

The process of merging is done by combining the two tables that have relations. Two tables are to be combined is the transaksi table and detail_lensa. Here is a merging process that can be viewed in Table 4.

Table 4 Table Merging

detail_lensa	
No	**Field**
1	no_trans
2	id_lensa

transaksi	
No	**Field**
1	no_trans
2	id_pembeli
3	id_frame
4	tgl_pesan
5	total
6	klaim_bpjs

→

transaksi	
No	**Field**
1	no_trans
2	id_pembeli
3	id_frame
4	id_lensa
5	tgl_pesan
6	total
7	klaim_bpjs

c. Conditioning

The process of conditioning is done by selecting table and attribute from the data source to the target (data mart). Conditioning process to the process of transformation is to change a date field into several field, that is date, month, and year. The main reason of this process is to facilitate the process of data analysis based on the distance of the desired time.

Here is conditioning process that can be viewed in Table 5

Table 5 Conditioning Table

tgl_pesan	dim_waktu		
	tanggal	**bulan**	**tahun**
1/03/2014	1	03	2014
20/03/2014	20	03	2014
5/09/2014	5	09	2014
15/09/2014	15	09	2014
7/11/2014	7	11	2014
30/11/2014	30	11	2014

3. Proses Loading

In this process, the data that has been extracted, cleaned, and changed its format, will be stored in the data mart. The process of loading the data will be performed automatically after the transform process is complete. The technique used is the update. This process will update immediately update the data mart without changing existing data.

2.1.6. Data Mart Layer Analysis

In this layer, the data that have been through the ETL process will be stored in a centralize storage logic that the data mart. Based on the information needs analysis that has been carried out it will be created a schema for the data mart. It will be needed two tables fact it is transaction fact table. Beside there will be a dimension table that will be used together in multiple fact tables. Based on the needs, the schema data mart to use is snowflake. For more

details data mart schema relation can be viewed in Figure 5.

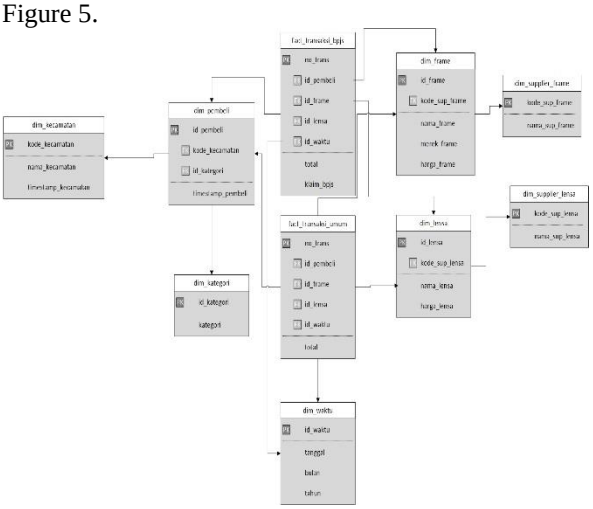


Figure 5 Data Mart Schema Relation

2.1.7. OLAP

For the OLAP process, analytical techniques to be used is roll-up and drill-down. The technique has been selected because later it will be assist in presenting data based on each dimension and derivatives of each dimensions. The selected technique can present data in general and detail of each dimension by a certain time. Besides that, the other technique is slice and dice will be helpful in a view of data from multiple perspective. Slice and dice and roll-up and drill-down can be seen in the pivot table. An example is the OLAP process for information of the most sold frame brands to BPJS customer in each month in each month and year. In this strategic information involves four dimension, frame dimension, customer dimension, time dimension, and category dimension. For more details can be viewed in Figure 6 belows :

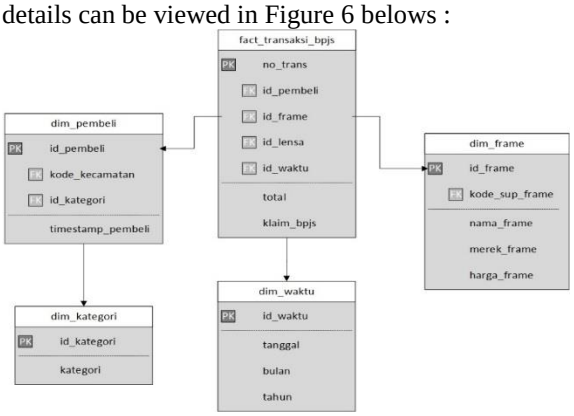


Figure 6 Information Strategic Relation

Example of OLAP cube that formed of strategic information can be viewed in Figure 7 below:

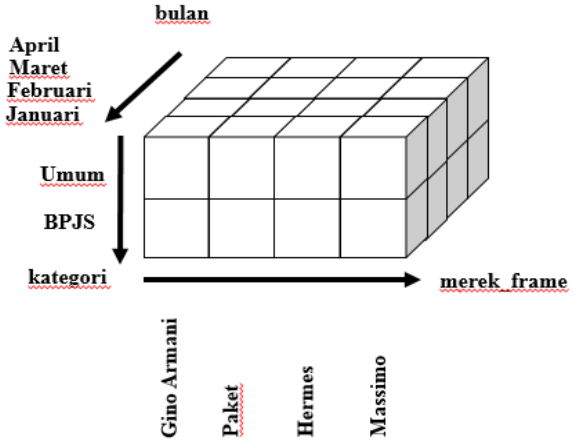


Figure 7 OLAP Cube

Techniques used to strategic information is slice and dice. An example is doing slicing and dicing of the frame gino Armani brand in March and April customer BPJS. Illustration slice and dice can be viewed in Figure 8 below:

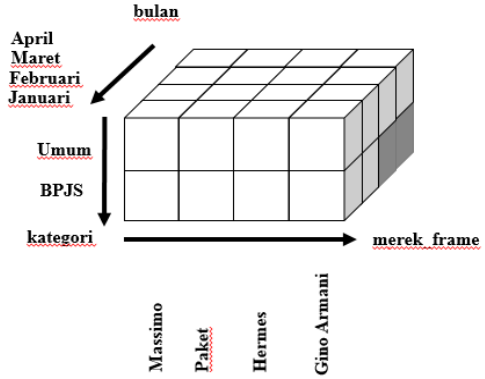


Figure 8 Slice and Dice Strategic Information

Examples result record of slicing and dicing the data from strategic information can be viewed in table 6 below:

Table 6 Slice and Dice Information

Frame Brands	Frame Sold	Quantity	Month
Gino Armani	3 pcs	Rp 1.050.000	March
Gino Armani	5 pcs	Rp 1.650.000	April

Roll-up and drill down technique to be used toward the dimension of the time is month.

2.1.8. Data Visualization

Data visualization techniques that will be used is in the form charts such as bar chart, pie chart, and line chart. Besides using a chart, other techniques to be used in data visualization is the dashboard. A dashboard is a summary of the existing data in Optik Yuda that have more simple data visualization, easily understood, and the main thing is to show information instantly.

In this case the visualization of data for each strategic information is going to use bar chart and

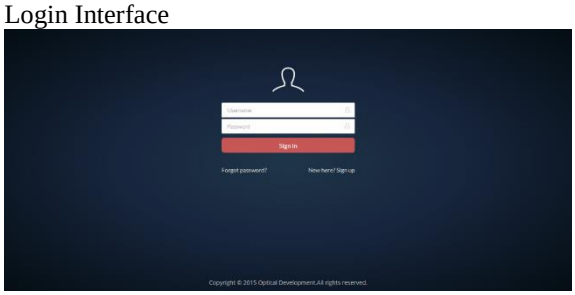
pie chart. Bar chart are used to compare the quantity of an item and see fluctuation in total transaction in each month. Pie chart are used to determine which item more often sold as compared to other. Table 7 show the strategic information and visualization to be used:

Tabel 1 Pemetaan Visualisasi Data

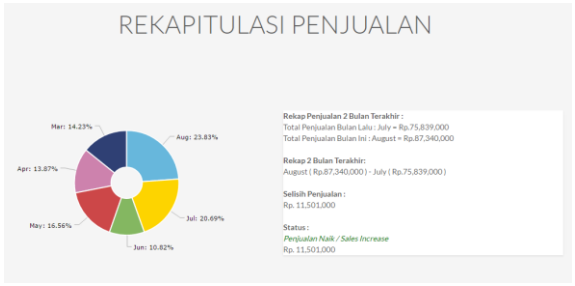
No	Informasi	Visualisasi
1	Information of the most sold frames brand information in each month and year.	Bar chart dan Pie chart
2	Information of the most sold lenses brand information in each month and year.	Bar chart dan Pie chart
3	Information of the most sold frames brands to the general customer in each month and year.	Bar chart dan Pie chart
4	Information of the most sold frames brands to the BPJS customer in each month and year.	Bar chart dan Pie chart
5	Information of the most sold lenses brands to the general customer in each month and year.	Bar chart dan Pie chart
6	Information of the most sold lenses brands to the BPJS customer in each month and year.	Bar chart dan Pie chart
7	Information of sales transaction amounts in general and by category of customers in each month and year.	Bar chart dan Pie chart
8	Information of sales of items total from each frames supplier and lenses in each month and year.	Bar chart dan Pie chart
9	Information of district of customer that make the most transaction in each month and year.	Bar chart dan Pie chart
10	Information of BPJS total claims in each month and year	Bar chart

2.2. Result

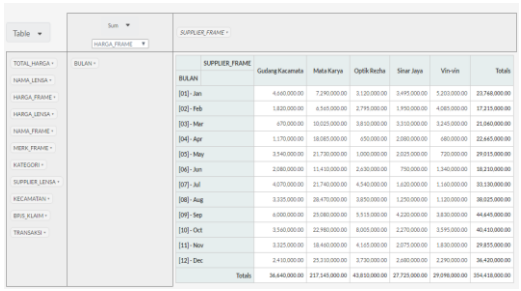
1. Login Interface



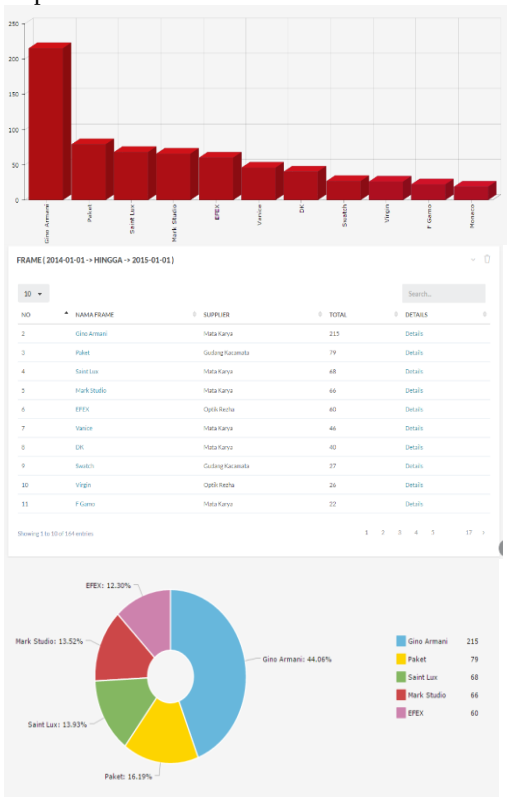
2. Dashboard Interface



3. OLAP Interface



4. Report Interface



3. CONCLUDING

3.1. Conclusion

Based on the result of the implementation and testing that have been done, it can be deduced from the thesis entitled Development of Independent Data Mart on Optik Yuda is as follows:

1. To assist administration and owner in the reporting process and analyzes quickly and also more interesting with the visualization of information.
2. To assist administration in the process of calculating the sales report.
3. To assist administration in the data management and data storage.

3.2. Suggestion

In the development of independent data mart is obviously still have a flaw and still have deficiency. Therefore, it is necessary to do more development and further improvement in order to meet user needs. As for suggestions towards development of data mart for the future are as follows:

1. Developed into dependent data mart where data from several units can be synchronized and integrated.
2. Developed into data warehouse so data in all units on Optik Yuda can be used properly.

REFERENCES

[1] M. Golfarelli and S. Rizzi, Data Warehouse Design : Modern Pricnciples and Methodologies, New York: McGraw-Hill, 2009.

[2] C. Ballard, D. Herreman, D. Schau, R. Bell, E. Kim and A. Valencic, Data Modeling Techniques for Data Warehousing, California: International Business Machines Corporation, 2000.

[3] W. Inmon, Building the Data Warehouse, New York: John Wiley & Sons, 2002.

[4] V. Rainardi, Building a Data Warehouse With Example in SQL Server, New York: Apress, 2008.

[5] P. Ponniah, Data Warehousing Fundamentals, New York: John Wiley & Sons, 2001.