DEVELOPMENT OF DATA WAREHOUSE ON PT.PUPUK ISKANDAR MUDA

Dinda Wilanda

Program Studi Teknik Informatika
Fakultas Teknik dan Ilmu Komputer Universitas Komputer Indonesia
Jl. Dipati Ukur. 112-116 Bandung
E-mail : Dinwil07@gmail.com

ABSTRAK

PT.Pupuk Iskandar Muda or commonly called as PT PIM is a subsidiary company of PT.Pupuk Indonesia (Persero) STATE-OWNED ENTERPRISES that working on chemical industry in particular produce urea fertilizer. The exixtence of PT.Pupuk Iskandar Muda can supply the needs of fertilizer to farmers and the vast plantations in some parts of the province of Aceh, North Sumatra, West Sumatra, Riau, Riau Islands, and West Kalimantan. Current amount of existing data on the PT. Pupuk Iskandar Muda is enormous , it makes delay in accessing the information, and every region has a different data type and naming. The form of the report is not yet support for the views of the various aspects of the company's decision-making, so it needed a system that can display information in a multi dimensiona. Therefore, the company needs data warehouse for managing production and sales data for accessing information to making a decision.

The company's data used for data warehouse including sales data, production data, stockroom, customer, product, factory, ingredients. Data warehouse makes the spread data on company to the integrated data, so that can helping company for analyze existing data for making decision that strategic to more quickly and accurately. The development of datawarehouse used modeling of Fact constellations and OLAP (On-Line Analytical Processing) design, for the processing and analysis of the data by using the technique of Roll-ups and Drill-Down.

The sofwatre of data warehouse can helping the company on integrate data. The result of information can be used for helping the analysis for company of PT.Pupuk Iskandar Muda.

Key Word : *Data warehouse,* Fact Constellations, *kRoll-Up, Drill-Down, OLAP*

1. INTRODUCTION

Data warehouse is a database that react with each other which can be used for query and analysis, is the orientation of the subject, integrated, time-variant, and has not changed, which is used to support decision making. Data warehouse is a system of extracting, cleaning, adjusting, and sends the data source into the data storage and then support the implementation dimensional query and analysis for decision-making purposes.

PT. Pupuk Iskandar Muda or commonly called the PT PIM is a subsidiary of PT. Pupuk Indonesia (Persero) under the state-owned enterprises engaged in the chemical industry in particular producing urea fertilizer. Hence the presence of PT. Pupuk Iskandar Muda fertilizer can meet the needs of the vast estate in North Sumatra and several other regions (provinces of Aceh, North Sumatra, West Sumatra, Riau, Kepulauan Riau, and West Kalimantan).

Based on interviews with the PT. Pupuk Iskandar Muda, the current storage of large amounts of data are still apart - separated. Companies that have large data sets to be analyzed for the continuity of the company's business processes, in conducting an analysis on the information required data display that can display a lot of data. Preparing reports on PT.Pupuk Iskndar Young conducted by collecting data from every part of the company is needed, problems in making this report the need for a long time. Form of the report does not yet support for the views of the various aspects of corporate decision-making, so we need a system that can display information in a multi-dimensional and dynamic.

Therefore, the company requires a data warehouse to manage production data and sales data to obtain information in making decisions.

1.1. Purpose and objectives

Based on the problems studied, the purpose of this research is to build a data warehouse software in the PT. Pupuk Iskandar Muda as a solution to the slow and difficult analysis of the previous system. While objectives to be achieved in the study are:

While the objectives to be achieved in the study are:

1. Assist the company in integrating data.

2. Assisting the company in analyzing the sales and production of goods in a given period is multidimensional.

2. LITERATURE

Data Warehouse can vary but have the same core, like the opinion of some experts the following:

Data Warehouse can vary but have the same core, like the opinion of some experts the following: The data warehouse is a collection of data that have a nature-oriented subject, integrated, time-variant, and is fixed on the collection of data in support of the decision making process management **[2]**.

The data warehouse is a relational database that is designed more to query and analysis of the transaction process, usually containing the data history of the transaction process and could also data from other sources. Data warehouses separate analysis workload from transaction workload and enables an organization to merge / consolidation of data from various sources **[2]**.

The data warehouse is a method in the design of the database, which support the DSS (Decission Support System) and EIS (Executive Information System). Physically data warehouse is a database, but the data warehouse and database design is very different. In traditional database design using normalization, while the normalization of the data warehouse is not the best way **[2]**.

From the definitions described above, it can be concluded that the data warehouse is a database that react with each other can be used to query and analisisis, is the orientation of the subject, integrated, time-variant, unchanged used to assist decision makers.

2.1 Basic Concepts Data Warehouse

The data warehouse is a collection of all sorts of data that is subject oriented, integrated, time variant, and nonvolatile in support of the decision-making process **[4]**.

Data warehouses are often integrated with various application systems to support the process of reporting and data analysis by providing historical data, which provides the infrastructure for the EIS and DSS.

a. *Subject Oriented*

The data warehouse is organized in major subjects, such as customers, items, and sales. Focusing on the model and analysis on the data to make decisions, so it's not on any transaction or process is not in the OLTP. Avoid useless data in taking a decision.

b. *Integrated*

Built by connecting or uniting different data. relational databases, flat files, and on-line transaction record. Ensuring consistency in the naming, coding structure, and structure attributes of data between each other.

c. *Datawarehouse time variant*

Data is stored to provide information from a historical perspective, the data that year - last year or 4-5 years. Time is a key element of a data warehouse at the time pengcapture's.

d. *Non Volatile*

Whenever the process of change, the data will be collected in each time. So it is not updated continuously. Data warehouse does not require transaction processing and recovery. There are only two operations initial loading of data and access of data.

2.2 ETL Process (*Extraction, Transformation, Loading*)

The three main functions that need to be done to make the data ready for use in the data warehouse is the extraction, transformation and loading. These three functions are in the staging area **[5]**.

In this staging of data, provided the place and area with multiple functions such as data cleansing, change, convert, and prepare the data to be stored and will be used in a data warehouse **[5]**.

a. *Extraction*

Data Extraction is the process of taking the necessary data from the source data warehouse and are then put on the staging area to be processed at a later stage. In this function are associated with different types of data sources such as data formats, different machines, software and architecture are not the same. So before the process is done, you should have to be defined requirement against data sources that will be used for the next process.

b. *Transformation*

In fact, the process of transactional data is stored in various formats so rare to find a consistent data between existing applications. Data transformation aimed at addressing this problem. With this data transformation process, we standardized the data on a consistent format. Some examples of such data inconsistencies can be caused by different types of data, the data length and so forth.

c. *Load*

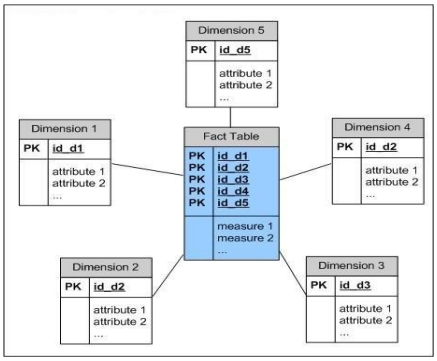
Data load is moving the data into the data warehouse. There are two loading data at the data warehouse. The first is the initial load, this process is done when it has completed design and build a data warehouse. The input data will be very large and takes a relatively longer. Second Incremental load, carried out when the data warehouse is operated. Incremental load can be carried out in accordance with a system built

2.3 Schema of *Data Warehouse*

The scheme is often used in data warehouse is a star or snowflake schema, the second scheme is very easy to understand and in accordance with business needs, supporting simple queries and provide superior query performance by minimizing the tables join **[7]**.

1. *Star Schema*

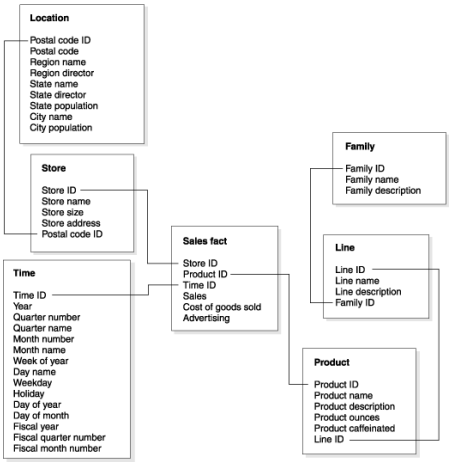
The star schema is a logical structure that has a fact table that consists of factual data in the middle, surrounded by tables dimension that contains reference data.



Picture 1 Skema Bintang

2. *Snowflake Schema*

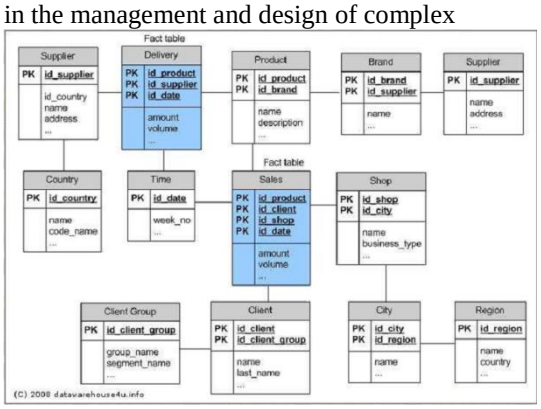
Is a variant of the star schema where table-table dimensions there are no data on the de-normalize. In other words, one or more dimension tables are not joined directly to the fact table but the table other dimensions.



Picture 2 *Snowflake Schema*

3. *Fact Constellation Schema*

Fact constellation schema is a dimensional models in which there are more than one fact table that divides one or more dimension tables. This scheme is more complex than the star schema that contain multiple tables facts. In fact constellation schema, one dimension tables can be used in several tables to the fact that the design is more complex. The advantage of the fact constellation schema is the ability to model more accurate business using multiple fact tables. But the disadvantage is difficult in the management and design of complex



Picture 3 *Fact Constellation Schema*

3. ANALYSIS AND DESIGN

3.1 Problem Analysis

The analysis conducted on the company PT.Pupuk Iskandar Muda got some problems that exist in the company, as follows:

1. The company currently difficult to obtain sufficient information to make a decision pengambilang.
2. Currently the company has yet to form a report on support for the views of the various aspects.

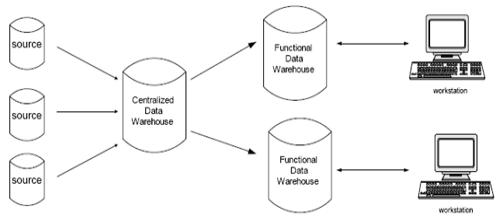
3.2 Information Needs Analysis

Analysis of information needs is the stage of analyzing what information is required by the Cooperative Award fortune of data warehouse that will be built. Based on interviews with Budi Adi Mulyo as cooperative owners have several needs information that would be required by the Cooperative Award fortune to achieve competitive advantage and improve business operations of cooperatives are as follows:

1. Information on the amount of fertilizer production plant per month and per year.
2. Information on the number of customers who most bought fertilizer per month and per year.
3. The information most amount of fertilizer sold in each region per month and per year.
4. The information most amount of fertilizer sold in each province per month and per year.
5. Information on the number of materials used in the factory every month and the year.
6. Information sales amount of manure produced by most every province in each month and year.

3.3 Development of Data Warehouse Architecture

The type of data warehouse is to be built is a functional kind of a data warehouse, where the source data to be stored in the data warehouse is external data, ie daily data of each activity in the form microsoft office excel file with format ".xlsx". Data warehouse type functional layer consists of a layer source, Data Staging, data warehouse layer and analysis. The following functional image data warehouse architecture.



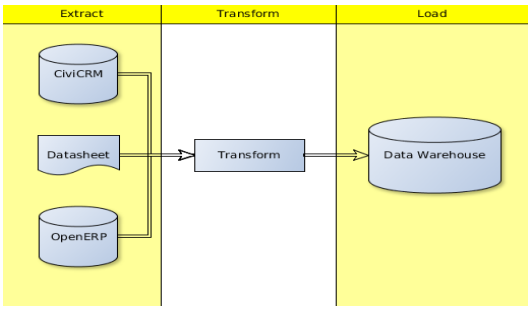
Picture 4 Data Warehouse Fungsional

3.4 Source Layer

Source layer is a layer of a data source, wherein the core layer of the data is still in the form of an external file. External data that will be used in the construction of a data warehouse is the data in the form of an excel file with xls format. Excel file will be imported into the database, Before importing excel file into the database, first column and the data content of each field or record that is analyzed in order to structure the table that will be built into the data warehouse in accordance with the file to be imported into the database.

3.5 Data Staging

At the core layer, the external data is already imported into the database will be extracted, transformed and then loaded into the data warehouse. This process is better known as the ETL process. ETL process is a process that is very important in building a data warehouse, the higher the level of truth ETL process more accurate information extracted from the data warehouse.



Picture 5 Langkah-langkah ETL

ETL process describes the steps that will be done in the process of staging. As explained below:

1. *Extraction* Process

The first step in the ETL process is to extract data from data sources. Data warehouses can combine data from different sources with separate systems that use different data formats. Extraction is to transform data into a format that is useful to the process of transformation. The process of extracting data from the data source into the data warehouse are as follows:

1) Process extract on the sales table.

Table 1 Sales *Extract*

Name of Table	Field
Penjualan	no_do
	kode_gudang
	kode_pelanggan
	kode_barang
	jumlah
	Total
	id_tanggal
	tanggal

2) Process extract on the production table

Table 2 *Extract* Produksi

Name of Table	Field
Produksi	no_produksi
	kode_produksi
	kode_barang
	kode_pabrik
	kode_bahan
	jumlah
	id_tanggal
	tanggal

3) Extract process in the storehouse table

Table 3 *Extract* Gudang

Name of Table	Field
Gudang	kode_gudang
	nama_gudang
	alamat_gudang

4) Extract process in the customer table

Table 4 *Extract* Type Customer

Name of Table	Field
Pelanggan	kode_pelanggan
	nama_pelanggan
	alamat_pelanggan

5) Process extract in items table

Table 5 *Extract* Items

Name of Table	Field
Barang	kode_barang
	nama_barang
	harga_barang
	Keterangan

6) Process extract in items facttory

Table 6 *Extract* Pabrik

Name of Table	Field
Pabrik	kode_pabrik
	nama_pabrik
	alamat_pabrik

7) Process extract in items ingredients

Table 7 *Extract* Bahan

Name of Table	Field
Bahan	kode_bahan
	nama_bahan
	keterangan

2. *Transformation* Process

The process of transformation is done cleaning and conditioning.

a) *Cleaning*

Cleaning process to clean the data or field of unnecessary and unused on the table that has been extracted. Here is a field name is omitted in the process of cleaning:

1. In the table does not require any storage field alamat_gudang.
2. Items in the table does not require a description field.
3. In the table of the plant does not require field alamat_pabrik province.
4. In the table caption field removes material.

Table 8 *Cleaning* Table Peminjam

Gudang		dim_gudang	
No	**Field**	**No**	**Field**
1	kode_gudang	1	kode_gudang
2	nama_gudang	2	nama_gudang
3	alamat_gudang		

In this table eliminates some fields "alamat_gudang" because it is not needed in the analysis of strategic information needs.

b) *Conditioning*

Conditioning process is done by selecting the attribute from the data source to the target (data warehouse). Explanation of conditioning in this transformation process that is changing the date field split into several fields (day, month, year) because when the process of analysis, the required data can be analyzed in a range based on the desired time. For more details, see the table below below:

Table 10 Table Conditioning

tanggal_penjualan
03/01/2013
04/01/2013
04/01/2013
07/01/2013
07/01/2013

dim_tanggal_penjualan

dim_tanggal_penjualan		
tanggal	**bulan**	**tahun**
03	01	2013
04	01	2013
04	01	2013
07	01	2013
07	01	2013

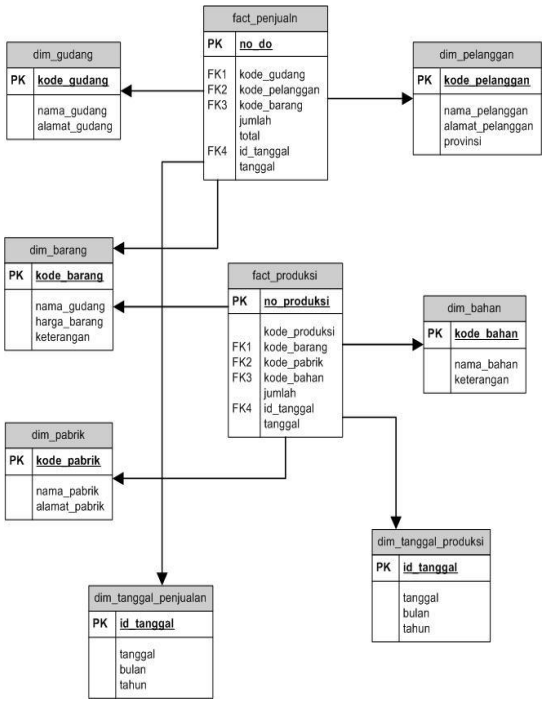
3. *Load* Process

In this process, the data is already in extract and transformation will be stored in the data warehouse. The technique used is the update. Existing data will not be deleted or changed because the data will be updated periodically. Later all the

data that passes through the extraction and transformation will be directly inserted into the data warehouse without changing existing data.

3.6 Data Warehouse Layer

In this layer, the data that have been through the ETL process will be stored in a centralized storage logic is the data warehouse. Will be needed two tables fact that the sales fact table and the fact table production. In addition there will be a dimension table that will be used together in multiple fact tables. Viewed from the requirement, then the schema data warehouse that will be used are the constellations because there is the dimension tables that are related to more than one other fact. For more details, relation schema data warehouse can be seen in Figure 6:



Picture 6 Skema Data Warehose

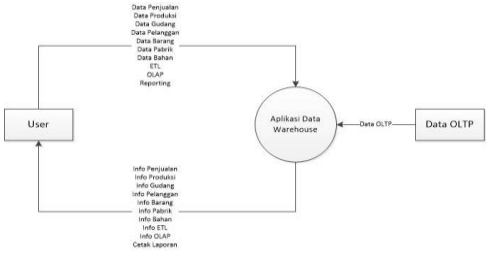
3.7 Functional Needs Analysis

Analysis of functional requirements describe the activities that will be implemented in a system. The analysis will be made to describe the functional model and information flows that context diagram and data flow diagram (DFD).

1. Diagram Context

Manufacture of the system context diagram illustrates the structural approach system in general which will then be broken down into more detailed parts.

The context diagram of the system to be built can be seen in the following figure:



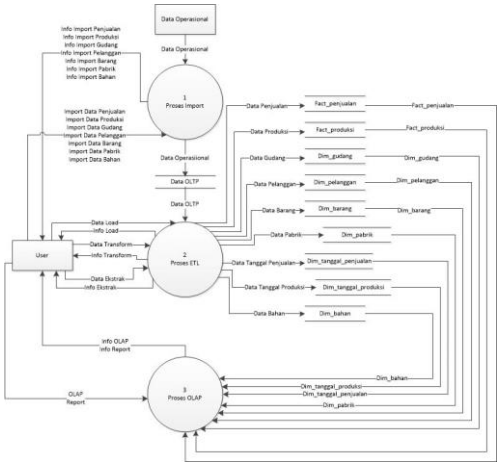
Picture 7 Diagram Konteks

2. Data Flow Diagrams

Data Flow Diagrams (DFD) is a process which describes a flow of information that is more detailed and granular which is the development of a context diagram. The following is a Data Flow Diagram of the system to be built:

a) DFD Level 1 *Data Warehouse Software*

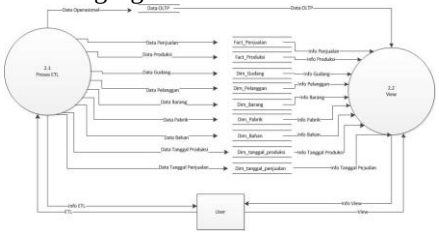
DFD Level 1 data warehouse application is a decrease from the context diagram to describe the processes that will occur in the system. Figure DFD Level 1 can be seen in the image below:



Picture 8 DFD Level 1 Aplikasi Data Warehouse

b) DFD Level 2 ETL Process

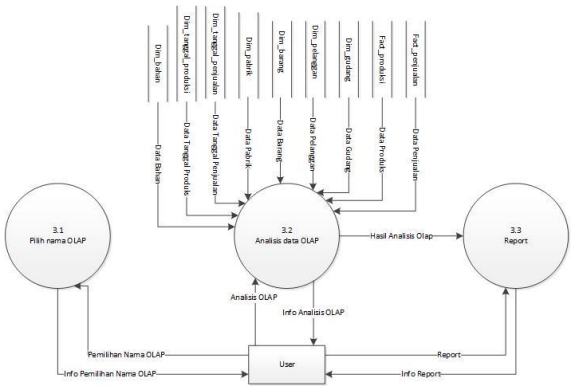
The DFD Level 2 process ETL data from the system to be built can be seen in the following figure:



Picture 9 DFD Level 2 ETL Process

c) DFD Level 2 OLAP Process

The DFD Level 2 process OLAP data from the system to be built can be seen in the following figure:



Picture 10 DFD Level 2 Proses OLAP

4. IMPLEMATION AND TESTING

4.1 Software Builders

The software used to build a data warehouse software PT.Pupuk Iskandar Muda, namely:

Table 11 Software Builder

No.	Perangkat Lunak	Keterangan
1	Windows 8	Sistem Operasi
2	SQL Server 2014	Penyimpanan *Database*
3	Visual Studio 2013	*Tool* Pembangun Aplikasi

4.2 Hardware Builders

Minimum requirement of hardware (hardware) required to build a data warehouse software is a computer hardware laptop with specs like this:

Table 12 Hardware Builders

NO	Perangkat Keras	Spesifikasi Minimum	Eksistensi	Keterangan
1	Processor	Intel Quad Core	Intel Core i3	Mendukung
2	RAM	2 GB	2 GB	Mendukung
3	Hardisk	64 GB	128 GB	Mendukung

4.3 Implementation Interface

I implementation of the interface is done by displaying each display and encoding system built in the form of program files. As for the implementation of data warehouse software to see built can be seen in the table below:

Table 13 Implementation Interface

No.Proses	Nama Proses	Nama File
1	ETL	Etl.cs
2	OLAP	Olap.cs
3.1	Information on the amount of fertilizer production plant per month and per year.	Olap1.cs
3.2	Information on the number of customers who most bought fertilizer per month and per year.	Olap2.cs
3.3	The information most amount of fertilizer sold in each region per month and per year.	Olap3.cs
3.4	The information most amount of fertilizer sold in each province per month and per year.	Olap4.cs
3.5	Information on the number of materials used in the factory every month and the year.	Olap5.cs
3.6	Information generated amount of fertilizer sales at most each month and year.	Olap6.cs

4.4 System Examination

Testing data warehouse built using black box testing methods. The test plan can be seen in the table below:

Table 14 Rencana Pengujian

Kelas Uji	Detail Pengujian	Jenis Uji
Import	Berhasil import atau gagal import	*Black Box*
ETL	Berhasil Ekstrak,Transform, Load atau gagal *Ekstrak,Transform, Load*	*Black Box*
OLAP	Berhasil melakukan analisis OLAP atau gagal melakukan analisis OLAP	*Black Box*

4.5 Alpha Testing Conclusion

Based on the results of Alpha testing has been done can be concluded that the application has been running quite well, in this test data integration can be said to be a success because of the stage of the import process, ETL and OLAP goes well, and output needs of the information produced are in accordance with the desired output so that the resulting report would also be appropriate for following output results OLAP analysis.

4.6 Conclusion Beta Testing

Based on the interview above concluded that the purpose of the development of data warehouse software has been fulfilled

5. END

5.1 Conclusion

From the results of tests and interviews that have been made to the company as the company PT.Pupuk Iskandar Muda can be concluded that:

1. The software's data warehouse can help the company in the integration of data.
2. The information generated can be a tool for the analysis of the company PT.Pupuk Iskandar Muda.

5.2 Suggestion

In the Data Warehouse Software development is still far from perfect and there are still many shortcomings. It is therefore necessary for the development and further refinement. As for suggestions that could be used in the further development of the system:

1. Developed to BI because the company has another subsidiary.

BIBLIOGRAPHY

[1] Roger S. Pressman, *Software engineering*, 7th ed .: Mc Graw. Hill. 2010

[2] Golfarelli, M., Hill, M.G., dan Rizzi, S.2009. *Data Warehouse Design, Modern Principles and Methodologies.New York: Mc Graw Hill*.

[3] Vincent Rainardi, *Building a Dta Warehouse;* Apress, 2008

[4] Ponniah Paulraj,2001.*Data, Warehouseing Fundamentals*, New York : A Wiley Interscience Publication.

[5] Rainardi, V. *Building A Data Warehouse, With Examples in SQL Server*. APRESS,2007.

[6] Inmon, William H. 2005. "*Building the Data warehouse* (4th ed.) Indianapolis :Wiley Publishing, Inc.

[7] Yudhi Hermawan, konsep OLAP. Yogyakarta; ANDI Yogyakarta, 2005