SOFTWARE DEVELOPMENT DATA MART IN PT. MATAHARI SENTOSA

Ihsan Mukhlish Shidiq[1]

[1] Department Of Informatics Enginering
Faculty of Enginering and Computer Science University Computer of Indonesia
Jl. Dipatiukur No 112-116 Bandung - Indonesia
E-mail : ihsanmukhlishshidiq@gmail.com

ABSTRACT

PT. Matahari Sentosa is a zipper company. The company is believed to be a supplier to other manufacturers. Nowadays, in making the multidimensional final report production becomes less effective due to there is not any systems which support the creation of multidimensional reports. In presenting information it requires a lot of time will to present it.

Based on the existing problems in the production division, it is necessary to build a data mart software that can ease in obtaining strategic information quickly so that it can be used for long-term planning, present a multidimensional and concise information, speed up the process of analyzing in order to maximize a decision made by production manager for handling stock of products and the handling of raw materials for further production.

By developing this data mart software, it may conclude that it eases in obtaining strategic information quickly which is required by the production manager of PT. Matahari Sentosa for long-term planning, present a multidimensional and concise information, and to accelerate the process of analyzing in order to maximize a decision made by the production manager for handling stock the products that have been produced and handling of raw materials for further production, and assist in making a multidimensional final report production which is required by the production manager at PT. Matahari Sentosa.

Keywords : Data Mart, Data Warehouse, Constellation Scheme

1. INTRODUCTION

PT. Matahari Sentosa is a zipper company. The company is believed to be a supplier to other manufacturers, such as the manufacturer of bags, jackets, pants, and the manufacture of goods that require a zipper.

Nowadays, in making the multidimensional final report production or have a time period become less effective and efficient becomes less effective due to there is not any systems which support the creation of multidimensional reports. In presenting information it requires a lot of time will to present it due to data retrieval and processing is slow and the lack of a system that can provide quick information and detail. For that necessary the construction of data mart to ease the production manager to get strategic information quicly so that it can be used for long-term planning and also can present a multidimensional and concise information, speed up the process of analyzing in order to maximize a decision made by production manager for handling stock of products and the handling of raw materials for further production.

The purpose of building data mart software in PT. Matahari Sentosa is to:

1. Ease the production manager to get strategic information quicly so that it can be used for long-term planning and can present a multidimensional and concise information, speed up the process of analyzing in order to maximize a decision made by production manager for handling stock of products that have been producted and the handling of raw materials for further production.
2. Assist the production manager in making a multidimensional final report production.

The problem of restrictions applied in the construction of this data mart software among other things, that is:

1. The data came from the Division of production, that is the data production, dyeing, stock production, and stock of products.
2. Data for testing the data mart that is years 2010 to 2014.
3. The process of making data mart with ETL (Extract Transform Loading) processes.
4. Testing the data mart using OLAP (On-Line Analytical Processing).
5. The DBMS using SQL Server 2012.
6. Software development using Visual Studio 2012.
7. Analysis and design software using object oriented analysis.

1.1 Data Mart

Data mart is a part of data warehouse which is supporting report making and analize data from an unit. Data mart's contain relevan information for user want to take bussiness decision.

There are four tasks that can be done with the data mart [1], four tasks are as follows:

1. Reporting
   Reporting is one of task from data mart wich is the most general thing to do. A simple query can handle daily, monthly, yearly reporting, and another periodly reporting.
2. On-Line Analytical Processing (OLAP)
   Data mart can task all needed information both detail or summary data easily and analize process will be begin. OLAP iis multidimension data concept and it can analyzing data till detail without type any SQL command.another facility of OLAP is roll-up and drill down. Drill-down is ability to see the detail of information and roll-up is reverse that ability.
3. Data Mining
   Data mining is process for dig more deep about knowledge and new information from a lot of data inside data mart.
4. Executive Information Process
   Data mart can make important information summary to take bussiness decision without all of data. So data mart can support user to take bussiness decision easily.

1.2 Data Mart Dimensional Model

Dimensional model of data mart consist of fact table and dimension table. Fact table contain a group of foreign key from primary key contained in each dimension table, beside that dimension table contain data detail that can explain fact table foreign key. There is some Scheme model on data mart modeling, there is star Scheme, snowflake Scheme, and constellation Scheme. Explanation of each Scheme model:

1. Star Scheme
   This scheme looks like a star, which is the fact table become centralizing of Scheme and arround by dimension table. All dimension table connected to the fact table. Fact table has some primary key on dimension table. Example of star Scheme can be seen in Figure 1:

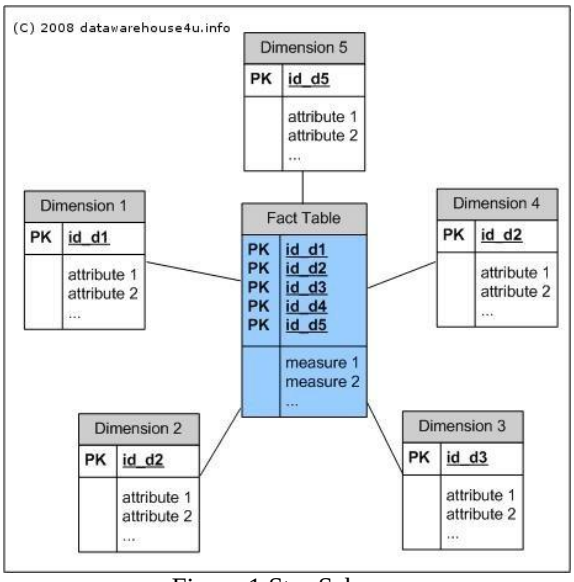


Figure 1 Star Scheme

2. Snowflake Scheme
   Snowflake Scheme is expansion of star Scheme with additional dimension table without connected with fact table. Additional dimension table any connected with another dimension table. Example snowflake Scheme can seen in Figure 2:

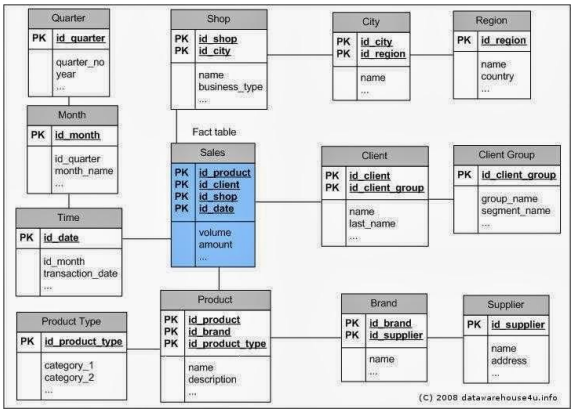


Figure 2 Snowflake Scheme

3. Constellation Scheme
   In the Scheme there are some fact table using one or more dimension tables. Example Scheme constellation can be seen in Figure 3:

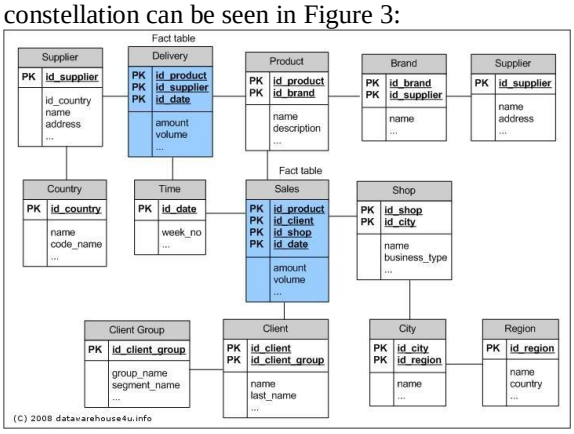


Figure 3 Constellation Scheme

1.3 ETL (*Extract, Transform, Loading*)

ETL (Eztract Transform Loading) Process is a process that must be traversed in the formation of the data mart [2]. ETL is a data processing phase from data source and then entered into the data mart. The purpose of ETL is collect, filter, manipulate and combine data from different sources to be stored into the data mart. The following is a description of each process of ETL:

1. Data Extraction (Extract)

Data extraction is the process of taking data from multiple operational systems, using either the query or ETL applications. There are several data extraction functions, that is:

a. Data extraction automatically from source applications.
b. Filtering or selection of data extracted.
c. Sending data from different application platform to the data source.
d. Format changes the layout of the original format.
e. Storage in file temporary for incorporation with the result of the extraction from other sources.

2. Data Transformation (Transformation)

Transformation is the process by which data extracted filtered and modified in accordance with business rules. Steps in data transformation is as follows:

a. Map the input data from the original data Scheme to Scheme data mart.
b. Convert data type.
c. Clean up and dispose of the same data (duplication).
d. Check the reference data.
e. Fill the empty values with default values.
f. Combine data.

3. Data Entry (Loading)

Data entry is the process of entering data obtained from the result of the transformation into data mart. How to insert data is to run SQL script on a periodic basis.

1.4 OLAP (On-Line Analytical Processing)

OLAP (On-Line Analytical Processing) is a technology that processes data into multidimensional structure, providing quick answer to complex analytical queries with the aim to organize large amounts of data, to be analyzed and evaluated quickly and provide the speed and flexibility to support analysis in real time [3].

There are several characteristics of OLAP, that is:

1. Allow bussiness to see the data from the standpoint of logical and multidimensional.
2. Facilitate complex queries and analysis for user.
3. Allow the user to make drill-down to display more detailed data or roll-up for the aggregation of dimension or same dimension.
4. Provide process calculation and comparative data.
5. Display result in tables or grahps.

Advantages of OLAP, that is :

1. Increase the productivity of the bussiness end users, developers, and the ovelass IT.
2. More supervision and timely access to strategic information can make decisions more quickly
3. Reducting application development for the IT staff to make end use can change the scheme and make your own model.
4. Storage control of the organization through corporate data integrity as OLAP application depends on the data warehouse and OLTP systems to update the data source level.

OLAP can be used to do like [3]:

1. Consolidation (roll-up)

Consolidation involves grouping data.

2. Drill-down

A form which is the opposite of consolidated data to describe succinctly be data in more detail. The description for the roll-up and drill-down can be seen in Figure 4:

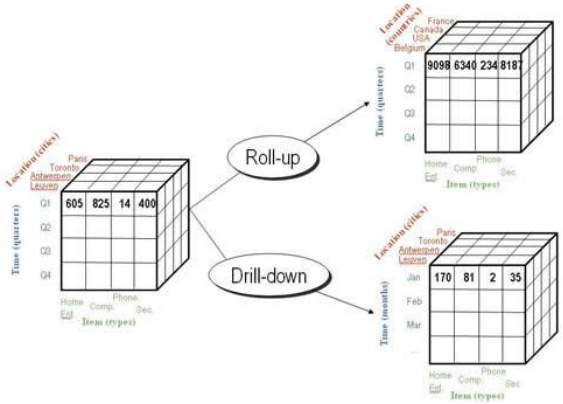


Figure 4 Roll-up and Drill-down

3. *Slicing* dan *dicing*

Describes the ability to see data from the viewpoint. Overview for slicing and dicing can be seen in Figure 5:

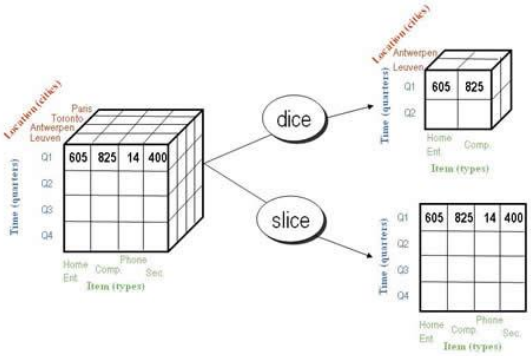


Figure 5 Slicing and Dicing

1.5 SSIS (SQL Server Integration Service)

SSIS (SQL Server Integration Services) is a platform to build a reliable system for data integration, extraction, transformation, and loading that is used in data warehousing [4]. SSIS offers solutions in dealing with the problem of data integration. In addition, theses tools to improve the efficiency of the petrified creation time.

SQL Server Integration Services architecture in general contain various components, such as:

1. SSIS Deginer. Is tools used to create and manage integration service package. In SQL Server 2012, this tools is integrated with Visual Studio 2010 which is part Bussiness Intelegence project.
2. Runtime Engine. This component is useful for running all the SSIS packages that have benn made.
3. Task and executable binary.
4. Data Flow Engine and Data Flow. Data flow component is an encapsulation of data flow engine that providing a buffer in memory ang charge of moving data from data source to destination data. While the data flow is data source, destination data, and transformation.
5. Integration Services service. Enables SQL Server Management Studio can be used to monitoring SSIS package and manage SSIS storage be used.
6. SQL Server Import and Export Wizard. These tools are used to copy the data from source to destination data.

2 RESEARCH CONTENTS

There are several stages in the development of analysis software data mart:

2.1 Information Requirements Analysis

Information requirements analysis is a step to analyze what is needed by PT. Matahari Sentosa to the data mart which will be built. The information will be presented in detail. Based on interview with the production manager, the information needed for the development of a data mart are as follow:

1. Information on number of products produced per year, month, and day.
2. Information on numbers of products produced by type of product per year, per month and per day.
3. Information on numbers of products that has dyeing processed year, month, and day.
4. Information on numbers of products that has dyeing processed based on the type of product per year, per month and per day.
5. Information on numbers of usage of raw material per year, month, and day.
6. Information of raw material consumption is based on the amount of certain raw materials per year, per month and per day.

2.2 Data Source

The Current OLTP databases at PT. Matahari Sentosa become the source of data to build a data mart. OLTP-relationship diagrams of PT. Matahari Sentosa can be seen in Figure 6 below:

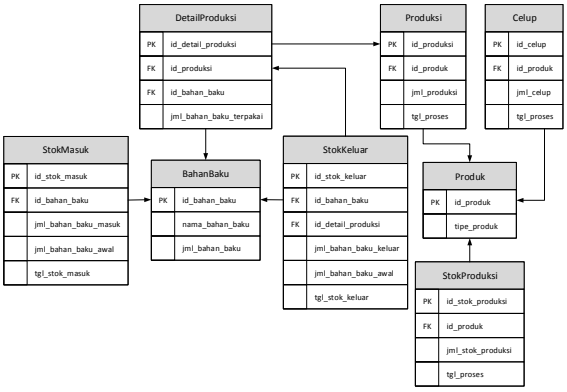


Figure 6 OLTP Relationship Diagram of PT. Matahari Sentosa

2.3 Data Mart Architecture Analysis

A used architecture for the construction of a data mart is a model of a two-layer architecture. Analysis of this architecture is divided into four layers, such as the layer source analysis, data analysis stagging, data mart layer analysis, and analysis using OLAP.

1. *Source Layer Analysis*

On this layer, data is still in the form of operational data. The data source used in the construction of the data is already in the form of a data mart logic in the database.

2. *Data Stagging Analysis*

On this layer, operational data will be extracted by the process of ETL (Extract Transform Loading) into the data mart.

a. *Extract*

This process is the selection of data from the data source for the creation of data marts, which are the product table, the production table, table dye, tables of raw materials, table stock out, and table stock production as well as tables that are not used for the creation of data marts, are table detail production and table entry stock because it is not needed in information needs. The attributes in the table that will be extracted has no changing in replenish or diminish its attributes, it still remains the same as the data source. The process of extracting data from the data source into the data mart are as follows:

Table 1 *Extract*

No	Nama Tabel	Field
1	Table of Product	id_produk
		tipe_produk
2	Table of Production	id_produksi
		id_produk
		jml_produksi
		tgl_proses
3	Table of Dye	id_celup
		id_produk
		jml_celup
		tgl_proses
4	Table of Raw Materials	id_bahan_baku
		nama_bahan_baku
		jml_bahan_baku
5	Table of Stock Out	id_stok_keluar
		id_bahan_baku

		id_detail_produksi
		jml_bahan_baku_keluar
		jml_bahan_baku_awal
		tgl_stok_keluar
6	Table of Production Stock	id_stok_produksi
		id_produk
		jml_stok_produksi
		tgl_proses

b. *Transform*

Transform process which is conducted consists of two processes as follow:

1) *Cleaning*

Cleaning process is performed to clean up unused data from a table that has been extracted, which eliminates unused field. Here are the field names which omitted in the process of cleaning:

a) In the table of Production does not require field id_produksi and tgl_proses.
b) In the table of dye does not require field id_celup and tgl_proses.
c) In the table of raw materials does not require jml_bahan_baku field.
d) In the table of stock out does not require field id_stok_keluar, id_detail_produksi, and tgl_stok_keluar.
e) In the table of stock production does not require field id_stok_produksi and tgl_proses.

2) *Conditioning*

Conditioning process is performed to change the format of the operational data into the format of data marts. Tables are conditioned are on table of production, table of dye, table of stock out, and table of stock production. Conditioning process is carried out to ensure no data redundancy, resulting in a fact table can have more than one dimension table. In Table 2 will be explained an example of the stages of conditioning on the table of product.

Tabel 2 *Conditioning* Table of Product

Tabel Produksi	
No	**Field**
1	id_produk
2	jml_produksi
3	tgl_proses

→

Fact_produksi	
No	**Field**
1	id_produk
2	jml_produksi
3	id_waktu

Tabel Produksi	
tgl_proses	date

→

Dim_waktu	
id_waktu	integer
tanggal	integer
bulan	integer
nama_bulan	nvarchar(50)
tahun	integer
full_date	date

c. *Loading*

Once the data extracted and transformed, then the data is inputted into the data mart. The process of loading on the data mart application will be performed automatically after the transform is complete. Using Update technique, where the process will immediately update the data mart without changing existing data.

3. *Data Mart Layer Analysis*

On this layer, the data that has been through the ETL process will be stored in a logic centralized storage, named a data mart. Scheme which is used in the construction of a data mart is a constellation scheme. Constellation scheme in the construction of a data mart can be seen in Figure 7 below:

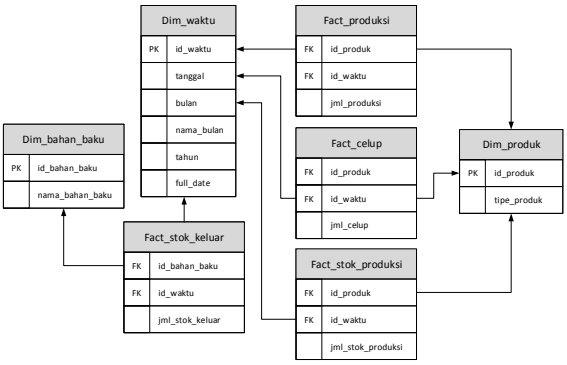


Figure 1 *Constellation Scheme*

Table 3 shows the type of each tables that exist in relation scheme shown in Figure 7.

Table 3 *Data Mart Scheme Explanation*

No	**Nama Tabel**	**Jenis Tabel**
1	Fact_Produksi	Fakta
2	Fact_Celup	Fakta
3	Fact_Stok_Keluar	Fakta
4	Fact_Stok_Produksi	Fakta
5	Dim_Produk	Dimensi
6	Dim_Bahan_Baku	Dimensi
7	Dim_Waktu	Dimensi

2.4 Software Requirements Specification

Software requirements specification Analysis contains a description of the software needs to be built both functional requirements and non-functional requirements. Software requirements specification table can be seen in Table 4 and Table 5.

Table 4 Functional Specifications Software Requirements

Requirement Code	Requirements
SKPL-F-1	Production Administrators and managers can access the application.
SKPL-F-2	The system can update the data in the data mart.
SKPL-F-3	The system can perform a process of ETL (Extract Transform Loading).
SKPL-F-4	Production managers can perform analysis against the data mart.
SKPL-F-5	The system can display the results of the analysis in the form of tables and graphs.
SKPL-F-6	The system can provide a printed report.

Table 5 Non-Functional Software Requirement Specification

Kode Kebutuhan	Kebutuhan
SKPL-NF-1	The system can run on Windows platforms.
SKPL-NF-2	The system can run on medium hardware specification.
SKPL-NF-3	The programming language used is C #.

2.5 Functional Requirements Analysis

Functional requirements analysis intend to outline the activities process and requirement that will be applied to the system to be built.

2.5.1 Use Case Diagram

Use case diagram is a description of the system functionality or behavior. *Use case diagrams* in the software can be seen in Figure 8 below:

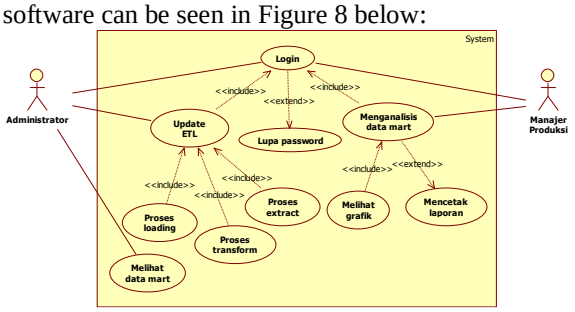


Figure 2 *Use Case Diagram*

2.5.2 Activity Diagram

Activity diagram models the workflow of the order activity in a process that refers to the existing use case diagram. *Activity diagram* of ETL updates can be seen in Figure 9 following:

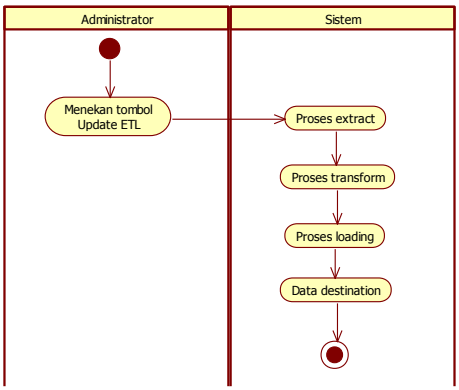


Figure 3 *Activity Diagram of Update ETL*

2.5.3 Sequence Diagram

Sequence diagram describes the interaction between each object in every use case in a time sequence. Sequence diagram to The ETL update can be seen in Figure 10 below:

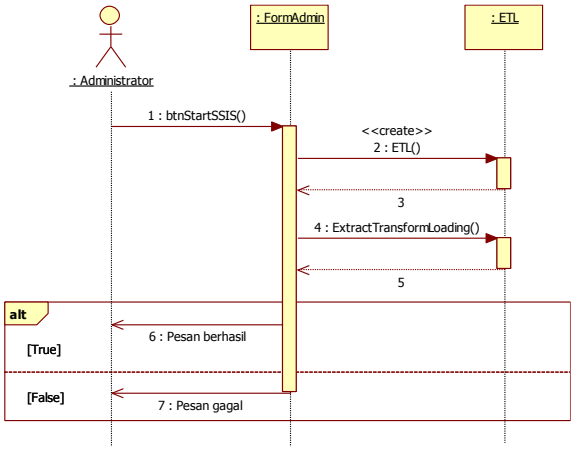


Figure 4 Sequence Diagram of Update ETL

2.5.4 Class Diagram

Class diagram describe the structure and relationships between objects in the system. Class diagram of the software to be built can be seen in Figure 11 below:

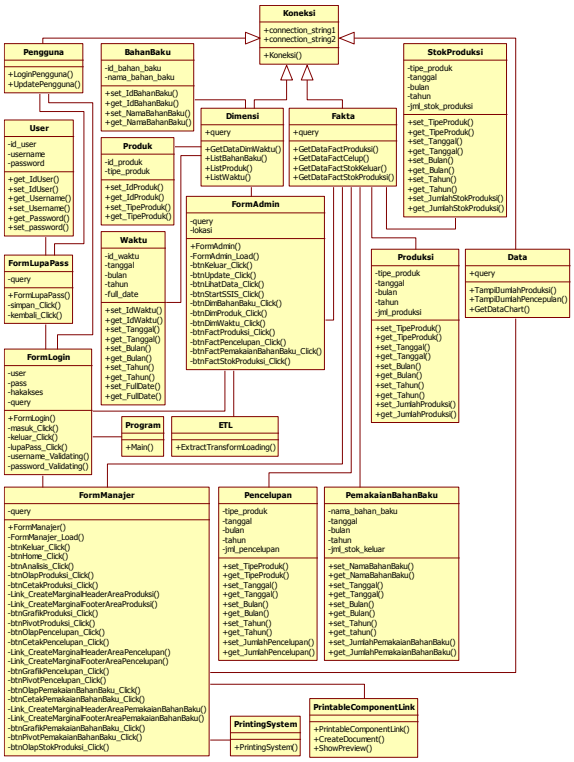


Figure 5 Class Diagram

2.6 Database Implementation

The database is done by using the DBMS Microsoft SQL Server 2012 application. The implementation of the database in SQL language is as follows:

Table 6 Table dim_produk Implementation

Dimensi Produk

```
CREATE TABLE [dbo].[dim_produk](
	[id_produk] [int] NOT NULL,
	[tipe_produk] [varchar](20) NOT NULL,
 CONSTRAINT [PK_dim_produk] PRIMARY KEY
CLUSTERED
(
	[id_produk] ASC
)WITH (PAD_INDEX = OFF,
STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY
= OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS
= ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Table 7 Table dim_waktu Implementation

Dimensi Waktu

```
CREATE TABLE [dbo].[dim_waktu](
	[id_waktu] [int] IDENTITY(1,1) NOT
NULL,
	[tanggal] [int] NOT NULL,
	[bulan] [nvarchar](50) NOT NULL,
	[tahun] [int] NOT NULL,
	[full_date] [date] NOT NULL,
 CONSTRAINT [PK_dim_waktu] PRIMARY KEY
CLUSTERED
(
	[id_waktu] ASC
)WITH (PAD_INDEX = OFF,
STATISTICS_NORECOMPUTE = OFF, IGNORE_DUP_KEY
= OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS
= ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Table 8 Table fact_produksi Implementation

Fakta Produksi

```
CREATE TABLE [dbo].[fact_produksi](
	[id_produk] [int] NOT NULL,
	[jml_produksi] [int] NOT NULL,
	[id_waktu] [int] NOT NULL
) ON [PRIMARY]
GO

ALTER TABLE [dbo].[fact_produksi]  WITH CHECK
ADD  CONSTRAINT [FK_fact_produksi_dim_produk]
FOREIGN KEY([id_produk])
REFERENCES [dbo].[dim_produk] ([id_produk])
GO

ALTER TABLE [dbo].[fact_produksi] CHECK
CONSTRAINT [FK_fact_produksi_dim_produk]
GO

ALTER TABLE [dbo].[fact_produksi]  WITH CHECK
ADD  CONSTRAINT [FK_fact_produksi_dim_waktu]
FOREIGN KEY([id_waktu])
REFERENCES [dbo].[dim_waktu] ([id_waktu])
GO

ALTER TABLE [dbo].[fact_produksi] CHECK
CONSTRAINT [FK_fact_produksi_dim_waktu]
GO
```

2.7 Interface Implementation

The implementation of the software interface consists of several pages. Following the implementation of the interface data mart software.



Figure 6 Login Page

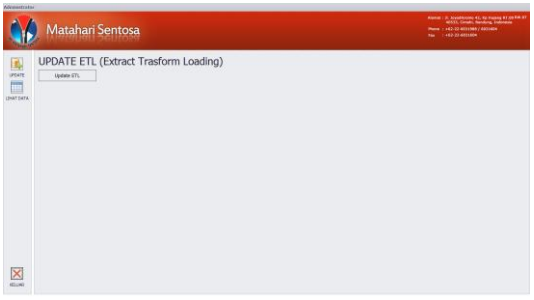


Figure 7 *Update* ETL Page

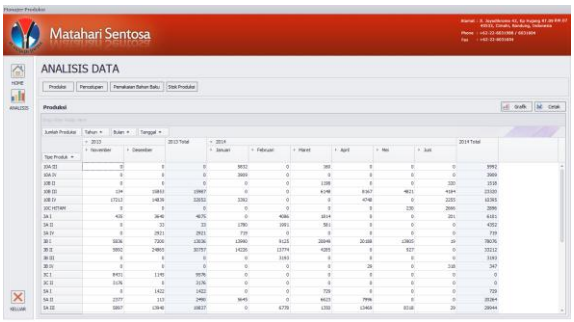


Figure 8 *Data Mart Analyzing* Page

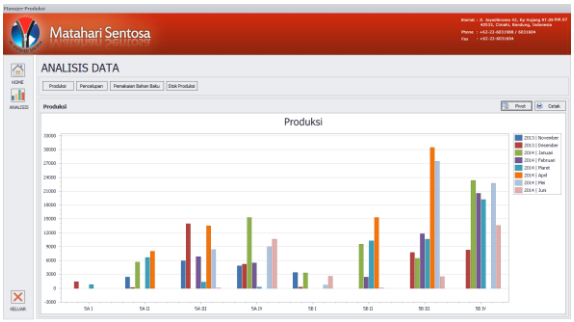


Figure 9 Graphic View Page

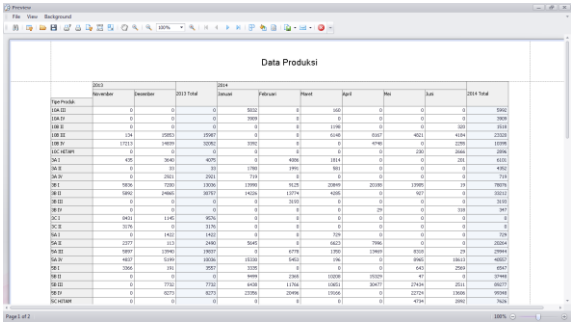


Figure 10 Print Report Page

3 CLOSURE

3.1 Conclusion

Based on research that has been done on the analysis, design, implementation, and testing of data mart software for the production division at PT. Matahari Sentosa, it can be concluded as follows:

1. Data Mart application provides convenience in obtaining strategic information quickly which required by the production manager of PT. Matahari Sentosa for long-term planning, can provide information that is multidimensional and concise, and to accelerate the process of analyzing in order to maximize a decision made by the production manager for the handling stocks of products that are produced and handling raw materials for further production.
2. Data Mart Application helps in making a multidimensional final report production required by the production manager at PT. Matahari Sentosa.

3.2 Advice

Here are some suggestions that can be used for further data mart software development are as follows:

1. The system not only analyzes the Production Division, but also analyzing all the divisions in PT. Matahari Sentosa in order to meet the company's strategic information.
2. For Data mart software development that have been made can be developed again for the manufacture of data marts in other divisions or data warehouse for PT. Matahari Sentosa.

BIBLIOGRAPHY

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

[2] R. Kimball and M. Ross, The Data Warehouse Toolkit, Indianapolis: John Wiley & Sons, Inc, 2013.

[3] M. Y. Pusadan, Rancang Bangun Data Warehouse, Yogyakarta: Graha Ilmu, 2013.

[4] B. Knight, D. Knight, M. Davis and W. Snyder, Microsoft SQL Server 2012 Integration Services, Indianapolis: John Wiley & Sons, Inc, 2013.