MODELING OF DATA WAREHOUSE ON RSUD H. HANAFIE MUARA BUNGO USING KIMBALL METHOD

M. Habibie Luktaputra[1], Dian Dharmayanti[2]

[1,2] Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112-116, Lebakgede, Coblong, Kota Bandung, Jawa Barat 40132
E-mail : mluktaputra@gmail.com[1], dian.dharmayanti@email.unikom.ac.id[2]

ABSTRACT

The development of information technology has greatly influenced the habits of an agency or organization in information processing. The use of applications that can process data to make report material has become one of the habits of agencies and organizations. H. Hanafie Hospital Muara Bungo is very difficult in making good reports. The report in question contains reports on all data on the personnel sector, medical services, pharmacies and cashiers. In making the final report takes a long time because all the data reports from each separate field and not integrated with each other in the Hospital H. Hanafie Muara Bungo still use the application partially which each field has its own application. Based on the above problems, an initiative emerged that H. Hanafie Muara Bungo Regional Hospital needed a data storage area such as Data Warehouse and OLAP & Reporting Tools that could improve the performance of the RSUD. With the data warehouse modeling and OLAP & Reporting Tools, it will be easier for the hospital in making the final report and analysis for the decision making of the RSUD in the future.

Keyword : data warehouse, kimball, Olap.

1. INTRODUCTION

The development of information technology has greatly influenced the habits of an agency or organization in information processing. Many hospitals have used information systems for processing data in order to accelerate the making of good recapitulation in the form of reports. With the development of the times as it is now, it is not enough only applications that can make a report recapitulation, but it is also needed an information that can be used to make managerial and executive decisions in the Hospital H. Hanafie Muara Bungo.

Based on the results of observation, RSUD H. Hanafie Muara Bungo is very difficult in making a good final report. The final report referred to contains reports on medical services, medical support, and drug sales. In making the final report, it takes quite a long time because all the final reports from each field are separate and not mutually integrated, because H. Hanafie Muara Bungo Regional Hospital uses partial applications, each field has its own application.

Based on the above problems, that RSUD H. Hanafie Muara Bungo needs a data storage place such as Data Warehouse and OLAP & Reporting Tools with the Kimball method that can improve the performance of the RSUD. The data warehouse itself is a collection of subject-oriented, integrated, time-variant and non-volatile data in supporting decision making [1].

The aim to be achieved from this research is to accelerate the RSUD in collecting, processing, and presenting data in the form of multidimensional and concise information so as to maximize the quality of decisions made by the managerial and executive of RSUD.

1. RESEARCH CONTENT

2.1 Theoretical Basic

The theoretical foundation aims to provide explanations from various sources and theoretical studies relating to the Data Warehouse Modeling of H. Hanafie Hospital Muara Bungo With the Kimball Method.

2.1.1 Data Warehouse

The data warehouse is multi-dimensional, meaning that the data warehouse consists of many layers, rows and columns. The data warehouse has screens that will show information delivered to different dimensions. When extracting information from many operational databases to create a data warehouse, the data warehouse combines information by summarizing and aggregating [1]. The collection of required information can be defined by the consent of the users in accordance with the information needed in making decisions to be made [4].

The characteristics of the data warehouse according to Inmon [1] are :

1. Subject Oriented

   The data warehouse is subject oriented, by analyzing data from designs based on specific subjects in agencies or organizations, not on specific processes or applications. the data warehouse is made on the main subjects such as medical services (inpatient and outpatient care), medical support and drug sales.

2. Integrated

The data warehouse can be a place for storing data from various separate sources into a format that has been defined and mutually connected to each other. Terms in the integration of data sources that can be fulfilled in such a way as consistent with naming variables, consistent with the size of the variables, consistent with the coding structure and the system in the physical attributes of the data source. With this in mind the data cannot be broken because the existing data is a unit that will support all the concepts in the data warehouse itself.

3. Time Varian

Time varian merupakan salah satu karakteristik dalam data warehouse karena seluruh data dapat dikatakan valid atau akuran pada rentang waktu tertentu. Untuk mengukur tingkat keakuratan data warehouse dapat dilakukan dengan melihat interval waktunya. Sebagai contoh yang paling sederhana untuk menyajikan data warehouse pada rentang waktu tertentu, misalnya antara 5 sampai 10 tahun kedepan.

4. Non Volatile

Time variants are one of the characteristics in the data warehouse because all data can be said to be valid or concise at a certain time span. To measure the accuracy of the data warehouse can be done by looking at the time interval. As the simplest example to present a data warehouse for a certain time, for example between 5 to 10 years..

2.1.2 Method of Building a Data Warehouse

In the construction of the data warehouse of RSUD H. Hanafie Muara Bungo the method used was the method from Ralph Kimball. The Kimball method has nine steps, as follows :

1. Choose the Process

Selecting the process is the first process in the Kimball method. Choosing a process means having to determine the main process refers to the business processes contained in the organization or agency to answer all important business questions and have certain characteristics [2].

2. Memilih Grain

Selecting a grain is determining what can be represented by a fact table. After you can determine the grain from the fact table, then you can also specify the dimension tables related to the fact table. Grains in the fact table can also determine the grain in the dimension table[2].

3. Identify Dimensions

Identifying dimensions is done to connect dimensions with fact tables. Dimension is an important point of view to describe the facts contained in the fact table [2].

4. Choose Fact

In selecting a fact table, it can be seen from the Grain that has been previously selected and can determine the facts that can be used [2].

5. Store Pre-Calculation in the Fact Table

At this stage the selected fact table will be checked again for an attribute for storage in the database. This is done to reduce the risk of errors in the program every time it calculates the attributes [2].

6. Round Out the Dimension Tables Dimensions that have been previously

selected at this stage will be made a dimensional description that will provide information on the structures in the dimension table and will be given a complete description to facilitate users [2].

7. Choose the Durations of the Database

Selecting the duration in the database is done by the author to determine the duration of time entered into the data warehouse [2].

8. Determine the Need to Track Slowly Changing Dimensions

Slow Changing Dimension is used to track changes that occur in dimensions that have been created. According to Kimball, there are three categories of SCD, namely: rewriting the changed attributes (type 1), creating a new record in dimensions (type 2), and making an alternative attribute to accommodate the new value (type 3) [2].

9. Decide the Physical Design

Determination of priority determination and query model is done to consider the influence of the physical design that has been made [2].

2.1.3 ETL Process

The ETL process (Extract, Transform, Load) is what must be in the data warehouse, ETL processes occur when data starts to exit the Staging Area [5]. At this stage, data will be extracted and sent to the Staging Area. The Staging Area is a temporary location for data that has been formatted (Transform) so that the data that has been formatted can be prepared to the next process. Then the data that has been changed. Will be loaded (Load) to the table in the data warehouse [8].

1. Extract

The first part of the ETL process is extracting from operational database sources, because not all data in the operational database will be taken to be entered into the data warehouse.

2. Transform

The Transform process is the transformation of the data that has been extracted in the previous stage into the formal required data warehouse. In the transform process it is also useful to standardize the previous different databases into the same data warehouse.

3. Load

The last process in ETL is the process of loading. In this process the data in the ETL process will be loaded into the data warehouse.

2.1.4 Multidimensional Modeling

Multidimensional Modeling is a technique for making visualization of a data model as a collection

of sizes described to aspects of a business. In this case it is very useful to compile and summarize data to show data in support of analysis of data.

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

1. Star Schema

Star Schema is the basic model of multidimensional modeling which can be seen in the 1 star schema model having a fact table as a parent table and surrounded by two or more dimension tables.

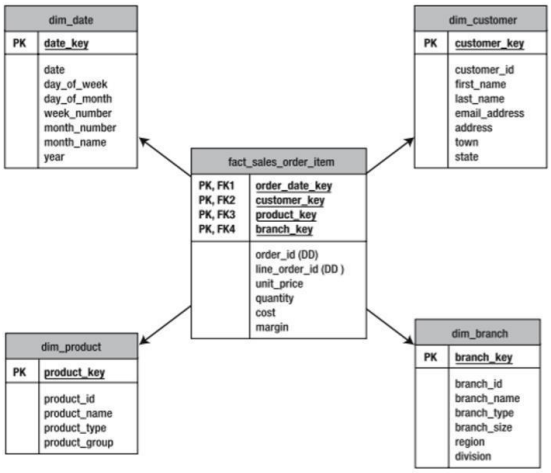


Figure 1 Star Schema [3]

2. Snowflake Schema

In addition to Star Schema, Snowflake is also the basic model of multidimensionality, the snowflake model as an example can be seen in Figure 2 in the snowflake model, there is a fact table as the parent table and surrounded by several dimension tables. The difference between Star Schema and Snowflake can be seen basically in Figure 2, which is as if the dimension table is a fact table surrounded by derivatives from the parent dimension..

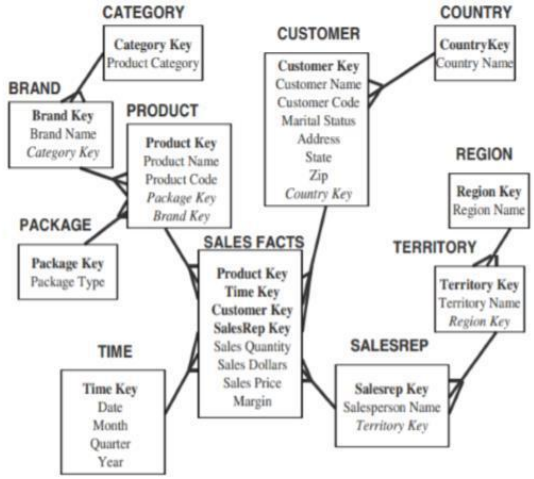


Figure 2 Snowflake [3]

2.1.5 On-line Analytical Processing (OLAP) OLAP is a technology that can accelerate the need for analysis by using multidimensional image technology in a number of data. In the need for data warehouse analysis, OLAP or Cube models classify the data so that they can make realization between the data and form a special understanding in the analysis in accordance with the information needs of the users [7,10].

Cube chooses three components such as a cube that has three sides and the three sides in the cube are dimension, time, dimension, and measure [9].

2.2 Analysis

2.2.1 Problem Analysis

Analysis carried out at RSUD H. Hanafie Muara Bungo found some problems in the RSUD. Making a report is a very common thing to do especially in an RSUD. Reports are used by hospitals as one of the work evaluation materials and also useful to provide important information on the development of hospitals and operational activities that have been carried out by hospitals. The following is an analysis of the problems that occur at RSUD H. Hanafie Muara Bungo :

1. With the number of reports in each division, in the preparation of reports, the collection and the need to review the data collected by the RSUD can take a very long time.
2. With the large number of data reports at the RSUD H. Hanafie Muara Bungo and done manually every division, it will be very difficult to analyze the data obtained to review what steps will be taken in the future by the H. Hanafie Hospital Muara Bungo.

2.2.2 Data Analysis

Data analysis is a stage of analysis of data processed in the system or procedure that is running. The following is the relation scheme in Figure 3 in RSUD H. Hanafie Muara Bungo.

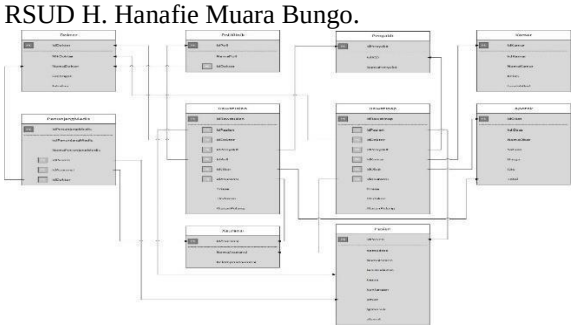


Figure 3 Relationship Scheme of H. Hanafie Hospital Muara Bungo

2.2.3 Choose Business Process

The main business processes in the RSUD H. Hanafie Muara Bungo according to the results of interviews with the executive of H. Hanafie Hospital Muara Bungo, the main business processes according to the organizational structure of the hospital are

medical fields and there are two parts of the medical field [6], as follows :

1. Medical Services
   a. Acceptance of Inpatients
   b. Acceptance of Outpatient Patients
2. Medical Support
   a. Acceptance of Medical Support Patients
   b. Purchasing Pharmacy Medicines

2.2.4 Indentify Grain

From the results of the identification of the field process, we obtained the grain from the identification of the field process as follows:

1. Inpatient information in every month and year
2. Outpatient information in every month and year
3. Number of medical support patients in every month and year
4. Number of drug sales at pharmacies in every month and year

The grain table explains what strategic information is available at the H. Hanafie Hospital Muara Bungo. There are 4 Grains and 20 strategic information, the following are samples from Grain and strategic information can be seen in table 1.

Table 1 Grain and Strategic Information Sample

No	Grain	Informasi
1	Number of inpatients	1. Total number of inpatients based on gender in each month and year 2. The total number of inpatients is based on the group of rooms in each month and year
2	Number of Outpatients	The total number of outpatients based on gender in each month and year
3	Number of Medical Supporting Patients	The total number of medical support patients based on gender in each month and year
4	Number of Drug Sales at the Pharmacy	The total number of drugs sold from each medical service in each month and year

2.2.5 Indentify Fact

After embarrassing grain selection, each fact can be designated for each strategic information needed by the RSUD based on the identification that has been done when determining the grain. The following is a sample of the fact table selection in table 2.

Tabel 2 Sample Choose Fact

No	Fakta	Informasi
1	Inpatient Facts	1. Total number of inpatients based on gender in each month and year
		2. The total number of inpatients is based on the group of rooms in each month and year
2	Outpatient Facts	The total number of outpatients based on gender in each month and year
3	Medical Support Facts	The total number of medical support patients based on gender in each month and year
4	Facts of Drug Sales	The total number of drugs sold from each medical service in each month and year

2.2.6 Indentify Dimensions

At this stage identification and adjustments will be made to the selected grain. The following sample in the 3-dimensional table is obtained from the identification and adjustment of the grain.

Table 3 Dimensions Sample

Informasi	Dimensi
Total number of inpatients based on gender in each month and year	Dim_pasien Dim_jeniskelamin Dim_waktu
The total number of inpatients is based on the group of rooms in each month and year	Dim_pasien Dim_kamar Dim_waktu
The total number of outpatients based on gender in each month and year	Dim_pasien Dim_kamar Dim_waktu
The total number of medical support patients based on gender in each month and year	Dim_penunjangmedik Dim_jeniskelamin Dim_waktu

2.2.7 ETL Process

In this ETL process, the RSUD database operations will be carried out into the data warehouse. The stages are as follows :

1. Extraction Process

When extracting data, it will form a new field to show the timestamp. Table 4 will load the table and the fields that will be extracted from the operational database into the data warehouse.

Table 4 Tables along with fields extracted

No	Nama Tabel	Field
1	Pasien	idPasien
		kdPasien
		NamaPasien
		JenisKelamin
		Kasus
		Kunjungan
		Umur
		Asuransi
		tglmasuk
		Alamat
2	Dokter	idDokter
		kdDokter
		NamaDokter
		JenisKelamin
		golongan
		Jabatan
3	Kamar	idKamar
		kdKamar
		NamaKamar
		Kelas
		JumlahBed
4	Asuransi	idAsuransi
		KelompokAsuransi
		NamaAsuransi
5	PenunjangMedik	idPenunjangMedis
		kdPenjangmedis
		NamaPenunjangMedis
		idPasien
		idAsuransi
6	Penyakit	idPenyakit
		kdICD
		NamaPenyakit
7	Klinik	idKlinik
		NamaKlinik
		idDokter
8	Obat	idApotek
		kdObat
		NamaObat
		Satuan
		Harga
		QTY
		Total

2. Transform Process

The Transform process is done after the extraction process is complete. In the transformation process, the selection process and conditioning process will be carried out according to the needs needed by the RSUD.

a. Selection Process

In this selection process is selected from the table that has been extracted in the previous process. In this process the selection of the field in question removes fields that are not useful in the analysis process. The following is a sample table of patients that will be removed from the field and become a new table that will be used as the analysis process needs.

Tabel 5 Sample Selection on Transform in Patient Tables

Pasien	
No	Field
1	idPasien
2	kdPasien
3	NamaPasien
4	JenisKelamin
5	Kasus
6	Kunjungan
7	Umur
8	Asuransi
9	tglmasuk
10	Alamat



Pasien	
No	Field
1	idPasien
2	kdPasien
3	NamaPasien
4	JenisKelamin
5	Kunjungan
6	IdUmur
7	IdAsuransi

b. Conditioning Process

In the conditioning process, we will select tables and attributes from operational data sources to the data warehouse. The conditioning process in the tranformation process in the analysis needed is to change the entry date field into several fields, namely date, month, and year. This conditioning process aims to facilitate data analysis based on the desired time interval in accordance with the characteristics of the data warehouse, namely time variant. The following is in table 5 is the result of the conditioning process.

Tabel 6 Conditioning Table Date of Patient Entry

TglMasuk
1/03/2017
20/3/2017
15/04/2017
30/05/2017
6/06/2017

Dim_waktu		
Tgl	Bulan	Tahun
1	03	2017
20	03	2017
15	04	2017
7	11	2014
30	11	2014

3. Loading Process

The Loading process is the last process in ETL, the data has been extracted, cleaned, and formatted according to the analysis needs in the data warehouse. The loading process in the data warehouse is done automatically after the transform process is complete. The technique used in the loading process is an update. This process will immediately update the data warehouse without changing data from operational databases.

2.2.8 Data Warehouse Layer Analysis

In this analysis, data that has gone through the ETL process will be stored in a centralized logic storage, the data warehouse. Based on the analysis that has been carried out in accordance with the

desired information needs, it will create a scheme for the data warehouse. By carrying out the analysis, the RSUD data warehouse needs to use the snowflake model for its multidimensional model. For more details, we can see the data warehouse relation scheme in Figure 4 below.

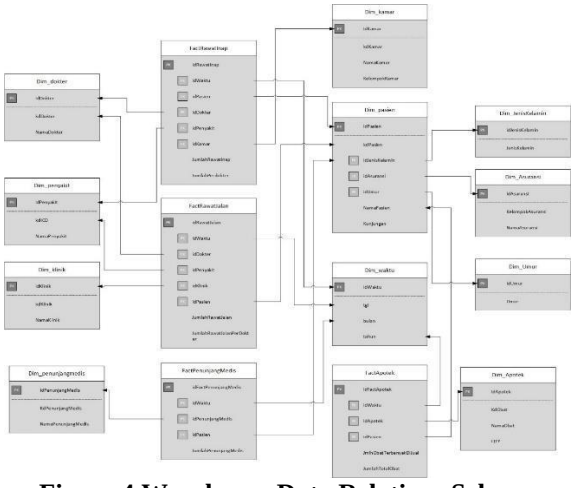


Figure 4 Warehouse Data Relations Scheme

2.2.9 OLAP

The analysis technique used in the OLAP process is drill-up and drill-down. Roll-up and drill-down techniques are chosen because they will help in the process of presenting data based on each dimension and derivative of each dimension. Besides this technique can present detailed and general data from each dimension based on the desired time interval. The other techniques used are slice and dice which will help in viewing from several perspectives. Roll-up and drill-down techniques and slice and dice can be seen in the pivot table. An example is doing the OLAP process for inpatient information based on the sex in each buldan and year. In this strategic information involves three dimensions namely patient dimensions, gender dimensions and time dimensions. For more details, see picture 5 below :

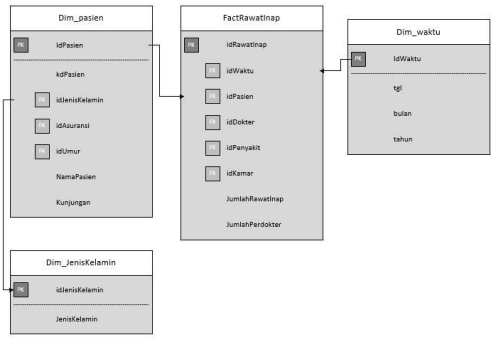


Figure 5 Strategic Information Relations

An example of OLAP Cube which is formed from strategic information can be seen in Figure 6 below.

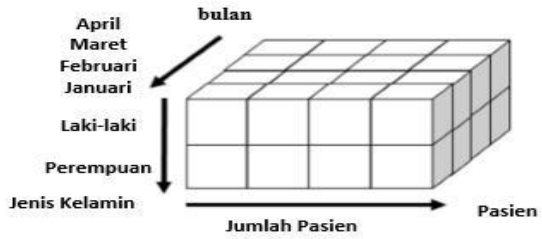


Figure 6 OLAP Cube

The technique used for the strategic information above is slice and dice. For example slice and dice for inpatients based on sex in February and March. Slice and dice illustrations can be seen in Figure 7 below.

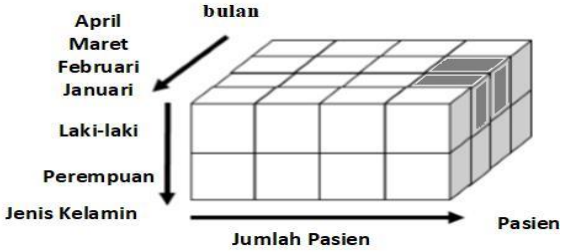


Figure 7 Slice and Dice Informasi Strategis Roll-up and drill-down techniques will be used for the time dimension, which is the month. Examples of record results from slicing and dicing data from strategic information can be seen in Table 7 below.

Tabel 7 Hasill Slicing and Dicing

Bulan	Jenis Kelamin	Jumlah Pasien per Jenis Kelamin
Februari	Laki-Laki	43
	Perempuan	72
Maret	Laki-Laki	38
	Perempuan	51

2.2.10 Data Visualization

At this stage data visualization techniques are used to form charts such as bar charts, pie charts, and line charts. In addition to the technical charts used in data visualization are dashboards. Dashboard is a summary of data that already exists in RSUD H. Hanafie Muara Bungo where data will be formed with data visualization to make it easier for users to understand and also be able to display information quickly.

In the case of this research data visualization that is used to display strategic information are bar charts and pie charts. Table 8 will show a sample of the overall strategic information and data visualization that will be used in this study.

Tabel 8 Data Mapping Sample Visualization

Informasi	Visualisasi
Total number of inpatients based on gender in each month and year	Bar Chart dan Pie Chart
The total number of inpatients is based on the group of rooms in each month and year	Bar Chart

The total number of outpatients based on gender in each month and year	Bar Chart dan Pie Chart
The total number of medical support patients based on gender in each month and year	Bar Chart dan Pie Chart
The total number of drugs sold from each medical service in each month and year	Bar Chart

2.3 Result

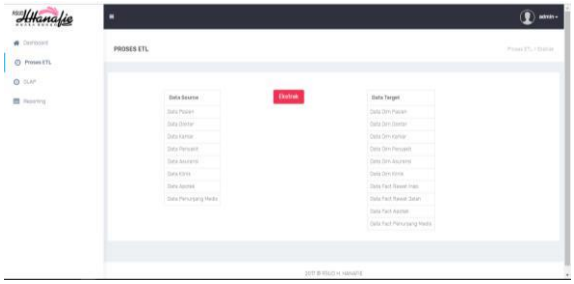
1. Login Display



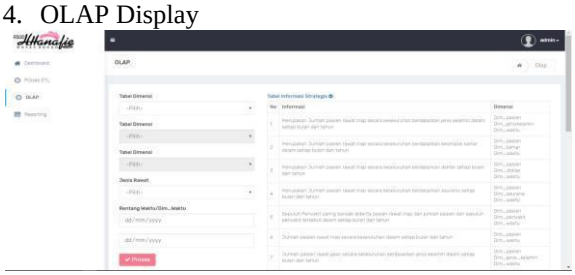
2. Dashboard Display



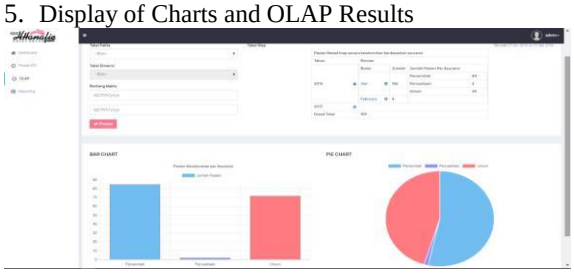
3. ETL Process Display



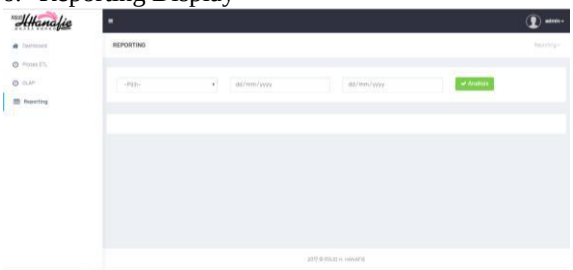
4. OLAP Display



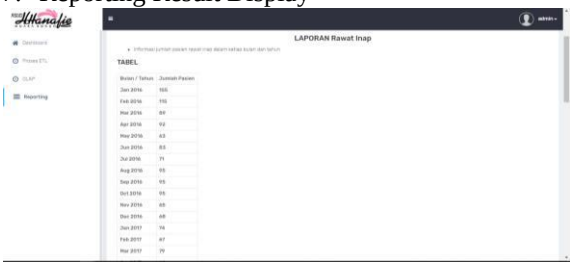
5. Display of Charts and OLAP Results



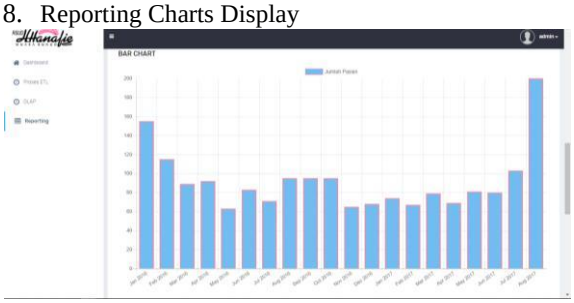
6. Reporting Display



7. Reporting Result Display



8. Reporting Charts Display



3. PENUTUP

3.1 Kesimpulan

Based on the results of the implementation and testing that has been carried out, it can be concluded from the thesis entitled Development of Data Warehouse RSUD H. Hanafie Muara Bungo With the Kimball Method as follows:

1. Can help the administration and the executive in the process of making reports and analysis quickly and also more interesting with the information visualization.
2. Can help the administration in the RSUD report process.
3. Can help the administration in data management and storage.

3.2 Saran

In the construction of Data Warehouse, RSUD H. Hanafie Muara Bungo is of course still far from perfect and still has shortcomings. Therefore, further development and refinement is needed to meet user needs. The suggestions for developing this data warehouse for the following are as follows::

1. The development of this data warehouse can be directly synchronized and integrated with the SIMRS program in RSUD H. Hanafie Muara Bungo.
2. The development of this data warehouse is expected not only to be applied to the main service process, but all other business processes in the H. Hanafie Hospital Muara Bungo.

DAFTAR PUSTAKA

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

[2] Kimball, R, The Data Warehouse 4th Edition, Canada: John Wiley & Sons Inc, 2002

[3] Ponniah, P, “Data Warehouse Fundamentals: a Comprehensive Guide for IT profesional.” John Wilay end Sons, Inc, New York, 2001

[4] Dhamaryanti, D, dkk. " Pemodelan Data Warehouse Pada Jurusan Teknik Informatika Unikom." Majalah Ilmiah Unikom, Vol.12 No.2, 2015.

[5] Priskawati,D, Dian Dhamaryanti, "Pembangunan Independent Data Mart Pada Optik Yuda." Jurnal Ilmiah Komputer dan Informatika, 2016

[6] Sutiyono, " Analisis dan Perancangan Data Warehouse Sebagai Alat Untuk Monitoring Jalannya Proses Bisnis " Tesis, Universitas Komputer Indonesia, 2015

[7] Arifin, dkk. " Perancangan dan Impelentasi Data Warehouse Pada CV. Diandra Primatitra Media Yogyakarta." Seminar Nasional Teknologi Informasi dan Multimedia, 2014

[8] Muyassir, A, “ Perancangan dan Implementasi Data Warehouse Pada Pondok Pesantren Salafiyah Syafi’iyahSukorejo Situbondo,” Tesis, Universitas Gajah Mada, 2011

[9] Giat, K, “Pengembangan Data Warehouse dan AplikasiOn-lineAnalytycal Processing(OLAP) Untuk Penemuan Informasi dan Analisis Data(studi kasus STMIK AMIKOM Purwokerto),” Tesis, STMIK AMIKOM Yogyakarta. 2011

[10] Purwanto, D, Pengertian OLTP, ETL, OLAP dan Data warehouse https://www.kompasiana.com/dhephe/552e1db26ea834f73d8b45b7/apa-itu-oltp-etl-olap-dan-datawarehouse. 20 Juli 2018 02.34