Designing Data Warehouse At the Inpatient and Outpatient (Case Study: Sanglah Hospital)

Putu Veda Andreyana[1], Putu Angelina Widya[2], and Made Suartika[3]

[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3] Department of Electrical and Computer Engineering, Udayana University
Email: vedaandre@gmail.com

Abstract— **The growth of patient data increasing the hospital resulted in even harder to compile data and analyze the data manually, so it takes a data warehouse that can perform this task automatically. The design of the data warehouse Sanglah hospital aims to build a data warehouse that can store data in structured and easier to analyze the data to make a decision. Design methods Fact Constellation Schema and method used is a Nine-step methodology consisting of nine stages, namely the Electoral Process, Selection of Grain, identification and adjustment of the dimensions, the Electoral Facts, Storage pre-calculation in the fact table, Ensuring the dimension tables, the Electoral duration database, Track changes of dimensions is slowly, prioritization and query model. Designing Data Warehouse Sanglah Hospital helpful enough for data processing in large enough quantities, so expect the needs and information about the patient can be met. Data warehouse Sanglah Hospital can be used to analyze patient data in order to get information on the number of patients of various dimensions.**

Index Terms— **Data Warehouse, Data Patient, ETL, Nine-step methodology, Fact Constellation Schema**

I. INTRODUCTION

The role of data in a hospital has become a very necessary thing to be able to promote the services provided. Along with the high demand for data needed by the hospital, the higher the technology needs to be achieved so that the need for appropriate information needs biperlukan by a hospital [1].

General Hospital Center Sanglah is a government agency that provides health services to the community. Along with the development of Sanglah Hospital, more numerous and complex the data medic yng owned and which should dioleh.

Great medical data difficult for the hospital to conduct an analysis of the data. To be able to handle large amounts of data and use it to the maximum, necessary to design information technology that is more to be overcome, namely the Data Warehouse, which can speed up the process of collecting data for presentation of information that is multidimensional and concise so as to maximize the quality of decisions made by the executive hospital.

Data Warehouse is a repository of information collected from various sources, which are kept in one place under the integrated scheme [10]. The data warehouse is a database that is specifically designed with a structure to perform query and analysis [12]. The data stored in the data warehouse is historical and can be used for long-term planning. With the data warehouse executives can more easily see the large amounts of data to support management decision making, and can also provide information of various media [2].

The most important thing in the process of building a data warehouse is the modeling process [3]. Data Warehouse development success greatly depends on the integration ofassurance qualitydata to all phases of the development of Data Warehouse [15]. From this data model is designed to be able to measure the extent to which the data warehouse is built capable of generating managerial information that is needed. Modeling the most common data warehouse is the dimensional model data. Modeling data warehouse can be done by various methods. One is a model Fact Constellation Schema.

II. STUDY LITERATURE

A. Literature Review

Various data warehouse modeling has been applied to help a business process in an organization. The data warehouse is used for the analysis of the needs of management at the Regional General Hospital Palembang Bari [2]. To better support the services that can be provided in the future, the design of the Data Warehouse Patients were also conducted at the Hospital Muhammadiyah Palembang [1]. The results of the analysis of the data warehouse also helps make it easier to analyze and complete information, quickly and accurately to support the process and planning, evaluation, and making the right decisions for the Library in College [4].

B. Database

Database is a collection of data organization to serve many applications efficiently by centralizing data and controlling data redundancy [5]. In designing a database, one thing to note

is efficiency. The number of redundant data that can lessen the efficiency of the database that needs to be normalized.

This database is used not only by one person or one department, the database can be used by all departments in the company. This database will be the source of the data used jointly in the company. The database is no longer owned by one department but the source of the company share. To get the database, the database simply not enough. Required Database Management System (DBMS) to be able to use the database.

C. Data Warehouse

Data Warehouse is a built to overcome the technical and business issues in cases related to the use of data and information to make business decisions and management [6]. Data Warehouse is not only focused on data storage but also a system with the ability to accept (retrieve), analyze the data (analyze), extract (extract), change (transform), load the data, organize the data dictionary. Data Warehouse is one part of Business Intelligence.

Data Warehouse can also be interpreted a set of data that is subject-oriented, integrated, time-variant, and nonvolatile in support of the decision making process [13]. Data Warehouse aims to enable companies to use the archive data to gain business advantage. The data warehouse allows the integration of various types of data from a wide variety of applications or systems. This ensures access mechanism "one door" for management to obtain information, and analyze it for decision-making [14].

Data Warehouse is not only used for decision support, data storage is also a current and historical data that can potentially be used managers in the company or organization.

D. Dimensional Modelling

Concepts Warehouse data modeling in modeling dimensionality known generally as follows:

1. Star Schema

The star schema structure is a structure that can be easily understood by the users and with which they can comfortably work. the structure mirrors how the users normally view their critical measures along their business dimensions [7].
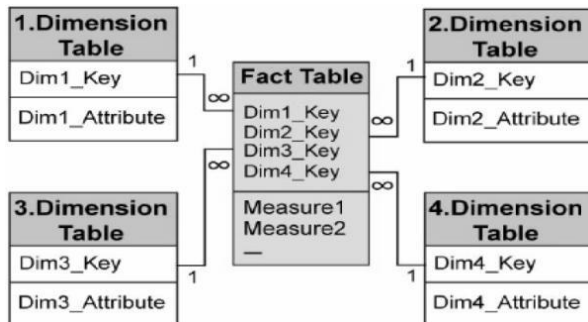

Fig 1. Star Schema

2. Snowflake Schema

Snowflake schema is method of normalizing the dimension tables in a star schema when you completely normalize all the

dimension tables, the resultant structure resembles a snowflake with the fact table in the middle [7].
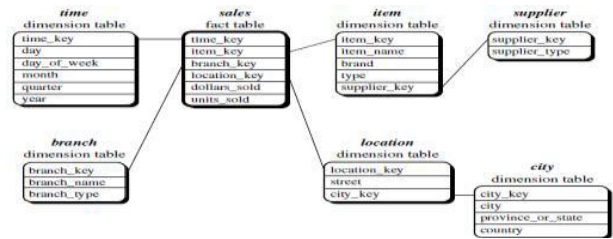

Fig 2. Snowflake Schema

3. Fact Constellation Schema

Fact constellation schema is the schema that contains more than one table to the fact that share dimension tables.
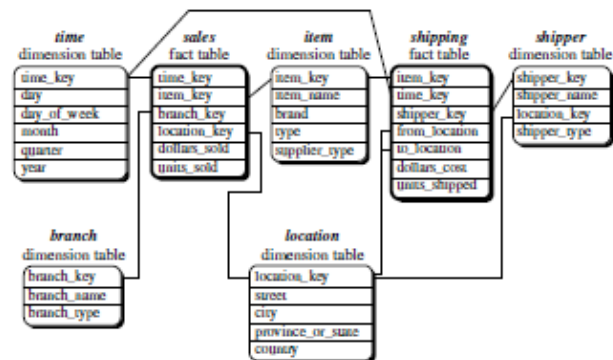

Fig 3. Fact Constellation Schema

E. ETL (Extraction, Transformation, and Loading)

ETL is the process of retrieving and changing the data from the source system and put it into the data warehouse [8]. In the process of data collection, the data must be clean in order to obtain good data quality. For example, no telephone number is invalid, there is a code book that does not exist anymore, there are some data that is null, and so forth. The traditional approach to the ETL processes retrieve the data from the data source, put on the staging area, and then transform and clicking loaded into the data warehouse.

In Figure 4 below alternative described in the ETL process. The first process using the staging area and the second process does not use staging areas within the meaning of staging area is not stored in the disc [8].
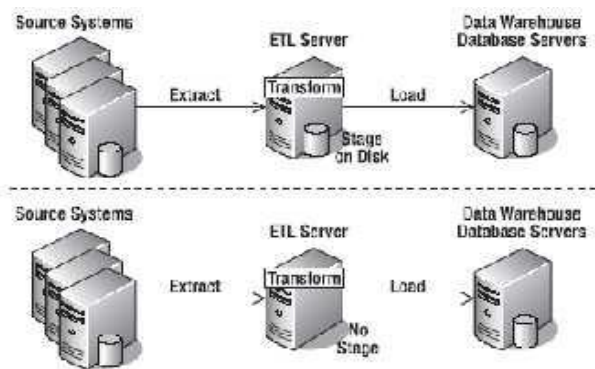
Fig 4. The Approach in the ETL process

F. Microsoft SQL Server 2008

SQL is a language used to access relational databases belonging, not limited only to retrieve data (query), but can also be used to create the table, delete data in the table, replace the data on the table, and various other Operational [1].

G. Microsoft SQL Business Intelligence Development Studio

SQL (Structured Query Language) is a language used to access databases belonging relational, not limited only to retrieve data (query), but can also be used to create the table, delete data, change Operational datal and various others.

Microsoft SQL Business Intelligence is a process to analyze the data, identify trends, and certain patterns contained in the data. Use to help generate analysis report which helps management in making business decisions.

H. Metodology Nine-step

Design method Data Warehouse according to Kimball used include nine stages known as the Nine-Step Methodology [9], the Ninth stage, namely:
1. Choosing the process.
2. Choosing the grain.
3. Identifying and conforming the dimensions.
4. Choosing the facts.
5. Storing pre-calculations in the fact table.
6. Rounding out the dimension tables.
7. Choosing the duration of the database.
8. Tracking slowly changing dimensions.
9. Deciding the query priorities and the query modes.

III. RESEARCH METHODOLOGY

A. Selection process

Based on the scope and conditions provided by the Sanglah Hospital, then some processes will be used in the design of data warehouse include:
1. Process Inpatient
Process hospitalized at Sanglah Hospital which is a process that starts hospitalization of patient enrollment, patients treated in inpatient units until the treatment is completed. The data is used as the code of the patient, gender of the patient, the patient's age, name of insurance used by a patient, the

name of the clinic where patients seek treatment, names of the rooms of hospitalization, the class of the room, on the inpatient admission, the disease suffered by the patient.

2. Process Outpatient
Outpatient process at Sanglah Hospital starting from patient registration, patient examination in the outpatient unit until the inspection is completed. The data is used as the code of the patient, the patient's sex, age of the patient, the name used by the patient's insurance, outpatient clinic name, date of admission to the street, as well as diagnosis.

B. Selection of Grain

Grain is a process to determine what is described by the records in the fact table. Grain contained in the design of data warehouse Sanglah Hospital, among others:
1. Inpatient
Election grain was conducted on the number of patients based on the patient's age, the number of patients by sex of the patient, the number of inpatients is based on the age of each gender of the patient, the number of patients by name room used, the number of patients by insurance used from each room.

2. Outpatient
Selection of grain was conducted on the number of outpatients by sex of the patient, the number of patients based on the patient's age, the number of outpatients is based on the age of each gender of the patient, the number of patients is based on the name of the diagnosis of the patient, the number of patients by insurance used by patients , the number of patients by insurance used from each clinic.

C. Identification and Adjustment Dimension

At this stage, identification and adjustment of dimensions related to the fact that is displayed in the form of a matrix. Here are the dimensions selected for each fact table maing:
1. Inpatient
Dimensions used in the analysis of hospitalization is the dimension of time, patients, insurers, diagnosis and rooms.

TABLE I
GRAIN OF DIMENSION TABLE FACT INPATIENT

Dimensions / Grain	Time	Patient	Rooms	Insurance
Number of patients by sex of patients	✓	✓		
Number of patients according to age category of patients	✓	✓		
Number of patients based on gender or age category of patients	✓	✓		
Number of	✓		✓	

patients by name room				
Number of patients based on class room	✓		✓	
Number of patients based on insurance name	✓			✓
Number of patients by insurance or room name	✓		✓	✓
Total Inpatient	✓	✓	✓	✓

2. Outpatient

Dimension used in the analysis is the dimension of time outpatient, patient, insurance, and clinical diagnosis.

TABLE II
Grain OF DIMENSION TABLE FACT OUTPATIENT

Dimensions / Grain	Time	Patient	Diagnosis	Clinic	Insurance
Number of patients by sex of patients	✓	✓			
Number of patients according to age category of patients	✓	✓			
Number of patients by sex of each age category of patients	✓	✓			
Number of patients based on the name of diagnosis	✓		✓		
Number of patients based on the name of the clinic	✓			✓	
Total patient by the name of the insurance	✓				✓
Number of patients by insurance name of each clinic	✓			✓	✓
Total outpatient	✓	✓	✓	✓	✓

D. Selction Facts

At this stage, the selection of facts that will be used in the table inpatient and outpatient fact tables. The relationship between the fact table and dimension tables can be seen in the fact constellation schema in Figure 5.
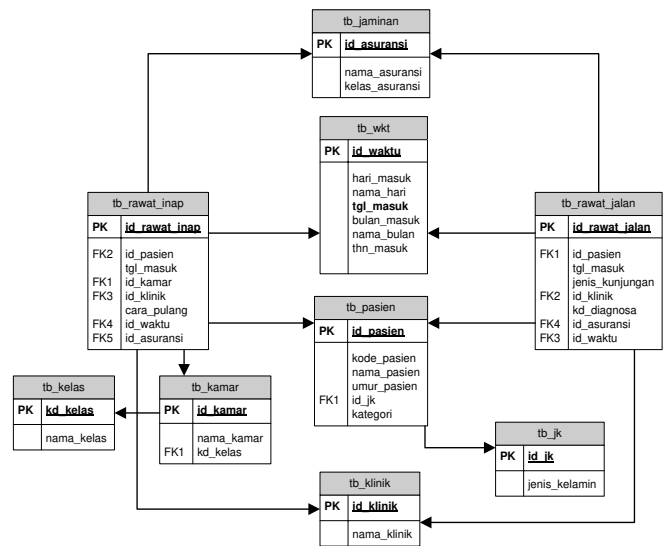

Fig 5. Fact Constellation Schema

E. Storing pre-calculations in the fact table

In fact there is a data table which is an initial calculation. The results of this initial calculation is stored in a fact table.

1. Fact inpatient

Calculatehospitalization fact is the number of inpatients is the number of patients for each patient care processes in inpatient units Sanglah Hospital are worth 1 (one) for each record in the table the fact of hospitalization.

2. Facts outpatient

Hospitalization Calculate fact is the number of outpatients who represent the number of patients for each patient examination process at the outpatient unit Sanglah Hospital are worth 1 (one) for each record there is a fact table outpatient.

F. Determination Table Dimensions

Dimension table is a table that contains data that shows an overview of the various perspectives. Their dimensions allows the creation of dynamic reports which can be viewed from various perspectives at the same level of details can be regulated [11].

In this stage, back in the dimension table and add a text description of the dimensions possible. Overview of text should be easy to use and understandable by the user.

G. Election Database Length

Length of data at Sanglah Hospital entered into the data warehouse as follows.

TABLE III
DURATIONDATABASE

Data Warehouse Name	Data base	Database there since the	Data is entered into Warehouse Data	Data Warehouse Data
RSUPSanglahDWH	RSUP	2001	2012-2014	3 Year

H. Tracking Changes of Dimensions Slowly

Attributes of the tables do not always have a fixed value or relative atatis, Attribute value changes can occur in a long time. Therefore it is necessary to do an update if necessary to maintain the consistency and accuracy of the data. Dimensions to be possible to change contained in the table 4.Durasi of data at Sanglah Hospital entered into the data warehouse as follows.

TABLE IV
COLUMN DIMENSION TO CHANGING

Dimension Name	Attribute Might Changed
dim_pasien	Umur, Kategori
dim_diagnosa	Nama_diagnosa
dim_kamar	Nama_kamar, kd_kelas
dim_klinik	Nama_klinik
dim_asuransi	Nama_asuransi

I. Deciding the query priorities and the query modes

In this stage to do is to consider the effect on the design of fiscal, such as the existence of the summary (summaries) and the sum of (aggregate). In the design of the data warehouse, storage capacity is one aspect that needs to be considered.

In this stage to do is to consider the effect on the physical design, such as the existence of the summary (summaries) and the sum of (aggregate).

J. Data Warehouse Analysis Results

From the analysis has been done, much the way the formation of reports, among others by using SQL Server 2008 Reporting Services (SSRS), Microsoft Office Excel and other apliaksi applications that can connect to the database. In making the analysis report on a data warehouse, used Microsoft Office Excel for use easier and quicker process. In Microsoft Office Excel data warehouse analysis results can be made in the form of a table as in the graphic display.

ETL process is done on the design of hospital data warehouse to map any data that would be required by the executive of the hospital as head of the hospital, chief medical and chief nursing aims to give a decision to be taken on the results of the data warehouse. Figure 6 shows the mapping of data warehouse ETL in the design of the hospital.
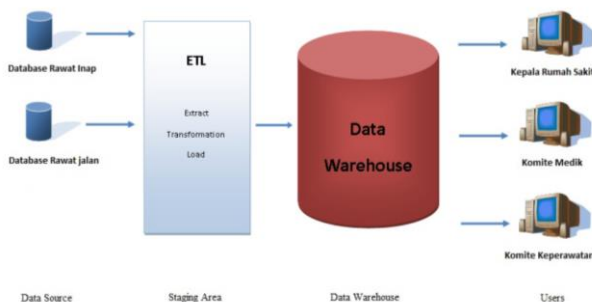
Fig 6. ETL Process

From the data warehouseoutpatient visits to the poly found that children are mostly taken by the patient. While a visit to the poly genital skin is the least performed.
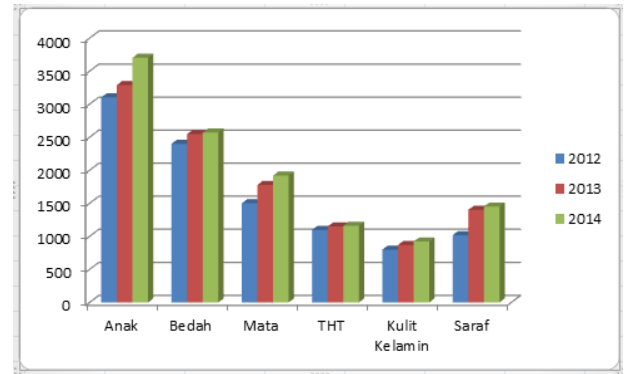

Fig 7. Report Number Outpatient

Visits data for hospitalized patients can be seen in Figure 8, where inpatient obstetrics highest rank, and inpatient cardiac occupy the lowest place.
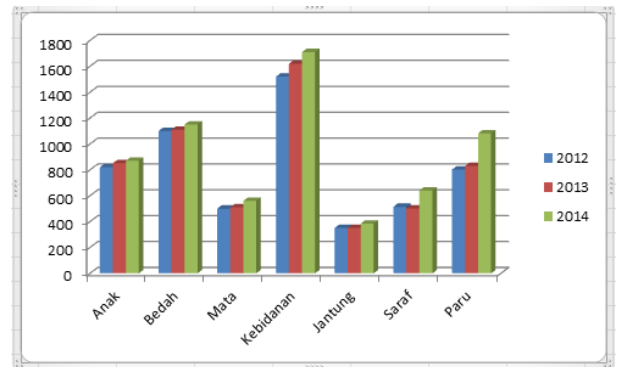

Fig 8. Report Total Inpatient

From the results given by the data warehouse, hospital executives can monitor patient visits and make decisions easier for steps to be done in providing services to patients in accordance with the vision and mission of the hospital.

IV. CONCLUSION AND FUTURE WORKS

Design Sanglah Hospital Data Warehouse quite useful for data processing in large enough quantities, so expect the needs and information about the patient can be met. Data warehouse Sanglah Hospital can be used to analyze patient data in order to get information on the number of patients of various dimensions (time, patients, insurance, diagnosis, rooms and clinics) are useful for executives in taking action in the future to fulfill the vision and mission of the hospital.

REFERENCES

[1]   Rahman, "Perancangan Data Warehouse Pasien pada RS. Muhammadyah Palembang," 11 April, 2013.

[2]   Meta Suzana, Jemakmun, Suyanto, "Analisi dan Perancangan Data Warehouse Rumah Sakit Umum Daerah Palembang Bari," Jurnal Ilmiah Teknik Informatika Ilmu Komputer, November 2013; Vol. 1, No. 14.

[3]   Patel, A. R, "Data Modeling Techniques For Data Warehouse," International Journal of Multidisciplinary Research, February 2012; ISSN 2231 5780, 240-246.

[4]   Akhmad Dahlan, Ema Utami, Emha Taufiq Luthfi, "Perancangan Data Warehouse Perpustakaan Perguruan Tinggi XYZ Menggunakan Metode Snowflake Schema," Jurnal Teknologi Informasi, 24 November 2013; Vol . 7, No. 24 : ISSN : 1907-2430.

[5]   Laudon, Kenneth C, Jane P, Sistem Informasi Manajemen, Mengelola Perusahaan Digital Edisi 10, Salemba Empat, Jakarta Selatan, 2007.

[6]   Heribertus Himawan, "Implementasi Sistem Business Intelligence Untuk Melakukan Analisis Data Guna Mendukung Pembuatan Keputusan Manajer," Techno.COM, Mei 2008; Vol. 7, No. 1.

[7]   Ponniah, Paulraj, Data Warehouseing, John Wiley & Sons Inc, Canada, 2011.

[8]   Rainardi, V., Building a Data Warehouse with Examples in SQL Server. Apress. New York, 2008.

[9]   Conolly, Thomas M., Carolyn E. Begg, Database System : A Practical Approach to design, Implementation, And Management 4th edition, Addison Whesley, Longman Inc. USA, 2005.

[10]   Munawar, "Perancangan Data Warehouse untuk Penerimaan Mahasiswa Baru," Jurnal Ilmu Komputer, September 2013; Vol . 9, No. 2.

[11]   Mukhlis Febriady, Bayu Adhi Tama, "Rancang Bangun Data Warehouse Untuk Menunjang Evaluasi Akademik Di Fakultas," Prosiding Konferensi Nasional Teknologi Informasi dan Aplikasinya (KNTIA) 2011 Palembang, 22 Oktober 2011.

[12]   Randy Oktrima Putra, "Rancang Bangun Data Warehouse untuk Analisis Kinerja Penjualan pada Industri dengan Model SPA-DW (Sales Performance Analysis - Data Warehouse)," Tesis, Semarang, 2012.

[13]   Adi Supriyatna, Mochamad Wahyudi, "Perancangan Data Warehouse pada Perpustakaan Bina Sarana Informatika," Seminar Nasional Aplikasi Teknologi Informasi 2012, Yogyakarta, 15-16 Juni 2012; ISSN: 1907-5022.

[14]   Armadyah Amborowati, "Perancangan dan Pembuatan Data Warehouse pada Perpustakaan STMIK AMIKOM Yogyakarta," Seminar Nasional Aplikasi Sains dan Teknologi 2008 – IST AKPRIND Yogyakarta, 2008.

[15]   Ballou D.P., Tayi G.K, "Enhancing Data Quality in Data Warehouse Environments," Communications of the ACM, January 1999; Vol . 42, No. 1 : pp 73-78.