Data warehouse Implementation on Denpasar City Online Community Complaints System

I Putu Ari Putra Wijaya[1], Wahyudin[2], and Made Mataram[3]

[1][2] Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3] Department of Electrical and Computer Engineering, Udayana University
E-Mail: ariputra@student.unud.ac.id

Abstract— **Department of Communications and Informatics, appointed as Executor of Handling Complaint Society in Denpasar City. Over time and more complex data complaints that come in so complicate the data analysis, to overcome the data it is necessary implementation of data warehouse. The method used is Snowflake Schema, this method is chosen because this method is the development of star schema where each dimension table can have sub-table dimension. It aims to minimize data overload. From the results of the implementation in get the conclusion that by implementing data warehouse can facilitate the user in view report in accordance with the desired from summary to detail..**

Keywords—Data Warehouse, Snowflake Schema, ETL.

I. INTRODUCTION

Department of Communications and Informatics, Appointed as the Implementer of Denpasar City Online Community Complaints System according to the Decree of the Mayor of Denpasar Number 188.45 / 206 / HK / 2013 dated March 7, 2013, with the duties and responsibilities of coordinating, providing facilities and infrastructure, receiving, delivering, monitoring and making reports of public complaints. Over time, more and more complex data complaints come in making it difficult to analyze the data, whereas the report of this complaint data will be a matter of consideration for decision making as well as an indicator of the success of the city government program denpasar. Solution to the above problem is to implement Data Warehouse.

The most important thing in the process of building a data warehouse is the modeling process [1]. Because designed data model will be measured the extent to which the built-in data warehouse is able to produce the necessary managerial information.

One method that can be used in the development of Data Warehouse is the Snowflake method. This method is a development of star schema, that is normalization method for dimension table that exist in star schema. In a star schema, a dimension table does not have sub-tables (sub-dimensions) so that data that can be used as material to produce information is not much. n snowflake schema, each dimension table can have a sub-table of dimensions. It aims to minimize redundancy data (redundancy data) [2]. Dimensions of this data is the subject of information to be a material in decision-making, because on each dimension of data is possible to do more detailed breakdown. Thus the data source that can be processed to become information can be more and more detail.

II. LITERATURE REVIEW

A. Database

The database is a data set that describes an activity of one or more related entities [3]. Database is a structure that is generally divided into 2 things, namely a flat database and and a relational database. Relational databases are better understood than flat databases because relational databases have simple database forms and easy to perform data operations [4]

B. Data Warehouse

Definition of Data Warehouse according to William H Inmon is a collection of subject-oriented data, integrated (integrated), non-changeable, has a time period (time variant), for information that is historical and can support decision making [5]. Meanwhile, according to Sean Kelly Data Warehouse is a collection of data that has several characteristics, namely separate, available, integrated, time-stamped, subject-oriented, non-volatile, and accessible [6].

According to Vincent Reinardi, Data warehouse is a system that takes and consolidates the data periodically from the source system into a normalized table of dimensions or data storage [2]. The data contained in it is usually updated within a certain period of time, not every time a transaction occurs within the source system.

Data warehouses are built to address technical and business issues related to data and information usage for decision making [7]. Data warehouse is not a product but an environment where users can find strategic information. The data warehouse is a collection of logical data separate from the operational database and is a summary.

Data Warehouse Architecture consists of a) operational Data, is a data source for data warehouse; B) operational Data Store, is a container of operational data that has been integrated, used for analysis; C) load Manager, perform all operations related to the extraction and loading of data into the data warehouse; D) Warehouse Manager, performs operations related to data management in the data warehouse; E) Query Manager, performs operations related to the query management

of the user; F) Detailed Data, keeps all detailed data in the database schema. In general, detailed data is available by aggregating; G) Lightly and Highly Summarized Data, storing lightly and highly summarized (aggregated) data generated by Warehouse Manager; H) Archive / Backup Data, save detailed data and concise data with backup and archiving purposes; I) Metadata, often called Metadata simplify the end user in doing the analysis and save time; And j) End-User Access Tools, a tool that leverages the usability of the data warehouse, such as in report generation, OLAP, data mining, executive information [8].

Adapun karakteristik dari *data warehouse* adalah sebagai berikut [9].

1) Subject-oriented.

Data warehouse is a storage place based on the subject rather than the application. The subject is part of a company. Examples of subjects in manufacturing companies are sales, consumers, inventory, and so on. The picture below is the difference between the data warehouse and the operational database.

2) Integrated data

The data sources that exist in the data warehouse not only comes from the operational database (internal source) but also comes from data outside the system (external source). Data on different sources can be encoded in different ways. For example, gender data can be encoded as 0 and 1 in one place and "m" and "f" elsewhere.

3) Nonvolatile

The data in the operational database will be periodically or periodically transferred into the data warehouse according to the predefined schedule n. For example per day, per week, per month, and so forth. Once entered into the data warehouse, the data is read-only.

C. Extract, Transform, Load (ETL)

The three main functions that need to be done to make the data ready for use in the data warehouse are extraction, transformation and loading. These three functions are found in the staging area [10]. In this staging data, provided places and areas with some functions such as data cleansing, change, convert, and prepare data to be stored and will be used in the data warehouse. In staging data is provided place and area with some functions such as data cleansing, change, convert, and prepare data to be stored and will be used in data warehouse In this staging data, provided places and areas with some functions such as data cleansing, change, convert, and prepare data to be stored and will be used in the data warehouse [10]. The ETL process is a process for converting, reformatting and integrating data from one or more OLTP systems [5].

D. Star Schema

Star Scheme is a specific type of database design that is used to support the analytical process as well as having a specific set of normalization tables. The star scheme has two kinds of tables: fact table and dimension table, fact table is the main table that contains actual data to be analyzed is dimension table is a

small table containing data which is further description of the data contained in the fact [11].
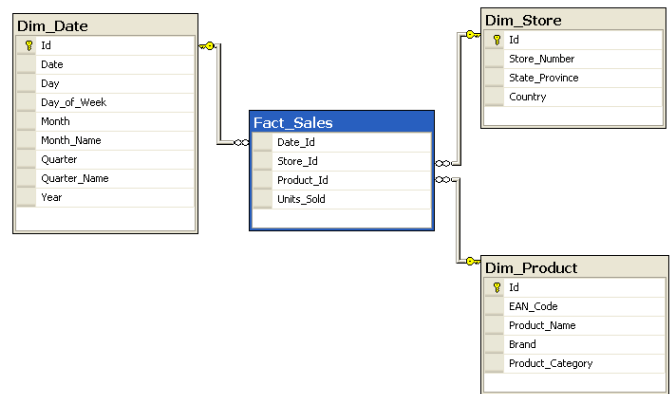

Fig. 1 Star Schema

E. Snowflake schema

According to Inmon snowflake schema is a combination of several modeled schema star schema data. In snowflake schema, each dimension table can have a sub-table of dimensions again [5]. It aims to minimize data redundancy (redudancy data). A scheme is called a snowflake scheme if one or more dimension tables are not directly related to the fact table but must be related through other dimension tables [12]. Dimension of this data is the subject of information to be the material in decision-making, because in each dimension of data is possible to do more detailed breakdown. Thus the data source that can be processed to become information can be more and more detail

F. MySQL

MySQL is a Relational Database Management System (RDBMS) that is distributed for free under the GPL (General Public License) license. Where everyone is free to use MySQL, but should not be a commercial derivative product. MySQL is actually a derivative of one of the main concepts in the database since a long time, namely SQL (Structured Query Language) [13]. MySQL

Mysql Derived from Scandinavia, MySQL includes SQL server, client program to access servers, useful things in administration, and a "programming interface" to write their own programs [14].

III. METHODOLOGY

A. Interview

An interview is a straightforward and systematic question-and-answer process to people who know about the issues being observed [15]. In this study the interview was done by directly coming to the Office of Communications and Informatics City of Denpasar and directly interviewed the relevant parties namely the Head of Analysis & Design Software Diskominfo Denpasar.

B. Literature Review

Literature study is the process of collecting data and information by reading and understanding various kinds of existing literature related to research conducted. The literature study conducted in this study comes from books, journals, and other sources in the form of print media or electronic media. This step is used to get reference data and literature about designing Data Warehouse.

IV. RESULT ANALYSIS

A. Data Source

The data source to be processed to build the datawarehouse is obtained from the complaints database based on a certain period. The data taken are complaint data from 2014 until 2016. Here is a conceptual database that becomes the source of data.
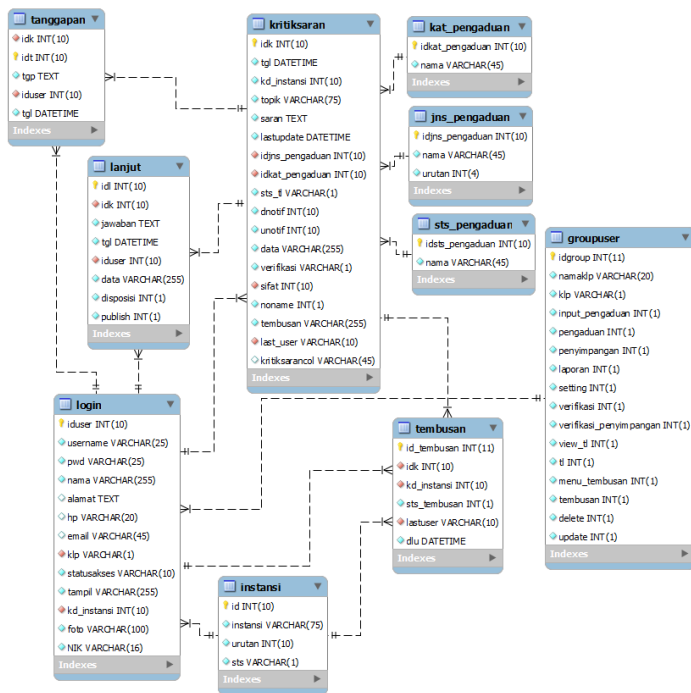

Fig. 2 Tabel Relationship Database

B. Designing Snowflake Schema

In this stage, the activity is to create a dimensional data model in the form of snowflake schema and proceed to make the design of ETL process. The main subject in this Data Warehouse complaint is the public complaint data which will become the center of the scheme. In this research is implemented in the form of a fact table named *fakta_aduan*. Whereas other data are implemented into dimension tables or sub-dimensions that are all related to the fact table. In this snowflake there are several dimension tables that is:

- *tbl_dim_instansi* which holds the destination intentions information
- *tbl_dim_type* that stores the type of complaint information
- *tbl_dim_asal* that stores the origin complain information
- *tbl_dim_sub_kat* which stores the complaint sub-category information
- *tbl_dim_kat* which stores the category information of the complaint.

Snowflake Schema from the above design can be seen in figure 3.

Fig. 3 Snowflake Schema

C. Results of Data Warehouse Implementation

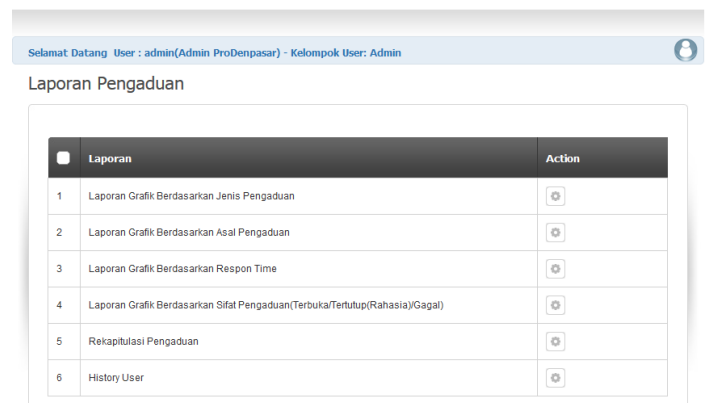Reports page dashboard of this system, can be seen in figure 4 below.

Fig. 4 Report Dasboard

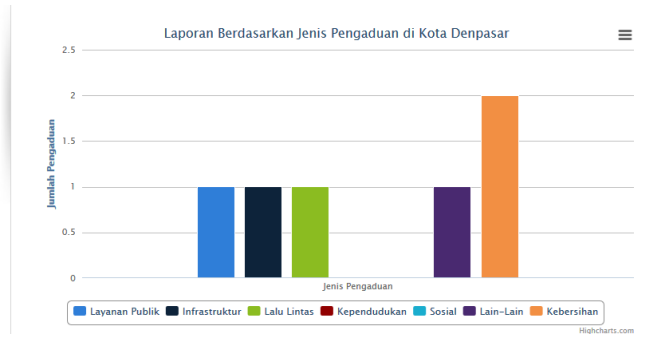Report pages layout in the system can be seen in Figure 5 below

Fig. 5 Tanpilan Halaman Laporan

D. Calculation Report Result

Several reports that can be generated from the implementation of data warehouses in the Denpasar city complaint system include, the number of complaints based on categories, the number of complaints based on the origin of the complaint, the number of complaints based on the nature and the number of complaints that have been followed up.
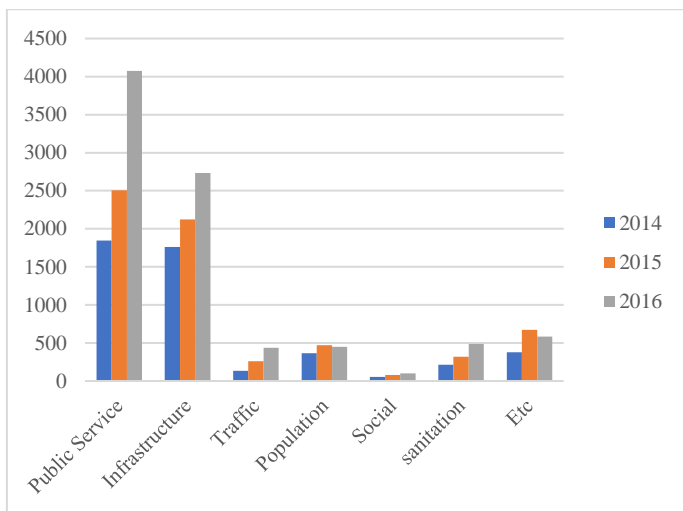

Fig. 6 Graph Number of Complaints by Category

The number of complaints based on the Platform or the origin of the complaint can be seen in Figure 7 below
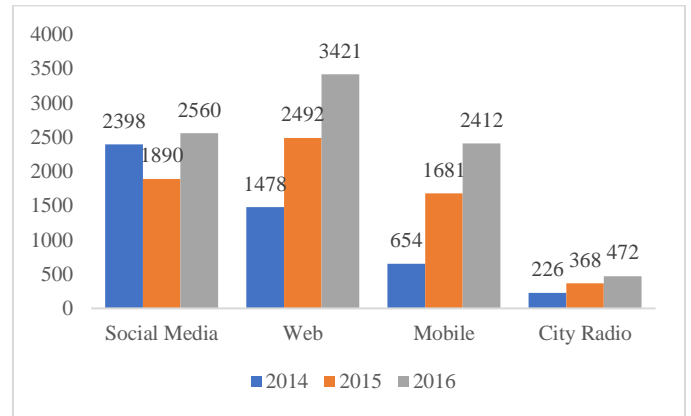

Fig. 7 Graph Number of Complaints Under the Platform

Report on the number of complaints based on the property of the complaint can be seen in Figure 8 below.
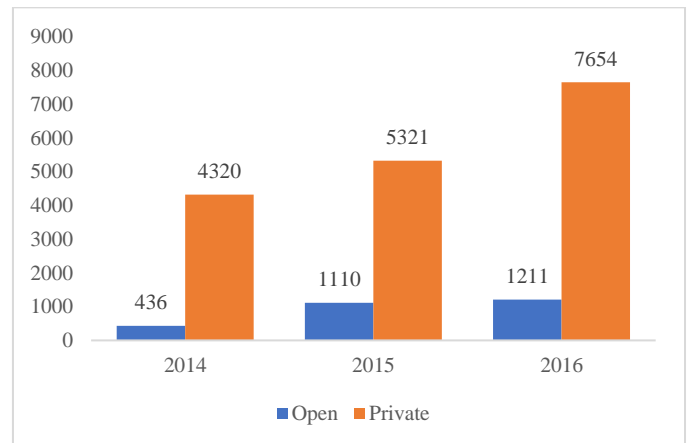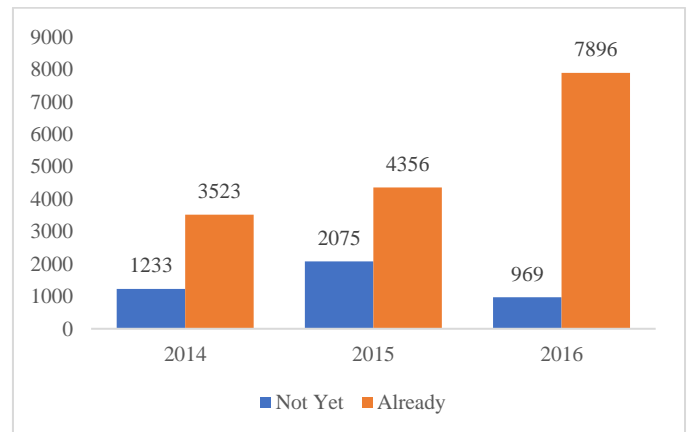

Fig. 8 Graph Number of Complaints Based on Property

Reports on the number of complaints that have been followed up can be seen in Figure 9 below

V. CONCLUSION

After identifying the problem, the design and implementation of the system obtained some conclusions as follows

1. Data Warehouse has been implemented into Denpasar online public complaints system using Snowflake Schemma method
2. Datawarehouse complaints can be used to analyze complaint data so that relevant department leaders can view information on the number of complaints from various dimensions of their daintara, category, nature, platform and follow-up
3. The information displayed is summary to detail that allows users to obtain information that is summary in accordance with the needs

VI. REFERENCES

[1] A. R. PATEL, "Data Modeling Techniques For Data Warehouse," *International Journal of Multidisciplinary Research,* vol. 2, no. 2, pp. 240-246, 2012.

[2] V. Reinardi, Building a Data Warehouse with Examples in SQL Server, New York: Apress, 2008.

[3] J. I. Maanari, "Perancangan Basis Data Perusahaan Distribusi dengan Menggunakan Oracle," *Teknik Elektro dan Komputer,* pp. 1-11, 2013.

[4] W. Komputer, Belajar MySQL Database Server, MediaKita, 2010.

[5] W. H. Inmon, Building the Data Warehouse - Fourth Edition, New York.: J.Wiley, 2005.

[6] S. Kelly, Data Warehousing: The Route to Mass Communication, New York: J.Wiley, 1996.

[7] M. K. Jiawei Han, Data Mining: Concepts and Techniques, San Fransisco: Morgan Kaufmann, 2006.

[8] T. Oktavia, "Perancangan Model Data Warehouse Dalam Mendukung Perusahaan Jasa Pengiriman," in *Seminar Nasional Informatika 2011*, Yogyakarta, 2011.

[9] A. A. Heni Jusuf, "Perancangan Data Warehouse Pada Perpustakaan Universitas NasionaL," *Jurnal Basis Data,* vol. 3, no. 1, pp. 2-8, 2008.

[10] D. Wilanda, "Pembangunan Data Warehouse Pada Pt. Pupuk Iskandar Muda," *KOMPUTA,* vol. 1, no. 2, pp. 1-8, 2016.

[11] K. A. S. Parsiyono, "Perancangan Data Warehouse Akademik Di Sekolah Tinggi Agama Buddha Semarang," *INFORMA,* vol. 1, no. 2, pp. 44-51, 2015.

[12] A. M. B. A. H. DIAN DHARMAYANTI, "Pemodelan Data Warehouse Pada Jurusan Teknik Informatika Unikom," *Majalah Ilmiah UNIKOM,* vol. 2, no. 12, pp. 151-168, 2014.

[13] U. A. S. T. M. Syaifudin Ramadhani, "Rancang Bangun Sistem Informasi Geografis Layanan Kesehatan Di Kecamatan Lamongan Dengan PHP MySQL," *Jurnal Teknika,* vol. 5, no. 2, pp. 479-483, 2013.

[14] D. H. Setiabudi, "Aplikasi E-Commerce Www.Komputeronline.Com Dengan Menggunakan Mysql Dan Php4," *JURNAL INFORMATIKA,* vol. 3, no. 2, pp. 88-95, 2002.

[15] Sunyono, "Teknik Wawancara (Interview) Dalam Penelitian Kualitatif," 2011.