Designing Data Warehouse for Analysis of Culinary Sales with Multidimensional Modeling – Star Schema Design

(Case Study: XYZ Restaurant)

Aggry Saputra[1*], Zulfachmi[2], and Made Sudarma[3]
[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
*Email: aggrysaputra@gmail.com

***Abstract*—Fast and accurate information is one of the factors that make a company can be more superior than the other. In order to meet the information needs of various data, it takes a data warehouse where data from data warehouse can be more useful information so that can be used in support of a decision quickly. Star Schema is one dimensional model of data warehouse that has fact tables in the middle and is surrounded by several dimension tables.**

Keyword— Culinary, Data Warehouse, MDM, Sales ,Star Schema.

I. INTRODUCTION

Utilization of technology in the business world is a necessity, because with the technology of an information can be retrieved in realtime. Fast and accurate information is one of the factors that make a company can be more superior than the other[1]. With the support of information quickly then a decision can also be quickly obtained which it is one important factor in the business world..

To meet the appropriate data needs of various data that is processed in the database, it takes data warehouse. Inmon (2002, p31), data warehouse is subject-oriented data collection, integrated, time-varying, and non-volatile to support management decision-making process. Data warehouses have the ability to perform complex queries, complete data models and emphasize support for data analysis.

Often an organization has difficulty in processing existing data, especially sales data that will be processed to become an information that can be used to support in decision making.

From the description above, it can be concluded the importance of a fast and accurate information processing. therefore, the authors are interested to design a system that can process sales data in the field of culinary to produce information relavan making it easier for the management in making decisions.

II. LITERATUR REVIEW

Data warehouse is not a tool, software or hardware, but a smart technology in processing a database that the results can be used quickly to support decision-making system[2]. As for some research that has been done in the field of data warehouse, such as data warehouse design for centralized medical record data by applying star scheme, data warehouse design is expected to be a solution in processing data that has been submitted in a systematic so that it can be used to simplify the process of analysis and reporting[3]. Implementation of data warehouse with OLAP Umbrella visualization model where this method has visualization and navigation support for several facts and various sizes related to different facts[4]. Designing a data warehouse using the nine-step method of ralph kimball, to analyze sales data and distributor status[5].

III. RESEARCH METHODOLOGY

A. Data Source

Data source on data warehouse design is culinary sales data which can be from XYZ Restaurant. The data obtained are visitor data, type of food and beverage sold, stock list data, and supplier data.

Data is taken from several parts of the restaurant, then processed in one data warehouse. The constraints that may occur are not synchronous data items, stocks that come out or enter. Therefore, before doing research the need to match the data taken to get maximum results.

B. Prototype Method

The prototype model is a technique for collecting certain information about user information needs quickly.

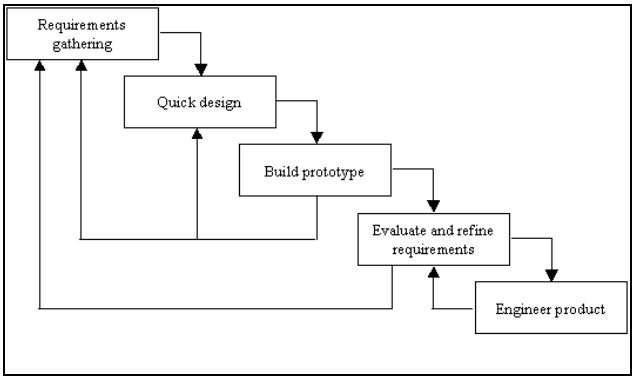

Fig. 1. Prototyping Method

From Figure 1, it appears that the prototyping paradigm begins with communication. Developers and users meet and define the overall goals of the software to be built, identifying what needs are desired. The prototype is submitted and evaluated by the user. User feedback is used to improve the software requirements criteria. This is done over and over where the prototype is customized to meet user needs, while at the same time the developer has a better understanding of what the user wants to be fulfilled.

C. Data Warehouse and Multidimensional Modeling (MDM)

The data warehouse is a place to archive summaries of historical data contained from existing databases within an organization[6].

Multidimensional Modeling (MDM) mdm is one approach of data warehouse built on OLAP (On-line Analytical Processing) method which has dimension, hierarchy, level and member concept[7]. Aggregate of the application of the concept can be displayed with star schema design. Compared with the concept of relational databases, the concept of multidimensional databases can perform analysis from various dimensional angles. The multidimensional data model is designed to facilitate analysis and not transactions. This model is commonly used in data warehouses because it has an intuitive concept from many dimensions or a business measurement perspective or facts. For example, to see sales from a customer perspective, product and time.

In modeling a data warehouse there are various schema, one of them is Star Schema. According to Connolly and Begg (2010:1227), star schema is a dimensional data model that has a fact table in the middle, surrounded by a dimension table consisting of reference data (which can be denormalized). There are several advantages and disadvantages of star schema, which is for system performance, star schema is more reliable than other schemes, this is caused by the use of join level between dimension tables and fact tables are little so as to facilitate the system to perform an aggregation operation, while the problem that arises is the emergence of redundancy in the dimension table.

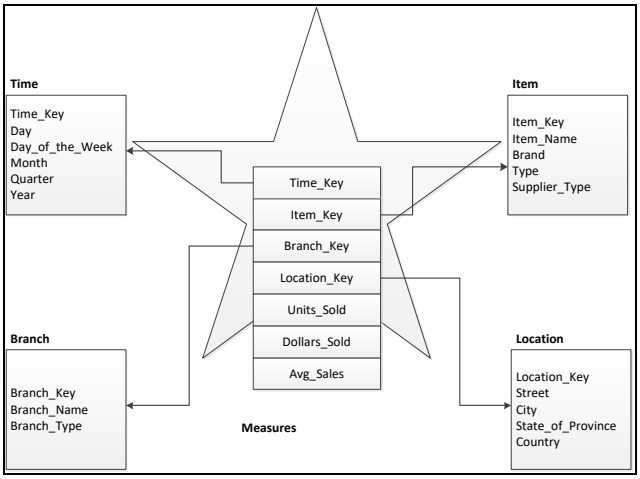

Fig. 2. Star Schema Design

The dimension table has the following characteristics:

- Key dimension table, is the primary key of the dimension table that identifies each row in the table uniquely.
- Is a wide table. The dimension table has a number of columns or many attributes, therefore the dimension table is wide.
- Attribute in the form of text. In the dimension table, numeric values are rarely found for calculations; the general attribute is text that represents the textual description of the components in the business dimension.
- Attributes are not directly related.
- Not normalized. For effective query performance, it is best if the query retrieves from the dimension table and directly to the fact table without going through the intermediate table to be formed if the dimension table is normalized.
- Drill-down and roll-up capabilities. The attributes in the dimension table provide the ability to get the detail of a high level of aggregation to a low level of detail.
- There are several hierarchies. Various parts of the company can classify dimensions in different ways, thus forming more than one hierarchy.
- The number of records is less. Tables of dimensions generally have fewer records or rows than fact tables.

As for the characteristics of fact tables, ie:

- Concatenated key. The rows in the fact table are identified by using the primary key of the dimension tables, then the primary key of the fact table is the combined primary key of all dimension tables..
- Grain data, is the level of detail for measurement. For example, the number of bookings relates to the number of specific products on an order, a specific date, to a specific customer and is obtained by a specific sales representative. If the number of orders is seen as the amount for a product per month, then the data grain is different and at a higher level.

- Fully additive measures. Aggregation of fully additive measures is accomplished by a simple summation of the attribute values.
- Semiadditive measures. Semiadditive measures are values that can not be directly summed, for example the percentage of profit.
- The tables are large, not wide. The fact tables generally have fewer attributes than the dimension table, but have a larger number of records.
- Sparse data. Tabel fakta tidak perlu menyimpan record yang nilainya null. Maka tabel fakta dapat memiliki gap.
- Degenerate dimensions. Terdapat elemen-elemen data dari sistem operasional yang bukan merupakan fakta ataupun dimensi, seperti nomor pesanan, nomor tagihan, dan lain-lain. Namun atribut-atribut tersebut dapat berguna dalam jenis analisis tertentu. Sebagai contoh, mencari rata-rata jumlah produk per pesanan, maka produk harus dihubungkan ke nomor pesanan untuk mendapatkan nilai rata-rata. Atribut-atribut tersebut disebut degenerate dimension dan disimpan sebagai atribut dari tabel fakta

IV. Result and Discussion

A. System Design

The design of data warehouse system involves two Actors, namely the actor Owner and Admin, where each actor has different authority as shown in the picture below:

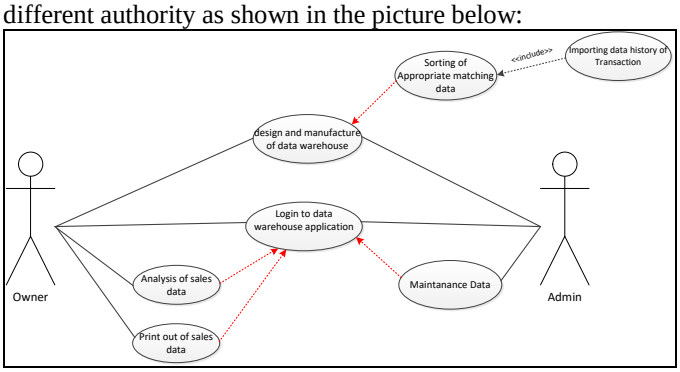

Fig. 3. Design of Data Warehouse System

Actor Owner is a priority in this culinary data warehouse system, where the role of actor owner is described in use case analysis of sales data, print out of sales data, and login to data warehouse application. The initial step of designing this system is the Owner actor asked for information about system requirements, then the actor can access the application that has been built. While the Admin actor has several use cases, such as use case design and manufacture of data warehouse, sorting of appropriate matching data, importing history of transaction data, login to data warehouse application, and maintanance data. Actor admin has tasks such as, update transaction data, transaction data sorting, design and implementation of data warehouses, login applications to perform data warehouse maintenance.

B. Data Warehouse Design

In the design of data warehouse there are several processes, such as:

- Choose the Process
  The process of XYZ Culinary Restaurant used to design data warehouses is the proeses of sales. Sales process in question is the sale of food / beverages. As for supporting documents such as sales invoices, goods purchase invoices, incoming and outgoing goods. Existing data include customer data, distributors, goods, item data and sales transactions.
- Choose Grain
  Grain on XYZ Culinary Restaurant used to design data warehouses is a sale. Where the analysis done on the sale is a lot of items sold or much in demand, the amount of sales, total sales. The analysis will be performed per time period (Day, Month, Year).
- Define Dimension Tables
  The data warehouse of culinary sales is designed to consist of 6 dimensional tables and 1 fact table connecting dimension tables.

TABLE I. Strukture of Dimension Table t_distribution

Column_Name	Data_Type	Length
Kd_Distributor	Varchar	10
Nama_Distributor	Varchar	30
Kategori	Varchar	15
Alamat	Varchar	30
Telp	Varchar	14

TABLE II. Structure of Dimension Table t_stock_list

Column_Name	Data_Type	Length
Kd_Barang	Varchar	10
Nama_Barang	Varchar	30
Jml_Masuk	Int	-
Jml_Keluar	Int	-

TABLE III. Structure of Dimension Table t_lokasi

Column_Name	Data_Type	Length
Kd_Lokasi	Varchar	10
Kelurahan	Varchar	30
Kecamatan	Varchar	30
Kota	Varchar	30

TABLE IV. Structure of Dimension Table t_customer

Column_Name	Data_Type	Length
Kd_Customer	Varchar	10
Nama_Customer	Varchar	30
Alamat	Varchar	30
Telp	Varchar	14

TABLE V. STRUCTURE OF DIMENSION TABLE T_ITEM

Column_Name	Data_Type	Length
Kd_Item	Varchar	10
Nama_Item	Varchar	30
Kategori	Varchar	10
Jumlah	Int	-

TABLE VI. STRUCTURE OF DIMENSION TABLE T_WAKTU

Column_Name	Data_Type	Length
Tgl	Date	-
Bulan	Varchar	15
Tahun	Varchar	5

- Specify a Face Table
  The fact tables are identified using the primary key of the entire dimension table, so that the fact table can be designed as follows:

TABLE VII. STRUCTURE OF FACT TABLE T_FACT_SALES

Column_Name	Data_Type	Length
No_Fak	Varchar	10
Tgl	Date	-
Kd_Customer	Varchar	10
Kd_Item	Varchar	10
Kd_Distributor	Varchar	10
Kd_Barang	Varchar	10
Kd_Lokasi	Varchar	10
Jml_Penjualan	BigInt	-
Tot_Penjualan	BigInt	-

There are two types of column types, namely foreign keys for dimension tables and primary keys as numeric fact tables[8]. Based on the analysis of culinary sales data at xyz restaurant, it can generate multidimensional data warehouse with star schema model as shown in figure 4.

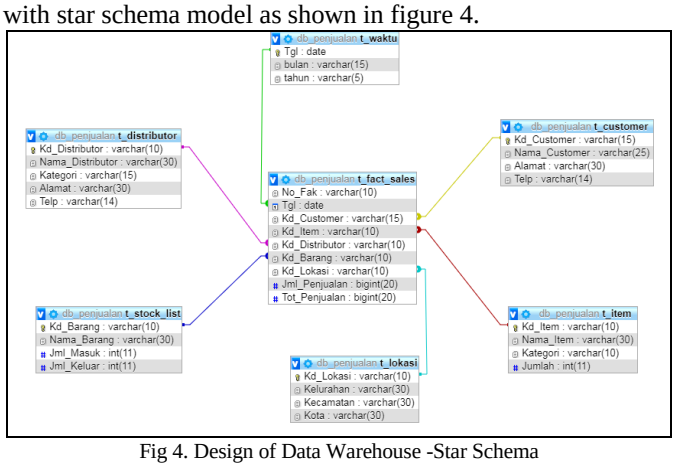

Fig 4. Design of Data Warehouse -Star Schema

V. CONCLUSSION

From the analysis and design data warehouse culinary sales can be concluded several things: (1) the information obtained more useful and effective because it has been done filtering from some previous dimensions. (2) with the construction of data warehouse reports generated more quickly and precisely. (3) Management is easier to see sales statistics, data on food and beverages that are in great demand.

REFERENCES

[1] S. Darudiato, "Perancangan Data Warehouse Penjualan Untuk Mendukung Kebutuhan Informasi Eksekutif Cemerlang Skin Care," in *semnasIF*, 2010, no. semnasIF, pp. 350–359.

[2] M. Subekti, H. Leslie, H. Spits, and Y. Heryadi, "The 3 steps of best Data Warehouse model design with Leaning Implementation For Sales Transaction in Franchise Restaurant," in *IEEE Conference on Cybernatics and Computational Intelligence*, 2017, pp. 170–174.

[3] K. Ary, B. Permana, G. B. Subiksa, and M. Sudarma, "Design Data Warehouse For Centralized Medical Record," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 2, pp. 47–51, 2017.

[4] T. Bhowmik, A. Sarkar, and N. C. Debnath, "OLAP umbrella: Visualization model for multidimensional databases," in *2010 ACS/IEEE International Conference on Computer Systems and Applications, AICCSA 2010*, 2010.

[5] G. Prajena, "Analisis Dan Perancangan Data Warehouse Untuk Penjualan , Identifikasi Status Distributor , Dan Perpanjangan Masa Aktif Distributor Pada Pt Harmoni Dinamik Indonesia," *ConTech*, vol. 4, pp. 475–483, 2013.

[6] Kusnawi, "Multidimensional Data Warehouse dengan menggunakan MySQL Multidimensional Database (Cube)," *J. Dasi*, vol. 11, no. 3, pp. 59–72, 2010.

[7] R. A. Nugroho, J. Tambotoh, and T. J. Hoetama, "Aplikasi Data Warehouse untuk Analisis Penjualan Mobil Berbasis Multidimensional Modeling (MDM) dan Star Schema Design (Studi Kasus PT. Asco Automotive)," *J. Teknol. Informasi-Aiti*, vol. 5, no. 2, pp. 101–200, 2008.

[8] M. S. Raza and A. K. Nayak, "A study on designing a Layered Star Schema for data mining optimization," in *Proceedings of the 2014 Conference on IT in Business, Industry and Government: An International Conference by CSI on Big Data, CSIBIG 2014*, 2014.