Decision support system using data warehouse for retail system

Abba Suganda Girsang¹²,³, Geri Arisandi¹, Calista Elysis¹, Michelle¹, Melva Hermayanty Saragih²⁴

¹Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia
²Management Department, BINUS Business School Undergraduate Program, Bina Nusantara University, Jakarta, Indonesia

E-mail: ³agirsang@binus.edu, ⁴melva.saragih@binus.ac.id

Abstract. CMK company is a retail company that sell and distributes miniature vehicle items to small stores. CMK has several branches in various countries that use for operational and storage warehouse. This company faces the problem for analysing sales to decide the useful strategy since this company only uses Online Transaction Processing (OLTP) database. This study implements Online Analysis Processing (OLAP) database and data warehouse to provide useful information for this company by using the nine-step method designed by Kimball & Ross. The data will be performed into dashboard to make easier for CMK Company to analyze data. Furthermore, many useful information can be provided in short and efficient time.

1. Introduction
CMK Company is a retail company that sells miniature or figure vehicles such as cars, buses, planes and others. CMK Company buys items from vendors and then distributes the items to smaller stores. The offices are spread across several branches in various countries as operational and storage warehouses. Nowadays, CMK Company is having a financial crisis due to inefficient management. To overcome this problem, the management of this company must be improved. Getting items or vendors that provide the highest profit and in terms of marketing, providing facilities for customers who have the highest value will help company to resolve this problem. To achieve that, it is necessary to know the products and top vendors with the most sales to give us information which products need to be stocked more, and from which vendors the products are most in demand by customers, and also information about top 10 customers can also assist in giving rewards for that top customers to attract customers.

In this case, CMK Company still uses an Online Transaction Processing (OLTP) database. This makes it difficult for CMK Company to analyze sales and make decisions in the future to solve the existing problems. Therefore, CMK Company attempt to implement a data warehouse that uses the Kimball methodology [1]. Data warehouse is a database that use Online Analysis Processing (OLAP) database that is created by integrating data from several heterogeneous sources, usually from Online Transaction Processing (OLTP) database [2][3].

2. Related Work
Data warehouse is a relational database designed for query processes and analysis for transactional process [4]. Data warehouse is created by integrating data from several heterogeneous sources which are usually database Online Transaction Processing (OLTP). Data Warehouse is a system that involves
cleaning, data integration, and consolidation data [1] [2]. Data warehouse helps organizations to analyse trends based on data repositories over a period a time. The main function from Data Warehouse is to facilitate the organization in carrying out strategic planning based on long term data stored and making good and fast decision. As the result, it easier for user to retrieve the information needed [5]. Data warehouse that hold large amounts of data are stored on mainframe server or cloud company [6]. Data warehouse using Online Analysis Processing (OLAP) which is a functional analysis techniques such as summarization, consolidation, and aggregation, who is able to see the information from a different perspective [3].

OLAP system allows user to query with large payload and automatically to collect the data from data warehouse that focuses on analytically oriented queries, specially for analysing data to decision making support [7]. OLAP consist of Cube to explore the information in the data warehouse. Cube on OLAP is a multidimensional data structure that allows the data analysis processes to be faster. Data setting on the cube servers to overcome the limitations of the transaction database. Data warehouse was modelled into dimension table and fact table it called star schema. This method is considered as an easier and faster step to execute queries [8]. The fact table [9]contains measureable data, while the dimension table as reference data for the data table is a fact. But in implementing the data warehouse ignores the normalization process because the design process only requires core data as a reference in the Decision Support System (DSS) [10].

3. Proposed Method
CMK company needs a new system that help them to analyse sales based on items, vendors, and customers that provide the most profit. From this system, it is possible for CMK to take a step forward for improving the financial crisis in the company. Currently the company analyzes based on OLTP databases which are basically ineffective if used in the purpose of analyzing data. OLAP database will give a better result in terms of time used to analyze. Therefore, it is important for using data warehouse that is able to process data from operational databases to identify and analyze decision.

Kimball & Ross [1] says that there are 9 steps to creating a data warehouse, this step is known as nine step design methodology [4] [11]. This 9 steps consist of:

Choose the process. Choose a process by determining the main subject which is a particular business activity in the company. Where this subject can answer the business questions needed in the analysis. In this case, the sales transaction becomes the subject to be used.

Choose the grain. Determining the grain means determine the level of detail from the fact table for the selected business process [1]. Transaction details per product become the grain that will be applied during the process of making CMK Company data warehouse.

Identify and conform the dimensions. At this step, dimension tables related to fact tables can be selected based on the relation of each row of data. This step will also connect dimension tables with fact tables.

Choose the facts. Determine what information or facts will be displayed in the fact table, including the required measure.

Store precalculations in the fact table. At this step, a review is needed to find out whether there are still attributes that must be pre-calculated first and whether it needs to be stored in the database.

Round out the dimension tables. What properties must be in the dimension table in order to describe the table. To describe, it can be done by loading structured information about the attributes in the dimension table, for example: giving information about the data types of each attribute. The dimension table must be clear and describe the hierarchy of attributes so that it is easy to use for analysis.

Choose the Durations of the database. Select time duration of the data that will be used in the data warehouse. The duration used in this case is 3 years starting from 2003-2005.

Determine the need to track Slowly Changing Dimensions (SCD) [12]. Changing dimensions cannot be predicted but it can cause problems. There are physical changes such as changes in the customer’s new address, but there are also changes that occur due to correction of data input errors [2].

Decide the physical design. The physical design of the data warehouse will be carried out at this step. Where great attention is needed to the problems that may arise in physical design that can influence the perception of the subject.
Figure 1 is a star schema [13] for the sales process at CMK Company. Star schema consists of 1 fact table: OrderFact and 5 dimension tables: ProductDim, CustomerDim, EmployeeDim, OfficeDim, and TimeDim.

![Star Schema Diagram]

**Figure 1. Star Schema**

ETL process [2] [14] [15] for selling process can be seen in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8.

![CustomerLoad Diagram]

**Figure 2. CustomerLoad**

![ProductLoad Diagram]

**Figure 3. ProductLoad**

![EmployeeLoad Diagram]

**Figure 4. EmployeeLoad**
4. Analysis Report
After ETL is created, the user can see the report from existing data. With this report, the user can get an overview of the analysis data about the company. This data can later be used in making decisions for the company’s progress. Pentaho report designer is used for generate report in Fig. 9 and Fig. 10.

Figure 9 shows a record of transaction data in detail, where each row of data refers to each item in each order number. Figure 10 is an example of a bar chart that shows the total transactions per month and per city. Year parameter is used for this chart. In this example, the year taken was 2003. Figure 11 displays the dashboard with 2 parameters that is year and country. User can display data according to their wishes based on these 2 parameters. Dashboard display information about product names, top 10 customers, and top 9 vendors that generate the most sales. This dashboard was made using Qlik Sense.
Figure 9. Record Data Transaction

Figure 10. Bar Chart total Transaksi perbulan perkota

Figure 11. Dashboard
5. Conclusion
By developing a data warehouse with this method, data can be reported faster and efficient for making decision. Using Pentaho BI tools, manager can see report about the percentage of most saleable items, both seen in terms of products, customers, or vendors. The result of this analysis, CMK can take action the strategic action to improve the benefit company. Some samples benefit are described as follows. For Knowing the products and top vendors with the most sales, it can give us an idea of which products need to be stocked more, and from which vendor the products are most in demand. Information about top 10 customer can also help in giving rewards to attract the customer.

Reference
[1] Kimball R and Ross M 2010 The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence
[2] Kimball R and Ross M 2011 The Data Warehouse Toolkit
[3] Martins D, Ramos C M Q, Rodrigues J M F, Cardoso P J S, LAM R, and Serra F 2015 “Challenges in Building a Big Data Warehouse Applied to the Hotel Business Intelligence,” Proc. 6th Int. Conf. Applied Informatics Computing Theory.
[4] Breslin M 2004 Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon Models,” Intelligence 9 pp 6–20
[5] Bizarro P and Madeira H 2002 Adding a Performance-Oriented Perspective to Data Warehouse Design, Data Warehou. Knowl. Discov., 2454 12 pp 232–244
[6] Rouse M 2015 What is data warehouse? - Definition from WhatIs.com, [Online]. Available: https://searchdatamanagement.techtarget.com/definition/data-warehouse. [Accessed: 13-Nov-2018].
[7] Dewan S, Aggarwal Y, and Tanwar S Review on Data Warehouse, Data Mining and OLAP Technology: As Prerequisite aspect of business decision-making activity, IJRIJ Int. J. Res. Inf. Technol. Int. J. Res. Inf. Technol. 1 10 pp 30–39
[8] Doreswamy, Gad I, and Manjunatha B R 2017 Hybrid data warehouse model for climate big data analysis, Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2017
[9] Dehne F, Kong Q, Rau-Chaplin A, Zaboli H, and Zhou R 2015 Scalable real-time OLAP on cloud architectures, J. Parallel Distrib. Comput. 79–80, pp 31–41
[10] Johnson J C 2015 OCP: Oracle9i™ Performance Tuning.
[11] Girsang A S, Isa S M, Adytiansya N, Utomo O K, and Simarmata J 2018 The data warehouse for down payment administration in the Constitutional Court of Republic of Indonesia, IOP Conf. Ser. Mater. Sci. Eng. 420 pp. 012104
[12] Ricky M Y 2014 Datwarehouse sales and supply of goods model based on HTML5, J. Theor. Appl. Inf. Technol. 61 1 pp. 175–182
[13] Chaudhuri S, 1998 An Overview of Data Warehousing and OLAP Technology Surajit
[14] Gill R and Singh J 2014 A Review of Contemporary Data Quality Issues in Data Warehouse ETL Environment, J. Today’s Ideas - Tomorrow’s Technol. 2 2 pp 153–160
[15] Hashem I A T, Yaqoob I, Anuar N B, Mokhtar S, Gani A, and Ullah Khan S 2015 The rise of ‘big data’ on cloud computing: Review and open research issues, Inf. Syst., 47 pp. 98–115