Design and Implementation Data Warehouse in Insurance Company

IOP Conf. Series: Journal of Physics: Conf. Series **1175** (2019) 012072 doi:10.1088/1742-6596/1175/1/012072

Design and Implementation Data Warehouse in Insurance Company

Ryan Ari Setyawan[1], Eko Prasetyo[2], Abba Suganda Girsang[3]

[1]Department of Informatics Engineering, Janabadra University, Yogyakarta, Indonesia 55231.
[2,3]Computer Science Department, Binus Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia.

agirsang@binus.edu

Abstract. *Insurance company are certainly has a rich of data from their business process. One of the data is related to business process sales. From this sales data can be used to analyze the condition of a company whether the condition of the company is in a good condition or not. The purpose of this research is to develop a technique to analyze this data from the company. The method used to implement the data analysis in this paper is to design a data warehouse. Use ninestep methodology as the method for design data warehouse and Pentaho as a tool for ETL (Extract, Transform, Load), OLAP analysis and reporting data, the results of this research concluded that the implementation of data warehouse for perform data analysis and reporting better and more effective.*

1. Introduction

Insurance is an agreement with which an insurer binds himself to an insured person by accepting a premium, to provide reimbursement to him because of a loss, damage or loss of expected profit which he may suffer from an event that is not certain. It is a form of risk management primarily used to hedge against the risk of a contingent, uncertain loss. An entity which provides insurance is known as an insurer, insurance company, or insurance carrier. A person or entity who buys insurance is known as an insured or policyholder. The insurance transaction involves the insured assuming a guaranteed and known relatively small loss in the form of payment to the insurer in exchange for the insurer's promise to compensate the insured in the event of a covered loss. The loss may or may not be financial, but it must be reducible to financial terms, and must involve something in which the insured has an insurable interest established by ownership, possession, or preexisting relationship. The insured receives a contract, called the insurance policy, which details the conditions and circumstances under which the insured will be financially compensated. The amount of money charged by the insurer to the insured for the coverage set forth in the insurance policy is called the premium. If the insured experiences a loss which is potentially covered by the insurance policy, the insured submits a claim to the insurer for processing by a claims adjuster. In insurance company, a data can represent the condition of the company whether the condition of the company is in a good condition or not. This insurance company, researched in this paper, have a lot of data from their activities in sales. So far the company does not have a method that can be used to analyze data, observe the trends of the sales and build a reporting from several variable. One of the impact of the absence of this applications is the company cannot maintain consistency sales from the sales insurance time by time.

One method that can be used to create data analysis, looking at the current trends and conditions of a data is to design a data warehouse, a database format specifically designed for data analysis and reporting. Through this data warehouse process data analysis and reporting process can be done quickly and accurately by using updated data. This will have an impact on the decision-making process in management can be more rapid and effective on matters relating to sales, for example, the decline in sales within a certain time can be immediately analyzed and given the policy of the management how to raise again the value of sales from the company. Data warehouse becomes an alternative solution that can give answer to existing problem in making report on transactional database. Data warehouse architecture is a database design that is focused for the utilization of reporting. With data warehouse architecture reporting process becomes faster, interactive and also can be in real time.

Research on data warehouse implementation has been done in many fields. Implementation of data warehouse to improve the quality of Customer Relationship Management (CRM) implementation in a company[1]. Data warehouse also implemented for research in medical, study has been studied for medical imaged [2], clinical immunology laboratory[3], development transactional bank system[4] and benchmarking in clinical rehabilitation[5]. Data warehouse also can be used for support data mining[6]. Data warehouse also can be used for astronomy, studied in earth observation satellites[7]. Research with data warehouse also has been studied in semantic web data[8]. In the telecommunications area, data warehouses have been used to analyze of data usage from telecommunications services[9]. In construction management, managers can view data from various perspectives with significantly reduced query time, thus making decisions fas and more comprehensive[10], also in another research in constructions data warehouse can be used to manage cost management[11]. More specific research on data warehouse has been done that is on the dimension table design, dimension table can be made historically by implementing the concept slowly changing dimension[12][15]. From some of these studies, the advantages of the design of data warehouses that have been made are the advantages in the support of managerial decisions, decision making can be faster and can present a comprehensive picture of the data.

2. Designing The Data Warehouse

The Data warehouse is made with several stages of integrating some data from various sources by using database integration tools then insert into the data warehouse schema that has been made before. Figure 1 explained about the process of designing data warehouse. The database that has been created is then accessed on data transformation, for example OLAP analysis, reporting data, dashboard etc. Processed results are then that will be accessed by end users as clients who use the data.

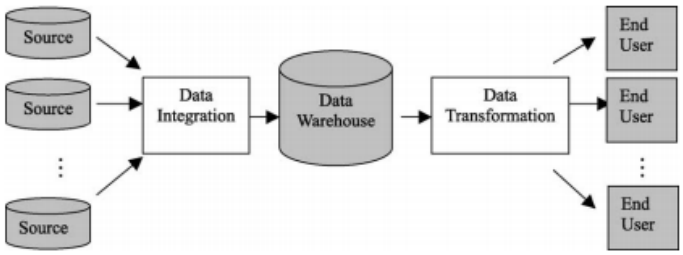

Fig. 1 Process of Data Warehouse Design[9].

In designing a data warehouse by Kimball & Ross provide a methodology commonly called nine-step design methodology on how to step in constructing a good data warehouse[16]. This nine steps implemented in this paper with the details as below :

a. Choosing the business process

The selection of business processes is to determine what business processes will be used for data warehouse creation. From the existing problems then determined the business process to be selected are:

i. The business process of selling insurance in cash.
ii. The business process of selling insurance sales installment.

b. Choosing the grain

Grain is the data line that will be stored in the data warehouse. Determining the items that specify the details of the data will be stored in the data warehouse. The details of the items to be selected in this study are one line of data from each sale of insurance sales.

c. Identifying and Conforming the dimensions

In this stage identification of dimension tables related to the business process. If in identification dimensions there is something similar that will be used in more than 1 business process then a table of dimensions can be used together for some fact tables. This dimension is often called conform dimension. The identification result in the sales process is found in several dimensional tables given in Table 1.

Table 1. Identification Dimension Table

Business Process	Dimension	Dimension Table
Sales Cash Insurance	Time	dim_time
	Customer	dim_customer
	Product	dim_product
	Polis	dim_polis
	Location	dim_location
	Branch	dim_branch
	Category	dim_category
Sales Installment Insurance	Time	dim_time
	Customer	dim_customer
	Product	dim_product
	Polis	dim_polis
	Location	dim_location
	Branch	dim_branch
	Category	dim_category

d. Choosing the fact

The selection of facts is to determine the fact table in accordance with the selection of business processes to be analyzed. The contents of this fact table are the transactional data for which each line has been determined by grain. From the business processes that have been determined in the previous stage obtained 2 fact tables. Table 2 explained details about the fact table.

Table 1. Identification Fact Table

Business Process	Fact Table	Attribute
Sales Cash Insurance	fact_sales_fullpayment_insurance	sk_polis
		sk_product
		sk_category
		sk_branch
		sk_customer_account

	sk_customer_client sk_fullpayment_date sk_location value_insurance_object gross_premium discount stamp_duty charges commision

e. Storing pre-calculation in the fact table
After obtaining some fact tables it is necessary to re-examine whether there are still things that need to be added back in the fact table. There could be a measure that in the future will be useful for the report. From stage 4 that has been done, no more calculations will be included in 2 fact tables.

f. Rounding out the dimension table
Once the dimension table has been completed, a re-examination of the dimension table needs to be done, such as specifying the data type of each attribute, adding text descriptions and defining the hierarchy of dimension attributes to simplify the analysis process. In this step includes information in the form of a list of dimension table descriptions and their table lists, and schematic drawings are already designed.

g. Choosing the duration of the database
Selection duration is to determine how long the time range of data to be included in the data warehouse. The longer and fuller the data that can be entered will be the better, especially in the future will be used for historical data analysis. In this research will be used 5-year data from 2010 until 2014.

h. Tracking slowly changing dimension
In slowly changing dimension, attribute values may change over time and must be tracked. They should maintain consistency and correctness of data, and show good query performance[17]. Slowly changing dimension commonly used is Type I replace old data with new data, type II adding new column for identification and new row for new data and type adding new column in same row old data[18]. From several dimension tables that have been made then obtained some dimensions that will implement slowly changing dimension. Table 3 provide a list of table that implement this SCD.

Table 3. Dimension Table That Use Slowly Changing Dimension

Dimension Type	SCD Type
Dim_branch	SCD Type II
Dim_product	SCD Type II
Dim_costumer	SCD Type II

i. Decide the physical design

Attention at this stage is about the physical form to be selected, such as database design, disk size where data will be stored and the type of database application to be used. Figure 2 show the design of star schema of this paper. As shown in the picture that the number of fact tables that exist there are 2 fact table that stores the data of insurance sales transactions in cash and another fact table that stores the transaction data insurance sales in installments. The two fact tables have 8 measured sizes to be displayed in the analysis: value_insurance_object, gross_premium, discount, stamp_duty, charges, commission, nett_premium and quantity. Each fact table related to the dimension table by relation many to one relation.

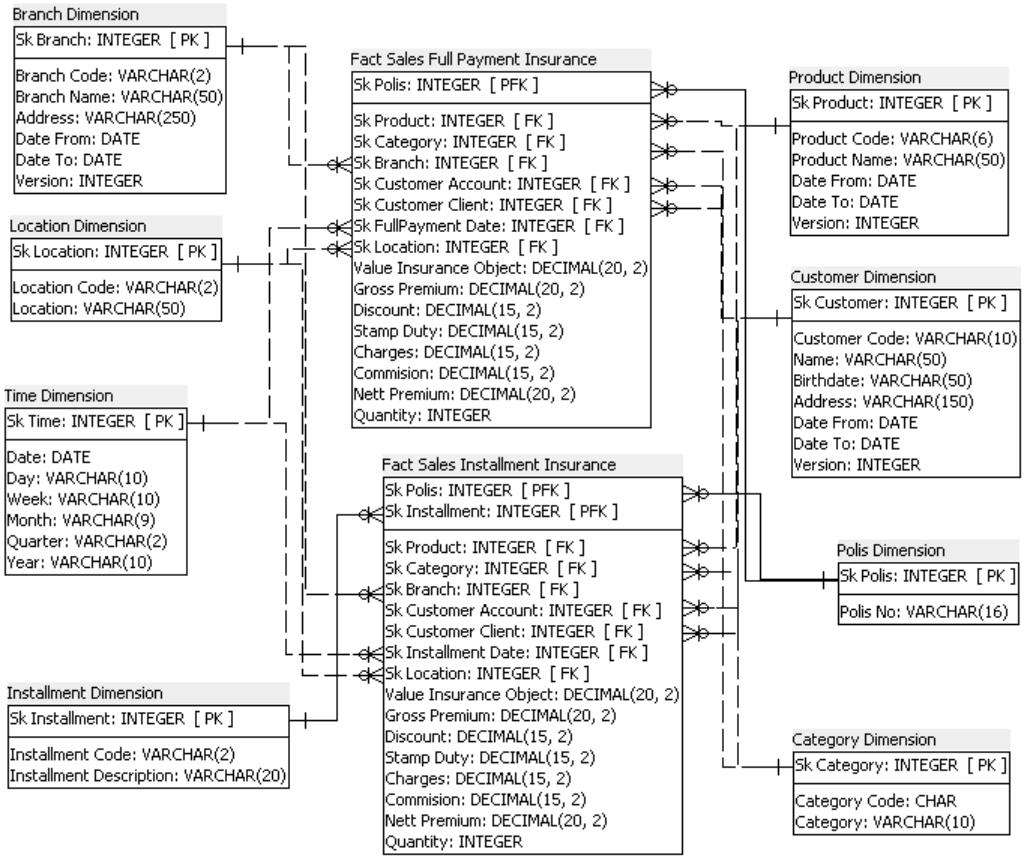

Figure 2. Star Schema Data Warehouse

3. Result

The results of the design of data warehouse that can be made in various user interfaces that will be used by the user. Figure 3 results of sales during the last 5 years from 2010 to 2014. In 2010 the company's sales results reached 473,533,272.00 and fell in 2011 to 425.670.140,00. In 2012 the number of sales of companies increase significant in 999.661.248,00. But one year later sales of the company has declined in 523.304.337,00. One year later in 2014 the company's sales increased better although not large enough at 572.693.909,00. This inconsistency in terms of sales is one of the things that management wants to change. The company wants the sales results to increase from year to year, and if at a certain moment cannot increase is expected not to decrease significantly.

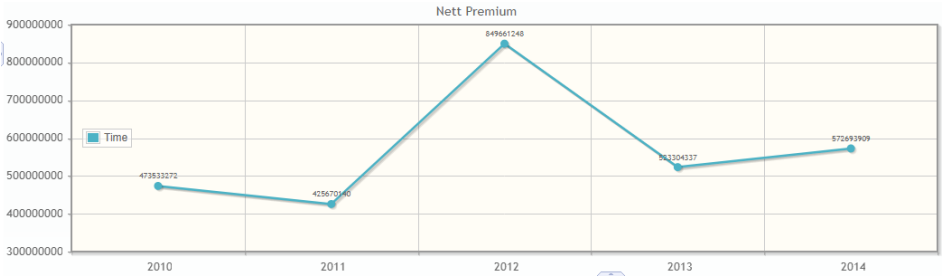

Figure. 3 **Sales by Time Chart.**

Another result can be shown is in OLAP analysis. Figure 4 shows the results of insurance sales and total insurance sales for 5 years using OLAP analysis. With OLAP analysis can be obtained flexible reporting which variables you want to display.

	Measures		
Time	Gross Premium	Nett Premium	Total Quantity
2010	536,404,843	473,533,272	106,965
2011	475,340,559	425,670,140	111,225
2012	898,447,048	849,661,248	112,251
2013	579,557,611	523,304,337	150,067
2014	616,184,631	572,693,909	134,732

Figure. 4 **OLAP Analysis Sales by Year.**

Figure 5 is the result of a more detailed OLAP analysis with time variables. Time variables are detailed again starting from the number of sales per year, per quarter up to per month. With a more detailed report this can be obtained better and more specific.

	Measures		
Time	Gross Premium	Nett Premium	Total Quantity
2014	616,184,631	572,693,909	134,732
Q1	107,906,094	98,997,687	36,142
Januari	47,491,680	43,672,995	11,609
Februari	24,736,326	22,344,994	11,044
Maret	35,678,088	32,979,697	13,489
Q2	154,139,904	144,377,741	33,395
April	46,298,613	42,966,253	10,963
Mei	32,385,811	29,444,643	11,223
Juni	75,455,480	71,966,845	11,209
Q3	121,062,508	109,317,012	32,284
Juli	51,415,352	46,343,100	11,046
Agustus	32,475,977	30,099,044	10,464
September	37,171,180	32,874,868	10,774
Q4	233,076,125	220,001,469	32,911
Oktober	39,507,031	37,200,139	9,529
November	84,177,442	78,586,564	10,437
Desember	109,391,651	104,214,766	12,945

Fig. 5 OLAP Analysis Sales by Quarter and Month.

OLAP analysis can also be create using more than one variable. Figure 6 is an OLAP analysis that using time variables and location variables, it can be seen that at the Batam location the sales numbers are

decreased while Denpasar has increased. Jakarta still cannot be predicted yet, sometimes decreased and also increased. From this result the management can doing action to the branch to analyze why this branch have sales declined.

		Measures		
Location	Time	Gross Premium	Nett Premium	Total Quantity
Batam	2010	6,809,210	6,354,355	958
	2011	6,721,548	6,294,070	524
	2012	5,834,110	5,439,499	513
	2013	5,841,636	5,373,807	1,278
	2014	6,485,604	5,921,362	671
Denpasar	2010	944,256	884,646	526
	2011	1,352,764	1,273,620	830
	2012	1,425,850	1,345,372	1,161
	2013	2,923,687	2,528,111	3,277
	2014	2,457,088	2,203,194	4,219
Jakarta	2010	500,184,902	441,094,438	87,538
	2011	438,276,593	392,325,899	90,010
	2012	434,840,099	389,380,852	87,956
	2013	525,629,433	474,311,195	96,646
	2014	564,845,232	526,525,645	86,965

Fig. 6 OLAP Analysis by Time and Location.

4. Conclusion

The design of data warehouse in this company can be used to analyze the results of insurance sales. By using OLAP analysis can be used to see trends in sales trends and can be used as a reference to the management to make decisions if there is a problem with the business results of the company's sales. The process of making reports either OLAP analysis or reports in other formats such as Excel, pdf becomes faster and easier than using the previous system. Decision making in the company become better and easier. Implementation of data warehouse that have been made can be developed for other business processes not only in terms of insurance sales because conceptually there are dimensions that can be used by other business processes such as customer dimension (customer), time dimension and branch dimension, product dimension can be made conform dimension. Security also became issue after data warehouse has been designed. There should be a separation of the role users for each analysis and report that has been created. For example, top management can access all the analysis and report, the human resource can only see in terms of employees not with related to finance and others

References

[1] A. Khan, N. Ehsan, E. Mirza, and S. Z. Sarwar, "Integration between Customer Relationship Management (CRM) and Data Warehousing," *Procedia Technol.*, vol. 1, pp. 239–249, 2012.

[2] J. W. Teixeira, L. P. Annibal, J. C. Felipe, R. R. Ciferri, and C. D. de A. Ciferri, "A similarity-based data warehousing environment for medical images," *Comput. Biol. Med.*, vol. 66, pp. 190–208, 2015.

[3] G. C. Miller, D. Ph, C. T. Dolan, M. D, A. N. Crowson, and S. R. Stout, "Data warehousing and information management strategies in the clinical immunology laboratory," *Clin. Appl. Immunol. Rev.*, vol. 3, pp. 73–86, 2002.

[4] P. L. Takecian *et al.*, "Methodological guidelines for reducing the complexity of data warehouse development for transactional blood bank systems," *Decis. Support Syst.*, vol. 55, no. 3, pp. 728–739, 2013.

[5] J. Completo, R. S. Cruz, L. Coheur, and M. Delgado, "Design and Implementation of a Data Warehouse for Benchmarking in Clinical Rehabilitation," *Procedia Technol.*, vol. 5, pp. 885–894, 2012.
[6] J. A. Lyman, K. Scully, and J. H. Harrison, "The Development of Health Care Data Warehouses to Support Data Mining," Clin. Lab. Med., vol. 28, no. 1, pp. 55–71, 2008.
[7] S. Liu, C. Han, S. Wang, and Q. Luo, "Data warehouse design for earth observation satellites," Procedia Eng.,vol. 29, pp. 3876–3882, 2012.
[8] V. Nebot and R. Berlanga, "Building data warehouses with semantic web data," Decis. Support Syst., vol. 52, no.4, pp. 853–868, 2012.
[9] Girsang, A., Prakoso, C., Data Warehouse Development for Customer WIFI Access Service at a Telecommunication Company, (2017) International Journal on Communications Antenna and Propagation (IRECAP), 7 (2), pp. 114-124. doi:https://doi.org/10.15866/irecap.v7i2.11736
[10] K. W. Chau, Y. Cao, M. Anson, and J. Zhang, "Application of data warehouse and decision support system in construction management," Autom. Constr., vol. 12, no. 2, pp. 213–224, 2003.
[11] S. W. Moon, J. S. Kim, and K. N. Kwon, "Effectiveness of OLAP-based cost data management in construction cost estimate," Autom. Constr., vol. 16, no. 3, pp. 336–344, 2007.
[12] C. Stanfill, "Type 2 slowly changing dimensions: A case study using the cooperating system," 15th ACM Int. Work. Data Warehous. OLAP, Dol. 2012 - Co-located with CIKM 2012, pp. 81–88, 2012.
[13] A. U. Sepulveda, R. C. Loyola, and M. W. Hernandez, "Performance comparison slowly changing dimensions using model relational and object-relational," in Proceedings - International Conference of the Chilean Computer Science Society, SCCC, 2016, vol. 2016–Febru.
[14] T. Manh Nguyen, A. Min Tjoa, J. Nemec, and M. Windisch, "An approach towards an event -fed solution for slowly changing dimensions in data warehouses with a detailed case study," Data Knowl. Eng., vol. 63, no. 1, pp. 26–43, 2007.
[15] R. C. Loyola, A. U. Sepulveda, and M. W. Hernandez, "Optimization slowly changing dimensions of a data warehouse using object-relational," in Proceedings - International Conference of the Chilean Computer Science Society, SCCC, 2016, vol. 2016–Febru.
[16] R. Kimball and M. Ross, Relentlessly Practical Tools for Data Warehousing and Business Intelligence. 2010.
[17] S. Faisal and M. Sarwar, "Handling slowly changing dimensions in data warehouses," J. Syst. Softw., vol. 94, pp. 151–160, 2014.
[18] Ralph Kimball, "The Data Warehouse Lifecycle Toolkit," p. 771, 1998.