Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

Data Warehouse Implementation Techniques In Data Processing (Case Study Data Sales at PT Spirit Sejahtera Bersama)

Made Hanindia Prami Swari[1], Fawwaz Ali Akbar[2], Martoni[3]

[1.2]Informatics Engineering, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" East Java, Rungkut Associate street, Surabaya, 60294, Indonesia
[3]Informatics Engineering, STMIK STIKOM Indonesia, Tukad Pakerisan street, Denpasar, 80225, Indonesia

E-mail: hanindia.pramiswari@gmaill.com

ARTICLEINFO

Article history:
Received: 17 Sept 2019
Revised: 24 Sep 2019
Accepted: 18 Nop 2019

Keywords:
data warehouse. ETL, pentaho

ABSTRACT

Data Warehouse is a technique that can provide a facility for the storage and management of historical data from various internal and external sources so that it can be used in making strategic decisions for the company. A characteristic that distinguishes ordinary data warehouse database was largely on the data source that can not necessarily uniform and comes from one source as happened in PT Sejahtera Bersama spirit. The company is engaged in the sale of medical equipment, up to now has more than 70 showrooms and is divided in 12 regions, of which the sales information system used by the company in each region has a separate database in excel format, this is what makes firms difficulty in the process analyzing the sales data of the company. This research aims to design and build a data warehouse system to support marketing performance assessment and achievement of sales at PT Sejahtera Bersama spirit. Data modeling methods used in the design of data warehouse in this study is the dimensional modeling star schemas. Results from this study is a web-based dashboard application that will display sales data from the data warehouse is equipped with charts and reports, and the results of tests with the method of black box testing can be seen that the system has been going well as expected. Data modeling methods used in the design of data warehouse in this study is the dimensional modeling star schemas. Results from this study is a web-based dashboard application that will display sales data from the data warehouse is equipped with charts and reports, and the results of tests with the method of black box testing can be seen that the system has been going well as expected. Data modeling methods used in the design of data warehouse in this study is the dimensional modeling star schemas. Results from this study is a web-based dashboard application that will display sales data from the data warehouse is equipped with charts and reports, and the results of tests with the method of black box testing can be seen that the system has been going well as expected.

1. Introduction

PT spirit Sejahtera Bersama has an extensive distribution network throughout Indonesia, and to this day has had more than 70 showrooms across major cities in Indonesia and some in Bali. The company has received many awards both nationally and internationally as a top brand award. To support the sales process PT spirit Sejahtera Bersama has been utilizing technology-based information systems of desktops

Journal Mantik
Volume 3 Number 3, November 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

that can be accessed online, the system saves the sales data in a database, but the data stored still has the disadvantage that when the data is to be presented back to the needs of the analysis by the manager, because the information generated is still shaped table that the data therein is still mixed a lot of data such as data showroom, marketing data, customer data and sales data, in addition to the many existing fields and figures increasingly baffled manager to analyze the information. This led to the lengthy process of analysis carried manager each month plus the amount of data that must be analyzed, with monthly sales branch for Bali and Nusa Tenggara reached an average of 300 to 500 transactions. Beginning of each month the manager must analyze sales data for the previous month to get some information needed by the company for the evaluation stage as both showroom maun target marketing, customer with the highest number of purchases or sales trend with previous months. The duration of the process of analyzing sometimes affect the time manager in decision making processes such as determining the showroom or marketing which need to be evaluated because sales are still below target the company set. Delay manager took the decision could have an impact on the progress of the company are less than optimal, because that company needs a proper system to help manage data and classify it into multiple dimensions so that it can be quality information that companies need in the decision making process. The system we know as the data warehouse. The duration of the process of analyzing sometimes affect the time manager in decision making processes such as determining the showroom or marketing which need to be evaluated because sales are still below target the company set. Delay manager took the decision could have an impact on the progress of the company are less than optimal, because that company needs a proper system to help manage data and classify it into multiple dimensions so that it can be quality information that companies need in the decision making process. The system we know as the data warehouse. The duration of the process of analyzing sometimes affect the time manager in decision making processes such as determining the showroom or marketing which need to be evaluated because sales are still below target the company set. Delay manager took the decision could have an impact on the progress of the company are less than optimal, because that company needs a proper system to help manage data and classify it into multiple dimensions so that it can be quality information that companies need in the decision making process. The system we know as the data warehouse.

*Data warehouse*is a concept and a combination of technologies that facilitate the company to manage and maintain historical data obtained from the system or operational applications. The data warehouse is a perfect solution for the company as a support in decision-making, because with this system the company can integrate various types of data from various different sources. This ensures access mechanism "one door" for management to obtain information, and analyze them for decision making.

Based on the problem above, PT Sejahtera Bersama spirit requires implementations to simplify enterprise data warehouse in the reporting process sales data. Through suistem to be made, the information generated by the sales information system PT Sejahtera Bersama spirit will be recycled and are grouped into several dimensions. Information will flow faster than ever before without the use of a data warehouse because the information obtained has been classified according to the needs and equipped with a chart that can be understood by anyone without having to train someone to understand a system.

2. Literature review

Some theories that form the basis of this research includes the concept of a data warehouse, dimensional modeling and ETL.

2.1 Data Warehouse

Data Warehouse is a collection of integrated databases, subject-oriented, designed to support the functions of DSS, where each unit of the relevant data for a time. The data warehouse contains a small fraction of the company's data [2]. According to [4], the data warehouse is a data storage area where the storage capacity of large-scale, data is accumulated by adding a new record and not to update the existing record with the new information, the data very easily towed, and the data is used solely for decision-making and not to operations of the company. According to [5], the data warehouse is a collection of integrated data extracted from operational databases, historical, and external, cleansed, transformed, and cataloged for withdrawal and analysis (data mining),

From the definition of a data warehouse according to the experts above, it can be concluded that the data warehouse is a database that interact and can be used for query and analysis, is the orientation of the subject, integrated, time-variant, unchanged (adhoc) the principal purpose of manufacture is to combine different data sources and diversified into a storage space so users can easily run a query data, generate

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

reports, and perform the analysis to assist decision-making by the decision maker. According to [1], data warehouse architecture can be seen in Figure 1

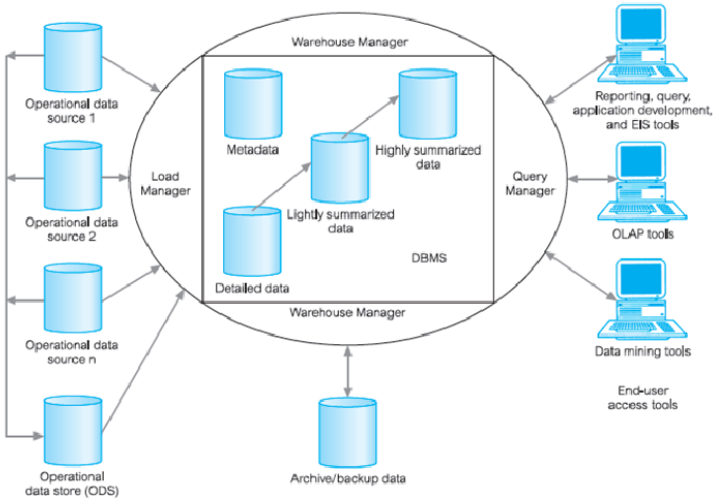

Fig 1. Data Warehouse Architecture

The characteristics of the data warehouse according to [2], namely:

a. *Subject Oriented* (Oriented subject) that the data warehouse is designed to analyze data based on subject-specific subject within the organization, rather than on the processes or functions of a particular application.
b. *integrated* (Integrated) is Data Warehouse can store data coming from separate sources into a consistent format and integrated with one another.
c. Range of Time (Time-Variant) pursuant to which all data in the data warehouse can be said to be accurate or valid at any given time.

There are four tasks of a data warehouse, the preparation of reports, OLAP, data mining, and executive information center [3]. One of the methodological design of the data warehouse is a Nine-Step Methodology by [1] which has 9 (nine) steps, namely selecting process, select the source, identifying and adjusting the dimensions, choose the facts, save the initial calculations in Table fact, looking back at the table dimension , select the duration dimension to track the changes of dimension, and decide priorities and the type of query.

2.2 Dimensional Modeling

Dimensional modeling is a logical design technique that aims to present the data in a standard, intuitive form that can be accessed with the high performance [1]. Each dimensional models consisting of a table with a composite primary key called the fact table, and a set of smaller tables called dimension tables. Each table has a primary key dimension (nonkomposit) corresponding simple right with a composite key in the fact table. In other words, the primary key of a fact table is made of two or more foreign keys. Characteristics of the structure as it is called star star or star schema join. In addition to a star schema, there are two other models namely snowflake schema structure, and starflake schema.

2.3 Extract, Transform, Load (ETL)

ETL is a process to be followed in the establishment of the Data Warehouse [3]. Has 3 Stages ETL processes ie data extraction is the process by which data is taken or extracted from various operational systems, using either a query or ETL applications. Transformation is the process by which the raw data (raw data) extraction results filtered and modified in accordance with the applicable business rules. The last process that needs to be done is the loading of the data obtained from the transformation into the data warehouse. one of the tools that can be used to perform this process is Pentaho.

3. Analysis and Design System

Design activities needed to guide the implementation into the software. The diagrams are designed in this study is a context diagram sserta star schema.

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

3.1 Context Diagram

*context diagram*is the highest level of data flow diagram that describes all activities that occur between the system and its entities. Context Diagram become the basic guidelines for system designers to know the technical basic foundation of a system design. Is in the form of current technical input, process and output. Context Diagram also describes the use of the system from the user in the access rights for the use of the system. The context diagram of the data warehouse system PT.SSB can be seen in Figure 2.

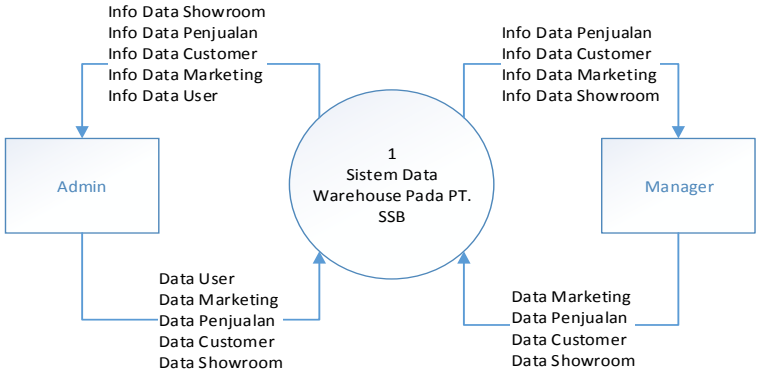

Fig 2. Context Diagram

3.2 Dimensional Modeling

Modeling dimensional is a design technique that aims to present data in a standardized framework that allows accessible with high performance, in which the data warehouse system at PT spirit Sejahtera Bersama, using Star schema or model of the dimensional data that has a fact table at its center surrounded tables dimensions are not normalized. The fact is most useful in a fact table is a numeric calculations on each record. While the dimension tables contain descriptive textual information. Figure 3 shows are Star schema in the data warehouse system PT Sejahtera Bersama spirit.

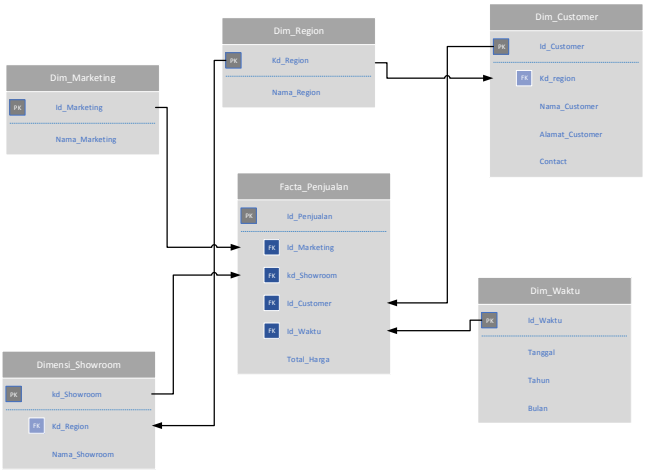

Fig 3. Star Schema

4. Implementation and Testing System

Stages of implementation systems consist of two activities, namely the implementation pentaho and implementation of web are used to display the results of data processing into a form that is readable by the user.

4.1 Implementation of Pentaho

To create a data warehouse pentaho tool is needed to perform the ETL process before the data is

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

stored into the data warehouse, and then make the appropriate dimensions of the system requirements. Figure 4 is the dimension and fact ang will be established in this research activity.

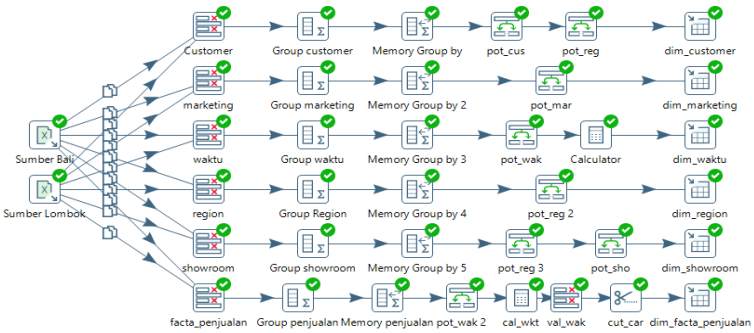

Fig 4. Dimensions and Facts

As seen in Figure 4, there are 5 pieces of dimensions formed, the dimensions customer, marketing, time, region, and a showroom. And there is one fact that is a fact of sales. Sales fact table is a table to store the entire data of the entities involved in the sales transaction process so that the process of querying the data all dimensions will be connected with the sales fact table.Sales facts of ETL processes will produce output that sales data will be stored in the data warehouse in the table fakta_penjualan as can be seen in Figure 5.

Examine preview data

Rows of step: Table output Facta Penjualan (1000 rows)

#	Invoice No	Customer	Total	Sales Organization	Sales Person	IdWaktu
1	SO/17K/GEL00018	GEL.17.0000096	1392000.0	GEL010001	ADVS005477	30112017
2	SO/17K/GEL00016	GEL.17.0000094	150000.0	GEL010001	ADVS005477	23112017
3	SO/17K/GEL00015	GEL.17.0000093	608000.0	GEL010001	ADVS005477	30112017
4	SO/17K/GEL00014	GEL.17.0000092	226000.0	GEL010001	ADVS005477	24112017
5	SO/17K/GEL00013	GEL.17.0000091	2211000.0	GEL010001	ADVS005477	24112017
6	SO/17K/GEL00012	GEL.17.0000090	40000.0	GEL010001	ADVS005477	22112017
7	SO/17K/GEL00011	GEL.17.0000089	198000.0	GEL010001	ADVS005477	22112017
8	SO/17K/GEL00010	GEL.17.0000088	242000.0	GEL010001	ADVS005477	23112017
9	SO/17K/GEL00009	GEL.17.0000087	100000.0	GEL010001	ADVS005477	25112017
10	SO/17K/GEL00007	GEL.17.0000085	398000.0	GEL010001	ADVS005477	17112017
11	SO/17K/GEL00006	GEL.17.0000084	1000000.0	GEL010001	ADVS005477	15112017
12	SO/17K/GEL00004	GEL.17.0000082	350000.0	GEL010001	ADVS005477	08112017
13	SO/17K/GEL00003	GEL.17.0000081	198000.0	GEL010001	ADVS005477	06112017
14	SO/17K/GEL00001	GEL.17.0000079	6678000.0	GEL010001	ADVS005477	22112017
15	SO/17K/GAC00309	GAC.16.0000064	500000.0	GAC010001	ADVS006125	30112017

Fig 5. Output Facts Sales

4.2 Implementation of Web

Website created based on the context diagram that has been built before. Some of the main functions of the system described in this section.

4.2.1 User Interface Sales

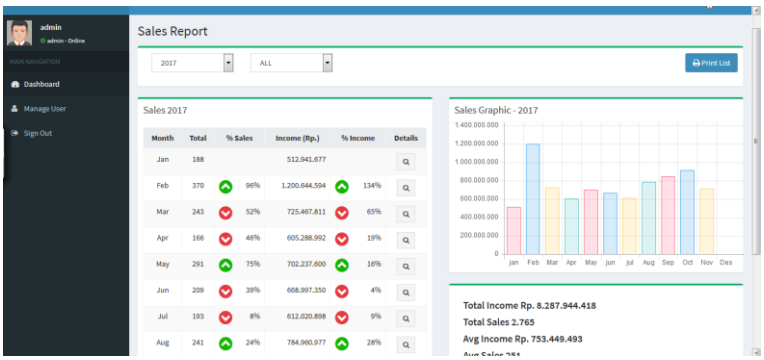

Fig 6. Home Sales

In this sales page will display the sales data for each period where there are also filters the year and region that serves to filter the data to be displayed. Pictures implementation of this page can be seen in Figure 6.

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

4.2.2 Sales Details page

On the detail page of this sales transaction detail will show sales on the month, the year, and the region that the user wants. GABAR 7 shows pages that display the sales report.

Fig7. Detail Page Sales

4.2.3 Home Sales Report

This report on sales page the user can print the appropriate data analysis panjualan year and the region that has been selected. The more details can be seen in Figure 8.

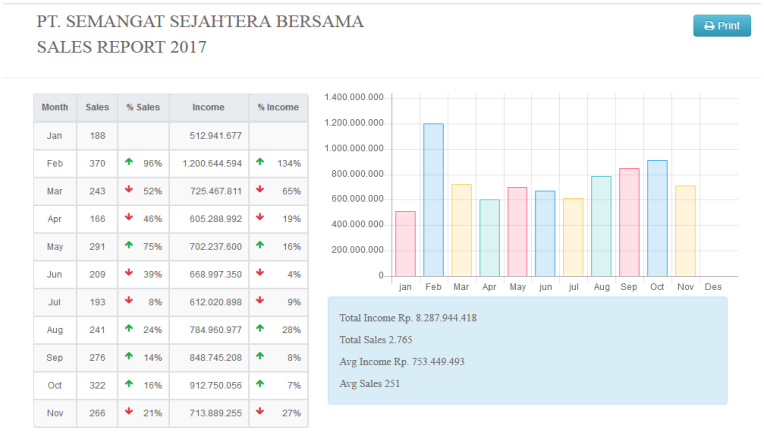

Fig 8, Sales Weather Report

4.2.4 Home Sales Detail Report

On this page the user can print sales data in accordance with the year, month, and the desired region. The more details can be seen in Figure 9.

Fig 9, Detail page Sales Report

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

View consisted of the page dimensions, details, reports, and detail reports made to keseluruahan dimensions formed.

4.2.5 Testing System

System testing phase is done in order to know the function of the system is running as expected. Testing method is by using black box testing. Table 1 lists the results of tests performed.

TABLE 1
USING THE BLACK BOX SYSTEM TESTING

No.	Aim	input	Output Expected	Information
1	Avoid empty data input	One of the empty fields	Warning appears with complete data content	succeed
2	*Login* failed	Input the username and password are incorrect	The message login failed	succeed
3	*Login* success	Input the correct username and password	Successful login message appears, and enter the user level in the menu page	succeed
4	Avoid empty data input	One of the empty fields	Warning appears with complete data content	succeed
5	Add your user data failed	Input e-mail that has been registered	Appearing warning e-mail address is already registered on the system	succeed
6	Add your user data successful	Input all fields correctly	New user message appears successfully saved	succeed
7	Avoid empty data input	One field emptied	Content warning appears with complete data	succeed
8	*Edit* user data successful	Changing all fields correctly	The message the user successfully revamped	succeed

4.3.4 Testing Total Data Sources and Data Warehouse

This test is performed to determine whether the ETL process has been done properly, by examining the amount of data in the data source if it is equal to the amount of data in the data warehouse and dashboard. The test results can be seen in Figure 10.

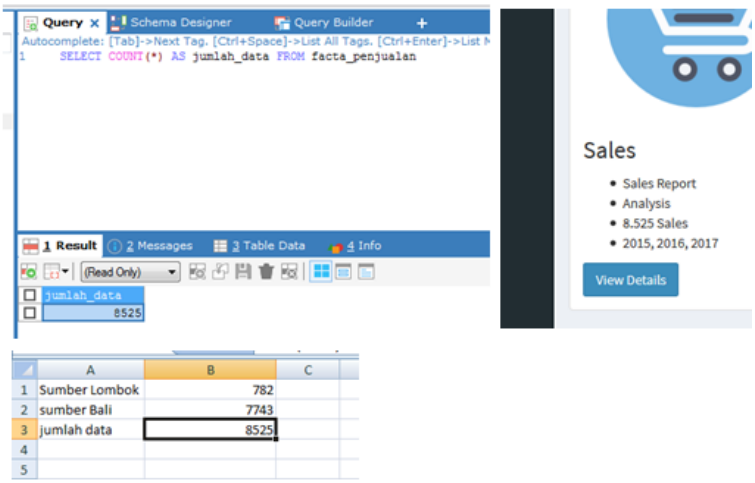

Fig 10. Testing Number of Data

5. Conclusion

Based on the results of the discussion that has been done, it can be concluded that:

a. Stages through to design and build a data warehouse system are the stages of planning, data collection, data integration, system design, system manufacturing and system testing.
b. From the results of testing black box testing, it can be concluded that the data warehouse system has been built according to the design stage and the system can function properly as expected.

Journal Mantik
Volume 3 Number 3, Novermber 2019 e-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik/index

c. Data stored in the data warehouse the same amount by the number of data sources, so the process of ETL (extract, transform, load) has been done correctly.

d. This web-based dashboard application, enables to analyze the data, because the data has been split several dimensions and also comes with charts and reports.

6. References

[1] Connolly, Thomas and Begg. 2005. Database Systems: A Practical Approach to Design, Implementation, and Managements. England: Addison Wesley.
[2] Inmon, WH 2002. Building the Data Warehouse, Third Edition. New York: John Wiley & Sons, Inc.
[3] Kimball, Caserta. 2011. The Data Warehouse ETL Tollkit: Pratical Techniques for Extracting Cleaning Conforming and Delivering Data. Mizan: Bandung.
[4] McLeod. 2004. Management Information Systems. Jakarta: PT. Index.
[5] O'Brien, James. A. 2005. Introduction to Information Systems Business and Managerial Perceptive: Salemba.