IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
ISSN (Online): 1694-0814
www.IJCSI.org

46

The Process of Data Tabulation Using Data Warehouse and OLAP Technology to Sales Analysis at Distributor Company

Febrian Wahyu Christanto[1], Wiranto. H Utomo[2], and Eko Sediyono[3]

[1] Information System Department, Faculty of Information Technology,
Satya Wacana Christian University, Salatiga, Central Java 50711, Indonesia

[2] Faculty of Information Technology, Satya Wacana Christian University,
Salatiga, Central Java 50711, Indonesia

[3] Faculty of Information Technology, Satya Wacana Christian University,
Salatiga, Central Java 50711, Indonesia

Abstract

Recently, the needs of data processing into information have a very important role, especially in the highly competitive corporate world. Here, the role of technology is required to produce information that is fast, safe, and appropriate in assisting the company process while making decision. As the development of OLAP technology in Indonesia, the study took place at PT. Sinar Niaga Sejahtera in Ambarawa, Indonesia. It will be constructed with the Data Warehouse using a Star Schema and OLAP technology based on Website to complete the company's needs about the summary of the information which was generated to analyze the company's sales results. The processing data into information which is emphasized in this paper is to describe in details about the data processing that will help in the learning and develop an application to build a better OLAP system.
Keywords: OLAP, Data Warehouse, Data Processing, Summary Report.

1. Introduction

Nowadays, the need of information and knowledge in the modern life become something important for human being even for today's organizations. Information is a collection of data made in such a way to be seen and consumed by the data owner as one factor to help the determining process. The meaning of knowledge is a compilation of information which is taken and used to solve a problem [1].

From the explanation above, thus, the validity and speed of cultivation from data become important to face the global competition. In accommodating that need, technology has a big impact to help in the process so that the information will bring some benefits in a short time.
Since the varieties of the data are getting bigger and varies, they should be supported with the technology which is capable to save data safely. Warehouse data is the best solution to save data in a large amount with diverse varieties. The solution offered by Warehouse Data is that there will be one and only trusted data source which can be used for all people who need it and it is also good for reporting, information analysis, even for decision making. The facilities that have been explained before about using Warehouse Data can save in the infestation which compare to the scope of every department, must build a process by building the operational application.

According to Suprianto [2], Data Warehouse is a data compilation from various source which is integrated, aggregated, and structured, saved in a data warehouse (repository) in a huge capacity so the user is capable of checking the data history and supporting the analysis made by the user himself. Therefore, the characteristic and structural design of Data Warehouse can be seen in Figure 1, which shows the process into Data Warehouse so that the information can be processed and result in fulfilling the user's needs.
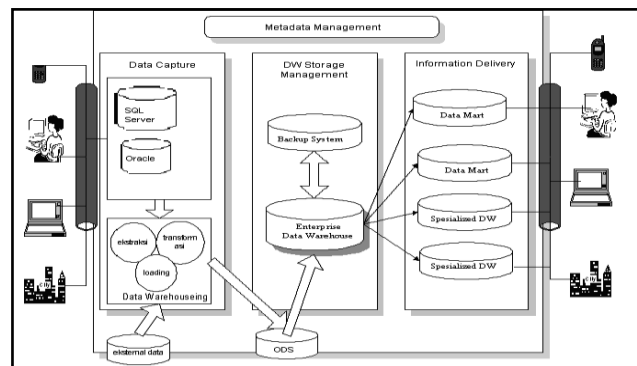

Fig. 1 Data Warehouse Architecture [2]

The structural design of Data Warehouse (Figure 1) shows that the data is taken from several extensions like SQL and Oracle. The data is then added into Data Warehouse, so that the data will be backed up for security. Taking the data from Data Warehouse can not be directly performed, but it must go through some Data Marts or some special applications in the Data Warehouse. The data inside the Data Mart is spatial as a sample data because the original data is in the Data Warehouse and it cannot be manipulated. The data and application inside the Data Mart can be manipulated and taken for some processes on it.

One of the special applications to process Data Warehouse is OLAP (On-Line Analytical Processing) that is an application which is used to present the result of the data processed taken from the Data Warehouse. OLAP is a system that has a task to change the data which is saved inside the Data Warehouse and then transform it into a multidimensional structure (cube). Later, the result can be used as a summary report in multidimensional [3]. Several applications produced by OLAP are business reports for selling, marketing, management report, management business process, budgeting and predicting, financial report and other similar reports. Generally, OLAP can be seen as the right approach in providing answers to analytic query which is multidimensional. The characteristic of OLAP based on Lin and Donald [4] said that: 1) Fast, which means information can be taken quickly to process the data. Take for example, 1.000.000 transactions would only need less than 2.5 minutes by 10 dimensions and 3 types of shapes, 2) Analysis, which means the system can overcome Business Logic and statistical analysis data that is relevant with the existing data, 3) Share, which means this system is implemented by a security system to protect the secret of information so, every users can arrange their right of access based on their need, 4) Multidimensional, which means the key of the need in using OLAP application always contains multidimensional unsure, and 5) Information, which means it conveys all data and derivative information needed by the application.

OLAP shows a class of technology that is created to analyze and access the data in a special way. When the general process of transaction is solely on the relational database, OLAP comes up with a point of view in the multidimensional data. The point of view of the multidimensional is supported by multidimensional database technology. Such a way gives a basic technique to calculate and analyze using an application business. OLTP has the characteristic that some users can do something like cheating, updating, and retrieving to every record data. Moreover, OLTP is very optimal to update the data. OLAP application is used for analyzing and arranging the level frequency from the aggregation (a summary of data amount in Data Warehouse which is spatially detailed simpler to increase the data warehouse performance) [5]. Database is usually updated on the compilation of data, it is hardly from multiple sources and it places the strength of analysis in the back-end to operate the application. That is why OLAP is very optimal to be used for analyzing something. When the data has been processed from Data Warehouse toward Data Mart or it is called as OLAP Cube (Figure 1), then that data will be distributed through several queries to an OLAP application as interface that directly connects to the user. The structural design of OLAP is in Figure 4.

2. Related Work

The development of technology in using OLAP that supported by Data Warehouse in developing countries, especially in Indonesia, is increasing from time to time. There has been several researches even in building OLAP application in different aspects. Taken from several literatures, they can be used as a reference in building the OLAP system and analyzing the data into information in PT. Sinar Niaga Sejahtera.

The research about OLAP and Data Warehouse in Indonesia have been carried out in focusing the research in designing and building OLAP and Data Warehouse in PLN Salatiga, Indonesia using Snowflake Scheme. This research discusses about the creation of Data Warehouse to accommodate the transaction data in every sub-districts in Salatiga and surrounds it. The research could help PLN Salatiga in analyzing the data automatically when the transaction was done. Therefore, while in the past the analyzing data was performed manually, now it can be easier and faster. Several data analyses that can be used for PLN's interest, such as counting the revenue from a different next dimension, can influence the adoption of policy. That policy is connected with the analysis of electric power, the ability to pay, and the misapplication product from PLN Salatiga by the people who use the electricity service. This research can be seen in Figure 2 [6].
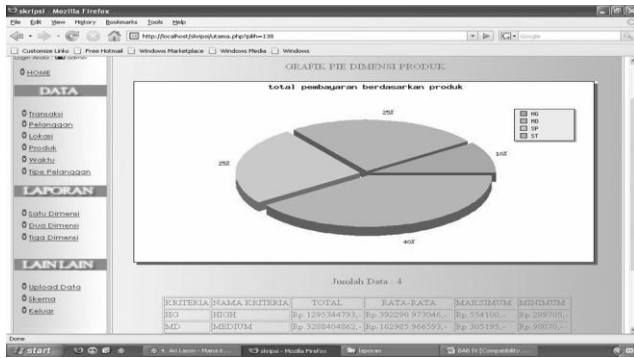
Fig. 2 OLAP PLN Salatiga [6]

Fig. 3 OLAP PT. Aneka Tuna Indonesia [9]

Furthermore, there is a research in educational world in Indonesia entitled EDUOLAP (Educational Online Analysis Processing) which is used for helping the process of decision making that serves as the representation of the education in Indonesia comprehensively because it is supported with accurate data and it is entailed with the tested methodology analysis. The main target in this system is the leader as the decision maker of the policy which is suitable with the scope of his responsibility area in the Department of National Education of Indonesia Republic [7]. There is also another reasearch about Data Warehouse entitled "Analysis and Data Warehouse Design in the Big Hall of Meteorology and Geophysics I Province Medan, Indonesia." Here, data warehouse is being built for consolidation and to provide the compilation of data in Meteorology, Climatology, and Geophysics based on the users' need. It can also be accessed quickly. Based on the design, Data Warehouse can provide the need for information which is detailed in giving information about weather and earthquake to fulfill the need in the future, the need for completeness, the need of fast access, and the need of integration process of data [8]. Nowadays, OLAP is being utilized to make a report in the process of production line. One of the examples is in the PT. Aneka Tuna Indonesia. PT. Aneka Tuna Indonesia is a company which operates in processing tuna fish, produces it into several different products and distributes it into some foreign countries in the world. The report shows the dynamic of special period in the production which started from the delivery of the fish by supplier into the staffing. This report will be presented in the form of table and graphic. Using the application of OLAP helps the management of PT. Aneka Tuna Indonesia in controlling and supervising the process and the results of the production that support the decision making for the company. This research can be seen in Figure 3 [9].

Next, another research talking about Data Warehouse based on OLAP which was also based on education was "Making Data Warehouse to Perform the Measurement in Teaching and Learning Process" conducted in the Information and Technology Department at Petra Christian University. It was about Data Warehouse that was used in helping provide data required in teaching learning process. The information in the Data Warehouse came from the lectures' and the students' performance, graduation level courses, and the summary of the results for each course. The multidimensional query process was done by using pivoting tables and charts. Users can manipulate the data that appears on each axis as it can be done in the pivot table, but the information will be seen like a graph [10].

3. Aim of The Research

Several definitions and the implementations have been given by several experts on the Data Warehouse and OLAP technology. This study will carry out an analysis data about processing data which will be processed into an information (Figure 4) that are relevant to the OLAP technology which is built in PT Sinar Niaga Sejahtera, Ambarawa , Indonesia (It is the distributor company of food and snack in the region of Central Java, Indonesia).
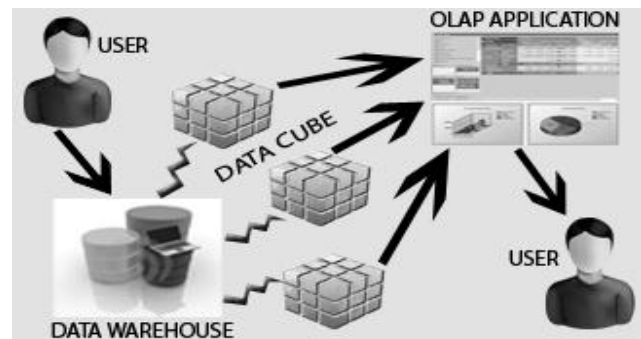

Fig. 4 Data Tabulation Process with OLAP Technology

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
ISSN (Online): 1694-0814
www.IJCSI.org

49

In Figure 4, the OLAP model in PT. Sinar Niaga Sejahtera explains that the data will be inputted by one of the users, as an administrator in the system through ETL process, so the data (extension .Xlsx) is stored in the Data Warehouse. Moreover, the retrieval and process of data will be carried later in the Data Cube for showing to other users via OLAP application.

4. Research Contribution

In terms of the company side, information expected is in the forms of statistical data and graphical data which are generated by OLAP technology that will help the company to make the decision and policy for the company's progress in the future. The competition is very tight, so the quick and accurate analysis information is very necessary to determine the strategy in order to obtain the maximum benefit.

Furthermore, the contribution to the learning and developing system for the results of this study will be used as a better guideline for the development of OLAP system because this study will describe the process of processing data using OLAP applications.

5. Star Schema and Snowflake Schema

Data Warehouse model is logic design that represents the data in a standard form, and supports the access to data quickly. There are several terms related to this model: the fact table, dimension tables, and hierarchies. Fact table is a table that consists of primary key from another table. Every fact table in the conceptual Data Warehouse model has a composite key. On the contrary, every table that has a composite key is the fact table. In other words, each table showing the "many to many relationships" must be the fact table. Dimension table is a simple table in which there is a primary key associated with a composite key in the fact table. Hierarchy defines that a sequence of mapping from the lower level to the higher level as a more general concept. The most popular dimensional model is the star schema, snowflakes schema, and the fact constellation schema.

In the implementation of the Data Warehouse, there are two types of relational database schemes; they are Star Schema and Snowflake Schema. Both of these schemes are common in data warehouse development processes. Star schema is one of a dimensional model where the fact table is located in the center of the other tables. It is the dimension tables which are situated around it. Most of the fact table in the star schema is the third form of database normalization, while the dimensional table is the second form of normalization. Star schema is the simplest model

from dimensional form. Then, Snowflake schema is an expansion form of the star schema where there is a normalization process of some or complete dimension table. Snowflake schema is usually used on large tables and in the condition when the star schema is not able to describe the complexity of the database [11].

This study uses a Star Schema as the schema in the Data Warehouse development. The selection scheme is based on the advantages of using Star Schema that is easily understood as a simple database structure and low maintenance. Database schema on this system is shown in Figure 5.
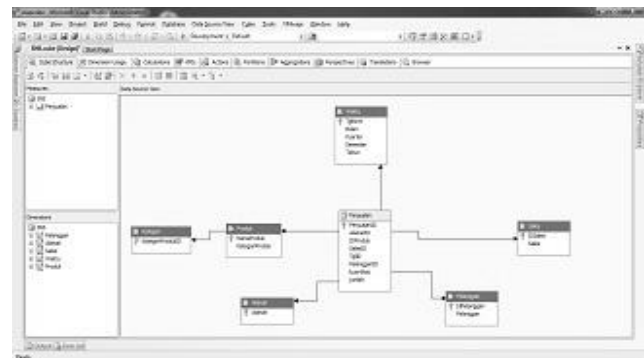

Fig. 5 Star Schema

Figure 5 explains that the Fact Table from this system is Sales (sale) because this system will be analyzed in the report of the sale of PT. Sinar Niaga Sejahtera. The Dimension Table for this scheme is a table of Products (Product), Time (Time), address (Address), CS (Customer), and Sales (Salesman). Products table here will be related to Categories table (Category) as Many to One. As a result, scheme in this system is a mixture of Star Schema and Snowflake Schema, but is more dominant to the Star Schema.

6. Data Processing using Data Warehouse and OLAP Technology

In this discussion, the process of the data is described to be more relevant information on the application of OLAP in PT. Sinar Niaga Sejahtera. It will first start from the company's sales data upload into the system (Figure 6). After that, the ETL process will be done by saving the file format extension .xlsx to .mdl (using DBMS SQL Server 2008).

Fig. 6 OLAP Interface

ETL process is performed by querying the application process through adjusting the contents of the data which are uploaded to the database field that has been made previously. This process will be shown in Figure 7.

Fig. 7 ETL Query Process

Then, the system will access data format .Xlsx to be read and stored in the Data Warehouse with OleDb Connection. Data that has been stored in the Data Warehouse can be seen in Figure 8.
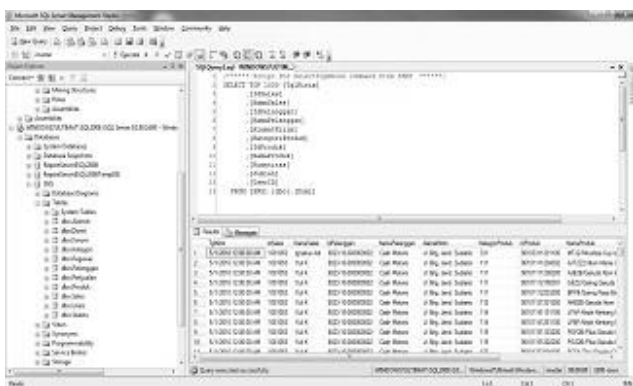

Fig. 8 Data Warehouse

Here, the saving data in the Data Warehouse results in a repository data that cannot be done in this process. Data process, however, can be carried out by creating some Cube from Dumi data, so that it can be processed into useful information (Figure 9).
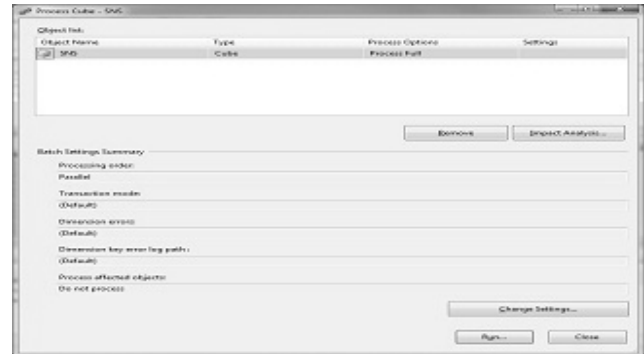

Fig. 9 Making a Cube

The Cube in this system, named SNS, stands for Sinar Niaga Sejahtera. Using the supporting feature from DBMS Microsoft SQL Server 2008, that is, Analysis Service Server Type, the making of Cube can thus be done more easily. After the Cube is formed, it can be determined what things needed to be displayed for users. This explanation is in Figure 10.
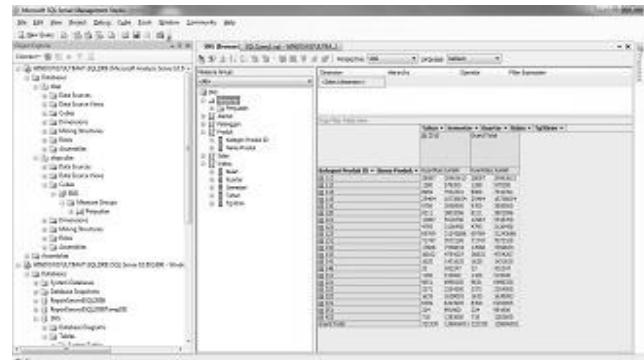

Fig. 10 Processing Cube

After the setting process of information in the Cube is completed, then, the next process is application setting. The application in OLAP System uses ASP.NET programming language in Microsoft Visual Studio 2010.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
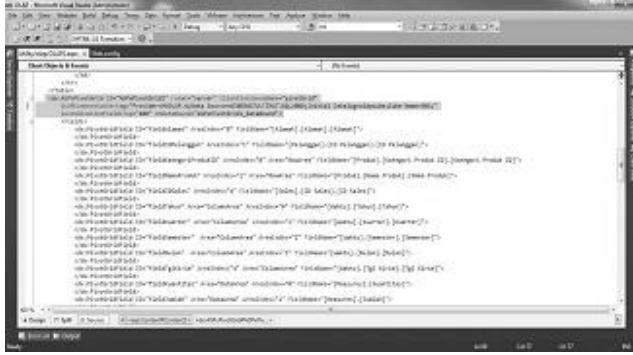ISSN (Online): 1694-0814
www.IJCSI.org

51

Fig. 11 OLAP's Connection String

Figure 11 explains about the formation of Connection String formation to display data that has been processed in the Cube to the application. This system uses OLAP Pivot Grid to display information in the application utilizing libraries that contain ADO.NET to get Connection String into Pivot Grid. To display a chart or a representative diagram of the numerical information in the form of graphic images, the writer uses a Web Chart Control by adding tools from other supporting software. Next, after all processes are accomplished, the result of OLAP system for PT. Sinar Niaga Sejahtera can be seen as displayed in Figure 12 and Figure 13.
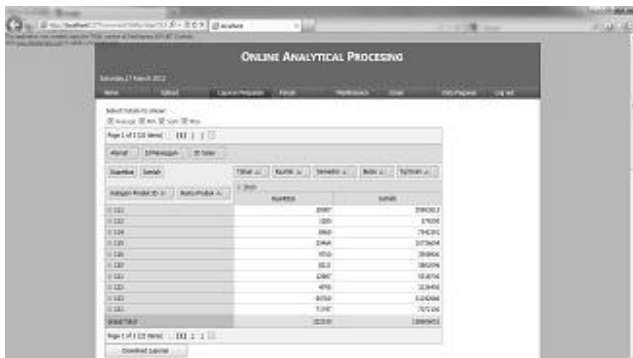

Fig. 12 OLAP's Pivot Grid

Figure 12 and Figure 13 are the results of the previous data processing, which is then processed into information by OLAP system. The information is the representative of the data, which has been stored for some period of time on the sales data of PT. Sinar Niaga Sejahtera. The information is represented in the form of numerical and graphical information in order to facilitate the reading. OLAP system meets certain criteria, named OLAP applications, such as: 1) Rotation / pivoting, which is an action to see information from a different point of view to make it easy in reading the data information; 2) Slice and Dice, which is to cut some of the data required by the user information (Slice), and to divide the information again to be smaller

part(s) (Dice); 3) Drill Down and Roll Up, it is the OLAP capability to view a subset of information more completely (Drill Down) and to view the global information or the summary of information (Roll Up) [9].
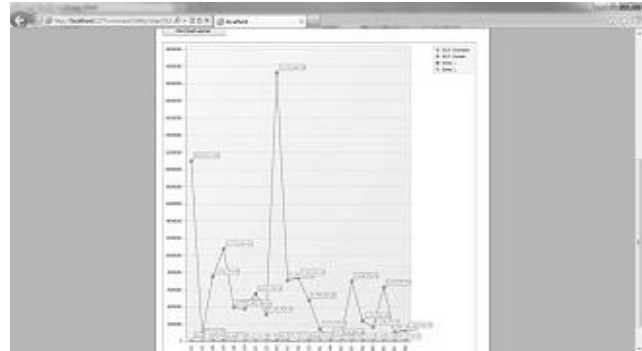

Fig. 13 Graphical Representation of OLAP

This OLAP application has to give a clear analysis about the sale result of the company. In this system, sales results analysis can be seen by referring to the sales' timing (per day, per week, per month, per quarter, per semester, or per year). More than that, it can also show the highest sales analysis, average sales, low sales, and total sales by simply pressing the menu buttons provided.

7. Future Work

The writer also proposed some suggestions for the development and improvement of the OLAP technology in PT Sinar Niaga Sejahtera. In this case, the OLAP technology can be more powerful and more stable by adding an application for making predictions, which is integrated into one system. These predictions can be calculated from the sales data using the free additional software provided by Microsoft by downloading SQL Server Data Mining Add-Ins application (Figure 14), in which Microsoft Office Excel will connect directly to Microsoft SQL Server 2008. The prediction process will be fully processed in the database for later be displayed in Microsoft Office Excel.

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
ISSN (Online): 1694-0814
www.IJCSI.org

52

Fig. 14 SQL Server Data Mining Add-Ins Installation

In this application, there are many predicting methods that allow a user in performing data prediction quickly and accurately. Several prediction methods which can be used in this software are the Classify, Estimate, Cluster, Associate, Forecast, and so forth. The results of these predictions can be in the form of graphs, diagrams, or in the form of numerical information as in Figure 15, in which the writer makes predictions using the Forecast method.
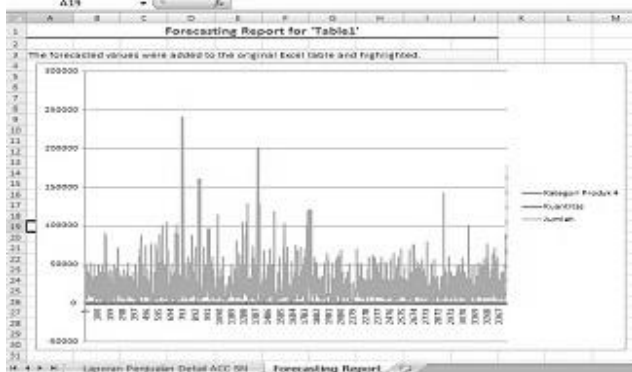

Fig. 15 Predicted Results

The research cannot be said perfect yet and still needs some further improvement to get a maximum result. Besides that, other features can also be added that will help in providing more information and, of course, to be an informative web service data warehouse system, which is better than the present.

Another applicable suggestion for further development is setting the server stabilization so that the server's performance will stay in the optimal conditions. Thus, the application can work faster and stable without any deferment when required. In addition, the present application not only processes sales and employee data, but also accesses the company's financial data through the addition of tables in the database, which makes the job of employees who work in the finance division more easily and the results better.

8. Conclusions

Finally, it can be concluded from the research that Data Warehouse and OLAP technology are very helpful for this current condition, especially in a highly competitive business world, which needs an instant and precise analysis to process the data into information that later will be used to help the company in making decision. This study is also conducted to analyze the way of data processing can be fast, secure, and precise, which later can advantage the learning and the application developer to build a better OLAP system.

This information represented in numerical and graphical form can also be seen from many different points of view, either time or number, so it is easy for PT. Sinar Niaga Sejahtera in reading and analyzing sales' results and determining strategies that will be used to sell products in order to obtain maximum profit.

References

[1]   Karvalics, L.Z., Information Society-What is it Exactly?, Budapest: European Commission, 2007.
[2]   Suprianto, Data Warehousing, http://student.eepisits.edu/DataWarehouse.html, 2008.
[3]   Ivanova, A, and Boris Rachev, "TU-Varna: Multidimensional Models-Constructing Data Cube", in International Conference on Computer System and Technologies-CompSysTech, 2004.
[4]   Lin, Song, and Donald E. Brown, Outliner-based Data Association: Combining OLAP and Data Mining. Virginia: Department of Systems Engineering, University of Virginia Charlottesville, 2002.
[5]   Edi, Doro, and Stevalin Betshani, "Analisa Data dengan Menggunakan ERD dan Model Konseptual Data Warehouse", Jurnal Informatika, Vol. 5, No. 1, June 2009, Maranatha Christian University.
[6]   Handoyo, and Sri Yulianto Joko Prasetyo, "Perancangan dan Pembangunan Data Warehouse pada PLN Salatiga Menggunakan Skema Snowflake", Aiti, Vol.5, No.1, 2008, Satya Wacana Christian University.
[7]   Balitbang, EDU-OLAP (Education Online Analysis Processing),http://www.balitbang.depdiknas.go.id/2008/?p=97, application link on http://www.eduolap.kemdiknas.go.id., 2008.
[8]   Safril, Agus, and Aries Tjahyanto, "Analisis dan Desain Data Warehousing di Balai Besar Meteorologi dan Geofisika Wilayah I Medan", in Prosiding Seminar Nasional Manajemen Teknologi III, 2006.
[9]   Prasetyo, Muhamad Adi, Ahmad, Saiku, and Sarwosri. "Pembuatan Aplikasi OLAP untuk Pelaporan pada PT. Aneka Tuna Indonesia Menggunakan SQL Server 2005", **Undergraduate Theses,** Institute of Technology, Surabaya, Indonesia, 2010.
[10]   Handojo, Andreas, and Silvia Rostianingsih, "Pembuatan Data Warehouse Pengukuran Kinerja Proses Belajar Mengajar di Jurusan Teknik Informatika

IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 2, May 2012
ISSN (Online): 1694-0814
www.IJCSI.org

53

Universitas Petra", Jurnal Informatika, Volume 5, No. 1, May 2004, Petra Christian University.

[11] Sharma, Umashanker, and Anjana Gosain, "Dimensional Modeling for Data Warehouse", Indraprastha University, Delhi, India, 2010.

Febrian Wahyu Christanto, a senior student at Magister of Technology and Information System, Faculty of Information and Technology at Satya Wacana Christian University Salatiga, Indonesia. Besides studying for his Magister to achieve M.Cs degree, he is also actively working on a software house "Data Frame" in Salatiga as software engineer. He got a bachelor's degree with a major in computer network, Faculty of Information and Technology Satya Wacana Christian University Salatiga in 2010. Research specialization of Febrian Wahyu Christanto, S.Kom is in the field of computer networking and business intelligence.

Wiranto. H Utomo, a lecture at Faculty of Information and Technology at Satya Wacana Christian University Salatiga, Indonesia. He got a Master Degree in Computer Engineering at the Gajah Mada University, Yogyakarta Computer Science in 2002 and his Ph.D at Gajah Mada University in 2011. He has published in several journals including IJCSI, IJWA, MASAUM Journals, IJCA, and International Conference including IWAS of the ACM. Research specialization of Dr. Ir. Wiranto Herry Utomo, M.Kom. is in the field of SOA, java EE, web services, and software engineering.

Eko Sediyono, a lecturer and professor at Faculty of Information and Technology at Satya Wacana Christian University Salatiga, Indonesia and Diponegoro University Semarang, Indonesia. He got a Master Degree in Computer Science at Indonesian University in 1994 and his Ph.D at Computer Science Department at Indonesian University in 2006. Prof. Dr. Ir. Eko Sediyono, M.Kom. is a member of the IEEE with member number 41605422. His research speciality is in the field of statistic, compilation techniques, algorithms, and data structure.