Data Warehouse Design with ETL Method (Extract, Transform, And Load) for Company Information Centre

Wulan Stau Fana[a,*], Rini Soviab[a,], Randy Permana[a] Md Ataul Islam[b]

[a] c Department Informatics Engineering , Faculty of Computer Science, University Putra Indonesia "YPTK" Padang, Indonesia
[b] Newton International Fellow, University of Manchester, United Kingdom
rini_sovia@upiyptk.ac.id*
* corresponding author

ARTICLE INFO

ABSTRACT

Data Warehouse is a technology use to analyze, extract and evaluate data into information which produce knowledge in the form of analysis to provide some advice in decision making process. Designing a Data Warehouse using ETL (Extract, Transformation and Load) process serves as the collection of data from different data sources into a multitude of integrated data sets. By using snowflake scheme for the design of the data warehouse make data prepare well and ready for analyze on Data Warehouse. The result of this research is to applied Data Warehouse that use to support company decision making progress make easier and has a good decision since it come from Data Warehouse

I. Introduction

One of the most important things in the field of consulting services that leads to the need for technology is needed, namely information technology, which all data and information is so important that it causes data needs to increase every year. The data warehouse holds the data as a whole and re-presents the data for reporting and analysis data that aims to support strategic decision making based on data analysis. The company's mapping data aims to explain the condition of information about the company from the various dimensions of the data obtained. The data warehouse will be used as the basis for strategic decision-making in the company. The purpose of the research in the design of data warehouse is to facilitate the company in decision making as well as action on the final results in the form of information due to the processing of data carried out, Facilitate the company in looking at the company's movement for several years as a whole from existing data, facilitate in the search for data that has been long, to monitor the development of the company, to improve the performance of more efficient and quality for the company , to predict the future state of the company, help the company to strategize against the challenges that exist in the company. Database is data systematically collected and stored on a computer that can be processed or altered using software to obtain information. Database is a system of hardware and software used in storing, deleting, processing and flowing data. Data can be an information with the management of a computer program. While the database management system is a system or software in which the system allows users to control, access, and maintain data practically. According to MySQL is a database server program capable of receiving and send data very quickly, multiuser and use SQL standard command (Structured Query Language). Data warehouse is a home for your high-value data, or data assets, that originates in other corporate applications, such as the one your company uses to fill customer orders for its products, or some data source external to your company, such as a database that contains sales information gathered from all your competitors.

The data used in the ETL process can come from a variety of sources including enterprise resource planning (ERP) applications, flat files, and spreadsheets. ETL (Extract, Transformation, and Load) is a very important process in establishing a data warehouse. The ETL process has the

function to collect data from different data sources making the entire data set integrated. Kimball or nine-step Kimball method was introduced by a warehouse and business intelligence data expert named Ralph Kimball who was used for the steps of building and developing a data warehouse namely Choosing the process, Choosing the grain, Identifying and conforming the dimensions, Choosing the fact, Storing pre-calculations in the fact table, Rounding-out the dimension tables, Choosing the duration of the dimension, Tracking slowly changing dimension , and Deciding the query priorities and the query modes. ETL (Extract, Transformation and Loading) is a critical process in the creation of data warehouse. The first process, extract, is the extraction of data from different data sources. Data is drawn from a variety of different operational sources and databases to flat files. After that, the data is cleaned, and transformed to fit the data warehouse scheme through the transformation process. The result of the process of ETL is it generates data that meets the criteria data warehouse such as historical data, integrated, encapsulated, static and has a structure that is designed for the purposes of business processes. snowflakes schema is a modeling on multidimensional data that is an extension of the star schema, snowflakes have the characteristic of not all dimensions contained in the fact table but can be connected to the main dimension.

Snowflake Schema is a variant of the star scheme model, some of the dimension tables are normalized so that it further breaks down the data into additional tables. The resulting schematic graph forms a shape like a snowflake. The snowflake scheme is a different scheme to the star scheme because the dimension table does not contain denormalization data. Snowflake is another variation of the star scheme, in which the dimensional table of star schemes is organized into a hierarchy by normalizing.

OLAP (Online analytical processing) is one way to process data in data warehouse. The OLAP process can be used as a medium to provide detailed and detailed information about the strategic information needed by the user. OLAP processes used include roll up & drill down. Online analytical processing is a combination of dynamic analysis and large-scale multidimensional data. OLAP consists of a combination of rules that provide a dimensional framework to support decision-making. OLAP is part of a more global category of business thinking, which also summarizes the relationship between reporting and data mining. OLAP is the data warehouse tools to perform data analysis. OLAP itself is a technology designed to provide superior performance for ad hoc business intelligence queries.

OLTP is referred to as On-line operational system work for processing transactions and queries. Data warehouse systems serve users or knowledge workers in data analysis and decision-making roles. OLAP is organizes and presents data in a variety of formats to accommodate a variety of different user needs. Data Warehouse technology is a solution for storing all data storage on OLTP. OLTP data is sent to the Online Analytical Process or OLAP as a data warehouse manifest, where a business analyst can identify all the needs to improve company's business optimization.

Data mart is a storage of data that comes from various databases or/and other sources of data that could support the business needs of the healthcare services. Data mart is a part of data warehouse to support the creation of reports and data analysis in the operations of a company. The data may be derived from various sources of information such as dedicated databases, other data warehouse and specific databases. Data marts can provide a conceptual view of dataset elements through different data ports. Data mart processes data using ETL (Extract, Transform, and Load) methods by selecting and selecting data processed from the data source, transforming the data into the appropriate data to be processed on the data mart and able to load or present the information to the user. The characteristics that distinguish a data mart and a data warehouse are as follows Data marts focus only on the needs of users involved in a department or business function and Data marts usually did not contain detailed operational data as in the data warehouse.

Data Warehouse is a method in the design of the database support the DSS (Decision Support System) and EIS (Executive Information System). Physical data warehouse is a database, but the design of the data warehouse and database are very different. In the database design using normalization, but on data warehouse normalization is not the best way. Data warehouse is very large data that is subject-oriented, integrated, time variant, and improved on stored data to support decision-making process management. Data warehouse is defined as a data set that has characteristics in the form of subject oriented, integrated process oriented, time variant, accessible,

and nonvolatile. The data warehouse contains merged data from many sources, plus summary information and covers a long period of time. Common workloads involving ad hoc queries, whose response times are complex and fast are important. These characters differentiate warehouse applications from OLTP applications and must use different DBMS designs and implementation techniques to achieve satisfactory results. Distributed DBMS with good scalability and high availability (achieved by storing tables in excess on more than one site) are required for a very large warehouse.

SQL Server is a central part of the Microsoft data platform. SQL Server is an industry leader in operational database management systems (ODBMS). This documentation helps you install, configure, and use SQL Server on both Windows and Linux. The content includes end-to-end examples, code samples, and videos. For SQL Server language topic. SQL Server is a key component of a Microsoft data platform. SQL Server as an industry leader for operational database management systems (ODBMS). SQL is a concept of database operation especially for data selection and entry that allows data operation to be done easily automatically.

Business intelligence is a set of theories, methodologies, processes, and technologies that convert raw data into meaningful and useful information aimed at business. Business Intelligence with Data Warehouse technologies are known in the literature as solutions that allow access to business data dynamically and analytical operations in company. Business intelligence is a set of theories, methodologies, processes, arithmetic's, and technologies that turn raw data into meaningful and useful information aimed at business. Business Intelligence (BI) is a collection of data warehouse, data mining, data analysis, reporting, and data visualization to collect, integrate, clean, and explore company information for decision making. BI handles a large amount of information to assist in identifying and developing a new opportunity. the Service is an activity given by one party or more to the other party that has a relationship with the purpose of providing satisfaction to both parties for the goods or services provided. The quality of service is explained by 5 gaps that cause differences in perceptions about the quality of services, namely covering the gap in management perception, the quality specification gap, the service delivery gap, the marketing communication gap, and the perceived service gap.

II. Research Method

The data source is raw data that has not been processed and has a wide range of formats. The data will be entered into the SQL Server database. Then the data is in the extract which is the process of parsing and cleaning the data to get the desired data structure. Transfrom is the second step in the process. In this process convert data from operational format to warehouse data format such as converting data type, cleaning and dumping duplication and data errors, performing calculation of derivative or initial values and data reference integrity checks. Loading is the last step that is the process to enter data into the last target that is the data warehouse. The data will be entered into the company's data warehouse. The result of the data warehouse is incorporated into OLAP to produce multidimensional information that the data can be viewed from various points of view.

Wulan Stau Fana et.al (Data Warehouse Design with ETL Method (Extract, Transform, And Load) for Company Information Centre)
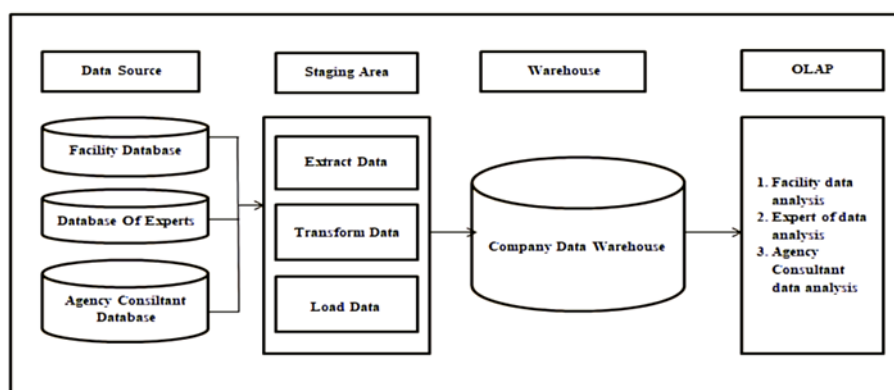
Fig. 1. Data Warehouse Architecture Design In Company

This data warehouse implementation is designed to import data into existing databases in SQL Server 2008 R2 through the data analysis process.

1. SQL Server Integration Services

SQL Server Integration Services (SSIS) is a data warehouse process using ETL (Extract, Transform and Loading) processes. Extract is for data collection from various sources, transform is the process of cleaning data and loading is the process of storage data into the database.
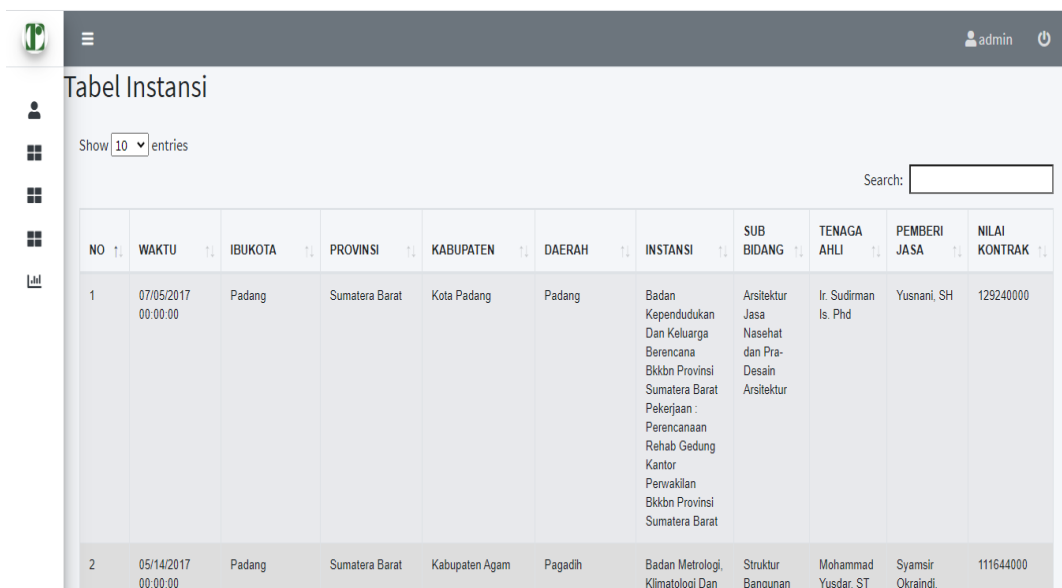
2. SQL Server Analysis services

SQL Server Analysis services (SSAS) is a process of data presentation and data analysis.

3. SQL Server Reporting Services

SQL Server Reporting Services (SSRS) is a part of the server-based report platform that can be used to provide functionality in report creation from data sources. SSRS makes it easy to create reports and can integrate reports with various settings and can also be used to view data from databases.

III. Result

Application testing is a test to find out errors in the data structure or access databases externally. View performance errors as well as existing functions in the application. This test aims at information in the creation of systems that have not been running perfectly, and this test can explain the usefulness of the applications that have been designed.

Fig. 2. Table Institutions

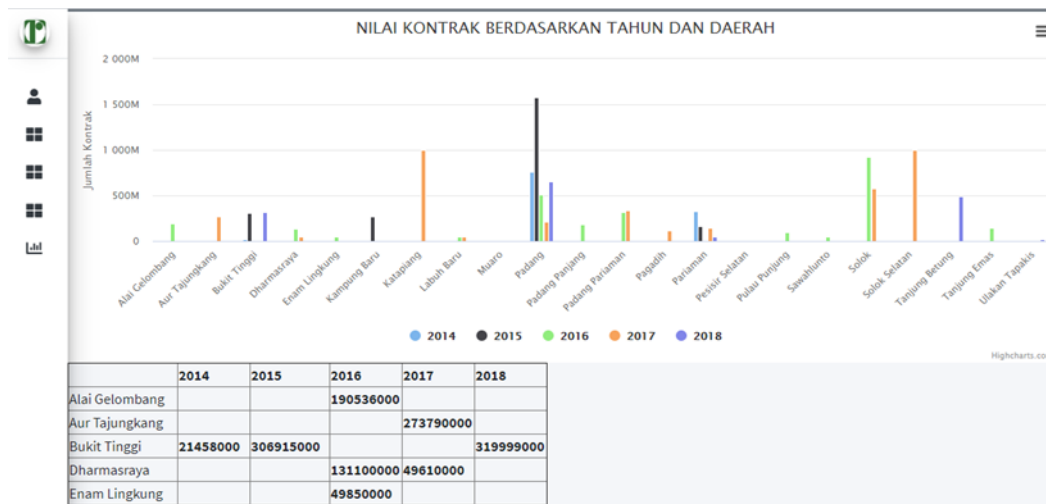The image above is a view of the institutions table in the test interface.

Fig. 3. Grapich Views

The graph to be displayed on the test interface. On the chart can see the data processed in the form of charts based on needs. To see how data is progressing based on numbers in detail and see a brief and clear comparison of interconnected data.

IV. Conclusion

Based on the results of problem identification, implementation, testing and designing data warehouse for company, we can get the following conclusions: Data warehouse can help select and identify data from various specified data sources. The data that has been retrieved and selected can be accessed based on the need for decision making on company. In the extraction process can store company data into a database for the merging of extraction results from other data sources from the company such as expert data, facility data and planning consultant data. Data warehouse can transform unneeded data and convert data into shapes that suit your needs. Map input data to a pre-defined snowflake schema form and perform duplication gaping and data errors using the pre-created primary key. Data warehouse can perform the process of presenting data (load) to be displayed to the application in the storage of data warehouse in the form of report for the company. Can assist the company in strategizing in the future in decision making in the company from the data of the information obtained.

References

[1] A. Filiana, A. G. Prabawati, M. N. A. Rini, G. Virginia, and B. Susanto, "Perancangan Data Warehouse Perguruan Tinggi untuk Kinerja Penelitian dan Pengabdian kepada Masyarakat," J. Tek. Inform. dan Sist. Inf., vol. 6, no. 2, pp. 174–183, 2020, doi: 10.28932/jutisi.v6i2.2557.

[2] I. G. W. Darma, K. S. Utami, and N. W. S. Aryani, "Data Warehouse Analysis to Support UMKM Decisions using the Nine-step Kimball Method," Int. J. Eng. Emerg. Technol., vol. 1, no. 1, pp. 65–68, 2019, [Online]. Available: https://ocs.unud.ac.id/index.php/ijeet/article/view/53633.

[3] D. Subuh, D. Harsono, and R. Sulistiyowati, "Implementasi Business Intelligence Pada Aplikasi E-Tilang Dengan Menggunakan Tableau Server," Semin. Nas. Geotik 2019, pp. 127–135, 2019.

[4] A. Yulianto, S. D. Putra, and R. Haroen, "JISICOM (Journal of Information System , Informatics and Computing) Vol . 3 No . 2 Desember 2019 p-ISSN : 2579-5201 (Printed) JISICOM (Journal of Information System , Informatics and Computing)," vol. 3, no. 2, pp. 103–109, 2019.

[5] A. Supriyatna, "Sistem Analisis Data Mahasiswa Menggunakan Aplikasi Online Analytical Processing (Olap) Data Warehouse," Pilar Nusa Mandiri, vol. 12, no. 1, pp. 62–71, 2016.

[6] G. Satyanarayana Reddy, R. Srinivasu, M. Poorna, C. Rao, and S. R. Rikkula, "Data Warehousing, Data Mining, Olap and Oltp Technologies Are Essential Elements To Support Decision-Making Process in Industries," Int. J. Comput. Sci. Eng., vol. 02, no. 09, pp. 2865–2873, 2010.

[7] I. N. A. Prabawa, D. Agung, K. Arimbawa, and I. G. N. Janardana, "Analysis and Design Data Warehouse For E-Travel Business Optimization," Int. J. Eng. Emerg. Technol., vol. 4, no. 1, 2019.

[8] Y. P. Sudarmojo, "Design of Library Data Warehouse Using OLTP Result of Services Analysis," vol. 3, no. 1, pp. 62–65, 2018.

[9] P. Widiadnyana, M. A. Maricar, I. N. Arnawan, and S. Ariyani, "Analysis of Data Warehouse for Transactional Analysis Methods Online Analytical Processing (OLAP) at Company XYZ Silver," Int. J. Eng. Emerg. Technol., vol. 2, no. 1, p. 72, 2017, doi: 10.24843/ijeet.2017.v02.i01.p15.

[10] D. Subuh and W. Yasman, "Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho," Pros. SENIATI, pp. 29–36, 2019.

[11] L. M. da L. Brum, V. do N. Lampert, and S. da S. Camargo, "Business Intelligence and Data Warehouse in Agrarian Sector: A Bibliometric Study," J. Agric. Sci., vol. 11, no. 2, p. 353, 2019, doi: 10.5539/jas.v11n2p353.

[12] M. Hilman and D. Djamaludin, "Analisis Faktor Optimalisasi Proses Etl Pada Data Warehouse Sebagai Pendukung Pengambilan Keputusan Management Dengan Business Intelligence," Fakt. Exacta, vol. 11, no. 1, p. 24, 2018, doi: 10.30998/faktorexacta.v11i1.2325.

[13 S. M. N. Huda And J. Sutrisno, "Analisa Perancangan Data Warehouse Dan Aplikasi Online Analytical Processing Pengajuan Kredit Pada Pt Bfi Finance Indonesia Tbk," Idealis Indones. J. Inf. Syst., Vol. 1, No. 3, Pp. 354–359, 2018.

[14] T. Connolly And C. Begg, Database System : A Practical Approach To Design, Implementation, And Management. Fourth Edition, Fourth Edi. England, 2005.

[15] K. Khotimah And Sriyanto, "Perancangan Dan Implementasi Data Warehouse Untuk Mendukung Sistem Akademik (Studi Kasus Pada Stkip Muhammadiyah Kotabumi)," J. Teknol. Inf. Magister, Vol. 2, No. 01, Pp. 94–107, 2017.

[16] A. Supriyatna, "Sistem Analisis Data Mahasiswa Menggunakan Aplikasi Online Analytical Processing (Olap) Data Warehouse," Pilar Nusa Mandiri, Vol. 12, No. 1, Pp. 62–71, 2016.

[17] M. K. A. Ghani, M. M. Jaber, And N. Suryana, "Telemedicine Supported By Data Warehouse Architecture," Arpn J. Eng. Appl. Sci, Vol. 10, No. 2, Pp. 415–417, 2015.

[18] R. Permana, "Perancangan Data Mart Bagian Penjualan Motor Bekas (Used Motor Cycle) Pada Cv . Atlas Motor," Vol. 3, No. 1, 2013.

[19] Z. Nur And I. Mukhlash, "Implementasi Business Intelligence Pada Manajemen Report Bank Xyz," J. Sains Dan Senni Pomits, Vol. 3, No. Bisnis Intelijen, Pp. 16–21, 2014.

[20] D. Marlius, "Loyalitas Nasabah Bank Nagari Syariah Cabang Bukittinggi Dilihat Dari Kualitas Pelayanan," J. Pundi, Vol. 1, No. 3, Pp. 237–250, 2018, Doi: 10.31575/Jp.V1i3.60.

[21] A. R. Simon And T. C. Hammergren, Data Warehousing For Dummies. John Wiley & Sons, 2009.

[22] I. P. A. E. Pratama, Handbook Data Warehouse : Teori Dan Praktik Berbasiskan Open Source, Informatik. Bandung, 2018.

[23]   Santoso, L. W., & Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. Procedia Computer Science, 124, 93–99. https://doi.org/10.1016/j.procs.2017.12.134

[24] D. Nurmalasari, M. Sari, and P. Hanifah, "Analisis kinerja pemodelan data star schema pada data perpustakaan," J. Komput. Terap., vol. 5, no. 2, pp. 44–53, 2019.

[25] Q. Aini, U. Rahardja, A. Moeins, and A. M. Wardani, "Penerapan Data Market Query (DMQ) pada Sistem Penilaian Berbasis Yii Framework," InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan), vol. 3, no. 1, pp. 26–31, 2018, doi: 10.30743/infotekjar.v3i1.565.

[26] A. Utami, B. R. Pratama, and S. R. Widianto, "Data Mart Design in Bkpp Bandung Using From Enterprise Models To Dimensional Models Method," JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer), vol. 5, no. 2, pp. 279–284, 2020, doi: 10.33480/jitk.v5i2.1219.

[27] B. Sudradjat, "Analysis and Design of Data Mart Decision Support Systems at PT Marlindo Tirta Nusantara," SinkrOn, vol. 3, no. 2, p. 133, 2019, doi: 10.33395/sinkron.v3i2.10062.

[28] I. WARMAN and R. RAMDANIANSYAH, "ANALISIS PERBANDINGAN KINERJA QUERY DATABASE MANAGEMENT SYSTEM (DBMS) ANTARA MySQL 5.7.16 DAN MARIADB 10.1," J. Teknoif, vol. 6, no. 1, pp. 32–41, 2018, doi: 10.21063/jtif.2018.v6.1.32-41.

[29] A. Heryanto and A. Albert, "Implementasi Sistem Database Terdistribusi Dengan Metode Multi-Master Database Replication," J. Media Inform. Budidarma, vol. 3, no. 1, p. 30, 2019, doi: 10.30865/mib.v3i1.1098.

[30] A. Heryanto and A. Albert, "Implementasi Sistem Database Terdistribusi Dengan Metode Multi-Master Database Replication," J. Media Inform. Budidarma, vol. 3, no. 1, p. 30, 2019, doi: 10.30865/mib.v3i1.1098.