

Data Warehouse Modeling Using Online Analytical Processing Approach

Galih Budiarto

Informatics Engineering, STMIK Tasikmalaya, Indonesian
galih.budiarto13@gmail.com

Abstract

Keywords: Analysis; Information; Online; OLAP; Technology;	Increasing business needs affect business competition in many companies that utilize information technology. The competition aims to gain advantages that utilize information technology in facilitating business processes. One of the conveniences offered is the use of information technology to support decision making in carrying out existing business processes in middle-to-upper scale companies that have large amounts of data. Online Analysis Processing (OLAP) is an approach method to present answers to the demand for a dimensional analysis process quickly, namely the design of applications and technologies that can collect, store, manipulate multidimensional data for analyst purposes. Information delivery generated from Online Analysis Processing (OLAP) can help executive information systems in generating sales reports because it has a response time of 0.0038769999519 seconds.
--	--

1.INTRODUCING

Increasing business needs affect business competition in many companies that utilize information technology. The competition aims to gain advantages that utilize information technology in facilitating business processes[1], [2]. One of the conveniences offered is the use of information technology to support decision making in carrying out existing business processes in middle-to-upper scale companies that have large amounts of data. The success factor of a company is its ability to analyze the market well. The behavior of consumers can be able to be captured well by the company in this case, of course, managerial parties to be evaluated, analyzed so as to produce a strategic policy. Evaluation, planning, and decision-making activities will be carried out better if an organization or company has completed, fast, precise, and accurate information.

Sales transaction data in the sales division related to goods sales data is so large and numerous that it requires large storage media and it is difficult to process these data for making reports used by management for decision making. Processing sales data to produce reports takes a long time because it is still processed with the help of Microsoft Excel applications. The amount of transactional sales data stored cannot be used as a basis for analysis before further data processing is carried out to be used as a basis by the top-level management in making a decision.

Online Analysis Processing (OLAP) is an approach method to present answers to the demand for a dimensional analysis process quickly, namely the design of applications and technologies that can collect, store, manipulate multidimensional data for analyst purposes[3]-[5]. OLAP is part of a more global category of business thinking, which also encapsulates the relationship between reporting and data mining. An OLAP database has its own schema structure and is usually a data warehouse. Data warehouse is a collection of data from various sources that are stored in a data warehouse (repository) in a large capacity and used for the decision-making process. However, it is possible that OLAP takes from this operational database with the record that this database already has an OLAP friendly design structure.

The use of data warehouse is expected to be able to assist the marketing division, sales division, and finance division in managing sales transaction data so that the decision-making process can be carried out quickly and precisely by the executive information system. Using OLAP can make it easier for

stakeholders to make decisions based on existing transactional data because OLAP will display data from various sides or multidimensional[3], [6], [7].

2. RESEARCH METHOD

The concept framework in this study, we first study current conditions, problem formulation, data collection, technology analysis, design, prototype, testing, and prototype results[8]-[10]. The conceptual framework of this research can be seen at Figure 1.

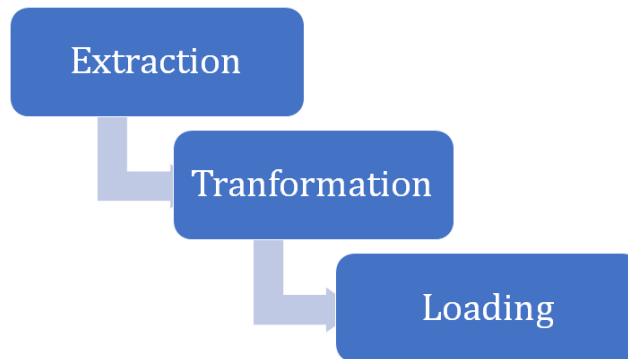


Figure 1. Research of Stage

ETL is a collection of processes for preparing data from operational sources to data. This process consists of extraction, transformation, loading, and several processes that are carried out before being published in the data warehouse. The ETL process is the phase of processing data from incoming data sources in a data warehouse. The purpose of ETL is to collect, filter, process and combine relevant data from various sources for storage in a data warehouse. ETL can also be used to integrate data. With a pre-existing system. The result of the ETL process is the production of data that meets data warehouse criteria such as historical, integrated, summarized, static data and has a structure designed for the purposes of the analysis process.

3. RESULTS AND DISCUSSION

Extraction

The sales database on CV AL is in the form of an excel file. Data used for staging data and data warehouse will use MySQL. The reason for using this database platform was chosen with consideration because the development and license currently owned is opensource. Users who can access the data warehouse system are users who are connected to the web server. Operational data stored in the database can be provided in the form of tables based on research needs on sales data on the basis of company confidentiality.

Data extraction from sales operational database in MySQL. With different platforms, changes to the data structure side are needed. The stages of the extract are carried out from the data source used through the process of selecting data which is then stored on the database. Database placement is placed on the same database storage, machine and platform as the data warehouse so that the ETL process can be done faster and does not interfere with operational processes. This extraction process takes the necessary sources in the form of data on goods, customers, drivers, districts, provinces, and sales to then create a database for sales warehouse data.

The data cleansing stage is carried out on the data because this study uses a data source that does not have a uniform data format used. The data was cleansed and organized consistently according to the needs in this study.

Transformation

The transformation stage in principle changes the form of data from data that has been extracted into a form that suits user needs. This process is carried out after the existing data has passed the extraction and cleaning process. The transformation process carried out is divided by level, namely record level and field level. In this process, the process of selection, merging and aggregation is carried

out to obtain summary data in accordance with the dimensions to be created. The transformation process takes time in creating procedures in SQL. In the sales table, the transformation process is carried out to take tables and de-normalize starting from customers, goods, drivers, districts, provinces and sales.

In this transformation stage, the function process and field level are carried out, namely:

- a. Map input data from the original data schema to the data warehouse schema, namely in the form of goods data to be transformed in `dim_barang`, customer data to be transformed in `dim_pelanggan`, driver data to be transformed in `dim_sopir`, district data to be transformed in `dim_kabupaten`, provincial data to be transformed in `dim_provinsi`, and sales data to be transformed in `fact_penjualan`.
- b. Converting data types or data formats, namely the type that is converted to data types with date, int, decimal types. Because in the previous data all data type formats are the same from one another.
- c. Cleaning and disposal of duplication and data errors, namely cleaning data that has the potential to cause redundancies during the loading process. This data cleaning process is checked against the value that will be the primary key in the dimension and fact table in the data warehouse.
- d. Calculation of derivative values or first, namely carrying out the process of adding calculations to the total price and total pay in the transaction data.
- e. Data merge is combining transaction data that occurs within 1 year.

Loading

Loading is moving OLTP data in a data warehouse. There are two data loading that we do in the data warehouse. First is the initial load, this process is done when we have finished designing and building the data warehouse. The data we enter will certainly be very large and take a relatively longer time. In the initial load process, enter data in the dimension table, namely `dim_barang`, `dim_pelanggan`, `dim_sopir`, `dim_provinsi`, `dim_kabupaten`, and `fact_penjualan` for processes that occur for a period of 1 year. Incremental loading, performed when the datawarehouse has been operated. We do data extraction, transformation, and loading of the data. For the initial load, it is done only once, but for incremental load we can schedule according to needs, it can be every day, month, quarter or year according to the needs of the transaction process that occurs in the company.

Star Schema

The star schema created from this prototype uses a data store dimension consisting of 5 (five) dimensions, namely `dimbarang`, `dimcustomer`, `dimdriver`, `dimkabupaten`, and `dimprovincial`. This star schema has 1 (one) fact, namely `factsales`. The design of the data warehouse star schema can be seen in figure 2 below.

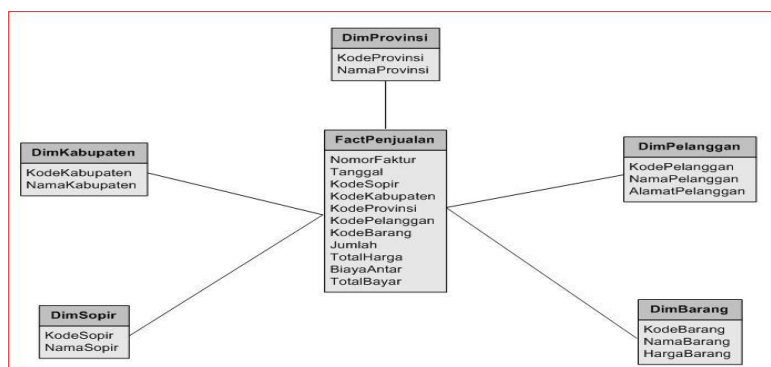


Figure 2. Star Schema

- a. `Dimbarang` Has an attribute of the item code which becomes the primary key, itemname, and price.
- b. `Dimdriver` Has an attribute of the driver code that becomes the primary key, and the driver's name.

- c. Dimcustomer Has a customer code attribute that becomes the primary key, customer name, and customer's address.
- d. Dimkabupaten Has an attribute of the district code which is the primary key, and the name of the regency.
- e. Dimprovinsi: Has an attributte, the provincial code, which is the primary key, and the name of the province.
- f. Factsales Has attributes invoice number, date, item code, driver code, customer code, district code, provincial code, amount, total payment, delivery cost, and total price.

OLAP Monthly Goods Sales Report

The first step at this stage displays the reports generated by the data warehouse. Here is what the monthly item sales report looks like.

Kode Barang	Jumlah Bel	Sub Total
H-0001	19	1543750
H-0002	13	1056250
H-0003	25	2031250
H-0004	20	1375000
H-0005	5	343750
H-0006	5	343750
H-0007	28	1925000
H-0008	24	1650000

Figure 3. Display of Monthly Item Sales Report

The next step in the monthly item sales report is to design the hypercubes report that has been described and efficiency will be carried out in each hypercubes, where the dimension table will be reduced and unified into a fact table with the aim of improving query performance where reducing the number of tables in the process of joining in the query will improve query performance.

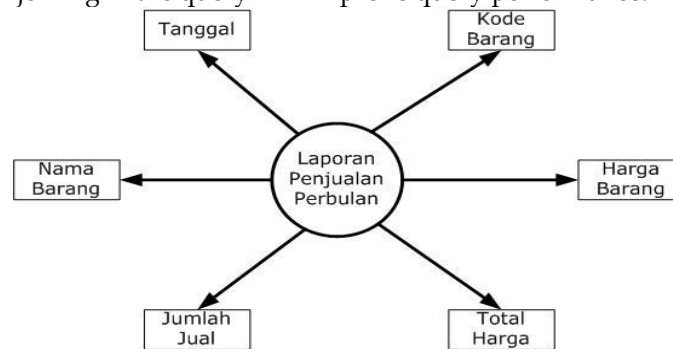


Figure 4. Hypercube of Monthly Item Sales Report

From the picture above, the hypercube design consists of date fields, item code, item name, item price, number of sales, and total for monthly item sales reports. Based on the design of hypercubes reports on monthly sales of goods, 1 fact table and 1dimension table will be formed. From the fact table fact_penjualan only take the item code, quantity, and total price fields. While the dimension table only takes the fields of the name of the goods and the price of goods, therefore 1 fact table and 1dimension table will be formed. Facts and dimension tables can be seen in the following figure.

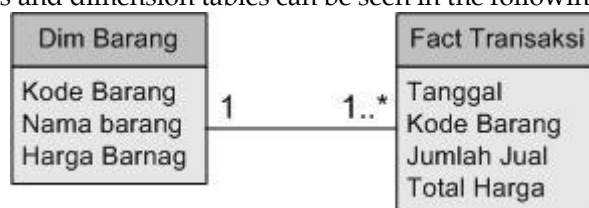


Figure 5. Table Design of Monthly Item Sales Report

OLAP Report on Sales of Goods by Region

The first step at this stage displays the reports generated by the data warehouse. Here's what the item sales report looks like by region.

Kode Barang	Jumlah Beli	Sub Total
H-0001	19	1543750
H-0002	13	1056250
H-0003	25	2031250
H-0004	20	1375000
H-0005	5	343750
H-0006	5	343750
H-0007	28	1925000
H-0008	24	1650000

Figure 6. Display of Sales Report View by Region

The next step in the report on sales of goods by region is to design hypercubes reports that have been described and efficiency will be carried out in each hypercubes, where the dimension table will be reduced and unified in a fact table with the aim of improving query performance where reducing the number of tables in the process of joining in the query will improve query performance.

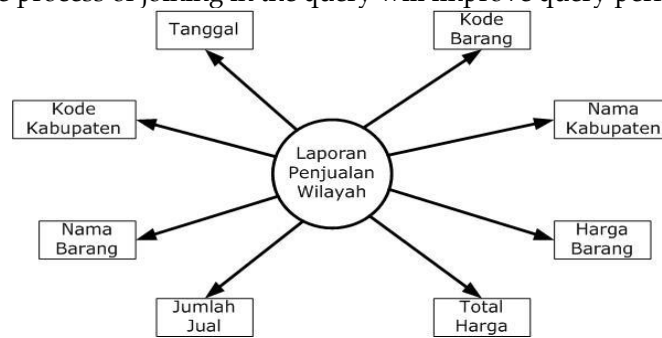


Figure 7. Hypercube of Sales Report by Region

Based on the design of hypercubes sales reports by region, 1 fact table and 2 dimension tables will be formed. From the fact table fact_penjualan only take the item code, quantity, and total price fields. While from the item dimension table only take the fields of the name of the goods and the price of the goods. The district dimension table takes the district code field, and the name of the district, therefore it will form 1 fact table and 2 dimension tables. Facts and dimension tables can be seen in the following figure.



Figure 8. Table Design of Sales Report by Region

Testing the Response Time of OLAP Prototype Results

This stage is tested on the prototype that has been made, with the purpose of this test is to test the response time of the hypothesis that has been made in the previous chapter to see whether the results of prototype testing can solve the problem. The results of this test by looking at how long it takes to display sales reports on CV AL so as to help the executive information system in producing sales reports. The following is a prototype display in displaying the goods sales report.

HOME	BARANG	PROVINSI	KABUPATEN	SOPIR	PELANGGAN	TRANSAKSI	GRAFIK	LAPORAN
P-002			8			600000		
TL-001			14			997500		
TL-002			12			510000		
TL-003			23			977500		
TL-004			18			765000		
TL-005			12			579600		
TL-006			8			386400		

Bulan Adalah : 1 Tahun Adalah : 2010

Time Response Adalah : 0.0038769999519

Figure 9. Goods Sales Report Response Time Test Results

Based on the picture above, it can be seen that the response time to display the report on the sales data report on the prototype of OLAP Development Sales Data on CV AL has a response time of 0.0038769999519 seconds. The test results show that the hypothesis of this research is correct, namely that with the development of OLAP, it will make it easier for executive information systems to produce sales data reports faster with a response time of less than 10 seconds to display goods sales reports.

This stage is carried out testing the prototype that has been made, with the purpose of this test is to test hypothetical items that have been made in the previous chapter to see whether the results of prototype testing can solve the problem of response time in displaying sales graphs. The results of this test by looking at how long it takes to display a graph of sales of goods on the CV AL so as to help the executive information system in decision making. The following is a prototype display in displaying a graph of sales of goods.

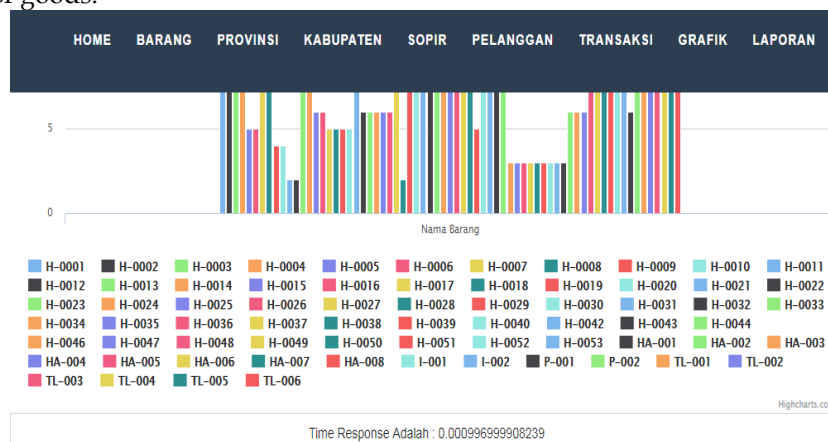


Figure 10. Goods Sales Report Response Time Test Results

Based on the picture above, it can be seen that the response time to display the item sales data report on the prototype of OLAP Development Sales Data on CV AL has a response time of 0.0009969999908239 seconds. The test results show that the hypothesis of this research is correct, namely that with the development of the OLAP prototype, it will make it easier for the executive information system to produce sales data graphs faster with a response time of less than 10 seconds to display sales graphs for decision making based on graph-shaped reports generated from faster prototypes.

4.CONCLUSION

The implementation of this prototype has hardware and software needs that are in accordance with those in non-functional needs, and for the response time of this prototype requires a different time for each query execution in displaying graphs and generating sales reports on CV AL. But the response time of each query executed is no more than 10 seconds to produce graphs and reports on sales of good.

The use of this prototype helps the executive information system in producing sales reports because it has a response time of 0.0038769999519 seconds. So that it can help generate sales data reports faster. The use of prototypes in displaying sales charts of executive information systems in decision making because the process of displaying graphs does not take long, only less than 10 seconds to produce sales graphs of goods. For data warehouse development, you can then add purchase data, and inventory data on CV AL to help companies in processing more complex sales data.

5. REFERENCES

- [1] R. Putra Setiawan, "Sistem Informasi Manajemen Presensi Siswa Berbasis Mobile Study Kasus Sma N 1 Sungkai Utara Lampung Utara," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 119-124, 2021.
- [2] H. Sulistiani, M. Miswanto, D. Alita, and P. Dellia, "Pemanfaatan Analisis Biaya Dan Manfaat Dalam Perhitungan Kelayakan Investasi Teknologi Informasi," *Eduatic-Scientific J. Informatics Educ.*, vol. 6, no. 2, 2020.
- [3] D. Subuh and W. Yasman, "Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho," *Pros. SENIATI*, pp. 29-36, 2019.
- [4] B. Santoso, "Perancangan aplikasi olap (online analytical processing) penjualan buku pada toko buku gramedia lubuklinggau dengan metode clustering," *J. Teknol. Inf. MURA*, vol. 9, no. 2, pp. 98-107, 2017.
- [5] A. Syam and A. R. Manga, "Sistem Tracer Study Alumni Fakultas Ilmu Komputer Universitas Muslim Indonesia Menggunakan Metode On-Line Analytical Processing (OLAP)," *Ilk. J. Ilm.*, vol. 9, no. 1, pp. 86-90, 2017.
- [6] M. A. A. Maulana, M. S.Kom, M.Kom., and W. Soeharso, "Analisis dan perancangan data warehouse evaluasi mahasiswa studi kasus pada jurusan Teknik Informatika UMM," *J. Repos.*, vol. 1, no. 1, p. 59, 2020, doi: 10.22219/repositor.v1i1.376.
- [7] M. Akbar and Y. Rahmanto, "Desain data warehouse penjualan menggunakan Nine Step Methodology untuk business intelegency pada PT Bangun Mitra Makmur," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 137-146, 2020.
- [8] S. Setiawansyah, H. Sulistiani, and V. H. Saputra, "Penerapan Codeigniter Dalam Pengembangan Sistem Pembelajaran Dalam Jaringan Di SMK 7 Bandar Lampung," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, pp. 89-95, 2020.
- [9] F. Irvansyah, S. Setiawansyah, and M. Muhaqiqin, "Aplikasi Pemesanan Jasa Cukur Rambut Berbasis Android," *J. Ilm. Infrastruktur Teknol. Inf.*, vol. 1, no. 1, pp. 26-32, 2020.
- [10] S. Setiawansyah, Q. J. Adrian, and R. N. Devija, "Penerapan Sistem Informasi Administrasi Perpustakaan Menggunakan Model Desain User Experience," *J. Manaj. Inform.*, vol. 11, no. 1, pp. 24-36, 2021.