

Data warehouse using Kimball approach in computer maniac

Kelvin Salim*, Lorita Damayanti, Marisya Puspita, Steven Liujaya, Abba Suganda Girsang

Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

*kelvin.salim@binus.ac.id

Abstract. It is difficult for small businesses to evolve without utilizing an adequate system which is able to manage information within the business entity efficiently. The reporting system commonly used utilizes query which consumes a large amount of time in order to generate results fulfilling requirements set by the user. In order to simplify and speed up reporting in digital world and eventually results in improvement of business, the business entity needs a system which can address problems quickly and easily. In this paper, data warehouse system which could simplify and speed up the reporting provisions that is needed by the business without having to modify the existing query being implemented. With Kimball approaches, the data to be reported could be customized in various forms in accordance with the requirements set by the entity.

1. Introduction

Computer Maniac manages its transactions as sales and purchase transactions. Purchase transaction is defined as a transaction that takes place when staff restocks a product from supplier. Sales transaction is defined as a transaction that occurs when customer buys a product being handled by the staff. After all processes related to restocking of goods and sales of goods are complete, the related transaction is then inputted into the system. After goods are received from the supplier by warehouse staff, the type and quantity of the goods, as well as all other related data pertaining to the purchase are then inputted to the system. Sales transactions are recorded as sales are made through POS system.

In order to have the data which are more structured and easier to be analyzed, a system which can analyze data quickly and efficiently is required. It is possible to use a query to retrieve the data; however it might consume a large amount of time in its process of analyzing the data. For the growing business, the amount of data to be analyzed will increase as time passes and more data is inputted. Implementing a data warehouse system is a way to make the data analysis becomes more efficient.

In this paper, a method to implement the data warehouse [1] for better analysis and creating report within the company requirement is needed. The method should be easy to understand and efficient. The Kimball life cycle [1], [2] method is chosen as it uses a bottom-up approach which is faster to implement.

2. Related Work

1. Data Warehouse

Data warehouse [1], [3] is a technology that aims at enabling the decision maker to make better and faster decisions [4]. Data Warehouse is a data structure that is optimized for distribution, mass storage and complex query processing [3]. In order to build a Data Warehouse [1], [3], it is required to run ETL tools which has three tasks: (1) extracting data from different data sources, (2) transforming data, (3) loading transformed data to data warehouse. There are two components in ETL tools, one component extracts raw data from different data sources (flat files, excel or csv files, web services, relational tables) and another component loads the data into the staging database, then cleans and transforms the extracted raw data and loads it into facts and dimension tables [5]. There are two phases involved in the extraction process: (1) initial extraction, (2) changed data extraction. The extraction process is only executed once after building data warehouse in order to fill it up with a huge amount of data from sources. In transforming process, all data are being cleaned and confirmed so that the data gained is correct, complete, consistent and unambiguous. The process includes data cleaning, transforming and integration. In loading process, all data that has been transformed will be loaded to multidimensional structure.

2. Business Intelligence

Business Intelligence (BI) is defined as the transformation of information into knowledge and it has the ability to provide the right information to the right user at the right time to support the decision-making process [6]. Business intelligence systems are very complex and expensive to design and implement [7]. The complexity and importance of BI system development necessitates a critical approach to successfully develop technically appropriate as well as usable (people-oriented) BI systems that meet user needs [7]. Some approach of Business Intelligence includes the Kimball lifecycle approach [2], Inmon's Corporate Information Factory [8] and Linstedts' data vault model [9]. When designing a new BI system, business users often restrict themselves within the performance limitations of their current (known) systems; hence, they only utilize current information and fail to explore new and improved key performance indicators that can enhance their decisions [7].

3. Proposed Method

In this paper, Kimball methodology that has nine steps for designing the data warehouse is used:

Choose Business Process and Analysis. In this step, the required tables which contain the data are chosen and transformed into the dimension tables. In this case there are eight tables such as Product, ProductType, Brand, Staff, Customer, Supplier, PurchaseDetail and SalesDetail table.

Choose the Granularity. In this step, the granularity or the relationship between data and information within transaction table and fact table is analyzed. In this case, the grain is number of sales quantities within a year based on customers region.

Create Dimension Tables. In this step, the dimension tables that are related to the fact tables are starting to be created. In this case seven dimension tables such as CustomerDim, StaffDim, ProductDim, ProductDim, ProductTypeDim, BrandDim, DateDim and SupplierDim are created.

Create Fact Tables. In this step, fact tables which contain measurable data are created. In this case there are 2 fact tables such as SalesFact, PurchaseFact tables.

Storing Pre-calculation in the Fact Table. In this step, after the fact table is formed, the function of each field that serves as a measurement field are determined.

Rounding out the dimension table. In this step, the dimension table to represent the attribute hierarchy to ease the analysis is determined.

Choosing the duration of the database. In this step, the duration for the data that will be presented is chosen so that the data behaviour is analyzable. In this case the data chosen is of 1 years backward.

Slowly changing dimension. In this step, it is made sure that dimension tables are not affected by transaction changes.

After all dimension tables and fact table are created, all dimension tables are connected with the fact table. A technique that can be used for this is Star Scheme [1]. Star Scheme [1] consists of a central data table, or fact table connected to one or more dimension tables. It is called a star scheme because this model resembles a star, with points centered from the center. The center of the star scheme consists of one or more fact tables and the points in the schema are dimension tables that contain information on certain attributes in the fact table [10].

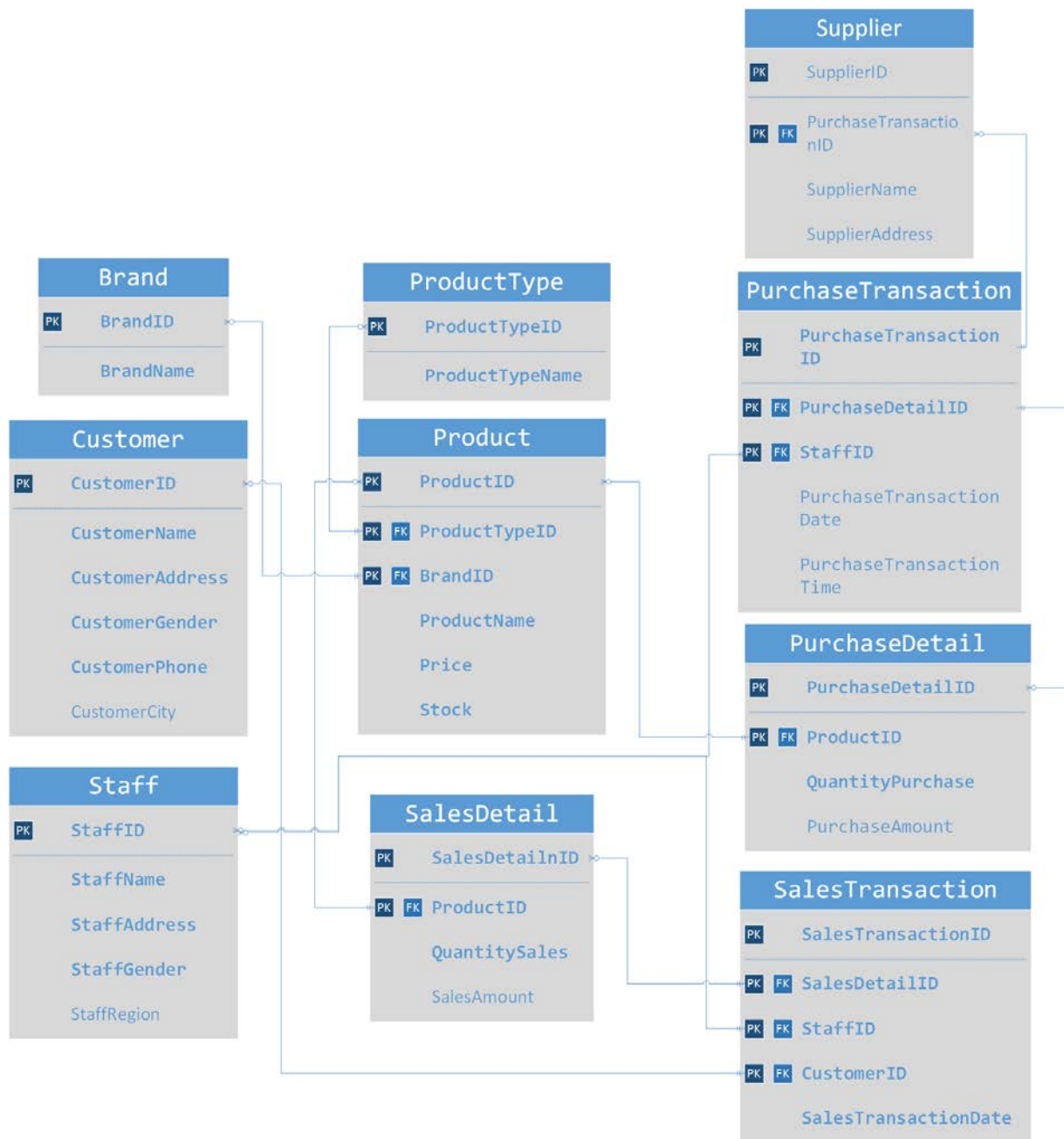


Figure 1. Entity Relational Diagram.

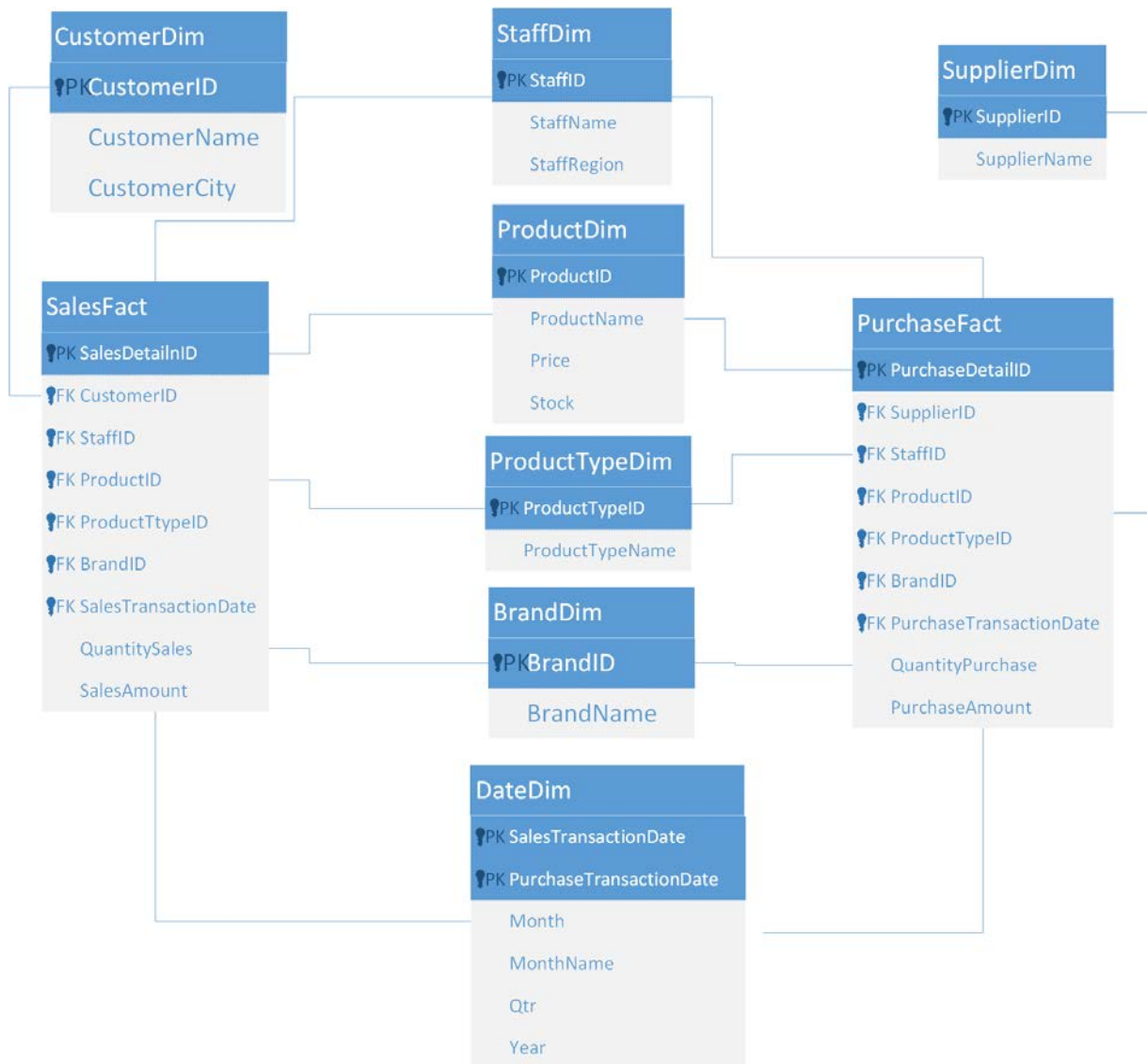


Figure 2. Star Scheme.

After creating all dimension tables and the fact tables, the data is stored using ETL process tool like Pentaho Data Integration. The ETL process is shown on Figure 3 – 11.



Figure 3. Load BrandDim table.



Figure 4. Load CustomerDim table.

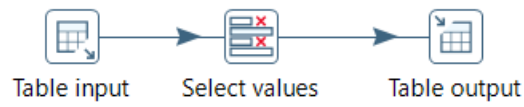


Figure 5. Load ProductDim table.



Figure 6. Load ProductTypeDim table.

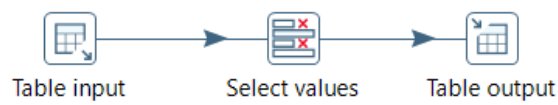


Figure 7. Load StaffDim table.

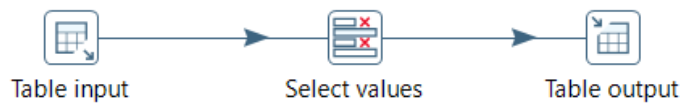


Figure 8. Load SupplierDim table.

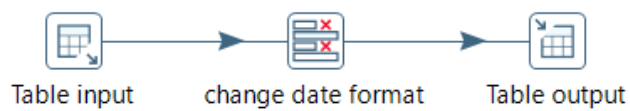


Figure 9. Load DateDim table.

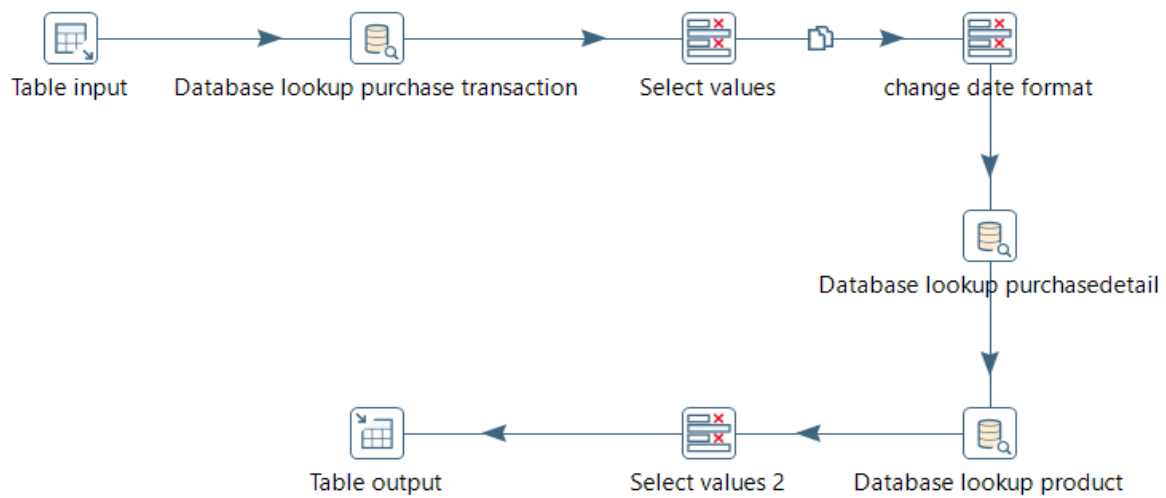


Figure 10. Load PurchaseFact table.

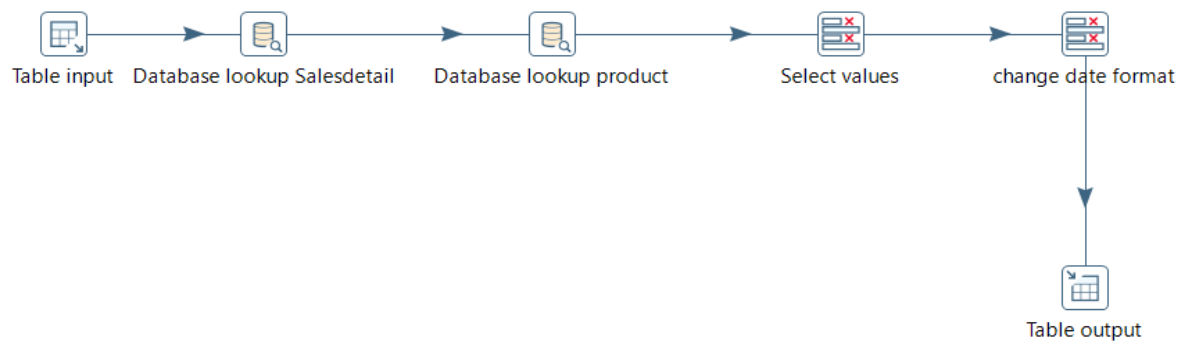


Figure 11. Load SalesFact table.

4. Analysis Results

After the ETL process is finished, all the dimension tables and the fact tables are filled with data which can be used as data warehouse source. In this section, the report dashboard is made using Qlik Sense. The result is shown on Figure 12 – 14.

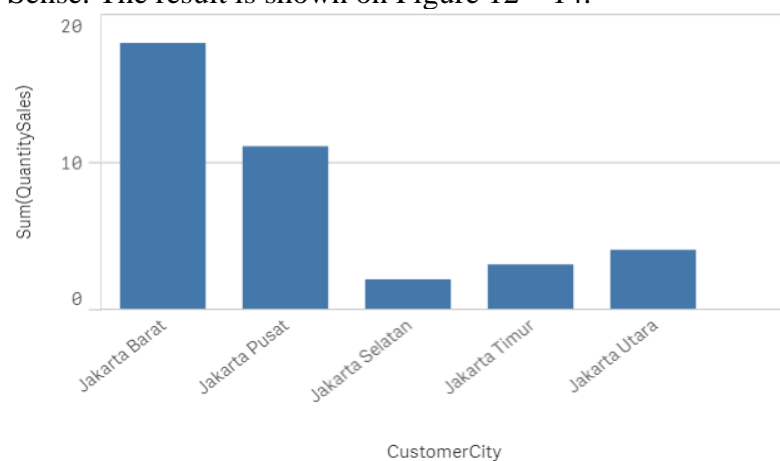


Figure 12. Chart of quantity sales for every customers region.

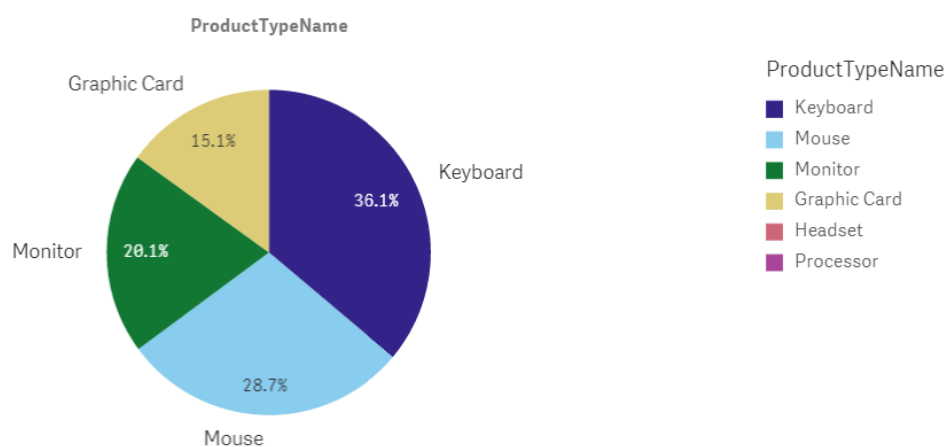


Figure 13. Round chart of product type that sold the most.

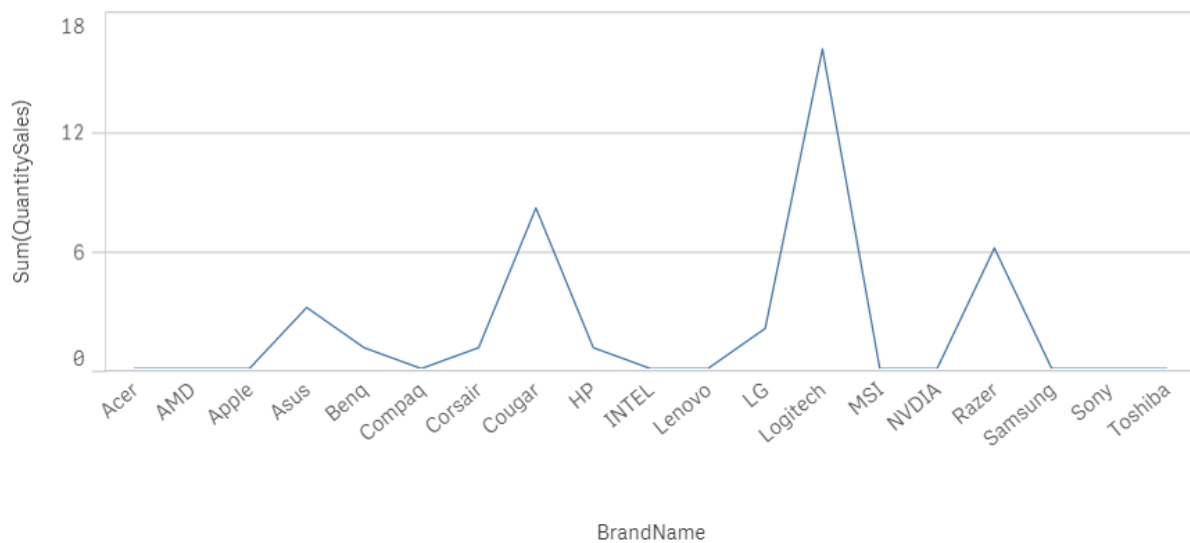


Figure 14. Chart of brand that sold the most.

There also another tool to make such report such as Pentaho Report Designer. The result can be seen on Figure 15.

Sales Report Group By Customer City

All

CustomerCity	Qty	SalesAmount
Jakarta Barat	8	4,820,000
Jakarta Pusat	6	4,750,000
Jakarta Selatan	1	200,000
Jakarta Timur	3	10,360,000
Jakarta Utara	2	1,100,000
Total	20	21,230,000

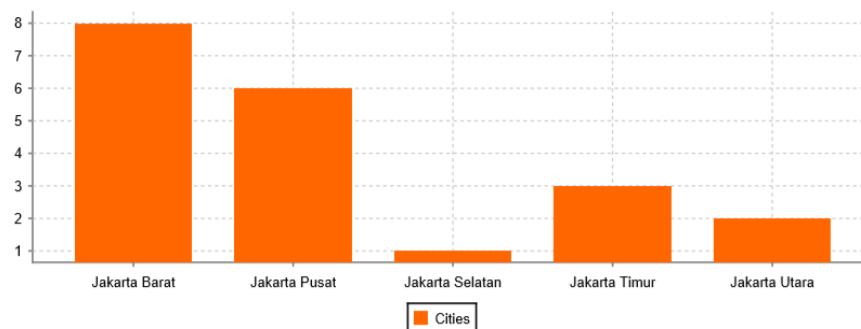


Figure 15. Report result using Pentaho Report Design

5. Conclusion

It is possible to make data easy to analyse and report faster to be made with a data warehouse system without the need to type a long query. All related table are processed by ETL process which processes all data needed in the data warehouse and separates all unnecessary data. As the result, utilizing the process, it is possible for staff to prepare report which fulfils company requirement within shorter period of time. In the future, can make it easier and more complex reporting system than what already exists and make a reporting system that is better and more efficient than now.

References

- [1] G. M and S. Rizzi, "From Star Schemas to Big Data: 20+ Years of Data Warehouse Research," *J. Biol. Chem.*, pp. 241–252, 2018.
- [2] R. Kimball and M. Ross, *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. Indianapolis, Wiley, 2010.
- [3] T. R. Sahama and P. R. Croll, "A Data Warehouse Architecture for Clinical Data Warehousing," *Russ. J. Gen. Chem.*, vol. 78, no. 11, pp. 2214–2219, 2008.
- [4] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. El Bastawissy, "A proposed model for data warehouse ETL processes," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 2, pp. 91–104, 2011.
- [5] S. Habte, K. Ouazzane, P. Patel, and S. Patel, "Generic Data Warehousing for Consumer Electronics Retail Industry," *Int. J. Comput. Inf. Eng.*, vol. 11, no. 7, pp. 828–831, 2017.
- [6] A. Brandao, E. Pereira, F. Portela, M. Santos, A. Abelha, and J. Machado, "Real-time Business Intelligence platform to maternity care," in *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, 2014, pp. 379–384.
- [7] C. Venter and R. Goede, "Critical systems approach to business intelligence systems development," *Proc. 59th Annu. Meet. ISSS-2015 Berlin Ger.*, pp. 1–18, 2016.
- [8] W. H. Inmon, C. Imhoff, and R. Sousa, *Corporate information factory*. New York, Wiley, 2001.
- [9] D. Linstedt, "Method and system of data warehousing and building business intelligence using a data storage model," vol. 1, no. 19, 2002.
- [10] S. M. Isa, E. C. Nugroho, D. Y. Gunarso, C. B. Process, K. C. Susena, and A. S. Girsang, "Business Intelligence for Analyzing Department Unit Performance in eProcurement System."