

Designing a Star Schema for Optimizing the Total Sales of Motorcycles

Eka Praja Wiyata Mandala^{1✉}, Randy Permana², Dewi Eka Putri³

^{1,2,3}University of Putra Indonesia YPTK Padang

ekaprajawm@upiypk.ac.id

Abstract

Motorcycle sales have increased significantly, motorcycle manufacturers are competing to produce the latest models which are then sold to consumers. As a result, motorcycle dealers are overwhelmed with more and more data, not knowing what to do with it. Motorcycle dealers also have difficulty calculating the total sales of motorcycles. We try to provide solutions to deal with data overflow. We propose designing a star schema as the basis for creating a data warehouse. To create a star schema, we propose a four-step sequence in creating an effective star schema, starting from requirements analysis and reporting, understanding business processes, connecting and matching business processes with suitable entities and determining the dimensions of the business processes. We get a star schema with 1 fact table, motorcycle_sales and 11 of dimension tables, such as brand, color, customer, customer_contract, distributor, district, motorcycle, repair_workshop, sell_location, type and time. The star schema is an optimized model that provides the best performance in presenting more complex information.

Keywords: motorcycle sales, big data, data warehouse, star schema, fact table, dimension table.

1. Introduction

The increase in motorcycles in Indonesia every year is very significant, so motorcycle dealers compete in selling their products. Each motorcycle dealer provides advantages for each product offered to consumers, thereby increasing market competition in motorcycle sales [1]. Motorcycles are a necessity in increasing economic development, one of which is in agriculture which is used as a practical transportation [2].

Motorcycle sales have great potential in increasing competition in the production of various motorcycle variants that can attract consumer interest [3]. Motorcycle sales in Indonesia have increased in the last ten years. In 2010, motorcycle sales reached 61 million units, increased to 92 million in 2014, then increased by 52% in the last five years [4].

High sales of motorcycles, creates a large and abundant volume of data, while data should not be discarded or deleted, because it will be needed for decision making. For motorcycle dealers who have been operating for more than ten years, this will be a big problem, because data has accumulated and is abundant. So it is necessary to find a solution to handle the abundant data so that the data can be used for decision making. One of the decisions that can be made from this abundant data is to be able to see an increase in total sales every day, week, month and even year.

To optimize the total sales of motorcycles, one solution that can be done is to collect historical data on motorcycle sales. The historical data that has been collected will become big data. Big data is a new paradigm that has resulted in a change in performing

data analysis using well-known schemas used to extract hidden patterns [5]. Big data analytics is the process of using analytical algorithms to uncover hidden patterns or unknown correlations [6]. Big Data refers to a complex problem and it is important to extract or retrieve information because the analysis process requires large computing power [7].

One technique in processing big data is a data warehouse. Data warehouse is a relational database to perform the query process to assist the analysis and reporting process. Using a data warehouse is more efficient than ordinary queries because it can process data into information which is then made into a report without requiring a lot of time [8]. Data warehouse processes have helped organizations provide a foundation for high-value information investments [9]. Data warehouse is used for data analysis such as identifying future trends or discovering new knowledge [10] and has been applied in the organization and processing of data and provides support to decision makers [11]. Data warehouse is a repository of various data sources organized under a unified schema that is useful for guiding decision making [12].

To build a data warehouse, an OLAP (On-Line Analytical Processing) approach is needed. OLAP can provide information access quickly and precisely because it can provide real-time access. OLAP displays data based on the time dimension and several other related dimension tables [13] that facilitates the consultation and aggregation of a given data set [14]. OLAP generates queries by traversing the data warehouse [15]. OLAP is a smart interface to avoid writing wrong SQL queries so that it can return data

properly [16]. OLAP consists of facts, dimensions, measures and dimensional hierarchies, which are used to represent schema [17].

OLAP works with multiple dimensions, so we need a scheme to design the OLAP. The schema used is a star schema. Designing a star schema takes a long time because it includes several tasks such as denormalizing, defining dimension tables and defining fact tables [18]. In designing a star schema, there are several steps and analysis of business needs to turn it into a dimension table [19].

In this paper, we focus on designing a star scheme for optimization of total motorcycle sales. We use data from one of the motorcycle dealers in Padang. The data we use is historical data on motorcycle sales for the last

ten years. The data warehousing process goes through three steps, known as extraction, transformation and loading (ETL) [20]. The data we get, we do the ETL process first. We extract data from several data sources, we transform data to get the same data format and we load data into the data warehouse.

2. Research Method

Star schema is a model that manages data in a multi-dimensional form. The multidimensional model will form data based on the characteristics of the data so that it will produce data groups (consisting of more specific dimensions and facts). To produce a star schema, dimensions and facts can be generated by analyzing the tables contained in the relational database can be seen in Figure 1.

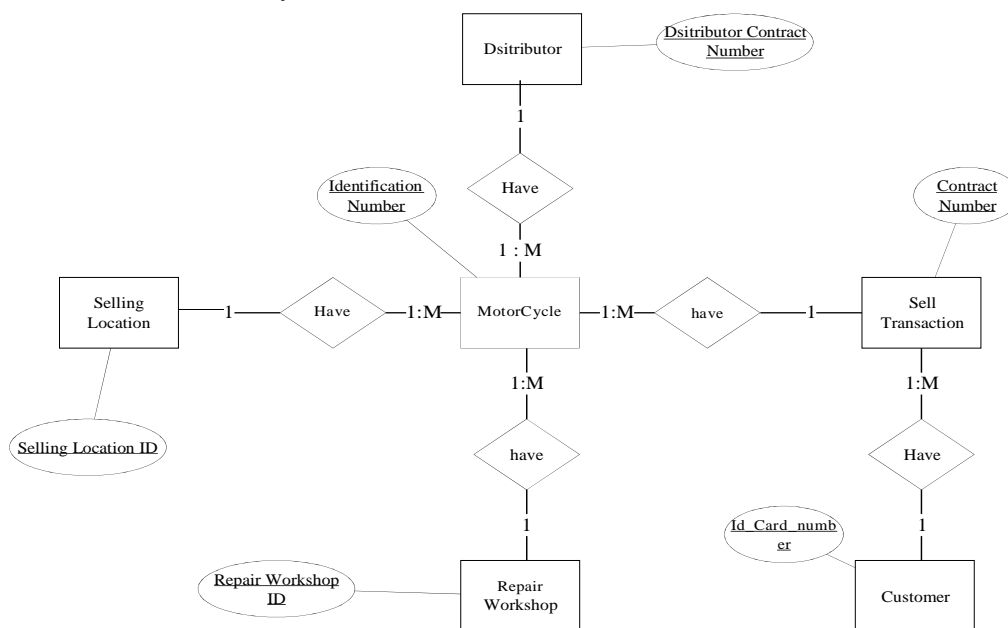


Figure 1. Relational Database of Motorcycles

Figure 1 illustrates a relational database consisting of 6 model in Figure 1, we propose a sequence of processes that can be carried out for the process of transforming the relational model into a star schema. The sequence of processes that we propose can be seen in Figure 2.

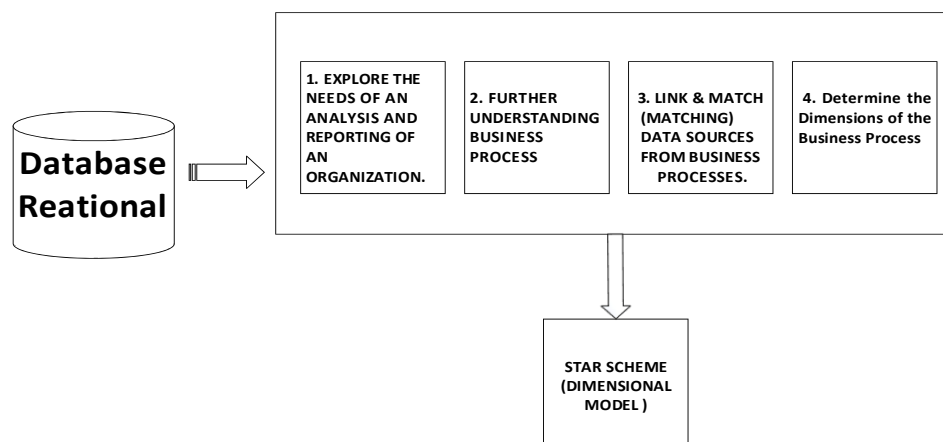


Figure 1. The Sequences of Processes that We Propose

From Figure 2, there are 4 process sequences that we propose.

3. Result and Discussion

We propose a sequence of processes for creating a star schema with 4 stages to complete. These stages must be sequential from beginning to end until a star schema is formed.

Step 1: Explore the needs of an analysis and reporting organization

Identification activities can be obtained through interviews with company owners or parties involved in decision-making. In addition, the identification process can also be done by analyzing and understanding the company's relational database.

For the case motorcycle sales, there are several needs analyses and reports that are written into the needs analysis and reporting form can be seen in Table 1.

Table 1. The Needs Analysis and Reporting

Type of Need	Analysis	Reporting
Increase in total sales of motorcycles	Motorcycle Sales Data	The sum of the total sales of motorcycles based on a certain time, selling location, customer, distributors and vehicle
The average of motorcycles sales	Motorcycle Sales Data	The average of the sales of motorcycles based on a certain time, and collaboration with location, customer, distributors and vehicle
Number of motorcycles repaired by the dealer's own workshop	Motorcycle Sales Data, Vehicle and Repair Workshop	The sum of motorcycle repaired by dealer's own workshop on a certain time
Number of motorcycles that supplied by distributor	Motorcycle Sales Data, Vehicle and Distributor	The sum of motorcycle that supplied by distributor on a certain time.

Step 2: Further understanding business process

This stage is carried out to ensure the suitability of the results of the demand prediction with the process business going on in the company. By understanding the business processes (transactions) that occur, a star schema designer can get some requirements that may not have been documented in the first process. The business process can be seen in Figure 3.

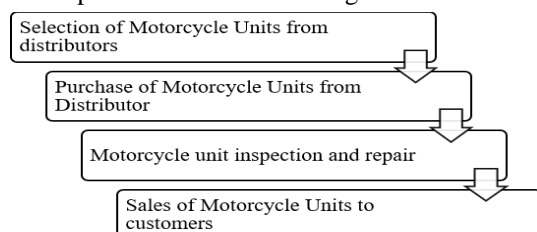


Figure 3. The Business Process

From Figure 3, there are 4 sequences of business processes. Here are the steps:

- Selection of Motorcycle Units from distributors. Since the company sells used motorcycle units, the company needs to make a selection of motorcycle units contained in the distributor. This process has an important role in achieving company profits, where the selected motorcycle unit is a unit in very good condition, or if it has damaged it can be repaired at a low cost.
- Purchase of Motorcycle Units from Distributor. After selecting the motorcycle unit, the next step of this business process lies in purchasing the unit from the distributor. Since the units are purchased collectively, the company implements a policy to make payments a few days after the motorcycle units are at the dealer
- Motorcycle unit inspection and repair. In this process, the dealer will make repairs to the motorcycle units that need to be repaired. This process will determine where the advantages or disadvantages of the company where repair costs will be considered. The large repair costs will cause the difference in profits obtained from the selling price of the vehicle to customer.
- Sales of Motorcycle Units to customers. Sales of motorcycle units to customers, which in this process will be made by purchasing in cash or credit. For purchases on credit, a contract agreement will be made between the buyer and the dealer partner.

Step 3: Link and match data sources from business processes

This process is related to the adjustment of business processes that have been identified with the entities contained in the ERD diagram can be seen in Table 2.

Table 2. The Business Process Related with Suitable Entity

Business Process	Suitable Entity
Selection of Motorcycle Units from distributors	Motorcycle, Distributor
Purchase of Motorcycle Units from Distributor	Motorcycle, Distributor
Motorcycle unit inspection and repair	Motorcycle, Repair Workshop
Sales of Motorcycle Units to customers	Motorcycle, Customer, Sell Transaction, Selling Location

Step 4: Determine the dimensions of the business process

After understanding the business process more deeply and knowing the data related to the process, the next

step is to start conceptualizing the dimensions of the business process. The star schema consists of several dimensions and consists of at least one fact. Dimensions are obtained from relational database entity extraction while facts are obtained from business process analysis processes can be seen in Table 3.

Table 3. The Candidate Dimension based on Business Process

Business Process	Candidate Dimension										
	Motorcycle	Brand	Type	Color	Customer	District	Distributor	Repair Workshop	Sell Location	Customer Contract	Time
Selection of Motorcycle Units from distributors	√	√	√	√			√				
Purchase of Motorcycle Units from Distributor	√	√	√	√							√
Motorcycle unit inspection and repair	√	√	√	√			√	√			
Sales of Motorcycle Units to customers	√	√	√	√	√	√	√	√	√	√	√

Based on the business processes that have been found previously, from the table the dimensions and facts that have the opportunity to determine the total sales of motorcycles are obtained. It can be seen in the table Figure 4. that the Sales of Motorcycle Units to customers business process determines the dimensions, while the facts are a set of dimension groups associated with the measure (calculated value) of total sales can be seen in Figure 4.

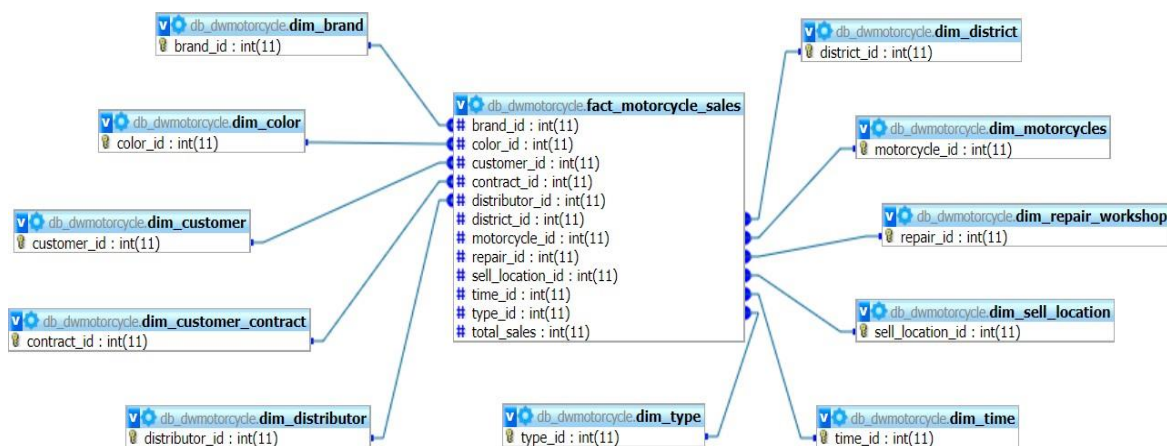


Figure 4. The Star Schema

The star schema forms a fact business process where there are measures and dimensions that act as attributes that describe the data contained in the facts. The star schema is a transformation of the form of a business process into a schema that allows users to perform analysis based on mapping which is more intuitive than using the relational model. The star schema is an optimized model that provides the best performance in presenting more complex information. Star schema is a scheme that is applied in the construction of Data Marts and Data Warehouses that allows for drill down, drill up, slice and dicing operations which are useful in flexibility in extracting information. We hope for further research to develop this star schema into a data mart or data warehouse by using tools that can visualize the results of the data mart or data warehouse.

4. Conclusion

Based on the star schema that we have designed we get several conclusions that the star schema is a transformation of the form of a business process into a schema that allows users to perform analysis based on mapping which is more intuitive than using the relational model. The star schema is an optimized model that provides the best performance in presenting more complex information. Star schema is a scheme that is applied in the construction of Data Marts and Data Warehouses that allows for drill down, drill up, slice and dicing operations which are useful in flexibility in extracting information. We hope for further research to develop this star schema into a data

References

- [1] A. D. Indriyanti, D. R. Prehanto, I. G. L. P. E. Prisma, and I. K. D. Nuryana, "The web-based estimation of motorcycles sales using linear regression method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 5, p. 052115, 2021, doi: 10.1088/1757-899x/1098/5/052115.
- [2] R. Fujinuma and Y. Asahi, "Proposal of Credit Risk Model Using Machine Learning in Motorcycle Sales," in *Human Interface and the Management of Information. Information-Rich and Intelligent Environments*, S. Yamamoto and H. Mori, Eds. Springer, Cham, 2021, pp. 353–363.
- [3] Windarto and Eridani, "Mathematical modelling of motorcycle sale competition in Indonesia," *Commun. Math. Biol. Neurosci.*, vol. 2021, pp. 1–10, 2021, doi: 10.28919/cmbn/5309.
- [4] W. Sutopo, I. Kurniyati, and R. Zakaria, "Markov Chain and Techno-Economic Analysis to Identify the Commercial

- Potential of New Technology: A Case Study of Motorcycle in Surakarta, Indonesia,” *Technologies*, vol. 6, no. 3, p. 73, 2018, doi: 10.3390/technologies6030073.
- [5] J. L. Monino, “Data Value, Big Data Analytics, and Decision-Making,” *J. Knowl. Econ.*, vol. 12, no. 1, pp.256–267, 2021, doi: 10.1007/s13132-016-0396-2.
- [6] K. Biesialska, X. Franch, and V. Muntés-Mulero, “Big Data analytics in Agile software development: A systematic mapping study,” *Inf. Softw. Technol.*, vol. 132, p. 106448, 2021, doi: 10.1016/j.infsof.2020.106448.
- [7] H. S. H. Abdel Azez, M. H. Khafagy, and F. A. Omara, “Optimizing Join in HIVE Star Schema Using Key/Facts Indexing,” *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 35, no. 2, pp. 132–144, 2018, doi: 10.1080/02564602.2016.1260498.
- [8] K. S. Ranti, D. Tuapattinaya, C. Chang, and A. S. Girsang, “Data warehouse for analysing music sales on a digital media store,” *J. Phys. Conf. Ser.*, vol. 1477, no. 3, 2020, doi: 10.1088/1742-6596/1477/3/032013.
- [9] P. Ghosh, D. Sadhu, and S. Sen, “A real-time business analysis framework using virtual data warehouse,” *Int. Arab J. Inf. Technol.*, vol. 18, no. 4, pp. 585–595, 2021, doi: 10.34028/18/4/11.
- [10] N. Sanprasit, K. Jampachaisri, T. Titijaroonroj, and K. Kesorn, “Intelligent approach to automated star-schema construction using a knowledge base,” *Expert Syst. Appl.*, vol. 182, no. December 2019, p. 115226, 2021, doi: 10.1016/j.eswa.2021.115226.
- [11] W. Jianmin, Z. Wenbin, F. Tongrang, Y. Shilong, and L. Hongwei, “An improved join-free snowflake schema for ETL and OLAP of data warehouse,” *Concurr. Comput. Pract. Exp.*, vol. 32, no. 23, pp. 1–11, 2020, doi: 10.1002/cpe.5519.
- [12] T. M. J. Al Taleb, S. Hasan, and Y. Y. Mahd, “On-line analytical processing (OLAP) operation for outpatient healthcare,” *Iraqi J. Sci.*, vol. 2021, pp. 225–231, 2021, doi: 10.24996/ijsc.2021.SI.1.32.
- [13] E. P. W. Mandala, D. E. Putri, D. W. T. Putra, and D. P. Mulya, “OLAP Approach in Searching Manufacturing Industries in West Sumatera,” *J. Phys. Conf. Ser.*, vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012052.
- [14] M. Bouakkaz, Y. Ouinten, S. Loudcher, and Y. Strekalova, “Textual aggregation approaches in OLAP context: A survey,” *Int. J. Inf. Manage.*, vol. 37, no. 6, pp. 684–692, 2017, doi: 10.1016/j.ijinfomgt.2017.06.005.
- [15] S. Chakraborty and J. Doshi, *Faster Result Retrieval from Health Care Product Sales Data Warehouse Using Materialized Queries*, vol. 1176. Springer Singapore, 2021.
- [16] M. Francia, E. Gallinucci, and M. Golfarelli, “Conversational OLAP in action,” *Adv. Database Technol. - EDBT*, vol. 2021-March, pp. 646–649, 2021, doi: 10.5441/002/edbt.2021.74.
- [17] S. Banerjee, S. Bhaskar, A. Sarkar, and N. C. Debnath, “A unified conceptual model for data warehouses,” *Ann. Emerg. Technol. Comput.*, vol. 5, no. Special issue 5, pp. 162–169, 2021, doi: 10.33166/AETiC.2021.05.020.
- [18] N. Sanprasit, T. Titijaroonroj, and K. Kesorn, “A semantic approach to automated design and construction of star schemas,” *Eng. Appl. Sci. Res.*, vol. 48, no. 5, pp. 518–528, 2021, doi: 10.14456/easr.2021.54.
- [19] M. R. Effendy, T. F. Kusumasari, and M. A. Hasibuan, “Star Schema Implementation For Monitoring in Data Quality Management Tool (A Case Study at A Government Agency),” *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, 2019, doi: 10.1109/ICIC47613.2019.8985695.
- [20] R. J. Salaki, J. Waworuntu, and I. R. H. T. Tangkawarow, “Extract transformation loading from OLTP to OLAP data using pentaho data integration,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 128, no. 1, 2016, doi: 10.1088/1757-899X/128/1/012020.