Dimensional Data Design for Event Feedback Data Warehouse

Ahmad Maulana Malik Fattah[1*)], Taufik Ridwan[2], Nina Sulistiyowati[3]

[1, 2, 3]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Email: [1]ahmad.maulana19003@student.unsika.ac.id, [2]taufik.ridwan@cs.unsika.ac.id, [3]nina.sulistio@unsika.ac.id

Abstract – Data is an important asset and a fundamental requirement for building valuable information for organizations. Association of Information Systems Students of Unsika (Himsika) as a university organization provides many events to develop student's academic and professional skills. A post-event evaluation through a feedback survey was conducted and stored in Google Sheets spreadsheet format. However, the current analysis process using spreadsheets lacks standardization, making it difficult to compare satisfaction rates over time and between events. Additionally, the lack of standardization leads to semi-structured data on spreadsheets, with varying question formats and meanings. To address these limitations, implementing a centralized data warehouse is proposed as a solution. The data warehouse would provide a structured and standardized approach to analyzing event feedback, enabling better comparisons and evaluation of management quality within Himsika. The research aims to design a data warehouse that supports multidimensional analysis. As a way to simplify and optimize analytical queries, the data structure is standardized in the data warehouse. The Four-step Dimensional Design method is applied in designing dimensional modeling on the data warehouse, consisting of four stages including selecting the business process, declaring the grain, identifying the dimensions, and identifying the facts. The design process resulted in 4 dimensions of events, dim_instances, dim_degree_programs, and dim_professions, and a fact table called fact_rates_by_responses. Overall, the proposed data warehouse and dimensional modeling approach aim to enhance the analysis and evaluation of Himsika's events.

Keywords – Data Warehouse, Dimensional Data, Event Feedback, Four-step Dimensional Design, Himsika

I. INTRODUCTION

Data has become an important asset for an organization to build information. The collected historical data in massive amounts raises opportunities to perform analysis from many dimensions or known as multidimensional analysis (Agapito, Zucco, & Cannataro, 2020). Besides, data could be the driving aspect while making business decisions and understanding the current condition of the organization by extracting its value with data analysis (Dollah & Aris, 2018). The activity of data collection and analysis is not limited to profit organizations, but in the academic area as well (Yulianto, 2019).

The Association of Information Systems Students of Unsika or Himpunan Mahasiswa Sistem Informasi Unsika (Himsika) is one of the academic associations at the University of Singaperbangsa Karawang. As a student organization focusing on developing students' academic and professional skills, Himsika holds many events (Himsika, 2022). At the end of the events, they have to spread out a feedback survey of how satisfied the participants are with the event. The survey is used to evaluate the management quality of Himsika which should be reported at the end of their management period. In the as-is process, the collected feedback is analyzed by screening the raw data in the form of spreadsheets. There are limitations to comparing the satisfaction rate by time as well as comparing between events. Furthermore, in consequence of there being no standardization in Himsika, the survey tends to have different numbers, phrases, and meanings of questions, resulting in semi-structured data on the spreadsheet. In order to tackle the current problems, centralized data storage in the form of a data warehouse could be one of the solutions.

There are two primary methods proposed by Bill Inmon and Ralph Kimball for designing data warehouses, respectively (Haryono et al., 2020). The Inmon method employs a top-down approach within an enterprise scope, emphasizing highly normalized, consistent, and regulated data. On the other hand, the Kimball method adopts a bottom-up approach in the business area, allowing frequent data redundancy and revisions. In terms of development time, the Inmon method typically requires a longer initial design and implementation time compared to the Kimball method (Wijaya, 2022). When it comes to data modeling, there are four stages in the Kimball method encompasses four stages known as the Four-step Dimensional Design method (Iswari, Fudholi, & Aditya, 2019).

Various studies have employed the Four-step Dimensional Design method to model data in the data warehouse. In the context of higher education, Togatorop (2018) explained how dimensional data design supports the analysis of new-year undergraduate students at a university. The research utilized a star schema that included 4 fact tables, the fact of registration, selection, registrant per school, and school alumni. The dimensional tables comprised 10 dimension tables that facilitated multidimensional analysis. Several dimensions, such as date, registrant, period, high school, and study program, were shared among the fact tables.

Another study conducted by Lapura, Fernandez, Pagatpat, & Dinawanao (2018) discussed the utilization of Kimball Modeling Techniques for building a data warehouse, which follows the same steps as the Four-step Dimensional Design method. The research focused on selecting the business process, specifically examining aggregated financial amounts across dimensions such as funds, categories, and clusters. The data model was designed with a granularity level that ensured each data point represented the allocation, expenditure, and balance for a specific fund on a given day. Dimension tables were created to incorporate dimensions of funds, finance units, accounts, and time, while the fact tables accommodated various measurements, including allocation, balance, obligation, and transferred funds.

Similarly, Yulianto (2019) utilized the Four-step Dimensional Design approach to structure and model data within a data warehouse, intending to employ it as the storage destination for an Extract-Transform-Load (ETL) pipeline. The resulting fact table included data related to new students, registrations, grades, graduation, and payments. Furthermore, the generated dimension tables comprised the time dimension, program of study dimension, and school dimension. This research demonstrates the integration of the Four-step Dimensional Design method as an effective data modeling technique in developing a data warehouse infrastructure.

Another data modeling method that can be utilized is the Nine Steps Methodology, which also covers the data warehouse implementation stages. The first four stages of this methodology align with the Four-step Dimensional Design method, which is employed for modeling dimensional data. Rahutomo, Putri, & Pardamean (2019) presented the additional stages used for implementing the data model and loading data into the data warehouse. These stages encompass storing pre-calculation in the fact table, tracking changes in dimensions with gradual change, and determining priorities and query modes. The research output is a data warehouse containing multiple storage units of historical data from various periods, enabling the analysis of trends.

In a study conducted by Al-Faris, Suharjito, Diana, and Nugroho (2018), the Nine Steps Methodology was applied in developing a data warehouse for an information technology services company. The methodology resulted in a physical data warehouse model comprising 8 fact tables and 9 dimension tables. The data warehouse employed an Extract-Transform-Load (ETL) process to load data from diverse sources. As a result, the company was able to track service requests requiring follow-up and predict sales performance.

This research aims to design a dimensional data model for the event feedback of Himsika. The data model supports the multidimensional analysis on a centralized and clear-structured data warehouse. As to focus on the data modeling aspect, the approach used in the research is the Four-step Dimensional Design method.

II. RESEARCH METHODOLOGY

A. Literature Review

The theoretical foundation used in this research includes data warehouse and data modeling. The research follows the Four-step Dimensional Design Kimball methodology to design the dimensional data model.

1. Data Warehouse

In "Data Engineering with Google Cloud Platform," a data warehouse is defined as a technology that transforms data from an operational system into another system used to support decision-making. The main principles of a data warehouse are to combine data from multiple sources in a centralized location and transform it into an analytically accessible and processable format. Data in a data warehouse can be analyzed through machine learning modeling, data visualization, and report generation (Wijaya, 2022). A data warehouse is also referred to as an

online analytical processing (OLAP) database due to its analytical nature (Garani & Butakova, 2019).

A data warehouse has characteristics that allow data to be organized in a subject-oriented manner. This means that data can be analyzed based on specific business subjects. The collected data tends to be non-volatile, resulting in a large amount of historical data. This historical data is identified by the period of its collection time, also known as 'time-variant'.

The other characteristic of a data warehouse is the data modeling schema used. While transactional databases model the data relationally (Yaqub, Kamel, & Aung, 2020), data warehouse uses a dimensional approach (Wijaya, 2022). Through this kind of modeling, data can be analyzed in general (roll-up) or in a more detailed view (drill-down). The analysis is also able to be focused on some data points of a dimension (slice) as well as multiple data points between dimensions (Challal et al., 2019). Dimensional data modeling in data warehouses can optimize data read operation as it is mainly used in the analytical activity (Wijaya, 2022).

2. Data Modeling

Data modeling is a process to represent database objects in the real world or business perspective. In a data warehouse, data modeling has to be representative of the real world because the end-users of a data warehouse are people and not a computer program. Data modeling is also used to ensure data consistency, optimize query performance, and increase efficiency in storage. One of the methods to model data in a data warehouse is the Kimball method (Wijaya, 2022).
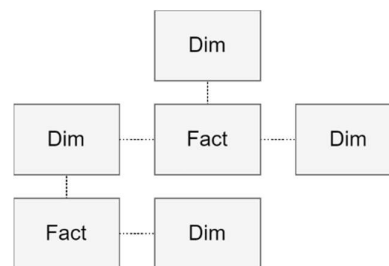

Figure 1. Star schema data model in the Kimball method

The Kimball method takes a bottom-up approach that focuses on answering business questions of a specific domain. The method resulted in several tables called 'fact' and 'dimension' (Turcan & Peker, 2022). The dimension table represents an entity along with its attributes. In the Kimball method, dimension tables are designed to be denormalized consequent in data redundancy. On the other hand, a fact table is a measurement table that has a certain level of granularity. The fact table is located in the center and is surrounded by dimension tables that determine its level of granularity (Wijaya, 2022). This design of fact and dimension tables is called a star schema (Garani & Butakova, 2019).

3. Four-step Dimensional Design Method

The Four-Step Dimensional Design Method is a method for designing a multidimensional data model in a data warehouse. This method emphasizes a bottom-up approach

JISA (Jurnal Informatika dan Sains) (e-ISSN: 2614-8404) is published by Program Studi Teknik Informatika, Universitas Trilogi
under Creative Commons Attribution-ShareAlike 4.0 International License.

70

where data is stored based on business needs. The resulting modeling will form a schema called a star schema. The Four-Step Dimensional Design Method consists of four stages as follows (Lapura et al., 2018):
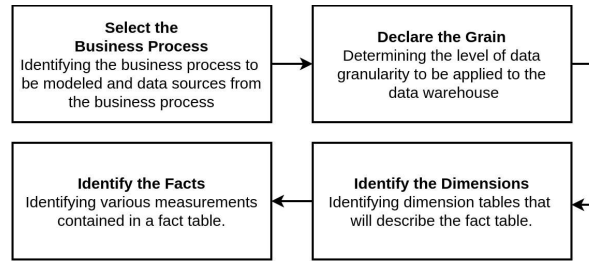

Figure 2. Four-step Dimensional Design method

a. Select the Business Process
This stage determines the operational activities that will be measured for performance. The selection of business processes will determine the output of the next stage. Some examples of business processes include financial activities (Lapura et al., 2018) and academic processes conducted by educational institutions. The selection of business processes is carried out by considering analysis needs and data availability (Yulianto, 2019).

b. Declare the Grain
Granularity is the level of detail in the data analyzed in a data warehouse. Granularity sets limits on what can be represented by a fact table. Each declared granularity will be implemented separately into a fact table. Through the declaration of granularities, the detailed information held by the data in the fact table is determined (Vincentdo, Pratama, Girsang, Suwandi, & Andrean, 2019). Additionally, the declaration of granularities also sets the candidate fact tables to be built (Fardhani, 2018).

c. Identify the Dimensions
This stage identifies entities and their attributes that are used to describe measurements in the fact table. Entities are represented in dimension tables to support the filtering and grouping of facts (Lapura et al., 2018). The identified dimensions are obtained based on the subject of the business process (Turcan & Peker, 2022). Dimension tables are commonly created in a data warehouse including time and date dimensions (Yulianto, 2019).

d. Identify the Facts
This stage identifies the measurements that exist in the fact table (Lapura et al., 2018). Measurements are obtained from the business process carried out by the organization. Each fact table implements a declared granularity that has been declared in the initial stage. A fact table is designed to be consistent with the granularity. This consistency ensures that analysis can be performed accurately (Yulianto, 2019).

B. Methodology
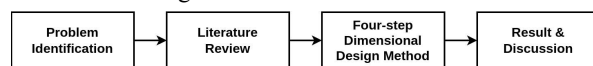The complete stages of the research process are shown in the below image.

Figure 3. Research workflow

1. Select the Business Process
In this research, the business process being measured is the event feedback of Himsika. The process produces event feedback data in multiple spreadsheets. This data is used as the data source of the dimensional model data warehouse.

2. Declare the Grain
This research identified one candidate of the fact table, which is the fact of participant satisfaction. The granularity in the fact table is in the level of feedback responses.

3. Identify the Dimensions
Based on the selected business process and data availability, there are several dimensions identified. These dimensions are the event dimension and participant profile dimensions, such as instances, study programs, and professions.

4. Identify the Facts
The fact table in this research at least includes a table called the fact of participant satisfaction. The table provides data analysis measurements, such as the total number of participants and the percentages of participant satisfaction rate with the speaker, committee/management, and the overall event.

III. RESULTS AND DISCUSSION

The Four-step Dimensional Design method is used to model the dimensional data. The results of each stage of the method are described as follows.

A. Select the Business Process
The selected business process is the event feedback from events held by the Association of Information Systems Students of Unsika or Himpunan Mahasiswa Sistem Informasi Unsika (Himsika). The data sources come from 11 events held from 2020/2021 to 2021/2022. The data is available in a spreadsheet format of Google Sheets and has various structures, called semi-structured data. As a way to simplify and optimize analytical queries, the data structure is standardized in the data warehouse. The below table shows 11 events used as the data source.

Table 1. Event feedback data description

Events	Held Year	Feedback
Pameran Expopedia	2021	4
Seminar Bisnis	2021	241
Talkshow Let's PKM	2021	9
Webinar Expopedia	2021	63
Webinar Online Career Talk (Oncat)	2021	130
Workshop *Program Belajar Bareng* (PBB)	2021	191
Talkshow Silogy Fest a. *Peserta* Online b. *Peserta* Offline	2022	14 15
Webinar Marathon *Acara* 1	2022	80
Webinar Marathon *Acara* 2	2022	58
Workshop PBB UI/UX	2022	69
Workshop PBB Python	2022	71
Total Feedback		945

B. Declare the Grain
Based on the selected business process and data

availability, a candidate fact table is determined, namely the fact of participant satisfaction. The granularity level in the candidate table is at the level of the event participant's response. This granularity declaration ensures that the fact table accurately represents the business process and provides meaningful information.

C. Identify the Dimensions

Dimension identification is based on the availability of data that can describe each measurement in the fact table. The following table contains 4 dimensions generated at this stage. The four dimensions are selected because they have categorical data sources. In terms of multidimensional analysis, the event dimension is employed to filter specific event data and measure the success rate of the event. Analysis of the participant's originating institution can be conducted through the institution dimension. The institution dimension can also be utilized to determine the number of participants from partner institutions with Himsika. Analysis through this dimension can be used to assess the success of Himsika's partnerships with organizations in other institutions. Through the study program dimension, analysts can observe the variation in the study programs of student participants. Meanwhile, the profession dimension allows for an analysis of event accessibility in terms of participants' occupational backgrounds, such as employees, lecturers or teachers, students, high school students, and the general public. Analysis of the study program and profession dimensions can also serve as a basis for how common the event is, whether it is focused on a particular domain or more broadly inclusive of diverse participants.

Table 2. Dimension table design

Dimensions	Attributes	Description
Event	Unique Identity	Unique identity of the event
	Event Name	Event name
	Organizing Cabinet	Name of organizing cabinet
	Held At	Event hold date
	Last Feedback Response	Last response date in the feedback survey
Institution	Unique Identity	Unique identity of the institution
	Institution	Name of participant's institution
Study Program	Unique Identity	Unique identity of the study program
	Study Program	Name of participants study program
Profession	Unique Identity	Unique identity of the profession
	Profession	Name of participant's profession

D. Identify the Facts

The identified fact table candidate is a single candidate, namely participant satisfaction with granularity in each response. Thus, the candidate selected as the fact table is the fact of participant satisfaction rate by responses. The measurements designed in the fact table consist of assessments of several aspects of the event, such as speakers, management, and overall event satisfaction. The following table details the design of analysis measurements in the fact table.

Table 3. Fact table design

Fact	Measurements	Description
Participant satisfaction rate by responses	Timestamp	The time of filling out the feedback survey
	Comment	Participant's comments in the form of suggestions or criticism
	Study Year	The academic year of the participant
	Revisit Expectation	Participant's statement of interest to attend future events
	Speakers Rate	Average rating of the speaker assessment
	Management Rate	Average rating of the management assessment
	Overall Event Rate	Average rating of the overall event
	Duration Rate	Average rating of the duration of the event
	Topic Rate	Average rating of the topic of the event
	MC Rate	Average rating of the host or master of ceremony (MC)
	Moderator Rate	Average rating of the event moderator

IV. CONCLUSION

The design result demonstrates the successful creation of a dimensional data model for the event feedback of Himsika. The semi-structured event feedback is now standardized and centralized within the data warehouse. The dimensional model enables analysis of participant satisfaction rates from multiple dimensions, including events, instances, study programs, and professions. The measurements provide insights into the satisfaction levels related to speakers, event management, and the overall event experience.

In the development of this research, a physical data warehouse can be implemented. The Extract-Transform-Load (ETL) process can be employed to load data from various semi-structured spreadsheets in Google Sheets into the data warehouse. Additionally, a presentation layer can be created to visualize the data in the data warehouse more interactively.

REFERENCES

[1] Agapito, G., Zucco, C., & Cannataro, M. (2020). COVID-WAREHOUSE: A Data Warehouse of Italian COVID-19, Pollution, and Climate Data. *Int. J. Environ. Res. Public Health*, *17*, 1–22. https://doi.org/10.3390/ijerph17155596

[2] Al-Faris, F. Z., Suharjito, Diana, & Nugroho, A. (2018). Development of Data Warehouse to Improve Services in IT Services Company. *International Conference on Information Management and Technology (ICIMTech)*, (September), 483–488. https://doi.org/10.1109/ICIMTech.2018.8528146

[3] Challal, Z., Bala, W., Mokeddem, H., Boukhalfa, K., Boussaid, O., & Benkhelifa, E. (2019). Document-

oriented versus Column-oriented Data Storage for Social Graph Data Warehouse. *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, 242–247. https://doi.org/10.1109/SNAMS.2019.8931718

[4] Dollah, R., & Aris, H. (2018). A Big Data Analytics Model for Household Electricity Consumption Tracking and Monitoring. *2018 IEEE Conference on Big Data and Analytics (ICBDA)*, 44–49. Langkawi, Malaysia: IEEE. https://doi.org/10.1109/ICBDAA.2018.8629769

[5] Fardhani, M. R. D. (2018). *Perancangan Data Warehouse Beasiswa Karya Salemba Empat dengan Nine Step Design Method* (Universitas Islam Negeri Syarif Hidayatullah Jakarta). Universitas Islam Negeri Syarif Hidayatullah Jakarta. Retrieved from http://repository.uinjkt.ac.id/dspace/handle/123456789/48402

[6] Garani, G., & Butakova, M. A. (2019). A Data Warehouse Approach for Business Intelligence. *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 70–75. https://doi.org/10.1109/WETICE.2019.00022

[7] Haryono, E. M., Fahmi, Tri W, A. S., Gunawan, I., Nizar Hidayanto, A., & Rahardja, U. (2020). Comparison of the E-LT vs ETL Method in Data Warehouse Implementation: A Qualitative Study. *Proceedings - 2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, 115–120. Jakarta, Indonesia: IEEE. https://doi.org/10.1109/ICIMCIS51567.2020.9354284

[8] Himsika. (2022). *Grand Design Program Kerja Himpunan Mahasiswa Sistem Informasi Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang Periode 2021-2022*. Karawang: Himpunan Mahasiswa Sistem Informasi Unsika.

[9] Iswari, L., Fudholi, D. H., & Aditya, S. K. (2019). Dashboarding the maternal and child health profiles for health supporting system. *International Conference on Information Technology and Digital Applications (ICITDA 2018)*, *482*(1), 012013. https://doi.org/10.1088/1757-899X/482/1/012013

[10] Lapura, E. V. F., Fernandez, J. K. J., Pagatpat, M. J. K., & Dinawanao, D. D. (2018). Development of a University Financial Data Warehouse and Its Visualization Tool. *Procedia Computer Science*, *135*, 587–595. https://doi.org/10.1016/j.procs.2018.08.229

[11] Rahutomo, R., Putri, R. A., & Pardamean, B. (2019). Building Datawarehouse for Educational Institutions in 9 Steps. *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, 128–133. https://doi.org/10.1109/INAPR.2018.8627010

[12] Togatorop, P. (2018). Perancangan Dimensional Model Pada Data Warehouse untuk Penerimaan Mahasiswa Baru di Institut Teknologi Del. *Jurnal Nasional Teknik Informatika Dan Elektro (JURNALTIO)*, *01*(01), 53–56. Retrieved from http://jurnaltio.del.ac.id/index.php/jurnaltio/article/view/12

[13] Turcan, G., & Peker, S. (2022). A multidimensional data warehouse design to combat the health pandemics. *Journal of Data, Information and Management*, *4*, 371–386. https://doi.org/10.1007/s42488-022-00082-6

[14] Vincentdo, V., Pratama, A. R., Girsang, A. S., Suwandi, R., & Andrean, Y. P. (2019). Reporting and Decision Support Using Data Warehouse for E-Commerce top-up Cell-phone Credit Transaction. *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, 1–4. https://doi.org/10.1109/CITSM47753.2019.8965349

[15] Wijaya, A. (2022). *Data Engineering with Google Cloud Platform*. Birmingham: Packt.

[16] Yaqub, W., Kamel, I., & Aung, Z. (2020). Toward watermarking compressed data in columnar database architectures. *Security and Privacy*, *3*(4), 1–17. https://doi.org/10.1002/spy2.84

[17] Yulianto, A. A. (2019). Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse. *APTIKOM Journal on Computer Science and Information Technologies*, *4*(2), 61–68. https://doi.org/10.11591/aptikom.j.csit.36