International Journal of Engineering and Applied Technology (IJEAT)
Vol. 5., No. 2, November 2022, pp. 7-15
ISSN 2620-9632
https://www.inter-journal.nusaputra.ac.id/IJEAT

Modelling of Data Warehouse With Making the Trend to Make Decision in Company XYZ

Dwi Sartika Simatupang[a,1*], Anggun Fergina[b,2], Bahadir Ozsut[c,3]

[a,b]Department of Informatic Engineering, Nusa Putra University, Jl. Raya Cibolang Kaler No.21, Kab. Sukabumi 43152, Indonesia
[c]Departement of Business Administration Cukurova University, Adana, Turkey
[1]dwi.simatupang@nusaputra.ac.id*; [2]angguun.fergina@nusaputra.ac.id; [3]bahadirozsut@gmail.com
* Corresponding Author

Received 29 October 2022; revised 01 November 2022; acceted 06 November 2022

ABSTRACT

The goals of the thesis were to modellig a data warehouse which was loaded from the operational database PT. XYZ, to analyze data warehouse with Tableau Software to know the trends of activity Trouble ticket, and to create reports and dashboards to facilitate PT. XYZ in viewing the trends. Data were compiled by observation to PT. XYZ from reporting team. Data were analyzed using software tableau. The results obtained were the trend of activity trouble ticket more than two hundred thousand data in sixteen weeks or four month in 2017. It can be concluded that developed data warehouse can be used to analyze in gaining information about the trends of activity trouble ticket and help to make a decision.

KEYWORDS
Data warehouse
Trend
Trouble ticket

Introduction

The development of technology is growing rapidly[1],[2]. Many companies use these technologies for the purpose of effectiveness in order to compete with other companies[3]. In general, every company has a database to store every information needed by the company.PT XYZ companyis a vendor engaged in telecommunications. This vendor has a managed service department that has one customer operator in Indonesia. As one vendor must have data relating to the activities or activities of managed services that are created in a trouble ticket[4]. Troble ticket is a ticket made when an alarm appears, this trouble ticket is a number in which there is a description of the alarm information that appears. With this Trouble Ticket you can find out how many activities have been done, how many have finished or closed, and how many are still open.This case study research starts from the needs of the PT XYZ company to be able to present reports or data quickly when needed. Especially every week, month and year a review is held to present a report to be reported to the customer. During this time the trouble ticket data is displayed in CSV and the data processing is done manually. The manual method takes a long time in the process because of the large number of troble tickets for each day. The solution offered in this case study is to compile and move data that was originally only available in the form of CSV to a data warehouse.

Method

2.1 Data warehouse

Data warehouse is data collection process that is subject-oriented, integrated, time varied, and non-volatile. It is used to support strategic decision-making process for an organization. Data warehouse contains the extraction of various corporate operating systems, each of which holds different records from every business transaction [5]. The data warehouse characteristics are as follows[6], [7], [8].

International Journal of Engineering and Applied Technology (IJEAT)
Vol. 5., No. 2, November 2022, pp. 7-15
ISSN 2620-9632

Subject-oriented, which covers the subject or main business entities in an organization such as the lecturers, students, subjects, grades, and curriculum. Data warehouse is designed to facilitate thorough analysis of data in a considerable amount. The data arranged by subject only contains important information for Decision Support System (DSS) processing. The information stored in the database is classified by particular subjects, for instance, in library case, members and books. The data in every subject is summarized into dimensions, such as time period; thus, historical data can be provided for analysis [9].

Integrated, with the data collected from operational data and external data which are integrated in a data warehouse in order to get a single data base to support a decision. The data in data warehouse can be obtained from several separate sources. This data will be stored in the same segment in a specific and consistent format. The data in data warehouse is sourced from operational database (internal source) and from outside of the system (external source). Data warehouse can store the data from separate sources in a consistent and integrated format[10].

Time-variant, with the data that are collected in data warehouse contain time dimension to identify trend, predict future operations, and controlling operational target. Data warehouse stores historical data useful for analysis and decision-making. The data in data warehouse is characterized as time-series data in the form of time-variant historical data. This function is geared to perform trend analysis of the data. Some ways of looking into time interval in measuring the accuracy of a data warehouse include the following: a) presenting data warehouse at a particular time interval, which is the simplest one, b) using time variance presented within the data warehouse, either explicitly using time units, such as day, week, months and particular time or implicitly and c) using time variance presented by data warehouse through a long set of snapshots[11], [12], [13].

4)Non-volatile, where the data available in data warehouse are not updated in real time, but refreshed regularly from the operational system. New data are always added to update the database.The data stored within a data warehouse cannot be changed. Unlike the data in the OLTP system, the data in a data warehouse is not updated in real time/continuously (by applying the insert, update and delete functions); the data can only be viewed or added with new data. In a data warehouse, only two data manipulation activities can be conducted, namely data loading (extracting data) and data access (accessing the data warehouse). The data in a data warehouse is uploaded on a periodic basis in the same period [14].

2.2 Nine-step methodology

Data warehouse planning method used was based on the Nine-step methodology from Kimball. The steps were[15]:

1. Selection of the process: The data mart that is built first must be the one that can be transferred in a timely manner and address all important business queries. Data mart is a database containing data that only describes segments of corporate operations [16].
2. Identification of the grain: To decide for certain what is represented by a fact table. Grain is a process in which what will be described by a record in a fact table is determined[17].
3. Identification and adjustment: A well-built dimension set, making it easy to understand and use data mart, this dimension is essential for describing the facts contained in a fact table. In this stage, an adjustment of dimension and grain presented in the form of a matrix is done [18].
4. Identification of facts: The source of a fact table determines which facts are usable in the data mart. All facts must be expressed at a level predetermined by the source.

   Storage of pre-calculation data in tables: Storing pre-calculation in the fact table.

Declaring the dimensional table: In this stage, complete information is inserted into the dimensional table. Textual description is added to possible dimensions too. The textual description must be easy for users to use and understand [19].

Selection of database duration: For instance, in an insuring company, data must be stored for a duration of 10 years or more. Selection of duration of historical data belonging to a hospital can be performed according to the information need. In most cases, the more the data is transferred into a data warehouse, the more complete the information is generated. Attention should be paid to the duration of

historical data by taking into account the existing data content and format. Care must be exercised so as to avoid transferring useless junk data [20].

Tracking the dimensional changes slowly: Tracking slow dimensional change. The following are three types of slow dimensional change:

- Type 1. The attributes of a dimension that has changed are rewritten
- Type 2. The attributes of a dimension that has changed make new dimensions
- Type 3. The attributes of a dimension that has changed make an alternative, allowing the values of the old and new attributes to be accessed simultaneously at the same dimensions.

Determining priority and query mode: In this stage, physical planning is required. The effect of physical planning, such as the sorting of the fact tables in the disk, and the location of initial storage of summaries or aggregate are taken into consideration [21]. All of the steps above should be taken before a data warehouse is implemented. The next stage is implementing simple data warehouse or data mart[22].

2.3 Data warehouse modeling

Dimensional Model scheme that was used to develop the data warehouse in this study was Fact Constellation Scheme. This scheme is considered as constellation scheme since there is a dimensional table used simultaneously by one or more fact tables [23]. Fact Constellation Schema is a schema comprising several fact tables that share the same dimensional table. This schema can be viewed as a star constellation, thus it is often called the galaxy schema [24]. Fact constellation schema is more complex than star schema as it contains various fact tables. In fact constellation schema, one dimensional table can be used in several fact tables, thus requiring a more complex design. An advantage of fact constellation schema is its ability to model business more accurately using some fact tables. However, it is difficult to manage and has a complex design[24].

2.4 ETL process planning

ETL process, or Extract, Transform, and Load, is a data processing that change it from OLTP database into data warehouse. ETL process is part of data staging. The process changes, re-formats and integrates data that are obtained from one or more OLTP systems.ETL is a critical process in data warehousing. With ETL, the data from the operational activity can be inserted into the data warehouse. ETL can also be used for integrating data with the existing system. The purpose of ETL is to collect, filter, process and combine relevant data from various sources to be stored in the data warehouse. ETL process results in data that meets the data warehouse criteria, such as being historical, integrated, summarized, static and in a structure designed for analysis purpose[25].

1. Extraction
   The first stage of ETL process is to extract data from data sources. Most data warehouse projects combine data from different sources. There is a high chance that separate systems use different data formats. Extraction is converting data into a format which will be useful for transformation process [26].
2. Transformation
   Data transformation is a phase where data that have been extracted are changed into raw data that are suitable to be used in data warehouse.
3. Transformation stage uses a set of rules or functions to extract data from the source and subsequently to input the extracted data into data warehouse. The following are what can be done in the transformation step: [27]
4. Select certain columns to be inserted into the data warehouse;
5. Translate values in the form of code;
6. Encode the values into a free form (for example, map “Male” as “M” and “Female” as “F”);
7. Conduct calculation of new values (for example, value = qty * unit_price);
8. Combine data together from various sources;
9. Make a summary from some data rows;
10. Generate surrogate key value;Conduct transposing or pivoting (convert a set of columns into a set of rows, and vice versa);
11. Split a column into some columns andUse various forms of data validation, both simple and complex.

International Journal of Engineering and Applied Technology (IJEAT)
Vol. 5., No. 2, November 2022, pp. 7-15
ISSN 2620-9632

12. Loading is a process to transfer data physically from OLTP system into data warehouse.Loading phase is a stage in which data is inserted into the final target, mostly a warehouse data. How much time taken for this process will depend on the organization's need [28].

2.5 Data warehouse tools

The following are the tools used by users for various purposes after a Data Warehouse is established: [29]

1. OLAP (On-Line Analytical Processing)
   OLAP is one of Data Warehouse tools used for data analysis. OLAP is a technology designed to provide superior performance for ad hoc business intelligence queries [21]. OLAP is designed to operate efficiently with data organized following general dimension models normally used in Data Warehouse. The following are ways in which OLAP is useful.
   a. OLAP improves productivity of manager, executive and business analysis.
   b. Effectively using OLAP enables users make their own analyses confidently without the help of IT assistance.
   c. OLAP greatly benefits IT developers, in that it is highly useful for improving the performance of their applications.
   d. OLAP improves work efficiency. OLAP may be used for the following:
2. Pentaho
   Pentaho Kettle is open-source software released by Pentaho corp, which is based in Orlando, the United States. The main elements of Pentaho Kettle are transformation and job. Transformation is a set of instructions for converting input into desired output, while job is a set of instructions for executing transformation.
   a. Reporting
      Reporting tools are the tools used to help users retrieve historical or current data and undertake some statistical analysis standards [24]. The data generated from the reporting tools can take the form of either normal report or graphics. The tool used for reporting is Tableau. Tableu is easy to use, especially in making data visualization, data analysis and reporting, owing to drag and drop system it uses. Tableu is capable of combining data from various data sources, such as spreadsheet, database, cloud data and big data into one program to be used in a dynamic analysis[30].

3. 3 Results And Discussion

3.1 Research steps

The description of the mindset in this study can be seen in this case study study at PT XYZ, beginning with the problems formulated in the solution to the existing problems. This formulation is supported by data collection and information and interviews. In addition, it is supported by some literature regarding several related studies that have been done before.

The development of this data warehouse is done by nine step design methodology [31]using Pentaho software. Furthermore OLAP is used to analyze data according to the needs of PT XYZ. The final stage in the research carried out is to make a report and dashboard, so that it can facilitate decision making.

The picture below is a mindset in research:

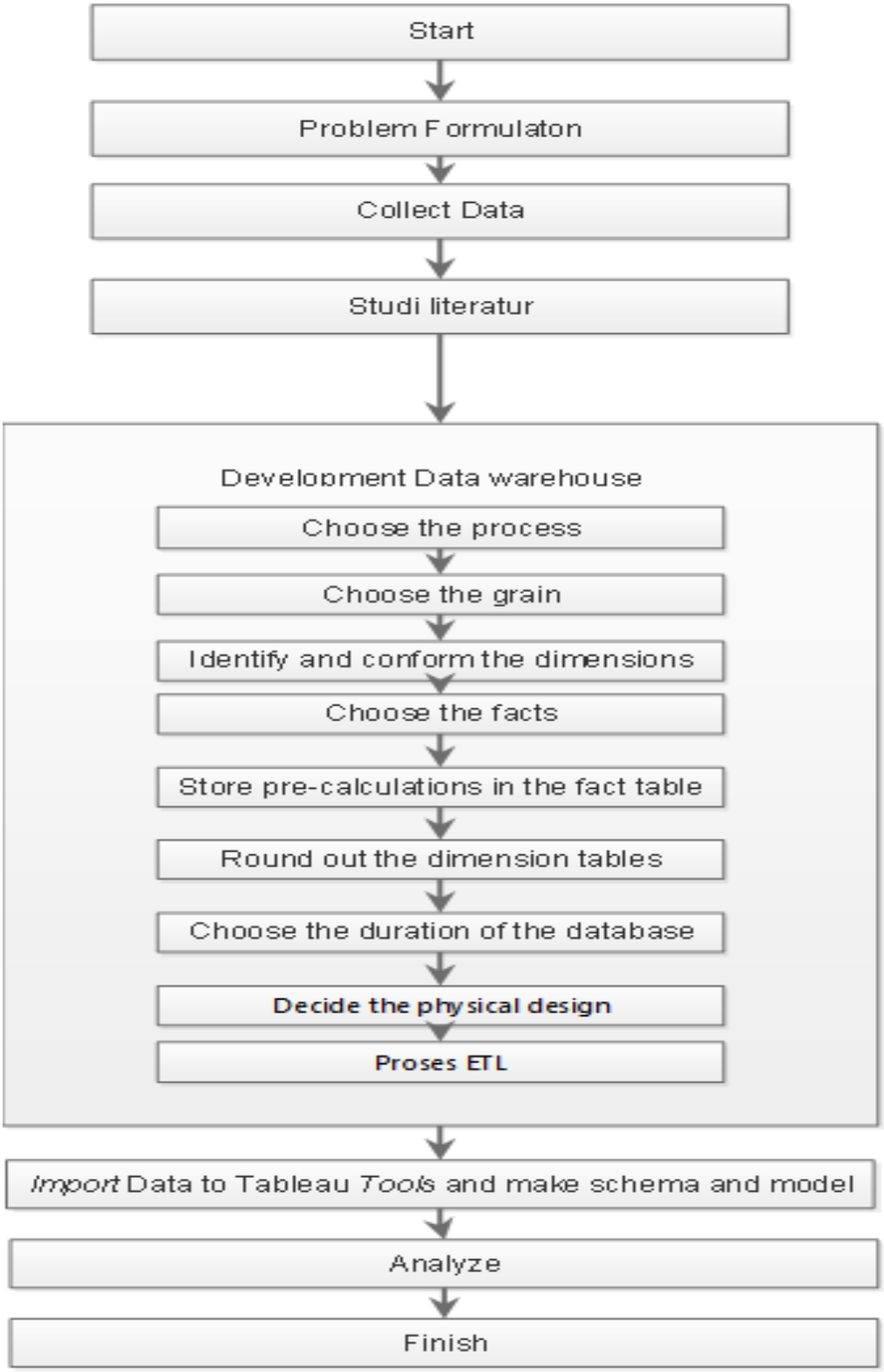

Fig. 1. Data warehouse development

International Journal of Engineering and Applied Technology (IJEAT)
Vol. 5., No. 2, November 2022, pp. 7-15
ISSN 2620-9632

The After nine step design methodology is done, the next step is the process of extracting, transform, and load (ETL) from the data source (operational source) to the data warehouse that has been designed. ETL process is done using pentaho data integration software. The process that will be carried out during the ETL process.

The stages of extraction, transformation, and loading were performed as follows:

1. Extraction: In this step, the data was imported from spreadsheet file into mysql database.
   - Creating new database and tables using mysql.
   - Importing the available data into the new database.
2. Transformation: there were 3 steps performed in this process:
   - Data selection: Selecting data from extraction results (student, subject, batch, and grade).
   - Data cleaning: Noise was cleaned at this step by manipulating empty or incomplete data.
   - Splitting/Joining: At this step, manipulation and joining of data were performed based on the fact tables developed previously (example: merging/joining students table with subjects table or students table with grades table, etc.).
3. Loading: This process is the last stage in ETL.

Cleaned data were stored in data warehouse. The output was in the form of transformed data from Pentaho that was stored in mysql database.

Fig. 2. Process ETL

3.1. 3.2Preparation stage

Preparation The first step taken in this research is to determine the problem formulation. Problems faced by PT. XYZ is the number of trouble tickets (TT) at l PT. XYZ increases as the number of TT increases. Transaction data is stored in an operational database, so information that can be obtained cannot be used as an analysis material to find out information in the form of TT transaction trends. Therefore, PT. XYZ has difficulty in making decisions on the number of TTs per week based on the trends that occur.

3.2. 3.3 Colecting Data

After the formulation of the problem is determined, the next step is collecting data. The data needed is a Trouble ticket data on December 2017, Data collection techniques used in this study are:

a. Observation
b. Interview
c. Study of literature

3.3. 3.3 Implementation of tableau tools

To see trends Trouble tickets can be visualized graphs using Tableau. The various information obtained is as follows.

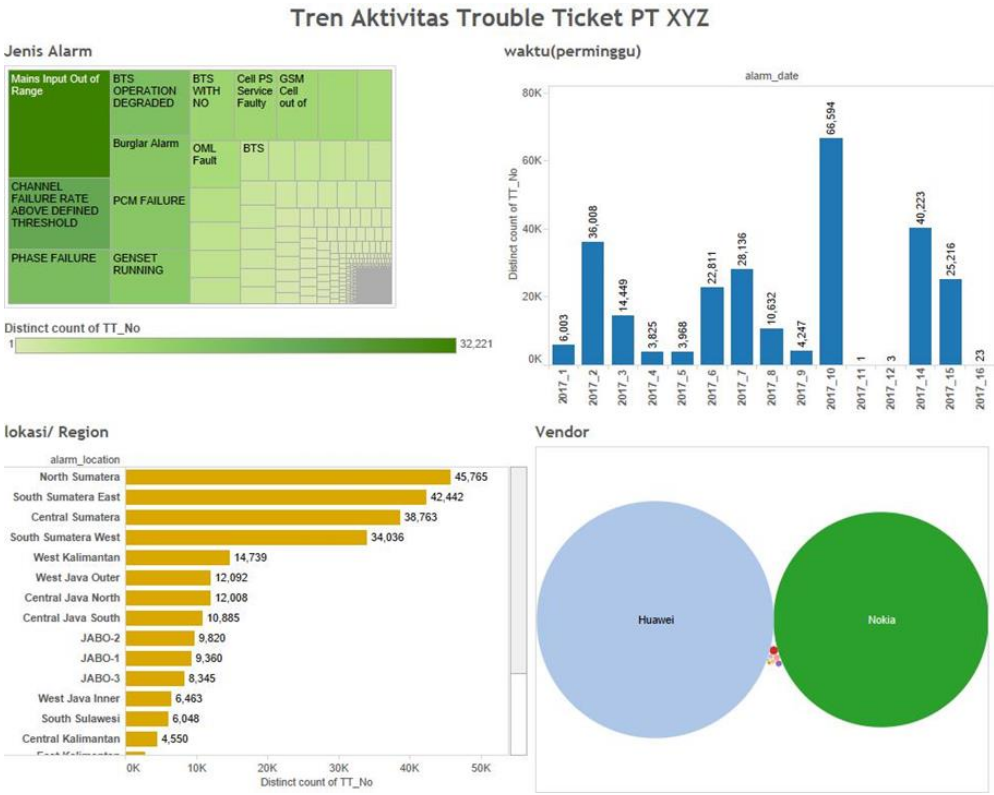

Fig. 3. Visualization of Tableau Tools

Figure 3 is a dashboard consisting of four trends in trouble ticket activity, based on the type of alarm, based on time, by region, and by vendor. The data is obtained based on four dimensions of data processed by ETL which are then entered into the database. This dashboard makes it easy to see and make decisions.

4. Conclusion

The process of data integration begins with extraction (extraction) and then is made uniform in accordance with the format used for analysis purposes. Data in a format that is suitable for evaluation and analysis purposes is then stored in the Data Warehouse (loading). The Data Warehouse in the Academic field has fourdimension tables (date dimensions, alarm dimensions, vendor dimensions, location dimensions) and one Fact table, namely the trouble ticket Fact.Based on Tableau's visualization it was found that trends that are obtained by information on the number of Trouble tickets based on the type of alarm, the most common alarm is playing input of range, trends obtained by information on the number of Trouble tickets based on time, obtained at week 2 and 3 each month have the most trouble tickets, trends that are obtained by information on the number of Trouble tickets based on vendors.

References

[1] G. Briganti and O. Le Moine, "Artificial intelligence in medicine: today and tomorrow," *Front. Med.*, vol. 7, p. 27, 2020.Google Schoolar

[2] B. Meskó and M. Görög, "A short guide for medical professionals in the era of artificial intelligence," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, 2020.Google Schoolar

[3] M. Falahat, T. Ramayah, P. Soto-Acosta, and Y.-Y. Lee, "SMEs internationalization: The role of product innovation, market intelligence, pricing and marketing communication capabilities as drivers of SMEs' international performance," *Technol. Forecast. Soc. Change*, vol. 152, p. 119908, 2020.Google Schoolar

[4] A. Polydoropoulou, I. Pagoni, A. Tsirimpa, A. Roumboutsos, M. Kamargianni, and I. Tsouros, "Prototype business models for Mobility-as-a-Service," *Transp. Res. Part A Policy Pract.*, vol. 131, pp. 149–162, 2020.Google Schoolar

[5] M. L. T. Situmorang, "PERANCANGAN ARSITEKTUR MICROSERVICES UNTUK DATA WAREHOUSE GOAPOTIK MENGGUNAKAN TEKNOLOGI BIG DATA (STUDI KASUS: PT. GLOBAL URBAN ESENSIAL)." Universitas Atma Jaya Yogyakarta, 2020.Google Schoolar

[6] W. Liang, Y. Fan, K.-C. Li, D. Zhang, and J.-L. Gaudiot, "Secure data storage and recovery in industrial blockchain network environments," *IEEE Trans. Ind. Informatics*, vol. 16, no. 10, pp. 6543–6552, 2020.Google Schoolar

[7] G. Nagasubramanian, R. K. Sakthivel, R. Patan, A. H. Gandomi, M. Sankayya, and B. Balusamy, "Securing e-health records using keyless signature infrastructure blockchain technology in the cloud," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 639–647, 2020.Google Schoolar

[8] B. Bode *et al.*, "Glycemic characteristics and clinical outcomes of COVID-19 patients hospitalized in the United States," *J. Diabetes Sci. Technol.*, vol. 14, no. 4, pp. 813–821, 2020.Google Schoolar

[9] J. Qi, P. Yang, L. Newcombe, X. Peng, Y. Yang, and Z. Zhao, "An overview of data fusion techniques for Internet of Things enabled physical activity recognition and measure," *Inf. Fusion*, vol. 55, pp. 269–280, 2020.Google Schoolar

[10] M. Armbrust, A. Ghodsi, R. Xin, and M. Zaharia, "Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics," in *Proceedings of CIDR*, 2021.Google Schoolar

[11] P. Kumar and H. H. Huang, "Graphone: A data store for real-time analytics on evolving graphs," *ACM Trans. Storage*, vol. 15, no. 4, pp. 1–40, 2020.Google Schoolar

[12] C. Tang *et al.*, "XIndex: a scalable learned index for multicore data storage," in *Proceedings of the 25th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, 2020, pp. 308–320.Google Schoolar

[13] A. X. Kohll *et al.*, "Stabilizing synthetic DNA for long-term data storage with earth alkaline salts," *Chem. Commun.*, vol. 56, no. 25, pp. 3613–3616, 2020.Google Schoolar

[14] Y. Yin, Y. Li, B. Ye, T. Liang, and Y. Li, "A blockchain-based incremental update supported data storage system for intelligent vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4880–4893, 2021.Google Schoolar

[15] K. S. Ranti, D. Tuapattinaya, C. Chang, and A. S. Girsang, "Data warehouse for analysing music sales on a digital media store," in *Journal of Physics: Conference Series*, 2020, vol. 1477, no. 3, p. 32013.Google Schoolar

[16] V. Belov, A. N. Kosenkov, and E. Nikulchev, "Experimental characteristics study of data storage formats for data marts development within data lakes," *Appl. Sci.*, vol. 11, no. 18, p. 8651, 2021.Google Schoolar

[17] C. Guo, M. Zhang, and S. Devahastin, "3D extrusion-based printability evaluation of selected cereal grains by computational fluid dynamic simulation," *J. Food Eng.*, vol. 286, p. 110113, 2020.Google Schoolar

[18] W. Hu, Z. Dong, L. Yu, Z. Ma, and Y. Liu, "Synthesis of W-Y2O3 alloys by freeze-drying and subsequent low temperature sintering: microstructure refinement and second phase particles regulation," *J. Mater. Sci. Technol.*, vol. 36, pp. 84–90, 2020.Google Schoolar

[19] M. B. Simanjuntak, M. S. Lumingkewas, and S. Sutrisno, "ANALYSIS OF TANGIANG ALE AMANAMI (OUR FATHER) USING THE TECHNIQUES OF TRANSLATION," *J. Adv. ENGLISH Stud.*, vol. 4, no. 2, pp. 70–75, 2021.Google Schoolar

[20] A. W. A. Geraghty *et al.*, "Exploring patients' experiences of internet-based self-management support for low back pain in primary care," *Pain Med.*, vol. 21, no. 9, pp. 1806–1817, 2020.Google Schoolar

[21] D. Ribeiro de Almeida, C. de Souza Baptista, F. Gomes de Andrade, and A. Soares, "A survey on big data for trajectory analytics," *ISPRS Int. J. Geo-Information*, vol. 9, no. 2, p. 88, 2020.Google Schoolar

[22] 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.Google Schoolar

[23] C. Milanés-Batista, H. Tamayo-Yero, D. De Oliveira, and J. R. Nuñez-Alvarez, "Application of Business Intelligence in studies management of Hazard, Vulnerability and Risk in Cuba," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 844, no. 1, p. 12033.Google Schoolar

[24] V. M. Ngo, N.-A. Le-Khac, and M.-T. Kechadi, "Data warehouse and decision support on integrated crop big data," *Int. J. Bus. Process Integr. Manag.*, vol. 10, no. 1, pp. 17–28, 2020.Google Schoolar

[25] R. R. Berahim and M. Iqbal, "Analysis and Design Optimize Data for the Depok Center Information System's Health Services on the E-Government Data Warehouse Application using an Olap Pivot Table and Star Schema".Google Schoolar

[26] K. Kaplan, Y. Kaya, M. Kuncan, M. R. Mĩnaz, and H. M. Ertunç, "An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis,"*Appl. Soft Comput.*, vol. 87, p. 106019, 2020.Google Schoolar

[27] E. Zdravevski, P. Lameski, C. Apanowicz, and D. Ślęzak, "From Big Data to business analytics: The case study of churn prediction,"*Appl. Soft Comput.*, vol. 90, p. 106164, 2020.Google Schoolar

[28] J. Heinonen, "From Classical DW to Cloud Data Warehouse," 2020.Google Schoolar

[29] L. Zhang, M. Peng, W. Wang, Z. Jin, Y. Su, and H. Chen, "Secure and efficient data storage and sharing scheme for blockchain-based mobile-edge computing,"*Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 10, p. e4315, 2021.Google Schoolar

[30] H. S. Munawar, S. Qayyum, F. Ullah, and S. Sepasgozar, "Big data and its applications in smart real estate and the disaster management life cycle: A systematic analysis," *Big Data Cogn. Comput.*, vol. 4, no. 2, p. 4, 2020.Google Schoolar

[31] Z. Wu, Y. Zhou, H. Wang, and Z. Jiang, "Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse," *Sci. Total Environ.*, vol. 716, p. 137077, 2020.Google Schoolar