

Extract, Transform, Load Monitoring Sales Multi Marketplace

Leonardus Adityo Toto Pratomo^{1✉}, Deni Mahdiana¹

¹Universitas Budi Luhur, Jakarta, Indonesia

*Corresponding Author: 2211602178@student.budiluhur.ac.id

Article Information

Article history:

No. 721

Rec. Mei 29, 2023

Rev. August 28, 2023

Acc. September 16, 2023

Pub. September 24, 2023

Page. 174 – 188

Keywords:

- ETL
- RDBMS
- MySQL
- Microsoft Access
- Multi Marketplace

ABSTRACT

This journal is created in the design of a database system and online sales file for sales on a medium business scale, in the Era of Society 5.0. Online marketplaces provide convenience for medium business entities to sell their products. The process conducted earlier involves calculating raw sales data using Microsoft Excel, which entails consolidating raw data from three marketplace sources, requiring effort until it takes the form of raw factual data. The Extract Transform Load process in MS Access accelerates the manual process, with the aim of enabling business entities to efficiently and quickly view sales reports. The goals of implementing ETL are as follows: Integrating a data mart in the form of factual data stored in MYSQL. Proposing a prototype method by creating a Microsoft Access program in the form of Extract Transform Load (ETL) that operates automatically with a button. Creating Analysis data into factual tables from the three-dimensional data mart sourced from three raw marketplace data, with the transformation process of tables in the MYSQL data warehouse taking the form of exporting daily and specific period reports. This journal designs a combination of open-source systems and Microsoft Access licenses for business entities that require sales analysis for multichannel sales (online or offline), in order to optimize inventory and consistently provide products through each sales channel.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. INTRODUCTION

The application of database technology began in the 1960s, At first, database systems were used to store data centrally in a table or file format, which could be accessed by multiple users. In the 1970s, Edgar F. Codd, a mathematician and computer scientist from IBM, developed relational models for databases, such as SQL (Structured Query Language). In the 2000s, database technology continued to evolve NoSQL databases [1]

The application of information technology has a role in managing raw data, even for large and small business scales, which can be imported, and and processed to the desired output at the management level. Opensource databases are used by business people as an alternative business solution that in terms of cost can be affordable[2]. Generally open-source databases such as MYSQL [1] and Postgresql. Design is simple in a Microsoft Access program connected to an open-source database, and dynamically can be changed quickly daily, including query builder tools and macros. [3]

A database is a structured data collection [4]. MySQL is a multi-user SQL database business management system and multithreaded DBMB (Database Management System). MySQL is derived from SQL (Structured Query Language) [5]. MySQL is a Database Management System that uses basic SQL commands. In its application, MySQL as data management in the system is certainly very easy to use and has fast performance in data storage [6].

Many companies today offer their goods through a variety of online and offline sales methods. A system that reliably and successfully combines and analyzes sales data from many sources is essential. By combining open-source technologies and Microsoft Access licensing, this research offers a solution. This study is vital for understanding sales patterns, demand trends, and consumer preferences, as well as for managing product inventories. This study provides a solution by combining sales data from several channels to assure product information uniformity. Product consistency across all channels is also an issue. Businesses may increase operational effectiveness, save time, and lower human error rates by using a system that combines and analyzes sales data. Additionally, properly integrated and evaluated sales data aids in decision-making, opportunity identification, and strategic decision-making for firms. Therefore, there is an urgent need for this study to help firms maximize multi-channel sales, manage inventory effectively, and deliver consistent products across all sales channels. According to the summary above, the goal of this project is to create a system that combines an open source platform and a Microsoft access license for use by businesspeople that require multichannel (online and offline) sales analysis in order to manage inventory and deliver goods reliably across all channels.

2. RESEARCH METHOD

Research Analyzes necessity Planning system information with ETL.ini be method prototype development. Prototype method with Accelerate pendekatan at Stages software development life cycle (SDLC) [7] deep development device soft with face CASE tools, Visual programming languages, generators yang can Use immediately Into shape report to of Analyzes and Implemented. Research Beginning with planning see files that are already in download from some marketplace, then data Collected, Analyze each column and Design architecture database. Design ETL uses logic flow The process sees data raw result Sales that

happen every day [8]. Language Programming that is Used with visual basic for application (VBA) in Microsoft access and connection Microsoft open database connectivity (ODBC). Testing Done Directly Gradual on data raw until baris serratus Thousands record on the line CSV [9] And excel [10]. Flow research like that is visible in Picture 1.

Figure 1 is the stages and research flow of the prototype [11] method. Starting from the planning stage of files that have been downloaded from several markets places then analyzing database variables, designing, implementing data processing, and transforming data, then processing data architecture design, entering The graphical user interface stage, and the resulting report form is in the form of raw data by paying attention to the columns in raw data. The ETL engine system can be implemented and tested by looking at column formats that change due to characters, cress, and spaces when it will be imported into the ETL system and view the contents which often change dynamically.

Case studies on ETL systems are divided into 3 designs: ETL engine systems, Entity relation diagram (ERD) design, logical records structure (LRS), and master report.[12]

Theoretical and researched practical aspects that were conducted earlier involve the merging of several tables and aligning different tables as illustrated in the image below:

The image shows a screenshot of an Excel spreadsheet with multiple columns and rows of data. The columns are organized into several sections, including product information, pricing, and status. The data appears to be a consolidated view of products from three different marketplaces. The spreadsheet is titled 'BLIBLI - Excel' and shows a detailed list of items with their respective attributes and values.

Figure 1. A segment of tables and processes carried out previously from 3 marketplaces

Subsequently, the alignment of the three raw marketplace data formats, which proved to be less efficient in terms of time and human resources when executed at the herbal store, transformed into a report. This led the author to propose a prototype method. Within the report, in this ETL system, a level in the ETL process begins with the identification of

files through renaming and incorporating the database into the Microsoft Access system, turning the data into a format that can be analyzed according to user needs. An overview of the system can be observed in Figure 2.

This is Levels deep ETL process Beginning from Indicates a file with a mere name and making the database enter the system Microsoft access becomes Data that then in the analysis appropriate necessity user. Picture common system get views at figure 1

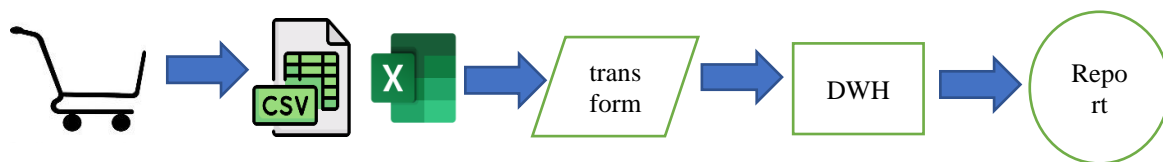


Figure 2. Strictly theory and practice that Researched deep system machine ETL

Figure 2 above shows a sequence of ETL processes from marketplace to report. Originally, the data used came from marketplace sales which were then downloaded, and then the process of extracting and transforming all data. In the ETL process, there is an equalization of data columns and a data scrubbing refinement process that is useful for seeing double data and different characters in each content of a data, in producing more quality data. Data that has gone through the scrubbing process is then loaded and saved to the herbal medicine store database which is a data warehouse database[2][13][14]

ETL Entity Relation Diagram System Sales data is divided into several tables which are RDBMS databases that can be used in prototype method processes in programs with query combination databases and denormalized, both query form tests can be used in the first generation Pentium i3 assembly computer, ETL has DDL and DML queries such as insert update delete which can run in 1 minute to 7 minutes from sales data at herbal stores, where the concept used is the concept of relational database management (RDBMS). [15][16]

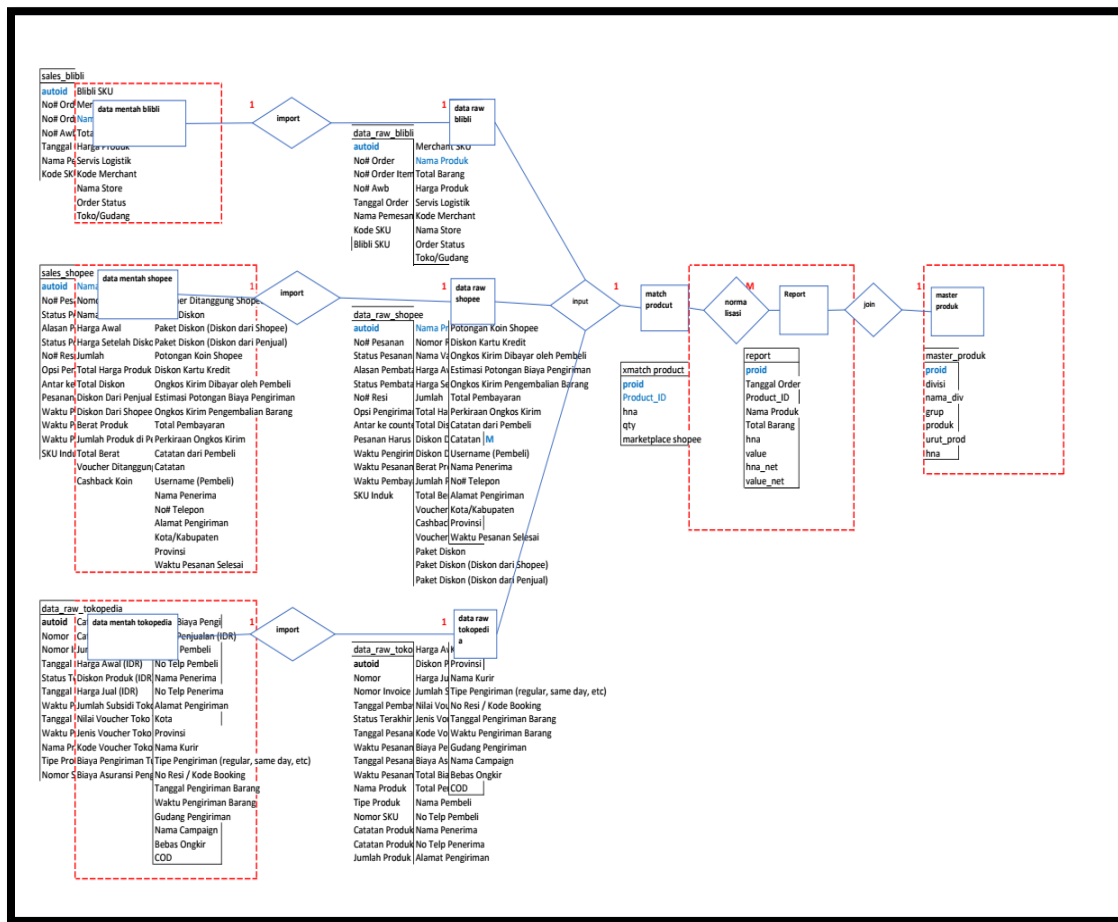


Figure 3. Data Warehouse in the form of ERD

Data Warehouse in the form of ERD figure 2, has 9 tables as in its function in each query and is interconnected by macros that can run automatically with 4 buttons graphical user interface, which are connected. To see the diagram in the Data Warehouse in Figure 3 stored data on all marketplaces, detailed tables of several marketplaces function to store prefix data more presented in column columns, and as material that serves to store data analysis materials.[14]

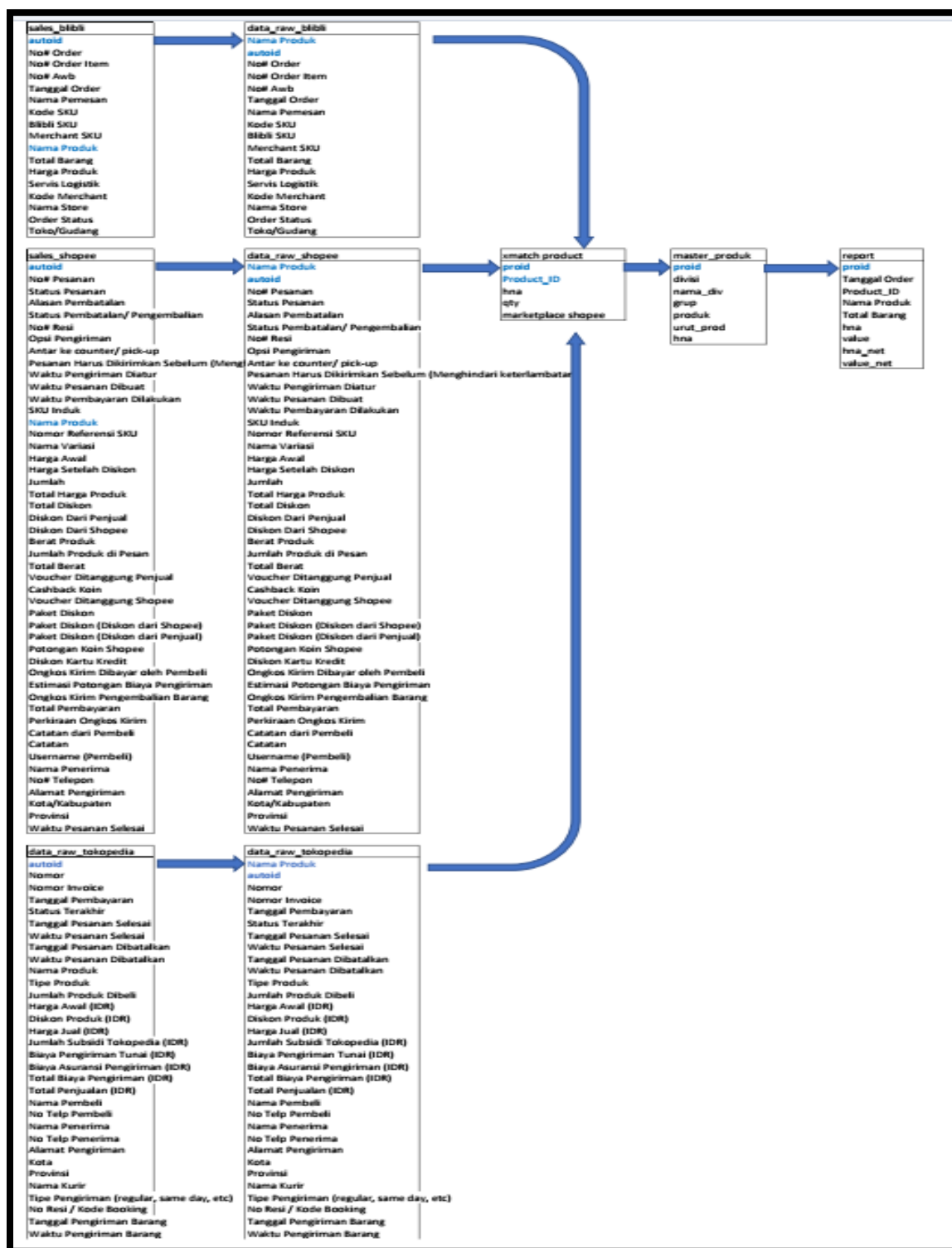


Figure 4. LRS ETL System

LRS ETL System, a flow of data from the ETL process from being stored in a database to the form of exported data raw, there are many contents of sales transactions, several marketplaces row by row and column by column.

Engineer RDBMS data warehouse software on LRS architecture. The data architecture consists of raw tables and master tables that are connected to raw tables from the marketplace. There are 2 raw tables, namely the raw sales table and raw table, and 6 normalization tables [17] where there are 3 raw fact data from the marketplace and 3 process data normalization data, product matching dimensions, combined raw fact dimensions from 3 marketplaces, product master dimensions, the report dimension. The data warehouse schema design can be seen in Figure 2.3.

Figure 2.3 above RDBMS data warehouse system design has a raw sales table and has several product fields as product codes, hna_net, and value_net that are connected (match/relation) between normalization-normalization tables as problem-solving [18]. In the product match dimension table, there are fields including proud, Product_ID as product, hna as net price, qty as several sales, market place code, which is divided between the normalization table of the product master table. The raw normalization table serves as a temporary table for all data connected in each field to make it easier to automate each data in the table.

3. RESULTS AND DISCUSSION

The implementation is designed to prevent duplicate data during the transformation process in each data mart from the three sources of raw data and dimension data. Therefore, each relation has its distinct function. The table master_group dimension data serves as the product master, and the table xmatch_product dimension data acts as the dimension for a collection of products, with a foreign key linking the dimension table to the fact table. The transformation query generates Analysis Data to support decision-making.

Below are three marketplace data sources for the data warehouse connected to the master data.

In figure 5, the fact data from Bibli is related to the MySQL data warehouse as shown below:

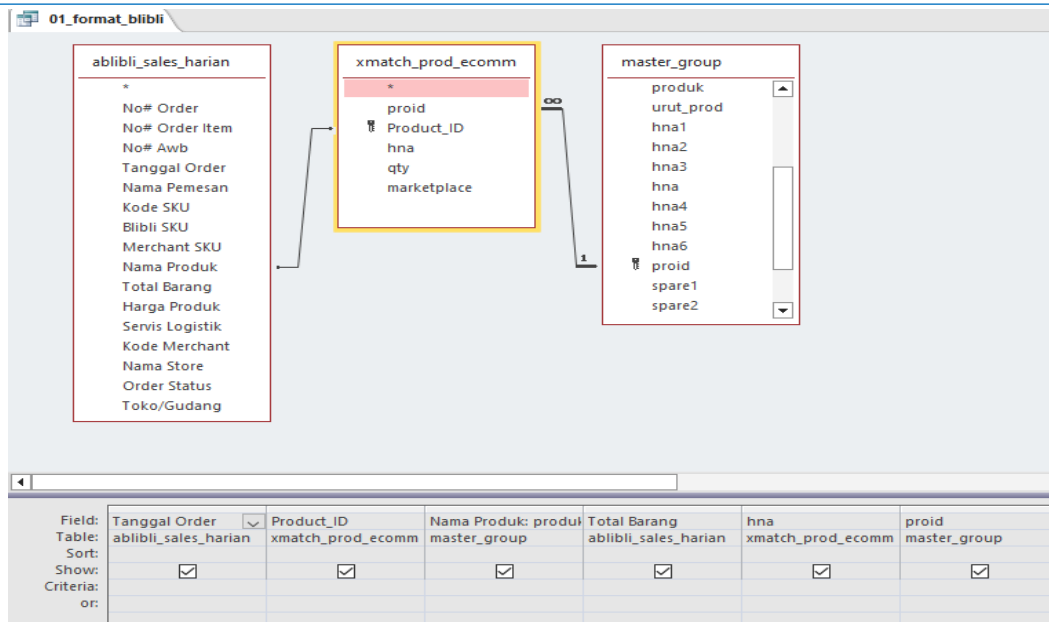


Figure 5. The data warehouse Blibli in the Microsoft Access ETL system

In the first example above, it presents data that includes sales and quantity analysis for products sold in the Blibli marketplace. In figure 6, the fact data from Shopee is related to the MySQL data warehouse as shown below:

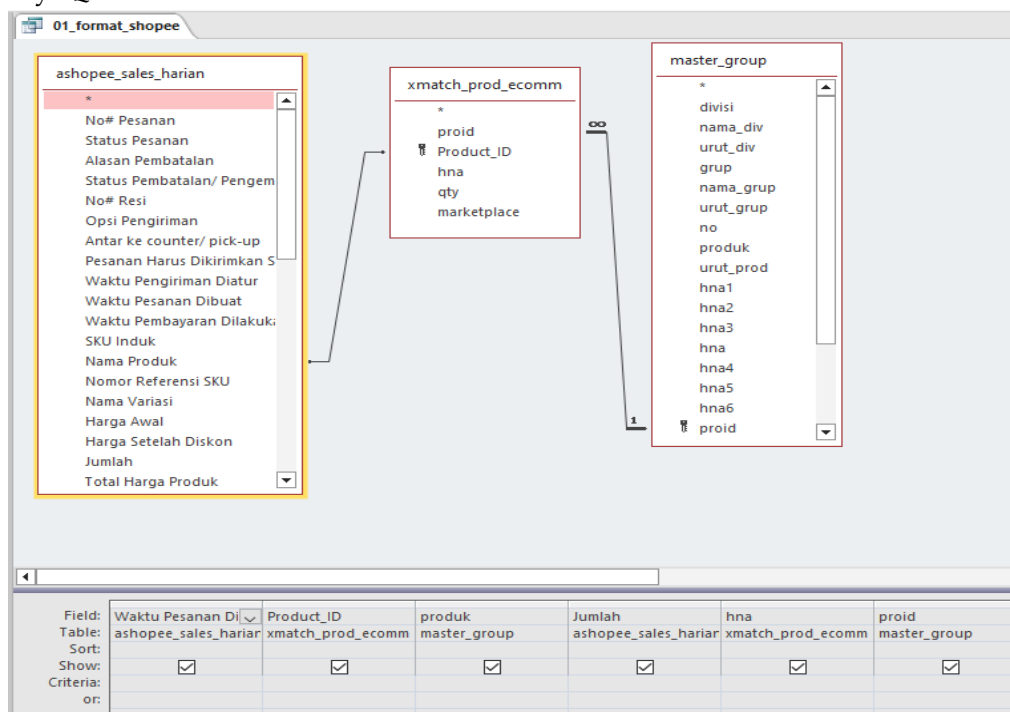


Figure 6. the datawarehouse Shopee in the Microsoft Access ETL system

In the second example above, it presents data that includes sales and quantity analysis for products sold in the Shopee marketplace. In Figure 7, the fact data from Tokopedia is related to the MySQL data warehouse as shown below:

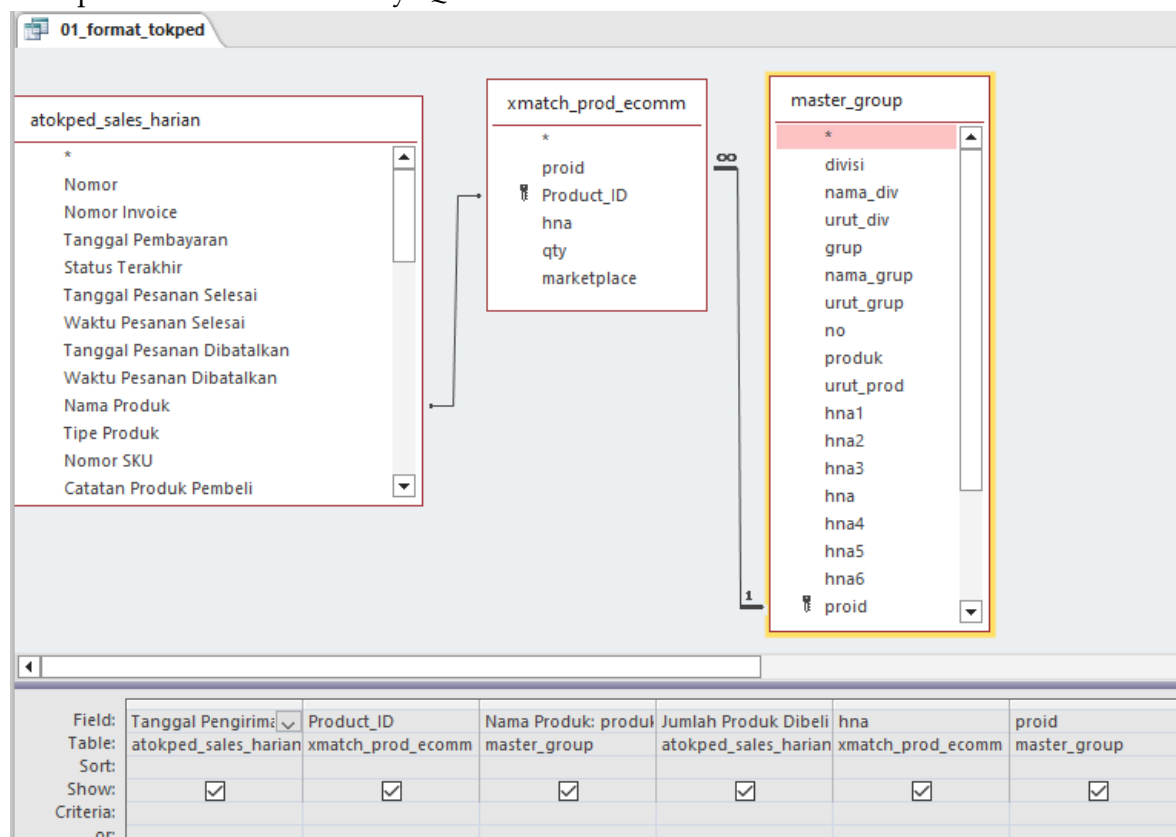


Figure 7. Datawarehouse Tokopedia in the Microsoft Access ETL system

In the third example above, it presents data that includes sales and quantity analysis for products sold in the Tokopedia marketplace.

All of these query data are processed automatically using Microsoft Access by macro tool, specifically using macro code. The results of this study implement ETL and produce reports, which have minimal costs in terms of hardware, and human resources because they daily use Microsoft Office, an application program that has been created with prototypes that have previously been designed with RDBMS data warehouse, so that in the previous discussion so that the ETL engine and raw report data can be used with pivot tables according to analysis needs. The ETL process works in 2 stages, namely the manual ETL process and the Automatic ETL process which works according to a button click (not yet scheduled) to process the delete import output process. Both of these processes have

previously been carried out in the scrub data process so that later the input data on the Report is neat, under what is desired in determining decision-making needs at the management level.

The form design feature of the ETL data warehouse in Toko Herbal serves to describe the data, defining the objects contained in each table in Toko Herbal. In this ETL form design feature, it can process data from several raw tables and normalized tables more accurately such as id, field name, data type, length, and description. [19]

This form design feature is made with the visual basic of application programming to make it easier for users to just click 4 buttons. Why is it made 4 to look manually to find out the origin table and destination table in doing the extract process at the herbal store if there is a difference between data and the query? This feature is also equipped with an automatic message field description that explains the completion of a process from every 4 buttons in each query, delete, import append, and export process. Keys are integrated into Microsoft Access macros, so that all query functions in each table, can be connected between tables, which contain primary keys and foreign keys. In this feature, users can also find out which fields will be carried out by the ETL process in the Toko Herbal data warehouse database. Form design and List table illustrated in 4.

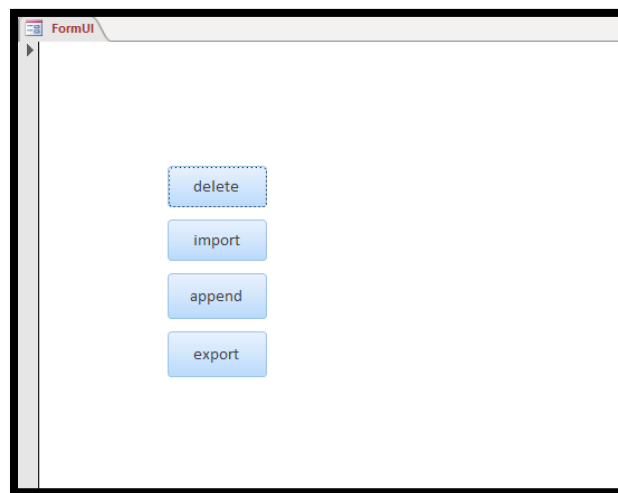


Figure 4. The form design feature of the ETL data warehouse

```
Private Sub imp_ecom_Click()  
On Error GoTo SubError  
  
DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel12, "sblibli_sales_harian", "C:\Users\leo\Downloads\e-commerce\BLIBLI.xlsx", True, ""  
DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel12, "stokped_sales_harian", "C:\Users\leo\Downloads\e-commerce\TOKOPEDIA.xlsx", True, ""  
DoCmd.TransferSpreadsheet acImport, acSpreadsheetTypeExcel12, "ashopee_sales_harian", "C:\Users\leo\Downloads\e-commerce\SHOPEE.xlsx", True, ""  
  
MsgBox " imported Success Lhooo!", vbInformation + vbOKOnly, "Success"  
  
SubExit:  
On Error Resume Next  
Exit Sub  
  
SubError:  
MsgBox "Error Number: " & Err.Number & " = " & Err.Description, vbCritical + vbOKOnly, _  
"An error occurred"  
GoTo SubExit  
End Sub  
  
Private Sub del_firast_Click()  
On Error GoTo Err_del_firast_Click  
  
Dim stDocName As String  
Dim stDocName1 As String  
Dim stDocName2 As String  
  
DoCmd.SetWarnings False  
  
stDocName = "00_mentah_blibli"  
stDocName1 = "00_mentah_shopee"  
stDocName2 = "00_mentah_tokped"  
  
DoCmd.OpenQuery stDocName  
DoCmd.OpenQuery stDocName1  
DoCmd.OpenQuery stDocName2  
  
MsgBox " Delete Success Lhooo!", vbInformation + vbOKOnly, "Success"  
  
Exit_del_firast_Click:  
Exit Sub  
  
Err_del_firast_Click:  
MsgBox Err.Description  
Resume Exit_del_firast_Click  
End Sub
```

Figure 5. ETL Features Automatic Process ETL produces an automated process

ETL Features Automatic Process ETL produces an automated process that is used to run ETL more efficiently without having to click through each query that runs all manual query process activities. [20]

The Automatic ETL process will run the data extract process, which queries delete and import files such as figure 4.2.1, transform data, and load data automatically from data that has been downloaded on a daily sales basis, [8]

```

Option Compare Database
Dim rs As New ADODB.Recordset
Private Sub app_ecom_Click()

    Dim stDocName As String
    Dim stDocName1 As String
    Dim stDocName2 As String

    stDocName = "01_format_blibli"
    stDocName1 = "01_format_shopee"
    stDocName2 = "01_format_tokped"

    DoCmd.OpenQuery stDocName
    DoCmd.OpenQuery stDocName1
    DoCmd.OpenQuery stDocName2

    MsgBox " Append Success Lhooo!", vbInformation + vbOKOnly, "Success"

Exit_app_ecom_Click:
Exit Sub

Err_app_ecom_Click:
MsgBox Err.Description
Resume Exit_app_ecom_Click

End Sub

Private Sub exp_ecom_Click()
On Error GoTo Err_exp_ecom_Click

    Dim stDocName As String

    DoCmd.SetWarnings False

    DoCmd.TransferSpreadsheet acExport, 9, "02_union_format_all", "C:\Users\leo\Downloads\e-commerce\sales_harian_format_" & Format(Date, "yyyymmdd") & ".xls", True, ""

    MsgBox " Export Success Lhooo!", vbInformation + vbOKOnly, "Success"

Exit_exp_ecom_Click:
Exit Sub

Err_exp_ecom_Click:
MsgBox Err.Description
Resume Exit_exp_ecom_Click

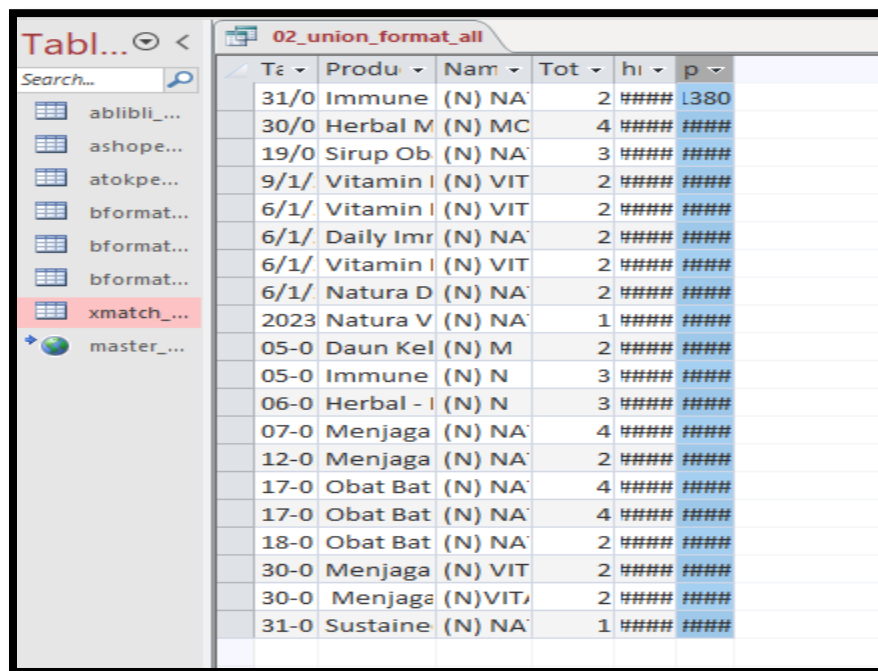
End Sub

```

Figure 6. Transform load macros visual basic of application

Figure 6 Transform load macros visual basic of application as above, after that, the ETL process runs by clicking the button that connects the vba syntax with Microsoft Access macros. The automatic ETL graphical user interface is a prototype of the display program of the Automatic ETL feature in the herbal marketplace store in Figure 4.1. There is an import auto number, the type in each field to determine ETL whether there are double transactions or not conditioned by creating a primary key and foreign key so that it will be automatically done when 4 buttons are clicked. After setting the Microsoft access popup and auto tools compact repair automatic, the ETL process at the herbal store will be automatically carried out according to what the user wants.

Report feature This research builds a data analysis system in an RDBMS data warehouse called ETL intended for the management level to determine and support decisions in a company that aims to help see decisions from pivot data. [21]



| Tε | Produ | Nam | Tot | hi | p |
|------|------------|---------|-----|------|------|
| 31/0 | Immune | (N) NA | 2 | #### | 1380 |
| 30/0 | Herbal M | (N) MC | 4 | #### | #### |
| 19/0 | Sirup Ob | (N) NA | 3 | #### | #### |
| 9/1/ | Vitamin I | (N) VIT | 2 | #### | #### |
| 6/1/ | Vitamin I | (N) VIT | 2 | #### | #### |
| 6/1/ | Daily Imr | (N) NA | 2 | #### | #### |
| 6/1/ | Vitamin I | (N) VIT | 2 | #### | #### |
| 6/1/ | Natura D | (N) NA | 2 | #### | #### |
| 2023 | Natura V | (N) NA | 1 | #### | #### |
| 05-0 | Daun Kel | (N) M | 2 | #### | #### |
| 05-0 | Immune | (N) N | 3 | #### | #### |
| 06-0 | Herbal - I | (N) N | 3 | #### | #### |
| 07-0 | Menjaga | (N) NA | 4 | #### | #### |
| 12-0 | Menjaga | (N) NA | 2 | #### | #### |
| 17-0 | Obat Bat | (N) NA | 4 | #### | #### |
| 17-0 | Obat Bat | (N) NA | 4 | #### | #### |
| 18-0 | Obat Bat | (N) NA | 2 | #### | #### |
| 30-0 | Menjaga | (N) VIT | 2 | #### | #### |
| 30-0 | Menjaga | (N) VIT | 2 | #### | #### |
| 31-0 | Sustaine | (N) NA | 1 | #### | #### |

Figure 7. This ETL system

This ETL system Figure 7 provides various types of reports such as sales reports for each marketplace every day and month or year which are united with certain append or union query systems in query builder tools formed by raw table data.[22]

4. CONCLUSION

The conclusion of the implementation and testing of the system that has been carried out in this study can be concluded that using an ETL engine that can work manually or automatically can make it easier for upper management to make decisions so that the ETL engine that is built is equipped with the scrub data feature, it is very important that this feature is to minimize the wrong input by the user, of course, it simplifies the ETL process, ETL engine software engineering is also equipped with a simple user interface button feature, the coding is also made simple according to the query builder and macros referring to naming in each process. This feature serves to find out existing table information to facilitate raw technical data users. The ETL engine that is built can extract, transform and load from the RDBMS Data Warehouse, and finalize data reports from the database illustrated when users can pivot tables as needed. The weakness of this system is that it cannot see process logs to be able to display the history or history of processes that have been carried out by the ETL engine. Report data view. However, at the small business level (Falchuk et al., 2021) built on this study, it can display sales transaction column data from

several marketplaces, and can easily add comparison data such as stock, costs, etc. to herbal stores.

REFERENCES

- [1] A. Rahmat Iskandar, A. Junaidi, and A. Herman, "Extract, Transform, Load sebagai upaya Pembangunan Data Warehouse," 2019.
- [2] I. Putu, W. Prasetia, I. Nyoman, and H. Kurniawan, "Implementasi ETL (Extract, Transform, Load) pada Data warehouse Penjualan Menggunakan Tools Pentaho," *TIERS Information Technology Journal*, vol. 2, no. 1, pp. 39–47, 2021, [Online]. Available: <https://journal.undiknas.ac.id/index.php/tiers>
- [3] E. Abdelhamid, N. Tsikoudis, M. Duller, M. Sugiyama, N. E. Marino, and F. M. Waas, "Adaptive Real-time Virtualization of Legacy ETL Pipelines in Cloud Data Warehouses Frequent Subgraph Mining View project Efficient Techniques for Modular Past State Systems View project Adaptive Real-time Virtualization of Legacy ETL Pipelines in Cloud Data Warehouses," 2023, doi: 10.48786/edbt.2023.64.
- [4] B. Rianto, F. Yunita, M. Amin, and R. Rahma, "Information Systems of Official Trips at General Election Commission Office (KPU)," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 14, no. 1, pp. 1–9, 2021.
- [5] D. Novaliendry and V. D. Puteri, "E-Retail Percetakan Anambaleh Desain Menggunakan Framework Laravel," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 13, no. 1, pp. 10–19, 2020.
- [6] F. Oktavia and Y. Hendriyani, "Perancangan E-Report Financial System Berbasis Web (Studi Kasus: Toko Kripik Balado Mahkota)," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 13, no. 1, pp. 87–95, 2020.
- [7] M. Hendayun, E. Yulianto, J. F. Rusdi, A. Setiawan, and B. Ilman, "Extract transform load process in banking reporting system," *MethodsX*, vol. 8, Jan. 2021, doi: 10.1016/j.mex.2021.101260.
- [8] D. Andriansyah, "Implementasi Extract-Transform-Load (ETL) Data Warehouse Laporan Harian Pool".
- [9] K. Y. Cheng, S. Pazmino, and B. Schreiweis, "ETL Processes for Integrating Healthcare Data - Tools and Architecture Patterns," in *Studies in Health Technology and Informatics*, IOS Press BV, Nov. 2022, pp. 151–156. doi: 10.3233/SHTI220974.
- [10] A. D. Barahama and R. Wardani, "Utilization Extract, Transform, Load for Developing Data Warehouse in Education Using Pentaho Data Integration," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Dec. 2021. doi: 10.1088/1742-6596/2111/1/012030.

-
- [11] F. L. Gaol, L. Abdillah, and T. Matsuo, "Adoption of Business Intelligence to Support Cost Accounting Based Financial Systems-Case Study of XYZ Company," *Open Engineering*, vol. 11, no. 1, pp. 14–28, Jan. 2020, doi: 10.1515/eng-2021-0002.
- [12] aljanabi alharqi haider, "An Optimal Warehouse Design for Crime Dataset," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 5, pp. 9080–9088, Oct. 2020, doi: 10.30534/ijatcse/2020/313952020.
- [13] G. Khan Ghajar, M. Plager, and S. Hallberg, "Conversational And Live Interactions Development And Marketplace Distribution System And Process," 2018, doi: 10.1109/I.
- [14] R. Venkatakrisnan, "Design, Implementation, and Assessment of Innovative Data Warehousing; Extract, Transformation, and Load(ETL); and Online Analytical Processing(OLAP) on BI," *International Journal of Database Management Systems*, vol. 12, no. 3, pp. 1–6, Jun. 2020, doi: 10.5121/ijdms.2020.12301.
- [15] J. C. Quiroz, T. Chard, Z. Sa, A. Ritchie, L. Jorm, and B. Gallego, "Extract, Transform, Load Framework for the Conversion of Health Databases to OMOP," 2021, doi: 10.1101/2021.04.08.21255178.
- [16] Mr. R. S. Ratnaparkhi, "The Extract Transform Load (ETL) Process and Optimization using Ab Initio," *Int J Res Appl Sci Eng Technol*, vol. 7, no. 5, pp. 1714–1717, May 2019, doi: 10.22214/ijraset.2019.5287.
- [17] Chatzistefanou Dimitra, "Data Warehousing in Business Intelligence and ETL Processes Chatzistefanou Dimitra," 2023.
- [18] alexander sychev, *Data Analytics and Management in Data Intensive Domains*, vol. 1427. in *Communications in Computer and Information Science*, vol. 1427. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-81200-3.
- [19] Ardhian Agung Yulianto, "Extract Transform Load process in distributed database academic datawarehouse," 2021.
- [20] S. M. Bouknight, "Automated Etl Workflow," 2022, doi: 10.1109/W1.2016.0083.
- [21] B.-E. B. Semlali, C. El Amrani, and G. Ortiz, "SAT-ETL-Integrator: an extract-transform-load software for satellite big data ingestion," *J Appl Remote Sens*, vol. 14, no. 01, p. 1, Jan. 2020, doi: 10.1117/1.jrs.14.018501.
- [22] A. Ta'a, N. Ishak, E. M. Elias, and N. Mahidin, "An Impact Analysis Of Extract Transform Load Process For Maintaining The System Of Data Warehouse," *Journal of Information System and Technology Management*, vol. 7, no. 27, pp. 168–186, Sep. 2022, doi: 10.35631/jistm.727014.