IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 2, June 2023, pp. 765∼775
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i2.pp765-775

Design and implementation of the web (extract, transform, load) process in data warehouse application

Seddiq Q. Abd Al-Rahman[1], Ekram H Hasan[2], Ali Makki Sagheer[1]
[1]Department of Computer Network Systems, University of Anbar, Anbar, Iraq
[2]College of Basic Education/Haditha, University Of Anbar, Anbar, Iraq

Article Info

Article history:

Received Jan 9, 2022
Revised Nov 3, 2022
Accepted Dec 3, 2022

Keywords:

Data warehouse
Dirty data
Fact table
Interim table
Web (extract, transform, load)

ABSTRACT

Owing to the recent increased size and complexity of data in addition to management issues, data storage requires extensive attention so that it can be employed in realistic applications. Hence, the requirement for designing and implementing a data warehouse system has become an urgent necessity. Data extraction, transformation and loading (ETL) is a vital part of the data warehouse architecture. This study designs a data warehouse application that contains a web ETL process that divides the work between the input device and server. This system is proposed to solve the lack of work partitioning between the input device and server. The designed system can be used in many branches and disciplines because of its high performance in adding data, analyzing data by using a web server and building queries. Analysis of the results proves that the designed system is fast in cleaning and transferring data between the remote parts of the system connected to the internet. ETL without missing any data consumes 0.00582 seconds.

Corresponding Author:

Seddiq Q. Abd Al-Rahman
Department of Computer Network Systems, University Of Anbar
Ramadi, Anbar, Iraq
Email: co.sedeikaldossary@uoanbar.edu.iq

1. INTRODUCTION

The past years witnessed an urgent need to increase the profits of and competitions between companies.so, companies have resorted to making quick and accurate decisions. Data warehouses (DWs) represent a modern technology that helps analyze and collect data and make appropriate decisions that are in line with the requirements of the work environment [1]. DW is a computer system designed to serve as a central repository of information generated from a set of data sources that represents the database application used to collect and analyze the historical information of an institution resulting from daily operations [1], [2]. Data flows from relational databases, such as from transactional systems to the target DW consist of multiple species of data, namely, structured, semi-structured and unstructured. These various data are subjected to processing, standardization and uploading and then employed in business intelligence. Users, such as business analysts, decision makers and data scientists, rely on various tools such as structured query language (SQL) clients, spreadsheets and business intelligence tools, to access data that are processed in DW [1].

In any organization, DW acts as a decision support system, and it relies on the historical data of the institution. Decision-makers' conclusions and decisions are primarily based on the results of the data analyzed. DW is used within the public domain across specific field [3]. DW was described by Inmon Bill as a time-variant, subject-oriented, non-volatile and integrated collection of data [1]. The key features of any DW are discussed as:

- Subject oriented: the first characteristic of a DW is that it is subject-specific and organized according to different commercial visions; in any DW, data can be analyzed in custom to any topic field [1], [3].
- Integrated: data are collected from multiple heterogeneous sources. In the DW environment, all these sources are combined and stocked in a mutual place [4]–[6].
- Non-volatile: all data in DW remain fixed; once then they are loaded and not permanently deleted as long as the historical data are relied upon [4], [5].
- Time variant: it is important to maintain historical data where historical positions of activities in time DW load. The data specified for each company have a retrieval limit, and data are retrieved according to the recovery limit from a period of three months to 10 years [7]–[9].

After introducing DW and its most important characteristics in this section, section 2 surveys some related studies. Section 3 presents DW applications and overview of the DW architecture. An explanation of the data extraction, transformation and loading (ETL) process of the proposed system is given in detail in section 4. The experiment results are described in section 5 and the conclusion is presented in section 6.

2. RELATED WORK

The DW and ETL have been adopted in many applications. A lot of research has been presented in this field. The following paragraphs provide details about recent studies and applications applied DW and ETL techniques.

In [2019] Jayashree and Priya proposed DW design principles for the supply chain. In a supply chain's environment, visibility is the most important factor because it helps reduce errors to achieve reliability for clients by the company. Visibility deals with a set of methods and strategies, including inventory, coordination and allocation strategies, that any company can employ to provide excellent service to its customers anywhere, any time and in the manner required. The advantages of the supply chain visibility are mostly realized as cost reductions. The case study for the proposed DW designed for order life cycle visibility involves sales representatives. It can reduce the time that sales representatives spend on pulling order information because they can access open orders by using mobile devices and emails. As a result, sales representatives can have more time to sell and provide customers with the information needed for order shipments. For the supply chain market, a previous case study presented the design architecture for order visibility. Logical data were employed to explain the systematic procedure of ETL series jobs. These data can be used in any order-processing system, such as the supply chain field [3].

In [2019] Homayouni *et al.* explained and developed a bank model. The developed bank model helps determine the reserve requirements of a bank, thereby helping increase customers' capabilities to obtain loans according to the bank's reserves. Loans are granted to customers. When they depend on the reverse, the performance of each branch is known, and the steps required to improve performance are determined depending on workhouse data reports, which also help the bank define a set of parameters to obtain abundant knowledge from the bank's DW [4].

In [2017] Fatima *et al.* recommended two approaches to access ideal applications for constructing an information distribution center for colleges. The first one was "Kimball approach (bottom-up)". It was recommended because of its high responses to customers' requests. The customers' needs are well developed from the beginning stage, which is the most accurate stage and characterized by the implementation of high demand. In this approach, data mart reports and analysis are created first then combined together to be the DW. This is a test for most relationships because client fundamentals and customers' requirements change often. In the second approach, the "Inmon approach (top-down design)", a relational data model collected from different resources is created first. The data pass through ETL stages. Afterwards, the dimensional data marts, reports and applications that contain the data needed for custom business processes or specified departments are outputted by the DW. The Kimball approach considers dimensional models, such as star schemas, that use dimensional tables and fact tables to order data in dimensional DWs, whereas the Inmon approach uses data mart as a separation between ETL and the final data [10].

In [2017] Khan *et al.*: in this study, an illustration of a query cache method was used to design an optimized DW before ETL processing, improve the developed design of the DW, and reduce the level of errors. The data coalition rule was used to identify errors, inappropriate data and defects. The purpose of this method is to store queries and the outputs related to them [11].

In [2017] Santoso and Yulia: in this study, the design and implementation of a useful educational DW for a higher education institution represented by the University of Basra in Iraq were presented. The proposed system was implemented based on two simulated databases, each of which depended on different constraints of ETL systems according to the type of data. The simulated databases were collected in two different periods from two different institutions: from the Department of Computer Science at the College of Science, University of Basra, for the last 10 years and from the University of Iraq for the past four years; then, they implemented the databases by using SQL server data tool (SSDT) 2012 and SQL Server 2014 [12].

The current study aims to design a DW system that contains a web ETL process that divides the work between the input device and server. The first step begins in the user's device to correct errors, where the input data are determined from different sources. Then, cleaning rules are selected, and cleaning preprocessing begins to detect and remove errors, standardize data formats and improve data quality. The data resulting from the cleaning processes are saved. Afterwards, the second step is performed to ensure non-repetition data. The data table is sent to a server, and a comparison is performed between the intermediate table and DW in the server. If data are not available, they will be loaded from Interim Table to DW in the server. This system avoids the problem of material depletion in the sub-warehouse, where the city warehouses that control the stock in the cities are linked to the general warehouse of the governorate's global store.

3. DW'S APPLICATIONS

DW technologies are employed in sundry fields for their efficient ETL methodology. DW has been widely used in various fields, where any organization or industry needs to conduct data analysis to determine the level of development and achieve growth [11], [13]. DW plays an active role in daily life and many other fields, such as medicine, business, banking, finance, web marketing, market segmentation and manufacturing, that consist of many stages, including process design, product planning, production and scheduling. Long-term strategic issues and profitability are closely related to and influenced by the decisions made, and many industries need decision-making systems [14]. DW technology is better than traditional decision-making techniques because it relies on different applications to collect, consolidate and store data, leading to the expansion of operations and increased efficiency. Notably, analyzing data in separate applications consumes much time. At this stage, to perform some manufacturing routines, transaction processing systems are used, and these are updated in a timely manner [15]–[17]. In this section, the most important issues related to ETL are addressed in terms of construction, design and stages. On this basis, an ETL is built, and the benefits of ETL are highlighted.

3.1. ETL process

The ETL process refers to a sequential set of operations that includes extracting, transforming and loading data [18]. ETL process is the core and backbone of the DW architecture [19]. It aims to standardize the processing of the source data and make the data available and appropriate for the intended use [7]. The role of ETL is to extract data from different sources, such as relational databases, flat files, web pages or any other type of data. The quality of data is consistent and good [1], [20], [21]. The separated resources are combined and used together to build a model for making strategic business decisions and to answer critical business intelligence queries [22]. Such data is processed, modified and subsequently loaded into another database [23]. The last pattern of data is the friendly user [24]. Web ETL is designed to communicate and interact with databases by reading diverse file formats within the organization [9]. Web ETL is a complex task that involves extracting many different types of data [11]. The extracted data go through an intermediate stage in which the data are scrubbed and cleaned to remove any inappropriate data and eliminate abnormalities (special characters and duplicate data) as well as to fill the missing values and attributes, identify and remove outliers, smooth and level the noisy data and determine and settle inconsistencies in transforming these data as needed [21], [3]. Business transformation rules (like applying calculations and concatenations) are then applied to the cleansed data, which are subsequently arranged consistently and taken for loading into the target DW [18]. Finally, the loaded data are made available for utilization. Such data can be generated and used as a part of the ETL process in several sources, including enterprise resource planning (ERP), centralized server application, document or an Excel spreadsheet and customer relationship management (CRM) [2], [25].

3.2. Stages of the ETL process

The ETL process involves three phases. These phases represent the main steps of our system. They should be implemented in a consecutive way. The phases are described in sequence as:

Extraction: external data are extracted from various source systems to an intermediate stage [3], [19]. In this stage, transformations are conducted to ensure that the performance of the source system does not deteriorate. In the staging layer, the data extracted from different sources undergo full validation before being loaded into DW to ensure that no corrupted data exist at the source because the procedure of retracting the data after loading from damaged data is a challenge. As a part of the ETL process, DW provides integration among systems that contain different hardware [17]. Before the data are extracted and physically loaded, logical mapping must be conducted on the data. Data mapping defines the relationships between the source and the target. The following shows examples of data extraction techniques: i) complete charge extraction, ii) partially extract load - without any update notification, and iii) partially extract load - accompanied by a notice. The following shows examples of the types of validations for data extraction: i) data mapping from source to target records, ii) removal of unwanted data, iii) verification of data kinds, and iv) removal of duplicate and split data [9], [10].

Transformation: the data extracted from the source layer are raw and not directly usable, so they must undergo cleaning and optimization so that they can be converted and easily dealt with. In this step, a set of functions or rules is applied to some extracted data, with the exception of some data that do not need conversions (direct transfer or pass-through data). During the transformation step, many customized operations are performed on the data. An example is when the end user needs the total sales income and it is not available in the database [22], [26].

Loading: it is the final step in ETL. After the conversion of data into a specific DW format, the data are loaded to the target DW scheme. In DW, data loading is time dependent because a huge volume of data is loaded within a short period of time. Developed performance optimization techniques are included in the loading process to improve performance. To maintain data integrity, failed data loading should be avoided. To ensure data preservation and avoid data loss, recovery mechanisms are prepared [3]. All stages mentioned above are clarified in Figure 1.

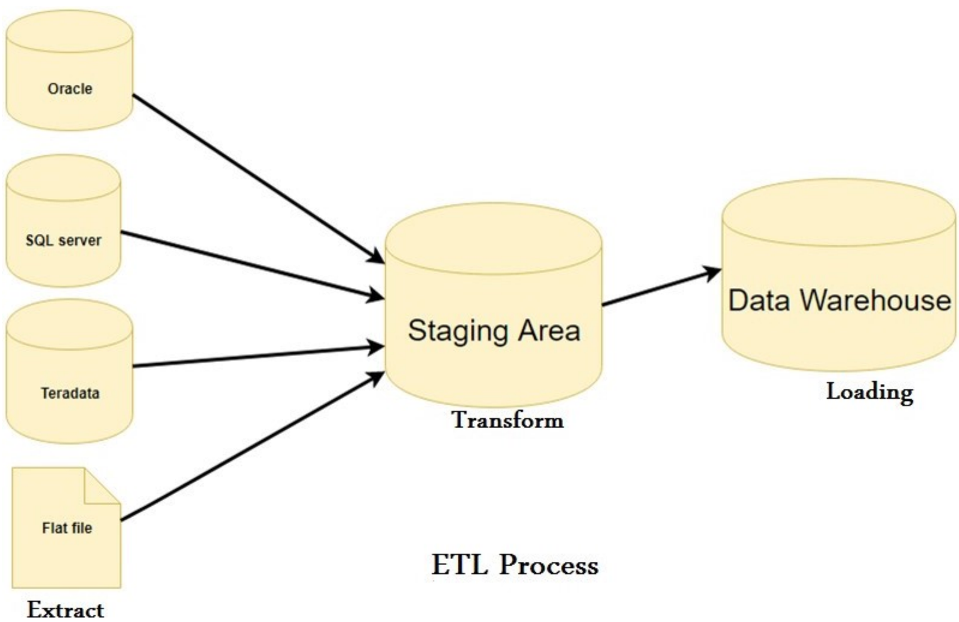

Figure 1. ETL process steps

3.3. Benefits of ETL

ETL has many benefits to the field it applied to. It has been adopted on a large scale in many institutions. Here we are listing some of the ETL benefits: i) one of the most important benefits of using ETL is to deal with complex business needs that the transac- tional databases cannot handle; ii) DW is a common repository for sharing and reusing data; iii) ETL transfers data to DW from different sources; iv) the modifications on the sources of data (e.g., adding, deleting and updating) are made automatically in DW; and v) ETL allows for integration between the sources of data and the final system because it provides a comparison of data samples with the source and the final system.

4. PROPOSED SYSTEM

This section presents additional details of the proposed system. It also shows the design of the DW that contains a we ETL process that divides its work between the input device and server. The proposed system solves the lack of work partitioning between the input and device and server.

4.1. General description of the proposed system

The proposed system is based on setting a DW in each city. The warehouse maintains the provision of a minimum level of paint products in each city distributed among the sales centers. The warehouses of the cities controlling the stock in the cities are connected to the general warehouse, which is connected to the global warehouse of the governorate. When ordering paint materials, the salesman supplies customers with products available at his center. The proposed system transfers the orders to other sales centers in the city in the event that any product reaches the minimum stock level. The DW within the city prepares weekly reports for administration officials. The reports contain materials whose stock has reached two folds the minimum level for their preparation. The reports also show what paint materials were spent during this week in each center. In addition, the reports include sale prices and profits arising from the sale. The most important thing the reports focus on is the colors spent in each city, their type and their use. City warehouse officials request to supply the shortage of paint materials from the governorate warehouse. the county store officials are provided with reports for each city. The city warehouse report contains the consumed materials, the desired colors, their quality and their use. It also contains the selling prices and profits obtained. Figure 2 illustrates scenario of data movement and coating material in the proposed system.

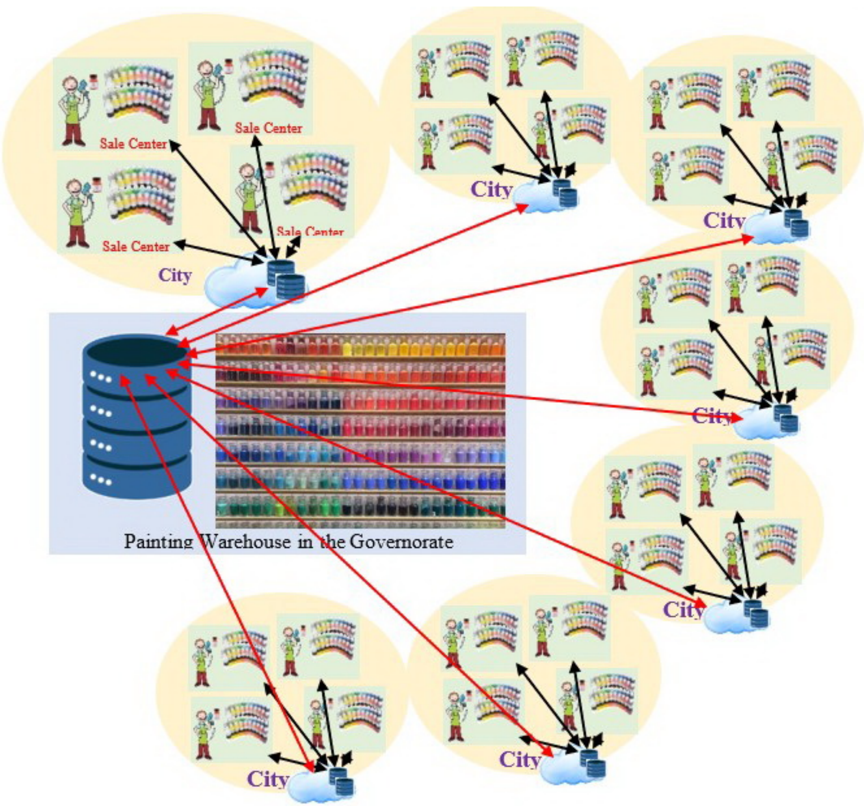

Figure 2. Scenario of proposed system

Within the proposed system, there is an administrator in each city and an administrator of the governorate warehouses. Therefore, the responsibilities of administrators differ according to the connection to the data warehouse that they deal with. The county store official enters the materials' data and prices according to the important information. In addition, it distributes dyes to cities according to their consumption. The city administrator distributes the dyes he received to his sales centers.

4.2. Data sources

The system has been tested on two different datasets. These datasets has been collected in two different ways. The process of collecting these datasets is explained as: the first source collects data from five offices and every office includes four centers to sell dyes in different areas in Al Anbar Governorate. The data sources differ according to the system or application used, such as spreadsheet, access database and data FoxPro) and the total number of data fields from the five offices in the various sales centers is approximately 5,809,600. Manual data entry is the second source, and the total number of data fields from this source is 1,691,611. The data sources used are varied. Data collection is conducted in eight months. After data collection, a DW has a total of 7,501,211 fields.

4.3. Web with ETL

The first stage of the ETL process extracts data from sources. The data extracted are incomplete and not usable in their original form, so after the extraction of all data from these sources, the data are forwarded to the second stage, which includes cleaning the input data (e.g., correcting spelling mistakes in entering data) and transforming them to the standardized form. Such transformation refers to the color name and date of product input in more than one form. For example, 01-01-2020, 01 January 2020 and 01/01/2020, should be unified into one form. All data after preprocessing and cleaning are aggregated in an interim table (ITbl), which represents the data staging area. Then, non-repetition preprocessing begins to ensure that the data are not duplicated. This table sends the match between the data in ITbl and the data in the server to the server to test whether the data have been previously entered or not. Data are loaded from ITbl to DW if they are unavailable. The final step of the ETL process a multidimensional DW is made through a star scheme with the data uploaded within a special server purchased from the site smarterasp.net, Which is characterized by high security by using the rivest-shamir-adleman (RSA) Algorithm 1 of web ETL.

Algorithm 1 Web-ETL design

Require: Data form input devices
Ensure: Cleaning data then Loading to Data Warehouse (DW) on web
In User Device: Determine input data and the cleaning rules are selected.
Start Cleaning Pre-processing.
Determine values from input with attributes of DW $Values \in DWinServer$.
Testing the mistakes and missing in input data.
The appropriate logical methods use to correct errors and missing of data.
End Begin.
Create interim table (ITbl), ITbl ← The data after cleaning.
Begin Non-repetition Preprocessing.
Connected with DW in the Server.
Sending ITbl to Server.
if (Data duplication between ITbl and DW = Yes) **then**
    Return Message ("Data are already entered") to User Device.
else
    Return Message ("Data are not entered") to System on User Device.
end if
End Begin.
if (System Receive Message) **then**
    Loading data from ITbl to DW.
end if
Close DW and Disconnect.
In User Device: Show the message "Adding data successfully".

4.4. Star scheme of the DW in the proposed system

DWs are multidimensional databases used in analytical processing. According to the types of DW scheme DW design can rely on the star data warehouse scheme in which the center represents the fact table.

It contains the ID of each of the other dimensions (three dimensions), fact table control and distributed data between tables. It also contains many attributes e.g., (date_in, time_in). Our DW scheme includes the following attributes, which are shown in Figure 3.

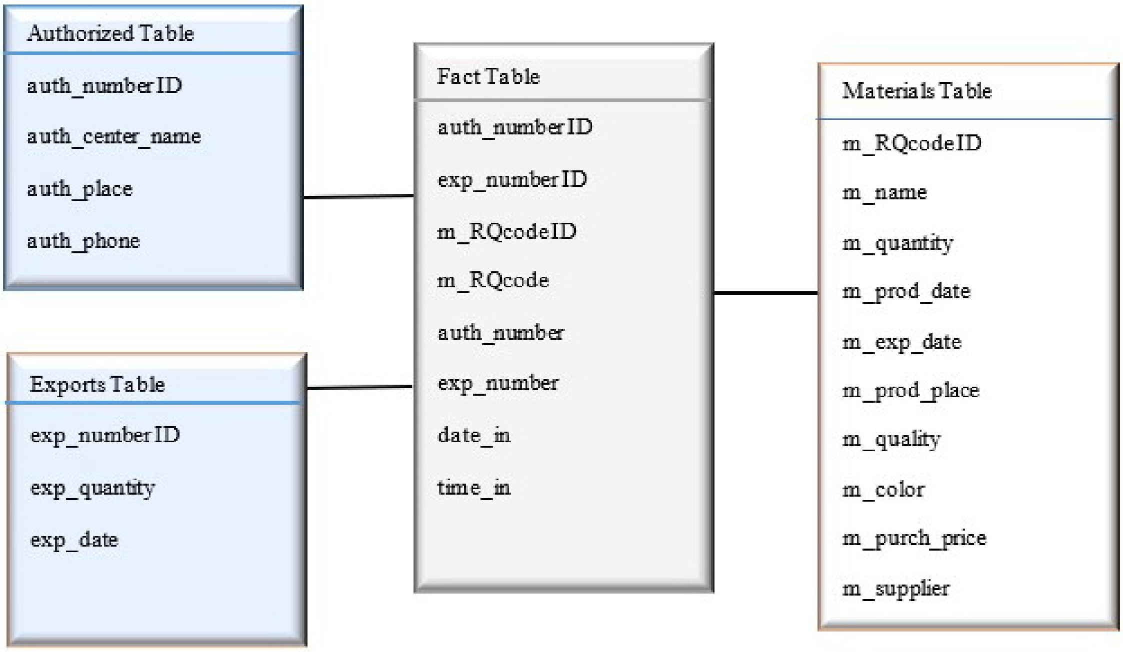

Figure 3. Data warehouse scheme

- Fact table (ID, m_RQcode, auth_number, exp_number, date_in, time_in).
- Materials table (m_RQcode, m_name, m_quantity, m_prod_date, m_exp_date, m_prod_place, m_quality, m_color, m_purch_price, m_supplier).
- Authorized table (auth_number, auth_center_name, auth_place, auth_phone).
- Exports table (exp_number, exp_quantity, exp_date).

4.5. Interface of the proposed system

The proposed system uses a mobile phone's applications programmed in Java language that works on android devices. This application is used by the owners of sales centers spread in cities. Barcode reader devices are also linked to mobile phones because they are dependent on the work. The proposed system uses program language C# from Visual Studio 2017. In addition, it uses (active server pages) ASP.net graphics interfaces for city and county warehouse administrators; this is attributed to the existence of numerous settings and wider workspace. Internet browsers on computers Figure 4 and smart devices Figure 5 can view them without any need to have special devices. This system distributes to the same office and its cell center.

5. RESULTS AND DISCUSSION

The system has been evaluated depending on three factors. These factors are widely used to show the performance of the system. The following sections explain all the details about these factors.

5.1. Execution time for ETL

Time is one of the most important factors that are considered to test the efficiency of any application. The proposed system is fast. At the data entry stage, the system does not consume much time. The data presented are characterized by speed that meets the needs of decision-makers and conforms to their conditions. At the analysis stage, a few seconds are used to extract the results and make the most appropriate decision. Table 1 clarifies these statements.

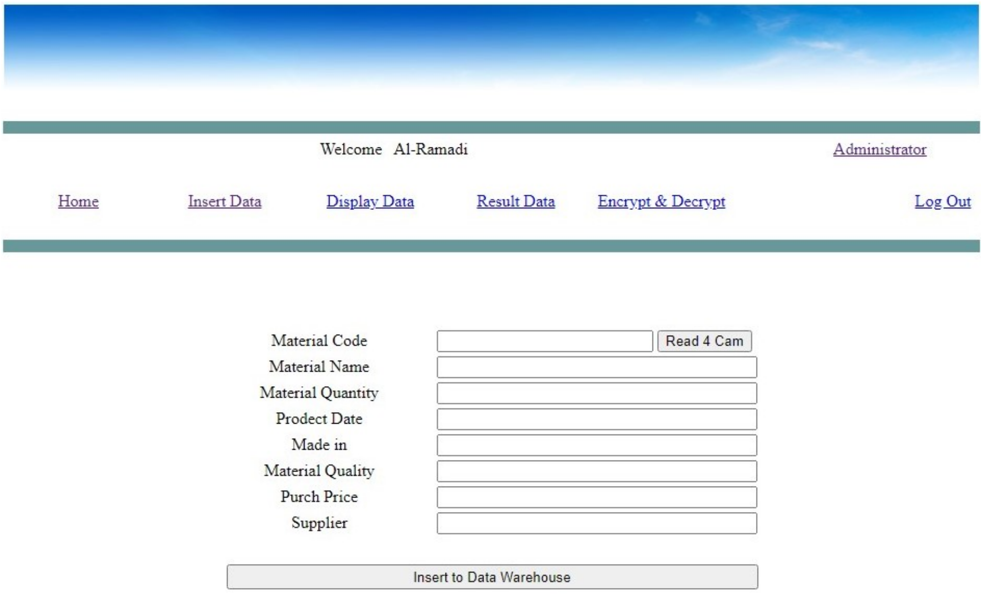

Figure 4. Computer interface for application

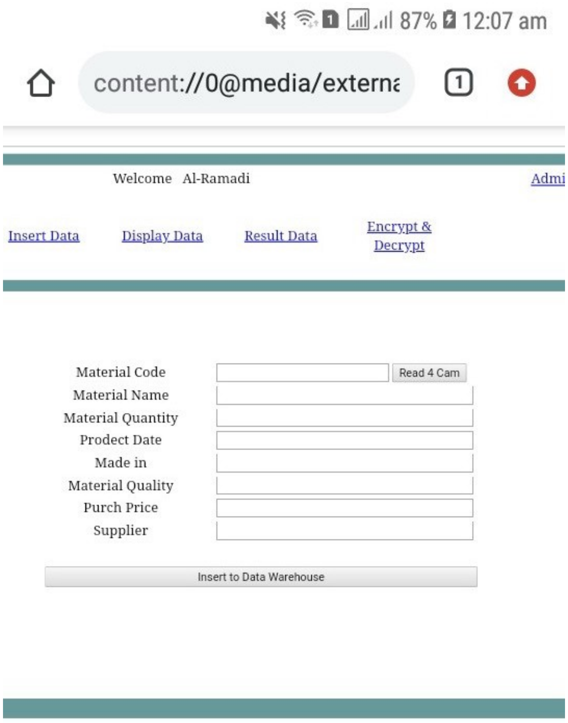

Figure 5. Mobile interface for application

Table 1. Execution time for ETL

Subject	Fields	Time in seconds
ETL without missing	Some	0.01570
	All	0.00582
ETL with missing	Some	0.01816
	All	0.03427

5.2. Speed data exchange between DWs

The speed of data exchange between the central DW (mother DW) in the governorate and the sub-warehouses (small DW) in each city is calculated, as illustrated in Table 2. The speed calculation is based on three data sizes of 10, 50 and 100 kB under the premise that the speed of internet used is constant on all sides. The internet speed is 512 download and 128 upload Figure 6.

Table 2. Speed data exchange between data warehouses

From	To	Data in KB	Time in seconds
Main input data	Mother DW	10	0.00298
Main input data	Mother DW	50	0.00735
Main input data	Mother DW	100	0.01098
Mother DW	Small DW	50	0.01163
Mother DW	Small DW	100	0.04710
Mother DW	Small DW	150	0.09026
Small DW	User interface	10	0.02196
Small DW	User interface	50	0.06004
Small DW	User interface	100	0.09714

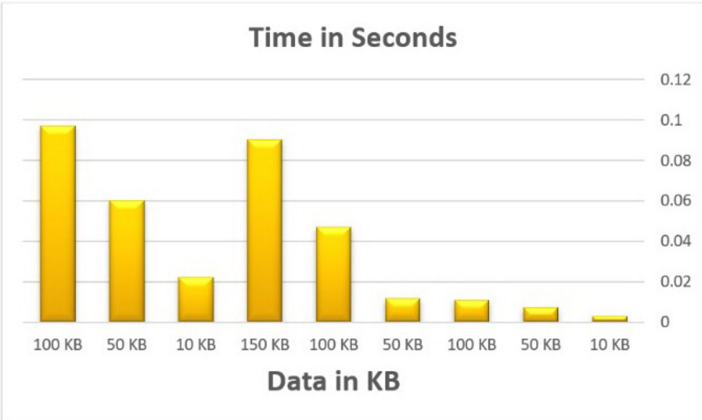

Figure 6. Time speed of data exchange between DWs

5.3. System evaluation

The system provided good results. However, it's not possible to proof its accuracy without an evaluation that shows how this system is doing. In this section we show the performance of our system. The system has many features and they are summarized as :

- Validity (business rule conformance): the proposed system is characterized by its conformity with working conditions of other institutions, where it can be implemented on different data systems whilst taking into account the needs of the institution with a slight change in the names and number of columns.
- Interactive with the user: the interfaces of the system are easy to use, clear to the user and very understandable. The user's interaction with these interfaces will lead to his interaction with the algorithm of the proposed system.
- Performance: the proposed system is characterized by its high performance in entering data, analyzing data and building queries by using a web server, so it can be used in many other disciplines and branches.
- Accuracy: The best display options can be relied on to improve the accuracy of queries. Queries are built with a new resolution technology that deals with web options and is the best way to improve performance.
- Ease of implementation: the proposed system can be implemented using other programming languages, which provides flexibility in implementing the system.

6. CONCLUSION

ETL is a major and vital part of DW. It fetches data from its main database to the destination DW. In this study, a mobile phone application representing a DW with web ETL is designed. The proposed system is based on two DW. The main DW (painting warehouse) is in the governorate, and the sub-warehouse in each city. The warehouse maintains the provision of a minimum level of paint products in each city distributed among the sales centers. The efficiency of this application in terms of speed of response to demand, speed of performance, interaction with the customer, ease of implementation and accuracy is proven in this work. In the future, the proposed system will be used in data analysis.

REFERENCES

[1] P. Gupta, "A book review on 'data warehousing, data mining, OLAP'," *JIMS8M: The Journal of Indian Management Strategy*, vol. 24, no. 2, p. 64, 2019.

[2] A. Bramantoro, "Data cleaning service for data warehouse: an experimental comparative study on local data," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 16, no. 2, pp. 834–842, 2018, doi: 10.12928/TELKOMNIKA.V16I2.7669.

[3] G. Jayashree and C. Priya, "Design of visibility for order lifecycle using datawarehouse," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, pp. 4700–4707, 2019, doi: 10.35940/ijeat.F9171.088619.

[4] H. Homayouni, S. Ghosh, and I. Ray, "Data warehouse testing," in *Advances in Computers*, vol. 112, Elsevier, 2019, pp. 223–273.

[5] C. G. Schuetz, S. Schausberger, and M. Schrefl, "Building an active semantic data warehouse for precision dairy farming," *Journal of Organizational Computing and Electronic Commerce*, vol. 28, no. 2, pp. 122–141, 2018, doi: 10.1080/10919392.2018.1444344.

[6] N. Rahman, "Data warehousing and business intelligence with big data," in *Proceedings of the International Annual Conference of the American Society for Engineering Management*, 2018, pp. 1–6.

[7] T. F. Efendi and M. Krisanty, "Warehouse data system analysis PT. Kanaan Global Indonesia," *International Journal of Computer and Information System (IJCIS)*, vol. 1, no. 3, pp. 70–73, 2020, doi: 10.29040/ijcis.v1i2.26.

[8] M. Golfarelli and S. Rizzi, "From star schemas to big data: 20+ years of data warehouse research," *A comprehensive guide through the Italian database research over the last 25 years*, pp. 93–107, 2018, doi: 10.1007/978-3-319-61893-7_6.

[9] P. Chandra and M. K. Gupta, "Comprehensive survey on data warehousing research," *International Journal of Information Technology (Singapore)*, vol. 10, no. 2, pp. 217–224, 2018, doi: 10.1007/s41870-017-0067-y.

[10] A. Fatima, N. Nazir, and M. G. Khan, "Data cleaning in data warehouse: a survey of data pre-processing techniques and tools," *International Journal of Information Technology and Computer Science*, vol. 9, no. 3, pp. 50–61, 2017, doi: 10.5815/ijitcs.2017.03.06.

[11] F. A. Khan, A. Ahmad, M. Imran, M. Alharbi, Mujeeb-ur-rehman, and B. Jan, "Efficient data access and performance improvement model for virtual data warehouse," *Sustainable Cities and Society*, vol. 35, pp. 232–240, 2017, doi: 10.1016/j.scs.2017.08.003.

[12] L. W. Santoso and Yulia, "Data warehouse with big data technology for higher education," *Procedia Computer Science*, vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.

[13] N. Rahman, "A simulation model for application development in data warehouses," in *Research Anthology on Agile Software, Software Development, and Testing*, vol. 2, IGI Global, 2022, pp. 819–834.

[14] N. A. Farooqui and R. Mehra, "Design of a data warehouse for medical information system using data mining techniques," in *PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing*, 2018, pp. 199–203, doi: 10.1109/PDGC.2018.8745864.

[15] I. Sutedja, P. Yudha, N. Khotimah, and C. Vasthi, "Building a data warehouse to support active student management: analysis and design," in *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech*, 2018, pp. 460–465, doi: 10.1109/ICIMTech.2018.8528196.

[16] R. Galici, L. Ordile, M. Marchesi, A. Pinna, and R. Tonelli, "Applying the ETL process to blockchain data. prospect and findings," *Information (Switzerland)*, vol. 11, no. 4, p. 204, 2020, doi: 10.3390/INFO11040204.

[17] V. M. Ngo, N. A. Le-Khac, and M. T. Kechadi, "Designing and implementing data warehouse for agricultural big data," in *International Conference on Big Data*, 2019, pp. 1–17, doi: 10.1007/978-3-030-23551-2_1.

[18] N. U. Ain, G. Vaia, W. H. DeLone, and M. Waheed, "Two decades of research on business intelligence system adoption, utilization and success – a systematic literature review," *Decision Support Systems*, vol. 125, p. 113113, 2019, doi: 10.1016/j.dss.2019.113113.

[19] A. Dhaouadi, K. Bousselmi, M. M. Gammoudi, S. Monnet, and S. Hammoudi, "Data warehousing process modeling from classical approaches to new trends: main features and comparisons," *Data*, vol. 7, no. 8, p. 113, 2022, doi: 10.3390/data7080113.

[20] S. L. Visscher, J. M. Naessens, B. P. Yawn, M. S. Reinalda, S. S. Anderson, and B. J. Borah, "Developing a standardized healthcare cost data warehouse," *BMC Health Services Research*, vol. 17, no. 1, pp. 1–11, 2017, doi: 10.1186/s12913-017-2327-8.
[21] C. A. Ul Hassan, R. Irfan, and M. A. Shah, "Integrated architecture of data warehouse with business intelligence technologies," in *ICAC 2018 - 2018 24th IEEE International Conference on Automation and Computing: Improving Productivity through Automation and Computing*, 2018, pp. 1–6, doi: 10.23919/IConAC.2018.8749017.
[22] C. Costa and M. Y. Santos, "Evaluating several design patterns and trends in big data warehousing systems," in *International Conference on Advanced Information Systems Engineering*, 2018, pp. 459–473, doi: 10.1007/978-3-319-91563-0_28.
[23] H. J. Watson, "Recent developments in data warehousing," *Communications of the Association for Information Systems*, vol. 8, no. 1, pp. 1–25, 2002, doi: 10.17705/1cais.00801.
[24] K. Dehdouh, O. Boussaid, and F. Bentayeb, "Big data warehouse: building columnar NoSQL OLAP cubes," *International Journal of Decision Support System Technology*, vol. 12, no. 1, pp. 1–24, 2020, doi: 10.4018/IJDSST.2020010101.
[25] P. S. Diouf, A. Boly, and S. Ndiaye, "Variety of data in the ETL processes in the cloud: state of the art," in *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, 2018, pp. 1–5, doi: 10.1109/ICIRD.2018.8376308.
[26] M. Souibgui, F. Atigui, S. Zammali, S. Cherfi, and S. Ben Yahia, "Data quality in ETL process: apreliminary study," *Procedia Computer Science*, vol. 159, pp. 676–687, 2019, doi: 10.1016/j.procs.2019.09.223.

BIOGRAPHIES OF AUTHORS

Seddiq Q. Abd Al-Rahman is a lecturer at the College of Computer Science and Information Technology, where he has been an employee since 2007. He was Rapporteur of the Computer Network Systems Department. From 2007–2016, he taught many laboratory materials and programming languages. Abd Al-Rahman completed his MSc degree in the University of Anbar at 2019 and his undergraduate studies in the same university. He teaches Computer Science in general, cryptography, visual programming and structure programming. His areas of interest are focused on the approaches for securing transfer data and storing them with intelligent methods. He is also interested in handling smartphone applications. He can be contacted at email: co.sedeikaldossary@uoanbar.edu.iq.

Ekram H Hasan received her BSc degree in Information Systems (2008) and MSc degree in Computer Science (2014) from the College of Computer and Information Technology, University of Anbar Iraq, Ekram since 2016, she has been working in the college of Basic Education/Haditha at University of Anbar. Her research interests include computer networks, data security and network security. She can be contacted at email: ekram.habeeb@uoanbar.edu.edu.iq.

Ali Makki Sagheer was born in Iraq-Basrah in 1979. He obtained his B.Sc. degree in Information System from the Computer Science Department of the University of Technology (2001)-Iraq, his M.Sc. in data security from the University of Technology (2004)-Iraq and his Ph.D. in Computer Science from the University of Technology (2007)-Iraq. He is interested in cryptology, information security, cyber security number theory, multimedia compression, image processing, coding systems and artificial intelligence. He has published more than 70 papers in different scientific journals and conferences and 24 of them have been published in Scopus. He obtained his professor of science degree in cryptography and information security on July 18, 2015. He can be contacted at email: ali.m.sagheer@gmail.com.