Journal of Engineering Science and Technology
Vol. 17, No. 1 (2022) 0275 - 0293

IMPLEMENTING SALES DECISION SUPPORT SYSTEM USING DATA MART BASED ON OLAP, KPI, AND DATA MINING APPROACHES

JASIM MOHAMMED DAHR[1,*], ALAA KHALAF HAMOUD[2], IHAB AHMED NAJM[3], MOHAMMED IMAD AHMED[2]

[1]Directorate of Education in Basrah, Basrah, Iraq
[2]Department of Computer Information Systems, University of Basra, Basrah, Iraq
[3]College of Applied Engineering, Tikrit University, Iraq
*Corresponding Author: Jmd20586@gmail.com

Abstract

Organizations and companies endeavour to expand their resources, margin of profit, and a robust management system. Managers seek tools that adapt the organizational data into information to support their strategic decisions to enhance the organization in terms of performance and profit. A challenging issue helps managers and analysts in making accurate departmental decisions on the basis of the analytical results. A large amount of historical data can be converted into information to support short and long-term strategic decisions. The decision support system (DSS) is one of the essential tools that can handle these data to provide analytical results that help in making the right decisions. This study presents a framework for designing and implementing sales. The framework module explains the DSS implementation. First, the framework establishes data pre-processing. Second, the framework extracts, transforms, and loads from the level of operational data from a data warehouse. Third, building sales cube and knowledge-driven sales is supported on the basis of three data mining approaches (decision tree, clustering, and neural network). The basic elements of the DSS (key performance indicator (KPIs), reports, and knowledge-driven sales supporter) are implemented in the next stage. Finally, the web interface is performed to provide access for sales managers. DSS used data mart on the basis of online analytical processing and KPI as fundamental tools. The design and implementation of a data mart utilize an SQL server database management system and SQL server data tools. The adopted source data were sales data of two years (2017 and 2018) of Basra Governorate. The proposed knowledge-driven sales DSS assists in decision-making. Such a system may assist managers in evolving new insights and strategies. The proposed technique, including the OLAP operation, presents the analytical results in addition to the KPIs that provide the indicators when a critical situation affects the overall profit or performance.

Keywords: Data mart, Data mining, DSS, KPI, Sales, OLAP.

1. Introduction

The organizations necessitate having adequate and agile access to strategic information to construct precise decisions. This strategic information extracted from the massive operational data is stored in the organization's databases [1]. The scope of the decision support system (DSS) functionalities may take two research aspects, namely, technical and theoretical. The technical aspect used the interactive information technological systems, while the theoretical aspect based on decision-making capabilities. DSSs are highlighted in the studies in mid 70s; however, the studies associated with DSSs increase in the '80s according to numerous disciplines, such as simulation, artificial intelligence, man to machine interaction, and databases [2, 3].

A DSS can be defined as an informatics system that assists organizational or business decision-making actions. DSSs aid various organizational levels (namely, planning, operations, and management) commonly for high and mid-management, and they also support people to take decisions concerning issues that may be quickly adapted and not mainly determined in advance. For instance, DSS offers the following features: expandable information technological tools plus techniques designed for analysing and processing data to help managers' decision-making activities. DSS as an information technology (IT) system can assist managers to take the accurate decision or select among various alternative options by providing the estimated value for each option to grant managers toward reviewing the conduct results. However, the DSS functionalities lack certain functions of the DSS, such as passive, active, collaborative, model-driven, communication-driven, data-driven, document-driven, and knowledge-driven [2, 4]. Consequently, significant outcomes are obtained by employing machine learning capabilities to optimize network proficiency [5] and predict student's academic performance [6-11]. The performance improvement of the mobile wireless communication is based on multi-criteria decision making to select the optimal parametric setting [12].

The evolution of e-commerce that is utilized by organizations generated a tremendous number of sales transactions. The data necessitate a robust storage mechanism besides an agile strategy for managing and analysing these data to support the making of a timely strategic decision. The characteristics of data in a data warehouse (DW) are subject-oriented. The selection is according to specific aspects to assist decisions in terms of product. This situation represents a challenging issue [13]. A strategic decision creation is based on the human resources of DW [14] to find hidden patterns in the education sector [15]. The consumer's complaint is handled through the DW [16], in addition to supporting critical clinical decisions [17-23]. The DW data are long term data; this notion means that the data are snapshots of data to perform a business analysis of the organizations [24-27]. DW is utilized to manage the back-end information storage of DSSs. DW is employed to conduct a rapid information retrieval from a massive amount of data that include data extraction from heterogeneous data sources. DW helps the managers in conducting their complex analytical queries for relevant information without exploring the operational databases [2, 28].

The knowledge driven DSS is categorized by two points, the model accessible by the managers and the interaction to the interface; the DSS is designed for the same decision-making and supporting purposes [29, 30]. The model simply represented outcomes and understandably presented them to the decision-maker [31]. The DSS dashboard is a crucial key to the system; the manager granted access into the knowledge-base, preferentially the cube that stores the complicated queries also unstructured information for the decision making [32, 33]. The proposed framework explains, designs, and implements the sales DSS implementation framework on the basis of data mart with OLAP and KPI as primary tools. Several cases are highlighted and discussed through the implementation process, such as the advantage of data mart in the short-term decision-making process. This study determines the optimal OLAP category for the proposed DSS and the benefits of using KPI in the short- and long-term implementation process, besides a potentiality to implement sales DW on the basis of this data mart. This study is organized into three sections. First, a literature review offers theoretical support for this research. Second, the methodology of the module framework is applied. Third, the results and conclusions of this research are discussed.

2. Literature Review

This study augmented the DSS capabilities and the adaptation of its features to serve the management necessities for organizations to have a convenient and accurate decision making. The following research trends are highlighted in various perspectives. The authors proposed a framework on the basis of DW to enhance the decision of the pharmaceutical distribution sector by predicting the medicine's sales. Elazeem et al. [34] used the neural network (NN) and autoregressive moving average model time series employed with the historical data to predict the sales. However, the design approach of the system was not presented. Many other machine learning approaches (such as naïve Bayes, decision tree (DT), and clustering) proved the model accuracy in the prediction sector not implemented. The data warehousing step is also unclear.

The DW schema is not defined, and the step of data cleaning is applied after the data warehousing. This process should be implemented in the staging area to prepare the data for the final loading into DW dimensions. Next, the research trend implemented a retail DW by using the Pentaho BI used for the CMK company that provides a piece of information in a short and efficient time to support decisions on the basis of OLAP, Girsang et al. [35]. The proposed architecture of the DW is Kimball and Rose. The star schema is selected for the DW architecture storage schema. However, the design approach of the DW is not presented because many other tools can be used to support online decisions after a refresh process, such as KPIs. Peng [36] proposed a DSS employed for administrative prediction purposes on the basis of DW and data mining. Researchers used multidimensional expressions (MDXs) and OLAP operations on the DW to present an analytical result to support decisions. The DW is implemented on the basis of the J2EE framework. However, the DW implementation framework is unclear. The size and origin of the source data are unclarified. The data mining approach and implementation framework are not clearly listed.

Katkar et al. [37] presented a novel method based on DW, Naïve Bayes, and fuzzy logic to forecast the sales. Many data sets were collected from different shops in various cities for 5 years to prove the mechanism's efficiency. However, the data warehousing process and the design methodology were not clearly explained, and the sales decisions were made on the basis of a classification approach using Naïve Bayes. OLAP was not used to analyse the store data. Other tools, such as KPI, were also not implemented because the future studies are not clearly discussed.

Finally, AbdAlrazig [38] investigated the advantage of using DW and data mining as a platform to enhance sales profit. The researchers found that the framework is helpful in terms of making decisions and predicting for future trends. However, the design methodology was not explained whether it is a top-down or a bottom-up approach. OLAP is not clearly presented because the proposed DW tends to be data mart instead of DW. The size of data sources is small, and the implementation cycle is also short of building the DW.

During the DSS assessment and evaluation, different assessment methods are used for evaluating the knowledge base in the DSS, such as static and dynamic methods [39]. DSSs, such as [40, 41], have been evaluated by experts to find the errors in the knowledge base. The static method is performed by experts to check the knowledge base and evaluate the DSS performance. This manuscript utilizes a model-driven DSS that employed OLAP and KPI to allow managers to investigate online reviews for patterns and present them around diverse levels and perspectives of analysis. The proposed framework of the decision support system (DSS) shown in this study is based on a previous literature; however, the preceding research gaps did not highlight OLAP and KPI into a DSS-based DW. However, the ETL details scheme to interact with unstructured then serves managers to build their systems by offering a unique dimensional model method. Thus, the model may be applied to provide data from multiple sources for later analysis on a dashboard. A group of dashboards for data analysis also offered so that managers may investigate and examined the functionalities model in search of possibilities and weaknesses in the market.

3. Methodology

The sales DSS framework methodology is implemented through the following five stages: data pre-processing; extract, transform, and load (ETL); building sales cube and KPIs; building reports; and deployment of the sales DSS dashboard. The DSS also holds a knowledge-driven sales supporter based on three data mining approaches (DT, clustering, and NN). These stages represent the general framework of the bottom-up approach. This approach suggests building low-level parts to integrate them later to create a large part. The proposed data mart is an independent data mart, and it can be combined to construct the enterprise sales DW. The bottom-up approach identifies all the star schema architectures that are proper to the data mart. This process starts with examining the entity-relationship (ER) model. An exhaustive analysis is performed to find the appropriate entities that can be facts than a large number of other entities around this fact entity [21, 42]. The bottom-up approach should make the characteristics of the model equivalent to the model implemented for the values of decision-makers. All the model options should be addressed in advance following the bottom-up approach. The model that followed the bottom-up approach in its design should minimize the changes in the information representation [43].

The proposed DSS architecture in Fig. 1 presents the system tiers starting from storage to presentation tier. The storage tier holds the source data (sales excel files) that have been collected from a trade market. This tier also holds the SQL server storage area where the data pre-processing is performed on the source data. Business intelligence (BI) tier holds DT, clustering, and NN, and sales cube after implementing ETL processes and performing data mining algorithms. Knowledge tier holds the knowledge-driven sales supporter conducted from implementing data mining algorithms besides the KPIs and reports. The final tier, presentation tier, holds a web interface to present all the reports, KPIs, and data mining algorithms results. The presentation area holds the MS excel pivot reports on the basis of the sales cube.

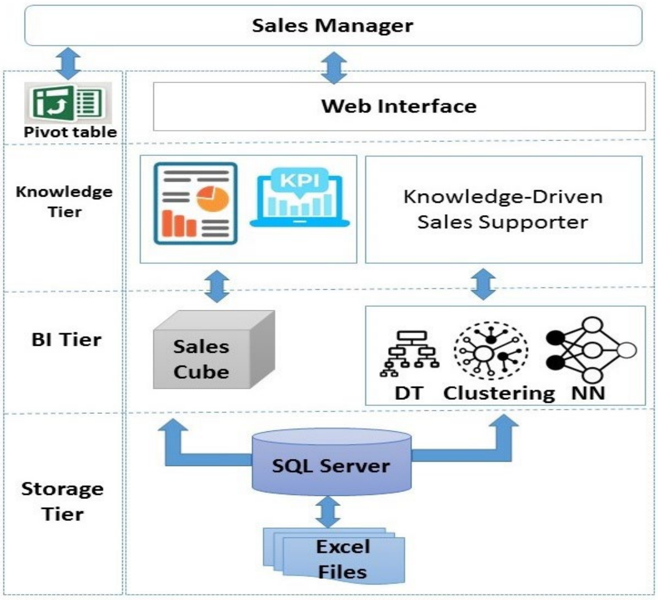

Fig. 1. Sales DSS architecture.

3.1. Data preprocessing

The data source of the sales data mart is a two-year data (2017 and 2018) taken from an operational system from a trade market in Basrah province, Iraq. The dataset holds more than nine attributes, such as invoice type, customer name, phone number, sell man, bill amount, quantity, item unit, market price, and date. The data have been imported as an excel file to manage and export this file to the SQL server. Before the data are exported into the staging area, many pre-processing processes are performed on the excel file, such as data cleansing, domain name misspelling, and some transformation techniques (e.g., deriving columns and adding new columns). The new derived columns are the date dimension columns (day, day of the week, month, quarter, and year) and address dimension columns (city, district, province, and country). After the excel file is exported into SQL server DBMS, the data are stored in the staging table to manage the transformation processes (part of ETL) to prepare the data for the final loading into data mart schema tables.

The schema of the sales data mart in Fig. 2 is a star schema where it consists of five dimensions (date, customer, item, Address, and invoice) dimension and sales fact tables. This schema is built-up by using the SQL server management server 2014. This DBMS is used to store and manage data mart data, cube, and reporting services. The star schema is chosen as a logical schema of the sales data mart because it is easy to implement and understand, exhibits fast query response, and new data in refresh can be easily added [44]. The dimension tables assist in making the complex queries later for the analysis and determine the projection of dimension columns with the fact table measurements. The date dimension consists of (day, day of the week, month, quarter, and year). For example, these columns will help in finding the projection of measurement (sales in US dollar) and the column (quarter) to produce the sales amount according to a quarter of the year for the other columns. The sales fact table consists of five surrogate keys and three additive measurements (item count, sales in dinar, and sales in USD) that can be added around all dimension tables.

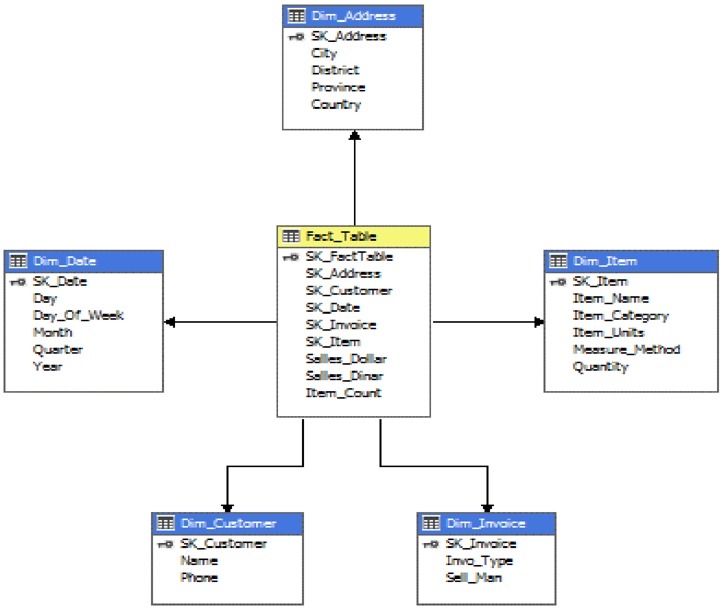

Fig. 2. Logical schema of sales data mart.

3.2. Extract, transform, and load (ETL)

ETL is the process of transforming data from the level of operational data to DW [45]. Extracting, transforming, and loading (ETL) are three main processes that include extracting data from a set of sources, followed by cleaning, allocating, rearranging, merging, and transferring to the DW. In the same context, the ETL processes can be among the challenging and biggest tasks in creating and configuring the DW, and they are regarded as complicated and time-consuming tasks in addition to the efforts to complete and implement a project, costs, and resources. Understanding the three main areas is one of the requirements for building a DW; these areas include the selected source region, destination area, and mapping area (ETL processes) [46, 47]. The first step of ETL processes is extraction. This step involves taking data from specific sources. These sources contain data with distinct characteristics that must be managed to extract data for the ETL process. This process is responsible for effectively integrating systems that are from different platforms, for instance, various operating systems OS, database management systems DMS, and communication protocols. Data transformation is the second scenario in any process of ETL. This step obtains correct, accurate, complete, consistent, and unambiguous data by performing some cleaning and compatibility with the source data. Moreover, this process is responsible for determining the accuracy in the fact table, dimension tables, DW layout, SCD and derived facts. The final step of ETL is loading the data to the repository. This process involves loading data to the fact and dimension tables [46, 48, 49].

The ETL strategy is shown in Fig. 3. The first step of the strategy holds the transformation processes, namely, data conversion, adding new columns, and resolving conflicts. The data conversion process converts the data from a string into data types in the dimension tables. The surrogate key is added in this stage. Accordingly, the conflicts and transformation processes are implemented. The transformation processes transform the domain data into another form and fill the null value with the values. The time dimension is present in the DW, which indicates the timeframe of the measures. The changes in other dimensions cannot be tracked by using the time dimension. In most cases, the changes in the dimensional data and time of occurrence are significant requirements for the purpose of analysis. These dimensions in which values of an attribute may change are called slowly changing dimensions (SCD) [50].

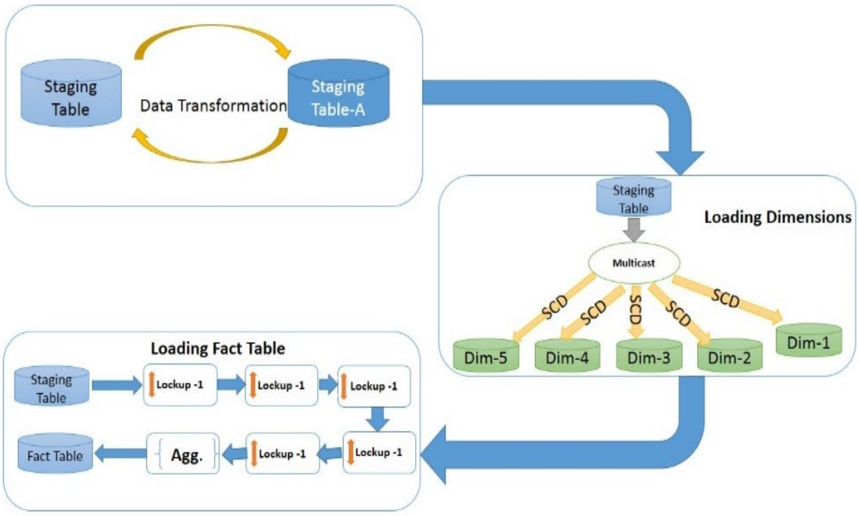

Fig. 3. ETL strategy.

The second step is loading dimension tables. The SCD tool is used to load the dimension tables. The three major types of SCDs are fixed, changing, and historical. With the fixed SCD type (of type 0), the attribute values never change. This type is proper for unchanging values, such as the customer birthdate. The changing SCD (type 1) overwrite the values of attributes. In this type, the old data is replaced with the new revision of data, thereby indicating that no history of the changing data is present. The third type is historical SCD (type 2), where the changing data are kept in a new row. The new data rows should be stored using surrogate keys to overcome the primary key problem [51].

The SCD loads of the dimension tables are arranged in a parallel form. The type of the SCD is fixed that load the data into tables, and the data in dimensions will not change. In the feature sales DW, the SCD will be turned into type 2 for some dimensions because historical data are required to make the hierarchies and store the changing data in the new rows. The measure type table is classified into three types depending on whether they can be summed across any of the dimensions linked with the fact table. This measure is also called “fully additive measure”. If the types can be summed across some dimensions, not all, then the measure is called “semi additive measure". The third type called "non-additive measure" is completely non-additive [52]. The third step is to load the fact table with the keys of the dimension tables and measurements. First, the lockup processes are performed to find the surrogate keys in the dimension tables (primary keys). Second, the key measurements are determined from the staging table to load them in the fact table. All these steps are implemented using SQL Server Integration Services (SSIS) 2014.

3.3. Building sales cube and KPIs

In this stage, the sales cube is created by using SQL Server Analysis Service 2014. The sales cube is implemented on the basis of the fact and dimension tables previously created. The sales cube (OLAP server) allows treating multidimensional expressions (MDX) on the dimensions and the measurements of the fact table. The MDX is an OLAP query language that enables the analyst to perform all the OLAP operations, such as roll-up, drill down, slice, and dice [21]. The analyst can navigate all the hierarchies and list the results by dropping down the dimensional columns and pre-calculated fact table measurements. After this step, the cube is available to all permitted analytical tools to navigate and implement charts. Microsoft Excel pivot table is one of the essential tools. This tool allows us to import the sales cube and perform all the OLAP operations. Moreover, this tool allows building customized reports that easily and flexibly present the result. The second part of this step is implementing sales KPIs. Many indicators are implemented to measure the current performance on the basis of the given data.

OLAP is a technology used by analysts to present multiple perspectives by extracting and analysing business data (BD). Analysts simultaneously examine data from various sets of database systems [53]. Three models of OLAP architecture are widely used depending on the literature: Relational (ROLAP), multidimensional (MOLAP) architectures, and Hybrid (HOLAP) [54, 55]. ROLAP uses the data in relational databases, where it stores the fact table and dimension tables as relational tables [56]. In MOLAO, the data are stored in a multidimensional cube to display different or multidimensional data. The method of storing data in architectures ROLAP and MOLAP makes MOLAP better than ROLAP in terms of time in accessing data in addition to response time to the query. This notion means that ROLAP requires a large amount of effort and resources despite its high efficiency and scalability for managing extensive data. Although MOLAP has a good response time to inquiries and access to data, it does not handle extensive data compared with ROLAP. Finally, HOLAP is an architecture that combines the useful features of ROLAP and MOLAP in an architecture. The HOLAP model provides extensive detailed data compared with the data in relational tables in ROLAP, in addition to the aggregations being stored in pre-calculated cubes; thus, it is more efficient than MOLAP. By contrast, HOLAP is complicated because it combines relational and multidimensional properties [55, 57].

OLAP is the process of simultaneously analysing the data by users (analysts) of a different set of database systems. This technology is used by analysts to extract and show BD from various viewpoints for decision-making. The two main OLAP architectures are as follows: relational (ROLAP) and multidimensional (MOLAP) architectures. The other categories of OLAP architectures include hybrid (HOLAP) and desktop (DOLAP) architectures [53, 54]. The OLAP server in our model is multidimensional OLAP (MOLAP). This category allows us to create and perform all OLAP facilities, such as a cube, dimensions, measurements, OLAP operations, hierarchies, and KPIs. MOLAP enables us to analyse the pre-summarized and pre-calculated data stored in the multidimensional cube form stored in the MOLAP server.

KPI is a measurable value that plays a vital role in monitoring and evaluating the processes carried out by the company to reach the required aims. The indicators or KPIs are used to provide accurate and rapid information that is used to compare the current performance with the goals needed to meet business goals. The KPIs are used at different levels by large enterprises to assess their performance in achieving the required goals. The overall corporate performance is measured by using high-level performance indicators, while low-level performance indicators are used to measure partial details [58, 59].

The main role of KPIs in sales DSS is to provide an alert when a specific group of items reaches the minimum profit or exceeds the previous amount at a particular time. The first KPI (quarter sales) provides an indicator when the profit amount exceeds or decreased the benefit of the last quarter. The value expression of this KPI is the sales in dinar (measurement). The goal of this KPI is the sales in dinar multiplied by 0.1. This notion means that the intended profit should increase by 10%. The KPI status is implemented using the MDX language to provide an indicator when the KPI value reaches the goal or not. When the KPI value exceeds the KPI goal, the status shows an up sign indicating that the state goes up. When the KPI value decreased the KPI goal, the status exposes a down sign indicating that the status goes down. When the KPI value divided by the KPI goal is between 1 and 0.85, the status is stable; otherwise, when the KPI value is less than the KPI goal, the status shows a down sign indicating that the status goes down. The KPI trend represents the changing status over time. Thus, the trend value is one when the sales in dinar for the current year is higher than that of the previous year; otherwise, the trend is −1. The value 1 represents the increasing trend of sales, −1 represents the decreasing trend of sales over time. The other KPIs are described and explained in Table 1.

Table 1. The proposed KPIs.

Name and Expr.	Value	Status	Trend
Name: Sales_Dinar **Expr.: [Measures].[Salles Dinar]**	[Measures].[Salles Dinar]*0.1	Case When KpiValue("Sales_Dinar") /KpiGoal("Sales_Dinar") > 1 Then 1 When KpiValue("Sales_Dinar") /KpiGoal("Sales_Dinar") <= 1 And KpiValue("Sales_Dinar") /KpiGoal("Sales_Dinar") >= .85 Then 0 Else -1 End	Case When (KPIValue("Sales_Dinar"),[Dim Date].[Year].CurrentMember) > (KPIValue("Sales_Dinar"),[Dim Date].[Year].CurrentMember.PrevMember) Then 1 When (KPIValue("Sales_Dinar"),[Dim Date].[Year].CurrentMember) < (KPIValue("Sales_Dinar"),[Dim Date].[Year].CurrentMember.PrevMember) Then -1 End

Name: Sales_Quarter Expr.: [Measures].[Salles Dollar]	[Measures].[Salles Dollar]*0.1	Case When KpiValue("Sales_Quarter") / KpiGoal("Sales_Quarter") > 1 Then 1 When KpiValue("Sales_Quarter") / KpiGoal("Sales_Quarter") <= 1 And KpiValue("Sales_Quarter") / KpiGoal("Sales_Quarter") >= .85 Then 0 Else -1 End	Case When (KPIValue("Sales_Quarter"),[Dim Date].[Quarter].CurrentMember) > (KPIValue("Sales_Quarter"),[Dim Date].[Quarter].CurrentMember.PrevMember) Then 1 When (KPIValue("Sales_Quarter"),[Dim Date].[Quarter].CurrentMember) < (KPIValue("Sales_Quarter"),[Dim Date].[Quarter].CurrentMember.PrevMember) Then -1 End
Name: Groups_Of_Item Expr.: [Measures].[Salles Dinar]	Case When [Dim Item].[Item Category] Is [Dim Item].[Item Category].&[Food] Then .20 When [Dim Item].[Item Category] Is [Dim Item].[Item Category].&[Toys] Then .10 When [Dim Item].[Item Category] Is [Dim Item].[Item Category].&[Drink] Then .30 Else .40 End	Case When KpiValue("Groups_Of_Item") / KpiGoal("Groups_Of_Item") > .90 Then 1 When KpiValue("Groups_Of_Item") / KpiGoal("Groups_Of_Item") <= .90 And KpiValue("Groups_Of_Item") / KpiGoal("Groups_Of_Item") > .80 Then 0 Else -1 End	Case When (KPIValue("Groups_Of_Item"),[Dim Date].[Month].CurrentMember) > (KPIValue("Groups_Of_Item"),[Dim Date].[Month].CurrentMember.PrevMember) Then 1 When (KPIValue("Groups_Of_Item"),[Dim Date].[Month].CurrentMember) < (KPIValue("Groups_Of_Item"),[Dim Date].[Month].CurrentMember.PrevMember) Then -1 End

3.4. Building reports

The reports are created on the basis of the cube analytical databases and KPIs. Two types of reports are implemented for the sales DSS dashboard, which are online and offline reports. The online reports consist of three types: reports that accept a parameter value to present the analytical results, pre-packaged reports, and KPI reports. The reports were created by using SQL Server Reporting Service 2014.

Meanwhile, the offline reports were created using MS excel pivot table 2010. The first type in the online reports is parametric reports type. The analyst can select a parameter value to present the chart results as a chart (Fig. 4).

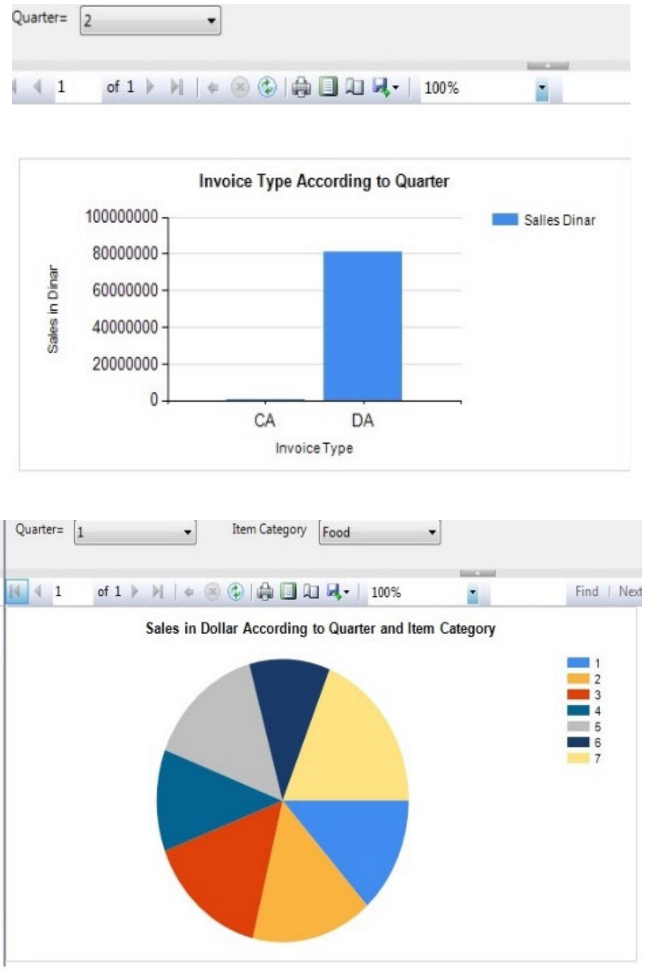

Fig. 4. Parametric reports.

The report on the left side accepts a quarter value as a parameter to present the sales in dinar according to invoice type. This type of selection can represent the dice operation of OLAP to provide the analytical result according to a specific quarter. The second report accepts two parameter values, quarter and item type, to present the sales in USD according to a day of the week. This type of report can represent the slice OLAP operation. The analysts can also use the reports to perform other OLAP operations, such as drill through and drill down, by using the hierarchy concept in any parameter values.

Figure 5 shows four pre-packaged reports that are implemented to instantly show the intended statistics. The first report lists the count of food sales according to the weekday. The seventh day of the week has the maximum food sales, while the sixth day holds the minimum food sales. The second report lists the sales in

USD according to month. The first month holds the maximum profit, while the fifth month has the minimum benefit.

Three KPIs are implemented to show the indicators related to three cases (Fig. 6). The first KPI is sales in the Iraqi dinar. This indicator calculates the profit and provides an indicator when it approaches the goal. The second KPI is the quarterly sales which measures and gives an indication related to sales in each quarter of year. While the third KPI is item group KPI which measures the sales for each group of items and give indication related to sales profit. These KPIs are explained before, and the reports of KPIs list the KPI details. The second type of reports is implemented by using an MS Excel Pivot table, where the charts can present the OLAP operation. This type of OLAP is desktop OLAP, where the analyst can instantly list the reports. Figure 7 records the sales of the four-item categories (cigarette, food, home application, and toys) according to the day of the week. The figure also shows that food sales take first place among all categories and an increase in the seventh day of the week. The toys group has a minimum number of sales among all types.

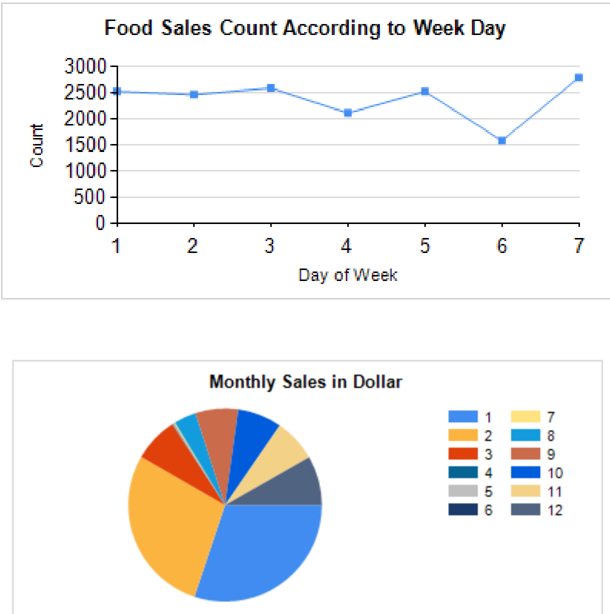

Fig. 5. Pre-packaged reports.

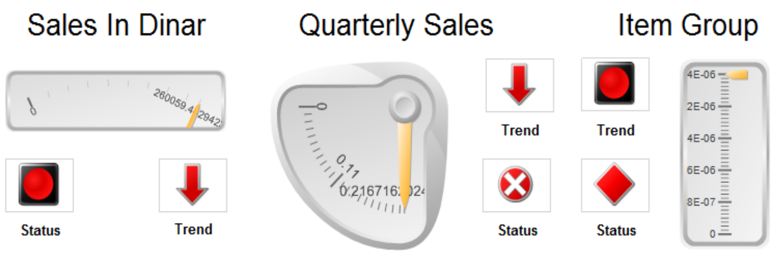

Fig. 6. Sales KPIs.

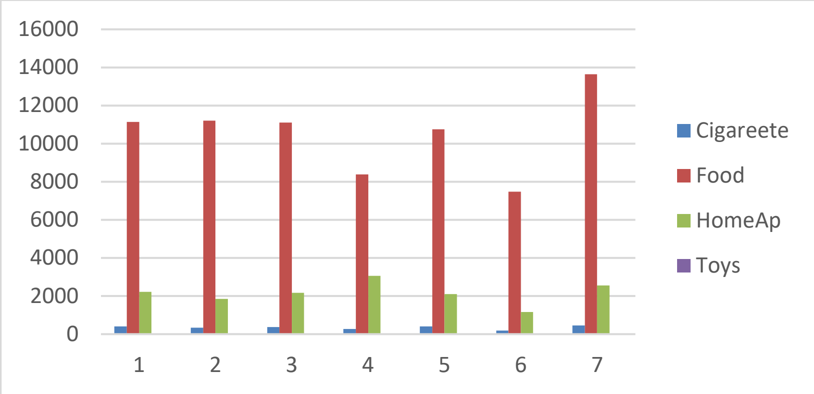

Fig. 7. Item categories of sales according to weekday.

3.5. Data mining knowledge

The knowledge-driven sales supporter was constructed on the basis of the three data mining approaches (DT, clustering, and NN). The algorithms are implemented and tested using SSAS package. SSAS provides a part to perform different data mining approaches. All data mining approaches help the decision-maker in supporting decision on the basis of the obtained results. The sales manager can use this basis to predict the sales rate, sales items, and quantities by using the analytical graphs and results obtained from the DT, clustering, and NN approaches. The results of the graphs and figures from data mining approaches are combined with the reports and KPIs in one dashboard.

3.6. Web interface

The web interface represents the DSS dashboard. The dashboard allows the analysts and managers to reach knowledge-driven sales supports and all the pre-packaged reports and KPIs either online or offline. The prebuilt reports and KPIs present the constant update of all reports after loading the data into the sales data mart. The sales DSS dashboard can be accessed online through a local area network, where the server is attached to the network, and the managers can access the system. The dashboard can help in monitoring the analytical information by different charts. The access to this dashboard is restricted to specific persons (analysts and managers). Only the permitted end users can access the dashboard and navigate the reports. The stakeholders can also perform OLAP operations, such as dice, slice, drill through, and drill down, through the reports. The second type of report enables a parameter value to present the analytical results according to the parameters' values.

4. DSS Evaluation

The sales data consist of 65,535 rows obtained from a trade market for two years (2017 and 2018). The proposed sales DSS provides a web interface to access all the reports of OLAP and KPIs besides the results of data mining algorithms. The sales DSS evaluation process passed throughout two evaluation processes (knowledge-driven evaluation process and reports of OLAP and KPIs). Three main data mining approaches are implemented using data mining part of SSAS package where the performance criteria are listed in the Table 2. DT, clustering based expectation maximisation (EM), and artificial NN are performed and examined. For the DT, the algorithm used to split nodes is a complete where the nodes are split into all branches resulting from split score. The split score algorithm is Bayesian Dirichlet Equivalent (BDE) with uniform prior. EM is the algorithm used for clustering approach where the variables or features are unobserved. EM tries using statistical model to find the maximum a posteriori (MAP) or maximum likelihood of the parameters. In artificial NN, the number of hidden nodes is 4, the sample size is 10000 instance of sample size to train the model. The percentage of test data set is 30% and 70% of train data set to compare the performance criteria of data mining algorithms.

Table 2. Performance criteria of data mining algorithms.

Algorithm	Precision	Recall	Sensitivity	Specificity
DT	69.64%	70.79%	70.79%	31.97%
Clustering	69.29%	73.23%	73.23%	34.01%
NN	65.85%	73.99%	73.99%	38.95%

With regard to the knowledge-driven sales supporter, the DT shows the following overall results: sensitivity of 70.79%, specificity of 31.97%, precision of 69.64%, and recall of 70.79%. Clustering shows results with sensitivity of 73.23%, specificity of 34.01%, precision of 69.29%, and recall of 73.23%. NN shows results with sensitivity of 73.99%, specificity of 38.95%, precision of 65.85%, and recall of 73.99%. The proposed system is implemented on a system with 8 GB of RAM, 2.4 GHz Core-i3 of processor and 140GB free space of hard. The platform is protected by authentication credentials where only the permitted users are allowed to access the system. The online access of the DSS dashboard is also restricted by authenticated users, The proposed DSS has been evaluated on the basis of the static method where the knowledge base of the DSS is manually checked to find errors. The assessment shows that no error occurred after the KPIs and OLAP reports are reviewed. The evaluation also shows a flexibility in viewing and constructing reports and charts in the MS excel pivot table on the basis of the sales cube. Some steps are skipped, such as determining data mining algorithm for each approach, post-processing, and graphs.

5. Conclusions

This study presents a knowledge-driven sales DSS implementation framework on the basis of data mart and by using OLAP and KPI as essential tools for analysis. The model implementation follows the bottom-up approach to create the DSS because the data mart is independent and to permit making sales DW by integrating the independent data mart. The sales data mart is the base of DSS that offers a short-term strategic decision making. The knowledge-driven sales supporter helps the analysts in predicting sales profits, quantities, and rates on the basis of the three data mining approaches (DT, NN, and clustering).

The static evaluation method of the knowledge base shows no error. Accordingly, the dynamic methods will be performed by using a questionnaire to collect all the information and feedback about DSS usability. The performance criteria of the implemented data mining algorithms in the DSS show a sensitivity of 70.79%, specificity of 31.97%, precision of 69.64%, and recall of 70.79% for

DT. Clustering has a sensitivity of 73.23%, specificity of 34.01%, precision of 69.29%, and recall of 73.23%. NN shows a sensitivity of 73.99%, specificity of 38.95%, precision of 65.85%, and recall of 73.99%. Two types of reports are implemented for the final dashboard, namely, offline and online reports. The online reports consist of three types of reports: KPI reports, pre-packaged reports, and reports that accept the parameter value to present the analytical results. Online reports are constructed using SSRS 2014, while the offline reports are made utilizing the MS excel pivot table. The offline reports are essential to managers to consistently access the sales cube. The online reports are significant for analysts to remotely access the reports and perform complex MDX through the network.

The MOLAP is used to execute all the OLAP operations, such as slice, dice, roll up, and drill down, on the sales data mart to discover the hidden patterns in the sales cube. The operations are implemented by either SSAS 2014, or MS excel pivot table 2010. The analyst can also view and perform complex MDX queries by using the sales DSS dashboard, where drop-down lists show the dimension selection. The KPIs are used to provide an indicator when a critical situation occurs or during performance enhancement. Three KPIs are implemented to measure the profit situation according to time. Many KPIs can be created according to the managers' needs. Offline and online MOLAP and KPIs are crucial tools in supporting decisions and exploring the hidden patterns in the sales data.

The OLAP operations can help the analysts and managers in viewing sales information from different aspects. The predefined KPIs can automatically compare and give an indication when a critical situation occurs. The web interface easily permits sales managers to navigate all reports and graphs resulting from the data mining approaches. The system will be configured using scheduled task to perform online ETL, data mining and KPI used the new fresh data. A new strategy will be configured to handle the new data in order to handle extraction, transformation and loading processes to present the new results properly. The system will also test against new data of a trade market to examine and re-evaluate the system.

References

1. Narayan, R.; and Meht, G. (2020). Design of customer information management system. *Data Communication and Networks*, 1049, 195-216.
2. Tria, F.D.; Lefons, E.; and Tangorra, F. (2017). Evaluation of data warehouse design methodologies in the context of big data. *Proceedings of the International Conference on Big Data Analytics and Knowledge Discovery*. Lyon, France.
3. Kasie, F.M.; Bright, G.; and Walker, A. (2017). Decision support systems in manufacturing: a survey and future trends. *Journal of Modelling in Management*, 12(3), 432-454.
4. Wagner, W.P. (2017). Trends in expert system development: A longitudinal content analysis of over thirty years of expert system case studies. *Expert Systems with Applications,* 76, 85-96.
5. Najm, I.A.; Hamoud, A.K.; Lloret, J.; and Bosch, I. (2019). Machine learning prediction approach to enhance congestion control in 5G IoT environment. *Electronics*, 8(6).

6. Hamoud, A.K.; Hashim, A.S.; and Awadh, W.A. (2018). Predicting student performance in higher education institutions using decision tree analysis. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(2), 26-31.
7. Hamoud, A.K.; Humadi, A.M.; Awadh, W.A.; and Hashim, A.S. (2017). Students' success prediction based on Bayes algorithms. *International Journal of Computer Applications*, 178(7), 6-12.
8. Hamoud, A.K. (2016). Selection of best decision tree algorithm for prediction and classification of students' action. *American International Journal of Research in Science, Technology, Engineering & Mathematics*, 16, 26-32.
9. Hamoud, A.K. (2018). Classifying students' answers using clustering algorithms based on principle component analysis. *Journal of Theoretical and Applied Information Technology*, 96(7), 1813-1825.
10. Hamoud, A.K.; and Humadi, A.M. (2019). Student's success prediction model based on artificial neural networks (ANN) and a combination of feature selection methods. *Journal of Southwest Jiaotong University*, 54(3).
11. Hashima, A.S.; Hamoud, A.K.; and Awadh, W.A. (2018). Analyzing students' answers using association rule mining based on feature selection. *Journal of Southwest Jiaotong University*, 53(5).
12. Najm, I.A.; Ismail, M.; and Abed, G.A. (2014). High-performance mobile technology LTE-A using the stream control transmission protocol: a systematic review and hands-on analysis. *Journal of Applied Sciences*, 14(19), 2194-2218.
13. Zhang, R.; Indulska, M.; and Sadiq, S. (2019). Discovering data quality problems. *Business and Information Systems Engineering*, 61(5), 575-593.
14. Burgard, M.; and Piazza, F. (2009). *Data warehouse and business intelligence systems in the context of e-HRM*. USA: IGI Global.
15. Triola, M.M.; and Pusic, M.V. (2012). The education data warehouse: a transformative tool for health education research. *Journal of Graduate Medical Education*, 4(1), 113-115.
16. Buttle, F. (2019). *Customer relationship management: concepts and technologies*. England: Routledge.
17. Hamoud, A.K.; Adday, H.; Obaid, T.A.S.; and Hameed, R.A. (2016). Design and implementing cancer data warehouse to support clinical decisions. *International Journal of Scientific and Engineering Research*, 7(2), 1271-1285.
18. Hamoud, A.K.; and Obaid, T. (2013). Building data warehouse for diseases registry: first step for clinical data warehouse. *International Journal of Scientific and Engineering Research*, 4(7), 636-640.
19. Hamoud, A.K.; and Obaid, T.A.S. (2014). Using OLAP with diseases registry warehouse for clinical decision support. *International Journal of Computer Science and Mobile Computing*, 3(4), 39-49.
20. Hamoud, A.K.; Hashim, A.S.; and Awadh, W.A. (2018). Clinical data warehouse: a review. *Iraqi Journal for Computers and Informatics*, 44(2), 1-11.
21. Hamoud, A.K. (2013). *Design and implementation data warehouse to support clinical decisions using OLAP and KPI*. Master thesis, University of Basrah.
22. Humadi, A.M.; and Hamoud, A.K. (2019). Liver hepatitis diagnosing based on fuzzy inference system. *Journal of Southwest Jiaotong University*, 54(4), 1-10.

23. Hamoud, A.K. (2017). Applying association rules and decision tree algorithms with tumor diagnosis data. *International Research Journal of Engineering and Technology*, 3(8), 27-31.
24. Muriana, C.; Piazza, T.; and Vizzini, G. (2016). An expert system for financial performance assessment of health care structures based on fuzzy sets and KPIs. *Knowledge-Based Systems*, 97, 1-10.
25. Rahman, N. (2017). An empirical study of data warehouse implementation effectiveness. *International Journal of Management Science and Engineering Management*, 12(1), 55-63.
26. Hamoud, A.K.; Hussien, H.N.; Fadhil, A.A.; and Ekal, Z.R. (2020). Improving service quality using consumers' complaints data mart which effect on financial customer satisfaction. *Journal of Physics Conference Series*, 1530(1).
27. Hamoud , A.K.; Ulkareem, M.A.; Hussain, H.N.; Mohammed, Z.A.; and Salih, G.M. (2020). Improve HR decision-making based on data mart and OLAP. *Journal of Physics Conference Series*, 1530.
28. Park, T.; and Kim, H. (2013). A data warehouse-based decision support system for sewer infrastructure management. *Automation in Construction*, 30, 37-49.
29. Rajaeian, M.M.; Cater-Steel, A.; and Lane, M. (2017). A systematic literature review and critical assessment of model-driven decision support for IT outsourcing. *Decision Support Systems,* 102, 42-56.
30. Wang, W.; Cui, Y.; Luo, Y.; Li, Z.; and Tan, J. (2019). Web-based decision support system for canal irrigation management. *Computers and Electronics in Agriculture*, 161, 312-321.
31. Dweiri, F.; Kumar, S.; Khan, S.A.; and Jain, V. (2016). Designing an integrated AHP based decision support system for supplier selection in automotive industry. *Expert Systems with Applications*, 62, 273-283.
32. Cobo, Á.; Llorente, I.; Luna, L.; and Luna, M. (2019). A decision support system for fish farming using particle swarm optimization. *Computers and Electronics in Agriculture*, 161, 121-130.
33. Humadi, A.M.; and Hamoud, A.K. (2017). Online real time fuzzy inference system based human health monitoring and medical decision making. *International Journal of Computer Science and Information Security*, 15(4), 197-204.
34. Elazeem, N.M.A.; Labib, N.M.; and Abdella, A.K. (2019). A proposed data warehouse framework to enhance decisions of distribution system in pharmaceutical sector. *Egyptian Computer Science Journal*, 43(2), 43-60.
35. Girsang, A.S.; Arisandi, G.; Elysisa, C.; Michelle; and Saragih, M.H. (2019). Decision support system using data warehouse for retail system. *Journal of Physics: Conference Series*, 1367(1), 1-6.
36. Peng, X. (2019). Analysis of administrative management and decision-making based on data warehouse. *Proceedings of the 11th International Conference on Measuring Technology and Mechatronics Automation*. Qiqihar, China, 527-530.
37. Katkar, V.; Gangopadhyay, S.P.; Rathod, S.; and Shetty, A. (2015). Sales forecasting using data warehouse and Naïve Bayesian classifier. *Proceedings of the International Conference on Pervasive Computing*. Pune, India, 1-6.

38. AbdAlrazig, H. (2018). *Designing a data warehousing model to support forecasting and decision making for sales*. Doctoral dissertation. Sudan University of Science and Technology.
39. Torres-Sanchez, R.; Navarro-Hellin, H.; Guillamon-Frutos, A.; San-Segundo, R.; Ruiz-Abellón, M.C.; and Domingo-Miguel, R. (2020). A decision support system for irrigation management: analysis and implementation of different learning techniques. *Water*, 12(2).
40. Baumeister, J.; and Striffler, A. (2015). Knowledge-driven systems for episodic decision support. *Knowledge-Based Systems*, 88, 45-56.
41. Hosio, S.; Karppinen, J.; Berkel, N.; Oppenlaender, J.; and Goncalves, J. (2018). Mobile decision support and data provisioning for low back pain. *Computer*, 51(8), 34-43.
42. Chakraborty, S.; and Doshi, J. (2018). Incremental updates using data warehouse versus data marts. *Proceedings of the 4th International Conference for Convergence in Technology*. Mangalore, India, 1-5.
43. Teotonio, I.; Cabral, M.; Cruz, C.O.; and Silva, C.M. (2020). Decision support system for green roofs investments in residential buildings. *Journal of Cleaner Production*, 249.
44. Adamson, C. (2012). *Mastering data warehouse aggregates: solutions for star schema performance*. New Jersey: John Wiley & Sons.
45. Almeida, D.R.; Baptista, C.S.; Andrade, F.G.; and Soares, A. (2020). A survey on big data for trajectory analytics. *International Journal of Geo-Information*, 9(2).
46. El-Sappagh, S.H.A.; Hendawi, A.M.A.; and El Bastawissy, A.H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University-Computer and Information Sciences*, 23(2), 91-104.
47. Vyas, S.; and Vaishnav, P. (2017). A comparative study of various ETL process and their testing techniques in data warehouse. *Journal of Statistics and Management Systems,* 20(4), 753-763.
48. Stonebraker, M.; and Ilyas, I.F. (2018). Data integration: the current status and the way forward. *IEEE Data Engineering Bulletin*, 41(2), 3-9.
49. Shadroo, S.; and Rahmani, A.M. (2018). Systematic survey of big data and data mining in internet of things*.* *Computer Networks,* 139, 19-47.
50. Atay, C.E.; and Garani, G. (2019). Maintaining dimension's history in data warehouses effectively. *International Journal of Data Warehousing and Mining*, 15(3), 46-62.
51. Kimball, R.; and Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling*. New Jersey: John Wiley & Sons.
52. Chandwani, G.; and Uppal, V. (2015). Implementation of star schemas from ER Model. *International Journal of Database Theory and Application*, 8(3), 111-130.
53. Widiadnyana, P.; Maricar, M.A.; Arnawan, I.N.; and Ariyani, S. (2017). Analysis of data warehouse for transactional analysis methods online analytical processing (OLAP) at company XYZ silver. *International Journal of Engineering and Emerging Technology*, 2(1), 72-75.
54. Jianmin, W.; Wenbin, Z.; Tongrang, F.; Shilong, Y.; and Hongwei, L. (2020). An improved join-free snowflake schema for ETL and OLAP of data warehouse. *Concurrency and Computation: Practice and Experience*, 32(23).

55. Moalla, I.; Nabli, A.; Bouzguenda, L.; and Hammami, M. (2017). Data warehouse design approaches from social media: review and comparison. *Social Network Analysis and Mining*, 7(5).
56. Singh, Y.S.; Singh, Y.K.; Devi, N.S.; and Singh, Y.J. (2019). Easy designing steps of a local data warehouse for possible analytical data processing. *ADBU Journal of Engineering Technology*, 8(7), 1-8.
57. Nanda, A.; Gupta, S.; and Vijrania, M. (2019). A comprehensive survey of OLAP: recent trends. *Proceedings of the 3rd International Conference on Electronics, Communication and Aerospace Technology*. Coimbatore, India, 425-430.
58. Khouri, S.; Ghomari, A.R.; and Aouimer, Y. (2019). *Model and data engineering*. Chapter: Thinking the incorporation of LOD in semantic cubes as a strategic decision. Springer, 287-302.
59. Domínguez, E.; Pérez, B.; Rubio, Á.L.; and Zapata, M.A. (2019). A taxonomy for key performance indicators management. *Computer Standards & Interfaces*, 64, 24-40.