

Implementing data-driven decision support system based on independent educational data mart

Alaa Khalaf Hamoud, Marwah Kamil Hussein, Zahraa Alhilfi, Rabab Hassan Sabr

Computer Information Systems Department, University of Basrah, Iraq

Article Info

Article history:

Received Nov 28, 2020

Revised Apr 8, 2021

Accepted Apr 26, 2021

Keywords:

Decision support system

Educational data mart

ETL

Independent data mart

KPI

OLAP

ABSTRACT

Decision makers in the educational field always seek new technologies and tools, which provide solid, fast answers that can support decision-making process. They need a platform that utilize the students' academic data and turn them into knowledge to make the right strategic decisions. In this paper, a roadmap for implementing a data driven decision support system (DSS) is presented based on an educational data mart. The independent data mart is implemented on the students' degrees in 8 subjects in a private school (Al-Iskandaria Primary School in Basrah province, Iraq). The DSS implementation roadmap is started from pre-processing paper-based data source and ended with providing three categories of online analytical processing (OLAP) queries (multidimensional OLAP, desktop OLAP and web OLAP). Key performance indicator (KPI) is implemented as an essential part of educational DSS to measure school performance. The static evaluation method shows that the proposed DSS follows the privacy, security and performance aspects with no errors after inspecting the DSS knowledge base. The evaluation shows that the data driven DSS based on independent data mart with KPI, OLAP is one of the best platforms to support short-to-long term academic decisions.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Alaa Khalaf Hamoud

Computer Information Systems Department

University of Basrah

61003 Karmat Ali Camp, Basrah Iraq

Email: alaa.hamoud@uobasrah.edu.iq

1. INTRODUCTION

The importance of data repositories has emerged with the existence of large institutions. Departments manage their own databases (marketing, financial, and administrative), which organise massive common data. Finding data related to a specific subject and organising such data into a single database called 'the data store' are required; another requirement is maintaining the special rules for stores with no modification or changing the rules on the basis of topics by using special software, especially in each subject, via a process called schema integration. This process also specifies how to transfer the merged data [1]. Data warehouse (DW) can be defined as a subject-oriented, time-variant, integrated and non-volatile data used to support strategic decision making [2]. DW holds a collection of permanent historical data that assists in administrative decision making to help in accessing data for the purposes of time analysis, knowledge discovery and decision making [3]. It is specifically designed to extract, process and represent data in a suitable format for this purpose. The data extracted from different sources, rules, systems and places are defined as a kind of database that contains a huge amount of existing data to help in making decisions within an organization [4]. The interior intention of a user requires indicators, which are systems needed for

studying, analysing and presenting enterprise data in a manner that enables senior management to make decisions [5], [6]. The type of DW that an organization should adopt depends on the way that the organization works and the type of decision support system (DSS) it needs. One of the simplest types of data repositories is the operational data store (ODS), a production database that has been replicated after errors have been processed. ODS is primarily used to complete standard process reports and provide details of business transactions for summary analysis [7], [8]. Another type of DW is called data mart, which is a limited use or single-use system that can be used to analyse specific information in a specific area or for a specific production line. Data stores usually contain only summary data but can be linked to ODS for diving into transaction details if needed. Data stores are sometimes managed by information technology (IT) departments in companies but can also be managed by users in a particular department or group [9]. Many types and applications use DWs and data marts to support strategic decisions, such as clinical path [10]-[12], cancer DW [13], educational health DW [14], HR DW [15], compliant DW [16], human resource DW [16] and security DW [17].

Educational staff members always need new technologies that support their decisions; these members seek different tools and applications to support their decisions on the basis of different algorithms and techniques, such as machine learning and data mining algorithms [18]-[25], mobile learning [26], DSSs [27] and web-based tools [28]. The decisions made on the basis of previous studies and solid analytical results are more considerable and reasonable than before. Educational DW is the best solution that provides online analytical results on the basis of a multidimensional view of OLAP queries for all educational stakeholders (professors, lecturers, managers and decision makers). Educational DW provides a large view of the performance of all students and allows them to detect the obstacles in their progress [29]. Educational DW can also be used for the future applications of data mining to implement all techniques and algorithms that can be implemented on huge educational historical records instead of a small dataset. The use of educational DW can effectively reduce educational errors that affect academic decision making [30], [31].

DSS is an essential platform for making right strategic decisions. The DSS type can be determined on the basis of the purpose of the DSS, the type of knowledge and target users. DSS can be categorised into five types: knowledge-, model-, data-, document- and communication-driven DSS. Model-driven DSS allows decision makers to choose among options on the basis of limited data and options. Knowledge-driven DSS covers many kinds of systems within organisations; it provides an advice management process or service and product selection. Data-driven DSS concentrates on a decision and manipulates the data to fit the decision where the data can be in different formats. Target users are managers and stakeholders. Document-driven DSS targets a specific group of users on the basis of a document search on the web; a mechanism is required for retrieving related documents to support decisions. Communication-driven DSS provides a shared dashboard for making decisions for more than one person in a single shared task [32].

Data mart is used in this model as the base of the data-driven educational DSS through the implementation of specific school data and analysis of executive information systems. This paper presents the stages of implementing an educational independent data mart, starting from selecting the proper approach of design to implementing OLAP queries and KPI. The implementation starts with converting the paper-based educational data of student records into electronic educational records. The data mart staging area is prepared to extract electronic students' records and implement all extraction, transformation and loading (ETL) processes. Star schema is chosen as an architecture of data mart for its ease of implementation, ease of load and fast OLAP query response. The educational cube is built to implement OLAP queries. Three types of result viewing are listed structured query language (SQL) server integration services (SSIS) cube view, online reports and offline excel pivot reports). SQL server management studio (SSMS) 2014 is used to store and implement educational data mart objects (data mart schema and staging area). ETL is implemented and designed by SSIS 2014, where SQL server analysis services (SSAS) 2014 is used to design and construct the educational cube. Likewise, SSRS 2014 is used for designing reports. All the implemented projects are deployed by SQL server data tools (SSDT), as this tool provides the ability to create and deploy (SSIS, SSAS and SQL server reporting services (SSRS)) projects.

The rest of this paper is organised as being as: Section 2 analyses and discusses the related works and presents the strong points in these works. Section 3 explains the model implementation roadmap and details each step of the model implementation. Section 4 concludes the points derived from the model result and presents the future works that can be implemented on educational data mart.

2. RELATED WORKS

Smyrnaki [33] proposed a model based on DW to implement a DSS in order to support decisions. The proposed system integrated a heterogeneous sources used for different purposes such as education, financial and managerial decisions. The results conducted the integration of multiple heterogeneous sources

of data will be a useful platform for both of educational leadership and quality assurance unit. Next, Suman *et al.* [34] proposed a DW to solve many challenges in the higher education center. These challenges are facing the designer throughout the designing process such as the overall system design, the main processes of extraction, transformation, and loading, and performing the analysis processes using multidimensional queries based OLAP. The tool used for developing DW is Mondrian and Pentaho business intelligence tool. However, the proposed systems did not implement KPI as an essential component in the decision making process.

Mihai *et al.* [35] proposed using enterprise data warehouse (EDW) to support academic decisions in educational institutions. The design and implementation process proceeded via two steps; the first one is implementing EDW to find the performance measurements of the management of the entire academic staff, whereas the second step is finding the result evaluation of the correlation between financial allocation and educational performance. The result obtained from implementing enterprise document management (EDM) is finding the indicator that measures the required financial efforts for education and helping decision makers estimate future efforts to enhance education. However, the paper presented a general overview of EDW implementation and neither went through design methodology nor explained how to select the proper approach of implementing EDM. Another model implemented by Abdullah and Obaid [36] combined educational records from two simulated databases of 10 years of data from the Department of Computer Science at Basrah University and four years of data from Al-Iraq University, Iraq. They unified the data under a single schema of EDW and then used OLAP to conduct a descriptive analysis and find student achievements through these years. The decision makers in the proposed model are lecturers and department heads. However, the approach of EDW design, the access method to EDW and the multidimensional cube were not clearly explained.

Kurniawan and Erwin [37] showed the advantages of using DW and DM in the prediction of students' performance. The model implementation also passed through two stages analysis and DW implementation; they used the Kimball approach to design DW with a star schema as a structure for the DW. However, the type of DW in the model was neither determined nor distinguished from the data mart. Mohammed *et al.* [38] designed an architectural approach on the basis of DW to combine databases from different Iraqi universities for increasing information sharing among all universities, colleges and departments. They used the beneficial characteristic of DW application to maximise information sharing among universities. However, the approach failed to show how to solve conflicts among databases from different universities. The difficulties in this approach are the huge data of a single university because many colleges exist with many departments and units in a single college. The paper also failed to explain how to deal with different standards and business rules among databases of a single college and how to handle this obstacle.

3. DSS IMPLEMENTATION

The data mart is the base of the proposed DSS. The flow chart of the implemented data mart is demonstrated in Figure 1. Six basic important steps are taken to implement the educational data mart; these steps are data pre-processing; data profiling; data mart area, which holds the staging area and the data mart schema; ETL; educational cube building; and OLAP, KPI and reports. The evaluation process is performed after completing the knowledge base of the DSS.

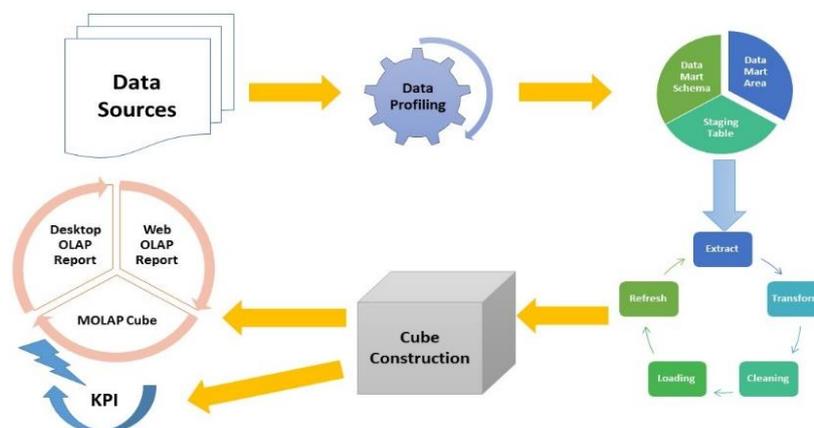


Figure 1. Flow chart of educational data mart implementation approach

The DSS architecture consists of four major areas, which are data preparation area; data mart area; KPI, OLAP and reports area; and decision-making area, as shown in Figure 2. The data preparation area involves all data source transformation and selection processes performed on the paper-based data source and their conversion into an electronic data source. The data mart area involves a data staging area where all the extracted data are stored and transformed to be loaded into data mart schema tables (fact tables and dimensions). The OLAP and query area provides many access methods to data mart analysis results from online to offline and multidimensional OLAP (MOLAP). KPI is an essential tool in the DSS that helps stakeholders in measuring progress. The last area is the decision-making area where all stakeholders (analysts, school managers, teachers and senior decision makers) reach the reports and make their decisions.

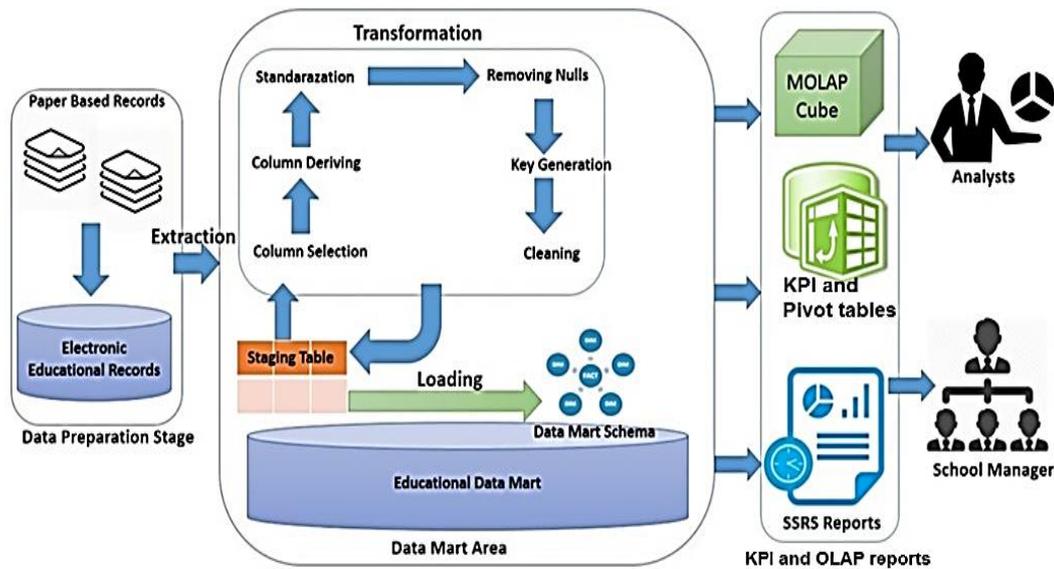


Figure 2. Model architecture

3.1. Data preprocessing

The base element of the proposed DSS is the collected data. The data source of the educational data mart is sourced from the Al-Isclandaria Private Elementary School; this school is in the Bahadria District, Basrah Province, Iraq. The data are paper-based data sheets, which hold all the required documentations for all students’ degrees. The school follows the paper-based procedure to calculate averages and success rates on different subjects and in different stages; thus, storing them in an electronic base takes a long time. The main table of student records contain the following details after storing them in the electronic base Table 1. In this table, the main attributes in data sources are explained in detail, content types, data types, and details of each attribute are presented.

Table 1. Details of students’ records

Field	Content Type	Data Type	Details
Student_ID	Continuous	Long	Represents the number of students currently attending the school.
Student_Number	Continuous	Long	The number of the paper record containing all the student's data that may be referred to in case of looking for more extensive data on the student (e.g: medical records).
Full Name	Discrete	Text	First three parts of the student's name as written in Arabic order (Student name + Father name + Grandfather name).
Birth	Continuous	Date	The student's birthdate presented in this format.
Enrollment	Discrete	Text	The year of which the data was taken presented in this format.
Address	Discrete	Text	Name of the district in which the student currently resides in.
Class	Continuous	Long	Number of the school year the student is in at the time of the recording of the data presented in an integer datatype.
Groups	Discrete	Text	Due to the huge number of enrolled students, a reasonable number of students of the same year are divided into groups (e.g: 5-A, 5-B, 5-C).
Subject	Discrete	Text	Name of the subject the class provides in each year, differing in number.
Grade	Discrete	Text	The grades are registered for each month (October, November, December, Mid-term, March, April, Final).

The second table as shown in Table 2 of data profiling presents the subjects that have been learned from class to class. Eight subjects taught are for students from classes 1 to 6. The subjects differ for each class; some subjects are taught in the first class, and different subjects are taught from the fourth class to the sixth class. For example, Islamic studies, Arabic language, mathematics, science, arts and physical education are taught from Classes 1 to 6, whereas social studies are taught from classes 4 to 6. Finally, English language is taught in classes 5 and 6.

Table 2. Subjects taken by classes

Subject Number	Subject	Class
1	Islamic Studies	1-6
2	Arabic Language	1-6
3	Mathematics	1-6
4	Science	1-6
5	Arts	1-6
6	PE	1-6
7	Social Studies	4-6
8	English Language	5-6

3.2. Data profiling

Data profiling, sometimes called data analysis, is the assessment and examination of data consistency, integrity and quality of data source. Using data profiling is a fundamental process to examine the data quality of DW data sources. Data profiling provides results that can be depended on when making a decision related to DW implementation. Data profiling concentrates on the individual attributes of the data source. It gives a complete summary that describes the length, data type, length, variance, uniqueness, null ratio, and domain range. It shows the full view of data quality related to all data source attributes [39], [40]. Data profiling is an important step in the building process of DW and data mart. Building data mart or DW does not actually succeed if this step is not performed. The results of data profiling can help in determining the dimensions and fact tables, the proposed primary keys, the null ratio in each column, mean and standard in each column, the maximum and minimum values in each column and the domain of each column. The result of the data profiling of student records is shown in Table 3. SSDT provides a data profiling tool that presents the results graphically for ease of understanding and use. Data profiling results are converted into table readings, as shown in Table 3.

Table 3. Data profiling results

Seq	Field Name	Minimum Value	Maximum Value	Number of Distinct Values	Null Ratio
1	Address	-	-	11	0
2	Birth	9-7-2003	27-2-2012	528	104
3	Class	1	6	6	0
4	Groups	-	-	5	0
5	Mark	-	-	613	0
6	Month	-	-	613	0
7	Number_of_Students	525	525	599	0
8	School_Address	-	-	1	0
9	School_Name	-	-	1	0
10	Student_Name	-	-	599	0
11	Student_Number	3	613	599	0
12	Subject	-	-	12	0
13	Year	2018	2018	1	0

3.3. Data mart area

The data mart consists of two areas: Staging and data mart schema tables. The staging area consists primarily of a staging table where the data are first extracted from the data source and loaded into it. The staging table faces all transformation processes and holds the final version of the table before loading it into dimensions and fact tables. Data mart schema is a star schema where five-dimension tables are connected to the fact table whilst their data are taken from the staging table. The staging table is an intermediate area used for storage between source systems and the DW; a temporary storage area is where data are successfully deleted after being uploaded to the repository. This area is used in many major processes, such as archiving and preparing data source, data extraction, cleaning, unifying, mirror, conversion, loading and indexing, quality assurance and updating [41], [42]. These processes are usually referred to as ETL. The staging area should be prepared and performed whilst all the intended OLAP queries are asked. The staging area also

holds the table with all columns and dimension tables. The staging table becomes the place where all data are manipulated. Data manipulation involves all data integrity processes, such as cleaning, transformation, enrichment and deletion. The data values of all columns must be the same as the data values of all columns in the dimension tables.

The data mart schema tables involve many approaches for designing DW, such as top-down, bottom-up, inside-out, and mixed approaches. Selecting an approach depending on the overall size and implementation duration period of DW can result in either enterprise DW or small data mart [43]. The top-down approach is used for long-term designing models and takes further analysis and redesigning to fit all enterprise goals. The bottom-up approach is used for short-term designing models where the results can be observed [44]. Given that the intended DW is an independent educational data mart, the best approach to build the data mart is the bottom-up approach. Based on the previously listed reasons, using this approach in implementing a data mart provides many facilities to build a solid solution that can provide answers to educational stakeholders.

Three famous schemas are used to implement the DW schema (i.e. star, snowflake and fact constellation). Each schema has its own advantages and disadvantages, which make designers prefer one schema over the other. Star schema is the famous one for its simplicity and wide usage; it consists of a central table called a fact table and many other tables called dimensions that surround the fact table. The fact table consists of many concatenated keys to the dimensions and many other keys called measurements, which represent the facts or functions that can be calculated along other dimension columns [45]. Figure 3 represents the proposed star schema of an educational data mart.

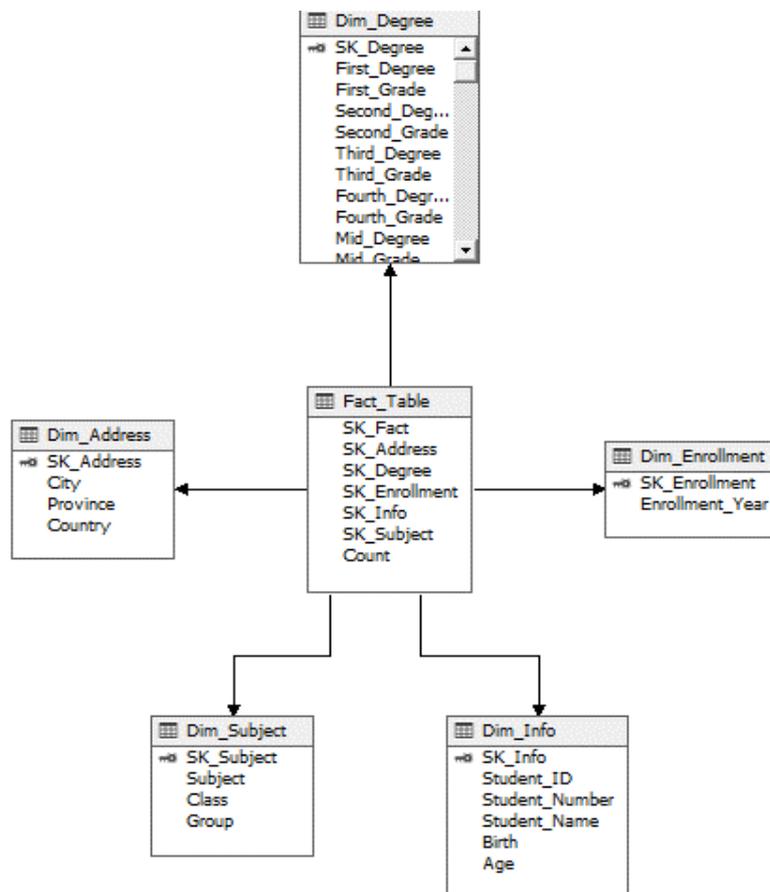


Figure 3. Star schema of educational data mart

The data mart schema is built using SSMS 2014. The schema consists of five dimensions (address, information, enrolment, subject, and degree) and the fact table. The fact table holds five keys to concatenate the dimension tables and one key measurement (count), which is the count function for finding the number of students represented by all the OLAP query answers. Using dimensions is one of the key factors that hasten

the OLAP responses. Concept hierarchy is normally used with dimensions such as date and address dimensions to permit OLAP operations, such as roll-up, drill down, slice, dice and pivot. The star schema can provide fast response OLAP queries, handle changes with time, have multiple hierarchies in dimensions, can build a simple DW schema and can easily be loaded with data [46], [47].

The next stage is performing ETL tasks, which take approximately 70% of the DW development time and cost spent on implementing the overall DW model. ETL involves many tasks used to manipulate data for obtaining their final cleaned and integrated version. The tasks of ETL (not limited to) are:

- Data extraction: The first step in the process of transferring data to the DW. It means reading the data and understanding them from different sources and then copying the necessary parts to the data submission area to continue the work later. The extraction step represents the greatest effort in the DW.
- Data cleaning is a task of detecting errors in the data and correcting them if possible; it involves tasks such as dealing with missing elements and reducing noise by defining extreme values and correcting data conflicts [48].
- Data transformation: When data are extracted from the source system, a series of actions is applied to convert the data into valid and meaningful formulas [49].
- Load: Load services need support before and after loadings, such as the regeneration of indexes and physical sections of the table. The specificity and structure of each goal when loading is also considered [50].
- Refresh: The last step of ETL where updates over time are transferred from data sources to repositories [51].

3.4. ETL

ETL is the stage where designers unconsciously prepare for fast-answered OLAP queries. In ETL, three important factors are prepared to make OLAP fast, namely, measurements, loading dimensions, and concept hierarchies. The first two stages (extract and transform) are implemented on the previously created staging table. The extraction process not only involves selecting the required data but also testing if the data are fulfilling the intended goals.

The two parts of extract and transform (ETL) are implemented on the staging table by using the SSIS package. The loading data mart strategy is divided into two stages, namely, loading dimensions and loading fact table. Figure 4 shows both stages of loading data mart schema tables. The first stage is loading the five-dimension tables with data. The multicast tool is used to make an image-like staging table to load dimensions by using a slowly changing dimension (SCD). Address and info dimensions are loaded using SCD with changing attributes. The next three dimensions (subject, enrolment and degree) are loaded with fixed attributes. The major difference between fixed and changing attributes is that fixed attributes do not detect changes in the staging table after loading data into dimensions, whereas changing attributes detect changes and reflect such changes in dimensions. Fixed attributes can be used for all dimensions' attributes.

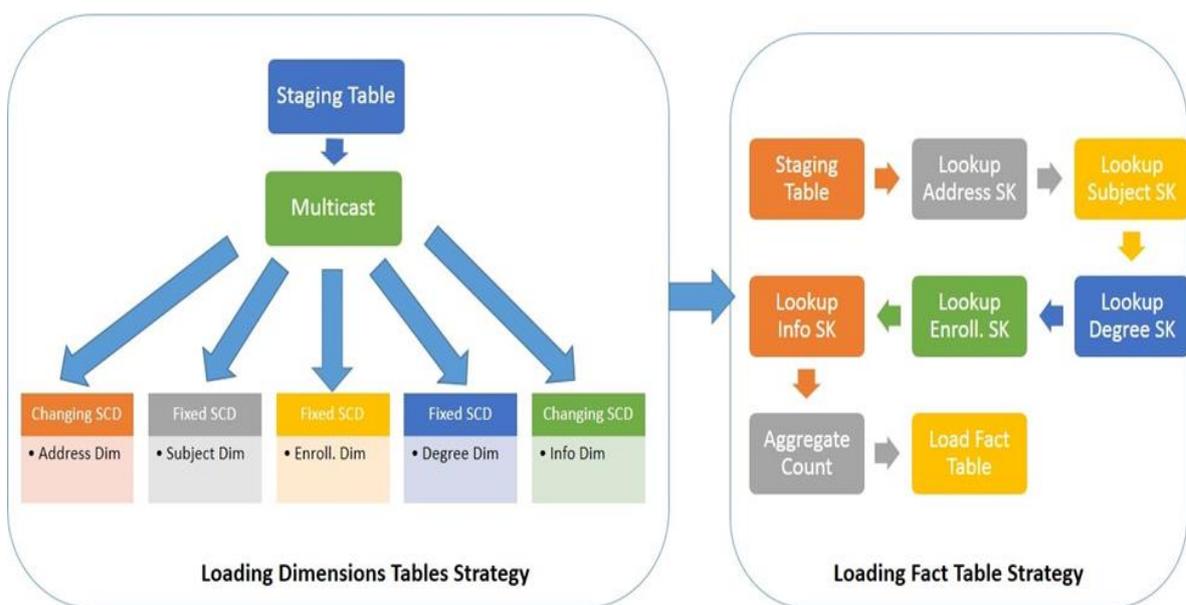


Figure 4. Loading dimensions and fact tables strategy

The second stage is building a cube to provide an analytical platform to perform OLAP queries. The cube is constructed to provide analysts with a platform where they can ask questions and find their answers as charts or tables. The educational multidimensional cube is implemented using SSAS 2014, which consists of the required dimensions and hierarchies. The cube consists of dimensions that can be used to answer OLAP queries on the basis of the measurement (count). Multidimensional cube is selected due to its advantages, such as fast complex query response and excellent performance [52], [53].

3.5. OLAP and KPI

The process of implementing OLAP and constructing reports is performed. OLAP is a technology used on DW architecture to obtain fast and accurate results as answers to complex queries [4]. The OLAP cube is primarily implemented to support complex queries on highly dimensional data structures. OLAP simultaneously processes dimensions and fact tables with the possibility to roll back if an error occurs. The OLAP cube is a popular important component of DW; the OLAP cube server stores security settings and complex calculations that can be integrated into data mining tools and algorithms [54]. The OLAP system is built on top of a relational database; OLAP has different categories, such as multidimensional OLAP (MOLAP), relational OLAP (ROLAP), hybrid OLAP (HOLAP), desktop OLAP (DOLAP), database OLAP and web OALP. MOLAP server is implemented by a multidimensional database, and all cube indexes are stored and retrieved. By contrast, ROLAP server sends query parameters and receives query answers to/from the relational database. HOLAP is a combination of the strengths of MOLAP and ROLAP features. DOLAP can be considered a variation of ROLAP. In DOLAP, users have the portability to perform OLAP complex queries by using existing DOLAP software on a pre-created multidimensional dataset on user desktop. In database OLAP, OLAP calculation can be performed on a relational database management system that supports the OLAP structure. Web OLAP allows OLAP calculations from web browsers. Three ways can be used to view OLAP results, namely, SSAS 2014 cube view, online reports using SSRS 2014 and offline reports using Microsoft Excel 2013 pivot table [16], [55].

The first access method is the SSAS cube view. After implementing a multidimensional educational cube by using SSAS 2014, the resulting cube is stored with separated locations, which can be accessed from the SSMS login screen locally or remotely. The SSMS login screen provides remote access to the database, ETL, cube and report. Cubes can be easily viewed by dragging and dropping dimension columns and the measurement. The educational cube is implemented on the basis of the multidimensional category because of its performance and flexibility in implementing complex queries. Figure 5 shows the grade counts in the mid-term examinations of all the classes of students who reside in Bahadria City.

The first class has the greatest number of failed students with 228, followed by Classes 2 and 4 each with 180 students. In Class 3, 150 students failed, whereas Classes 5 and 6 hold 30 and 20 failed students, respectively. The figure also illustrates the other grades with the number of students according to class and grade. The figure represents the slice OLAP operation, where a slice of data is selected on the basis of Bahadria City as a dimension with other dimensions (class and mid-grade).

Dimension	Hierarchy	Operator	Filter Expression	Parameters
Dim Subject	Class	Equal	{ 1, 2, 3, 5, 4, 6 }	<input type="checkbox"/>
Dim Address	City	Equal	{ Bahadria }	<input type="checkbox"/>
<Select dimension>				

Class	Mid Grade	Count
1	F	228
2	F	180
3	F	150
4	F	180
5	E	65
5	F	30
5	G	53
5	M	59
5	P	44
5	V	53
6	E	70
6	F	20
6	G	36
6	M	29
6	P	25
6	V	20

Figure 5. Grades of students of all classes in Bahadria district in mid-term

The second method used to view reports is SSRS 2014 for constructing web OLAP. The reports are designed using SSRS 2014 to provide pre-defined web OLAP answers and remote access for analysts to view statistical and complex multidimensional query answers. Reports are stored in reporting servers that can be accessed remotely or locally within the intranet network. Figure 6 presents the results as a report using SSRS where these results are conducted in the previous Figure 5.

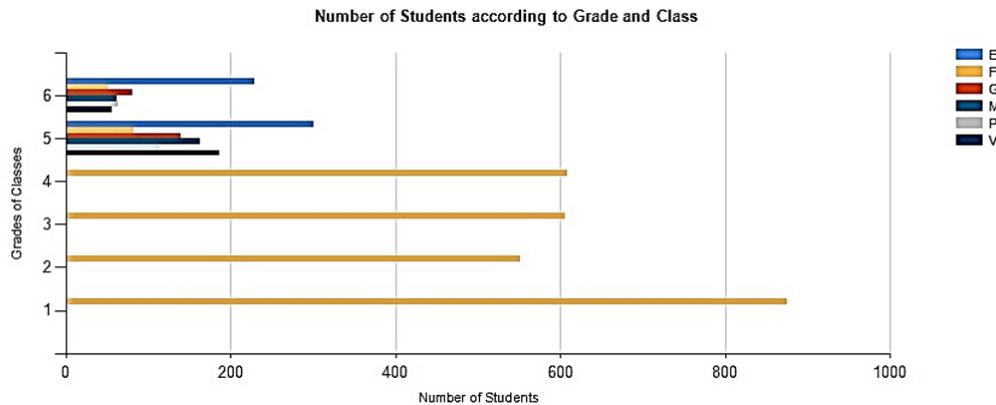


Figure 6. Grades of students in the mid-term exam

The last method of viewing reports is using the offline Microsoft Excel 2013 pivot table to implement desktop OLAP. This method allows offline access for instant access to the cube. In this method, the final staging table is imported to an excel file after the transformation processes, and the pivot table report chart is used to view the results. The pivot table provides the flexibility to select the required chart to view the results on the basis of the selected columns and a measurement. Excel pivot provides fast calculation functions of a selected table. After importing the staging table from SSMS 2014 and selecting the imported table, a pivot table can easily provide a chart to view the selected columns with a count aggregate function.

Figure 7 illustrates the examples of excel pivot tables. The figure lists the number of students according to their ages (10–15) as classified on the basis of the three selected grades (Excellent (E), good (G), and very good (V)). Two ages (11, 12) have the highest grades among the other ages. Students between the ages of 11 and 12 have the highest grade (E) among all students (with >300 cases), followed by those aged 12 (with >200 cases); these results can help analysts in finding the factors that improve student grades in these two age groups, including the factors that influence the grades in other ages.

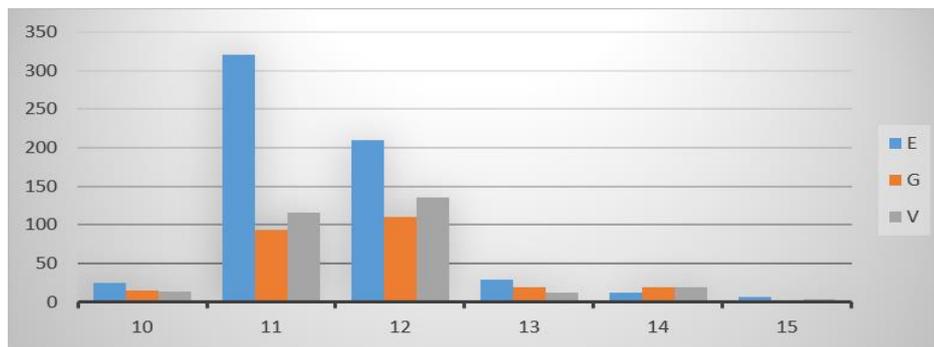


Figure 7. Number of students according to age and grade (E, G, and V)

Key performance indicator (KPI) is one of the most powerful tools in any DSS. KPI helps in measuring performance according to specific criteria. SSAS provides a powerful environment to build and implement KPIs that help in constructing DSSs [10]. Figure 8 shows the KPI of the success rate that helps in measuring the performance according to the number of successful students. The value cursor represents the number of successful students in the current year, and the goal is to increase the success rate by 10%. The

status indicator can be represented by three types of cursors where the cursor in the figure represents the current status. The status indicator colour changes according to the status. The trend indicator represents the performance measure according to time. When the KPI value changes over time, the trend indicator changes. The figure represents the implemented report on the basis of the KPI of the success rate. SSAS also allows to drag and drop dimensions to measure KPI value according to the selected dimensions.

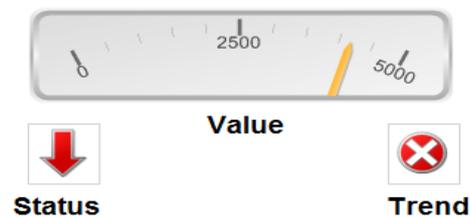


Figure 8. KPI of success rate

3.6. Data-driven educational DSS evaluation

In this stage, the proposed DSS is evaluated to find the framework feasibility. The proposed system is implemented on a platform with 8 GB of random-access memory (RAM), intel core i3, 2.4 GHz of processor, Intel (R) 82579 LM Gigabit network adapter, and 140 GB free sized hard. The access to platform is restricted and the online access throughout network is restricted to permitted users only. These restrictions to access DSS satisfy the security aspects of the system. The privacy of the DSS is ensured by hiding all the personal information of the students in the system. Educational DSS holds KPI and three kinds of OLAP reports (Web OLAP, MOLAP, and desktop OLAP). Each OLAP report holds many strength points and serves a specific kind of user. In the proposed educational DSS, the user (manager, analyst, decision maker) has the flexibility to use and navigate any type of reports according to needs. Analysts can use MOLAP reports. MOLAP permits to select more than one dimension to perform a multidimensional expression on the selected dimensions. Managers can use web OLAP reports to view the analyst results on the predefined OLAP reports through the network. Desktop OLAP reports can be used by decision makers and any other users to perform all OLAP operations instantly and show the analytical results according to the selected dimensions. OLAP reports allow users to navigate the educational cube on the basis of the OLAP operations. KPI is a powerful tool in any DSS that helps in giving status and trend indicators and presenting the current value and how near it is to the goal.

The static evaluation method is performed to check the DSS errors by inspecting knowledge base without using DSS. The proposed DSS confirms the security aspects where the dashboards are accessed by permitted users only. These restrictions also confirm the privacy aspects where students' personal information is hidden from analysts. The evaluation process of DSS is a static method where the knowledge base is reviewed to check the errors and mistakes in the knowledge base of the DSS without using DSS. The presentation is checked to confirm the static evaluation requirements. The DSS usability, cost effectiveness and effectiveness are checked on the basis of the system usage [56]-[61]. Subsequently, the dynamic evaluation is performed over test cases after applying the DSS in the school. In this step, the experts check the obtained results from the test cases. All the usability factors are examined, such as learnability, understandability, error prevention, accuracy, operability, efficiency, attractiveness, and effectiveness [62]-[67].

4. CONCLUSION AND FUTURE WORKS

In this study, a roadmap for implementing an independent educational data mart is explained. The independent educational data mart is the base of the resulting DSS. Many difficulties are dealt with in implementing an educational data mart, such as handling the paper-based data source, implementing the ETL package, applying OLAP, KPI, and finally designing and deploying reports. The data source pre-processing stage involves converting paper-based student information and degrees into electronic data sources by using SSMS. The ETL package is implemented using SSIS, which involves many ETL tasks, such as deriving columns, filling missing values, constructing surrogate keys and constructing concept hierarchies. Four approaches can be used to design and implement models (i.e. top-down, bottom-up, inside out and mixed). The best approach to design the data mart is a bottom-up approach because the result is required in a short period to support decisions for constructing enterprise DW. The educational data mart provides many

benefits, such as the provision of a fast-implemented platform to support educational decisions related to the academic performance of students; it also provides academic decision makers, teachers and school managers with solutions to the factors that improve students' success on the basis of their age, class, address and subject. In addition, the educational data mart provides educational stakeholders with a comprehensive overview of all data to detect outliers.

Three famous schemas exist for implementing DW, namely, star schema, snowflake schema and fact constellation. Implementing a schema is a complementary step with the OLAP cube implementation because the OLAP cube completely depends on the previously constructed fact tables and dimensions. The proposed schema for the educational data mart is the star schema, which is selected on the basis of many reasons, such as the fast responses to OLAP queries, easiness to apply hierarchies on dimensions and easiness to adapt changes in the schema. The educational data mart can be considered a solid platform for data mining applications that need the information collected from the entire school. The use of a data mart at the school level is preferred in this type of analysis and aggregation of large data. Given the complexity and breadth of its scope of work, the DW of this institution is usually managed by the central information technology (IT) departments. ETL tasks take approximately 70% of time and effort during data mart implementation. Given that the data source of student degrees and information is converted into an electronic data source, the extraction task is performed smoothly.

The transformation task consists of many processes, such as deriving columns, converting data, filling missing values and constructing surrogate keys. The loading task primarily consists of two stages-loading dimensions and loading fact tables. SCD has three types that are used for loading dimensions, namely, historical, fixed, and changing attributes. However, only two types (fixed and changing attributes) are used to load dimensions. A fixed attribute is also used to load dimensions for one time and cannot allow changes, whereas a changing attribute detects changes in staging tables and reflects such changes on the dimensions. Loading fact tables involves looking up for keys in the dimensions and calculating the count measurements to load these looked-up keys and calculated measurements into the fact tables. Many OLAP categories can be implemented throughout this model, such as MOLAP, desktop OLAP and web OLAP. OLAP queries; MOLAP is especially fast because of its use of dimensions, concept hierarchies, and measurements. These three factors, in addition to storing a cube as indexed, make OLAP responses fast. KPI is an essential tool in any DSS where the KPI of a proposed success rate is implemented to measure the progress of increasing the success rate in a school. Based on the proposed educational data mart, data marts can be considered reliable platforms for performing machine learning and data mining algorithms, such as decision tree, neural networks, clustering and association rules, to detect and find hidden patterns that affect student performance in the knowledge base of DSS.

The static evaluation method is performed to check the DSS errors by inspecting the knowledge base without using DSS. The proposed DSS confirms the security aspects where the dashboards are accessed by permitted users only. These restrictions also confirm the privacy aspects where the personal information of students are hidden from analysts. The evaluation process of DSS is a static method where the knowledge base is reviewed to check the errors and mistakes in the knowledge base of DSS without using DSS. The presentation is checked to confirm the static evaluation requirements. The DSS usability, cost effectiveness and effectiveness are checked on the basis of the system usage. Subsequently, a dynamic evaluation is performed over test cases after applying the DSS in the school. In this step, experts check the obtained results from test cases. All the usability factors are tested, such as learnability, understandability, error prevention, accuracy, operability, efficiency, attractiveness and effectiveness. The proposed system can be depended as an analysis platform where it can be used to implement a mobile application to transmit online analytical results to all stakeholders. These results can be used to measure the progress to satisfy the school goals set earlier. This system can be used for implementing data mining algorithms in order to find the hidden patterns, relationship among attributes, and students failure causes. Using these algorithms can enhance students' performance and though improve academic performance of the school. Besides, a knowledge mining will be implemented to conduct the data mining algorithms results to allow users to find the relationship among features in the data.

REFERENCES

- [1] J. Han, J. Pei, M. Kamber, and J. Pei, "Data mining: Concepts and techniques third edition," *The Morgan Kaufmann Series in Data Management Systems*, vol. 5, no. 4, pp. 83-124, 2011.
- [2] S. Bouaziz, A. Nabli, and F. Gargouri, "Design a data warehouse schema from document-oriented database," *Procedia Computer Science*, vol. 159, pp. 221-230, 2019, doi: 10.1016/j.procs.2019.09.177.
- [3] S. Sundari, M. N. Fadli, D. Hartama, A. P. Windarto, and A. Wanto, "Decision Support System on Selection of Lecturer Research Grant Proposals using Preferences Selection Index," in *Journal of Physics: Conference Series*, vol. 1255, no. 1, Art. No. 012006, 2019, doi: 10.1088/1742-6596.

- [4] I. A. Najm, A. K. Hamoud, J. Lloret, and I. Bosch, "Machine Learning Prediction Approach to Enhance Congestion Control in 5G IoT Environment," *Electronics*, vol. 8, no. 6, Art. No. 607, 2019, doi: 10.3390/electronics8060607.
- [5] W. H. Inmon, "Building the data warehouse," John Wiley & Sons, 2005.
- [6] T. H. Davenport, "Putting the enterprise into the enterprise system," *Harvard Business Review*, vol. 76, no. 4, pp. 121-131, 1998.
- [7] A. Baghal, M. Zozus, A. Baghal, S. Al-Shukri, and F. Prior, "Factors Associated with Increased Adoption of a Research Data Warehouse," in *ITCH*, vol. 257, pp. 31-35, 2019, doi: 10.3233/978-1-61499-951-5-31.
- [8] R. Kimball and M. Ross, "The data warehouse toolkit: the complete guide to dimensional modelling," John Wiley & Sons, 2011.
- [9] T. Z. Ali, T. M. Abdelaziz, A. M. Maatuk, and S. M. Elakeili, "A Framework for Improving Data Quality in Data Warehouse: A Case Study," *2020 21st International Arab Conference on Information Technology (ACIT)*, 2020, pp. 1-8, doi: 10.1109/ACIT50332.2020.9300119.
- [10] A. K. Hamoud and T. A. S. Obaid, "Design and implementation data warehouse to support clinical decisions using OLAP and KPI," Department of computer science, University of Basrah, 2013.
- [11] A. Hamoud and T. Obaid, "Building data warehouse for diseases registry: first step for clinical data warehouse," *International Journal of Scientific and Engineering Research*, vol. 4, no. 7, pp. 636-640, 2013, doi: 10.2139/ssrn.3061599.
- [12] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Clinical data warehouse: a review," *Iraqi Journal for Computers and Informatics*, vol. 44, no. 2, 2018.
- [13] A. Hamoud, H. Adday, T. Obaid, and R. Hameed, "Design and implementing cancer data warehouse to support clinical decisions," *International Journal of Scientific and Engineering Research*, vol. 7, no. 2, pp. 1271-1285, 2016, doi: 10.2139/ssrn.3061594.
- [14] M. M. Triola and M. V. Pusic, "The education data warehouse: a transformative tool for health education research," *Journal of graduate medical education*, vol. 4, no. 1, pp. 113-115, 2012, doi: 10.4300/JGME-D-11-00312.1.
- [15] A. K. Hamoud, M. A. Ulkareem, H. N. Hussain, Z. A. Mohammed, and G. M. Salih, "Improve HR Decision-Making Based on Data Mart and OLAP," In *Journal of Physics: Conference Series*, IOP Publishing, vol. 1530, no. 1, 2020, Art. no. 012058, doi: 10.1088/1742-6596/1530/1/012058.
- [16] A. K. Hamoud, H. N. Hussien, A. A. Fadhil and Z. R. Ekal, "Improving service quality using consumers' complaints data mart which effect on financial customer satisfaction," in *Journal of Physics Conference Series*, vol. 1530, no. 1, 2020, Art. no. 012060, doi: 10.1088/1742-6596/1530/1/012060.
- [17] R. Chowdhury, P. Chatterjee, P. Mitra, and O. Roy, "Design and implementation of security mechanism for data warehouse performance enhancement using two tier user authentication techniques," *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 3, no. 2, pp. 165-172, Feb. 2014.
- [18] Hashim, A. Salah, W. A. Awadh, and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 928, no. 3, p. 032019, 2020, doi: 10.1088/1757-899X/928/3/032019.
- [19] A. Hamoud, "Applying association rules and decision tree algorithms with tumor diagnosis data," *International Research Journal of Engineering and Technology*, vol. 3, no. 8, pp. 27-31, 2017, doi: 10.2139/ssrn.3028893.
- [20] A. Hamoud and A. Humadi, "Student's Success Prediction Model Based on Artificial Neural Networks (ANN) and A Combination of Feature Selection Methods," *Journal of Southwest Jiaotong University*, vol. 54, no. 3, 2019.
- [21] A. Hamoud, A. Humadi, W. A. Awadh, and A. S. Hashim, "Students' success prediction based on bayes algorithms," *International Journal of Computer Applications*, vol. 178, no. 7, pp. 6-12, Nov. 2017, doi: 10.2139/ssrn.3080633.
- [22] A. Hamoud, "Selection of best decision tree algorithm for prediction and classification of students' action," *American International Journal of Research in Science, Technology, Engineering & Mathematics*, vol. 16, no. 1, pp. 26-32, 2016.
- [23] A. S. Hashima, A. K. Hamoud, and W. A. Awadh, "Analyzing students' answers using association rule mining based on feature selection," *Journal of Southwest Jiaotong University*, vol. 53, no. 5, pp. 1-15, Oct. 2018, doi: 10.3969/j.
- [24] A. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, pp. 26-31, 2018.
- [25] A. K. Hamoud, "Classifying students' answers using clustering algorithms based on principle component analysis," *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 7, 2018.
- [26] G. Fulantelli, D. Taibi, and M. Arrigo, "A framework to support educational decision making in mobile learning," *Computers in human behavior*, vol. 47, pp. 50-59, Jun. 2015, doi: 10.1016/j.chb.2014.05.045.
- [27] T. Pradhan and S. Pal, "A multi-level fusion based decision support system for academic collaborator recommendation," *Knowledge-Based Systems*, vol. 197, no. 7, Art. No. 105784, Jun. 2020, doi: 10.1016/j.knosys.2020.105784.
- [28] T. Feghali, I. Zbib, and S. Hallal, "A web-based decision support tool for academic advising," *Journal of Educational Technology and Society*, vol. 14, no. 1, pp. 82-94, 2011.
- [29] F. Di Tria, E. Lefons, and F. Tangorra, "Research Data Mart in an Academic System," *2012 Spring Congress on Engineering and Technology*, 2012, pp. 1-5, doi: 10.1109/SCET.2012.6341952.
- [30] I. Mekterovic, L. Brkic, and M. Baranovic, "Improving the ETL process and maintenance of higher education information system data warehouse," *WSEAS transactions on computers*, vol. 8, no. 10, pp. 1681-1690, 2009.

- [31] M. E. Zorrilla, "Data warehouse technology for e-learning," in *Methods and Supporting Technologies for Data Analysis*, ed: Springer, vol. 225, pp. 1-20, 2009, doi: 10.1007/978-3-642-02196-1_1.
- [32] S. Belciug and F. Gorunescu, "A brief history of intelligent decision support systems," *Intelligent Decision Support Systems-A Journey to Smarter Healthcare*. Springer, Cham, 2020, pp. 57-70.
- [33] O. Smyrnaki, "Data warehousing in higher education. A case study of the Hellenic Mediterranean University," Thesis, Hellenic Mediterranean University, 2020.
- [34] S. Suman, P. Khajuria, and S. Urolagin, "Star Schema-Based Data Warehouse Model for Education System Using Mondrian and Pentaho," in *International conference on Modelling, Simulation and Intelligent Computing*, vol. 659, pp. 30-39, 2020, doi: 10.1007/978-981-15-4775-1_4.
- [35] M. Paunica, M. L. Matic, A. L. Manole, and C. Motofei, "Measuring the performance of educational entities with a data warehouse," *Annales Universitatis Apulensis: Series Oeconomica*, vol. 12, no. 1, pp. 176-184, 2010.
- [36] Z. A. Abdullah and T. A. Obaid, "Design and implementation of educational data warehouse using OLAP," *International Journal of Computer Science and Network-IJCSN*, vol. 5, no. 5, 2016.
- [37] Y. Kurniawan and E. Halim, "Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits)," *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, 2013, pp. 98-103, doi: 10.1109/TALE.2013.6654408.
- [38] M. A. Mohammed, A. R. Hasson, A. R. Shawkat, and N. J. Al-khafaji, "E-government architecture uses data warehouse techniques to increase information sharing in Iraqi universities," *2012 IEEE Symposium on E-Learning, E-Management and E-Services*, 2012, pp. 1-5, doi: 10.1109/IS3e.2012.6414947.
- [39] I. F. Ilyas and X. Chu, "Data cleaning," *ACM*, 2019.
- [40] R. Zhang, M. Indulska, and S. Sadiq, "Discovering data quality problems," *Business & Information Systems Engineering*, vol. 61, no. 5, pp. 575-593, 2019, doi: 10.1007/s12599-019-00608-0.
- [41] T. Rujirayanyong and J. J. Shi, "A project-oriented data warehouse for construction," *Automation in Construction*, vol. 15, no. 6, pp. 800-807, November 2006, doi: 10.1016/j.autcon.2005.11.001.
- [42] A. Sen and A. P. Sinha, "A comparison of data warehousing methodologies," *Communications of the ACM*, vol. 48, no. 3, pp. 79-84, 2005.
- [43] P. A. Sabatier, "Top-down and bottom-up approaches to implementation research: a critical analysis and suggested synthesis," *Journal of public policy*, vol. 6, no. 1, pp. 21-48, 1986.
- [44] C. Ghezzi, "Designing data marts for data warehouses," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 10, no. 4, pp. 452-483, 2001, doi: 10.1145/384189.384190.
- [45] P. Ponniah, "Data warehousing fundamentals for IT professionals," John Wiley & Sons, 2011.
- [46] N. Tryfona, F. Busborg, and J. G. B. Christiansen, "StarER: a conceptual model for data warehouse design," in *Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP*, November 1999, pp. 3-8, doi: 10.1145/319757.319776.
- [47] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM Sigmod record*, vol. 26, no. 1, pp. 65-74, 1997, doi: 10.1145/248603.248616.
- [48] J. Caserta and R. Kimball, "The data warehouse toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data," Wiley, 2013.
- [49] A. Sellami, A. Nabli, and F. Gargouri, "Graph NoSQL Data Warehouse Creation," *Proceedings of the 22nd International Conference on Information Integration and Web-based Applications & Services*, Nov. 2020, pp. 34-38, doi: 10.1145/3428757.3429141.
- [50] S. Thulasiram and N. Ramaiah, "Real Time Data Warehouse Updates Through Extraction-Transformation-Loading Process Using Change Data Capture Method," in *International Conference on Computer Networks and Inventive Communication Technologies*, vol. 44, pp. 552-560, 2019, doi: 10.1007/978-3-030-37051-0_62.
- [51] S. Chakraborty and J. Doshi, "Materialized Queries with Incremental Updates," in *Information and Communication Technology for Intelligent Systems*, ed: Springer, vol. 106, pp. 31-40, 2019, doi: 10.1007/978-981-13-1742-2_4.
- [52] A. Hamoud and T. Obaid, "Using OLAP with diseases registry warehouse for clinical decision support," *International Journal of Computer Science and Mobile Computing*, vol. 3, no. 4, pp. 39-49, 2014, doi: 10.2139/ssrn.3061597.
- [53] R. T. Rimi and K. A. Hasan, "Efficient Query Processing for Multidimensional Data Cubes," in *International Conference on Cyber Security and Computer Science*, Springer, Champp, vol 325, pp. 647-658, 2020, doi: 10.1007/978-3-030-52856-0_51.
- [54] M. C. Trivedi, V. K. Yadav, and A. K. Gupta, "A Proposed DDS Enabled Model for Data Warehouses with Real Time Updates," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 7, no. 1, pp. 31-38, 2018, doi: 10.11591/ijict.v7i1.pp31-38.
- [55] E. Pourabbas, "Providing accurate answers to OLAP queries based on standardized moments of data cubes," *Information Systems*, vol. 94, 2020, Art. no. 101588, doi: 10.1016/j.is.2020.101588.
- [56] A. Krtalić and M. Bajić, "Development of the TIRAMISU advanced intelligence decision support system," *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 40-55, 2019, doi: 10.1080/22797254.2018.1550351.
- [57] M. el Yaakoubi, P. Ravesteijn, A. Prinsen, H. Hooimeijer, and M. van der Ven, "Data Driven Decision Support: The Role of the Controller in Decision-Making Processes," in *16th European Conference on Management, Leadership and Governance*, 2020, pp. 73-80, doi: 10.34190/ELG.20.026.
- [58] G. Jayashree and C. Priya, "Comprehensive Guide to Implementation of Data Warehouse in Education," in *Intelligent Computing and Innovation on Data Science*, ed: Springer, vol. 118, pp. 1-8, 2020, doi: 10.1007/978-981-15-3284-9_1.

- [59] G. Roccasalva, "Towards a DSS: A Toolkit for Processes of Co-designing," in *Project and Design Literacy as Cornerstones of Smart Education*, ed: Springer, vol. 158, pp. 49-52, 2020, doi: 10.1007/978-981-13-9652-6_4.
- [60] H. M. Iqbal, R. Parra-Saldivar, R. Zavala-Yoe, and R. A. Ramirez-Mendoza, "Smart educational tools and learning management systems: supportive framework," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 14, pp. 1179-1193, 2020, doi: 10.1007/s12008-020-00695-4.
- [61] A. Häggman-Laitila, P. Saloekkilä, and S. Karki, "Integrative review of the evaluation of additional support programs for care leavers making the transition to adulthood," *Journal of pediatric nursing*, vol. 54, pp. 63-77, September-October 2020, doi: 10.1016/j.pedn.2020.05.009.
- [62] K. Shafinah, M. Selamat, R. Abdullah, A. Muhamad, and A. Noor, "System evaluation for a decision support system," *Information Technology Journal*, vol. 9, no. 5, pp. 889-898, 2010.
- [63] A. B. Deraman and F. A. Salman, "Managing usability evaluation practices in agile development environments," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 1288-1297, April 2019, doi: 10.11591/ijece.v9i2.pp1288-1297.
- [64] T. Wahyuningrum and K. Mustofa, "A systematic mapping review of software quality measurement: Research trends, model, and method," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 5, pp. 2847-2854, Oct. 2017, doi: 10.11591/ijece.v7i5.pp2847-2854.
- [65] H. V. Gamido and M. V. Gamido, "Comparative review of the features of automated software testing tools," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, pp. 4473-4478, Oct. 2019, doi: 10.11591/ijece.v9i5.pp4473-4478.
- [66] D. Delen, K. Topuz, and E. Eryarsoy, "Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition," *European journal of operational research*, vol. 281, no. 3, pp. 575-587, March 2020, doi: 10.1016/j.ejor.2019.03.037.
- [67] K. Fahd, S. J. Miah, K. Ahmed, S. Venkatraman, and Y. Miao, "Integrating design science research and design based research frameworks for developing education support systems," *Education and Information Technologies*, pp. 1-22, 2021, doi: 10.1007/s10639-021-10442-1.