Gümüşhane University Journal of Science

GUFBD / *GUJS* (2023) 13(3): 750-766 doi: 10.17714/gumusfenbil.1239468 Research Article

Implementing data warehouse infrastructure for an e-learning system

Uzaktan eğitim sistemi için veri ambarı altyapısının geliştirilmesi

İhsan GÜNEŞ*[1], Mustafa Kemal BİRGİN [2]

[1]*Eskisehir Technical University, Open and Distance Education Application and Research Center, 26555, Eskişehir*

[2]*Anadolu University, Open and Distance Education Faculty, 26000, Eskişehir*

• Received: 19.01.2023 • Accepted: 23.06.2023

Abstract

The use of a data warehouse (DW) in e-learning applications helps us evaluate students from different perspectives. Depending on user behavior, we may change the content or appearance of the e-learning platform to achieve the best results. The aim of this study is to build a data warehouse infrastructure, which is a central information system for Anadolu University's distance education system. Thanks to the data warehouse project, all learners' activities in the Learning Management System (ÖYS: Anadolum e-Kampüs) will be recorded regularly. Various analytical reports were obtained about students' study styles, preferred e-learning resources, and the time spent in the system. Additionally, the data warehouse system regularly pulls information from other information systems, such as the student information system related to the open education system. Data warehousing is designed to combine data from multiple sources into a single unified view, rather than performing operational operations like a database. Generally, data warehouses are used for analytical processing as opposed to transaction processing. They are used to analyze historical data and trends, as well as to identify patterns and insights that can aid in informing business decisions. When analysis and reporting is done in a different environment using data warehousing, there is no additional load on the database, which is already busy with daily operational processes. The data warehouse has made it easy to obtain student analytics for various periods, including daily, monthly, and yearly. The reports provided through this infrastructure are used in studies to increase the quality of the LMS system.

Keywords: Data analytics, Data mining, Data warehouse, Open and distance learning

Öz

E-öğrenme sistemlerinde veri ambarı (DW) kullanımı, öğrencilerin çeşitli açılardan değerlendirilmesine yardımcı olur. En iyi sonuçları elde etmek için, e-öğrenme platformunun içeriğini veya tasarımını kullanıcı davranışına göre değiştirebiliriz. Bu çalışmanın amacı, üniversitenin uzaktan eğitim programının ana bilgi sistemi olarak hizmet verecek bir veri ambarı altyapısı oluşturmaktır. Veri ambarı altyapısı sayesinde öğrencilerin öğrenim yönetim sistemindeki (ÖYS: Anadolum e-Kampüs) tüm etkinlikleri rutin olarak kayıt altına alınmaktadır. Öğrencilerin çalışma stilleri, tercih edilen e-öğrenme kaynakları ve sistemde harcanan zamanla ilgili çeşitli analitik raporlar almak mümkün olacaktır. Veri ambarı sistemi ayrıca açıköğretim sistemi ile ilgili öğrenci bilgi sistemi gibi diğer bilgi sistemlerinden de düzenli olarak veri çekmektedir. Veri ambarı, veritabanı gibi anlık operasyonel işlemleri yapmak yerine birden çok kaynaktan gelen verileri, tek bir birleşik görünümünde birleştirmek için tasarlanmıştır. Genel olarak, veri ambarları, işlem işleme yerine analitik işleme için kullanılır. Geçmiş verileri ve eğilimleri analiz etmenin yanı sıra iş kararlarını bilgilendirmeye yardımcı olabilecek modelleri ve görüşleri belirlemek için kullanılırlar. Veri ambarı kullanılarak analiz ve raporlama işlemleri farklı bir ortamda yapıldığında, günlük operasyonel işlemler ile meşgul olan veri tabanı üzerine daha fazla yük getirilmemiş olunur. Veri ambarı, günlük, aylık ve yıllık dahil olmak üzere çeşitli zaman dilimleri için öğrenci analitiği elde etmeyi kolaylaştıracaktır. Bu altyapı üzerinden sağlanan raporlar ÖYS sisteminin kalitesini artırmak amacıyla yapılan çalışmalarda kullanılmaktadır.

***Anahtar kelimeler**: Veri analitiği, Veri madenciliği, Veri ambarı, Açık ve uzaktan eğitim sistemi*

* İhsan GÜNEŞ; igunes@eskisehir.edu.tr

ISSN: 2146-538X e-ISSN: 2146-538X https://dergipark.org.tr/tr/pub/gumusfenbil

1. Introduction

There are approximately 3 million students, greater than 1 million active, in the Anadolu University Open Education System. Most of these active students use the LMS web and mobile applications. Student movements are stored independently in LMS and mobile application databases. In addition, data, such as the courses taken by the student, success grades and profile information, are also kept in the student information system. Apart from this, applications are related to many LMS systems, such as help desks, e-learning material tracking systems and gamification applications. There is no direct relationship between the information held in these systems. For example, to measure the effects of e-learning material on student success, data from LMS and student information systems should be taken separately, and a relationship should be established in a different database. Apart from this, in the current situation, when the LMS system is changed, there is no relation between the data in the old LMS system and the new one. In this case, data continuity will be lost because the data of the old and new LMS will be stored in different databases and the data structures will be different; therefore, it may not be possible to merge the data. Analyzing and reporting processes including past years may not be done correctly. Therefore, the data to be used in analytical and reporting processes should be stored in a data warehouse structure in a certain format instead of in operational databases. To solve all these problems, a data warehouse infrastructure is needed to establish the relationship between the LMS system and other information systems and to store the desired data in a certain format over the years.

Data warehouse and database are two different data storage methods. Data warehousing is the design and implementation of processes, tools, and facilities to manage and distribute complete, timely, accurate, and understandable information for decision making. It encompasses all the actions that make it possible for an organization to establish, operate, and maintain a data warehouse or data mart. In addition, a data warehouse is frequently understood to be an architecture that integrates data coming from a wide variety of sources in order to provide support for analytical reports, structured and/or unstructured queries, and decision-making processes. Data warehousing accesses disparate data sources, cleans, filters, and transforms it, and stores it in an easy-to-use format. Data are queried, reported, and analyzed. Thus, transaction-oriented operational environments have varied access, use, technology, and performance requirements. In data warehousing, historical data analysis requires a lot of data. Data analysis programs must scan large amounts of data, which could slow operational applications. To reduce operational conflicts and performance degradation, the two environments must be separated (Chuck et al., 1998, Golfarelli et al., 1998).

A database is a structured collection of data for efficient retrieval and management. Databases can handle everything from data entry forms to enterprise-level solutions. Operational applications that need real-time data access store data in databases designed for read write, update and delete activities. Data warehouse and databases differ mostly in purpose and design. Data warehouses support business intelligence and store massive amounts of historical data for analysis, while databases support operational applications and hold current, often updated transactional data. Data warehouses use denormalized data structures for read-intensive queries, whereas databases use normalized data structures for both read and write operations. As a consequence of this, databases are typically employed for the processing of operational transactions, while data warehouses are used for report generation.

In this study, data analytic operations were performed on a data warehouse. Data warehousing and data analytics are related but distinct data management concepts. In data analytics, data sets are analyzed to draw conclusions. Data analytics on a data warehouse involves using advanced techniques and tools to extract insights and value from large volumes of structured and semi-structured data stored in a data warehouse. Data analytics on a data warehouse can help organizations gain a deeper understanding of their operations, identify trends and patterns, make informed decisions, and improve business performance. However, the task of a data warehouse is to centrally collect, store and manage very large amounts of data. A data warehouse centralizes enterprise-wide data to improve business intelligence and decision-making.

In the open education system, students are offered many course materials, such as e-books, e-audio books, summary lecture videos, TRT school videos, unit summaries, summary voice-overs, leaf tests, end-of-unit practice exams, one-question and one-answer videos, animation and video tutorials. Based on the need for a learning management system (LMS) that offers open and distance education services as a whole, has a modular structure and provides the highest level of interaction. The Anadolum eKampus system, built on Blackboard learning infrastructure, was put into use in 2016. Later, in 2019, the LMS was developed by the software developers of the Open Education Faculty. As mentioned above, different LMS systems have been used in e-learning processes. However, since the data related to the LMS were not collected in a central data warehouse, a relationship could not be established between the old data and the new data in case of switching LMS software. Since the data of previous years resided in the database of the old LMS and the data structure of the new LMS was different, merging of data was not possible. Thus, there were difficulties in preparing reports and analyses that included historical data.

The educational data warehouse collects a lot of data such as student profile, demographics, class attendance or data about activities intended to influence student outcomes, which are recorded within universities' information systems, repositories of learning management systems (LMSs), communication/collaboration platforms, or separate files created by the teaching staff (Gladić & Petrovački, 2021). The educational data warehouse developed within the scope of this study stores many student and course data such as web and mobile lms access, material access, course access, trial exam results, forum data, course enrollments, demographics, course information and course material information.

The main purpose of this study is to collect all student, instructor, and staff movements in the learning management system and other related information systems in a certain format. Then, in the data warehouse, the requested analytical reports are regularly created based on the accumulated data. These reports can include different analytics, such as material usage frequencies, student interactions, and the relationship between LMS system usage and student success. In this context, there is a need for tools to help monitor and measure the activities performed by all students to evaluate the effectiveness of e-learning actions and the structure of the courses. One of the technologies that can be used to meet this need is data warehouse technology. Unlike the LMS database, whose sole purpose is to support the operation of the application, all information in the data warehouse must be consistent, organized, and standardized. Therefore, a data warehouse that follows the dimensional model is necessary to extract the data and retain them after transformation. To this end, the Learning Management System as a data source integrated with the business intelligence tool is a good starting point to provide the answers needed. The true potential of such databases lies in retaining data and hidden patterns. The main concurrent addition for most service providers is the ability to find, understand and use these patterns with intelligent analytics in numerous databases.

The main contributions of this study are listed below:

1. Determining the reports and analyses to be gathered through a meeting with open education system field experts.
2. Determining the information systems from which to pull data based on the needed reports and analyses.
3. Developing the data warehouse infrastructure and executing data extraction, conversion, and loading operations
4. Creating web interfaces for displaying reports to authorized users.

The remainder of the paper is structured as follows. In the next section 2, related studies are reviewed. In Section 3, the e-learning data warehouse design stages are described. We describe technical data warehouse infrastructure in Section 4. Finally, in Section 5, the conclusion and further studies are reported.

2. Related work

The data warehouse is the central data storage area where the data stored in the different information systems of the organizations are kept in an organized manner. In addition, since the data in the data warehouse is kept simplified in accordance with the data warehouse's objective, the analysis and reporting processes are more efficient than those in the operational database. A data warehouse is a repository where related data can be queried and analyzed. If the desired queries and reports are created in the data warehouse, the databases of the information systems will not be forced in terms of performance. A data warehouse performs the necessary operations to analyze relevant data easily, fast, and accurately. Data from transactional systems are copied and stored in a data warehouse usefully decision-making. Data and information can be obtained in different information systems and associated with each other. The processes of extracting, transforming, and loading the requested data from information sources are called the ETL Extraction-Transformation-Loading process. The major activities listed below serve as an overview of how these tools work: A dataset is cleaned up based on database and business rules, (a) by identifying relevant information on the source side, (b) by removing that information, (c) by customizing and integrating information from various sources in a common format, (d) by cleaning the final dataset, and (e) by distributing the data into data warehouses and/or data marts (Simitsis et al., 2005).

Nebi & Mahni (2010) proposed a data warehouse solution for a variety of purposes, including information on LMS system utilization, the degree of institutional adoption, course activity, and monitoring and analysis of platform usage. The authors offer models for a potential data warehouse that would enhance decision-making and offer resources for using OLAP and data warehouse technologies to evaluate and explain outcomes. The authors used the Moodle e-learning system at the University of Rijeka to explain how to develop a functioning data warehouse solution and several associated analyses. The E-learning Data Warehouse Development Lifecycle model created by the authors, is depicted in the picture below in Figure 1.

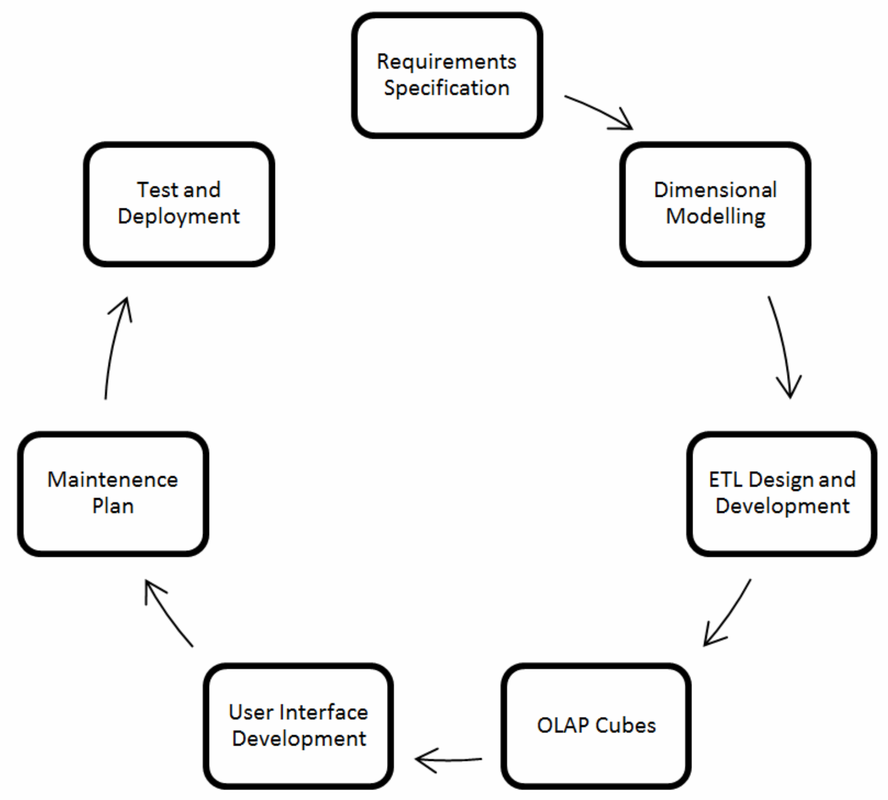

Figure 1. E-learning data warehouse development lifecycle (Nebić & Mahnič, 2010)

According to the authors, the data warehouse development lifecycle steps are described as follows:

- Requirements specification: The requirements document, which outlines the questions that must be resolved before the e-learning data warehouse can be considered complete, is the product of the first step of the development process. It usually includes specifying the reports and analyses required to create the data warehouse schema. Student activities by day, month, and year as an example.

- Dimensional modeling: A dimensional schema is composed of a central fact table and analytical dimensional tables. Dimensional modeling is creating a multidimensional schema based on user needs analysis and importing it into the base system, which is often a relational database. The dimensional model begins with two different kinds of tables: Fact tables are followed by dimension tables. Fact tables contain metrics, or items that we wish to measure, whereas dimensional tables contain the attributes of the metrics, or descriptions of what is being measured. Each fact table is associated with multiple dimensions, some of which are shared by multiple fact tables, but the dimensions are unrelated.
- Extraction-Transformation-Loading (ETL) Design and Development: This includes establishing the data stage schema, specifying the ETL processes, and configuring synchronization settings between the data source and the data warehouse.

- OLAP Cubes: The major goal of this phase is to develop and set up online analytical processing (OLAP) cubes needed to supply information for reports. An OLAP cube is a multidimensional data structure that enables the organization and analysis of data based on multiple dimensions or perspectives. OLAP cubes provide quick and versatile data access for analysis and reporting and can be used to answer complex business questions requiring multiple levels of aggregation and drill-down.

- User interface development: creates a user interface that enables end users to browse and create reports from data in the data warehouse.

- Maintenance Plan: This covers backing up and maintaining the database that makes up the data warehouse.

- Test and Deployment: At this stage, the analytics and reports are tested for accuracy and the application is activated.

Goyal & Vohra (2012) used data warehousing and data mining methods to increase the efficiency of higher education institutions. Thanks to these methods, they tried to learn student performance and behavior carefully. The authors designed the data warehouse infrastructure in four sections in their studies. These:

- Data sources: The operational systems that support the organizations main business operations are where the data in the data warehouse is obtained from. Production data, internal data, external data, and archived data are some examples of the four primary categories of source data that enter the data warehouse.

- Data preparation: Operational data are cleaned and transformed at this layer so that they may be stored and quickly retrieved. Data extraction, data transformation, and preparation for loading are the three basic tasks carried out during the data preparation.

- Data storage: The data are now kept in the data store. The operational system of a company can only support data that is now available, but, since the data warehouse is designed for large data storage and fast reporting, it will be easier to access historical data.

- Information distribution: At this point, a large data warehouse user is given helpful information using various platforms, including online, intranet, internet, and e-mail.

By fusing data from university administration information systems with WebCT learning management system log information, Solodovikova & Niedrte (2005) presented a centralized data warehouse model. The model enables several e-learning points of view. These are for senior management, academic administration, teachers who create courses, and departments in charge of the standard and organization of the educational process. For image definition, measurements characteristic of business functions is suggested. Analysis of measurement results in terms of WebCT usage at the university throughout the study is then presented. Gladi & Petrovaki (2021) present a pilot project of a DWS in the paper to track how well students are meeting course requirements in both traditional and distance learning settings. A DWS that provides data analysis and the generation of numerous reports in the portion of the teaching process that deals with student achievement was developed as part of the project. The analysis findings and information reveal potential drawbacks that must be quickly discovered to minimize the effects. To help teachers, students, and administrators make decisions, Araque et al. (2007) conducted studies on the use of the data warehouse as an integrated module in an e-learning platform. The authors have also demonstrated how OLAP analysis can be used to collect data that will be used to enhance the interaction between teachers and students. The authors wanted to ensure that the e-learning platform automatically adjusted to each student who used it for future studies; therefore, they planned to add adaptive hypermedia concepts into the platform. In a case study conducted at the Catholic University of Murcia, Cantabella et al. (2019) described how student behavior throughout the previous four academic years was examined in relation to learning modalities. To speed up the statistical analysis of the data, statistical and association rule techniques have been employed using a Big Data framework. This is due to the difficulties of managing enormous volumes of user-generated data in the LMS. The Apache Hive data warehouse architecture is used to store data. To swiftly extract the data, the authors used the Hadoop MapReduce technique. Visual analytical approaches were used to demonstrate the obtained results.

An adaptive e-learning system named AE-LS was proposed by Janati et al. (2019) to track and analyze student behavior. This strategy offers teachers and administrators thorough reports to assess, perform, and evaluate student activities. It enables teachers to see, examine, and explore student behavior to enhance the instruction and reach the best decisions. The authors' research has demonstrated that students spend less time on courses and tutorials and instead prefer to engage in group projects such as participating in forums and virtual classrooms. To develop an analytical model for AE-LS, the authors suggested a data warehouse model implemented in the PostgreSQL database management system (DBMS). All data and analysis concerns addressed by adaptive e-learning systems are handled by the proposed model. The primary goal of the data warehouse system is to gather information from various AE-LS sources that are required to enable analysis and facilitate decision-making. This data warehouse solution allows us to query the data and saves a considerable amount of data cumulatively over the years. To predict student performance in schools, Kurniawan & Halim (2013) created a model that may be used in data warehousing and data mining techniques. The authors examined the correlations between the variables housed in the data warehouse and extracted fundamental information from it using data mining techniques. In the study, it was examined how low-achieving students can benefit from the data mining and data warehouse model, how they can assess the applicability of a course or module, and how the interventions required to improve students' academic performance in schools can be used. Student retention is being used as a gauge for academic success and record management (Zhang et al., 2010). The authors of the study tracked the student, examined the student's academic behavior, and established a foundation for effective intervention tactics using data mining and natural language processing technologies. The study aims to identify potential issues as soon as feasible and follow-up with alternatives for intervention to improve student retention. The authors talked about how data mining might help students determine their risk status, gauge the applicability of a course or module, and modify necessary interventions to boost student retention.

Solodovikova & Niedrīte (2005) combined WebCT log files with a university management information system and WebCT data. They suggested storing integrated data in a data warehouse. This study supported senior management, faculty management, course designers, and learning process quality, and planning departments' e-learning opinions. The authors established four viewpoints to determine the indices needed to measure course consumption at different management levels and instructor levels. The authors presented data warehouse star schema information regarding the course structures and application. In their study, the star schema consisted of three fact tables: Structure Fact, Usage Fact, and Activity Fact, which comprise measurements, and of dimensions, which comprise descriptive data.

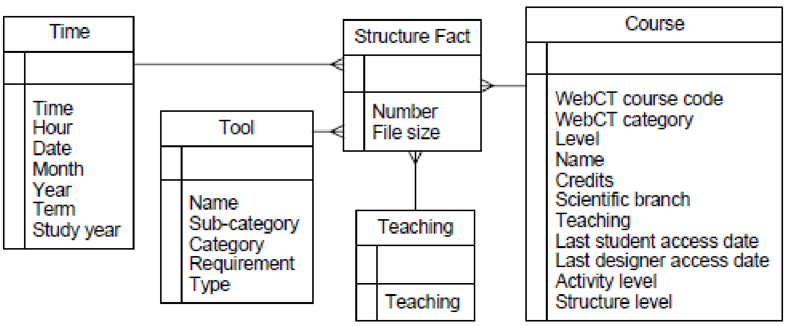

Figure 2. The Structure star schema (Solodovikova & Niedrīte, 2005)

Li et al. (2022) developed a multidimensional teaching and learning analytics data warehouse system (TLDWAS) to evaluate massive amounts of teaching and learning data. Senior management/educators could use TLDWAS's simple/interactive interface to view university students' applications, admissions, registration, and academic progress/status until graduation. This new analytic technology let policymakers and senior management see the broad picture and obtain insight from massive student course and teacher assessment data. Student, program, academic year, and course data were modeled using a multidimensional model in their paper. As shown in Figure 6, time, student, and course are dimension information in TLDWAS, while grade data is fact information in the data warehouse. Dimensional tables describe the dimension hierarchy for each dimension.

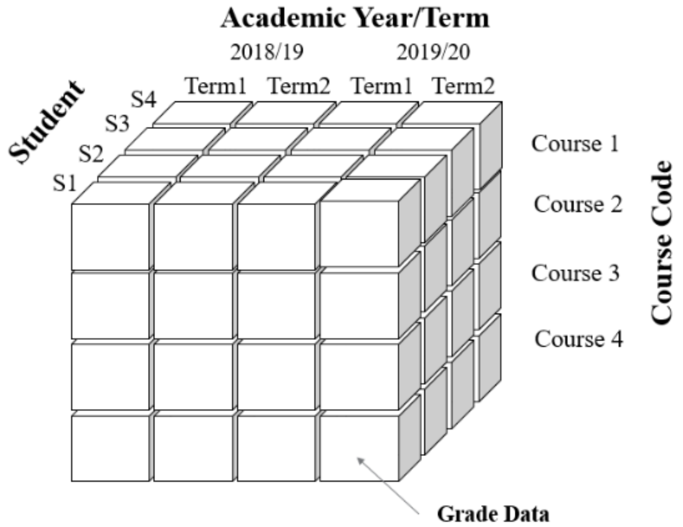

Figure 3. 3D data cube for the student performance tracking (Li et al., 2022).

Li et al. (2022) used multidimensional lenses to analyze the massive amount of data in the course student evaluation area, a crucial aspect of the teaching and learning quality assurance in higher education. On the basis of the data warehouse system, a novel analytical framework for categorizing courses into various domains (Student Effort, Outcome Achievement, Course Rating, and Teacher Rating) to address domain-specific course issues was proposed. Their framework enabled senior administrators and program leaders to quickly identify courses that require closer examination. Their streamlined system determined whether course enhancement measures are warranted and in what areas. The flow diagrams of the data warehouse architecture developed by the authors are shown in Figure 4. Figure 4 demonstrates that the operational database's raw sensor data are aggregated by ETL and then processed by the data warehouse, which divides them into various datamarts for the reporting service.

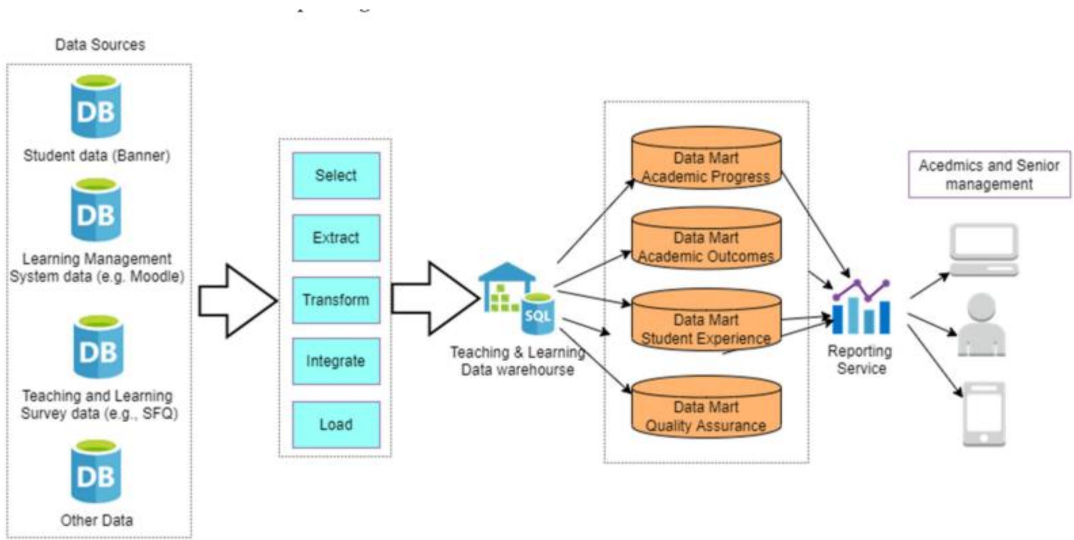

Figure 4. The data flow in the data warehouse (Li et al., 2022)

In study (Zorrilla, 2009), it was demonstrated that the technologies of data warehouse and OLAP are the most appropriate for developing this software application. The author detailed the implementation process from inception to the creation of the user interface. The author provided a summary of her experience with the design and implementation of MATEP, the Monitoring and Analysis instrument for E-learning Platforms, an instrument developed at the University of Cantabria.

According to the author, the fact that this application was built on a BI/DW architecture has the following technological advantages:

- The software is autonomous of the e-learning platform it audits.
- The application collects academic and demographic information, and instructor-specified context.
- The application provides instructors with up-to-date information based on the frequency with which the administrator executes ETL procedures. Generally speaking, daily.
- Even though the volume of data is large, the system manages the information effectively.
- The use of OLAP technology not only provides fast answers, but also permits users to manage the data by increasing the level of aggregation along one or more-dimension hierarchies, decreasing the level of aggregation, or reorienting the multidimensional view of the data.
- The system is readily expandable to accommodate the needs of other academic environment users.
- In addition, this can be used to retain the information that intelligent algorithms should obtain patterns and serve as their repository.
- The solution's modular design and adherence to industry standards facilitate the replacement of any component.

Figure 5 depicts the E-learning Data Warehouse architecture presented by the author. Presented are the four components of a data warehouse environment: data sources, data staging area, data area, and data access tools.

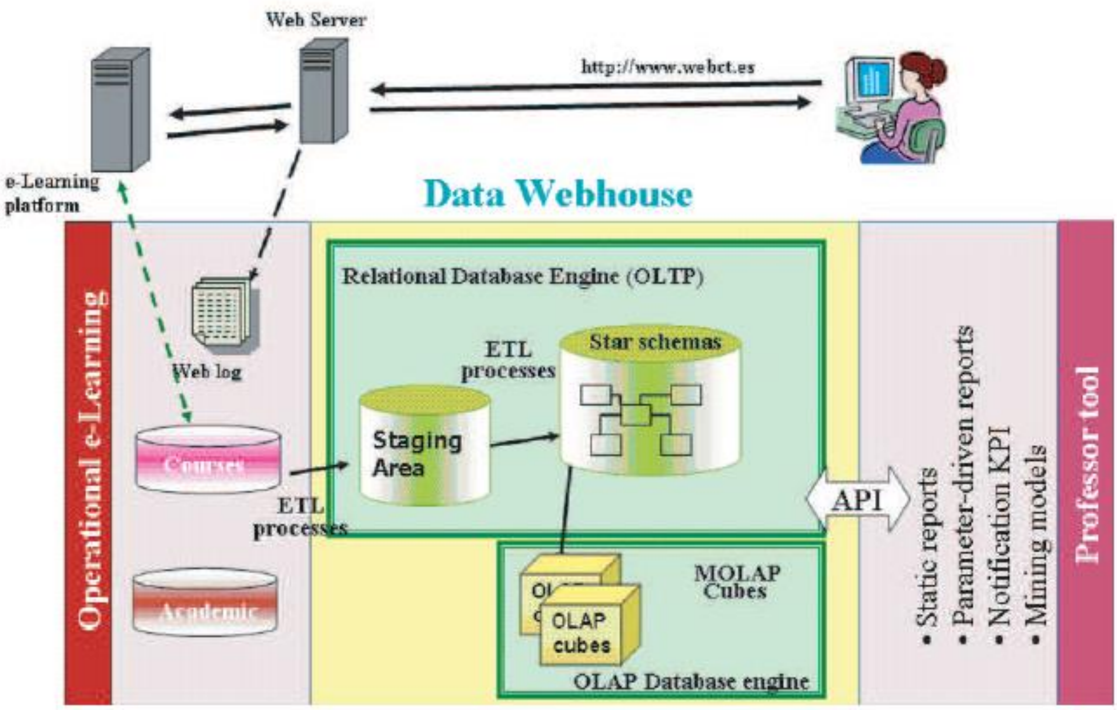

Figure 5. E-learning Data Webhouse (Zorrilla, 2009)

3. E-learning data warehouse design

An e-learning data warehouse (EDW) is a type of data warehouse designed expressly to meet the needs of educational institutions, such as schools, colleges, and universities. An EDW aggregates data from diverse sources, such as student information systems, learning management systems, assessment tools, financial systems, and other operational systems, to provide an all-encompassing view of the institution's data. In conclusion, an EDW is a specialized type of data warehouse that is tailored to the requirements of educational institutions, whereas a data warehouse is a more general-purpose system that can be applied in a variety of industries and settings.

At the stage of designing the data warehouse infrastructure, the opinions of experts who have administrative or academic experience in the distance education system were consulted to determine which analyzes and reports would be needed to improve the e-Learning processes and increase the quality of the Open Education system. At the same time, the opinions of experts in data mining and statistics were taken on which methods to apply for these analyses. Within the framework of the opinions, integrations to the information systems needed for the data warehouse was carried out. Data obtained from the participants, especially the administrators in the Open Education system and ÖTAG unit, were used.

The main purpose of this study is to collect all student, instructor and staff movements in the learning management system and other related information systems in a particular format. Then, the data warehouse regularly creates the requested analytical reports based on the accumulated data. These reports can include different analytics, such as material usage frequencies, student interactions, and the relationship between LMS system usage and student success. Tools are required in this situation to aid in monitoring and measuring the actions taken by all students to assess the efficacy of the course structure and e-learning activities. All of the information in the data warehouse needs to be consistent, arranged, and standardized, unlike the LMS database, which exists solely to facilitate the operation of the application. To extract the data and keep it after transformation, a new data warehouse that adheres to the dimensional model is needed. Data warehouse technology is one of the technologies that can be applied to fulfill this demand. Alternately, technologies like data lakes and NoSQL databases can be used for data storage, but because data is kept unstructured in such environments, analytics and report retrieval can be more challenging. The main contributions of the data warehouse system are the ability to locate, interpret, and exploit these patterns with intelligent analytics in oversized databases. Such databases have the genuine potential to hold data as well as hidden patterns.

In this context, this study sought answers to the following questions:

- The analytics and reports that will be requested should be determined in consultation with the administrators of the open education system. The data to be extracted from other information systems should be determined based on the requested analytics and reports
- Determining which data will be retrieved and how often they will be retrieved in the LMS system
- Determining which data will be retrieved and how often they will be retrieved from other applications, such as the Student Information System, support desk, and material management system (ETS) related to the LMS system within the scope of the Open Education System
- Transforming and simplifying data in accordance with the data warehouse
- Creating table structures in the database and determining the relationship between tables, combining and associating data from the LMS system and other systems
- Setting up the database server and preparing the infrastructure
- Preparation of works for extracting data from LMS and other systems
- Determining the reports requested after the creation of the data warehouse system
- Developing interfaces for writing queries and displaying reports required for the production of reports

The desired result of the study

- The accumulative storage of data in the LMS system and other related information systems in a secure environment over the years

- Improving the benefits and effects of e-learning based on analytic reports extracted using data from the data warehouse
- Optimizing e-learning materials in the LMS system
- Measuring the relationship between the use of the LMS system and student achievement, using simplified data
- Monitoring student interactions in the LMS system
- Individual guidance specific to the student can be provided by thanks to the data in the data warehouse

Within the scope of this study, it regularly stores the e-learning activities of all students in a secure data warehouse. In addition, the data of these students were correlated with the LMS system and different reports were provided. Through this study, complex queries that may cause performance problems on the LMS system can now be run on the data warehouse. In addition, this data warehouse infrastructure works independently of the LMS. Therefore, in the case of switching the LMS system in the future, there will be no disruption in the reporting system.

4. Data warehouse infrastructure

Enterprise-licensed Microsoft MSSSQL Server 2019 was employed as a database engine in the data warehouse system. The .NET environment has been used for the web service project that pulls the data and for the software of the web environment where the reports will be displayed. Also, as the programming language, C# is used within .NET environment.

The data warehouse stores data in three distinct forms and levels. These are known respectively as operational data stores (ODS), data warehouses (DW), and data marts (DM). The operational system's data are replicated to create the ODS layer. Providing the following data layers is the objective here. The primary responsibility of the ODS layer is to feed the DW layer. In the data warehouse, data moved from operational systems to operational data stores (ODS) is transformed to create the DW layer. The DM layer is composed of aggregated representations of the multidimensional data received via the DW layer. The necessary SQL codes were generated and then executed at specific intervals to transfer the necessary data to the aforementioned levels of the data warehouse, ensuring that the data warehouse was continually updated. This procedure is known as ETL (Extract, Transform, Load). During the extraction process, data from multiple sources are compiled in a single location. Transformation refers to cleaning and transforming the retrieved data into the appropriate format. The process of transferring updated data into the relevant resource is known as loading. In the data warehouse, the SQL codes required by ETL operations have been generated and implemented. MSSQL Server was used to set up linked servers, and the necessary connections were established to pull data from operational systems. The ETL procedure was executed daily in order to maintain an up-to-date data warehouse. MS SQL Server's job scheduling function was used to perform this operation.

4.1. Data warehouse schema

In data warehouse architecture, there are two possible designs. These are known as the star and snowflake schemas. This study employed a star schema architecture. In star schema architecture, a fact table is constructed within a structure fed by dim tables. In this structure, dim tables are not interconnected. These results in a denormalized structured architecture. This denormalized structure of the star schema necessitates more disk space; however, because it requires fewer joins, SQL expressions are significantly simpler and more efficient. The material usage activity fact table depicted in the figure is populated with data from numerous dim tables. Fact tables contain metrics, or items that we wish to measure, whereas dimensional tables contain attributes of metrics, or descriptions of what is being measured. Each fact table is connected to a number of dimensions, some of which are common to multiple fact tables; however, the dimensions are not connected to one another in any way.

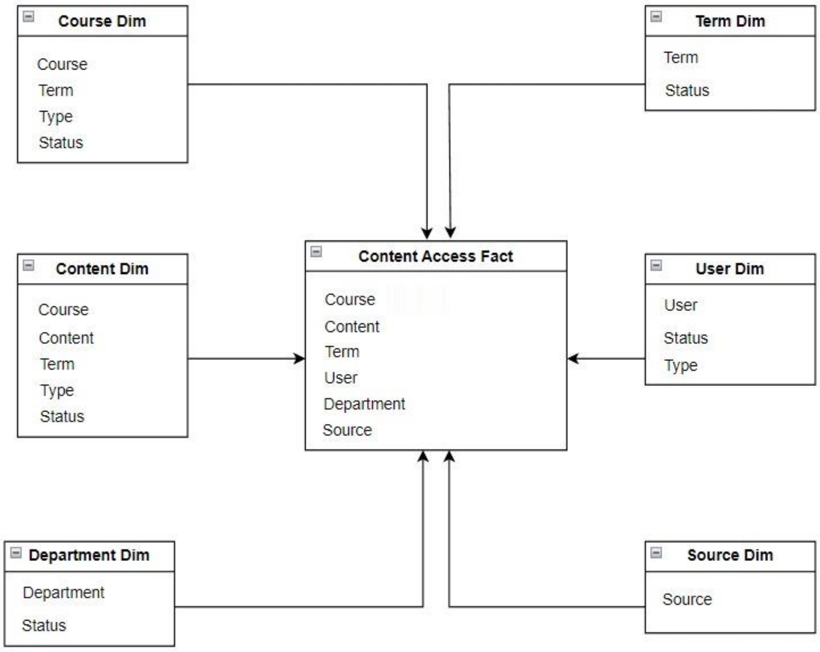

Figure 6. A sample fact table in the star schema

4.2. Datamart

In this study, many datamarts were used to obtain some standardized reports and analytics faster. A datamart is a subset of a data warehouse (DW) that serves a particular business function or department, such as sales, marketing, or finance. Typically, a datamart comprises a subset of DW data that has been pre-aggregated and organized to support the analytical requirements of the business function it serves. As it comprises a smaller subset of the data from the DW, the primary advantage of using a datamart is that it can provide faster access to the data required for a particular business function. As the data is already aggregated and organized for a particular business function, this can improve query performance and reduce the complexity of data analysis. Additional advantages of using a datamart may include:

- Improved data quality: By focusing on a certain business function or department, a datamart may ensure data accuracy, completeness, and consistency.

- Better decisions: Datamarts enable faster access to pertinent data, which can improve decision-making.

- Lower cost: A datamart requires less hardware, software, and development work than a full-scale data warehouse.

- Increased agility: Datamarts can adapt to change business requirements and analytical demands by focusing on a certain business function or department.

Using a datamart can improve the effectiveness and efficacy of data analysis for particular business functions, while also reducing costs and enhancing agility.

In the study, many datamarts were created when needed. Different datamarts such as course information, material information, demographics, general usage, temporal dimension of access to the system, course and material access summaries, semester usage statistics, general usage of mock exams, semester enrollments were created according to the needs.

The data warehouse is fed from the student information system (ÖBS), learning management system (ÖYS), material management system (ETS), and various flat file sources. LMS and other related databases are defined as OLTP (Online Transaction Processing) databases, the operational system in which the processes in the organization are managed. Data are pulled from these OLTP sources and transferred to the data warehouse system.

4.3. Stages of the data warehouse infrastructure

In general, there are three stages of the data warehouse infrastructure to be developed. These are ETL (Extract Transform Load), data storage, and data processing (analysis and reporting), respectively.

The ETL process:

a) Extract: It is called the process of extracting the data from the source system. It specifies the process of obtaining data from the LMS system and related information systems at certain times.

b) Transform: Putting the collected data into the desired format in accordance with the purpose, that is, cleaning it and increasing its quality,

c) Load: This means to load the data into the created central information system data warehouse.

Data Retention: The data organized with the ETL process are stored in the database that forms the infrastructure of the data warehouse. The data pushed into the data warehouse are permanent and cannot be modified. The data warehouse has a time dimension. Data are stored in chronological order.

Data analysis reporting: Data mining methods enable analysis of student movements in e-learning environments to enable administrators and educators to have an idea about the student's learning methods. Many analyzes can be performed, such as statistical analysis of training data, the most popular pages that students access, the number of downloads of e-learning resources, the number of different pages accessed, and the total time they spend in the system. These analytics are visualized for display to end users.

Thanks to the data warehouse infrastructure to be developed, student success can be predicted by running analyze on the data. With the recommendation system algorithms, suggestions, such as e-learning materials and study methods, can be offered to the students to enhance their performance. In this study, the data warehouse architecture developed for the distance education system is shown in Figure 7. The figure shows data sources, ETL processes and outputs that can be obtained from the system.

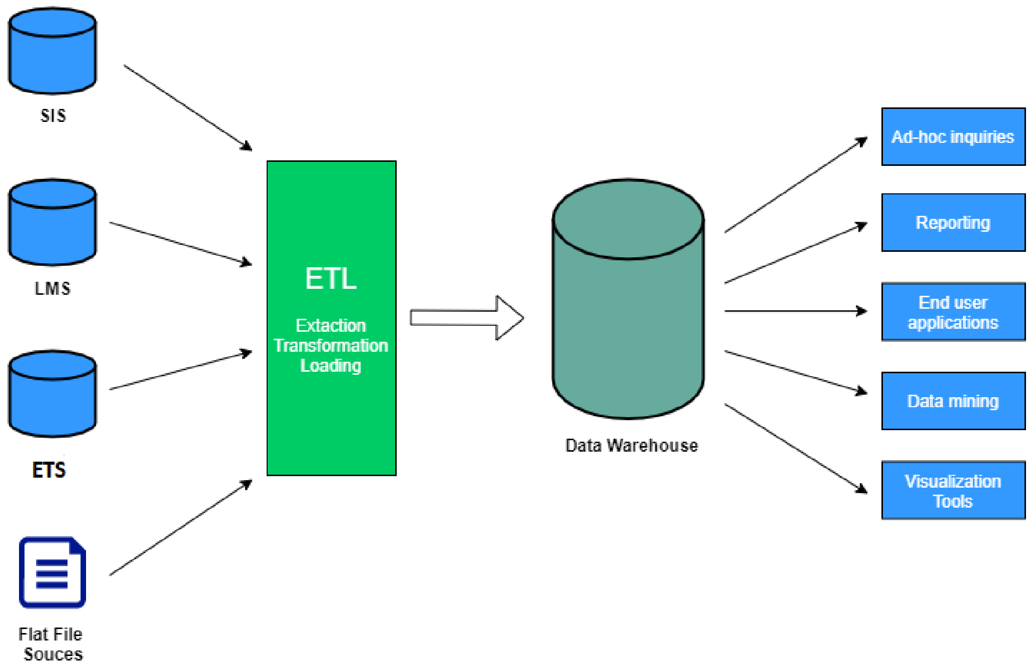

Figure 7. Open education system data warehouse architecture

4.4. Features of the designed data warehouse system

The designed system has the following features:

- Ensuring that the data in the LMS system and other related information systems are stored in a safe environment cumulatively over the years.
- Data can be retrieved at certain intervals from data sources, such as ÖYS, ETS and Student Information System, using web services.
- A web application has been developed to display the reports from the web interface. The web application is published at https://aofanalytics.anadolu.edu.tr/. People in the open education system can access the web page.
- Presenting reports have been made according to the role.
- A dashboard has been prepared for the administrators in the Open Education system to show the summary of the whole current semester.
- With the admin panel, administrative staff can access summary information, such as the number of student entries, the use of LMS according to the student profile status, and the material usage status with the LMS system.
- Reports created in the web interface can be downloaded as CSV files or images.
- Reports, such as material usage, individual usage numbers, and course access information, can be obtained.
- Reports can be received based on the Open Education System, faculty and program.
- The system allows us to generate reports for intervals, such as daily, weekly, or custom periods.
- By connecting the data warehouse system with visualization tools, such as Tableau and Power BI, instant reports and graphs, can be created based on their needs.
- In the case of a need for instant data not presented as the administrators, reports can be provided by SQL queries.
- Within the scope of data up-to-dateness, data mining and artificial intelligence algorithms can be run by connecting to a data warehouse system with software, such as Python, R or MATLAB, to be used in academic and administrative studies for the Open Education System.
- The data warehouse system is designed to be flexible and open to development. Optionally, new reports can be created.
- With the responsive interface design, transactions can be made by easily accessing mobile devices.
- Feedback about the usage of the LMS system can be shown to students from the data warehouse system. Summary reports showing the students work status in the LMS system can be presented. In addition, the usage of course materials and the scores in the practice exams are compared to the class average. Feedback is reflected on the students page in the LMS system (ekampus.anadolu.edu.tr).

Some of the reports available from the web interface are shown below. The dashboard report contains summary information such as users' web/mobile usage status, age, gender and number of unique users per week. The reports on the dashboard screen are organized according to the request of the institution administrators.

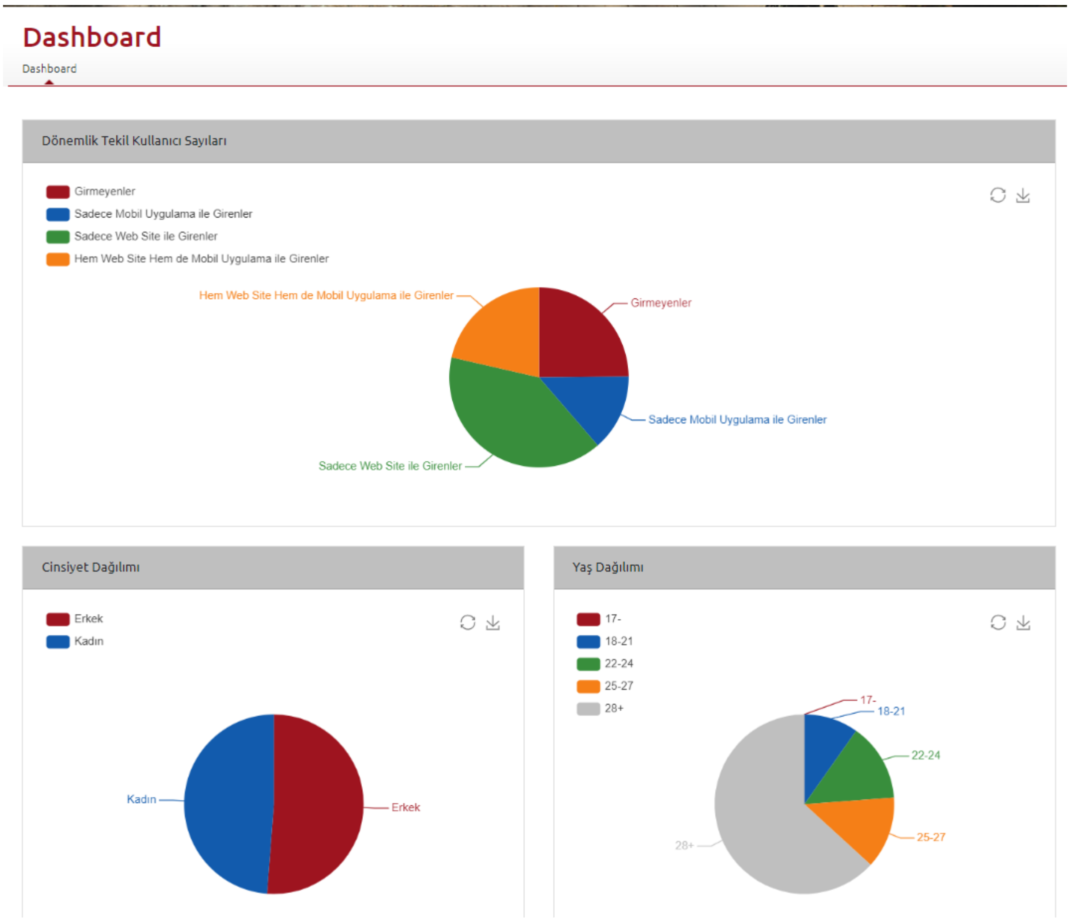

Figure 8. Summary Reporting View

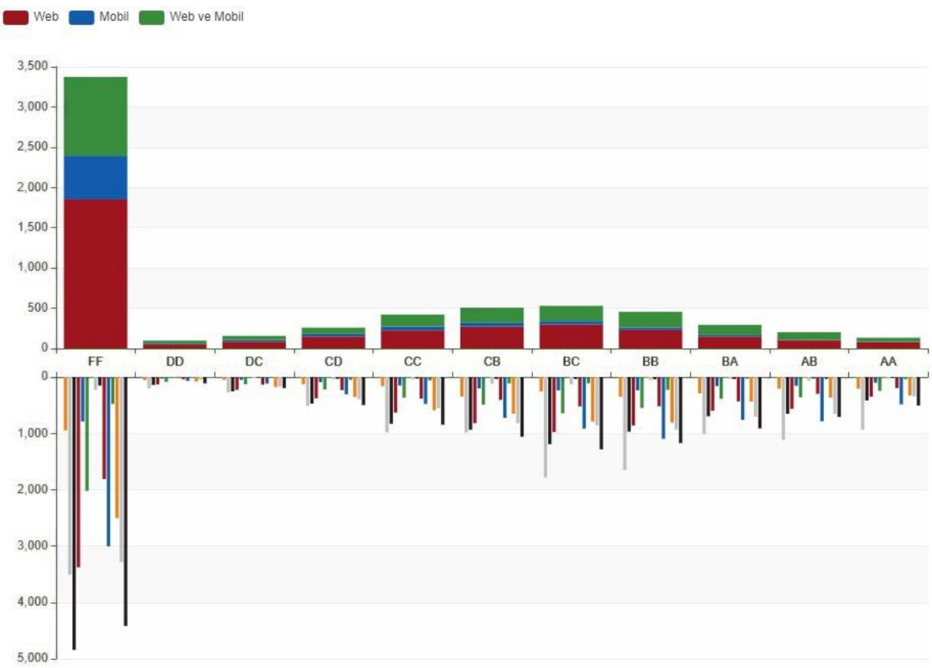

Figure 9. The relationship between material use and student achievement in a sample course

The graph in the Figure 9 shows an example of the multidimensional data obtained in the data warehouse. The data on the top axis shows the number of students classified according to their letter grades within a course. The platforms from which these students access the system are also shown in the same graph. On the other hand, the types of materials used by students belonging to each letter grade class are also shown on the lower axis. In this way, the platforms from which successful or unsuccessful students of a course accessed the system and which materials they used can be shown in a single graph. It is not possible to claim this kind of data from any other database other than the data warehouse.

The graph in the Figure 10 shows aggregate data on the usage of materials for a course. The total access to the materials of a course and the number of individual students accessing each material are given in a single graph. In this way, it is possible to make various inferences about a material. Questions such as "Do students continue to use materials that can be used offline by accessing them online or do they prefer to use them offline?" or "Is the high usage of materials in video format due to the nature or is it really used a lot?" can be answered through this graph.

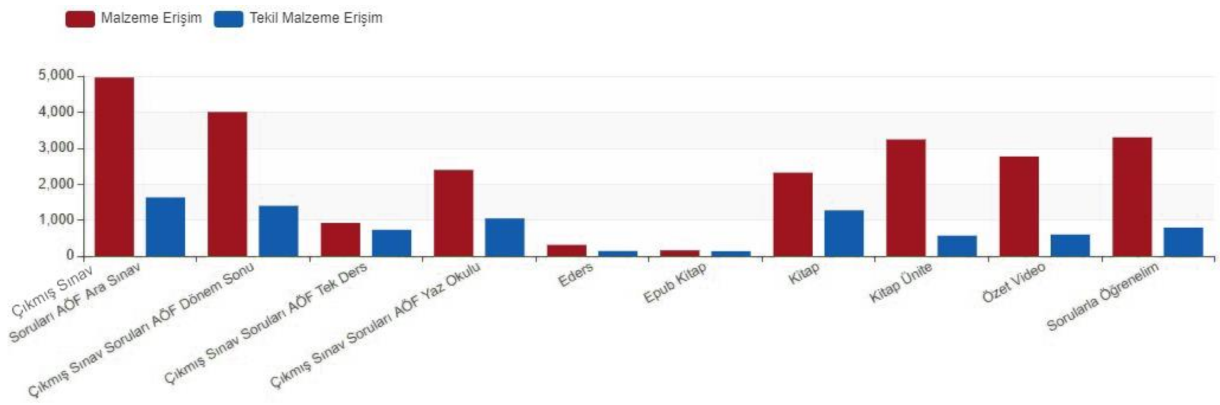

Figure 10. Number of material accesses in a sample course

5. Discussion and conclusion

With this study, the infrastructure of the environment where administrators, academicians, and related unit managers can easily receive reports on the e-learning system has been created. Thanks to this study, data from different data sources such as LMS, student information system, mobile applications have been simplified by combining them in a single environment. In this way, in-depth and faster analytical reports about the distance education system will be produced using simplified data. Reports obtained through the data warehouse help managers decide on processes for improvement with these reports, improvements can be made in areas such as LMS interfaces, material types and question quality to increase student success. In the study, datamarts were created in order to get some complex reports faster. New datamarts can be added to the system as needed. The data warehouse system is open to the continuous expansion. New reports can be added as new data related to Open Education is added. Since the interfaces are developed by the software developers of the institution, operations such as adding, removing and modifying reports can be done easily. Data is stored on the data warehouse server with secure and daily backups. The server can be accessed by authorized users over the local network. Maintenance operations are performed daily on the database server and operations such as backup, data integrity, and indexing is performed.

It is expected that the academic studies conducted using these arranged data will contribute to the scientific performance of our university. It is thought that this system developed for the open education system will set an example for other institutions implementing the distance education system in Turkey.

Today, as the data grows, it becomes difficult to process data with classical SQL queries. Since the data in the data warehouse will increase over the years, it is planned to integrate with technologies such as Hadoop, Apache Spark, which are used to process larger data in future studies. The e-learning recommendation system, which provides personalized recommendations to students based on their learning styles and past behaviors, is planned to be integrated into the data warehouse system. In addition, according to student behavior, it is planned to predict the student success status in advance and alert the system when necessary.

Acknowledgments

This work was partially supported by the Anadolu University Scientific Research Projects Coordination Department under Grant No 1808E291.

Author contribution

- Ihsan Gunes performed literature search, database design, data processing, prepared figures, authored or reviewed drafts of the article, and approved the final draft.

- Mustafa Kemal Birgin performed database design, coding, data collection, processing, authored or reviewed drafts of the article, and approved the final draft.

Declaration of ethical code

The author of this article declares that the materials and methods used in this study do not require ethical committee approval and/or legal-specific permission.

Conflicts of interest

The authors declare that they have no conflicts interests.

References

Araque, F., Salguero, A., Martínez, L., Navarro, E., & Calero, M. (2007). Data warehousing for improving web-based learning sites. *International Journal of Emerging Technologies in Learning*, *2*(4). https://doi.org/10.3991/ijet.v2i4.81

Ballard, C., Herreman, D., Schau, D., Bell, R., Kim, E., & Valencic, A. (1998). *Data modeling techniques for data warehousing* (p. 25). San Jose: IBM Corporation International Technical Support Organization.

Cantabella, M., Martínez-España, R., Ayuso, B., Yáñez, J. A., & Muñoz, A. (2019). Analysis of student behavior in learning management systems through a Big Data framework. *Future Generation Computer Systems*, *90*, 262-272. https://doi.org/10.1016/j.future.2018.08.003

Gladić, D., & Petrovački, J. (2021, March). Using a data warehouse system to monitor and analyze student achievement in teaching process: student paper. In *2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-6). IEEE. https://doi.org/10.1109/INFOTEH51037.2021.9400685

Golfarelli, M., Maio, D., & Rizzi, S. (1998, January). Conceptual design of data warehouses from E/R schemes. In *Proceedings of the Thirty-First Hawaii International Conference on System Sciences* (pp. 334-343). IEEE.

Goyal, M., & Vohra, R. (2012). Applications of data mining in higher education. *International Journal of Computer Science Issues (IJCSI)*, *9*(2), 113.

Janati, S. E., Maach, A., & El Ghanami, D. (2019). Learning analytics framework for adaptive E-learning system to monitor the learner's activities. *International Journal of Advanced Computer Science and Applications*, *10*(8).

Kimball, R., & Ross, M. (2010). *The Kimball group reader: relentlessly practical tools for data warehousing and business intelligence*. John Wiley & Sons.

Kurniawan, Y., & Halim, E. (2013). Use data warehouse and data mining to predict student academic performance in schools: A case study (perspective application and benefits). *In Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 98-103). IEEE. https://doi.org/10.1109/TALE.2013.6654408

Lane, P., & Schupmann, V. (2007). *Oracle Database Data Warehousing Guide*. 11g Release 2 (11.2) E16579-01.

Li, Q., Duffy, P., & Zhang, Z. (2022). A novel multi-dimensional analysis approach to teaching and learning analytics in higher education. *Systems*, *10*(4), 96. https://doi.org/10.3390/systems10040096

Li, Q., Li, Z., Han, J., & Ma, H. (2022). Quality assurance for performing arts education: A multi-dimensional analysis approach. *Applied Sciences*, *12*(10), 4813. https://doi.org/10.3390/app12104813

Nebić, Z., & Mahnič, V. (2010). Data warehouse for an e-learning platform. *Latest Trends on Computers*, *2*, 415-420.

Seker, S. E. (2015). Büyük veri ve büyük veri yaşam döngüleri. *Ybs Ansiklopedi*, *2*(3), 10-17.

Simitsis, A., Vassiliadis, P., & Sellis, T. (2005). Optimizing ETL processes in data warehouses. In *21st International Conference on Data Engineering (ICDE'05)* (pp. 564-575). IEEE. https://doi.org/10.1109/ICDE.2005.103

Solodovnikova, D., & Niedrite, L. (2005). Using data warehouse resources for assessment of e-learning influence on university processes. *In: Proc. of the 9th East-European Conf. on Advances in Databases and Information Systems (ADBIS)*, (pp. 233. – 248), Tallinn.

Zhang, Y., Oussena, S., Clark, T., & Kim, H. (2010). Use data mining to improve student retention in higher education. *In Proceeding of the 12th International Conference on Enterprise Information System (ICEIS 2010)*, (pp. 8-12), Funchal.

Zorrilla, M.E. (2009). Data warehouse technology for E-learning. In: D., Zakrzewska, E., Menasalvas, & L., Byczkowska-Lipinska (Eds.), *Methods and Supporting Technologies for Data Analysis* (ss. 1-20). Springer. https://doi.org/10.1007/978-3-642-02196-1_1