IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 26, 2018, accepted October 1, 2018, date of publication October 22, 2018, date of current version November 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2876753

A Holistic View of Data Warehousing in Education

OSWALDO MOSCOSO-ZEA[1], JOEL PAREDES-GUALTOR[1], AND SERGIO LUJÁN-MORA[2]
[1]Facultad de Ciencias de la Ingeniería e Industrias, Universidad Tecnológica Equinoccial, Quito 170509, Ecuador
[2]Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, 03690 Alicante, Spain

Corresponding author: Oswaldo Moscoso-Zea (omoscoso@ute.edu.ec)

This work was supported by Universidad Tecnológica Equinoccial.

ABSTRACT Data warehousing (DW) is a widespread and essential practice in business organizations that support the data analytic and decision-making process. Despite the importance of DW in complex organizations, the adoption of a data warehouse (DWH) in education is apparently lower compared with other industries. To clarify this situation, this paper presents a systematic mapping that includes the study of empirical research papers from 2008 to 2018 on the topic of DW in education. For this paper, we applied a qualitative and quantitative approach based on a four-stage research method with the objective to have a holistic view of DWHs in education. After filtering and applying the proposed method, 34 relevant papers were identified and studied in detail. The study revealed interesting facts; for example, Kimball's approach is the most applied methodology for DWH design in education. In addition, a mapping between this comprehensive collection of research papers covering educational DW and six dimensions of analysis (schema proposal, analysis of the user requirements, analysis of the business requirements, effectiveness, implementation, and data analysis) was performed. From this analysis, we discovered that the star schema is the most implemented approach. The purpose of the mapping was to explore and identify the priority areas of research and the research gaps within the academic community. These gaps are a source of opportunities to start new lines of research.

INDEX TERMS Business intelligence, data warehouse, educational data warehouse, systematic mapping.

I. INTRODUCTION

Data warehousing (DW) is the process of storing, managing and analyzing large amounts of historical, summarized and non-volatile data. These data are extracted from multiple heterogeneous data sources into a single multi-dimensional repository called data warehouse (DWH). The core objective of DW is to provide greater insights into the performance of an organization and improve decision-making [1]. The complementary fields that study the analysis of the data in this repository are data analytics (DA) and on-line analytical processing (OLAP). On the one hand, DA is the process of analyzing the data in the DWH using technological and statistical tools with the purpose to draw conclusions and generate knowledge from the information it contains [2]. On the other hand, OLAP is the process of exploiting the DWH for multidimensional analysis by applying data cube operations as roll-up, drill-down, slicing and dicing on the dimensions and fact tables [3]. DA and OLAP, plus additional technological tools, are part of what data scientists call business intelligence (BI).

Even though DW and BI are widely used in business organizations and have been exhaustively analyzed from the industry standpoint for many years [4]–[7], its use is still low in educational institutions. The paper from Shahid *et al.* [8] presents a set of case studies performed to determine the percentage of use of DWHs within different industries. The results of the case study analysis show that the industry which uses DW the most is the medical industry (hospitals, clinics and physician offices) with 23.3% of usage. Following that, the finance and banking industry with 6.2% of DW usage, whereas one of the industries that less use DW is education with only 3.8% of usage [8]. Although these facts reveal that not much effort has been conducted to overcome the barriers of adoption of DW in education, nowadays boards of governance and directors in educational institutions are recognizing the potential and the leading role that BI and DW plays in improving educational and organizational processes [9]–[11]. Additionally, some studies suggest to carry out an implementation of a DWH in educational scenarios to improve decision-making and knowledge management [12], [13].

Moreover, different studies are giving major importance to the educational data mining (EDM) topic which bases its analytical process in a well-designed DWH to store and maintain the information of their application systems [14], [15].

Therefore, the need of DWHs in education is clear, but it is not clear the reason for a low adoption. One of the objectives of our paper is to provide some insight into this situation. Consequently, the scope of the adoption process studied in this paper primary focuses in institutions that have implemented a DWH for academic purposes.

According to Markets and Markets, the worldwide BI market is forecast to reach $26.88 billion in 2021, an increase of 63.58% from 2016 [16]. The scope of this market includes all BI and DA platforms, management suites and modern analytics solutions [17]. One of the key applications of BI and DA is the generation of knowledge to enable informed decision-making [18]. Knowledge is a key resource which can strengthen the positioning of an organization [19]. Knowledge can be defined as experience, facts, know-how, processes and beliefs that increase an organizational or individual's capability [20]. The creation of knowledge involves capturing information from people, processes and technology and establishing mechanisms for data analysis. Capturing all dimensions of knowledge on a technological platform is a huge challenge. One of the key technologies that can support the knowledge capturing process is a DWH.

BI and DA have evolved in the educational world mainly through EDM [21], [22] and learning analytics [23], primarily to evaluate the teaching-learning process (when, what, how and how much the students learn). Nevertheless, there are other processes and units inside the educational institutions which can be benefited as well, for example, admission, research, finances, planning, and student welfare. The DWH is in the core of the analytics process and is the instrument that assures that the information extracted, transformed and loaded meets the quality standards that are necessary to have a quality output in the generation of organizational knowledge [24].

With the adoption of the internet of things and cloud computing, the BI and the DA tasks are currently challenging in educational institutions, mainly due to a large amount of information generated every day by heterogeneous administrative and academic applications and devices. For this reason, the complexity to store, process and analyze data in these institutions increases. An optimal solution to tackle this complexity is summarized in the following steps: the implementation of a DWH; the definition of an extraction, transformation, and loading (ETL) process to populate the DWH; and the setting up of BI tools to generate knowledge from the DWH.

A DWH is a multidimensional data repository used for data analysis [5]. The DWH conceptual design consists of a set of dimension tables, fact tables, and their relations. The populated DWH schema can be analyzed using BI and data mining tools to discover knowledge. BI is defined as ''a system comprised of both technical and organizational elements that present historical information to its users for analysis and enables effective decision-making and management support, for the overall purpose of increasing organizational performance'' [25, p. 161]. BI in educational institutions builds upon a set of tools and applications that enables intensive calculation in millions of rows of data and the analysis of vast amounts of a variety of academic information (Big Data) to improve decision-making and performance [26]. Despite the fact of the importance of setting up a DWH in business and financial industries, much more attention has been paid lately in educational research to the study of BI, DA, and EDM over DW. Data analysis that does not meet quality standards will not produce quality reports and therefore, explicit knowledge. Managers must realize that in this knowledge era, information flow comes from a myriad of devices and social networks platforms. Here lies the importance of setting up a novel data infrastructure in which the DWH is still one of the core technologies for the success of the implementation of BI and Big Data projects in educational institutions [27].

Educational institutions that implement these suggested technologies will improve information management and decision-making and become smarter universities [28]. Data infrastructure in universities need to be reexamined due to three reasons: first, board of directors' demand improved data management strategies to support strategic planning; second, information available and updated improves administrators in recruitment and retention of students; and third, accrediting institutions demand information about the performance of higher education institutions [29]. Moreover, the big extent of information from social networks loaded into a central DWH can give university administrators insights to improve decision-making.

There are different methodologies proposed by researchers when designing a DWH, but there are two that stand out (Bill Inmon's and Ralph Kimball's). Practitioners and researchers can choose which one better suits a particular scenario based on the requirements of the project. Inmon [4] proposes a top-down approach (see Figure 1) whereas Kimball proposes a bottom-up approach [5] (see Figure 2). On the one hand, Inmon's methodology provides a very consistent dimensional view of data across data marts, as all data marts are loaded from the central DWH. The top-down approach is flexible to support change management as it looks at the organization as a whole [30]. On the other hand, Kimball's methodology is recommended when the time and budget are short [12]; this is because data marts are developed gradually.

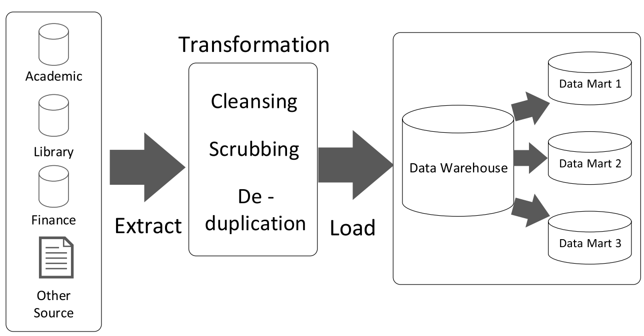


FIGURE 1. Bill Inmon's top-down approach to DWH design.

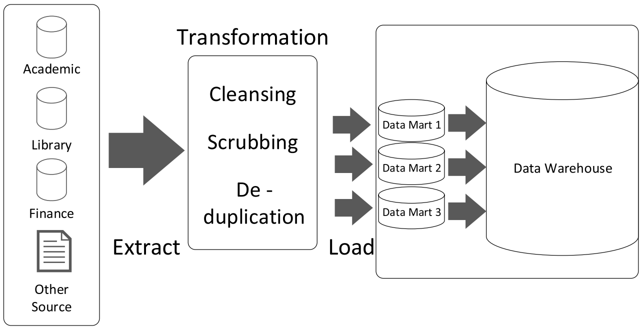


FIGURE 2. **Ralph Kimball's bottom-up approach to DWH design.**

The implementation of data marts can be carried out according to the strategic priorities obtained from an information need analysis. The approach to be used when building a DWH for an education institution greatly depends on the business objectives, business model, budget, and the level of dependencies between different business units. Figure 1 and Figure 2 illustrate a summary of the two approaches. In Figure 1 it is shown the ETL process of the Inmon's top-down approach, in which data are extracted from internal and external data sources. These data are then transformed by performing cleaning, scrubbing and deduplication activities, among others. Finally, the cleaned data are loaded into a central DWH where it is later fragmented into different data marts of the diverse departments or business units.

By comparison, in Figure 2, it is shown the ETL process of Kimball's bottom-up approach. The extraction and transformation are the same as Inmon's approach, whereas the loading process differs since cleaned data is loaded directly into data marts and later loaded into a central DWH. A case study performed in a university [12] recommends the use of Kimball's DWH design methodology for educational environments. This recommendation is based on the fact that in most universities the departments and units are not integrated and operate as information silos. One of the objectives of this paper is to clarify which is the most implemented methodology in the catalog of studied papers.

As seen in the methodologies presented in this section, the development of a DWH is a challenging task. We have reviewed that there are published research papers in the last years on DW in high impact journals in different areas, but we still found a lack of research on DW on education. For example, all of the following papers are published in high impact journals: 1) "Medical Big Data Warehouse: Architecture and System Design, a Case Study: Improving Healthcare Resources Distribution" was published in 2018 in the Journal of Medical Systems [31]; 2) "Atrak: a MapReduce-based data warehouse for big data" was published in 2017 in the Journal of Super Computing [32]; and, 3) "Augmenting Data Warehouses with Big Data" was published in 2015 in the Journal of Information System Management [33]. These examples illustrate that DW still arouses interest in the scientific community.

In this paper, we present the results of a systematic mapping (SM) on the topic of DW in education institutions from 2008 to 2018. The goal of this study is to collect and analyze a set of relevant research papers that study DWH in an educational scenario. With this collection, we can catalog available evidence and discover research gaps. To discover these gaps and to present a research roadmap for future initiatives, six dimensions are analyzed in detail: Schema Proposal, User Requirements, Business Requirements, Effectiveness, Implementation, and Data Analysis. These analyses have introduced us to greater insights on the DWH topic, for instance, we have noticed that even though the traditional DWH technology for multidimensional relational databases is still being used, many institutions already have upgraded their infrastructures to a hybrid model for Big Data. For example, in the paper from Santoso, an implementation and the characteristics of a modern DWH are presented [34].

Furthermore, this SM intends to discover the areas of highest and least interest for researchers and academics in the topic DWH in education. The rest of the paper is organized as follows: Section 2 describes the method used for the search and mapping of the papers; Section 3 presents the results of the SM on the DWH in education topic; finally, Section 4 provides conclusions and discussion of the work.

II. METHOD

The research method adopted in this study is SM. An SM aims to present and categorize the scientific publications published within a time frame and those which are available on a specific topic or research trend. The SM provides a text or graphic summary of the results which is called "the map". It requires less effort to build than a systematic literature review (SLR), with the advantage that readers can have a general vision of the examined topic. Additionally, researchers need less effort to drive an investigation [35]. The main differences between an SLR and an SM are presented in a paper by Kitchenham *et al.* [36] and are summarized in Table 1. As seen in Table 1, an SM is a more general study defined by a topic area. An SM is not designed to answer a specific research question, whereas, in an SLR, researchers' search process is extremely stringent to discover quality evidence that allows authors to answer the research question.

Furthermore, in this work, we have applied the snowballing technique which consists of the scanning of the reference list from the selected papers to decide and judge if there are valuable papers that can be further reviewed [37]. With the incorporation of this technique, relevant papers are added to the documentary corpus that otherwise would not be found with the keywords from the original primary search.

There are different methods and methodologies presented in the literature for performing an SM. Most of the existing studies, methods, and methodologies for SMs are found in medical research [35]. In this study, we have proposed a

TABLE 1. Differences between systematic mappings (SM) and systematic literature reviews (SLR). Source: [36].

SLR Process	SM	SLR
Research question	General – related to research trends. Which researchers, how much activity, what type of studies etc.	Specific - related to outcomes of empirical studies. Of the form: Is technology/method A better or not than B?
Search process	Defined by topic area	Defined by the research question
Search strategy requirements	Less stringent if only research trends are of interest	Extremely stringent – all relevant studies must be found
Quality evaluation	Not essential	Important to ensure that results are based on the best quality evidence
Results	Set of papers related to a topic area and counts of the number of papers in various categories	The answer to the specific research question, possible with qualifiers (e.g. results apply to novices only)

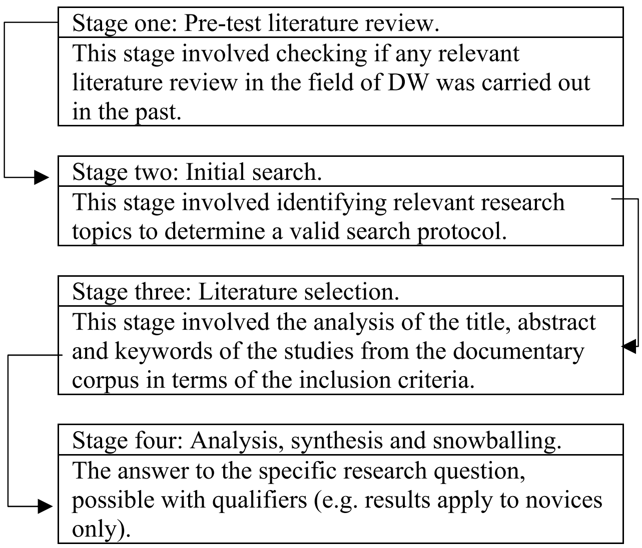


FIGURE 3. Four-stage method for systematic mapping.

four-stage method for the SM based on the best practices suggested in the literature, as shown in Figure 3.

The proposed method incorporates the results part from the SLR method shown in Table 1. We consider that the count of the number of papers in various categories is not sufficient to unveil a research outlook. Therefore, we have incorporated in the proposed four-stage method the answer to specific research questions, as shown in the results part of the SLR displayed in Table 1.

A. STAGE ONE: PRE-TEST LITERATURE REVIEW

This stage involved checking if any relevant literature review in the field of DW was carried out in the past. The results of this analysis will help researchers to assess the viability of the investigation. To do so, a general keyword search was performed. The bibliographic databases used for the search were Scopus (SCO), Web of Science (WOS), Education Resources Information Center (ERIC), Directory of Open Access Journals (DOAJ) and Google Scholar (GS). SCO and WOS were selected because they are considered the largest and the most widespread databases used for searching the scientific literature [38]. Both databases present a complete coverage of high quality and high impact journals from different disciplines, including science education and educational technology.

ERIC, DOAJ, and GS were used in this stage for its completeness and a large number of accessible open papers. The search string used is shown in Table 2. This initial search was performed without including the word ''education'' or ''educational'' in the search string with the purpose to discover all the research performed in the DWH field in all industries.

TABLE 2. Pretest search string.

Search String
title: ((“data warehouse” OR “datawarehouse”) AND (“literature review”))

B. STAGE TWO: INITIAL SEARCH

This stage involved identifying relevant research topics (RT) to determine a valid search protocol. This protocol includes the keywords used to build the search string and the bibliographic databases used for performing the search. The research databases considered in this study were again: SCO, WOS, ERIC, DOAJ, and GS.

Only studies of up to eleven years (since 2008) were chosen for the preliminary documentary corpus. This period was selected due to the fact that the International Working Group on EDM established an annual International Conference on EDM in 2008 [21]. Since that conference, an increase in research in EDM and DWH in education has been detected.

C. STAGE THREE: LITERATURE SELECTION

This stage involved the analysis of the title, abstract and keywords of the studies from the documentary corpus regarding the inclusion criteria: only peer-reviewed conference papers and journal articles written in English. Studies with unrelated research topics were rejected. Studies that talk about data infrastructures or DWHs implemented in higher education institutions were included for further analysis.

D. STAGE FOUR: ANALYSIS, SYNTHESIS, AND SNOWBALLING

This stage involved reviewing the full content of the selected studies from stage three. Furthermore, in this stage, the contents of each study regarding the inclusion criterion were analyzed. The inclusion criterion comprises only studies related to the educational or academic field and long or short papers with more than five pages. Consequently, this phase supports authors to address the research issues. In this phase, we also applied the snowballing technique [37] to discover if any additional papers could be added to the final documentary corpus. This technique can also be applied to detect new keywords for improving the search string. We will incorporate the new keywords found with the snowballing technique in future research.

III. SYSTEMATIC MAPPING RESULTS

The results of the SM analysis are presented in this section. Each stage from the method proposed in the previous section is described in detail. This approach (the four-stage method) allowed us to fully understand the current state of DWH in education and the research that is being carried out in this technological field.

A. PRE-TEST LITERATURE REVIEW

A set of 10 literature reviews was found and extracted from bibliographic databases. This search was performed with the search string presented in Table 2. The period of search corresponds to the time frame of January 2008 – June 2018. Table 3 shows the number of studies found per year. We found out that these previous studies neither performed an exhaustive review nor defined a search protocol that includes all major bibliographic databases to perform the searches. For example, Jourdan *et al.* [39] worked over a group of papers from 10 specific journals, not including conference or workshop papers. Wibowo [40] did not establish any research protocol and Choudhary *et al.* [41] selected the papers arbitrarily.

TABLE 3. Systematic mappings on data warehousing per year.

Year	2008	2009	2011	2015	2016	2017	2018
Literature Reviews	1	1	1	2	4	1	0

The analysis of the papers from Table 3 revealed that none of the literature reviews were related to the field of educational or academic DW. The fields covered by these studies were: social media, quality of DW, Internet of Things, the ETL process, the decision-making process, data analysis, manufacturing, and the World Wide Web.

From these results, we conclude that there is a research gap in the topic DWH in education. There are not exhaustive and methodical literature reviews performed in the past. Therefore, we recommend using our SM as the basis for further research on this area.

B. INITIAL SEARCH

After observing in the previous stage that there were no previous literature reviews on the topic of DWH in education, the next task was to perform a mapping to explore papers of interest individually. At this stage, we defined two research topics for the mapping. Table 4 details these topics. The goal of the first research topic is to discover details about the papers related to DWH in education (titles, authors, affiliations, and countries). The goal of the second research topic is to unveil the dimensions of analysis studied by the different authors of the selected papers.

The next step was to perform several searches in the bibliographic databases previously defined applying the search string shown in Table 5. The search string was created to limit the search scope to papers on DWH in education and to exclude other studies related to education as EDM or learning analytics.

We fetched 101 papers distributed from January 2008 to June 2018. Table 6 details the results per year and bibliographic database. At this stage, we did not eliminate duplicate papers between sources, this was done at a later stage.

From the results shown in Table 6, it is interesting to point out the following observations:

- At least three papers were published each year from 2008 to 2018.
- A similar number of papers was found in SCO, WOS, and GS. These results show the importance that GS is acquiring as an alternative to SCO and WOS.
- The year with most studies was 2016 with 19 papers.
- The years with least studies were 2013 with 3 papers and 2010 with 5 papers.
- The search performed in ERIC and DOAJ yielded only 3 papers of the research topic. Therefore, we do not recommend to investigate the DW topic only in this databases, because the results are not comprehensive.

C. LITERATURE SELECTION

In stage three, we filtered 61 papers from the 101 previously selected studies, based on the inclusion criteria. There were several reasons to discard the studies. The most important reason was the fact that the discarded studies did not propose the use of an educational data repository or a DWH.

Furthermore, it is interesting to notice that a big quota of the documentary corpus on this review proposed DWHs focused on medical evidence to support doctors and health staff on how to use historical data to make decisions in the future. For example, the paper entitled "A data warehouse system to help assist breast cancer screening in diagnosis, education, and research" has the main three key terms in its title (data, warehouse, and education); however, the terms breast, cancer, and diagnosis led us to know that the paper focused on medicine and therefore, it was discarded [42].

Table 7 shows the total number of selected studies distributed per year and source. Again, duplicates between sources were not eliminated at this stage. The results of this classification show that researchers have presented studies on the DWH topic for all years. In addition, the year with the highest number of studies was 2016 with 18 studies.

D. ANALYSIS, SYNTHESIS, AND SNOWBALLING

Among the 61 remaining studies, several duplicates between bibliographic databases were found. Hence, we eliminated 33 studies, leaving us with 28 studies. At this point the snowballing technique was applied. The combination of keywords

TABLE 4. Research topics and motivations.

Research Topic (RT)	Motivation
RT1. Number of significant studies with the words "data warehouse" and "education" or "academic". What are their titles, authors, institutional affiliations, and countries? Which are the journals and conferences that have published them?	These questions provide a starting point for new researchers in the field of data warehousing in education using factual information.
RT2. Research dimensions that have been tackled in the relevant studies. What are the most researched dimensions? What are the least researched dimensions?	These questions let researchers identify unbalanced research efforts and the research needs of the field.

TABLE 5. Search string.

Search String
title: (("data warehouse" OR "datawarehouse") AND ("education" or "educational" or "academic"))

TABLE 6. Studies fetched from the initial search.

Source	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	All
SCO	2	5	2	2	2	1	3	3	6	4	2	32
WOS	3	5	1	3	3	1	3	5	6	0	0	30
ERIC	1	0	0	0	0	0	0	0	0	0	0	1
DOAJ	0	0	0	1	0	0	0	0	1	0	0	2
GS	3	3	2	4	2	1	3	4	6	4	4	36
Total	9	13	5	10	7	3	9	12	19	8	6	101

TABLE 7. Selected studies.

Source	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	All
SCO	0	0	1	2	0	1	2	2	6	2	2	18
WOS	2	0	0	2	0	1	2	1	5	0	0	13
ERIC	1	0	0	0	0	0	0	0	0	0	0	1
DOAJ	0	0	0	1	0	0	0	0	1	0	0	2
GS	3	2	0	4	1	1	2	2	6	2	4	27
Total	6	2	1	9	1	3	6	5	18	4	6	61

used in the snowballing search was: data warehousing, data warehouse, education, educational, academic, heterogeneous databases, university. We reviewed the list of references for each of the remaining 28 papers and found out 31 additional papers related to the research topic. These papers did not appear in the original search because they did not have the search terms in the title. From this list, 22 of the 31 papers were discarded. We discard these papers due to the following reasons: papers were published out of the time frame (written before the year 2008), papers were no completely relevant to the research topic (some papers described BI in education, but they did not describe the implementation of a DWH in an educational institution), some were poster papers (less or equal than 5 pages) [43], some were not written in English, or the papers were previously identified, leaving us with nine papers.

With the nine remaining papers, we investigated if they were published in the research databases used during the initial search: SCO, WOS, ERIC, DOAJ, and GS. With this last filter, we discarded three more papers. Finally, six additional papers were added to the final documentary corpus. Therefore, the final catalog was established in 34 papers. Table 8 shows the final documentary corpus distributed by year for further analysis including the snowballing phase.

An example of a paper that was added using the snowballing technique is the paper from Aziz and Wahid [44] ''Integration of Heterogeneous Databases in Academic Environment Using Open Source ETL Tools'', this paper was added after reviewing the list of references from Abdul Aziz *et al.* [45]. Despite the paper describes the DW process in education, the title does not have the search terms.

The purpose of the first research topic of this study was to aggregate and explore evidence of the implementation and design of a DWH in the educational field. With the results of this first classification, we wanted to identify the journals and conferences that are focused on this line of investigation. Additionally, we wanted to explore the communities, institutions and the countries which are leading the studies on this topic. Therefore, the first research topic in the SM study is:

TABLE 8. Relevant studies per year (final documentary corpus).

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	All
Number of studies	2	2	0	4	2	1	3	5	8	3	4	34

RT1. Number of studies with the words "data warehouse" and "education" or "academic". What are their titles, authors, institutional affiliations, and countries? Which are the journals and the conferences that have published them?

Table 9 and Table 10 show the results of the search for RT1. Table 9 shows the selected studies classified by the authors' names, the source of the paper (journal or conference), the year of publication and the number of citations in SCO, WOS, and GS. The table is ordered from the most recent year to the oldest year of publication.

Table 9 shows that Miranda and Miranda [46] and Miranda and Suryani [47] presented two studies on the research topic, the rest of the authors only presented one paper. Similarly, the Journal of Theoretical and Applied Information Technology was the only journal from the documentary corpus which appeared twice on the list of papers [45], [47]. The most cited paper from the list was "Use of data warehouse and data mining to predict student academic performance in schools: A case study (perspective, application, and benefits)" from Kurniawan and Halim [48] with a total of 26 cites (9 citations in SCO, 2 citations in WOS and 15 citations in GS). The next most cited paper was "E-Governance in Higher Education: Concept and Role of Data Warehousing Techniques" from Bhanti *et al.* [49] with 19 citations, all of them found in GS.

In Table 10 we have analyzed the affiliation of the first author of the paper and the country of origin. It is important to mention the variety of regions where the research subject "DWH in education" has been studied. The country with the most studies is Indonesia with six papers [34], [46]–[48], [50], [51]. From this group of papers, three were written in Bina Nusantara University [46]–[48]. The next two countries with the most published papers were India [13], [49], [52], [53] and Malaysia [44], [45], [54], [55] with four studies each. These results bring to light a surprising interest in DW in education in South Asia.

After performing the analysis, we could not identify a wide research community that is working on the topic "DWH in education". In any case, the study of the research topic has not declined over the years and we expect that with the resurgence of DA, EDM, and the potential that Big Data brings to BI, DW will maintain its importance over the next years in the handling and the analysis of academic information. After presenting the findings of the first research topic, DWH in education was further explored from multiple aspects defined as research dimensions. The first studied dimension in this phase of the investigation was "Schema Proposal". The goal was to synthesize the most used schemas in the design of DWHs. Following that, the next goals of the research were to validate if the papers considered the user and business requirements, ensured operation effectiveness, proposed an implementation of the DWH or performed a data analysis. Therefore, the second research topic to be answered in this SM study is:

RT2. Research dimensions that have been tackled in the relevant studies. What are the most researched dimensions? What are the least researched dimensions?

The research dimensions studied in the final documentary corpus were diverse. Table 11 shows the most important dimensions identified in the papers. For the characterization of this table, a literature review matrix was created. A literature review matrix is a graphic organizer used to understand and compare the author's ideas [56] The top of the matrix shows the dimensions of analysis. These dimensions are matched versus the aspects studied in each one of the selected papers. Initially, 13 dimensions were identified while reviewing the papers. Later, they were reduced to 6 dimensions so that there would be a significant number of articles per dimension (minimum of 4 papers per dimension). We choose the threshold of four after performing the analysis with the 13 dimensions and finding out that most of those dimensions were not studied in the papers. The omitted dimensions that did not meet the minimum of four studies were marketing, academic performance, scientific production, educational assessment, ETL automation, model validation and readiness analysis.

The most discussed dimension in the papers was Implementation studied in 27 papers. The next most studied dimension was Schema Proposal studied in 26 papers. The third most studied dimension was Data Analysis, proposed in 19 papers, the Effectiveness dimension was studied in 11 papers and the Business Requirements dimension in 8 papers. The least studied dimension was User Requirements which were studied in 4 papers. These dimensions are explained in detail in the following subsections.

The schema types used for the DWH development were star, snowflake, constellation, and relational schema. Even though there were articles that did not specify the method used for the design of the DWH, we have analyzed the schema and implementation proposed to identify the design methodologies. Kimball's Bottom-Up approach was proposed in 15 papers [12], [13], [45]–[48], [53], [54], [57]–[63]. Whereas, Inmon's Top-Down approach was proposed in 11 papers [44], [49]–[51], [55], [65], [68], [70], [72], [73], [75]. The remaining paper proposed an approach to design a Big Data warehouse [34].

TABLE 9. Relevant studies on the topic of DWH in education (2018-2008).

ID	Authors	Source	Years	Citations (SCO/WOS/GS)
[54]	Salaki R., Ratnam K.	6th World Conference on Information Systems and Technologies, WorldCist	2018	(0/0/0)
[55]	Abu Bakar M., Suwannit Chareen A., Mohd Soid M.	e-Academia Journal, vol. 6(2)	2018	(0/0/0)
[53]	Khan A., Ghosh S., Ghosh S. K.	International Conference on Intelligent Systems Design and Applications, ISDA	2018	(0/0/0)
[64]	Pasyeka M.	Software Engineering, vol. 6(1)	2018	(0/0/0)
[34]	Santoso L., Yulia.	4th Information Systems International Conference, ISICO	2017	(2/0/1)
[50]	Budiarta K., Ananta P., Indrapartha C.	International Journal of Engineering and Emerging Technology, vol. 2(1)	2017	(0/0/0)
[65]	Cheowsuwan T., Rojanavasu P., Srisungsittisunti B., Yeewiyom S.	International Journal of Geoinformatics, vol. 13(2)	2017	(0/0/0)
[57]	Song Y., Pramudianto F., Gehringer E.F.	46th Conference Frontiers in Education, FIE	2016	(1/0/3)
[66]	Debus K., Kickmeier-Rust M., Albert D.	24th International Conference on Computers in Education: Think Global Act Local - Workshop Proceedings, ICCE	2016	(0/0/0)
[12]	Moscoso-Zea O., Sampedro A., Luján-Mora S.	15th International Conference on Information Technology Based Higher Education and Training, ITHET	2016	(3/0/2)
[58]	Aljawarneh I.	Journal of Information Development, vol. 32(5)	2016	(0/0/1)
[51]	Iskandar A., Taufik I.	4th International Conference on Cyber and IT Service Management, CITSM	2016	(0/0/0)
[67]	Bondarev A., Zakirov D.	12th International Conference on Electronics Computer and Computation, ICECCO	2016	(1/0/4)
[59]	Abdullah Z., Obaid A. S.	International Journal of Computer Science and Network, vol. 5(5)	2016	(0/0/0)
[52]	Panchal R.	We'Ken International Journal of Basic and Applied Sciences, vol. 1(4)	2016	(0/0/0)
[60]	Di Tria F., Lefons E., Tangorra F.	Journal Computer Science and Information Systems, vol. 12(1)	2015	(3/1/4)
[46]	Miranda R., Miranda E.	International Journal of Multimedia and Ubiquitous Engineering, vol. 10(4)	2015	(2/0/0)
[61]	Moturi C. Emurugat A.	British Journal of Applied Science & Technology, vol. 8(6)	2015	(0/0/2)
[68]	Gao L., Chen Y.	Journal of Computer Science and its Applications, vol. 330	2015	(0/0/2)
[62]	Mirabedini S.	Journal of Novel Applied Sciences, vol. 3(5)	2015	(0/0/2)
[47]	Miranda E., Suryani E.	Journal of Theoretical and Applied Information Technology, vol. 64(3)	2014	(0/0/3)
[69]	Ahmed N., Anwar M., Ameen A.	8th International Technology, Education, and Development Conference, INTED	2014	(0/0/0)
[45]	Abdul A. et al.	Journal of Theoretical and Applied Information Technology, vol. 69(1)	2014	(3/0/3)
[48]	Kurniawan Y., Halim E.	IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE	2013	(9/2/15)
[44]	Aziz A. et al.	International Conference on Informatics & Applications, ICIA	2012	(0/0/1)
[70]	Triola, M., Pusic, M.	Journal of Graduate Medical Education, vol. 4(1)	2012	(0/5/11)
[71]	Schaefer B., Tanrikulu E., Breiter A.	Procedia - Social and Behavioral Sciences, vol. 28	2011	(2/1/3)
[72]	Paunica M. et al.	Annales Universitatis Apulensis - Series Oeconomica, vol. 12(1)	2011	(1/0/12)
[49]	Bhanti P. et al.	International Journal of Computer Applications, vol. 18(1)	2011	(0/0/19)
[13]	Manjunath T. et al.	International Journal of Computer Science, vol. 8(2)	2011	(0/0/4)
[73]	Mekterović I., Brkić L., Baranović M.	WSEAS Transactions on Computers, vol. 8(10)	2009	(5/2/9)
[74]	Čeliković M., Aleksić S., Luković I.	The International Conference on Information Technology, ICIT	2009	(0/0/1)
[75]	Dimokas N. et al.	12TH PAN-HELLENIC Conference on informatics, PCI	2008	(4/3/8)
[63]	Van Dyk L.	Electronic Journal of e-Learning, vol. 6(3)	2008	(0/0/10)

1) DATA WAREHOUSE SCHEMA PROPOSAL FOR EDUCATIONAL INSTITUTIONS

This dimension describes if any educational DWH schema was proposed in the paper. Schema proposal is the second most studied dimension from the documentary corpus with 26 papers. Papers that studied this dimension presented the types of methods followed for the modeling of the multidimensional data schema. These implementations can serve as roadmaps for future practitioners or researchers in the field to support their DWH design decisions with best practices

TABLE 10. First Author and affiliation country from selected papers (2018-2008).

ID	First author affiliation	First author country
[54]	Asia Pacific University of Technology and Innovation	Malaysia
[55]	College of Arts and Sciences	Malaysia
[53]	Indian Institute of Technology Kharagpur	India
[64]	Ivano-Frankivsk National Technical University of Oil and Gas	Ukraine
[34]	Petra Christian University	Indonesia
[50]	Udayana University	Indonesia
[65]	University of Phayao	Thailand
[57]	North Carolina State University	USA
[66]	Graz University of Technology	Austria
[12]	Universidad Tecnológica Equinoccial	Ecuador
[58]	University of Business and Technology	Saudi Arabia
[51]	Universitas Widyatama	Indonesia
[67]	International Ataturk Alatoo University	Kyrgyzstan
[59]	University of Basra	Iraq
[52]	Institute of Business Management & Rural Development	India
[60]	Università degli Studi di Bari Aldo Moro	Italy
[46]	Bina Nusantara University	Indonesia
[61]	University of Nairobi	Kenya
[68]	Shanghai University of Finance and Economics	China
[62]	Payame Noor University	Iran
[47]	Bina Nusantara University	Indonesia
[69]	Al Ghurair University	United Arab Emirates
[45]	Sultan Zainal Abidin University	Malaysia
[48]	Bina Nusantara University	Indonesia
[44]	Universiti Sultan Zainal Abidin	Malaysia
[70]	New York University	USA
[71]	University of Bremen	Germany
[72]	Academy of Economic Studies	Romania
[49]	MITS Deemed University	India
[13]	Bharathiar University	India
[73]	University of Zagreb	Croatia
[74]	University of Novi Sad	Serbia
[75]	Aristotle University of Thessaloniki	Greece
[63]	University of Stellenbosch	South Africa

from the case studies. The schemas proposed in the papers are of four types: star schema (Figure 4.a), snowflake schema (Figure 4.b), constellation schema (Figure 4.c) [76] and a relational schema (Figure 4.d).

The most implemented schema was the star schema with 18 implementations [12], [13], [44], [46]–[50], [55], [57]–[59], [61], [62], [65], [69], [73], [75]. The dimensions in the star schema are denormalized, therefore, querying is faster, accurate and consistent since use fewer joins. However, this schema requires more storage capacity due to redundancy. The next schema modeled in the papers is the constellation schema with four implementations [45], [60], [72], [74]. The constellation schema is a hybrid model composed of stars and snowflake schemas and has the advantage to provide a better scenario for sophisticated application analysis. The third schema modeled in the papers is the snowflake schema with three implementations [53], [66], [67]. The snowflake schema is a normalized

TABLE 11. Dimensions tackled in each study (2018-2008).

Id	Schema proposal	User requirements	Business requirements	Effectiveness	Implementation	Data analysis
[54]					X	
[55]	X		X		X	X
[53]	X				X	X
[64]	X					
[34]					X	X
[50]	X		X	X	X	
[65]	X		X		X	
[57]	X				X	
[66]	X					X
[12]	X			X	X	X
[58]	X		X	X	X	X
[51]					X	X
[67]	X					X
[59]	X				X	X
[52]		X	X			
[60]	X	X			X	X
[46]	X		X		X	X
[61]	X				X	X
[68]			X	X	X	
[62]	X				X	X
[47]	X				X	X
[69]	X			X		
[45]	X			X	X	X
[48]	X			X	X	
[44]	X				X	
[70]					X	
[71]		X		X		
[72]	X				X	X
[49]	X				X	
[13]	X		X	X	X	
[73]	X			X	X	X
[74]	X	X				
[75]	X				X	X
[63]				X	X	X
TOTAL	26	4	8	11	27	19

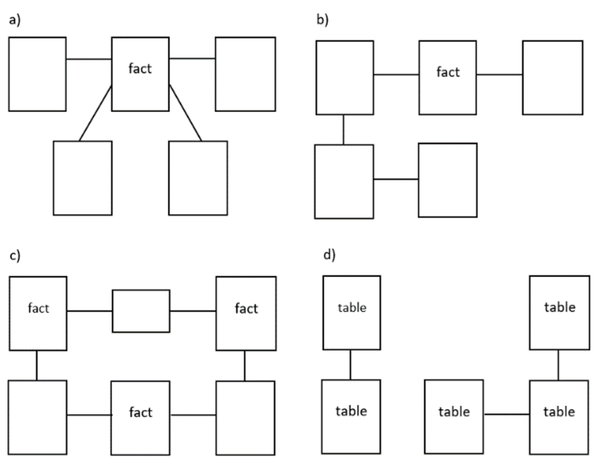


FIGURE 4. Proposed schemas. (a) Star schema. (b) Snowflake schema. (c) Constellation schema. (d) Relational schema.

schema which improves data storing and eliminates data redundancy but could produce a major latency in querying the database. There was also a paper that proposed a relational schema which did not present a multidimensional approach [64]. It is important to note that although in three papers [63], [68], [70] there was a general DWH implementation presented, there were no details of the schema proposed. Therefore, these studies were discarded from this dimension.

2) ANALYSIS OF THE USER REQUIREMENTS TO BUILD AN EDUCATIONAL DATA WAREHOUSE

One of the core phases of systems development is the analysis of information requirements. Likewise, a quality design and development of a DWH requires that all the stakeholders share their needs with the data designers. There are many issues derived from a poor analysis of information requirements. Some of these issues are the provision of irrelevant information, inexistence of data sources, or that requirements are not aligned with the corporate strategy of the organization. This dimension shows if the selected studies have focused on the user needs before building the DWH. Another fact studied in this dimension is if the papers take into account the correct capturing of information from the right data sources.

Analysis of users requirements is a critical step in building the DWH and only in four studies it is taken into account [52], [60], [71], [74]. The study from Panchal [52] describes the different challenges derived in the gathering of students and business process requirements. Di Tria *et al.* [60] elaborate on two approaches in the design of the DWH. The first design approach presented is data-oriented, in which the DWH is designed based on a reengineering process over the existing data sources with little involvement of the stakeholders. The second approach presented in the paper is requirement–oriented based on the user need without a technical view of the data sources.

Authors of the paper state that both approaches present deficiencies and therefore, they introduce and hybrid approach which depicts the design of multidimensional schemas based on user requirements and then, they validate if the data for the requirements is stored on the existing data sources. The paper from Schaefer *et al.* [71] presents a method for eliciting DWH requirements based on the key question "which users need which data for what purposes". Methods proposed in this paper for the information need analysis were document analysis, surveys, group task analysis, card sorting and interviews [71]. The paper from Čeliković describes a complex analysis of accomplished course requirements [74].

3) ANALYSIS OF THE BUSINESS REQUIREMENTS TO BUILD AN EDUCATIONAL DATA WAREHOUSE

This dimension expresses if the selected studies performed any business requirements analysis prior to build the educational DWH or if institutions aligned the design approach with the mission and vision of the institution. Some researchers call this analysis objective information needs analysis. The objective information needs analysis is derived from the strategic plans and the goals of the organization. Furthermore, this analysis requires a complete understanding of the business processes and the business model of the institution. Another fact analyzed in this dimension is the way educational institutions are facing different challenges regarding the myriad of requirements from students, teachers, processes and boards of managers. The business requirements have to be coherent to the transactional databases existing in the organization. Moreover the business requirements, the user requirements and the information supply must be aligned. Thus, the required DWH must be implemented according to the business goals of the organization [77]. There were eight papers which studied these dimension from the documentary corpus [13], [46], [50], [52], [55], [58], [65], [68].

4) EFFECTIVENESS OF THE EDUCATIONAL DATA WAREHOUSE

This dimension covers the post-implementation process and shows if the researchers have analyzed the effectiveness of their DWH proposal according to the requirements. This dimension is important to measure the effectiveness of the implementation process and studies the drawbacks of implementation to take corrective actions. This topic was studied in 11 papers [12], [13], [45], [48], [50], [58], [63], [68], [69] [71], [73] Some of the findings show that in an ideal scenario the ETL process should be automated. However, this is very hard to achieve due to information silos that are very common in universities due to a large number of non-integrated applications [12].

Additionally, a design framework is proposed by Aljawarneh [58]. This framework has improvements from traditional design methodologies and is validated in a case study performed in a university. The paper from Gao and Chen [68] introduces an application example to show the value of a decision support system based on a DWH. Other studies implement the DWH to improve the quality of the curriculum program [69], to design reports intended to analyze students' data [45] and to predict students' performance using EDM techniques [48], [71]. Other researchers show the effectiveness of the methods used for the design, for example by defining a method for eliciting user requirements [71].

Furthermore, some researchers study adoption barriers of a DWH. For example to overcome financial barriers researchers propose the use of open source tools [13] which they show are effective in a DW project. Another barrier is the lack of a maintenance process. To overcome this issue of maintenance, researchers propose a novel maintenance plan to have updated information for a sound data analysis [73]. Additionally other authors propose that information must be loaded to the DWH and tracked continuously [50], [63].

5) IMPLEMENTATION OF THE EDUCATIONAL DATA WAREHOUSE

Implementation is the most studied dimension with 27 papers from the documentary corpus. This dimension studies if the researchers have implemented a DHW in a real educational scenario. This dimension intends to analyze the phases that were followed for implementation and the outcomes of the DWH project. If the paper did not show the steps of implementation, then it was not considered. The analysis of this dimension will offer researchers a roadmap for future implementations. It is important to compare which was the methodology or approach used for the development of a DWH in an educational institution to identify best practices of implementation. Some of the shared phases used in these papers and that could define a common fivestep method for a DWH project implementation are:

1. Information needs analysis and requirements analysis [12], [46], [48], [50], [54], [55], [58], [60], [61].
2. Data source and data supply analysis [13], [49], [51], [55], [59], [72], [73].
3. DWH design and multidimensional modeling [44]–[48], [53], [54], [55], [57]–[60], [62], [63], [65], [68], [70], [75].
4. ETL process [12], [13], [44], [45], [53]–[55], [60], [62], [63], [68], [73].
5. System, application, reporting, dashboard and OLAP development [45], [47]–[50], [53]–[55], [61], [62], [65], [72], [75].

The criteria for selection of the common phases found in the 27 papers was that at least a quarter of papers propose the step in their implementation (25% of the total of documents that study implementation). Studies that propose data mining, data analytics, learning analytics or EDM were discarded. This decision was taken due to the fact that these are specialized topics which need their own method for implementation as knowledge discovery in databases (KDD) [78] or CRISP-DM [79].

There are other approaches proposed in the papers, however, they do not meet the criteria for inclusion. Some of these approaches are DWH processing [50], system testing [65], selection of technological infrastructure [12], schema validation [60], analysis of security [49] and statistical analysis [75].

6) DATA ANALYSIS AFTER THE ETL PROCESS

This dimension indicates if the authors have performed data analysis using the DWH. Some of the approaches to analysis and reporting are BI, DA, EDM or learning analytics. There were 19 papers [12], [34], [45]–[47], [51], [53], [55], [58]–[63], [66], [67], [72], [73], [75] that went beyond the DWH design process and presented how the information from the DWH was analyzed. The research papers examined in this dimension used their data to analyze assignments [66], grades [75] and the average marks [67] obtained by the students in their courses. Others analysis performed in the papers are: analysis of academic or organizational data for the accreditation process of the university [12], analysis with the goal to improve strategic planning [58] or to enhance change management [51], analysis of educational systems [59], analysis for the evaluation of quality of didactics and research [60], marketing analysis to increase admission rates [46], financial and revenue analysis [61], registration, evaluation, graduation, research and grant analysis [47], scholar performance [72], analysis of scholarship grants [73] , analysis of data generated from a learning management system [63]. Most of the papers describe the implementation of reports, dashboards and details of visualization.

IV. DISCUSSION AND CONCLUSIONS

This paper presents an SM carried out on the topic of DW in education. The review was performed between January 2008 and June 2018. For this review, a four-stage method was applied. This review supported the researcher team to answer the proposed research questions. To answer the questions formulated in the first research topic (details and number of significant studies with the words ''data warehouse'' and ''education'' or ''academic''), this paper presents the relevant papers found on the main scientific and research databases on the topic DWH in education. This first analysis includes the authors, the institutional affiliation, and their country of origin. Furthermore, this analysis lists the journals or conferences in which these papers were presented.

The second research topic was further analyzed (research dimensions that have been tackled in the relevant studies). Six dimensions of analysis were discussed in this paper. The most studied dimensions among researchers are in this order: Implementation, Schema Proposal, Data Analysis, Effectiveness, Business Requirements and User Requirements. Within this analysis we have identified best practices in DWH design and implementation and have suggested a fivestep method to develop a DWH that can be used in future implementation projects. We have also identified that user requirements and business requirements are not taken into account in many of the DWH implementation initiatives. Failures to obtain requirements might be an important cause of the failure of BI projects which cannot provide quality reports.

The summary of the percentages of papers addressed in the research dimensions' analysis showed that 27 of 34 (79.4 %) studies described the implementation phase of the DWH 26 of the 34 studies (76.5%) proposed a DWH schema (18 of them implemented star schemas). Additionally, 19 of 34 studies (55.9%) presented information about DA and 11 of 34 studies (32.3%) presented information about effectiveness. It is also important to notice that very few studies analyzed the user and business requirements before the DWH construction. Only 8 of 34 studies (23.5%) established the business requirements and 4 of 34 studies (11.8%) established the user requirements, as shown in Figure 5.

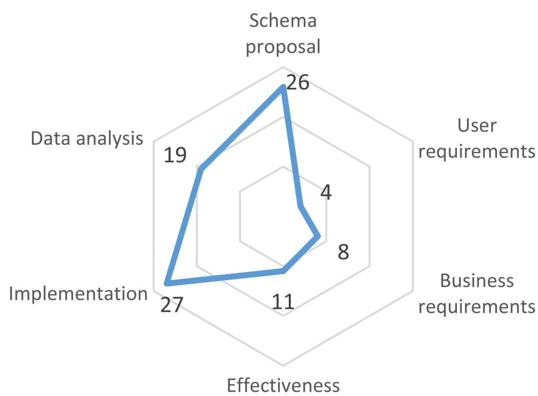


FIGURE 5. Number of studies per dimension of analysis.

The final results of the SM show that the ''DWH in education'' topic is being discussed and analyzed in different geographical regions with promising results. The validation and application of the DWH implementation ensure quality results in the visualization phase of the BI process. This paper shows that there are different universities which are implementing a DWH as a strategic initiative. The outputs from the DWH are used to improve the learning and teaching process. The knowledge created from the repository give directions to stakeholders to improve the academic processes.

Some of the main findings of this investigation show that although there is an interesting number of researchers and educational institutions working on DWH issues, there is still a lack of DWH implementations in institutions performing EDM and learning analytics. Many institutions experimenting with EDM or learning analytics do not perform an adequate ETL process to load the refined information into a DWH. The DWH in these cases is not the main repository of analysis. Information analysis is performed without a proper data preprocessing and transformation which could lead to an inconsistent analysis and erroneous decisions of managing boards. To complement this study and understand the reasons for not adopting a DWH, we propose as future work to survey a group of educational institutions. Moreover, we believe it is important to establish a framework and guidelines on how to implement a DWH that handles transactional data from operational data sources and unstructured data (Big Data) which comes from a myriad of devices. A sound framework could smooth the process in educational institutions with a clearer route map for implementation.

From all the literature reviewed it seems that nowadays much attention is paid to the mining of educational data and the analysis of students' learning, but not much attention is paid to the data storage infrastructure that allows having quality results in the BI and the data analysis' process. The barriers of adoption could be attributed to complexity in the implementation process, uncertain requirements, lack of a data maintenance plan, lack of guidance, lack of staff expertise, decentralized nature of data, resistance from data owners and financial costs [13], [52], [58]. Usually, a successful implementation of a DWH project in a university is measured by comparing the financial benefits versus the investment. This is hard to accomplish since most of these organizations are non-profit. Therefore, Panchal [52] recommends evaluating the success of a DWH implementation based on the ability to meet user and business requirements.

This paper could be used for researchers and practitioners as a guide to establish the best mechanism for implementing a DWH in an educational institution. On the one hand to support academic applications and learning systems. On the other hand, to support the introduction of Big Data. Thus, a mechanism should be thought for the extraction, loading, collection and integration of unstructured data. Unstructured data should be captured from the devices and social networks used by current and prospect students to improve decision making.

Despite the decrease of studies on the DWH topic in business organizations in the last years, the topic in education has revived due to the educational analytic revolution [80]. There is a large number of studies that consider educational data mining and learning analytics a panacea for academic improvements. These two trendy topics plus Big Data are major influencers in the reborn of DWH research in education. Researchers, academics and practitioners in education see the value that a sound implementation of a DWH could bring to the learning and teaching process. A sound method for DWH and ETL implementation is an enabler to obtain quality data as the input of the analytical process. The output of the analytical process is quality information and trustworthy knowledge which can be visualized with different BI tools to improve academic management.

Finally, in this paper we present additional arguments from the ones presented in the paper from Rizzi *et al.* [81] to affirm that DWH research is not dead. Many interesting approaches on DW are being implemented for example in the context of large-scale systems in biology or in the context of climate change. We suggest data scientists and researchers in educational institutions to seriously consider the implementation of freely available DWH systems in education to facilitate the analysis of academic data. We have also detected in this investigation that there are only a few quality studies on the DWH in education topic which opens great opportunities for further research in this field.

REFERENCES

[1] C. S. Jensen, M. Schneider, B. Seeger, and V. J. Tsotras, *Advances in Spatial and Temporal Databases*. Berlin, Germany: Springer-Verlag, 2001.

[2] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quart.*, vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[3] O. Romero and A. Abelló, "A survey of multidimensional modeling methodologies," *Int. J. Data Warehousing Mining*, vol. 5, no. 2, pp. 1–23, 2009.

[4] W. H. Inmon, *Building the Data Warehouse*, 4th ed. Indianapolis, IN, USA: Wiley, 2005.

[5] R. Kimball and M. Ross, *The Data Warehouse Toolkit*, 3rd ed. Indianapolis, IN, USA: Wiley, 2013.

[6] J. Lechtenbörger, G. Vossen, and S. Straße, "Conceptual data warehouse design," in *Proc. 2nd Int. Workshop Design Manage. Data Warehouses (DMDW)*, vol. 28, 2000, pp. 1–11.

[7] S. Luján-Mora, J. Trujillo, and S. Il-Yeol, "A UML profile for multidimensional modeling in data warehouses," *Data Knowl. Eng.*, vol. 59, no. 3, pp. 725–769, 2006.

[8] M. B. Shahid, U. Sheikh, B. Raza, M. A. Shah, A. Kamran, and A. Anjum, "Application of data warehouse in real life: State-of-the-art survey from user preferences' perspective," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 4, pp. 415–426, 2016.

[9] J. Bichsel. (2012). Analytics in higher education benefits, barriers, progress and recommendations. EDUCAUSE Center for Applied Research. Accessed: Jan. 9, 2018. [Online]. Available: https://library.educause.edu/ /media/files/library/2012/6/ers1207.pdf?la=en

[10] L. V. Morris, S.-S. Wu, and C. L. Finnegan, "Predicting retention in online general education courses," *Amer. J. Distance Educ.*, vol. 19, no. 1, pp. 23–36, 2005.

[11] P. Baepler and C. J. Murdoch, "Academic analytics and data mining in higher education," *Int. J. Scholarship Teach. Learn.*, vol. 4, no. 2, pp. 1–17, 2010.

[12] O. Moscoso-Zea, A. Sampedro, and S. Luján- Mora, "Datawarehouse design for educational data mining," in *Proc. 15th Inf. Technol. Higher Educ. Training (ITHET)*, Sep. 2016, pp. 1–6.

[13] T. N. Manjunath, R. S. Hegadi, I. M. Umesh, and G. K. Ravikumar, "Design and analysis of DWH and BI in education domain," *Int. J. Comput. Sci.*, vol. 8, no. 2, pp. 545–551, 2011.

[14] O. Moscoso-Zea and S. Luján-Mora, "Educational data mining: An holistic view," in *Proc. 11th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2016, pp. 1–6.

[15] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, pp. 135–146, Jul. 2007.

[16] Markets and Markets. (2017). *Business Intelligence Market Worth 26.88 Billion USD by 2021*. Accessed: Feb. 15, 2018. [Online]. Available: http://www.marketsandmarkets.com/PressReleases/social-business-intelligence-bi.asp

[17] Gartner Inc. (2016). *Gartner Says Worldwide Business Intelligence and Analytics Market to Reach 16.9 Billion in 2016*. Accessed: Feb. 12, 2018. [Online]. Available: http://www.gartner.com/newsroom/id/3198917

[18] J. Lloyd, "Identifying key components of business intelligence systems and their role in managerial decision making," Appl. Inf. Manage., Univ. Oregon, Eugene, Oregon, Tech. Rep. 11389, 2011.

[19] C. Curado, "The knowledge-based view of the firm and its theoretical precursor," *Int. J. Learn. Intellectual Capital*, vol. 3, no. 4, pp. 1–18, 2006.

[20] K. Karemente, J. R. Aduwo, E. Mugejjera, and J. Lubega, "Knowledge management frameworks: A review of conceptual foundations and a KMF for IT-based organizations," *Strengthening Role ICT Develop.*, vol. 5, pp. 35–57, Aug. 2009.

[21] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.
[22] R. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–17, 2009.
[23] R. A. Huebner, "A survey of educational data-mining research," *Res. Higher Educ. J.*, vol. 19, pp. 1–13, Apr. 2013.
[24] O. Moscoso-Zea and S. Lujan-Mora, "Knowledge management in higher education institutions for the generation of organizational knowledge," in *Proc. 12th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2017, pp. 1–7.
[25] O. Isik, M. Jones, and A. Sidorova, "Business intelligence (BI) success and the role of BI capabilities," *Intell. Syst. Accounting, Financial Manage.*, vol. 18, no. 4, pp. 161–176, 2011.
[26] Gartner. (2018). *Gartner Business Intelligence*. Accessed: Jul. 24, 2018. [Online]. Available: http://goo.gl/LmJRG3
[27] S. Luján-Mora and J. C. T. Mondéjar, "A comprehensive method for data warehouse design," in *Proc. 5th Int. Workshop Design Manage. Data Warehouses (DMDW)*, 2003, pp. 1–14.
[28] B. Williamson, "The hidden architecture of higher education: Building a big data infrastructure for the 'smarter University,'" *Int. J. Educ. Technol. Higher Educ.*, vol. 15, no. 1, pp. 1–26, 2018.
[29] J. Guan, W. Nunez, and J. F. Welsh, "Institutional strategy and information support: The role of data warehousing in higher education," *Campus-Wide Inf. Syst.*, vol. 19, no. 5, pp. 168–174, 2002.
[30] Edgewood Solutions. (2016). *Comparing Data Warehouse Design Methodologies*. Accessed: Jul. 11, 2018. [Online]. Available: https://www.mssqltips.com/sqlservertip/2976/comparing-data-warehouse-design-methodologies-for-microsoft-sql-server/
[31] A. Sebaa, F. Chikh, A. Nouicer, and A. Tari, "Medical big data warehouse: Architecture and system design, a case study: Improving healthcare resources distribution," *J. Med. Syst.*, vol. 42, no. 4, p. 59, 2018.
[32] M. Barkhordari and M. Niamanesh, "Atrak: A MapReduce-based data warehouse for big data," *J. Supercomput.*, vol. 73, no. 10, pp. 4596–4610, 2017.
[33] N. Jukić, A. Sharma, S. Nestorov, and B. Jukić, "Augmenting data warehouses with big data," *Inf. Syst. Manage.*, vol. 32, no. 3, pp. 200–209, 2015.
[34] L. W. Santoso, "Data warehouse with big data technology for higher education," in *Procedia Comput. Sci.*, vol. 124, pp. 93–99, Dec. 2017.
[35] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proc. 12th Int. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2008, pp. 68–77.
[36] B. Kitchenham, D. Budgen, and O. P. Brereton, "The value of mapping studies—A participant-observer case study," in *Proc. 14th Int. Conf. Eval. Assessment Softw. Eng. (EASE)*, 2010, pp. 25–33.
[37] T. Greenhalgh and R. Peacock, "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources," *Bmj*, vol. 331, no. 7524, pp. 1064–1065, 2005.
[38] A. A. Chadegani *et al.*, "A comparison between two main academic literature collections: Web of science and scopus databases," *Asian Soc. Sci.*, vol. 9, no. 5, pp. 18–26, 2013.
[39] Z. Jourdan, R. K. Rainer, and T. E. Marshall, "Business intelligence: An analysis of the literature," *Inf. Syst. Manage.*, vol. 25, no. 2, pp. 121–131, 2008.
[40] A. Wibowo, "Problems and available solutions on the stage of extract, transform, and loading in near real-time data warehousing (a literature study)," in *Proc. 1st Int. Seminar Intell. Technol. Appl. (ISITIA)*, May 2015, pp. 345–350.
[41] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: A review based on the kind of knowledge," *J. Intell. Manuf.*, vol. 20, no. 5, pp. 501–521, 2009.
[42] S. Demigha, "A data warehouse system to help assist breast cancer screening in diagnosis, education and research," in *Proc. 2nd Int. Conf. Comput. Sci. Appl. (CSA)*, 2009, pp. 1–6.
[43] N. Pasyeka and M. Pasyeka, "Construction of multidimensional data warehouse for processing students' knowledge evaluation in universities," in *Proc. 13th Int. Conf. Mod. Problems Radio Eng., Telecommun. Comput. Sci. (TCSET)*, Feb. 2016, pp. 822–824.
[44] A. Aziz and A. Wahid, "Integration of heterogeneous databases in academic environment using open source ETL tools," in *Proc. 1st Int. Conf. Inform. Appl. (ICIA)*, 2012, pp. 433–439.
[45] A. A. Aziz *et al.*, "A framework for educational data warehouse using business intelligence technologies," *J. Theor. Appl. Inf. Technol.*, vol. 69, no. 1, pp. 50–58, 2014.
[46] R. Miranda and E. Miranda, "Management report for marketing in higher education based on data warehouse and data mining," *Int. J. Multimedia Ubiquitous Eng.*, vol. 10, no. 4, pp. 291–302, 2015.
[47] R. Miranda, E. Miranda, and E. Suryani, "Implementation of datawarehouse, datamining and dashboard for higher education," *J. Theor. Appl. Inf. Technol.*, vol. 64, no. 3, pp. 710–717, 2014.
[48] 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)," in *Proc. 2nd IEEE Int. Conf. Teach., Assessment Learn. Eng. (TALE)*, Aug. 2013, pp. 98–103.
[49] P. Bhanti, U. Kaushal, and A. Pandey, "E-governance in higher education: Concept and role of data warehousing techniques," *Int. J. Comput. Appl.*, vol. 18, no. 1, pp. 15–19, 2011.
[50] K. Budiarta, P. A. A. Wijaya, and C. G. I. Partha, "Analysis and design of data warehouse on Academic STMIK STIKOM bali," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, pp. 35–39, 2017.
[51] A. R. Iskandar and I. Taufik, "Readiness for implementing data warehouse of higher education system with Sodano's perspective," in *Proc. 4th Int. Conf. Cyber IT Service Manage. (CITSM)*, Apr. 2016, pp. 1–6.
[52] R. Panchal, "Study on barriers in adoption of data warehouse for the higher educational institutes," *We'Ken-Int. J. Basic Appl. Sci.*, vol. 1, no. 4, pp. 148–153, 2016.
[53] A. Khan, S. Ghosh, and S. K. Ghosh, "eDWaaS: A scalable educational data warehouse as a service," in *Proc. Int. Conf. Intell. Syst. Design Appl. (ISDA)*, 2018, pp. 998–1007.
[54] R. J. Salaki and K. A. Ratnam, "Agile analytics: Applying in the development of data warehouse for business intelligence system in higher education," in *Proc. 6th World Conf. Inf. Syst. Technol. (WorldCist)*, 2018, pp. 1038–1048.
[55] M. S. A. Bakar, A. Ta'a, S. C. Chit, and M. H. M. Soid, "Data warehouse system for blended learning in institutions of higher education," *e-Academia J.*, vol. 6, no. 2, pp. 144–155, 2018.
[56] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.
[57] Y. Song, F. Pramudianto, and E. F. Gehringer, "A markup language for building a data warehouse for educational peer-assessment research," in *Proc. 46th IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2016, pp. 1–5.
[58] I. M. Aljawarneh, "Design of a data warehouse model for decision support at higher education: A case study," *Inf. Develop.*, vol. 32, no. 5, pp. 1691–1706, 2016.
[59] Z. A. S. Abdullah and T. A. S. Obaid, "Design and implementation of educational data warehouse using OLAP," *Int. J. Comput. Sci. Netw.*, vol. 5, no. 5, pp. 2277–5420, 2016.
[60] F. Di Tria, E. Lefons, and F. Tangorra, "Academic data warehouse design using a hybrid methodology," *Comput. Sci. Inf. Syst.*, vol. 12, no. 1, pp. 135–160, 2015.
[61] C. A. Moturi and A. Emurugat, "Prototyping an academic data warehouse: Case for a public University in Kenya," *Brit. J. Appl. Sci. Technol.*, vol. 8, no. 6, pp. 550–557, 2015.
[62] S. Mirabedini, "The role of data warehousing in educational data analysis," *J. Novel Appl. Sci.*, vol. 3, no. 5, pp. 1439–1445, 2014.
[63] L. van Dyk, "A data warehouse model for micro-level decision making in higher education," *Electron. J. e-Learn.*, vol. 6, no. 3, pp. 235–244, 2008.
[64] M. Pasyeka, "Development of data warehouse structure model for educational process management," *Softw. Eng.*, vol. 6, no. 1, pp. 1–6, 2018.
[65] T. Cheowsuwan, P. Rojanavasu, B. Srisungsittisunti, and S. Yeewiyom, "Development of data warehouses and decision support systems for executives of educational facilities in northern Thailand to increase educational facility management capacity," *Int. J. Geoinf.*, vol. 13, no. 2, pp. 35–43, 2017.
[66] K. P. Debus, M. D. Kickmeier-Rust, and D. Albert, "Time well spent: A data warehouse solution for temporal analyses of educational data," in *Proc. 24th Int. Conf. Comput. Educ. (ICCE)*, 2016, pp. 160–168.
[67] A. Bondarev, D. Zakirov, and D. Zakirov, "Data warehouse on Hadoop platform for decision support systems in education," in *Proc. 12th Int. Conf. Electron. Comput. Comput. (ICECCO)*, Sep. 2015, pp. 1–4.
[68] L. Gao and Y. Chen, "Application research of University decision support system based on data warehouse," in *Computer Science and its Applications*. Berlin, Germany: Springer-Verlag, 2015, pp. 631–637.
[69] N. Ahmed, M. A. Anwar, and A. M. Al Ameen, "Program outcome data warehouse for assessing the effectiveness of academic program," in *Proc. 8th Int. Technol., Educ. Develop. Conf. (INTED)*, 2014, pp. 6570–6576.

[70] M. M. Triola and M. V. Pusic, "The education data warehouse?: A transformative tool for medical education research," *J. Graduate Med. Educ.*, vol. 4, no. 1, pp. 113–115, 2012.

[71] B. C. Schaefer, E. Tanrikulu, and A. Breiter, "Eliciting user requirements when there is no organization: A mixed method for an educational data warehouse project," *Procedia-Social Behav. Sci.*, vol. 28, pp. 743–748, Jan. 2011.

[72] M. Paunica, L. M. Matac, A. L. Manole, and C. Motofei, "Measuring the performance of educational entities with a data warehouse," *Ann. Univ. Apulensis, Ser. Oeconomica*, vol. 12, no. 6, pp. 176–184, 2010.

[73] I. Mekterović, L. Brkić, and M. Baranović, "Improving the ETL process of higher education information system data warehouse," in *Proc. 9th WSEAS*, vol. 8, no. 10, 2009, pp. 265–270.

[74] M. Čeliković, S. Aleksić, and I. Luković, "A data warehouse system for monitoring University education process," in *Proc. Int. Conf. Inf. Technol. (CIT)*, 2009, pp. 1–7.

[75] N. Dimokas, N. Mittas, A. Nanopoulos, and L. Angelis, "A prototype system for educational data warehousing and mining," in *Proc. Panhellenic Conf. Inform. (PCI)*, Aug. 2008, pp. 199–203.

[76] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM SIGMOD Rec.*, vol. 26, no. 1, pp. 65–74, Mar. 1997.

[77] R. Winter and B. Strauch, "A method for demand-driven information requirements analysis in data warehousing projects," in *Proc. 36th Annu. Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2014, pp. 1–9.

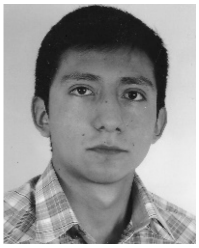
[78] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: A survey," *IEEE Trans. Neural Netw.*, vol. 13, no. 1, pp. 3–14, Jan. 2002.

[79] P. Chapman *et al.*, "CRISP-DM 1.0 step-by-step data mining guide," CRISP-DM Consortium, Brussels, Belgium, Tech. Rep CRISPMWP-1104, 2000.

[80] A. Collins and R. Halverson, "The second educational revolution: Rethinking education in the age of technology," *J. Comput. Assist. Learn.*, vol. 26, no. 1, pp. 18–27, 2010.

[81] S. Rizzi, A. Abelló, J. Lechtenbörger, and J. Trujillo, "Research in data warehouse modeling and design: Dead or alive?" in *Proc. 9th ACM Int. Workshop Data Warehousing OLAP (DOLAP)*, 2006, pp. 3–10.

OSWALDO MOSCOSO-ZEA received the Engineering degree in computer science from Pontifical Catholic University, Quito, Ecuador, in 2001, and the M.Sc. degree in information systems from Westfälische-Wilhelms Universität, Münster, Germany, in 2013. He is currently pursuing the Ph.D. degree in computer science with the University of Alicante, Spain. Since 2003, he has been an Aggregate Professor with Universidad Tecnológica Equinoccial, Quito, where he has been the Head of the Computer Science Department since 2016. He is also a Microsoft Certified Professional in software development and a Subject Matter Expert in data management at CompTIA, Chicago, IL, USA. He has participated in several projects as an IT consultant in different countries. His main research topics include educational data warehousing, big data, enterprise architecture, and knowledge management.



JOEL PAREDES-GUALTOR has studied in computer science and economics with full scholarships. He is currently a Research Assistant with the Computer Science Department, Universidad Tecnológica Equinoccial. His main research interests are data warehousing, data analytics, model-driven software development, and enterprise architecture. He has participated in different research projects. He has been part of the organizing committee of several international scientific conferences and meetings. He has received mentions as the Best Grade Point Average in the Engineering Faculty and in the Computer Science Department.

SERGIO LUJÁN-MORA received the degree in computer science and engineering from the University of Alicante, Spain, in 1998, and the Ph.D. degree in computer engineering from the Department of Software and Computing Systems, University of Alicante, in 2005. He is currently a Senior Lecturer with the Department of Software and Computing Systems, University of Alicante. He has authored several books and numerous articles published in various conference proceedings (ER, UML, and DOLAP) and high-impact journals (DKE, JCIS, JDBM, JECR, JIS, JWE, IJEE, and UAIS). His main research interests include Web applications, Web development, and Web accessibility and usability. In recent years, he has focused on e-learning, massive open online courses, open educational resources, and the accessibility of video games.