Southeast Europe Journal of Soft Computing ISSN 2233 – 1859
Available online: http://scjournal.ius.edu.ba

Vol 11, No. 1, March 2022, pp. 50-61

COVID-19 Data Warehouse: A Systematic Literature Review

[1*]Ahmed Khaled AbdelLatif, [1]Ahmed Naji Abdullah, [1]Ahmed Munther Abboud, [1]Zahraa Abdulkareem Mohammed, [1,2]Hisham Noori Hussain, [1*]Alaa Khalaf Hamoud

[1]College of Computer Science and Information Technology, University of Basrah, Iraq,
[2]School of Housing, Building and Planning, Universiti Sains Malaysia, 11800, Penang, Malaysia.
Corresponding Author: alaa.hamoud@uobasrah.edu.iq.

Article Info

Article history:

Article received on 13 July 2022

Received in revised form 28 July 2022

Keywords:

COVID-19 Data Warehouse; Data Warehouse; SARS-Cov-2; COVID-19 Data Mart; COVID-19 Infection.

ABSTRACT: The SARS-CoV-2 disease (COVID-19) affects the whole world and led clinicians to use the available knowledge to diagnose or predict the infection. Data Warehouse (DW) is one of the most crucial tools that may enhance decision-making. In this paper, three main questions will be investigated according to using DW in the COVID-19 pandemic. The effect of using DW in the field of diagnosing and prediction will be investigated, besides, the most used architecture of DW will be explored. The sectors that faced a lot of researchers' attention, such as diagnosing, predicting, and finding the correlations among features will be examined. The selected 27 studies out of more than 2700 are explored where the papers that have been published between 2019-2022 in the digital libraries (ACM, IEEE, Springer, Science Direct, and Elsevier) in the field of DW that handle the COVID-19 are selected. During the research, many limitations have been detected, while some future works are presented. Enterprise DW is the most used architecture for COVID-19 DW while finding correlation among features and prediction are the sectors that had taken the researchers' attention. Using DW as a platform requires historical data, which can be a limitation for using DW in the COVID-19 pandemic.

1. INTRODUCTION

After being first discovered in Wuhan, China, the COVID-19 outbreak quickly spread to all nations. According to a report from the World Health Organization, COVID-19 will be a major public health issue in the world in 2020.After that, the increasing infectious diseases are increasing in rapid spread, endangering the health of the entire world and thus transmission of infection is carried out to large numbers of people, and therefore, immediate measures to prevent COVID-19 must be taken when the local and global community level and the number of infected people reached the largest number before February 20, 2020. The percentage of infected people expected to be infected with them. It is expected that this emergency case of the disease will reach its highest statistic in late May 2020 and is expected to start declining in early July 2020. Based on news queried in the questionnaires that were worked on globally, and it was found that there are more negative articles than positive articles, as the psychological factor was one of the main reasons for the outbreak of the disease in addition to the speed of its transmission in the air and climate fluctuations and the lack of commitment of all people in the world to quarantine and preventive precautions COVID-19 is still a contagious disease that is not clear, which means that we cannot get a reason for its existence or how to get rid of it. At the end of 2020 and the beginning of 2021 the vaccine was discovered that reduced the spread of corona and did not eliminate it completely [1][2][3]. This pandemic makes all the researchers around the world working to find solutions. The medical records considered as the start to find the best solution to help in this situation. While the medicine and medical

treatment costs and the death ratio are increasing, the proposing systems and tools that may help in the pre-diagnostic may reduce the costs and help the medical staff and institutions.

The Clinical Data Warehouse (CDW) may be one of the best solutions in the field of decision-making and analyze the medical data to find the hidden trends and discover the knowledge. Decisions made based on bad or inconsistent data may lead to disastrous results and then to wrong decisions [4], [5]. The Data Warehouse (DW) is considered as one of the best platforms to support decision making in different disciplines since it contains integrated data which can be viewed in a multidimensional data view that allows a deep visualization and analyzing process. The stored data allows obtaining high quality results in order to support decisions [6][7][8]. The medical decision related to COVID19 is due to the need for zero error in the decision making. DW can serve as a platform for the study, analyzing, reporting, and supporting clinical decisions. The integrated data in different sectors can make the multi-decision making criteria according to different sectors related to the pandemic [9][10]. The importance of online analytical processing (OLAP) and Key Performance Indicator (KPI) is the ability in supporting strategic decisions. The ability of OLAP can be summarized in presenting the information from different view points and visualizing the analytical results in a way that helps in finding the hidden patterns in data. There are three famous OLAP server types: Multidimensional OLAP (MOLAP), Relational OLAP (ROLAP), and Hybrid OLAP (HOLAP) where each type can handle a specific type of stored information. ROLAP utilizes the relational database as a platform to store the data, while MOLAP stores the data in a multidimensional cube. HOLAP server stores the large data in a relational database while the analyzed aggregated data in the multidimensional cube. HOLAP type is more efficient than MOLAP in handling the large size of data and more efficient than ROLAP in the query processing [11][12][13]. The KPI on the other hand, can help in giving instant alert when the trend changes and not achieving the goals. The KPI value plays a vital role in the evaluation and visualizing of the performance of any company by comparing the current performance with the goals that have been set [14]. This paper will investigate using DW in the field of visualization and analyzing of COVID19 data based on DW. OLAP and KPI will be investigated as supporting tools for analyzing the COVID-19 data. Three main questions will be answered according to the study sample of papers from five databases : IEEE, ACM, Springer, ScienceDirect, and Elsevier.

The rest of the paper is arranged as follows: section two lists the literature review in the topic of COVID-19 DW and presents the selected studies. Section three explains the systematic literature review and how to formulate the questions. Section four presents the results and discussions, while section five shows the concluded points, limitations, and future works.

2. LITERATURE REVIEW

The next table (1) shows the literature review and the data type of the data used in each paper and DW architecture type that has been implemented. Enterprise DW is mostly used for implementing DW for COVID-19. Text and image data types are the available types of data in the pandemic. Text data type is mostly used to implement DW to either find the correlation among features, predict infection, or diagnose the infection.

Table 1 Summaries of Literature.

Ref	Abstract	Data Type	DW Type
[15]	The goal of the study is to design, build, and test the Extract, Transform, and Load (ETL) tool that converts data in order to be used in COVID-19 data integration.	Text	Data Mart
[16]	The goal is to use a national dataset to look at trends in drug usage among COVID-19 patients across the US.	Text	Data Mart
[17]	The goal of the study was to create models that can be used easily to access data from hospital admission in order to stratify patients by likelihood of severe outcomes following COVID-19 hospitalisation.	Text	Enterprise
[18]	The innovative proposed Data Mart technology is scalable and appears to hold promise as a broadly applicable approach for simultaneously recommending individuals from a pool of patients versus a pool of trial possibilities which are time-sensitive.	Text	Enterprise
[19]	To objective is to assess the accuracy of diagnostic	Text	Enterprise

	entry in issue lists in a hospital Electronic Health Record (EHR) system in the COVID-19 pandemic.		
[20]	In order to measure the current level of travel-related activity relative to the pre-pandemic baseline and investigate its relationship with confirmed COVID-19 cases, this study proposes a Community Activity Score (CAS) based on inter-community traffic characteristics (in and out of community traffic volume and travel distance).	Text	Enterprise
[21]	A spatial network approach is suggested in this study for locating the SSPs that disproportionally spread COVID-19. Using activity information at the individual level for the Hong Kong confirmed instances.	Images	Enterprise
[22]	Using panel data to examine the differences between the psychological intentions of Japanese consumers toward using e-commerce and the corresponding causes before and after the COVID-19 outbreak.	Text	Enterprise
[23]	A visualization (data mining approach) is proposed to support the strategic decisions in the field of airlines by extracting the knowledge of competitors database.	Text	Enterprise
[24]	A machine learning and OLAP approach is proposed for epidemiological data processing and analysis. Big data is obtained from a variety of data; an example of this is epidemiological data such as (Covid-19). OLAP helps researchers get better results to reveal new ways to treat disease.	Text	OLAP
[25]	To diagnose COVID-19, it is necessary to develop tools that treat different	Text	
	data on the disease. Information technology is an important area for making decisions about COVID-19 by analyzing regional data in countries.		
[26]	A large spatial data analysis is proposed for COVID-19 diagnosis in various geographical locations, which leads researchers and scientists to understand large amounts of data to understand the disease COVID-19 and to discover ways to control the epidemic.	Text	
[27]	The Covid-19 epidemic has caused the global economy to contract, forcing governments and businesses to rely on DW and Decision Support System (DSS) data analytics to make better decisions and profitable businesses to draw in customers by utilizing service continuity measures.	Images	Enterprise
[28]	A data mining approach is proposed to derive important data and also evaluate data designs through one or more applications. Also, the Orange tool was used to depict the relationships between COVID19 data, as it relies on groups to extract accurate data.	Images	
[29]	A mortality prediction model was created for COVID-19 diagnosis. The clinical characteristics of the patient were used to create the prediction model. Analysis was done on patient information stored in the Mount Sinai Health System database in New York City, New York, USA.	Images	
[30]	The case study for this study concerns university students in Jordan during the COVID-19 epidemic, and a new educational	Text	

	system, the e-learning system, has been implemented in Jordanian universities.		
[2]	By relating the data from this study to the overall number of COVID-19 cases in Texas on a given day, it is able to derive numbers that have significance. Many gadgets It can deliver an item location data stream over time (in GPS format). automobiles, cell phones, etc.	Images	
[3]	Computational techniques are integrated the current and expanding COVID-19 and its related Data (eg pollution, demographics, climate, etc.) in order to contradict COVID-19 data.	Images	Data Mart
[31]	The impact of COVID-19 on healthcare molarity and utilisation by enthancity and race of patients who have been tested for SARS-Cov-2 based on DW usage was studied by the researchers.	Text	Enterprise
[32]	The association between in-hospital mortality and the antihypertensive agent exposure with COVID-19 is investigated to explore the COVID-19 diagnosing.	Text	Enterprise
[33]	DW is utilized for SARS-Cov-2 data to compare the relationship of adverse outcomes with burden in adults of different age ranges.	Text	Enterprise
[34]	The researchers aimed to find the the association between the molarity and chlorpromazine use among hospital patients with COVID-19 based on DW.	Text	Enterprise
[35]	DW utilized to investigate the limitations of flu surveillance system in France.	Text	Enterprise
[36]	DW is used to provide the COVID-19 clinical information, demographic and social information, and vaccination. The vaccination information, such as the number of doses, or types of doses are the outcome of this study.	Text	Enterprise
[37]	The researchers used DW to utilize the obtained data related to employees, such as demographic information, burnout, and employee engagement for those who work in primary care institutions. They linked these data with burnout of historical data in 2019 and death rates of COVID-19 in 2020.	Text	Enterprise
[38]	DW data is utilized in order to compare with the historical activities in 2019 regarding COVID-19 pandemic. The analyzed activities included many activities such as the first consultation, therapy assignment, disease re-assignment, tele examination, and follow-up visits.	Text	Enterprise

3. RESEARCH METHOD

The Systematic Literature Review (SLR) can be characterized by covering many previous studies and depicts the comprehensive results depending on a specific strategy. SLR is widely used to summarize the research evidence rather than narrative reviews and expert commentaries. SLR follows a rigorous approach in collecting the literature in order to get accurate information [39][40][41][42][43]. SLR should have a protocol where it can be followed to collect and extract data, and then synthesis of the extracted data as shown in Figure (1).

Figure 1: Systematic Literature Review Steps [39].

3.1 Formulating the systematic review questions

The formulation of the research question and finding the answers is the first step in SLR. In order to find the impact of using DW in the field of discovering COVID-19 infection, the researchers identified the questions as follows:

What is the impact of using Data Warehouse on discovering trends in COVID-19 pandemic?

After formulating the first question to collect and summarize the selected papers that utilized DW or any tools such as KPI, and OLAP in discovering the hidden patterns. There are many architectures for DW that can be adopted in the field of implementing DW. According to that, the researchers formulated another research question:

What is most DW architecture used for COVID-19 data?

The COVID-19 pandemic that was handled by DW has been faced by many sectors. The field that faced a lot of attention related to COVID-19 DW will be explored. According to that, the researchers formulated the next question:

What is the sector that faced a lot of attention in the field of COVID-19 DW?

3.2 Constructing the search

The relevant keywords that have been adopted in this SLR are determined in order to investigate the previous studies. The digital libraries are determined to search for the keywords in a specific range of years according to the basic topic of SLR.

3.2.1 Search process

The digital libraries that used for the searching process are: ACM, IEEE, Science Direct, Springer, and Elsiever digital libraries. These digital librarieshave been chosen according to the papers availability to the researchers. The years from (2019 to 2022) have been selected according to start of COVID-19 panademic and ending with the year 2022 according to last paper that heve been published in the same topic.

3.2.2 Terminology

The digital libraries that are used for the search process are: ACM, IEEE, Science Direct, Springer, and Elsiever digital libraries. These digital libraries have been chosen according to the papers available to researchers. The years from (2019 to 2022) have been selected according to start of COVID-19 pandemic and ending with the year 2022, according to last paper that has been published in the same topic.

3.3 Study selection

The process of selection of the papers has been made according to many criteria that make the researcher select or exclude the papers. These criteria are the correlation between using DW and COVID-19 pandemic, the researchers of the selected papers used DW or its tools to handle the COVID-19, the paper date falls in the range (2019-2022), the paper is not duplicated in many libraries. Otherwise, papers that didn't match the criteria have been excluded. Figure (2) represents the SLR search strategy where it shows the years range that has been selected for search, the keywords, and the sources of libraries.

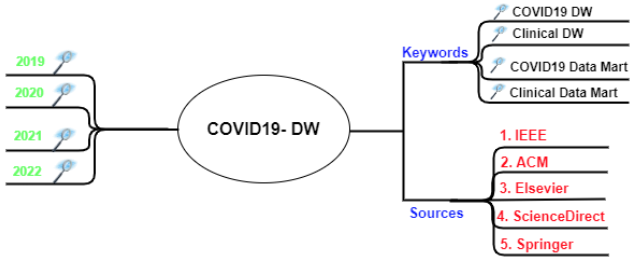

Figure 2: Search strategy of systematic literature review

3.4 Data Extraction

The table (2) represents and includes all papers in this SLR according to title, authors, publication type, publication date, and digital library. The table includes the information of each paper according to the details that have been set before. This SLR relies on strategies that have been presented by Salvado, Nakasone, and Pow-Sang (2014) in data extraction.

Table 2 SLR Selected Study Papers

Ref	Authors	Publication Type	Date	Library
[24]	Carson K. Leung et al.	Conference	2020	IEEE
[25]	Oleksii Duda et al.	Conference	2020	IEEE
[26]	Siyuan Shang et al.	Conference	2020	IEEE
[27]	Romona M. Harris	Conference	2020	IEEE
[28]	Uzma Thange et al.	Journal	2021	IEEE

[16]	Stroever, Stephanie J et al.	Journal	2021	ScienceDirect
[17]	Thomas W. Campbell et al.	Journal	2021	ScienceDirect
[18]	Tara T. Helmer et al.	Jouranl	2021	ScienceDirect
[19]	Jordan Poulos et al.	Journal	2021	ScienceDirect
[20]	Yuntao Guo et al.	Journal	2021	ScienceDirect
[21]	Jianwei Huang et al.	Journal	2021	ScienceDirect
[15]	Yue Yu et al.	Journal	2022	ScienceDirect
[22]	Tomoya Kawasaki et al.	Journal	2022	ScienceDirect
[23]	Darío Pérez-Campuzano et al.	Journal	2022	ScienceDirect
[29]	Arjun S Yadaw et al.	Journal	2020	Elsevier
[30]	Manaf Al-Okailya et al.	Journal	2020	Elsevier
[1]	Lucy Lu Wang et al.	Journal	2020	ACM
[44]	Mohammad A. Shaito et al.	Conference	2021	ACM
[3]	Giuseppe Agapito et al.	Conference	2020	ACM
[31]	Razjouyan, Javad et al.	Journal	2022	Springer
[32]	Chouchana, Laurent et al.	Journal	2022	Springer
[33]	O'Hare, AM et al.	Journal	2021	Springer
[34]	Hoertel, Nicolas et al.	Journal	2021	Springer
[35]	Fartoukh, Muriel et al.	Journal	2021	Springer
[36]	K. H. Seal et al.	Conference	2022	Springer
[37]	M. S. J. Smirnova et al.	Conference	2022	Springer
[38]	G. Targato et al.	Conference	2021	Springer

3.5 Synthesis of the extracted data

The last step of SLR is collecting and summarizing the obtained results from literature studies. Three rounds of filtering have been made; the first round is selecting the papers from digital libraries; the second round is determining the studies that utilized DW in the study, while the final round is determining the studies that match the criteria. There are 27 studies that have been extracted in the methodology stage that achieved the criteria. Figure (3) shows the number of the selected studies that have been investigated in this SLR according to digital libraries. All the studies can be accessed in the link:

https://drive.google.com/drive/folders/14r1pNjMNRMwWxlTknIBFY4lrkP5uP732?usp=sharing

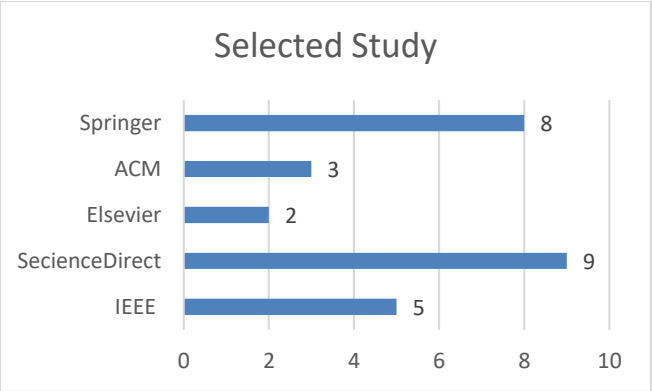

Figure 3: The total number of studies selected

4. RESULTS AND DISCUSSION

DW is a database management system that contains a large group of data and performs many tasks, the most important of which is queries and analyzes. The sources that the data warehouse produces for the purpose of utilizing the analyzes generated through case processing. It is also able to integrate huge amounts of data through different sources for the purpose of

producing a clear vision for the institutions to facilitate the decision-making process. Big data includes epidemiological and health data collected in multiple sources accurately and quickly, such as the data of patients infected with the Corona virus-19, which is required of scientists to transmit it, so that DW will help obtain results for the purpose of discovering and inventing appropriate treatments to control the disease. It provides a scientific solution for the analysis of epidemiological data. A virus analysis is carried out through the data of confirmed cases of infection, death and recovery cases that have been scheduled and studied on a weekly basis to refine the data to extract positive results regarding the disease. DW was used, which led to a rapid productivity in this regard. DW always produces an easy policy because it is based on the available data, which is available through data services, which leads us to a framework for data mining and inquiries, the creation of a different set of data and ETL design, which helps companies, organizations, corporations or governments increase their ability to make decisions from Through DW, which gives you many tools to detail the data in any form that everyone needs in the analysis process. We conclude that it is the only and powerful solution to support decision makers [45][46][47].

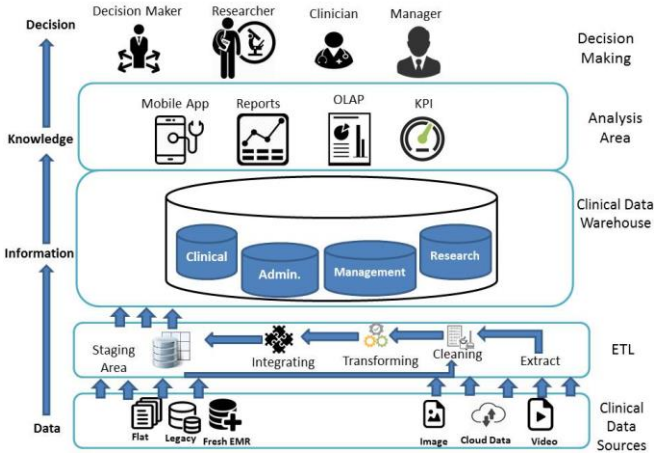

Figure 4: Clinical DW [48]

CDW architecture, and the data flow from each step of ETL is shown in figure (4). The figure shows the tools that can be used to support strategic decision-making by clinical stakeholders. The architecture of CDW reduces the time for the purpose of collecting, validating, and processing the clinical operational data in order to ensure the data quality that ensures the right decision making [49][50][51][52]. Many researchers in the field of research on COVID-19 virus disease depend on the data warehouse, which contains statistics by taking samples on a certain number of infections with the virus, the possible treatment for them, or the number of deaths for the purpose of predicting the results and what are the effects that occurred due the COVID-19 pandemic [16][18][20][21][22][23]. Prediction of the effects of the pandemic depend on the data warehouse, which contains a large number of samples and statistics, and it is not possible to reach final results on some samples because the pandemic is continuous and changing the number of injuries or deaths during the time , so a specific sample is taken from the data warehouse for the injured or dead due to the COVID-19 pandemic[23][20][17]. Most of the data is also taken from the hospital data warehouse [18], regarding the various studies based on the data warehouse of COVID-19 infection, approval must be made through the Transparency Portal in accordance with the European Regulation on Data Protection from the National Authority for Information Technology and Civil Liberties (CNIL) in order to ensure patient information. By using Electronic Health Records (EHR) derived data from records at the time of hospital admission can predict the stage of COVID-19 on the patient and his health condition and what are the methods of treatment or mitigation from the virus [15][17][18][19].

The process of extracting data from different data sources consists of three main steps Extract, Transform, and Load (ETL). It collects data from different sources and copies it into a single system, which represents a different set of data from different sources and in a different context. It can be converted and stored in a data warehouse. One of his tasks was to clean up the data and turn it into an easy-to-understand model. In order to perform data checks and to get good extractions in scientific and medical experiments. ETL makes it easy to index and quickly access the required data. Scientists have benefited greatly from the use of ETL, where samples of suspected infections, confirmed cases, and fatalities have been collected for a large data set to examine reliable data from inaccurate data that cannot be trusted, while developing a mechanism to detect what is wrong with the data and process it to reach treatment faster . The scientists added data tracking by domains associated with the data source and history to achieve accurate data with a high layer of quality. The necessity of using ETL in the data where the Extract process is used is the first and important step and the feature allows us to extract all the data or extract part of the data with or without notice of bold updates. As for Transform, which is done with several data extraction tools to convert it into an understandable format, and it can convert multiple data or through DW. Load inserts

the data into the final database with several completions of updating, in addition, ETL pipelines allows us to through which high-level accuracy data is obtained through data transmission, processing and retention in a system with different efficiency, which facilitates the process of sorting and obtaining accurate data for use in loading and conducting scientific tests on them [53][54][55].

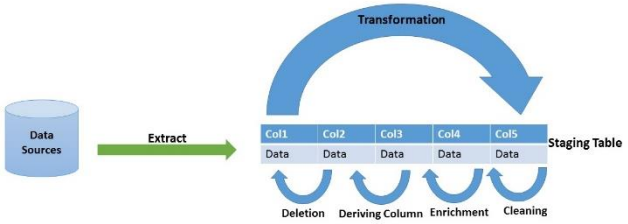

Figure 5: ETL [8]

As shown in figure (5), an ETL tool is designed to perform tasks such as extracting, transforming, and loading, the procedure of copying data from one or more sources to a destination system that represents the data differently from the source . The ETL tool is used to convert the records of patients infected with COVID-19 virus from a certain number of tables to a smaller number, and this process may be accompanied by the loss of some information, which is very little compared to the accuracy of the results. Working with ETL will work with information recovery, correspondence, sharing and investigation between various organizations. To ensure that there are no records missing during the ETL interaction, a correlation of the pre-owned table records is performed to guarantee that all information from the first data set has been effectively moved to the objective data set. An ETL instrument is frequently assembled utilizing SQL contents to accomplish explicit errands for instance information change, for example, esteem change, rule-based change, and idea planning, Contrasts in information definitions might cause information quality issues e.g planning Loss of data because of contrasts in data exactness, contrasts in favored clinical phrasing, may prompt calculated jumble issues during the standard idea code task process, ETL execution is assessed utilizing the CDM when utilized in genuine applications [15][17].

The multidimensional data model serves as the foundation of the (Online Analytical Processing) server. Because it enables managers and analysts to access information quickly, consistently, and interactively, OLAP plays a crucial role in data warehousing. Relational OLAP (ROLAP), multidimensional OLAP (MOLAP), and hybrid OLAP (HOLAP) are the three basic types of OLAP. Other OLAP types include (WOLAP), (DOLAP), (MOLAP), and (HOLAP) (SOLAP). In a relational data warehouse, snowflake schema, star, or specialized data management system, OLAP data is typically stored.

All OLAP systems consist of a multi-dimensional cube or the so-called hyper-cube, and within it there is a set of numerical facts, which are called measures and are divided and categorized according to the dimensions as in the pivot tables. Examples of a cube are store sales, which are based on a measure and date/time, as each sale has a date/time that explains more about the sales process . The OLAP processes show data as a cube with hierarchical dimensions, allowing for simple operations to speed up analysis. Cutting, milling, dice, rolling, and spinning are frequent activities. OLAP is used to swiftly and accurately extract information, analyze multi-dimensional data, and produce reports on the results. It is also used in several areas such as marketing operations, business administration, writing administrative reports, and the emergence of other fields such as agriculture [56][57][58].

There are six different types of OLAP servers, with Multidimensional OLAP (MOLAP), also referred to as the traditional OLAP type being the first. This type stores and processes data using a Multidimensional Database (MDDB). Summaries may be efficiently stored in MDDB, giving users a mechanism to swiftly query and get data from the database for processing. The second one is ROLAP server where it can handle more data than the MOLAP server can, whether they are written, or digital, productively. Allows customers to be directed to the lowest rung of the hierarchy. The third type of OLAP is Hybrid OLAP (HOLAP) where it incorporates both MOLAP and ROLAP. HOLAP combines the advantages of both systems by making use of information stores from both ROLAP and MOLAP. The fourth type is Web OLAP (WOLAP), also referred to as OLAP that supports the web, is accessed through a web browser. The fifth type is Cabinet Desktop Online Analytical Processing (DOLAP) is a single-layer OLAP solution that runs on desktop computers. 6. SOLAP (SOLAP) to deal with issues in the geo-business intelligence industry [2][3].

The COVID-19 pandemic, is an ongoing global infectious disease; the virus was first identified in China, specifically in Wuhan, in December 2019, and they attempted to contain the virus but failed. The impact of the disease is greater on children and the elderly, especially those who have diabetes and heart disease, where the disease leads to a fluctuation in the

level of sugar in the blood from the expected level. The World Health Organization declared a state of emergency in January 2020 and an epidemic in March 2020; the pandemic has caused more than 6.04 million deaths out of 458 million cases of the disease, and that makes it one of most deadliest viruses in the world. The pandemic has affected the social and economic situation all over the world, resulting in widespread food and supply shortages, and the closure of public places and educational institutions causing political disruption and public health. The pandemic disrupted food systems around the world, causing an estimated 690 million undernourished people to remain food insecure in 2019 due to reduced incomes, reduced food production and high prices around the world. The pandemic has affected educational systems around the world, as many educational institutions have been closed and rely on online education. The pandemic has caused enormous tension in the economy, as it has caused an increase in online commerce; increased orders for home delivery, restaurants have been closed, sales have decreased, and prices have followed suit [59][60][61][30].

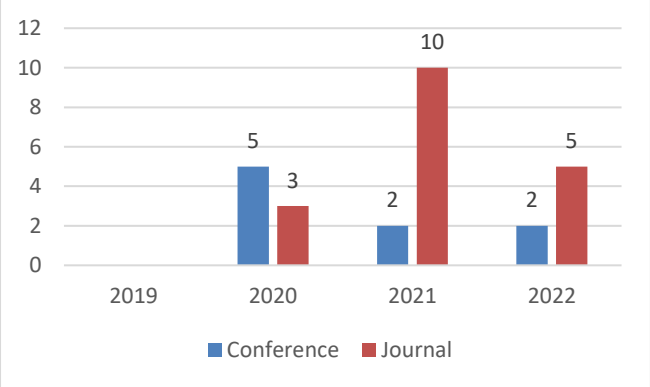

Figure 6: Journal vs conference COVID-19 DW papers

Figure (6) shows the number of selected studies according to years and classified by publication type (whether it be in a journal or in a conference). The year where it shows the peak of pandemic has the most published papers where it holds 12 papers (two conference papers and ten journal papers). The first year of the pandemic (2019) didn't hold any study in the main topic since there is no data available for the COVID-19. The first year (2019) didn't face any research in the field of COVID-19 DW due to the low availability of data. The second year starting with publishing studies (five conference papers and three journal papers). The current year faced seven study articles (two conference papers and five journal papers).

5. CONCLUSION AND FUTURE WORKS

DW in the COVID -19 pandemic helped to predict future outcomes, which in turn may reduce the number of infections or diagnose the condition of the injured by taking previous results for the number of injured and analyzing these results [16][17][19][20]. DW had a positive effect on COVID-19 diagnosing and predicting by helping researchers and doctors to analyze large data quickly and accurately to obtain results that help discover treatment and reduce the spread of the disease [45][26][62]. DW can be considered a tool and platform to support clinical strategic decisions related to diagnosing and predicting COVID-19 infections. COVID-19 DW is used to store the demographic, clinical, and personal information of patients and find the relationship of factors among patients information [31][32]. Most of the data used in the DW structure is clinical data, which is mostly electronic health care records (EHR) data that typically measures and records temperature, pulse rate, respiratory rate, and blood pressure [15][16][17][31][19].

Data sources, integration, warehousing, analysis, and data visualization are the five independent and interdependent levels that make up the COVID-Warehouse architecture. In addition to the business DW, the corporate DW was used in the majority of the research. Numerous considerations relating to patient data led to the use of these architectural designs. The other reason for using enterprise and corporate DW is the need to integrate different kinds of data [36][37]. Besides, Enterprise Data Warehouse (EDW) was used mostly to conduct analyzes of patients infected with the COVID-19 using electronic health care records [29][30]. The data mart used for the COVID-19 allows a specific line of discrete data that has been thoroughly vetted and can be obtained and viewed to infer positive results for more accurate treatment, giving you quick access rather than spending time a more data warehouse is complex [46].

Most of the sectors that have faced interest in the field of the COVID-19 pandemic for the DW are the hospital sector, or in particular, the data of the infected patients, their condition, the various effects on them, the number of injuries and deaths, etc [16][18][19][20]. The sectors that have been interested in the field of COVID-19 for DW are the medical experts, as they insisted on analyzing the data of infected patients more extensively to get an answer for the world to benefit from it [27][26]. Most of the sectors that have received attention in the field of the Covid-19 pandemic for the DW are the countries in which the epidemic began,

North Korea and America, in order to study the causes of disease outbreaks and hospitals in those countries and then the rest of the world, or in particular the data of infected patients and their condition and the various effects on them and the number of injuries and deaths [1][2]. Feature selection, finding, the correlation among factors, prediction, and diagnosing are the most important sectors in the field of COVID-19 DW [34][35]. The most sector that have a lot of attention in the field of COVID-19 data warehouse it's the health care and hospital sector, Where studies were conducted in hospitals in order to analyze patients' cases and obtain the numbers of patients and deaths inside and outside the hospital [29][30].

Daily COVID-19 statistics are provided, along with a variety of other pollution and climate-related data. A dimensional fact model was used to combine and organize data on COVID-19 instances, pollution actions, and climatic data utilizing two main dimensions: time and geographical position. All of this data is utilized in a public health context to highlight the ways in which an epidemic spreads over time and space and to link the outbreak to statistics. Climate and pollution in a certain location [1]. The effect of using the data warehouse was very beneficial as it was used in hospitals and health centers to collect and analyze large data and predict the outcomes and to quickly access data for planning and providing high-quality care and decision support [29][30].

There are many limitations in the field of COVID-19 DW, such as the difficulties in the data collection in the beginning of the pandemic. The data availability, the difficulties in understanding the behavior of COVID-19 virus led to the difficulties in determining the right data. The different kinds of limited and small data makes it difficult to use the DW as a platform since DW requires historical and huge data to integrate and implement. The field of feature extraction from (images) and using the extracted data to analyze the behavior of the virus, then linking the data with the historical data of patients may led to discover new trends which led to accurate decision-making process [48].

REFERENCES

[1] L. L. Wang et al., “Cord-19: The covid-19 open research dataset,” ArXiv, 2020.

[2] M. Shaito and R. Elmasri, “Map visualization using spatial and spatio-temporal data: Application to covid-19 data,” in The 14th PErvasive Technologies Related to Assistive Environments Conference, 2021, pp. 284–291.

[3] G. Agapito, C. Zucco, and M. Cannataro, “COVID-warehouse: A data warehouse of Italian COVID-19, pollution, and climate data,” Int. J. Environ. Res. Public Health, vol. 17, no. 15, p. 5596, 2020.

[4] S. R. Gardner, “Building the data warehouse: the tough questions project managers have to ask their" companies’ executives--and themselves--and the guidelines needed to sort out the answers,” Commun. ACM, vol. 41, no. 9, pp. 52–61, 1998.

[5] H. Cao, M. Markatou, G. B. Melton, M. F. Chiang, and G. Hripcsak, “Mining a clinical data warehouse to discover disease-finding associations using co-occurrence statistics,” in AMIA Annual Symposium Proceedings, 2005, vol. 2005, p. 106.

[6] M. M. Jaber, M. K. Abd Ghani, N. Suryana, M. A. Mohammed, and T. Abbas, “Flexible data warehouse parameters: Toward building an integrated architecture,” Int. J. Comput. Theory Eng., vol. 7, no. 5, p. 349, 2015.

[7] A. K. Hamoud, M. Abd Ulkareem, H. N. Hussain, Z. A. Mohammed, and G. M. Salih, “Improve HR decision-making based on data mart and OLAP,” in Journal of Physics: Conference Series, 2020, vol. 1530, no. 1, p. 12058.

[8] A. K. Hamoud, H. N. Hussien, A. A. Fadhil, and Z. R. Ekal, “Improving service quality using consumers’ complaints data mart which effect on financial customer satisfaction,” in Journal of Physics: Conference Series, 2020, vol. 1530, no. 1, p. 12060.

[9] C. Saegerman et al., “Clinical decision support tool for diagnosis of COVID-19 in hospitals,” PLoS One, vol. 16, no. 3, p. e0247773, 2021.

[10] S.-Y. Shin, W. S. Kim, and J.-H. Lee, “Characteristics desired in clinical data warehouse for biomedical research,” Healthc. Inform. Res., vol. 20, no. 2, pp. 109–116, 2014.

[11] J. M. DAHR, A. K. HAMOUD, I. A. NAJM, and M. I. AHMED, “IMPLEMENTING SALES DECISION SUPPORT SYSTEM USING DATA MART BASED ON OLAP, KPI, AND DATA MINING APPROACHES,” J. Eng. Sci. Technol., vol. 17, no. 1, pp. 275–293, 2022.

[12] A. Nanda, S. Gupta, and M. Vijrania, “A comprehensive survey of OLAP: recent trends,” in 2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 425–430.

[13] I. Moalla, A. Nabli, L. Bouzguenda, and M. Hammami, “Data warehouse design approaches from social media: review and comparison,” Soc. Netw. Anal. Min., vol. 7, no. 1, pp. 1–14, 2017.

[14] E. Dom\’\inguez, B. Pérez, A. L. Rubio, and M. A. Zapata, “A taxonomy for key performance indicators management,” Comput. Stand. Interfaces, vol. 64, pp. 24–40, 2019.

[15] Y. Yu et al., “Developing an ETL tool for converting

the PCORnet CDM into the OMOP CDM to facilitate the COVID-19 data integration," J. Biomed. Inform., vol. 127, p. 104002, 2022.

[16] S. J. Stroever et al., "Medication Use Among Patients With COVID-19 in a Large, National Dataset: Cerner Real-World DataTM," Clin. Ther., vol. 43, no. 6, pp. 173–196, 2021.

[17] T. W. Campbell et al., "Predicting prognosis in COVID-19 patients using machine learning and readily available clinical data," Int. J. Med. Inform., vol. 155, p. 104594, 2021.

[18] T. T. Helmer et al., "Creating and implementing a COVID-19 recruitment Data Mart," J. Biomed. Inform., vol. 117, p. 103765, 2021.

[19] J. Poulos, L. Zhu, and A. D. Shah, "Data gaps in electronic health record (EHR) systems: An audit of problem list completeness during the COVID-19 pandemic," Int. J. Med. Inform., vol. 150, p. 104452, 2021.

[20] Y. Guo, H. Yu, G. Zhang, and D. T. Ma, "Exploring the impacts of travel-implied policy factors on COVID-19 spread within communities based on multi-source data interpretations," Health Place, vol. 69, p. 102538, 2021.

[21] J. Huang, M.-P. Kwan, and Z. Kan, "The superspreading places of COVID-19 and the associated built-environment and socio-demographic features: A study using a spatial network framework and individual-level activity data," Health Place, vol. 72, p. 102694, 2021.

[22] T. Kawasaki, H. Wakashima, and R. Shibasaki, "The use of e-commerce and the COVID-19 outbreak: A panel data analysis in Japan," Transp. Policy, vol. 115, pp. 88–100, 2022.

[23] D. Pérez-Campuzano, L. R. Andrada, P. M. Ortega, and A. López-Lázaro, "Visualizing the historical COVID-19 shock in the US airline industry: A Data Mining approach for dynamic market surveillance," J. Air Transp. Manag., vol. 101, p. 102194, 2022.

[24] C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang, and A. Cuzzocrea, "Machine learning and OLAP on big COVID-19 data," in 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 5118–5127.

[25] O. Duda, V. Pasichnyk, N. Kunanets, R. Antonii, and O. Matsiuk, "Multidimensional Representation of COVID-19 Data Using OLAP Information Technology," in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), 2020, vol. 2, pp. 277–280.

[26] S. Shang, C. K. Leung, Y. Chen, and A. G. M. Pazdor, "Spatial data science of COVID-19 data," in 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2020, pp. 1370–1375.

[27] R. M. Harris, "Data Warehousing and Decision Support System Effectiveness Demonstrated in Service Recovery During COVID19 Health Pandemic," in 2020 14th International Conference on Open Source Systems and Technologies (ICOSST), 2020, pp. 1–5.

[28] U. Thange, V. K. Shukla, R. Punhani, and W. Grobbelaar, "Analyzing COVID-19 Dataset through Data Mining Tool 'Orange,'" in 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2021, pp. 198–203.

[29] A. S. Yadaw, Y. Li, S. Bose, R. Iyengar, S. Bunyavanich, and G. Pandey, "Clinical features of COVID-19 mortality: development and validation of a clinical prediction model," Lancet Digit. Heal., vol. 2, no. 10, pp. e516--e525, 2020.

[30] M. Al-Okaily, H. Alqudah, A. Matar, A. Lutfi, and A. Taamneh, "Dataset on the Acceptance of e-learning System among Universities Students' under the COVID-19 Pandemic Conditions," Data Br., vol. 32, p. 106176, 2020.

[31] J. Razjouyan et al., "Differences in COVID-19-related testing and healthcare utilization by race and ethnicity in the veterans health administration," J. racial Ethn. Heal. disparities, vol. 9, no. 2, pp. 519–526, 2022.

[32] L. Chouchana et al., "Association of antihypertensive agents with the risk of in-hospital death in patients with Covid-19," Cardiovasc. drugs Ther., vol. 36, no. 3, pp. 483–488, 2022.

[33] A. M. O'Hare et al., "Age differences in the association of comorbid burden with adverse outcomes in SARS-CoV-2," BMC Geriatr., vol. 21, no. 1, pp. 1–10, 2021.

[34] N. Hoertel et al., "Observational study of chlorpromazine in hospitalized patients with COVID-19," Clin. Drug Investig., vol. 41, no. 3, pp. 221–233, 2021.

[35] M. Fartoukh et al., "Seasonal burden of severe influenza virus infection in the critically ill patients, using the Assistance Publique-Hôpitaux de Paris clinical data warehouse: a pilot study," Ann. Intensive Care, vol. 11, no. 1, pp. 1–11, 2021.

[36] K. H. Seal, ; ; Jennifer K. Manuel; Natalie Purcell; A. Rani Elwy Beth, and DeRonne, "DEMOGRAPHIC, SOCIAL DETERMINANTS, AND CLINICAL FACTORS ASSOCIATED WITH COVID-19 VACCINATION COMPLETION AMONG 6.2 MILLION VETERANS IN THE VETERANS HEALTH ADMINISTRATION FROM JANUARY 1, 2020 TO DECEMBER 22, 2021.," in 2022 Annual Meeting of the Society of General Internal Medicine, 2022.

[37] M. S. J. Smirnova, L. W. L. Weisong, Robert;, K. J. Reisman, and ; Hong Yu; William Becke, "THE ASSOCIATION OF PRESCRIBED LONG-ACTING VERSUS SHORT-ACTING OPIOIDS AND MORTALITY

AMONG OLDER ADULTS," in 2022 Annual Meeting of the Society of General Internal Medicine, 22AD.

[38] G. Targato et al., "THE IMPACT OF COVID-19 PANDEMIC ON ONCOLOGY WORKLOAD IN AN ITALIAN REFERENCE CANCER CENTER," in MASCC Annual Meeting on Supportive Care in Cancer, 2021.

[39] J. H. Littell, J. Corcoran, and V. Pillai, Systematic reviews and meta-analysis. Oxford University Press, 2008.

[40] P. Hemingway and N. Brereton, "What is a systematic review', What is Series," Bandolier, April, 2009.

[41] A. M. O'Connor, K. M. Anderson, C. K. Goodell, and J. M. Sargeant, "Conducting systematic reviews of intervention questions I: writing the review protocol, formulating the question and searching the literature," Zoonoses Public Health, vol. 61, pp. 28–38, 2014.

[42] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[43] D. Gough, S. Oliver, and J. Thomas, An introduction to systematic reviews. Sage, 2017.

[44] T. S. Qaid, H. Mazaar, M. Y. H. Al-Shamri, M. S. Alqahtani, A. A. Raweh, and W. Alakwaa, "Hybrid deep-learning and machine-learning models for predicting COVID-19," Comput. Intell. Neurosci., vol. 2021, 2021.

[45] G. Garani, A. Chernov, I. Savvas, and M. Butakova, "A data warehouse approach for business intelligence," in 2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2019, pp. 70–75.

[46] C. K. Leung, Y. Chen, S. Shang, and D. Deng, "Big data science on COVID-19 data," in 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE), 2020, pp. 14–21.

[47] B. K. Seah and N. E. Selan, "Design and implementation of data warehouse with data model using survey-based services data," in Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014), 2014, pp. 58–64.

[48] A. Khalaf Hamoud, A. Salah Hashim, and W. Akeel Awadh, "CLINICAL DATA WAREHOUSE A REVIEW," Iraqi J. Comput. Informatics, vol. 44, no. 2, Dec. 2018.

[49] M. G. Seneviratne, T. Seto, D. W. Blayney, J. D. Brooks, and T. Hernandez-Boussard, "Architecture and implementation of a clinical research data warehouse for prostate cancer," eGEMs, vol. 6, no. 1, 2018.

[50] O. E.-S. Sheta and A. N. Eldeen, "Building a health care data warehouse for cancer diseases," arXiv Prepr. arXiv1211.4371, 2012.

[51] E. Roelofs, L. Persoon, S. Nijsten, W. Wiessler, A. Dekker, and P. Lambin, "Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial," Radiother. Oncol., vol. 108, no. 1, pp. 174–179, 2013.

[52] J. G. DeWitt and P. M. Hampton, "Development of a data warehouse at an academic health system: knowing a place for the first time," Acad. Med., vol. 80, no. 11, pp. 1019–1025, 2005.

[53] J. Sreemathy, S. Nisha, G. P. RM, and others, "Data integration in ETL using TALEND," in 2020 6th international conference on advanced computing and communication systems (ICACCS), 2020, pp. 1444–1448.

[54] J. Sreemathy, S. Priyadharshini, K. Radha, K. Sangeerna, and G. Nivetha, "Data validation in ETL using TALEND," in 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 1183–1186.

[55] B. Chen, X. Weng, B. Wang, and X. Hu, "Analysis and solution of data quality in data warehouse of Chinese materia medica," in 2009 4th International Conference on Computer Science & Education, 2009, pp. 823–827.

[56] A. Cuzzocrea, "SPPOLAP: computing privacy-preserving OLAP data cubes effectively and efficiently algorithms, complexity analysis and experimental evaluation," Procedia Comput. Sci., vol. 176, pp. 3831–3842, 2020.

[57] A. S. Chauhan et al., "Predictive Big Data Analytics for Service Requests: A Framework," Procedia Comput. Sci., vol. 198, pp. 102–111, 2022.

[58] C. Kamga, R. Tchamna, P. Vicuna, S. Mudigonda, and B. Moghimi, "An estimation of the effects of social distancing measures on transit vehicle capacity and operations," Transp. Res. Interdiscip. Perspect., vol. 10, p. 100398, 2021.

[59] D. B. Knox, E. L. Hirshberg, J. Orme, I. Peltan, and M. J. Lanspa, "Effect of COVID 19 pneumonia on hyperglycemia: Is it different from non COVID pneumonia?," Diabetes Metab. Syndr. Clin. Res. Rev., vol. 16, no. 2, p. 102407, 2022.

[60] A. Zakaria et al., "Determinants of all-cause in-hospital mortality among patients who presented with COVID-19 to a community teaching hospital in Michigan," Heliyon, vol. 7, no. 12, p. e08566, 2021.

[61] H. Tong, F. Aletta, A. Mitchell, T. Oberman, and J. Kang, "Increases in noise complaints during the COVID-19 lockdown in Spring 2020: A case study in Greater London, UK," Sci. Total Environ., vol. 785, p. 147213, 2021.

[62] B. Pérez, C. Castellanos, and D. Correal, "Predicting student drop-out rates using data mining techniques: A case study," in IEEE Colombian Conference on Applications in Computational Intelligence, 2018, pp. 111–125.