JSCDM

http://penerbit.uthm.edu.my/ojs/index.php/jscdm
e-ISSN : 2716-621X

Journal of Soft Computing and Data Mining

Analyses the Performance of Data Warehouse Architecture Types

Karwn Jameel Merceedi[1*], Abdulmajeed Adil Yazdeen[1], Abass Kh. Ibrahim[2], Maiwan Bahjat Abdulrazzaq[3], Mayyadah Ramiz Mahmood[3]

[1]Duhuk Polytechnic University,
Information Technology Management, Duhok, IRAQ

[2]University of Zakho,
Petroleum Engineering, Duhok, IRAQ

[3]University of Zakho,
Computer Science, Duhok, IRAQ

*Corresponding Author

DOI: https://doi.org/10.30880/jscdm.2022.03.01.005
Received 01 February 2022; Accepted 01 April 2022; Available online 01 June 2022

Abstract: The concept of storing historical data for retrieving them when needed has been conceived, and the idea was primitive to build repositories for historical data to store these data, despite the use of a specific technique for recovering these data from various storage modes. The data warehouse is the most reliable and widely used technology for scheduling, forecasting, and managing corporations. Also, it is concerned with the data storage facility that extensive collection of data. Data warehouses are called ancient modern techniques; since the early days of relational databases, the critical component of decision support, increasing focus for the database industry. Many commercial products and services are now available, and almost all of the primary database management system vendors provide them. When opposed to conventional online transaction processing applications, decision support puts slightly different demands on database technology. This paper analyzes the performance of the data warehouse architectures by studying and comparing many research works in this field. The study involves extracting, transforming, and loading the data from different recourses and the important characteristic of the architecture types. Furthermore, the tools and application service techniques used to build data warehouse architecture.

Keywords: Data warehouse, data warehouse component, data warehouse architecture

1. Introduction

The Data Warehouse system has two main architectures, the first is called the data flow architecture, and the second is called the device architecture. The data flow architecture is concerned with describing the storage data warehouse in the data stores as well as the structure in which this data are stored, also representing the sequence of the data flows from the source systems to the users. The physical configuration of the servers, network, applications, storage, and clients is referred to as device architecture [1].

Data Warehouse (DW) generalizes and consolidates data in multidimensional spaces, which construction entails data cleansing, integration, and transformation. Data can be viewed in many ways as a necessary phase in data mining, especially for visualization. Business executives may use data warehouse architectures and software to systematically coordinate, understand, and use their data to make strategic decisions. The data collected in a warehouse can be applied

*Corresponding author: karwan.jamil@dpu.edu.krd

to any of the following fields: Customer analysis, customer behaviors, and Operations analysis are all areas where production methods can be fine-tuned [2].

DW has become a critical piece of Information Technology (IT) today's world; it is one of the best technologies to consider if you want to make fast and precise decisions. Educational institutes, businesses, and corporations must advance their information record system to continue in a modest setting in planning, decision-making, and analytical processes. Organizations must increase the efficiency and effectiveness in controlling the flow of activities. [3].

When designing or building DW, architecture is a set of rules to follow. The use of architecture is needed for DW in a large organization that considers a complex task [4]. DW architecture can be represented as the overall structure of the data, communication, processing, and presentation for end-user computing within the enterprise. Using application tools, architectures, information services, and communication networks. The importance of the DW process led to the synthesizing of helpful knowledge for decision-making from distributed heterogeneous data sources. As a result, vendors acknowledge that DWs are not off-the-shelf products but must be custom-designed and customized to each customer's unique requirements [5].

Online transaction processing (OLTP) applications address an organization's operating data requirements, which are essential to the functioning of the enterprise on a day-to-day basis. However, they are unsuitable for sustaining decision-support inquiries or business questions that managers often encounter. Online analytical processing (OLAP) systems are ideally suited for analytics tasks such as aggregation, drill down, and data slicing and dicing. Data warehouses facilitate OLAP applications by storing and maintaining data in a multidimensional format. ETL tools are used to retrieve and load data into an OLAP warehouse from various OLTP data sources (including DB2, Oracle, SQL Server, and Flat files) [6].

DWHs module depends on the type of the architectures with other components to achieve the best accurate result through the data quality issues and data integration [7]. To primarily serve DWHs, data warehouse architectures (DWHAs) are built synchronously with DWH development. Several DWHAs are suggested to meet a variety of needs and circumstances, which can be just addressed as five different forms of DWHAs (Data Mart Bus, Hub-and-Spoke, Centralized, Independent Data Marts, and Federated) [8]. The most recent DWHA proposes the classification, but some DWHAs are still removed, such as Virtual DWHA, Big DWHA, and DWH with data lake architecture (DLA). These excluded DWHAs also appear regularly in literature and industry, solving unique problems in resolving major data issues and performing fast data analyses. At the same time, the DHWAs listed a few advantages in solving these problems, which would take more time and be less effective. As a result, it is worthwhile to discuss and describe them together to get the benefit of having more options to meet their needs, extensive data, and high performance [9].

All organizations can get the benefit of using DW capability on the old save data to help with decision-making by consolidating, collecting, researching, and analyzing data. The data that is properly unified and associated with a specific topic can get varied over time and provides the necessary and not volatile data collection support the data mining that allows the process management of an ecosystem than a specific product. The architectural design of an information system offers historical and current decision support data to business users, which is difficult to collect or access from an operational data store [10].

2. Data Warehouse of Architecture

DW architecture is a technique for specifying the overarching architecture of data communication, collection and display within an organization for end-client computing. - data warehouse is unique, but they all share some critical components. The DW framework can be built in three different ways. The number of levels in the architecture is used to classify these approaches: single-tier architecture, Two-tier architecture, and Three-tier architecture [11].

2.1 Single-tier Data Warehouse Architecture

Single-tier architecture is a relatively uncommon practice. The primary goal of this structure is to avoid redundancy by storing the least amount of data available. Its primary shortcoming is the lack of a section defining the analytical and transactional processing divisions [12].

2.2 Two-tier Data Warehouse Architecture

A two-tier architecture needs a staging area for all data sources before the data warehouse layer. By establishing a staging area between the origins and the storage repository, you will guarantee that all data loaded into the warehouse is appropriately cleansed and formatted. This technique may have certain drawbacks in terms of network coverage. You obviously will not expand it to handle further people. Data warehouse architecture with three levels. The three-tier architecture is the most popular for data warehouse systems. There are three stages of it:

- The warehouse's folder, which contains the cleansed and converted documents, is on the bottom tier.
- The framework layer provides an abstracted view of the database to the middle stage. It organizes the information in a manner that makes it easy to understand. An OLAP server that follows the ROLAP or MOLAP paradigm is used to achieve this.

- The top tier is where the user communicates with the information. It represents the front-end client layer. There are applications for reporting, querying, analysis, and data mining [12].

2.3 Best Practices of Data Warehouse Architecture

The following are the best practices of data warehouse architecture:

- Create dimensional, de-normalized, or hybrid Information retrieval models based on data warehouse models.
- Choose a specific approach to building data centers, choose one approach, select a methodology, such as top-down or bottom-up, and adhere to it.
- Always use an ETL tool to cleanse and convert data before It is loaded into a data warehouse.
- Create an automatic data cleaning process that uniformly cleans all data before loading.
- To ensure a smooth retrieval process, enable metadata to be shared between various data warehouse components.
- When transferring data from data stores to the data warehouse, ensure it is appropriately incorporated, not just consolidated. This will necessitate 3NF data model normalization [13].

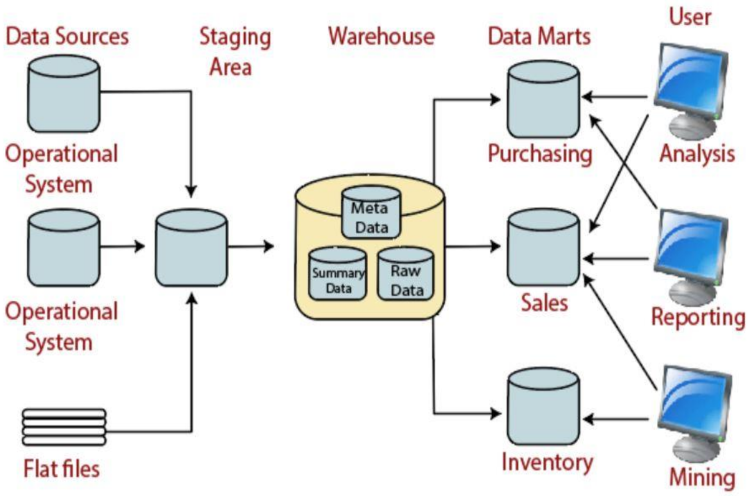

Fig. 1 - The architecture of data warehouse

3. Data Warehouse Architecture Classification

There are three types of data warehouse architectures, Single-layer, two-layer, and three-layer. The architecture's layer count determines the number of layers in a structure-oriented architecture. The second classifications are independent data mart infrastructure, bus, hub-and-spoke, clustered, and distributed architecture, in which the main layers are mixed in a multitude of ways.

3.1 Single-Layer Architecture

In practice, one-layer architecture is seldom used. The aim of removing data redundancies is to reduce the amount of data stored. The virtual data warehouse is implemented as a multidimensional view of operational data. The fundamental weakness of a single-layer architecture is its inability to differentiate between analytical and transactional data processing [14].

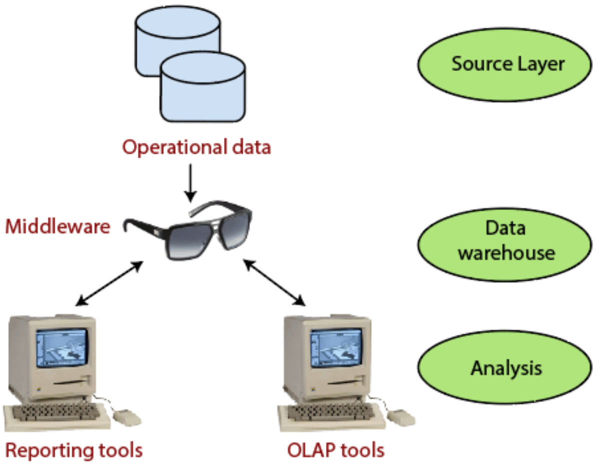

Fig. 2 - Single-layer architecture

3.2 Two-Layer Architecture

There are two distinct layers: the data source layer and a data warehouse layer. The four levels of data flow ate the source layer, data staging, data warehouse layer, and research layer [14]. To highlight the contrast between the two layers, it is referred to as a two-layer architecture. In contrast to single-layer architecture, there is a difference between analytical and transactional data processing [15].

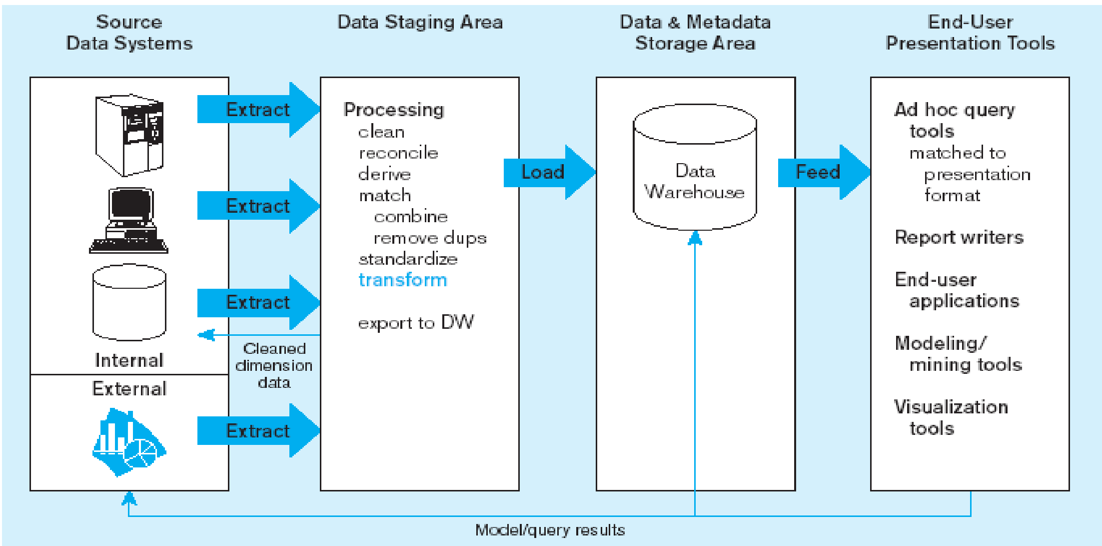

Fig. 3 - Two-layer architecture [15]

3.3 Three-Layer Architecture

Three layers are physically implemented in three layers: the source layer, the reconciled layer, and the data warehouse layer. After merging and cleansing source data, the reconciled data layer materializes operational data. Data is incorporated, reliable, and detailed after passing through the reconciled layer. This type of Data warehouse is shown in fig. 4 [16].

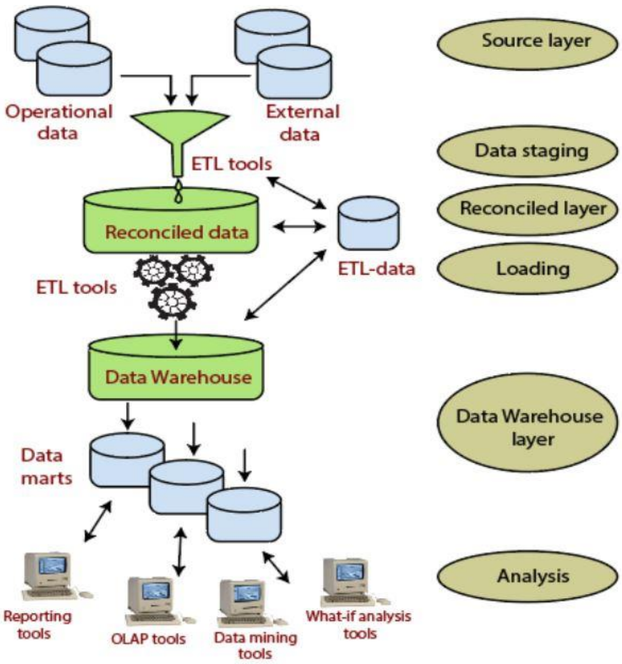

Fig. 4 - Three-layer architecture [16]

3.4 Independent Data Marts Architecture

The independent data mart architecture consists of different data marts planned and implemented separately and not incorporated (see fig. 5). Data marts are notorious for having inconsistencies in data descriptions and disparate dimensions and scales, rendering data analysis difficult. In order to achieve improved data integration and cross-reporting, this architecture is usually replaced by another [17].

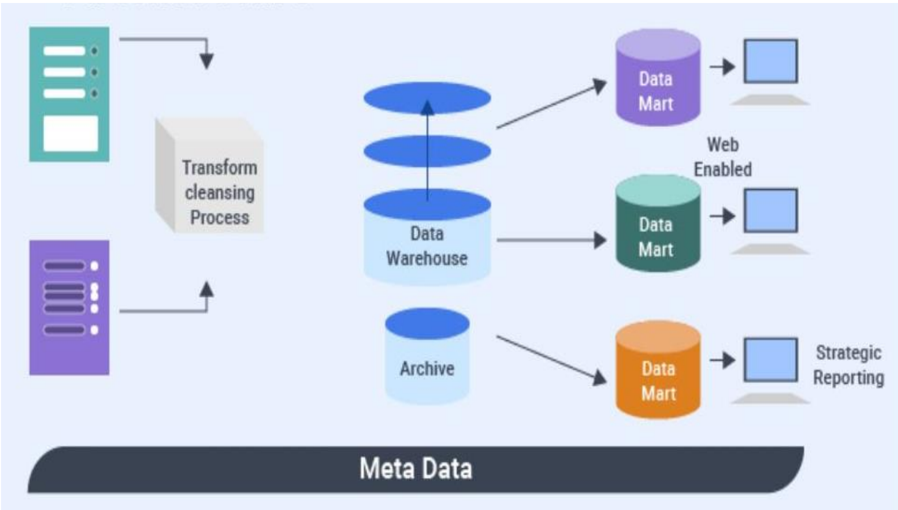

Fig. 5 - Independent data mart architecture [17]

3.5 Bus Architecture

Ralph Kimball recommends the bus architecture, which is identical to independent data mart architecture with one significant difference: data marts are logically interconnected, as shown in fig. 6. There is a knowledge perspective that spans the whole organization [18].

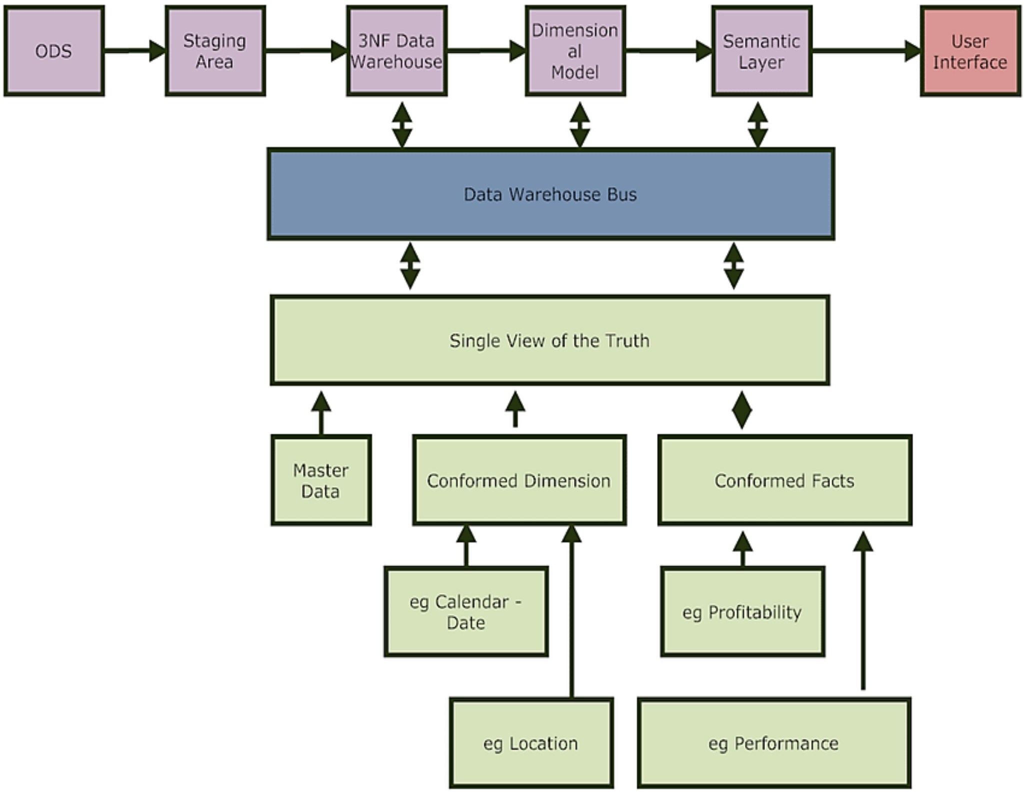

Fig. 6 - Bus architecture [18]

3.6 Hub-And-Spoke Architecture

Data sources, reconciled data, and data marts make up the hub-and-spoke architecture. The center, or enterprise data warehouse, is built from data marts known as spokes. A reconciled layer stores atomic, normalized data that feeds a series of data marts made up of summarized data in the multidimensional form. The focus of hub-and-spoke architecture is scalability, extensibility, and retrieving large volumes of data [19].

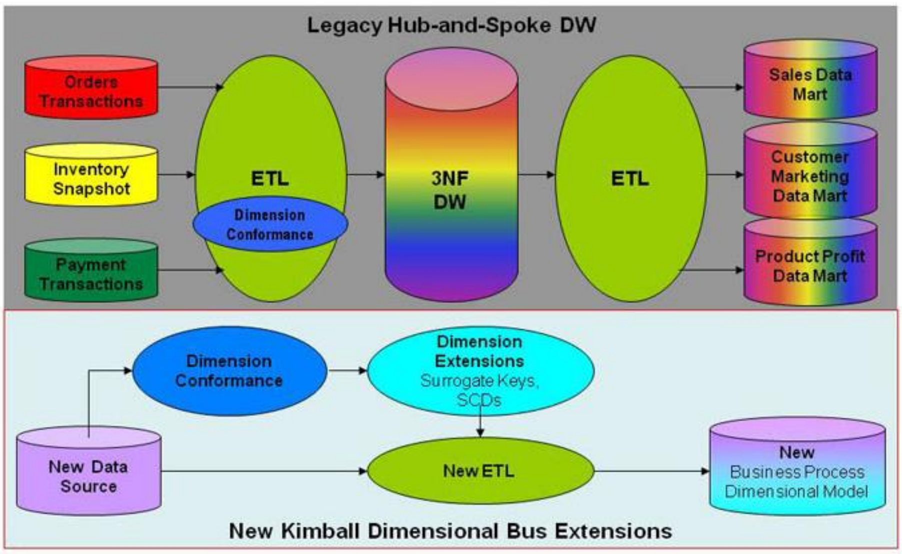

Fig. 7 - Hub and spock architecture [19]

3.7 Centralized Architecture

Bill Inmon recommends centralized architecture, as shown in fig 8. This architecture can be considered a variant of hub-and-spoke architecture, except that there are no dependent data marts. It comprises a single unified data warehouse with integrated data and data marts [20].

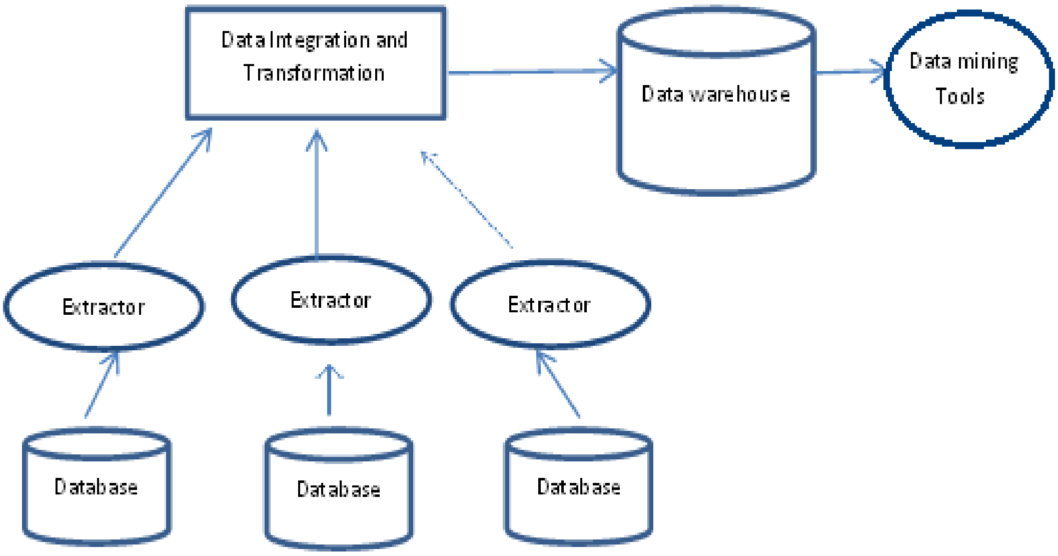

Fig. 8 - Centralized architecture [20]

3.8 Distributed Architecture

The distributed architecture is often used in dynamic scenarios where data warehouses can be combined to provide a single solution. Using joint keys, global metadata, distributed queries, and other methods, each data warehouse/data mart is logically or physically incorporated into this architecture [21].

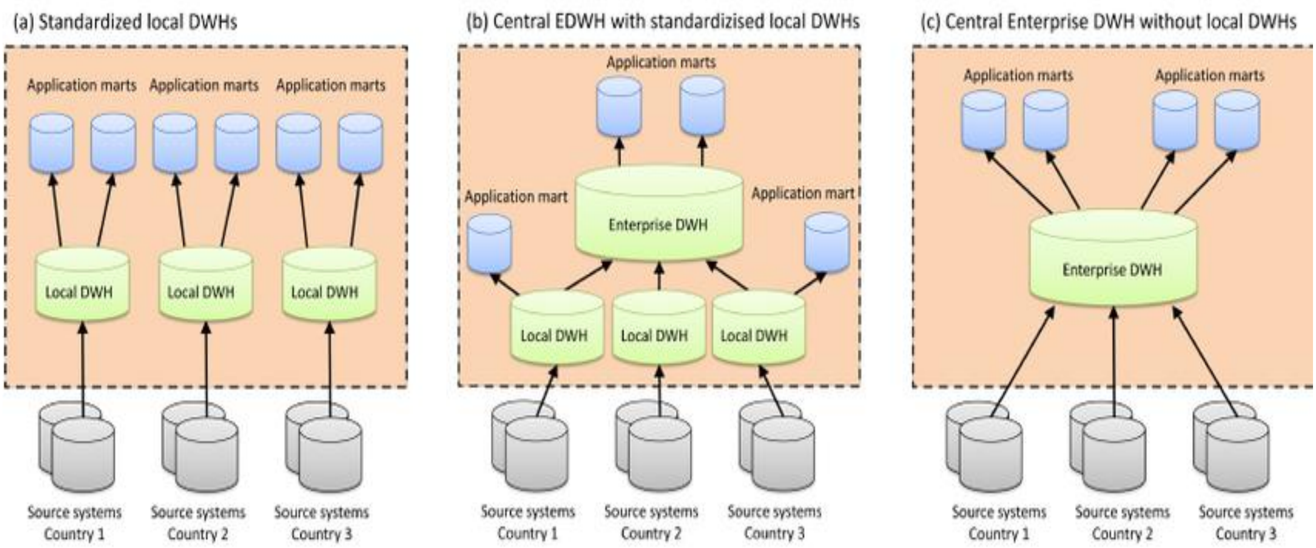

Fig. 9 - Distributed architecture [21]

4. Data Warehouse Design Consists of Six Main Vital Components

4.1 Data Warehouse Database

The most critical aspect of a data warehousing architecture is a databank, which preserves and manages all organizational data for monitoring. This means can choose the database to be used in the warehouse to do the occupation of storing records. [22]. The four database types are:

- The classical one that familiar with row-oriented databases, which are the most common relational databases like Microsoft SQL Server, SAP, Oracle, and IBM DB2.

- Analytics databases are designed specifically for storing and processing data in order to facilitate and maintain analytics.
- While data warehouse systems are not technically a storage database, some vendors also provide both software and hardware for data storage, like SAP Hana, Oracle Exadata, and IBM Netezza.
- Cloud-based applications are stored and accessible in the cloud, like Amazon Redshift, Microsoft Azure SQL, and Google Big Query.

4.2 Extraction, Transformation, and Loading Tools (ETL)

The architecture of a data warehouse relies heavily on ETL software. These tools aid in extracting data from various sources, transforming that data into a usable format, and loading that data through a warehouse of data [23]. Depending on which ETL tool you use, you will be able to:

- The time spent extracting data.
- The techniques used to extract data.
- Filling in incomplete data.
- The types of transformations used and how easy they are to implement.
- Defining data authentication and cleaning market rules in order to boost end-product analytics.

4.3 Metadata

Metadata defines the data warehouse and provides a framework for data in the data warehouse architecture. It facilitates the development, maintenance, management, and usage of data warehouses [23]. It is classified into two types:

- Technical Metadata consists of information that engineers and administrators may use to perform warehouse development and administration activities.
- Business Metadata is the material that depicts the data in the warehouse as a whole. For businesses and technical teams to understand and convert data in the warehouse to content, metadata is essential.

4.4 Data Warehouse Access Tools

A data warehouse is constructed upon the foundation of a database or set of databases. If they employ database administrators, data warehouse organizations are unable to handle databases without using the software. This is not true for all company divisions. As a result, they depend on several no-code data warehousing tools, including the following:

- Users may generate operational reports using query and reporting software in spreadsheets, forecasts, or immersive visualizations for analysis.
- Application development tools simplify the process of creating customized documentation and presenting them in context-sensitive interpretations.
- Data warehousing data mining applications simplify finding arrays and connections in massive amounts of data using cutting-edge statistical simulation techniques.
- OLAP solutions simplify the process of developing a multidimensional data repository and analyzing business data from a range of perspectives [22].

4.5 Data warehouse Bus

It assumes the use of a data mart and defines data flow within a data warehousing bus architecture. A data mart is a specialized form of access level that enables users to gain access to data. It is used to segment data created for a particular user population [24].

4.6 Data Warehouse Reporting Layer

End-users can access the business intelligence portal or the business intelligence network architecture through the data warehouse's reporting layer. The reporting layer is intended to be a hub for data visualization, report generation, and information extraction [24].

5. Literature review

The work of Solodovnikova et al. [25] has mentioned the key contribution of a data warehouse architecture that allows various analytical tasks, such as OLAP-like analysis, to be performed on big data loaded from multiple heterogeneous data sources with varying latency on one side. On the other hand, the proposed architecture can process changes in data sources as well as changing research requirements. Components of the architecture, relevant metadata, and examples of changes supported by the architecture and their implementations within the architecture have all been identified. The study's main goal is to propose a significant data warehousing architecture that can adapt to user needs and requirements and changes in the underlying data sources, automatically or semi-automatically.

Agapito et al. [26] proposed a novel method for developing a data warehouse from sparse data sets, such as Italy's COVID-19 data. We developed an integrated methodology for enriching Italian COVID-19 data with regional air quality and climate data. Additionally, because the data sources used to build our data warehouse model were heterogeneous, we needed to develop multiple merger models to manage, explain (eliminate ambiguity), and merge data. The data warehouse registry stores and renders available merged models.

Bouaziz et al. [27] presented the results of a survey on real-time data warehousing and the challenges and strategies that can be used to get closer to this RTDW. The report then discusses how the ETL process has been improved to fit real-time data warehouse systems and how the integration data in the RTDW is managed.

Farooqui et al. [28] proposed using data mining to build a data archive for a medical information system. The data center is built on top of a transformation of an operating database into an informational database. A data analyst can analyze data and make decisions based on it. The platform is designed to serve as a database for the medical data system. Saddad et al. [29] proposed a novel data warehouse model called the Lake Data Warehouse Architecture. The Lake Data Warehouse Architecture is a hybrid data warehouse that uses Hadoop Framework and Apache Spark to merge traditional data warehouse strategies with Hadoop Framework. It puts the average DW at ease. It uses big data technology and techniques to complement and extend the existing architecture (Hadoop, Apache Spark, Data Lake, and Delta Lake). It will also improve scalability and reduce the time it takes to construct traditional data warehouse architectures.

Yah et al. [30] presented a Data Warehouse architecture for an Electronic Manufacturing Company. The spiral approach and the Kimball method are used to build the data warehouse. Oracle Database XE is used to execute the template. The User Acceptance Test and the Integrity Check are included in this study. The successful outcome demonstrates that the data warehouse meets the users' needs and, as a result, assists them in identifying basic patterns and trends. However, further research is needed to enhance the output outcome. Integrity review and User Acceptance test are included. According to this report, the data warehouse meets the user's requirements. In Hamoud et al. [31], the goal of providing a concise description of CDWs, including their features, requirements, data sources, the collect, convert, and load (ETL) process, protection and privacy concerns, design strategy, architecture, and the complexities and difficulties that come with successfully implementing a CDW.

The Gunawan et al. [32] model is PT Autochem Industry, a HealthCare retail company. The company has 24 branches in Indonesia's major cities. When 150 advertisers are responsible for more than 100 items, an issue occurs. In a healthcare product, marketing has some difficulty recognizing 100 items with their uniqueness and marketing strategy. Marketing will be more effective and profitable in managing the market for healthcare goods by using Knowledge Management, and companies can be more successful in the globalization period.

In Kortüm et al. [33], a DW was effectively implemented in an ophthalmologic learning setting to encourage and facilitate research by increasing EMR and measurement evidence. The procedure for defining research subjects has been simplified. To build a near-real-time data warehouse (DW) in an academic ophthalmology center to allow scientific use of the growing amount of digital data provided by EMRs and diagnostic devices.

Puppala et al. [34] suggested a rigorous combination of technology and best practices, such as technological de-identification of data, restricted data access, and security controls in the underlying technical platforms, is the best way to protect patient privacy. Our findings indicate that the proposed security model makes data security breaches and unauthorized access to protected patient health data highly unlikely. The findings indicate that data security breaches and unauthorized access to protected patient health information are highly unlikely.

Although natural language processing engines such as the clinical Text Analysis and Knowledge Extraction System have been proposed for processing study notes, the paper states that maximizing their performance for a clinical data warehouse remains a challenge. The clinical Text Analysis and Knowledge Extraction System have been used by Afshar et al. [35] to develop a high-throughput natural language processing infrastructure and to show how statistical models can be used. The NLP architecture processed 83 867 802 clinical documents in 13.33 days and produced 37 721 886 606 CUIs from eight formal medical vocabularies. Using 30 parallel instances of the clinical text analysis and knowledge extraction system, the design could handle more than 500 000 documents per hour. Ten of them are reserved for records more than 20 000 bytes in size.

Aftab et al. [36] offered a thorough examination of big data, big data challenges, analytics, and warehouse. It also describes the need for big data and a data warehouse. It also delves into the applications that sustain big data, data warehouses, and the problems that come with them Challenges. Maji et al. [37] aim to develop a framework for storing CDR data in a suitable Data Warehouse (DW) schema and analyzing it using OLAP tools in order to gain a better understanding of prepaid customers' use spending and willingness to accept marketing offers. Customers have been correctly segmented based on their consumption habits. They have been classified to facilitate the distribution of various tailored marketing offers and benefits to retain and acquire new customers.

Gerbel et al. [38] discussed the Enterprise Clinical Research Data Warehouse (ECRDW) at Hannover Medical School (MHH). Since 2013, ECRDW has been used at the MHH as an intra-discipline binary forum for research-related queries. It was developed using Microsoft SQL Server Data Warehouse and Business Intelligence technologies. ECRDW integrates disparate data sets incrementally and contains over 2,1 million distinct patients and over 500 million unique data points (8/2018). (diagnoses, laboratory reports, vital signs, medical history, and metadata associated with related data, such as biospecimens or images).

Poenaru1 et al. [39] mention some of the research issues associated with the storage of medical data in Health Information Systems (HIS), including sophisticated data modeling functions, advanced classification processes, and the integration of highly complex data. It demonstrates how data warehousing functionality can support this area. Additionally, a case study is provided that illustrates how to configure a data warehouse designed using multi-agent technology and integrating data from disparate sources in a healthcare unit.

The query cache method was first used in Tiwari et al. [40] study to increase the speed and efficiency of ETL within the data warehouse by reducing response time. This method's primary goal is to save queries and their associated output. If any user is faced with the same database, query cache memory will generate the output. Later, in data mining, we used the coalition rule based on the association rule and provided results for error detection in various data. The results were better because there were fewer errors and a quick answer time. Xavier Palacios-P et al. [41] described a method for implementing a business intelligence system that includes data treatment, analysis, and presentation and can react to any problem that arises in a school. Our structure provides a realistic way for universities to start their BI journey by crystallizing their most critical processes and identifying the knowledge needed to support these processes.

Mu et al. [42] proposed a data warehouse paradigm for a customer service business environment with a strong emphasis on data integration. It incorporates the customer service center's business processes to conform to the OLAP standards defined in the 95598-customer service structure for business data, thus increasing its service efficiency and decision-making capabilities. Venditti et al. [43] suggest a systematic method for assisting students and engineers in selecting a data warehouse architecture that considers the needs of the particular context in which the data warehouse will be used. This method necessitates a prior assessment of the significance of the parameters defining the various architectures in the reference sense. Then, for each architecture, a global value is specified, allowing us to compare them. We also present an empirical assessment of the proposed approach's efficacy.

Gavrilov et al. [44] present a novel model for a healthcare data warehouse based on a restructured Extract–Transform–Load mechanism. Describe a gateway architecture that provides a comprehensive set of interoperability tools that allows national health platforms to provide cross-border health information networks that conform to the European Patients Smart Open Services standard. The suggested approach combines technological and operational interoperability by linking the Health Level Seven standard to the Open National Contact Points network, resulting in a versatile, flexible, and interoperable architecture.

6. Discussion and Results

Table 1 summarizes many projects' most used distributed data warehouse and hybrid data warehouse architectures. Due to their simplicity in dealing with viruses type of data sources and dealing with a huge number of stores data, restoring data, and back up from archive and refreshing, most the researcher depends on integrated and enterprise data warehouse architecture. The table provides information on the type of data warehouse architecture, the tools used to program it, and the application service projects employed.

Table 1 - Types of data warehouse architecture with application and services

Ref.	DWH Type	Tool	Applications and services	Description
[25]	Distributed data warehouse	SQL	management big data	A data warehousing architecture for big data capable of adapting automatically or semi-automatically to user requirements and changes in the underlying data.
[26]	Five independent and cooperating levels	open-sources BIRT	COVID-WAREHOUSE supports OLAP	Using OLAP for analysis of data set discover combinations of pollution and climate. The multidimensional model is known as a snowflake.
[27]	Real-time Data Warehouse.	(SQL) to RTDW	Business Web Services Based Real-Time Data Warehouse	Means of Online Analytical Processing (OLAP) techniques. Construction of an RTDW based on Web services. RTDW High number of users accessing. Alterations to the ETL method to accommodate a real-time data warehouse
[28]	DW model, which is based upon data mining techniques, different stages	Oracle10g	proposed the Medical Information System	Implementing a data warehouse for a medical information system would be used in sixteen different columns, using modern methodologies. Therefore, make use of Techniques for data processing
[29]	DW hybrid (integrated traditional DW technique, Hadoop and Apache Spark)	data lake as JSON, XML files. Use Spark SQL and Python	A case study using a dataset of IoT sensors. Data is a simulation of data taken from a heart rate monitor.	The Lake Data Warehouse Design enables complementary big data technology and tools to sustain and expand current architectures (Hadoop, Apache Spark, Data Lake, and Delta Lake).
[30]	Kimball's Method. Multidimensional model	Oracle Database XE	Electronic Manufacturing Company	The performance result shows that the data warehouse matches the user needs and eventually helps them discover critical patterns and trends. This study obtained seven-dimension tables and eight fact tables, and 11 draft reports.
[31]	Distributed DW (data integration clinical and IT team)	Different Open-source tools	clinical information.	Contribute to the resolution of CDW-related issues and make recommendations for adopting a good CDW. Utilize therapeutic ODS.
[32]	data integration, OLAP services	Pentaho. (Spoon)	Knowledge Management in Organization	Implement knowledge management from online transaction processing (OLTP) data or online-based enterprise resource planning (ERP) data.
[33]	-	SQL database And QlikView for visualization.	Building an ophthalmologic data warehouse with electronic health records	To create a near-real-time data warehouse (DW) in an academic ophthalmology centre to make scientific use of the growing amount of digital data generated by electronic medical records (EMR) and diagnostic devices.

[34]	enterprise data warehouse	SQL Server 2008.	Clinical health data Management	METEOR environment, including the information model of data security and privacy management. In an enterprise-wide clinical data warehouse and analytic environment,
[35]	DW Workflow Architecture	CDW based on Microsoft SQL server (Clarity)	clinical data warehouse	The high-throughput NLP architecture used by the health system could serve as a blueprint for large-scale clinical trials using a CUI-based approach.
[36]	Enterprise Data Warehouse	SQL based Hadoop	For a set of applications and workflows	Analytics over big data, Data Warehouse with Data Lake (offload the processing)
[37]	the relational database based (ROLAP) DW schema of snowflake	Different applications	Commercialization (customers usage patterns as well as calling behaviors based on CDR data).	CDR Analysis for Retaining and Acquiring Customers Via Targeted Marketing. Additionally, a DW schema is specified to create OLAP cubes for easy visualization and analysis. Proposed data warehouse schema for analyzing voice call expenditures.
[38]	Enterprise Clinical Research Data Warehouse (ECRDW)	Microsoft SQL Server	Hannover Medical School's (MHH) Clinical Research Data Warehouse (ECRDW) stores patient data.	Due to the amount of data and various clinical information systems, reusing routine healthcare data for research purposes is problematic.
[39]	Multi-agent Data Marts, CDW architecture	NewSQL or NoSQL.	Clinical Data Warehouse for health information system	Medical Data Storage. A data warehouse built with multi-agent technology that collects data from a healthcare unit's disparate sources.
[40]	Enterprise distributed	MySQL	Improved Performance of Data Warehouse	Enhancing data warehouse processes use ETL inside the data warehouse by reducing response time and fetching the data from different sources.
[41]	data warehouse using Kimball's Method and data mining technics	MySQL	Business Intelligence analyzing Educational Data	A business intelligence system that covers data treatment, analysis, and presentation can respond to any problem that arises in a school and many clarifications. Data mining algorithms are used in databases that store socioeconomic and academic data about students.
[42]	integration of multi-heterogeneous data	data warehouse and statistical analysis tools	customer service business	The future advantage of data drives workflow optimization and increases the business center's service quality and performance. Using the 95598-customer-service system's OLAP specifications for business data.
[43]	Dependent and Independent data marts architecture.	Context in which the data warehouse	systematic approach to individuate the optimal architecture of the data warehouse.	Defined a simple systematic approach to obtaining a weight for each data warehouse and helpful architecture to compare them. Weight assignments take place depending on the context in which the data warehouse will be used.
[44]	Web services and, service-oriented architectures	(SQL) Server	Healthcare data warehouse system	have proposed a novel concept for the staging area of a healthcare data warehouse that facilitates cross-border data sharing by servicing national Open NCPs.

7. Conclusion

The automatic decision support system's foundation is data warehousing. While much research has been done in the last decade, there are still many topics that need to be addressed in the future. One of the most common study topics today is a concern with managing a huge amount of data by using data warehouse tools that classified them logically. The data warehouse architecture is often separated logically, and a standard model of the architecture is provided. Data cleaning, data optimization, management, and effective query processing are some of the key research fields that have been examined further. Identified extensive data warehousing study areas and stuff to do in the future to get the most out of our data warehousing. It provides an overview of data warehousing architecture, focusing on the latest criteria. It specifies back-end tools for processing, cleaning, and loading data into a data warehouse, as well as front-end client tools for querying and data analysis, metadata management, and warehouse management. This paper also included different types of data warehouse tools and application service techniques used to build data warehouse architecture. The analyses of the performance show that this field is very active in business intelligence, healthcare, and many other fields.

Acknowledgement

The author would like to acknowledge the Duhuk Polytechnic University, Information Technology Management, University of Zakho, Petroleum Engineering and University of Zakho, Computer Science, Duhok, Iraq.

References

[1] X. Huang, "Data Warehouse Architecture," Evol. Appl. Domains Data Warehouse. Min., pp. 1–15, 2010, DOI: 10.4018/978-1-60566-816-1.ch001.

[2] V. S, M. Srinath, A. C. Kumar, and N. A.S, "Data Warehousing Architecture and Pre-Processing," Ijarcce, vol. 6, no. 5, pp. 13–18, 2017, DOI: 10.17148/ijarcce.2017.6503.

[3] W. Tebourski, W. Ben Abdessalem Kara, and H. Ben Ghezela, "New data warehouse designing approach based on principal component analysis," 2014 IEEE/ACIS 15th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2014 - Proc., 2014, DOI: 10.1109/SNPD.2014.6888696.

[4] I. Inuwa and E. J. Garba, “An Improved Data Warehouse Architecture for SPGS, MAUTECH, Yola, Nigeria,” West African J. Ind. Acad. Res., vol. 14, no. 1, pp. 39-48–48, 2015.
[5] N. Elgamal and A. El-bastawissy, “An Architecture-Oriented Data Warehouse Testing Approach,” no. March, 2016.
[6] T. Oketunji and O. Omodara, “Design of Data Warehouse and Business Intelligence System,” Master Thesis, no. June, 2011.
[7] Yang, Q., Ge, M., and Helfert, M. (2017). Data quality problems in tpc-di based data integration processes. In International Conference on Enterprise Information Systems (pp. 57-73). Springer, Cham.
[8] Blaˇzi´c, G., Poˇsˇci´c, P., and Jakˇsi´c, D. (2017). Data warehouse architecture classification. In Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2017 40th International Convention on (pp. 1491-1495). IEEE.
[9] Inmon, B. (2016). Data Lake Architecture: Designing the Data Lake and Avoiding the Garbage Dump. Technics.
[10] S.-L. Peng, L. Hoang, S. G. Suseendran, and D. Balaganesh, Lecture Notes in Networks and Systems 118 Intelligent Computing and Innovation on Data Science. 2019.
[11] R. R. Nadikattu, “Data Warehouse Architecture – Leading the next generation Data Science,” Int. J. Comput. Trends Technol., vol. 67, no. 9, pp. 78–80, 2020, doi: 10.14445/22312803/ijctt-v67i9p113.
[12] I. R. H. T. Tangkawarow, J. P. A. Runtuwene, F. I. Sangkop, and L. V. F. Ngantung, “Three Tier-Level Architecture Data Warehouse Design of Civil Servant Data in Minahasa Regency,” IOP Conf. Ser. Mater. Sci. Eng., vol. 306, no. 1, 2018, doi: 10.1088/1757-899X/306/1/012061.
[13] M. A. Farnum et al., "A dimensional warehouse for integrating operational data from clinical trials," Database, vol. 2019, no. 1, pp. 1–13, 2019, DOI: 10.1093/database/baz039.
[14] H. Lv, L. Zhou, and Y. Zhao, "Classification of data granularity in data warehouse," Proc. - 9th Int. Conf. Intell. Human-Machine Syst. Cybern. MSC 2017, vol. 2, pp. 118–122, 2017, DOI: 10.1109/IHMSC.2017.142.
[15] Azwa Abdul Aziz, J.A.J., Hasni Hassan,wan mohd rizhan wan idris, <two00.pdf>. 2014. 2.
[16] Sebaa, A., et al., Medical Big Data Warehouse: Architecture and System Design, a Case Study: Improving Healthcare Resources Distribution. J Med Syst, 2018. 42(4): p. 59.
[17] Hamoud, A.K., et al., Implementing data-driven decision support system based on independent educational data mart. International Journal of Electrical and Computer Engineering (IJECE), 2021. 11(6).
[18] Yang, Q., M. Ge, and M. Helfert, Developing Reliable Taxonomic Features for Data Warehouse Architectures, in 2020 IEEE 22nd Conference on Business Informatics (CBI). 2020. p. 241-249.
[19] Helfert, M., M. Ge, and Q. Yang, Analysis of Data Warehouse Architectures: Modeling and Classification, in Proceedings of the 21st International Conference on Enterprise Information Systems. 2019. p. 604-611.
[20] Allae Erraissi, A.T., Digging into Hadoop-based Big Data Architectures. International Journal of Computer Science Issues, 2017. 14(6): p. 52-59.
[21] Amelec Viloriaa*, D.N.R., Omar Bonerge Pineda Lezamac, <1-s2.0-S1877050919306477-main.pdf>. 2019.
[22] A. Mathematics, "ijpam.EU," vol. 116, no. 13, pp. 273–275, 2017.
[23] I. Singh, “Study of Data Warehouse Architecture,” Int. J. Eng. Sci. Res. Technol., vol. 2, no. 7, pp. 1675–1680, 2013.
[24] R. Kimball, “Kimball Dimensional Modeling Techniques,” pp. 1–24, 2013.
[25] D. Solodovnikova and L. Niedrite, "Towards a Data Warehouse Architecture for Managing Big Data Evolution," DATA 2018 - Proc. 7th Int. Conf. Data Sci. Technol. Appl., no. Data, pp. 63–70, 2018, DOI: 10.5220/0006886100630070.
[26] 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, pp. 1–22, 2020, DOI:
[27] S. Bouaziz, A. Nabi, and F. Gargouri, "From traditional data warehouse to real-time data warehouse," Adv. Intell. Syst. Comput., vol. 557, no. February, pp. 467–477, 2017, DOI: 10.1007/978-3-319-53480-0_46.
[28] N. A. Farooqui and R. Mehra, "Design of a data warehouse for medical information system using data mining techniques," PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput., no. July 2019, pp. 199–203, 2018, DOI: 10.1109/PDGC.2018.8745864.
[29] E. Saddad, A. El-Bastawissy, H. M. O. Mokhtar, and M. Hazman, "Lake data warehouse architecture for big data solutions," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 8, pp. 417–424, 2020, DOI: 10.14569/IJACSA.2020.0110854.
[30] F. yah, M. Mustikasari, and I. Bastian, "Data Warehouse Design for Electronic Manufacturing Company," Int. J. Comput. Trends Technol., vol. 56, no. 1, pp. 47–53, 2018, DOI: 10.14445/22312803/ijctt-v56p107.
[31] A. Khalaf Hamoud, A. Salah Hashim, and W. Akeel Awadh, "Clinical Data Warehouse a Review," Iraqi J. Comput. Informatics, vol. 44, no. 2, pp. 1–11, 2018, DOI: 10.25195/2017/4424.
[32] A. Gunawan and S. G. Kurnia, "Knowledge Management Study in Data Warehouse," Proc. 2018 Int. Conf. Inf. Manag. Technol. ICIMTech 2018, no. September, pp. 1–5, 2018, DOI: 10.1109/ICIMTech.2018.8528121
[33] K. U. Kortüm et al., "Using Electronic Health Records to Build an Ophthalmologic Data Warehouse and Visualize Patients' Data," Am. J. Ophthalmol., vol. 178, pp. 84–93, 2017, DOI: 10.1016/j.ajo.2017.03.026.

[34] M. Puppala, T. He, X. Yu, S. Chen, R. Ogunti, and S. T. C. Wong, "Data security and privacy management in healthcare applications and clinical data warehouse environment," 3rd IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2016, no. December 2018, pp. 5–8, 2016, DOI: 10.1109/BHI.2016.7455821.
[35] M. Afshar et al., "Development and application of a high throughput natural language processing architecture to convert all clinical documents in a clinical data warehouse into standardized medical vocabularies," J. Am. Med. Informatics Assoc., vol. 26, no. 11, pp. 1364–1369, 2019, DOI: 10.1093/jamia/ocz068.
[36] U. Aftab and G. F. Siddiqui, "Big Data Augmentation with Data Warehouse: A Survey," Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, pp. 2785–2794, 2019, DOI: 10.1109/BigData.2018.8622206.
[37] G. Maji and S. Sen, "Data warehouse based analysis on CDR to retain and acquire customers by targeted marketing," 2016 5th Int. Conf. Reliab. Infocom Technol. Optim. ICRITO 2016 Trends Futur. Dir., no. February 2018, pp. 221–227, 2016, DOI: 10.1109/ICRITO.2016.7784955.
[38] S. Gerbel, H. Laser, N. Schönfeld, and T. Rassmann, "The Hannover medical school-enterprise clinical research data warehouse: 5 years of experience," Lect. Notes Comput. Sci. (including Subsea. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11371 LNBI, pp. 182–194, 2019, DOI: 10.1007/978-3-030-06016-9_17.
[39] Cristina Elena Poenaru, Daniel Merezeanu, Radu Dobrescu, Eugenie Posdarascu, Advanced Solutions for Medical Information Storing: Clinical Data Warehouse 2018.
[40] P. Tiwari, A. C. Mishra, S. Kumar, V. Kumar, and B. Terfa, "Improved performance of data warehouse," Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017, no. Icicct, pp. 94–99, 2017, DOI: 10.1109/ICICCT.2017.7975167.
[41] W. Villegas-Ch, X. Palacios-Pacheco, and S. Luján-Mora, "A business intelligence framework for analyzing educational data," Sustain., vol. 12, no. 14, pp. 1–21, 2020, DOI: 10.3390/su12145745.
[42] S. Mu, Q. Zhu, Y. Zhang, and Y. An, "Data warehouse dimensional modeling for customer service business," MATEC Web Conf., vol. 309, p. 05010, 2020, DOI: 10.1051/matecconf/202030905010.
[43] A. Venditti and F. Fasano, "A systematic approach to choose the data warehouse architecture," ICISSP 2019 - Proc. 5th Int. Conf. Inf. Syst. Secure. Priv., no. Icissp, pp. 711–718, 2019, DOI: 10.5220/0007700307110718.
[44] G. Gavrilov, E. Vlahu-Gjorgievska, and V. Trajkovik, "Healthcare data warehouse system supporting cross-border interoperability," Health Informatics J., vol. 26, no. 2, pp. 1321–1332, 2020, DOI: 10.1177/1460458219876793