INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 9, Issue 2 - V9I2-1224)
Available online at: https://www.ijariit.com

Review on: Modern Data Warehouse & how is it accelerating digital transformation

Ashish Dibouliya
ashish.dibouliya@gmail.com
Webster Bank, Stamford CT USA

ABSTRACT

Modern data warehousing is the practice of storing, organizing, and analyzing large volumes of data generated by modern data sources such as cloud-based applications, social media platforms, mobile devices, and the Internet of Things (IoT). This approach to data management has gained increasing importance in recent years as organizations seek to gain insights from their data to drive business value and improve decision-making. Modern data warehousing solutions offer a range of advanced features, including scalability, cloud-based architecture, real-time data processing, data integration, and data visualization. These features enable organizations to store and analyze large volumes of structured, semi-structured, and unstructured data from a wide range of sources, providing a single source of truth for decision-making. Overall, modern data warehousing has become an essential component of data-driven decision-making for organizations across a wide range of industries. By leveraging modern data warehousing solutions, organizations can gain valuable insights from their data, improve their competitive position, and drive business success. In this paper we review on modern data warehousing key concept, technology and difference about traditional data warehousing and modern data warehousing.

Keywords *Business Intelligence (BI), Data Warehousing (DW), Data Lake, Modern Data Warehouse, Data Warehouse.*

1. INTRODUCTION

Data warehousing is a technology that has been around for several decades, but it has continued to evolve over time to keep up with the ever-increasing volume and complexity of data generated by modern organizations. In recent years, modern data warehousing has emerged as a more flexible and scalable solution that can handle a wide range of data types and workloads. In this review paper, we will discuss the key characteristics of modern data warehousing, its architecture, and the various tools and technologies used in the modern data warehousing environment.

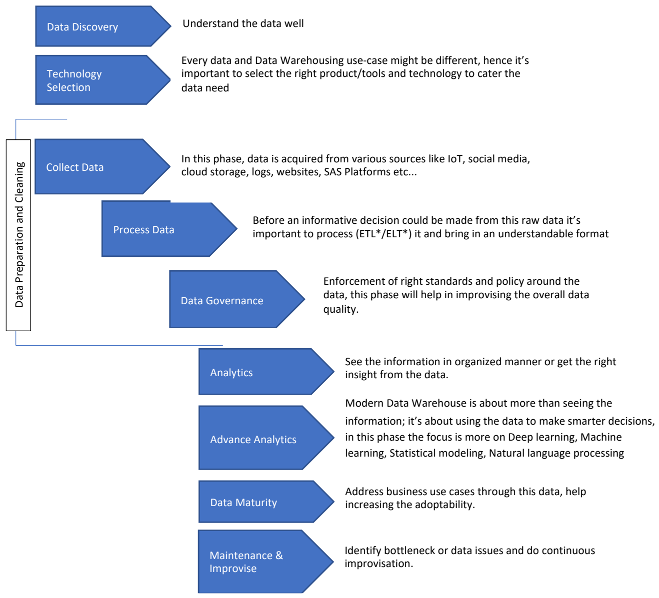

Figure: 1 different components of a Data Warehousing.

Key Characteristics of Modern Data Warehousing

Cloud-Based Infrastructure: One of the key features of modern data warehousing is its cloud-based infrastructure. Cloud data warehouses have become increasingly popular due to their scalability, flexibility, and cost-effectiveness. With cloud data warehouses, organizations can quickly and easily scale up or down as needed to accommodate changes in data volume and processing requirements.
Multi-Tenant Architecture: Modern data warehousing also typically employs a multi-tenant architecture, which allows multiple users to access and work with the same data warehouse simultaneously. This makes it easier for organizations to share data and collaborate on analytics projects.
Support for Structured and Unstructured Data: Modern data warehousing is designed to support both structured and unstructured data, including text, images, audio, and video. This makes it easier for organizations to integrate and analyze data from a wide range of sources.
Advanced Analytics Capabilities: Modern data warehousing also typically includes advanced analytics capabilities, such as machine learning, natural language processing, and predictive analytics. These tools allow organizations to gain deeper insights from their data and make more informed business decisions.

Modern Data Warehousing Architecture

Modern data warehousing typically consists of three layers: the data storage layer, the data processing layer, and the data access layer.

Data Storage Layer: The data warehouse's storage layer is in charge of archiving information. This layer is often implemented in modern data warehousing by means of a distributed file system like Hadoop Distributed File System (HDFS) or a cloud-based data storage service like Amazon Simple Storage Service (S3).
Data Processing Layer: The data warehouse's data is processed by the data processing layer. This layer is often implemented in modern data warehousing using a cloud-based data processing service like Amazon Elastic MapReduce (EMR) or a distributed processing framework like Apache Spark.
Data Access Layer: It is the job of the data access layer to grant users access to the information stored in the data warehouse. This layer is often deployed in modern data warehousing with the help of a business intelligence tool like Tableau or Power BI, or a data science platform like Dataiku or Databricks.

The Snowflake data warehouse is built on a multi-cluster, shared data architecture that separates compute and storage, enabling users to scale each component independently. This separation also allows users to store, analyze, and share data across multiple clouds and regions, making it highly flexible and accessible.

One of the primary advantages of Snowflake's architecture is its ability to handle massive data volumes and complex queries at a rapid pace. Its virtual warehouses can spin up and down in seconds, providing fast and scalable performance for workloads of any size. Additionally, Snowflake's architecture provides high availability and disaster recovery capabilities, ensuring data availability and business continuity in case of any unexpected issues. Overall, the Snowflake data warehouse architecture offers a powerful, flexible, and highly scalable solution for modern data warehousing needs. By leveraging Snowflake's cloud-based architecture, businesses can easily manage their data, analyze it more effectively, and make more informed decisions to drive success.

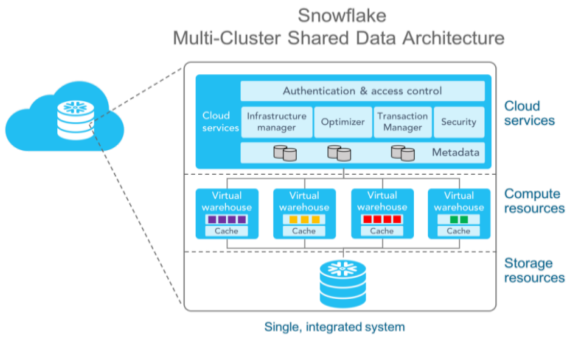

Figure 2: Snowflake Architecture

Tools and Technologies Used in Modern Data Warehousing

Cloud Data Warehouses: Cloud data warehouses, such as Amazon Redshift, Google BigQuery, and Snowflake, are becoming increasingly popular due to their scalability, flexibility, and cost-effectiveness.

Distributed Processing Frameworks: Distributed processing frameworks, such as Apache Spark, Apache Flink, and Apache Beam, are used for processing large volumes of data in parallel.

Data Integration Tools: Data integration tools, such as Apache NiFi and Talend, are used for integrating data from various sources into the data warehouse.

BI Tools: BI tools, such as Tableau, Power BI, and QlikView, are used for visualizing and analyzing data stored in the data warehouse.

Scope of Modern Data Warehousing

Modern data warehousing encompasses the storage, organization, and analysis of large volumes of data generated by modern data sources such as cloud-based applications, social media platforms, mobile devices, and the Internet of Things (IoT). This data is typically stored in a centralized data warehouse, which acts as a single source of truth for organizations seeking to make data-driven decisions.

The scope of modern data warehousing has expanded significantly in recent years, thanks to advancements in data processing, storage, and analysis technologies. Modern data warehousing solutions can handle structured, semi-structured, and unstructured data, making it possible for organizations to integrate data from a wide range of sources.

Some of the key features of modern data warehousing include:

Scalability: Modern data warehouses can scale horizontally and vertically to accommodate large volumes of data and users.

Cloud-based architecture: Cloud-based data warehousing solutions provide organizations with a scalable and cost-effective way to store and analyze data.

Real-time data processing: Modern data warehouses can process data in real-time, enabling organizations to make real-time decisions.

Data integration: Modern data warehouses support data integration from a wide range of sources, including traditional databases, cloud-based applications, and IoT devices.

Data visualization: Modern data warehousing solutions offer advanced data visualization capabilities, making it easier for organizations to gain insights from their data.

Overall, modern data warehousing plays a critical role in helping organizations leverage their data to drive business value, improve decision-making, and gain a competitive advantage in today's data-driven world.

II. LITERATURE REVIEW

Helmut Spengler et. All (2022) Authors have outlined a framework for enhancing clinical and translational data warehousing infrastructures' data quality. For critical medical research platforms like i2b2 and tranSMART, as well as warehouses based on the OMOP CDM, our method enables monitoring the evolution of data quality over time using customisable dashboards and alerting mechanisms. In addition, this design allows for consolidated oversight of various data marts. An event logging API is made available to ETL process developers, which may be used to log and organize information about data quality issues. Authors demonstrate the feasibility of this approach by releasing an implementation in the form of open-source software that is already being put to use in ongoing medical research initiatives. [1]

Aaron Groulx et. All (2018) Managing the shadow economy and enhancing tax administration are only two of the many potential benefits of the social media tax data warehouse. Detecting tax fraud rings, smuggling, offshore business agreements, money laundering, and other financial crimes are all possible uses for the warehouse. Taxation and the distribution of social services are just two examples of other potential applications. Prior to the complete realization of the capabilities outlined in this article, a number of hurdles must be cleared on the technical, political, ethical, and legal fronts. [2]

Ivan Alexander et. All (2018) The field of marketing has never been more exciting than it is in today's rapidly developing globe. The pay is proportional to the amount of merchandise sold to each customer. How to determine which products are now in demand is a key concern at PT. XYZ. The purpose of this article is to build a method and report that will aid marketing departments in determining which products will best meet the needs of their target audience. This paper proposes Kimball and Ross's methodology for building data warehouses. In addition, OLAP analysis can be performed on the data before it is displayed on the dashboard. This company solution will allow PT. XYZ to more quickly sort through product options and determine what would sell best to their clientele. The outcome demonstrates the system's ability to select products that are appropriate for offering to customers based on data from the dashboard. [3]

Abderrazak Sebaa et. All (2017) Till recently, decision support systems relied heavily on conventional data warehouses. New data warehousing systems are necessary because of the exponential growth of data produced by modern applications in terms of dataset number and format, data source diversity, the integration of unstructured data, and the capability of analytical processing. It's crucial to keep up with the fast-paced world of Big Data by modernizing warehouse management systems to meet the problems posed by the era of Big Data. The emphasis of this paper is on data warehousing rather than big data. Authors point out the shortcomings of the more common ones. Authors discuss potential future developments in alternate technologies for data warehousing. [4]

Abobakr Aljuwaiber (2022) In reaction to the information explosion of the 21st century, businesses have developed extensive data repositories to store the vast volumes of data available in a wide variety of formats. This has led to the development of the term "data warehouse" to characterize a massive collection of data. Data warehousing was first defined in this article as the process of aggregating data from many sources and making it available to consumers in a standardized format. In the context of data warehousing, it is crucial for businesses to enhance various information systems. The steps involved in extracting, combining, filtering, transforming, cleansing, converting, and aggregating raw data into a usable data warehouse have also been shown. A data warehouse project will be successful if its members work hard and follow through on their plans. [5]

Taghrid Z. Ali et. All (2020) A strategy for data hygiene has been provided in this research. Each discipline has its own unique set of challenges when it comes to ensuring high-quality data, but all of them can be addressed with the right set of tools, techniques, procedures, and methodologies. Current information systems were developed with meeting operational requirements in mind, however they fail to include mechanisms for enhancing data quality. Authors came to the conclusion that the design of information systems needs to be rethought if data quality is to be achieved. Providing the operational aspects and functional needs of the system jointly is crucial for achieving data quality. All of the enterprise's applications now share a unified and consistent set of metadata. The framework concluded with instructions on how to maximize data cleaning effectiveness within ETL prior to loading it into a data warehouse. Authors also made it clear how the framework's progress could be measured using the quality of the data reported to us. And the results were weighed against the goals that had been set. [6]

Asanka et. All (2019) Users have the option of using the same table for the analysis because it is employed in the same table structure. Non-functional needs like performance, configurability, and type 2 slowly changing dimensions have been taken into account in this study. It was established that this is usable in practical situations by taking into account the non-functional necessity. This study provides evidence that linguistic data analysis can be performed on top of a fuzzy data warehouse. Non-functional needs were met as well, which aided in practical application. Parameters in linguistics are now pre-defined, but a data-driven approach might be used to discover them. [7]

Patric Beinschob et. All (2013) Manual forklifts are still widely used in today's factories and warehouses. SICK AG is leading an EU-funded research project called PAN Robots, which intends to replace robots in factories on a wide scale with autonomous guided vehicles due to their poor efficiency, high energy consumption, and high prevalence in fatal incidents (AGV). High upfront costs, such as registering all pallet placements and reflectors, which must be done by highly experienced specialists in order to produce an accurate navigation map for the AGVs, present significant challenges for operators using current AGVs. The scientific method for first mapping is called simultaneous localization and mapping (SLAM). In this work, we discuss specific 3D mapping methodologies for usage in a warehouse setting using a number of Laserscanners. Measurement data collected on-site at our research partner Casbega in Madrid, Spain, is used to demonstrate first findings. [8]

Leo Willyanto Santoso et. All (2017) As universities already collect enormous volumes of data, "big academic data" has evolved into a more accurate description of what goes on in a university setting. These data sets are extensive and expanding. Tools to generate data from records are essential for university administration. The data produced should aid in the deliberation of upper-level management. In this study, Authors investigate the feasibility of integrating big data technologies with a data warehouse to facilitate the decision-making procedure. Hadoop, a set of large data analytic tools, is proposed here for use with data ingestion and staging. The report finishes by detailing potential next steps for creating and launching a Big Data initiative at an organization. [9]

Fattakhova N et. All (2019) This article explores the many approaches currently in use to consolidate data from disparate systems into a unified data warehouse. Using this data, a unified information environment for the machine-building company can be created. In this essay, we provided evidence on why businesses need to create a unified data repository. They devised a system to bring together unrelated data at the machine manufacturing firm. Information system integration is an old problem. The levels of component integration take into account the issue as well, proposing the incorporation of a universal module into a software product and the establishment of an environment for the interaction of single-function modules via standard data transfer formats. The use of ESB is prevalent in many types of information sharing (Enterprise Service Bus). [10]

Ranak Ghosh et. All (2015) Business intelligence (BI) provides historical, current, and predictive perspectives on business operations through the use of various technologies like reporting, online analytical processing, analytics, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics, and prescriptive analytics. Business intelligence (BI) packages often include OLAP because of the critical nature of analytics to BI. The utilization of a data warehouse is the most typical approach to creating an OLAP system. The suggested architecture unifies reporting, knowledge production, query processing, data center management, and other facets of business intelligence with the data warehouse environment. [11]

Abdul Moktadir et. All (2019) In this study, we provide the groundwork for a new approach to data warehouse modeling. This structure goes against the conventional wisdom of data warehouse design, which is to employ dimensions to characterize data. The operational facts are used to characterize certain characteristics of a dimension and are then distributed among the records that make up that dimension in this schema. This data structure breaks the data into multiple smaller fact tables by dividing it horizontally. The availability of data in a dimension table can be improved by linking it to smaller fact tables that provide supplementary characteristics for summarizing facts. Again, the data warehouse can return results to specific queries in a fraction of the time it takes to process a minimal data set. As each piece of information has its own distinct schema, it can be quickly and easily retrieved using this method. As a result, it achieves higher performance in providing high-quality data and quicker responses throughout a range of query processing. [12]

M. Asif Naeem et. All (2014) In this study, Authors provide a new semi-stream join, CSSJ, which may be used to connect data from a stream to a master data table stored on disk. Authors evaluate it in comparison to MESHJOIN, a classic algorithm that serves a similar purpose. CSSJ was developed to take use of the asymmetric and non-normally distributed information typical of real-world contexts. In particular, we think about the possibility of foreign keys in the stream data having a Zipfian distribution. CSSJ, in contrast to MESHJOIN, keeps these most-used tuples of R permanently in memory, which eliminates the need for disk I/O and dramatically improves the algorithm's performance. We have supplied an experimentally verified cost model for the new approach. We have conducted a comprehensive experimental investigation demonstrating that CSSJ is superior to the previous MESHJOIN algorithm. [13]

Gao Ningning et. All (2010) Functions in university teaching and management have been given new meanings as higher education has become more accessible around the world. The old understanding of everyday administration is being phased out in favor of a focus on ideological and political education, advice for students' mental and physical growth, and management of student affairs. Hence, in the 21st century, professors in higher education need to adapt to the role of study guides and course organizers, or shift their focus to the business of teaching. They need that teachers utilize IT in their lessons and lessons planning. As was noted above, the author incorporates data warehouse and data mining techniques from cutting-edge IT into his pedagogical and administrative practices. It is hoped that this would help bring in insightful feedback, and that I will learn something from my 21st century coworkers. [14]

Xiaofeng He et. All (2005) When dealing with massive amounts of data with intricate relationships, data warehouse (DW) technology is the most efficient tool at your disposal. The document provides an overview of the SCADA/EMS DW architecture. The system is comprised of a node for enterprise application integration (EAI), a center for real-time data processing (RTDPC), an operational data warehouse (ADW), and management tools. Metadata, node configuration data, and system management data are all under the control of the management tools. The importance of frameworks in EAI and the architectural topologies behind successful implementations are highlighted. The DW is a modern control center, data mining, and decision support system's reliable and safe information exchange and storage platform. Due to privatization and deregulation, new requirements have arisen for interpreting data from many DW sources. The paper outlines the new data warehouse of SCADA/EMS system, including its characteristics and structure, as a solution to these problems. Utilities have started using this new method, and others are planning to soon. The SCADA/EMS system will continue to expand with the advent of new applications as the industry acquires expertise with this innovative technology. [15]

Illia Sokolov et. All (2018) The proposed method, data storage reengineering, is a partial and intermediate solution on the way to contemporary data processing technologies, given that current technology for Big Data allows organizations to greatly boost return on investment from their existing data warehouse environment. For decades, corporate architects in industries such as banking, retail, manufacturing, energy, telecommunications, and insurance have relied on data warehouse designs to provide a wide range of consumer analytics, integrated analytics, advanced analytics, and predictive analytics. With an average age of 25 years, the technologies and architectures used to build traditional data warehouses were originally developed with an eye on online transaction processing (OLTP). The immediate processing of today's massive, diverse, and poorly organized data was never meant to be a requirement of data warehousing. The demands of business analysts have been a major driver of the long-term trend toward exponential increase in the need for data warehousing. [16]

Frank S.C. Tseng et. All (2020) There can be either an object-centric or a people-centric approach to integrating data from various sources. The former classifies things according to their identifiers and the people, places, times, and events involved with them. Objects are "honest" and do not try to conceal themselves if they each have a unique identifier (UID), making this task straightforward. It is not an easy operation to link personal IDs across networks and collect resources, together with the associated locations, times, and events, especially when people may use many identifiers to conceal their online presence. That's why we think this paradigm will be useful for creating user-centered technology to meet the demands of integrating digital footprints and social business data across industries. [17]

Hichem Dabb`echi et. All (2016) Management choices are aided by business intelligence (BI) tools. New possibilities for BI and data analytics systems can be found at the intersection of two of the most essential technologies today: big data and cloud computing. Yet, in order to offer business intelligence services based on cloud computing from large data sources, the conventional data warehouse must be updated. This information in these systems comes from many different places and is stored in many different formats. Therefore, they require an advanced IT system with high processing efficiency and large amounts of storage space. The cloud computing paradigm could be used as a means of dealing with the new data warehouse architecture design. This latter provides helpful strategies, infrastructures, and offerings for handling this vast data effectively during its processing, computation, storage, and analysis phases. In this study, Authors advocate for a novel cloud-based data warehouse structure from the vantage point of big data analytics. Specifically, we describe the layers suggested for enabling big data analytics, including the data warehouse architecture, the platform, and the analytics software as a service. [18]

Ramesh Babu Palepu et. All (2012) It's common knowledge that the Extraction, Transformation, and Loading (ETL) processes are fundamental to building a Data warehouse. First, we'll need to extract the information from a wide variety of small operational databases, then we'll need to transform the data into the schema model of Data Warehouse, and finally we'll need to load the data into Data Warehouse while keeping its quality in mind. Meta data repository, which keeps every feature of the data objects loaded in the Data Warehouse, is a crucial part of this procedure. In this way, the quality concerns in the Data Warehouse framework may be fixed by first screening the Meta data, then detecting the quality control issues, and then applying quality standards to the data. Furthermore, modifications to the Data Warehouse design are proposed in accordance with Meta database quality control criteria. The results are a Meta database and Data Warehouse architecture that are in sync with one another. In this study, we suggest a new Meta data quality architecture for fixing the fundamental data quality issues that plague Data Warehouses. [19]

Nayem Rahman (2016) An organization will need to invest a lot of money into a data warehouse implementation. Companies have a corporate data warehouse so that their employees and executives always have access to the latest and greatest data. To make important business decisions, companies require access to this data at the correct moment. An empirical analysis of data warehouse implementation success is discussed here. The goal of this research was to pin down the factors that make a difference. Eight predictor variables and four response variables were used in this analysis. Study results demonstrate the importance of a subset of criteria in determining data warehousing success response variables. These results may prove useful for IT experts, professors, and researchers. Authors present the results of an empirical study into the success of enterprise data warehouse implementations in this article. Based on the findings of prior research, Authors isolated eight independent variables that have been shown to significantly affect data warehousing outcomes. Based on the literature review, we also isolated four response variables that have been shown to be crucial in retaining the interest of data warehouse project stakeholders, including customers. Authors tested hypotheses about the associations between each predictor variable and each responder variable using correlation and regression analysis. [20]

Matteo Golfarelli et. All (2009) Data warehouses are specialized data repositories used to aid in managerial judgment. Managing time becomes critically important since making decisions usually involves looking back at patterns in the past. In this work, we examine the wide range of problems that arise from the requirement of providing an accurate description of how information evolves over time in data warehousing systems, problems that are generally bundled together under the name temporal data warehousing. Authors acknowledge that these problems can be broken down into several categories with reference to a three-levels architecture, including: querying temporal data, creating temporal data warehouses, and dealing with data/schema changes in the data warehouse and data mart. In this poll, we organized the challenges of temporal data storage by category and held in-depth discussions about each one. A thorough literature review found that these factors do not always receive complete attention from the scientific community. In reality, modifications in dimensional data are the single point of agreement among the possible design solutions. While numerous novel approaches have been proposed for dealing with factual data and schema updates, no widely accepted paradigm has yet been developed. [21]

Ablimit Aji et. All (2013) Despite having many of the same features as corporate data, "big" spatial data from imaging and spatial applications presents its own distinct challenges due to the data's multi-dimensionality and the computational complexity of spatial query processing. In this paper, Authors provide Hadoop-GIS, a system that combines the advantages of MapReduce's scalability and low cost and spatial access methods' ability to quickly and accurately answer spatial queries. Hadoop-GIS accomplishes this via spatial partitioning, partition-based parallel processing over MapReduce, efficient boundary object handling for accurate query results, and a user-configurable spatial indexing and querying engine that supports multilevel spatial indexing. Our results show that Hadoop-GIS is a scalable and effective solution for analytical geographic queries over massive spatial datasets, as we tested it out on two real-world use cases. [22]

III. WORKING ARCHITECTURE MODERN DATA WAREHOUSE

The working architecture of real-time Modern Data Warehouse is mentioned below:

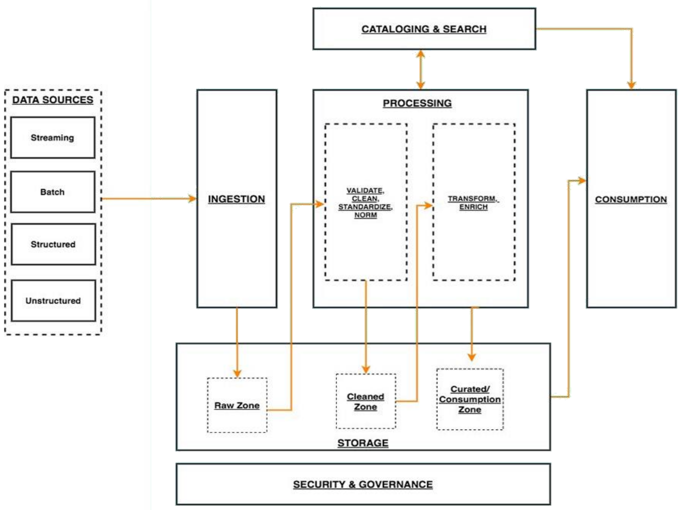

Figure 3: Modern Data Warehousing Architecture

Multiple Parallel Processing (MPP) Architectures

- “MPP architecture enables a mighty scale and Distributed Computing”.
- “Resources add for a linear scale-out to the largest Data Warehousing projects”.
- “Multiple parallel processing architecture uses a "shared-nothing". There are numerous physical nodes, each runs its instance. This results from performance many times faster than traditional architectures”.

Multi-Structured Data

- “Define Big Data & Analytics Infrastructure for multiple storage data with a polyglot persistence strategy”.
- “Integrate portions of the data into the Data Warehouse”.
- “Federated query access”.

Lambda Architecture

In lambda, architecture defines three layers -

- **“Speed Layer -** Low latency data”.
- **“Batch Layer -** Raw Data processing to support complex analysis”.
- **“Serving Layer -** Response to queries”.

Hybrid Architecture

Scale up MPP compute nodes during -

- “Peak ETL data loads”.
- “High query volumes”.
- “Utilize existing On-Premises data structures”.
- “Use Cloud services for Advanced Analytics”.

FEATURES OF MODERN DATA WAREHOUSE

- “Variety of subject areas & data sources for analysis with the capability to handle the large volume of data”.
- “Expansion beyond a single relational DW/Data Mart structure to include Data Lake”.
- “Logical design across multi-platform architecture balancing performance & scalability”.
- “Data virtualization in addition to Data Integration”.
- “Support for all type & levels of users”.
- “Flexible deployment decoupled from the tool used for development”.
- “Governance model to support security and trust, and Master Data Management”.
- “Support for promoting the self-service solution to the corporate environment”.
- “Ability to facilitate Real-Time analysis of high-velocity data”.
- “Support for Advanced Analytics”.
- “Agile Delivery approach with the fast delivery cycle”.
- “Hybrid Integration with Cloud services”.
- “APIs for downstream access to data”.
- “Some DW automation to improve speed, consistency, business terminology”.
- “An analytics sandbox or workbench area to facilitate agility within a BI environment”.
- “Support for self-service BI to augment corporate BI; Data discovery, Data Exploration, Self-service Data preparation”.

IV. TRADITIONAL DW CONCEPTS

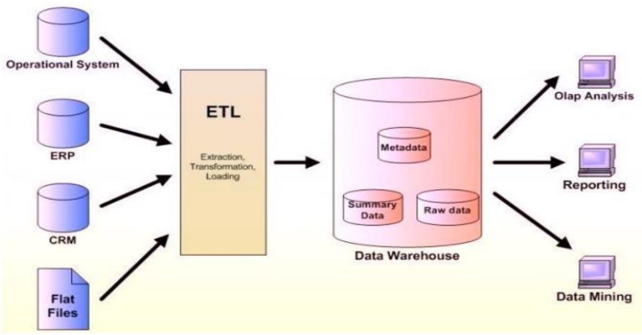

Figure 4: General architecture of a data warehouse

By grouping information according to categories, "a data warehouse (DW) is a collection of data, arranged to be utilized as a decision support" [23]. Such as Manufacturing, Retail, Promotion, and so on. Decision-making is facilitated by this structure since it allows for the collection of all relevant data on a given topic. Information in a warehouse is typically accessed in a read-only capacity, with few if any updates or deletions made by end users; this preserves data provenance for long-term analysis. As data are put into a DW, integration of the data resolves any representation, naming, or contextual issues that would otherwise arise [24]. Figure 1 shows a DW's basic structure. It's a repository for dispersed and diverse data sets. Databases, data files, external sources, etc., may all fall under this category. It is important to sanitize data sources before storing them. Data cleansing involves selecting and purifying data to fix errors and settle semantic discrepancies. The data will be cleansed and then added to the warehouse. To integrate data from various sources, one must perform the loading of data sources at the logical warehouse schema transformations, a process known as ETL (extract, transform, and load) [25]. The first step is "extraction," when information is gathered from various sources. The extraction process must be synchronized at this stage to prevent data corruption. Transformation (2) is the process of reformatting the extracted data in accordance with the desired warehouse schema. Data source semantics assignment and source field combination are two examples. After data is loaded into a target database, DW, or data mart, analysis may begin. In the warehouse, there is a special directory for storing information related to its development, administration, and use. Data sources, associated data, the integration schema, refresh rules, user profiles, and user groups are all part of the "Metadata" [26]. Several data Marts could make up a data Warehouse. These selections are taken from the warehouse to serve a specific audience's needs. They focus solely on OLAP (Online Analytical Processing) research and analysis for all of their business decisions. When it comes to decision-making aids, OLAP's many data visualization options are invaluable.

V. STEPS FOR TRANSFORMING RAW DATA INTO DATA WAREHOUSE

A data warehouse processes unstructured information into a structured database for the purpose of analysis and reporting. Online Transaction Processing (OLTP) systems track typical business activities such sales, purchases, and stock adjustments [27, 28]. The next phase involves incorporating the operations of extracting, consolidating, filtering, transforming, cleansing, aggregating, and loading [29, 30, 31] raw data into a data warehouse where different formats can be strictly united. Figure 2 depicts the flow of information from its primary sources into the data warehouse.

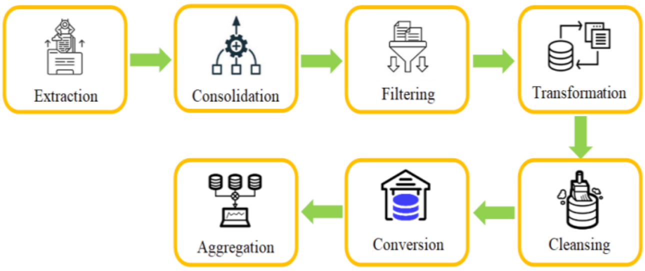

Figure 5: Processes of transforming raw data into a data warehouse

Extraction: In most cases, this phase involves working with multiple data sets. In this phase, the responsible party may extract data from the database and change it into a data warehouse setup.

Consolidation: What we're talking about here is the method by which various data sets are combined into a single repository.

Transformation: Putting the data through a series of transformations in order to derive new values for loading into the target system or verifying the source data.

Cleansing: For reliable results, it's crucial to take this extra measure. It makes the information stored in the warehouse more reliable. Inconsistent field lengths, inconsistent descriptions, inconsistent value assignments, missing entries, and violation of integrity constraints are just a few examples of the kinds of mistakes or anomalies that could be present in the warehouse's massive amount of data if numerous sources were used. Therefore, it is crucial that the information stored in the warehouse is accurate.

Conversion: Type and translate the raw data into the format utilized by a warehouse, and then plan it onto a new data field inside the warehouse.

Aggregation: Finally, the aggregation stage classes and connects data into useable metrics for analysis, which is where mh of a data warehouse's value lies.

VI. TRADITIONAL DATA WAREHOUSE AND MODERN DATA WAREHOUSE

The following sections cover both traditional data warehouses and newer ones. We also discuss their similarities and differences. To store and retain historical data generated from operational and transactional applications, businesses may take use of a data warehouse, which is a combination of ideas and technology. This will help executives, managers, and analysts make better, faster judgments. Data warehousing is a revolutionary step forward for corporate strategy. The provision of access to a warehouse of strategic data is the service that we call a data warehouse [33]. Information useful in making choices may be stored in a data warehouse. In contrast to the operational database, the data warehouse is a repository for aggregated logical data. Information from many sources may be brought together with the help of the data warehouse. This ensures that management has a streamlined process for obtaining the information they need to make well-informed choices. Data warehouses have several characteristics, including subject-oriented, integrated data; persistence; time-variable; not homogenized data; and so on [34, 35]. The data warehouse was modeled using dimensional analysis. Dimensional modeling employs a call-based methodology that facilitates sophisticated querying. The Star Schema is one kind of dimensional modeling system; it centers on a fact table and surrounds it with dimension tables. In order to query and link to the dimension table, the fact table must include a descriptive property. Everything from performance indicators to operational data to aggregate sizes and beyond may be included into a decision's consideration. Tables showcase the data warehouse's decision-making potential. Attributes from the dimension table may be used to characterize data in the fact table.

Data warehouses hold information gathered from a variety of sources using a process known as "Extract," "Transform," and "Load" (ETL). Data used in ETL may be retrieved from a wide variety of sources, such as flat files, spreadsheets, and ERP programs. An aid to decision-making based on a data warehouse. Decision support system (DSS) refers to a computer-based system that helps decision makers make better use of available data and models. DSS functions combine human knowledge with the computational power of computers to provide more accurate outcomes from analysis and decision-making. DSS need information from many different sources to solve a problem. Each problem requires an answer, and each strategy a set of data. The DSS methodology starts with information gathering. The DSS model, the second component of the architecture, represents the state that the data represent. The third component of DSS design is the information certain systems have. The fourth user of the DSS interacts with the system via a user interface, the fifth component of the DSS design. A well-thought-out strategy and seamless component integration are essential for the DSS's growth. Data warehouses have found widespread application in the academic community as well. Data warehouse solution is useful for the usual university information system [36]. The academic data warehouse is useful for conducting research and making administrative decisions across the board. The data warehouse plays a pivotal role in the analysis of academic information.

Characteristics	Traditional Data Warehouse	Modern Data Warehouse
“Purpose”	“Treatment of collected data for a specific business area that is integrated, non-volatile and time-varying. It supports decision-making process”.	“Processing of structured, semi-structured, and unstructured data, from diverse sources and the volume of data exceeds the ability of traditional tools to capture, store, manage and analyze them”.
“Data source”	“Usually transactional and operational databases”.	“Various sources and data types (social media, sensors, blog, video, and audio)”.
“Scope”	“The integrated structured data to support Business Intelligence (BI) and Online Analytical Processing (OLAP)”.	“Analyze and discover knowledge from large volumes of data characterized by the 4Vs (volume, velocity, variety and veracity)”
“Architecture”	“Oriented to processes of extraction, transform and load (ETL). Star schema is the appropriate solution for the architecture”.	“The architecture is depending on the problem. There is still no reference architecture or standardized terminology. They are some proprietary and product oriented architectures from the vendor”
“Technology”	“The technology is mature and tested tools in large amount applications, both free and licensed software”.	“The technology is still growing. Hadoop is one of the open-source software frameworks used for distributed storage and processing of dataset of big data”
“End-user”	“Business analysts or top managements who do not require specific knowledge of technologies or data exploration”.	“Data scientists with knowledge in technologies, algorithms, mathematics and statistics”.

Table 1: The characteristic of traditional data warehouse and modern data warehouse

Big data is here, and traditional data warehouses aren't equipped to deal with it [37]. Historically, schools have relied on academic information systems and time-honored tests to compile their student data. However, this information is now being obtained mostly through digital channels such as online courses, educational games and simulations, and social media platforms. Due to the massive amount of labor, users, and data, the logical and physical design must be optimized. That's why we need parallel data processing. In addition, information from unstructured data with a varied data structure cannot be extracted by a conventional data warehouse. Structured data from transactional sources is integrated in a traditional data warehouse, and OLAP-based analysis is provided as a result. This presents a chance for big data technology to provide a solution. Hadoop and other big data technologies need to be better integrated with traditional data warehousing practices. Hadoop is used in most solutions because it allows for parallel and distributed processing of massive volumes of data. Hadoop can analyze massive, disparate datasets at lightning speed. Compare and contrast the features of a classic data warehouse with those of a state-of-the-art one in terms of their goals, data sources, scope, architecture, technology, and end users in Table 1.

VII. HOW MODERN DATA WAREHOUSING HELPING IN DIFFERENT SECTORS

The banking and healthcare sectors, as well as the agriculture, forestry, fishing and hunting sectors, the arts, entertainment and recreation sectors, the finance and insurance sectors, the healthcare and social assistance sectors, the manufacturing sector, and the mining sector are all benefiting greatly from modern data warehousing:

- **Banking**: Financial institutions may benefit from data warehousing because it provides a single location for storing and analyzing vast amounts of customer, transaction, and market data. This is useful for a variety of business operations, including fraud prevention, client segmentation, and individualized advertising.
- **Healthcare**: Data warehousing facilitates the integration and analysis of patient data from multiple sources, such as electronic health records (EHRs), clinical trials, and medical research. This enables better patient care, predictive analytics for disease prevention, and population health management.
- **Agriculture**: Data warehousing helps in collecting and analyzing data from various sources such as weather patterns, soil conditions, and crop yields. This enables farmers to make data-driven decisions regarding planting, irrigation, and fertilization, leading to optimized crop production and increased yields.
- **Accommodation and Food Services:** Data warehousing enables the collection and analysis of customer data, reservation data, and transaction data in the hospitality and food services industry. This helps in personalized marketing, customer segmentation, and improving overall customer experience.
- **Administration, Business Support, and Waste Management Services:** Data warehousing helps in consolidating and analyzing data related to waste management, resource utilization, and operational efficiency. This enables better decision-making, optimization of resources, and cost reduction.
- **Agriculture, Forestry, Fishing, and Hunting:** Data warehousing helps in collecting and analyzing data related to harvesting, processing, and distribution in the agriculture, forestry, fishing, and hunting industries. This enables better supply chain management, product traceability, and compliance with regulations.
- **Arts, Entertainment, and Recreation:** Data warehousing helps in analyzing customer data, ticket sales data, and marketing data in the arts, entertainment, and recreation industry. This enables better targeting of customer preferences, optimized pricing, and improved customer experience.
- **Finance and Insurance:** Data warehousing enables the storage and analysis of large volumes of financial data, including customer transactions, market data, and risk assessment data. This helps in fraud detection, risk management, and financial forecasting.
- **Healthcare and Social Assistance:** Data warehousing enables the integration and analysis of patient data, billing data, and operational data in the healthcare and social assistance industry. This helps in improving patient care, optimizing resource utilization, and streamlining administrative processes.
- **Manufacturing**: Data warehousing helps in collecting and analyzing data related to production processes, supply chain management, and product quality in the manufacturing industry. This enables better production planning, inventory management, and quality control.
- **Mining**: Data warehousing enables the storage and analysis of data related to mining operations, geological data, and environmental impact data. This helps in optimizing mining processes, ensuring compliance with regulations, and minimizing environmental impact.

Modern data warehousing is transforming various industries by enabling data-driven decision-making, improving operational efficiency, enhancing customer experience, and optimizing resource utilization. It plays a crucial role in driving innovation and growth in these industries.

VIII. HOW IS MODERN DATAWAREHOUSE TRANSFORMING BANKING SYSTEM

The modern data warehouse is transforming the banking system by empowering banks with advanced data management, analytics, real-time processing, scalability, flexibility, and data security capabilities. These transformations can lead to improved customer experiences, enhanced operational efficiency, better risk management, and increased competitiveness in the banking industry.

The modern data warehouse has the potential to significantly transform the banking system by revolutionizing how data is stored, processed, and analyzed. Here are some key ways in which the modern data warehouse is transforming the banking system:

- **Enhanced Data Management**: Modern data warehouses allow banks to store and manage large volumes of data in a structured, semi-structured, and unstructured format. This enables banks to collect and store data from various sources, such as customer transactions, social media, and external market data, and process it in real-time or near-real-time. This enhanced data management capability allows banks to gain deeper insights into customer behavior, market trends, and risk assessment, which in turn can drive better decision-making and improve operational efficiency.
- **Advanced Analytics**: Modern data warehouses offer advanced analytics capabilities, such as machine learning, artificial intelligence (AI), and predictive analytics, which can help banks uncover hidden patterns and trends in data. Banks can use these advanced analytics capabilities to analyze customer data, identify potential fraud or risk factors, and personalize offerings based on customer preferences. Advanced analytics can also help banks optimize pricing, manage risk, and forecast demand, thereby improving their overall performance and competitiveness in the market.
- **Real-time Data Processing**: The modern data warehouse allows banks to process data in real-time or near-real-time, which can significantly impact the banking system. For example, banks can monitor transactions in real-time to detect and prevent fraudulent activities, assess credit risk in real-time to make quicker lending decisions, and deliver personalized offers and recommendations to customers in real-time, enhancing their overall banking experience.
- **Scalability and Flexibility**: Modern data warehouses are highly scalable and flexible, allowing banks to handle large and complex data sets with ease. Banks can scale up or down their data warehousing infrastructure based on their needs and requirements, and also integrate data from diverse sources, such as internal systems, third-party data providers, and external data sources. This scalability and flexibility enable banks to adapt to changing market conditions, comply with regulatory requirements, and leverage new data sources and technologies for business insights.
- **Data Security and Compliance**: Data security and compliance are critical in the banking industry, and modern data warehouses provide robust security features to protect sensitive customer data. Banks can implement encryption, access controls, and other security measures to ensure data confidentiality, integrity, and availability. Modern data warehouses also enable banks to comply with various regulatory requirements, such as GDPR, Basel III, and KYC/AML, by providing audit trails, data lineage, and data governance capabilities.

IX. CURRENT CHALLENGES IN MODERN DATA WAREHOUSING HOUSING SYSTEMS

Modern data warehousing systems offer numerous benefits, they also face challenges related to data integration, scalability, data security, data governance, data complexity, cost management, data quality, data analytics, and evolving technology landscape. Addressing these challenges requires careful planning, robust strategies, and ongoing efforts to ensure efficient and effective data warehousing operations. While modern data warehousing systems have brought numerous benefits to industries, they also face several challenges. Some of the current challenges in modern data warehousing systems include:

- **Data Integration**: Data warehousing systems often need to integrate data from various sources, such as structured and unstructured data, internal and external data, and data from different formats and systems. Data integration can be complex and time-consuming, requiring data cleansing, transformation, and validation to ensure data quality and consistency.
- **Scalability and Performance**: As data volumes continue to grow exponentially, data warehousing systems need to handle large-scale data processing and analysis efficiently. Ensuring high performance and scalability can be challenging, especially when dealing with complex queries and real-time data processing requirements.
- **Data Security and Privacy**: Data warehousing systems store sensitive and valuable data, making data security and privacy critical concerns. Protecting data against unauthorized access, data breaches, and complying with data privacy regulations such as GDPR and HIPAA require robust security measures, encryption, access controls, and audit trails.
- **Data Governance**: Data governance is essential in data warehousing systems to ensure data integrity, accuracy, and compliance. Establishing data governance policies, data lineage, data cataloging, and data lineage can be challenging, especially when dealing with multiple data sources and data integration.
- **Data Complexity and Variety**: Modern data warehousing systems need to handle diverse data types, including structured, unstructured, and semi-structured data. Data variety and complexity, such as data from IoT devices, social media, and multimedia, can pose challenges in data integration, storage, and analysis.

- **Cost Management**: Implementing and managing data warehousing systems can be costly, including hardware infrastructure, software licenses, and maintenance costs. Optimizing costs while ensuring performance and scalability can be challenging, especially for smaller organizations or startups with limited resources.
- **Data Quality and Consistency**: Ensuring data quality and consistency across different data sources can be challenging in data warehousing systems. Data discrepancies, data duplication, and data inconsistencies can impact data analysis and decision-making.
- **Data Analytics and Insights**: Extracting meaningful insights and value from large and complex data sets stored in data warehousing systems can be challenging. Advanced analytics techniques, such as machine learning and data mining, require specialized skills, tools, and expertise.
- **Evolving Technology Landscape**: The technology landscape for data warehousing is constantly evolving, with new technologies, platforms, and approaches emerging. Keeping up with the latest advancements and selecting the right technologies for specific use cases can be challenging.

X. COST CONSTANTS WITH MODERN DATA WAREHOUSING PLATFORM

It's important for organizations to carefully consider these cost constants when planning for and implementing modern data warehousing platforms. Proper cost management strategies, such as optimizing resource usage, leveraging cost-effective storage options, and monitoring and optimizing data processing tasks, can help organizations mitigate the costs associated with modern data warehousing platforms.

- Modern data warehousing platforms can offer several cost-related benefits, but they also come with certain cost considerations that organizations need to be aware of. Some of the cost constants with modern data warehousing platforms include:
- Infrastructure Costs: Modern data warehousing platforms typically require robust hardware infrastructure to handle large data volumes and processing requirements. This can include costs for servers, storage, networking equipment, and other hardware components. Organizations need to plan for the initial investment in infrastructure as well as ongoing maintenance and upgrades.
- Software Licensing Costs: Many modern data warehousing platforms require licensing fees for the software used in the data warehousing stack, including database management systems, data integration tools, data analytics tools, and other software components. Organizations need to consider the licensing costs associated with the chosen data warehousing platform.
- Data Storage Costs: Data warehousing platforms involve storing and managing large amounts of data, which can result in ongoing data storage costs. This can include costs for data storage resources such as cloud storage, data warehouses, and data lakes, as well as data backup and disaster recovery solutions.
- Data Transfer Costs: Moving data into and out of data warehousing platforms, especially in cloud-based environments, can incur data transfer costs. These costs can arise from data ingestion, data extraction, and data transfer between different data warehousing components or across different regions or data centers.
- Data Processing Costs: Modern data warehousing platforms often involve data processing tasks such as data transformation, data aggregation, and data analysis. These data processing tasks can consume computing resources, such as CPU and memory, which may result in processing costs, especially in cloud-based environments where organizations pay for the resources used.
- Data Governance and Security Costs: Ensuring data governance and security in data warehousing platforms may require additional investments. This can include costs associated with data encryption, access controls, audit trails, data lineage tracking, and compliance with data privacy regulations.
- Training and Skill Development Costs: Modern data warehousing platforms often require skilled resources to operate and manage effectively. Organizations may need to invest in training and skill development programs for their employees to ensure they have the necessary expertise to work with the chosen data warehousing platform.
- Maintenance and Upgrades Costs: Data warehousing platforms may require ongoing maintenance and upgrades to ensure smooth operations, performance optimization, and security patching. Organizations need to plan for the costs associated with regular maintenance and upgrades of the data warehousing platform.
- Vendor Support and Service Costs: Organizations may rely on vendor support and services for their data warehousing platforms, which can incur additional costs. This can include costs for technical support, consulting services, and other vendor-provided services.

XI. CONCLUSION

In reaction to the information explosion of the 21st century, businesses have developed extensive data repositories to store the vast volumes of data available in a wide variety of formats. Hence, the term "data warehouse" was coined to characterize a massive data storage facility. Data warehousing was first defined in this article as the process of aggregating data from many sources and making it available to consumers in a standardized format. To sum up, modern data warehousing is an adaptable, scalable, and economical method of handling and analyzing massive amounts of data. It's built to handle any workload, any sort of data, and comes equipped with cutting-edge analytics tools to help businesses learn more from their information. In this study, we looked at a contemporary data warehouse that may replace the antiquated ones that are unable to process massive data in the classroom. The challenges of conventional data analysis will be mitigated by the data warehouse method of big data technologies. In addition, this has the potential to improve the efficiency and effectiveness of policymaker decision making and to expand educational opportunities through novel approaches to learning.

XII. REFERENCES

[1] Helmut Spengler, Ingrid Gatz, Florian Kohlmayer "Improving Data Quality in Medical Research: A Monitoring Architecture for Clinical and Translational Data Warehouses" IEEE 2020.
[2] Aaron Groulx, Carolyn McGregor "A Social Media Tax Data Warehouse to Manage the Underground Economy" IEEE 2018.
[3] Ivan Alexander, Rian Rassetiadi, Samuel Garcia "Business Solution for Choosing Products Using Data Warehouse in Payment Solution" IEEE 2018.
[4] Abderrazak Sebaa, Fatima Chikh, Amina Nouicer "Research in Big Data Warehousing using Hadoop" Journal of Information Systems Engineering 2017.
[5] Abobakr Aljuwaiber "DATA WAREHOUSING AS KNOWLEDGE POOL : A VITAL COMPONENT OF BUSINESS INTELLIGENCE" IJCSEIT 2022.
[6] Taghrid Z. Ali, Tawfig M. Abdelaziz "A Framework for Improving Data Quality in Data Warehouse: A Case Study" IEEE 2020.
[7] P. P. G. D. Asanka, A. S. Perera "Linguistics Analytics in Data Warehouses Using Fuzzy Techniques" Smart Computing and Systems Engineering 2019.
[8] Patric Beinschob and Christoph Reinke "Strategies for 3D Data Acquisition and Mapping in Large-Scale Modern Warehouses" IEEE 2013.
[9] Leo Willyanto Santoso, Yulia "Data Warehouse with Big Data Technology for Higher Education" Procedia Computer Science 2017.
[10] Fattakhova N., Ponomareva O., Kalmykov A., Koromyslov I "Ways to Collect Disparate Information in a Single Data Warehouse at a Machine-Building Enterprise" IEEE 2019.
[11] Ranak Ghosh, Sujay Halder , Soumya Sen "An Integrated Approach to Deploy Data Warehouse in Business Intelligence Environment" IEEE 2015.
[12] Abdul Moktadir, N. M. Istiak Chowdhury "Subject Oriented Data Partitioning – A Proposed Data Warehousing Schema" ICASERT 2019.
[13] M. Asif Naeem "A Caching Approach to Process Stream Data in Data Warehouse" IEEE 2014.
[14] Gao Ningning "Proposing Data Warehouse and Data Mining in Teaching Management Research" IEEE 2010.
[15] Xiaofeng He, Gang Wang, Jiancang Zhao "Research on the SCADA /EMS System Data Warehouse Technology" IEEE 2005.
[16] Illia Sokolov, Ihor Turkin "Resource Efficient Data Warehouse Optimization" IEEE 2018.
[17] Frank S.C. Tseng, Annie Y.H. Chou "Spatiotemporal Multi-Dimensional Modeling of Data Warehouse for Event Tracing Applications" International Computer Symposium 2020.
[18] Hichem Dabb`echi, Ahlem Nabli "Towards Cloud-Based Data Warehouse as a Service for Big Data Analytics" Springer International Publishing Switzerland 2016.
[19] Ramesh Babu Palepu, Dr K V Sambasiva Rao "META DATA QUALITY CONTROL ARCHITECTURE IN DATA WAREHOUSING" IJCSEIT2012.
[20] Nayem Rahman "An empirical study of data warehouse implementation effectiveness" International Journal of Management Science and Engineering Management 2016.
[21] Matteo Golfarelli, Stefano Rizzi "A Survey on Temporal Data Warehousing" International Journal of Data Warehousing & Mining 2009.
[22] Ablimit Aji, Fusheng Wang, Hoang Vo, Rubao Lee "Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce" Proceedings of the VLDB Endowment 2013.
[23] Inmon, W. H. (2005). Building the Data Warehouse. John Wiley and Sons, New York, NY, 4th edition, ISBN-13: 978-0764599446,
[24] Shah, S.P., Huang, Y., Xu, T., Yuen, M.M., Ling, J. and Ouellette, B.F. (2005). Atlas – A data warehouse for integrative bioinformatics. BMC bioinformatics, 6(1), 34
[25] Golfarelli, M. (2010). From User Requirements to Conceptual Design in Data Warehouse Design – a Survey. Data Warehouse Design and Advance Engineering Applications: Methods for Complex Construction. Part of the Advances in Data Warehousing and Mining (ADWM) Book Series (IGI Global Publication).
[26] Coronel, C., Morris, S. and Rob, P. (2013). Database Systems: Design, implementation, and Management, 10th edition, Cengage Learning. ISBN-10: 1285099672 ISBN-13: 978-1285099675

[27] Palepu, R. B., & Rao, D. K. S. (2012). Meta data quality control architecture in data warehousing. International Journal of Computer Science, Engineering and Information Technology, Vol. 2, No. 4, pp 15-24.
[28] Rainardi, V. (2008). Introduction to Data Warehousing. Building a Data Warehouse: With Examples in SQL Server, 1-27. United States, Apress.
[29] Coronel, C., & Morris, S. (2016). Database systems: design, implementation, & management. Cengage Learning.
[30] Khan, A. (2012). Business Intelligence and data warehousing simplified: 500 questions, answers, and tips. Mercury Learning and Information.
[31] Lin, H. C. (2012). Real-time clinical decision support system. In Medical Informatics, 111-136.
[32] Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. ACM Sigmod record, Vol. 26, No.1, pp 65-74.
[33] Ponniah, Paulraj (2010) "Data Warehousing: a Comprehensive Guide for IT Professional." 2nd ed., New York:The McGraw-Hill Companies.
[34] Reddy, G. Satyanarayana, Rallabandi Srinivasu, M. Poorna Rao, and Srikanth R. Rikkula. (2010) "Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements To Support Decision-Making Process In Industries". Int. Journal on Comp. Sci. and Eng. 2(9): 2865-2873
[35] Gour, Vishal, S.S. Sarangdevot, G.S. Tanwar, and A. Sharma. (2010) "Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse". Int. Journal on Comp. Sci. and Eng. 2(3): 786-789.
[36] Youssef, Bassil (2012) "A Data Warehouse Design for A Typical University Information System", Journal of Comp. Sci. & Research 1(6): 12-17, Dec. 2012.
[37] Salinas, Sonia Ordonez and Alba C.N. Lemus. (2017) "Data Warehouse and Big Data integration" Int. Journal of Comp. Sci. and Inf. Tech. 9(2): 1-17.

