

Development of Data Cleansing and Data Storage for Modern Analytics and Reporting through Data Warehousing

Elenia Ghosrau*

Department of Computer Science and Engineering, National Technical University of Athens, Athens, Greece

DESCRIPTION

Businesses depend on analytics as well as reporting in today's data-driven environment to make wise decisions and obtain a competitive edge. A key component of this process is data warehousing, which offers a centralized location for gathering, arranging, and evaluating huge amounts of data. As organizations accumulate ever-increasing volumes of data from various sources, the need for effective data management and analysis becomes more important. Data warehousing serves as the foundation for modern analytics and reporting by offering a comprehensive solution for storing, organizing, and accessing data. By centralizing disparate data sources into a unified repository, data warehousing enables organizations to derive actionable insights, enhance decision-making processes, and drive business growth [1,2]. Unlike transactional databases, which are optimized for transaction processing, data warehouses are designed for analytical queries and reporting. Key components of a data warehouse include:

Data sources

Data warehouses integrate data from various internal and external sources, including transactional systems, enterprise applications, cloud services, and external data feeds [3]. A data source could be the place where physical information is first converted to digital form or where data initially originates, although even the most refined data can be a source as long as another process can access and use it.

Extract, Transform, Load (ETL) processes

Data is extracted from source systems through ETL operations, which then format and load the data into a data warehouse. This process involves data cleansing, normalization, and aggregation to ensure data quality and consistency [4]. To clean and arrange raw data in preparation for storage, data analytics, and Machine Learning (ML), ETL employs a set of business rules.

Data storage

Data warehouses use specialized database systems optimized for analytical workloads. These databases are designed to handle large volumes of data and support complex queries for analytics and reporting purposes [5].

Query and analysis tools

Data warehouses provide tools and interfaces for querying, analyzing, and visualizing data. These tools enable users to generate reports, create dashboards, and perform ad-hoc analysis to extract valuable insights from the data [6].

Data warehousing offers several benefits for organizations seeking to leverage data for analytics and reporting. By consolidating data from disparate sources into a centralized repository, data warehousing provides a single source of truth for analytics and reporting. This ensures data consistency, eliminates silos, and facilitates cross-functional analysis across the organization. Data warehousing improves the quality and consistency of data by utilizing procedures for data transformation, cleansing, and standardization [7,8]. By standardizing data formats and resolving inconsistencies, data warehouses ensure that analytical results are accurate and reliable. It enables organizations to analyze data in real-time or near-real-time, allowing for faster decision-making processes. By providing timely insights into business performance, market trends, and customer behavior, data warehousing empowers organizations to respond quickly to changing market conditions and emerging opportunities [9]. It also provides tools for creating customizable reports, dashboards, and visualizations that enable users to explore data and communicate insights effectively. Organizations can better grasp their business measurements and performance indicators by presenting data in a dynamic, aesthetically pleasing manner.

In order to handle expanding data volumes and analytical workloads, data warehouses are made to scale both vertically and horizontally. This scalability ensures that organizations can expand their data warehousing infrastructure to meet evolving

Correspondence to: Elenia Ghosrau, Department of Computer Science and Engineering, National Technical University of Athens, Athens, Greece, E-mail: elegho@NTUA.gr

Received: 22-Feb-2024, Manuscript No. JITSE-24-30626; **Editor assigned:** 26-Feb-2024, PreQC No. JITSE-24-30626 (PQ); **Reviewed:** 11-Mar-2024, QC No. JITSE-24-30626; **Revised:** 18-Mar-2024, Manuscript No. JITSE-24-30626 (R); **Published:** 25-Mar-2024, DOI: 10.35248/2165-7866.24.14.379

Citation: Ghosrau E (2024) Development of Data Cleansing and Data Storage for Modern Analytics and Reporting through Data Warehousing. J Inform Tech Softw Eng. 14:379.

Copyright: © 2024 Ghosrau E. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

business needs without compromising performance or reliability. Despite its many benefits, data warehousing also poses challenges for organizations [10]. Integrating data from heterogeneous sources into a data warehouse can be complex and time-consuming. Organizations must develop robust ETL processes and data integration strategies to ensure seamless data flow and consistency. Analytical queries and reporting operations can put a strain on data warehouse performance, especially as data volumes grow. To optimize performance, organizations should implement indexing, partitioning, and query optimization techniques to streamline data access and retrieval.

Data warehouses contain sensitive and confidential information, making data governance and security critical concerns. Organizations must implement robust security measures, access controls, and data encryption techniques to protect data privacy and comply with regulatory requirements. Initiatives for user acceptance and training are necessary for data warehousing to be successfully used for analytics and reporting [11]. For users to gain the ability to derive insights and make well-informed decisions from data, organizations should allocate resources towards training initiatives and intuitive tools.

CONCLUSION

Directories contain files, and drives include folders. When a computer is turned off, a storage device is one that may continue to store and retain data. Devices that read and write data to the hard disk are called hard disk drives. By consolidating disparate data sources, data warehousing enables organizations to derive actionable insights, enhance decision-making processes, and drive business growth. Despite its challenges, data warehousing offers many benefits, including centralized data repository, improved data quality, faster decision-making, enhanced reporting and visualization, and scalability. By implementing best practices and overcoming challenges, organizations can harness the full potential of data warehousing to gain competitive advantages and achieve strategic objectives in today's data-driven landscape.

REFERENCES

1. Lieberman HR, Stavinocha TB, McGraw SM, White A, Hadden LS, Marriott BP. Use of dietary supplements among active-duty US Army soldiers. *Am J Clin Nutr.* 2010; 92:985-995.
2. Batarseh FA, Gonzalez AJ. Predicting failures in agile software development through data analytics. *Softw Qual J.* 2018;26:49-66.
3. Faundez-Zanuy M, Elizondo DA, Ferrer-Ballester MA, Travieso-González CM. Authentication of individuals using hand geometry biometrics: a neural network approach. *Neural Process Lett.* 2007;26(3):201-216.
4. Claps GG, Berntsson Svensson R, Aurum A. On the journey to continuous deployment: technical and social challenges along the way. *Inf Softw Technol.* 2015; 57:21-31.
5. Birch I, Raymond L, Christou A, Fernando M.A, Harrison N, Paul F. The identification of individuals by observational gait analysis using closed circuit television footage. *Sci Justice* 2013;53(3): 339-342.
6. Nachman I, Friedman N, Elidan G. Ideal parent structure learning for continuous variable Bayesian networks. *Art J Mach Learn Res.* 2007; 8:1799-1833.
7. Falavarjani K.G, Modarres M, Joshaghani M, Azadi P, Afshar AE, Hodjat P. Interocular differences of the Pentacam measurements in normal subjects. *Clin Exp Optom.* 2010; 93(1):26-30.
8. Kantor ED, Rehm CD, Du M, White E, Giovannucci EL. Trends in dietary supplement use among US adults from 1999-2012. *JAMA.* 2016; 316(14):1464-1474.
9. Chandra A, Calderon T. Challenges and constraints to the diffusion of biometrics in information systems. *Commun ACM.* 2005;34(12):101-106.
10. Kushwaha A, Singh SK, Tewari S, Sinha A. Empirical approach for designing of support system in mechanised coal pillar mining. *Int J Rock Mech Min Sci.* 2010; 47(7):1063-1078.
11. Birch I, Raymond L, Christou A, Fernando M.A, Harrison N, Paul F. The identification of individuals by observational gait analysis using closed circuit television footage. *Sci Justice* 2013;53(3): 339-342.