

Chapter 90

Exploring Graph Data Warehouses: Systematic Literature Review

Maja Cerjan

 <http://orcid.org/0009-0003-8825-6050>

Faculty of Organization and Informatics, University of Zagreb, Croatia

Kornelije Rabuzin

Faculty of Organization and Informatics, University of Zagreb, Croatia

ABSTRACT

The necessity to implement new methods of data processing and storage is a result of the global phenomena of data growth. In this article, the idea of graph data warehouses will be presented. A systematic review of the literature, i.e., published studies pertinent to this topic, will be used to give a fair and impartial overview of study findings for this subject. The purpose of this paper is to provide an objective review of graph data warehouses and the most significant studies on the subject. The summarized findings from the research will be presented together with a rating of the quality of the chosen studies.

1. INTRODUCTION

The massive rise in the amount of data that needs to be managed daily is driving the growing popularity of graph databases. Graph databases excel at handling interconnected data and outperform relational databases in queries that involve many related nodes. While relational databases were historically the most popular option for storing and processing, NoSQL (Not Only SQL) database management systems, including graph databases, are now gaining traction.

Traditional relational data warehouses (DWs) further highlight the challenges posed by large data sets. A DW represents an integrated collection of data from several, potentially heterogeneous sources, resulting in massive data volumes and large tables with hundreds of millions or even billions of records. Whether using star or snowflake models, typical DWs always contain extensive data and interconnected records. Despite using well-known data analysis technologies like Online Analytical Processing (OLAP)

DOI: 10.4018/978-1-6684-7366-5.ch090

This chapter published as an Open Access Chapter distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

to provide relevant information, DWs often lack the performance needed to effectively analyze and handle vast amounts of data.

DWs can be thought of as cleaned and integrated data collections designed to offer useful information for decision-making. The term “data warehouse” was coined by W. Inmon in 1992 and has since been widely used in both academic and practical contexts. As DWs and business intelligence systems have been established (or are being installed) in many firms, it has become clear that cleaned and integrated data can provide valuable decision-making information. Initially, these tools were expensive, specialized companies were scarce, and professionals were difficult to find. As a result, only businesses that could afford them implemented such systems in the 1990s and early 2000s.

However, these technologies became more affordable in the second decade of the 2000s. As more people gained knowledge, more businesses began offering these services, leading to broader deployment of DW systems. Compared to the volumes of data managed today, the earlier amounts were relatively small. The data volume is now doubling every two years, making it evident that the size of the data was not previously problematic. Instead, issues stemmed from constrained hardware resources, including limited main memory (RAM), mechanical hard drives (HDDs), insufficient CPU power, and expensive server solutions. While relational databases can handle large amounts of data volumes, they face challenges with today's data sizes. The goal is to reduce data warehousing time, employing various methods to achieve this. One emerging concept is NoSQL databases, which offer advantages over relational databases.

When examining NoSQL databases, it is crucial to discuss the three Vs: Volume, Velocity, and Variety. The first V, Volume, refers to the large data volumes that NoSQL databases can store, surpassing the capacity of relational databases. NoSQL databases are designed to handle heterogeneous data efficiently. The second V, Velocity, pertains to the speed at which data is generated and processed. NoSQL databases should process data faster than relational databases. However, it is important to acknowledge the disadvantages of NoSQL databases, such as their lack of theoretical underpinnings and the absence of a universal language like SQL.

The third V, Variety, refers to the diverse types of data that can be used. NoSQL databases come in four basic forms: column-oriented, document-oriented, key-value, and graph databases. This paper focuses on graph databases and data warehousing and will not discuss the other categories in detail. Different databases support various models, which can be explored by reviewing relevant documentation and articles (Akid & Ben Ayed, 2017; Guminska & Zawadzka, 2018; Liu & Vitolo, 2013; Robinson et al., 2013; Skhiri & Jouili, 2013).

Graph databases are based on graph theory, where nodes in a graph are connected by edges. In graph databases; nodes typically store information about entity instances, while edges link these nodes. Understanding basic graph theory concepts, such as null graphs, pathways, and cycles, is beneficial, although not strictly necessary for comprehending graph databases.

One of the most significant graph database systems in use today is Neo4j. The field of graph databases is continually evolving, with some older systems no longer available. Relational database management systems have evolved into multi-model systems, supporting both relational and graph models.

The purpose of this paper is to examine advancements in graph data warehousing, as graph databases are excellent options for storing and managing large amounts of interconnected data.

The paper aims to provide an objective overview of significant studies on graph DWs. A quality assessment analysis follows a summary of the chosen research findings. Despite the limited attention the subject of graph databases and DWs has received, the most relevant works are included. These works

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/exploring-graph-data-warehouses/354060

Related Content

Issues and Opportunities in E-Business Research: A Simonian Perspective

Ye-Sho Chen, Guoqing Chen and Soushan Wu (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1759-1779).

www.irma-international.org/chapter/issues-opportunities-business-research/19072

Managing Resilient Renewable Supply Chains: Based on the Information of Patient Capital in Post-Subsidy China

Tianxu Lan and Boqiang Lin (2026). *Journal of Global Information Management* (pp. 1-24).

www.irma-international.org/article/managing-resilient-renewable-supply-chains/408723

The Role of Governments in Driving Industry 4.0 Adoption in Emerging Countries: Mediating Effect of Organizational Structure

Muhammad Mohiuddin, Mohammad Nurul Hassan Reza, Sreenivasan Jayashree, Md Samim Al-Azad and Slimane Ed-Dafali (2023). *Journal of Global Information Management* (pp. 1-31).

www.irma-international.org/article/the-role-of-governments-in-driving-industry-40-adoption-in-emerging-countries/323439

Intangible Determinants in the Introduction and Development of Information Technology: Mediterranean Evidence

Ana R. del Aguila, Sebastián Bruque and Antonio Padilla (2003). *Advanced Topics in Global Information Management, Volume 2* (pp. 207-224).

www.irma-international.org/chapter/intangible-determinants-introduction-development-information/4518

Impact of Covert Advertising Characteristics in Short Video E-Commerce on Impulse Buying Behavior: A Dual-Drive Mechanism of Emotion and Cognition

Yi Cui, Yi Shi, Jinni Chen and Jingrong Zhu (2025). *Journal of Global Information Management* (pp. 1-27).

www.irma-international.org/article/impact-of-covert-advertising-characteristics-in-short-video-e-commerce-on-impulse-buying-behavior/389668