International Journal of Data Warehousing and Mining
Volume 21 • Issue 1 • January-December 2025

Research on Application of Electromagnetic Environment Data Warehouse Based on Big Data

Jie Wang
22nd Research Institute of China Electronics Corporation, China

Zhenxing Li
https://orcid.org/0009-0005-7528-8146
22nd Research Institute of China Electronics Corporation, China & Xiamen University, China

Liping Zhou
Qingdao Preschool Education College, China

ABSTRACT

The widespread use of electromagnetic space utilization technology across various fields—including maritime, terrestrial, aeronautical, orbital, electrical, and telecommunications—has generated vast amounts of electromagnetic environment data. To manage challenges, such as the storage of large raw datasets, data discrepancies, isolated data across multiple pathways, and low data value density, a big data-based electromagnetic environment data warehouse is proposed. This warehouse standardizes data from diverse sources, integrates and reconstructs it according to business themes, and uses a mix of relational and non-relational databases for storage. It meets the needs for high data reliability, fast access, and massive storage capacity, offering a solution to data overload while supporting data mining and knowledge discovery in the electromagnetic field.

KEYWORDS

Electromagnetic Environment, Data Warehouse, Big Data

INTRODUCTION

With the continuous improvement in the informatization of electromagnetic environment applications and the long-term operation of related businesses, various departments have accumulated vast amounts of electromagnetic environment data (Chen & Zhao, 2020). This data has played an important role in radio management, spectrum control, electronic warfare, and other areas (Cheng et al., 2019; Ding, 2015; Ha & Jia, 2015). However, there are also many challenges in terms of comprehensive storage and application of the data (Group, 2017):

1. Difficulty in storing massive raw data: Electromagnetic environment sensing devices have powerful data collection capabilities, with fast data output frequencies and diverse data types. Over long periods, these devices accumulate large amounts of data, and existing information systems are unable to meet the storage demands.
2. Large differences in data elements: Currently, China has been building electromagnetic environment sensing devices for different business applications. Since these devices are produced

DOI: 10.4018/IJDWM.373715

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creative-commons.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.

by different manufacturers and follow different data transmission protocols, the resulting electromagnetic environment data varies in source and structure.

3. Data silos: The existing electromagnetic environment data is non-integrated, with massive amounts of data stored independently in each business system. This leads to inefficiencies in data sharing and interoperability, and there is a lack of integrated data applications.
4. Low data value density: Short-term data is insufficient to reflect the state of the electromagnetic environment, and effective analysis can only be achieved using long-term, large-scale data. Therefore, developing a standardized solution with fast storage and highly efficient processing for electromagnetic environment data has become an urgent need to enhance the value of these data applications.

As a subject-oriented, integrated, stable, and time-variant collection of data used to support management decisions, a data warehouse enables the integration of multiple heterogeneous data sources. It can reorganize data according to themes to meet online analytical processing (OLAP) needs, supporting data mining and management decision-making (He et al., 2022). A data warehouse based on Oracle and its related components was constructed to address the storage and processing requirements of radio data, as detailed in relevant studies (Hu, 2017). This data warehouse design was well-suited for scenarios characterized by low data volume, structured data storage, and stable business needs. A solution based on Oracle database and data warehouse technologies had been proposed, specifically tailored for the efficient storage and processing of radio monitoring data (Imran et al., 2021). This solution integrates functionalities, such as clustering analysis, unknown signal prediction, and pattern mining, utilizing intelligent data mining techniques to improve the comprehensive coverage of signals. Furthermore, by incorporating OLAP analysis within a Browser/Server model, the system enables more intuitive and visual data presentation, offering robust support for decision-making. However, with the increasing variety and volume of monitoring data, relying on relational databases, such as Oracle, can result in excessive database load. Additionally, the table structure design may become suboptimal, posing challenges in effectively supporting diverse data retrieval and processing demands.

Big data storage technologies excel at managing data with high timeliness, fast storage speeds, large capacities, and varying quality. One study designed a distributed storage system architecture for massive radio monitoring data, leveraging the Hadoop cloud computing platform and the HBase distributed database to address the storage, retrieval, and analysis requirements of radio regulatory agencies (Inmon, 2006). Another approach employed a combination of Redis and MongoDB, two non-relational databases, to achieve distributed storage for large-scale spectrum data (Lu, 2019). An efficient data storage solution leveraging the distributed characteristics of HBase has been proposed, enabling dynamic design of storage table structures based on query requirements and offering the flexibility to meet diverse storage needs (Pan et al., 2023). Another approach combines MongoDB's replica sets and sharding technology to create a distributed storage architecture that efficiently stores spectrum data across multiple storage server nodes (Tian et al., 2017). Additionally, an integration of HBase with Elasticsearch was introduced to address the challenges of storing and querying large volumes of electromagnetic environment data. This solution provided a robust and efficient framework for managing complex data requirements (R. Wang, 2023). Moreover, the application of cognitive radio technology and spectrum sensing algorithms, such as GSRED, was proposed to enhance the spectrum management process in smart grids, offering significant improvements in real-time data processing and decision-making in dynamic environments (Y. Wang, 2020). A comprehensive framework for an electromagnetic environment data warehouse was proposed, addressing the challenge of managing large-scale electromagnetic data through a multidimensional model (Q. Yang et al., 2020). One study explored the use of open-source NoSQL columnar databases to build data warehouse solutions in such environments, further supporting the need for flexible, scalable storage technologies (X. Yang, 2023).

19 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/article/research-on-application-of-electromagnetic-environment-data-warehouse-based-on-big-data/373715

Related Content

Sentiment Mining Approaches for Big Data Classification and Clustering
Ashok Kumar J, Abirami Sand Tina Esther Trueman (2018). *Modern Technologies for Big Data Classification and Clustering (pp. 34-63).*
www.irma-international.org/chapter/sentiment-mining-approaches-for-big-data-classification-and-clustering/185978

Recommender Systems Using Collaborative Tagging
Latha Banda, Karan Singh, Le Hoang Son, Mohamed Abdel-Basset, Pham Huy Thong, Hiep Xuan Huynhand David Taniar (2020). *International Journal of Data Warehousing and Mining (pp. 183-200).*
www.irma-international.org/article/recommender-systems-using-collaborative-tagging/256169

Experimental Study I: Automobile Dataset
(2018). *Predictive Analysis on Large Data for Actionable Knowledge: Emerging Research and Opportunities (pp. 91-110).*
www.irma-international.org/chapter/experimental-study-i/196390

The Power of Sampling and Stacking for the PAKDD-2007 Cross-Selling Problem
Paulo J.L. Adeodato, Germano C. Vasconcelos, Adrian L. Arnaud, Rodrigo C.L.V. Cunha, Domingos S.M.P. Monteiroand Rosalvo F. Oliveira Neto (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments (pp. 297-306).*
www.irma-international.org/chapter/power-sampling-stacking-pakdd-2007/40412

TripRec: An Efficient Approach for Trip Planning with Time Constraints
Heli Sun, Jianbin Huang, Xinwei She, Zhou Yang, Jiao Liu, Jianhua Zou, Qinbao Songand Dong Wang (2015). *International Journal of Data Warehousing and Mining (pp. 45-65).*
www.irma-international.org/article/triprec/122515