



Exploration of various Models used for Online Analytical Process (OLAP) : A Review

Anjana Yadav

Department of Computer Science
P. K. University
Shivpuri, India
yadavanjana3485@gmail.com

Balveer Singh

Department of Computer Science
P. K. University
Shivpuri, India
adm.pk@gmail.com

Abstract— The Online Analytical Processing (OLAP) is highly suitable for examination and processing of healthcare data by utilizing the consistent models in IoT devices. The OLAP models are introduced on large medical data warehouses for computation and decision making to reduce the processing overhead of data in IoT devices. The multidimensional, relational and hybrid representation of healthcare data are specifically used for describing the OLAP models. The multidimensional data cube model is examined in sturdy time for processing of OLAP queries over IoT devices. The query processing is highly expensive and time consuming for large medical data warehouses. This cost and time of query processing are reduced by using several strategies and heuristics to offer efficient OLAP examination by improving the performance materialized medical data cube selection strategy in IoT devices. The query processing is also improved by using the data clustering over huge data warehouses storing the different types of information such as medical, defence, industrial and confidential transmitted through IoT devices. These numerous amounts of medical information are easily and quickly accessed and processed by IoT users, if the data is arranged into groups using clustering rather than in raw form. The main motive of this research paper is given an extensive collection of valuable healthcare data about OLAP analysis, models, query processing techniques, problems and optimization methods in various environments using IoT devices. Hence, the researchers can be utilized the important medical data over IoT devices to implement novel and efficient techniques for OLAP data analysis. A brief literature of prior researches is further illustrated in textual and tabular arrangements to give the inspiration to researchers for implementing innovative hypothesis for OLAP data examination.

Keywords—Data Clustering, IoT, Materialized Data Cube, OLAP, Optimization, Query Processing, Warehouse

I. INTRODUCTION

The OLAP [1] is developed for business data examination by performing the tough computations, trend examination, and complicated data modelling. The OLAP is composed of a wider grouping of business intellect, which in addition combines concerned database, details script and data clustering. The OLAP tools facilitate customers to examine multidimensional information collectively from various possibilities [2]. The OLAP lies of 3 common models Relational Online Analytical Processing (ROLAP), Multi-dimensional Online Analytical Processing (MOLAP), and Hybrid Online Analytical Processing (HOLAP).

A. MOLAP Model

The MOLAP is a typical variety of OLAP [3] which is occasionally indicated as simply OLAP. It saves the information in multi-dimensional array space optimally, not in related tables. The efficiency of query processing strategy is enhanced by utilizing the index and cache in multi-dimensional space. Several compression methods are introduced for reducing the information size in MOLAP model to reduce the query processing time and cost.

B. ROLAP Model

The ROLAP functions straight by means of relational information which is not involve any prior evaluations. The relational tables are used for saving the base and dimensional information. The recent tables are formed to seize the clan data stooping on expert schema architecture. It remains on stacking the saved information in relational tables to provide the emergence of slicing and washing up processes of conventional OLAP [4]. It is measured in the direction of highly flexible in managing huge data warehouses, mainly dimensional models having extremely elevated cardinality. The ROLAP design is provided several benefits: (i) it is simply incorporated into former previous relational data architectures and (ii) the relational information is saved perfectly and precisely as compare to multidimensional information [5].

C. HOLAP Model

The detrimental transaction between extra price and deliberate query processing has checked that largely mercantile OLAP tool currently utilize a HOLAP model, which permits the model architecture to make a decision which fraction of information would be saved in MOLAP and which fraction in ROLAP [6].

The data warehouse is an accumulation of enormous details of various data regarding numerous societies exploiting for appraisals. These appraisals are occupied through multifarious queries introduced to data warehouse and models for decreasing the reply instant. The OLAP models are exploited to process the query for data warehouse and excerpting the needful information for examination [7]. The query running cost and time is also major concerns for data models to illustrate the efficiency and power of OLAP. The business intelligence [8, 9] utilizes the intellectual properties, rules and laws for examining the personal

and social information for solving the problems considering in future aspects. The data information is combined into groups (clusters) for fast accessing and query processing through OLAP models. All the types of OLAP models are achieved different level of query running speed through several optimization strategies and clustering methods [10, 11].

II. LITERATURE REVIEW

The trajectory information permits the revise of nature of affecting elements, from person to creatures by utilizing the wireless transmission, portable machines and techniques. The key concern of the spatial database management system is saving and processing the huge volume of informative data through OLAP models. The indexing, saving and extracting the spatial information from big data warehouses is performed by using various data management techniques. The problems arising in trajectory database and the solutions of problems in big data analysis are combined in tabular format for explaining the use of several OLAP data management techniques to improve the quality of decision making [12].

The contemporary day progression is gradually more digitizing the real life scenarios with quick enhancement of information. The novel and important knowledgeable information are extracted from data warehouses with the help of multidimensional data models. Although the Hadoop architecture is well defined policy to deal with big datasets and it has several computing architectures for multiple application fields. This creates a strategy for dividing the big data into several groups over cloud computing platform. The key concerns and challenges of big data analysis are identified and solutions are proposed for data analysis over cloud environment [13].

The organization of COVID-19 epidemic shows various extraordinary issues in numerous areas from drug to biology, which might advantage from examination techniques capable to amalgamate the enhancing existing COVID-19 and associated information such as effluence, demographics, and weather. On the basis of this information, a COVID-WAREHOUSE is developed for saving and processing the COVID-19 data of effluence and weather. The time and environment position are two major factors of dimensional fact model of OLAP examination [14].

A console application is anticipated and urbanized to proceed as an identical twin which can be notated the precise significance of sharing responsibility for few future crashes. The present learning illustrates a technique to utilize the historical information associating with construction presently and clan from IoT for detracting the future activities of the construction, even as representing the calculated values which are conscientious for negative construction presentation, devoid of guidance. This console application is implemented in java language and key performance indicators are used to visualize the OLAP model for verification and revelation purpose [15].

The precise feeling is the major concern to anatomical fitness sensing of subversive petroleum mines, although utilizing nerve Bragg strident sensors. On the other hand, the earlier urbanized machines for architectural invigilating of subversive mines contain restricted to invigilating devoid of some strength of harm exposure. Consequently, this paper incorporates an extremely precise invigilating machine on an Internet of things (IoT) framework over Web 2.0 server. The prime element examination, besides hierarchical clustering, is utilized to locate the scratch pointer of the mine. The scratch manifestation is verified, representing the least value for rigidity diminution. Therefore, incorporation of this method with Internet would be efficiently introduced for premature security measurement of subversive petroleum mines and data contribution in real time [16].

The insidious medical services are recognized several components like internet, ad-hoc environment, and transmission techniques to provide better solutions of challenges in medical system. The health information is taken from the IoT devices and

utilized with the help of machines and OLAP models. The OLAP and data mining are used healthcare data information taken from various heterogeneous sources. The platform utilizes several rules and laws to calculate quality and processing speed of medical data by using Hadoop Map/Reduce interface [17].

The solar energy data are generated from sensors introduced in various geographic positions and systems of weather organizations. Still, the Portable Document Format (PDF) and Hyper Text Mark-up Language (HTML) formats of files are not given meaningful solar energy data extracting from various resources. So, a query platform is developed with solar data processing, where the data is taken and combined from various heterogeneous resources [18].

At present era the industry environment demands contribute sequences to be as compared to theoretical, which need a novel analysis method for data mining sortilege. A sortilege model is provided the combination of progression, routine and data mining models. This model is also calibrated with the business intelligence to evaluate the efficiency and performance through rules and key performance indicators. The real life datasets are used for modelling the sortilege model with business intelligence. A particular methodical web entrance offering, which proposes concerted efficiency invigilating and assessment creation, is offered. The outputs present that the models confer extremely precise key performance indicators throws and give expensive presence into novel promising chances and problems [19].

The model having relational information is possibly the maximum habitually utilized database structure; still, complex queries for huge dataset are not preferably run and examined by relational model. The OLAP concept is introduced to evaluate the multidimensional data for online processing and examination. The business intelligence has grown with improving the OLAP model facilitating the efficient data cube selection. The ROLAP and MOLAP have improved their performance against query processing time and cost to store the reporting time of working hours over economical data. The outcomes represent the better quality performance of ROLAP as compared to OLAP for performing data cubing [20].

The telecom organizations have to inflate the services with cheapest price on the basis of customer requirement information along with call detail record and behaviour of purchase. This model is developed and designed abstractly, reasonably and physically to solve the problems of data mart in sufficient time. The OLAP model is developed to provide superior performance nearer to customers purchase nature and enhance the marketing of goods [21].

The conceptual models [22] are major concerns in OLAP models using the data warehouse, which improves the use of logical models for better performance. The conceptual methods have several limitations like maximum learning arc, not easily understandable and flexible in user friendly environment. These techniques are more complex to study as well as analysis for knowledge workers. Then the cube algebra is introduced as a conceptual structural model providing the maximum level of knowledge about the database and OLAP to workers. The undesired information is hidden from the unknown users for security purpose [23].

The OLAP is utilized for multidimensional data representation and processing for various application areas. The cube presentation model is presented and examined over unified modular language to show the data cube more precisely and accurately. Hence needful data are extracted for different organizational systems and represented through data cubes using extensible mark-up language and unified modular language to enhance the visualization strength of information of multidimensional data cubes [24].

A major amount of power is addicting in industrial building zone, ensuing in several undesirable concerns. A data cube model is introduced with relationship rule mining applying over

industrial buildings power expenses survey dataset (6700 industrial buildings) to diminish the power expenses and enhance the power efficiency in industrial buildings. The OLAP is applied to power expenses industrial data to analyze the power based on atmospheric conditions, amalgamated equipment, construction types and cooling systems [25].

The query running costs and time is decreased by using an optimal group of materialized data cubes in the data warehouse. The Particle Swarm Optimization (PSO) [26] is well suitable algorithm utilizing for optimal selection of the data cubes. The speed of PSO is higher as compare to other greedy and heuristic techniques. The global optimal results are also achieved by the PSO in terms of materialized data cube selection. The PSO is applied to a collection of data cubes to find out the best cubes to reduce the query running time cost. The exploration and exploitation power of PSO is superior for searching the local and global optimum values of data cubes enhancing the accuracy and performance of the system. The results represent the better quality efficiency of PSO against the optimization technique like Genetic Algorithm (GA) based on multiple performance factors [27].

The data clustering is introduced to examine the numerous data, where the data are divided into multiple partitions for further processing. Therefore, the data is easily accessible in least time for users to save the extra cost of query processing in data warehouse. The K-Means are a famous method to partition the data into clusters for data analysis. Several optimization algorithms are also introduced for data clustering to generate optimal clusters of data for reducing the computational cost and time over huge data information of data warehouses. The PSO is one of the best utilized approaches to improve the strength of data clustering with least error rate and highest convergence speed as compared to other clustering techniques [28].

Table 1 represents the various works of researchers in comparative way.

TABLE I. COMPARATIVE ANALYSIS OF OLAP BASED RESEARCH WORKS

Author s	Comparative Analysis				
	Approach/M ethod	Applicatio n Area	Platfor m	Advantag es	Limitatio ns
G. Agapito et. al. [14]	Dimensional Fact Model & OLAP	Medical (COVID-19)	Python	Used automatic extraction, transformation and loading	Not used graphical user interface
A. Papacharalam popoulos et. al. [15]	Key performance indicator & OLAP	Production System	Java	Used the IoT to indicate the future failures	Not used differential order and knowledge based libraries
B. W. Jo et. al. [16]	Fiber Bragg Grating	Coal Mines	Web 2.0	Hierarchical clustering utilized for harm pointer of mine	Subversive mines unsympathetic situation
J. N. S. Rubi et. al. [17]	Internet of medical things	Healthcare	e-Health Sensor Kit	Automatic preparation of data and knowledge extraction method	Through put, and average time is not evaluated
J. L. S. Carvantes et. al. [18]	Solar radiation extraction and query platform	Weather Stations	Sensor Web Enablement	Reuse the data and develop the web application	Sensors not used like Thermometer, hydrometer etc.

Author s	Comparative Analysis				
	Approach/M ethod	Applicatio n Area	Platfor m	Advantag es	Limitatio ns
N. Stefano vic [19]	Supply Chain Model & Key performance indicator	Business Intelligence	Web Portal	Global, Collaborative, predictive analysis	Not support visual intelligence.
P. Westerlund [20]	ROLAP & MOLAP	Business Intelligence	Structured Query Language (SQL)	Server side data analysis	Time Consuming and costly
D. Camilovic et. al. [21]	OLAP	Data Mart	SQL	Dynamic data processing	Not provide any time and cost model
C. Ciferri et. al. [23]	Cube algebra model & OLAP	Pollution Control	SQL	User friendly model with flexible design	Not include spatial and multimedia data
A. S. Maniatis [24]	Cube presentation model and OLAP	Rational Rose	XML and UML	Stereotype extension of data	Not provide visualization and automatic generation
B. Noh et. al. [25]	Data cube model with relationship rule mining	Industrial building data	R Tools	Evaluate the power strength	The time and harmful effects not considered
A. Gosain et. al. [27]	PSO	Sales data	MATLAB Tool	Several frequencies and dimensions are taken	Processing time is not calculated

III. CONCLUSION

The reliable OLAP models are employed to analyze and deal with important informative data of data warehouse. The overhead of handling the huge data and making the decision for data processing is diminished through OLAP models. The ROLAP, MOLAP and HOLAP are three popular models for data warehouse to explain and organize the data into multidimensional, relational and hybrid structures. The query processing expenses and time are abridged by introducing various methods and optimization techniques with multidimensional data model to enhance the OLAP performance. The efficiency of OLAP model is also enhanced by utilizing the data clustering methodologies, in which the important huge data information such as medical, industrial and secret data is combined into clusters. So, data clustering is improving the accuracy and accessing capability of OLAP models over data information. The key concern of this research paper is specified a widespread compilation of precious data about OLAP examination, models, query performing, problems and optimization techniques in multiple eras. Therefore, the researchers can be used the necessary information to develop novel and proficient strategies for OLAP data examination. The text and tabular arrangements of previous works are further utilized for innovative OLAP strategy selection.

REFERENCES

- [1] N. Jukic, B. Jukic, and Malliaris, "Online Analytical Processing (OLAP) for Decision Support," pp. 1-25, 2008.
- [2] E. Emmanuel, A. Obiageli and V. Osinachi, "Design and Implementation of Multidimensional Students Result Analytical Processing for Tertiary Institutions," *International Journal of Engineering and Computer Science*, vol. 8, Issue 8, pp. 24814-24828, 2019.
- [3] M. Eltabakh, "OLAP & Data Mining," Spring, WPI, pp. 1-38, 2012.
- [4] D. Bulos and S. Forsman, "OLAP Database Design," *Symmetry Corporation*, pp. 1-19, 2019.
- [5] I. L. Cruz, R. Berlanga, and M. J. Aramburu, "Modelling Analytical Streams for Social Business Intelligence," *Informatics, MDPI*, vol. 5 (53), pp. 1-17, 2018.
- [6] L. Shen, S. Liu, S. Chen, and X. Wang, "The Application Research of OLAP in Police Intelligence Decision System," *International Workshop on Information and Electronics Engineering (IWIEE)*, Elsevier, vol. 29, pp. 1-6, 2012.
- [7] A. Vaisman and E. Zimanyi, "Mobility Data Warehouses. *International Journal of Geo-Information*," MDPI, vol. 8 (170), pp. 1-22, 2019.
- [8] S. Venkatraman, "A Proposed Business Intelligent Framework for Recommender Systems," *Informatics, MDPI*, vol. 4 (40), pp. 1-12, 2017.
- [9] W. Fuertes, F. Reyes, P. Valladares, F. Tapia, T. Toulkeridis, and E. Perez, "An Integral Model to Provide Reactive and Proactive Services in an Academic CSIRT Based on Business Intelligence," *Systems, MDPI*, vol. 5 (52), pp. 1-20, 2017.
- [10] J. G. Aher and V. A. Metre, "Clustering Multidimensional Data with PSO based Algorithm," *Soft Computing and Artificial Intelligence*, pp. 1-6, 2014.
- [11] W. Q. Qwaider, "Apply On-Line Analytical Processing (OLAP) With Data Mining For Clinical Decision Support," *International Journal of Managing Information Technology (IJMIT)*, vol. 4 (1), pp. 1-13, 2012.
- [12] D. R. D. Almeida, C. D. S. Baptista, F. G. D. Andrade, and A. Soares, "A Survey on Big Data for Trajectory Analytics," *International Journal of Geo-Information, MDPI*, vol. 9 (88), pp. 1-24, 2020.
- [13] S. Ullah, M. D. Awan and M. S. H. Khiyal, "Big Data in Cloud Computing: A Resource Management Perspective," *Scientific Programming, Hindawi*, pp. 1-18, 2018.
- [14] G. Agapito, C. Zucco, and M. Cannataro, "COVID-WAREHOUSE: A Data Warehouse of Italian COVID-19, Pollution, and Climate Data," *International Journal of Environment Research and Public Health, MDPI*, vol. 17 (5596), pp. 1-22, 2020.
- [15] A. Papacharalampopoulos, C. Giannoulis, P. Stavropoulos, and D. Mourtzis, "A Digital Twin for Automated Root-Cause Search of Production Alarms Based on KPIs Aggregated from IoT," *Applied Science, MDPI*, vol. 10 (2377), pp. 1-16, 2020.
- [16] B. W. Jo, R. M. A. Khan, Y. S. Lee, J. H. Jo and N. Saleem, "A Fiber Bragg Grating-Based Condition Monitoring and Early Damage detection System for the Structural Safety of Underground Coal Mines Using the Internet of Things," *Journal of Sensors, Hindawi*, pp. 1-16, 2018.
- [17] J. N. S. Rubi and P. R. L. Gondim, "IoMT Platform for Pervasive Healthcare Data Aggregation, Processing, and Sharing Based on OneM2M and OpenEHR," *Sensors, MDPI*, vol. 19 (4283), pp. 1-25, 2019.
- [18] J. L. S. Cervantes, M. Radzimski, C. A. R. Enriquez, G. A. Hernandez, L. R. Mazahua, C. S. Ramirez, and A. R. Gonzalez, "SREQP: A Solar Radiation Extraction and Query Platform for the Production and Consumption of Linked Data from Weather Stations Sensors," *Journal of Sensors, Hindawi*, pp. 1-19, 2016.
- [19] N. Stefanovic, "Proactive Supply Chain Performance Management with Predictive Analytics," *The Scientific World Journal, Hindawi*, pp. 1-18, 2014.
- [20] P. Westerlund, "Business Intelligence: Multidimensional Data Analysis," *ECTS Credits*, pp. 1-58, 2008.
- [21] D. Camilovic, D. B. Vujaklija and N. Gospic, "A Call Detail Records Data Mart: Data Modeling and OLAP Analysis," *ComSIS*, vol. 6, No. 2, pp. 87-110, 2009.
- [22] A. Tsois, N. Karayannidis and T. Sellis, "MAC: Conceptual Data Modeling for OLAP," *International Workshop on Design and Management of Data Warehouses*, pp. 1-12, 2001.
- [23] C. Ciferri, R. Ciferri, L. Gomez, M. Schneider, A. Vaisman, and E. Zimanyi, "Cube Algebra: A Generic User-Centric Model and Query Language for OLAP Cubes," *International Journal of Data Warehousing and Mining*, pp. 1-23, 2012.
- [24] A. S. Maniatis, "OLAP Presentation Modeling with UML and XML," pp. 1-10, 2005.
- [25] B. Noh, J. Son, H. Park, and S. Chang, "In-Depth Analysis of Energy Efficiency Related Factors in Commercial Buildings Using Data Cube and Association Rule Mining," *Sustainability, MDPI*, vol. 9 (2119), pp. 1-20, 2017.
- [26] J. Loureiro and O. Belo, "A Discrete Particle Swarm Algorithm for OLAP Data Cube Selection," *8th International Conference on Enterprise Information Systems-DISI*, pp. 46-53, 2006.
- [27] A. Gosain and Heena, "Materialized Cube Selection using Particle Swarm Optimization algorithm," *7th International Conference on Communication, Computing and Virtualization, Elsevier*, pp. 2-7, 2016.
- [28] J. G. Aher and V. A. Metre, "PSO based Multidimensional Data Clustering: A Survey," *International Journal of Computer Applications*, vol. 87, No. 16, pp. 41-48, 2014.