A Comprehensive Survey of OLAP: Recent Trends

Adhish Nanda (*Student*)
Dept. of Computer Science, Amity University Haryana, Haryana, India
adhish.nanda@gmail.com

Swati Gupta, Meenu Vijrania (*Asst. Prof.*)
Dept. of Computer Science, Amity University Haryana, Haryana, India
Sgupta.ggn@amity.edu, mvjarania.ggn@amity.edu

Abstract— **A Multi-Dimensional Database (MDD) is a type of database for performing Data Warehousing and OLAP Operations. They are often made contribution for existing relational databases. The On-Line Analytical Processing (OLAP) is a Multidimensional database that is being used for discovery of information, which includes the abilities of complex analytical estimations, endless report views and prescient situation (conjecture) outline. OLAP discovers its applications in different Business Intelligence (BI) Objectives. Its main parameter is multidimensional facts and data. In this paper, we cover the architectural components and existing OLAP technologies. We have also figured out the usage of each existing technologies. Further, we have framed out the detailed analysis of research on OLAP.**

*Keywords— **OLAP, Business Intelligence, Multi-Dimensional, Database, MOLAP, ROLAP***

I. INTRODUCTION

The Data Warehouse (DW or DWH) is a framework that is being utilized for revealing the information for analyzing the historical data to derive out special intelligence [1]. The Data warehouse are local archives that contains incorporated information that are being collected from multiple sources. They store current and verifiable information at one single place [2] that are being utilized in making investigative reports by the specialists. [3] This information which is stored in the stockroom (DW) is transferred from the operational frameworks, (for example, showcasing or deals). This information may go through data cleansing [2] for extra tasks to guarantee quality of information before it is utilized in the given Data warehouse for broadcasting. The procedure of Extracting, Transforming, Loading (ETL)-based data warehouse [4] utilizes arranging the information, combining and access layers to house its key capacities. One has to stack their data warehouse routinely with the goal that it can fill its need to aid in business analytics.

To do this, data from at least one of the operational frameworks should be separated and duplicated into multi-dimensional warehouse. The challenges that are being incorporated in data warehouse settings is to coordinate, revise and unite huge volumes of data over numerous frameworks. Thus, giving another consolidated database for business intelligence.

In Extraction the data is being fetched from the different source systems such as (SAP, ERP other operational systems). The data which is being extracted from the given different source systems is than converted into an appropriate consolidated warehouse format which is ready for transformation processing. The business rules are being applied in transforming the data which involves data cleaning, filtering of the given data, splitting the column into multiple columns and vice-versa, joining together data from multiple sources, apply the transposing of rows and columns, it also involves applying any kind of simple or complex data processing etc. In loading phase, the data is being loaded or inserted into data warehouse or repository for the different reporting applications. Some advantages of ETL Phase [4] include enhanced business intelligence, increase query and system performance, timely access to data, enhanced quality and consistency with high returns on investments.

A data warehouse is basically a combination of facts and dimensions. A fact is defined as a value that represents given value. A dimension represents a structure that contains the certainties and measures in order to empower clients so as to address business questions. Ordinarily utilized measurements are often referred as individuals, items or it may refer to any place and time.[5][6] In the given scenario a dimension represent a structured labelling information. Basically, a dimension is a combination of multiple data sets which are composed of individual, non-overlapping data elements. The primary functions of dimensions include characterizing the data set, applying a filtering technique and grouping the data set.

In data warehousing, a fact table is a combination of certain estimations, measurements or metrics for a given business procedure. It lies at the central position for a star schema or a snowflake schema comprises of several dimensional tables representing each dimension. In a scenario where there are numerous facts tables such schemas are organized as a fact constellation schema.

Basically, a fact table comprises of two kinds of segments: those that contain certainties also known as measures and those that act as a foreign key to dimension tables. The primary key of a fact table is normally a composite key which majorly comprises of its foreign keys. It consists of the main contents of data warehouse and store diverse kinds of measures like aggregated, non-aggregated and semi-aggregated measures. The OLAP data can be primarily used for data mining technique such as regression analysis and to preserve consistency in database [4, 5].

Dimensions classify and depict data warehouse actualities and measures in manners that significantly help to get responses to business queries [7] A data warehouse composes descriptive traits or attributes as sections or columns in dimension tables. For instance, a client dimension's characteristics could incorporate first and last name, birth date, sexual orientation, and so forth., or a site dimension would incorporate site name and URL qualities. It has a primary key segment that particularly recognizes each dimension record (row). The dimension table is related to a fact table utilizing this key. Information in the fact table can be sifted and assembled ("cut and diced") by different combinations of characteristics, traits, and attributes. While making a dimension table in a data warehouse, a key generated by the system is utilized to particularly distinguish a row in the tables.

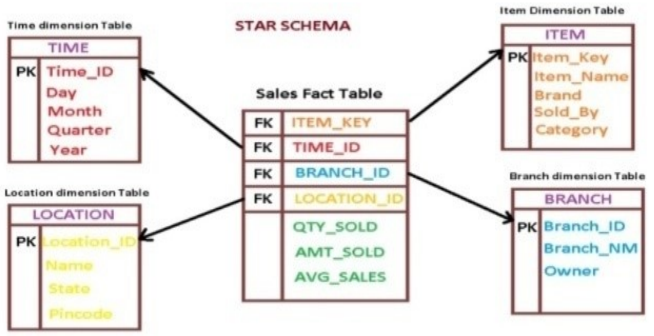

Fig. 1. A Fact and Several Dimension Tables [9]

The special key also known as surrogate key acts as the primary key for the dimension table. This surrogate key is present in the fact table and a foreign key is characterized between the two tables. At the point when the information has to be aggregated, it does as such similarly as some other joins inside the database. Like fact tables, dimension tables are frequently very de-standardized, in light of the fact that these structures are not built to oversee exchanges, they are built to empower clients to examine information as effectively as could reasonably be expected.

II. OLAP ARCHITECTURE

An OLAP systembasically consists of four stages. The most basic one is the process of locating the data sources which might include operational databases, relation tables, flat files and other external sources. [10] The data collected from these sources are heterogeneous in nature and may be structured, unstructured or semi-structured.

Thus, the data is passed through the ETL tools that convert it into form suitable for processing. Then the data is passed on to the second stage, which is also known as Bottom Tier. It consists of data warehouse, data marts and metadata repository. It also performs monitoring and administrative functions. The data is organized here and arranged for processing.

The third stage or the Middle Tier is the place where all the transactions take place. It consists of OLAP servers that are responsible for performing various operations on the multidimensional data cubes. The data is ordered and forwarded to the last stage i.e. Top Tier where it is utilized for analysis, reporting, summarizing, visualization, data mining and other applications as shown in figure.
Basically, a Data warehouse architecture comprises of following main components

- Bottom Tier: Database and Data Marts
- Middle Tier: OLAP Servers which stores the huge massive amount of data.
- Top Tier: User view level which specify how the user uses the warehouse data.

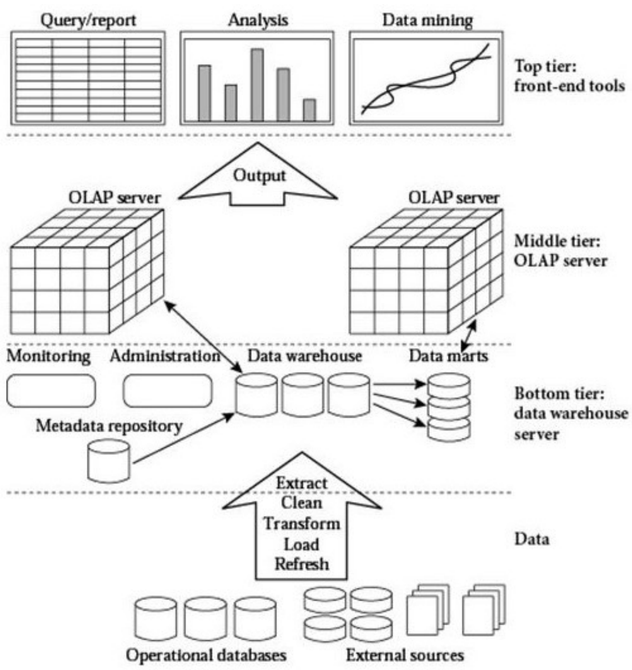

Fig. 2. Architecture of an OLAP System [11]

III. OLAP TECHNOLOGIES

An organization maintains huge amounts of data which is a critical resource that needs a powerful tool to fetch query information. Currently, there are three OLAP technologies which are as follows:

A. MOLAP

MOLAP (Multi-dimensional Online Analytical Processing) as shown in Fig. 3 is the exemplary type of OLAP and is now and again alluded to as just OLAP [10]. In MOLAP the information is being stored in a streamlined multi-dimensional exhibit stockpiling, instead of social database. Some MOLAP apparatuses needs the pre-calculation and storage of determined data, for example, solidifications – the activity known as handling. Such MOLAP instruments, for the most part, use a pre-determined informational index alluded to as data cubes.

Advantages: -

1. Multidimensional Indexing and Optimize Storage provides fast processing of queries.
2. It automatically computes an aggregation at higher level of data and is compressed for low dimensional data sets.
3. Efficient data elicitation can be achieved due to pre-processing of aggregated data and natural indexing is provided by the array model.

Disadvantages: -

1. The data load in the processing step is quite lengthy, especially, for large volume of data.
2. Data redundancy is also present.

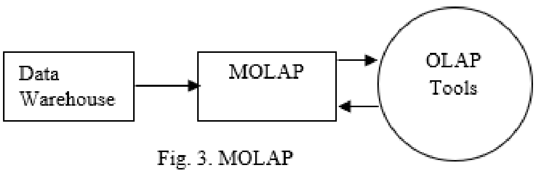

Fig. 3. MOLAP

B. ROLAP

ROLAP (Relational Online Analytical Processing) directly works with Relational databases as it does not require any pre-processing of data as shown in Fig. 4[10]. The relational tables contain the base information and the dimensional tables and new tables are made to hold the amassed data. It relies upon a specific schema structure. This philosophy depends on controlling the information stored in the relational database to give the presence of conventional OLAP's slice and dice usefulness. Generally, each activity of slicing and dicing is proportionate to including a "WHERE" provision in the SQL explanation. ROLAP apparatuses do not utilize pre-determined information from data cubes.

Advantages: -

1. It is very scalable in nature and can handle large volumes of data and dimensions with high cardinality.
2. Data can be accessed by any kind of SQL Tool.
3. To perform ETL a variety of data loading tools are available.
4. They are better at handling non-agreeable facts. The ROLAP approach can use database approval controls, for example, providing security at row, where the query results are separated relying upon preset some criteria.

Disadvantages: -

1. The tools used in ROLAP are slower in performance than MOLAP in most of the cases and does not help with managing the loading of aggregate tables.
2. Since ROLAP devices depend on SQL for the majority of the calculations, as a result they are not appropriate when the model is overwhelming on figuring's which does not make an interpretation well in SQL.

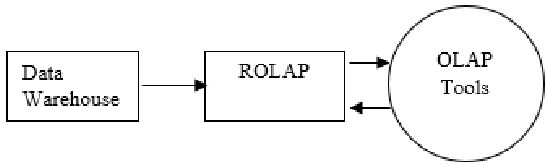

Fig. 4. ROLAP

C. HOLAP

In HOLAP (hybrid online analytical processing) the unwanted exchange between extra ETL cost and moderate query execution resulted in a "Hybrid OLAP" (HOLAP) approach as shown in Fig. 5. It enables in deciding which part of the information can be placed in MOLAP and ROLAP. There is no fundamental understanding over the business with respect to what establishes "Hybrid OLAP", then again, actually a database will isolate information among relational and specially designed storage.[12]

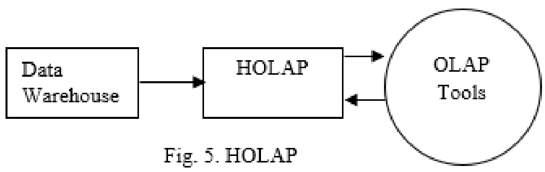

Fig. 5. HOLAP

For instance, the certain merchants, in a HOLAP database will utilize relational tables to hold the less detailed information or aggregated data and utilize particular functional storage for probably a few parts of the less-detailed info or more-aggregated data.

HOLAP overcomes over the shortcomings of MOLAP and ROLAP by joining the abilities of the two methodologies. It instruments can be use both pre-determined cubes and relational information sources by two given techniques

- Vertical Partitioning

In this mode HOLAP stores conglomerations in MOLAP for quick query execution, and remaining information in ROLAP to upgrade cube handling.

- Horizontal Partitioning

In this mode of HOLAP certain slice of information is being stored for fast query execution. In addition, we can also store a few dices in MOLAP and others in ROLAP [13][14]

IV. PREVAILING TRENDS: OLAP TECHNOLOGIES

The choice of OLAP Technologies depends upon the incurred benefit of OLAP like improving decision making which is dependent upon the existing problem [7]. It is also used when consistent information is required for period of time. For making decisions based on current data ROLAP is used. The table below shows OLAP technologies and their usage in enterprises.

TABLE I. Usage of Different OLAP Technologies

OLAP Technology	Features	Adopted By
MOLAP	High performance, Less scalable	Adopted by Microsoft SQL Server 2005, Essbase Server from Hyperion [7]
ROLAP	Low performance, More scalable	Microsoft SQL, Micro Strategy DSS Server, Informix Meta Cube
HOLAP	High performance and Scalable	SAS Server

V. CURRENT ONGOING RESEARCH ON OLAP: LITERATURE REVIEW

Plenty of research work has been done in the area of OLAP. Some of the important work in the area is as follows: -

1. P. Vassiliadis [14]

This paper discusses about storage of multi-dimensional data on different structures and operations that can be performed on them. On-Line Analytical Processing (OLAP) represents a pattern in database innovation, which was as of late presented and has thrown the light of a legitimate concern for a great deal of research work. It depends on the multidimensional perspective on information, which is represented either by multidimensional databases (MOLAP) or relational databases (ROLAP).

In this paper, a model is proposed for multidimensional databases. Measurements, dimension chains of importance and cubes are formally presented. They additionally present cubic operations (changing of levels in the dimension order, work application, route and so on.). The methodology depends on the idea of the base 3D square, which is utilized for the computation of the aftereffects of operations on data cubes.

2. A. D. Patil and N. D. Gangadhar [16]

Online Analytics Processing (OLAP) create multidimensional operations which empowers inquiries and visualization for Business Intelligence (BI). The OLAP systems comprises of a user interface for query and then portray the given information without the use of core OLAP tasks. The BI applications can be created and made to complex work flows for the OLAP tasks that are accessible as an API. This paper addresses the structure and prototyping of an OLAP based Platform as a Service, named OLAP as a Service (OLaaS) which uncovers the core OLAP tasks of OLAP Cube configuration

3. Zhang et al. [17]

OLAP models can be ordered with two techniques namely: MOLAP (multidimensional OLAP) and ROLAP (Relational OLAP). For effective multi-dimensional processing the MOLAP is used while ROLAP lessens the information. In this paper, the authors proposed a novel fusion for the given OLAP model to meld the multidimensional processing model and relational stockpiling model together to make the best parts that work for both MOLAP and ROLAP architectures.

4. Bellatreche et al. [2]

OLAP clients vigorously depend on the representation of answers of queries for their intelligent analysis and interactive examination of gigantic measures of data. Frequently, these answers can't be envisioned altogether, and the client needs to explore through them to find significant actualities. In this paper, the author proposes a system for customizing OLAP inquiries. In this structure, the client is approached to give his (her) inclinations and a representation imperative that can be for example the impediments forced by the gadget used to show the response to a question. Given this, for each inquiry, their technique figures the piece of the appropriate response that regards both the client inclinations and the representation limitation.

Likewise, a customized structure for the perception is proposed.

In this paper, the author has proposed a methodology for customizing answers to OLAP inquiries over Cubes. All the more definitely, they have appeared to show the most fascinating subset of realities with regards to the response to an OLAP query, just as a perception for it. Both the subset of certainties and the representation are processed w.r.t. the client profiles, communicated as a pre-requesting over individuals of Cubes and as a visualization limitation. They have likewise considered the outcomes of personalization on OLAP query enhancement.

5. Gómez et al. [6]

In current Big Data schemes, data warehousing and Online Analytical Processing (OLAP) procedures on Cubes are not adequate to address the present data processing prerequisites. However, OLAP operations and model search expand the conceivable outcomes of graph examinations. In spite of this, a much work on the issue of OLAP analysis to the graph data model is not adequate. In their last work, the authors have proposed a multidimensional (MD) data model for graph scrutiny, that considers both the basic and background data as dimension hierarchies.

A comparative analysis of research work has been carried out specifying the major contributions summarized in table 2.

TABLE 2. Overview of Research on Olap

SNo.	Author	Major Contribution
1	P. Vassiliadis	Mapping of Multidimensional Model to Multidimensional Arrays
2	A. D. Patil and N. D. Gangadhar	Created an OLAP based PaaS Model
3	Zhang et al.	Proposed a Novel Fusion of an OLAP Model which fuses the Multidimensional Computing Model and Relational Storage Model
4	Bellatreche et al.	Framework for personalizing OLAP queries and a personalized structure for Visualization
5	Gómez et al.	Proposes a Multidimensional (MD) data model for graph analysis

VI. CONCLUSION

In this paper, we discussed OLAP Technologies and their different accessing methods used for multi-dimensional database. Though MOLAP and ROLAP comprises of different features, both are considerably extending their services to cater the real-time decision-making capabilities. We have successfully summarized the most important research available on OLAP Operations in the Data Warehouse and Multi-dimensional databases in the current scenario. We have reviewed the work of various researchers and have summarized their major contributions. However, we have found that the available research is not sufficient to solve the different problems. As part of future work more research has to be carried out in the field of Data Warehousing and Multi-dimensional databases.

REFERENCES

[1] Dedić, Nedim; Stanier, Clare (2016). Hammoudi, Slimane; Maciaszek, Leszek; Missikoff, Michele M. Missikoff; Camp, Olivier; Cordeiro, José, eds. An Evaluation of the Challenges of Multilingualism in Data Warehouse Development. SciTePress. pp. 196–206 ISBN 978-989-758-187-8.

[2] Bellatreche, Ladjel & Giacometti, Arnaud & Marcel, Patrick & Mouloudi, Hassina & Laurent, Dominique. (2005). A personalization framework for OLAP queries. 9-18. 10.1145/1097002.1097005.

[3]"Exploring Data Warehouses and Data Quality". spotlessdata.com. Retrieved 2017-04-30.

[4] Swati Gupta, Shalini Bhaskar Bajaj 'Roll back Mechanism in Multi Version Locking in Distributed Database 'Published in second International Conference on Electronics, Communication and aerospace Technology, ICECA 2018.

[5] Swati Gupta 'Multiple Regression Technique in Data Mining' International Journal of Computer Applications (0975 – 8887) Volume 126 – No.5, September 2015

[6] Gómez, Leticia & Kuijpers, Bart & Vaisman, Alejandro. (2017). Performing OLAP over Graph Data: Query Language, Implementation, and a Case Study. 1-8. 10.1145/3129292.3129293.

[7] Q. Hanlin, J. Xianzhen and Z. Xianrong, "Research on Extract, Transform and Load (ETL) in Land and Resources Star Schema Data Warehouse," 2012 Fifth International Symposium on Computational Intelligence and Design, Hangzhou, 2012, pp. 120-123.

[8] Vassiliadis, Panos. (2009). A Survey of Extract-Transform-Load Technology, International Journal of Data Warehousing and Mining. 5. 1-27.

[9] Image from https://stackoverflow.com/questions/20036905/difference-between-fact-table-and-dimension-table

[10] Surajit Chaudhuri and Umeshwar Dayal. 1997. An overview of data warehousing and OLAP technology. SIGMOD Rec. 26, 1 (March 1997).

[11] Bach Pedersen, Torben; S. Jensen, Christian (December 2001). "Multidimensional Database Technology". Distributed Systems Online: 40–46. ISSN 0018-9162.

[12] Owen Kaser and Daniel Lemire, Attribute Value Reordering for Efficient Hybrid OLAP, Information Sciences, Volume 176, Issue 16, pages 2279-2438, 2006.

[13] A. Levy, A. O. Mendelzon, Y. Sagiv, D. Srivastava, "Answering Queries using Views", In. PODS, 1995.

[14]P. Vassiliadis, Nat. Tech. Univ. of Athens, Greece, Modeling Multidimensional Databases, Cubes and Cube Operations, Proceedings. Tenth International Conference on Scientific and Statistical Database Management (Cat. No.98TB100243), ISBN: 0-8186-8575-1.

[15] M. Rafanelli, F.L. Ricci, "A functional model for macro-databases", SIGMOD Record, 20(1), March 1991.

[16] A. D. Patil and N. D. Gangadhar, "OLaaS: OLAP as a Service," 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, 2016, pp. 119-124.

[17] Zhang, Yansong & Zhang, Yu & Wang, Shan & Lu, Jiaheng. (2018). Fusion OLAP: Fusing the Pros of MOLAP and ROLAP Together for In-memory OLAP. IEEE Transactions on Knowledge and Data Engineering. PP. 1-1. 10.1109/TKDE.2018.2867522.