ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

A Survey on Analysis of Bitemporal Data Warehouse using Querying and Modeling Techniques

Rushabh Balpande[1], Prof. Amreen Khan[2]

M.Tech Student, Dept. of Computer Science & Engineering, G.H.R.C.E, Nagpur, India[1]

Professor, Dept. of Computer Science & Engineering, G.H.R.C.E, Nagpur, India[2]

ABSTRACT: Success of any Decision Support System (DSS) depends on the Data Warehouse (DW). Important results have been produced in the field of DW by the temporal database research. DW store data of historical importance, and hence could be clearly beneficial from this research. Temporal data warehouse combines two areas of temporal database research and data warehouse. The multidimensional Bitemporal relational database generates objects and cubes of DW. With the help of set of queries, the performances of the relational database and the Bitemporal database are compared and evaluated in terms of the size or disk space consumption and execution time. The idea proposed in this system is to perform an analysis on the relational and Bitemporal extension of the database on the basis of time required to perform queries and the size of the database. Based on the analysis, it is easier to understand the nature of the traditional and Bitemporal database. In the proposed work, we have developed a system which will be useful for comparing the relational and Bitemporal databases and optimizing the performance of these databases by query optimization. The proposed system will also be able to compare four databases at a single time.

KEYWORDS: Decision Support System (DSS), Data Warehouse (DW), Temporal database, Bitemporal database, Relational database, Query, Disk space consumption, Execution time, Optimization

I. INTRODUCTION

In recent times, research on Big Data and Data Warehouses has gained popularity. Technology is an absolutely essential part of today's lifestyle. People have changed their habits of shopping, communication and transaction. Due to this reason, lot of information is left unknowingly when surfing on social networking sites. This data is processed to make recommendations by various e-commerce web sites or utilized by them to increase their profits. This data is stored in databases and eventually in Data Warehouses.

Data warehouse is an important research area and it is used to access historical data. A data warehouse can be defined as a combination of subject-oriented, integrated, non-volatile, and time-variant data that is useful in making decisions in corporate world. The sole purpose is to gain required information or data in an easy and fast way. This type of data is used in reports of management, certain business queries, DSS, MIS, and applications of data mining by associating the various databases which are scattered throughout an organization. Important results have been produced in the field of Data Warehouse by the temporal database research. Temporal data warehouse combines the two areas of temporal database and data warehouse research. The multidimensional bitemporal relational database generates objects and cubes of Data warehouse. With the help of set of queries, the performance of the relational database and the bitemporal extension is compared and evaluated in terms of the disk space consumption and execution time.

Time-variant and non-volatile data features of data warehousing proposes that changes are permissible to data values without duplicating existing values[1]. Combination of data warehouses and temporal database as described in [1] has been considered by researchers. However, only the tuple time-stamped temporal database approach [1] has been considered by them. But the combination of attribute time stamping and data warehousing is yet to be studied by researchers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

A. ***CHARACTERISTICS OF DATA WAREHOUSE:***

- Data warehousing stresses on capturing data or information from various sources which can be helpful in analysing and accessing, but doesn't usually start from the view of the end user who generally needs access to definite field of local databases known as the data mart.
- Top down and bottom up are two approaches of data warehousing.
- Data marts for specific group of users after the formation of data warehouse are created by the top down approach.
- First building the data marts and then combining them into a single and enclosed data warehouse is created by bottom up approach.
- User queries and analytical applications are used to abstract data from different online transaction processing (OLTP) applications. Typically the data warehouse is stored on the mainframe server of enterprise or on the cloud.

B. ***CHARACTERISTICS OF TEMPORAL DATABASE:***

- A temporal database stores a time series of data efficiently, generally by having some fixed scale of time (such as seconds or milliseconds) and then storing changes only in the measured data.
- A timestamp in a relational database is a privately stored value for each measurement, which is very indisposed.
- A temporal database either logs the history of the transactions in database (transaction time) or history of the significant objects and their attributes (valid time).
- A valid time is essentially used in order to ideal the time-varying states of an object and its attributes. These states may be in the past, present, or even in the future.
- A transaction time indicates the timestamp of any change as it is logged in the database. It may not increase beyond the present time and is generated by the system.

C. ***NEED FOR ANALYZING THE PERFORMANCE OF BITEMPORAL DATABASE:***

- A Bitemporal database design is a process of storing data logs which are relying on time to portray both the history of the facts and the history of changes to the logs in the database.
- A Bitemporal database allows queries over two irrelevant dimensions of time: transaction time and valid time.
- Time when a fact is operative in reality it is called valid time. Time when the log is operative in the database is called transaction time.
- Only possessing the valid time, captives the history of an object, but it does not jam the history of any post active or retroactive changes.
- Only possessing the transaction time, you cannot jam the historical data, i.e., the validity period of data values.

D. ***PROBLEM DEFINITION:***

The issues that are related to the existing system are feasible and can be overcome in the proposed system. The main problems that are faced in the existing system can be described as follows:

- A temporal database either logs the history of the transactions in database (transaction time) or history of the significant objects and their attributes (valid time) but not both.
- A valid time is essentially used in order to ideal the time-varying states of an object and its attributes and these states may be in the past, present, or even in the future.
- A transaction time indicates the timestamp of any change as it is logged in the database which may not increase beyond the present time and is generated by the system.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

- The traditional relational database consumes more disk space and takes more time to execute queries.

E. ***OBJECTIVES AND SCOPE:***

The objectives or goals of the proposed system are to overcome the problems in the existing system. The main objectives of the proposed system can be described as follows:

- To minimize the execution time required to execute a query.
- To minimize the disk space consumption of tables in database.
- To capture both valid and transaction time in bitemporal database.
- To evaluate available relational database and bitemporal database on the basis of their performance.
- To compare these databases in terms of consumption of disk space and execution time with the help of set of queries.

The scope of the proposed system:

- Business Intelligence (BI) tool and this will be used in IT industry and business firms.
- Also useful for work from home companies for part time and full time employees.

II. RELATED WORK

Canan Eren Atay, et al. [1] proposed a bitemporal data warehouse model. According to them the temporal data warehouse combines the two areas of temporal database and data warehouse research. The multidimensional bitemporal relational database generates objects and cubes of Data warehouse. With the help of set of queries, the accomplishment of the relational database and the bitemporal extension is compared and evaluated in terms of the disk space consumption and execution time.

Burdakov A., et al. [2] proposed a new ripple method for inaccurate processing of query that manages randomly sized multidimensional file of related records with minimum additional calculations and gives a attached fault of single or clustered value reconstruction. Authors have demonstrated that the new method permits evaluating a trusting intermission of the error in a query depending on given compaction ratio of the data warehouse or executing an opposite task, i.e. evaluating the required data warehouse compression ratio for a given allowable error.

Leopoldo Zepeda, et al. [3] proposed a meta-model approach for the implementation and design of Data Warehouses (DW). The approach is designed of a set of rules of transformation as a mechanism to originate multidimensional structure of databases from the reasonable representation of the Database and for users requirements definition they propose a Computation Independent Model (CIM). As a result they have implemented an Eclipse based prototype which evolved from the logical representation of the operational Database and requirements of the users for multidimensional structure of databases from the reasonable representation of the Database which reflects user's requirements nicely.

Michelangelo Ceci, et al. [4] proposed a method that is based on a density-based hierarchical clustering algorithm, to support drill-down and roll-up operations over OLAP data cubes with continuous dimensions. The work of the method is to clusters dimension instances hierarchically by considering fact-table measures into account. With the help of this method, the authors have enhanced the effect of clustering w.r.to the possible analysis.

Martin Kaufmann, et al. [5] proposed a benchmark proposal that provides comprehensive coverage of the bitemporal data management. It is designed on the base of TPC-H but elaborates it with certain sets of queries and update scenarios. The proposed work has roots both from real-life temporal applications from SAP's customer base and a systematic coverage of temporal operators. The authors have presented preliminary results of the benchmark on many temporal database systems.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Suman Mann, et al. [6] proposed usage of multidimensional model as they are explicitly designed to support data analyses. Data is classified as collection of facts and dimensions, in this model. Facts having numeric measures stating they are quantitative while dimensions are comprised of textual attributes that categorize facts hence they are qualitative.

Boon Keong Seah, et al. [7] proposed a survey-based service data with the implementation and design of Data Warehouse framework for reports of business intelligence and data mining. The authors have developed a multidimensional data model for the generation of multiple data marts and design of an ETL process from data source for increasing the data marts.

Robert Wrembel [8] proposed some challenges in building, designing and managing data warehouse and their evolution w.r.to time. The survey is based on the Multiversion Data Warehouse (MVDW). The author in his work has come up with some issues like the concept of the MVDW, structure for transferring data in the MVDW, a framework for indentifying changes in data source, index structures for the data indexing in the MVDW.

Wided Oueslati, et al. [9] proposed that the Data Warehouse which is the core component of the modern Decision Support System (DSS) has to be updated according to the different types of information evolution sources to show the real world subject to analysis. Authors also propose various works related to the evolution of Data Warehouse and have done comparative studies on these works.

Matteo Golfarelli, et al. [10] considered certain issues which they have grouped under temporal data warehouse, and have described how information changes with the time in data warehousing systems. Authors identified that these issues can be classified into certain topics with reference to three-level architecture. These topics are namely: handling data/schema changes in data marts and data warehouse, querying temporal data and designing temporal data warehouses.

Elzbieta Malinowski, et al. [11] proposed a temporal database for the multidimensional ER model. The proposed model permitted the addition of valid and transaction time which are fetched from the source system, inclusive of the loading time of data warehouse. The proposed model allowed a conceptual representation of time-varying levels, hierarchies and attributes. Later authors have also discussed various cases to show the impact of changes in levels on the relationship between them.

Alberto Abello, et al. [12] considered two research areas, i.e. Data Warehouses and Temporal Databases. The authors surveyed their results in order to achieve the goals. The authors found more similarities than differences between these two areas. Thus, the authors focused on contributions of the temporal database research that benefits the data warehouses.

III. PROPOSED METHODOLOGY

The phases of Software Development Life Cycle (SDLC) used in the system:
1. Planning Phase: In this phase, we planned different aspects of our system such as what kind of data is needed, how to generate user interface, what sorts of testings have to be performed.
2. Requirement Analysis Phase: In this phase, we performed requirement gathering, understanding and finalizing.
3. Design Phase: In this phase, we design and assemble application and database to generate a working model.
4. Development Phase: In this phase, we performed build system and test systems build.
5. Integration and Testing Phase: In this phase, we perform different independent testing viz. unit testing to evaluate the system.
6. Implementation Phase: In this phase, we established product baseline and deploy the system.
7. Maintenance Phase: In this phase, we migrate system to production environment and carry out system enhancements.
Figure 1 show the whole Software Development Life Cycle (SDLC) used in the system:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

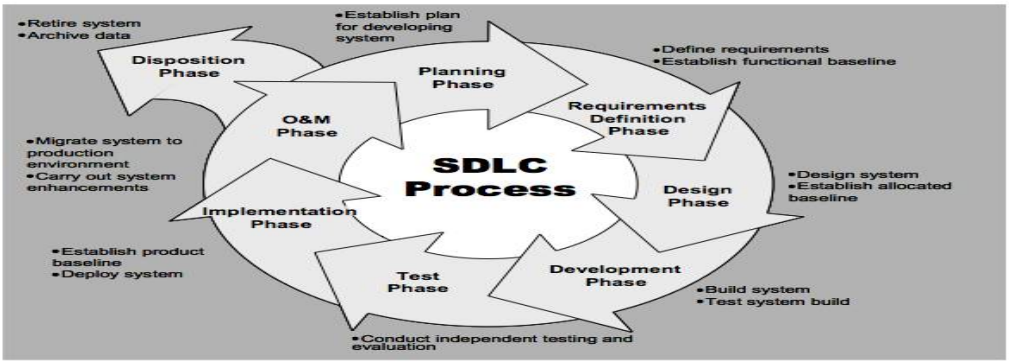

Fig.1: Software Development Life Cycle (SDLC)

Query execution time algorithm:

To calculate the execution time of the query, we will use the following algorithm:

- Input query by the user
- SQL will fire the query in the form of stored procedure dynamic query
- Execute the query in stored procedure
- Stored procedure will return execution time or output parameter

Record count algorithm:

To calculate the records in the database, we will use the following algorithm:

- Split the entered query in arrays
- Find the tables and views from the parts of the query
- Fetch the structure from database server using object in query from relational database
- Fetch the structure from database server using object in query from bitemporal database
- Extract keys, indexes and record statistics of the tables

Figure 2 show the architecture of the proposed system:

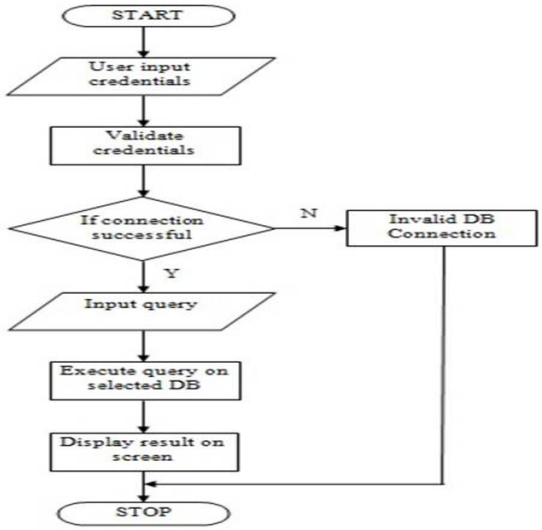

Fig.2: System Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

IV. CONCLUSION AND FUTURE WORK

To the best of our knowledge we proposed a system which compares two or more databases simultaneously. The database can be relational, temporal or bitemporal in nature. The system we proposed can be used to perform queries on these databases simultaneously and generate a result which gives the time required to execute the query, the number of tables and their records found and the disk space consumption of these databases and their tables. The proposed system can also generate reports on the basis of the execution speed and disk space consumption. The proposed system maintains the log of the users who logged in the system and also the databases they have worked on. The system also proposes a mechanism to optimize the performance of the executed query by providing to user, the suggestions and intelligence to write an optimized query. The future enhancements can be visual representation of the schema of the database while optimizing the query and some intelligence which can automatically modify the query which is manual at present.

REFERENCES

[1] Canan Eren Atay and Gözde Alp, "Modeling and Querying Multidimensional Bitemporal Data Warehouses", *International Journal of Computer and Communication Engineering* (IJCCE), Vol. 5, No. 2, pp. 110-119, 2016.
[2] Burdakov A., Grigorev U., Ukharov A., and Ploutenko A., "Approximate Query Processing Using Wavelets in OLAP with Arbitrarily Sized Data and Bounded Errors", *24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*, pp. 66-73, 2016.
[3] Leopoldo Zepeda, Luis Santillan, Elizabeth Cecena, Emir Manjarrez, Liliana Vega, and Clemente Garcia, "A meta-model based approach for Data Warehouses design and implementation", *26th International Workshop on Database and Expert Systems Applications*, pp. 212-216, 2015.
[4] Michelangelo Ceci , Alfredo Cuzzocrea, and Donato Malerba, "Effectively and efficiently supporting roll-up and drill-down OLAP operations over continuous dimensions via hierarchical clustering", *J Intell Inf Syst, Springer Science+Business Media New York*, pp. 309-333, 2015.
[5] Kaufmann, M., Fischer, P. M., May, N., Tonder, A., and Kossmann, D. "TPC-BiH: A benchmark for bitemporal databases", *Springer International Publishing*, pp. 16-31, 2014.
[6] Suman Mann, Bharti, and Perminder Singh, "Empirical Validation of Multidimensional Model for Data Warehouse", *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*, pp. 1-6, 2014.
[7] Boon Keong Seah and Nor Ezam Selan, "Design and Implementation of Data Warehouse with Data Model using Survey-based Services Data", *4th edition of the International Conference on the Innovative Computing Technology (INTECH 2014)*, pp. 58-64, 2014.
[8] Wided Oueslati and Jalel Akaichi, "A Survey on Data Warehouse Evolution", *International Journal of Database Management Systems (IJDMS)*, Vol.2, No.4, pp. 11-24, 2010.
[9] Wrembel, R., "A survey on managing the evolution of data warehouses", *International Journal of Data Warehousing and Mining*, Vol. 5, Issue 2, pp. 24-56, 2009.
[10] Matteo Golfarelli and Stefano Rizzi, "A Survey on Temporal Data Warehousing", *International Journal of Data Warehousing & Mining*, Vol. 5, Issue 1, pp. 1-17, 2009.
[11] Elzbieta Malinowski and Esteban Zimanyi, "A Conceptual Solution for Representing Time in Data Warehouse Dimensions", *3rd Asia-Pacific Conference on Conceptual Modelling (APCCM2006), Conferences in Research and Practice in Information Technology*, Vol. 53, pp.45-54, 2006.
[12] Alberto Abello and Carme Martin, "The Data Warehouse: An Object-Oriented Temporal Database", *8th International Conference of Software Engineering and Databases (JISBD 2003)*, pp. 675-684, 2003.

BIOGRAPHY

Rushabh Balpande is an M.Tech student from G.H. Raisoni College of Engineering, Nagpur (An Autonomous Institute Affiliated to RTM Nagpur University). He has completed his Bachelor of Engineering (B.E.) in Information Technology from Dr. D. Y. Patil College of Engineering, Ambi, Pune, India in 2015. He is interested in Data mining, Cloud computing, Networking and has keen interest in programming languages.

Amreen Khan is an Assistant Professor at the G.H. Raisoni College of Engineering, Nagpur (An Autonomous Institute Affiliated to RTM Nagpur University). She has completed her Masters of Technology (M.Tech) in Computer Science and Engineering from G.H. Raisoni College of Engineering, Nagpur, India. She has 7 years of experience in teaching. She took up her post as an Asst. Professor at the CSE Department, G.H. Raisoni College of Engineering, Nagpur in 2014. She is currently pursuing her PhD in the domain of Particle Swarm Optimization (PSO) from RTMNU, Nagpur.