A STUDY OF HYPERCUBE-BASED ACCESS CONTROL TECHNIQUES FOR INFORMATION PRIVACY AND CYBER SECURITY

Kulwinder Kaur

Assistant Professor in Computer Science, Punjabi University T.P.D. Malwa College, Rampura Phul, Dist. Bathinda, Punjab

Hardeep Kaur

Assistant Professor in Computer Science, Punjabi University T.P.D. Malwa College, Rampura Phul, Dist. Bathinda, Punjab

Abstract

The study is focus on the business intelligence, business analytics, multidimensional modeling with various privacy accesses control techniques: further research objectives and statement of purpose will be outline in this study. Apart from concepts also had been defines so as to provide a clear picture of the context of the study. Survey of literature work on the related domains of the problem will be discussed. A study to build a cube and secure against unauthorized access to sensitive information.

Introduction

This paper includes introduction about Business analytics & Intelligence, Multidimensional modeling, (OLAP) Online Analytical Processing, and the statement of purpose, objective of research is stated as well as the development environment during the implementation of proposed study.

Business Analytics & Intelligence

Multidimensional databases are a key technology in the enabling of interactive analyses of large amounts of data for decision-making purposes. However, accompanying the growth in business analytics and increasing demand by users for more powerful tool that provides advanced analytical capabilities. That's why

business analytics empowers analysts with online analytical processing to rapidly acquire, transform and analyze data and secure against unauthorized access.

So far, standard reporting, ad hoc query and reporting applications have run directly from relational tables while more complicated Business Intelligence applications have used specialized analytical databases. These specialized analytical databases typically provide support for multi-dimensional modeling and various predictive functions [23].

Multidimensional Modeling

Multidimensional data models have three important applications for dataanalysis.

First, multidimensional model is used in data warehousing. A data warehouse is a large repository of data integrated from several sources in an enterprise for the specific purpose of data analysis. Typically, this data is modeled as being multidimensional, as this model is best for dataanalysis.

Second, multidimensional models lie at the core of On-Line Analytical Processing systems. Such systems provide fast answers for queries that aggregate large amounts of detail data to find overall trends, and they present the results in a reports, which renders a multidimensional data organization ideal for OLAP.

Third, multidimensional data is increasingly becoming the basis for data mining, where the aim is the semi-automatic discovery of unknown Facts in large databases, as it turns out that multi- dimensionally organized databases are particularly well suited for the queries that are posed by data mining tools.

Online Analytical Processing (OLAP)

OLAP is a technique for summarizing, consolidating, viewing, applying formulate and synthesizing data according to multiple dimensions in data warehouses. OLAP describe a technology that uses a multi- dimensional view of aggregate data to provide quick access to calculated information for the purposes of advanced analysis. It needed for executing complex queries on large amount of data that are processedon aggregated level, not on individual level of records as for OLTP.

OLAP allows the user to view summarized data in such a way that it is a better

model of the true dimensionality of the creativity. While OLAP systems can easily answer 'who?' and 'what?' questions, it is their ability to answer 'why?' and 'what if?' type questions that distinguish them from general-purpose query tools. OLAP enables decision-making about future actions. A typical OLAP calculation can be more complex than simply aggregating data.

Statement of Purpose

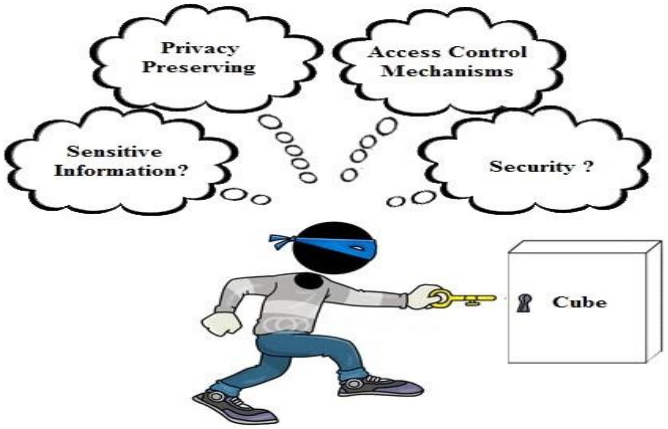
In today's world of analytics, it is impossible to find out relevant facts from relational databases. Because it requires complex joins to link data from different tables and evaluate its result based on the query triggered to access the relevant information. Therefore, it is required to perform demographic analysis of relational databases to evaluate relevant facts effectively and to convert into multi-dimensional databases.

The multidimensional cube is now available with such kind of relevant facts (sensitive information) which are crucial and still susceptible to linking attacks by the unauthorized users.

Therefore, the main purpose of this study, is to secure multidimensional cube via privacy preserving access control mechanismfrom unauthorized access.

Privacy Preservations and leaks of Sensitive data

Business Intelligence (BI) applications are gaining popularity, consistent with the desire of officials of public and private companies to monitor, analyze, understand, and eventually improve business processes and provide better Service to their



customers and users.

BI applications typically extract data from multiple data sources, then apply data profiling to ensure data quality and consistency to the possible extent, transform them, and then generate various kinds of reports used by managers and officials to analyze the performed processes.

With the rapid increase in the amount of an organization's data which is gathered and exchanged electronically, the problem of information privacy is rapidly gaining attention. Now it must be admitted that there are situations in which a relational database is simply not an appropriate tool. In fact, by providing a staging point for data integration, manipulation and analysis, multidimensional database technology can be seen as a complement to relational database technology. Multidimensional technology is not designed to take the place of relational and other database technologies employed for transaction intensive applications.

We explore different and possibly complementary options, includingvarious ways to identify sensitive data for analyzing and then applying various access control mechanism for privacy preserving of sensitive dataof cube, and to secure against the unauthorized access.

Difference between Relational modeling and Multidimensional modeling.

Sr. No	Relational Modeling	Multi-Dimensional Modeling
1	Relational models can be very complex joins of relationship among tables.	The Direct and In-direct relationship of dimension with the fact table define Star or Snowflake Schema.
2	Normal data modeling is quiteflexible.	The Multidimensional modeling has a rigid structure.

3	One of the goals of relational modeling is to confirm to the rules of normalization. In a normalized database each datavalue is stored only once.	Multidimensional modeling is radically de-normalized. The dimension tables have a high number of repeated values in their fields.
4	Standard relational models are optimized for On Line Transaction Processing. OLTP needs the ability to efficiently update data. This is provided in a normalized database that has each value stored only once.	Multidimensional modeling is optimized for On Line Analytical Processing. OLAP needs the ability to retrieve data efficiently. This is provided with the simple structure of relationship in a Multidimensional modeling, where each dimension tableis only a single join away from the facttable.
5	SQL queries required to manipulate relational data.	OLAP queries required to manipulate multi-dimensional data.
6	Tables are units of relationaldata storage.	Cubes are units of multi-dimensionaldata storage.
7	Table fields of particular data type store the actual data.	Dimensions and measures stores actualdata.
8	Response time of query is low.	Efficient in query performance.
9	It is difficult to apply privacy on relational databases	Multi-dimensional databases are best suited for privacy preservation.
10	Data is still susceptible for linking attacks	Unauthorized access can be prohibited due to privacy preserving access controlmechanism.

Objectives of the study

- To perform demographic analysis of a dataset and identifying various dimensions, levels of dimension and building a cube based on basic elements required for the structure of cube.

Research Methodology

Criteria for literature survey

Accordingly, the research study started with a systematic investigation of the extensive existing literature. The study also had analyzed the structure and characteristics of various multidimensional databases with privacy preserving mechanisms for related studies. Many of research studies have been studied based on multidimensional modeling to enhance basic knowledge in the area of multidimensional modeling that were introduced in the literature. The literature survey focused uponprivacy preserving of data cube from unauthorized access.

Data Collection, Analysis and Findings

To achieve the first objective stated, taking some important factors in consideration, the proposed study had facilitated for demographic analysis of an organization's data which is in relational form and prerequisite to accomplish dimensioning.

Framework Design and studying Feasibility

Identify the various dimensions, levels of dimension and building a cube based on these findings and anonymized to avail sensitive information protected against unauthorized access.

Development Environment

The tools used for the implementation of proposed study based on building a multidimensional cube and securing privacy preserving of sensitive details listed below.

- Microsoft SQL Server 2005
- Microsoft Visual Studio 2010
- Microsoft .NET Framework

An overall implementation of this project has been carried out using these tools.

Microsoft SQL Server is a relational database management system (RDBMS) produced by Microsoft. SQL Server 2005 introduces "studios" to help the development and management tasks such as SQL Server Management Studio and Business Intelligence Development Studio. In Management Studio, to develop and

manage SQL Server Database Engine and notification solutions, manage deployed Analysis Services solutions, manage and run Integration Services packages, and manage report servers and Reporting Services reports and report models.

Microsoft Visual Studio is a collection of tools that allow a programmer to edit, compile, debug, test and deploy applications on a variety of operating systems such as Microsoft Windows. Visual Studio supports different programming language and allows the code editor anddebugger to support any programming language, provided a language- specific service exists. Built-in languages include C, C++ and, VB.NET.

Microsoft .NET Framework is a software framework that is available with several Microsoft Windows operating systems. It includes a large library of pre-coded solutions to common programming problemsand a virtual machine that manages the execution of programs written specifically for the framework.

Review Of Literature

The review of literature is discussed in this study refers to topic onbuilding and securing multidimensional cube. This study introduces multidimensional design methods, a comprehensive survey and finally appraises summary based on various research studies, designs, and methods; it must to clearly depict the current situation of the proposed study.

Literature review emphasized on various research papers and articles ondimensional modeling, relational databases to multidimensional databases, privacy preservation mechanism and various comparative studies and innovative ideas for building a cube of data and secureagainst unauthorized access of data.

Many of research studies have been studied based onmultidimensional modeling to enhance basic knowledge in the area of multidimensional modeling that were introduced in the literature. After the analysis, this study able to find an implicit agreement on how multidimensional data should be manipulated and eventually, identify the multidimensional integrity constraints preserving a correct data manipulation. Thus, the schemas produced by various approaches guarantee the multidimensional constraints identified in proposed study and consequently make

multidimensional sense.

Multidimensional Design Methods

According to Winter & Strauch multidimensional modeling methods may be classified within a demand-driven, a supply-driven or a hybrid framework [30].

Supply-driven approaches: It is also known as data-driven which just start from a detailed analysis of the data sources to determine the multidimensional concepts in a reengineering process.

Demand-driven approaches: It is well known as requirement- driven or goal-driven, focus on determining the user multidimensional requirements to later map them onto data sources.

Hybrid approaches: This approach combines both paradigms in order to design the data warehouse from the data sources but bearing in mind the end-user requirements. Here we distinguish as well between interleaved hybrid approaches and sequential hybrid approaches. The main difference is that sequential approaches perform the demand- driven and supply-driven stages independently and later on conciliate results got in a final step, whereas interleaved approaches perform both stages simultaneously benefiting from feedback retrieved by each stage all over the whole process and obtaining better results at the end.

A Comprehensive Survey

This section presents an insight into current multidimensional design and Privacy preservation methods. These methods were selected according to three factors: reference papers with a high number of citations according to Google Scholar, papers with novelty contributions and in case of papers of the same authors, survey included the latest version of their works.

Cabibbo and Torlone (1998) present one of the most cited multidimensional design methods. This approach generates a logical schema from Entity-Relationship (ER) diagrams and it may produce multidimensional schemas in terms of relational databases or multidimensional arrays [1]. The authors suggest identifying facts and dimensions and restructuring the ER diagram. Indeed the authors suggest performing

them in an iterative way to refine results obtained [2].

Pedersen and Christian (1999) describes fundamentals concept in multidimensional data models that have three important applicationareas within data analysis. First, multidimensional models are used in data warehousing. Typically, this data is modeled as being multidimensional, as this best supports data analysis. Second, multidimensional models lie at the core of On- Line Analytical Processing (OLAP) systems. Such systems provide fast answers for queries thataggregate large amounts of detail data. Third, multidimensional data is increasingly becoming the basis for data mining, where the aim is the semi- automatic discovery of unknown knowledge in large databases [3].

Moody and Kortink (2000) present a method to develop multidimensional schemas from ER models. Although it is not the first approach working over ER schemas, they present a structured and formal method to derive multidimensional logical schemas. The authors introduce two different operators to produce logical schemas as Collapse hierarchy and Aggregation. According to the resulting schema level of denormalization and the granularity of data, they introduce rules to derive flat schemas, terraced schemas, star schemas, snowflake schemas or star cluster schemas [36].

Wijesekera et al. (2004) introduced a generic overview of the existing access control model in a relational database is unsuitable for multidimensional data cube. In this approach the partitioning of a data cube both vertically based on dimension hierarchies and horizontal based on data. In this study the author proposed method eliminates bothunauthorized accesses and inference in the data cube [4].

Wang et al. (2007) explains the main purpose of an OLAP system is to enable analysts to construct a mental image about the underlying data by exploring it from different perspectives, at different level of generalizations, and in an interactive manner. The data to be analyzedby OLAP are usually stored based on the relational model in the backend data warehouse. Each dimension has a dimension table associated with it, indicating a dimension hierarchy. The dimension tables may contain redundancy, which can be removed by splitting each dimension table into

multiple tables, one per attribute in the dimension table [5].

Cuzzocrea and Sacca (2011) introduced a framework on novel privacy OLAP notion, which following consolidated paradigms of OLAP research, privacy of aggregate patterns defined in a multidimensional range rather than the privacy of individual data cell [6].

Thavavel and Sivakumar (2012) introduce a solution to managing unstructured data into structured data using a legacy system and distributed data partitioned method for giving distributed data for mining multi text documents. In this framework authors provide testing of similarities among text documents and privacy preserving metadata hiding technique [7].

Srikant et al. (2000) describes a technique for privacy-preserving computation of multidimensional aggregates on data partitioned across multiple clients. Data from different clients is randomized in order to preserve policy before it is integrated at the server. In this study, author proposed an algorithm for reconstructing counts of sub cubes over perturbed data [8].

Gehrke et al. (2003) describes a new formulation of privacy breaches, together with a methodology, "amplification", for limiting them. In this approach author proposed this methodology for the problem of mining association rules, and modify the algorithm to limit privacy breaches without knowledge of the data distribution [9].

Xiong et al. (2005) describes the approach a query finds the aggregated values over all selected cells of an online analytical processing (OLAP) data cube where the selection is specified by the ranges of contiguous values for each dimension. An important issue is how to preserve information in a data cell while providing an accurate estimation of original values for queries. In this study, author proposed an effective solution, called the zero-sum method, to this problem. This study result shows that privacy preservation and query accuracy has fulfilled goals: security, accuracy, and accessibility [10].

Sung et al. (2006) present cubic-wise balance method to provide privacy preserving queries on data cubes in a data warehouse. This approach is motivated by the following observation: analysts are usually interested in summary data rather than

individual data values. In this approach provide summary data for queries without providing access to actual individual data values. In this method achieve both better privacy preservation and better query accuracy than random data perturbation alternatives [11].

Chaudhuri et al. (2007) introduced Fine Grained Access Control for datasecurity. In this approach the requirement the granularity of fine grainedaccess control can be on directories, database level, table level, even on individual records and data field level. The access control is enforced by a mechanism implementing regulation established by a security policy. In this study, a novel access control model is proposed, which provides fine grained access control to share data to authorized users [12].

Gangopadhyay et al. (2008) focused on multiple sensitive attributes. In this approach K-anonymity model can be trivially extended to multiple sensitive attributes, instead of L-diversity model. The reason is that achieving L-diversity for each individual sensitive attribute does not guarantee L-diversity over all sensitive attributes. In this study, author proposed a new model that extends L-diversity and K-anonymity to multiple sensitive attributes a practical method to implement this model[13].

Vasudevan et al. (2008) define a new technique that is integrating the advantage of role based access control approach and cryptographic technique with the view of minimizing information loss and privacy loss. In this technique store sensitive data and providing access to the stored data based on an individual's role, and ensure that the data is safe from privacy breaches [14].

Rokach et al. (2010) introduced different approach for achieving k- anonymity by partitioning the original dataset into several projections such that each one of them adheres to k-anonymity. In this study, author proposed data mining privacy by decomposition (DMPD) algorithm which uses a generic algorithm to search for the optimal feature set partitioning. The result with DMPD better than existing k-anonymity based algorithm and there is no necessity for applying domain dependent knowledge [15].

Vignesh and Senthil (2012) focused on privacy preserving OLAP focuseson a single aggregate function and multiple aggregate functions which deal with both exact disclosure and partial disclosure of the data. This model performed using the combination of the simple aggregate functions which guarantees the level of privacy disclosure as required bythe user. In this the various malicious users can exploit the correlation among data to infer sensitive information from a series of seemingly innocuous data accesses based on data dependency, database schema and, the system constructs a semantic inference model that represents the possible inference channels and the violation detection system keeps track of the user history when the infer sensitive information exceeds the pre specified threshold the current query will be rejected [16].

Lamba and Abbas (2013) describe how to preserve sensitive informationusing fuzzy logic. Clustering is done on the original data set, and then adds noise to the numeric data using the fuzzy membership function that results is distorted data. Set of Clusters generated using the fuzzified data is also equivalent to the original cluster as well as privacyis achieved. It is also provided that the processing time of the data is considerably reduced when compared [17].

Srimani and Rajasekharaiah (2014) focused on service oriented and data centric privacy preserved model for Business Intelligence. In the approach the author suggests that the C5.0 algorithm with the Business Intelligence model would provide better performance with privacy preservation facility for Business Intelligence applications [18].

Pervaiz et al. (2014) suggests Access control mechanisms for protecting sensitive information from unauthorized user. When sensitive information is shared and a Privacy Protection Mechanism is not in placed, an authorized user compromises the privacy of a person leading to identify disclosure. In this study the author purposed an accuracy- constrained privacy-preserving access control framework and heuristics for anonymization algorithm and show empirically that the approach satisfies imprecision bound for more permissions and has lower total imprecision than the current state of art [19].

Cuzzocrea and Bertino (2014) introduced the CUR matrix decomposition technique as the elementary component for computing privacy preserving of two-dimensional OLAP views effectively and efficiently. In this approach, a widespread experimental analysis of the framework is provided, which fully confirm to reader the major practicalachievements, in the privacy preserving mechanism [20].

Summary

In order to provide a summary for multidimensional design methods and privacy preserving mechanism, the aim is to provide a best of these methods discussed in review of literature. Setting a basis for discussion will facilitate the mapping of the surveyed methods to a common framework from which inspect each approach, detect trends such as features in common or analyze the evolution of assumptions made by themodeling methods.

According to proposed study, dimensional concepts have beenidentified by means of identification of relevant fact. In fact, many methods use requirements to identify factual data and later they analyze the data sources looking for functional dependencies to identify dimensional data. Maybe for this reason, the use of requirements to identify dimensional concepts has not been that relevant as to identify factual data. After identification of factual information we must have to secure the identified facts by using privacy preserving mechanisms against unauthorized access.

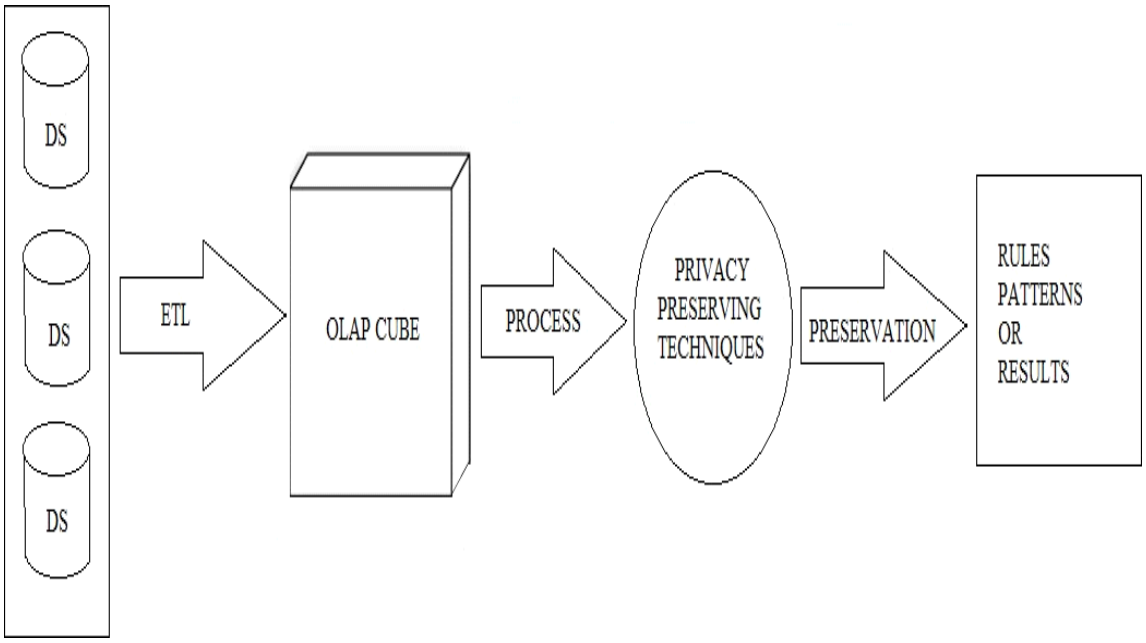
For the completion of proposed study and as per the academic interest, an innovative step is derived towards accomplishment of study by implementing lattice structure technique to aggregate various levels of dimensions of summarized data which is an important application of discrete mathematics and find out various facts. Then, annoymization technique is selected from various available techniques to secure sensitive information against unauthorized access and it is implemented using Sql-Server Business intelligence development studio.

Design & Implementation

This study describes the design and implementation process step by step, how the work was done. It describes the context view of proposed study, various available approaches, implementation of selected approach and complete working overview with tools and finally presents the multidimensional cube.

Context view of proposed study

This phase describes the design and implementation of proposed work, how the work has been done and various problems occur during the implementation phase. In this proposed study, we build a cube and secure against the unauthorized access to sensitive information. The context diagram defines the complete overview of the study which is going to be accomplish by using various tools as mentioned in development environment.



Context diagram

Figure describes the context view of implementation work to accomplish the proposed study. As shown in the figure, a huge amount of data is available in relational sources in summarized form. To identify dimensions, levels of dimensions and their hierarchies within cube, the Data profiling must be applied for dimensional modeling [29]. After dimensional modeling users are ready with a multidimensional cube for querying to access the most relevant facts based on the requirements of an

organization.

Due to concerns of privacy on cube, the organization want to prohibit an unauthorized user to access all the information stored in a data cube. Because multidimensional cube is somewhat like a part of dashboard of an organization. Any of the organization never permits an unauthorized user to access their factual data because it includes some sensitive information about their current policies.

OLAP Architecture/ OLAP Approaches

One may distinguish among various approaches to implementing multidimensional databases. OLAP approaches are categorized according to the architecture used to store and process multidimensionaldata.

There are four main categories of OLAP approaches as defined by Berson and Smith (1997) [31] and Pendse and Creeth (2001) [32] including:

Multidimensional OLAP (MOLAP) systems store data on disk in specialized multidimensional structures. These typically includeprovisions for handling sparse arrays, and they apply advanced indexingand hashing to locate the data when performing queries [21].

Relational OLAP (ROLAP) systems use relational database technology for storing the data, while employing also specialized index structures, such as bit-mapped indices, to achieve good query performance. Generally, MOLAP systems provide faster query response times and more space-efficient storage, while ROLAP systems scale better in the number of facts, are more flexible with respect to cube redefinitions, and provide better support for frequent updates [22].

2.3 Hybrid OLAP (HOLAP) The hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the greater scalability ofROLAP and the faster computation of MOLAP. For example, a HOLAP server may allow large volumes of detail data to be stored in a relational database, while aggregations are kept in a separate MOLAP store.

3.2.4 Desktop OLAP (DOLAP) This approach store the OLAP data in client-based files and support multi-dimensional processing using a client multi-dimensional

engine. DOLAP requires that relatively small extracts of data are held on client machines. This data may be distributed in advance through the Web. As with multi-dimensional databases on the server, OLAP data may be held on disk or in RAM, however, some DOLAP products allow only read access.

Table 3.1 Summarizes advantages & disadvantages of ROLAP,MOLAP, HOLAP.

	ROLAP	**MOLAP**	**HOLAP**
Query performance	Slowest	Fastest	Moderate
Scalability	High	Low	Dependsdesign on
Ease of data access	Any SQL tool	OLAP tools only	OLAP tools only
Data redundancy	No redundancy	Whole dataset is Duplicated	Some data redundancy
Security	Robust standard RDBMS security facilities, like row level access control	Weaker than ROLAP	WeakerROLAP than
Schema	Uses star schema	Uses data cube	Uses both
Loading data	No loading needed	Not efficient for huge datasets	Moderate
Speed	Slow	Fast	Fast
handling non-aggregatefacts table	Best	Low performance	Depend on design

Selection of Approach

OLAP is a requirement for large scale data analysis. Most of the studies concerning data cubes and data warehouses seem to agree that ROLAP is the most suitable architecture for OLAP [28]. The benefit that MOLAP brings is better performance, but the question is if it is worth reduced flexibility. ROLAP is more flexible than

MOLAP in the sense that not all cells in the data cube has to be pre-aggregated, this is also an important aspect of the scalability of OLAP applications. When the number of dimensions increases the data explosion problem grows, primarily with MOLAP, and at some point it becomes overwhelming.

Why multi-dimensional cube using Relational-OLAP?

In Multi-dimensional cube each dimension is a collection of related levels and then levels again having hierarchies. Data is being organized by using star schema to enhance the aggregations in relational database. The star design exposed by OLAP cube is similar to table based star model. Fact table in star schema include that kinds of details which can'tbe changed. Factual information in Fact table consist of context and measure data. The measures are the numeric attributes of the facts that are queried against data cube. DSS queries may access the facts from many different dimensions, levels and hierarchies. The levels in each dimension enable the recovery of facts from different dimensions. The data warehouse should have summarized collection of data attributes that makes information retrieval more efficient.

Multi-dimensional Cube

A Multi-dimensional cube is allow data to be modeled and viewed in multiple dimensions. The Multidimensional cube contains cells that are measure which represent a set of dimensions. Cubes represent data in a form which is easy for the user to understand.

A multidimensional database is modeled through the concept of dimensions and fact tables. Dimensions define the structure of the cube and constructed from hierarchies of dimension levels, whereas fact tables are repositories for the factual data. Each cell in the data cube corresponds to a unique combination of values for the dimensions. In other words it can be say that a cube is a dimensioning form of summarized data encloses in relational resources [33].

Defining Dimensions

Dimensions are lists of unique members that identify and categorize data. They form the edges of a cube, and thus the measures within the cube. Dimensions may contain

levels, hierarchies, and attributes.

Defining Levels

Levels are the somewhat kind of layers within cube or further It is quite easy to define levels at the same time that create a dimension, or you may define the levels later as dynamically changes in relationalresources.

Defining Hierarchies

A hierarchy is a way to organize data at different levels of aggregation.

Defining Measures

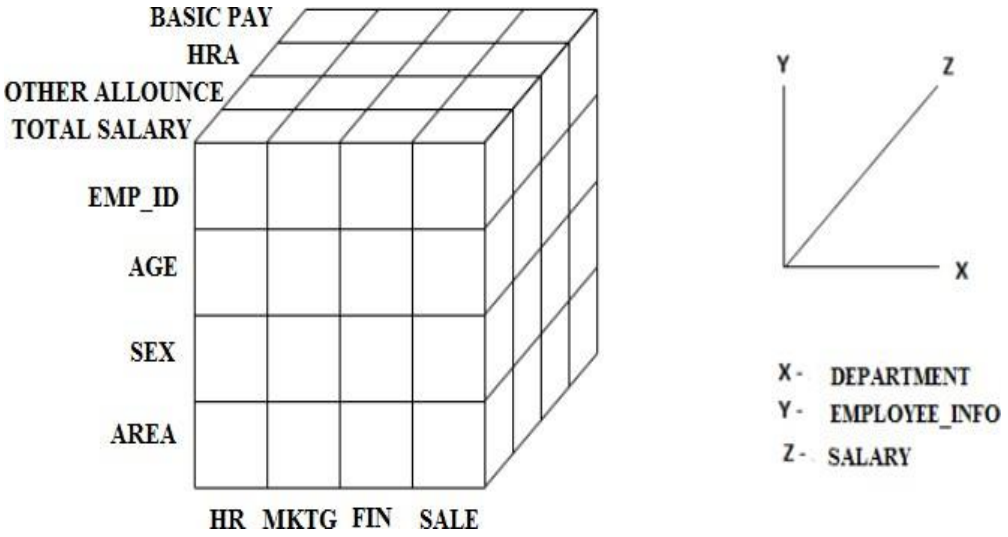
Define measures based on the criteria to identify various facts from the summarized data. Basically measures involve the various cubes' operations.

Defining Attributes

An attribute provides additional information about the data.

Implementation of Selected approach

Identify various dimensions, levels and the hierarchies of levels: Dimensions are lists of unique members that identify and categorize data. They form the edges of a cube, and thus the measures within the cube. Dimensions may contain levels, hierarchies, and attributes.



Multi-dimensional cube of data

Figure describes the dimensional modeling having different view to the data in multidimensional cube. This view may be used in a Decision Support System as a part of business intelligence in associate with data mining errands. Decision support applications often require that information obtained along many dimension. For example, if we want to list out the detail of male_employee from marketing_department and belongs to rural_area having above 15% hra allowance in their gross salary. This query requires the three perspectives as Employee_Info, Salary and Department.

Each dimension is a collection of logically related levels and then levels again having hierarchies and attributes. From which dimensions viewed as an axis for modeling the data then levels for further partitioning. Within each dimension, these entities form levels, on which various questions may be asked. The specific data stored is known as facts and mainly numeric data. Facts consist of context and measure data. The measures are the numeric attributes of the facts that are queried against data cube. Decision Support System queries may access the facts from many different dimensions and levels and hierarchies. Thelevels in each dimension enable the recovery of facts from different dimensions. The data warehouse should have summarized collection of data attributes that makes information retrieval more efficient.

Lattice Structure for Multi-dimensional cube

Multidimensional models view data as consisting of business facts, with associated measures, that are characterized by descriptive data values organized in multiple dimensions. Dimension values are organized in containment-type hierarchical structures that enable the computation of aggregate queries at different levels of granularity. This data organization lends itself towards graphical formulation of aggregate queries and collapse hierarchies in lattice manner with various cuboid levels for the efficient evaluation of aggregate queries.

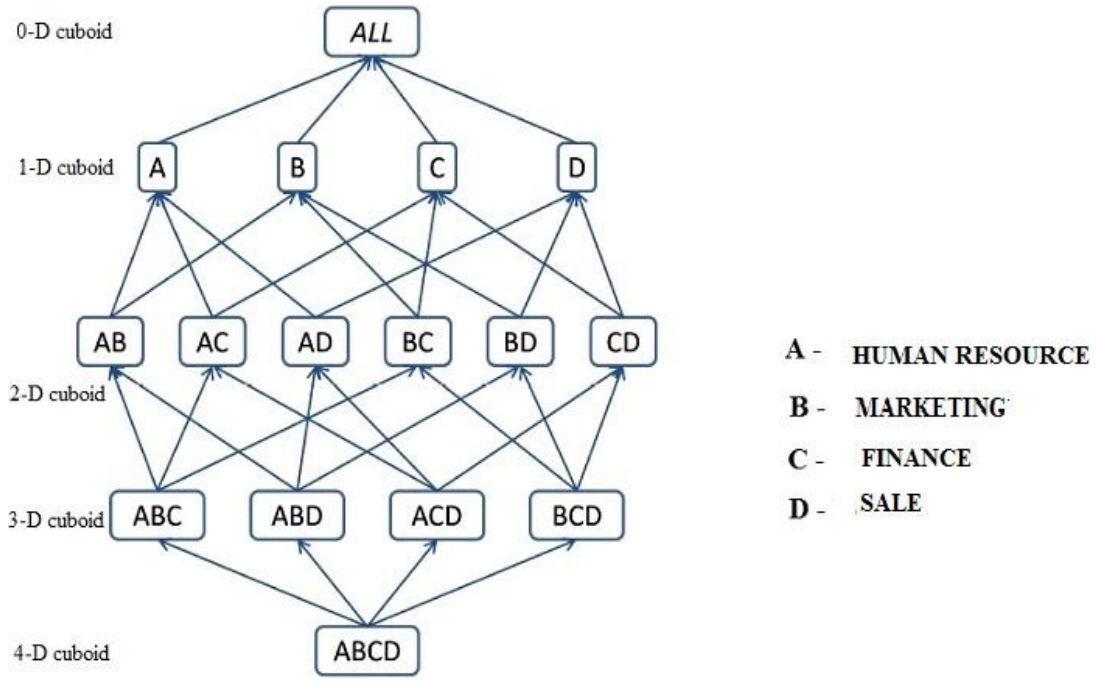


Figure : Lattice Structure of multi-dimensional cube

Facts are also known as key performance indicator in case of data warehousing. The aggregated information of fact can be viewed as 0-D level which is top most hierarchical level. The same info can be viewed at 1-D level as Department or Employee_Info or Salary. Further the combination of two dimensions can be analyzed the 2-D level and 3-D level as combination of three dimensions [34]. The combination of different dimensions with along various axis browse the OLAP cube which helps us upto the granularity level analysis of fact data for efficient decision

making. The developed cubes are mainly for Data analysis. Data analysis is the process to evaluate predictive information from large databases to find out information about the present status and predict future trends in the concerned sector of economy.

Implementation of Study

Now to start work with implementation of proposed study the implementation phase contains various phases to accomplish the objectives of proposed study. The various steps are given below:-

- Create database connection using Sql Server
- Import summarized dataset by using import wizard
- Create dimensions
- Create levels of dimensions
- Define hierarchies of levels within dimensions
- Design Schema design based on cube's structure
- Defining Cube
- Creating Measures for fact findings
- Assign dimensions to Multidimensional Cube
- Load and aggregate data to Cube
- Applying Anonymization to evaluate Sensitive Information
- Applying various access control to Sensitive Information fromunauthorized access

Identify Dimensions and Levels

In the initial evaluation of this study, a dataset is available insummarized form. This dataset required demographic analysis to identify various dimensions and levels of dimensions. After demographic analysisof the dataset of an organization, three of the dimensions are identifiedi.e.

Employee_Info which contains various levels as Emp_id, Age, Sex andArea.

Salary which contains various levels as Basic Pay, HRA, Other Allounce and Total Salary.

Department which contains various levels as Human resource, Marketing, Finance and Sale.

Identify the Grain

The "grain" refers to the level of details or finess upto which the data can be analyzed. The grain in a dimensional model is finest level of detail implied by the joining of fact and dimension tables. In this study, a lattice structure schema has been implemented which is an important application of discrete mathematics to enhance aggregation at grain level from different perspectives.

Identify the Facts

Once the schema is designed then it is very much easier to find the relevant facts from multidimensional cube upto the grain level of details. Before identifying the relevant facts/measures should identify the various elements of fact tables in the database design.

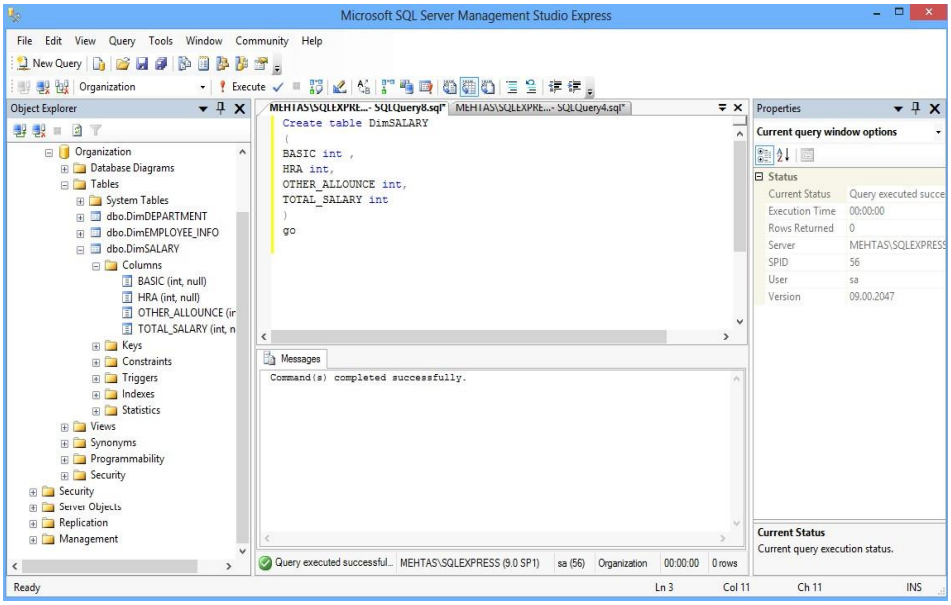


Figure : Creating Dimensions, levels of dimensions for multi-dimensional cube

To identify the facts upto the granularity level of analysis it must be required to apply aggregations of various dimensions into cube. The star design exposed by cubes is very similar to traditional table-based star models. The dimension views form a constellation around one or more cube views. However, there are two key differences:

- Fact tables in a star schema store detail data, while the cube views reveal many summary levels.
- Calculations in a cube are simply exposed as columns in the cube view, and the computation for the equations occurs in the inference engine.

The implementation of star schema is a most important step of proposed study to fulfill all the defined objectives. The figure states the structure of cube by defining as fact table and Department, Employee info and Salary as dimension table or view. The structure lies under the rule of star schema because all of the dimensions are directly attached to fact table rather than through other dimension.

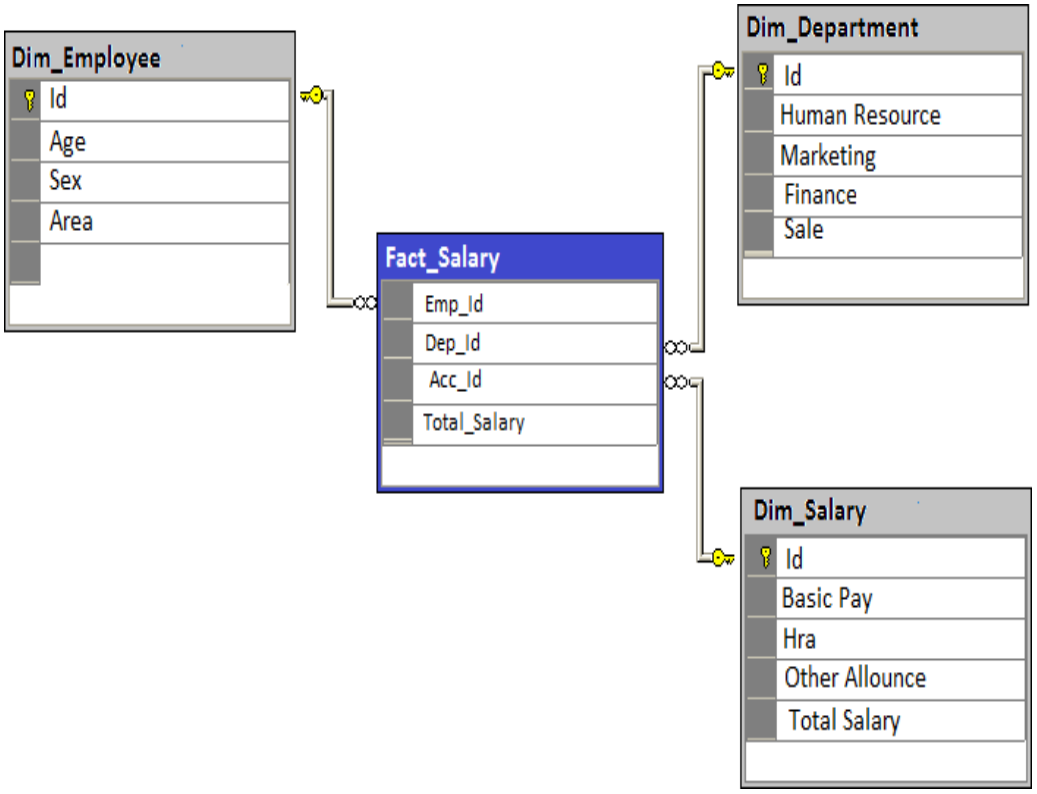


Figure : Star schema of proposed study

Conclusion

After completing the first objective of the study, it clarifies that the multi-dimensional cube is true model of dimensionality.

The multidimensional cube of data that is designed in this study is a best suitable for dimensioning of summarized data. OLAP provide better performance for accessing summarized data. The most important mechanism in OLAP which allows it to achieve such performance is the use of aggregations. Aggregations are built from the fact table by changing the granularity on specific dimensions and aggregating up data along these dimensions. The number of possible aggregations is determined by every possible combination of dimension granularities by using lattice structure.

After completing all of the implementation of the each and every detail of the first objectives of the research, the researcher come to a conclusion which clarifies that the Multi-dimensional cube of is highly efficient in query performance and evaluation of results.

OLAP's multidimensional view of data provides the foundation for analytical processing through flexible information access. It's very important benefit, because more control and timely access to strategic information equal more effective decision-making.

To fulfill the requirements of second objective, the multidimensional cube of data required various types of privacy preserving access control mechanism because it contains sensitive information. Sensitive information is somewhat like relevant facts which are too important for the privacy of organization's policies. Therefore, anonymization on sensitive information is applied in this study by using a threshold value

REFERENCES

1. P. P. S. Chen. The Entity-Relationship Model: Toward a Unified View of Data. ACM Transactions on Database Systems, 1(1):9–36,1976.
2. R. Kimball, L. Reeves, W. Thornthwaite, and M. Ross. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. John Wiley &

Sons,Inc., 1998.

3. Pedersen, Torben Bach, and Christian S. Jensen. "Multidimensional data modeling for complex data. "Data Engineering, 1999. Proceedings., 15th International Conference on. IEEE, 1999
4. Wang, L., Jajodia, S., & Wijesekera, D. (2004, May). Securing OLAP data cubes against privacy breaches. In Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on (pp. 161-175). IEEE.
5. L.Wang, Jajodoa. Sand Wijesekera .D. OLAP and Data Cubes. In ISBN:978-0-387-46273-8,p-180,2007.
6. Cuzzocrea, A., & Saccà, D. (2011).A Constraint-Based Framework for Computing Privacy Preserving OLAP Aggregations on Data Cubes. In ADBIS (2) (pp. 95-106).
7. Vignesh, M. R., & Kumar, C. S. (2012). OLAP Online Privacy Control.Bonfring International Journal of Software Engineering andSoft Computing, 2(Special Issue Special Issue on Communication Technology Interventions for Rural and Social Development), 06-10.
8. Agrawal, R., Srikant, R., & Thomas, D. (2005, June).Privacy preserving OLAP. In Proceedings of the 2005 ACM SIGMOD international conference on Management of data (pp. 251-262). ACM.
9. Evfimievski, A., Gehrke, J., & Srikant, R. (2003, June). Limiting privacy breaches in privacy preserving data mining. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium onPrinciples of database systems (pp. 211-222).ACM.
10. Sung, S. Y., Liu, Y., Xiong, H., & Ng, P. A. (2005). Privacy preservation for data cubes. Knowledge and Information Systems,9(1), 38-61.
11. Liu, Y., Sung, S. Y., & Xiong, H. (2006). A cubic-wise balance approach for privacy preservation in data cubes. Information Sciences, 176(9), 1215-1240
12. S. Chaudhuri, T. Dutta, and S. Sudarshan, “Fine Grained Authorization through Predicated Grants,” Proc. IEEE 23rd Int’l Conf. Data Eng., pp. 1174-1183, 2007
13. Gal, T. S., Chen, Z., & Gangopadhyay, A. (2008). A privacy protection model for patient data with multiple sensitive attributes. IGI Global, 28-44.
14. Vasudevan, L., Sukanya, S. D., & Aarthi, N. (2008). Privacy Preserving Data Mining Using Cryptographic Role Based Access Control Approach. In Proceedings of the International Multi Conference of Engineers and Computer Scientists (Vol. 1).
15. Matatov, N., Rokach, L., & Maimon, O. (2010). Privacy-preservingdata mining: A feature set partitioning approach. Information Sciences, 180(14), 2696-2720.
16. Thavavel, V., & Sivakumar, S. (2012). A Generalized Framework ofPrivacy Preservation in Distributed Data mining for Unstructured Data Environment. IJCSI International Journal of

Computer ScienceIssues, 9(1), 1694-0814.

17. Lamba, M. S., & Abbas, S. Q. A Model For Preserving Privacy Of Sensitive Data
18. P.K. Srimani, & Rajasekharaiah K.M. (2014). A Data Centric Privacy Preserved Mining Model for Business Intelligence ApplicationsPervaiz, Z., Aref, W. G., Ghafoor, A., & Prabhu, N. (2014). Accuracy-Constrained Privacy-Preserving Access Control Mechanism for Relational Data.Knowledge and Data Engineering, IEEE Transactions on, 26(4), 795-807.
19. Cuzzocrea, A., & Bertino, E. (2014, January).A ComprehensiveTheoretical Framework for Privacy Preserving Distributed OLAP. In On the Move to Meaningful Internet Systems: OTM 2014 Workshops(pp. 117-136). Springer Berlin Heidelberg.
20. E. Thomsen. OLAP Solutions: Building MultidimensionalInformation Systems. Wiley, 1997.
21. R. Kimball. The Data Warehouse Toolkit. Wiley Computer Publishing, 1996.
22. Schrader, M., Vlamis, D., Nader, M., Claterbos, C., Collins, D., Campbell, M., & Conrad, F. (2009). Oracle Essbase & Oracle OLAP. McGraw-Hill, Inc..
23. Pedersen, T. B., Christian S. J. (1999). "Multidimensional data modeling for complex data." Data Engineering, Proceedings., 15thInternational Conference on. IEEE. Laurent, B. Bouchon-Meunier, A. Doucet, et. al., “Fuzzy Data Mining from Multidimensional Databases”, Int. Symp. On Computational Intelligence, Studies in Laurent, B. Bouchon-Meunier, A. Doucet, “Towards Fuzzy- OLAP Mining”, in Proc. Work. PKDD "Database Support for KDD",pp. 51-62, 2001.
24. Connolly, T., & Begg, C. (2005). A Practical Approach to Design, Implementation, and Management.
25. E. F. Codd. (1993).Providing OLAP (on-line analytical processing) touser-analysts: An IT mandate. E.F. Codd and Assoc.
26. Acharya, S., Prasad, R.N. Fundamentals of Business Analytics Wiley India. In ISBN: 978-81-265-3203-2, pp-166-203.
27. Winter, R., Strauch, B. (2003).A Method for Demand-Driven Information RequirementsAnalysis in DW Projects. In Proc. of 36th Annual Hawaii Int. Conf. onSystem Sciences, pages 231–239. IEEE.
28. Berson A. and Smith S.J. (1997). Data Warehousing, Data Mining, & OLAP. New York, NY: McGraw Hill Companies Inc.
29. Pendse N. and Creeth R. (2001). The OLAP Report. Available at www.olapreport.com
30. P. Vassiliadis. (1998). Modeling multidimensional databases, cubes and cube operations. In 10th Int.Conference on Scientific and Statistical Database Management (SSDBM’98), IEEE Computer Society Press, pages 53–62.
31. Chaturvedi, K.K., Rai, A., Dubey,V., Malhotra,P.K. (2006).OLAP in Agriculture using Multidimensional Cubes IASRI, New Delhi.

32. Samarati. P, "Protecting Respondents Identities in Microdata Release, IEEE Trans. Knowl. Data Eng. 13(6):1010-1027(2001).

33. Moody D.L., Kortink, M.A. (2000). From Enterprise Models to Dimensional Models:A Methodology for Data Warehouse and Data Mart Design. In Proc. of 2nd Int.Workshop on Design and Management of DataWarehouses.CEUR-WS.org.

34. K. LeFevre, D.J. DeWitt and R. Ramakrishnan, "Workload- AwareAnonymization Techniques for Large-Scale Datasets," ACM Trans.Database Syst., vol. 33, no. 3, pp. 1-47, 2008.