ISSN: 2321-7782 (Online)
Volume 3, Issue 11, November 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Advanced Data-Intensive Applications Using Multidimensional Analysis and Generalization of Class Composition Hierarchies

Yannam Apparao
Associate Professor in CSE,
Computer Science and Engineering Department,
Marri Laxman Reddy Institute of Technology & Management (MLRITM),
Dundigal, Quthbullapur Mandal, Hyderabad, Telangana, India.

Abstract: ***This paper gives scientific research and engineering design is advanced data-intensive applications which are used for storing, indexing, accessing and manipulating complex data objects. It is a very difficult task to represent those complexes objected as simple and consistent structured records. To serve the application requirements, such as efficient storage capacity and accessing large amount of dist which deals with complex structure data objects motivated to design and develop database system.***

Keywords: Data, Data Base, storing, indexing, and accessing, Object-relational data base systems, object-oriented data base systems.

I. INTRODUCTION

This paper explains the application requirements such as storage capacity and accessing large amount of space provided by Object-oriented Relational Data Base systems and massive volume of data which deals with complex structure data objects **[1][2].**

II. PROBLEM STATEMET

Top level management want to know the information about their organizational different products information in the form of income by yearly ,off yearly and quarterly sometimes month wise along with country ,state ,city and related customers **[3]** if it is regular structure query language may not get relevant data for management so management may not get the knowledge about their products whether the product manufacture data increase or decrees but in Multi-Dimensional Analysis is an Informational Analysis on data which takes into account many different relationships, each of which represents a dimension in this paper providing the advanced technique called object oriented and object relational data base systems with generalization of structured data .

III. LITERATURE SURVEY

3.1. Introduction to data mining

Data mining is a process of extracting knowledge from massive volume of data. It refers to a way of finding significant and useful information from an organization's data base, the knowledge which is extracted can include pattern types, associated rules and different trends. Organizations that make use of data mining techniques are benefited in their corresponding business area by identifying the significance trends and anomalies that were not possible to be detected by human analyst **[4].**

3.1.1. Reasons for using data mining

a. Knowledge discovery
b. Data Visualization

c. Data correction

Knowledge discovery: *The* objective of knowledge discovery process is to identify the invisible correlation, patterns, and trends available in the database.

Data Visualization: The objective of Data *Visualization is to 'harmonize 'large volume of data so as to find a sensible way of displacing data*

Data correction: The processes is used to identify and correct incomplete, erroneous, inconsistent data

3.1.1.1. Knowledge Discovery (KDD)

KDD is refers to a process of extracting use full knowledge which may include correlation patterns between different data objects. it is highly repetitive and influential approach [5]. "Knowledge Discovery is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [Osama Fayyad et al., 1996].

a. Data cleaning and preprocessing stage

b. Data integration stage

c. Data selection stage

d. Data transformation and reduction stage

e. Data mining Discovery stage

f. Pattern interpretation and analysis stage

g. Knowledge visualization stage

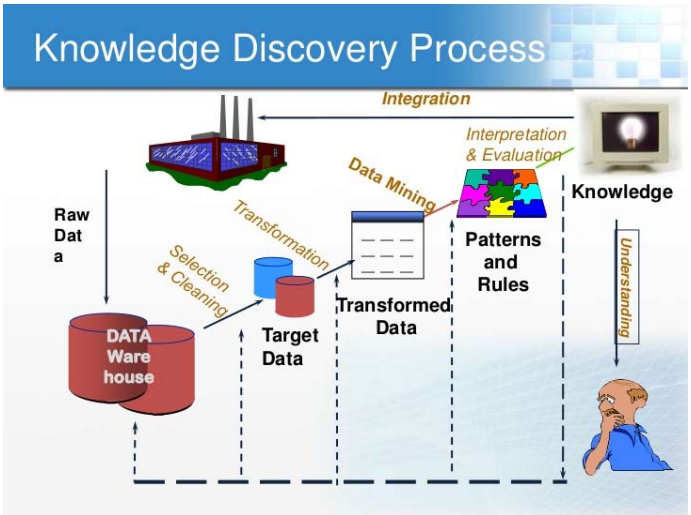

Figure 3.1.1.1: steps of Knowledge Discovery

3.2. Architecture of data mining

Data mining system consists of following components

1. Data source or repositories: This component presents multiple and heterogeneous source of data such as set of databases
2. Database server or data warehouse server: This component is responsible for extracting potentially useful, desired data from data repositories depending on the data mining request made by the user
3. Knowledge information base: it is an area of knowledge that is used to guide the search and to perform analysis of the desire pattern

4. Data mining engine: it is a core component to data mining system which includes different individual operational units for each function like characterization, classification, prediction, outlier analysis, cluster analysis.

5. Pattern analysis module : this component uses various interestingness procedures and communication with different data mining modules so as to concentrate on searching the desired pattern

6. User interface module: the communication between the user and data mining system is made possible with the help of user interface module.

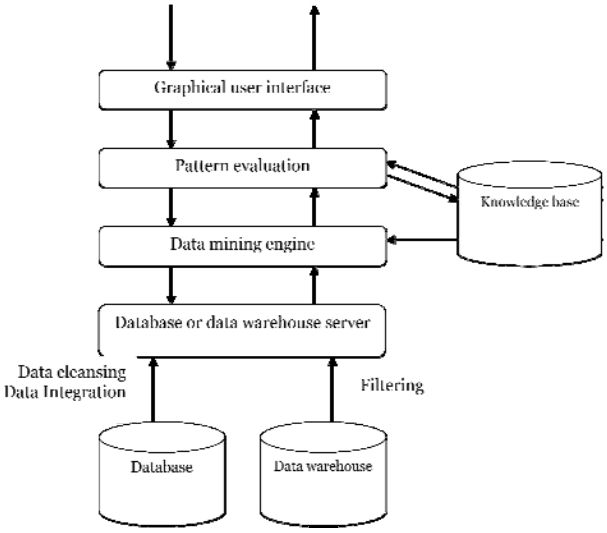

Figure 3.2: Architecture of data mining

IV. SYSTEM DESIGN

Scientific research and engineering design is advanced data-intensive applications which are used for storing, indexing, accessing and manipulating complex data objects. It is a very difficult task to represent those complex objected as simple and consistent structured records .To serve the application requirements ,such as efficient storage capacity and accessing large amount of dist which deals with complex structure data objects motivated to design and develop database system .

There are two kinds of data base systems,

1. Object relational data base systems
2. Object oriented data base systems

In the field of data base systems ,rigorous research is conducted on object-relational and object oriented database systems to know how efficiently complex object can be indexed ,stored ,accessed and manipulated ,In these systems ,a large set of complex data objects are organizes into classes and subclasses .

A class containing object is associated with

a. An identifier for each object i.e object- identifier

b. A set of attribute that include list-valued data, set valued data, hierarchies of classes and subclasses, multimedia data.

c. A set of methods, A method specifies the evaluation rules associated with the object class.

A systematic analysis and mining of large-scaled complex structured data objects is said to be a complex object data, It consists of two impotent tasks.

1. Construction of multidimensional data warehouses for complex object data and performing OLAP operations.

2. Efficient methods are developed for performing knowledge discovery by extracting relevant knowledge from different data warehouses.

4.1. OLAP Operations

Data in multidimensional model is arranged at different level of granularity of defined by concept hierarchies. These hierarchies specifies mapping from a specialized low level concept to more general high level concept. Multiple concepts hierarchies are not directly expressed with in data base schema. Because of this granularity, users are provided with a benefit of viewing data from different point of views. In order to materialize these views numerous OLAP data cube operations are executed. These operations allow flexible interactive querying and data analysis **[9].**

The different OLAP operations are,

1. Drilling operation :
2. Slicing operations
3. Dicing operation
4. Rotation operation

4.2. Limitations of Data warehouse and OLAP

1. Multidimensional data analysis is confined to use limited numbers of data types associated with dimensions and measures
2. Data cube implementations restricts dimensions to ,
    a. Categorical data.
    b. Measures which are simple aggregated values.

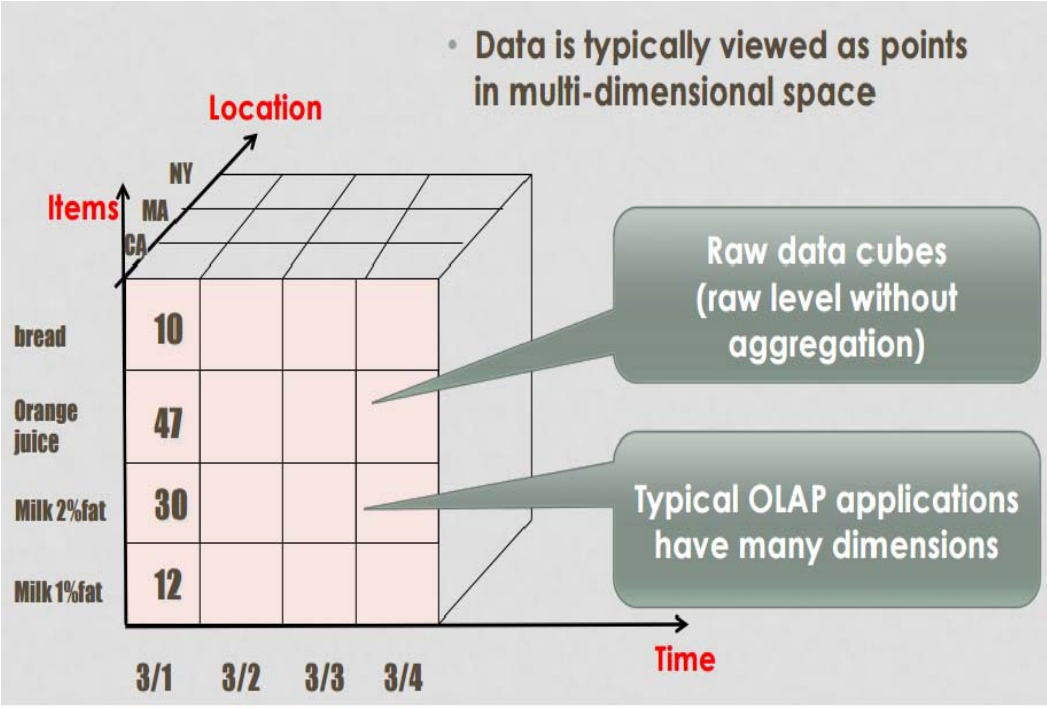

Figure 4.1.OLAP Multidimensional view

gender	age	accident
Male	27	3
Male	37	1
Male	37	0
Male	37	1
Male	49	2
Male	39	4
Male	43	0
Male	41	2
Male	49	1
Male	44	2
Male	43	3
Male	53	4
Male	60	0
Female	26	0
Female	39	0
Female	45	2
Female	41	2
Female	39	1
Female	37	0
Female	43	1

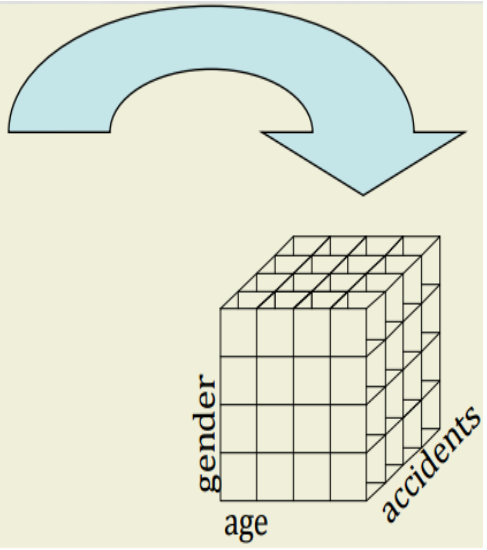

Figure 4.1.Example of OLAP Multidimensional view

V. System Architecture

5.1. Generalization of structured data

The essential characteristics of object-oriented and object relational database systems are their ability to store, access and model complex structured valued data. The structured data can be one of the following **[10]**.

1. **Set-valued data:** The type of set-valued attribute can be either heterogeneous or homogeneous.
   - **a.** Generalization each value in a set into their respective higher-level concepts.
   - **b.** Determining general behavior of the set like, number of elements, types or value ranges, weighted mean value for numerical data.

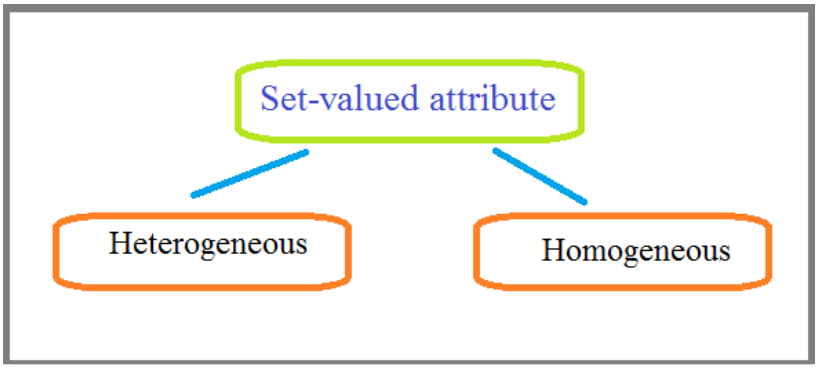

Figure 5.1. Set-valued attribute

Example: Consider that, the co-curricular activity of a student is a set-valued attribute containing the set of values {chess, table-tennis, carroms, painting, cricket, football, and hockey}, each value in the set can be generalized to (indoor-games, art, outdoor games) and this set is consider as higher level concept.

List valued data:

List valued attribute is also follow similar generalization approach as that of set-valued attributes, but, the difference is that the elements order in the list valued attributes should be observed during generalization. It is possible to generalize individual value present in the list into respective higher-level concept.

Complex structure –valued data:

A complex structure valued attributes comprises following data structure elements sets, tuples, Lists, Free and records

5.2. *Generalization of class composition Hierarchies*

A class composition hierarchy is formed, when an attribute of an object may be described by another by another object. Some of whose attributes may be in turn described by another objects, generalization class composite hierarchy is similar to the generalization performed on set of nested structured data. Sometimes, class components hierarchy may lead to recursive which can be viewed the bellow figure [10]. In the figure 5.2 object A describes two attributes B and A1,one of the attributes of A i.e B acts as an object which describes the attributes A2 and C .object C which is an attribute of B describes attributes A3 and A. Now ,the attribute A is again describeing the same object A. Sequence is reapeting agaian and again,thereofore ,resulting in infinity possibility of recurion .

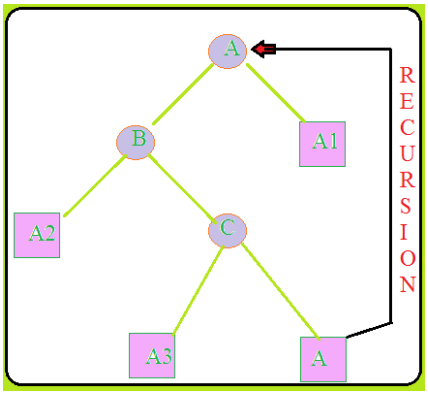

Figure 5.2. Generalization of class composition Hierarchies

VI. IMPLEMENTATION

6.1 Generalization-based mining of plan databases by divide –and-conquer:

Divide –and-conquer approach is emplyed for mining sophisiticated databases.the purpose of this approch is to initially search relenet generalized plan present in a plan database. A plan consists of variable sequence of actions,A plan databse is also known asa planbase.is a database whivh consists of numerous plans ,Once the plans are accumulated,plan database is divided into small distict subplan databases.The partitioning is done depending on the patterns which ae extracted for finfidng out the features of subplan databases [11] [12]. The task of mining a pattern from the plan database is refered to as plan mining.plan mining is a very crucial step required for attaining th epartitions of paln databse . The objective of plan mining is to retrrive a trival pattern that is present in the plan database. This type of mining differs from the way sequential mining is performed. Divide –and-conquer strategy works as follows a. Starts the mining of plan databases by using multidimensional generalization to get the shor sequence of plans .b. Partitions the plan base is depending on the mined sequences in order to get the features of sub plan base.

Example: A Railway plan Base: Suppose a railway plan base consists of the following tables.

Plan_Number	Action_Number	Dep	Dep_Time	Arr	Arr_time
1	1	A	8:00	B	9:00
2	2	B	10:00	C	12:30
2	2	C	13:00	E	16:00
3	3	D	17:00	A	18:00
4	4	E	9:00	D	9:50
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Table [1]:A Railway planbase

Description of table [1]: “A Railway plan base” table stores the following information

1. Plan_Number: This number indicates a particular plan to be taken, It assists the passengers by giving necessary information associated with arrivals ,departure time of the trains A plan_Number can be used more than once to indicate that a single plan has a sequence of actins
2. Action_Number: The Action_Number is a primary key in another database that shows the sequence of actions to be taken for a particular Action_Number
3. Dep: This column indicates a railway station involved
4. Arr: Arr column means railway station involved

VII. CONCLUSION

Paper gives about the advanced data-intensive applications which are used for storing and retrieving and manipulating complex data objects, generally use front end as the application and data storing have to use back end, Hence the front end application uses object oriented or object based concept where in back end also uses the same concept where the concepts are Object-relational data base systems, object-oriented data base systems. Both fronts, back end applications uses same class and object concepts, hence storing and retrieving of complex data will be efficient.

References

1. http://web.engr.illinois.edu/~hanj/cs512/bk2chaps/chapter_10.pdf
2. http://www.slideshare.net/akhilachilukuri1/dwdm-unit8-jntuworld-1
3. http://www.learn.geekinterview.com/data-warehouse/data-analysis/multi-dimensional-analysis.html
4. http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm
5. http://www.usc.edu/dept/ancntr/Paris-in-LA/Analysis/discovery.html
6. https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd
7. Data Mining Concepts and Techniques –Jiawei Han and Micheline Kamber, 2003
8. Handbook of Data Mining and Discovery- Willi Klosgen and Jan M Zytkow, 2002
9. http://www.cs.ccsu.edu/~markov/ccsu_courses/datamining-2.html
10. http://jmlr.org/proceedings/papers/v37/eggeling15.pdf: **Ivo Grosse**: Martin Luther University Halle-Wittenberg, Germany German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
11. https://www.cs.berkeley.edu/~vazirani/algorithms/chap2.pdf
12. http://web.cse.ohio-state.edu/~gurari/course/cis680/cis6802.html#QQ2-50-118

AUTHOR(S) PROFILE

Yannam Apparao, received the M.Tech Post Graduation Degree in Software Engineering from Jawaharlal Nehru Technological University Hyderabad in 2010 and Currently working as Associate Professor, Marri Laxman Reddy Institute Of Technology & Management, Dundigal, Quthbullapur (M), R.R. District, Telangana-500043.INDIA.(yapparao@mlritm.ac.in)

