iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** IV **Month of publication:** April 2014

DOI:

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Online Analytical Processing System Providing spatial information to the data warehouse by using Geographical Cube Methodology

[1]A.V. S. M AdiSeshu, [2]B Kumara Swamy, [3]Errolla Venkataiah

[1]Assistant professor [2]Assistant professor, [3]Assistant professor
CSE Department, Sree Dattha Institute of Engineering & Science

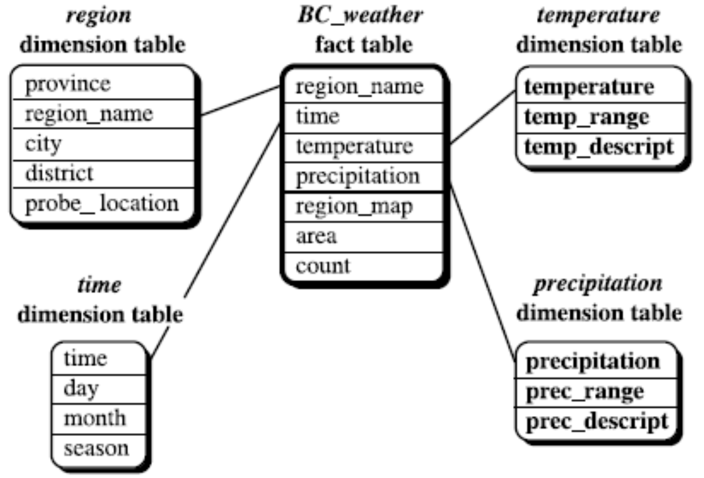
Abstract- Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial databases. A spatial database stores a large amount of space-related. They carry topological and/or distance information, usually organized by sophisticated, multidimensional spatial indexing structures that are accessed by spatial data access methods and often require spatial reasoning, geometric computation, and spatial knowledge representation techniques. The main objective of this paper is to deal with the categorical and non categorical attributes by providing spatial and standard online analytical processing systems to the data warehouse. The major issues we are discussing in this paper is, spatial or continuous values and representing and updating by applying efficient query processing with indexing. Finally we present Geographical Cube storage and indexing procedure to aggregate the spatial information/Continuous values with significant performance.

Key Word: data warehouse, Geographical Cube, indexing, spatial data, aggregation, OLAP

1. INTRODUCTION

A spatial database stores a large amount of space-related data, such as maps, preprocessed remote sensing or medical imaging data, and VLSI chip layout data. Spatial databases have many features distinguishing them from relational databases. They carry topological and/or distance information, usually organized by sophisticated, multidimensional spatial indexing structures that are accessed by spatial data access methods and often require spatial reasoning, geometric computation, and spatial knowledge representation techniques.

Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial databases. Such mining demands an integration of data mining with spatial database technologies. It can be used for understanding spatial data, discovering spatial relationships and relationships between spatial and non spatial data, constructing spatial knowledge bases, reorganizing spatial databases, and optimizing spatial queries. It is expected to have wide applications in geographic information systems, geo marketing, remote sensing, image database exploration, medical imaging, navigation, traffic control, environmental studies, and many other areas where spatial data are used. A crucial challenge to spatial data mining is the exploration of *efficient* spatial data mining techniques due to the huge amount of spatial data and the complexity of spatial data



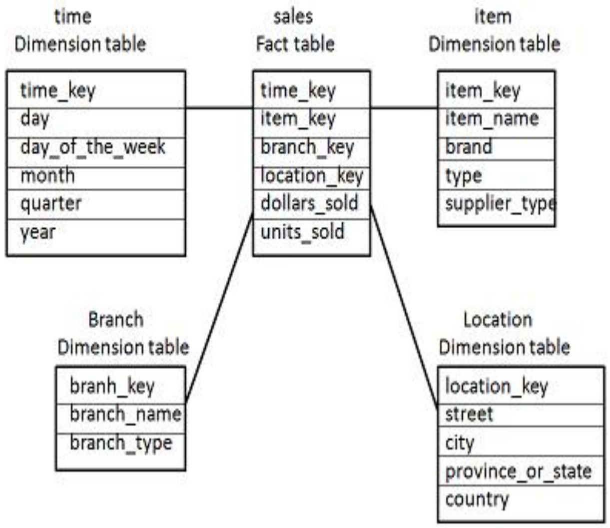
Vol. 2 Issue IV, April 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

types and spatial access methods.

Indexing of multi-dimensional point is a well studied topic. The R-Tree is a spatial indexing structure that allows the indexing and efficient query of points, polygon shapes and it is considered to be the most commonly recommended and accepted indexing. The basic difference between traditional OLAP System and SOLAP System is that features in OLAP are categorical. Features can exploit in several ways, such as categorical data stored in *Star Schema.*



A star schema of the *BC weather* spatial data warehouse and Star Schema of a typical data warehouse

Categorical dimensions can be normalized in each dimension leading to the importance of fixed cardinality. The basic purpose of defining dimension hierarchy is to support drill down and roll up operations. It is not compulsory that every captured data is categorical or structured, unstructured data is also captured often. It is not compulsory that every captured data is categorical or structured, un structured data is also captured often Dealing with non categorical dimensions is real challenge for researcher and developers.

This describes the original continuous dimension in data dependent manner and once continuous dimension have been transformed in this way, recompilation of the complete view becomes mandatory to view updates with respect to the original Space Research and business interest in OLAP with mixed dimensions has proposed various representations and indexing procedures for various things such as sequential parallel and peer to peer. In this paper, we propose GCUBE for representation and indexing of mixed dimensions in a Relational OLAP setting. Our proposed approach is intended for the indexing of views in ROLAP setting as the resulting data structure only represents those sections of the multidimensional view that data records associated with them.

2. RELATED WORK

Over the many number of years, the problem of indexing multidimensional data has been studied and the most commonly employed approach for effective indexing is R-Tree and its variants like R+-Tree, R*-Tree, SS-Tree and SR-Tree. The R-tree is a spatial indexing structure that allows the indexing and efficient querying of points and polygonal shapes. Real time range queries are supported by R-Tree and it is inspired by B-Tree. Problems could be calculation, methodology, complexity which intern lead to the difficulty of updates. The efficient aggregation in OLAP queries over categorical dimensions often crucially relies on these dimensions being integer-valued, and the use of space-filling curves as a locality preserving mapping from higher-dimensional space into one dimension. This approach organizes multi-dimensional categorical OLAP views by using the Hilbert space filling curve to generate a linear ordering of the records in multi-dimensional space and then indexing this linear ordering with a data structure similar to a B-tree. This method exploits that the Hilbert curve strongly preserves spatial locality.

3. ROLAP

The objective of this paper is to address the issue of the representation and indexing of mixed data is a ROLAP setting. This approach fails leading when massive data is available to work with. Sequential, parallel and peer to peer are sure of the employed representation and indexing techniques. Then the indexing is analyzed based on one space filling curves can be extended to mixed dimension setting with mapping of Multi-Dimensional Data into Linear Ordering using Hilbert Curve, use the Linear Order to distribute the data over the available storage and Construct an index structure on top of the Ordered Data for efficient Query Processing. R-Tree is being generated; however one of space

Vol. 2 Issue IV, April 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

filling curves to fold the multi-dimensional space into a one dimensional space well suited for even distribution of records over multiple disks processors or peers. It facilitates the batch updates into the original view with a single linear scan. The proposed approach uses Hilbert Curve as the space filling curve that indexes our method.

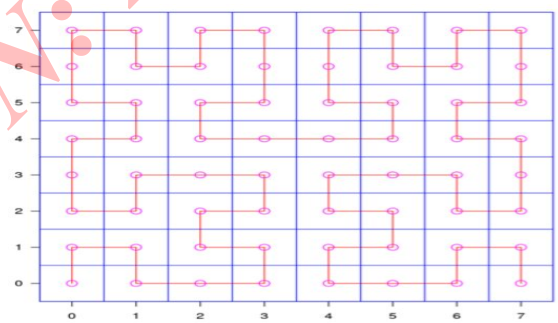
4. HILBERT CURVES

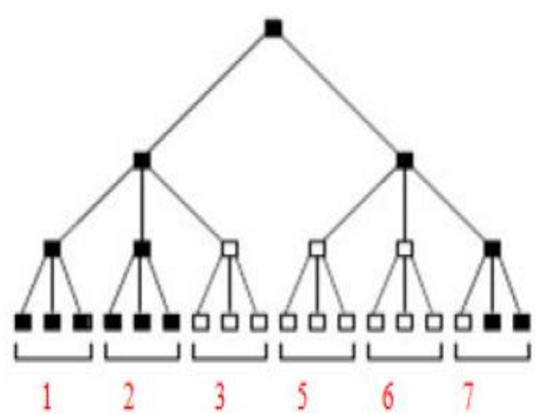
Hilbert curve provides the facility of mapping d-dimensional Space to 1-dimesional Space without losing the information of higher dimension [4].This new approach is flexible with respect to the resolution of the grid as it attempts to optimally utilize the grid space for continuous dimensions. It also reduces the grid cells compare to pre-descretization approach and new records are allowed with previously unknown attribute values without re-computing the entire view. Pair wise exclusion is necessary for predetermination of Grid resolution and duration time between records plays most crucial role when dealing with large records. Based on the recursive definition and the self-similarity of the Hilbert Curve, the Hilbert Ordering of records can be efficiently computed in continuous space by using an adaptive resolution for the underlying descretization grid. It is observed that the relative order of records is preserved by Hilbert ordering when the resolution of the Hilbert curve is increased.

5. INDEXING

Once the Hilbert ordering of the records has been determined, the records are sequentially written to disk in a block-wise manner. Each block on disk stores a constant number B of records that are consecutive in Hilbert order. Over the ordered records an indexing structure is formulated which provides features comparable to those of a combination of B-Tree of an R-Tree. The following figures illustrate the construction of such a tree structure with specific example. While each intermediate node of the tree is very similar to node in conventional B-Tree. It is also annotated with minimum bounding box of the records in its sub tree, similar to nodes in R-Tree. Due to the self similarity of the Hilbert curve and its properties to not favor any specific dimensions, the bounding boxes of the records in each level of indexing structure are usually flat and mutually overlap only very little and helps in improving performance of the query. To help aggregate OLAP queries efficiently, the aggregate information to be stored at each node is pre computed, while iterating over the children of the node during the bottom up construction of index tree. Note that this pre-computation of aggregate information works well for distributive (e.g. COUNT, SUM) and algebraic (e.g. AVERAGE) aggregation functions. Holistic aggregation functions (e.g. MEDIAN) on other hand require the retrieval of the actual records enclosed in the query region in order to compute the aggregate value. This is supported with direct reference sequence of lines in its sub tree. When evaluating a query the records in a node's sub tree can be reported immediately once the node has been reached, without traversing remaining sub tree below this node.

Records in Hilbert order are stored in consecutive blocks on disk and from multidimensional regions in the original space.





ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

6. CONCLUSION

This paper presents the dynamic system for representation and indexing relational OLAP System with mixed categorical and continuous dimensions. The proposed approach is dynamic in nature and flexible enough with respect to dimension cardinality. Therefore, it allows from for the indexing of continuous space while building on top of established mechanisms for index construction, querying representation of mixed categorical /continuous data at storage level is the contribution of this research work. The outcome of the current research is the efficient SOLAP system that is capable of handling massive data. Empirical results imply the practical benefits of the proposed indexing approach. Hilbert curve space may also be applicable to multidimensional storage of Disks (MOLAP) as future direction of current research.

7. REFERENCES

[1] Guttmann, A., (1988) "R-trees: a dynamic index structure for spatial searching", Readings in database systems, pp599–609.
[2] Gutting, R. H., (1994) "An introduction to spatial database systems", The VLDB Journal, Vol.3, Issue 4, pp357–399.
[3] Rigaux, P, M. Scholl, M, and Voisard, A (2002) "Spatial Databases with Application to GIS", Morgan Kaufmann.
[4] Jagadish, H, V., (1990) "Linear clustering of objects with multiple attributes", In Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, pp332–342.
[5] Moon, B, Jagadish, H, Faloutsos, C and Saltz, J, H.,(2001) "Analysis of the clustering properties of the Hilbert space-filling Curve", Knowledge and Data Engineering, Vol.13,Issue 1, pp124–141.
[6] Edzer Pebesma, (2012) "Space-time: Spatial-Temporal Data in R-Tree", Journal of Statistical Software, Vol.51 Issue 7 pp123-130.
[7] Schmidt, C and Parashar, M.,(2004) "Enabling flexible queries with guarantees in P2P systems",IEEE Internet Computing,Vol. 08 Issue3,pp19–26.
[8] Papadis, D, Kalnis P, Zhang, J and Tao Y.,(2001) "Efficient OLAP Operations in spatial data warehouses", Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases, Springer, pp443–459.
[9] Papadis, D, Kalnis P, Zhang, J and Tao Y.,(2002)"Indexing spatio-temporal data warehouses", Proceedings 18th International Conference on Data Engineering, IEEE, pp166–175.
[10] Tao, Y and Papadis, D., (2001) "MV3R-tree: A spatio-temporal access method for timestamp and interval queries",Proceedings of the 27th International Conference on Very Large Data Bases, Morgan Kaufmann, pp431–440.
[11] Dehne, F, Eavis, T and Rau-Chaplin, A, (2003) "Parallel multi-dimensional ROLAP indexing",
[12] Kamel, I and Faloutsos,C,(1994)"Hilbert R-tree: improved R-tree using fractals", Proceedings of the 20th International Conference on Very Large Databases, Morgan Kaufmann, pp500–509.
[13] Lewder, J,K and King ,P, J, H.,(2001) "Querying multi-dimensional data indexed using the Hilbert space-filling curve", ACM SIGMOD Record, Vol.30 Issue 1, pp 19–24.
[14] Fidalgo, R, N, Times, V, C, Silva, J and Souza, F, F (2004) "GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas', Data Warehousing and Knowledge Discovery, Springer, pp26-37.
[15] Gomez, L, I, Vaisman, A and Zimanyi ,E., (2010) "Physical design and implementation of spatial data warehouses supporting continuous fields", Data Warehousing and Knowledge Discovery, pp25–
[16] Han, J, Stefanovic, N and Koperski, K., (1998) "Selective materialization: An efficient method for spatial data cube construction", Proceedings of the 2nd Pacific Asia Conference on Research and Development in Knowledge Discovery and Data Mining, Springer, pp144–158.
[17] Rivest, S. Bedard, Y and Marchand, P., (2001) "Toward better support for spatial decision making:Defining the characteristics of spatial on-line analytical processing (SOLAP)', Geomatica, Vol.55 Issue4, pp539–555.

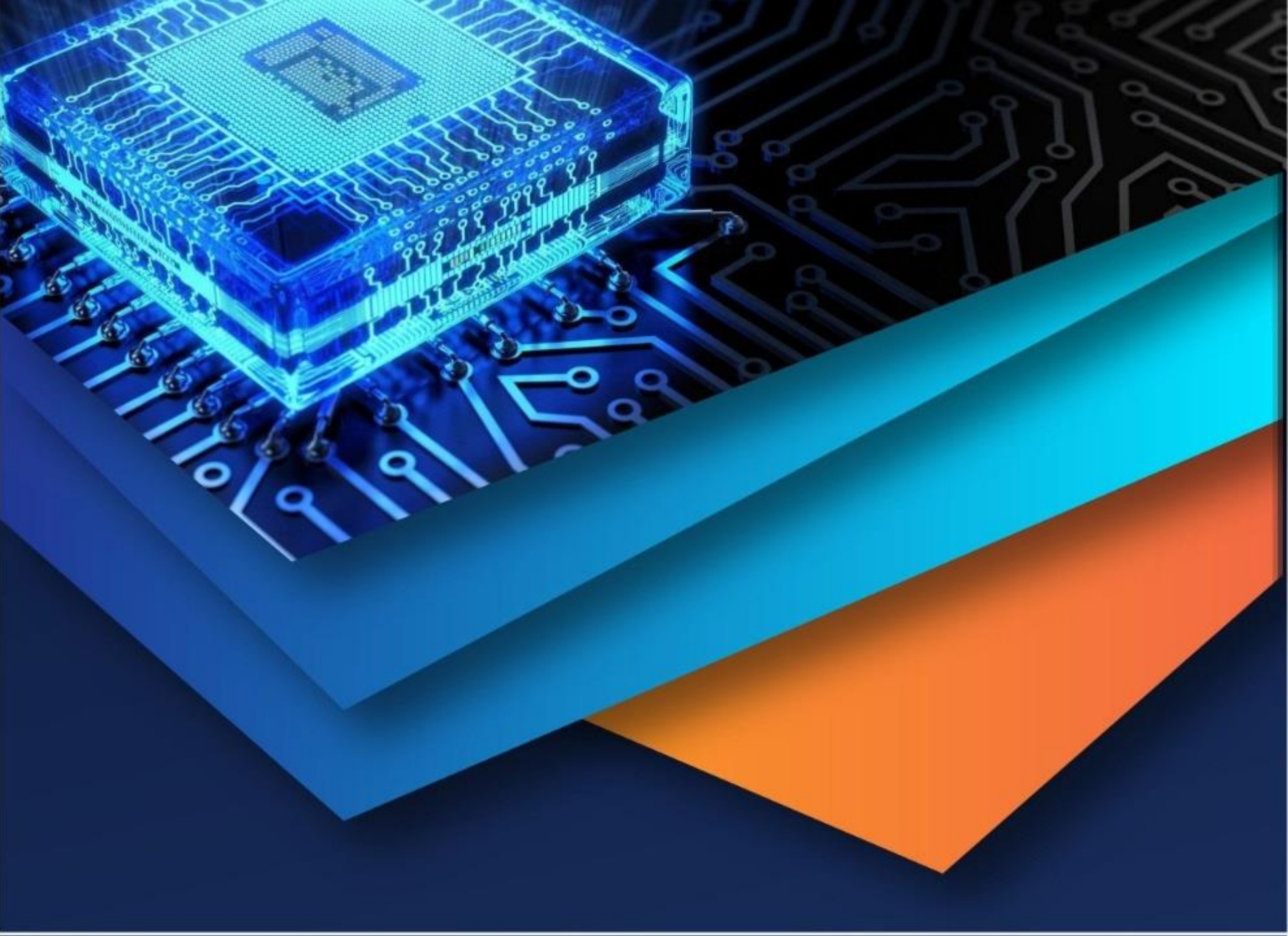
ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Mr. A. V. S. M AdiSeshu received his M.Tech (CSE). Present he is working as an Assistant Professor in the CSE Dept at Sree Dattha Institute of Engineering & Science, Hyderabad.

Mr. B Kumara Swamy received his M.Tech (CSE). Present he is working as an Assistant Professor in the CSE Dept at Sree Dattha Institute of Engineering & Science. Hyderabad.

Mr.E Venkataiah received M.Tech(CSE) from HCU. Present he is working as an Assistant Professor in the CSE Dept at Sree Dattha Institute of Engineering & Science, Hyderabad.. His research includes Automata & Compiler Design.



10.22214/IJRASET

INDEX COPERNICUS 45.98

ISRA JIF IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY