Enhancing Performance of GIS on Cloud Computing

Ahmed Teaima A.T

Institute of Statistical Studies and
Researches.
Cairo Univ.
Cairo, Egypt

Hesham Ahmed Hefny. H. F. H

Institute of Statistical Studies and
Researches.
Cairo Unive.
Cairo, Egypt

BahaaShabana B.S

Misr Higher Institute for Commerce
& Computer, Mansoura Univ.
Mansoura, Egypt

Abstract—**Cloud computing provides a way of determining dynamically scalable and virtualized resources as a service over the Internet. GIS is a technology, which could use Cloud Computing for distributed parallel processing of a large set of data, store and share the results with users around the world. GIS is beneficial and works well when it be available to everyone, everywhere, anytime and with downcast fee of minimal sized in terms of technology and outlay. Cloud Computing used to portray and help users to use GIS applications in an easy way. This paper will study some example of a data structure like a K-d tree and Quad trees of GIS application and compare between them when storing these data structures on Cloud computing, the paper also portrays the results of the study of data structure on cloud computing platforms to retrieve data from cloud computing. The paper provides an application for "finding neighborhood from existing data stored.**

Keywords—Cloud Computing; GIS; Kd-tree; Quadtree

I. INTRODUCTION

Geographical information system (GIS) is a group of Tools that analyzes, stores, manages, captures and presents visual data that are associated with geographical locations" this assumes a definition of the acronym of GIS [11]. GIS or geospatial information studies play a prominent role in many fields and widely adopted nowadays. In another view, it is any information system merging of statistical analysis, cartography hardware, software, and special types of DB (huge size-different shapes -…) and data to provide information and present the result of all these operations. GIS used in decision –making as in public health [12] which describe the relation between distribution diseases and concentration vary in different locations for making best possible decision by using spatial relations between it, visualizing the data to produce information and processing these data. In addition, in a pilot project designed to explore the potential for an information tool and educate sector engagement model to benefit the sector and its communities in the transport corridor to the north of Brisbane. By allowing participants, community, government and non-government organizations (NGO) to access information at a regional level to assist with decision making and the evaluation of shared cross-sector service provision and planning initiatives. [15]

Over a few decades, efforts made to upgrade applications of GIS in order to provide huge spectrum services to the users through the globe. For example, but not limited to, application of integration, GIS and hydrology, by monitoring of Surface water and Groundwater resources is dependent on dynamic and static parameters of these water systems as well as meteorological data sets. All this information is large in volume and spatial as well as temporally varying in Nature. [13] Another one of using GIS in watershed management. By studying the basic characteristic of watershed likes, drainage network and flow paths derived from readily available Digital Elevation Models (DEMs) and USGS's National Hydrography Dataset (NHD) program. [14]. Cloud computing has emerged as a paradigm to deliver on demand resources (e.g., infrastructure, platform, software, etc.) to customers similar to other utilities (e.g., water, electricity and gas). [16]

Cloud Computing can be used across the challenges in GIS applications. GIS is a complete System of Hardware, Software, and Spatial Data (topographic, demographic, graphic image, digitally...) performs processing and analysis operations on those data to produce reports, graphics, statistics, and controls geographic data processing workflows. [1] [8]

II. PREVIEW ON DIFFERENT GIS DATABASES

The GIS has a special natural due to the large amount of data, the way of storing this data, and the experts who deal with it. The recent emergence of cloud computing brings new possibilities in service deployment. Services deployed in environments that made to scale up or down as required, with the service provider only charged for actual usage. Many types of database can be used for storing the spatial data like Quad tree, R tree and K-d tree I will look for those types as a background.

A *quad tree* is a tree data structure used to represent a picture successively in deeper level represented the best subdivisions of picture areas. Each node represents and links to the quadrant of its parent. A process of subdividing an image matrix into four quadrants parts recursively until every part has unique color fills up the tree. A Quad tree is trees whose nodes are either leaf (no children) or have four children Fig 1 show the shape of quad tree structure. The children are arranged one, two, three, and four. The region of the quad tree describes a piece of space in two dimensions (like X and Y) by dividing the region into four equal quadrants, sub quadrants, and so on with each leaf (which mean the node is the last one) node containing data corresponding to a specific sub region. Each node in the tree must have exactly four children, or have leaf node. [17]

The region quad tree is not strictly a 'tree' as the positions of subdivisions are independent of the data. They precisely called 'tries'. A region quad tree with a depth of n (number of levels) used to represent an image consisting of $2n \times 2n$ pixels,

where each pixel value is zero or one [2]. The advantage of the quad tree lies in reducing the complexity of the intersection process by enabling the pruning of certain objects or portions of objects from the query. The disadvantage of quad tree, it must planed scope beforehand. [18]
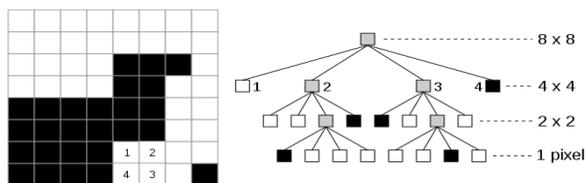

Fig. 1.   Quad tree structure

Rtree is a real extension of B-trees (Comer 1979), which refers to binary search tree, in that a node can have more than two children, unlike self-balancing binary search trees, the B-tree optimize for systems that read and write large blocks of data. B-trees are a good example of a data structure for external memory. A B - tree commonly used in databases and file systems. B-tree does not need re-balancing as frequently as other self-balancing search trees, but may waste some space [19]. The *R tree* (**R**ectangle tree) is a data structure, use in multiple dimensions. Which is height-balanced tree [2]. It consists of two levels of storage, medium and last node (leaf node). The stats of last nodes and medium nodes stored the data objects built by gathering rectangles at the lower level. The collected data in rectangle shape as shown in Fig 2, which illustrate the structure of the R tree. The nodes can be covered, overlapping, or completely disjoint, no assumption about their properties. The Minimum Bounding Rectangles (MBRs) of the actual data objects assume stored in the last node of the tree. Each medium node is associated with some rectangle, which completely encloses all rectangles that correspond to lower level nodes. [3]
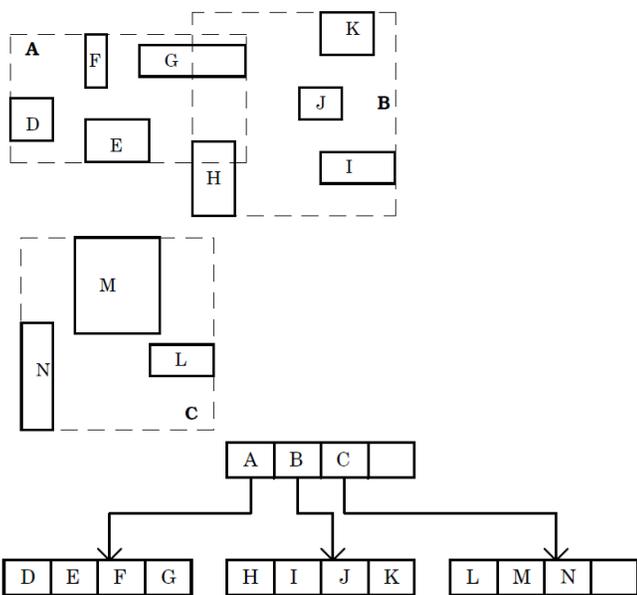

Fig. 2.   R tree structure

k-d tree (short for k-dimensional tree) is a space-partitioning data structure for organizing points in a k-dimensional space. k-d trees are a useful data structure for several applications, such as searches involving a multidimensional search key (e.g. range searches and nearest neighbor searches). k-d trees are a special case of binary space partitioning trees [20].k-d tree also used in computer vision and machine learning algorithms consists of finding nearest neighbor matches to high dimensional vectors that represent the training data.[21]. However, a key problem of data driven tree structures is the capability of data update. Each point insertion or deletion requires the modification of large parts of the actual tree structure. [22]
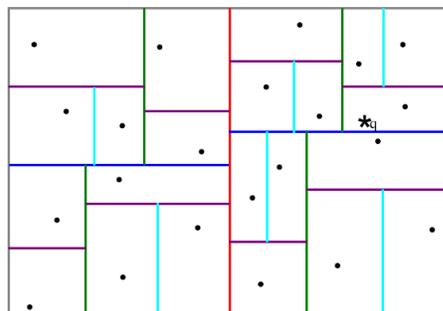

Fig. 3.   k-d tree structure

III.   PREVIOUS WORK

To implement spatial applications efficiently requires the use of spatial data structures, which used to store data objects that linked with location and an important class of data structures used in computer graphics, geographic information systems, and many other fields.

To improve performance of K-d tree and Quad tree have different shapes. For example, the building of the data structure which represent in the using of mathematical mean by using median of data [9]. Another shape of improving performance, in this paper the author use the same data structures K-d tree and Quad tree addition to Tile arrays and by ignoring the unnecessary objects, the time of retrieving data is decreased as in [10]

IV.   METHODOLOGY

Fig 4 illustrates the GUI of The application .It builds to display the performance of different types of GIS data structures, quad tree, and K-d tree. The application has the ability to search, add, delete, and update the points. In addition to, it has to division the map to different scale from 1 to 9 and prints the name of site beside the location on the tree. The application contains the ability to find the nearest neighbor from any record exits in DB. All the process occurs on the cloud computing on the internet. The database structures storage on the "SOMEE.com." The application can play online and off line.

The database of application uploaded to the cloud computing which represented by SOMEE.COM as a hosting DB. The application loaded the database in the first running of application Fig 5 shows the time of searching in database.
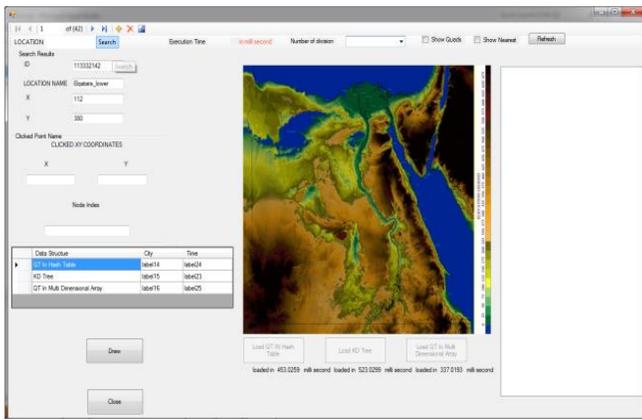
Fig. 4.   GUI of the project

Hash table, and a Multidimensional array storing the data in data structure, the application illustrates the time response in Fig 5 represents the search time in DB.

Fig. 5.   Search time in DB

Fig 6 illustrates the division in (4) degree. Moreover, the red point is the result of search for the point under search and its location in the tree.
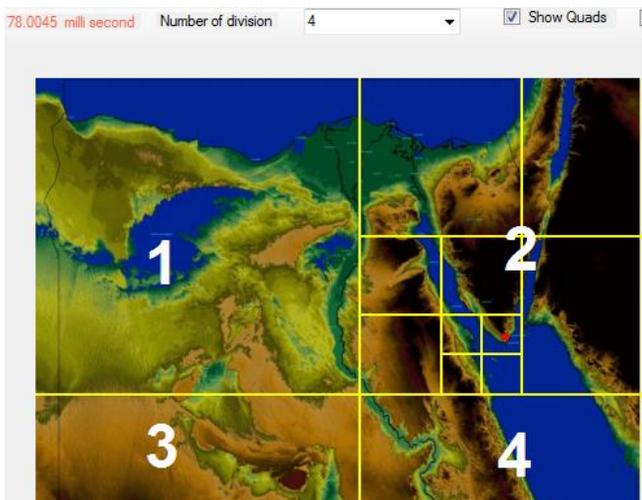

Fig. 6.   Snap shoot for the result on map

We have one database but with different data structure.Actually, we use hash tables and multidimensional array.

A hash function is any function that used to map data of arbitrary size to data of fixed size. The values returned by a hash function called hash values, hash codes, hash sums, or simply hashes. One use is a data structure called a hash table, widely used in computer software for rapid data lookup. Hash functions accelerate table or database lookup by detecting duplicated records in a large file.
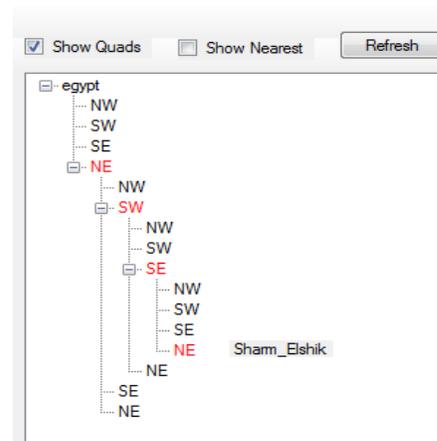

Fig. 7.   locations in the tree

Hash functions used in hash tables, to find out a data record (a dictionary definition) given its search key (the headword). Specifically, the hash function used to plane the search key to an index. The index gives the address in the hash table where the suitable record should be stored. Hash tables, sequenced, used to implement associative arrays and dynamic sets. [23]

Typically, the domain of a hash function (the set of possible keys) is larger than its range (the number of different table, indexes), and so it will plan several different keys to the same index. Therefore, each slot of a hash table is associated with (implicitly or explicitly) a multi of records, rather than a single record. For this reason, each slot of a hash table often called a "bucket," and hash values called "bucket indices." The hash function only hints at the record's location — it tells where we can start looking for it. Still, in a half-full table, a good hash function will typically decrease the search down to only one or two entries. Hash table used in many applications like an approximate nearest neighbor.

Searching in large databases has become popular owing to its computational and memory efficiency. The famous hashing methods, e.g., *Locality Sensitive Hashing (LSH) and Spectral Hashing (SH),* construct hash functions based on random or principal projections. [4]. The *complementary hashing approach,* is an approach used hash table, which is able to balance the precision and recall in a more effective way. The key idea is to employ multiple complementary hash tables, which are learned sequentially in a boosting manner, so that, given a query, its true nearest neighbors missed from the active bucket of one hash table are more likely to be found in the active bucket of the next hash table. [5]

Now we look for another type of data structure "*multidimensional array.*" It is recognized in the past that, traditional database Management Systems (DBMSs) does not handle   efficiently   multi-dimensional   data   (which   are geometrical shapes in our search) such as squares, Polygons, or even points in a multi-dimensional space. Multidimensional data arise in many applications, such as the most important fields:

1) Cartography, Maps could be stored and searched electronically answering efficiently geometric queries.

2) Computer-aided Design (CAD).

3) Computer vision and robotics.

4) Rule indexing in expert database systems. [6].

On some applications the focus on increase the memory storage and reduce the conflict of data access to use one dimensional array, but by use the automatic partitioning memory scheme for multidimensional arrays based on linear transformation to provide high data throughput of memory storage the experimental illustrate that saving in memory banks and digital signal processing (DSP). [7]

The application starts as in Fig 8.

Fig. 8. Start of the project

We have 42 records which are the cities and famous place in Egypt, ID represent the index of location in DB Location name, x and y the coordinate of the location.

Fig. 9. coordinates of point form

Fig 8 gives the coordinate of the point when clicked on the map directly and also the index number of the location.

V. RESULTS

First, we select the name of city from location name and put it in the location in the first label and press "search" which bring the name of the location and display it on the map Fig 9 show the choosing process of "Kafr_Elshik" city from a database.

The name of a city under the search "Kafr_Elshik" Fig 11 and Fig 12 shown the result of choosing process.

Fig 13 shows the time response for the different type of database structures.

Fig. 10. information of "Kafr_Elshik" point

Fig. 11. The location on the map

Fig. 12. The location in the tree

Fig. 13. time response for data structures

TABLE I.   TIME RESPONSE

X	1	2	3	4	5
QT HT (mic.s)	979	1232	982	1334	1111
K-DT (mic.s)	999	4484	4481	880	4566
QTMD (mic.s)	724	897	724	976	1085

The curves illustrate the numbers of search (X-axis) with the time in (microsecond). The curves represent that the biggest time the system taken is in the data structure of K-d tree(K-DT)which green one and next it the Quad tree(QT HT) with data structure in hash table represented by red one and the smallest one is the Quad tree in multidimensional array(QTMD) in blue.

VI.   CONCLUSION

Our goal is studying the performance of different kind of DB structure for GIS which storage in Cloud computing. The data type of GIS is huge and need to store in the data structure with a method to provide the performance goal:-
-   Min. Mass storage.
-   Min. Time search.

So, during store data map we use two data structures:-

-   K-d tree to store points.
-   Quad tree to store regions.

We find that quad tree is more useful and can store large regions with small data as shown in Fig 14.

Fig. 14. the map in quad tree division

The gauge of any query is the query time. We use 3 data structures to search in database Hash table (has a constant time "hash function") for regions, K-d tree (proportional to the length of the tree) query for XY points, and Multidimensional array (has constant time) to search for regions. As we say before we will treat with regions so we compare between hash table and multidimensional array. We find a multi dimensional array is the fastest one. The disadvantage of multidimensional array large memory size, but this memory is local on the computer not in the server. Maybe not as effective if the amount of divisions increase. The result illustrates that the data type of GIS should store in a "quad tree in multidimensional array" which give better performance than the two other types, K-d tree and Quad tree in the hash function.

REFERENCES

[1]   Muzafar Ahmad Bhat,RazeefMohd Shah , Bashir Ahmad ."Cloud Computing: A solution to Geographical Information Systems (GIS)"ISSN: 0975-3397Vol. 3 No. 2 Feb 2011

[2]   Lakshmi Balasubramanian, M. Sugumaran"A State-of-Art in R-Tree Variants for Spatial Indexing", International Journal of Computer Applications;Mar2012, Vol. 42, p35

[3]   YannisTheodoridis, TimosSellis" Optimization Issues in R-tree Construction",http://link.springer.com/book/10.1007/3-540-58795-0http://link.springer.com/bookseries/558Volume 884, pp 270-273, 1994

[4]   Jun Wang, Sanjiv Kumar "Semi-Supervised Hashing for Large Scale Search", Ieee Transactions On Pattern Analysis and Machine Intelligence Vol, 34, No 12, December 2012

[5]   HaoXu , JingdongWang, Zhu Li Gang Zeng, Shipeng Li Nenghai Yu "Complementary Hashing for Approximate Nearest Neighbor Search",ICCV '11 Proceedings of the 2011 International Conference on Computer Vision Pages 1631-1638

[6]   TimosSellis, Nick Roussopoulos, Christos Faloutsos"The R Tree: A Dynamic Index for Multi- Dimensional Objects", In The VLDB Journal (1987)

[7]   Yuixnwang, Pengli,Peng Zhang,Jasf Con Cong " Memory partitioning for Multidimensional Arrays in High-level Synpaper". May 29 - June 07 2013, Austin, TX, USA. Copyright 2013 ACM

[8]   P.S.V.S. Sridhar, T. N. Jowhar, AnantBhaskarGarg, U. Kedareswarudu"A Framework of Information Technology for Water Resources Management" International Journal of Computer Applications (0975 – 8887) Vol. 30– No.5, September. 2011

[9]   Logamani. K, Punitha. S. C "Density Based Clustering using Enhanced K-d tree" Logamani. K, et al | International Journal of Computer Science, Engineering and Technology (IJCSET) | November 2014 | Vol 4, Issue 11,314-318

[10]   TilmannZäschke, ChristophZimmerli, Moira C. Norrie "The PH-Tree – A Space-Efficient Storage Structure and Multi-Dimensional Index" SIGMOD'14, June 22–27, 2014, Snowbird, UT, USA. Copyright is held by the owner/author (s). Publication rights licensed to ACM. ACM 978-1-4503-2376-5/14/06

[11]   "Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions". Kenneth E. Foote and Margaret Lynch, The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. Retrieved 21 Apr 2015.

[12]   "GIS Applications in Public Health as a Decision Making Support System and It's Limitation in Iran" M.S. Mesgari, K. N. Toosi World Applied Sciences Journal 3 (Supple 1): 73-77, 2008

[13]   "A Framework of Information Technology for Water Resources Management" P.S.V.S. Sridhar, T. N. Jowhar, Anant Bhaskar Garg, U. Kedareswarudu, International Journal of Computer Applications (0975 – 8887) Volume 30– No.5, September 2011

[14]   "Application of GIS Technology in Watershed-based management and Decision Making" U. Sunday Tim, Sumant Mallavaram, AWRA Hydrology & Watershed Management Technical Committee Vol. 1, No. 5, July - August 2003

[15]   "GIS as a Tool for Decision Making and Evaluation" Delforce, Wayne and Rich, Bruce and Savage, Bernadette and Adkins, Barbara and Hall, Carly and Caldeweyher, Daniel (2005) GIS as a Tool for Decision Making and Evaluation. In Bailey, C. and Barnett, K., Eds. Proceedings Social Change in the 21st Century, Brisbane, Australia.

[16]   "A framework for ranking of cloud computing services" Saurabh Kumar Garg, Steve Versteeg, bRajkumar Buyyaa Cloud Computing and Distributed Systems Laboratory, Department of Computing and Information Systems, University of Melbourne, Australia, June 2012

[17]   " Image Rotation Using Quad Tree" Journal of Engineering, Computers & Applied Sciences (JEC&AS) Volume 1, No.1, October 2012

[18]   "Using a distributed quadtree index in peer-to-peer networks" Egemen Tanin • Aaron Harwood • Hanan Samet, Egemen Tanin • Aaron Harwood • Hanan SametPublished online: 4 April 2006 c Springer-Verlag 2005

[19]   "AS B-tree: A Study of an Efficient B+-tree for SSDs", Hongchan Roh, Sungho Kim, Daewook Lee and Sanghyun Park, journal of information science and engineering 30, 85-106 ,2014

[20]   "A Tool based Edge Server Selection Technique using Spatial Data Structure" Global Journal of Computer Science and Technology, Vol. 14 Issue 3 Ver. 1.0, 2014

[21]   "Scalable Nearest Neighbor Algorithms for High Dimensional Data" Marius Muja, David G. Lowe, ieee transactions on pattern analysis and machine intelligence, vol. 36, no. 11, 2014

[22]   "filtering of point clouds from photogrammetric surface reconstruction" K. Wenzel, M. Rothermel, D. Fritsch, N. Hale, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5, 2014 ISPRS Technical Commission V Symposium, 23 – 25 June 2014, Riva del Garda, Italy

[23]   https://books.google.com.eg/books?id=k6Yfuxaqp2QC&printsec=frontcover