I J C T A, 9(26) 2016, pp. 233-239

A Review on Challenges of Storage and Retrieval for Spatial Applications of GIS and Cloud Computing

Karthi S[1] and Prabu S[2]

ABSTRACT

Geographic Information System (GIS) is an enormous area which helps people to present nature changes, discover the law of social and economic development with analysis of massive spatial and attributed data of the earth surface.To provide connectivity and to handle them effectively, GIS deals with varied types of spatial data: attribute data, geometric data, relational data that expresses the mutual relationship of map elements, auxiliary data which facilitates the map processing.In the real time applications, the data changes relatively with time and to handle the huge volume of data with time are a challenging task.As the management and maintenance of spatial data storage costs a lot of hardware, human and financial resources which are a major limitation, researchers have to focus towards to handle them efficiently.As a whole, Cloud computing which is a new internet-based supercomputing principle and practice, which supplies dynamic, scalable, and pay-per-use services and has huge computing and storage capacity with high reliability and security has been advised to proceed to handle data in GIS. Of course, the cloud storage, which assembles different kinds of storage devices to work together by software with functionality of clustering, grid or distributed file systems, is a complex system. This is composed of various parts such as network device, storage equipment's, servers, applications, public access interface, the access network, and client programs. This paper first summarizethe concept, features of GIS and the development of the cloud computing. This paper specifies an approach where a triple security is provided in cloud computing and hence ensuring confidentiality, availability and integrity to the data. This is embedded with the application of cloud computing techniques in spatial analysis, decision support and the storage of spatial data, and the impact on software development and product patterns of GIS. Finally, the paper has proposed a model to handle effectively data storage with time relatively which deals with GIS and Cloud Computing with open source tools.

Keywords: Geographic Information System (GIS), Cloud Computing,Triple Security Confidentiality, Availability, Integrity.

1. INTRODUCTION

GIS is having huge growth in all kinds of industries. It enables users to visualize, analyse, issue, store, process and interpret spatial data in order to understand the relationships, patterns, and trends. GIS gives an electronic representation about the Earth's natural and man-made features. This deals with the real-world spatial data elements so that it can coordinate system through high computing resources. Large size of spatial repositories and the geospatial models is involved thus increasingthe time complexity and high computing resources.

Cloud computing is an Internet based computing providing servers, storage applications and resources to many organizations and enterprises. Security in cloud computing is very important and it has been handled properly and effectively. The users do not have any idea where their data is placed. Thus, triple Security in Cloud Computing is made by using related algorithms sothat availability, confidentiality, integrity to the data is provided.

[1] Research Scholar, VIT University, Vellore, TN, India. *E-mail: s.karthi2014@vit.ac.in*

[2] Associate Professor, VIT University, Vellore, TN, India. *E-mail: sprabu@vit.ac.in*

GIS performs controlling and analytical operations on the data to produce reports, graphics, statistics and to control geographic data processing workflows. At the same time Cloud Computing is a parallel and distributed system that consists of a collection of interconnected and virtualized computers that are dynamically provisioned and presented based on the service-level agreements. GIS Cloud infrastructure providers include Amazon, Microsoft and IBM which provides reliability and security towards cloud technology to the end-users. GIS cloud is a service where huge volume of data can be stored and accessed securely over the Internet through computer's hard drive.

2. RELATED WORK

The author [1] describes GIS and Cloud Computing will provide full featured GIS available over the web in any browserwith powerful data processing capability. These web browsers provide tools that work with maps includingspatial analysis. Clients do not need to install any updates delivered over the network, with cloud GIS users share singlegeographical data storage that is updated centrally. With this kind of usage, update propagations can said to be usertransparent. All users and the editors of GIS have consistent spatial data while sufficient implementations aremade.Moreover, cloud provides low user training costs for non-expert GIS users.

The author [2] describes GIS functions and services are geographically and logically distributed due to the source of data, location of computing facilities and organizations. The spatial analysis on this large amount of data is complex and it is computationally intensive. In order to share and collaborate GIS data and the computational results among geographically dispersed users, a scalable and low cost computation platform, such as cloud computing- is required for GIS applications. For this purpose, a new technology called GeoCloud has emerged.

The author [3] describes GIS is beingimplemented not only as a standalone system butintegrated into other systems. It is a collection of tools that helps in capturing, storing, analyzing,managing and utilizing large volumes of geospatialdata and delivers thegeodata and services for hugeconcurrent users. Thus, one of the requirementsis the ability to handle the huge volume of spatialdata and ensure the required performance is provided withoperational flexibility.

The author [4] describes that proposed new architecture for spatial data processing based on cloud computing technology. In particular, the problems of large volume of spatial data in some of the disaster cases which make the needs for stretch way for storing and analysis and computing all of these resources. Proposed architecture for spatial data processing based on cloud computing technology then results showed better presentation in assessment to previous works. To enhancing the proposed model to support raster data of GIS system also using full use of computing services in the cloud computing infrastructure to swarm large volume of data as possible using Amazon storage services such as Amazon S3.

The paper [5] deals with the cloud computing which is the most admired information technology for data processing. This enables users to process large amount of information without having their own computing power and this also apt for geospatial data processing. To examine how the cloud computing can improve geospatial data processing with comparing its advantages and disadvantages and discussing cloud computing features.

The author [6] deals with the CC which is the internet based technique provide various resources. The accessibility of these resources is very flexible in nature and few are obtainable to customers free of cost but some are on a pay-as-use basis. Then the customers also allowed to access information and utilize the computer resources from anywhere having internet access. This cloud computing technology uses some security technology like SHA 512, AES, and stenography to overcome the various security issues (stealing, hacking, unauthorized access etc

This author [7] says about the Orfeo Toolbox (OTB) which is the tool for the operational development of the future sub-metric optic and radar image. The OTB is useful for all public working in the Remote sensing imagery community. Using a open source license, CNES hopes to gain from charity of many specialists to develop the particular use of satellite imagery.

The paper [8] deals with the most important feature of cloud computing is capability in powerful computing and dynamic storage. Also they discussed that cloud computing is intended for data infrastructure which in turn be able to develop the relation for spatial data infrastructure in the upcoming.

The author [9] told Geospatial data is an essential element in decision making processes and planning efforts across a variety of industries and information sectors. This quantity and variety of data is rapidly increasing and where as more of this data is risk oriented of being lost or becoming unusable. There is a increasing recognition of the significance of being able to access historical geospatial data.

The author [10] describes about the database capable of dealing with image together with other spatial and non-spatial data types. To describe a solution for efficient handling of large image data sets in a standard object-relational database management system. By means of sufficient indexing, density and retrieval techniques, acceptable performances can be achieved using a standard DBMS, even for very large satellite images. The part of the development which aims to provide a complete environment for the development of GIS applications.

The author [11] said geo information agencies are managed, maintained and build the geospatial information platform. They combine the various geo-spatial data to give data analysis services for supporting government decision making. This big data is difficult to address the data and computing intensive issues by established platform. This uses HDFS for managing image data and MapReduce-based computing service. Especially for IaaS layer, need to improve the efficiency of DC2's resource schedule, and optimize workload balancing and auto-scaling algorithm, as well as increase platform stability and reliability.

The paper [12] says about Hadoop raised area for image processing rather than for its unique purpose of text processing. It has never been proved that Hadoop can be efficiently utilized for high-volume image files. The purpose of Hadoop for image processing has been researched using eight different practical image processing algorithms. It expand the file loom in Hadoop to deem the entire TIFF image files as a unit by increasing the file format that Hadoop uses. This technique is scalable and resourceful in processing various large images; used frequently for remote sensing applications.

The author [13] described about evaluation of multi-datacenter Hadoop deployment with single-datacenter Hadoop deployment to identify the performance issues that inherent in a geographically distributed cloud.

The author [14] discuss about the Hadoop which is a java based programming framework that supports the storage and process of large data sets in a distributed computing environment. It is suitable for high volume of data. It is using with HDFS for data storage and MapReduce to process the data. The main aim of Mapreduce programming model is to parallelize the job implementation across multiple nodes for execution. All center of the researchers and companies toward to Hadoop. due this, many scheduling algorithms have been proposed in the past decades.

This paper [15] deals with the MapReducea system that gives suitable due to their superior scalability, fault tolerance, and flexibility to handle unstructured data. It discovers the viability of building a hybrid system that takes the best features from both technologies. Hadoop and Hive are relatively young open-source projects.

The author [16] discuss about Hadoop data where users face with sensible advice on how to protect these environments. This method of data storage is increasing extremely in lots of folds. There has been increasing data security and isolation concerns for people who outsource data on this Hadoop clusters. The

spatial data analysis of big datasets using distributed method. Experimented the cloud computing technology in spatial fields over single spatial databases and proved the performance and efficiency of operations on spatial data in Hadoop environment

The paper [17] discusses about Apaches Hadoop- HDFS which is used to store the streaming data that is too big in size most organization. Using Hadoop Map Reduce for computing and HDFS for storage. This Hadoop technique is most trendy for analysis, storage and to process very large data which does not require lots of changes in hadoop system.

The author [18] describes the cloud data base that is based on hadoop technologies that is hadoop distributed file system (HDFS). Data is simulated in different data nodes which can be accessed by name of the node using logs that are present in them. They use Mapreduce method to process the data on cloud with various types of systematic to perform using map reduce codes. As storage mechanism on cloud to the sender subscriber to a cloud DaasHadoop enables surplus data to be streamlined for any distributed processing system across clusters of computers using simple programming models.

The author [19] discuss about the development of a scalable spatial data supervision system and their various spatial queries with MapReduce which cannot be supported using MapReduce without difficulty. Then they focus on problems of difficult spatial applications and current methods correspondingly. It implements two distinctive spatial applications, all nearest neighbor and astronomical cross-match which face the same complex problem where distance computing is essential. MapReduce is a key-value based programming model and an associated implementation for processing large data sets. It has been adopted in various scenarios and seems promising. However, when spatial computation is expressed straightforward by this key-value based model, difficulties arise due to unfit features and performance degradation.

The author [20] says that the Hadoop Distributed File System (HDFS) is used to store the huge amount of data set consistently to stream the data sets at high bandwidth to user applications. In this type of cluster, thousands of servers mutually deals with storage and implement user application tasks. Through distribution of storage and computation across many servers, the resource can develop at an economical growth at each size. In a large cluster, thousands of servers both host directly attached storage and execute user application tasks. By distributing storage and computation across many servers, the resource can grow with demand while remaining economical at every size.

The paper [21] deals that the MapReduce framework and Hadoop are temporarily introduced in this segment. It describes about the gradual execution process of this programming model, as well as its core modules. This also analyzes the distributed file system of Hadoop (HDFS) and its specialized nodes. The outlined the importance of spatial data, since they are used in almost every research field that needs to depict multidimensional data. They illustrated how they can be efficiently processed in a parallel manner. Furthermore, due to ever-growing data sets, the need arises for technological advancements towards information management. It is clearly understandable that parallel computing is more beneficial over sequential methods. As a result, the way towards parallelization – MapReduce and its implementation, Hadoop – will be the solution to a variety of difficult computational problems

The author [22] describes about the Spatial Hadoop is a comprehensive expansion to Hadoop that injects spatial data awareness in each Hadoop layer, namely, the language, storage, MapReduce, and operations layers. This method used to improve the cloud architecture which is hybrid for object storage based for geospatial data processing. Due to increase in retrieval of the geo-spatial data by various fields, the necessity for delivery of the geospatial process is also increased.

The author [23] says that Encryption techniques are the most secured method to transfer the perceptive information from sender to the intended receiver. The main Objective of this method is to create a perceptive information illegible to all other except the receiver. The secure transmission of information AES encryption has been used which provides most secure way to transfer the sensitive information from sender to the intended receiver. The main purpose of using this technique is to make sensitive information unreadable to all other except the receiver. The data thus compressed enables utilization of storage space in cloud environment. It has been augmented with Hadoop's map-reduce paradigm which works in a parallel mode. The experimental results clearly reflect the effectiveness of the methodology to improve the security of data in cloud environment.

The paper [24] deals with the cloud computing technology which is the tremendous next generation information technology. This technology is used to access remote data and provide a large amount of storage space .It provides a secured cloud storage system and also maintains privacy between Third party auditor and Cloud service providers. In cloud computing the Third Party Auditor guarantee that the cloud service provider & also itself TPA would not learn any knowledge about the data that is stored on the cloud server. During the efficient auditing process, it not only eliminates the burden of cloud user from the tedious and possibly expensive auditing task. The partitioning of data will enables storing of the data in easy and effective manner. It also gives way for flexible access and there is less cost in data storage

3. COMPARISON OF SPATIAL DATABASE AND SPATIAL CLOUD DATABASE

Table 1
Shows the comparison of spatial database and spatial cloud database using Hadoop

S. No.		*Spatial DB*	*Spatial Cloud DB using hadoop*
1.	User access	Limited	Unlimited
2.	Security	Normal security	High security
3.	Storage capacity	Limited	Unlimited
4.	Auto scaling	No	Yes
5.	Normalization	No	Yes

The above table is deals with the comparison between spatial database and spatial cloud database using hadoop. Cloud database main advantage is to access easily to data anywhere and anytime. Security is the one of an important issue in the database, but cloud database concentrate on security issues also. It has high security using efficient and innovative techniques and methodologies. Cloud database includes auto scaling method which is used for data storage capacity as a specification towards the proposed methodology. Hadoop mainly concentrates on data storage ina normalized form for further process of varied applications.

As a mention, our proposed architecture has been given in [25] An Investigation on Hybrid Computing for Competent Data Storage and Secure Access for Geo-Spatial Applications.Which discuss about the hyperspectral images wherein the main sources for these images are acquired by remote sensing satellites. In turn, this will become as an input to the proposed architecture. The processing includes three steps as initially, the input images are processed by using Orfeo toolbox to create Geo-Spatial Database.Secondly, the geo spatial database is normalized which are reduced using Hadoop to store it into the cloud. Thirdly, the Map Reduced Geospatial data are processed for various applications in the cloud environment. A simple authentication mechanism is used for providing access to the data which are stored in cloud. Based on the application and query given by the client the output will be given in either raster or in vector form efficiently.

4. CONCLUSION

The paper is proposing a hybrid model by studying analysing other existing stenography and cryptographic algorithms. Many tools have been studied related to the detection of hidden information over the internet in order to secure the data from steganalyst. Also, Hadoop is used to minimize the storage area which is mapped to GIS database in a cloud using normalization techniques.Thereby the proposed model may result with delivery of information without any loss of data and with minimum time consumption. Finally the integration of GIS and cloud gives efficient data retrieval and other E-Commerce services that reduce the manual work by outsourcing to the cloud GIS Effective data balancing is made by applying load balancing strategy.

REFERENCES

[1] Aly, Ahmed Gamal, and NevineMakramLabib., "Proposed model of gis-based cloud computing architecture for emergency system",*International Journal of Computer Science*, **1(4)**, 17-28, 2013.

[2] Nabil EL KADHI and Walid (oualid) Ben ALI., "Trust based security framework model for geocloud",*International Journal of Power Control Signal and Computation*, **5(1)**, 56-62,2013.

[3] Eman Mahmoud, Osman Hegazy, and Mohamed Nour El-Dien., "Integration of GIS and Cloud Computing for Emergency System", *International Journal of Engineering and Computer Science*, **2(10)**, 2889-2893, 2013.

[4] Yaser Khalilizangelani and Saman Ghaffarian., "A Study of Geospatial Data Processing Based on Cloud Computing", *Proceedings of ACRS*, 228-231, 2013.

[5] Ms. Manisha, Dr. Kamleshsahrma and RidhikaSharma., "A Technique To Increase Integrity Of Cloud Data Using Hybridalgorithms",*International journal of engineering sciences & research Technology*, **4(7)**, 369-373,2015.

[6] Emmanuel Christophe, Jordi Inglada and Alain Giros., "AOrfeo toolbox: A complete solution for mapping from high resolution satellite images", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1263-1268, 2008.

[7] Naghavi Mohammad., "Cloud computing as an innovation in GIS & SDI: methodologies, services, issues and deployment techniques", *Journal of Geographic Information System*, **4**, 597-607, 2012.

[8] Guy McGarv Steve Morris and Greg Janée., "Technology Watch Report: Preserving Geospatial Data", *Digital Preservation Coalition*, 2009.

[9] Lubia vinhas, Ricardo cartaxo modesto de souza and Gilberto camara., "Image Data Handling in Spatial Databases", *In GeoInfo*, 2003.

[10] W.W. Song, B.X. Jin, S.H. Li, X.Y. Wei, D Li and F. Hu., "Building Spatiotemporal Cloud Platform for Supporting GIS Application. ISPRS Annals of Photogrammetry", *Remote Sensing and Spatial Information Sciences*, 55-62, 2015.

[11] Mohamed H. Almeer., "Hadoopmapreduce for remote sensing image analysis", *International Journal of Emerging Technology and Advanced Engineering*, **2(4)**, 443-451, 2012.

[12] Zhang Q, Liu L, Lee K, Zhou Y, Singh A, Mandagere N and Alatorre G., "Improving Hadoop Service Provisioning in A Geographically Distributed Cloud", *IEEE 7th International Conference on cloud computing*, 432-439, 2014.

[13] Seyed Reza Pakize., "A Comprehensive View of HadoopMapReduce Scheduling Algorithms", *International Journal of Computer Networks and Communications Security*, 308-319,2014

[14] Azza Abouzeid,Kamil Bajda-Pawlikowski, Daniel Abadi, Avi Silberschatz and Alexander Rasin., "Hadoop DB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads", *Proceedings of the VLDB Endowment*,2(1), 922-933, 2009.

[15] Rajesh Laxman Gaikwad, Prof. Dhananjay M Dakhane and Prof. Ravindra L Pardhi., "Network Security Enhancement in Hadoop Clusters", *International Journal of Application or Innovation in Engineering And Management*, **2(3)**, 151-157, 2013.

[16] Sayali Ashok Shivarkar., "Speed-up Extension to HadoopSystem", *International Journal of Engineering Trends and Technology*, **2(12)**, 105-108, 2014

[17] N. Brahmanaidu, Shaik Riaz., "Distributed Data Storage and Retrieval on Cloud by using Hadoop", *International Journal of Science and Research*, **3(6)**, 2768-2769, 2012.

[18] Wang K, Han J, Tu B, Dai J, Zhou, W and Song, X., "Accelerating spatial data processing with mapreduce", *16th International Conference on Parallel and Distributed Systems*, 229-236, 2010.

[19] Shvachko K, Kuang H, Radia S and Chansler R., "The hadoop distributed file system", *In Mass Storage Systems and Technologies,IEEE 26th Symposium on*, 1-10, 2010.

[20] Georgios Economides, Georgios Piskas and Sokratis Siozos-Drosos., "Spatial Data and Hadoop Utilization", 2013.

[21] Ahmed Eldawy and Mohamed F. Mokbel., "SpatialHadoop: A MapReduce framework for spatial data", *In Proceedings of the IEEE International Conference on Data Engineering*, 2015.

[22] Mehak and Gagandeep., "Improving Data Storage Security in Cloud using Hadoop" *International Journal of Engineering Research and Applications*, **4(9)**, 133-138, 2014

[23] Yogesh V Bhapkar, Rakesh S Gaikwad and Milind R. HegadeYogesh., "Providing Security And Privacy To Cloud Data Storag", *International Journal of Computer Science and Information Technologies*, **6(2)**, 969-971, 2015

[24] Kumar M, Gupta A, Shah K, Saurabh A, Saxena P and Tiwari VK., "Data Security Using Stegnography and Quantum Cryptography", *Network and Complex Systems*, **2(2)**, 46-55, 2012.

[25] KarthiSankar, PrabuSevugan., "An investigation on hybrid computing for competent data storage and secure access for geo-spatial applications", *Institute of Integrative Omics and Applied Biotechnology*, **7(5)**, 139-149, 2016.