ISSN: 0976-3104
SPECIAL ISSUE (SCMDSA)
Sankar and Sevugan

ARTICLE OPEN ACCESS

AN INVESTIGATION ON HYBRID COMPUTING FOR COMPETENT DATA STORAGE AND SECURE ACCESS FOR GEO-SPATIAL APPLICATIONS

Karthi Sankar* and Prabu Sevugan

School of Computing Science and Engineering, VIT University Vellore, Tamil Nadu, INDIA

ABSTRACT

The GIS and cloud computing which is a hybrid method, a proposed technique is made on par with existing stenography and cryptographic algorithms. In modern years different steganography methods has been used to construct data in a more secured way along with special steganalysis methods. There are various tools that may be used for detecting hidden information which is easily available over the internet in order to secure the data from steganalyst. This is considered to be a gap from the review of the paper. Internet based computing provides security to cloud users as encrypted data in the cloud that protects the data from many attacks. Cryptography using GIS is collective geographic information for geographic or locational component purposes. It discharges Infrastructure as Service (IaaS), Software as a Service (SaaS) and Platform as a Service (PaaS) to computers. The existing papers meet the services individually, but as a whole, this will be done by GIS and CC which is a proposed hybrid model. Many open source tools will be used for hybrid model which is secured for cryptography. Also MapReduce is used to minimize the storage area which is mapped to GIS database in a cloud. The author proposed hybrid model to deal with information in a secured way that will store and hide the information. This happens along with some normalization techniques. Finally, the proposed model yields results in an efficient way of delivery of information without any loss of data and with minimum time through effective load balancing

Received on: 30th-Nov-2015
Revised on: 11th-March-2016
Accepted on: 26– March-2016
Published on: 20th–May-2016

KEY WORDS

Steganalysis, Geo-Spatial, Cloud Computing, GeoDB, GeoCloud

***Corresponding author: Email:** cskarthi11@gmail.com **Tel:** +91 9943369964

INTRODUCTION

Nowadays, Geographic Information System (GIS) is having enormous growth in all kinds of industries. It enables users to visualize, issue, analyze, process store and interpret spatial data to understand relationships, patterns, and trends. A GIS provides an electronic representation about the Earth's natural and man-made features which in turn deal with real-world spatial data elements to a coordinate system through high computing resources. It involves large size of spatial repositories and the complicacy of the geospatial models with an increase in time complexity and high computing resources.

Cloud Computing is evolving as a key computing platform for sharing resources that include infrastructures, software, applications, and business processes. Cloud computing is an Internet based computing which provide servers, storage applications and resources to many organizations. Cloud Computing is a model to enable expedient, on-demand network access to a shared pool of configurable computing resources which can be rapidly provisioned and released with minimal management effort or service provider interaction. Security in cloud computing is very necessary so that it would be more effective and useful. The users do not have any idea where their data is placed. Hence, triple Security in Cloud Computing is made using related algorithms to provide availability, confidentiality, integrity to the data.

Cloud computing technology includes Hadoop platform and MapReduce parallel computing model into the domain of geographical information system to solve issues like spatial data storage, spatial data index and spatial operation in various applications of GIS. MapReduce is applicable for computing-intensive spatial applications and the lead in scalability and data storage efficiency is done by Hadoop. Thus large-scale distributed data management provides an efficient way for big data storage based on cloud computing technology.

THE IIOAB JOURNAL
ISSN: 0976-3104

GIS is an Integrated System of Computer Hardware, Software and Spatial Data that performs controlling and analytical operations on this data to produce reports, graphics, statistics and controls geographic data processing workflows. In turn, Cloud Computing is a parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements. Geographic Information Systems applications have been moving into the cloud with improved drive, Global organizations like ESRI, GIS Cloud Ltd etc have already taken the quantum leap and taken a technical shift to Cloud Computing Paradigm. GIS Cloud infrastructure providers are Amazon, Microsoft and IBM which provides reliability and security towards cloud technology to the end-users. GIS cloud is a service where large datacan be stored and accessed securely over the Internet through computer's hard drive

RELATED WORK

The author [1] describes that GIS and spatial data techniques lead to distributed data method and computing environment. This made the cloud computing technology to focus on extensive resource sharing and low cost for large data storage technology. They 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 [2] 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 cloud computing technique's growth in the field of geospatial data seem to become a hot spot and has to be more extensive in the further tasks as the valuable consequences produced and seen from various organizations.
The author [3] 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.). The author planned SHA-512 is used for verify integrity of data. For encryption, use AES and then apply Steganography which can change presentation of data so that unauthorized person could not access the hidden data. Hence this is essential for providing maximum security to the data. Receiver can get original plain text by reversing the, SHA-512, AES, steganography. Outlook implemented SHA-512, AES and STEGANOGRAPHY to provide maximum security in cloud computing. By implementing these three algorithms we are intended to provide maximum security to the data. The technique used also increased the time complexity which should be reduced and hence we will try to improve the same

The author [4] said the enlargement of spatial technology and the spatial data have been collected through different approaches. Raising the anxiety is being put forward on use of spatial information in emergence system. This GIS technique and spatial information have lead to more distributed process and computing environment. Then the cloud computing technology is focused on extensive resource sharing and low cost for large data storage technology. Cloud based service to compute the large spatial data technique in emergence management and provides efficient spatial service to the spatial platform. Cloud means infrastructures that provide resources and services over the Internet. Frequently the services are layered to create a stack of cloud services that serves as a platform for developing cloud-based appearance management applications. Face new challenges as try to manage our spatial data that might be stored in a variety of devices.

This author [5] 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 IIOAB JOURNAL
ISSN: 0976-3104

The paper [6] deals with the cloud computing technology which is an important issue of scientific community to the spatial data. It refers the computing infrastructure for an illustration of network. Then the main individuality of cloud computing is being dynamic, high power in computing and storage. It is cost effective for web based spatial data and for difficult analysis. This cloud technology is an easiest way of distributed computing that handles varied data. One of 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 [7] 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 [8] describes about the enlargement of new generation of spatial databases, where the DBMS is able to manage spatial and non-spatial data types together. This spatial database deal with vector geometries but it has limited facilities for managing image data. It is important to make 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 of which aims to provide a complete environment for the development of GIS applications.

The author [9] 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. Further optimization of platform architecture. 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. For PaaS layer, it will expand GIS service functions for service chaining with thematic application services, such as land use and planning, government emergency response, and geographical condition analysis, etc. On the other hand, need to consider the unified framework of spatial data cloud storage, which implements integrated management based on the spatial data access interface for spatiotemporal data, such as managing spatial vector data and integrating HDFS with geodatabase.

The paper [10] 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. This creates variation between the single PC runtime and the Hadoop runtime which is evidently noticeable.

The author [11] 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. A simplification of the problem description in the context of geographically distributed cloud datacenters is also provided with deliberations on general optimization strategies. It describes about the design and realization of a suite of system-level optimizations for improving performance of Hadoop service provisioning in a geo-distributed cloud. This also deals with the prediction-based job localization that configures HDFS data placement and data perfecting.

The author [12] 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. 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. There are three important scheduling issues in MapReduce such as locality, harmonization and equality. The most common objective of scheduling algorithms is to minimize the completion

THE IIOAB JOURNAL
ISSN: 0976-3104

time of a parallel application and also achieve to these issues. In describe the overview of Hadoop MapReduce and their scheduling issues and problems. then, it have studies of most popular scheduling algorithms in this field. finally, highlighting the implementation Idea, advantages and disadvantage of these algorithms.

This paper [13] deals with the MapReduce systems 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. HadoopDB will automatically benefit from these improvements. HadoopDB is therefore a hybrid of the parallel DBMS and Hadoop approaches to data analysis, achieving the performance and efficiency of parallel databases, yet still yielding the scalability, fault tolerance, and flexibility of MapReduce-based systems. The ability of HadoopDB to directly incorporate Hadoop and open source DBMS software makes HadoopDB particularly flexible and extensible for performing data analysis at the large scales expected of future workloads

The author [14] 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 main feature of hadoop is to partition the data by calculating 1000's of hosts in similar to the remaining data. Geospatial data analysis has been done through Hadoop and MapReduce in cloud computing technology. The author concludes that the execution of Spatial query in mapreduce involves join transactions on small scale yields efficient results with minimum time on comparison with hadoop and distributed DB systems.

The paper [15] discuss 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. This Hadoop application requires streaming access to data files. Data storage and data processing try to solve which helps hadoop system to improve processing speed and reduce time to execute the task. Hadoop application requires stream access to data files. During placement of data files default assignment of Hadoop does not consider any data characteristics. If the related set of files is stored in the same set of nodes, the efficiency and access latency can be increased. Hadoop uses Map Reduce framework for implementing large-scale distributed computing on unexpected data sets. There are possible duplicate computations being performed in this process. No mechanism is to identify such duplicate computations which increase processing time. Solution for above problem is to co-locate related files by considering content and using locality sensitive hashing algorithm which is a clustering based algorithm will try to locate related file streams to the same set of nodes without affecting the default scalability and fault tolerance properties of Hadoop and for avoiding duplicate computation processing mechanism is developed which store executed task with result and before execution of any task stored executed tasks are compared if task find then direct result will be provided . By storing related files in same cluster which improve data locality mechanism and avoiding repeated execution of task improves processing time, both helps to speed up execution of Hadoop.

The author [16] 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 Daas Hadoop enables surplus data to be streamlined for any distributed processing system across clusters of computers using simple programming models. It truly is made to scale up from single servers to a large number of machines, each and every offering local computation, and storage space. Instead of depending on hardware to provide high-availability, the library itself is built to detect and handle breakdowns at the application layer, so providing an extremely available service along with a cluster of computers, as both versions might be vulnerable to failures. This module described the Map Reduce execution platform at the heart of the Hadoop system. By using Map Reduce, a high degree of parallelism can be achieved by applications. The Map Reduce framework provides a high degree of fault tolerance for applications running on it by limiting the communication which can occur between nodes, and requiring applications to be written in a "dataflow-centric" manner
The author [17] 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

THE IIOAB JOURNAL
ISSN: 0976-3104

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 present methods as splitting method for balancing workload, pending file structure and redundant data partition dealing with relation between spatial objects, a strip-based two-direction plane sweeping algorithm for computation accelerating. Based on these methods, ANN(All nearest neighbors) query and astronomical cross-certification are developed. Performance evaluation shows that the MapReduce-based spatial applications outperform the traditional one on DBMS. We implement two typical spatial applications, all nearest neighbor and astronomical cross-match, which faces the same difficult problem that distance computing is required

The author [18] 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. We describe the architecture of HDFS and report on experience using HDFS to manage 25 peta bytes of enterprise data at Yahoo.

The paper [19] 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. The 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 [20] describes about the SpatialHadoop and a full-fledged MapReduce framework with native support for spatial data. SpatialHadoop 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. This has led to hybrid cloud architecture to handle geospatial data compared to object based data storage. The hybrid model allows retrieval of geospatial data from public cloud. The execution is possible over a distributed large scale devices which are virtually reserved by the organization of cloud

The author [21] 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 [22] 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. The space and time will also effectively reduce during storage. Also the remote data integrity checking detects the threats and misbehaving server while storing the data in cloud ensuring data security. Calculating digital signature may secure file more efficiently.

THE IIOAB JOURNAL
ISSN: 0976-3104

The author [23] discuss about the Steganography technique to secure the data dealing with number of attacks. Tools have been considered for detection of information over Internet. Instead of secured algorithms of the existing algorithms to hind the image with data , still public key like Deffie-Hellman and RSA is susceptible in mathematics. Hence the paper [] proposed quantum cryptography with steganography which provides security in an efficient way. They use of quantum cryptography has added another layer of defense to our data. Even if we use most secure encryption algorithm and best stegno technique to hide our data, if the keys gets compromised these security will be of no use. Quantum cryptography addresses current as well as emerging threats and it definitely has “competitive advantage” over other public key cryptosystems. They used quantum cryptography only for key distribution rather than for entire messages because of limitations of transmission speeds and hardware expenses. The representation of bits through polarized photon is the foundation of quantum cryptography that serves as the underlying principle of quantum key distribution. They are concentrates on theory of quantum cryptography and the use of quantum cryptography for key distribution and how the use of quantum cryptography contributes to the field of steganography. They using quantum mechanics and photon polarization we can provide perfect security

The paper [24] deals with the Advanced Encryption Standard (AES) which is a secured authentication technique for cloud computing environment. In Advanced Encryption Standard, the key deliberation dealt in this application is the encryption schema for protected data by making it incoherent for all. . Implementing AES for security over data provides less memory utilization and less computation time as compared to other algorithms. It also provides security to cloud users as encrypted data in the cloud that protects the data from many attacks

The paper [25] deals with cloud computing which is an internet based computing technology. It is used to share the large amount of software information and resources to the world. In this cloud environment, resources are shared to all servers to separate users. This computing system supports distributed services, multi-domain Infrastructure, and multi-users to secure data efficiently. It also guarantees the secured data storage system by using several encryption algorithms with digital signature. The security model in a cloud computing environment, here file one encrypted with RSA algorithm in which keys are created sequence one by one to the system. This ensuring a major secures and also solves the main security issues like a new login user data hacking to the attacker. Login into the main system is compulsory and download, store the files. The encrypted a file is hide from unauthorized users. The files already store the main system server. It only single user multiple servers. The user forgets a password and not able to access same user name have key value to identify unique values. Once login the entry detail is cannot access the same user name login. The new user security for cloud computing platform includes RSA and encryption algorithm. The user login execution period is not a part of higher.(ie) implementation of each algorithm is perform different servers, and download, upload a files to take overall system is stop difficult. The RSA algorithm and digital signature with encryption model high secured and light encryption system information. We want to work ensure safe communication computers between systems to user.
The paper [26] deals with cloud computing that is used to resolve the day to day computing problems like Hardware, Software and Resource accessibility unhurried by Computer users. It is used to provide unchallenging and non-ineffectual solution for computational issues. They implement the digital signature with RSA algorithm to provide the Cloud Storage Methodology and Data Security in cloud.

The author [27] discussed three algorithms to secure the communication on the network. These algorithms provides authentication of the document by the sender through digital signatures. Data encryption algorithm provides encrypted data to secure data from outsider attacks. Steganography can change the presentation of data to make the data secured and to avoid the interpretation of data by users in order to provide data security on the network. This security related issues are the greatest obstacle in the popularity of cloud. Therefore we are going to use the combinations of three different algorithms- DSA, DES and Steganograohy. These algorithms help to reduce the problems of security on cloud. Digital signature Algorithm, Data Encryption Standard and Steganography to improve the security in cloud computing. We find that the Time complexity is high because it is a one by one process but in future this time complexity could be reduced. We try to improve the time complexity by using other security algorithms.

The author [28] describes integration of GIS and cloud computing solves many issues based on large-scale scalable server cluster. The author discusses about issues like storage of spatial data, index of spatial data and its operations. They evaluated the presentation and effectiveness of operations of spatial data in the platform of Hadoop with the original data available. Normally spatial objects are bigger and more complex structure with

THE IIOAB JOURNAL
ISSN: 0976-3104

various properties. Every object not only has spatial properties such as points, lines and polygons in itself, but also group of similar objects with their relationship in clustering of spatial data. These clusters give easy access to process the data where the access does not use the Hadoop Distributed File System (HDFS) data management interface directly to access and operate the spatial data. Data files are divided into a lot of small discrete objects which are stored in different processing nodes in HDFS file system using MapReduce. This achieves storage efficiency and minimum computing power in accordance with scope and the space size. Author proposed a hybrid index process of data preprocessing which is spatial search and spatial operation of Hadoop a distributed search method. Hadoop has done dynamic spatial data index. The index data split to basic unit handled by each node and Map/Reduce node can concurrently handle the data file and its equivalent index file. The author done preliminary research on storing and indexing of spatial data. For this, work on spatial data analysis, processing and mass spatial data storage is required to attain the objective of geospatial cloud service.

The author [29] explained that the Hadoop – GIS is a high performance system that deals with big data by modifying query output (execution of query on MapReduce, etc).This paper proved the high scalability and performance of Hadoop-GIS with global partition indexing system on comparison with parallel SDBMS which performs efficient SDBMS to calculate intensive queries. Hadoop-GIS is available in terms of package in Hive, a software package, a set of library to process queries of spatial data. The author proposed a technique on index of building which partition the data to execute queries. The paper concludes with efficient query processing of spatial data. They have proposed an effective solution for analytical spatial queries over big datasets but they did not concentrate much on complicated spatial queries in their proposed system.

The author [30] developed a spatial Hadoop in MapReduce frame for spatial data which overcomes Hadoop-GIS drawbacks. This paper yields efficient throughput than Hadoop for k-nearest-neighbor queries and triple presentation for spatial joins.

The author [31] describes that Hadoop is a programming platform used to support the processing of large data sets in a distributed computing environment. Authors described the evolution of Hadoop which is a framework that divides an application into different parts. The current Hadoop ecosystem consists of the Hadoop Kernel, MapReduce, HDFS and numbers of various components like Apache Hive, Base and Zookeeper.

The author [32] discussed about geospatial data in a distributed environment for Big Data analysis in WebGIS. This involves cloud-based storage and calculation of commodity cluster that lead to increase in performance of Input and Output. Also, parallelized performance is achieved simultaneously. This paper concludes with performance improvement of 14.6 % and 57%-153% better than traditional database management system.
The author [33] proposed techniques for handling the large size of images acquired by airborne and remote sensing satellite sensors. Also explained various techniques used for effective management and process to handle large data sets.
The author [34] described the HDBS based Distributed cache system in their work. The cache services are designed with three access layers an in-memory cache, a snapshot of the local disk, and the actual disk view provided by HDFS.
The author [35] discussed three algorithms to secure the communication on the network. These algorithms provides authentication of the document by the sender through digital signatures. Data encryption algorithm provides encrypted data to secure data from outsider attacks. Steganography can change the presentation of data to make the data secured and to avoid the interpretation of data by users in order to provide data security on the network.

The author [36] describes the security related problems like hacking, stealing, misusing etc. and also deals with the security related problems in cloud computing. The author used combinations of three different algorithms such as DSA, DES, and Steganography. These algorithms are used to rectify the issues on cloud to store the data. Some unwanted activities can damage the data that is cybercrime. The security issue in cloud is that, it has a single point of failure. One mistake or failure can impact the whole group. The hacker not only hack the cloud data it also hack the user account. The main intent of security is to provide availability, confidentiality, integrity to the data. The author used triple security in cloud computing by using three different security algorithms. These algorithms are based on signature of digital based encryption of data and data hiding beyond the file of audio. The author implement Digital signature Algorithm, Data Encryption Standard and Steganography to improve the security in cloud computing with the disadvantage of time complexity.

THE IIOAB JOURNAL
ISSN: 0976-3104

The author [37] describes the method to pass information with unknown existence to repel attention of the potential attacker. This method of information hiding can be given to copyright protection for digital media along with hiding data for confidentiality. The paper focuses on the Least Significant Bit (LSB) method to hide messages in an image. The author enhanced the LSB technique by randomization to disperse the messages in an image to make harder for unauthorized users to get the original messages.

The author [38] describe data security initiate the threats and the unpleasant trend that has necessitated the need for an advanced approach to data security. Serpent encryption algorithm and distributed steganography are already proven techniques for securing the data. This paper describes an developed mechanism to ensure data security by strategically combining serpent cryptographic algorithm and distributed steganography. The unified approach is aimed at leveraging the strength of these two proven techniques to achieve a robust mechanism for ensuring confidentiality and integrity of data in the cloud. In this paper, they focus on securing the data in the cloud to ensure the data confidentiality and integrity as a proposed work.

The author [39] described data storage and management technology because of cloud computing are getting more extensive attention in various applications. Security will be a serious challenge to store the data in the cloud environment. This paper describes the advantages of spatial data management in cloud computing environment. The paper sets a spatial data security management model under the cloud environment, and reviews the key security technologies of the spatial data storage and management. But the application of cloud computing in spatial data storage and management is favorable to many IT enterprise and users. It has a lot of advantages in cloud computing technology which is still to be improvised further.

The author [40] describes the collection of thousands of servers host directly that are connected to store and execute the client request tasks with distribution of storage of data and computation across many servers which the resource can produce with demand which are economical at every size. The author explains the architecture of HDFS and manages 25 peta bytes of enterprise data at yahoo.

The author [41] describes the Hadoop Distributed File System (HDFS) storage files. HDFS gives scalable, fault-tolerant storage at low cost. HDFS stores files across a collection of servers in a cluster. Files are splitted into a blocks and each block is written to more than one servers. This copying provides both fault-tolerance and performance. HDFS make sure data availability by continually monitoring the servers in a cluster and the blocks that they manage.

The author [42] describes the fault occurs in large scale distributed systems such as Hadoop clusters. Native Hadoop provides basic support for Fault tolerance. However, simply re-processing the whole task decreases the efficiency of job execution, especially when the task is almost done is discussed in this paper.

The author [43] describes the inspects Hadoop cluster and security for Hadoop clusters using Kerberos. Then security enhancement using role based access control, reviewing built in protections and weaknesses of these systems. The goal is to explore security problems in which Hadoop data users face with pragmatic advice on how to secure these environments. Since this new technology of data storage is increasing tremendously in many folds, there have been increasing data security and privacy concerns for people who outsource data on this Hadoop clusters.

The author [44] deals with confidentiality for encrypting large data. As it is a time consuming process, this is controlled by an efficient application of the process in parallel mode. This paper yields efficient cost solution to process large scale data by encryption which undergoes compression to save the space. The persistent storage compared to network storage that gives high performance which is cost-effective than other traditional alternatives. And the cloud infrastructures should support the use of persistency which is attached locally for efficient support of Big Data and other I/O intensive applications. As there are existing methods which supports networked storage for availability of data may not be applicable to big data systems.

THE IIOAB JOURNAL
ISSN: 0976-3104

COMPARISON OF SPATIAL DATABASE AND SPATIAL CLOUD DATABASE

Table: 1. shows the comparison of spatial database and spatial cloud database

Sno		Spatial Db	Spatial Cloud Db
1	security	Normal security	High security
2	user access	Limited	Global access
3	Storage capacity	limited	unlimited

PROPOSED METHODOLOGY

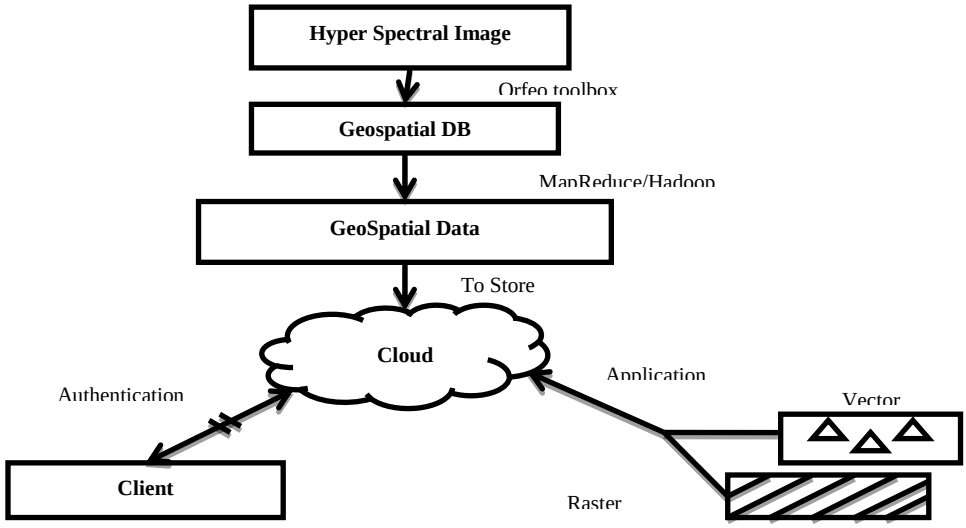

Fig: 1. proposed Architecture

For the proposed architecture, the hyperspectral images acquired by remote sensing satellites are given as input. The input images are already preprocessed. The input images are processing by using Orfeo toolbox to create Geo-Spatial Database.

Secondly, the geo spatial database is normalized and reduced using MapReduce and Hadoop to store it into the cloud.

Thirdly, the Map Reduced Geospatial data are processed for various application in the cloud environment. A simply authentication mechanism is used for providing access to the data which is 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

CONCLUSION

The paper is concluded by proposing a hybrid model through survey of papers on other existing stenography and cryptographic algorithms. Many tools also have been surveyed to detect hidden information over the internet. Also, MapReduce is used to minimize the storage area which is mapped to GIS database in a cloud using normalization techniques. The paper concludes with the detection of hidden information available over the internet in order to secure the data from steganalyst .Thereby the proposed model may result with delivery of

THE IIOAB JOURNAL
ISSN: 0976-3104

information without any loss and with minimum time consumption. Finally the integration of GIS and cloud gives efficient data retrieval and other E-Commerce services that reduce any manual work by outsourcing to the cloud GIS Effective data balancing is made by applying load balancing strategy.

FINANCIAL DISCLOSURE

No financial support was received to carry out this research.

ACKNOWLEDGEMENT

Authors would like to thank school of computing science and engineering, vit university vellore for providing resource and support to carry out this research work.

CONFLICT OF INTERESTS

Authors declare no conflict of interest.

REFERENCES

[1] Eman Mahmoud, Osman Hegazy, Mohamed Nour El-Dien. [2013] Integration of GIS and Cloud Computing for Emergency System. International *Journal Of Engineering And Computer Science*.ISSN:2319-7242

[2] Yaser Khalilizangelani , Saman Ghaffarian.[2013] A Study of Geospatial Data Processing Based on Cloud Computing. Proceedings of ACRS 2013

[3] Ms. Manisha, Dr. Kamlesh sahrma, Ridhika Sharma Manisha.[2015] A Technique To Increase Integrity Of Cloud Data Using Hybridalgorithms. ISSN: 2277-9655

[4] Ge, X., & Wang, H. [2009] Cloud-based service for big spatial data technology in emergence management. In Proceedings of the ISPRS 38(7):C4.

[5] Christophe, E., Inglada, J., & Giros, A. (2008). Orfeo toolbox: a complete solution for mapping from high resolution satellite images. International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 37: 1263-1268.

[6] Mohammad Naghavi.[2012]Cloud Computing as an Innovation in GIS & SDI: Methodologies, Services, Issues and Deployment Techniques. JGIS: 597-607

[7] McGavra G, Morris S, Janée G. [2009] Technology Watch Report: Preserving Geospatial Data.

[8] Vinhas L, De Souza RCM, Câmara G. [2003] November). Image Data Handling in Spatial Databases. In GeoInfo.

[9] Song WW, Jin BX, Li SH, Wei XY, Li D, Hu F. [2015] Building Spatiotemporal Cloud Platform for Supporting GIS Application.ISPRS Annals of Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 1:55-62..

[10] Almeer MH. [2012] Hadoop mapreduce for remote sensing image analysis.*International Journal of Emerging Technology and Advanced Engineering*,2(4): 443-451.

[11] Zhang Q, Liu L, Lee K, Zhou Y, Singh A, Mandagere N, Alatorre G. [2014] Improving Hadoop Service Provisioning in A Geographically Distributed Cloud. In Cloud Computing (CLOUD), 2014 IEEE 7th International Conference on (pp. 432-439). *IEEE*..

[12] Seyed Reza Pakize.[2014] A Comprehensive View of Hadoop MapReduce Scheduling Algorithms. *International Journal of Computer Networks and Communications Security* ISSN 2308-9830

[13] Abouzeid A, Bajda-Pawlikowski K, Abadi D, Silberschatz A, Rasin A. [2009] HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads. *Proceedings of the VLDB Endowment*,2(1):922-933..

[14] Gaikwad, R. L., Dakhane, D. M., & Pardhi, R. L. (2013). Network Security Enhancement in Hadoop Clusters. International Journal of Application or Innovation in Engineering And Management (IJAIEM), 2:151-157.

[15] Sayali Ashok Shivarkar.[2014] Speed-up Extension to Hadoop System. *International Journal of Engineering Trends and Technology (IJETT)* ISSN: 2231-5381

[16] Brahmanaidu N, Riaz S. [2012]. Distributed Data Storage and Retrieval on Cloud by using Hadoop. *International Journal of Science and Research*..

[17] Wang K, Han J, Tu B, Dai J, Zhou, W, Song, X. [2010]. Accelerating spatial data processing with mapreduce. In Parallel and Distributed Systems (ICPADS), 2010 IEEE 16th International Conference on(pp. 229-236). *IEEE*..

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

[19] Economides G, Piskas G, Siozos-Drosos S.[2013] Spatial Data and Hadoop Utilization..

[20] Eldawy A, Mokbel MF. [2015]. SpatialHadoop: A MapReduce framework for spatial data. In Proceedings of the IEEE International Conference on Data Engineering (ICDE'15). *IEEE*..

[21] Mehak , Gagandeep .[2014] Improving Data Storage Security in Cloud using Hadoop. Mehak Int. Journal of Engineering Research and Applications ISSN : 2248-9622.

[22] Yogesh V Bhapkar, Rakesh S Gaikwad, Milind R. Hegade Yogesh V. Bhapkar et al,[2015] Providing Security And Privacy To Cloud Data Storag. *International Journal of Computer Science and Information Technologies*, ISSN:2231-5381

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

[24] Abha M, & Bhansali, M. (2013). Enhancing cloud computing security using AES algorithm. *International Journal of Computer Applications*, 67(9).

[25] T. Sivasakthi,Dr. N Prabakaran[2007] Applying Digital Signature with Encryption Algorithm of User Authentication for Data Security in Cloud Computing. *International Journal of Innovative Research in Computerand Communication EngineeringISSN(Online):* 2320-9801 ISSN (Print): 2320-9798

[26] Somani U, Lakhani K, Mundra, M. [2010] Implementing digital signature with RSA encryption algorithm to enhance the Data

THE IIOAB JOURNAL
ISSN: 0976-3104

Security of cloud in Cloud Computing. In Parallel Distributed and Grid Computing (PDGC), 2010 1st International Conference on (pp. 211-216). *IEEE*.

[27] Khushboo Gupta, Neha Goyal, Puneet Rani.[2014] Study of Security Algorithm to Provide Triple Security in Cloud Computing. *International Journal of Scientific and Research Publications* ISSN 2250-3153.

[28] Arya, Richa. [2014] Triple Security of File System for Cloud Computing. *International Journal of Computer Science and Engineering*, 2(3):148-154.

[29] Bhosale Harshawardhan S., and Devendra P. Gadekar. [2014] Review Paper on Big Data and Hadoop. *International Journal of Scientific and Research Publications*,4(10):1-7.

[30] Hu P, Dai W.[2014] Enhancing Fault Tolerance based on Hadoop Cluster, *International Journal of Database Theory and Application*, 7(1):37-48

[31] Lethby Toby, and A Jagan. [2015] Spatial Data Analysis using Map-Reduce Technique. International Journal of Electronics Communication and Computer Technology, 5(1):94-97.

[32] Izevbizua PO. [2015] Data Security In The Cloud Using Serpent Encryption and Distributed Steganography, *European Scientific Journal*, ESJ, 11(18):347-359.

[33] Mukhi, Parul, and Bhawna Chauhan. [2014] Triple System Security in Cloud Computing. *International Journal Of Engineering And Computer Science* 3(7):7364-7374.

[34] Rama Naga Durga Rao Khaja, Venkateswara Rao Kota. [2014] Hybrid Cloud framework for Object Storage based Geo Spatial Remote Sensing Data Processing. International Journal of Engineering Sciences & Research Technology, 449-454.

[35] Olson, Mike. [2010] Hadoop: Scalable, flexible data storage and analysis. IQT Quart 1(3):14-18.

[36] C Wanga,b, F Hub, X Hua, S Zhaoc, W Wena, C Yangb. [2015] A Hadoop-Based Distributed Framework for Efficient Managing and Processing Big Remote Sensing Images. ISPRS Annals of Photogrammetry, *Remote Sensing and Spatial Information Sciences*, 1:63-66.

[37] Eldawy Ahmed, and Mohamed F Mokbel. [2015] SpatialHadoop: A MapReduce framework for spatial data. Proceedings of the IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea.

[38] Li Xiang, Zi Fan Liu, Wen Bing Liu, An Xu, and Ling Ma. [2013] A spatial data security model under the cloud environment. *In Advanced Materials Research*, 765:1267-1270.

[39] Akshay MS, Suhas Mohan, Vincent Kuri, Dinkar Sitaram, HL. Phalachandra. [2014] Efficient Support of Big Data Storage Systems on the Cloud. arXiv preprint arXiv, 1411-7507.

[40] Shvachko, Konstantin, Hairong Kuang, Sanjay Radia, and Robert Chansler. [2010]. The hadoop distributed file system. 2010 IEEE 26th Symposium Mass Storage Systems and Technologies (MSST), 1-10.

[41] Wang, Yonggang, and Sheng Wang. [2010] Research and implementation on spatial data storage and operation based on Hadoop platform. Geoscience and Remote Sensing (IITA-GRS), 2010 Second IITA International Conference on IEEE, 2: 275 – 278.

[42] Zhang J, Wu G, Hu X, Wu X. [2012] A distributed cache for hadoop distributed file system in real-time cloud services. 2012 ACM/IEEE 13th International Conference on Grid Computing (GRID), 12-21).

[43] Zhong, Yunqin, Jizhong Han, Tieying Zhang, and Jinyun Fang. [2012] A distributed geospatial data storage and processing framework for large-scale WebGIS. 20th International Conference on Geoinformatics (GEOINFORMATICS), 1-7.

[44] Aji A, Wang F, Vo H, Lee R, Liu Q, Zhang X, Saltz J. [2013] Hadoop GIS: a high performance spatial data warehousing system over mapreduce. *Proceedings of the VLDB Endowment*, 6(11):1009-1020

ABOUT AUTHORS

Karthi Sankar *has completed his BE in CSE from Muthayammal Engineering College Rasipuram. And then he completed his ME in CSE from Jerusalem College of Engineering Chennai. Currently he is doing his PhD in School of Computing Sciences and Engineering, VIT University, Vellore, India. His area of research includes Geo-Spatial Application and Cloud computing.*

Dr. Prabu Sevugan *have completed Bachelor of Engineering in Computer Science and Engineering from Sona College of Technology (Autonomous) and Master of Technology in Remote Sensing from College of Engineering Guindy, Anna University Chennai and one more Master of Technology in Information Technology at School of Computer Science and Engineering, Bharathidasan University Trichy. Did his Doctoral studies on Integration of GIS and Artificial Neural Networks to Map the Landslide Susceptibility from College of Engineering Guindy, Anna University, Chennai. He was a Post-Doctoral Fellow at GISE Advanced research lab, Department of Computer Science and Engineering, Indian Institute of Technology Bombay. He has more than 50 publications in national and international journals and conferences. He organized 3 International Conferences which includes one IEEE Conference as chair and also participated in many workshops and seminars. He is a member of many professional bodies and senior member of IACSIT, UACEE and IEEE. He is having more than eleven years of experience in teaching and research. Currently I am working as a Division Chair for Parallel and Distributed Computing, School of Computing Science and Engineering, VIT University Vellore.*