BUILDING SPATIOTEMPORAL CLOUD PLATFORM FOR SUPPORTING GIS APPLICATION

W.W. Song [a], B.X. Jin [b,*], S.H. Li [b,c], X.Y. Wei [b,d], D. Li [b], F. Hu [e]

[a] Department of Geoinformation Science, Kunming University of Science and Technology, 68 Wenchang Road, Kunming, Yunnan, China-55192452@qq.com
[b] Yunnan Provincial Geomatics Centre, 404 West Ring Road, Kunming, Yunnan, China- (jinbx163, lsh8010)@163.com, (19423221, 29738890) @qq.com,
[c] College of Tourism & Geographic Sciences, Yunnan Normal University，768 Juxian Street in Chenggong District, Kunming, Yunnan, China-lsh8010@163.com, wang_jinliang@hotmail.com
[d] College of Geographic Sciences, Nanjing Normal University,No.1,Wenyuan Road，Xianlin University District,Nanjing,China-19423221 @qq.com
[e] Center for Intelligent Spatial Computing, George Mason University, 4400 University Dr., Fairfax, VA, USA- hufei68@gmail.com

KEY WORDS: Spatiotemporal Cloud Platform, HDFS, MapReduce, GIS Application, Geospatial Analysis

ABSTRACT:

Traditional geospatial information platforms are built, managed and maintained by the geoinformation agencies. They integrate various geospatial data (such as DLG, DOM, DEM, gazetteers, and thematic data) to provide data analysis services for supporting government decision making. In the era of big data, it is challenging to address the data- and computing- intensive issues by traditional platforms. In this research, we propose to build a spatiotemporal cloud platform, which uses HDFS for managing image data, and MapReduce-based computing service and workflow for high performance geospatial analysis, as well as optimizing auto-scaling algorithms for Web client users' quick access and visualization. Finally, we demonstrate the feasibility by several GIS application cases.

1. INTRODUCTION

Modern information technology achievements, especially the rapid popularity of the Internet, 3G and 4G wireless communication networks in the global scale, led to the fundamental changes in geospatial information services mode. The service mode has been transformed from providing the traditional data into providing comprehensive geospatial information, and the scope of services has been gradually enlarged from traditional specialization to social popularization (Ye 2008; Luo et al., 2009). For example, people need navigate routes, search location and obtain surrounding information through the mobile map, as well as easily enjoy geospatial information services at any place and time. Therefore, geospatial information platforms are built by authoritative geo-information agencies for integrating the fundamental data (e.g. DLG, DOM, DEM, etc.), LiDAR data, street-view data, oblique photogrammetry data, and other data collected for supporting government decision-making and public social services.

The traditional geospatial information platforms manages structured, unstructured, and semi-structured data types, which are generated dynamically and fast. However with the dramatic increasing of data volumes, it exposed the following issues:

(1). Big data storage and management (Ji et al., 2012; Yang et al., 2015): the platform needs to manage multi-type, multi-scale, multi-resolution and multi-temporal big geo-database with hundreds of terabyte capacity, that makes the big data management much more complicated. Storing big data requires high scalable storage devices that are able to easily scale up to fulfill the request, and easily scale down to minimize the cost. Furthermore, it need the visualization of 2-dimension (2D), 3-dimension (3D) and 4-dimension (4D, 3D+time) data to realize the information integration amongst different application systems through platform. 2D and 3D representation are relatively simple because the techniques are more mature. Due to the addition of time dimension, it is difficult to visualize 4D data. Therefore, how to effectively process, store, manage and visualize big data poses severe challenges for the existing platform.

(2). High performance geospatial analysis: the platform will provide various geospatial computing and analysis services, such as map projection, spatial data editing, map cache, spatial indexing, spatial interpolation, data modeling, and fundamental spatial analysis (e.g. topology, statistics, buffer, overlay, and route analysis, etc.). When all geospatial analysis are executed at the same time, it will put great computing pressure on platforms and leads to low compute efficiency. At present, fundamental geospatial data updating cycle is gradually shortened, and the novel data from all kinds of sensors need be uploaded into platform to update database, so the data processing work strength will be increased. The latest geo-database will be applied to update the map-service published on GIS website. Before publishing the map-service, the map data should be partitioned into tiled map. It's a very time consuming process as well as typical computing-intensive process. Therefore, if you want to raise the computing capability for platform, the compute-intensive challenge needs to be resolved.

* Corresponding author

(3). High concurrent access: The main purpose for building the platform is to help different government departments and social communities share geo-information. By accessing platform, users can acquire valuable map services and spatial analysis services. However, with the increasing number of the applications from early several to hundreds running on platform, it will be difficult to avoid the high concurrent access issue. In the existing platform, when there are a large number of concurrent accesses, users need wait too long to get the response, and sometimes users cannot even get the right response from the platform. It is because both hardware and software have some problems in carrying capacity, such as concurrent access could lead to out of memory and high CPU usage. In some cases, excessive concurrent access will cause software issues, such as deadlock (Yang and Huang, 2013). These issues, if seriously, will cause the collapse of the entire platform. If the platform can't address these issues, such as concurrent access very well, it will lose the meaning of existence, in other words, the platform will lose vitality. Obviously, the high concurrent access is another significant challenge that needs to be resolved for platform. Fortunately, the emerging cloud computing technology and MapReduce computing framework can offer reliable solutions for traditional geospatial information platform.

In this research, the main contribution includes: First, we design a new 3-layer architecture for spatiotemporal cloud platform. Second, we develop and implement spatiotemporal cloud platform based on Eucalyptus for IaaS, HDFS for big data management, GIS services chaining and the cluster of ArcGIS server for geospatial analysis, as well as Hadoop MapReduce for high performance spatial computing. Third, we demonstrate the feasibility by GIS application cases based on spatiotemporal cloud platform.

The remaining of this paper is structured as follows. In Section2, we overview cloud computing and big data technologies, and introduce Eucalyptus, MapReduce, Hadoop Distributed File System, and CloudGIS to help understand the challenges involved in the presented research. In Section3, we design platform architecture and implement platform with Eucalyptus, Hadoop cluster and ArcGIS server cluster, and describe the prototype system functions; Section 4 evaluates the feasibility of platform by GIS application. In Section5, we conclude this paper with discussions and future work.

2. OVERVIEW OF BIG DATA AND CLOUD COMPUTING, CLOUDGIS TECHNOLOGIES

Big Data is defined as the representation of the progress of human cognitive processes, which generally includes data sets with sizes beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time (Graham-Rowe et al., 2008). Douglas and Laney (2012) define Big Data as the high volume, high speed and/or high variety of information that require new ways of processing to allow better decision making, new knowledge discovery and process optimization. Provost and Fawcett (2013) define Big Data as datasets that are too large for traditional data-processing systems and that therefore require new technologies, like Hadoop, Hbase, MapReduce, MongoDB or Couch-DB. IBM researchers (Zikopoulos et al., 2012) describe Big Data in terms of four dimensions:1.volume, 2.velocity, 3.variety and 4.veracity. From the definition of big data described above, we note that the traditional platform built involves typical big geospatial data with characteristics of large volume, various data types (structured, semi-structured and unstructured data), and fast generating speed. How to mine valuable information from big geospatial dataset for geospatial analysis has brought the great challenges such as big data management, processing and analysis, and visualization (Ji, Li, Qiu et al., 2012; Yang et al., 2015). The birth of Cloud computing technology provides a scalable and cost efficient solution to the big data challenge in geospatial area.

The National Institute for Standards and Technology (NIST) defines Cloud computing as a pay-per-use model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud services can be categorized into at least four types: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) and Data as a Service (DaaS). The first three are defined by NIST (Mell and Grance, 2011) and DaaS is essential for geospatial sciences (Yang et al., 2011). IaaS provides the capability of provisioning computation, storage, networks, and other fundamental computing resources on which the operating systems and applications can be deployed. Being benefits of cost-efficiency, automation, scalability and flexibility of cloud services, many organizations, especially commercial IT enterprises, are moving from traditional physical IT systems to cloud computing, and increasingly providing cloud services in their products (Armbrust et al. 2010). Amazon Web Services (AWS), with some estimates suggesting that AWS holds 70% of the total IaaS market share, are the leaders in IaaS. AWS provides Elastic Compute Cloud (EC2), which enables users to rent servers to build their VMS based on Windows or Linux from a pre-built specific image, and Simple Storage Service (S3) for online storage service. By paying fee, users can get access to the world's largest data centre. In recent years, open-source solutions are available for deploying and managing a large number of virtual machines to build a highly available and scalable cloud computing platform, such as Nimbus, Eucalyptus, OpenNebula, and CloudStack (Huang et al. 2012). Eucalyptus IaaS cloud services give consumers the capability to run and control virtual machine instances deployed across a variety of physical resources (Nurmi et al. 2009). In the geospatial sciences area, cloud computing based on EC2 or Eucalyptus has solved many geospatial problems, such as data intensity, computing intensity, concurrent intensity, and spatiotemporal intensity problems (Yang et al., 2011). PaaS provides a computing platform on which consumers can develop software using the tools and libraries from the cloud provider, and deploy the software onto cloud services. Within the technology field, platforms include Google App Engine, Microsoft Azure and so on. MapReduce is a data processing and analytics technology that has been revolutionary in the field of computer science and is one of the popular technologies in the big data space (O'Driscoll et al. 2013). MapReduce organizes computation into two key operations: the 'map' function that is responsible for splitting huge dataset into sub-datasets and processing each of them independently, and the 'reduce' function that collects and merges the results from the 'map' function (Dean and Ghemawat, 2008; Del R ó et al., 2015). The original MapReduce technology is a proprietary system developed by Google, and therefore, it is not available for public use. However, the Apache Hadoop project develops an open-source software for reliable, scalable massive data processing with the MapReduce model. It contains Hadoop Distributed File System (HDFS) that is distributed storage

system for reliably storing and streaming petabytes of both unstructured and structured data on clusters, and Hadoop MapReduce, a software framework for distributed processing a large volume of datasets, as well as providing job schedules for balancing data, resource and task loads on compute clusters. Cloudera is Apache Hadoop-based software, and Cloudera manager is an industry standardized administration package for the Hadoop ecosystem. With Cloudera Manager Web User Interface, we can deploy and centrally operate the Hadoop infrastructures. In addition, it gives us a cluster-wide, real-time view of nodes and monitors the running services, and enables configuration changes across the cluster.

With the development of Cloud computing technologies, GIS also experienced from traditional GIS to Cloud GIS development. Esri ArcGIS is operated on desktops and local servers, and focuses on single users, mostly GIS experts. It requires the installation and maintenance of both hardware and software, and lacks the ability to support the needs of large-scale concurrent access (Yang and Huang, 2013). Cloud computing provides the computing capability to build and deploy GIS as a service, which can be referred as Cloud enabled GIS or Cloud GIS (Mann 2011). GIS Software as a Service provides centralized, cloud-based clients and applications that can easily solve complex problems using GIS tools and data. ArcGIS Online is a typical SaaS application, it provides the opportunity to gain insight into data quickly and easily without installing and configuring GIS software. For example, in 2010, when flooding endangered residents and inundated homes, businesses, and farmland of Queensland, Australia, Esri Australia published and disseminated GIS services by ArcGIS Server on Amazon EC2 for fast emergency responding to the disasters. Based on Cloud computing technologies, Chen et al. (2008) built a high performance workflow system MRGIS using MapReduce clusters to execute GIS applications efficiently; Park et al. (2010) used Hadoop HDFS and MapReduce to do massively parallel processing of 3D GIS data, they found the computing time is vastly reduced with a cluster of computing nodes. In GIS PaaS application, Gong et al. (2010) proposed to integrate GIS geoprocessing functions with scalable Microsoft Cloud computing platform Azure for providing geoprocessing capabilities; Aji et al. (2013) presented Hadoop GIS for running large scale spatial queries, it's a scalable and high performance spatial data warehousing system; Lin et al. (2013) proposed and implemented an architectural design for a novel Cloud computing platform based on two Web Coverage Service and Web Map Service interfaces from the Open Geospatial Consortium (OGC), cloud storage from Hadoop HDFS, and image processing from MapReduce; Gao et al. (2014) built a scalable distributed platform and a high performance geoprocessing workflow based on the Hadoop ecosystem to harvest crowd-sourced gazetteer entries.

3. BUILDING SPATIOTEMPORAL CLOUD PLATFORM

3.1 Spatiotemporal Cloud Platform Architecture

Spatiotemporal cloud platform discussed in this paper is a private cloud platform. The goal of the platform is to solve the issues faced by traditional geospatial information platform, such as data-intensive, computing-intensive, and concurrent-intensive problems, so that to implement big geo-data analytics and management, to provide geospatial information services for multi-departments of government, and to facilitate information sharing.

Figure 1 demonstrates the architecture of spatiotemporal cloud platform. It's composed of three layers from bottom to top, which respectively is IaaS, PaaS, and SaaS.

(1). IaaS layer is located in the bottom of spatiotemporal cloud services model, this type of software provides customers with virtual machine (VM), virtual storage, virtual network and other infrastructure resources. Based on the open source cloud computing software Eucalyptus4.0, we develop a novel software Dynamic Computing Cloud (DC2) for the specific requirements of geospatial domain. DC2 optimizes some cloud service functions, such as resource scheduling, auto-scaling and load balance. Through DC2, all existing different types and properties of physical servers and storages resources are virtualized, so they can be adapted into a unified, dynamic and scalable pool of resources for computing and storage, and provide API interface for PaaS or SaaS.

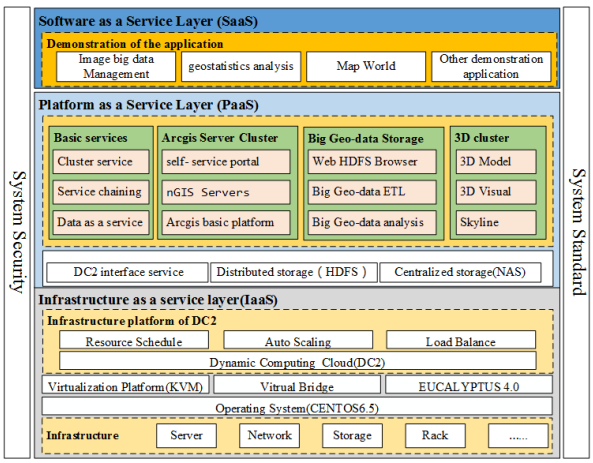

Figure 1. Spatiotemporal cloud platform architecture

(2). PaaS layer is located in the middle of architecture, which provides the environment of development, running and testing for users. PaaS realizes a series of key technologies for Cloud GIS services and provides software and application platform support for all kinds of business application of SaaS. For publishing map services, DaaS (Data as a Service) and geospatial analysis services chaining, PaaS builds ArcGIS Server cluster. In order to implement 3D scene modeling and visualization, PaaS offers 3D GIS software Skyline cluster environment. Through HDFS, PaaS provides the reliable big geo-data distributed storage. Integrating DC2 with Hadoop ecosystem, PaaS achieves high performance statistical analysis.

(3). SaaS layer provides users with more than one application. These applications can run on a physical machine, or may run on multiple physical machines, or can run on VMs. Users don't care about the specific running of applications. In this paper, we mainly offer application framework such as big iamge data management, geostatistics analysis, and geospatial visualization for users.

3.2 Design for Spatiotemporal Cloud Platform

In this paper, we mainly discuss serval key technologies designed for spatiotemporal cloud platform, such as HDFS-based big data management, MapReduce-based geospatial

analysis, and auto-scaling mechanism for high concurrent access.

3.2.1 HDFS-based Big Data Management: Big geo-data has the large volume of various types of data (e.g. structured, unstructured, and semi-structured data). The traditional way of data storage is to use geodatabase (e.g. Oracle Spatial) and disk array. With the increase of data volume, it becomes difficult and brings high cost. In this paper, a HDFS-based big data management system is designed for big image data storage and its application. The scalability of spatiotemporal cloud platform provides a flexible resources extension, and the system can increase or decrease the nodes on HDFS cluster. The figure 2 shows the architecture of the system. It is composed of three modules: a Data management User Interface &API, a Big geo-data process unit, and a HDFS system.

1). Data management User Interface & API: the Web HDFS APIs are integrated to provide a web page for client users to browse and manage the data stored on HDFS. They can upload data from local disk to HDFS, move data and delete data. A set of data operation APIs, based on HDFS API, are designed for user applications to pump large data from database or NAS storage to HDFS or query a subset of big data and download them to user application system.

2). Big geo-data process unit: The process unit includes Web HDFS APIs, MapReduce APIs, HDFS APIs and spatiotemporal index generator. In order to efficiently store and access big image data in the HDFS, a spatiotemporal index generator is designed. Through creating spatiotemporal index in the <key, value> type which is suitable for MapReduce framework and defining the appropriate image segmentation, the system can store and locate image data more efficiently and accurately. The algorithm based on Mapreduce can be created to help users or applications query image data from big image dataset stored in HDFS. A mechanism called PMMI (predefined multiple indices mechanism) provides the optimal indices to access image data.

3). HDFS: the HDFS-based big data is built on IaaS layer of spatiotemporal cloud platform. A Hadoop cluster is constructed with numbers of data nodes on physical server and virtual server. The HDFS uses all nodes server as DataNode. The NameNode and Secondary NameNode are installed on two separate physical servers.

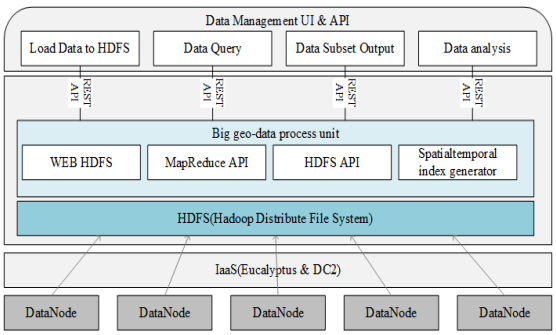

Figure 2. HDFS-based big data management architecture

3.2.2 MapReduce-based Workflow and Services Chaining for Geospatial Analysis: The traditional way of geospatial analysis usually uses a single process (e.g. ArcGIS Desktop) to execute a single task of spatial analysis (e.g. equals, crosses, within and contains of spatial objects). But current GIS applications need more and more spatial analysis functions to be executed at the same time. Service chaining is an effective method for compositing spatial analysis functions simultaneously. Based on spatiotemporal cloud platform, spatial analysis functions and resources can be packed as services, which can be called by different users and applications. If a GIS application wants to execute data processing, model building, visualization, and spatial analysis as a workflow, the service chain engine can complete this task automatically and efficiently. Figure 3 designs the service chaining mechanism for geospatial analysis. Details are listed as below:

1). Service chain combination engine：In this paper, the engine is based on the existing workflow engine Oozie. Computing resources (cloud service) are built as important services and integrated into service chain. For example, the big data reading service based on Hadoop can be called and executed by the service chain engine.

2). Service chain execution and monitoring environment: the engine is constructed on the spatiotemporal cloud platform. The environment can be dynamically built by REST API. The monitoring module can monitor the workload information and adjust resource usage quota.

3). Service of chain node: Different types of services are maintained by the DC2 service chaining driver. These service can be called and executed by workflow builder.

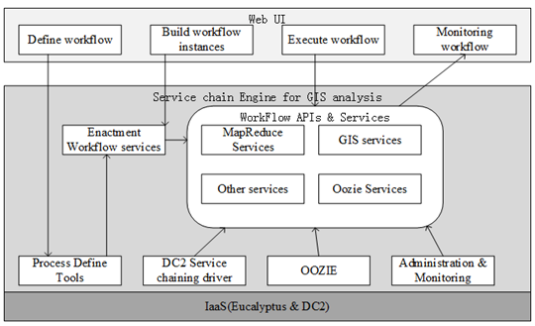

Figure 3. Service chaining for geospatial analysis

The service chaining mechanism is suitable for spatiotemporal analysis. For example, geographic condition statistics analysis involves multiple services: ① read data from geo-data source; ② convert data format: the vector data is converted from standard ArcGIS format to GeoJSON;③ upload data to HDFS; ④ call and execute Oozie workflow; ⑤ generate reports and charts of statistical results.

3.2.3 Auto-scaling Mechanism for Concurrent Access: Auto scaling and workload balancing are two key advantages of cloud computing. To meet high concurrent access and high performance computing based on spatiotemporal cloud, the auto-scaling mechanism is designed from two aspects:

1). DC2 auto-scaling for computing resources. It helps to increase or decrease computing resources. On one hand, it provides enough computing resources for smoothly and fast calculation, to alleviate the burden of accumulated load and to

support multiple concurrent access. On the other hand, the computing resources can be dynamically adjusted.

2). Auto-scaling for ArcGIS Server cluster: the ArcGIS Server cluster model called nGIS servers uses a point-to-point (P2P) way, which defines every GIS server nodes as the same role. This model can deal with the situation that if a GIS server node crashes accidently, it will not lead to GIS services stop (e.g. Map Service stop). Similarly, when a new GIS server node needs to be added into the cluster, a plug-in way can add a new node to improve services load capacity. Figure 4 describes auto-scaling mechanism of spatiotemporal cloud platform.

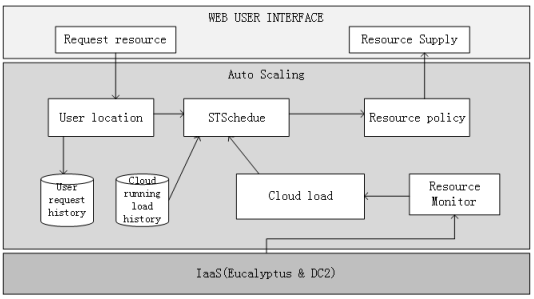

Figure 4.1. DC2 auto-scaling mechanism

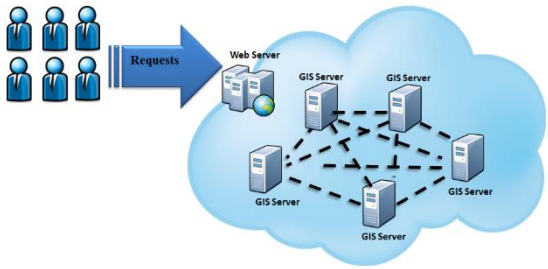

Figure 4.2. ArcGIS nServer model

Figure 4. Auto-scaling mechanism (Figure 4.1 and Figure 4.2)

The auto-scaling mechanism provides high concurrent access and high performance computing for applications built on spatiotemporal cloud platform. For example, the elastic ArcGIS Server cluster provides GIS services and DC2 provides computing resources for the Map World system which is constructed based on spatiotemporal cloud platform. When 100 concurrent users request the service of geoprocessing, 10 VMs will be automatically added to ArcGIS Server cluster for improving the efficiency of application. During the holidays, if there are less users, the platform can turn off some nodes and detach them from cluster, so to reduce the cost of resources.

3.3 Spatiotemporal Cloud Platform Prototype System

Based on the architecture and key technologies designed above, a spatiotemporal cloud platform is constructed by using 10 X86 servers and 10GB network infrastructure. The operation system installed on all servers is CentOS 6.5. Linux Kernel-based Virtual machine (KVM) is the virtualization platform. Eucalyptus 4.0 is installed and managed by DC2 to construct the IaaS layer of cloud platform. On the PaaS layer, an ArcGIS Server cluster, Hadoop cluster and 3D-GIS cluster are built. The platform supports HDFS-based big data management and high performance computing. Services for chaining managed by DC2 meet the requirement of complex and multi-step spatial analysis. All SaaS applications (e.g. land use, image management, geostatistics analysis, Map World, etc.) are deployed on spatiotemporal cloud platform. Table1 shows the platform prototype system configuration information. Figure 5 shows main functions of spatiotemporal platform.

Server name	Roles	server info
clc01	cloud controller	centos 6.5 64bit; 2.0GHz*32;128GB memory; 1TB SSD + 1.2TB SAS
cc01	cluster controller	centos 6.5 64bit; 2.0GHz*32;128GB memory; 1TB SSD + 1.2TB SAS
NC(1-7)	cloud node; data node, node manager	centos 6.5 64bit; 2.0GHz*32;128GB memory; 1TB SSD + 1.2TB SAS
HMaster	Name node, HDFS, Resource Manager	centos 6.5 64bit; 2.0GHz*32;128GB memory; 1TB SSD + 1.2TB SAS

Table 1. Prototype system configuration information

The main functions of spatiotemporal cloud platform are described as follows:

3.3.1 Portal of Spatiotemporal Cloud Platform: It provides platform management and resource services, including user/role management, physical resources management, virtual resources management, network security and policy, auto-scaling, Hadoop UI, cloud monitor, service chaining, etc.

3.3.2 Cloud GIS Services Portal: It provides GIS services (e.g. map service, buffer, overlay, spatial query, geostatistics, etc.) and allows users to work on a web interface to use all GIS services without installing any GIS tools on local machine. Through Cloud GIS portal, users can utilize cloud resources for building GIS applications.

3.3.3 Application Management: It helps users to build, deploy and manage applications on SaaS layer. For example, spatial analysis based on service chaining mechanism can be defined, executed and monitored by cloud users.

Figure 5.1. Cloud nodes management

Figure 5.2. Portal of spatiotemporal cloud platform

Figure 5.3. Cloud GIS services

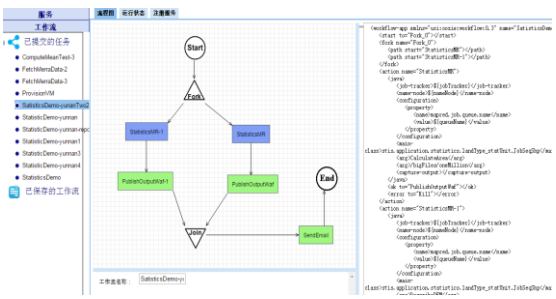

Figure 5.4. Service chaining engine

Figure 5. Spatiotemporal Cloud Platform (Figure 5.1., Figure 5.2., Figure 5.3., Figure 5.4.)

4. GIS APPLICATION CASES BASED ON SPATIOTEMPORAL CLOUD PLATFORM

4.1 Image Big Data Management System

Based on HDFS management strategy, the spatiotemporal cloud platform with redundant storage mode ensures the reliability of the data and makes the data for the expansion of the "infinite" ability. Using the cloud platform, we implement image big data management system with massive image data storage capability. The system manages multi-source, multi-type, multi-resolution, and multi-period image data of Yunnan province and provides data dissemination for users.

In the spatiotemporal cloud environment, we use 22 TB image data to test the I/O efficiency of the distributed file system. When the cluster has 5 nodes, writing data speed is 77.3 Mb/s, and reading speed is 229 Mb/s. At the same time, for the windows OS (Configured with 2 * CPU, 8 GB memory), 1 TB image file is read from data source and wrote to local disk. The reading rate is 4.3MB/s and writing rate is 1.3MB/s. The result shows that HDFS-based image file system can get high efficiency of reading and writing. Figure 6 presents the UI of the big image data management system.

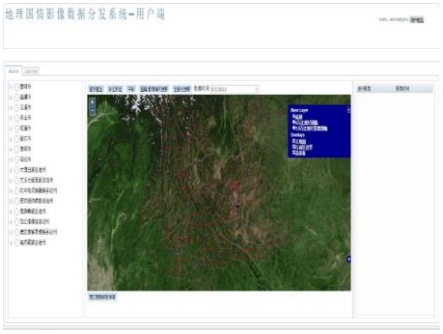

Figure 6. Image big data management system

4.2 Hadoop-Geostatistics GIS System

After data acquisition of the 1st China national geographic condition, we will conduct spatiotemporal statistics analysis task for supporting government decision-making. Traditional geostatistics analysis methods and computation patterns cannot satisfy the requirement because it takes hours, even days to obtain the statistics results, so we propose to build a high performance Hadoop-Geostatistics GIS system based on spatiotemporal cloud platform, which uses MapReduce and Oozie workflow to obtain statistics analysis results quickly.

We compare the performance between Hadoop-Geostatistics GIS system and traditional geostatistics analysis system. Figure 10 shows the runtime for a single computer, the cluster with 5 computing nodes and that with 10 computing nodes. The experimental results illustrate that the Hadoop-Geostatistics GIS system has a processing time in linear growth with the increase of data volume and computational complexity. It demonstrates the feasibility of high performance geostatistics analysis based on spatiotemporal cloud platform.

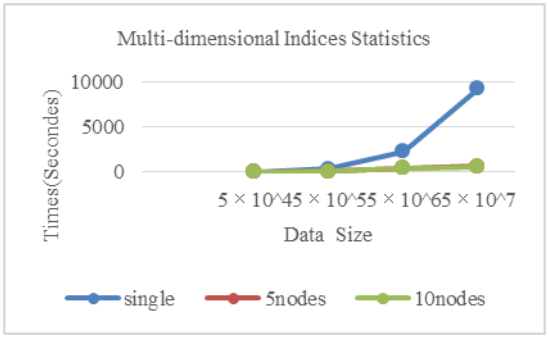

Figure 7. Computing time-consuming comparison between Hadoop-Geostatistics GIS system and Traditional statistics analysis system with PC, 5 nodes cluster, and 10 nodes cluster.

4.3 Map World System

Map World Yunnan system, as a Web application, which is based on the spatiotemporal cloud platform to provide the public with comprehensive and efficient geospatial information

services (e.g. map services, navigation, POI query, spatial analysis, etc.). Generally, this system will handle large number of users' concurrent access over Internet. In the traditional system, with the increase of concurrent accesses, users wait too long to get the response. Sometimes users cannot even get the right response from the system. In this research, we re-design Map World system based on the elastic cloud framework and deploy the system on the spatiotemporal cloud platform. By DC2 with the scalable computing resource and ArcGIS server cluster with rapid visualization and high performance geoprocessing, the system can satisfy high concurrent access. For example, when users request the GIS services (e.g. map service, gazetteers query, etc.) from the system portal, the web server will parse the requests and transfer tasks to the ArcGIS server cluster to process the request. Normally, a GIS Server node will suffice the user's request when the workload or the number of concurrent access is small. When concurrent access or workload begin to increase, DC2 will provide compute nodes (VMs) to meet the requirements of high concurrent access. When the number of concurrent accesses decreases to be a certain degree, some GIS server compute nodes will be removed from the cluster. Such continuous dynamic adjustment for resources will be used to meet the low concurrent access, so the redundant server machines will be deleted from the cluster to achieve the most efficient using of resources. Figure 11 shows the UI of the Map World Yunnan system.

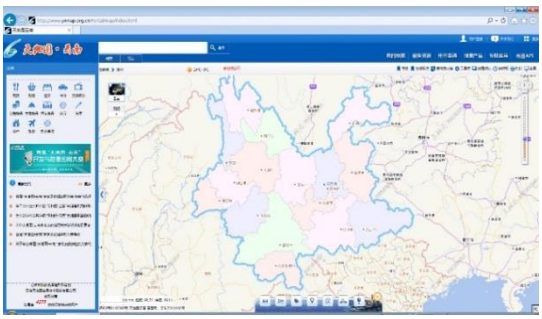

Figure 8. Map World system user interface

5. CONCLUSIONS AND FUTURE WORK

In recent years, the authoritative geoinformation agencies have built geospatial information platforms to support government decision-making and geospatial information sharing. However, with the increase of geospatial data volume and GIS applications, the platform faces great challenges, such as data-intensive, compute-intensive, concurrent-intensive, and application-intensive issues. Fortunately, cloud computing technologies provide relevant solutions for these challenges. In this paper, we propose a novel architecture for spatiotemporal cloud platform, which includes three layers (IaaS, PaaS, and SaaS) from the bottom to top. IaaS layer provides customers with VMs, virtual storage, virtual network and other infrastructure resources. PaaS layer provides the environment of development, running and testing for users. SaaS layer provides users with more than one software and application. Moreover, we design and implement serval key technologies for the spatiotemporal cloud platform. For geospatial big data management, we design the distributed storage scheme based on HDFS. For high performance geospatial analysis, we develop and chain the geospatial services based on MapReduce driven by Oozie engine. For high concurrent access, we design the auto-scaling mechanism based on ArcGIS Server and DC2. At the end, we demonstrated the feasibility by several GIS application cases. Through Big Image Data Management System, we implement effective image management and dissemination of the geospatial data for Yunnan province, China. By Hadoop-Geostatistics GIS system, we implement high performance geostatistics analysis. Furthermore, it can meet high concurrent access requirement and GIS visualization.
Future work includes the further optimization of platform architecture. Especially for IaaS layer, we 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, we 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, we 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 (e.g. Oracle spatial database, ArcGIS SDE geodatabase, etc.).

ACKNOWLEDGEMENT

We are grateful to all members of the Center of Intelligent Spatial Computing for Water/Energy Science (CISC) at George Mason University for providing Dynamic Computing Cloud (DC2) to establish spatiotemporal cloud platform for spatiotemporal statistics analysis.

REFERENCES

Ablimit, A., Wang, F.S., Hoang, V., Rubao, L., Liu, Q.L., Zhang, X.D., Joel, S., 2013. Hadoop gis: a high performance spatial data warehousing system over mapreduce. *Proceedings of the VLDB Endowment*, 6(11), pp.1009–1020.

Armbrust, M., Fox, R., Griffith, et al., 2010. A view of cloud computing. *Communications of the ACM* 53, no. 4, pp. 50-58.

Chen, Q., Wang, L., Shang, Z., 2008. MRGIS: A MapReduce-Enabled high performance workflow system for GIS//*eScience, 2008. eScience'08. IEEE Fourth International Conference on. IEEE*, pp. 646-651.

Dean, J. and Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), pp.107-113.

Del R ó S, L ópez V., Ben ítez J.M., et al., 2015. A MapReduce Approach to Address Big Data Classification Problems Based on the Fusion of Linguistic Fuzzy Rules. *International Journal of Computational Intelligence Systems*, 8(3), pp.422-437.

Douglas and Laney, 2012. "The importance of 'big data': A definition." *Gartner (June 2012)*.

Gao, S., Li, L., Li, W., *et al*., 2014. Constructing gazetteers from volunteered big geo-data based on Hadoop. *Computers, Environment and Urban Systems*.

Gong, J., Yue, P., Zhou, H., 2010. Geoprocessing in the Microsoft Cloud Computing Platform-Azure. *Proceedings the Joint Symposium of ISPRS Technical Commission IV & AutoCarto*. pp.6.

Graham-Rowe D, Goldston, D., Doctorow C, *et al*., 2008. Big

data: science in the petabyte era. *Nature*, 455(7209), pp. 8-9.

Huang, Q., Xia, J., Yang, C., et al.,2012. An experimental study of open-source cloud platforms for dust storm forecasting. *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*. ACM, pp. 534-537.

Ji, C., Li, Y., Qiu, W., et al., 2012. Big data processing in cloud computing environments. *Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks*, I-SPAN. 2012.

Lin, F.C., Chung, L.K., Wang, C.J., et al., 2013. Storage and processing of massive remote sensing images using a novel cloud computing platform. *GIScience & Remote Sensing,* 50(3), pp. 322-336.

Luo, J., Wang, D.H., Zu, X.F., et al., 2009. Popularized Spatial Information Service Platform Based on Multi-Major GIS. *Knowledge Discovery and Data Mining, 2009. WKDD 2009. Second International Workshop on. IEEE*, pp. 890-894.

Mann, K., 2011. Esri Spring 2011 Edition, *http://www.esri.com/news/arcuser/0311/cloud-gis.html*

Mell, P. and Grance, T., 2011. The NIST definition of cloud computing.

Nurmi, D., Wolski, R., Grzegorczyk, C., et al., 2009. The eucalyptus open-source cloud-computing system. *Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on*. *IEEE*, pp. 124-131.

O'Driscoll, A., Daugelaite, J., Sleator, R.D., 2013. 'Big data', Hadoop and cloud computing in genomics. *Journal of biomedical informatics*, 46(5), pp. 774-781.

Park, J.W., Lee, Y.W., Yun, C.H., et al.,2010. Cloud computing for online visualization of GIS applications in ubiquitous city. *CLOUD COMPUTING 2010, The First International Conference on Cloud Computing, GRIDs, and Virtualization.* pp. 170-175.

Provost, F. and Fawcett, T., 2013. Data science for business: Fundamental principles of data mining and data-analytic thinking.

Yang, C., Goodchild, M., Huang, Q., et al., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing. *International Journal of Digital Earth,* 4(4), pp. 305-329.

Yang, C. and Huang, Q., 2013. Spatial cloud computing: a practical approach. *CRC Press.*

Yang, C., Sun, M., Liu, K., et al., 2015. Contemporary computing technologies for processing big spatiotemporal data//*Space-Time Integration in Geography and GIScience. Springer Netherlands*, pp. 327-351.

Ye, D., 2008. The evolution of geographic information systems from my view. *Geoinformatics 2008 and Joint conference on GIS and Built Environment: The Built Environment and its Dynamics. International Society for Optics and Photonics*, pp. 714423-714423-8.

Zikopoulos, P., Parasuraman, K., Deutsch, T., et al.,2012. Harness the Power of Big Data The IBM Big Data Platform. *McGraw Hill Professional*.