ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Hybrid Cloud framework for Object Storage based Geo Spatial Remote Sensing Data Processing

Rama Naga Durga Rao Khaja, Venkateswara Rao Kota
Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Andhra Pradesh, India

Abstract

With the increase in the collection and usage of the geo spatial data by several application developers and users, the demand to deliver the geospatial data and the application processing is growing day by day. The data collected from several geo spatial sensors needs to be stored, processed and analysed to extract the value out of the data. Since the geospatial data volumes are increasing, hence there is a need to explore the better storage mechanisms and the automated processing mechanisms for the application and data delivery to the end users. In this work, we design a Hybrid Cloud computing framework for organizing the Geospatial data over Object based storages for the data preserving and a platform to quickly build and publish the geospatial services as cloud based services. This framework allows accessing the geospatial data even from public cloud storages. For processing the public cloud data, the data to be retrieved at local machine and processed by the services running on private cloud environment. The services are executed over a large scale distributed virtual machines provisioned by the underlying cloud infrastructure.

Keywords: Cloud computing, Hybrid cloud, Platform as a Service (PaaS), Web services, Geo spatial data processing, Ground Control Points (GCP)

Introduction

Geospatial data systems collect the large volumes of data from several space borne sensors and use several data processing methods to extract the information from such collected data. The data needs to be stored and indexed properly such that it could be effectively retrieve when required. Traditional file systems are limited by the metadata attributes and the volumes cannot grow to more than the peta bytes. Hence, new storage technologies are emerging, that could store the user defined metadata contents to the data. This also has advantages in the large volume storage and the data could be accessed via web service technologies. The Most often, the data which collected from several sensors and generated out of the data Object storage is a storage architecture that manages data as objects. It can store both structured and unstructured data like files, trees, images and data files etc. Objects not only include the data, it also contains information regarding itself called metadata. Object storage explicitly separates metadata from data to support additional metadata for capturing application specific or user specific information to better indexing purposes and centralized management of storage across many individual nodes and clusters.

Some differences between Traditional file system and Object file system are given as follows. Transactional file system only supports only fixed file system attributes whereas object file system support to add additional attributes to the object. Traditional file system best suited to store shared file data. Object file system suited for storing relatively static file data. Traditional file system use CIFS and NFS to access files. Object file system use REST and SOAP protocols over HTTP to access objects. So object storage is best suited for cloud storage. Open stack swift storage is one of the object storage mechanisms in which we can create, modify, and get objects and metadata by using object storage API, which is implemented as a set of Representational State Transfer (REST) web services.

A hybrid cloud is a composition of at least one private cloud and at least one public cloud. A hybrid cloud is typically offered in one of two ways: a

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

vendor has a private cloud and forms a partnership with a public cloud provider, or a public cloud provider forms a partnership with a vendor that provides private cloud platforms. In this work we implement hybrid cloud storage by integrating private cloud storage of an organization and Google map public cloud storage provided by Google map cloud provider. An implemented open stack swift storage used as a private cloud storage. To process the private data we use RESTful web service to access the data. For processing public cloud data the data need to be retrieved at the local machine.

RESTful web service architectural style, consider data and functionality as resources and are accessed using Uniform Resource Identifiers (URIs), typically links on the web. The resources are acted upon by using a set of simple, well-defined operations. REST constraints to client-server architecture and designed to use a stateless communication protocol, typically HTTP. RESTful web services maps four main HTTP methods to the operations they perform on the resources: create, retrieve, update and delete. In this work, we use REST services to access the data in open stack swift storage.

Canonical URL (CURL) is a command line tool used to transfer data between servers in form of URL syntax. CURL supports various protocols like HTTP, HTTPS, and FTP etc. By using CURL API we can access open stack swift storage. CURL API is not only used to perform operations on data but also on metadata. We can create, update and retrieve metadata by using this API. This API also supports to filter the objects based on metadata. In this, work we use this CURL API in the development of cloud services to access object storage.

Cloud services means services made available to users on demand via the internet from a cloud computing provider's server. Cloud services are designed to provide easy, scalable access to applications, resources and services. A cloud service can dynamically scale to meet the needs of its users. Cloud services are deployed on servers in virtually created machines. Virtual machines need to scale up or down to process large number of requests.

In this work, we develop data platform services and cloud services for geo spatial data processing. A case study is implemented which develops Ground Control Points (GCPs) identification module as a cloud service and data publishing, viewing data and processing data as data platform services. The GCP module is implemented as a web service and deployed on apache axis server in virtual machine. The GCP module uses an executable file which takes geo spatial image as input from object storage and results in GCP file. Now we develop data platform services for data publishing, retrieving and data processing. Using Data publishing service user can upload geo spatial image along with its metadata to object storage. This service uses Extended Java Script (ExtJS) library to develop web pages. ExtJS provides a high level of interaction with the website. Using Data retrieving service user can retrieve and view the data. User can view the data based on metadata. In this case study we used cloud percentage metadata value to filter the data. This service uses Open Layers library to display data on the map. Open Layers is an open source, client side JavaScript library that provides you with tools to develop your own web maps. Data processing service accessing the geospatial web services which are deployed on cloud to process the data.

Related work

As cloud technology becomes immensely popular among these businesses, the question arises: Which cloud model to consider for your business? There are four types of cloud models available in the market: Public, Private, Hybrid and Community. All the four models are defined, discussed and compared with the benefits and pitfalls, thus giving a clear idea, which model to adopt for your organization [1]. In this paper we used hybrid cloud model to implement cloud environment.

Platform as a Service (PaaS) is one of the key services in cloud computing. PaaS is the delivery of a computing platform and solution stack as a service without software downloads or installation. An investigation of information technology (IT) enablement for global service organization is reported with focus on feasibility of an implementation with platform as a service using cloud computing. Based on this feasibility report, an acquisition proposal for a cloud based portal with a PaaS implementation is developed [2]. A framework for managing PaaS in a virtual cloud computing lab is developed that implements the user management, resource management and access management so that the system has good expandability and can improve resources sharing and utilization [3].

Windows Azure is Microsoft's platform as a service cloud that provides on demand computation and storage to developers host, scale and manages web

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

applications on the internet through Microsoft data centres. Architecture of Windows Azure platform and different aspects of Azure based development for those who are interested in adopting Windows Azure within their enterprise IT landscape are presented [4]. Cloud computing enables web processing service (WPS) framework for Earth Observation data (EOD) processing that consists of two layers: client layer, which is the application layer where users are distributed all over the web and use WPS and WPS layer, which provides processing service with again two layers: WPS server layer, which is deployed on WPS server, receives a WPs request and returns a standard response and cloud computing layer, which provides the cloud computing environment for executing all EOD processing services [5].

A cloud computing framework [6] is developed for implementing geospatial web services to store, reuse and process geospatial data. It also includes data acquisition services that are essential especially when dealing with satellite images and applications in the area of remote sensing. A comparative analysis of geo processing [7] is provided for two cloud computing platforms- Microsoft windows azure and Google App engine. This analysis compares differences in the data storage architectural model, and developed environment based on the experience to develop geo processing services in the two cloud computing platforms.

An approach for large sized, very high resolution image registration, which is accomplished based on course-to-fine strategy and block wise Scale-Invariant Feature Transform (SIFT) matching is proposed in [8]. Course-to-fine strategy reduces the high resolution original image pairs to a low resolution level by performing direct sub sampling or pyramid decomposition. Block wise SIFT extraction and matching collects matched key point set and then removes outliers. Finally accurate transformation parameters can be estimated by the global optimization of the whole matched key points based on iterative reweighted least squares. Ground control points (GCP's) database is established for high resolution remote sensing satellite images in [9]. Firstly, the selection of GCP's on satellite imagery, secondly, the GPS observation for selected GCP positions, thirdly establishment of GCP's database and making it open to the public and finally evaluation of GCP's database.

System architecture

The system implements a hybrid cloud architecture which considers the organization cloud storage as private cloud storage and Google map cloud storage as a public cloud storage provided by Google map cloud provider. System provides the user to access both local and public cloud storages to store and access geo spatial data.

The system architecture is illustrated in figure 1. It comprises of three layers. First, compute infrastructure layer contains computing resources like servers and virtual machines (VM). Geo processing methods such as Ground Control point (GCP's) identification, Orthorectification, and DEM etc... are implemented as web services and deployed in different virtual machines which are running on various servers in cloud computing environment. Second, Data platform layer in which data platform services are implemented and deployed. Data platform services are performed operations on the data stored in cloud storage such as data publishing, data retrieval, and data processing. Data publishing uploads geospatial data into cloud storage, data retrieval retrieves and displays the data from cloud storage and data processing performs geo spatial data processing on a selected data. Data processing platform service uses the web services GCP's identification, Orthorectification, and DEM which are deployed on virtual machines in cloud computing environment to process the geo spatial data. Third, User layer provides interfaces for end users and developers. For developers it provides API for geo spatial data web services to implement their applications and for end users it provides geo spatial applications as services.

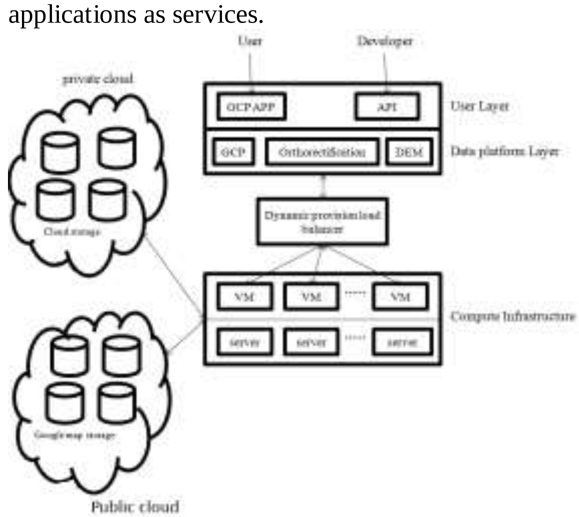

Figure 1: System Architecture

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Dynamic provision load balancer is one of the key features in cloud computing architecture. It dynamically scales up or scales down the virtual machines based on the user requests. It can also adjust the load between virtual machines.

Private cloud or cloud storage is a storage structure that centralizes the data stored by all users. User can access the data anywhere from the internet or intranet. In this paper we used object storage structure as cloud storage. This structure stores data as objects. We can also store information about data within object called metadata. This metadata can helpful to search the required data in retrieval process. In this paper we use this metadata in the implementation of data retrieval platform service to get the required data from object storage.

Public cloud is provided by Google map cloud providers. We use the same methodology for accessing Google map cloud storage like in private cloud. This system supports both storing and retrieving the geospatial data from Google map cloud storage.

Figure 2 illustrates the accessing of published services in cloud computing environment. The geo spatial data services are deployed in virtual machines on servers. The structure provides API's and interfaces to access the services in cloud computing environment. Developers can access the web services through API's to develop their applications as services. A third party authentication system always monitors the user whether he is authenticated or not by using multi-level authentication.

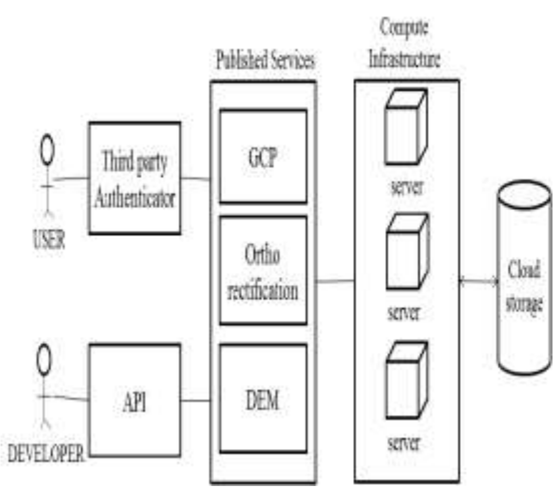

Figure 2: Platform for accessing published services

Case Study

In this paper we implement a case study for the geo spatial processing module Ground control point identification which automatically extracts the ground control points from a selected image. Ground control point is a known point on surface of the earth. GCP's identification module uses multilevel block wise scale invariant feature transform (SIFT) matching algorithm to extract points by comparing the input raw image with a referenced image.

First we created a virtual machine using Xen server virtualization. A web service named GeoServices is implemented which contains geo spatial data processing methods as operations. Here we implement GCP's identification module as operation in the web service GeoServices. This operation takes the raw image as input, and then extracts the GCP points to a file and finally stores that file into object storage. The implemented web service is deployed on Apache axis server in virtual machine.
Open stack swift storage structure is an object storage structure. In this paper we are using already installed open stack storage as private storage. The geo spatial data collected from several external sources are stored in this cloud storage in the organization. We used CURL API to access the open stack swift storage.

Google map cloud storage is used as a public storage. In this paper Bing Maps API is used to access the data from Google map storages. Bing Maps API offers developers access most of the mapping and satellite images that are available through their main map site. By using object stack swift storage as a private cloud storage and Google map cloud storage as a public cloud, it presents a hybrid cloud structure.

Now we implement data platform services data publishing, data retrieval, data processing. Data publishing is uploads geo spatial images and metadata. In this paper we adding metadata along with the image is date, satellite ID, sensor, segment, orbit, top left latitude, top left longitude, top right latitude, top right longitude, bottom left latitude, bottom left longitude, bottom right latitude, bottom right longitude and cloud percentage. We used extended java script (ExtJS) to provide a better interaction with the user for data publishing. Data retrieval service extracts the data from object storage. User can retrieve entire data or data filtered by using metadata. In this paper we used cloud percentage to filter the data. In the implementation of data retrieval service we used open layers API to provide better

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

interaction with the usage. Data processing service acts as client to apache axis server for calling GCP's identification operation in the web service GeoServices that is deployed in apache axis server. The operation results a GCP file which can be downloaded by the user from object storage. In this paper we use a Third party authentication system that has 2 levels, login credentials and face recognition system to authenticate the user.

Performance evaluation

To evaluate the performance of the Cloud services, we have conducted two experiments. One will evaluate the performance in time by considering the number of jobs (user requests) with constant number of virtual machines. Other will evaluate the performance in time by considering number of virtual machines with constant number of jobs.

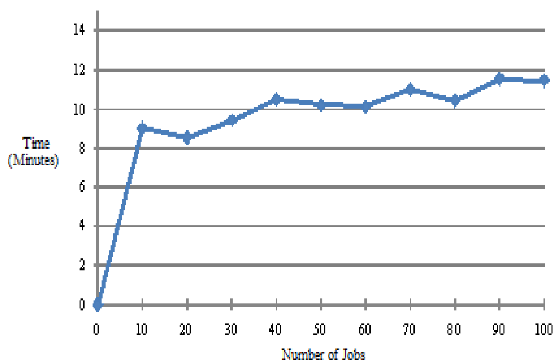

Figure 3: Performance test with constant number of virtual machines with different number of user requests

Figure 3 evaluates the performance of cloud computing framework by considering the user requests. It contains hundred virtual machines as constant value for this evaluation. Each virtual machine can process three jobs at a time. The input data (number of jobs) is increased from 0 to 100. For 10 jobs, 4 virtual machines will be allotted. Three jobs are allotted for first three virtual machines and one job for fourth machine. Now load balancer adjusts the load by reassigning one job from any of the first three virtual machines to fourth one. One virtual machine will complete the 3 jobs in approximately in 9.05 minutes. So for 10 jobs it will take 9.05 min. This process is same for any number of jobs. It can be seen that, as number of jobs to be handled are increasing, time taken for processing jobs is also increasing.

Figure 4 evaluates the performance of cloud computing framework by considering varying number of virtual machines. It considers hundred jobs as constant value for this evaluation. Each virtual machine can process three jobs at a time. The input data (number of virtual machines) is increased from 0 to 100. For 1 virtual machine it will take approximately 320 minutes. If we increase the virtual machines count to 10 then for each virtual machine 3 jobs will be assigned. It takes 9.05 minutes to complete the assigned jobs. The next 60 jobs completes in 18.10 minutes. The last 10 jobs assigned to 10 virtual machines and remaining machines shutting down automatically. The last 10 jobs will executes approximately in 9.10 minutes. Therefore 10 virtual machines execute 100 jobs approximately in 31.56 minutes. This process is same for any number of virtual machines consider in cloud environment. It can be seen that, as number of virtual machines are increasing, time taken for processing jobs is decreased.

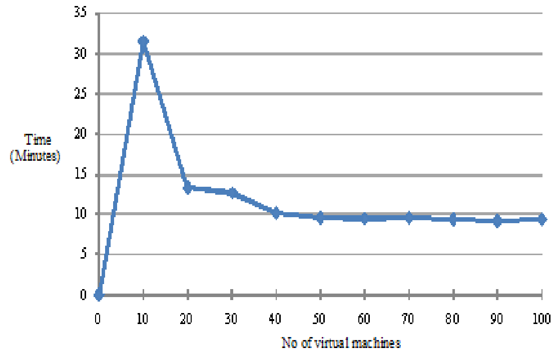

Figure 4: Performance test with constant number of jobs with different number of virtual machines

Conclusion

A better framework is proposed in this project to deliver and process the geo spatial data over object based storages in cloud computing environment. Organization vendors can store their private geo spatial data in private cloud storage and can access satellite images in Google map cloud storage. This framework allows any number of services to be deployed flexibly in cloud environment. It provides APIs to the developers for easily developing their web applications. It also provides green computing by automatically shutting down the virtual machines which are in idle state. In the future work, the developers can implement the community cloud in which group of organizations can access the geo spatial services published by any organization.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

References

[1] "Public vs Private vs Hybrid vs Community-Cloud Computing: A Critical Review", Sumit Goyal, Member, IDA, New Delhi, India, I.J. Computer Network and Information Security, 2014, 3, 20-29
[2] "Feasibility of a Platform-as-a-Service Implementation using Cloud Computing for a Global Service Organization", Keke Gai, Annette Lerine Steenkamp, 2013 Proceedings of the Conference for Information Systems Applied Research ISSN: 2167-1508 San Antonio, Texas, USA.
[3] "Implementing and Managing framework for PaaS in Cloud Computing", Junaid, Qayyum, Faheem Khan, Muhammed LaL, Fayyaz Gul, Muhammed sohaib, Fahad Masood, Gandhara University of Sciences, Peshawar, 25000, Pakistan, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.
[4] "Windows Azure PaaS Cloud: an Overview", Rabi Prasad Padhy, Manas Ranjan Patra and Suresh Chandra Satapathy
[5] "Cloud Computing Enabled Web Processing Service for Earth Observation Data Processing", Zeqiang Chen, Nengcheng Chen, Chao Yang, Liping Di, Applied Earth Observations and Remote Sensing, VOL. 5, NO. 6, December 2012.
[6] "Geo Spatial services in cloud", Konstantinos Evangilidis, Konstantinos Ntouros, Stathis Makridis, Constantine papatheodorou, Computers and Geo Sciences, volume 63, February 2014,pages 116-122.
[7] "Geo processing in cloud computing platforms- A comparative analysis", Peng Yue, Hong Xiu Zhou, Jianya Gong and Lei Hu, International Journal of Digital Earth ,volume 6, issue 4 , 2013.
[8] "Multilevel SIFT matching for Large-Size VHR Image Registration", Chunlei Huo, Member, IEEE, Chunhong Pan, Leigang Huo, and Zhixin Zhou, Multilevel SIFT matching for Large-Size VHR Image Registration
[9] "Establishment of Ground Control Point Database for Satellite Remote Sensing",Kojima Mitsuhiro, Masya Tanouchit, Kyaw Sann Oo and Masataka Takagi , Kochi University of Technology Kami-shi, Kochi, 782-8502, JAPAN.