CS 694. SEMINAR.

CLOUD COMPUTING AND GIS: FEATURES AND APPLICATIONS

by
Bharat Singhvi [10305912]

Under Guidance of
Prof. Umesh Bellur

Department of Computer Science and Engineering
Indian Institute of Technology, Bombay
April, 2012
# Contents

1 Background Study: Cloud Computing

1.1 What is cloud computing? .................................................. 3

1.2 Basics of Cloud Computing ................................................. 4

1.2.1 Architecture of Cloud Computing .................................. 4

1.2.2 Categories of Cloud Services ....................................... 4

1.3 Cloud Computing: Merits and Issues ................................. 5

1.3.1 Merits of Cloud Computing ......................................... 5

1.3.2 Issues with Cloud Computing ...................................... 6

2 Literature Survey

2.1 Paper Reviews ............................................................... 8

2.1.1 CLOUD-BASED SERVICE FOR BIG SPATIAL DATA TECHNOLOGY IN EMER-GENCE MANAGEMENT .................................................. 8

2.1.2 Cloud Computing: A Solution to Geographical Information Systems (GIS) ... 9

2.1.3 Geoportal - A Spatial Cloud Information Service .................... 10

2.1.4 SPATIAL CLOUD COMPUTING ..................................... 11

2.1.5 Massive spatial data processing model based on cloud computing model ... 13

3 Cloud Based GIS Systems ................................................... 15

3.1 Current GIS Systems ....................................................... 15

3.2 Characteristics of GIS Systems ......................................... 15

3.3 Features of a cloud based GIS system ................................ 17

3.4 GIS Systems: Applications .............................................. 17

3.5 Why Cloud based GIS? ................................................... 18
3.6 Case Study 1. OpenGeo Architecture .................................................... 19
  3.6.1 OpenGeo web mapping architecture ............................................. 20
  3.6.2 OpenGeo in the cloud ................................................................. 20
3.7 Case Study 2. ESRI in the Cloud ....................................................... 21
  3.7.1 Proposed Architecture for deploying ArcGIS Server on Amazon EC2 .... 21
  3.7.2 ArcLogistics and Business Analyst in the cloud ............................. 22
3.8 Case Study 3. www.giscloud.com ...................................................... 22
  3.8.1 Features ...................................................................................... 23

4 Architecture Design for Cloud Based GIS .............................................. 24
  4.1 Public Cloud Based GIS System Architecture ................................. 24
    4.1.1 Model ......................................................................................... 24
    4.1.2 Pros and Cons ........................................................................... 25
  4.2 Private Cloud Based GIS System Architecture ............................... 25
    4.2.1 Model ......................................................................................... 26
    4.2.2 Pros and Cons ........................................................................... 26

5 Conclusions and Possible Extensions .................................................. 28
Abstract
A GIS system is an information system which has tools to collect, manage, analyze and display geographically referenced data. GIS plays an essential role in a wide range of areas. As an example, Ushahidi is an open source web based GIS system which was developed as a disaster management application and is a crowd-sourced collaborative platform. GIS can be seen as the blend of cartography, statistical modeling, software, hardware and data. However, fully capable GIS systems are expensive. Moreover, management of the large volume of associated data is another characteristic of spatial analysis. For providing a highly available GIS services to users distributed across the globe at the expense of less resources, a cloud computing based platform to host and develop GIS applications is needed. Cloud computing enables the users to harness abstracted and virtualized resources and permit computations over huge amount of information without having their own processing power. In this study, an attempt has been made to specify the features of a cloud based GIS system and architectural design for cloud based GIS systems has been proposed.

Introduction
Cloud computing is a novel deployment model which has emerged recently. Strongly supported by the pioneers in IT industry, it provides an interoperable way of providing and sharing services such as computational resources or data storage over internet etc. Cloud computing is quite different as compared to other distributed computing paradigms like grid computing [6] which uses the resources of several loosely coupled, heterogeneous and geographically spread systems in a network to a single problem at same time while maintaining strict admission and access control policies over its members. Another paradigm is high performance computing or HPC which uses supercomputers and computer clusters to solve problems which require high computational power. Cloud computing, on the other hand, takes advantage of economy of scale by using a large number of computers to distribute the computing power and data storage over the networks and offers easy access to these resources. Cloud computing offers flexible configurations which allow the user to increase or decrease the occupied hardware resources dynamically, depending on the real time demands of hosted application.

A geographic information system (GIS) integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information. A GIS system allows us to analyze and interpret the data in ways that reveal relationships, patterns, and trends. These analysis can help us in answering questions and potentially solve problems in various related fields like oceanography, land-use, natural disasters etc [2]. GIS applications generally include data acquisition and pre-processing from multiple sources which is followed by exhaustive spatial analysis. The same data is hosted in different locations and it is needed to be processed in the same way
many times when used by different users. In many cases, the data processing and spatial analysis over the acquired data requires expensive investment in terms of hardware, software and personnel training. Thus the two major challenges associated with a GIS system are:

1. Data acquisition which is usually expensive and time consuming and
2. Software and hardware needed to run GIS applications which are usually costly and also require an expertise for use,

which makes it quite impractical for occasional users to operate a GIS system. It is evident that cloud computing offers a solution to these challenges as it provides a solution to host large volumes of data as well as provide powerful computing resources, yet GIS presents a challenge for cloud computing based platform. Firstly, the high data volumes can make the data transfers between cloud infrastructure and end users extremely slow. Secondly, the algorithmic nature of spatial processes demand development of algorithms for spatial processes specifically suited for the distributed domain.

This study is an attempt to review the cloud computing technologies and propose some prototype system designs for cloud computing based GIS systems. Chapter 1 presents the background study on cloud computing. Chapter 2 is a compilation of technical critiques on the literature survey done to investigate the open research areas related to cloud based GIS systems. In Chapter 4, an attempt has been made to emphasize the need of cloud based GIS systems and define them. Three case studies have been performed to compare the existing cloud based GIS solutions with the proposed definition. Chapter 5 presents possible architectures for cloud based GIS systems.
Chapter 1

Background Study: Cloud Computing

1.1 What is cloud computing?

While the idea of cloud computing is not really new, the term cloud computing itself is relatively young and therefore it is not surprising to see little consensus among the experts on the definition of the term cloud computing. There are a large number of published studies which introduce the notion of cloud computing, each in a different style. Authors Zhao and Foster in their paper [6] define cloud computing as “A large scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the Internet”. In paper [12], the definition is “A computing cloud is a set of network enabled services, providing scalable, QoS guaranteed, normally personalized, inexpensive computing platform on demand which could be accessed in a simple and pervasive way.” A more general view is provided in [4] which defines cloud computing as “Cloud Computing refers to both the application delivered as services over the internet and the hardware and systems software in the datacenters that those services. The services themselves have long been referred to as the Software as a Service (SaaS), so we use that term. The datacenter hardware and software is what we will call a cloud.”

In order to form a complete overview about the term cloud computing, it is necessary to identify the key attributes associated with it. Some of the attributes which can be inferred are:

1. Service abstraction and encapsulation. This means that the system in the cloud which is used to deliver the service to the end user is abstracted from the end user. The users would not know and neither would there be a need to know that which particular system in the cloud is serving their requests. Any information pertaining to the technical specifications of the serving system would be kept hidden from the user and neither would it be of any interest to the user.

2. Isolation in delivered services. Cloud computing model adopts virtualization to create several
OS and storage blocks over the same physical machine or disk. Each user is provided these virtual machines. The users using the virtual machines would not even notice that other users are using the same piece of hardware which brings in fault isolations among the virtualized users.

3. Dynamic scalability of the system. This attribute allows the user of cloud technology to have dynamic resource allocation based on the current demand of the user. This means that the users do not have to engineer the solutions for handling peak loads and the resource allocation is taken care of dynamically by the model itself.

4. Pricing on utility computing basis. This means that the providers of the cloud technology monitor per user usage of the services and charge them accordingly which reduces the cost by converting capital expenditure to operational expenditure.

1.2 Basics of Cloud Computing

1.2.1 Architecture of Cloud Computing

Cloud architecture, the systems architecture of the software systems involved in the delivery of cloud computing, typically involves multiple cloud components communicating with each other over a loose coupling mechanism such as a messaging queue [1]. One of the widely discussed cloud architecture is presented by Zhao and Foster [6] which consists of four layers:

1. Fabric layer which basically represents the entire hardware which comprises the cloud computing infrastructure.

2. Unified resource layer which is a service layer which provides the users with abstracted resources like file systems, virtualized operating systems and storage spaces. The services provided by this layer can be used to serve upper service layers or the end users.

3. Platform layer, sits on top of the unified resource layer, provides the tools, middleware and services to developers for development and deployment. The platform layer services can be considered as an Application Programming Interface [API] in most cases.

4. Application layer is the top layer in the cloud computing architecture and it provides applications and services which can be executed in the cloud by the end users. These applications and services are built based on the lowers layers like the platform layer or the unified resource layer.

1.2.2 Categories of Cloud Services

Today, cloud computing offers a wide spectrum of services ranging from simple web log processing to heavy scientific computations. In general, most of the cloud computing services fall into one of the following three categories: [6, 9]
Infrastructure as a Service (IaaS): It is a provision model which provides computational infrastructure. This means that virtual servers, desktops, storage units, network resources etc. are delivered as a service over a network. One example of IaaS is Amazon’s Elastic Cloud Computing [EC2] service and Simple Storage Service [S3]. A typical EC2 service to the user is to allow the user to create and host virtual machines over the cloud.

Platform as a Service (PaaS): It provides an Application Programming Interface for the users to develop and deploy applications and services. In the PaaS model, cloud providers deliver a computing platform and/or solution stack typically including OS, programming language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers. An example of platform as a service is the Google App Engine API.

Software as a Service (SaaS): In this model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. The software is delivered to the end users over network. An example of SaaS provider is Salesforce which provides CRM software and services to end users.

1.3 Cloud Computing: Merits and Issues

1.3.1 Merits of Cloud Computing

Some of the advantages of cloud computing technology are:

1. Availability: Cloud computing technology is based on enormous infrastructure which generally resides in geographically distributed locations. This means that if a part in a cloud infrastructure gets faulty, it can easily be replaced by a backup hardware. This makes the cloud based services highly available by cutting the costly downtime and recovery.
2. Scalability: With cloud computing, users are allowed to scale the resources allocated to their application dynamically. This means that rather than paying for additional infrastructure to handle peak time load, the users can quickly scale the resources in a cloud environment by avoiding the risk of idling the infrastructure when the demand is low.

3. Reliability: Major cloud providers redundantly back up the data stored in their cloud infrastructure which makes the cloud based services highly reliable in cases of disaster recovery. An example is Microsoft Azure which keeps the backup copies in different data centers.

4. Economy of scale: If a comparison is made between having an in-house computing infrastructure with the one over cloud, it is evident that the user has to spend much less in the cloud based services.

5. Efficient application/data update model: In a cloud environment, the user do not have to worry about updating the applications as compared to the normal scenario when the application is present on the personal system of the user since everything is maintained on the cloud server side. Similar is the case for data in the cloud environment.

6. Easy to share and distribute: Cloud infrastructure can provide massive network throughput which allows the data distribution to large number of users across the globe.

7. Better utilization of resources: In a traditional scenario, only certain servers run at high demand. Moreover, the demand fluctuates with time. This not only results in the waste of electricity, but also the waste of hardware resources. Cloud computing provides resource virtualization which maximizes the utilization of available resources. Same hardware is shared by a number of users and resource allocation is managed to serve the demand rates.

8. Data volume friendly: Cloud computing is built on massive shared infrastructure which is capable of handling large volumes of data. Some cloud computing based frameworks have been developed specifically to deal with high data volumes. [8]

1.3.2 Issues with Cloud Computing

1. Privacy and data security: This is one of the major issues with cloud based environments. Because of the abstraction in services, the users do not know where their data is getting stored and who would possibly have the privilege to access the data inside the cloud infrastructure. Also, because of the geographic distribution of data centers, the data stored in these data centers are protected or unprotected according to the local jurisdiction which raises a privacy issue.

2. Performance Stability: There may be times when the cloud infrastructure is under heavy load. The performance of an application in the cloud depends on network traffic and also on the resources utilized by the other virtual machines running on the same physical machine as the host application. This may result in performance instability.

3. Service provider switching flexibility: Different cloud providers offer different level of services and APIs which makes it difficult for the users to switch from one cloud provider to another. This
has also been referred to as the lock in problem in some of the studies surveyed. In general, IaaS provides easier possibilities in switching than PaaS which SaaS is the most difficult to switch from.
Chapter 2

Literature Survey

2.1 Paper Reviews

2.1.1 CLOUD-BASED SERVICE FOR BIG SPATIAL DATA TECHNOLOGY IN EMERGENCE MANAGEMENT

Authors: Xiaosan Ge*, Huilian Wang

Paper Summary
1. An attempt to identify the characteristics of big spatial data which makes it suitable to use cloud technology to manage it.
2. Description of drawbacks in existing emergence management systems. [Lack of data acquisition and processing system planning, not real time.]
3. Explains the advantages of using cloud based services for managing spatial data [on demand, scalable, reliable, cost effective] and explains the flow of spatial data in emergence system.
4. Paper can be classified as the one where the authors attempt to bring out the features of spatial data which makes it important to enable processing platform for this data as a service.

Technical Critique
While the paper explains the advantages of cloud computing scenario and the characteristics of big spatial data, it fails to clarify the road of union for the two. The only explicit sentence in the paper that tries to justify the use of cloud computing technology is for better network infrastructure. While the authors have described the steps of data flow in an emergence system, it is not clear as to how will this be different in a cloud based environment.

The major contribution of this paper is exploring the features of big spatial data, like complex structure, correlation in closely located data, temporal patterns, huge amount of data to be processed, the computing resources required to do so and how the data can be scaled. The paper then identifies the characteristics of cloud computing technology which supplement these features. Finally, the authors identify the process which may be followed in building an emergence system.

The paper lacks the support of experimental data. However, it is understandable that the issues being addressed are of theoretical nature and the description of application flow as described in the
2.1.2 Cloud Computing: A Solution to Geographical Information Systems (GIS)

Authors: Muzafar Ahmad Bhat, Razeef Mohd Shah, Bashir Ahmad

PAPER SUMMARY
1. After a short background on cloud computing, the paper moves discuss the term GIS Cloud, which has been defined as an approach to provide GIS applications as a service to users around the world.
2. The paper identifies the principles in GIS Cloud and explains them in excruciating detail. These principles can be thought of as the key features of a typical GIS cloud application. They include:
   a. Application infrastructure. Web services and application that enable easy usage of the data.
   b. Support technology infrastructure. The applications which allow creation and updating of data as infrastructure.
   c. Location independent resource pooling.
   d. Interoperability support. Applications which allow the users to share data in different technology across different platforms. Flattening the diversity in nutshell.
   e. Maintenance service.
3. The authors propose architecture for GIS cloud applications. They have broadly categorized it into two components: thin client based interaction interface and the server side.
4. While the web interface for the users has the basic requirements of fulfilling the features discussed above, the server side has the traditional requirements pertaining to security, robustness, cost, maintenance and time abreast. The server side architecture has been proposed where the key components are:
   a. Communication Layer
   b. Repository Layer
   c. Utilities Layer
   d. Logic Layer
   e. Configuration Layer
Each of these layers is very relevant. A simple analogy would be the interaction between the configuration layer and all other layers on the fly where each user has customized the interface side. While all of this is seems very natural, the point to note is the development aspect associated.
5. The purpose of each layer and communication requirements between them have been discussed. An overall flow diagram for the architecture has been proposed.

TECHNICAL CRITIQUE
This paper provides an insightful overview of the requirements in any cloud based system. The specific configuration for GIS services arise from the features of GIS applications. Combining these two, the authors have proposed an underlying architecture to power a GIS Cloud. While the paper does not discuss any implementation, it definitely stands out in the detail it provides to explain what it takes to build a system in a limited word-space.

Developing a GIS Cloud requires a basic architecture to thrive on. The architecture proposed in the paper is loosely coupled and relates to everything which an application must provide the users. This solves the problem of what should be architecture that may be followed while implementing a GIS Cloud.

One way that I think of if I were to write a GIS cloud application [without worrying about the hardware configuration] is to have an asynchronous communication between client and server where the server gathers the activities of client and generates a possible set of operations that the user may request. So, another layer say a proactive layer should be present at the server side which minimizes the response time to a request being made.

However, the disadvantage of this study is the robustness which the architecture lacks. A single layer failure would lead to a total crash. Secondly, the architecture does not provide any scope for storing the processing data. There should be another layer where the already processed data is lying, similar to the cache concept, except that this one is more of a permanent unit.

One of the interesting works that can be done is defining the implementation for proactive layer as mentioned above. This will require the analysis of features of spatial data and any algorithms that may be written to harness the specificity for this implementation.

2.1.3 Geoportal - A Spatial Cloud Information Service

J Srinivas, K Kanti Kumar, B. Sankara Babu, N. Subhash Chandra, G. Charles Babu

PAPER SUMMARY
1. Introduces the term spatial cloud computing [SC2]. SC2 can be understood as a technology which provides geographic intelligence as a web service. Geographic intelligence in turn may refer to information processed from spatial data specific to the enterprise using the service.
2. Authors present advantages of using SC2. These advantages are the cost efficiency of using the technology while getting spatially enabled.
3. Proposes a geoportal which essentially is a web based service that provides visualization of data on maps along with a data viewer which enables querying the underlying data and a document viewer which manages the content used by the data viewer and visualizer.

This paper attempts to describe a simple application which can enable an organization to geo-tag their data without going into the hassles of an enterprise solution to do so.
CHAPTER 2. LITERATURE SURVEY

TECHNICAL CRITIQUE
This paper identifies the problem described subsequently and attempts to provide a solution to it. The problem, as I understand, is given a company which provides different kinds of solutions wishes to display this information on a map customizing the information depending what solutions are provided where and then the associated details, but does not want an enterprise solution like Oracle Spatial or ArcGIS to host and maintain this data and neither wishes to spend development time for it how should it spatially enable itself without paying much? The solution is intended to be the geoportal described in the paper.

While the advantage to proposed solution is clearly the cost, there is a serious disadvantage as far as privacy is concerned Geoportal is a third party application hosting the contents of several organizations. If one chooses to use it as a solution, privacy is compromised.

The solution would have been more convincing if the architecture for geoportal would have been described. It looks like a content management system with no discussion on security aspect.

The paper is introductory. It lacks research on the feasibility of proposed solution and its implications.

However, it would be interesting to see what becomes of such an application if one is built.

2.1.4 SPATIAL CLOUD COMPUTING
Authors: Yang, C., Goodchild M., Huang Q., Nebert D., Raskin R., Xu Y., Bambacus M., Fay D.

PAPER SUMMARY
Part 1. Intensity challenges
Paper explains the challenges which the geoscience domain poses for information technology. The challenges are in terms of
1. Data. Massive data coming from different sensors needs to be stored which raises the storage problem and also the same data needs to be processed to find information which requires huge processing resources which brings us to the next point.
2. Computing. Computing not only requires resources but also the algorithms and applications that can process the data.
3. Concurrency. This refers to the availability of service to a large number of users. Millions of users may access a service concurrently which requires the processing to be fast and application to be robust.
4. Spatio-temporal. All geosciences applications are Spatio temporal in nature and which requires the application to be aware of dynamicity.

These challenges are fundamental to geoscience applications. This answers the basic question of why use cloud computing for an application. Such intense applications need huge infrastructure for
CHAPTER 2. LITERATURE SURVEY

working and cloud begins by providing the same. There are several other features of cloud computing form the basis of using this as a platform to provide GIS solutions.

Part 2. Cloud Computing and Geo-Sciences

Cloud computing is essentially providing computing as a service. There are four types of services through which cloud computing can be provided: Infrastructure as a Service, Platform as a Service, Software as a Service, and Data as a Service. If we look at different components of geo-spatial sciences, then based on their characteristics, we may employ cloud computing services to make the most of them. Some of them may be:

1. Earth Observation [EO] Data Access. Can be implemented as DaaS.
2. Parameter Extraction. Can be implemented as PaaS as this involves using the EO data for processing.
3. Knowledge and Decision Support can be implemented as SaaS.

This leads us towards the term Spatial Cloud Computing and explaining as to how the challenges in part 1 are actually met.

Part 3. Spatial Cloud Computing [SC2] and SC2 scenarios

SC2 simply refers to the computing paradigm for the applications which are specific to geospatial sciences. This means that the challenges briefed in part 1 are answered by SC2. In this paper, the authors have created a scenario corresponding to each of the challenges and implemented [or are implementing] it over a cloud environment. Each of the scenarios is briefly listed here:

1. Data Intensity Scenario. The massiveness of the data and its distribution globally calls for the use of DaaS.
2. Computing Intensity Scenario. This has been tested as PaaS.
3. Concurrent access scenario. This has been tested as IaaS.

The results of the tests conducted lead me to conclude that the implementation has been quite successful. The response time of the compute scenario is very promising and the results of concurrent access scenario prove that SC2 offers a viable solution for providing geo-intense services.

Part 4. Discussion of issues

The concluding quarter of this paper is dedicated for discussing issues that may arise in use of SC2. The obvious one is security. To address this, a security requirement baseline has been proposed and any SC2 service must comply by that baseline. This baseline includes responsibilities for:

1. Functionality and availability of cloud services.
2. Provision for solutions to protect data loss.
3. Users can access and control only their own jobs.
4. Identity management system for different type of users.

The other issues include:

1. Privacy. Ensuring protecting the data and services deployed over internet.
2. Trustworthy. Data and services provided are legitimate.
3. Ethical. Use of location technologies must ensure ethical sharing of information without breaching individual privacy.

This paper offers a complete understanding of the term spatial cloud computing. It takes a step ahead to identify the opportunities in SC2 and the related social concerns.

### 2.1.5 Massive spatial data processing model based on cloud computing model

Authors: Dong Cui, Yunlong Wu, Qiang Zhang

**PAPER SUMMARY**

Paper Classification: Talks about cloud computing for computational power. PaaS.

1. The paper begins by justifying the need of using a cloud model by pointing out that with the advent of satellite technology, the amount of data received is too big for the traditional means to carry out the required processing as it is based on a stand-alone machine which performs this processing. [Processing like classification of received data which is time intense.]
2. The authors propose a cloud computing model for processing this data. The model proposes using a workstation responsible for distributing the received data to processing unit which itself is a cloud. The processing units after performing the required actions would store the data to central storage system.
3. The authors also propose a simple workflow for the machines under this architecture. According to this model, the distribution units create a task list and then try to connect to various processing units delivering the work to them upon availability. Each of the systems in the processing unit then perform the required actions, store the image in central storage and informs the distribution unit that assigned the task has been completed and that it is available for the next task.
4. Further, the paper proposes a model for spatial data processing model for image processing of remote sensing data. The experimental results in the paper suggest that the proposed model works better than the traditional technique of performing processing.

**TECHNICAL CRITIQUE**

**Positives:**
1. The cloud model has actually been experimented.
2. Attempts to characterize a generic model for processing remote sensing data.

**Negatives:**
1. The paper does not describe anything about the mode of communication between various units in the cloud architecture. How the data is transferred from the distributing unit to the processing unit? Over TCP/IP is mentioned, but associated risks have not been addressed.
2. In the proposed scheme, it has not been mentioned how the processed data would fit together? Using several machines in the processing unit means using lot of machines to process the data being
received instead of a single stand-alone machine. Is it really a cloud model?

3. Experimental results are not comparative and hence less convincing.

Scope for further work:

One of the interesting problems that can be addressed is how to make distribution of tasks intelligent. If processing units has n systems and there is different kind of processing requirements on same image the design of the processing unit itself is a problem. Specific algorithms can be loaded on specific machines suited to run that algorithm. The task distribution unit can then be intelligently designed which is aware of individual capabilities of machines in processing unit and assigns task accordingly. An empirical study can then be performed to understand the frequency of different processing requirements and based on the number n, how many machines should be capable of executing more than 1 algorithm for efficient task distribution. This would make the architecture sound and scalable.
Chapter 3

Cloud Based GIS Systems

3.1 Current GIS Systems

Currently, GIS systems can be classified into three categories based on how they distribute the services, which are: Desktop GIS, Client/Server GIS and public internet GIS.

1. Desktop GIS: Here, the computing software as well as the data storage units are located in the computer of the end user. No network is necessary for this type of GIS system. Example of desktop GIS are ArcGIS Desktop, IDRISI, MapInfo, QuantumGIS etc.

2. Client/Server GIS: In this model, most of the storage space resides on the server side while the data processing can happen on both the client and server side depending upon the processing operations. This model requires atleast an intranet to function. The users are from a specific group which his authorized to access and use the resources on a server. Client machines are used to display data and perform queries and other analysis. An example of a client/server model based GIS system is ArcGIS Server with ArcGIS Explorer. However, it is evident that although client/server model based GIS systems can be used to deliver services and data over internet, they are not scalable to a large number of users due to software constraints.

3. Public Internet GIS: In this model, both the storage and the computing units reside on the server side. GIS systems in this category generally provide only the facility of data visualization and limited operations such as attribute queries. However, such systems are usually scalable to large number of users. Google Maps, Bing Maps and Google Earth are the examples of public internet GIS systems.

3.2 Characteristics of GIS Systems

According to [10], “A geographic information system is a computer based information system that provides tools to collect, integrate, manage, analyze, model and display data that is referenced to an accurate cartographic representation of objects in space.”

Given this definition, it should be pointed out that GIS system posses certain characteristics which
distinguish them from other information systems. In this section, these characteristics have been discussed.

1. **Production of Maps.** GIS are designed to support production of maps. There can be several ways in which maps can be made. For example, on way to produce them is by digitizing paper maps. Another commonly used technique is to mosaics the aerial photographs or satellite imagery for map generation. Collection of co-ordinates using a GPS device is yet another method. These techniques are labor intensive and they need specialized training and equipment which makes the process of data acquisition costly. As a result, the cost of data for the end user is quite high.

2. **Spatial Database Management Tools.** GIS systems are used to collect and manage spatially defined data. The term geocoding refers to the process of linking attribute data with the geo-referenced co-ordinates on the map. This process creates attributes like latitude and longitude which are of spatial nature. The data can also have associated non-spatial attributes. The task of a GIS system is to be capable of defining and managing this kind of data.

3. **Support for Query on Spatial Data.** GIS system should be capable of query processing over spatial data. Both spatial and attribute queries can be used to answer questions like: “What is located at this intersection?” or “What is the nearest facility from this location?”

4. **Visualization of Spatial Data.** GIS system should be capable of displaying the spatial data. In particular, it should be possible for the system to display more than one set of data simultaneously by overlaying different data sets. This overlay should be compatible with different projection systems. It should be possible to create new layers which contain relationships in the overlayed data. For example, consider an application for overlaying the market boundaries and the location of prospective customers. Then, it should be possible for the GIS system to store these two layers separately, allow the visualization by overlaying the two layers and create a new layer which marks the suitable areas for optimal location for a new facility.

5. **Spatial Analysis.** GIS system must contain statistical tools and data manipulation functions which can be used to implement models and transform the data according to the user requirements. Spatial analysis is quite similar to the decision modeling capabilities in decision support systems, enabling the users to perform sophisticated ‘what if’ analysis. For example, it should be possible to perform analysis in order to answer queries like: “How many people will look at an advertisement placed at some location X in a single day?” Some other examples of the required spatial analysis support can be related to transportation where path planning or routing strategy related queries can arise, environmental management where modeling of impact of erosion rates may be required or marketing analysis where the proximity modeling is of high importance.

6. **Geo-editing of Spatial Data.** The GIS system must be capable of editing the existing data and saving the changes made to it. Data editing means that it should be possible for users to collaborate new data, correct the existing data or perform data analysis and create new data out of it.
3.3 Features of a cloud based GIS system

A GIS system which is built on top of a cloud computing infrastructure should have the following attributes:

1. It should be able to use the cloud infrastructure to dynamically scale its storage and computing requirements.
2. Should be able to provide parallelized services for serving various user bases. The users can be authorized users or public users.
3. The services should be capable of providing a set of GIS functionalities corresponding to the characteristics of GIS systems discussed in the previous section. These functionalities are:
   a. Spatial Imaging which makes the system capable of representing data and information with respect to a co-ordinate system.
   b. Database Management function which enables the system to store, manipulate and query the data.
   c. The Decision Modeling function which allows the use of analytical tools that can be used to support decision making.
   d. Data visualization and editing function which allows the user to view and change the existing data.

3.4 GIS Systems: Applications

The functionalities of GIS systems can be used in several applications. Some of them have been discussed in this section.

1. Surveying and Mapping. Commonly known as automated mapping, it represents one of the very early application of GIS systems. It enables the organizations to generate spatial data. Moreover, remote sensing and GPS can be used to generate the maps with higher accuracy. However, acquisition of this data can be very costly. According to [11], the cost of data can exceed 20% of the cost of a GIS implementation. Moreover, errors can arise because of several reasons like problem with positional accuracy, improper classification of attributes, incompleteness of spatial objects, i.e. not all the relevant objects are present on the map. This requires the process of data cleaning which requires domain expertise.

2. Facility Placement/Management. GIS systems are capable of providing tools for real time monitoring of resources and facilities. In the FM applications, the spatial visualization and database management functions of GIS system aid organizations that have geographically distributed resources or the ones which are undergoing restructuring. FM applications also rely highly on the imaging capabilities of GIS systems to represent the data elements spatially. Often, both the AM and FM ap-
applications are combined to design applications which allow organization to generate, manage and utilize maps and other spatial data.

3. **Market Analysis.** GIS is a powerful market analysis tool because it has the ability to provide the platform for representation of spatial relationship between the customers, suppliers and competitors. The key GIS functionalities used in the market analysis applications are database management and decision modeling. Most market analysis applications use historical as well as transaction data in combination with decision modeling tools to analyze the marketing environment of an organization.

4. **Logistics and Transportation.** Logistics problems involve a lot of spatial data. In this context, the GIS systems can be used for performing the decision modeling and visualizing the results of such analysis. There are several specific GIS applications in this area such as vehicle routing, inventory management, production control etc. Algorithms such as the route finding algorithms for transport networks, facility layout models, adjacency and flow models are implemented in these applications.

5. **Strategic Decision Making.** Although these are not fully developed, it has been pointed out in [7] that the term spatial decision support system has been coined to describe the systems which incorporate GIS functionalities along with the analytical tools that are found in decision support systems.

6. **Design and Planning.** Several information systems have been developed which are used for drafting and designing purposes. CAD systems are the classic examples of systems which are routinely used to digitize the map and other data. GIS can be used to design plans, maps and layouts in a way similar to CAD systems. A recent advancement in such applications has been Indoor GML based building modeling system from Virtual Builders.

One of the challenges for building GIS applications in the cloud which has not been explored in this study is to differentiate between the system architecture of GIS applications in the cloud as compared to the architecture for a desktop based GIS system or a client server based GIS system.

### 3.5 Why Cloud based GIS?

Cloud based GIS systems might gain many benefits from what the cloud environment infrastructure offers in general. Adopting a cloud based infrastructure for GIS can bring a specific advantage - lower barrier to entry, particularly for the non regular users. GIS systems are usually associated with extreme upfront costs on hardware and softwares, but cloud based GIS can provide users with low cost access to GIS systems. Unlike a desktop based GIS system or a client server based system, a cloud system has no requirements for upfront investment in hardware and software since it allows utility based pricing. Another specific advantage of a cloud based GIS system is that it can provide the users with services without giving them access to actual data. A simple example to take in our scenario would be the one meter resolution satellite imagery data which is not open for researchers.
due to privacy or security concerns. With a cloud based GIS system, researchers can actually request the analysis on such data which can be performed by an intermediate regulating agency which may as well be automated and the results of such analysis can be provided back to requesters without actually giving them the access to data. It can be argued that such measures can also be taken in the conventional GIS systems domain but the central point remains the ease of management with cloud based environment.

However, there are some disadvantages of associated with the cloud based GIS systems. Besides the obvious disadvantages arising from the issues with cloud computing discussed earlier in section 1.3.2, another one is the network speed bottleneck of cloud computing. Most of the GIS analysis involve large volumes of data. For example, the LandSat Thematic Mapping Data available from the EROS Data Center of US Geological Survey for a specific city is typically of size over 30 GigaBytes. Even if the computing nodes inside the cloud are connected using high speed network, the speed of connection between the client’s system and cloud computing infrastructure is limited by the internet service providers. Therefore, when there is a need to upload or download the data to/from the cloud system, the data transfer time can be a big problem.

### 3.6 Case Study 1. OpenGeo Architecture

This section is based on [3] OpenGeo architecture models the technologies which work together to generate an internet map. It provides a complete solution for internet mapping by following a modular approach in architecture. This architecture breaks internet mapping into different functional layers - storage, application server and support for existing systems.

1. Storage: It is required to store the raw data in a consistent manner. Relational database systems can be used to achieve this. In OpenGeo architecture, PostGIS spatial database over PostgreSQL is used in the back end. It also supports importing data directly from the ESRI’s shapefiles format.

2. Application Server: In order to access the data from data store, an application server is needed. OpenGeo uses GeoServer for the same. GeoServer also implements an application cache for improved performance.

3. Development Framework: OpenGeo uses GeoExt/ExtJS for application development. This allows the developers to write vertical applications which can be integrated to form a complete GIS system on web.

4. Support for existing systems: This is another important requirement. Users must be able to export the data from the existing GIS systems and integrate it with the new systems which are developed. OpenGeo architecture provides support for existing systems like Google’s KML, ESRI’s ArcSDE. OpenGeo architecture can be used to serve data from databases and files to mapping portals and build desktop like applications, by embedding maps and providing utilities for capturing data, that can be accessed using web browsers.
CHAPTER 3. CLOUD BASED GIS SYSTEMS

3.6.1 OpenGeo web mapping architecture

The web mapping architecture has five open source components:

1. Storage: PostGIS on PostgreSQL database which processes spatial queries as well as attribute queries.
2. Application Server: GeoServer map/feature server which provides visualization of the underlying data as well as processed data either by WPS or directly from the database.
3. Application Cache: GeoWebCache which intelligently stores and servers map tiles.
4. User interface framework: GeoExt library over ExtJS is a framework which allows the developers to write applications which bind spatial features to powerful UI components.
5. User interface map component: OpenLayers which can consume maps from multiple sources and provide ability to edit and capture data.

Fig 4.1 explains the interactions between various functional components. The database and the application server interact using SQL. The application server and user interface layers communicate using standard web encodings like XML, JSON and images over HTTP protocol. OpenGeo provides flexibility to replace the existing components in the architecture with other components which can server the same purpose.

![Figure 3.1: OpenGeo Web Mapping Architecture](image)

3.6.2 OpenGeo in the cloud

OpenGeo Suite is a software package which is based on the OpenGeo architecture discussed above. The suite comprises of the following modules:

1. Geoserver which is the map/feature server for data visualization.
2. Geoexplorer which is a framework built using GeoExt and it makes it easy to assemble a browser
based mapping application.

3. Geoeditor which allows editing the existing spatial data by adding new spatial content as well as attributes to existing data.

4. Styler which allows the user to decide the layer visualization based on the attributes.

3.7 Case Study 2. ESRI in the Cloud

This section is based on [5] ESRI’s ArcGIS server is a platform which enables the users to deliver GIS services to softwares on other systems. These services can be consumed by clients written in cross platform languages like javascript, flex and silverlight. The users typically deploy ArcGIS server on a privately maintained system. However, ArcGIS can also be deployed in a cloud environment using the Amazon Web Services. The advantages of deploying ArcGIS using AWS include:

1. Easy and fast deployment: ESRI provides preconfigured Amazon Machine Images containing the ArcGIS Server which enables the user to create a EC2 based VM directly. This enables faster deployment since the user does not have to deal with the software installation which is quite tricky and requires expert support for proper configuration.

2. Lower Cost: This is the direct consequence of using AWS. The user have to pay as per use. This might be favorable to users depending upon the usage and operating cost of the same server over a privately maintained machine.

3. Better Availability: The software running on Amazon Web Services can be accessed by public over internet which makes the services provided by ArcGIS server broadly available. Although it raises the issue of security for the users who do not want their services to be public.

The paper also mentions simpler development and testing as an advantage, but that essentially remains the same. Whether the development is done on cloud or on the local system hosting ArcGIS, it does not make a difference.

3.7.1 Proposed Architecture for deploying ArcGIS Server on Amazon EC2

The architecture presents running of an ArcGIS Server on the virtual machines on Amazon EC2 cloud. The ArcGIS server software installed on EC2 is the same as the one which is installed on a single system. The scalability is achieved by creating more virtual machines with pre-installed ArcGIS Server on Amazon EC2. This architecture therefore provides good scalability if we consider the number of individual servers that can be deployed at the same time. However, in terms of performance, the architecture is not scalable because there is no centralized management system which automatically does the load balancing. It should be mentioned that this architecture comes
closest to meet the requirements for cloud based GIS systems.

3.7.2 ArcLogistics and Business Analyst in the cloud

ArcLogistics and Business Analyst Online fall in the category of Software as a Service. ArcLogistics provides capabilities to perform route computing and optimization and generating schedules based on a number of factors. A typical example can be providing solution to an organization that maintains a fleet of vehicles which co-ordinates pickups and drop-offs of numerous people. On the other hand, business analyst is a web based service that provides the functionality to generate reports, location based maps and other forms of data from various sources like demographic factors, business data, consumer spending etc. An example would be to find optimal location for opening a new shop. Both, ArcLogistics and Business Analyst are present in the cloud based environment. The also enable the users to perform complex spatial functions. However, it is not possible to classify them as a true cloud based GIS systems based on the requirements discussed previously. This is because of the lack of knowledge about how the data is processed - does it use a distributed paradigm for processing?

3.8 Case Study 3. www.giscloud.com

'giscloud' is one of the interesting innovator in the recent times. www.giscloud.com is a cloud based GIS system which has the following features:
1. Easy Web Publishing: GIS cloud offers a simple web based solution to create, edit, process and publish data. It also provides easy maintenance of published data.
2. Easy Data Sharing: GIS cloud, being a web application, simplifies the online data sharing. In order to use a published map, all that a user needs to do is use the map’s URL.
3. Improved GIS collaboration: The sharing features provided by GIS cloud makes collaboration easy. There is centralized access to projects, like Google Docs which allows data to be assembled in
real time.
4. Integration support with existing systems: GIS cloud enables the user to integrate their applications with existing GIS systems.
5. HTML 5 canvas for web mapping: This canvas is capable of vector data rendering as vector graphics in real time. It has been shown to perform better than the existing raster rendering engines. Besides these, all the advantages of cloud computing are also applicable here. It also provides profound security option by allowing the user to place the data in the server of their own choice.

3.8.1 Features

The important features of GIS Cloud are described here:
1. Database Manager: The database manager allows the users to connect to PostGIS database and import tables as layers which can later be visualized using the HTML 5 based web mapping engine.
2. Source projection detection and automatic reprojection: These features are important for data visualization. Detection allows mapping to set the corresponding spatial extent while reprojection allows the user to view same data under different project systems. This capability is also present in GeoServer but is known to generate undesired errors.
3. Layer grouping: This feature allows the users to group the layers. For example, all the buildings like academic, residential, recreational etc can be grouped under one layer category - buildings.
4. Thematic map creation.
5. Attribute Information display on maps.
7. Virtually unlimited storage and resources.
8. HTML 5 map engine: The mapping engine is capable of both vector as well as raster graphics rendering. Moreover, it allows the users to interact with the map display.
10. Editing of spatial features.
11. Collaboration by sharing maps over permalinks.

It is important to mention that the full list of features for GIS cloud match with that of the OpenGeo Suite, with GIS cloud having some additional features. GIS cloud offers the user to perform computational actions like buffer analysis, range and area analysis, hotspot analysis, coverage operations and spatial selection operations like intersects, contains etc. An additional module is the fleet management module which is similar to ArcLogistics.

All the features and capabilities of GIS Cloud makes it look like a prototype cloud based GIS system. However, it is again not explicitly explained that how it works as a cloud system because there is no evidence to support that GIS cloud has the ability to take advantage of the distributed data storage and parallel computing, which should be the most important attributes of a cloud based GIS system.
Chapter 4

Architecture Design for Cloud Based GIS

[The idea of architecture design which has been presented in this chapter is based on simple observations. As a part of my RA work, I have been developing an application for NSDI which is a government organization. The application takes input from a CSV file which contains attributes as well as spatial information and loads this file in the PostgreSQL database. It creates spatial data using the attributes. Now, there are tools which can directly load the DGN file from which the CSV is generated in the first place and load it in PostgreSQL, but the requirement is to know what the application is actually doing which is not the case for proprietary tools. This leads to a simple observation that at times, there is a need for transparency in the operations that are being performed on the data and at times there are not. This simple thought can be used to give a high level description of architectures for GIS systems in cloud. In what follows next, this is what has exactly been done.] Depending on the factors that who is providing the services and where the data is hosted, a cloud based GIS system can take one of the three basic forms - public cloud based, private cloud based and hybrid cloud based. Each of these forms have been formulated in this chapter.

4.1 Public Cloud Based GIS System Architecture

4.1.1 Model

The model for a public cloud based GIS has been described here:
1. In a public cloud based GIS system, services will be created and maintained by the vendor who provide the service.
2. Services will be delivered either as a software or as a platform.
3. Users must execute all the applications in the provider’s infrastructure.
4. Data will be provided by the vendor and would be shared by all the authorized users.
5. Data maintenance and updation will be the responsibility of vendor providing the service.
6. All applications can only be executed on the server side.

There can be some refinements to the model. For example, user could be allowed to store private data on the cloud server. This data would only be accessible to the users authorized by the owner of that data. If the user wants to process the data using a custom application, then he/she can be allowed to create the application, upload it to the cloud and use the cloud resources to perform processing.

### 4.1.2 Pros and Cons

Using the above model for public cloud based GIS system has the following advantages:

1. Since all the applications can only be executed on the server side, the vendor may provide a basic set of applications. These can be used by the users without actually having to develop them.

2. It is maintenance free from the user’s perspective. The vendor will have to take care of all the issues related to data, infrastructure and softwares/applications.

The possible disadvantages with such an architecture would include:

1. Users do not have control over the privacy aspect of their own data or applications.

### 4.2 Private Cloud Based GIS System Architecture

This architecture can be understood by placing the user in place of the vendor in public cloud based architecture. The basic architecture may look like the one in Fig 5.2.
4.2.1 Model

1. Private cloud based GIS system is built, operated and maintained by a user.
2. Infrastructure can be purchased or it can also be a virtual cluster residing in a third party cloud computing technology provider’s infrastructure.
3. Cloud GIS servers needs to be created by the user.
4. Data acquisition and maintenance, and application management has to be done by the user. Applications can be self developed or it can be third party extensions.

4.2.2 Pros and Cons

The advantages of using a private cloud based GIS architecture should include:

1. Data privacy is in the control of user.
2. The user has complete control over the application deployment.

Thus privacy and security are enhanced here as compared to the public cloud based architecture. It should be mentioned that in a public cloud GIS system, users share physical hardware with other users and it is difficult for the users to keep track of the condition of its data storage. Moreover, the physical locations of the data stores would be unknown to the users since it is upto the vendor as to how the incoming data and applications are administered. However, in case of a privately owned infrastructure, the user has full control over the storage which makes it easier to monitor and secure the system. Thus, while a public cloud based GIS is suited for people who do not care much about the data privacy and are only interested in data analysis, private cloud based architecture is more suited in the cases when data security is a major concern. One example would be the management
of data relevant to military organizations.

Disadvantages with the private cloud based GIS system architecture are:

1. Too expensive to maintain.

2. Data acquisition itself can be expensive because the economy of scale does not work in private cloud environment.

3. The scalability will be limited in case of privately owned infrastructure.
Chapter 5

Conclusions and Possible Extensions

This study tries to summarize the characteristics of cloud computing technology and GIS systems separately and then tries to bring them together by identifying the features for cloud based GIS systems. The study also attempts to present architectural designs for cloud based GIS systems. However, the literature survey makes it clear that the notion of cloud computing is still new for GIS systems which is the reason for absence of systems which satisfy the requirements mentioned in the features of cloud based GIS systems. However, this presents the GIS domain with a great opportunity to explore the possibility of hosting complete GIS systems in a cloud. One possible extensions in the field of cloud based GIS can be to come up with the guidelines or recommendations for hosting a GIS system in the cloud by defining the resource requirements for hosting GIS system in cloud infrastructure or coming up with the development and testing standards for providing GIS services as a platform. Another interesting work that can be done is performance benchmarking of spatial database management systems. Thus, it can be concluded that application of cloud computing deployment model in the field of GIS is a topic which requires further investigation which can help the existing GIS cloud systems develop into true cloud based GIS systems.
Bibliography