# Geospatial Services in Cloud Infrastructure

Soumadip Biswas, Arindam Dasgupta, Soumya K. Ghosh
*School of Information Technology, Indian Institute of Technology Kharagpur*
*India*
soumadip.cse@gmail.com
adgkgp@gmail.com
skghosh.iitkgp@gmail.com

***Abstract*—Cloud computing is rapidly emerging as a technology trend which any industry that provides or consumes software, hardware, and infrastructure can leverage. The architecture and technology that cloud service and deployment models offer are main areas of research and development for Geographic Information System (GIS). The rise of cloud computing just might be the most important thing happening in the industry of information technology these days. While not everything will move into the cloud, it's fair to say that nearly every industry will use this new approach in some way or another and so is Geographical Information System. It is unnecessary to mention that GIS technology is one of the most technological advances that human has achieved and lives of the people on the world is effected significantly by the same. Given this reality, cloud computing has the potential to change many aspects of our world as well as for users of geographic information system (GIS). In this Thesis work we will explore our work of creating a *GIS-in-the-cloud* environment by incorporating GIS and Cloud together in a meaningful way. This thesis work gives a model about how to collaborate basic service models of cloud to represent the business architecture completely in the Cloud. In this work *Enterprise–GIS* has been used as the Geographical Information System and an Open Source Cloud system namely Eucalyptus is used as the Cloud structure.**

## I. Introduction

### A. Cloud Computing

National Institute of Standards and Technology(NIST) [1] has posted a working definition of cloud computing: Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management of service provider interaction. This cloud model promotes availability and is composed of five essential characteristics (Rapid Elasticity, Measured Service, On-Demand Self-Service, Ubiquitous Network Access, Location-Independent Resource Pooling), three delivery models (Software as a Service, Platform as a Service, and Infrastructure as a Service), and four deployment models (Public Cloud, Private Cloud, Community Cloud and Hybrid Cloud). As per NIST(USA) Cloud Computing is a system which enable on-demand network access to a shared pool of configurable IT resources, which may include networks, storage, applications and services. These resources are managed in such a way so that it can be rapidly provisioned and released with minimum effort of service providers or the system managers. These types of systems now gaining a high popularity in software researcher community because it is new and offers numerous challenges to programmer. However problem with these systems are most of them are private and not available for research community. Above all as the name suggests for cloud computing, it has no standardized shapes or size similar as a cloud, so we have to define cloud as ourselves as per the requirements of the organizations. In our case we define cloud as an integration of Service Oriented architecture and Virtualization [2], [3], [4], [5].

Conceptually, in Cloud Computing everything is assumed as a service (XaaS) [6], such as TaaS (Testing as a Service), SaaS (Software as a Service), PaaS (Platform as a Service), HaaS (Hard-ware as a Service). To this end, a large number of cloud service providers and middleware suits have emerged, each providing different Cloud Computing services. These providers include Amazon EC2, Google App Engine (GAE), SalesForce.com (SFDC), Microsoft Azure, IBM Blue Cloud, 3Tera, to name a few. As we will discuss throughout the paper, the major challenge of the current Cloud Computing evolution, with multiple offerings and providers, is the lack of standardized APIs and usage model.

### B. Enterprise GIS Framework

The enterprise GIS is a platform which is able to integrate enterprise wise geospatial data sources and delivers these data in a uniform format or in form of map. In order to generate information from heterogeneous datasets, Web service is the key technology to integrate different geospatial data repositories. To develop enterprise GIS framework, the Web Service technology is utilized as a platform of the work. In this work, an enterprise GIS platform has been developed. Through this platform, the mobile devices are able to access geospatial data in a uniform format from the diverse and heterogeneous data repositories. The mobile devices are also able to access the geospatial data in form of map [7].

## II. Why GIS and Cloud together makes sense

Geographical Information System is a huge data intensive system. For large geographic area a little information gaining can cause enormous amount of data storage. Another very important issue for a GIS system is that as the data is huge in volume, every time a request comes for processing a service demands significant amount of computation. Clearly

GIS is a data as well as computation intensive system. For a popular and important Geospatial applications it can easily cause server overload due to computation or storage. Of course one can have enough memory and computational power beforehand in store which simply cause quite amount of capital investment beforehand and at the same time the statistical analysis over a long time in such a system can show that the use of the available resources is not efficient as the resource consumption in concerned. So it is always preferable that if the computational power can be supplied as per requirement i.e. as a service that can dynamically scale up the storage as required(scaling down for storage in this particular case is merely significant) and also if multiple server can be there to balance the load of the GIS only when it is at its Peak-Usage. Clearly from this discussion two things are clear that a GIS system needs to have (1) Dynamically scale Up/Down of storage and (2) replicating server to balance load (or increase computational power overall). As far as the requirement is concerned these two requirements are the area of specialization of the Cloud Computing Technology. So if we can submerge these two i.e. Cloud and GIS in a meaningful way it is completely convenient to have one dedicated Geospatial Cloud [8].

## III. Integration of Standard Service Models of Cloud

### A. Standard Service Models of Typical Cloud

For a typical system to be called a cloud it needed to deploy itself to provide at least one of the standard service models namely Infrastructure as a Service, Platform as a Service and Software as a Service. Following is a small description on each.

*1) Infrastructure as a Service [IaaS]:* In this most basic cloud service model, cloud providers offer computers as physical or more often as virtual machines , raw (block) storage, firewalls, load balancers, and networks. IaaS providers supply these resources on demand from their large pools installed in data centers. Local area networks including IP addresses are part of the offer. For the wide area connectivity, the Internet can be used or - in carrier clouds - dedicated virtual private networks can be configured.

To deploy their applications, cloud users then install operating system images on the machines as well as their application software. In this model, it is the cloud user who is responsible for patching and maintaining the operating systems and application software. Cloud providers typically bill IaaS services on a utility computing basis, that is, cost will reflect the amount of resources allocated and consumed [9].

*2) Platform as a Service [PaaS]:* In the PaaS model, cloud providers deliver a computing platform and/or solution stack typically including operating system, 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. With some PaaS offers, the underlying compute and storage resources scale automatically to match application demand such that the cloud user does not have to allocate resources manually [9].

*3) 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 cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support. What makes a cloud application different from other applications is its elasticity. This can be achieved by cloning tasks onto multiple virtual machines at run-time to meet the changing work demand. Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user who sees only a single access point. To accomodate a large number of cloud users, cloud applications can be multitenant, that is, any machine serves more than one cloud user organization. It is common to refer to special types of cloud based application software with a similar naming convention: desktop as a service, business process as a service, Test Environment as a Service, communication as a service. The pricing model for SaaS applications is typically a monthly or yearly flat fee per user [9].

### B. Proposed Model for Geospatial Cloud

Before going to the discussion of the model let us consider one example, basically its the idea of coding style. In conventional codes we first visualize the problem and solve the same by applying logic with the code snippets. As time goes, we improved the our coding style by dividing the problem into sub-problems and code for each of them, these sub-problems are called modules. Later we tried to make these modules and tried to make them as independent to main problem as possible. Once this is done now we can see these modules to be independent code snippets which can be used with some other problem as per similar requirements. We now call the collection of the code snippets library and the idea is called Modularization. In this particular discussion we will be linking to the similar concept of modularization, but this time the idea is to be realized into the physical world entities. The big challenge here is whether it is possible if so then how. That is where the idea of cloud comes into the picture. In computer science, general users stick to the most abstract end of the whole system i.e. they are concerned with the jobs to be done, all they care about is the output. But for the output to come properly there are certain background phenomenon that needs to be taken care of by the providers. In the background first to serve a request from an user first thing that is needed is an application. Now to run an application it needs an environment typically called the run-time (e.g. jvm, jre, Operating Systems, Database services). Now to run such an run-time or operating system it needs some physical infrastructure to run on. So broadly it needs three main modules to serve a request, they are in reverse order infrastructure, platform and the software. Thus if we can separates out each of the modules and serve each of them independently, conversely we can achieve the modularization in the physical level. Cloud computing has three main modes of deployment which are respectively Infrastructure as a Service, Platform as a Service and Software as a service, which are as the name suggests can give us the three level of independent modular services as discussed earlier.

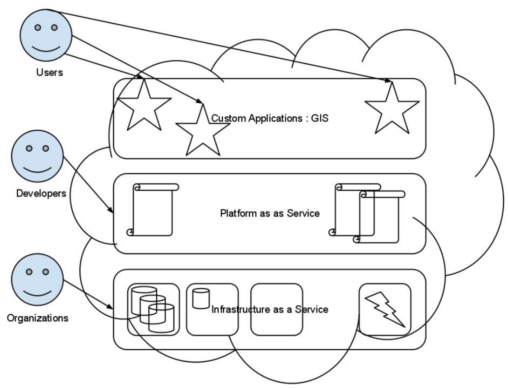


Fig. 1. Actors at Different Levels in the Cloud

Now let us put together all these models. This particular case is shown in the Figure 1 when we put one model over another one complete business models can be represented and the figure shows the positions of different level of users of the system.

In a cloud all different kind of actors are having its place,[Figure 1] such as an organization normally tend to buy computing infrastructures which can be obtained from the IaaS service paradigm of the cloud. Now having a company established it will have workers who will be working on those resources of the company but obviously been provided a platform to work on, which again can be fetched as PaaS service over a cloud. Finally one has to deploy products/Web applications more specifically, to achieve the goal of productivity, it is now as simple as deploying on the SaaS environment. So all the requirement an organization is having can be met with one service – The Cloud Service.

### *C. Modeling a Scenario for Geospatial Cloud*

In case of GIS there can be many back-end data services available for providing the different spatial data layers for the applications to be fetched. For instance let us consider there are 3 such services there. Now for each application it may not be necessary to contact all the services so required services (or a caching server for each service) and another host can be bind together to complete the system. Once this is done each custom application can be made available to the public for using.

In case of GIS Technology, as it demands huge amount of data processing in a efficient time and it is an on-demand service, so that the requirement of the resource can differ in different times. Clearly Cloud in this kind of scenario comes as quite promising infrastructure for organization that has the service. Now that they don't have to care much about the management of the infrastructure they can just concentrate on the data only. At the same time when an organization wants to provide web services developer shouldn't need to worry about the infrastructure either, here the organization can provide them only with the platform to use, similarly cloud can provide that with the PaaS service, so that they can concentrate only on the developing. Following is general case to understand the problem. Figure 2 illustrates the problem.

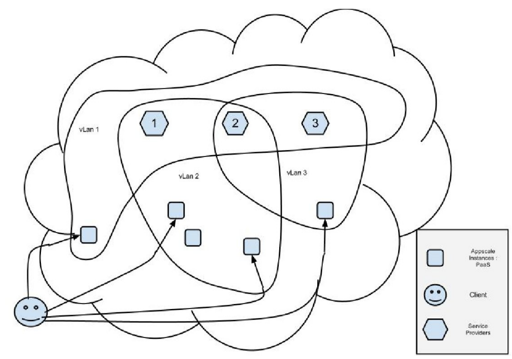


Fig. 2. Problem Illustration for Geospatial Cloud

Now that an organization can have their services running over instances inside the cloud, by using grouping facilities they can provide layer 2 or layer 3 security to the instances. Thus only required services can be connected together and can be bind with the desired services. Again the services can be run through the PaaS service model, so that load balancing, distributed replica and dynamic scaling can be done in a transparent way to the developers who are the client for the platform service. In the end deployed application/services can now be viewed to the client as a software as a service. Thus it can be integrated as a SaaS-PaaS-IaaS module altogether, and providing a complete industry structure in the cloud itself. Now in addition to that if a service needs to be fetched from outside the cloud it can also be done easily by simple network accesses. Moreover a dedicated Catalog service can be incorporated for searching required data server or any other service, so that depending on the requirement of the customer at different level of the model can be entertained easily.

## IV. Description of Work Accomplished

To implement the proposed model as discussed in previous section the first thing that is needed to accomplish is provision of Infrastructure as a Service. In this case the open source project Eucalyptus has been selected to fulfill the role. Once the IaaS has been provided successfully, now the next task is to provide Platform as a Service and then the Software as a Service on top of that to fulfill the model requirement. In this section the next discussion section is on what a platform really stands for and one way of demonstrating the same. Afterwards how application(SaaS) can be deployed on the platform is shown.

### A. Set Up Cloud Infrastructure to Provide IaaS

Eucalyptus is a software platform for the implementation of private cloud computing on computer clusters. There is an open-core enterprise edition and an open-source edition. Currently, it exports a user-facing interface that is compatible with the Amazon EC2 and S3 services but the platform is modularized so that it can support a set of different interfaces simultaneously. Eucalyptus has be set up successfully over the given infrastructure. Eucalyptus can be accessible over the web-interface and also via the command line using commands provided by Euca2ools.

We have used 3 Dell rack servers with a total of 56 CPU cores, 80 GB of RAM and 1.5 TB of Hard Drive as the backing resource/Infrastructure to the GIS cloud. Our Eucalyptus configuration uses two different network for its configuration. First is the enterprise network which can be public for users of the system. Second, is the internal network of the eucalyptus components with which components are privately connected to each other. There is another network is used for the configuration which another private network to be assigned to the instances internally, and a set of public ip to be assigned to the instances for public access. Thus each instance running under the cloud infrastructure will be having assigned two IPs. And Eucalyptus can be accessed over the web-interface and also via the command line using commands provided by Euca2ools for full control of the system.

### B. Set up a PaaS to provide Infrastructure level independence.

In fair understanding of the platform as a service architecture the platform can actually be as vast as an Operating system or as simple as a run-time environment. In our case to run a Geospatial Web application it basically requires required resources to compute that is data (WFS service to be specific) and a run-time environment (in our case it is apache2). Now run-time can be run on some operating system which can be running on a machine(in our case virtual machine/instances). So clearly on the IaaS architecture that we have the PaaS service is provided. A web portal is there which can show you the current status of the cloud. and one can deploy their web application suitable for the environment just by some clicks and the application will be running. Until now the basic framework has been built which can deploy web applications.

## V. Deploying Enterprise GIS framework over the Infrastructure.

### A. Enterprise GIS Framework

The enterprise GIS is a platform which is able to integrate enterprise wise geospatial data sources and delivers these data in a uniform format or in form of map. In order to generate information from heterogeneous datasets, Web service is the key technology to integrate different geospatial data repositories. To develop enterprise GIS framework, the Web Service technology is utilized as a platform of the work. In this work, an enterprise GIS platform has been developed. Through this platform, the mobile devices are able to access geospatial data in a uniform format from the diverse and heterogeneous data repositories. The mobile devices are also able to access the geospatial data in form of map [7].

## VI. System Behavior Models

As the problem stated earlier, it is required to deploy the geospatial services onto the cloud, which as we encountered can be done primarily in two different ways. Both the cases is described below.

### A. Model to illustrate 'Everything in the Cloud'

Here we consider that we need to deploy the whole geospatial system in the cloud. In this case that means we need to provide the data services(WFS) as well as any other services that are intended to provide by the cloud(e.g. WMS- Web Map Service, CWS- Catalog Service, WPS- Web Processing Service etc.). For this we deployed the instances of Ubuntu Desktop Edition in the cloud and configure that image with required parameters such as java, apache-tomcat etc. After that the image is ready to go once we deployed the service over there. In the following figure the idea is illustrated.

Deployment Scenario 1 is shown in the Figure 3.

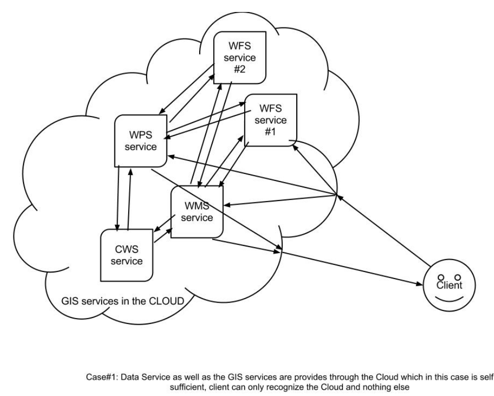


Fig. 3. Deployment Scenario 1: Everything in the Cloud

This is a typical case for a Cloud System. All the services such as WFS, WMS, CWS etc. are provided through the cloud itself. All those services are deployed over one or more individual instance which runs from inside the cloud. In this case binding different services are very easy to handle, thus just by providing security from inside the cloud image isolation can be done and for each different Web Service provided to the public network can be made secure at the infrastructure level. But sometimes it is not possible for the service providers to give up their data solely to the cloud without hesitation, for such cases the following model has been prepared.

### B. Model to illustrate 'Data Providers are Outside the Cloud'

In contrast to the previous case here we consider that not all the basic sources are available with the cloud. As the Enterprise GIS Application uses many different sources of data from different providers and make integrate all those data by sitting on top of them, it has been seen that the data providers are not willing to giving up all the raw data to the organization but as a service. Now this problem can be solved by remodeling the problem where we have considered that the WFS of the data services are outside at their places and the cloud needs to take services from them when required by the other geospatial services which are still provided from inside the cloud. Thus the client side still dosen't know any more than contacting the cloud but receiving the reply for its request.

Deployment Scenario 2 is shown in the Figure 4.

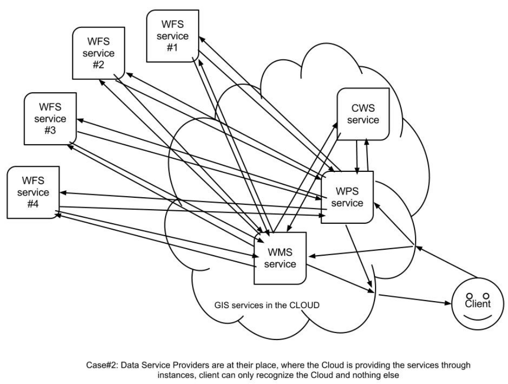


Fig. 4. Deployment Scenario 2: Situation when Cloud has to take Other Services

In this figure as we can see that all the data services(which can be any other service that is required to stay outside) are outside of the cloud. In this case the computation such as creating maps or other computationaly expensive processing of the spatial data that is fetched from the data providers will be done in the cloud by the applications running inside the cloud (which can be done efficiently given that they use as much resource as required within the availability from the cloud) and then the results can be forwarded to the client. Again here also the transparency for the user is preserved. But as cloud doesn't have any other access to the service providers outside than only fetching the data the security model for this case has to made in the application level.

## VII. An Example Scenario of the Geospatial Cloud

To understand the solution lets take a generic example as shown in the following figure. This particular example shows a Cloud as a wrapper to the Geospatial Information System and the effect of the same to the existing system.

Geospatial data are huge in size and collected & maintained by different organizations depending on the attributes of the data e.g. layers, type, region. All these vendor/ organizations provide Data as a Service to the customers of them, which also known as WFS or Web Feature Service. To view the same information in a view-able format GIS has a Map Service known as WMS and another important service is the Web Processing Service (WPS) which deals with large amount of computation on the spatial data and produce desired information out of it.

Given the situation as above in general scenario all these organization are scattered over the web in different physical locations, which essentially means whenever any service is triggered a huge amount of data needs to be transferred from one host to another over the network . It can be easily seen by the problem statement now that the more it needs cascading of services the more bitter it will go to produce the end result and essentially that effects the end user. Moreover it gets worse with the increment in service requests in terms of computation capability and the network bandwidth limitation.

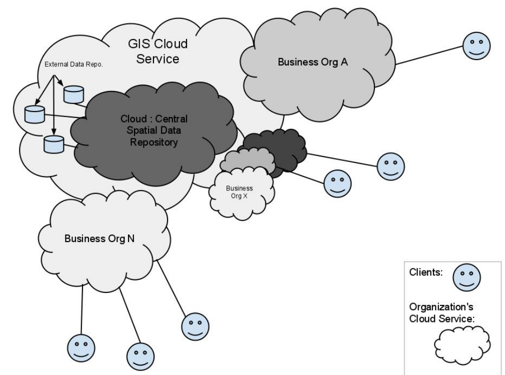


Fig. 5. A Geospatial Cloud Representation

The model in the Figure 5 shows a scenario that fits the Cloud to the GIS and creates a solution to the above mentioned problems known as GIS Cloud. As discussed earlier the model SaaS over PaaS over IaaS is the exact solution in this particular situation. For a GIS central organization it needs to have the data repository as the central repository in the figure, but not all organization will surrender their data to the same so all together these organization includes with the central repository integrated with a cloud service creates the desires GIS cloud service. At this level this particular Cloud will provide the Infrastructure as a Service which can be availed by all the basic service repositories that means mainly the WFS providers. once this is set up, other business organizations like to offer WPS, WMS of other services to their end user customers will be availed infrastructure as their requirement from the cloud itself and can be coupled with their desired WFS or other requirements and will be ready to go.

Now from each business organization point of view they have the resources those are elastic as per the computation requirements, at this point only thing they need is to run their applications. So they need developers to create their web apps and deploy the same on the cloud. For this the Organization can provide the Platform as a Service on the infrastructure they have acquired to the developers to develop and run their applications. Now developers can now write their application in the form of Software as a Service and can deploy in there. At this point all the three models of deployment has been used one over another and the end result is each organization having their business running with no difference as the existing infrastructure but with lesser headache such as maintenance, similarly developers doesn't need to care about the infrastructure behind but to concentrate on the application development and of course the end users do not feel a difference at that level. So to make a gist of the situation following is the list of providers and consumers at different stage of this arrangement.

- IaaS (Generic Hardware)
  - Vendor: Cloud Service Provider(IITkgp Geospatial Cloud)
  - Client: Business Organization
- PaaS (Operating System/ Runtimes)
  - Vendor: Business Organization
  - Client: Developers
- SaaS (Application-web apps/geospatial apps)
  - Vendor: Developers
  - Client: Customer Base

***Now let's have a look at the bright sides of this arrangement:***

- All the data are maintained by a central maintenance
- All end user request are streaming through the front end and actually served at the central Cloud
- As all the requests are served at a single point so not much moving of large data is required
- Computation Power as a Service gives the flexibility to large computation whatsoever
- All similar requests are streaming at a single point leads to better routing

## VIII. Conclusion

The idea of putting everything into the cloud, in this case the whole GIS gives the big picture of the *GIS-cloud or GIS-in-the-cloud*. As described earlier in this thesis the Cloud provides the infrastructure to the owners. So that way the main GIS service providers will deploy their data service over the cloud in one way or another. At the same time using the Platform as a service paradigm, access will be given to developers to design and upload custom application for the organization itself which uses the data provided the the data service from single or multiple organizations from the cloud. Now the applications being the web service based Software as a Service will provide the final wrapping over the Cloud. At this point we can see that the whole GIS system has been deployed in the cloud in this manner, thus our goal succeeds as *GIS-in-the-cloud* or can be abbreviated as *GIS Cloud.*

## Acknowledgment

ACKNOWLEDGMENT

## References

[1] M. P. and G. T., "Perspectives on cloud computing and standards," tech. rep.

[2] A. Weiss, "Computing in the clouds," *netWorker*, vol. 11, pp. 16–25, Dec. 2007.

[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599 – 616, 2009.

[4] "The new age of cloud computing and gis." Website. http://www.esri.com/news/arcwatch/0110/feature.html.

[5] "Twenty experts define cloud computing." Website. http://cloudcomputing.sys-con.com/read/612375_p.htm.

[6] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy, survey, and issues of cloud computing ecosystems," in *Cloud Computing* (N. Antonopoulos and L. Gillam, eds.), vol. 0 of *Computer Communications and Networks*, pp. 21–46, Springer London, 2010. 10.1007/978-1-84996-241-4_2.

[7] A. Dasgupta and S. Ghosh, "An enterprise gis framework for mobile device based on orchestration of geospatial services," tech. rep., SIT, Indian Institute of technology, Kharagpur.

[8] E. I. S. A. Victoria Kouyoumjian, "Gis in the cloud," tech. rep., ESRI, 6 2011. http://www.esri.com/library/ebooks/gis-in-the-cloud.pdf.

[9] "Cloud computing." Website. http://en.wikipedia.org/wiki/Cloud_computing.