



# MARKET-ORIENTED CLOUD COMPUTING – DELIVERING IT SERVICES AS COMPUTING UTILITIES

B. Tulasi\*

## ABSTRACT

*This paper provides information about 21st century vision of computing, identifies various computing paradigms promising to deliver the vision of computing utilities, defines Cloud computing and provides the architecture for creating market-oriented Clouds by leveraging technologies such as VMs, Cloud platforms especially those developed in industries along with our current work towards realizing market-oriented resource allocation of Clouds and concludes with the need for convergence of competing IT paradigms for delivering our 21st century vision.*

## Introduction

We are witnessing a seismic shift in information technology — the kind that comes around every decade or so. It is so massive that it affects not only business models, but the underlying architecture of how we develop, deploy, run and deliver applications.

---

\* Lecturer, Department of MCA, Christ University, Bangalore, Karnataka. Ph: +919845755295

In 1969, Leonard Kleinrock [1], one of the chief scientists of the original Advanced Research Projects Agency Network (ARPANET) project which seeded the Internet, said: "As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of 'computer utilities' which, like present electric and telephone utilities, will service individual homes and offices across the country." This vision of the computing utility based on the service provisioning model anticipates the massive transformation of the entire computing industry in the 21st century whereby computing services will be readily available on demand, like other utility services available in today's society. Similarly, computing service users (consumers) need to pay providers only when they access computing services. In addition, consumers no longer need to invest heavily or encounter difficulties in building and maintaining complex IT infrastructure.

Software practitioners are facing numerous new challenges toward creating software for millions of consumers to use as a service rather than to run on their individual computers. Over the years, new computing paradigms have been proposed and adopted, with the emergence of technological advances such as multi-core processors and networked computing environments, to edge closer toward achieving this grand vision. These new computing paradigms include cluster computing, Grid computing, P2P computing, service computing, market-oriented computing, and most recently Cloud computing. Computing services need to be highly reliable, scalable, and autonomic to support ubiquitous access, dynamic discovery and composability. In particular, consumers can determine the required service level through Quality of Service (QoS) parameters and Service Level Agreements (SLAs). Of all these computing paradigms, the two most promising ones appear to be Grid computing and Cloud computing.

Grid computing [2], most simply stated, is distributed computing taken to the next evolutionary level. The standardization of communications between heterogeneous systems created the Internet explosion. The emerging standardization for sharing resources, along with the availability of higher bandwidth, are driving a possibly equally large evolutionary step in grid computing. Grid [3] computing appears to be a promising trend for three reasons: (1) its ability to make more cost-effective use of a given amount of computer resources, (2) as a way to solve problems that can't be approached without an enormous amount of computing power, and (3) because it suggests that the resources of many computers can be cooperatively and perhaps synergistically harnessed and managed as a collaboration toward a common objective. In some grid computing systems, the computers may collaborate rather than being directed by one managing computer. One likely area for the use of grid computing will be pervasive computing applications - those in which computers pervade our environment without our necessary awareness. Due to its

potential to make impact on the 21st century as much as the electric power Grid did on the 20th century, Grid computing has been hailed as the next revolution after the Internet and the Web.

Today, the latest paradigm to emerge is that of Cloud computing [4] which promises reliable services delivered through next-generation data centers that are built on compute and storage virtualization technologies. Cloud computing is a computing paradigm in which tasks are assigned to a combination of connections, software and services accessed over a network. This network of servers and connections is collectively known as "the cloud." Computing at the scale of the cloud allows users to access supercomputer-level power. Consumers will be able to access applications and data from a "Cloud" anywhere in the world on demand. In other words, the Cloud appears to be a single point of access for all the computing needs of consumers. The consumers are assured that the Cloud infrastructure is very robust and will always be available at any time.

## Definition and Trends

A number of computing researchers and practitioners have attempted to define Clouds in various ways [5]. "A Cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers."

The cloud is a virtualization of resources that maintains and manages itself. There are of course people resources to keep hardware, operation systems and networking in proper order. But from the perspective of a user or application developer only the cloud is referenced. The majority of cloud computing infrastructure consists of reliable services delivered through next-generation data centers that are built on compute and storage virtualization technologies. The services are accessible anywhere in the world, with The Cloud appearing as a single point of access for all the computing needs of consumers. Commercial offerings need to meet the quality of service requirements of customers and typically offer service level agreements. Open standards and open source software are also critical to the growth of cloud computing.

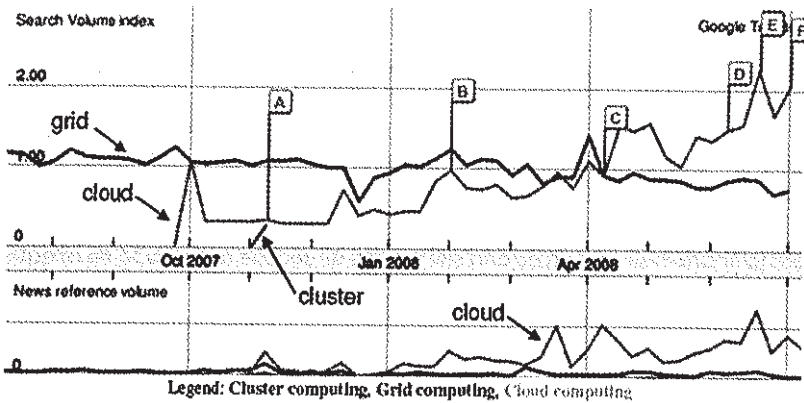


Figure 1: Google search trends for the last 12 months.

## Web Search Trends

The popularity of different paradigms varies with time. The Web search popularity, as measured by the Google search trends during the last 12 months, for terms “cluster computing”, “Grid computing”, and “Cloud computing” is shown in Figure 2. From the Google trends, it can be observed that cluster computing was a popular term during 1990s, from early 2000 Grid computing become popular, and recently Cloud computing started gaining popularity. Spot points in Figure indicate the release of news related to Cloud computing as follows:

- A** IBM Introduces ‘Blue Cloud’ Computing, CIO Nov 15 2007
- B** IBM, EU Launch RESERVOIR Research Initiative for Cloud Computing, IT News Online -Feb 7 2008
- C** Google and Salesforce.com in Cloud computing deal, Siliconrepublic.com -Apr 14 2008
- D** Demystifying Cloud Computing, Intelligent Enterprise -Jun 11 2008.
- E** Yahoo realigns to support Cloud computing, ‘core strategies’, San Antonio Business Journal -Jun 27 2008.
- F** Merrill Lynch Estimates “Cloud Computing” To Be \$100 Billion Market, SYS-CON Media -Jul 8 2008.

## Market-Oriented Cloud Architecture

As consumers rely on Cloud providers to supply all their computing needs, they will require specific QoS to be maintained by their providers in order to meet their objectives and sustain their operations. Cloud providers will need to consider and meet different QoS parameters of each individual consumer as negotiated in specific SLAs. To achieve this, Cloud providers can no longer continue to deploy traditional system-centric resource management architecture that do not provide incentives for them to share their resources and still regard all service requests to be of equal importance. Instead, market-oriented resource management [6] is necessary to regulate the supply and demand of Cloud resources at market equilibrium, provide feedback in terms of economic incentives for both Cloud consumers and providers, and promote QoS-based resource allocation mechanisms that differentiate service requests based on their utility.

Figure 2 shows the high-level architecture for supporting market-oriented resource allocation in Data Centers and Clouds. There are basically four main entities involved:

1. **Users/Brokers:** Users or brokers acting on their behalf submit service requests from anywhere in the world to the Data Center and Cloud to be processed.
2. **SLA Resource Allocator:** The SLA Resource Allocator acts as the interface between the Data Center/Cloud service provider and external users/brokers. It requires the interaction of the following mechanisms to support SLA-oriented resource management:
  - o **Service Request Examiner and Admission Control:** When a service request is first submitted, the Service Request Examiner and Admission Control mechanism interprets the submitted request for QoS requirements before determining whether to accept or reject the request. Thus, it ensures that there is no overloading of resources whereby many service requests cannot be fulfilled successfully due to limited resources available. It also needs the latest status information regarding resource availability (from VM Monitor mechanism) and workload processing (from Service Request Monitor mechanism) in order to make resource allocation decisions effectively. Then, it assigns requests to VMs and determines resource entitlements for allocated VMs.
  - o **Pricing:** The Pricing mechanism decides how service requests are charged. For instance, requests can be charged based on submission time (peak/

off-peak), pricing rates (fixed/changing) or availability of resources (supply/demand). Pricing serves as a basis for managing the supply and demand of computing resources within the Data Center and facilitates in prioritizing resource allocations effectively.

- Accounting: The Accounting mechanism maintains the actual usage of resources by requests so that the final cost can be computed and charged to the users. In addition, the Service Request Examiner and Admission Control mechanism to improve resource allocation decisions can utilize the maintained historical usage information.
  - VM Monitor: The VM Monitor mechanism keeps track of the availability of VMs and their resource entitlements.
  - Dispatcher: The Dispatcher mechanism starts the execution of accepted service requests on allocated VMs.
  - Service Request Monitor: The Service Request Monitor mechanism keeps track of the execution progress of service requests.
3. VMs: Multiple VMs can be started and stopped dynamically on a single physical machine to meet accepted service requests, hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests. In addition, multiple VMs can concurrently run applications based on different operating system environments on a single physical machine since every VM is completely isolated from one another on the same physical machine.
4. Physical Machines: The Data Center comprises multiple computing servers that provide resources to meet service demands. In the case of a Cloud as a commercial offering to enable crucial business operations of companies, there are critical QoS parameters to consider in a service request, such as time, cost, reliability and trust/security. In particular, QoS requirements cannot be static and need to be dynamically updated over time due to continuing changes in business operations and operating environments. In short, there should be greater importance on customers since they pay for accessing services in Clouds. In addition, the state-of-the-art in Cloud computing has no or limited support for dynamic negotiation of SLAs between participants and mechanisms for automatic allocation of resources to multiple competing requests. Recently, we have developed negotiation mechanisms based on alternate offers protocol for establishing SLAs [7]. These have high potential for their adoption in Cloud computing systems built using VMs.

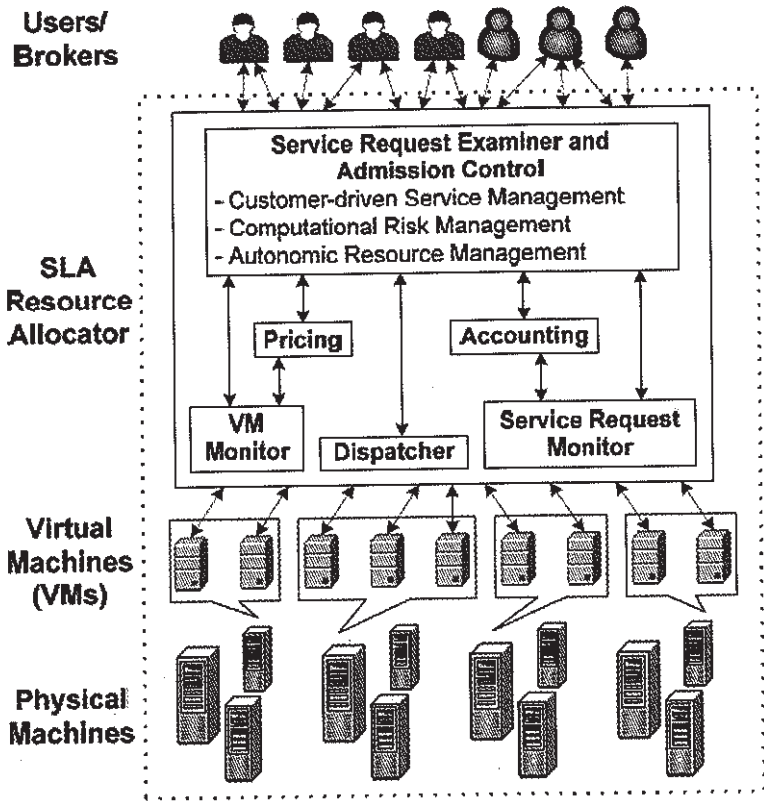


Figure 2: High-level market-oriented cloud architecture.

## Cloud Platforms

The coming shift to cloud computing is a major change in our industry. One of the most important parts of that shift is the advent of cloud platforms. As its name suggests, this kind of platform lets developers write applications that run in the cloud, or use services provided from the cloud, or both. Different names are used for this kind of platform today, including on-demand platform and platform as a service (PaaS). Whatever it's called, this new way of supporting applications has great potential.

To see why, think about how application platforms are used today. When a development team creates an on-premises application (i.e., one that will run within an organization), much of what that application needs already exists. An operating system provides basic support for executing the application, interacting with storage, and more, while other computers in the environment offer services such as remote storage. If the creators of every on-premises application first had to build all of these basics, we'd have many fewer applications today.

Similarly, if every development team that wishes to create a cloud application must first build its own cloud platform, we won't see many cloud applications. Fortunately, vendors are rising to this challenge, and a number of cloud platform technologies are available today. The following are few of them:

- Amazon Elastic Compute Cloud (EC2) [8] provides a virtual computing environment that enables a user to run Linux-based applications.
- Simple Storage Service (S3), before he can start, stop, and monitor instances of the uploaded AMIs. Amazon EC2 charges the user for the time when the instance is alive, while Amazon S3 charges for any data transfer (both upload and download).
- Google App Engine [9] allows a user to run Web applications written using the Python programming language.
- Microsoft Live Mesh [10] aims to provide a centralized location for a user to store applications and data that can be accessed across required devices (such as computers and mobile phones) from anywhere in the world.
- Sun network.com (Sun Grid) [11] enables the user to run Solaris OS, Java, C, C++, and FORTRAN based applications.
- GRIDS Lab Aneka [12], which is being commercialized through Manjrasoft, is a .NET-based service-oriented platform for constructing enterprise Grids.

## Summary and Conclusion

Cloud computing is a new and promising paradigm delivering IT services as computing utilities. As Clouds are designed to provide services to external users, providers need to be compensated for sharing their resources and capabilities. The state-of-the-art Cloud technologies have limited support for market-oriented resource



management and they need to be extended to support: negotiation of QoS between users and providers to establish SLAs; mechanisms and algorithms for allocation of VM resources to meet SLAs; and manage risks associated with the violation of SLAs. Furthermore, interaction protocols needs to be extended to support interoperability between different Cloud service providers. As Cloud platforms become ubiquitous, there is need for internetworking them to create a market-oriented global Cloud exchange for trading services. Several challenges need to be addressed to realize this vision. Finally, it is also required to address regulatory and legal issues, which go beyond technical issues.

## References

- 1 L. Kleinrock. A vision for the Internet. *ST Journal of Research*, 2(1):4-5, Nov. 2005.
- 2 Viktors Berstis Fundamentals of Grid Computing, *IBM Redbooks*, November 11, 2002.
- 3 What is grid computing? <http://www.redbooks.ibm.com/redpapers/pdfs/redp3614.pdf> [16 October 2008].
- 4 A. Weiss. Computing in the Clouds. *netWorker*, , Dec. 2007.
- 5 Twenty Experts Define Cloud Computing, [http://cloudcomputing.syscon.com/read/612375\\_p.htm](http://cloudcomputing.syscon.com/read/612375_p.htm) [16 October 2008].
- 6 R. Buyya, D. Abramson, and S. Venugopal. The Grid Economy. *Proceedings of the IEEE*, 93(3): 698-714, *IEEE Press*, USA, March 2005.
- 7 S. Venugopal, X. Chu, and R. Buyya. A Negotiation Mechanism for Advance Resource Reservation using the Alternate Offers Protocol. In *Proceedings of the 16th International Workshop on Quality of Service (IWQoS 2008)*, Twente, The Netherlands, June 2008.
- 8 Amazon Elastic Compute Cloud (EC2), <http://www.amazon.com/ec2/> [16 October 2008]
- 9 Google App Engine, <http://appengine.google.com> [16 October 2008]
- 10 Microsoft Live Mesh, <http://www.mesh.com> [16 October 2008]
- 11 Sun network.com (Sun Grid), <http://www.network.com> [16 October 2008]
- 12 X. Chu, K. Nadiminti, C. Jin, S. Venugopal, and R. Buyya. Aneka: Next-Generation Enterprise Grid Platform for e-Science and e-Business Applications. In *Proceedings of the 3th IEEE International Conference on e-Science and Grid Computing (e- Science 2007)*, Bangalore, India, Dec. 2007.