Volume 3, Issue 8, August 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Resource Provisioning for Elastic Applications on Hybrid Clouds Environment

Ms. S.G.Chavan
Lect. Govt Polytechnic
India

Prof. R.P.Mogre
Principal Govt Polytechnic
India

Abstract— Computing is being transformed to a model consisting of services that are commoditized and delivered in a manner similar to traditional utilities such as water, electricity, gas, and telephony. In such a model, users access services based on their requirements without regard to where the services are hosted or how they are delivered. Several computing paradigms have promised to deliver this utility computing vision and these include cluster computing, Grid computing, and more recently Cloud computing. The latter term denotes the infrastructure as a ``Cloud'' from which businesses and users are able to access applications from anywhere in the world on demand. Thus, the computing world is rapidly transforming towards developing software for millions to consume as a and find out advantage and limitation to increase resources utilization and profit in future proposed work.

Keywords:-Cloud, PaaS, IaaS SaaS

1. Introduction

Cloud computing [1,4] led to an innovative approach in the way in which IT infrastructures, applications, and services are designed, developed, and delivered. It fosters the vision of any IT asset as a utility, which can be consumed on a pay-peruse basis like water, power, and gas. This vision opens new opportunities that significantly change the relationship that enterprises, academia, and end-users have with software and technology. Cloud computing promotes an on-demand model for IT resource provisioning where a resource can be a virtual server[4], a service, or an application platform. Three major service offerings contribute to defining Cloud computing: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). Infrastructure-as-a- Service providers deliver on-demand components for building IT infrastructure such as storage, bandwidth, and most commonly virtual servers, which can be further customized with the required software stack for hosting applications. Platform-as-a- Service providers deliver development and runtime environments for applications that are hosted on the Cloud. They allow abstraction of the physical aspects of a distributed system by providing a scalable middleware for the management of application execution and dynamic resource provisioning. Software-as-a-Service providers offer applications and services on-demand, which are accessible through the Web. SaaS applications are multi-tenant and are composed by the integration of different components available over the Internet. The offer of different models on which computing resources can be rented creates new perspectives on the way IT infrastructures are used, because Cloud offers the means for increasing IT resource availability whenever necessary, by the time these resources are required, reducing costs related to resource acquisition and maintenance. A case for exploring such a feature of Clouds is in Desktop Grids, which are platforms that use idle cycles from desktop machines to achieve high-throughput computing [2]. Typically, applications are executed in such platforms on a best-effort basis, as no guarantees can be given about the availability of individual machines that are part of the platform. If Desktop Grid resources are combined with Cloud resources, a better level of confidence about resource availability can be given to users, and so it is possible to offer some QoS guarantees related to the execution time of applications at a small financial cost.

Cloud Computing [3] is a recent technology trend whose aim is to deliver on demand IT resources on a pay per use basis. Previous trends were limited to a specific class of users, or focused on making available on demand a specific IT resource, mostly computing. Cloud Computing aims to be global and to provide such services to the masses, ranging from the end user that hosts its personal documents on the Internet, to enterprises outsourcing their entire IT infrastructure to external data centers. Never before an approach to make IT has a real utility been so global and complete: not only computing and storage resources are delivered on demand but the entire stack of computing can be leveraged on the Cloud.

Figure 1 provides an overall view of the scenario envisioned by Cloud Computing. It encompasses so many aspects of computing that very hardly a single solution is able to provide everything that is needed. More likely, specific solutions can address the user needs and be successful in delivering IT resources as a real utility. Figure 1 also identifies the three pillars on top of which Cloud Computing solutions are delivered to end users. These are: *Software as a Service (SaaS), Platform as a Service (PaaS)*, and *Infrastructure/Hardware as a Service (IaaS/HaaS).* These new concepts are also useful to classify the available options for leveraging on the Cloud the IT needs of everyone. Examples of Software as a Service are Salesforce.com2 and Clarizen.com3 , which respectively provide on line CRM and project management services. PaaS solutions, such as Google AppEngine4 [4], Microsoft Azure5, and Manjrasoft Aneka provide users with a development platform for creating distributed applications that can automatically scale on demand. Hardware and Infrastructure as a Service solutions provide users with physical or virtual resources that are fitting the requirements of the user applications in term of CPU, memory, operating system, and storage. These and any others QoS parameters are established through a Service Level Agreement (SLA).

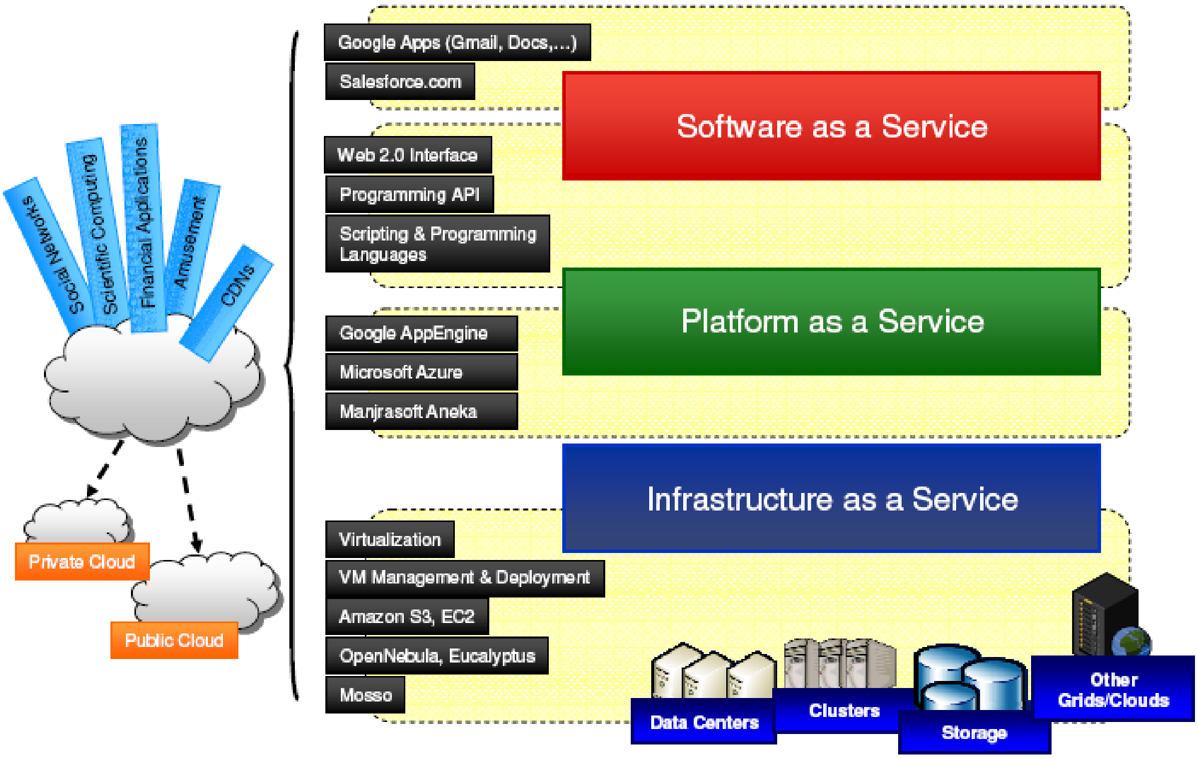

Figure 1. Cloud Computing Environment.

It is very unlikely that a single solution provides the complete stack of software, platform, infrastructure and hardware as a service. More commonly, specific solutions provide services at one (or more) of these layers in order to exploit as many as possible the opportunities offered by Cloud Computing. It simplifies the development of distributed applications by providing: a collection of different ways for expressing the logic of distributed applications, a solid infrastructure that takes care of the distributed execution of applications, and a set of advanced features such as the ability to reserve and price computation nodes and to integrate with existing cloud infrastructures such as Amazon EC2[5,6].

2. Brief Review of Research Papers

In the area of distributed computing, from past many years Grid[6] computing vastly interested scientific community due to its advantages in delivering high-performance services for compute- and data-intensive scientific applications. Although these environments are capable of effectively and comprehensively. modeling Grid environment and applications, none of them provide a clear abstraction of application, virtual and physical machines required by Cloud computing environment. These abstractions are essential to model multi-layer services (SaaS, PaaS and IaaS) of Cloud computing. In these tools, there is almost no support for modeling virtualized resources and application management environment.

InterGrid[3] applied virtualization technology to allow resource sharing among Grids. Resources where sought in remote Grids whenever an incoming allocation request could not be served by local available resources. Grid and Cloud resources are typically differently allocated. While a Grid federation is typically based in cooperation and sharing of resources with time-constrained allocations, Cloud allocations are typically not timely constrained (customers explicitly request resources release) and are based on financial compensation from customers to resource providers. Therefore, solutions for federation aimed at Grid platforms cannot be directly applied to Clouds, and thus cannot be directly applied in the context of InterCloud[3].

Lee et al. [4] propose a queueing network to model SaaS mashup applications whose goal is to maximize profit or reduce costs of the SaaS provider by finding an optimal number of instances for the application. where provisioning is based on performance indicators and elasticity rules defined by users. In both approaches number of instances vary reactively to incoming request rate,

Quiroz et al. [6] propose a mechanism for dynamic VM provisioning in IaaS data centers based on clustering. Specifically, this paper presents a decentralized, robust online clustering approach that addresses the distributed nature of these environments, and can be used to detect patterns and trends, and use this information to optimize provisioning of virtual (VM) resources. It then presents a model-based approach for estimating application service time using long-term application performance monitoring, to provide feedback about the appropriateness of requested resources as well as the system's ability to meet QoS constraints and SLAs. Specifically for high-performance computing workloads, the use of a quadratic response surface model (QRSM) is justified with respect to traditional models, demonstrating the need for application-specific modeling. To deal with the underutilization of physical resources and increasing cost resulting from over provisioning, it proposed the use of a decentralized, robust online clustering approach to drive VM provisioning. It also demonstrated a model-based approach for estimating the application service time given its provisioning in order to deal with inaccurate resource requirements provided by application job requests.

Bi et al. [7] propose a model for provisioning multitier applications in Cloud data centers based on queuing networks. However, such a model does not perform recalculation of number of required VMs based on expected load and monitored performance. Paper present a cloud data center based on virtual machine to optimize resources provisioning. Using simulation experiments of three-tier application, we adopt an optimization model to minimize the total number of virtual machines while satisfying the customer average response time constraint and the request arrival rate constraint. The goal is to minimize the using of resources under a workload while satisfying different customer for the constraints of average response time. Chieu et al. [8] propose a reactive algorithm for dynamic VM provisioning of PaaS and SaaS applications, scaling approach with a front-end load-balancer for routing and balancing user requests to web applications deployed on web servers installed in virtual machine instances. A dynamic scaling algorithm for automated provisioning of virtual machine resources based on threshold number of active sessions will be introduced. The on-demand capability of the Cloud to rapidly provision and dynamically allocate resources to users will be discussed. work has demonstrated the compelling benefits of the Cloud which is capable of handling sudden load surges, delivering IT resources on-demands to users, and maintaining higher resource utilization, thus reducing infrastructure and management costs.

3. Limitation

When considering a large-scale system such as the InterGrid[3], a number of challenges arise, such as resource management among different Grids, varying resource usage across Grids, different security policies, resource reservation and co-allocation by research communities in peered Grids, agreements on QoS requirements and SLAs, and formation and management of VOs in the InterGrid. Paper [4] propose cannot support model proactively applies adaptive provisioning to deliver negotiated QoS to requests whose request arrival rate varies along with the time. proposed scheduling algorithm can bring higher profit while meeting SLAs constraints compared with other revenue-aware scheduling algorithm. In paper [6] , such a work, it is necessary not only to determine the number of virtualized application instances but also their types. In our approach, type of instance is not part of the problem, thus deployed instances can always be used to serve requests. Furthermore, clients (enterprises and SMBs) can easily under or overestimate their needs because of a lack of understanding of application requirements due to application complexity and/or uncertainty. In paper [7] , Working in a cloud result can be change by integrate load prediction method technique to fit workload characteristics. As adopt Service Level Agreement (SLA) based negotiation of prioritized applications to determine the costs and penalties by the achieved performance level. If the entire request cannot be satisfied, some virtualized applications will be affected by their increased execution time, increased waiting time, or increased rejection rate. Paper [8] proposed approach is not proactive in the sense that number of instances is changed based on the expected arrival rate of requests. A dynamic scaling algorithm for automated provisioning of the virtual machine resources based on number of active sessions only.

4. Proposed Problem Statement

The major players in the IT industry such as Google, Amazon, Microsoft, Sun, and Yahoo have started offering cloud computing- based solutions that cover the entire IT computing stack, fromhardware to applications and services. These offerings have become quickly popular and led to the establishment of the concept of "Public Cloud," which represents a publicly accessible distributed system hosting the execution of applications and providing services billed on a pay-per-use basis. After an initial enthusiasm for this new trend, it soon became evident that a solution built on outsourcing the entire IT infrastructure to third parties would not be applicable in many cases, especially when there are critical operations to be performed and security concerns to consider. Moreover, with the public cloud distributed anywhere on the planet, legal issues arise and they simply make it difficult to rely on a virtual public infrastructure for any IT operation. As an example, data location and confidentiality are two of the major issues that scare stakeholders to move into the cloud—data that might be secure in one country may not be secure in another. Private clouds are virtual distributed systems that rely on a private infrastructure and provide internal users with dynamic provisioning of computing resources. Differently from public clouds, instead of a pay-as-you-go model, there could be other schemes in place, which take into account the usage of the cloud and proportionally bill the different departments or sections of the enterprise. Private clouds have the advantage of keeping in house the core business operations by relying on the existing IT infrastructure and reducing the burden of maintaining it once the cloud has been set up. In this scenario, security concerns are less critical, since sensitive information does not flow out of the private infrastructure. Moreover, existing IT resources can be better utilized since the Private cloud becomes accessible to all the division of the enterprise. Another interesting opportunity that comes with private clouds is the possibility of testing applications and systems at a comparatively lower price rather than public clouds before deploying them on the public virtual infrastructure.

5. Conclusion

In this paper, we review the work depend upon cloud environment and find advantages and limitation of works , In proposed idea, we develop hybrid framework for developing distributed applications on the cloud. It harnesses the computing resources of a heterogeneous network of workstations and servers or data centers on demand. This can be a public cloud available to anyone through the Internet, a private cloud constituted by a set of nodes with restricted access within an enterprise, or a hybrid cloud where external resources are integrated on demand, thus allowing applications to scale

References

[1] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging IT platforms: vision, hype, and reality for delivering IT services as the 5th utility, Future Generation Computer Systems 25 (6) (2009) 599–616.delivering IT services as the 5th utility, Future Generation Computer Systems 25 (6) (2009) 599–616.

[2] D. Kondo, A. Chien, H. Casanova, "Scheduling task parallel applications for rapid turnaround on enterprise desktop Grids", Journal of Grid Computing 5 (4) (2007) 379–405.

[3] L. Vaquero, L. Rodero-Marino, J. Caceres, M. Lindner, A break in the clouds: towards a cloud definition, *SIGCOMM Computer Communication Review,* 39 (2009), 137–150.

[4] Y. C. Lee, C. Wang, A. Y. Zomaya, and B. B. Zhou, "Profit-driven service request scheduling in clouds," in Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGrid'10), 2010.

[5] J. Varia, "Best practices in architecting cloud applications in the AWS Cloud," in Cloud Computing: Principles and Paradigms. Wiley, 2011.

[6] A. Quiroz, H. Kim, M. Parashar, N. Gnanasambandam, and N. Sharma, "Towards autonomic workload provisioning for enterprise grids and clouds," in Proceedings of the 10th IEEE/ACM International Conference on Grid Computing (GRID'09), 2009.

[7] J. Bi, Z. Zhu, R. Tian, and Q. Wang, "Dynamic provisioning modeling for virtualized multi-tier applications in cloud data center," in Proceedings of the 3rd International Conference on Cloud Computing (CLOUD'10), 2010.

[8] T. C. Chieu, A. Mohindra, A. A. Karve, and A. Segal, "Dynamic scaling of web applications in a virtualized cloud computing environment," in Proceedings of the 6th International Conference on e-Business Engineering (ICEBE'09), 2009.

