ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Stochastic Model to Investigate Data Center Performance and QOS in IAAS Cloud Computing Systems

Kavitha. P
P.G. Scholar (M. Tech),
Department of CSE,
Modugula Kalavathamma Institute of Technology for Women, Rajampet, Kadapa District.

Sailaja D.L
Assistant Professor,
Department of CSE,
Modugula Kalavathamma Institute of Technology for Women, Rajampet, Kadapa District.

Abstract:

Cloud data center management is a key problem due to the numerous and heterogeneous strategies that can be applied, ranging from the VM placement to the federation with other clouds. Performance evaluation of cloud computing infrastructures is required to predict and quantify the cost-benefit of a strategy portfolio and the corresponding quality of service (QoS) experienced by users. Such analyses are not feasible by simulation or on-the-field experimentation, due to the great number of parameters that have to be investigated. In this paper, we present an analytical model, based on stochastic reward nets (SRNs), that is both scalable to model systems composed of thousands of resources and flexible to represent different policies and cloud-specific strategies. Several performance metrics are defined and evaluated to analyze the behavior of a cloud data center: utilization, availability, waiting time, and responsiveness. A resiliency analysis is also provided to take into account load bursts. Finally, a general approach is presented that, starting from the concept of system capacity, can help system managers to opportunely set the data center parameters under different working conditions.

KEYWORDS:

Performance evaluation, Cost-benefit, Quality of service, Cloud-specific, Resiliency analysis.

1. INTRODUCTION:

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation.

Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers The goal of cloud computing is to apply traditional supercomputing orhigh-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer oriented applications such as financial portfolioos, to deliver personalized information, to provide data storage or to power large, immersive computer games [3]. The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.Cloud computing is a promising technology able toresources will be accessed in the near future. Throughthe provision of on-demand access to virtual resourcesavailable on the Internet, cloud systems offer services atthree different levels: infrastructure as a service (IaaS),platform as a service (PaaS), and software as a service(SaaS). In particular, IaaS clouds provide users withcomputational resources in the form of virtual machine(VM) instances deployed in the provider data center, whilePaaS and SaaS clouds offer services in terms of specificsolution stacks and application software suites, respectively.To integrate business requirements and application-levelneeds, in terms of quality of service (QoS), cloud serviceprovisioning is regulated by service-level agreements(SLAs): contracts between clients and providers that expressthe price for a service, the QoS levels required during theservice provisioning, and the penalties associated with theSLA violations. In such a context, performance evaluationplays a key role allowing system managers to evaluate theeffects. of different resource management strategies on thedata center functioning and to predict the correspondingcosts/benefits.

Cloud systems differ from traditional distributed systems. First of all, they are characterized by a very large number ofresources that can span different administrative domains.Moreover, the high level of resource abstraction allows us toimplement particular resource management techniques such as VM multiplexing or VM live migration [3] that, even iftransparent to final users, have to be considered in the designof performance models to accurately understand the systembehavior. Finally, different clouds, belonging to the same orto different organizations, can dynamically join each other toachieve a common goal, usually represented by theoptimization of resources utilization. This mechanism,referred to as cloud federation [4], allows us to provide andrelease resources on demand, thus providing elastic capabilitiesto the whole infrastructure.

Characteristics and Services Models:

The salient characteristics of cloud computing based on the definitions provided by the National Institute of Stan¬dards and Terminology (NIST) are outlined below:

•On-demand self-service: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requir¬ing human interaction with each service's provider.

•Broad network access: Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client plat¬forms (e.g., mobile phones, laptops, and PDAs).

•Resource pooling: The provider's computing resources are pooled to serve multiple consumers using a multi-ten¬ant model, with different physical and virtual resources dynamically assigned and reassigned according to con¬sumer demand. There is a sense of location-independence in that the customer generally has no control or knowl¬edge over the exact location of the provided resources but may be able to specify location at a higher level of ab¬straction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

•Rapid elasticity: Capabilities can be rapidly and elasti¬cally provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning of¬ten appear to be unlimited and can be purchased in any quantity at any time.

•Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering ca¬pability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be managed, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

LITERATURE SURVEY:

Live migration of virtual machines (VM) across physical hosts provides a significant new benefit for Administrators of data centers and clusters. Previous memory-to-memory approaches demonstrate the effectiveness of live VM migration in local area networks (LAN), but they would cause a long period of downtime in a wide areanetwork (WAN) environment [9]. This paper describes the design and implementation of a novel approach, namely, CR/TR-Motion, which adopts check pointing/recovery and trace/replay technologies to provide fast, transparent VM migration for both LAN and WAN environments. With execution trace logged on the source host, a synchronization algorithm is performed to orchestrate the running source and target VMs until they reach a consistent state. CR/TRMotion can greatly reduce the migration downtime and network bandwidth consumption [8].

Cloud computing aims to power the next generation data centers and enables application service providers to lease data center capabilities for deploying applications depending on user QoS (Quality of Service) requirements. Cloud applications have different composition, configuration, and deployment requirements. Quantifying the performance of resource allocation policies and application scheduling algorithms at finer details in Cloud computing environments for different application and service models under varying load, energy performance, and system size is a challenging problem to tackle [11].Cloud computing is an emerging infrastructure paradigm that promises to eliminate the need for companies to maintain expensive computing hardware. Through the use of virtualization and resource time-sharing, clouds address with a single set of physical resources a large user base with diverse needs.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Thus, clouds have the potential to provide their owners the benefits of an economy of scale and, at the same time, become an alternative for both the industry and the scientific community to self-owned clusters, grids, and parallel production environments. For this potential to become reality, the first generation of commercial clouds needs to be proven to be dependable. In this work we analyze the dependability of cloud services. Towards this end, we analyze long-term performance traces from Amazon Web Services and Google App Engine, currently two of the largest commercial clouds in production .Advanced computing on cloud computing infrastructures can only become viable alternative for the enterprise if these infrastructures can provide proper levels of non functional properties (NPFs). A company that focuses on service-oriented architectures (SOA) needs to know what configuration would provide the proper levels for individual services if they are deployed in the cloud. In this paper we present an approach for performance evaluation of cloud computing configurations. While cloud computing providers assure certain service levels, this it typically done for the platform and not for a particular service instance. Cloud Computing is emerging today as a commercial infrastructure that eliminates the need for maintaining expensive computing hardware. Through the use of virtualization, clouds promise to address with the same-shared set of physical resources a large user base with different needs. Thus, clouds promise to be for scientists an alternative to clusters, grids, and supercomputers. However, virtualization may induce significant performance penalties for the demanding scientific computing workloads. In this work we present an evaluation of the usefulness of the current cloud computing services for scientific computing. We analyze the performance of the Amazon EC2 platform using micro benchmarks and kernels. While clouds are still changing, our results indicate that the current cloud services need an order of magnitude in performance improvement to be useful to the scientific community [18].

III. PROBLEM IDENTIFICATION:

In cloud environment the server utilization is not properly maintained. For example, a web site has three web servers; each server is capable of handling 1 lacks requests at a time. It is not mandatory that all servers need to beonline even when there is no enough request. This leads to high cost, because deploying the Iaas need more cost. This project explains the way of solving this problem in the efficient way.

Another problem occurs if the request limit rises immediately, this may lead to crash all the active servers at a time. Recently “Flipcart.com” site has been crashed due to high bandwidth at a time [14].

IV. EXISTING SYSTEM:

In order to integrate business requirements and application level needs, in terms of Quality of Service (QoS), cloud service provisioning is regulated by Service Level Agreements (SLAs) : contracts between clients and providers that express the price for a service, the QoS levels required during the service provisioning, and the penalties associated with the SLA violations. In such a context, performance evaluation plays a key role allowing system managers to evaluate the effects of different resource management strategies on the data center functioning and to predict the corresponding costs/benefits Cloud systems differ from traditional distributed systems. First of all, they are characterized by a very large number of resources that can span different administrative domains. Moreover, the high level of resource abstraction allows to implement particular resource management techniques such as VM multiplexing or VM live migration that, even if transparent to final users, have to be considered in the design of performance models in order to accurately understand the system behavior.

DISADVANTAGES OF EXISTING SYSTEM:

•On-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider.
•Simulation does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated.

V. PROPOSED SYSTEM:

Here is the proposed system based onStochastic Reward Nets (SRNs), that exhibits the above mentioned features allowing capturing the key concepts of an IaaS cloud system [29]. The proposed model is scalable enough to represent systems composed of thousands of resources and it makes possible to represent both physical and virtual resources exploiting cloud specific concepts such as the infrastructure elasticity.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

With respect to the existing literature, the innovative aspect of the present work is that a generic and comprehensive view of a cloud system is presented. Low level details, such as VM multiplexing, are easily integrated with cloud based actions such as federation, allowing to investigate different mixed strategies. An exhaustive set of performance metrics is defined regarding both the system provider (e.g., utilization) and the final users (e.g., responsiveness).

SYSTEM ARCHITECTURE:

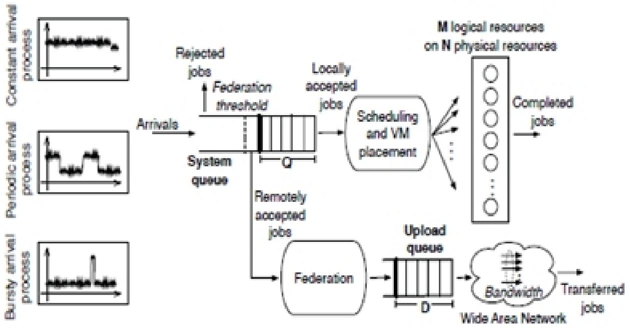

Fig. 1. An IaaS cloud system with federation.

ADVANTAGES OF PROPOSED SYSTEM:

To provide a fair comparison among different resource management strategies, also taking into account the system elasticity, a performance evaluation approach is described. Such an approach, based on the concept of system capacity, presents a holistic view of a cloud system and it allows system managers to study the better solution with respect to an established goal and to opportunely set the system parameters.

VI. MODULES:

There are five modules in this system:
1. System Queuing
2. Scheduling Module
3. VM Placement Module
4. Federation Module
5. Arrival Process

A. System Queuing:

Job requests (in terms of VM instantiation requests) are en-queued in the system queue. Such a queue has afinite size Q, once its limit is reached further requests are rejected. The system queue is managed according to a FIFO scheduling policy.

B. Scheduling Module:

When a resource is available a job is accepted and the corresponding VM is instantiated. We assume that theinstantiation time is negligible and that the service time (i.e., the time needed to execute a job) is exponentiallydistributed with mean1/μ.

C. VM Placement:

According to the VM multiplexing technique the cloud system can provide a number M of logical resourcesgreater than N. In this case, multiple VMs can be allocated in the same physical machine (PM), e.g., a core in amulticorearchitecture. Multiple VMs sharing the same PM can incur in a reduction of the performance mainly due toI/O interference between VMs.

D. Federation Module:

Cloud federation allows the system to use, in particular situations, the resources offered by other public cloudsystems through a sharing and paying model. In this way, elastic capabilities can be exploited in order to respond toparticular load conditions. Job requests can be redirected to other clouds by transferring the corresponding VM disk images through the network .

E. Arrival Process:

Finally, we respect to the arrival process we will investigate three different scenarios. In the first one (Constantarrival process) we assume the arrival process be a homogeneous Poisson process with rate λ. However, large scaledistributed systems with thousands of users, such as cloud systems, could exhibit self-similarity/long-range dependencewith respect to the arrival process. The last scenario (Bursty arrival process) takes into account the presence of a burstwhit fixed and short duration and it will be used in order to investigate the system resiliency.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

VII. RESULTS:

Figure 2 .Acting Server

Figure 3. Secure Details Editing

VIII.CONCLUSION AND FUTURESCOPE:

A stochastic model was presented here to evaluate the performance of an IaaS cloud system. Severalperformance metrics have been defined, such as availability, utilization, and responsiveness, allowing investigatingthe impact of different strategies on both provider and user point-of-views. In a market-oriented area, such as the CloudComputing, an accurate evaluation of these parameters is required in order to quantify the offered QoS and opportunelymanage SLAs. Future works will include the analysis of autonomic techniques able to change on-the fly the systemconfiguration in order to react to a change on the working conditions. This can also extend the model in order to represent PaaS and SaaS Cloud systems and to integrate the mechanisms needed to capture VM migration and datacenter consolidation aspects that cover a crucial role in energy saving policies.

REFERENCES:

1. R. Buyyaet al., "Cloud computing and emerging it platforms :Vision, hype, and reality for delivering computing as the 5thutility," FutureGener. Comput. Syst., vol. 25, PP. 599–616, June 2009.
2. X. Menget al., "Efficient resource provisioning in compute clouds via vm multiplexing," in Proceedings of the 7th internationalconference on Autonomic computing, ser. ICAC '10. New York, NY, USA:ACM, 2010, pp. 11–20.
3. H. Liu et al., "Live virtual machine migration via asynchronous replication and state synchronization," Parallel and Distributed Systems,IEEE Transactions on, Vol. 22, No. 12,PP. 1986 –1999, December - 2011.
4. B. Rochwergeret al., "Reservoir - when one cloud is not enough, "Computer, vol. 44, no. 3, pp. 44 –51, march 2011.
5. R. Buyya, R. Ranjan, and R. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloud sim toolkit:Challenges and opportunities," in High Performance Computing Simulation, 2009. HPCS '09. International Conference on,PP. 1 – 11,June - 2009.
6.
6. A. Iosup, N. Yigitbasi, and D. Epema, "On the performance variability of production cloud services," in Cluster, Cloud and GridComputing (CCGrid), 2011 11th IEEE/ACM International Symposiumon, may 2011, pp. 104 –113.
7. K.G.S. Venkatesan and M. Elamurugaselvam, "Design based object oriented Metrics to measure coupling & cohesion", International journal ofAdvanced & Innovative Research, Vol. 2, Issue 5, PP. 778 – 785, 2013.
8. Teerawat Issariyakul • Ekram Hoss, "Introduction to Network Simulator NS2".
9. S. Sathish Raja and K.G.S. Venkatesan, "Email spam zombies scrutinizer in email sending network Infrastructures", Internationaljournal of Scientific & Engineering Research, Vol. 4, Issue 4, PP. 366 – 373, April 2013.
10. G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE J. Sel. Areas Communication., Vol. 18, No. 3, PP. 535–547, Mar. 2000.
11. K.G.S. Venkatesan, "Comparison of CDMA & GSM Mobile Technology", Middle-East Journal of Scientific Research, 13 (12),PP. 1590 – 1594, 2013.
12. P. Indira Priya, K.G.S.Venkatesan, "Finding the K-Edge connectivity in MANET using DLTRT, International Journal of Applied EngineeringResearch, Vol. 9, Issue 22, PP. 5898 – 5904, 2014.