International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Review on Provisioning of Services in Cloud Computing

Mridul Paul[1], Ajanta Das[2]

[1, 2]Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Kolkata Campus, *Kolkata – 700 107, India*

Abstract: ***Cloud computing has been one of the most profound discoveries of this century. It has posed society with solution to extend computing beyond boundaries within organizations and relook at leveraging outsourced infrastructures for their computational needs. With scale and mass, comes the advantage of flexible infrastructure at reduced costs. At the same time, the complexity of service provisioning has increased exponentially. Many researchers have come up with novel techniques and algorithms in order to maintain Quality of Service (QoS) with reduced provisioning costs. With ever increasing adoption of Cloud services by consumers, demands have multiplied and pose challenges around Service Level Agreement (SLA) measurement and monitoring to maintain QoS. In this paper, a detailed review on some of the state-of-the-art of service provisioning techniques and algorithms in cloud computing is presented. First few sections focus on the cloud evolution and taxonomy to bring clarity on expectations from and complexity of Cloud Computing and remaining sections synthesize and summarize different provision techniques, approaches, and models through a comprehensive literature review.***

Keywords: Cloud Computing, Service Provisioning in Cloud, Resource Management, Service Level Agreement and Quality of Service

1. Introduction

Cloud computing is becoming increasingly popular for its high availability, economic viability and scalability. Numerous researchers and scientists are focusing on devising revolutionary techniques to take cloud computing to the next level. Major area of interest has been service provisioning where by several hypotheses have been put forward, tested and implemented by service providers. Service providers are looking forward on relying on newer techniques to reduce underlying costs. Underlying resources are common to any service provider which is server, storage, network and VMs. Service providers are seeking state of the art algorithms and techniques to enable seamless orchestration to provide optimal service. At the same time maintaining Quality of Service (QoS) as agreed with the end consumers is always challenging. Needless to say, Service Level Agreements (SLAs) are becoming increasing stringent and service providers are focusing on automations and predictive analysis to avoid penalties. Hence the service providers are seeking to refine SLAs before signing those with the end consumers.

Elastic platforms are becoming more and more popular and start to be a viable alternative to host distributed applications and web applications in particular [1]. In a typical cloud infrastructure, a user can rent virtual machines (VM) and allocate dedicated resources (such as CPU cores, RAM, disk space, etc.) to them in order to closely match the application needs. Moreover, through cloud infrastructure application programming interface (API), users are able to programmatically increase or decrease the resources allocated to a virtual machine, and can also start new VMs or stop unused ones. Virtualization technologies that have greatly contributed to the success of cloud computing also come with some drawbacks. For example, in today's public cloud infrastructures, physical resources are controlled and managed by virtual machines. These virtual machines may or may not have all physical resources in one location or hub. This may result in decreased performance as expected from condition when physical resources were in same location. In effect, an end consumer does not have full control over the underlying infrastructure.

The key expectations of cloud services [1] are as follows –

- Scalability and agility: From inception till retirement, any service is expected to scale to any level during its lifetime. This non-functional requirement has underlying principle of agility, flexible enough to start/stop at any location and replicate/move from one location to another (portability).
- Availability and reliability: In today's competitive environment where end consumers are located in different geographies accessing services any time. "Lights on" services running in 24 X 7 timelines need robust infrastructure and techniques to make services reliable.

The remainder of this paper is organized as follows: Section II presents evolution of cloud computing and cloud service is described in Section III. Section IV describes cloud service related taxonomy. A review of some existing research work in the area of service provisioning is explained in Section V. Section VI describes the importance of Service Level Agreement, followed by that Section VII describes research challenges. Section VIII, concludes the paper.

2. Evolution of Cloud Computing

The growth of high speed computing power and connecting network led to the evolution of Grid computing. Grid computing changed the way the society managing and processing information services. It enabled geographically distributed resources such as supercomputers, storage systems, data sources; instruments and special devices to not only connected together, but also perform co-ordinated meaningful execution of set of activities. These activities could be treated as utility which users can consume (same as electricity, gas and water). In 1995, National Centre of

Supercomputing Applications, Argonne National Lab, the San Diego Supercomputing Centre and Sandia National Lab, collaborated to form I-WAY (International Wide-Area Year) which was a testbed for distributed virtual supercomputing. Followed by subsequent trial and tribulations, this led to further spinning of several research projects (such as Globus) which took grid computing to next level [2]. Eventually with the Internet taking a hot seat and boundaries of services getting further refined, the Cloud Computing started to foray into distributed computing. Thus Cloud Computing addressed both applications and hardware delivered as a service over internet [3].

Cloud computing as defined by The National Institute of Standards and Technology's [4] has five basic and essential characteristics which are as follows:

- On-demand self-service: Cloud Computing needs to allow consumers to choose unilaterally provision computing capabilities, such as server time and network storage, without any human intervention.
- Broad network access: Computing capabilities need should be available to the consumers over standard network protocols on any type of client platforms (e.g., mobile phones, tablets, laptops, and workstations)
- Resource pooling: Cloud providers should be able to serve multiple consumers in a multi-tenant model whereby resources that are released by one consumer can be utilized for serving another
- Rapid elasticity: Computing capabilities should be easily provisioned on request for demand and released on demand fulfillment which should be taken care automatically. This must be transparent for end consumer for whom this would appear to be indefinite capacity computing at his/her disposal
- Measured service: Cloud services should automatically manage, control and optimize underlying resources such as storage, processing, network bandwidth and administer services to consumers on usage basis. This metering mechanism should be transparent and mutually agreed with both parties (provider and consumer).

With Cloud Computing, organizations are undergoing major transformation in the way computing is performed. Industries such as Manufacturing, Pharmaceutical & Life sciences, Travel & Logistics, Banking, Finance, Insurance, are impacted with this powerful concept. Nevertheless, organizations are either already pursuing cloud computing or are in the fray of evaluating how the benefits can be obtained from such a revolution. Companies such as Amazon, Google, and Microsoft are investing heavily on the research in this area. Given the momentum Cloud computing has gained over the years, commercialization of the services on cloud computing has been gained prevalence. With several layers of services – infrastructure, platform and software, computing organizations are focusing on expanding services and reducing pay as you go costs. Figure 1 depicts cloud environment and key components of the cloud. Service consumers connect over internet to request for service and Service providers in turn deploy infrastructure to provide services as requested. However as the complexity is arising with this service layers, areas such as Service Provisioning and Service Level Agreements need to be relooked at.

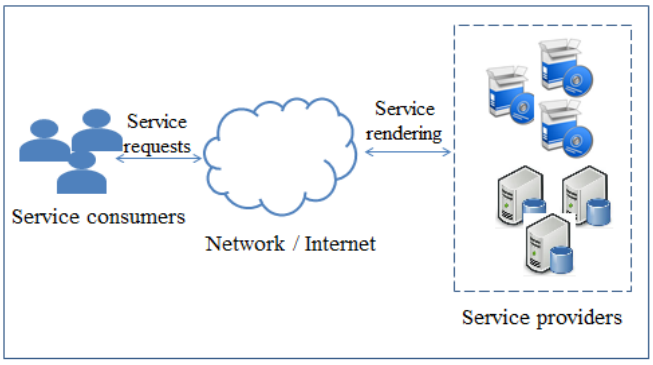

Figure 1: Cloud Environment

3. Cloud Services

As Cloud computing evolved, different categories of services emerged which could be tailor made for consumers. Service providers found ways to organize aggregate and bundle resources to provide such services. Such services were only made reality through virtualization which was a key. Figure 2 describes three distinct layers of service provisioning.

Infrastructure as a Service [5] are most basic and mature market for cloud offering. Also known as Hardware as a Service, the main technology used to deliver such a service is the hardware virtualization. One or more virtual machines configured and interconnected define the distributed system on top of which applications are installed and deployed. Virtual machines constitute atomic components having specific features: memory, number of processors, and disk storage. A mature example of IaaS is Elastic Computing Cloud of Amazon (Amazon EC2) [6] and storage by Elastic Book Store (EBS) and Simple Storage Services (S3); a platform developed by Amazon which was carved out of the existing infrastructure the company had which was underutilized.

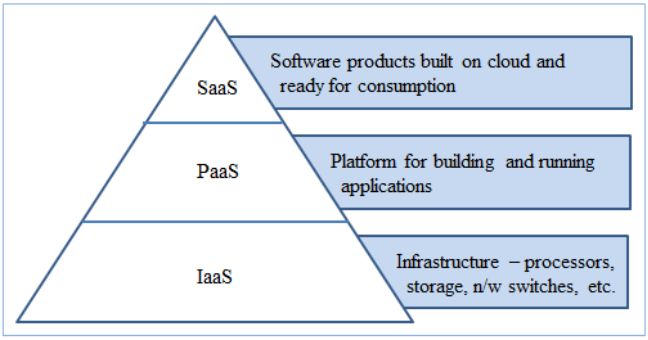

Figure 2: Cloud Service Layers [4]

Platform as a service [5] defines a class of cloud service provider that offers an additional layer of abstraction above the virtualized infrastructure. PaaS solutions provide a development and deployment platform for running applications in the Cloud. These constitute the middleware on top of which applications are built. Application management is the functionality of the middleware. PaaS implementations provide applications with a runtime environment and do not expose any service for managing the underlying infrastructure. Hence service consumers have to

rely on PaaS service providers on providing optimal hardware infrastructure for PaaS layer. The layer automate the process of deploying applications to the infrastructure, configuring application components, provisioning and configuring supporting technologies such as load balancers, databases and managing system change based on the policies set by the consumer. An example of PaaS is Google AppEngine [7] serves as Platform as a Service which provides hosting platform for Web applications. AppEngine leverages underlying distributed and scalable runtime infrastructure which is in the form of servers available within Google datacenters.

Software as a Service [5] is a software delivery model providing access to applications through the Internet as Web based service. This service provides a means to liberate users from complex hardware and software management by offloading such tasks to third parties, who build applications accessible to multiple users through a Web browser. The consumers neither install any software on their premises nor have to pay considerable upfront costs to purchase softwares and required licenses. Salesforce.com is one the best examples of SaaS. Salesforce.com [8] as a company has matured over years in providing software services over the cloud. This company provides platform for managing customer relationships. Its offering include Service Cloud which aims to facilitate creation, tracking and routing of customer cases within organizations. It also provides AppExchange which a based on marketplace concept to provide web application for consumers and Collaboration Cloud (including Chatter) to connect among employees within organizations.

4. Service Provisioning Taxonomy

Cloud computing paradigm revolves around services. Any function that can be consumed by end users that is hosted by a cloud service provider can be considered as a Service. Figure 3 captures various taxonomies that are applicable to Cloud services. A service can be classified into three broad areas of Cloud topology which are: Software, Platform and Infrastructure which is explained in details in previous section. These three layers are stacked in that order from top to bottom. However when an consumer approaches a cloud service provider for Software as a Service offerings, then the service provider offers SaaS services with an abstraction of below two layers (PaaS and IaaS), providing transparency to the consumer and total ownership cost (shifted to service provider) with complete freedom to pay for what is consumed.

Usually, service providers offer services on single cloud where resources lay within one domain. Such service providers manage resources well within defined boundary and ensure end to end service to the consumers. In certain cases, instead of consuming services from single cloud, the consumer may choose to consume services in multiple clouds (federated cloud). Such a topology is viable to assimilate best of breed cloud services hosted by various cloud service providers.

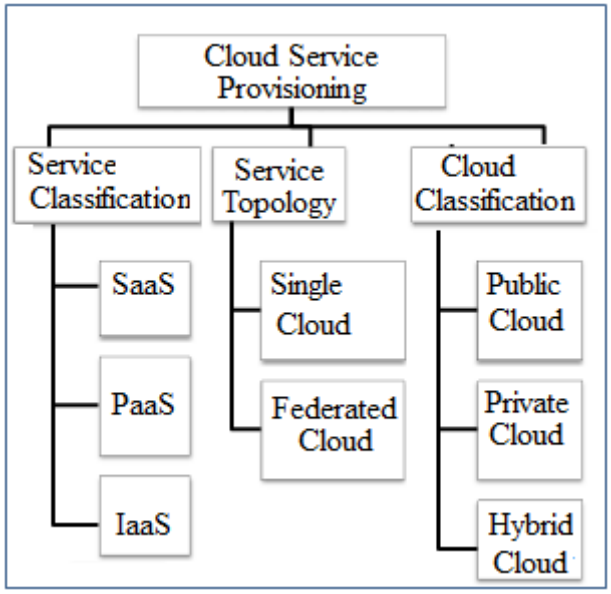

Figure 3: Cloud Service Provisioning Taxonomy [5]

Cloud is further classified into Private, Public and Hybrid cloud. Private cloud has infrastructure which is established with the aim of catering requirements for a single organization, whether managed internally or by a third-party, and hosted either within organization boundaries or outside. A Private cloud initiative needs a significant level and degree of engagement to virtualize the business environment, and requires the organization to reevaluate decisions about existing resources. However most widely used cloud platforms used today are Public cloud. Public cloud requires similar setup as needed for a Private cloud; however the target audience is public in case of former. One notable difference between Private and Public cloud is the security considerations which may differ for services rendered. Hybrid cloud answers to some questions where organizations want to maintain tighter security to its mission critical data and at the same time leverage public cloud to expose services to general users.

5. Review of Service Provisioning Models

In this section, a study of some existing service provisioning models and techniques is discussed.

Some researchers focused on improving provisioning on Web applications on cloud. A noted work from Nicolas Bonvin et. al. [9, 10] proposed and demonstrated that adaptive mechanism of service provisioning has higher benefits than static mechanism. The components that form Web application was taken into consideration in this research. In static service provisioning, the number of instances of underlying infrastructure becomes fixed. Hence the components which reside on virtual machines (which in turn can reside on same or different physical servers) have fixed on capacity. Usually this capacity is derived from the peak load which Web application may go through. The drawback of such provisioning is that the underutilization of resources becomes evident and during load spikes, further resource planning for underlying resources needs to be done every time. This research is centered on concept of economic

fitness of the components defined as utility provided by that components versus cost of retaining that component in the cloud. Based on this fitness, it is decided on whether that component needs to stay in the cloud or be moved from one VM to another or be destroyed. In order to accomplish this mechanism, an approach was proposed which has agents that run on physical server. These agents are responsible for handling lifecycle of the components and also perform health checks.

Mathais Bjorkqvist et. al. [11, 12] proposed opportunistic way of managing resources for Web applications. The approach centered on maintaining target system utilization through predictive workload and VM performance. Web applications are based on Service Oriented Architecture where applications are composed of atomic services which are typically Web services and service composition execution engines (middleware for executing composition of atomic services) that reside on Virtual Machines (VMs). The objective of this research was to ensure use of lesser number of faster virtual machines (VMs) to be used at any given time rather than more number of slower VMs for servicing requests from the consumers. Though through put for both options mentioned earlier should be the same however the cost of maintaining former will certainly be lesser than the latter option. The architecture which was considered, took into account of a Virtual Machine Broker which decides on number of VMs where the servers will be running. VM Controller controls the VMs where the services and its replicas are running. Load balancers typically distribute incoming service requests among service replicas. The replication policy is targeted to be implemented on the VM broker. The policy enable VM broker to first decide number of VMs for a service and then to select VMs based on performance of the active VMs and billing periods.

Rodrigo N Calheiros et. al. [13] focused on the problems encountered in the area of virtualization, workload and performance modeling, deployment and monitoring of applications on cloud. These problems are driven by unpredictable behavior of virtualized IT resources and network elements and eventually lead to overprovisioning of underlying resources and negative impact on Quality of Services (QoS). The solution proposed was modeled on analytical performance model and workload information. Both these areas feed in data to Application provisioner that handles application and Virtual Machine provisioning. Analytical performance model was used to predict effectiveness of provisioning schedule on desired Quality of Service (QoS). Workload information provides application provisioner demands for resourcing needs, removing uncertainties and over/under estimation of Cloud resources. The algorithm proposed by them to calculate optimal number of virtual machines is used by the Load predictor and Performance modeler.

Andreas Lodde et al. [15] stressed on service response time as a parameter for measuring and controlling service provisioning. A history of requests and associated response time is maintained where services are hosted. The underlying framework devised, has a key component - request classifier which takes in service requests, collects required information and appends requests with header information. The request scheduler defines the sequence of request processing and a dispatcher acts on the guidelines to send requests to the computing resource. In order to calculate optimal service provisioning resources, dimension evaluator sends number of VM instances needed which a processing simulator verifies using history of processed requests, and SLA evaluator checks on whether SLA is complied with. This process is iterative till SLA evaluator confirms on SLA compliance, once confirmed, dimension evaluator ends its process and this new number is applied to the cloud resources for servicing requests. The key aspect that drives above framework is the history of requests that is stored which forms basis of calculating optimal resourcing. Be it processed requests or requests waiting to be processed, history data is maintained for processing simulator to verify numbers proposed by dimension evaluator. The dimension evaluator uses binary search algorithm to determine minimum possible number of instances needed to service requests.

Kuo-Chan Huang et. al [16] highlight issues that arise from software based distributed workflows and online concurrent user access and addresses important issues pertaining to resource allocation and dynamic provisioning of services in cloud. A service deployment strategy was devised based on application service flows and a resource provisioning technique based on future service requests. The strategy detects queued application service flow instances and calculates number of remaining service type to be executed. Then the resources are adjusted based on the each service type. Two scenarios were taken into consideration, one where number of virtual machines was greater that the number of service types deployed. Therefore each service type can be assigned one or more VMs. In the other computing scenario, multiple service types are assigned to one VM. The provisioning mechanism estimates based on three different policies which are based on the current, short term and long term system workload and nature of remaining service flows.

At times web applications are hosted on cloud. Even though service providers offer SLAs to the consumers, however that is not enough to guarantee response time for web applications. It also becomes complex when there is multitier applications hosted which have Web server tier for consumer requests, application tier for business specific calculations and logic and data tier for storing persistent data. Currently neither the commercial cloud providers nor the existing open source providers support maximum response time guarantees. Waheed Iqbal et. al. [17] researched on this area to provide guarantees on maximizing response time for web applications with minimum resource utilization. The algorithms designed as a part of this research can detect bottleneck in multi-tier Web applications hosted on a cloud. This was done through profiling of the CPU, memory, and I/O resource usage of each tier with a combination of real time monitoring and processing of each tier's log files. Usually web applications need configuration of concurrency levels for Database tier connections, threads to Application tier and worker processers in Web tier. Any inaccuracy on these levels create bottleneck. The algorithm focuses on heuristic analysis done for multi-tier web application and on detection of a bottleneck the solution is either administered through

horizontal scaling the tier with load balancing or by vertical scaling by dynamically increasing underlying resources.

Mahyar Movahed Nejad et. al. [18] focus on problems faced by cloud service providers on virtual machine provisioning, especially in the area of auction based models. The cloud service providers provision resources either on static or dynamic provisioning basis. In static provisioning, the service provider provisions set of virtual machines even before user requests come in. In dynamic provisioning, the service provider considers user demands before provisioning virtual machines. To sell its services, the service providers either use fixed pricing approach or auction based approach. In fixed pricing, the price for each type of VM is fixed and pre-determined by the service provider which hardly changes over time. However, auction based pricing enables service providers to bundle available VMs and use auction to fix price and allocation to the users. This creates a win-win model from service provider and consumer. The service consumer can get resources at lower price than fixed price model and service providers get increased revenue for unutilized VMs. Auction based model was considered in this research and formulated an integer program that considers different types of resources while performing dynamic VM provisioning. This model setup is slightly different from service providers such as Amazon where auction is carried out for short time and for individual VM instances (rather than bundles). Also that wining users pay same price (per unit) for VMs instances. The model that was proposed will bundle available VMs instances and open it for auction to set of users. Each user shall assign private price for bundle and pays if the bid is successful. The underlying uniqueness of this model is that it considers different resources such as processer cores, memory, storage, etc. which is a reality in today's world.

Another interesting area of solving service provisioning problems is application of game theory. G. Wei et. al. [19] proposed game theory to solve resource allocation in cloud services. The proposition had two steps, first each participant tries to solve the problem independently without considering resources getting allocated by other participants. This was achieved through Binary Integer Programming method for individual optimization of resource allocation. Second step is to combine multiplexed strategies of those individual optimal solutions with minimum loss of efficiency. It is also demonstrated that Nash equilibrium is attained during this optimal resource allocation game. In fact generalized Nash equilibria were further extended by Danilo Ardagna et. al. [20] in solving problems faced by SaaS providers who host applications at an IaaS provider. Each SaaS provider has to conform to QoS standards as per agreed SLA which determines net revenues for the services provided, and also at same time it has to ensure minimum cost of using resources provided by IaaS providers. So in order to balance between services rendered by SaaS and service consumed from IaaS provider, this research attempted to devise efficient distributed algorithm based on General Nash Equilibrium for runtime allocation of IaaS resources among competing SaaS providers. The cost model of the algorithm consists of utility functions, which has both revenues received and penalties incurred during the achieved performance of the resource allocation and also includes infrastructural costs associated with IaaS resources. In similar lines, Valerio Di Valerio et. at. [21] also focused on SaaS provider problems and used Stackelberg game approach to devise solution which derives equilibrium price and allocation strategy through Mathematical Program with Equilibrium Constraints(MPEC) problem. The scenario considered here was an IaaS provider offers resources to several SaaS providers, which is on flat pricing basis, is on demand, and for spot VM instances. In turn, SaaS providers offer to end users Web applications with Quality of Service (QoS) guarantees, using the IaaS facilities to host and run the provided applications. For SaaS provider, revenues and penalties depend on provisioning of resource at an adequate performance level, which is specified in a Service Level Agreement (SLA) contract that each SaaS provider conspires with its end users. Therefore, each SaaS provider fasces the problem of determining the optimal number of VMs to satisfy the SLA with his end users while maximizing revenue. A two stage provisioning scheme was proposed. In the first stage, the SaaS providers determine the number of required flat and on demand instances by means of standard optimization techniques. In the second stage the SaaS providers compete, by bidding for the spot instances which are instantiated using the unused IaaS capacity. It was assumed that the SaaS providers want to maximize a suitable utility function which accounts for both the QoS delivered to users and the associated cost. This stage is modelled on Stackelberg game theory which provides way to derive equilibrium solution for maintaining QoS of service delivered and keeping the allocations to optimal.

6. Importance of Service Level Agreement in Cloud

From service design to service rendering, Service Level Agreement (SLA) plays a pivotal role. It is pertinent to have ground rules set in before a service gets consumed. These rules are used by service providers to define nature of service based on the available capacity and vulnerability for hardware layers. Various aspects that need to be considered while designing are network, security, storage, processing power, and database. SLA is legal agreement or a contract which bound service providers on clause pertaining to Quality of Service [15]. It forms basis of how a service is rendered by service provider and consumed by service consumer. The agreement has following important characteristics:

- describes clearly a service so that consumer can understand functionalities of the service
- articulates the performance levels of the service
- defines mode by which the service parameters can be monitored and reported
- imposes penalties in case the service does not meet service requirements

In order to set specific SLAs, service providers need to capture service requirements. The requirements take the form of parameterized set of data which is provided by service consumers. This parameterized set consists of start/stop time of the service, duration of uptime, quantity, throughput/speed, etc. The service provider has to consider

each of these data and render service such that these data is complied with. These SLAs can be categorized based on the following criteria -

- Speed: performance criterion that captures rate at which a service can be consumed
- Availability: service characteristic that encapsulates likelihood of responding to service consumers whenever accessed
- Accuracy: completeness and worthiness of the results from the service

Figure 4 depicts various parameters that are communicated between a service provider and consumer. While service is getting delivered over a period of time, the Quality of Service (QoS) can be measured through verification of SLA based on the criteria mentioned in the following-

- Reliability: ability to perform required function under stated function
- Flexibility: options provided by service provider on this service offered
- Capability: ability to meet demand of a given size under internal conditions
- Usability: effectiveness and efficiency with which a service consumer can accomplish specified tasks

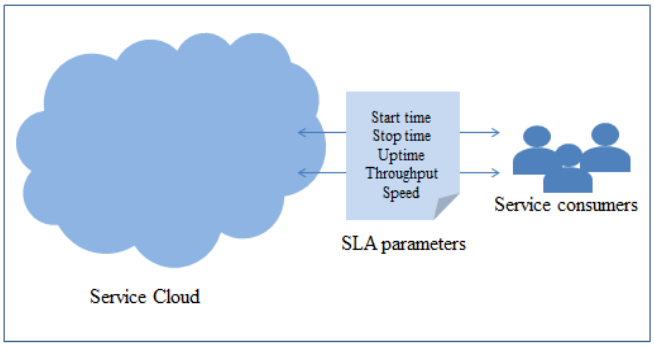

Figure 4: Service Cloud and associated SLAs

Once the service has been designed and deployed based requirements of service consumers, service providers measure SLAs and track in the form of metrics which is shared with service consumers from time to time. However an SLA cannot convert a good service from a bad service, but can mitigate risk from choosing a bad service. There are several tools and framework to manage various aspects of SLA lifecycle such as for SLA specification and modelling, tools such as Cloudscale, Viola can be leveraged, for SLA enforcement, tools such as CloudSOA, Stream can be chosen and for SLA management, Plane, SmartLM can be used.

7. Research Challenges

The research papers reviewed focus primarily on virtual machine provisioning as almost all service providers do not let underlying cloud infrastructure visible to outside world. The algorithms proposed by the researchers on economic fitness of the existing VMs and deciding on whether to abort slower running VMs or to start new instances. From the already completed work, it could be found that service availability is one of primary challenges in current research on this topic. Service consumption is occurring between service providers and consumer over a network. Hence service disruption, network congestion, poor signal or even a node failure are highly intolerable in service provisioning. It has been also noted that with the evolution of cloud infrastructures, supports for complex mathematical models for virtual machine allocation and maintain high Quality of Service (QoS), provides an opportunity to implement sophisticated scheduling algorithms. In addition to enhancements to existing scheduling algorithms, new methodologies could be applied, such as the adaptive virtual machine scheduling and SLA driven models for virtual machine allocation.

Since the complexity of Cloud computing is phenomenal given the size of consumer requests and underlying infrastructure that supports such requests, it will interesting to see how existing algorithms behave in unstable cloud environment where underlying resources are changed. Also another key aspect of service provisioning is Service Levels Agreements (SLAs) which need to take into account while refining service provisioning algorithms. SLA measurement and monitoring has to be considered while purview of service provisioning algorithms to cover 360 degree view of the service provided and consumed.

8. Conclusion

Cloud computing is an emerging paradigm that is revolutionizing the way computing is performed and administered. With virtualization of physical resources, infrastructure, and applications, cloud service provisioning has become a reality. The growing adoption of cloud services by all industries indicates underlying value proposition which cloud computing carries. Due to rise in expectations in service provisioning, providing effective and continuous service in cloud is getting difficult. Therefore it is pertinent to understand service provisioning fundamentals, taxonomy, and several expectations that must be considered to evaluate the provisioned services in terms of user requirements and costs. The Service Provisioning algorithms and techniques reviewed in this paper provide insights on optimizing underlying resources for efficient provisioning of services. Besides, Service Level Agreement which forms basis of service provisioning formulations and consumption has been articulated. Finally, open research challenges are categorized and identified for future research directions.

References

[1] Md Whaiduzzaman, Mohammad Nazmul Haque, Md Rejaul Karim Chowdhury, and Abdullah Gani, “A Study on Strategic Provisioning of Cloud Computing Services”, eScientific World Journal, Volume 2014, Article ID 894362, 16 pages

[2] Ian Foster, Carl Kesselman, Jeffrey M. Nick and Steven Tuecke, “Grid Computing: Making of Global Infrastructure a Reality”, Wiley 2003, pp. 217-249

[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia, “Above the Clouds: A Berkeley

View of Cloud Computing", Advanced Computing Machines, pp. 50-58
[4] http://www.nist.gov/itl/cloud/index.cfm
[5] Borko Furht, Armando Escalante , "Handbook of Cloud Computing", Springer, 2010, pp 3-8
[6] Simson L. Garfinkel, "An Evaluation of Amazon's Grid Computing Services: EC2, S3 and SQS", Center for Research on Computation and Society, Harvard University, Technical Report, 2007
[7] Malawski, M. Kuźniar, M. Wójcik, P. Bubak, M. , "How to Use Google App Engine for Free Computing", Internet Computing, IEEE , Volume:17 , Issue: 1 , pp. 50 - 59
[8] Jerry Gao, Xiaoying Bai, W.T. Tsai, Tadahiro Uehara, "SaaS Testing on Clouds-Issues, Challenges and Needs", IEEE 7th International Symposium on Service-Oriented System Engineering, pp.409-415, 2013
[9] Nicolas Bonvin, Thanasis G. Papaioannou and Karl Aberer, "Autonomic SLA-driven Provisioning for Cloud Applications", 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp.434-443, 2011
[10] Nicolas Bonvin, Thanasis G. Papaioannou and Karl Aberer, "An economic approach for scalable and highly-available distributed applications" IEEE 3rd International Conference on Cloud Computing, Pg. 498-505, 2010
[11] Mathias Bjorkqvist, Lydia Y. Chen, Walter Binder, "Opportunistic Service Provisioning in the Cloud", IEEE 5th International Conference on Cloud Computing, pages 237-244, 2012
[12] Mathias Bj¨orkqvist, Lydia Y. Chen, Walter Binder, "Dynamic Replication in Service-Oriented Systems", 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Pages 531-538, 2012
[13] Rodrigo N. Calheiros, Rajiv Ranjan, and Rajkumar Buyya, "Virtual Machine Provisioning Based on Analytical Performance and QoS in Cloud Computing Environments", International Conference on Parallel Processing, pp.295-304, 2011
[14] Andreas Lodde, Antoine Schlechter, Pascal Bauler, Fernand Feltz, "SLA-Driven Resource Provisioning in the Cloud", 1st International Symposium on Network Cloud Computing and Applications, pp. 28-35, 2011
[15] Mohammed Alhamad, Tharam Dillon, Elizabeth Chang, "SLA-Based Trust Model for Cloud Computing", 13th International Conference on Network-Based Information Systems, pp. 321-324, 2010
[16] Kuo-Chan Huang, Bo-Jyun Shen, Tsung-Ju Lee, Hsi-Ya Chang, Yuan-Hsin Tung, Pin-Zei Shih, "Resource Allocation and Dynamic Provisioning for Service-Oriented Applications in Cloud Environment", IEEE 4th International Conference on Cloud Computing Technology and Science, Pg. 839-84, 2012
[17] Waheed Iqbal, Matthew N. Dailey, David Carrera, "SLA-Driven Dynamic Resource Management for Multi-tier Web Applications in a Cloud", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Pg. 37-46, 2010
[18] Mahyar Movahed Nejad, Lena Mashayekhy, Daniel Grosu "Truthful Greedy Mechanisms for Dynamic Virtual Machine Provisioning and Allocation in Clouds", IEEE Transactions on Parallel and Distributed Systems, Pg. 188-195, 2013
[19] G. Wei, A. V. Vasilakos, Y. Zheng, and N. Xiong, "A game theoretic method of fair resource allocation for cloud computing services," The Journal of Supercomputing, pp. 1–18, 2009
[20] Danilo Ardagna, Barbara Panicucci, and Mauro Passacantando, "Generalized Nash Equilibria for the Service Provisioning Problem in Cloud Systems", IEEE Transactions on Services Computing, Vol. 6, No. 4, Oct-Dec 2013
[21] Valerio Di Valerio, Valeria Cardellini , Francesco Lo Presti, "Optimal Pricing and Service Provisioning Strategies in Cloud Systems: A Stackelberg Game Approach", IEEE 6th International Conference on Cloud Computing, Pg. 115-122, 2013

Author Profile

Mridul Paul is currently pursuing PhD in Computer Science and Engineering from Birla Institute of Technology, Mesra, a deemed University. He received B.E. in Computer Science & Engineering from same University and also possesses MBA degree in Finance. He has been associated with Information Technology field for last 15 years and his area of interest is in Cloud Computing. He is currently with Cognizant Technology Solutions as Associate Director-Projects.

Dr. Ajanta Das is working as Associate Professor in the department of Computer Science & Engineering in Birla Institute of Technology, Mesra, a deemed University. She is having altogether eighteen years of experience including six years of industry experience. She had worked for reputed, famous, pioneer companies like, Tata Steel, Jamshedpur, India and Lexis Nexis Inc., Boston, USA. She has proven herself confident and capable of handling real-life projects according to the deadline. Later, she joined academic and started teaching and research simultaneously. She has been awarded PhD in Engineering from Jadavpur University in 2009. Her major interest of teaching includes database, software engineering and distributed computing. Her focused research area includes Grid Computing, Cloud Computing and Wireless Sensor Network. She is life member of Computer Society of India.