INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 6 (2016) PAGES 1331 - 1335
RECEIVED : 29/10/2016. PUBLISHED: 7/11/2016
November 7, 2016

OPTIMAL WAY OF RESOURCE ALLOCATION FOR CUSTOMER CLOUD SERVICES

[1]S.Nithyabharathi, [2]S.Prasanna

[1]PG Scholar, Dept of CSE, Mailam College Of Engineering, Villupuram,

[2]Associate Prof, Dept of CSE, Mailam College Of Engineering, Villupuram.

Abstract:

Cloud computing, an amalgam of existing technologies ranging from distributed computing to cluster computing, to grid computing, to virtualization (which forms the foundation of these technologies) has changed the way organizations use Information and Communication Technology (ICT). Instead of acquiring resources for onpremise ICT departments, these resources are provisioned as service. It usually involves a pool of resources that multiple users can tap into and make use of (in parallel) whenever there is need to. These resources are also provisioned dynamically and are scaled up/down depending on demand. In addition, like any other utility, payment is done on a pay-per-use model thus reducing the huge initial cost of acquiring on-premise IT infrastructure. Since inception, there has been a steady increase in the number of users migrating to the clouds. Based on this increase, there is need to optimally allocate cloud resources so as to ensure that users perceived satisfaction is guaranteed. This work is an exposé on the challenges of resource allocation in cloud computing and works done in order to surmount these challenges. The work further goes on to juxtapose the various resource allocation strategies in order to identify their strengths and weakness based on how well they avoid situations such as resource under provisioning, over provisioning, contention, fragmentation and scarcity.

Keywords: Cloud Computing, Resource Allocation, Service Oriented Architecture, Quality of Service, Service Level Agreement.

1. INTRODUCTION

Cloud computing which evolved as an amalgamation of various existing technologies, virtualization being the key technology, enables users to have access to a pool of resources as a service. Other technologies that power this relatively new paradigm of computing include automated provisioning (servers have software installed automatically) and internet connectivity technologies to deliver the services [549]. This paradigm is gradually relinquishing control from in-house Information Technology infrastructures (necessary hardware and software installed within the premises) to service providers (that provision these hardware and software resources), enabling business organizations to focus more on core business activities and strategic development for their organizations. Furthermore, the cost of setting up a standard IT department is drastically reduced. Only the resources needed at any given time by Cloud computing users are received as services from the provider on a pay-as-you-go basis. This attribute makes Cloud computing elastic. According to a technical report published by the University of California, Berkley and [2] there is no precise definition for cloud computing. It is defined according to changes in services rendered by different organizations that provide the cloud solutions. The main reason for this indefinite description according [3] is as a result of statistical multiplexing of datacenters made possible with the efforts of researchers in areas such as Distributed Computing, Grid Computing, Web Technologies, Service Computing and Virtualization. Based on this, [28] concluded that the existence of different perceptions of Cloud Computing is that it is not a new technology, but rather a new model that brings together a set of existing technologies, identified in [3] to develop and run applications in a different way. However, [4] defined the Cloud as “a large

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 6 (2016) PAGES 1331 - 1335
RECEIVED : 29/10/2016. PUBLISHED: 7/11/2016
November 7, 2016

pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to optimum resource utilization. This pool of resources is typically exploited by a pay-per-user model in which guarantees are offered by the Infrastructure Provider by means of customized SLA's". Despite the numerous advantages that the Cloud offers, business organizations hesitated and expressed a level of doubtfulness about adopting the services. This was because the thought of their data residing on someone else' server heighted their fear concerning security of these data. Also, because resources in Cloud computing are obtained as a service, questions about Quality of Service arose. However recent studies [4] [5] [6] show that there has been a steady increase in Cloud adoption since 2009. According to a current research carried out by RightScale, Infrastructure as a Service (IaaS) grew 45 percent from 2012 to 2013. Currently it is estimated to be a $15 billion business and it expected to grow to about $31 billion by 2015 while in 2012, Software as a Service (SaaS) was a $13 billion market, but is now predicted to grow to over $30 billion by 2016. In 2008, revenue from worldwide cloud services was $46.4 billion; in 2013, it is expected to reach $150 billion, a jump of just over 225 percent.

2. TRENDS IN CLOUD COMPUTING

- **Private cloud:** this is the deployment of Cloud computing service for exclusive use by a single organization consisting of multiple users. The private Cloud may be owned and operated by the organization using it or a third party. It also could an amalgam of both.
- **Public cloud:** this refers to the deployment of cloud services on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources and bills on a fine-grained utility computing basis for the consumption of the general public [2] [10].

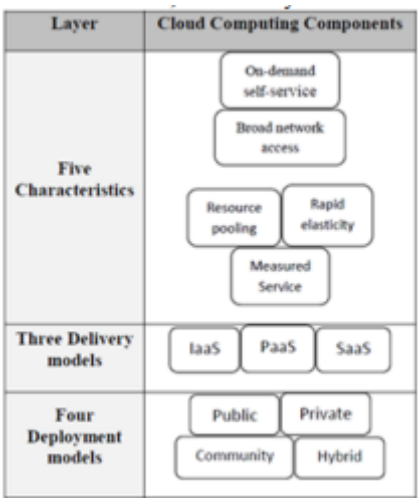

Fig.1. Cloud Environment Architecture

- **Hybrid cloud:** this is a concatenation of two or more distinct cloud service deployment models.
- **Community cloud:** this is deployed to serve several organizations that have a common theme and share a common concern. It may be owned, organized, and functioned by one or more of the organizations in the community, a third party, or some blend of them, and it may exist on or off premises.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 6 (2016) PAGES 1331 - 1335
RECEIVED : 29/10/2016. PUBLISHED: 7/11/2016
November 7, 2016

3. RELATED WORK

Very little literature is available on this survey paper in cloud computing paradigm. Shikharesh et al. in paper [30] describes the resource allocation challenges in clouds from the fundamental point of resource management. The paper has not addressed any specific resource allocation strategy. Patricia et al., investigates the uncertainties that increase difficulty in scheduling and matchmaking by considering some examples of recent research. It is evident that the paper which analyzes various resource allocation strategies is not available so far. The proposed literature focuses on resource allocation strategies and its impacts on cloud users and cloud providers. It is believed that this survey would greatly benefit the cloud users and researchers. Different kinds of resource allocation mechanisms are proposed in cloud. In the work by Jiani , actual task execution time and preemptable scheduling is considered for resource allocation. It overcomes the problem of resource contention and increases resource utilization by using different modes of renting computing capacities. But estimating the execution time for a job is a hard task for a user and errors are made very often [30]. But the VM model considered is heterogeneous and proposed for IaaS.

4. ANALYSIS

Experimental results shows that the most-fit policy has higher time complexities but the time overheads are negligible compared to the system long time operation. This policy is practical to use in a real system. A system which can automatically scale its infrastructure resources is designed . The system composed of a virtual network of virtual machines capable of live migration across multi-domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources. But the above work considers only the nonpreemptable scheduling policy.

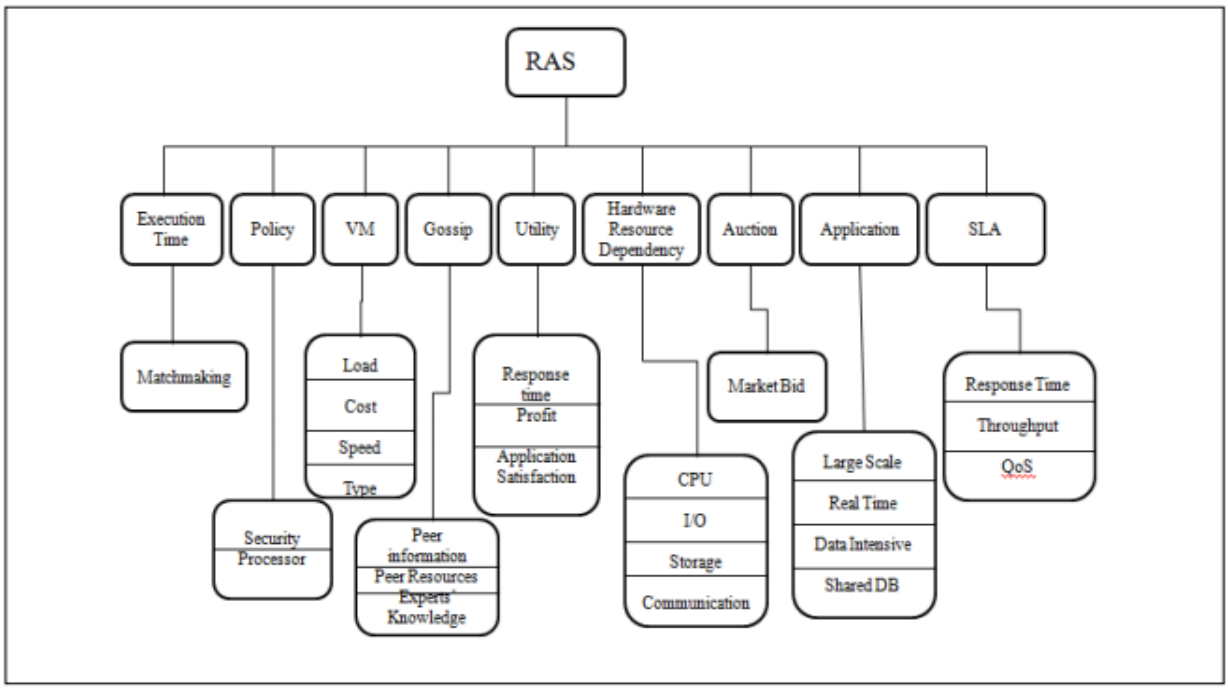

Fig.2.System Analysis

Cloud resources can be seen as any resource, be it physical or virtual, that users may request from the Cloud. These include network requirements, storage, computational needs such as CPU time, or even software applications . These resources are usually placed in multi-tenant data center that are able to match the resources with the volume of work being done at any point in time such that an expansion in business activities leads to more resources being provisioned and a contraction leads to less resources being provisioned. Cloud is defined as both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services [3]. According to this, delivery of application as services (SaaS - Software as a Service) over the Internet and hardware services (IaaS - Infrastructure as a Service) are both parts of cloud computing phenomena. From hardware service (utility computing) point of view, there are few new aspects in cloud, the most prominent being the illusion of infinite computing resources and the ability to pay for use of computing resources on a short-term basis as needed.

Real time application which collects and analyzes real time data from external service or applications has a deadline for completing the task. This kind of application has a light weight web interface and resource intensive back end . To enable dynamic allocation of cloud resources for back-end mashups, a prototype system is implemented and evaluated for both static and adaptive allocation with a test bed cloud to allocate resources to the application. The system also accommodates new requests despite a-priori undefined resource utilization requirements. In cloud, the works related to the SaaS providers considering SLA are still in their infancy. Therefore in order to achieve the SaaS providers' objective, various RAS specific to SaaS in cloud has been proposed.

CONCLUSION

Cloud computing technology is increasingly being used in enterprises and business markets. In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the classification of RAS and its impacts in cloud system. Some of the strategies discussed above mainly focus on CPU, memory resources but are lacking in some factors. Hence this survey paper will hopefully motivate future researchers to come up with smarter and secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

REFERENCES

[1] A.Singh ,M.Korupolu and D.Mohapatra. Server-storage virtualization:Integration and Load balancing in data centers. In Proc.2008 ACM/IEEE conference on supercomputing (SC'08) pages 1-12, IEEE Press 2008.

[2] AndrzejKochut et al. : Desktop Workload Study with Implications for Desktop Cloud Resource Optimization,978-1-4244-6534-7/10 2010 IEEE.

[3] Atsuo Inomata, TaikiMorikawa, Minoru Ikebe, Sk.Md. MizanurRahman: Proposal and Evaluation of Dynamin Resource Allocation Method Based on the Load Of VMs on IaaS(IEEE,2010),978-1-4244-8704-2/11.

[4] D. Gmach, J.RoliaandL.cherkasova, Satisfying service level objectives in a self-managing resource pool. In Proc. Third IEEE international conference on self-adaptive and self organizing system.(SASO'09) pages 243-253.IEEE Press 2009 .

[5] David Irwin, PrashantShenoy, Emmanuel Cecchet and Michael Zink:Resource Management in Data-Intensive Clouds: Opportunities and Challenges .This work is supported in part by NSF under grant number CNS-0834243.

[6] Dongwan Shin and HakanAkkan :Domain- based virtualized resource management in cloud computing.

[7] Dorian Minarolli and Bernd Freisleben: Uitlity –based Resource Allocations for virtual machines in cloud computing(IEEE,2011),pp.410-417.

[8] DusitNiyato, Zhu Kun and Ping Wang : Cooperative Virtual Machine Management for Multi-Organization Cloud Computing Environment.

[9] FetahiWuhib and Rolf Stadler : Distributed monitoring and resource management for Large cloud environments(IEEE,2011),pp.970-975.

[10] HadiGoudaezi and MassoudPedram : Multidimensional SLA-based Resource Allocation for Multi-tier Cloud Computing Systems IEEE 4th International conference on Cloud computing 2011,pp.324-331.