



Survey on Existing Resource Provisioning Techniques Available in Cloud for Effective Utilization of Hybrid Cloud Infrastructure

Rohit Vijayan

PG Scholar, M-Tech Network and Internet Engineering,
School of Information Technology, Karunya University,
Coimbatore, Tamilnadu, India

Bijolin Edwin.E

Assistant Professor/ Department of IT,
Karunya University, Tamilnadu,
India

Abstract-- Collection or pool of data centers is known as “cloud”. Cloud computing refers to a paradigm where services are offered via internet in a pay-as-you go manner, and service are provided with data centers and. The use and popularity of hybrid cloud is on the rise in all fields of computing where computation is required, as utilization of both private and public clouds can be done effectively for the requirement of the user. As requirement of are on rise, we need effective technique to optimally allocate resources to the require jobs seeking resources. The selection of a particular technique for resource provisioning depend on the needs of the specific job and parameters such as scalability, turnaround time, fault tolerance, load balancing and quality of service. This paper is concentrate on giving an overview about types of clouds and about few resource provisioning technique available for better utilization of resource and to provide high quality of service.

Keywords— Cloud, Resource provisioning, scheduling, Grid

1. Introduction

In this fast moving technology depended world use of resource in a more efficiently and according to the need is a very important. Resource availability and utilization constitute a major role in any field may it be scientific, campus experiments etc. A resource is purchased considering the requirement of the present and estimating a maximum of what it can be in near future, but the demand is not static nor progressive in a defined manner, it is always in a wave form sometimes require more resources than we would have anticipated to be the maximum peak and sometimes much less than the average need hence determining the actual peak need of the resource is out of the question for example a website traffic in holiday will always be higher than the rest of the time so the resources can be made available as needed using utility computing which is termed as the upcoming biggest revolution in computing world. So as to make the resource available as needed according to the need we need to have an efficient resource provisioning schedulers, this paper is considering the various schedulers there advantages disadvantages methodology used and prepare a comparative study so as to use the correct scheduler in the correct manner where it suits. A proper definition of cloud computing as mentioned by George Reese [1] is as follows “Cloud is where you go to use technology when you need it, for as long as you need it, and not a minute more”. Utility computing is also known as cloud computing is an way of providing IT infrastructure and application as a service to end-users under a usage based payment model better known as pay-as-you-go or pay-for what-you-use. Further cloud computing can be classified as private clouds, public clouds and hybrid clouds.

2. An Introduction into Types of Cloud

A look into the types of clouds and its utilization models available to be leveraged for better and efficient utilization IT resources. We have private cloud, public cloud hybrid cloud and community cloud as the major types of clouds. The major type of platform available for the users to leverage the cloud in an useful, effective and as required manner, are in the form of platform-as-a-service(PaaS),software-as-a-service(SaaS),infrastructure-as-a-service(IaaS) these three constitute the major divisions. We also have IT-as-a-service, network-as-a service, or in simple terms can summarize it as anything-as-a-service.

2.1. Types of Clouds

a. Private clouds

The private cloud is a cloud for a group of individuals under an organization, this is mostly utilized in the cooperate world where a cloud infrastructure is maintained for the internal working of the institution and are accessed by the employees with access rights, the private cloud is considered to be more secure when compared with public cloud as private cloud is accessed internally and no access form outside is provided without access privileges.

b. Public clouds

Public cloud is available for all those who are in need of resources as service under a usage based payment model better known as pay-as-you-go or pay-for what-you-use. The main advantage of public cloud is it can be utilized by anyone as they require in a scalable manner.

c. Hybrid clouds

The hybrid cloud is getting popular as computational needs are rising, the main reason for the popularity of the hybrid cloud infrastructure is the utilization of private and public cloud infrastructure as required. Combining the goodness of two entities into a single entity is more beneficial and cost effective. As some might already have infrastructure up and running but as the computational requirement increases it isn't a good option to expand so we make use of the public cloud and when requirement can be handled using the existing infrastructure we utilize the private cloud which in long term proves to be cost effective and efficient.

Types	Scalability	Security	Availability
Public	High	Low	High
Private	Low	High	Low
Hybrid	Very High	Medium	High

3. An Overview of Resource Provisioning Techniques

Cloud computing has emerged over the years as the provider of IT resources and application as service to end users in a usage based payment model. With the emergence of cloud computing Resource provisioning has become a major factor in this energy efficient and green computing world, more over the need to use the resource in a much effective and efficient manner is an important aspect in every field utilizing computer systems. This paper concerns with the survey of different techniques and methods being deployed in different systems for resource provisioning to provide an effective and efficient utilization of computing resources as and when required in a cost effective manner.

3.1 Inter-Operating Grids through Delegated Matchmaking[3]

The main idea with this technique is to provide a solution by combining the resource available in a hierarchical and decentralized approach for interconnecting grids. The hierarchy of grid sites is augmented with peer-to-peer connections between sites under the same administrative control. To make this work in an efficient manner we utilize the delegation matchmaking, which temporarily binds resource from remote sites to the local environment.

3.2 VioCluster: Virtualization for Dynamic Computational Domains [2]

The idea of emergence of VioCluster is from the conflicts happening in computational domain between dynamic workload and static capacity, to this problem the solution provided by this technique is by dynamically adapting the capacity of clusters by borrowing idle machines of peer domains which is done using design implementation of VioCluster, a virtualization based computational resource sharing platform. Through machine and network virtualization.

3.3 Elastic Site Using Clouds to Elastically Extend Site Resources[4]

As cloud computing has emerged and is taking over the IT environment and providing with new possibilities in the similar fashion it also provides with new possibilities to the scientific communities as well. This paper discusses about the ability to elastically provision and utilize new resources in response to change in demand. To avoid the over or under utilization of resources the paper proposes three different policies on demand, steady stream and bursts to efficiently schedule resource deployment based on demand.

3.4 Evaluation of gang scheduling performance and cost in a cloud computing system[5]

The gang scheduling is an efficient job scheduling algorithm for time sharing which is applied in parallel and distributed systems. In this the virtual machine acts as the computational need of the job, but as per the computational needs and requirement new VM's can be leased and later released dynamically, the two gang scheduling being considered are Adaptive first come first server(AFCFS) and Largest job first served (LJFS).

3.5 Backfilling using system-generated predictions rather than user runtime estimates[6]

Backfilling allows to run short jobs in queue to run ahead of their time provided they do not delay previous jobs in queue. To get such estimation the user need to provide estimate of how long jobs will run and those which violate are killed. But user estimates of a job runtime are inaccurate, and system-generated based on history are significantly better, but underprediction is technically unacceptable i.e. user will not accept a job termination because system predictions were too short. This paper solves this problem by considering both user prediction and system prediction and if the job gets completed within the system predicted time frame the job is terminated and resources is freed for next job in the queue without considering the users prediction but if job do not get completed within the system predicted time frame user prediction is utilised and on completion its terminated.

4. Comparison Table

SL.NO	PAPER NAME	SCALABILITY	UTILIZATION	TURNAROUND TIME	FAULT TOLERANT	LOAD BALANCING	QoS
1.	Inter-Operating Grids through Delegated Matchmaking	Y	Y	N	N	Y	N
2.	VioCluster: Virtualization for Dynamic Computational Domains	Y	Y	Y	N	Y	Y
3.	Elastic Site Using Clouds to Elastically Extend Site Resources	Y	Y	Y	N	Y	Y
4.	Evaluation of gang scheduling performance and cost in a cloud computing system	Y	Y	Y	N	Y	Y
5.	Backfilling using system-generated predictions rather than user runtime estimates	N/A	N	Y	N	Y	Y

5. Conclusion

This paper is an introduction and comparative study in the field of cloud computing concentrating on resource provision techniques available, to make it helpful for decide upon which is better suitable for the situation so as to provide better services to the end user. This paper give a brief idea about few available resource provision techniques, to make it easy to choose according to need of the user to provide with better suited technique and provide efficient Quality of service.

Reference

- [1]. George Reese, "Cloud Application Architectures," Pub. O'Reilly Media, it-ebooks.info/book/286/, pp.1-10, 2009.
- [2]. P. Ruth, P. McGachey, D. Xu, VioCluster: virtualization for dynamic computational domain, in: Proceedings of the 7th IEEE International Conference on Cluster Computing, Cluster 2005, IEEE Press, Piscataway, NJ, Burlington, MA, 2005, pp. 1–10.
- [3]. A. Iosup, D.H.J. Epema, T. Tannenbaum, M. Farrellee, M. Livny, Inter-operating Grids through delegated matchmaking, in: Proceedings of the 18th ACM/IEEE Conference on Supercomputing, SC 2007, ACM, New York, NY, Reno, Nevada, 2007, pp. 1–12.
- [4]. P. Marshall, K. Keahey, T. Freeman, Elastic site: using clouds to elastically extend site resources, in: Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, CCGrid 2010, IEEE Computer Society, Washington, DC, Melbourne, Australia, 2010, pp. 43–52.
- [5]. I. Moschakis, H. Karatza, Evaluation of gang scheduling performance and cost in a cloud computing system, The Journal of Supercomputing 1 (2010) 1–18.
- [6]. D. Tsafirir, Y. Etsion, D.G. Feitelson, Backfilling using system-generated predictions rather than user runtime estimates, IEEE Transactions on Parallel and Distributed Systems 18 (2007) 789–803.
- [7]. M. Islam, P. Balaji, P. Sadayappan, D.K. Panda, QoPS: a QoS based scheme for parallel job scheduling, in: Proceedings of the 9th International Workshop on Job Scheduling Strategies for Parallel Processing, JSSPP'03, Springer-Verlag, Berlin, Seattle, WA, 2003, pp. 252–268.
- [8]. D. Kondo, B. Javadi, P. Malecot, F. Cappello, D.P. Anderson, Cost-benefit analysis of Cloud computing versus desktop grids, in: Proceedings of the 23rd IEEE International Parallel and Distributed Processing Symposium, IPDPS 2009, IEEE Computer Society, Washington, DC, Rome, Italy, 2009, pp. 1–12.
- [9]. D. Oppenheimer, A. Ganapathi, D.A. Patterson, Why do Internet services fail, and what can be done about it? in: Proceedings of the 4th Conference on USENIX Symposium on Internet Technologies and Systems, USENIX Association, Berkeley, CA, Seattle, WA, 2003, pp. 1–15.
- [10]. L.F. Orleans, P. Furtado, Fair load-balancing on parallel systems for QoS, in: Proceedings of the 36th International Conference on Parallel Processing, ICPP 2007, IEEE Computer Society, Los Alamitos, CA, XiAn, China, 2007, pp. 22–30.
- [11]. S. Ostermann, A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer, D. Epema, A performance analysis of EC2 Cloud computing services for scientific computing, in: Proceedings of the 1st International Conference on Cloud Computing, CloudComp 2009, Springer-Verlag, Berlin, Beijing, China, 2009, pp. 115–131.
- [12]. D.G. Feitelson, Workload Modeling for Computer Systems Performance Evaluation, e-Book. <http://www.cs.huji.ac.il/~feit/wlmod/>, 2009.