ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 249 – 252 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

Deploying Virtual Cluster Nodes Distributed Over Different Cloud Providers for On-Demand Resource Provisioning

Saira Arjumand[1], Saleha Farha[2], M. S. Qaseem[3]
[1]M.tech(SE), Jntu, Nizam Institute of engineering and technology, Nalgonda, A.P, India, saira_arjumand@yahoo.com
[2]M.tech(SE), Jntu, Nizam Institute of engineering and technology, Nalgonda, A.P, India, saleha_farha@yahoo.com
[3]Assoc.Professor, Jntu, Nizam Institute of engineering and technology, Nalgonda, A.P, India, ms_qaseem@yahoo.com

Abstract : Cloud computing technologies can offer important benefits for IT organizations and data centers running MTC applications such as elasticity and rapid provisioning, enabling the organization to increase or decrease its infrastructure capacity within minutes, according to the computing necessities; pay as you go model, allowing organizations to purchase and pay for the exact amount of infrastructure they require at any specific time; reduced capital costs, since organizations can reduce or even eliminate their in-house infrastructures, resulting on a reduction in capital investment and personnel costs; access to potentially "unlimited" resources, as most cloud providers allow to deploy hundreds or even thousands of server instances simultaneously; and flexibility, because the user can deploy cloud instances with different hardware configurations, operating systems, and software packages. Computing clusters have been one of the most popular platforms for solving MTC problems, especially in the case of Cluster overloading and insufficient Computational resources during peak demand periods. Regarding these limitations, cloud computing technology has been proposed as a viable solution to deploy elastic computing clusters, or to complement the in-house data center infrastructure to satisfy peak workloads.

Key words : Cloud computing, Cluster, data centers, MTC applications.

INTRODUCTION

Computing clusters have been one of the most popular platforms for solving MTC problems, especially in the case of loosely coupled tasks. Unlike traditional utilities where a single provider scheme is a common practice, the ubiquitous access to cloud resources easily enables the simultaneous use of different clouds. This scenario is used to deploy a computing cluster on the top of a multi cloud infrastructure, for solving loosely coupled Many-Task Computing (MTC) applications[1]. In this way, the cluster nodes can be provisioned with resources from different clouds to improve the cost effectiveness of the deployment, or to implement high-availability strategies. The viability of this kind of solutions is proved by evaluating the scalability, performance, and cost of different configurations of clusters.

The main goal of this work is to analyze the viability, from the point of view of scalability, performance, and cost of deploying large virtual cluster infrastructures distributed over different cloud providers for solving loosely coupled MTC applications[5].

The main advantages of the proposed work are as follows:

High availability and fault tolerance: the cluster worker nodes can be spread on different cloud sites, so in the case of cloud downtime or failure the cluster operation will not be disrupted. Furthermore, in this situation, we can dynamically deploy new cluster nodes in a different cloud to avoid the degradation of the cluster performance.

Infrastructure cost reduction: Since different cloud providers can follow different pricing strategies, and even variable pricing models the different cluster nodes can change dynamically their locations, from one cloud provider to another one, in order to reduce the overall infrastructure cost.

DEPLOYMENT OF VIRTUAL CLUSTER

The following figure shows distributed used at the top of a multi cloud infrastructure. This kind of multicloud deployment involves many challenges. The challenges and viability of deploying a computing cluster on top of a multicloud infrastructure is analyzed spanning four different sites for solving loosely coupled MTC applications. Here a real testbed cluster (based on a SGE queuing system) is implemented that comprises computing resources from our in-house infrastructure, and external resources from three different clouds: Amazon EC2 (Europe and US zones) and Elastic Hosts.

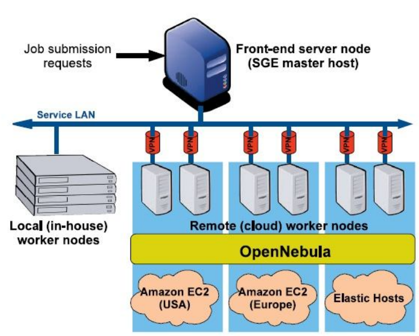

Fig 1: Experimental framework

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 249 – 252 (2013)

Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

This work is to analyze the viability, from the point of view of scalability, performance, and cost of deploying large virtual cluster infrastructures distributed over different cloud providers for solving loosely coupled MTC applications. This work is conducted in a real experimental testbed that comprises resources from in-house infrastructure, and external resources from three different cloud sites: Amazon EC2 (Europe and US zones1) and Elastic Hosts. On top of this distributed cloud infrastructure, we have implemented a Sun Grid Engine (SGE) cluster, consisting of a front end and a variable number of worker nodes, which can be deployed on different sites (either locally or in different remote clouds).It analyze the performance of different cluster configurations, using the cluster throughput (i.e., completed jobs per second) as performance metric, proving that multi cloud cluster implementations do not incur in performance slowdowns, and we also analyze the performance/cost ratio.

Due to hardware limitations of the local infrastructure, and the high cost of renting many cloud resources for long periods, the tested cluster configurations are limited to a reduced number of computing resources (up to 16 worker nodes), running a reduced number of tasks (up to 128 tasks).However, as typical MTC applications can involve much more tasks, we have implemented a simulated infrastructure model, that includes a larger number of computing resources (up to 256 worker nodes), and runs a larger number of tasks (up to 5,000). The simulation of different cluster configurations shows that the performance and cost results can be extrapolated to large-scale problems and cluster infrastructures. More specifically, the contributions of this work are the following:

1. Deployment of a multi cloud virtual infrastructure spanning four different sites: our local data center, Amazon EC2 Europe, Amazon EC2 US, and Elastic Hosts; and implementation of a real computing cluster testbed on top of this multi cloud infrastructure.
2. Performance analysis of the cluster testbed for solving loosely coupled MTC applications proving the scalability of the multi cloud solution for this kind of workloads.

Several studies have explored the use of virtual machines to provide custom cluster environments. Some recent works have explored the use of cloud resources to deploy hybrid computing clusters, so the cluster combines physical, virtualized, and cloud resources. There are many other different experiences on deploying different kind of multitier services on cloud infrastructures, such as webservers database appliances or web service platforms among others. However, all these deployments only consider a single cloud, and they do not take advantage of the potential benefits of multi cloud deployments. Regarding the use of multiple clouds, Keahey et al. introduce in the concept of "Sky Computing," [2] which enables the dynamic provisioning of distributed domains over several clouds, and discusses the current shortcomings of this approach, such as image compatibility among providers, need of standards at API level, need of trusted networking environments, etc. This work also compares the performance of two virtual cluster deployed in two settings :a single-site deployment and a three-site deployment, and concludes that the performance of a single-site cluster can be sustained using a cluster across three sites. However, this work lacks a cost analysis and the performance analysis is limited to small size infrastructures (up to 15 computer instances, equivalent to 30 processors).

COST ANALYSIS

The cost of cloud resources also has important impact on the viability of the multi-cloud solution. From this point of view, it is important to analyze, the total cost of the infrastructure, in order to find the most optimal configurations. The cost of leased resources from cloud providers are mainly derived from three sources: computing resources, storage, and network data transfer[4].

In this work, we have analyze the deployment of a computing cluster in a multi-cloud environment, using resources from three different cloud sites. These providers can offer different pricing schemes for computing resources, e.g. on demand instances, reserved instances, or spot instances and monthly subscription instances, or hourly burst. However, in this work we have used a single pricing method, based on a pay per compute capacity used scheme, which is available in most of cloud providers

This pay-per-use pricing scheme is the most flexible one, since a user can start or stop compute instances dynamically as and when needed, with no long-term commitments, that are charged by the cloud provider at a given price per hour of use. Table 1 displays the hourly prices (using the pay-per-use pricing method) for the type of resources used in this work charged by the three cloud providers considered in this work.

Table 1- Characteristics of different cluster nodes

Site	Arch.	Processor (single core	Mem. (GB)	Cost (USD/hour)	Cost per second* (USD)
Cloud 1	i686 32-bits	Xeon 1.2GHz	1.7	0.10	2.8×10^{-5}
Cloud 2	i686 32-bits	Xeon 1.2GHz	1.7	0.04	1.1×10^{-5}
Cloud 3	i686 32-bits	Xeon 1.2GHz	1.7	0.11	3.1×10^{-5}

* *cost per second=cost per hour/3,600*

In this section ,the analyze and compare the cost offered by different configurations of the computing cluster. In particular, we have chosen 6 different cluster configurations (with different number of worker nodes from the three cloud providers), and different number of jobs(depending on the cluster size), as shown in Table 2.In the definition of the different cluster configurations, we have use the following acronyms: *c1*: cloud 1; c2:cloud 2; *c3:*cloud. The number

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 249 – 252 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

preceding the site acronym represents the number of worker nodes. For example, *2c1* is a cluster with two worker nodes deployed in the cloud 1; and *2c1+3c3* is a five nodes cluster, two deployed in the cloud 1 and three in the cloud 3.

Although cloud resources are charged by the provider on a per-hour basis, we assume that, in a general case, our multi-cloud cluster is running and working for long periods of time (maybe days, weeks, or even months) and is continuously queuing and executing works from different users. Hence, the cost of computing resources imputed to a given work or experiment is calculated using a per-second basis, i.e., multiplying the time spent by each worker node (in seconds) in running this experiment by the price per second of this node, and adding the resulting cost of all the worker nodes. The price per second of a computing resource is simply calculated by dividing the hourly price of the resource by 60×60, as displayed in the last column of Table 1. Using this per-second cost model, the resulting cost of computing resources for the different experiments achieved in this work is summarized in Table 2.

Table 2 -Total cost of computing resources for various experiments with different cluster configurations

Cluster Configuration	Jobs	Total cost	Cost per job
Exp .1 (2c1)	20	0.56	0.028
Exp .2 (3c2)	12	0.13	0.011
Exp .3 (3c3)	15	0.46	0.030
Exp .4 (2c1+3c3)	35	1.02	0.029
Exp .5 (3c2+3c3)	27	0.59	0.021
Exp .6 (2c1+3c2+3c3)	47	1.15	0.024

* *cost per job* = total cost of each experiment by the number of jobs in the experiment.

The different cluster configurations and their cost per jobs are taken from the above table and a bar chart and a line graph are drawn.

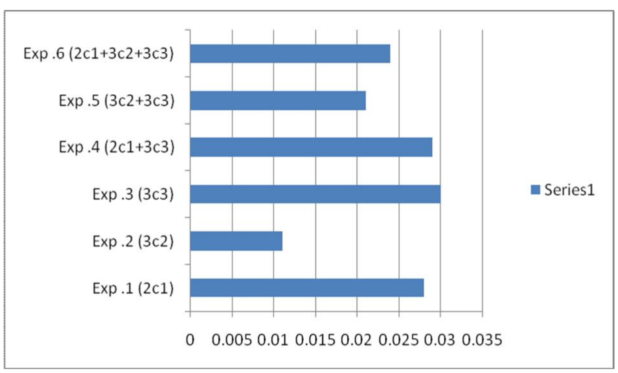

Fig 2: Cost per job for different configurations (bar chart)

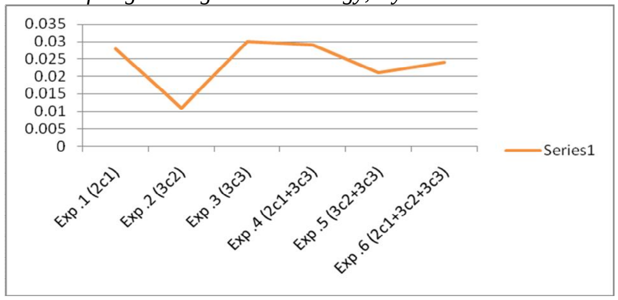

Fig 3: Cost per job for different configurations (line graph)

The above figure depicts the cost Analysis of the experimental setup which is used to compare the different cluster configurations, and proving the viability of the multi-cloud solution from a cost perspective and showing that multi cloud deployment is efficient when compared to single cloud deployment.

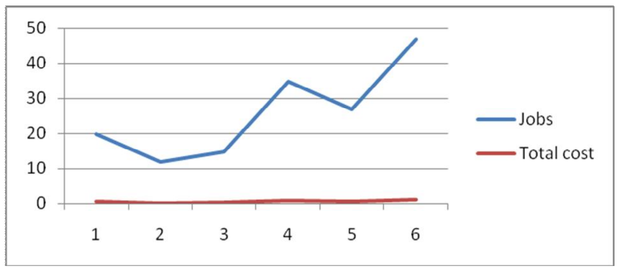

Fig 4: Cost Analysis for the jobs provided by different cluster configuration (line graph)

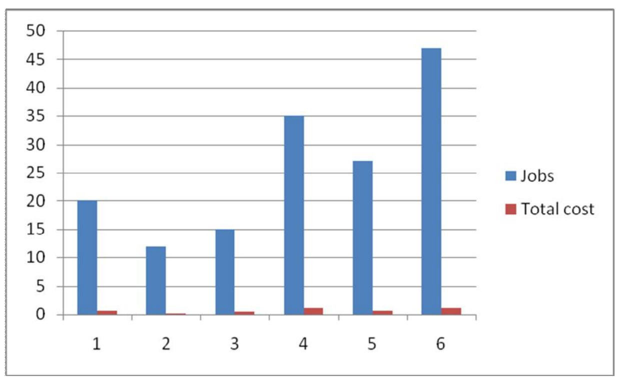

Fig 5: Cost Analysis for the jobs provided by different cluster configurations (bar chart)

The above graph and bar chart shows that as the no. of jobs increases the cost is not linearly increased by using different cluster configuration and the no. of jobs done using multi cloud deployment is more when compared to single cloud deployment so there is high availability of jobs(resources) with a minimum cost.

RELATED WORK

Efficient management of large-scale cluster infrastructures has been explored for years, and different techniques for on-demand provisioning, dynamic partitioning, or cluster virtualization have been proposed. Traditional methods for the on-demand provision of computational services consist in overlaying a custom software stack on top of an existing

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 249 – 252 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

middleware layer. The Falkon system [6] provides a light high throughput execution environment on top of the Globus GRAM service. The dynamic partitioning of the capacity of a computational cluster has also been addressed by several projects. For example, the Cluster On Demand software enables rapid, automated, on-the-fly partitioning of a physical cluster into multiple independent virtual clusters. Similarly, the VIO [7]cluster project enables to dynamically adjust the capacity of a computing cluster by sharing resources between peer domains.Several studies have explored the use of virtual machines to provide custom cluster environments. In this case, the clusters are usually completely build up of virtualized resources, as in the Globus Nimbus project , or the Virtual Organization Clusters (VOC) proposed in. Some recent works [3] have explored the use of cloud resources to deploy hybrid computing clusters, so the cluster combines physical, virtualized, and cloud resources. There are many other different experiences on deploying different kind of multitier services on cloud infrastructures, such as web servers [8], database appliances, or web service platforms, among others. However, all these deployments only consider a single cloud, and they do not take advantage of the potential benefits of multicloud deployments. Regarding the use of multiple clouds, Keahey et al. introduce in [2] the concept of "Sky Computing," which

enables the dynamic provisioning of distributed domains over several clouds, and discusses the current shortcomings of this approach, such as image compatibility among providers, need of standards at API level, need of trusted networking environments, etc. This work also compares the performance of two virtual cluster deployed in two settings: a single-site deployment and a three-site deployment, and concludes that the performance of a single-site cluster can be sustained using a cluster across three sites. However, this work lacks a cost analysis and the performance analysis is limited to small size infrastructures

CONCLUSION

The simulation of different cluster configurations shows that performance and cost results can be extrapolated to large-scale problems and clusters.

The different cluster configurations considered in this work have been selected manually, without considering any scheduling policy or optimization criteria, with the main goal of analyzing the viability of the multi cloud solution from the points of view of performance and cost. Although a detailed analysis and comparison of different scheduling strategies is out of the scope of this work, and it is planned for further research.

REFERENCES

[1] Rafael Moreno-Vozmediano , Ruben S. Montero , and Ignacio M. Llorente, Member , "Multicloud Deployment of Computing Clusters for Loosely Coupled MTC Applications ", Proc IEEE Tranctions on Parallel and Distributed Systems , vol 22,No.6, pp.924-930,June 2011.

[2] K. Keahey , M. Tsugawa , A. Matsunaga , and J. Fortes, "Sky Computing," IEEE Internet Computing vol. 13, no. 5, pp. 43-51, Sept./Oct. 2009.

[3] Llorente, R. Moreno-Vozmediano , and R. Montero, "Cloud Computing for On-Demand Grid Resource Provisioning," Advances in Parallel Computing, vol. 18, pp. 177-191, IOS Press, 2009.

[4] John W.Rittinghouse,James F.Ransome,CRC Press,Talor & Francis group"Cloud Computing implementation , management and security" ,2010.

[5] Raicu,I.Foster,Y.Zhao, Many-Task Computing for Grids and Supercomputers ,Workshop on Many- Task Computing of Grids and Supercomputers , pp.1-11,2008.

[6] Raicu, Y. Zhao , C. Dumitrescu, I. Foster, and M. Wilde, "Falkon : A Fast and Light-Weight TasK Execution Farmework," Proc.IEEE/ACM Conf. SuperComputing , 2007.

[7] P. Ruth , P. McGachey, and D. Xu, "VioCluster: Virtualization for Dynamic Computational Domains ," Proc. IEEE Int'l Conf. Cluster Computing, 2005.

[8] J. Fronckowiak, "Auto-Scaling Web Sites Using Amazon EC2 and Scalr,"Amazon EC2 Articles and Tutorials, 2008.