Journal of Computer Applications ISSN: 0974 – 1925,
Volume-5, Issue EICA2012-5, February 10, 2012

Deployment of Multicloud for Solving Loosely Coupled Applications Using Cost Optimizing Technique

Mr.Venkatesa Kumar V
Asst.Professor,
Dept. of Computer Science and Engineering,
Anna University of Technology Coimbatore,
Coimbatore, Tamil Nadu
E-mail: mail2venkatesa@gmail.com

Ms.Sindhu M,
PG Scholar,
Dept. of Computer Science and Engineering,
Anna University of Technology Coimbatore,
Coimbatore, Tamil Nadu
E-mail:sindhu.shanu@gmail.com

Abstract - Cloud Computing has recently received considerable attention, as a promising approach for delivering Many-Task Computing (MTC) application services by improving the utilization of data centre resources. In this paper, Computing Cluster on top of a multi-cloud infrastructure is developed, to solve the loosely-coupled MTC applications has proposed. In the existing system, it has been analyzed that the challenges and viability of deploying the computing clusters for loosely coupled MTC applications based on a cost effective control. However, no specific scheduling algorithm was considered for solving the loosely coupled MTC applications. In our model, the cluster nodes are required which is composed of resources from different Clouds to improve the cost-effectiveness of the deployment, and to apply the high-available strategies. For the initial scheduling scheme, PCH (Path Clustering Heuristic) algorithm is used. The proposed approach will improve the cost optimization compared with the other optimization techniques.

Keyword — *Multicloud, Computing clusters, MTC Applications, Cost – optimization.*

I. INTRODUCTION

Cloud computing has emerged as an important paradigm for accessing distributed computing resources. Commercial providers such as Amazon, Rackspace, and Microsoft, all offer environments for developing and deploying applications in the cloud. There are many definitions of cloud computing, but some characteristics exist in most definitions, e.g., virtualized environments and on-demand provisioning of compute resources. Cloud computing is driving a significant part of various above mentioned products' development plans from enterprise applications to middleware, business intelligence technology, databases servers and storage devices as well as cloud management systems.

These developments build long leadership in underlying technologies like – grid computing, clustering, server virtualization, dynamic provisioning, SOA, identity management and large scale management automation. These new concepts have created much of interest in the recent past. This was a result of its huge potential for substantiating other technological advances while presenting a superior utilitarian advantage over the currently under-utilized resources deployed at data centers.

In this sense, cloud computing can be considered a new computing paradigm that allows users to temporarily utilize computing infrastructure over the network, supplied as a service by the cloud-provider at possibly one or more levels of abstraction. The core cloud computing services, and their inter-relations are ambiguous and the feasibility of enabling their inter-operability has been debatable. Furthermore, each cloud computing service has a distinct interface feature that employs a different access protocol. A unified interface provides integrated access to cloud computing services that does not exist, although its portals and gateways can provide this unified web-based user interface. Many-Task Computing (MTC) paradigm embraces different types of high-performance applications involving many different tasks, and requiring large number of computational resources over short periods of time. These tasks can be of very different nature, with sizes from small to large, loosely coupled or tightly coupled, or compute-intensive or data-intensive. Many-Task Computing leads to a stronger appreciation of the fact that applications that are not tightly coupled via MPI are not necessarily embarrassingly parallel: some have just so many simple tasks that managing them is hard, some operate on or produce large amounts of data that need sophisticated data management in order to scale. There also exist applications that involve MPI ensembles, essentially many jobs where each job is composed of tightly coupled MPI tasks, and there are loosely coupled applications that have dependencies among tasks, but typically use files for inter-process communication Cloud computing technologies can offer important benefits for various IT organizations and data centers that run MTC applications such as elasticity and rapid provisioning enabling the organization to increase their infrastructure capacity within minutes according to the computing necessities. The baseline of the Path Clustering Heuristic Algorithm (PCH) here is to work on scheduling through MTC applications which can be programmed perfectly through the PCH algorithm that yields effective results especially in optimizing the cost for the organizations.

A. Significant Individualities of the Clouds

- **Insisting the self – services**: It can unilaterally provide the computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each of the service provider.

- **Wider access to network**: Through standard mechanisms heterogeneous thin or thick client platforms can be accessed throughout.

- **Reserve pooling**: Computing resources are pooled to use multi – tenant model, with different virtual resources which are dynamically assigned according to the demand. Examples of resources are storage, processing, memory, network bandwidth and virtual machines. Sense of location independence may be able the location at a higher level of abstraction such a datacenter.

- **Speedy elasticity**: These cloud systems can be rapidly and elastically provisioned automatically to quickly scale out and rapidly released to quickly scale in.

- **Service measurement**: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriately.

B. Deployment of Multi – Cloud Transmitting in Virtual Cluster

Cloud is a large pool of easily usable and accessible virtualized resources, which can be dynamically reconfigured. Many-task computing (MTC) can deliver much large numbers of computing resources over short period of time to accomplish many computational tasks, and high throughput computing (HTC) can deliver large amounts of processing capacity over long period of time. Various researchers have proposed that each staffs of the organization directly leases virtual machine resources in a specified period for running various applications. Resource management are widely researched in the context of cloud computing and grid to be specific.

Cloud computing is always closely connected to virtualization because of their major hypervisor vendors such as – VMware, Microsoft and Citrix Systems who inputs many of the emphasis on cloud. Moreover cloud computing does not axis on particular virtualization or any one technology. Virtualization is a single computer, controlling multiple machines generally to analyse a database. It also supports more users, delivers and runs the application rapidly and accurately.

Virtualization is basically one physical computer pretending to be many computing environments whereas cloud computing is many different computers pretending to be the one computing environment. Since it provides greater flexibility it is said to be an immense match for cloud computing. Multiple operating system instances running on one hardware device are far more economical than a single piece of hardware for every server task. Virtualization provides more servers on the same hardware and cloud computing provides measured resources while paying for what we use.

C. Types of Clouds and their Significance

i. **Public Cloud** – It is based on the standard cloud computing model, in which a service provider makes the following resources, such as applications and storage, available to the general public over the Internet. Public cloud services may be free or offered on a pay-per-usage model.

ii. **Private Cloud** – It is an infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally.

iii. **Hybrid Cloud** – Hybrid cloud is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. It can also be defined as multiple cloud systems that are connected in a way that allows programs and data to be moved easily from one deployment system to another.

II. RELATED WORK

MTC applications with the help of three different cloud networks like private, public and hybrid. And also analyze the performance of different cluster configurations, using the cluster throughout as performance metric. Deployed Multicloud virtual infrastructure is undergone in this stage. This kind of Multicloud deployment involves several challenges, related to the lack of a cloud interface standard; the distribution and management of the service master images; and the interconnection links between the service components. The clusters are deployed in a hybrid setup, which combines local physical nodes with virtual nodes deployed in another compute cloud. Compare the different cluster configurations, and proving the viability of the Multicloud solution also from a cost perspective. This approach is described by Rafael Moreno-Vozmediano et al. in July 2010.

Sky computing provides combined use of Multiple Clouds. In Infrastructure-as-a-service (IaaS) Cloud Computing a remote user "leases" a resource, the service provider turns control of that resource over to the user. Before virtualization, turning over control to users was fraught with danger because users could easily subvert a site. Virtualization isolates the leased resource from the site in a secure way was illustrated by K. Keahey et al. in Oct 2009.

An approach which automatically scales the Web service applications while maintaining the availability & scalability guarantees at an optimum cost. The Web service application developer should need to write the application once, and simply deploy it on the cloud. Web service applications on Amazon EC2 are a very appealing idea from a business point of view was proposed by A. Azeez in 2009. VioCluster is Virtualization for Dynamic Computational Domains. VioCluster, a novel architecture allows a cluster to dynamically grow and shrink based on resource demand. VioCluster creates software-based network components which seamlessly connect physical and virtual machines to create isolated virtual domains cluster. Administrators are able to configure borrowed machines as required, while not granting root privileges to others making use of their nodes was described by P. Ruth in 2005.

A. *Scheduling the Workflow in Hybrid Cloud using PCH Algorithm*

Here it is introduced with Path Clustering Heuristic (PCH) algorithm in hybrid cluster for the initial schedule scheme. The hybrid cloud systems are a novel research challenge that comes together with the merging of private and public clouds. It checks the private resources whether it already satisfies the deadline. If the deadline is not satisfied, the algorithm starts the process of deciding which resources it will request to the public cloud. In this method, the different cluster configurations are considered with PCH algorithm dynamically and 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. We propose an approach to schedule service workflows that consider a deadline and execution costs. Deploying a hybrid cloud that offers support or automatic service installation in the resources dynamically provided by the grid or by the cloud to execute the PCH algorithm. In PCH algorithm all the information necessary to compute these attributes is given by the programming model or by the infrastructure. Initially, it is assumed as a virtual homogeneous system composed of an unbounded number of the best processor available connected by links with the highest bandwidth available. Each task is scheduled on a different processor on the virtual system, and then the algorithm computes the initial attribute values of each node. This decision is based on performance, cost, and the number of services to be scheduled in the hybrid cloud using PCH algorithm.

The initial step in scheduling process is based on the default number of service mangers in which it manages the virtual machines.

(i) Computation Costs

$$w_{i,r} = \frac{instructions}{p_r}$$

wi,r represents the computation cost (time to execute the node) of the node i in the resource r, and pr is the processing capacity of resource r in instructions per second.

(ii) Communication Costs

$$c_{i,j} \frac{data\ i,j}{lr.p}$$

ci,j represents the communication cost (time to transfer data) between nodes ni and nj using the link l between resources r and p. If r = p, then ci,j = 0.

(iii) Priority

$$P_i = \begin{cases} w_{i,r} & , \quad if\ i\ has\ no\ successor \\ w_{i,r} + \max_{\forall n_j \in suc(n_j)} \left(C_{i,j} + P_j\right) & , otherwise \end{cases}$$

Pi is the priority level of node i at a given time instant during the scheduling process.

(iv) Estimated Start Time

$$EST(n_i|n_k) = \begin{cases} Time(r_k) & , \quad if\ i = 1 \\ max\{Time(r_k), ST_i\} & , \quad Otherwise \end{cases}$$

$$\text{where } STi = \max_{\forall n_h \in pred(n_i)} (EST(n_h, r_k) + w_{h,k} + c_{h,i})$$

EST(ni, rk) represents the earliest start time possible for node i in resource k at a given scheduling instant. Time (rk) is the time when resource k is available to execute node i.

(v) Estimated Finished Time

$$EFT(n_i, r_k) = EST(n_i, r_k) + w_{i,k}$$

EFT(ni, rk) represents the estimated finish time of node i in resource k.

Based on the attributed value, similar tasks are assigned to each group. Once the initial scheduling is completed, the resources in Private cloud checks whether the required resources has meet the deadline. Also it ensures the availability of resources in the Public cloud. If there are sufficient resources exists in Private cloud then PCH algorithm has no necessity to take part. Else the required resources have to be acquired from Public cloud. Until the deadline is met in Public cloud the iteration of the resources will take place until it is equivalent to the number of nodes in the service manager.

Now it will select the node which will have the maximum $P_i + EST_i$ and that will be assigned to schedule in the private cloud. This will be added to the existing and rescheduled nodes from the Private cloud to Public cloud. Now it calculates the number of nodes which will result the number of resources that is obtained from Public cloud.

An iteration is performed in the computer cluster and thus, selects the smallest $\frac{price}{num\_cores_i \times P_i}$ with num_cores < num_clusters. This will now schedule each node in the resources which has smallest EFT.

*B. **PCH Algorithm for Hybrid Cloud***

Schedule G in the Private cloud using PCH

R=all resources in the Private cloud

If Makespan (G) > deadline (G) then
While Makespan (G) > D AND iteration < size (G)
do
iteration = iteration +1
select node n_i such that $P_i + EST_i$ is maximum and node n_i is currently scheduled in the Private cloud.
T= T U t
Num_clusters = number of clusters of nodes in T

While um-cluster > 0 do

Select resource r_i from the public clouds such that $\frac{price}{num\_coresi \times Pi}$ is minimum and num_core s_i <= num_clusters.
R = R U r_i
Num_clusters = num_clusters – num_core s_i
end while
for all n_i ε T do
schedule n_i in r_n ε R such that EFT_i is minimum
recalculate EST s and EFT s
end for
end while
end if

III. PERFORMANCE EVALUATION

The proposed approach is evaluated and compared with the existing approach for the job scheduling problem. The evaluation comparison by the parameter metrics such as

- Scalability,
- Execution Time and
- Cost.

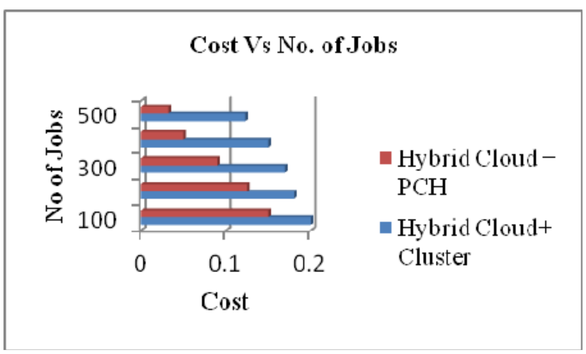

Figure1. Cost Vs Number of Jobs

Figure1. Analyse and compare the performance offered by different configurations of the computing cluster with proposed PCH algorithm. Here if the number of jobs increased the cost is decreased linearly. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing system.

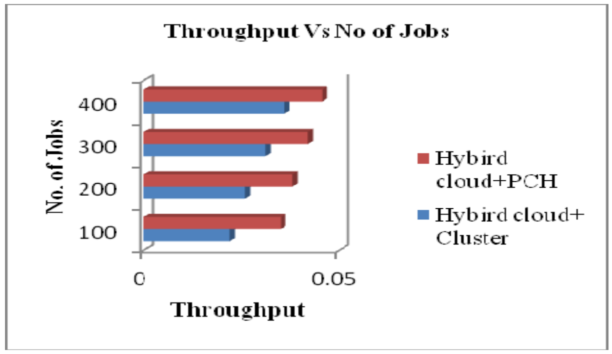

Figure. 2. Throughput Vs Number of Jobs

Fig.2. Analyses and compares the performance offered by different configurations of the computing cluster with proposed PCH algorithm. Here if the number of jobs increased the throughput of the system is increased linearly. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing systems.

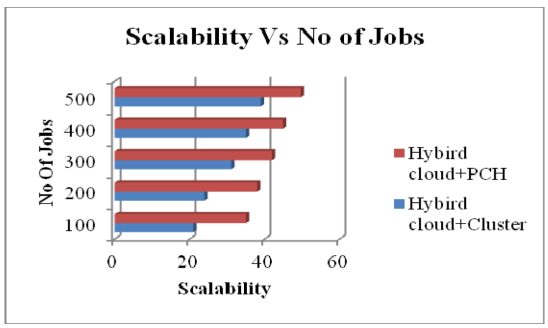

Figure. 3.Scalability Vs Number of Jobs

Fig.3. Analyses and compare the performance offered by different configurations of the computing cluster with proposed PCH algorithm. Here if the number of jobs increased the scalability of the system is increased linearly. Based on the comparison and the results from the experiment show the proposed approach works better than the other existing systems.

IV. CONCLUSION

The approach have analysed that the challenges and viability of deploying computing clusters for loosely coupled MTC applications with the help of three different cloud networks like private, public and hybrid. And it also examines the performance of different cluster configurations, using the cluster throughput (i.e., completed jobs per second) as performance metric. The clusters are deployed in a hybrid setup, which combines local physical nodes with virtual nodes deployed in another compute cloud. Compare the different cluster configurations, and proving the viability of Multicloud solution also from a cost perspective. In this method, the different cluster configurations are considered with PCH algorithm dynamically and 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. Future work will focus on new approach for hybrid cluster. Here we introduce hybrid cluster using Path Clustering Heuristic (PCH) algorithm. This is used for the initial schedule scheme. It checks the private resources whether it already satisfies the deadline. If the deadline is not satisfied, the algorithm starts the process of deciding which resources it will request to the public cloud. This decision is based on performance, cost, and the number of services to be scheduled in the hybrid cloud using PCH algorithm. The Private cloud can release its own resources under any appeal.

REFERENCES

1. Aboulnaga, K. Salem, A. Soror, U. Minhas, P. Kokosielis, and S. Kamath, "Deploying Database Appliances in the Cloud," Bull. of the IEEE Computer Soc. Technical Committee on Data Eng., vol. 32, no. 1, pp. 13-20, 2009.
2. Azeez, Autoscaling Axis2 Web Services on Amazon EC2: ApacheCon Europe, 2009.
3. Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2, 2010.
4. BioTeam "Howto: Unicluster and Amazon EC2," technical report, BioTeam Lab Summary, 2008.
5. Chase, D. Irwin, L. Grit, J. Moore, and S. Sprenkle, "Dynamic Virtual Clusters in a Grid Site Manager," Proc. 12th IEEE Symp. High Performance Distributed Computing, 2003.
6. Huedo, R.S. Montero, and I.M. Llorente,"The GridWay Framework for Adaptive Scheduling and Execution on Grids," Scalable Computing—Practice and Experience, vol. 6, pp. 1-8, 2006.
7. Walker, "The Real Cost of a CPU Hour," Computer, vol. 42, no. 4, pp. 35-41, Apr. 2009.
8. Walker, J. Gardner, V. Litvin, and E. Turner, "Creating Personal Adaptive Clusters for Managing Scientific Jobs in a Distributed Computing Environment," Proc. IEEE Second Int'l Workshop Challenges of Large Applications in Distributed Environments (CLADE '06).
9. ElasticHosts, http://www.elastichosts.com/, 2010.
10. Foster, T. Freeman, K. Keahey, D. Scheftner, B. Sotomayor, and X. Zhang, "Virtual Clusters for Grid Communities," Proc. Sixth IEEE Int'l Symp. Cluster Computing and the Grid, 2006.
11. 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.
12. 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.
13. J. Fronckowiak, "Auto-Scaling Web Sites Using Amazon EC2 and Scalr," Amazon EC2 Articles and Tutorials, 2008.
14. K. Keahey, M. Tsugawa, A. Matsunaga, and J. Fortes, "Sky Computing," IEEE Internet Computing vol. 13, no. 5, pp. 43-51, Sept./Oct. 2009.
15. M. Murphy, B. Kagey, M. Fenn, and S. Goasguen, "Dynamic Provisioning of Virtual Organization Clusters," Proc. Ninth IEEE Int'l Symp. Cluster Computing and the Grid, 2009.
16. M.A. Frumkin and R.F. Van der Wijngaart, "NAS Grid Benchmarks: A Tool for Grid Space Exploration," J. Cluster Computing, vol. 5, no. 3, pp. 247-255, 2002.
17. P. Ruth, P. McGachey, and D. Xu, "VioCluster: Virtualization for Dynamic Computational Domains," Proc. IEEE Int'l Conf. Cluster Computing, 2005.
18. R.S. Montero, R. Moreno-Vozmediano, and I.M. Llorente, "An Elasticity Model for High Throughput Computing Clusters," to be published in J. Parallel and Distributed Computing, doi: 10.1016/ j.jpdc.2010.05.005, 2010.
19. Raicu, I. Foster, and Y. Zhao, "Many-Task Computing for Grids and Supercomputers," Proc. Workshop Many-Task Computing on Grids and Supercomputers, pp. 1-11, 2008.

BIOGRAPHY

V.Venkatesakumar is currently working as an Assistant Professor in Anna University of Technology Coimbatore. He received Bachelor and Master Degree, B.E and M.E from Anna University Chennai. He has 6 years of Teaching Experience. Currently he is an academician and doing his doctorate. His research area includes Operating System, Software Engineering and Web Technologies.

M.Sindhu received B.Tech degree in Information Technology from Anna University, Chennai in 2010. She is pursuing her M.E in Computer Science and Engineering, Anna University of Technology, Coimbatore.