International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

Resource Provisioning Policies in Hybrid Cloud on Resource Failures: A Practical Study

Roqaiya Fatima

Abstract: Hybrid cloud is increasing attention in recent days, and it requires considerable exploration. 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 has an important and challenging problem in the large scale distributed system such a cloud computing environments. As resource failures may take place due to the increase in functionality and complexity in hybrid cloud. So a resource provisioning algorithm that has ability of attending the end user quality of service (QoS) is paramount. In this paper resource provisioning policies are studied to assure the QoS targets of the user. These policies are also examined in the paper against the backdrop of including the workload model and the failure correlations to redirect user's requests to the appropriate Cloud providers. While applying real failure traces and a workload model, the proposed resource provisioning policies are evaluated in order to demonstrate their performance, cost as well as performance-cost-efficiency. Under circumstances, they become able to improve the user's QoS about 32% in terms of deadline violation rate and 57% in terms of slowdown with a limited cost on a public Cloud.

Index Terms: Aurin, CloudSim, Hybrid Cloud, Inter-Grid Gate-way, Quality of Service (Qos), SLA, Virtual Machines.

1 INTRODUCTION

Cloud computing is a process in which a large number of computers are connected through real time communication network via Virtual Machines (VMs). Cloud computing completely depends on sharing of resources so that it achieves coherence and scalability, like the electricity distributed over a large network. The backbone of cloud computing is to focus on maximizing the impact of the resources that are shared. These resources are dispensed among multiple users and also functionally redistributed on demand. When their users satisfies by the cloud services like applications, data storage, software and other processing capabilities, Organizations improve their efficiency and give responses fast according to the user requirement. Apart from attractive characteristics of cloud computing it is still in starting phase and has many research challenges and issues to be addressed such as, how resource provisioning are automated, virtual machine migration, energy management, security for data and server consolidation [1].

Generally, Cloud has different deployment models like Public, Private and Hybrid Clouds. Public Cloud offer services on large-scale data center that includes huge number of servers and data storage system. The motive of public Clouds are to provide IT capacity related to open market offerings. Users can easily access their applications from anywhere and pay for their usage. Amazon's EC2 and Go Grid are such examples of public Cloud. Private Cloud is different from Public Cloud that allows users locally to manage loads in their own domain in an agile & fixed infrastructure. Additionally, private Clouds have small scalability of cloud systems that is usually a holder of a single organization. Some Private Clouds are like Go Front's Cloud and NASA's Nebula. The Hybrid Cloud is the services utilization & integration of both public and private Clouds (Fig 1). Recently Hybrid Cloud is the most popular Cloud computing deployment models by employing public Cloud services with local resources of private Cloud [2].

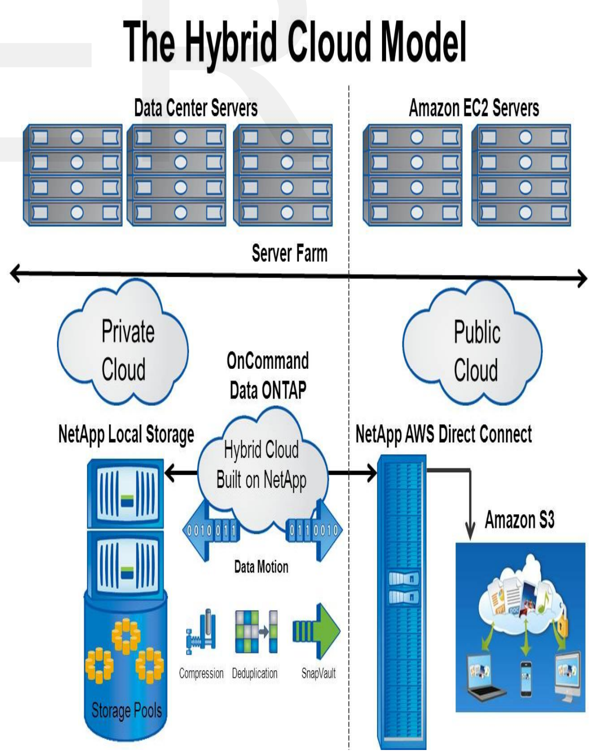

Fig 1: The Hybrid Cloud Model

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

The platform of hybrid cloud not only increases scalability and reduces cost of the public Cloud by paying for IT resources consumed i.e. server, storage & connectivity but also by delivering the performance levels and control availability in private Cloud environments with no change in their IT setup. Thus an integration of public and private clouds should be the major issue for computing hybrid cloud infrastructure. As increment of complexity and functionality in the hybrid cloud make resources failures on rise. These failures may result to degradation in performance, loss and corruption of data, dreadful loss to customers, premature termination of execution, violation of Service Level Agreements (SLAs). So resource provisioning approaches are essential for the enhancement of hybrid Cloud.

Hence the paper aims to introduce some strategies that would consider a resource failure correlations and a workload model while choosing the selected Cloud providers. Many research papers have been presented to adopt the public cloud. Such works improves the level of performance as well as cost benefits for this application. Recently how companies opt local clusters to improve the performance level by allocating resources given by Assunc¸ ao et al [3]. These strategies are not considered for the type of workload and the failures for resource to make feedback for request redirection.

This paper remarkably tends to consider that resource failure could cause due to software and hardware faults. Also, assuming the fact that the proposed strategies take benefits of a knowledge free approach which does not need any information about the model. These strategies had been proposed and studied by a project named "AURIN" [28] for the inception of developing the e-infrastructure to research urban build environment. To support multi urban research activities a visualized collaborative environment is built, which will federate the access of heterogeneous data sources through a browser infrastructure. This paper also observed that the applications of workflow in the "AURIN" [28] project as a workload for the similar jobs have been requested for resources in aspects of Virtual Machines and assumed that the end-user Quality of Service requirements are as a deadline.

2 System Overview

This portion explains the overview of the architecture and implementation of the "AURIN" project as well as the hybrid Cloud system that is studied in certain papers.

2.1 The Aurin Architecture

The "AURIN" [28] Project is applied by evolving an online infrastructure that could support certain research activities for a broad range of urban environment [7]. This architecture is based on single point of entry. As the sign on ability is applied to mix Australian Access Federation that give support throughout the Sector of Australian University. A portal facilitates access to a different data set. Portals identify component of user interface lies within a service oriented loosely coupled architecture, disclose discovery and search of data, maps services and enhance other properties of visualization

Large local libraries like Java, federated services like "REST / SOAP" tools have been unveiled by a workflow environment related to working of "OMS" (Object Modeling System). These technical tools allow advancement of analysis in spatial as well as non-spatial data so that a complex environment is built that could provide urban research activities. "AURIN" s Users could able to integrate various workflows that may be data intensive. "AURIN workflows" are allowed to work on cluster & cloud computing environments for user workflows by using OMS framework [8].

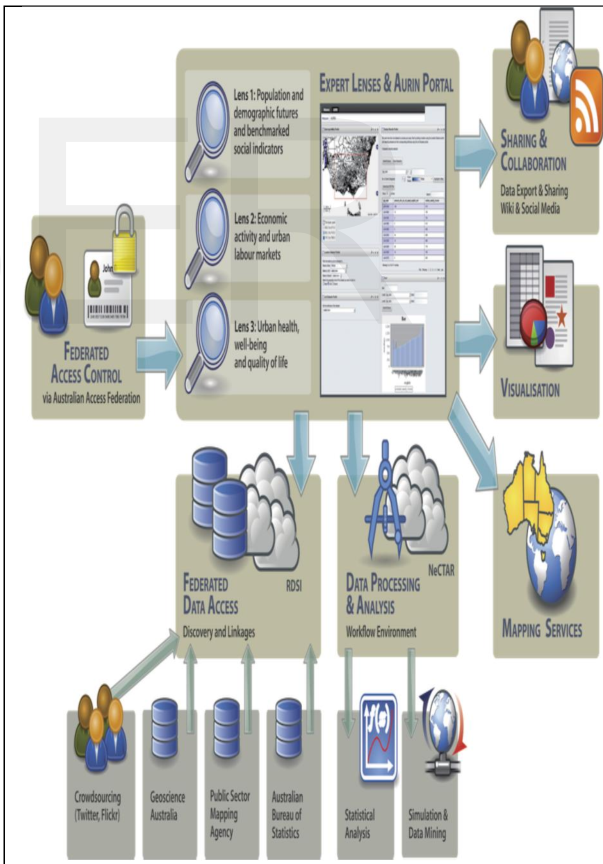

Fig 2: "The Aurin Architecture" [28]

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

2.2 System of Hybrid Cloud

When the local resources from private cloud are failed short to meet the user's requirements then the hybrid cloud system takes the advantage of public cloud to supply the computing capacity to their users. In the Hybrid cloud system architecture (in Fig 2), Inter-Grid components are structured and applied in the "Cloud Bus5" research group for its greater efficiency. The main component of this architecture is the broker that chooses appropriate resource provider providing various computing services for input requests. The broker allows interconnecting different types of resource manager as an "Inter-Grid Gateway". In order to operate local resources it is feasible to link with Open Nebula [10] / Eucalyptus [11]. Moreover there are two interfaces that are evolved to connect with a grid & Iaas Providers and with "Nectar" research group [12].

2.3 Workload Model

"AURIN" [35] users operate and access data in different situations [8], [7]. Usually that is the case when different tasks need huge no. of resources in short interval of time. As model workflows with various jobs are dependent on communication networks that lead to resource failures. These workflows declare that they should be tightly coupled for single cloud architecture. So works are assumed to determine resources virtually from one provider. Users can help the workflows by sending a VM's request to the broker (in fig 2). Each request exhibits different properties depending on the number of Virtual Machines, types of Virtual Machines, estimated time for request and the request deadline. When such request is arrived to brokers it help them to decide which resource provider should be used. Thus different resource provisioning policies & strategies are described for managing such user's requests by brokers for efficient utilization.

3 The Resource Provisioning Policies

The resource provisioning policies are introduced which include some scheduling algorithms and some brokering strategies which will surely help to contribute the incoming load on public cloud providers. The providers of public cloud select appropriate engineered modules so that it could not include those irrelevant components which will result to resource failures. Moreover it is too costly for a private cloud to opt this design style which makes it a less reliable. Hence it needs to concentrate more on private cloud for resource failure.

These Policies are the part of the broker like Inter-Grid Gateway which is described here.

The Strategies depend on the Workload Model as well as the Failure Correlation that are good for a knowledge-free approach, so the failure model does not require any further information.

3.1 User Request

Each User's request is assumed as a rectangle where length represent the duration of user' request (T) and the breadth is no. of Virtual Machines (Q). Since the Private cloud resources are failure prone so it is possible for occurrence of some failure events (E) in the nodes when a request has been serviced. However some correlations like spatial & temporal occur in the failure events that are dependent on the workload type and capacity of rate of the failure [14], [15], [16]. When several failures take place on various nodes in small duration i.e. called as Spatial Correlation, while Temporal Correlation in failures give dissymmetry of the failure distribution over a period. So there is possibility of some overlapped failure events also to be occurring in the system.

Let 'Ta()' & 'Te ()' are the functions that are the arrival and closing times of a failure event. Let 'O' be the order of overlapped failure occurrence in start time on increasing order,

$$O = \{Fi|Fi = E1, E2, \dots, En, Ta(Ei + 1) \leq Te(Ei)\} \quad (1)$$
$$\text{where } 1 \leq i \leq n - 1$$

As workflows are tightly coupled, so all VMs should be present for the whole time estimated for a request. If any failure events occur in any VM then that could make the whole request to stall for further execution. So for this case, the downtime of the service will be;

$$D = \forall Fi \in O(\max\{Te(Fi)\} - \min\{Ta(Fi)\}) \quad (2)$$

The analyses demonstrated above that even if the Private Cloud has the mechanism of fault tolerant as optimal then also there is the condition of delay in request time of 'D' units which may possibly exceed its deadline.

Additionally if any request is stopped for a certain interval due to overlapping failures, a large delay makes its services unapproachable. To make the Private Cloud more reliable there are three different strategies which will help to face failures working on the workload model for such failure happenings.

3.2 Size-based Strategy

Spatial Correlations are the failures which occur multiply on the various nodes in a short interval of time in the distributed systems [14][15]. This characteristic could be harmful if every request requires whole VMs for whole span of time. From Equation 2 it can be said that downtime is dependent on no. of VMs. So more VMs are requested then there is more likelihood of jobs failures. To deal with these situations redirection strategy has been evolved to send large requests with larger Q to Public Cloud Systems (more

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

reliable), reducing the requests to send in Private Cloud (failure-prone). The core property of this redirection strategy is to calculate the mean no. of Virtual Machines needed for per request. Hence to calculate the mean no. of VMs, probabilities of various no. of VMs for the incoming requests are determined. Let P1 is the probability of request with 1 VM, P2 is the probability with 2 VM. Mean no. of VMs for the request is:

$$Qm = P1 + 2\text{\^{}}x(P2) + 2\text{\^{}}x(1 - (P1 + P2)) \quad (3)$$

Parallel workloads are based on size of requests with uniform distribution of parameters (a, b, c, d). Distribution has two parts. 1st one has the probability of‘d’ with interval [a, b] whereas 2nd has ‘1-d’ probability of interval [b, c]. ‘b’ is the mid-point between ‘a’ & ’b’. Hence ‘x’ is the mean value for uniform distribution as:

$$x = (da + b + (1 - d)c)/2 \quad (4)$$

Public Cloud provider can now only service these requests if the Q>Qm otherwise the request will not be submitted.

3.3 Time-based Strategy

Additionally Spatial Correlation has the failure of events that are usually the domain of time which can lead to deflect the failure rate [14]. Consequently, the failure distribution rate is time-dependent. Few periodic failure patterns can be executed in time-scales [16]. Requests are usually longer get affected by temporal correlations as they live long and cause more failures to the systems. Downtime & Duration for request are strongly correlated. In contrast to others, real time distributed systems have large duration request [18] [19] which is meant for a small fraction of whole requests to take part in the main load. This strategy provides an efficient behavior to deal with above characteristics. The duration for mean request is the point of decision for a gateway is to re-direct the requests into the cloud providers. So it can be said that if the request takes less or equal time than the mean request, the request will utilize private resources. Most of the short requests would find the deadline of workload as they have fewer chances to face failures. Longer requests will meet the deadline under scaled availability of resources and are allocated to the authentic provider i.e. Public Cloud. The duration of mean request can be easily found from the fittest method of distribution on the workload model. Parallel workloads have the request duration as a log normal distribution with parameters (α, β).
So the mean request is:

$$Tm = e^{\alpha+(\beta^2)/2} \quad (5)$$

If the request duration T>Tm then the redirection strategy address the claim to public cloud provider otherwise will be served by the private cloud resources.

3.4 Area-based Strategy

Above mentioned two strategies are completely focused on one aspect of request i.e. request duration (T) & no. of VMs (Q). Third strategy is a compromise of both the strategies. This is using area of a request as a checkpoint for a gateway as a rectangle with length as ‘T’ and breadth as ‘Q’. A mean request can be calculated by integrating the mean no. of VMs with request duration mean as:

$$Am = Tm * Qm \quad (6)$$

Public Cloud provider accepts the request if this redirection strategy prepares requests to make the request’s A> Am otherwise will be appended to private resource providers. Long & broad range of requests can now be send to providers of Public Cloud. So this strategy is less conservative than time based but more than Size based Strategy.

3.5 Scheduling Algorithms

Earlier it was described that the resource provisioning policies include scheduling algorithms and brokering strategies for addressing to both public as well as local resources. So there is two popular algorithms are applied to schedule the requests. First one is Conservative Backfilling [20] & other one is Selective Backfilling [21]. “Conservative Backfilling” can work with each request. These requests are forwarded to queue if other requests are not stopped.
“Selective Backfilling” offers priorities to those requests whose estimated slowdown rate is greater than the threshold value. So it is good for those requests that have been waiting for long time in the queue.
The slowdown required for a request is determined by “Expansion Factor” [35] and given as:

$$XFactor = (Wx+Tx) / Tx \quad (7)$$

Where ‘Wx’ is the waiting time by the request x and ‘Tx’ is the run time for a request x.
When the selected requests are placed in the scheduler, each runs on the Virtual Machines’ nodes available in scheduler. If resources are failed in between the execution, checkpoints are placed to check the applications from where these applications are stopped so that it could restart again when the new node is available at desired location.

4 EVALUATING PERFORMANCE

The performance level of provisioning policies are studied and observed that they are checked by a discrete event simulator called Cloudsim [23]. The performance metrics which are taken into account all simulation situations where rate of violation and the slowdown [24] is bounded. The slowdown is bounded

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

for the ‘N’ requests are defined as:

$$\text{Slowdown} = 1/M\sum_{x=1}^{M}\frac{\{Wi+\max(Tx,bound)\}}{\max(Tx,bound)} \qquad (7)$$

where bound is fixed to a very less time to remove the effect of short requests [24].

4.1 Results

The results are shown in Figure 3 on evaluation of violation rates versus several workloads for various policies regarding resource provisioning. Each figure describes three brokering strategies for a scheduling algorithm. Size is Size based, Time is time based, Area is Area based brokering strategies. CB refers to Conservative Backfilling and SB refers to Selective Backfilling.

In Figure 3 as there is increment in the workload intensity like rate of arrival, duration and size of requests there is increase in the rate of violation for these provisioning policies. The size-based brokering strategy gives a very low violation rate, Time-based strategy has very worst performance of violation rates and the Area-based strategy demonstrates average performance when there is increment of workflows is shown by figures.

Size-based brokering strategy has low violation rate as it has indirect dependency on the request size, the number of deadlines can be increased on reducing the size of requests. This characteristic is because of the increase in the number of re addressing the requests to the Public Cloud (failure-prone) in the size-based brokering strategy. Besides, the policies using a selective backfilling scheduler improves the performance than the conservative backfilling [25][26][27].

Fig 4 shows the requests slowdown for all provisioning policies versus several workloads. However in figure 4(b) slowdown versus duration of request has been shown which clearly point out that the slowdown can be less sensitive to the duration of request. Figure 4(c) describes that the slowdown decreases on reducing the size of request for Time based and Size based brokering strategies [28]

The policies can have various cost and performance level. Required level of Quality of Service (Qos) and budget constraints should be kept in mind to choose appropriate resource provisioning policy[24[25][29]. Although Time-based Strategy can be opted effectively to decrease the rate of violation by 20%.

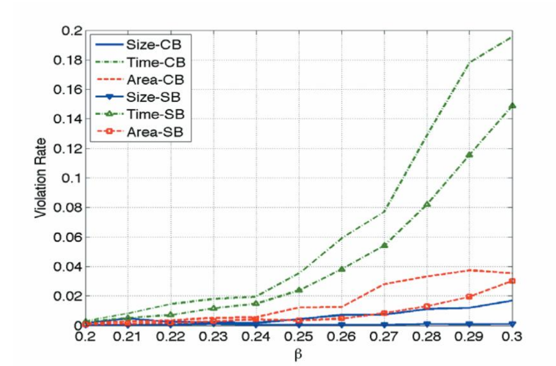

Fig 3.1: Arrival Rate

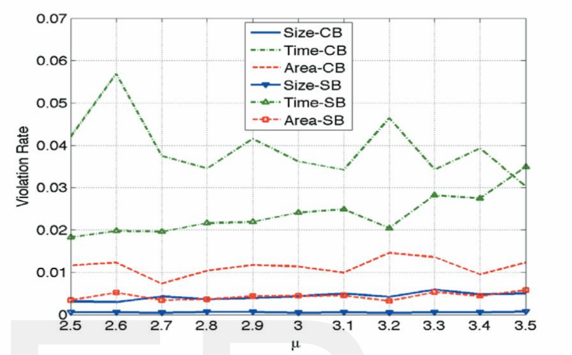

Fig 3.2: Request duration

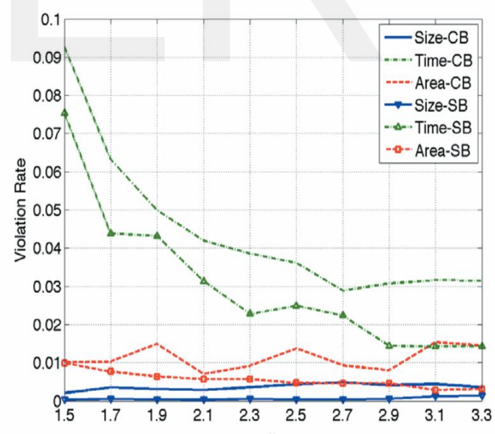

Fig 3.3: Request size

Fig 3: Violation Rate for all provisioning policies v/s different workloads

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

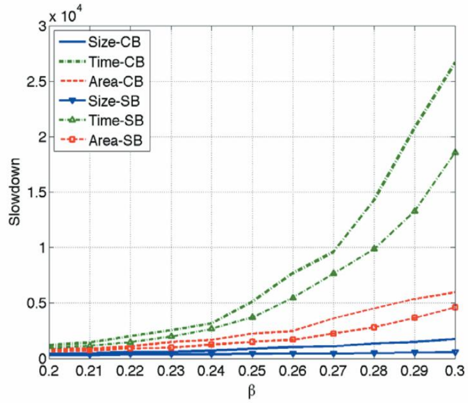

Fig 4.1: Arrival Rate

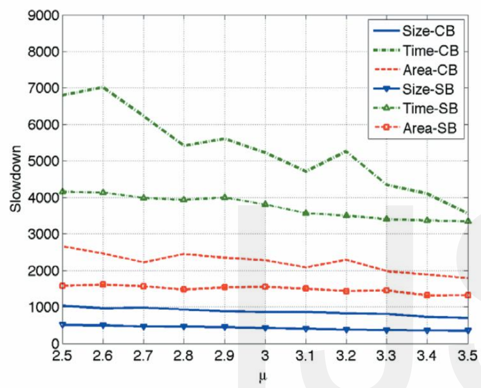

Fig 4.2: Request Duration

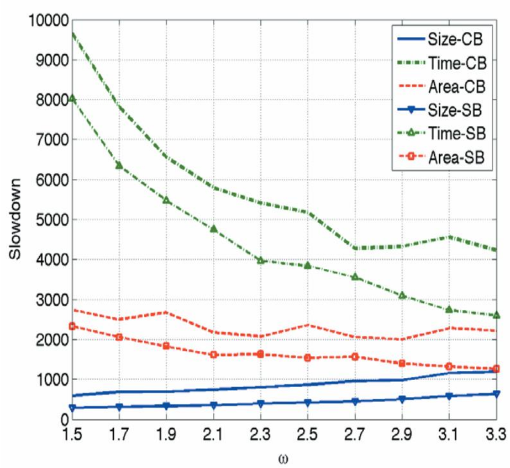

Fig 4.3: Request Size

Fig 4: Slowdown for all provisioning policies versus different workloads [28]

5 RELATED WORK

The work which is studied and observed in research papers are divided into two parts: sharing of load in the distributed systems and solutions to utilize the Cloud resources properly without any resource failure so to upgrade the potential of computing infrastructures. Various load sharing processes have been discussed for various distributed systems.

- In [2], virtual infrastructure management is designed via two open source projects: "OpenNebula" and "Haizea8". Apart from that Inter-Grid infrastructure is built on the working of virtual machine technology and can be connected to any distributed systems via "Virtual Machine Manager (VMM)"[9].
- The authors in [3] have suggested an organization to take a local cluster to provide proper benefits to enhance the performance level of their user requests.
- In [4], the authors tried to find the cost of running a scientific workflow over a Cloud. They found that the computational costs are outweighed by storage costs for the applications.
- In [6], the authors have tried to prove that the services like Amazon are used for data-intensive applications. They conclude that costs in monetary are higher than the collective costs for storage groups such as durability, availability and access performance. The applications which have characteristics of data are usually not requiring most of its properties.
- Iosup et al. [27] has introduced a match-making technique to enable the allotment of resources in Grids computing.
- Balazinska et al. [28] has proposed one technique for transferring the processing operators in a system distributed. Leased resources are used to enhance the level of Qos as a Resource provisioning policy on resource failure.
- Montero et al. [30] used Inter-Grid Way to contribute the virtual machines and support the interoperability on a Globus Grid.
- The author in [32] used a "VioCluster", computing infrastructure system in which the broker is liable for operating and allocating their resources dynamically by taking the systems in advance among clusters in a virtual domain. did a cost-benefit analysis of Clouds. However, no distinct variety of scientific applications is observed.
- In [33], the authors proposed a model of an Elastic Site that properly used the services offered to a site.
- In [34] it is discussed that the gang scheduling are done to send parallel jobs to a cluster of VMs hosted on "Amazon EC2". Workload model and correlation of failures are considered to take public resources in

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

advance.

6 CONCLUSION

This paper discusses the problems of Hybrid Cloud System involving local Cloud as failure prone. Different brokering strategies are discussed to manage local resources as well as public resources in the hybrid cloud where an organization should focus to improve the Quality of Service (Qos) of User & its own organization. These strategies which are discussed here are adopting the workload model and failure correlations. These policies are knowledge free approach that is not requiring any details of resources of private cloud (failure-prone). They can have different effect and influence factor on performance including the rate of violation and the slowdown.

Thus in order to opt for appropriate strategy, organizations should consider their Qos requirement level of user as well as its own and the budget availability.

ACKNOWLEDGEMENT

This work is supported and guided by the Founder, Dr. Ashok Chauhan, Head of Department, Dr Abhay Bansal, Program Leader Mr Abhishek Singhal & other faculties of Amity University, Noida, India.

Author wishes to thank her parents, brothers, sisters, and friends with their unfailing humor and warm wishes.

REFERENCES

[1] Amazon Inc. Amazon Elastic Compute Cloud (Amazon EC2). http: / /aws.amazon.com /ec2.

[2] B. Sotomayor, R. S. Montero, I. M. Llorente, and I. Foster. Virtual nfrastructure management in private and hybrid clouds. IEEE Internet Computing, 13(5):14 – 22, Sep. 2009.

[3] M. D. de Assunc¸ ˜ao, A. di Costanzo, and R. Buyya. Evaluating the cost-benefit of using cloud computing to extend the capacity of clusters. In 18th International Symposium on High Performance Parallel and Distributed Computing (HPDC 2009), pages 141–150, Garching,

[4] E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good. The cost of doing science on the Cloud:The montage example. In 19thACM / IEEE InternationalConference on Supercomputing (SC 2008), pages 1–12, Piscataway, NJ, USA, 2008.

[5] D. Kondo, B. Javadi, P. Malecot, F. Cappello, and 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), pages 1–12, Rome, Italy, 25-28 May 2009. IEEE Computer Society, Washington, DC.

[6] M. R. Palankar, A. Iamnitchi, M. Ripeanu, and S. Garfinkel. Amazon S3 for science Grids: a viable solution? In 1st International Workshop on Data-aware Distributed Computing (DADC'08) in conjunction with HPDC 2008, pages 55–64, New York, NY, USA, 2008.

[7] R. O. Sinnott, G. Galang, M. Tomko, and R. Stimson. Towards an einfrastructure for urban research across Australia. In eScience, pages 295–302, 2011.

[8] B. Javadi, M. Tomko, and R. O. Sinnott. Decentralized orchestration of data-centric workflows using the object modeling system. In CCGRID, pages 73–80, 2012.

[9] A. di Costanzo, M. D. de Assunc¸ ˜ao, and R. Buyya. Harnessing cloud technologies for a virtualized distributed computing infrastructure. IEEE Internet Computing, 13(5):24–33, 2009.

[10] J. Font´an, T. V´azquez, L. Gonzalez, R. S. Montero, and I. M. Llorente. OpenNEbula: The open source virtual machine manager for cluster computing. In Open Source Grid and Cluster Software Conference – Book of Abstracts, San Francisco, USA, May 2008.

[11] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman, L. Youseff, and D. Zagorodnov. The Eucalyptus open-source cloud-computing system. In 1st Workshop of Cloud Computing and Its Applications, October 2008.

[12] T. Fifield. NeCTAR research Cloud node implementation plan. Research Report Draft-2.5, Melbourne eResearch Group, The University of Melbourne, October 2011.

[13] U. Hoelzle and L. A. Barroso. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines. Morgan and Claypool Publishers, San Rafael, CA, 2009.

[14] S. Fu and C.-Z. Xu. Quantifying event correlations for proactive failuremmanagement in networked computing systems. Journal of Parallel and Distributed Computing, 70:1100–1109, November 2010.

[15] M. Gallet, N. Yigitbasi, B. Javadi, D. Kondo, A. Iosup, and D. Epema. A model for space-correlated failures in large-scale distributed systems. In Euro-Par 2010, volume 6271 of Lecture Notes in Computer Science, pages 88–100. 2010.

[16] N. Yigitbasi, M. Gallet, D. Kondo, A. Iosup, and D. Epema. Analysis and modeling of time-correlated failures in large-scale distributed systems. In 8th IEEE / ACM International Conference on Grid Computing, October 2010.

[17] H. Li, D. Groep, and L. Wolters. Workload characteristics of a multi-cluster supercomputer. In 10th International Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP), pages 176–193, New York, USA, 2004.

[18] D. G. Feitelson. Workload Modeling for Computer Systems Performance Evaluation. 2009.

[19] L. F. Orleans and P. Furtado. Fair load-balancing on parallel systems for QoS. In 36th International Conference on Parallel Processing (ICPP 2007), pages 22 –22, 2007.

[20] A. W. Mu'alem and D. G. Feitelson. Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. IEEE Transactions on Parallel and Distributed Systems, 12(6):529–543, 2001.

[21] S. Srinivasan, R. Kettimuthu, V. Subramani, and P. Sadayappan. Selective reservation strategies for backfill job

International Journal of Scientific & Engineering Research, Volume 5, Issue 5, May-2014
ISSN 2229-5518

scheduling. In 8th International Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP '02), volume 2537 of LNCS, pages 55–71, London, UK, 2002.

[22] M. Bouguerra, T. Gautier, D. Trystram, and J.-M. Vincent. A flexible checkpoint/restart model in distributed systems. In 9thI nternational Conference on Parallel Processing and Applied Mathematics, volume 6067 of LNCS, pages 206–215. Springer, 2010.

[23] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41(1):23–50, 2011.Germany, June 2009.

[24] D. G. Feitelson, L. Rudolph, U. Schwiegelshohn, K. C. Sevcik, and P. Wong. Theory and practice in parallel job scheduling. In 3rd Job SchedulingStrategies for Parallel Processing (IPPS'97), pages 1–34, London, UK, 1997.

[25] D. Kondo, B. Javadi, A. Iosup, and D. H. J. Epema. The Failure Trace Archive: Enabling comparative analysis of failures in diverse distributed systems. In 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (CCGRID), pages 398–407, 2010.

[26] M. Islam, P. Balaji, P. Sadayappan, and D. K. Panda. QoPS: A QoS based scheme for parallel job scheduling. In 9th International Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP '03), volume 2862 of LNCS, pages 252–268, Seattle, WA, USA, 2003.

[27] I. Moschakis and H. Karatza. Evaluation of gang scheduling performance and cost in a cloud computing system. The Journal of Supercomputing, pages 1–18, 2010.

[28] Bahman Javadi, Jemal Abawajy, Richard O. Sinnott. Hybrid Cloud Resource Provisioning Policy in the Presence of Resource Failures. In 4th IEEE International Conference on Cloud Computing Technology and Science, 2012.

[29] Ye Hu, Johnny Wong, Gabriel Iszlai and Marin Litoiu. Resource Provisioning in cloud computing. University of Waterloo, IBM Toronto Lab, York University, 2009.

Roqaiya Fatima is a research fellow in area of 'Cloud computing' pursuing Master's Degree Program (M.Tech) in Computer Science & Engineering at Amity University, Noida, India.
E-mail- roqaiya13oct@gmail.com