SERVICE BROKER ROUTING POLICES IN CLOUD ENVIRONMENT: A SURVEY

Rekha P M[1] and M Dakshayini[2]
[1]Department of Information Science & Engineering, VTU, JSS academy of technical Education, Bangalore, Karnataka
[2]Department of Information Science & Engineering, VTU, BMS college of Engineering, Bangalore, Karnataka

ABSTRACT
Cloud computing refers to hosted online services accessed via the internet, which can help in the faster deployment of user applications for less cost. Cloud computing services offer great opportunities for clients to find the maximum service and best pricing. Algorithms, policies and methodologies are necessary for user application evaluation before actual deployment in cloud to select the best service with minimum cost. Hence to model and analyze such user applications Cloud-Analyst, a Cloud-Sim based toolkit is used. Simulator consists of different service broker policies, load balancing algorithms which give report as per user requirements. With the help of simulation report, we can modify the cloud environment as per the requirement of user. The service broker policy selects data centre located in different regions considering cost-effectiveness of Virtual machine, data centre processing and other parameters. We have analyzed various service broker policies for selecting data centre and compared their cost. This paper describes some of the issues and challenges related to the response time, cost effective data centre selection in cloud environment which would benefit both cloud users and researchers. In this paper we describe various service broker policies for effective data centre selection in cloud environment to reduce data centre processing cost and response time.

KEYWORDS: *Cloud Computing, Cloud-Sim, Cloud Analyst, Cost, Service Broker policy.*

I. INTRODUCTION

Cloud computing is an economic model to manage IT resources. Clouds data centers are designed by architecting them as networks of virtual services so that users can access and deploy applications from anywhere in the world on demand at reasonable costs depending on their QoS (Quality of Service) requirements. Cloud computing presents significant benefits to IT companies by freeing them from the low-level task of setting up hardware and software infrastructures. The availability of high end processors and advanced communication technology with the numerous interconnected hosts resulted in cloud computing. Different metrics of cloud computing are fault tolerance, availability, scalability, flexibility, reduced overhead for users, performance, on demand services etc. It provides resources, applications and system software as a service to the end user based on pay per use model. As the cloud is made up of data centers which are very much powerful to handle large numbers of users applications ranging from those that run for a few seconds to those that run for longer periods of time on shared hardware platforms. Hence the resource allocation policies and scheduling algorithms in cloud environments for various applications and service models is essential. However a technique has to be designed for distributing the user application workload among various data centre so as to minimize the response time, minimize the cost, minimize the resource utilization, and minimize the overhead. The data centers are targeted according to the availability and best response time with respect to location of data centre in the same region or different region. [11]
Data centers are expensive since it is usually built to serve irregular peak loads resulting in low average utilization of the resources. [20] User applications in cloud environment have different

structure, configuration and deployment requirements. Based on user requirements data centre can be designed for deploying applications. Measuring the performance of such applications is essential. Several Cloud providers are available, each one offering different pricing models, services and are located in different geographic regions. Application developers are concerned in selecting providers and data centre locations for typical large scale application on the Internet. Access to such services varies with respect to the time of the day, geographic location and service requests. With increase in requests Cloud allows infrastructures to react by dynamically increasing application resources, and reduces available resources when the number of requests reduces. To model the behavior of social networking applications such as Face book evaluation of costs and performance is essential.

Data centre is the main resource of the cloud which holds the computing and storage server with number of host machine. The main aim of data centres is to maximize the utilization of computing resources such as storage, CPUs, and network bandwidth as service by service providers at less cost. The optimization models aims to optimize both resource-centric such as utilization, availability, reliability and user-centric like response time, budget spent fairness. The aim of this paper is to briefly discuss about various efficient service broker policies and enhanced service broker policies using the tool called cloud Analyst.

This paper is organized as follows. Section II describes Cloud Analyst tool. Section III describes the various service broker policy and defines the environment parameters. Section IV describes related work and future directions. Finally, we conclude the paper in Section V and the future scope is suggested in the Section VI.

1.1 Factors For Response Time Improvement

Response time is the amount of time it takes when a request was submitted until the first response is produced. Response time has to be reduced while keeping tolerable delays.

1.11 Optimal Bandwidth

To run a user application service we require enough bandwidth to get better response time which in turn performance will be affected. Therefore, optimum bandwidth plays a vital role in the improvement of response time.

1.12 Protocol Selection

Protocol selection suited for the required application will improve the response time and hence will take out the extra operating cost. Various internet applications require specific protocols to run them. Response time and data center processing time act as a performance evaluation parameter.

Hence tool based simulation of cloud computing environment may help the users to access and deploy applications from anywhere in the world.

II. CLOUD ANALYST

Cloud Analyst, built on top of Cloud-Sim, allows a description of application workloads, has information of geographic location of users generating traffic and location of data centres, number of users, data centres, and number of resources in each data centre. [5] Data centre manages virtual machines configuration and destruction and does the routing of user requests received from user base by the use of the internet to virtual machines. [21]

By using Cloud Analyst, application developers are able to determine the best plan for allocation of resources among available data centres, strategy for selecting data centres to serve user requests, and costs related to applications. Since Cloud infrastructures are distributed, applications can be deployed in different geographic locations and the chosen distributions of the application impact its performance for users that are far from the data centre. Impact of number of simultaneous users, geographic location of relevant components and network in applications is hard to achieve in real environment. To allow control and repeatability of experiments, simulators are used. Cloud Analyst generates response time of requests, processing time of requests, and other metrics. It deploys different service brokering policies depending on the requirements; proximity based routing policy selects nearest data centre. [5]

Using Cloud Analyst simulation results can be produced in the form of tables and charts. The simulated output will have the response time of the simulated application, overall and average response time, virtual machine cost, data centre processing cost of all the user requests. The simulated output gives the mean time taken by each data centre to satisfy each user request, average, minimum and maximum request processing time by each data centre.

III. PROXIMITY AWARE CLOUD INFRASTRUCTURE SELECTION

When selecting the cloud infrastructures for an application, we need to consider the location of cloud data centres, the location of client applications, the relationship of the components of the application as well as the location of the related applications. If the distance between two communicating entities is less the network latency, bandwidth will be higher.

Cloud providers often deploy many data centers around the world which can greatly meet the requirements of distributed applications. Based on location of client requests providers can select cloud infrastructures that are close to their clients. The distances between the cloud infrastructures and the client are necessary once the location is obtained. When deploying the applications in the cloud platforms it is important to consider the proximity among the application components. The interaction among the components of user applications plays an important role in the Quality of Service. To minimize the network latency it is better to place all the components in the same location.

3.1 Parameters to control usage and optimize user workload performance

- Geographic locations where the user is located with respect to regions, zones, data centres
- Cost limitations based on per hour, on quotas, etc.
- Maximum latency that can be tolerated by user
- Virtual machine requirements such as CPU, memory for user applications
- Maximum provisioning time that is acceptable by the user
- Minimum SLA required for availability

3.2 Service Broker

A cloud service broker serves as a mediator between user and cloud service provider as depicted in figure 1. Service broker collects multiple cloud services, integrates with in-house applications and customize them to meet user needs. Gartner refers to these mediators as cloud service brokerages. Cloud service brokers use particular tools to recognize the cloud resource, and then map the requirements of a user application to it. Service brokers automatically route data, user applications and infrastructure needs based on criteria such as price, location, latency needs, SLA level, supporting operating systems, cloud scalability, backup, disaster recovery capabilities and regulatory requirements.

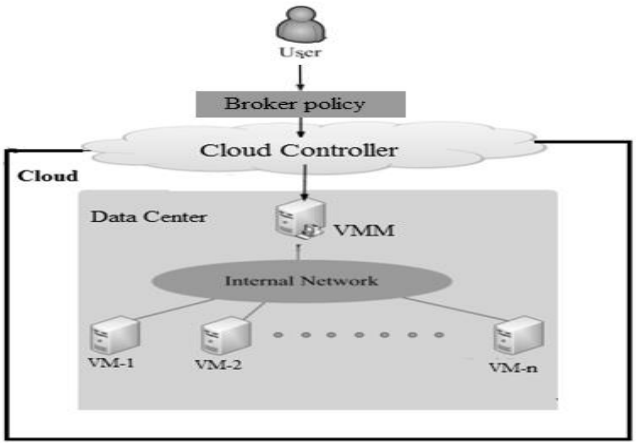

Figure 1. Cloud Service Broker

3.21 Service broker policies

In Cloud environment data centres provide services to users. In case of many data centres service broker policies are used to select best data centre which is beneficial to users and service providers. Service Broker in Cloud Analyst is responsible for routing the traffic from the user requests from the user base based on the service brokering policies as depicted in figure 2. Service Broker Policies controls the user base that is being serviced by a data centre at a given time.

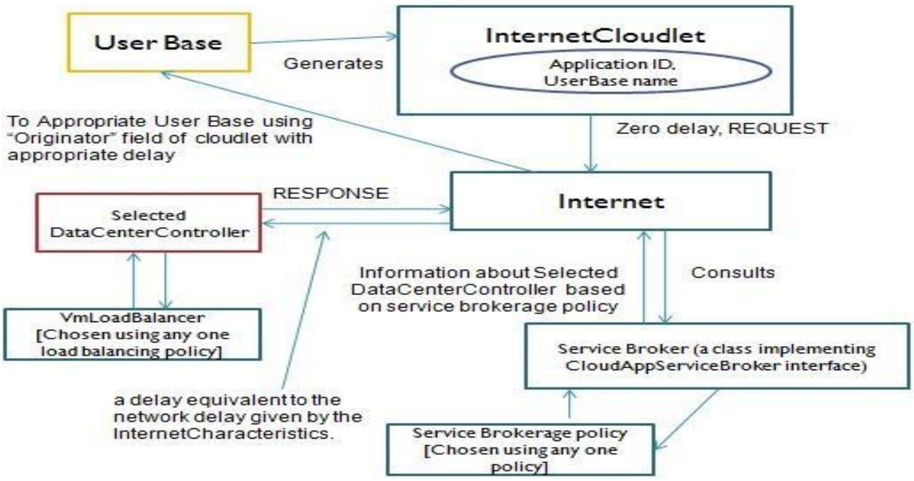

Figure 2. Routing of user request in Cloud Analyst tool

The main components of cloud Analyst are:

User Base - This component models a user base and generates traffic representing the users. The user base represents the single user but ideally a user base should be used to represent a large numbers of users for efficiency of simulation.

Internet - This component models the Internet and implements the traffic routing behaviour.

Internet Characteristics - This component maintains the characteristics of the Internet during the simulation, including the latencies and available bandwidths between regions, the current traffic levels, and current performance level information for the data centres.

VmLoad Balancer - This component models the load balance policy used by data centres when serving allocation requests.

Data Centre Controller - This component controls the data centre activities.

CloudAppService Broker- This component models the service brokers that handle traffic routing between user bases and data centres. The default traffic routing policy is routing traffic to the closest data centre in terms of network latency from the source user base.

Service broker policy - In the service broker policy the desired policy can be selected such as closest data centre, optimal response time and dynamic service broker. Six different regions can be chosen around the whole world with different data centres and users in different locations depending upon the condition. Number of concurrent user accessing the data centre and number of data centres simultaneously providing services to users can be selected. Cloud infrastructures are distributed; applications can be deployed in different geographic locations. The chosen distribution of the application impacts its performance for users that are far from the data centre. At data centre level various service broker algorithms choose which data enter is taken into consideration for the handing out of the incoming request.

Objectives of Service broker policy:

- Minimize the deployment cost.
- Improvement of performance data centers.
- Minimize the overall distances between user and data centres.

Three different broker algorithms in service broker policy are Service Proximity Based Routing, Performance Optimized Routing, and Dynamic Service Broker.

3.22 Service Proximity based Routing

When the Internet receives a request it queries the service proximity service broker for the destination data center. It maintains a listing all data centers indexed by their region. Algorithm firstly retrieves the region of the incoming request and queries for the region proximity list for that region. This list gives the remaining regions in the order of lowest network latency first when calculated from the given region. The algorithm picks the data center at the earliest/highest region in the proximity list. If more than one data center is in the same region, then random data center is selected.

3.23 Performance Optimized Routing

It is based on Response Time Service Broker which maintains a list of all data centers available. When the Internet receives a request, it queries the Best Response-Time Service Broker for the destination data center. Using the Service Proximity-Service Broker algorithm the closest (in terms of latency) data center is identified. Then same is repeated for the list of all data centers and the approximate current response time at each data center is found. If the least estimated response time is found in the closest data center, then the Service Broker selects the closest data center else, it picks either the closest data center or the data center with the least response time.

3.24 Dynamic Service Broker

Service Broker maintains a list of all data centers and another list with the best response time recorded so far for each data center. It extends either Service Proximity Service Broker or the Best Response Time Service Broker. When request arrives it queries the Dynamic Service Broker for the destination datacenter. The Dynamic Service Broker uses the Service Proximity Service Broker/ Best Response-Time Service Broker algorithm to identify the destination. The Dynamic Service Broker updates the best response time records if the current response time is better than previous.

The cloud analyst tool is used to evaluate the policies for the closest data center by using user base located in different regions.

IV. RELATED WORK

Based on existing three different broker algorithms, comprehensive technical studies and survey, we notice variations of service broker policy gives better performance than existing service proximity based policy. The author in [6] presents an enhanced proximity-based routing policy that avoids the direct selection of nearest data centre. If more than one data centre is in the same region, then the data centre having less cost will be selected. In [7], the author describes the cost-aware selection of data centre from the data centres within same region. In [8], the author presents proximity-based routing policy that avoids the overloading of the nearest data centre by routing the traffic to the neighbouring data centre in the same region with improved response time. The author in [9] describes the random data centre selection and virtual machine cost in each data centre is compared with other data centres in the same region. The data centre with lowest virtual machine cost is selected. The author in [10] describes policy that eliminates the sequential selection of inter region data centre with improvement in overall performance and the data centre with less number of users is selected when network latency is same for all data centres.

4.1 Comparison

Most data centre routing policy schemes are modified to user specific requirements. Each Service broker algorithms has its own pros and cons with different data centre processing cost as shown in figure 3. For example, reducing the service level agreement violation algorithm aims to reduce overloading of the closest data centre by redirecting a part/whole of the user requests to the next neighbouring data centre in the same region. While an enhanced proximity-based routing policy aims to Virtual Machine (VM) cost which avoids the direct selection of nearest data centre. The selection is not made randomly and VM cost in each data centre is compared with other data centres in the same region. Finally data centre with lowest VM cost is selected. Now the requests will be sent to this data centre to be processed. We identified some of the key research directions in data centre network routing policy schemes to specific requirements. Although current proposals improve cost, provide mechanisms for best response time, Minimum Virtual Machine cost there are challenging and important issues that are yet to be explored.

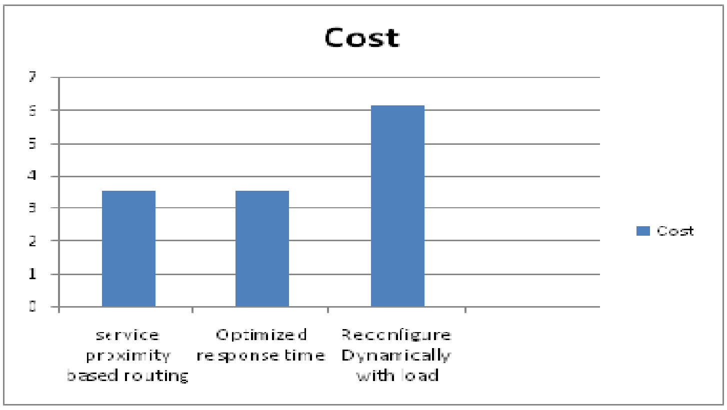

Figure 3. Cost comparison of service broker policies

V. CONCLUSIONS

The response time and data transfer cost is a challenge in cloud environment that can increase the business performance in the IT industry. In this paper, a survey is made on three service broker policies for user application request located in different regions, compared data centre processing cost. We discussed the proposed schemes of different service broker policies, highlighting the cost effective data centre, selection of data centre. From the previous work done, we can conclude that the simulation process can be still improved by modifying or adding new strategies for traffic routing, load balancing. In order to balance the user requests of the resources it is essential to recognize cost effectiveness to achieve an overall improvement in system performance at a reasonable cost.

VI. FUTURE WORKS

In our future work, we will use the simulation framework to develop and evaluate new service broker policy to improve the data centre performance. Furthermore, we will design routing policy that routes the user application for the next nearest data centre region considering peak time hours. Cost effectiveness, performance guarantees, devising effective business and pricing models are important directions for future research.

REFERENCES

[1] R. Buyya, C. S. Yeo, and S. Venugopal, “Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities”, Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC 2008, IEEE CS Press, Los Alamitos, CA, USA), Sept. 25-27,

[2] “Google App Engine,” http://code.google.com/appengine/

[3] R. Buyya, R. Ranjan, and R. N. Calheiros, “Modeling and Simulation of Scalable Cloud Computing Environments and the Cloud Sim Toolkit: Challenges and Opportunities,” Proc. of the 7th High Performance Computing and Simulation Conference (HPCS 09), IEEE Computer Society, June 2009

[4] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, and Rajkumar Buyya, The Cloud Computing and Distributed Systems (CLOUDS) Laboratory Department of Computer Science and Software Engineering The University of Melbourne, Australia. Cloud Analyst: A Cloud Sim-based Visual Modeler for Analysing Cloud Computing Environments and Applications.

[5] Wickremasinghe, Bhathiya. "Cloud Analyst: A Cloud Sim-based Tool for Modeling and Analysis of Large Scale Cloud Computing Environments." Project Report. 2009.

[6] Devyaniba Chudasama and Naimisha Trivedi, “Cost effective selection of Data center by Proximity-Based Routing Policy for Service Brokering in Cloud Environment” International Journal of Computer Technology & Applications,Vol 3 (6), 2057-2059.

[7] Dhaval Limbani and Bhavesh Oza, “A Proposed Service Broker Strategy in Cloud Analyst for Cost-Effective Data Center Selection” International Journal of Engineering Research and Applications, India, Vol. 2, Issue 1, Jan-Feb 2012, pp.793-797.

[8] Enhanced Proximity-Based Routing Policy for Service Brokering in Cloud Computing A. Sarfaraz Ahmed, A. Sarfaraz Ahmed / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622

[9] Dhaval Limbani and Bhavesh Oza, "A Proposed Service Broker Policy for Data Center Selection in Cloud Environment with Implementation" International Journal of Computer Technology & Applications,Vol 3 (3), 1082-1087.

[10] Vibhavari Patel, Prof. Nisha Shah , "A Proposed Service Broker Policy for Inter region Data Center Selection in Cloud Environment" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 3, Issue 4, Jul-Aug 2013, pp.1699-1702.

[11] Ashraf Zia, M.N.A. Khan, "A Scheme to Reduce Response Time in Cloud Computing Environment" I.J.Modern Education and Computer Science, 2013, 6, 56-61 Published Online July 2013 in MECS http://www.mecs-press.org/)DOI: 10.5815/ijmecs.2013.06.08.

[12] Y. C. Lee, A.Y. Zomaya and M. Yousif, "Reliable Workflow Execution in Distributed Systems for Cost Efficiency", 11th IEEE/ACM International Conference on Grid Computing, IEEE(2010). DOI:10.1109/GRID.2010.5697959.

[13] Y. Yuan and W. Liu, "Efficient resource management for cloud computing", 11th IEEE/ACM International Conference on System Science, Engineering Design and Manufacturing.

[14] M. Litoiu and M. Litoiu, "Optimizing Resources in Cloud, a SOA Governance View", GTIP, USA (Dec. 7, 2010). DOI:10.1145/1920320.1920330.

[15] M. Alhamad, T. Dillon, C. Wu and E. Chang,"Response Time for Cloud Computing Providers",WAS2010, France, (8- 10 November, 2010).DOI:10.1145/1967486.1967579.

[16] Marcos Dias de Assunção, Alexandre di Costanzo and Rajkumar Buyya, "Evaluating the Cost-Benefit of using Cloud Computing to Extend the Capacity of Clusters", HPDC'09, Germany, (June 11–13, 2009). DOI:10.1145/1551609.1551635.

[17] V. Sekar and P. Maniatis, "Verifiable Resource Accounting for Cloud Computing Services", CCSW'11, USA, (October 21, 2011).DOI:10.1145/2046660.2046666.

[18] R.K. Jha and U. D. Dalal, "A performance comparison with cost for QoS application in on demand cloud computing", International Conference on Recent Advances in Intelligent Computational Systems (RAICS), IEEE, (2011).DOI:10.1109/RAICS.2011.6069264.

[19] D. Bein, W. Bein and S. Phoha, "Efficient datacenters, cloud computing in the future of distributed computing", Seventh International Conference on Information Technology, IEEE, (2010). DOI:10.1109/ITNG.2010.31.

[20] Deepali Koul and chaitali patel , "Analysis of Various Energy Efficient DataCenters Using Cloud Systems",International Journal of Computer Science and Mobile Applications,Vol.1 Issue. 5, November-2013, pg. 57-64 ISSN: 2321-8363.

[21] Manoranjan Dash,, Amitav Mahapatra & Narayan Ranjan Chakraborty, "Cost Effective Selection of Data Center in Cloud Environment", International Journal on Advanced Computer Theory and Engineering (IJACTE), Volume-2, Issue-1, 2013, ISSN (Print) : 2319 – 2526.

AUTHORS

Rekha PM holds M.Tech degree in Computer Science and engineering. Working as Assistant professor in the Department Information Science and Engineering, JSS College of Technical Education, and Bangalore, India. She has a decade experience in teaching field. Her research interests include computer networks, Mobile computing.

.

M Dakshayini holds M.Tech (VTU Belgaum) and PhD in computer science. She has two decades experience in teaching field. She has published many papers. Currently she is working as Professor in the department of Information science and engineering at BMS College of Engineering, Bangalore, India