Comparison of Dynamic Load Balancing Policies in Data Centers

Sunil Kumar
Department of Computer Science, Faculty of Science, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, India.

Manish Kumar Pandey
Department of Computer Science, Faculty of Science, Banaras Hindu University, Varanasi-221005, Uttar Pradesh, India.

ABSTRACT
Cloud computing is an emerging advanced technology which provides the computing facilities through the Internet to end users by supplying an on demand basis as per usage like water, electricity etc. This could be considered as 5th utility of human need in this new cloud era. Earlier computing technique facilities were developed by the developer at the organization and purchased by users and deployed as per the compatibility of applications in available infrastructure which usually is not sufficient, e.g. it was not able to handle the demand at peak traffic period. Again, servers were not fully utilized as peak traffic only happens in some period of time. All big organizations are paying their attention to utilize their server/infrastructures capability through cloud implementation. This technology can sort out the problem at both levels, i.e. at service provider level as well as at user level by facilitating more infrastructures and less infrastructure respectively by providing the cloud-service-provider (CSP) on pay-as-you-go (PAYG) model. Next problem is cost as per requirement of Virtual Machines (VMs) on Data Center and the response time. This paper investigates the optimized synchronization between DCs (Data Centers) and UBs (User Bases) for enhancing the application performance and response time for the same cost to the vendors and end users by using a tool called CloudAnalyst. The reliability of cloud computing is maintained through load balancing of VMs on a Data Center. Load balancing technique improves the performance of VMs, reduces overall response time, processing time and cost of VMs. In this paper a comparative study was performed among available service broker policies (Closest Data Center, Optimize Response Time and Reconfigure Dynamically with Load) and available load balancing algorithms (Round Robin, Equally Spread Current Execution Load, Throttled etc.). The objective of this study is to analyze how these policies help to coordinate between Data Centers to optimize the applications performance and the cost to the user.

Keywords
Internet applications, CSP, Load-balancing, Service broker, Reliability

1. INTRODUCTION
In 1969, Leonard Kleinrock, one of the chief scientists of the original Advanced Research Projects Agency Network (ARPANET) which seeded the Internet, said: "As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, I will probably see the spread of "computer utilities" which, like present electric and telephone utilities, will service individual homes and offices across the country [1][2]." The cloud computing was evolved from distributed computing, cluster computing, grid computing, utility computing and software [3]. In March 2006 various applications were introduced such as Amazon S3, Sales Force and AppExchange. In August 2006 (Amazon EC2), May 2007 (Face book Platform), April 2008 (Google App Engine) and October 2008 (Microsoft Azure Platform) were launched in market. These are the most popular examples of cloud computing. Main technologies behind cloud computing are SOA, Virtualization and Web 3.0 which were basically introduced towards the end of year 2007.

Cloud Computing has attracted a lot of attention in recent times. In May 2008, Merrill Lynch [4] estimated the cost advantages of Cloud Computing to be three to five times for business applications and more than five times for consumer applications. According to a Gartner press [4] release from June 2008, Cloud Computing will be "no less influential than e-business". The positive attitude towards the importance and influence of Cloud Computing resulted in optimistic Cloud-related market forecasts. In October 2008, IDC [5][6] forecasted an almost threefold growth of spending on Cloud services until 2012, reaching $42 billion. The current trends of cloud computing research is going towards energy efficient green IT or computing [7] which will ultimately reduce the overall cost for the end users defined by same firm IDC.

Buyya et al. [7] have defined **"pay as you use"** model for cloud computing as: Cloud computing is a type of parallel and distributed computing system. It is an extension of grid and utility computing that are collection of federated networks of multiple clouds and virtualized computers which are dynamically provisioned and integrated computing resources based on service-level-agreements (SLAs) established through agreements between the service vendors and end users.

IT resources like hardware, software and services that are abstracted from the underlying infrastructure or fabric layer of hardware and provided "on-demand" basis and "at-scale-in or out" in a multitenant environment based on Service level Agreement (SLA)– ***CISCO*** [8].

Cloud computing is a platform for enabling ubiquitous, convenient, on-demand network access to a shared group of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be swiftly provisioned and unconstrained with negligible management endeavor or service provider interaction. This cloud model is a collection of five essential characteristics (such as virtualization, scalability, on demand services, abstraction and Service Level Agreements)**,** three service models (such as SaaS, PaaS and IaaS), and four deployment models (such as private, public, hybrid and federated) – ***NIST*** [9].

The service scheduling algorithm establishes to handle some fairness constraints. The first constraint is to categorize user applications by QoS preferences parameters such as execution time, size and cost of the expected applications on the internet, and establish the general functions in accordance with the categorization of tasks to contain the fairness of the resources in matching process. The second constraint is to define fair resource providence justified with some functions of the resource allocation.

There are a number of simulation tools available in the market such as CloudSim [10], GridSim [11], CloudReports [12] and CloudAnalyst [13]. Cloudsim and Gridsim are programming based simulation tools. One can make changes in these and get outputs of simulated values such as no. of VMs, DCs, PEs, MIPSs, BWs, etc. but due to the programming complexities it is hard to implement. In CloudLabs, Melbourne, Australia GUI based simulation tools CloudReports and CloudAnalyst were developed. CloudAnalyst is used in the current study to simulate the internet based incoming cloudlets and applications flow.

CloudAnalyst tool have three types of service broker policies namely Closest Data Center, Optimize Response Time and Configure Dynamically with Load. In the advanced configuration, one can select any one of the given three Load Balancing Policies namely Round Robin (RR), Equally Spread Current Execution Load (ESCEL) and Throttled. The present work compares the load balancing policy in different service broker policy on one or more Data Centers with various configurations. The aim of this comparative study is to select the optimal service broker policy robustly tested on different parameters. The impact of this work is in selection of the best policy among them and finding out one optimal algorithm which is most profitable to the user as well as provider.

2. ARCHITECTURE OF CLOUD ANALYST

CloudAnalyst [13] a cloud simulation tool which takes apart the simulation, experimental set up exercise from a programming exercise and enables a modeler to concentrate on the simulation parameters rather than the technicalities of programming. It also provides the facility of repeated and quick experiments of simulations with different parameter values in a very easy way. The advantage of the CloudAnalyst is the graphical output of the simulations which can be easily and efficiently analyzed by the users. It also helps to detect any problem in simulation logic with the performance and accuracy.

CloudAnalyst is a technique which extends its environment to study the behavior of large scaled Internet applications such as Facebook, Google Scholar, and other social sites in a cloud computing environment. This simulation tool can also be extended with some novel approach of load balancing algorithms which could be tried to improve the behavior of simulation for large scaled. The CloudAnalyst is built by using the framework of CloudSim with some additional features. [13]

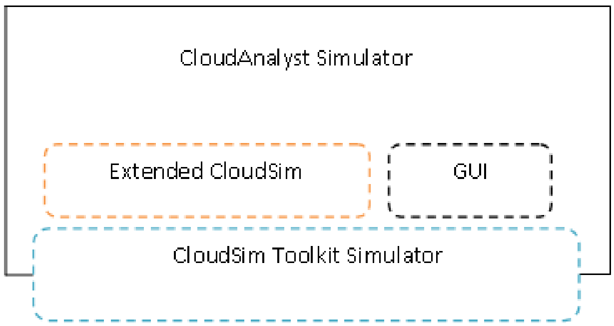


Fig 1: CloudAnalyst built on top of CloudSim Toolkits

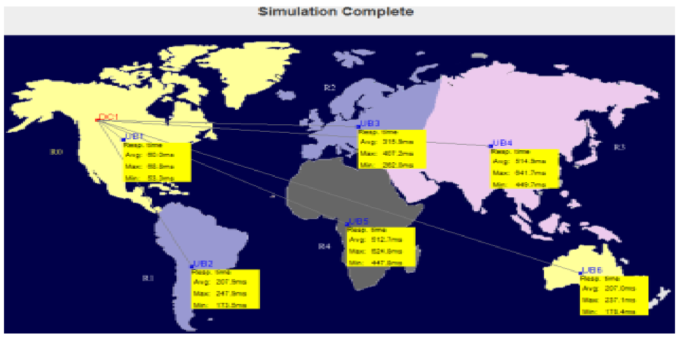


Fig 2: Graphical Output of Simulation with Colored Regions

It produces output as Response time of the simulated application, the usage patterns of the application, the time taken by data centers to service a user request and the cost of operation.

The CloudAnalyst is fully developed in Java language, using Java SE 1.6. Its GUI component is built using Java Swing components and CloudSim is the background used for modeling data centers in CloudAnalyst. It is developed on the platform of CloudSim and CloudSim is developed on the platform of SimJava which is the core part of underlying simulation framework of CloudSim and responsible for managing simulation clock and handle basic discrete event simulation of cloud environments. So some features of SimJava are directly used in CloudAnalyst. It is an open source cloud simulation tool developed in year 2009. Its VM allocation policy works on Time-Shared basis and useful for heterogeneous environments.

2.1 Features of the Simulator

- Ease of use
- Simulations can be run with high configurationally flexibility
- Graphical output
- Repeatability
- Ease of extension

3. RELATED WORKS

Bhathiya et al.[13] proposed/described the CloudAnalyst and developed this tool to simulate the best / optimal selection of cloud provider in the respect of / term of overall response time, Data Center processing time and costs. There have been many studies using simulation techniques to calculate the performance of large scale distributed system. These simulations are GridSim[13], MicroGrid[13], GangSim[13], SimGrid[13] and CloudSim[13] based on heavy java programming and provides the environment to develop the cloud, design cloud and deploy cloud on its. But CloudAnalyst is also a simulation tool to simulate the performance and cost on the basis of basically six divided regions all over the world. Because, Data Centers are located in the any of regions and cost paying scheme also varies from vendor to vendor. Thus one should decide that which of the scheme is profitable for the application deployment and accessing the cloud. This facility is provided by CloudAnalyst. [13]

Brototi et al. [14] proposed the new load balancing technique in cloud computing using Stochastic Hill Climbing approach and compared with the given load balancing policy in CloudAnalyst. A comparative study is also done with Round Robin (RR) algorithm and First Come First Serve (FCFS) and results are found to be encouraging.

In this paper, a comparison was done among given load balancing policy and service broker policy with different configuration parameters. The analysis of the simulation result can help to infer the best approach to calculate and select profitable vendors/ providers and in which regions one should select to deploy for applications and data.

4. EXPERIMENTAL SETUP

This section presents the experiments and evaluations of three dynamic load-balancing policies for fixed set of parameters in CloudAnalyst simulation tool. These experiments were conducted on a Intel (R) Core (TM) i5-3210M CPU having configurations like : 2.50 GHz processor speed, 4 GB RAM, 64-bit OS Windows 7 Ultimate, JDK 1.7 and 500 GB HDD. Simulation was set with One Data Center with two physical hardware units and six User Bases (UBs) with six different regions. VM policy is selected as TIME_SHARED with one Data Center having five Virtual Machines. Simulation duration was set to 60.0 in minutes (we can select hours and days also). Advanced simulation configuration chosen as 100,000 users grouping factor in UBs and 100,000 request grouping factor per Data Center. The power of DCs set as executable instructions length 1000 Bytes per request. Internet characteristics configures as well as defined in the simulation tool, there are specified Delay Matrix and Bandwidth Matrix Region by Region for six Regions.

Table 1. Simulation Result in Respect of Service Broker Policy: *Closest Data Center*

Load Balancing Policy Across VM's in a Single Data Centers		**Average (ms.)**	**Minimum (ms.)**	**Maximum (ms.)**
Round Robin (RR)	Overall Response Time	303.15	52.36	672.62
	Data Center Processing Time	10.23	7.21	13.26
Equally Spread Current Execution Load (ESCEL)	Overall Response Time	303.34	52.36	672.62
	Data Center Processing Time	10.23	7.21	13.26
Throttled	Overall Response Time	303.39	52.36	672.62
	Data Center Processing Time	10.23	7.21	13.26
Data Center	VM Cost $	Data Transfer Cost $		Total Cost $
DC1	0.50	0.35		0.85

Table 2. Simulation Result in Respect of Service Broker Policy: *Optimize Response Time*

Load Balancing Policy Across VM's in a Single Data Centers		**Average (ms.)**	**Minimum (ms.)**	**Maximum (ms.)**
Round Robin (RR)	Overall Response Time	303.36	53.26	641.72
	Data Center Processing Time	10.25	7.21	13.51
Equally Spread Current Execution Load (ESCEL)	Overall Response Time	303.38	53.26	641.72
	Data Center Processing Time	10.25	7.21	13.51
Throttled	Overall Response Time	303.33	53.26	641.72
	Data Center Processing Time	10.25	7.21	13.51
Data Center	VM Cost $	Data Transfer Cost $		Total Cost $
DC1	0.50	0.35		0.85

Table 3. Simulation Result in Respect of Service Broker Policy: Reconfigure Dynamically with Load

Load Balancing Policy Across VM's in a Single Data Centers		**Average (ms.)**	**Minimum (ms.)**	**Maximum (ms.)**
Round Robin (RR)	Overall Response Time	333.80	56.26	672.62
	Data Center Processing Time	40.90	8.41	136.51
Equally Spread Current Execution Load (ESCEL)	Overall Response Time	333.30	56.26	672.62
	Data Center Processing Time	40.53	8.41	89.13
Throttled	Overall Response Time	333.46	56.26	672.62
	Data Center Processing Time	40.53	8.41	89.13
Data Center	VM Cost $	Data Transfer Cost $		Total Cost $
DC1	3.16	0.35		3.51

5. RESULTS AND DISCUSSIONS

Case 1:

After fixing simulation configuration by choosing service broker policy as "Closest Data Center", the "Round Robin" load balancing policy is producing best result. In this service broker policy every load balancing policy produces the same result of Data Center Processing Time (Average/Minimum/Maximum) and Costs (VM/Data Transfer). Overall Response Time is also same in each load balancing policy for the minimum and maximum conditions. Only average overall response time varies in different load balancing policy. Overall cost is also same in each load balancing policy. VM costs and Data Transfer Costs are also similar in each case of simulation configuration of different load balancing policy.

After simulation, one can see that Round Robin load policy is producing best result in average case of overall Response Time. If no. of DCs, no. of VMs, etc. are increasing then its overall response time, data center processing time and costs also increases.

Case 2:

When selected service broker policy is "Optimize Response Time", it is producing same results for overall response time (minimum/maximum) as well as for Data Center processing time (average/minimum/maximum). It is generating same costs of the VM and data transfer for all load balancing policies.

The only difference which arises is of average simulation time of overall response time in case of different load balancing policies such as: Round Robin, ESCEL and Throttled. On making changes in no. of DCs, no. of VMs and simulation period such as minutes, hours and days different results will be produced in same pattern as is mentioned above.

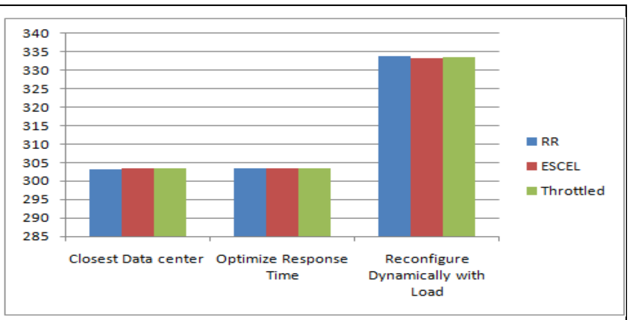
Case 3:

When selected service broker policy is "Reconfigure Dynamically with Load", the Overall response time (minimum, maximum), minimum Data Center processing time and costs (VM, Data Transfer) produced the same result in all load policy.

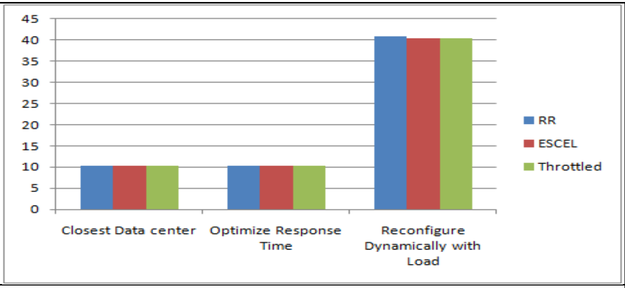
But slight difference is between/among the overall response time (average) for different given load policy. Average / Maximum Data Center processing time are same for the ESCEL and Throttled.

6. COMPARISON

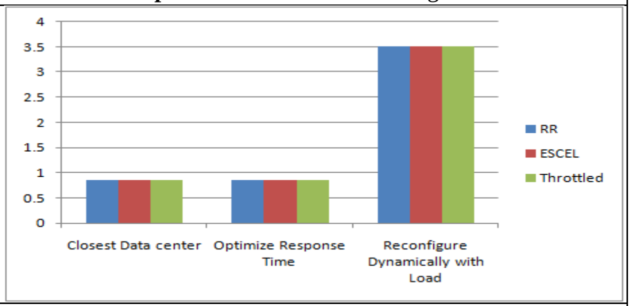
After various rounds of simulation it can be summarized that the best service broker policy and the best load balancing policy are Optimize Response Time and Round Robin as they are giving the best result for simulating the internet coming requests/cloudlets/services/applications and user requests on the social sites.



Graph 1: Overall Response Time



Graph 2: Data Center Processing Time



Graph 3: Total Costs (VM + Data Transfer)

7. CONCLUSION AND FUTURE SCOPE

In this paper, a comparison of the given policies of load balancing at the Data Center level on VMs was performed. By enhancing the given scheduling algorithms and then implementing those in cloud computing environment using CloudSim toolkit, one can get better results. By visualizing the cited parameters in graphs and resulted tables it can be easily identify that which one is the best policy to produce improved overall response time and data centre processing time. The future work includes overcoming the problem of deadlocks and server overflow. New service broker policy in the simulator can also be implemented. We will try to extend this CloudAnalyst tool with some new added parameters and include some novel policy of load balancing algorithms to optimize the results.

8. ACKNOWLEDGEMENT

I [first Author] am very much thankful to Prof. Umesh Singh, Coordinator, DST-CIMS, Faculty of Science, Banaras Hindu University, Varanasi, India who has extended all financial as well as supportive guidance for carrying out this research work.

9. REFERENCES

[1] L. Kleinrock, A Vision For The Internet St Journal Of Research, Nov, 2005, (1), Pg4-5.

[2] http://www.livinginternet.com/w/wi_online.htm.

[3] R.Buyya, C. Yeo, S.Venugopal, J.Broberg, I.Brandic, Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility, in: Future Generation Computer Systems, vo1.25, 2009, pp. 599–616.

[4] Jane Anderson, Assessing Cloud Computing Challenges and Opportunities for Network Providers, strategic white paper, May 2008.

[5] Frank Gens, IT Cloud Services Forecast – 2008, 2012: A Key Driver of New Growth on October 8th, 2008.

[6] Frank Gens, Enterprise IT in the Cloud Computing Era New IT Models for Business Growth & Innovation, IDC, 2008.

[7] R. R.Buyya, R.Ranjan, Intercloud: Utility-oriented federation of cloud computing environments for scaling of application services, in: ICA3PP 2010, Part I, LNCS 6081, 2010, pp. 13–31.

[8] Kapil Bakshi, Cisco Cloud Computing -Data Center Strategy, Architecture, and Solutions , Point of View White Paper for U.S. Public Sector 1st Edition 2009 Cisco Systems, Inc.

[9] Peter Mell, Timothy Grance, The NIST Definition of Cloud Computing, NIST Special Publication, 800- 145, September 2011.

[10] R.N.Calheiros, R. Ranjan, A. Beloglazov, C. Rose, R. Buyya, Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms, in: Software: Practice and Experience (SPE), Volume 41, Number 1, ISSN: 0038-0644, Wiley Press, New York, USA., 2011, pp. 23–50.

[11] Anthony Sulistio, Uros Cibej, Srikumar Venugopal,Borut Robic and Rajkumar Buyya, A toolkit for modelling and simulating Data Grids: An extension to GridSim, Concurrency And Computation: Practice And Experience, 0123; 34:1, Version: 2002/09/19 v2.02.

[12] B.Wickremasinghe, R.N.Calheiros, R. Buyya, Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications, in: Proceedings of the 24th International Conference on Advanced Information Networking and Applications (AINA 2010), Perth, Australia, 2010.

[13] Bhathiya Wickremasinghe, “CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments” MEDC Project Report, 2009, 44 p.

[14] Brototi Mondal, Kousik Dasgupta, Paramartha Dutta, Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach. Procedia Technology, vol. 4, pp.783-789, 2012.