International Journal of Computer and Communication Technology

Volume 6 | Issue 1 Article 15

January 2015

MINIMIZATION OF LOAD BASED RESOURCES IN CLOUD COMPUTING SYSTEMS

ZULFIKHAR AHMAD
College of Engineering and Technology, BBSR, zulfikharahmad@gmail.com

ASHIS KU. MISHRA
College of Engineering and Technology, BBSR, ashishkumishra@gmail.com

BIKASH CHANDRA ROUT
IIMT, BBSR, bikashchandrarout@gmail.com

B. BIKRAM KUMAR
College of Engineering and Technology, BBSR, bbikramkumar@gmail.com

Recommended Citation
AHMAD, ZULFIKHAR; MISHRA, ASHIS KU.; ROUT, BIKASH CHANDRA; and KUMAR, B. BIKRAM (2015) "MINIMIZATION OF LOAD BASED RESOURCES IN CLOUD COMPUTING SYSTEMS," *International Journal of Computer and Communication Technology*: Vol. 6 : Iss. 1 , Article 15.
DOI: 10.47893/IJCCT.2015.1277
Available at: https://www.interscience.in/ijcct/vol6/iss1/15

MINIMIZATION OF LOAD BASED RESOURCES IN CLOUD COMPUTING SYSTEMS

[1]ZULFIKHAR AHMAD, [2]ASHIS KU. MISHRA , [3]BIKASH CHANDRA ROUT & [4]B. BIKRAM KUMAR

[1, 2, & 4]College of Engineering and Technology, BBSR
[3] IIMT, BBSR

Abstract-“Cloud computing” is a term, which involves virtualization, distributed computing, networking, software and Web services. Our Objective is to develop an effective load balancing algorithm using Divisible Load Scheduling Theorem to maximize or minimize different performance parameters (throughput, latency for example) for the clouds of different sizes. Central to these issues lays the establishment of an efficient load balancing algorithm. The load can be CPU load, memory capacity, delay or network load. Load balancing is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work. Load balancing ensures that all processor in the system or every node in the network does approximately the equal amount of work at any instant of time.

1. INTRODUCTION

Cloud Computing is an on Demand service in which shared resources, information, software and other devices are provided according to the clients to the requirements at specific time. It’s a term which is generally used in case of Internet. The whole Internet can be viewed as a cloud. Load balancing in cloud computing systems is really a challenge now. Always a distributed solution is required. Because it is not always practically feasible or the most efficient to maintain one or more idle services just as to fulfill the required demands. Jobs can’t be assigned to appropriate servers and clients individually for efficient load balancing as cloud is a very complex structure and components are present throughout a widespread area.

2. CLOUD COMPUTING

2.1 Cloud Computing Architecture

The architecture behind Cloud Computing is a massive network of “Cloud Servers” interconnected as if in a grid running in parallel, sometimes using the technique of virtualization to maximize computing power per server. Front-end interface allows a user to select from a catalog. This request gets passed to the system management which finds the correct resources, and then calls the provisioning services which carves out resources in the Cloud. The provisioning service may deploy the requested stack or web application as well.

User interaction interface: This is how users of the cloud interface with the cloud to request services.

Services Catalog: This is the list of services that a user can request.
System Management: This is the piece which manages the computer resources available.

Provisioning tool: This tool carves out the systems from the cloud to deliver on the requested service. It may also deploy the required images.

Monitoring and metering: This optional piece tracks the usage of the cloud so the resources used can be attributed to a certain user.

Servers: The servers are managed by the system management tool. They can be either virtual or real.

3. REVOLUTIONARY PHASE OF CLOUD COMPUTING:

Two key enabling technologies based on the practices in the areas of service provisioning and solution design would play a very significant role in the revolutionary phase of cloud computing:
Virtualization technique: This technology works on the handling of how the image of the operating system, middleware, and application procreated and allocated to a physical machine or part of the server stack away. This technology also provides assistance in reuse licenses of operating systems, middleware, or software applications, as soon as the user releases their service from the Cloud Computing platform.

Service Oriented Architecture (SOA): SOA is that software which assists in addressing multi-component, reusability, extensibility and flexibility.

4. SOA ARCHITECTURE

5.

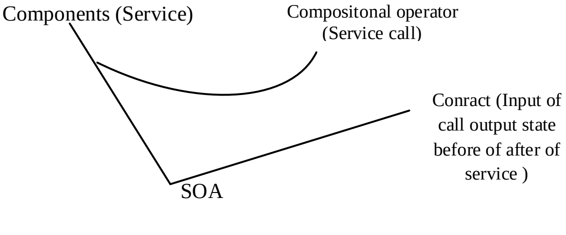


Fig-1

Types of Clouds

Based on the domain or environment in which clouds are used, clouds can be divided into 3 categories:

-Public Clouds also referred to as 'External Cloud' describes the conventional mean of cloud computing: scalable , dynamically provisioned, often virtualized resources available over the internet from the offsite third party provider, Which divides up resources and bills its customer on a 'utility' basis

-Private Clouds also referred to as 'corporate' or 'internal' cloud is a term used to denote a proprietary computing architecture providing hosted services on private networks. This type of Cloud computing is generally used by large companies, and allows their corporate network and data center administrators to effectively become in-house 'service providers' catering to 'customers' within the corporation. However, it negates many of the benefits of Cloud computing, as organization still need to purchase, set up and manage their own clouds.

-Hybrid Clouds environment combining resources from both internal and external providers that will become the most popular choice for enterprises. For example, a company could choose to use a public cloud service for general computing, but store its business critical data within its own data centre.

Services Provided by Cloud Computing

Service means different types of applications provided by different servers across the Cloud. It is generally given as "as a service". Services in a cloud are of 3 types as given in:

Software as a service (SaaS)- A SaaS provider gives subscribers access to both resources and application. Saas makes it unnecessary for you to have a physical copy of software to install on your devices.Saas also makes it easier to have the same software on all of your devices at once by accessing it on the cloud.In a Saas agreement, you have the least control over the cloud.

Platform as a service (PaaS)- A PaaS system goes a level above the software as a service setup. A PaaS Provider gives subscribers access to the components that they require to develop and operate applications over the internet.

Infrastructure as a service(IaaS) or Hardware as a service(HaaS)-An IaaS agreement, as the name states, deals primarily with computational infrastructure. In an IaaS agreement, the subscriber completely outsources the storage and resources, such as hardware and software, that they need.

4. LOAD BALANCING

It is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. A load balancing algorithm which is dynamic in nature does not consider the previous state or behavior of the system, that is. It depends on the present behavior of the system. The important things to consider while developing such algorithm are: estimation of load comparison of load. Stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones. Interaction among components of a dynamic load balancing algorithm

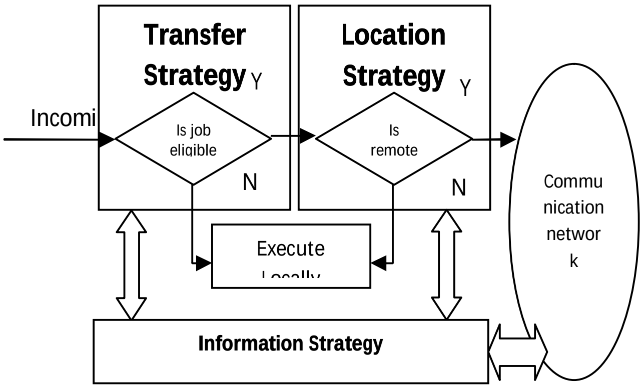


Fig-2

5. DISTRIBUTED LOAD BALANCING FOR THE CLOUDS

We are discussing three types of solutions which can be applied to a distributed system:

Honeybee Foraging Algorithm. Biased Random Sampling on a random walk procedure and Active Clustering.

Honeybee Foraging Algorithm

```
[A] Initialization-s in VI serving Q. Revenue rate interval Tr
Advert: posting prob P, reading prob n, read interval T,
[B] forever
[c] While Q, Not empty
        Do serve request; // service queue
If tp expired Then compute revenue rate;Adjust n from lookup table;
If Flip (p) ==TRUE
Then post Advert;
If T,expired &&==TRUE
Then
If forager
Then Select/ Read Advert ID Vi// Randomly Select
Else virtual Server id Vi // Randomly Select
If Vi Not. Eq V
Then Switch (Vi) // migrate to virtual server Vi
End while
End forever
```

Biased Random Sampling

Here a virtual graph is constructed, with the connectivity of each node (a server is treated as a node) representing the load on the server. Each

Server is symbolized as a node. In the graph, with each in degree directed to the free resources of the server.

Active Clustering

Active Clustering works on the principle of grouping similar nodes together and working on these groups. The process involved is:

“ A node initiates the process and selects another node called the matchmaker node

“ the So called matchmaker node then forms a connection between a neighbored it which is of the same type as the initial node.

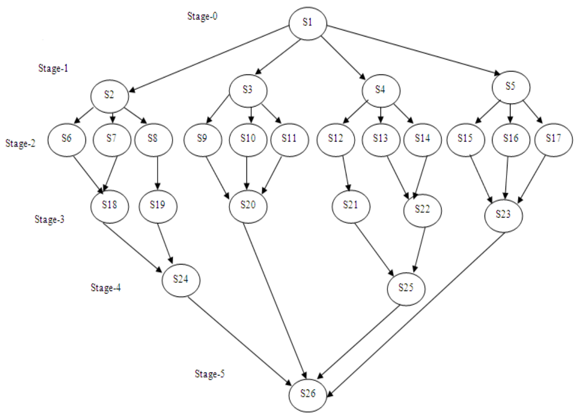
“ The matchmaker node then detaches the connection between itself and the initial node.

The above set of process is followed iteratively.

6. PROPOSED WORK

The time required for completing a task within one process is very high. So the task is divided into number of sub-tasks and each sub-task is given one job. Let the task S is divided into number of sub-tasks S1, S2, S3,.Sn, out of these some are executed sequentially and some are executed parallel. So the total time period for completing the task decreases and hence the performance increases. These sub-tasks can be represented in a graph structure known as state diagram. An example is given below:

The distributed network may follow different topologies. The tasks are divided over the whole network. One topological network connects with the other through a gateway.



7. DIVISIBLE LOAD SCHEDULING THEORY IN CLOUDS

Divisible load scheduling theory (DLT) in case of clouds is an optimal division of loads among a number of master computers, slave computers and their communication links. Our objective is to obtain a minimal partition of the processing load of a cloud connected via different communication links such that the entire load can be distributed and processed in the shortest possible amount of time.

System Model

The Cloud that we have considered here is a single level tree (star) topology consisting of K no. of master computers and each communicating N no. of slave computers K no. of master computers each joining N no. of slave computers in single level tree network(STAR Topology).

Measurement and Reporting Time

Parameters, Notation and Definitions:

?ki The fraction of load that is assigned to a slave i by master K. aki A constant that is inversely proportional to the measuring speed of slave i in the cloud. Bki A constant that is inversely proportional to the communication speed of link i in the cloud.

Tms Measurement intensity constant. This is the time it takes the ith slave to measure the entire load when aki=1. The entire assigned measurement load can be measured of ith slave in time aki Tms. Tcm Communication intensity constant. This is the time it takes to transmit the entire measurement load over a link when bki=1. The entire load can be transmitted over the ith link in time bki Tcm. Tki The total time that elapses between the beginning of the scheduling process at t=0 and the time when slaver i completes its reporting to the master k, i=0,1,2……N. This includes, in addition to measurement time, reporting time and idle time. Tfk This is the time when the last slave of the master k finishes reporting (Finish time or make-span) Tfk=max(Tk1;Tk2;Tk3…………Tkn)

Tf This is the time when the last master receives the measurement from its last slave. Tf = max(Tf1;Tf2;Tf3…….Tfn) When Measurement starts simultaneously and reporting is done sequentially:

Initially when time t=0, all the slaves are idle and the master computers start to communicate with the first slave of the corresponding slaves in the cloud. By time t=t1, each slave will receive its instructions for measurement form the corresponding master. It is assumed that after measurements are made, only one slave will report back to the root master at a time(or we can say that only a single link exists between them).The slaves here receive a fraction of load from their corresponding master sequentially and the computation will start after each slave completely receives its load share. The minimum measuring and reporting time of the network will then be given as

Tf1=t1 + (a11 Tms + b11 Tcm)

K(1+?Ni=2 ?ij=2 Sij)

Similarly we can obtain the generalized equation for master computer r as

Tfr = t1 + (ar1 Tms+ br1 Tcm)

K(1 + ?Ni=2 ?ij=2 Sri)

The minimum measuring and reporting time of the homogeneous network will then be given as where i=1, 2, 3………n

Tf1 = t1 + (a1 Tms + b1 Tcm)(1-S1)

K(1-S1N)

Where the Measurement starts and Reporting ends simultaneously:

Here each of N slave computers corresponding to a master computer in the cloud finishes reporting at the same time. The cloud will have the same report finishing time for each slave corresponding to the master. That is each slave has a separate channel to its master as shown in the timing diagram of the network. In this case the slaves receive their share of load form the master concurrently and start computation after completely receiving their share of load. Each slave begins to measure its share of the load at the moment when all finishes receiving their measurement instructions from the corresponding master.

The minimum measurement and reporting time of the network will then be given as

Tf1 = T11 = t1 + (a11 Tms + b11 Tcm)1/r11
K (? Ni= 1 (1/r1i)

Similarly for the master computer p, the generalized equation will be

Tfp = T1p=t1 + (ap1Tms + bp1Tcm) 1/rp1
K(?Ni= 1(1/rpi)

For the case of a homogeneous network, each slave corresponding to a master in the network shares the load equally. That is, ?1i=1/(KN), for i=1,2,3,…………..N. So, the minimum measuring and reporting time of the network will be

Tf1 = t1 + ((a1Tms + b1Tcm)/KN)

Similarly, we can obtain the above expression for rest of the master computers.

8. CONCLUSION AND FUTURE WORK

In this topic we have discussed on basic concepts of Cloud Computing and Load balancing and studied some existing load balancing algorithms, which can be applied to clouds. In addition to that, the closed-form solutions for minimum measurement and reporting time for single level tree networks with different load balancing strategies were also studied. The performance of these strategies with respect to the timing and effect of link and measurement speed was studied. Cloud computing is a vast concept and load balancing plays a very important role in case of clouds. There is a huge scope of improvement in this area. We have discussed only two divisible load scheduling algorithms that can be applied to clouds, but there are still other approaches that can be applied to balance the loads in clouds. The performance of the given algorithms can also be increased by varying different parameters.

REFERENCES

[1]. "Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, TATA McGRAW-HILL Edition 2010.

[2]. "Martin Randles, David Lamb, A. Taleb-Bendiab, A Comparative Study into Distributed Load balancing Algorithms for Cloud Computing 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.

[3]. "Ali M.Alakeel , A guide to Dynamic Load Balancing in Distributed Computer Systems, IJCSNS International Journal of Computer Science and Network Security, VOL. 10 No.6, June 2010.

[4]. "Martin Randles, Enas Odat, David Lamb, Osama Abu-Rahmeh and A.Taleb-Bendiab," A Comparative Experiment in Load Balancing",2009 Second International Conference on Developments in systems Engineering ."http://www-3.ibm.com/press/us/en/pressrelease/22613.wss"

[5]. Lewis, Grace. *Cloud Computing: Finding the Silver Lining, Not the Silver Bullet*. http://www.sei.cmu.edu/newsitems/ cloudcomputing.cfm (2009).

[6]. Dormann, Will & Rafail, Jason. *Securing Your Web Browser*. http://www.cert.org/tech_tips/securing_browser/ (2006).

[7]. Jansen, Wayne & Grance, Timothy. *Guidelines on Security and Privacy in Public Cloud Computing*. National Institute of Standards and Technology, 2011.

[8]. Strowd, Harrison & Lewis, Grace. *T-Check in System-of-Systems Technologies: Cloud Computing* (CMU/SEI-2010-TN-009). Software Engineering Institute, Carnegie Mellon University, 2010. http://www.sei.cmu.edu/library/abstracts/reports/10tn009.cfm

[9]. Lewis, Grace. *Basics About Cloud Computing*. http://www.sei.cmu.edu/library/abstracts/whitepapers/cloudcomputingbasics.cfm (2010).