Dynamic Resource Allocation Using Latent Semantic Ranking in Cloud Computing Environment

N.Krishnaveni,
PG Scholar, Dept. of CSE
Erode Sengunthar Engineering College
Thudupathi
Erode (DT), Tamilnadu.

Prof.G.Sivakumar (Ph.D.)
Associate Professor, Dept. of CSE
Erode Sengunthar Engineering College
Thudupathi
Erode (DT), Tamilnadu.

Abstract-Cloud computing becomes quite popular among cloud users by offering a variety of resources. This is an on demand service because it offers dynamic flexible resource allocation and guaranteed services in pay as-you-use manner to public. The different resource allocation technique is compared for resource allocation process in cloud environment. The skewness algorithm was developed in the existing process to reduce the server overload. The new technique called latent semantic ranking is applied to the resource allocation process.

Keywords: Cloud computing, Dynamic Resource Allocation, Data center, Virtual machine, Ranking

I INTRODUCTION

Cloud computing allows customers to scale up and down their resources based on needs. Cloud computing technology makes the resources as a single point of access to the client and cost is pay per usage. Cloud computing is a computing technology, where a pool of resources are connected in private and public networks and to provide these dynamically scalable infrastructure for application. Cloud computing is not application oriented and this is a service oriented. It offers the virtualized resources to the cloud users. Cloud computing provide dynamic provisioning and thus can allocate machines to store data and add or remove the machines according to the workload demands. Cloud computing platforms such as, those provided by Microsoft, Amazon, Google, IBM. Cloud computing is an environment for sharing resources without the knowledge of the infrastructure and can makes it possible to access the applications and its associated data from anywhere at any time.

Cloud environment provide the four types of cloud.

* Public cloud
* Private cloud
* Hybrid cloud
* Community cloud

Cloud computing offers three types of services

* Software as a service(Saas)
* Platform as a service(Paas)
* Infrastructure as a service(IaaS)

The dynamic resource allocation process must reduce the user access time and overload of the server. In dynamic resource allocation process using latent semantic ranking technique avoid the resource contention and reduce the server overload and also reduce the latency. The existing process does not provide the scalability and server overload is increased during resource allocation process. This will be overcome by the ranking technique. The ranking provide the high priority for high user bandwidth.

Virtualization Technology

Cloud computing is based on the virtualization technology. Virtualization technology is used to allocate the data center resources dynamically based on the application demands [11].

Live Migration

Virtual machine live migration technology makes it possible to mapping between the virtual machines (VMs) and the physical machines (PMs) while applications are running. Live migration increase the resource utilization and provide the better performance result.

II RESOURCE ALLOCATION

In cloud computing, Resource allocation is the process of assigning available resources to the needed cloud applications. Cloud resources can be provisioned on demand in a fine-grained, multiplexed manner. In cloud the resource allocation is based on the infrastructure as a service (IaaS).In cloud platforms, resource allocation takes place at two levels [1]:

* when an application is uploaded to the cloud, the load balancer assigns the requested instances to physical computers, to balance the computational load of multiple applications across physical computers
* When an application receives multiple incoming requests, these requests should be assigned to a specific application instance to balance the

computational load across a set of instances of the same application

Resource allocation techniques should satisfy the following criteria:

* Resource contention arises when two applications try to access the same resource at the same time
* Resource fragmentation arises when the resources are isolated. There would be enough resources but cannot allocate it to the needed application due to fragmentation.
* Scarcity of resources arises when there are limited resources and the demand for resources is high
* The multiple applications needed different types of resources such as CPU, Memory, I/O devices and the technique should satisfy that request
* Over provisioning of resources arises when the application gets surplus resources than the demanded one

III EXISTING WORK

In existing system the skewness algorithm is developed[9].

The Virtualization technology dynamically allocates data center resources based on application demands and support green computing by optimizing number of servers. The load prediction algorithm to capture future resource usages of applications accurately and capture rising trend of resource usage patterns .

Skewness is introduced to measure unevenness in multidimensional resource utilization of a server.Skewness algorithm executes periodically to evaluate resource allocation status based on predicted future resource demands of VMs. Here two types of threshold was used to measure the server overload status.

* If utilization of any of its resources is above a hot threshold and its indicates server is overloaded and VMs running on it migrated away
* If utilizations of all its resources are below a cold threshold indicates server is mostly idle

IV PROPOSED WORK

The proposed technique for cloud resource allocation process is Latent Semantic Ranking (LSR). LSR governs semantic relativity of user job requirements and resource availability. LSR provide generic capability of collecting log data about internal jobs. Unify the collected data to reveal global coherence semantivity.

Scheduling of resource allocation for jobs is done through ranking. Ranking procedure is adapted for semantic relation of requested jobs to available resources with its corresponding user bandwidth. There are three classification methods in Latent Semantic Ranking (LSR) Model for Cloud Resource Allocation. The three models are,

* Distributed Cloud with Virtual Servers
* Semantic Relation of Requested Jobs
* Ranking of Resources to Semantic Jobs

Distributed Cloud with Virtual Servers

Distributed Cloud with virtual servers allocates resources based on predicting the historical data**.** Virtual servers do management of requested jobs and allocating resources to corresponding jobs. Virtual servers are distributed across multiple locations deploying containers on each of them. The messaging system allows communication between clouds servers located on different spheres improve scalability by processing in parallel on multiple hosts. This is managing user required resources.

Semantic Relation of Requested Jobs

Execution has different states some indicate execution is running correctly; others indicate problems in the execution. Depending on the semantic state of the job execution resource scheduler has to act in a different way. Jobs are separated in several sub-goals activated depending on the job status.

In the running state scheduler activates the goal for monitoring job execution and evaluating if required performance is fulfilled and if the job is finished by the scheduler execute plans for de-allocating resource.

Ranking of Resources to Semantic Jobs

Resource allocation for a particular job is decided between semantic relation of requested jobs and resource available ranking. Ranking is done on the resource size availability matching to the requested job. Message exchange between requested jobs to the allocated resource coordinates the job scheduling. Ranking of resources activated a goal for allocation to a job, whose resources match with the job requirements. Rank scheduler initiates a negotiation sending to the selected resources call for scheduling proposals.

The rank assigning to resources proposals and job allocation evaluations are done using scheduler module. Rank assignment makes scheduling proposals evaluating scheduling rules over resource and assigned job descriptions.

V RESULTS AND DISCUSSION

In this section we evaluate performance of Latent semantic ranking (LSR) model for cloud resource allocation through CloudSim simulation in java environment. The performance of is evaluated by the following metrics.

- Resource Availability Rate
- Overload
- Latency

Resource Availability Rate

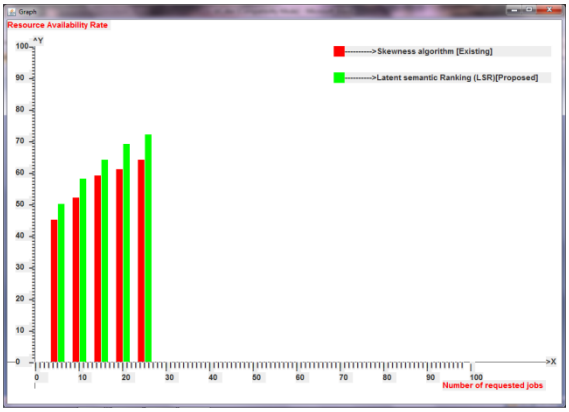

Fig 1.Number of Requested jobs Vs Resource availability rate

This figure shows the resources availability rate for both skewness and ranking technique the resource availability rate is control how much a work resource is available to work throughout the life of the project. In existing process (skewness) the resource availability rate was very less if the number jobs is increased. In proposed process (ranking) the resource availability rate is more when the number jobs is increased.

Overload of the process

This figure shows the overload of the server.the overload is reduced compare to the existing process using ranking technique.The proposed system allocate the resources based on the user bandwidth and it allows the more user at a time.This was reduce the overload of the server.

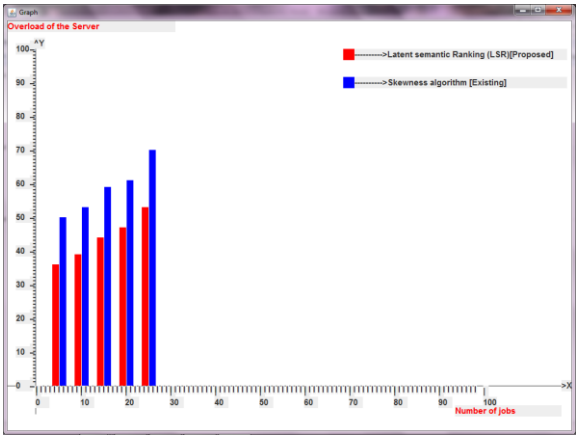

Fig 2.Number of jobs Vs Overload of the server

Latency

The following figure(fig 3.) shows the latency of the process.In existing process using the threshold value of the server and during resource allocation process this value is checked every time.so,this will take the more time for resource allocation.In proposed technique,there is no use of threshold value but using the ranking technique.this was allocated the resources based upon the user bandwidth,so this minimize the latecy of the process.

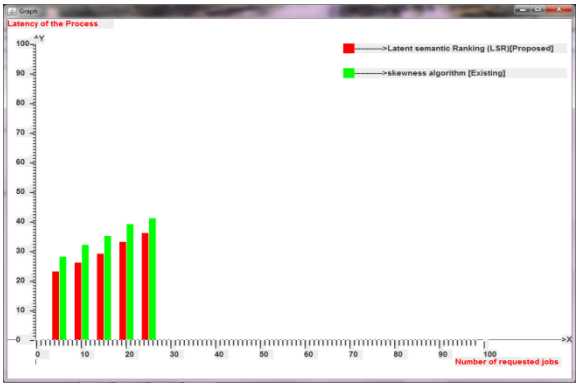

Fig 3.Number of jobsVs Latency

VI CONCLUSION

By resource allocation process the server overload is reduced. The detail description of the techniques is summarized and also summarizes the advantages with parameters of the various techniques in cloud computing environment. The latent semantic ranking was introduced to allocate the resources dynamically and its reduce the overload compare to the existing skewness algorithm. The latent semantic ranking also minimizes the latency of the process.

VII REFERENCES

[1] Ronak Patel, Sanjay Patel" Survey on Resource Allocation Strategies in Cloud Computing" International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, Feb- 2013

[2] K C Gouda, Radhika T V, Akshatha M,"Priority based resource allocation model for Cloud computing" International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 1, January 2013

[3]Gunho Leey, Byung-Gon Chunz, Randy H.Katzy,"Heterogeneity-Aware Resource Allocation and Scheduling in the Cloud", IJERT-2012

[4] Nilton Bilay, Eyal de Laray, Kaustubh Joshi_, H. Andres Lagar-Cavilla ,Matti Hiltunen_ and Mahadev Satyanarayananz," Efficient Idle Desktop Consolidation with Partial VM Migration", Journal of computer application-2012

[5]Venkatesa Kumar, V. And S. Palaniswami," A Dynamic Resource Allocation Method for Parallel data processing in Cloud Computing", Journal of Computer Science 8 (5): 780-788, 2012

[6] Sheng Di and Cho-Li Wang," Dynamic Optimization of Multi-Attribute Resource Allocation in Self-Organizing Clouds", IEEE Transactions on parallel and distributed systems, - 2013

[7] V.Vinothina, Dr.R.Sridaran, Dr.padmavat higanapa thi," A Survey on Resource Allocation Strategies in Cloud Computing "International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012

[8] Qi Zhang, Eren G¨urses, Raouf Boutaba, Jin Xiao," Dynamic Resource Allocation for Spot Markets in Clouds", Journal of computer science-2012

[9] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen," Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment",

IEEE Transactions on parallel and distributed systems, vol. 24, no. 6, June 2013
[10] Ts`epomofolo, R Suchithra," Heuristic Based Resource Allocation Using Virtual Machine Migration: A Cloud Computing Perspective", International Refereed Journal of Engineering and Science (IRJES) Volume 2, Issue 5(May 2013)
[11]Ratan Mishra & Anant Jaiswal (2012), "Ant Colony Optimization: A Solution of Load Balancing in Cloud", International Journal of Web & Semantic Technology, Vol: 3, No: 2, pp: 33-50.

