International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-2 Issue-5, November 2013

Review of Delay and Cost Efficient Methods in Cloud Computing

Ninad Shinde, J. Ratnaraja Kumar

Abstract: ***Now days the approach of cloud computing is providing utility oriented computer services to end users all over the world. At same time, cloud users can make number of requests for different cloud services and their resources. With help cloud we can able to access our personal or public data or applications from anyplace and anywhere. Many commercial companies from the world can able to take the required resources on rent from the cloud for storage of their private data as well as for other purposes. This can help companies to reduce their costs in using extra infrastructures significantly. Cloud can offer companies security to their applications or data and allow its access to them based on pay-as-you-go model. As the number of requests to cloud computing for their usages resulted into the disadvantages as well. The main challenge in cloud computing is related to the efficient allocation of their resources. If there is no optimized technique for resource allocation in cloud computing then it resulted into long delays, end user hang out, request failure. Hence it is required to use optimized resource allocation technique with aim of minimizing resource allocation delays and their costs. In this paper we are aiming to present different techniques for resource allocation in cloud computing.***

Keywords: ***Cloud Computing, Resource Allocation, Cost Optimization, resource management, Quality of Service.***

I. INTRODUCTION

Cloud computing is an emerging technical area. The cloud computing is a popular technology, which gives the resources and applications to their users. The multinational companies like Microsoft, IBM, Google, and Oracle provide the different cloud computing services to their clients. At present, the cloud computing is a business area. Therefore, money and quality of service are the main important things in this area. The cloud computing providers delivers resources and, software to their clients in terms of Service Level Agreement (SLA). It is a deal between a user and cloud provider. This includes the user's requirement of resources, service time limit, and cost of the service [1].

Cloud computing emerges as a new computing paradigm which aims to provide reliable, customized and QoS (Quality of Service) guaranteed computing dynamic environments for end-users. Distributed processing, parallel processing and grid computing together emerged as cloud computing. The basic principle of cloud computing is that user data is not stored locally but is stored in the data center of internet [2] [3]. The companies which provide cloud computing service could manage and maintain the operation of these data centers.

Revised Manuscript Received on 30 November 2013.
* Correspondence Author
Mr. Ninad Shinde, M.E. (Comp.Sci.), Genba Sopanrao Moze College Of Engineering, Balewadi, Pune India.
Prof. J. Ratnaraja Kumar, Genba Sopanrao Moze College Of Engineering, Balewadi, Pune , India.

Every person can operate the stored data at any time by using Application Programming Interface (API) provided by cloud providers through any terminal equipment connected to the internet. Cloud computing has been established as the most recent and most flexible delivery model of supplying information technology. Cloud computing can be seen as an innovation in different ways. From a technical point of view calculation history can be traced back to the machine, which is an advancement of computing. While from a technical perspective, we can manage the cloud computing challenges, but from a strategic point of view, with operations both on a business level, there are a number of challenges [4]. A fundamental advantage of the cloud paradigm, the computational power of cloud customers is no longer limited by their resource constraint devices where outsourcing is calculated. By outsourcing the workloads into the cloud, customers could enjoy the literally unlimited computing resources in a pay-per-use manner without committing any large capital outlays in the purchase of hardware and software and or the operational overhead there in. The cloud computing workloads to outsource their large customers with limited computational resources enables, and financially overwhelming computational power, bandwidth, storage, and even a pay-per-use way to share enjoy that appropriate software. Cloud computing is the next generation of calculations. Perhaps people can have everything they need on the cloud. On-demand cloud computing services and products, development of information technology is the next natural step. Cloud Computing is an emerging computing technology that is rapidly consolidating itself as the next big step in the development and deployment of an increasing number of distributed applications [5] [6].

II. RELATED WORK

Very little literature is available on this survey paper in cloud computing paradigm. Shikharesh et al. [7 in paper describes the resource allocation challenges in clouds from the fundamental point of resource management. The paper has not addressed any specific resource allocation strategy. Patricia et al., [8] investigates the uncertainties that increase difficulty in scheduling and matchmaking by considering some examples of recent research. It is evident that the paper which analyzes various resource allocation strategies is not available so far. The proposed literature focuses on resource allocation strategies and its impacts on cloud users and cloud providers. It is believed that this survey would greatly benefit the cloud users and researchers. The dynamic resource allocation in cloud computing has attracted attention of the research community in the last few years. It is one of the most challenging problems in the resource management problems. Many researchers around the world have come up with new ways of facing this challenge.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

In authors propose a model and a utility function for location-aware dynamic resource allocation. A comprehensive comparison of resource allocation policies is covered in. Hua's paper proposed an ant colony optimization algorithm for resource allocation, in which all the characteristics in cloud are considered. It has been compared with genetic algorithm and annealing algorithm, proving that it is suitable for computing resource search an allocation in cloud computing environment.

This paper is not intended to address any specific resource allocation strategy, but to provide a review of some of the existing resource allocation techniques. Analyzes of different resource allocation strategies that are not many papers being the most recent technology available in the form of cloud computing. Literature survey on cloud users and cloud providers focused on resource allocation strategies and impacts. This survey is a cloud that will benefit users.

III. RESOURCE ALLOCATION AND METHODS

A. Significance of Resource Allocation

In cloud computing, resource allocation (RA) needed cloud on the Internet is the process of assigning resources to applications. If not managed properly allocate resource allocation services starve. Resource provisioning service providers for each individual module are allowing you to manage the resources that solve this problem [11].

Resource Allocation Strategy (RAS) is all about integrating cloud provider activities for utilizing and allocating scarce resources within the limit of cloud environment so as to meet the needs of the cloud application. It requires the type and amount of resources needed by each application in order to complete a user job. The order and time of allocation of resources are also an input for an optimal RAS. An optimal RAS should avoid the following criteria as follows:

a) Resource contention situation arises when two applications try to access the same resource at the same time.
b) Scarcity of resources arises when there are limited resources.
c) Resource fragmentation situation arises when the resources are isolated. [There will be enough resources but not able to allocate to the needed application.]
d) Demanding applications when compared to a surplus of resources, excessive provisioning resources arise.
e) Under-provisioning of resources occurs when the application is assigned with fewer numbers of resources than the demand.

Resource users' (cloud users) estimates of resource demands to complete a job before the estimated time may lead to an over-provisioning of resources. Resources, resource providers' allocation of resources may lead to an under-provisioning [10]. The angle of the cloud user, application and service level agreement is required (SLA) as shown in Table I for a RAS cloud providers and users need both anomalies mentioned above, the major inputs to overcome ness are invested. The offerings, resource status and available resources are the inputs required from the other side to manage and allocate resources to host applications by RAS. The outcome of any optimal RAS must satisfy the parameters such like throughput, timing of response & latency. And also after cloud delivering reliable resources, it also poses a crucial problem in allocating and managing resources dynamically across the applications.

TABLE 1. INPUT PARAMETERS

Parameter	Provider	Customer
Provider Offerings	√	-
Resource Status	√	-
Available Resources	√	-
Application Requirements	-	√
Agreed Contract Between Customer and provider	√	√

From the perspective of a cloud provider, predicting the dynamic nature of users, user demands, and application demands are impractical. For the cloud users, the job should be completed on time with minimal cost. Hence due to limited resources, resource heterogeneity, locality restrictions, environmental necessities and dynamic nature of resource demand, we need an efficient resource allocation system that suits cloud environments [12].

Cloud resources consist of physical and virtual resources. The physical resources are shared across multiple compute requests through virtualization and provisioning. The request for virtualized resources is described through a set of parameters detailing the processing, memory and disk needs which is depicted in Fig.1. Provisioning satisfies the request by mapping virtualized resources to physical ones. The hardware and software resources are allocated to the cloud applications on-demand basis. For scalable computing, Virtual Machines are rented [13].

The complexity of finding an optimum resource allocation is exponential in huge systems like big clusters, data centers or Grids. Since resource demand and supply can be dynamic and uncertain, various strategies for resource allocation are proposed. This paper puts forth various resource allocation strategies deployed in cloud environments.

B. Ant Colony Optimization Algorithm for Resource Allocation

Like the storage space allocation strategy in the memory or cache of a PC, the client requirement of the hardware infrastructure following the content of agreement should be allocated dynamically from the node pool. Since the specific condition of resource is unknown under cloud circumstance, and the networks do not have a fixed topology, the structure and the resource allocation of the whole cloud environment is unpredictable.

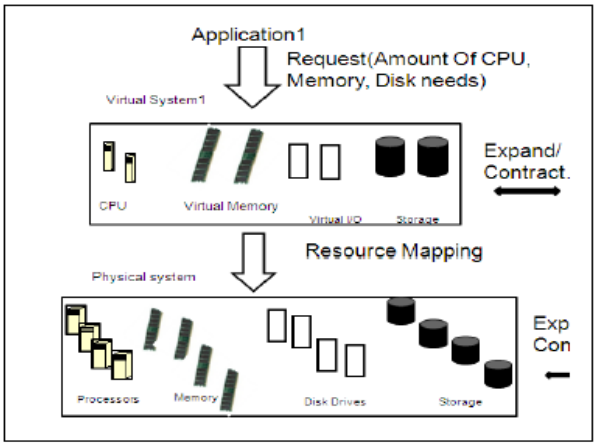

Figure1. Mapping of virtual to physical resources

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

This means to allocate resources to the user all according to the instant need of the client program instead of the commission described in the agreement, all the time each unit in the cloud computing node cluster uses a Master/Slaves structure as shown in Fig.2.

Using Ant colony algorithm we can discover the computing resources in an unknown network topology and select the most appropriate one or more resources to user's job until it meets user's requirements. There is a Master node responsible for controlling and supervising all the Slave nodes. The slave node only responses to the tasks the master node is distributed to.

When the search begins, query messages will be sent by slave node, and they will play the role of ants and these ants will follow the formula that choosing the point with the probability as more pheromone, more possibility. Furthermore, those ants will leave a certain dose of pheromone on the point that will be passed. In order to reflect the change of the pheromone, researchers adopt a local update strategy to modify the pheromone intensity onto the node.

When a user task comes, the master job tracker node is responsible for all assignments, of which data resources may be contained by the user image slices spreading in different storage nodes of their slave task tracker node. When Sent by the master node to a slave node task tracker receives the appropriate computing nodes as the storage nodes will begin to search. In the first phase of the slave node is useable computing resources to scan begins.

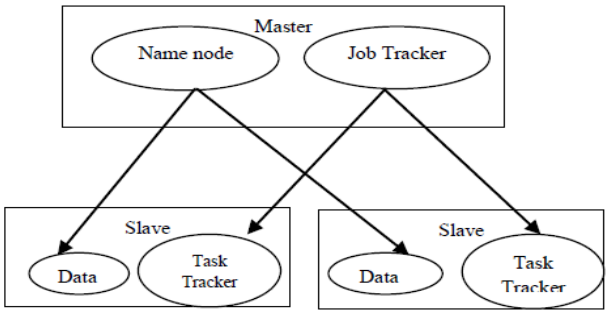

Fig. 2 Master Slaves Structure

If it can satisfy the user's needs which are guaranteed by the provider, slave node will allocate them first to assignments from master node as local computing resource has high demand. If not, then it will create resources in the cloud environment. These discovering should be done in a certain scale to manage the network cost. Ultimately, if not getting resources, slave tracker will announce the master tracker to move the user image slice to other slave trackers. The entire cloud environment is unknown, because the details of the resources are in cognizable and the construction of network topology is mutable. In this situation, the location and quality of the computing resource points are in cognizable to the storage points [14].

Hua's paper provides a detailed description about ant colony algorithm in resource allocation and uses Gridsim to simulate local domain of cloud computing to inspect the operating conditions of the algorithm under cloud network environment.

Procedure: ACO Algorithm
Begin
While (ACO has not been stopped) do
Schedule activities
Ant's allocation and moving (ant distribution and movement)
Local pheromone updates (local pheromone update)
Global pheromones update (global pheromone update)
End Schedule activities
End While
End Procedure

Through Grid Resource class and a series of helper classes in Gridsim, researchers simulate the computation and network resources of cloud computing and constructs a relatively real cloud layout.

After much experimentation, researchers found that ant colony algorithm is more effective in the case that there are more nodes and fewer resources, which is just the characteristic of cloud environment. Ant colony algorithm aims at the large-scale, shared, dynamic and other characteristics of cloud environment. It assigns search and allocates computation resources to user's job dynamically. And it shows more advantages in cloud environment.

C. Dynamic Resource Allocation for Parallel Data Processing

Dynamic Resource Allocation for Efficient Parallel data processing introduces a new processing framework explicitly designed for cloud environments called Nephele. Most notably, Nephele is the first data processing framework to include the possibility of dynamically allocating/de-allocating different compute resources from a cloud in its scheduling and during job operation. One common task for perform a job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution.

1) Architecture: Nephele's architecture follows a classic master-worker pattern as illustrated in Fig 4. Before submitting a Nephele computes job, a operator want to be perform a VM in the cloud which runs the so called Job Manager (JM).The Job Manager receives the client's jobs, manger is the responsible person to manage them, and coordinates their execution. The cloud operator to control the start of VMs provides the interface is able to communicate with. We call this interface cloud controller. Job Manager via the cloud controller allocated or de-allocated according to current job execution phase can VMs. Nephele work consists of an actual execution of tasks that are performed by a set of examples. Each example is a so-called Task Manager (TM) runs .A Task Manager receives one or more tasks from the team manager at this time, used them, and after report to Job Manager about their completion or possible errors.

2) Job Description: Jobs in Nephele are expressed as a directed acyclic graph (DAG). Each vertex in the graph represents a task of the overall forwarding job. Communication flow between these functions defines the edges of the graph. Example parameter assignments of subtasks per job description, for example, the number of instances of the subtasks, are based on the number of data sharing.

3) Job Graph: Once the job graph is specified, the user's credentials to get her together with cloud operator, present it to the Job Manager. During the execution of the job from the Job Manager User instances allocated / D must allocate the necessary credentials.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

4). Advantages and Limitations of Resource Allocation: Regardless of the size of the organization and the business market using cloud computing has many advantages in resource allocation. It is an emerging technology, but there are some limitations. The advantages and limitations of resource allocation in the cloud is a comparative look.

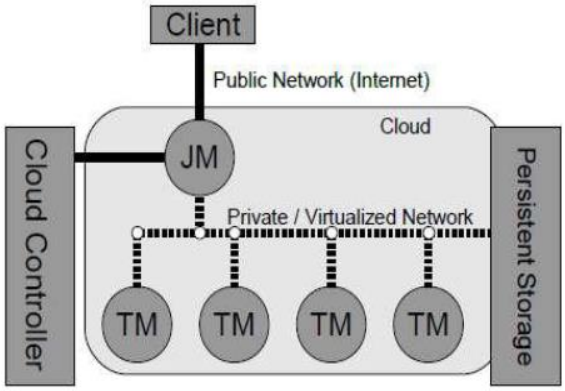

Fig 3. Design Architecture of Nephele Framework

Advantages:

- The higher advantage of resource allocation is that user neither has to install software nor hardware to access the applications, to develop the application and to host the application over the internet.
- The next major benefit is that there is no limitation of place and medium. We can get our data with applications anywhere in the world, on every system.
- The operator not necessary to exclude on hardware & software systems.
- Cloud services can provide their resources over the internet during resource scarcity.

Limitations:

- Since users rent resources from remote servers for their purpose, they don't have control over their resources.
- Migration problem occurs, when the users wants to switch to some other provider for the better storage of their data. It's not easy to transfer huge data from one provider to the other.
- In public clouds, customers' data may be susceptible to hacking or phishing attacks. Since the servers are connected to the cloud, it is easy to spread malware like printers or scanners.
- Peripheral devices cannot work with cloud. Many of them require software to be installed locally at the external network problems. Knowledge about the functioning of cloud mainly depends on the cloud service provider
- More and profound knowledge, resource allocation and management in the cloud are necessary.

IV. CONCLUSION

Cloud computing technology is increasingly being used in enterprises and business markets. A review shows that dynamic resource allocation is growing need of cloud providers for more number of users and with the less response time. Resource allocation can influence the quality and cost estimation in cloud computing system. In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the classification of RAS and its impacts in cloud system. Some of the techniques communicating above mostly focus on CPU, memory resources but are lacking in some factors. Hence this paper will exactly motivating next future researchers to come up with smarter and secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

REFERENCES

1. V. Vinothina, Dr. R. Shridaran, & Dr. Padmavathi Ganpathi, A survey on resource allocation strategies in cloud computing, International Journal of Advanced Computer Science and Applications, 3(6):97--104, 2012.
2. Gunho Lee, Niraj Tolia, Parthasarathy Ranganathan, and Randy H. Katz, Topology aware resorce allocation for data-intensive workloads, ACM SIGCOMM Computer Communication Review, 41(1):120--124, 2011.
3. Abirami S.P. and Shalini Ramanathan, Linear scheduling strategy for resource allocation in cloud environment, International Journal on Cloud Computing: Services and Architecture(IJCCSA), 2(1):9--17, 2012.
4. Daniel Warneke and Odej Kao, Exploiting dynamic resource allocation for efficient parallel data processing in the cloud, IEEE Transactions On Parallel And Distributed Systems, 2011.
5. Atsuo Inomata, Taiki Morikawa, Minoru Ikebe and Md. Mizanur Rahman, Proposal and Evaluation of Dynamic Resource Allocation Method Based on the Load Of VMs on IaaS, IEEE, 2010.
6. Dorian Minarolli and Bernd Freisleben, Uitlity–based Resource Allocations for virtual machines in cloud computing, IEEE, 2011.
7. Gunho Lee, Niraj Tolia, Parthasarathy Ranganathan, and Randy H.Katz, Topology aware resource allocation for data-intensive workloads, ACM SIGCOMM Computer Communication Review, 41(1):120--124, 2011.
8. Li Li, Niu Ben. Particle swarm optimization [M]. Beijing: Metallurgical Industry Press, 2009.
9. Daniel Warneke and Odej Kao, Exploiting dynamic resource allocation for efficient parallel data processing in the cloud, IEEE Transactions On Parallel And Distributed Systems, 2011.
10. A.Singh ,M.Korupolu and D.Mohapatra. Server-storage virtualization:Integration and Load balancing in data centers. In Proc.2008 ACM/IEEE conference on supercomputing (SC'08) pages 1-12, IEEE Press 2008.
11. AndrzejKochut et al. : Desktop Workload Study with Implications for Desktop Cloud Resource Optimization,978-1-4244-6534-7/10 2010 IEEE.
12. Atsuo Inomata, TaikiMorikawa, Minoru Ikebe, Sk.Md. MizanurRahman: Proposal and Evaluation of Dynamin Resource Allocation Method Based on the Load Of VMs on IaaS(IEEE,2010),978-1-4244-8704-2/11.
13. D. Gmach, J.RoliaandL.cherkasova, Satisfying service level objectives in a self-managing resource pool. In Proc. Third IEEE international conference on self-adaptive and self organizing system.(SASO'09) pages 243-253.IEEE Press 2009 .
14. Arkaitz Ruiz-Alvarez, Marty Humphrey, A Model and Decision Procedure for Data Storage in Cloud Computing, in Proceedings of the IEEE/ACM International Symposium on Cluster, Ottawa Canada, 2012.
15. Qi Zhang, Quanyan Zhu, Raouf Boutaba , Dynamic Resource Allocation for Spot Markets in Cloud Computing Environments, Fourth IEEE International Conference on Utility and Cloud Computing (UCC), Melbourne Australia, 2011.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)