Techniques for Task Reallocation in Cloud Computing

Amandeep Kaur
Research Scholar
Punjabi University Regional Centre for Information Technology and Management, Mohali, Punjab, India

Dr. Rekha Bhatia
Associate Professor
Punjabi University Regional Centre for Information Technology and Management, Mohali, Punjab, India

ABSTRACT
The cloud computing is the architecture which has decentralized in nature. The cloud computing is the architecture in which virtual machines, hosts and data centers are involved in the communication. The virtual machines are responsible to execute the cloudlet which is assigned by the virtual machine. The virtual machines get overloaded when more task is assigned to the virtual machines. In this paper, various techniques of virtual machines overloading are reviewed and discussed.

KEYWORDS
Cloudlet, VM, fault, overloading

INTRODUCTION
Cloud Computing:
A large pool of systems where the private as well as the public networks are connected to each other is known as cloud computing. This technology provides a dynamically scalable infrastructure for the applications, data as well as the file storage. The cost of computation, the content storage and the delivery are reduced using this evolving technique. Through this technique, direct cost benefits are to be experienced and the data center can be transformed from a capital-intensive set up.

Public Cloud: Public clouds are owned and operated by outsiders; they convey better economies of scale than customers, as the infrastructure costs are spread among a mix of users, giving every individual customer an attractive minimal cost, "Pay-as-you-go" model.

Private Cloud: This type of cloud is particularly used for a single enterprise and has an objective for concerning on data security. It provides a great control over the data storage and this facility is absent in the public cloud.

Hybrid Cloud: Hybrid Clouds join both public and private cloud models. With a Hybrid Cloud, service providers can use outsider Cloud Providers in a full or partial manner in this way expanding the flexibility of computing.
Regulatory and Compliance Restrictions: In some areas of the European countries, Government regulations don’t permit customer’s personal information and other sensitive information to be physically located outside the state or country.

Resource Allocation Techniques in Cloud Computing:

Optimized Resource scheduling algorithm: To accomplish the optimization for cloud scheduling tribulations, an optimized resource scheduling algorithm is proposed based on the profound research on Infrastructure-as-a-Service (IaaS) cloud systems.

Resource allocation strategy based on market (RAS-M): A resource allocation strategy based on market (RAS-M) is proposed here, consecutively to advance resource consumption of bulky data centers while giving services higher QoS to Cloud consumers.

Scheduling with multiple SLA parameters: In Cloud computing methodology, the vital trademark is dispensing resources in an adaptable on-demand approach. Services are given in Clouds based on Service Level Agreements (SLAs).

Rule Based Resource Allocation Model (RBRAM): Resource arbitration allows multiple independent components safe access to a resource, without including any extra maintenance cost. In cloud computing, services are owed based on customer computing constraints and henceforth enabling optimal consumption of the requested resources by the customers is a challenge.

Resource allocation using Scalable computing: Cloud computing gives the user IaaS service, of leasing computing resources over the Internet. Based on the necessities, the client can choose from diverse types of computing resources.

Literature Review

Amit Nathani, et.al, 2011 In present situation, the greater part of the Infrastructure as a Service (IaaS) clouds utilize simple resource allocation policies like immediate and best effort. Immediate allocation policy allocates the resources if accessible, generally the request is rejected. Haizea supports four sorts of resource allocation policies: immediate, best effort, booking ahead of time and deadline sensitive. This work provides a superior approach to support deadline sensitive leases in Haizea while minimizing the total number of leases rejected by it. Proposed dynamic arranging based scheduling algorithm is implemented in Haizea that can admit new leases and prepare the schedule at whatever point another rent can be accommodated. Experiments results demonstrate that it maximizes resource utilization and acceptance of leases compared to the existing algorithm of Haizea.

Anton Beloglazov et.al.”, 2010

Fast growth of the interest for computational power by logical, business and web-applications has prompted to the creation of large-scale data centers consuming enormous measures of electrical power. An energy efficient resource management system is proposed for virtualized Cloud data centers that reduces operational costs and provides required Quality of Service (QoS). The results demonstrate that the proposed system brings substantial energy savings, while ensuring reliable QoS. Other than the significant reduction of operational costs, the project is socially valuable as it decreases carbon dioxide footprints and general energy consumption by modern IT infrastructures. This legitimizes further examination and development of the proposed resource management system.

Haohao Zhou, et.al, 2016

Cloud computing provides users a shared pool of configurable computing resources. In this paper, a cloud computing system is regarded as a queuing system, where users touch base as indicated by a stochastic process and request resources, including CPU, memory, storage space. and so on. To improve the utilization of the system under stable state, some theoretical results are given about the relationship between the utilization and the stability of the cloud computing system [14]. Exploring the stability region of existing scheduling algorithms can be considered as a piece of future work. Moreover, scheduling in real cloud systems dependably have various factors that ought to be considered, for example, utilization, QoS (Quality of Service), et cetera. QoS can likewise be considered in the stability analysis of specific forms in future works.

Huangke Chen, et.al , 2015

Green cloud computing has turned into a major concern in both industry and academia, and efficient scheduling approaches demonstrate promising approaches to reduce the energy consumption of cloud computing platforms while guaranteeing QoS.
requirements of tasks. Existing scheduling approaches are insufficient for real-time tasks running in uncertain cloud environments, in light of the fact that those approaches assume that cloud computing environments are deterministic and pre-computed schedule decisions will be statically followed amid schedule execution [15]. In this paper, this issue is addressed. Extensive experiments are conducted to compare PRS with four typical baseline scheduling algorithms. The experimental results demonstrate that PRS performs superior to those algorithms, and can effectively improve the performance of a cloud data center.

Doulamis ND, et.al, 2014

Resource selection and task assignment are essential operations in dispersed computing environments, similar to the grid and the cloud, where tasks compete for resources. The decisions made by the corresponding algorithms ought to be judged based not just on metrics related to client satisfaction, for example, the percentage of tasks served without violating their quality-of-service (QoS) requirements, additionally based on resource-related performance metrics, for example, the number of resources used to serve the tasks and their utilization efficiency [16]. The partitioning is performed utilizing a spectral clustering methodology through normalized cuts. Experimental results demonstrate that the proposed algorithm outperforms other scheduling algorithms for various values of the granularity and the load of the task requests.

Abdul Hameed, et.al, 2014

In a cloud computing worldview, energy efficient allocation of various virtualized ICT resources (servers, storage disks, and networks, and so forth) is a complex problem because of the nearness of heterogeneous application (e.g., content delivery networks, MapReduce, web applications, and so forth) workloads having contentious allocation requirements in terms of ICT resource capacities (e.g., network transfer speed, processing speed, response time, and so on) [17]. Furthermore, accessible techniques as of now presented in the literature are summarized based on the energy-efficient research dimension taxonomy. The focal points and inconveniences of the existing techniques are comprehensively analyzed against the proposed research dimension taxonomy namely: resource adaption policy, objective function, allocation method, allocation operation, and interoperability.

Young Choon Lee, et.al, 2012: The energy consumption of under-used resources, especially in a cloud environment, accounts for a substantial measure of the genuine energy utilize. Inherently, a resource allocation strategy that considers resource utilization would lead to superior energy efficiency; this, in clouds, extends further with virtualization advancements in that tasks can be effectively merged. Task consolidation is an effective method to increase resource utilization and thus reduces energy consumption. In this paper, two energy-cognizant task consolidation heuristics are presented, which aim to maximize resource utilization and explicitly consider both dynamic and sit without moving energy consumption. Our heuristics assign every task to the resource on which the energy consumption for executing the task is explicitly or implicitly minimized without the performance degradation of that task. Based on the experimental results, the heuristics demonstrate their efficient energy-saving capability.

Zhanjie Wang, et.al, 2015: The development of big data challenges the computing power and communication capability of cloud architecture; however traditional resource-allocation algorithms perform poorly because of the large-scale communication among cloud nodes [19]. In this paper, a dynamically hierarchical, resource-allocation algorithm is proposed for multiple cloud nodes collaborating in big data environment. Both theoretical and experimental results represent that the proposed algorithm outperforms the MinMin algorithm in terms of communication traffic and makespan. The results demonstrate that DHRA can reduce message number and communication traffic significantly, with the equal or even less tasks finish time compared with MinMin. DHRA is proven to be efficient for resource allocation in cloud computing environment.

Guiyi Wei, et.al, 2010: As cloud-based services turn out to be more numerous and dynamic, resource provisioning turns out to be more and more challenging. A QoS constrained resource allocation problem is considered in this paper, in which service demanders intend to take care of sophisticated parallel computing problem by requesting the utilization of resources over a cloud-based network, and a cost of each computational
service depends on the measure of computation. The algorithms in the evolutionary mechanism consider both improvement and fairness. It is demonstrated that Nash equilibrium dependably exists if the resource allocation game has feasible solutions. The advancement problem considered in this paper relates to a large proportion of cloud-based computing services. The method might be a valuable analytical tool for shedding light on seeking optimal scheduling solution for the complex and dynamic problems that can be divided into multiple cooperative subtasks in many cloud-based computing and data store services.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Description</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amit Nathani, et.al</td>
<td>2011</td>
<td>It shows the greater part of the Infrastructure as a Service (IaaS) clouds utilize simple resource allocation policies like immediate and best effort. Immediate allocation policy allocates the resources if accessible, generally the request is rejected.</td>
<td>Experiments results demonstrate that it maximizes resource utilization and acceptance of leases compared to the existing algorithm of Haizea.</td>
</tr>
<tr>
<td>Anton Beloglazov et.al.,”</td>
<td>2010</td>
<td>Fast growth of the interest for computational power by logical, business and web-applications has prompted to the creation of large-scale data centers consuming enormous measures of electrical power.</td>
<td>It result in real cloud systems dependably have various factors that ought to be considered, for example, utilization, QoS (Quality of Service), et cetera. QoS can likewise be considered in the stability analysis of specific forms in future works.</td>
</tr>
<tr>
<td>Huangke Chen, et.al</td>
<td>2011</td>
<td>Green cloud computing has turned into a major concern in both industry and academia, and efficient scheduling approaches demonstrate promising approaches to reduce the energy consumption of cloud computing platforms while guaranteeing QoS requirements of tasks.</td>
<td>The experimental results demonstrate that PRS performs superior to those algorithms, and can effectively improve the performance of a cloud data center.</td>
</tr>
<tr>
<td>Doulamis ND, et.al</td>
<td>2014</td>
<td>Resource selection and task assignment are essential operations in dispersed computing environments, similar to the grid and the cloud, where tasks compete for resources.</td>
<td>The partitioning is performed utilizing a spectral clustering methodology through normalized cuts. Experimental results demonstrate that the proposed algorithm outperforms other scheduling algorithms for various values of the granularity and the load of the task requests.</td>
</tr>
<tr>
<td>Abdul Hameed, et.al</td>
<td>2014</td>
<td>In a cloud computing worldview, energy efficient allocation of various virtualized ICT resources is a complex problem because of the nearness of heterogeneous application (e.g., content delivery networks, MapReduce, web applications, and so forth) workloads having contentious allocation requirements in terms of ICT resource capacities (e.g., network transfer speed, processing speed, response time, and so on)</td>
<td>The focal points and inconveniences of the existing techniques are comprehensively analyzed against the proposed research dimension taxonomy namely: resource adaption policy, objective function, allocation method, allocation operation, and interoperability.</td>
</tr>
</tbody>
</table>

CONCLUSION
In this work, it has been concluded that task migration is the major issue of cloud computing. In the network, hosts assign tasks to the virtual machines through the brokers. The virtual machines are selected from the various machines on the basis of their available resources. The available resources are assessed by the broker. The virtual machine which is capable to execute the task is selected as the best machine for task execution. In the network, some uncertainties may arise which overload the virtual machines. In this work, technique will be proposed which re-allocate the tasks of the overloaded virtual machines. The task reallocation will be based on bio-inspired greedy techniques. The proposed improvement leads to reduce energy consumption and reduce execution time of the network.

References:
[12] Amit Nathani, Sanjay Chaudharya, Gaurav Somani, “Policy based resource allocation in IaaS cloud”, 2011 Elsevier B.V. All rights reserved