International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-1, June 2013

Comparison of Load Balancing and Scheduling Algorithms in Cloud Environment

Karthika M.T., Neethu Kurian, Mariya Seby

Abstract*— The importance of cloud computing is increasing nowadays. Cloud computing is used for the delivery of hosted services like reliable, fault tolerant and scalable infrastructure over Internet. A variety of algorithms is used in the cloud environment for scheduling and load balancing, thereby reducing the total cost. The main algorithms usually used include, optimal cloud resource provisioning (OCRP) algorithm and hybrid cloud optimized cost(HCOC)scheduling algorithm These algorithms will formulate the optimized cost of resources in the cloud environment.

***Index Terms*— Cloud computing, load balancing, scheduling**

I. INTRODUCTION

In cloud computing, a resource provisioning mechanism is needed to supply cloud consumers a set of varying resources for processing the jobs and storing the data. Cloud providers can give cloud consumers two resource provisioning plans. They are: a short term plan which is on-demand and a long-term reservation plan.

The pricing in on-demand plan is based on pay-per-use mode. For reservation plan, pricing is charged by a onetime fee (e.g., 1 year) usually before the computing resource will be utilized by cloud consumer. With the reservation plan, the cloud consumers reserve the resources in advance. As a result, the under provisioning problem can occur when the reserved resources are unable to fully meet the demand because of the uncertainty. On the other hand, the over provisioning problem can occur if the reserved resources are more than the actual demand in which part of a resource pool will be underutilized.

It is important for the cloud consumer to minimize the total cost of resource provisioning by reducing the on-demand cost and oversubscribed cost of under provisioning and over provisioning. To achieve this goal, the optimal computing resource management is the critical issue.

In particular, an optimal cloud resource provisioning (OCRP) algorithm is used to minimize the total cost for provisioning resources in a certain time period. In order to make an optimal solution, the demand uncertainty from cloud consumer side and price uncertainty from cloud providers are considered to adjust the tradeoff between on-demand and oversubscribed costs. In this paper, an optimal cloud resource provisioning (OCRP) algorithm is proposed to minimize the total cost for provisioning resources in a certain time period.

Manuscript published on 30 June 2013.
***Correspondence Author(s)
Karthika M T, Computer Science and Engineering, Vedavyasa Institute of Technology, India.
Neethu Kurian, Computer Science and Engineering, Vedavyasa Institute of Technology, India.
Mariya Seby, Computer Science and Engineering, Vedavyasa Institute of Technology, India.

Bender's Decomposition [1] method is used to obtain the solution of OCRP.

The HCOC algorithm will give the resources from public cloud to private cloud with reduced cost.

II. RELATED WORKS

Until recently, most of the scientific tasks were run on clusters and grids, and many works explored how to optimize the performance of scientific applications in such specific contexts. However, cloud is not a completely new concept with respect to grids, it indeed has intricate connection to the grid computing paradigm and other technologies such as utility and cluster computing, as well as with distributed systems in general. Jiyayin Li [2] proposed a resource optimization mechanism with preemptable task execution. In Infrastructure-as-a-Service (IaaS) cloud computing, an organization can outsource various resources including networking components, hardware and storage to remote users. Usually a cloud user can request multiple cloud services simultaneously. In this case, performance can be improved by parallel processing. The optimal virtual machine placement algorithm (OVMP) [3] that can yield the optimal solution for both resource provisioning and VM placement in two provisioning stages. Traditional data centers offer many services hosted on dedicated physical servers, which the resources are often under-utilized. Virtualization helps these data centers to maintain their services onto a lesser number of physical servers than originally required. There may be a number of such services running in a single data center. In traditional data centers, applications are tied to specific physical servers that are often overprovisioned in order to serve the large resource requirements of enterprise services and in the hope of handling unexpected surges in resource demands. As a consequence, the level of resource utilization on any server is typically very low. This needs high operational costs and large investments, not to mention wasted power and floor space and a significant management overhead. Such a situation can be called as a 'server sprawl', in which there are many underutilized servers taking up large space and energy than can be justified by their workloads.

The OVMP algorithm makes a decision based on the optimal solution of stochastic integer programming (SIP) to lease resources from cloud providers. It views the provisioning of resources in three phases. They are: reservation, utilization and on-demand. The algorithm make use of stochastic integer programming and works in two stages: the first stage calculates the number of provisioned VMs (demand) in the reservation phase and the second stage calculates the number of allocated VMs in both utilization and on-demand phases.

It considers all the possible demands and prices, called realizations. Using these, an objective function is developed considering the cost of resources charged by each provider in different phase.

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

Then the probability of each demand will be realized and the probability of each set of resource prices offered by each provider being realized. Decision variables are X(p)ij the number of VMs in class Vi to be allocated to provider Pj in phase p. Although this approach considers all possible demands and prices,that it will not be scalable once the number of possible demands and prices goes high which happen in a real world scenario only. Also, it ultimately formulates the problem as an integer linear program which is proved to be NP-hard. So, this approach will not work with problems of large sizes.

The performance of OVMP algorithm can be evaluated by numerical studies and simulation. The evaluation results show that the OVMP algorithm can minimize users' expenditure. This algorithm can be applied to provide resources in cloud computing environments.

The architecture of a fault-tolerant, novel scalable, and distributed consolidation manager called Snooze that can dynamically consolidate the workload and conserve energy and reduce the operating costs. It performs dynamic consolidation depends on constraint programming and takes migration overhead into account.

III. FORMULATION OF OCRP

The architecture of cloud computing environment has four main components. They are: cloud consumer, cloud providers, virtual machine (VM) repository, and cloud broker. The cloud consumer want to execute jobs. The cloud provider has to provision the resources before the jobs are executed. In order to obtain such resources, the consumer has to creates VMs integrated with software required for the jobs. The created VMs are then stored in the VM repository. Then, the newly created VMs can be hosted on cloud providers' infrastructures from where the resources can be utilized by the VMs. The cloud broker can allocate the VMs originally stored in the VM repository to appropriate cloud providers.

In OCRP, there are multiple VM classes used to classify different types of VM. Let $I \in IN_1$ denote the set of VM classes where IN1=$\{1,2,3 \ldots\}$.It is assumed that one VM class represents a distinct type of jobs (e.g., one class for web application and the other for database application). Let $J \in IN_1$ denote the set of cloud providers. Each cloud provider supplies a set of resources to the consumer. Let R denote the set of resource types which can be provided by cloud providers. Resource types can be storage (in unit of GBs/ month), computing power (in unit of CPU-hours), and bandwidth for Internet data transfer (in unit of GBs/month). The VM class specifies the total amount of resources in each resource type. Let b_i^r be the amount of resource type r required by the VM in class $i \in I$.

There are three provisioning phases in cloud environment: reservation, expending, and on-demand phases. In the reservation phase, the cloud broker provisions resources with reservation plan in advance without knowing the consumer's actual demand. In the expending phase, is one in which the demand and price of the resources are determined, and the reserved resources can be used. So, the reserved resources could be observed as either over provisioned or under provisioned. If the demand is greater than the amount of resources reserved, the broker can pay for additional resources with on-demand plan, and the on-demand phase starts then.

A provisioning stage can be defined as the time epoch when the cloud broker is ready to provision resources by understanding reservation and/or on-demand plans, and also allocates VMs to different cloud providers for utilizing the resources that are provisioned. Therefore, each provisioning stage can consist of one or more provisioning phases. Let $T \in$ IN1 denote the set of all provisioning stages where $T \geq 2$. Also $K \in$ IN1 denote the set of all reservation contracts which are offered by cloud providers.

The optimal solution used by the cloud broker is obtained from the OCRP algorithm based on stochastic integer programming [4]. Stochastic programming takes a set of uncertainty parameters , described by a probability $p(w)$.

Let $c_{jkr}^{(R)}$ denote the unit price (i.e., costs to the consumer) of resource type r subscribed to reservation contract k provided by cloud provider j in reservation phase of the first provisioning stage. Usually the price of reservation plan in the first stage is charged by a fixed one-time fee. The reservation cost (c_{ijk}^{R}) is the cost for provisioning every resource type defined as follows:

$$c_{ijk}^{(R)} = \sum b_{ir} C_{jkr}$$

Let $c_{ijkt}^{(r)}$, $c_{ijkt}^{(e)}$ and $c_{ijt}^{(o)}$ denote the reservation ,expending and on-demand costs for provisioning resource type r with reservation contract k provided by cloud provider j, respectively. The parameter Early Start Time can be used to give priorities to the cloud consumers.

The stochastic programming with multistage recourse is presented as the core formulation of the OCRP algorithm. The Benders decomposition algorithm [5] is applied to solve the formulated stochastic programming problem.

Minimize:

$$Z_v^{(r)} = \sum\sum\sum c_{ijk}^{(R)} + \sum\sum\sum\sum p(w)\, c_{ijkt}^{(r)}$$

$$Z_v^{(o)} = \sum\sum\sum p(w)\, c_{ijt}^{(o)}$$

$$Z_v^{(e)} = \sum\sum\sum\sum p(w)\, c_{ijkt}^{(e)}$$

The optimal objective function value Z will give the minimum cost.

IV. CONCLUSION

The OCRP and HCOC algorithm can provision computing resources for short term as well as a long-term plan, e.g., twelve provisioning stages in a yearly plan. The demand and price uncertainty of resources are considered in OCRP. Numerical studies shown that with the OCRP and HCOC algorithm, total cost of resource provisioning in cloud computing environments can be minimized.

The optimal solution obtained from OCRP is obtained by formulating and solving stochastic integer programming with multistage recourse. We have applied the Benders decomposition approach to divide an OCRP problem into sub problems which can be solved parallelly. Both the algorithms can be used as a resource provisioning tool for the emerging cloud computing market in which the tool can effectively save the total cost.

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

V. ACKNOWLEDGMENT

This work has done in the Department of Computer Science & Engineering of Vedavyasa Institute of Technology, Calicut University, India under the guidance of Kavitha Murugesan (HOD Computer Science & Engineering, Vedavyasa Institute of Technology, India).

REFERENCES

1. A.J. Conejo, E. Castillo, and R. Garcı´a-Bertrand, "*Linear Programming: Complicating Variables*," Decomposition Techniques in Mathematical Programming, chapter 3, pp. 107-139, Springer, 2006.
2. Jiayin Li a, Meikang Qiu a, Zhong Mingb, Gang Quanc, Xiao Qin d, Zonghua Gue, "*Online optimization for scheduling preemptable tasks on IaaS cloud systems"*..
3. S. Chaisiri, B.S. Lee, and D. Niyato, "*Optimal Virtual Machine Placement across Multiple Cloud Providers*," Proc. IEEE Asia-Pacific Services Computing Conf.2009.
4. Amazon EC2 Reserved Instances, http://aws.amazon.com/ec2/ reserved-instances, 2012.
5. F. Hermenier, X. Lorca, and J.-M. Menaud, "*Entropy: A Consolidation Manager for Clusters*," Proc. ACM SIGPLAN/ SIGOPS Int'l Conf. Virtual Execution Environments (VEE '09), 2009.
6. D. Kusic and N. Kandasamy, "*Risk-Aware Limited Lookahead Control for Dynamic Resource Provisioning in Enterprise Computing Systems*," Proc. IEEE Int'l Conf. Autonomic Computing, 2006.M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
7. F.V. Louveaux, "*Stochastic Integer Programming*," Handbooks in OR & MS, vol. 10, pp. 213-266, 2005.

AUTHOR PROFILE

Karthika M.T, student of Vedavyasa Institute of Technology, India ,pursuning M.tech(CSE). She has received B Tech(CSE) from Sree Buddha College of Engineering ,India in 2011.

Neethu Kurian, student of Vedavyasa Institute of Technology, India ,pursuning M.tech(CSE). She has received B Tech(CSE) from College of Engineering Poonjar ,India in 2011.

Mariya Seby, student of Vedavyasa Institute of Technology, India ,pursuning M.tech(CSE). She has received B Tech(CSE) from Baselios Mathews II College of Engineering ,India in 2011.