INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Economic Resource Provisioning Techniques Using OCRP Algorithm In Cloud

M.Antro Monica Sanjas
Department of CSE, Srinivasan engineering college, Perambalur, India
C.Subramanian
Department of CSE, Srinivasan engineering college, Perambalur, India

Abstract:

The cloud computing provides users an efficient way to allocate the computing resources to the cloud consumers by providing them two resource provisioning plan namely, reservation and on-demand plan. The resource provisioned by reservation plan is cheaper than that of the on-demand plan but it is difficult to achieve the best advance reservation of resource due to uncertainty of consumer's future demand and provider's resource prices. To solve this problem an Enhanced optimal cloud resource Provisioning algorithm is proposed by formulating a stochastic programming model. The Enhanced Optimal Cloud Resource Provisioning algorithm can provide Computing resources for being used in multiple provisioning stages as well as a long-term plan. The demand and price uncertainty is considered in Enhanced Optimal Cloud Resource Provisioning. Here, different approaches are considered to obtain the solution of the Optimal Cloud Resource Provisioning algorithm namely, deterministic equivalent formulation, sample-average approximation and Benders decomposition.

Keywords: *Cloud computing, computing resources, resource provisioning, virtual machine, provisioning stages, provisioning plans, stochastic programming.*

1.Introduction

Cloud computing is the use of computing resource like hardware, software, processing power, storage, network bandwidth that are delivered as a public utility service over a network in a cloud-shaped symbol as an abstraction for the complex infrastructure. Infrastructure as a service is the most basic cloud service model, providers offer computers, as physical or more often as virtual machines, and other resources. The virtual machines are run as guests by hypervisor. Management of pools of hypervisors by the cloud operational support system leads to the ability to scale to support a large number of virtual machines. Infrastructure as a service is the most basic cloud service model, providers offer computers, as physical or more often as virtual machines, and other resources. The virtual machines are run as guests by hypervisor. Management of pools of hypervisors by the cloud operational support system leads to the ability to scale to support a large number of virtual machines. In the software as a service model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. The cloud users do not manage the cloud infrastructure and platform on which the application is running. This eliminates the need to install and run the application on the cloud user's own computers simplifying maintenance and support.

In cloud computing, two resource provisioning plans is used to provide the cloud consumers a set of computing resources for processing the jobs and storing the data namely, on-demand and reservation plans. Reservation plan is the long-term plan where the plan is subscribed in advance (eg:1 or 3 years) to reduce the provisioning cost because for reservation plan, pricing is charged by a onetime fee (e.g., 1 year) typically before the computing resource will be utilized by cloud consumer. With the reservation plan, the price to utilize resources is cheaper than that of the on-demand plan. The On-demand plan is the short-term plan, since it can be purchased for a short period of time when resource reserved by the reservation plan is insufficient. Pricing in on-demand plan is charged by pay-per-use basis (e.g., 1 day). Therefore, purchasing this on-demand plan, the consumers can dynamically provision resources at the moment when the resources are needed to fit the fluctuated and unpredictable demands.

The under provisioning problem occur due to the uncertainty, when the reserved resources cannot meet the demand. To solve this problem, the extra demand can be satisfied by providing the resources in on-demand plan. But the price will be expensive when the resources are provisioned with on-demand plan. The over provisioning problem occur, when the reserved resources are more than the actual demand. By reducing the on-demand cost and oversubscribed cost, the cloud consumers can reduce the total cost of the resource provisioning.

2.Related Work

Here we discuss about the existing methods for minimizing the under provisioning and over provisioning problems due to the uncertainty of future demands and resource prices.

2.1.Profile-Based Approach

It is used to capture expert's knowledge of scaling different types of applications. The profile-based approach automates the deployment and scaling of applications in cloud. Just-in-time scalability is achieved without binding to specific cloud infrastructure. A profile is an embodiment of knowledge and best practice of a commonly adopted computing environment with inherent just-in-time scalability. The main purpose of a profile is to automate the instantiation of a computing environment that inherits the just-in-time scalability from the profile by straightforward configuration. Therefore a profile declares what constitutes a computing environment and how a computing environment is configured to enable just-in-time scalability.

The component definition of profile defines the software components involved in computing environments that are created by a profile. There are two components namely, virtual and physical component.

Each virtual component is defined by an image file and the configuration file for the image. The image records the execution environment, including the operating system, the minimal and maximal memory and the hard disk capacity. The configuration file is used to automate the provisioning of a virtual image, including which virtual technology the image is based, where the image is located, the host operation system for the image.

Physical component is defined as a URL and a configuration file. The URL denotes the place where all required software is, the configuration file specifies the execution environment that is required for installing software in a physical machine. Application integration defines how to deploy application packages so that the application could leverage the just-in-time scalability offered by the scalable topology. It just defines only the just-in-time scalability, it does not consider the uncertainty of future consumer demands.

2.2.Optimization Framework

For resource provisioning, an optimization framework was developed. The multiple client QOS classes are considered under uncertainty of workloads i.e., demands of computing resources. By using the online forecasting techniques, the arrival pattern of workloads are estimated.

This existing approach accounts for the switching costs incurred during resource provisioning and explicitly encode risk in the optimization problem. The provisioning problem of interest is to decide an optimal allocation of computing resources to multiple client QoS classes under a dynamic workload, this discrete optimization problem may need continuous re-solving with observed environmental events such as time-varying client workload patterns and computer failures and cannot be applied directly and a closed-form expression for a feedback control map cannot be established.

2.3.K-Nearest-Neighbors Algorithm

The K-nearest-neighbors algorithm was developed to predict the demand of resources. Here the prediction of demands was performed to define the reservation prices. The KNN algorithm uses lightweight monitoring of essential system and application metrics in order to decide how many databases it should allocate to a given workload. The pro-active algorithm also incorporates awareness of system stabilization periods after adaptation in order to improve prediction accuracy and avoid system oscillations.

It compares the pro-active self configuring scheme for scaling the database tier with a reactive scheme. The experiments using the industry-standard TPC-We-commerce benchmark demonstrate that the pro-active scheme is effective in reducing both the frequency and peak level of SLA violations compared to the reactive scheme. Furthermore, by augmenting the pro-active approach with awareness and tracking of system stabilization periods induced by adaptation in our replicated system, we effectively avoid oscillations in resource allocation.

Due to the excessive personnel costs involved in managing these complex systems and dynamic content workload in response to load or failure-induced SLA violations.

2.4.SLA-Aware Virtual Resource Management

A key challenge for cloud providers is to automate the management of virtual servers while taking into accounts both high-level QoS requirements of hosted applications and resource management costs. It proposes an autonomic resource manager to control the virtualized environment which decouples the provisioning of resources from the dynamic placement of virtual machines. This manager aims to optimize a global utility function which integrates both the degree of SLA fulfillment and the operating costs. We resort to a Constraint Programming approach to formulate and solve the optimization problem. Results obtained through simulations validate our approach.

As virtualization is a core technology of cloud computing, the problem of virtual machine placement (VM placement) becomes crucial. A resource management consisting of resource provisioning and VM placement was used. It does not consider the uncertainty of future demands and prices and it also does not guarantee the optimal solution.

2.5.Optimal Cloud Resource Provisioning Algorithm (OCRP)

In OCRP algorithm, first the problem is generalized into the multiple stage formulation. Second, the different approaches to obtain the solution of computing resource provisioning are considered and finally, the performance evaluation is extended to consider various realistic scenarios. This OCRP algorithm is proposed to minimize the total cost for provisioning resources in a certain time period. To make an optimal decision, the demand uncertainties from cloud providers are taken into account to adjust the tradeoff between on-demand and oversubscribed costs. This optimal decision is obtained by formulating and solving a stochastic integer programming problem with multistage recourse. Benders decomposition and sample-average approximation are also discussed as the possible techniques to solve the OCRP algorithm. Here an extensive numerical studies and simulations are performed, and the results show that OCRP can minimize the total cost under uncertainty.

The major contributions of the OCRP algorithm lie in the mathematical analysis which is summarized as: The optimal cloud resource provisioning algorithm is proposed for the virtual machine management. The optimization formulation of stochastic integer programming is proposed to obtain the decision of the OCRP algorithm as such the total cost of resource provisioning in cloud computing environment is minimized. The formulation considers multiple provisioning stages with demand and price uncertainties. The solution methods based on benders decomposition and sample-average approximation algorithms are used to solve the optimization formulation in an efficient way. The performance evaluation is performed which can reveal the importance of optimal computing resource provisioning. The performance comparison among the

OCRP algorithm and the other approaches is also presented. The proposed mathematical analysis will be useful to the cloud consumers for the management of virtual machines in cloud computing environment. The proposed OCRP algorithm will facilitate the adoption of cloud computing of the users as it can reduce the cost of using computing resource significantly.

3.Problem Definition

The proposed system aims to reduce the resource provisioning cost in cloud by providing two provisioning plans for computing the resources, namely reservation plan and on-demand plans. In the reservation plan, the cloud consumer will reserve the resources in advance without knowing their future demand. Here the reservation plan is considered as the long-term planning since the resources are reserved for a long time use by the cloud consumer in future. When the demand exceeds the amount of reserved resources the broker can pay for additional resources with on-demand plan. This on-demand plan is considered as the short-term planning since the resource can be purchased for a short period of time when the resource reserved by the reservation plan are insufficient.

4.Algorithms

4.1.Stochastic Integer Programming

The stochastic programming is the framework for modeling optimization problems involving uncertainty. The stochastic programming with multistage recourse is presented as the core formulation of the Enhanced Optimal Cloud Resource Provisioning algorithm. First, the original form of stochastic integer programming formulation is derived. Then, the formulation is transformed into the deterministic equivalent formulation (DEF) which can be solved by traditional optimization solver software.

The general form of stochastic integer program of the Enhanced Optimal Cloud Resource Provisioning algorithm is formulated:

$$z = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk}^{(R)} x_{ijk}^{(R)} + \mathrm{IE}_{\Omega}[Q\left(x_{ijk}^{(R)}, \omega\right)]$$

Subject to:

$$x_{ijk}^{(R)} \in \mathrm{IN}_0, \qquad \forall i \in I, \forall j \in J, \forall k \in K$$

The objective function of stochastic integer programming is to minimize the cloud consumer's total provisioning cost.

4.2.Deterministic Equivalent Formulation

Here the probability distribution of all scenarios in set Ω, and formulations are transformed into the deterministic integer programming called deterministic equivalent formulation. Here to solve the deterministic integer programming, probability distribution of both price and demand must be available.

$$
\begin{aligned}
Z_{\Omega} = & \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} c_{ijk}^{(R)} x_{ijk}^{(R)} \\
& + \sum_{\omega\in\Omega}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\sum_{t\in T_k} p(\omega) c_{ijkt}^{(R)}(\omega) x_{ijkt}^{(r)}(\omega) \\
& + \sum_{\omega\in\Omega}\sum_{i\in I}\sum_{j\in J}\sum_{t\in T} p(\omega)(\sum_{k\in K} c_{ijkt}^{(e)}(\omega) x_{ijkt}^{(e)}(\omega) + c_{ijt}^{(o)}(\omega) x_{ijt}^{(o)}(\omega)
\end{aligned}
$$

4.3.Benders Decomposition

The benders decomposition algorithm is applied to solve the stochastic programming problem formulated. The goal of this benders decomposition algorithm is break down the optimization problem into multiple smaller problems which can be solved independently and parallely. As a result, the time to obtain the solution of the OCRP algorithm can be reduced. The benders decomposition algorithm can decompose integer programming problems with complicating variables into two major problems: master problem and sub-problem.

4.4.Sample-Average Approximation

When the number of scenarios is numerous, it may not be efficient to obtain the solution of the OCRP algorithm by solving the stochastic programming formulation, if all scenarios in the problem are considered. So to address this complexity issues, the sample-average approximation (SAA) approach is applied.

This approach selects a set of scenarios, e.g., N scenarios, where N is smaller than the total number of scenarios |Ω|. Then, these N scenarios can be solved in a deterministic equivalent formulation. The optimal solution can be obtained if N is large enough which can be verified numerically.

5.Conclusion

From the existing systems the problems like uncertainties of provider's resource price and consumer's future demand is addressed using the Enhanced OCRP algorithm and the price of the resources provisioned by the providers is reduced by using the optimal pricing scheme. The performance evaluation has been performed by simulations and numerical studies. These algorithms optimally adjust the tradeoff between allocation of on-demand and reservation of resources. These algorithms can minimize the total cost of resource provisioning in cloud computing environments.

6.Acknowledgment

I take this chance to express my deep sense of gratitude to our Management, our principal Dr. B. Karthikeyan , for providing an excellent infrastructure and support to pursue this research work at our college. I express my profound thanks to Head of the department Prof. J. Mercy Geraldine M.E.,(Ph.D) for her administration and keen interest, which motivated me along the course as well as research work, and also thank to all staff members.

7.Reference

1. F.V. Louveaux, “Stochastic Integer Programming,” Handbooks in OR & MS, vol. 10, pp. 213-266, 2003
2. Y.Jie, Q.lie, and L.Ying, “A Profile-Based Approach to Just-in-Time Scalability for Cloud Apllications,” Proc. IEEE Int’l Conf. Cloud Computing (CLOUD ’09), 2009
3. Y. Kee and C. Kesselman, “Grid Resource Abstraction, Virtualiza-tion, and Provisioning for Time-Target Applications,” Proc. IEEE Int’l Symp. Cluster Computing and the Grid, 2008
4. D. Kusic and N. Kandasamy, “Risk-Aware Limited Lookahead Control for Dynamic Resource Provisioning in Enterprise Com-puting Systems,” Proc. IEEE Int’l Conf. Autonomic Computing, 2006
5. K. Miyashita, K. Masuda, and F. Higashitani, “Coordinating Service Allocation through Flexible Reservation,” IEEE Trans. Services Computing, vol. 1, no. 2, pp. 117-128, Apr.-June 2008
6. J. Chen, G. Soundararajan, and C. Amza, “Autonomic Provision-ing of Backend Databases in Dynamic Content Web Servers,” Proc. IEEE Int’l Conf. Autonomic Computing, 2006
7. L. Grit, D. Irwin, A. Yumerefendi, and J. Chase, “Virtual Machine Hosting for Networked Clusters: Building the Foundations for Autonomic Orchestration,” Proc. IEEE Int’l Workshop Virtualization Technology in Distributed Computing, 2006
8. H.N. Van, F.D. Tran, and J.-M. Menaud, “SLA-Aware Virtual Resource Management for Cloud Infrastructures,” Proc. IEEE Ninth Int’l Conf. Computer and Information Technology, 2009
9. F.Hermenier, X,Lorca, and J-M.Menaud, ”Entropy: a Consolidation Manager for Clusters,” Proc ACM SIGPLAN/SIGOPS Inf’l Conf. Virtual Execution Enviroments (NEE ’09), 2009
10. N.Bofroff, A.Kochut, and K.Beaty, “Dynamic Placement of Virtual Machines for Managing SLA Violations,” Proc. IFIP/IEEE Int’l Symp. Integrated Network Management (IM ’07), pp.119-128, May 2007
11. Sivadon Chaisiri, Bu-Sung Lee, Dusit Niyato , ”Optimization of Resource Provisioning Cost in Cloud Computing,” IEEE Transactions on Services Computing, VOL.5, NO.2, April-June 2012