

A review on various optimization techniques of resource provisioning in cloud computing

K. Sumalatha¹, M. S. Anbarasi²

¹Department of Computer Science and Engineering, Annamacharaya Institute of Technology and Sciences, India

²Department of Information Technology, Pondicherry Engineering College, India

Article Info

Article history:

Received Jan 4, 2018

Revised Jul 17, 2018

Accepted Aug 2, 2018

Keywords:

IaaS

Optimization techniques

Resource provisioning

Uncertainty

Virtual machines

ABSTRACT

Cloud computing is the provision of IT resources (IaaS) on-demand using a pay as you go model over the internet. It is a broad and deep platform that helps customers build sophisticated, scalable applications. To get the full benefits, research on a wide range of topics is needed. While resource over-provisioning can cost users more than necessary, resource under provisioning hurts the application performance. The cost effectiveness of cloud computing highly depends on how well the customer can optimize the cost of renting resources (VMs) from cloud providers. The issue of resource provisioning optimization from cloud-consumer potential is a complicated optimization issue, which includes much uncertainty parameters. There is a much research avenue available for solving this problem as it is in the real-world. Here, in this paper we provide details about various optimization techniques for resource provisioning.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

K. Sumalatha,

Department of Computer Science and Engineering,

Annamacharaya Institute of Technology and Sciences,

Tirupati, India.

Email: sumalathacs@yahoo.com

1. INTRODUCTION

Cloud Computing is a broad and deep platform that helps customers build sophisticated scalable applications. Cloud provisioning is the allocation of a cloud provider's resources to a customer [1]. When a cloud provider accepts a request from a customer, it must create the appropriate number of virtual machines (VMs) [2] and allocate resources to support them. In this context, the term provisioning simply means "to provide".

The Provisioning has been done in several different ways. a) Advance Provisioning: The customer requests the provider for services and the provider prepares the appropriate resources [2] in advance. The customer is charged a flat fee or is billed on a monthly basis. b) Dynamic Provisioning: The provider allocates more resources when they are needed and removes them [3] when they are not needed. The customer is billed on a pay-per-use basis.

Provisioning allows optimal allocation of resources to consumers in a finite time to achieve desired quality of service. Here, the problem is either the user gets over-provisioning or under-provisioning. Formally, provisioning problem involves uncertainty parameters while choosing resources subject to some constraints to optimize some objective function. The aim is to develop an optimized method that reduces over provisioning and under provisioning problems. It has remained a topic of research in various fields for decades, may it be supply of electricity or water to consumers [4].

In recent years, distributed computing paradigm [5] has gained much attention due to high scalability, reliability, and flexibility. Thus stochastic based techniques deal with these problems by

providing near optimal solutions within the reasonable amount of time. In this paper, we present a review of various optimization techniques based on various parameters.

2. RESOURCE PROVISIONING PRICING MODELS

Reserved instances are cheapest resources. Price is based on the period of subscription (static). Customer must reserve the resources in advance. Customer might overpay for the resources reserved if he/she doesn't use them extensively and he might under pay for the resources reserved for long time [6].

On-demand instances are the highest priced resources. Price is set by the service provider and remains constant. Consumer has to pay per use. Customer is aware of the exact price to be paid. Resources are reserved for the customer for the paid period of time [7]. Service provider might reserve the resources for longer than the customer's utilized. Service provider cannot raise the price when demand is high; when demand is low, the user pays higher than the market price.

Spot Instances [8] allow you to specify the maximum hourly price that you are willing to pay to run a particular instance type, usually lower than the On-Demand rate. These are suitable for both customers and the service provider because the price is set according to the level of supply and demand. Less scalability of high demand in the market than fixed pricing, the spot instances are the unused on-demand instances.

The Spot Price fluctuates based on supply and demand for instances, but customers will never pay more than the maximum price they have specified. If the Spot Price moves higher than a customer's maximum price, the customer's instance will be shut down by the cloud provider [9]. Figure 1 shows three pricing models.



Figure 1. Pricing models

3. OPTIMIZATION TECHNIQUES

In Computer Science and operation research, Optimization is referred to as the selection of best element from a set of alternatives with regard to some criteria. Two types of entities involved in cloud are cloud service provider and cloud consumer. Cloud service providers provision their resources on rental basis to cloud consumers and cloud consumers submit their requests for provisioning the resources. They both have their own motivations when they become part of cloud environment. Consumers are concerned with the performance of their applications, whereas providers are more interested in efficient utilization of their resources. Thus these Optimization Techniques [10] can be classified into two types: Static techniques and Dynamic techniques. Following are some of the optimization criteria followed while provisioning resources in cloud environment as shown in Figure 2.

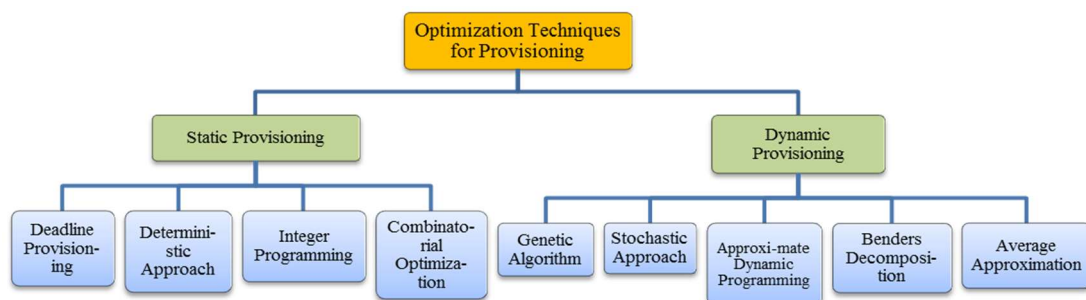


Figure 2. Optimization techniques

3.1. Static provisioning

Cloud service provider allocates resources to consumer in advance. That is cloud consumer has to decide how much capacity of resources he/she wants statically means before using them. The deterministic approach, integer programming, linear programming and deadline provisioning algorithm show how resources are allocated statically for deadline based workflows. Here we consider only the expected values for all the parameters. In static provisioning we may get over provisioning or under provisioning problem.

3.2. Dynamic provisioning

Cloud service provider allocates resources to consumer as needed and when required and they have to pay per use. This is also called as pay as you go model. Here, genetic algorithm, stochastic approach, approximate dynamic programming, benders decomposition and average approximation are used for provisioning resources dynamically.

3.2.1. Genetic algorithm

Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. Genetic algorithms do not scale well with complexity, operating on dynamic data sets is difficult.

3.2.2. Stochastic approach

In this section, the stochastic programming with multistage recourse is presented as the core formulation. 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.

3.2.3. Approximate dynamic programming

Approximate Dynamic Programming (ADP) is a powerful technique to solve large scale discrete time multistage stochastic control processes, i.e., complex Markov Decision Processes (MDPs). These processes consists of a state space S , and at each time step t , the system is in a particular state $St \in S$ from which we can take a decision xt from the feasible set Xt . This decision results in rewards or costs, typically given by $Ct(St, xt)$, and brings us to a new state $St+1$ with probability $P(St+1|St, xt)$, i.e., the next state is conditionally independent of all previous states and actions. Therefore, the decision not only determines the direct costs, but also the environment within which future decisions take place, and hence influences the future costs. The goal is to find a policy. A policy $\pi \in \Pi$ can be seen as a decision function $X\pi(St)$ that returns a decision $xt \in Xt$ for all states $St \in S$, with Π being the set of potential decision functions or policies. The problem of finding the best policy can be written as (1).

$$\min_{\pi \in \Pi} E^{\pi} \{ \sum_{t=0}^{T-1} \gamma^t C_t(St, X\pi_t(St)) \} \quad (1)$$

where γ is a discount factor, and T denotes the planning horizon, which could be infinite.

3.2.4. Benders decomposition

The goal of this algorithm is to break down the optimization problem into multiple smaller problems which can be solved independently and parallel. The Benders decomposition algorithm can decompose integer programming problems with complicating variables into two major problems: master problem and sub problem.

3.2.5. Sample average approximation

It may not be efficient to get the solution of the algorithm if the number of scenarios is more by solving the stochastic programming formulation defined in directly if all scenarios in the problem are considered. The sample-average approximation (SAA) approach is used to address this complex problem. This approach is applied on a set of scenarios.

4. COMPARISON OF RESOURCE PROVISIONING TECHNIQUES

In [11], proposed an algorithm called optimal cloud resource provisioning algorithm (OCRP) to overcome resource under provisioning and over provisioning. Authors applied various optimization techniques to minimize the user's cost. This approach includes an algorithm called optimal cloud resource provisioning algorithm (OCRP) to overcome resource under provisioning and over provisioning. Authors applied various optimization techniques to minimize the user's cost with more scenarios. Minimizing both

under -provisioning and over provisioning problems under the demand and price uncertainty in cloud computing environments is our motivation to explore a resource provisioning strategy for cloud consumers. The following Table 1 shows proposed optimization techniques and their merits and demerits.

Table 1. Comparison of Various Optimized Provisioning Techniques based on Uncertainty, QoS, Time, Cost, Heterogeneity and SLA violations

Parameter	Problem Proposed	Optimization Technique Applied	Inputs	Merits	Challenges
Cost & QoS	[12] Resource provisioning and scheduling on IaaS Cloud	Particle Swarm Optimization (PSO) & Meta Heuristic Optimization.	Workflows	Well suited for work-flows with smaller size	It Takes More Cost to meet the deadlines of larger workflows.
	[13] Resource provisioning for adaptive applications in Grid	Integer Programming	Bandwidth, CPU cycles	Maximize Revenue	Implement on cloud environment
	[14] Autonomic Resource Provisioning	Match Making Technique	User requests	Provides QoS	More Energy consumption
	[15] Dynamic Resource Provisioning for Data Streaming Applications	CPU Allocation Algorithm	CPU Cycles	Reduce the processing time and budget	Need an implementation on real scenario
	Optimization of Resource Provisioning Cost	OCR algorithm with Stochastic Optimization	Resource requests	Minimize the cost & avoid over & under provisioning	More Scenarios
	[16] Optimization Approach for Resource Allocation for IoT	Combinatorial auction approach	Network, Data-base	QoS & Maximize profit	Not suitable for all users
	[17] Joint Optimization of Resource Provisioning	Stochastic Optimization, Sensitivity Analysis	Bandwidth & Virtual Machines	Reduce over and Under provisioning	Need to consider network delay and VM migration.
	[18] Joint Virtual Machine and Bandwidth Allocation in Software Defined Network (SDN)	Deterministic formulation	Bandwidth & Virtual Machines	Minimize user's provision cost	Applicable for only Network intensive applications
Heuristics & Time	Resource provisioning and scheduling on IaaS Cloud [19]	Particle Swarm Optimization (PSO) & Meta Heuristic Optimization.	Workflows	Well suited for work-flows with smaller size	It Takes More Cost to meet the deadlines of larger workflows.
	[19] Adaptive Resource Provisioning	Static Algorithm based on Heuristics	Workflows	Low cost for execution of each task.	Not suitable for complete workflow structure
	[20] SLA-Based Resource Provisioning for Hosted Software-as-a-Service Applications	Customer driven heuristic algorithms	Customer profile and response time	Less cost & reduce SLA violations	Applicable only for SaaS providers.
	[21] Dynamic Provisioning in Hybrid Cloud using Aneka Platform	Map Reduce model, Deadline Priority Provisioning algorithm	User requests,	Reduces total execution time	Applicable only for hybrid clouds.
	[22] Application of Selective Algorithm for Effective Resource Provisioning	min-min or max-min algorithm	File size of user requests	Minimize make span	Dependencies among tasks can be modeled
	[23] An Efficient Architecture and Algorithm for Resource Provisioning in Fog Computing	Efficient resource allocation (ERA) Algorithm	Use Requests for resources	Reduces the response time	Applicable for only reserved instances
	[24] Towards Understanding Uncertainty in Resource Provisioning	stochastic scheduling	Sources of uncertainty	Introduced various solutions to resolve uncertainty	Need implementation on real Scenarios

5. CONCLUSION AND FUTURE WORK

The ultimate goal of cloud computing is to satisfy both cloud consumer and cloud service provider. Broad research is needed to find best optimal provisioning technique. This paper presents various optimized provisioning variants with their merits and challenges while considering different parameters and different algorithms. The efficient dynamic optimization provisioning is one of the primary challenges in cloud environment, because a tradeoff between professional SLA [25] and QoS constraint like max resource

utilization, cost etc. In future, we will propose a method for Dynamic optimized resource provisioning. It has dynamic nature so that the cost of provisioning resources may be reduced without SLA violations.

REFERENCES

- [1] www.buyya.com/MasteringClouds/ToC-Preface-TMH.pdf.
- [2] Taskeen Zaidi, "Rampratap Rampratap.Virtual Machine Allocation Policy in Cloud Computing Environment using CloudSim," *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 8, No. 1, pp. 344-354, February 2018, ISSN: 2088-8708, DOI: 10.11591/ijece.v8i1.pp344-354.
- [3] Frederic Nzanywayingoma, Yang Yang. "A Literature Survey on Resource Management Techniques, Issues and Challenges in Cloud Computing," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, Vol. 15 No 4, pp. 1918-1928, 2017. DOI: <http://dx.doi.org/10.12928/telkomnika.v15i4.6574>.
- [4] Moreno Marzolla, Stefano Ferretti, Gabriele D'Angelo. "Dynamic Resource Provisioning For Cloud-Based Gaming Infrastructures," Volume 10 Issue 1, Article No. 4 in ACM Digital library, October 2012.
- [5] <https://www.ijcai.org/Proceedings/15/Papers/038.pdf>
- [6] <https://aws.amazon.com/ec2/pricing/>
- [7] Song, Jiayi and Guérin, Roch A. "Pricing and Bidding Strategies for Cloud Computing Spot Instances." Report Number: 2017.
- [8] Agmon Ben-Yehuda, Ben-Yehuda, M, Schuster, A, and Tsafirir, D, T. F. "Deconstructing Amazon EC2 Spot Instance Pricing." *ACM Trans. Econ. Comp.* V, N, Article A., 2012.
- [9] Bhavani B H, H S Guruprasad. "Resource Provisioning Techniques in Cloud Computing Environment: A Survey." *International Journal of Research in Computer and Communication Technology*, Vol 3, Issue 3, March-2014.
- [10] Chaisiri, S.; Bu-Sung Lee; Niyato, D. "Optimization of Resource Provisioning Cost in Cloud Computing. Services computing," *IEEE Transactions on Service Computing*, Vol. 5, No. 2, pp. 164-177, April-June 2012 doi: 10.1109/TSC.2011.7
- [11] <http://www.cloudbus.org/students/MariaThesis2016.pdf>
- [12] A. Filali, A. S. Hafid. "Adaptive Resources Provisioning for Grid Applications and Services," *IEEE Communications Society subject matter experts for publication in the ICC 2008 proceedings*.
- [13] Pooyan Jamshidi, Aakash Ahmad and Claus Pahl. "Autonomic Resource Provisioning for Cloud-Based Software." *IEEE Communications Society subject matter experts for publication in the ICC 2008 proceedings, SEAMS'14*, June 2-3, 2014. Hyderabad, India. Copyright 2014 ACM 978-1-4503-2864-7/14/06.
- [14] Smita Vijayakumar, Qian Zhu, Gagan Agrawal. "Dynamic Resource Provisioning for Data Streaming Applications in a Cloud Environment," *2nd IEEE International Conference on Cloud Computing Technology and Science*.
- [15] Yeongho Choi, Yujin Lim. "Optimization Approach for Resource Allocation on Cloud Computing for IoT," *International Journal of Distributed Sensor Networks* Volume: 12 issue: 3.
- [16] Jonathan Chase and Dusit Niyato. "Joint Optimization of Resource Provisioning in Cloud Computing," DOI 10.1109/TSC.2015.2476812, *IEEE Transactions on Services Computing*,
- [17] Jonathan Chase and Dusit Niyato. "Joint Virtual Machine and Bandwidth Allocation in Software Defined Network (SDN) and Cloud Computing Environments," *IEEE ICC 2014 - Next-Generation Networking Symposium*.
- [18] Waheed Iqbal, Matthew N. Daileya, David Carrerab, Paul Janecek. "Adaptive Resource Provisioning for Read Intensive Multi-tier Applications in the Cloud," *Future Generation Computer Systems*.
- [19] K. Kalyana Chakravarthi, Vaidehi Vijayakumar. "Workflow Scheduling Techniques and Algorithms in IaaS Cloud: A Survey," *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 8, No. 2, April 2018, pp. 853-866, ISSN: 2088-8708, DOI: 10.11591/ijece.v8i2.pp853-866.
- [20] Linlin Wu, Saurabh Kumar Garg, Steve Versteeg, and Rajkumar Buyya. "SLA-Based Resource Provisioning for Hosted Software-as-a-Service Applications in Cloud Computing Environments," *IEEE transactions on services computing*, vol. 7, no. 3, july-september 2014
- [21] <http://www.manjrasoft.com/download/2.0/AnekaDynamicProvisioning.pdf>
- [22] Mayanka Katya and Atul Mishra. "Application of Selective Algorithm for Effective Resource Provisioning In Cloud Computing Environment," *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, Vol. 4, No. 1, February 2014.
- [23] Swati Agarwal, Shashank Yadav, Arun Kumar Yadav. "An Efficient Architecture and Algorithm for Resource Provisioning in Fog Computing," *I.J. Information Engineering and Electronic Business*, 2016, 1, 48-61.
- [24] A. Tchernykh, et al., "Towards Understanding Uncertainty In Cloud Computing With Risks Of Confidentiality, Integrity, And Availability," *J. Comput. Sci.*, 2016, <http://dx.doi.org/10.1016/j.joes.2016.11.011>
- [25] K.Sumalatha, M.N Sinduri, C.BhanuPrakash. "A Novel Method Of Directly Auditing Integrity On Encrypted Data," *International Research Journal of Engineering and Technology (IRJET)*, Volume: 03 Issue: 01 | Jan-2016 .X. S. Li, et al., "Analysis and Simplification of Three-Dimensional Space Vector PWM for Three-Phase Four-Leg Inverters," *IEEE Transactions on Industrial Electronics*, vol. 58, pp. 450-464, Feb 2011.

BIOGRAPHIES OF AUTHORS

Ms.K.Sumalatha is currently working as Assistant Professor in Annamacharya Institute of Technology and Sciences and Research Scholar in the Department of Computing Science and Engineering at Pondicherry Engineering College. Her research work focuses Cloud Computing, Optimization Techniques, and Networks.



Dr. M.S.Anbarasi is currently working as Assistant Professor in the Department of Information Technology at Pondicherry Engineering College. Her area of interest mainly focuses Data Mining, Databases and Cloud Computing.