



INTERNATIONAL JOURNAL OF
ADVANCE RESEARCH, IDEAS AND
INNOVATIONS IN TECHNOLOGY

A Review on Resource Allocation in Cloud Computing

Shivani Sharma
School of Computer Science
and Engineering, Bahra University
shivani.sharma27nov@gmail.com

Dhanshri Parihar
School of Computer Science
and Engineering, Bahra University
dhanshriparihar@gmail.com

Abstract: This paper is based on the literature review of the allocation of different resources such as applications, data and servers, over the cloud in cloud computing technology. Cloud computing is a new generation technology that allows the users to share resources over any communication network by using virtualization technique. The server computer plays the major role in the clouds, as it stores all the data within itself. The data on the server can be accessed by any cloud client by using a web browser. In cloud computing different resources are provided to the users with the help of dynamic resource allocation. Resource allocation is an integral part of Infrastructure-as-a-service (IaaS) model of cloud computing and also is one of the issues in cloud computing. Resource allocation is the process of allocating resources to the users according to their requirement.

Keywords- Cloud Computing, IaaS, Virtualization, Scheduling, Resource Allocation, Resource Sharing

I.INTRODUCTION

Cloud Computing can be used as a synonym for distributed computing over the internet, by which a similar program or application can be made to run on different computers at the same time. Cloud computing involves adaptation & evolution of existing technologies and paradigms. Cloud computing is a type of grid computing, which was evolved by addressing the QoS (Quality of Service) reliability problems [5]. In cloud computing resources as well as data and applications can be shared. For sharing different resources in cloud computing different resource sharing algorithms have been used.

According to the definition of cloud computing given by NIST (National Institute of Standards & Technology):

“Cloud computing is a model for enabling ubiquitous, convenient, on demand network access to shared pool of configurable computing resources (for example: Networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models and four deployment models [6].”

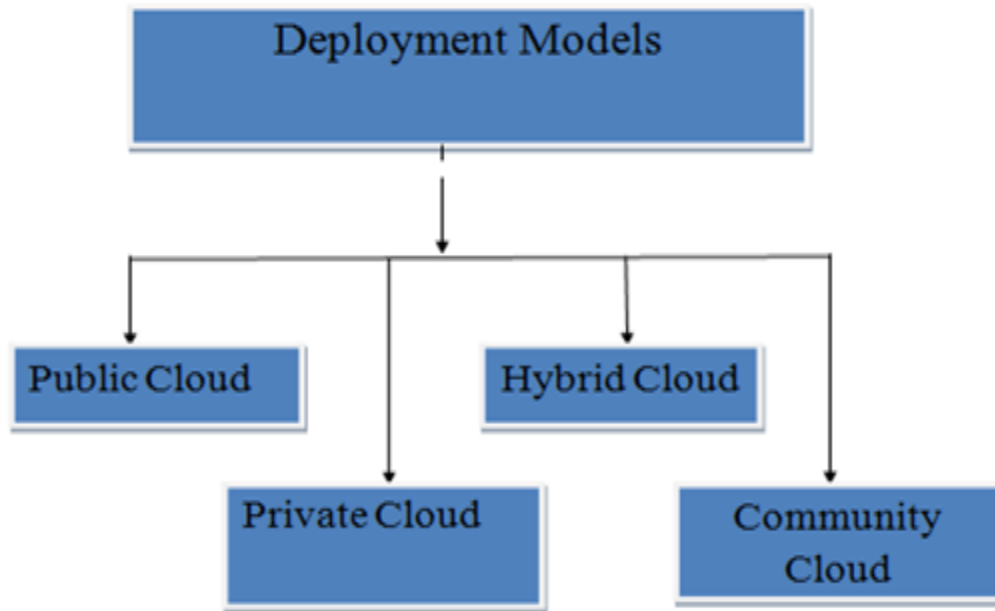


Figure 1: Deployment Models in Cloud Computing

Cloud computing consists of four deployment models: Public Cloud provides storage and applications for general use over the internet. For example: IBM's Blue Cloud. Private Cloud is owned by some private organization or companies for their private use. Hybrid Cloud is combination of Private and Public Clouds. Community Cloud is a kind of Private Cloud which is shared by certain organizations having similar requirements.

Cloud computing provides three types of service models to its users. These are:

- Software-as-a-Service
- Platform-as-a-Service
- Infrastructure-as-a-Service

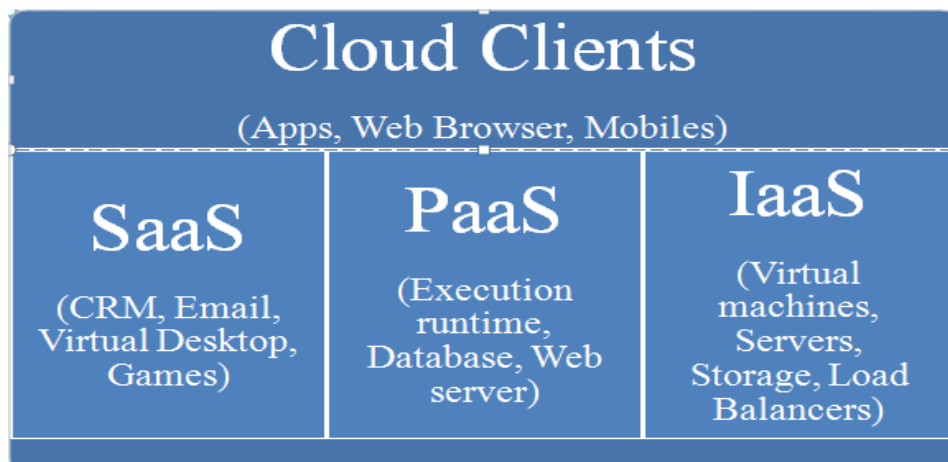


Figure2: Service Models in Cloud Computing

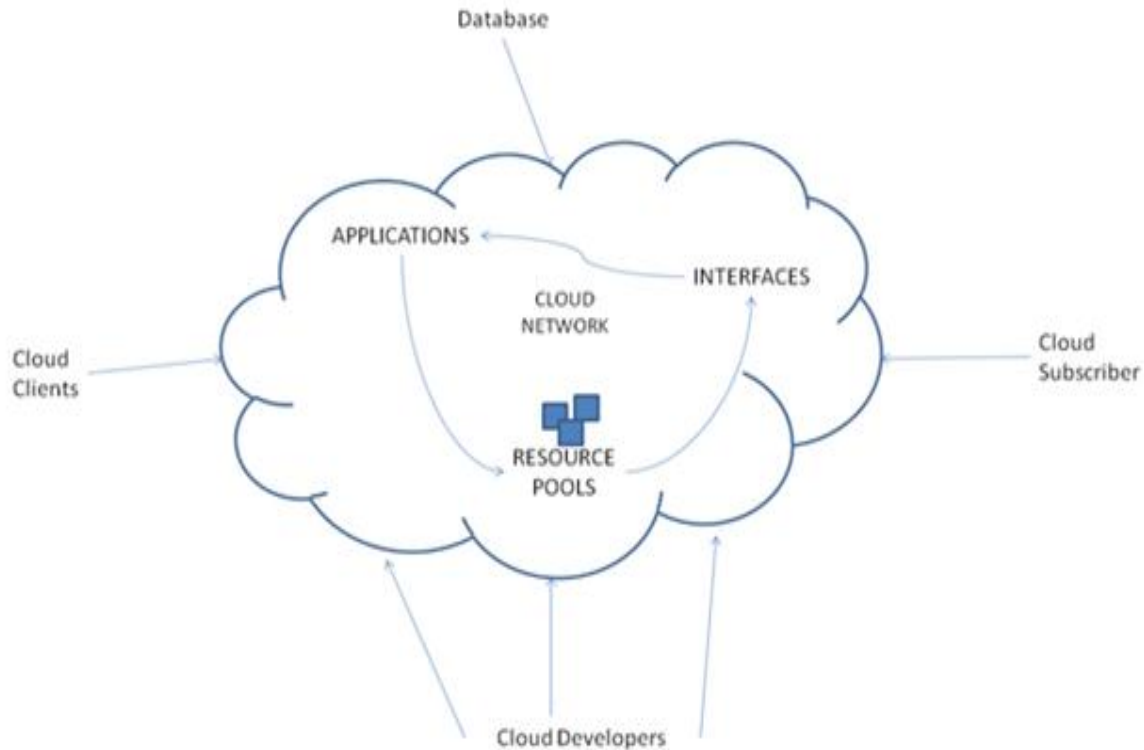


Figure 3: Basic Structure of cloud computing Network

In cloud computing different resources are provided to the users with the help of dynamic resource allocation. Resource allocation is an integral part of Infrastructure-as-a-Service model. Resource allocation is one of the major issues in cloud computing.

Resource allocation is the process of allocating the resources to the clients according to their need. There are various algorithms which are being used for resource allocation in cloud computing. These algorithms help in scheduling virtual machines on the server at various data centers. Some of these algorithms are ACO (Ant Colony Optimization Algorithm) [7], Bee's Algorithm [3], Bin Packing Algorithm [2], Non-preemptive and Preemptive Scheduling Algorithm [2], Priority algorithm [2] and Choco Based (CB) algorithm [1]. These algorithms are used for efficient resource allocation in cloud computing.

Resource allocation algorithms:

- a) Ant Colony Optimization algorithm- Ant Colony Algorithm is based on the behavior of the ants collecting food. Ants form groups to search and collect food in a reliable way. The basic principle of this algorithm is the behavior of ant while moving from source i.e. nest to the destination i.e. food and from source to destination. This algorithm firstly checks the available resources, then selects a set of optimal nodes after analyzing the factors like response time and then allocates the jobs to the suitable nodes [7].

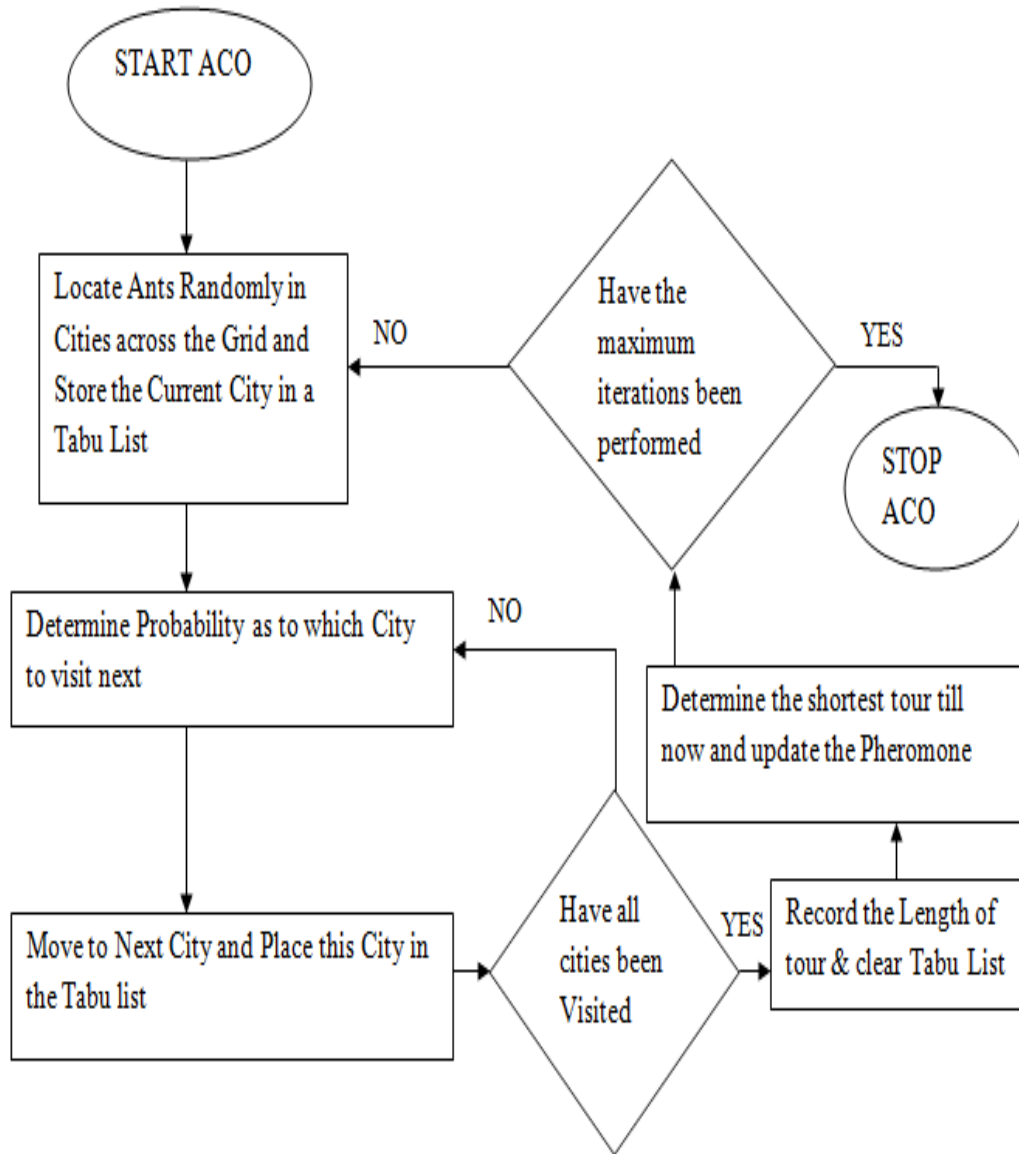


Figure 4: A Flow-Chart of Ant Colony Optimization Algorithm [14]

- b) Bee's Algorithm- This algorithm is based on the action of bee's to get their food. In this algorithm, a meta-scheduler finds a job with lowest memory, input-output, and processor requirement. This job acts as a scout bee to find the suitable site. The scout job is sent to the location at which the task requires the resource at present. The scout job identifies the location by using a fitness function. This fitness function runs the task in a particular instance and evaluates that the task is memory dependent or processor dependent. Fitness is the progress of the particular job with assigned resources. After identifying the resources and location, scout job returns back to the meta-scheduler and perform a waggle function. Waggle function segregates the tasks which are present in meta-scheduler on the basis of the information provided by scout job such as cost, processor and memory requirements [3].

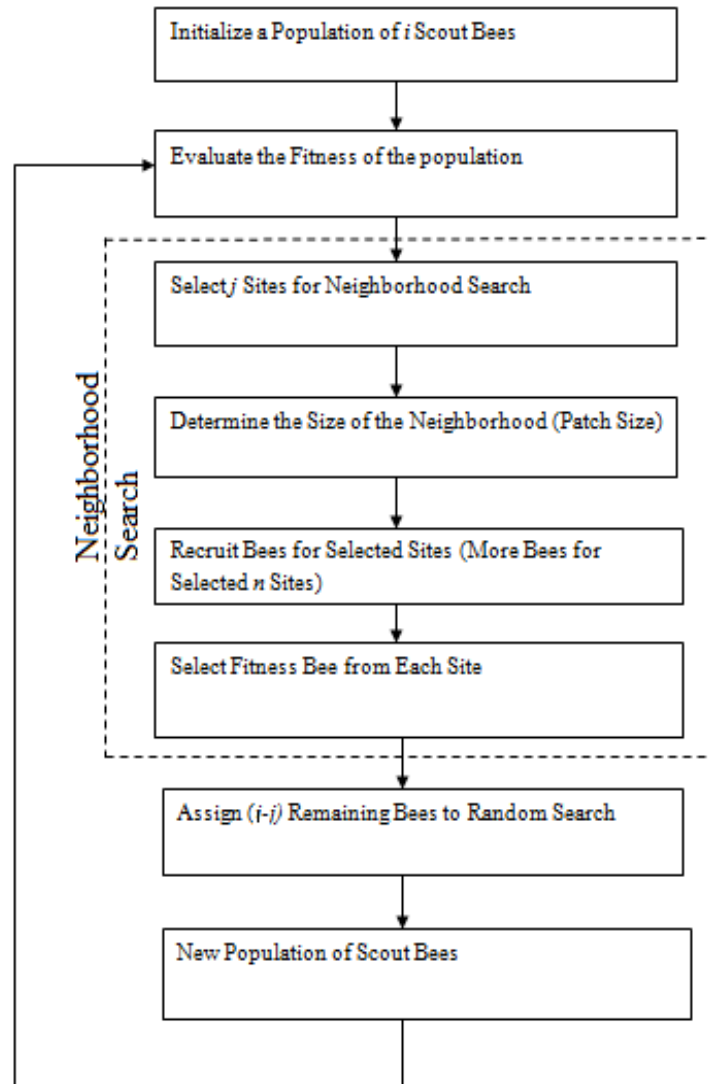


Figure 5: Flow Chart Diagram of Bee's Algorithm [13]

- c) Priority Algorithm: The ideas behind the dynamic resource allocation for preempt able jobs in cloud is allocation of resources to the users according to their demands, Priority based scheduling algorithm performs better than the Cloud min-min scheduling algorithm. In priority algorithm when a job arrives at Cloud scheduler, it divides it into tasks according to their dependencies and then algorithm is called to form the list of the tasks according to their priorities. Then it forms a list of available resources and virtual machines which can be allotted to the tasks. The algorithm then decides the appropriate virtual machine & resource and allocates them to the tasks in the list [2].
- d) Bin-Packing Algorithm: Bin Packing problem involves the packing of the objects of given size into bins of given capacity. In one-dimensional bin packing the size of each object is the real number between 0 and 1 and size of each bin is same, provided that the sum of the number of objects in the bin must not exceed 1. Bin packing algorithms used best fit algorithms of resources in cloud. A formal definition of BPP can be defined as a given list of objects and their weights, bin size, find the least number of bins so that all the resources are assigned to that bin [2].

II.LITERATURE REVIEW

Mohamad Abu, Sharkh, Mahar Jammal, A. Bdallah Shami and Abdel Kader Ouda of the Western University in their paper of Resource Allocation in Network Based Cloud Computing Environment: Design Challenges said that there is a need of a proper solution for allocation of resources in the field of cloud computing for the cloud service provider. The main aim of resource allocation algorithms is to schedule the virtual machines on the main server that resides in data centers, while dealing with the problem on the cloud network. This paper also discusses the challenges faced in the designing the cloud as well as aims to provide proper references for designing comprehensive energy aware resource allocation modes for cloud computing data centers [10].

Resource allocation has been mentioned to be an integral, extending part of many data center management problems, which consists of virtual machine placement in data centers, multipath network routing as well as network virtualization. The paper of generalized resource allocation for the cloud presents an approach to allocation of resources that allows easy evolution of problem specification. In this research an extensible and generic tool, known as WRASSC was designed that could be used by the cloud environment for solving the specific allocation problem [8].

According to the efficient resource allocation for different jobs in cloud, a research paper given by K. Dinesh, G.Poornima and K.Kiruthika many resource scheduling algorithms have been implemented and analyzed for minimizing resource utilization. Job scheduling that uses Berger model is one of the job scheduling algorithms [11].

According to Weiwei Lin's, Baoyun Peng's, Cheem Liang's and Boliu's research in 2013, the scheduling of resources is one of the most important cloud computing problem. The designing of an effective and efficient resource allocation algorithm is a big challenge, because of the reason that scheduling problems occur in NP-Complete. Most of the scheduling algorithms designed till now only consider the CPU, memory and has no emphasis on the requirement of the bandwidth. The optimization of resource allocation has been represented as CSP (Constraint Satisfaction Problem), where three types of resources have been considered by them, namely RAM, CPU and bandwidth & Choco Based (CB) algorithm for VM resource allocation in virtual cloud data centers has been designed [1].

III.CONCLUSION

Cloud computing is a new generation technology which is getting attention these days. Cloud provides various software and hardware resources to the customer according to their need. For efficient sharing of resources various resource allocation algorithms are designed. Some of these algorithms are discussed in this paper. In future as the size of the cloud will increase, more efficient resource allocation algorithms will be required.

REFERENCES

- [1] Weiwei Lin, Baoyun Peng, Chen Liang & Bo Liu, "Novel Resource Allocation Model & Algorithm for Cloud Computing", 2013.
- [2] M.Gokilavani, S.Selvi, C.Udhayakumar, "A Survey on Resource Allocation and Task Scheduling Algorithms in Cloud Environment", 2013.
- [3] Pradeep.R, Kavinya.R, "Resource Scheduling in Cloud Using Bee Algorithm for Heterogeneous environment", 2012.
- [4] V.Vinothina, Dr. R.Sridaran, Dr. Padmavathi Ganapathi, "A Survey on Resource Allocation Strategies in Cloud Computing", 2012.

- [5] Hamdaqa Mohammad, Ladan Tahvildari “Cloud Computing Uncovered-A Research Landscape”, 2012.
- [6] Peter Mell, Timothy Grance, “The NIST definition of Cloud Computing”, 2011.
- [7] Mario Ventresca, Beatric M. Ombuki, “Ant Colony Optimization for Job Scheduling Problem”, 2004.
- [8] Anshul Rai, Ranjta Bhagwan, Saikat Guha, “Generalized Resource Allocation for Cloud”.
- [9] Chandra Shekhar S. Pawar, Rajnikant B. Wagh, “Priority Based Dynamic Resource Allocation in Cloud Computing”.
- [10] Mohamed Abu Sharkh, Manar Jamwal, Abdullah Shami, Abdelkader Ouda, “Resource sAllocation in a Network Based Cloud Environment: Design Challenges”.
- [11] K. Dinesh, G.Poornima, K. Kiruthika, “Efficient Resource Allocation for Different Jobs in Cloud”.
- [12] Mayanka Katyal, Atul Mishra, “Application of Selective Algorithm for Effective Resource Provisioning in Cloud Computing Environment”.
- [13] <http://www.slideshare.net/njoudomar/bee-algorithm>.
- [14] http://www.slideshare.net/mcradc/lecture-9-aco?qid=c42f7411-5520-47e6-a1cc-a320c07804d5&v=default&b=&from_search=1