WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study on Resource Allocation Strategies of Cloud Computing

Ayush Usare[1], Dr. Harsh Lohiya[2]

[1]Research Scholar, Department of CSE, SSSUTMS, Sehore, Madhya Pradesh, India,
[2]Associate professor, Department of CSE, SSSUTMS, Sehore, Madhya Pradesh, India

ABSTRACT

This work is based on a literature review on the allocation of various resources such as applications, data and servers over the cloud in cloud computing technology. Cloud computing is a new generation technology that allows users to share resources in any communication network using virtualization techniques. The main role in the cloud is played by the server computer, which stores all data in it. Data on the server can be accessed by any cloud client using a web browser. In cloud computing, different resources are provided to users using dynamic resource allocation. Resource allocation is an integral part of the Infrastructure-as-a-service (IaaS) model of cloud computing and is also one of the challenges of cloud computing. Resource allocation is the process of allocating resources to users according to their requirements.

Keywords- Cloud Computing, IaaS, Virtualization, Quality of Service, Resource Allocation, Resource Sharing.

1. Introduction

Cloud computing can be used as a synonym for distributed computing on the Internet, where the same program or programs can run on different computers at the same time. Cloud computing involves adapting to the evolution of existing technologies and paradigms. Cloud computing is a form of advanced mesh computing by solving QoS (Quality of Service) reliability issues. Data and applications can be shared across cloud computing resources. Different resource sharing algorithms have been used to share different resources in cloud computing.

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

"Cloud computing is a model where a collection of common computing resources (for example: networks, servers, storage, applications, and services) can be provided quickly and with minimal management effort or maintenance in a network anywhere, conveniently, on-demand... supplier interaction This cloud model consists of five key features, three service models and four deployment models [6].

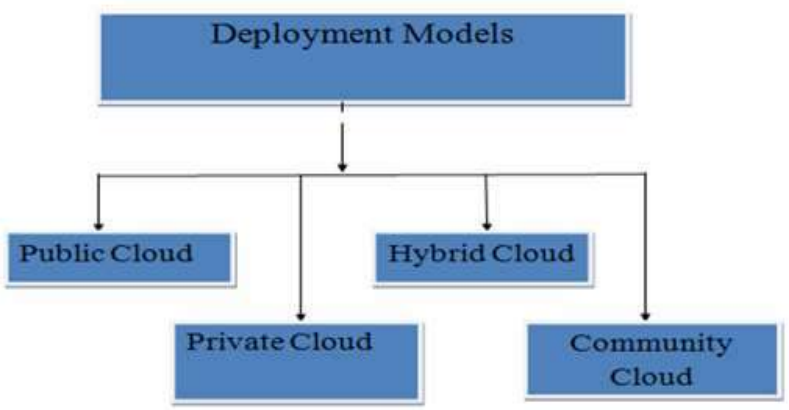

Fig. 1: Deployment Models in Cloud Computing

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 isshared 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

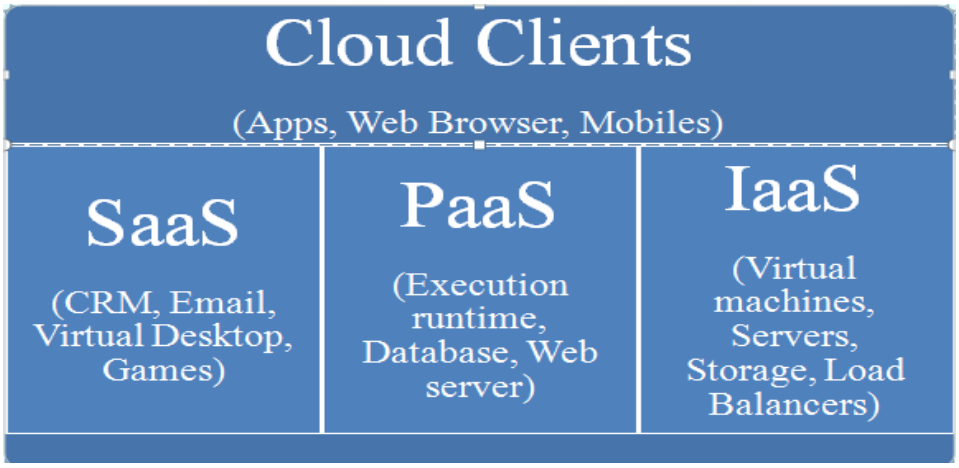

Figure2: Service Models in Cloud Computing

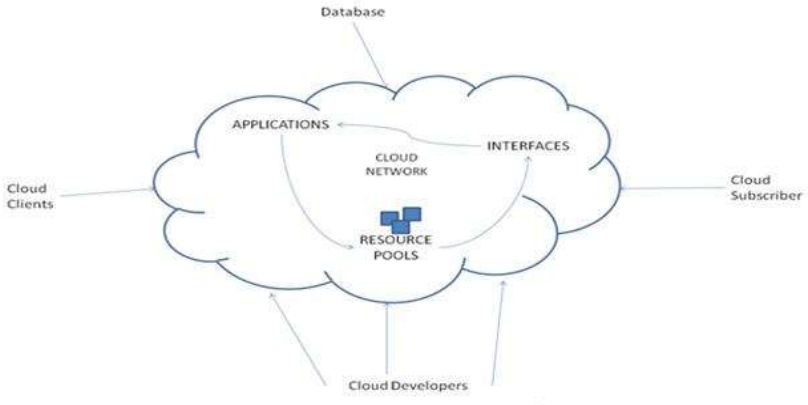

Fig. 3: Basic Structure of cloud computing Network

In cloud computing, users are allocated different resources using dynamic resource allocation. Resource allocation is an integral part of the infrastructure-as-a-service model. Resource allocation is one of the main challenges of cloud computing.

Resource allocation is the process of allocating resources to customers according to their needs. There are several algorithms used for resource allocation in cloud computing. This algorithm helps in scheduling virtual machines on servers in different data centers. Some of these algorithms are ACO (Ant Colony Optimization Algorithm) [7], Bee's Algorithm [3], Packet Packing Algorithm [2], Non-Preemptive and Preemptive Scheduling Algorithm [2], Priority Algorithm [2] and Choco-Based (CB).algorithm [1]. This algorithm is used for efficient resource allocation in cloud computing.

1.1 Resource allocation algorithms:

a) **Ant Colony Optimization algorithm-** The ant colony algorithm is based on ants collecting food. Ants form groups to find and gather food. The basic principle of this algorithm is that ants behave when they move from their source, i.e. nest, food and resources. This algorithm first checks the available resources, then selects the optimal node set after analyzing factors such as response time and allocates the workload to the appropriate node [7].

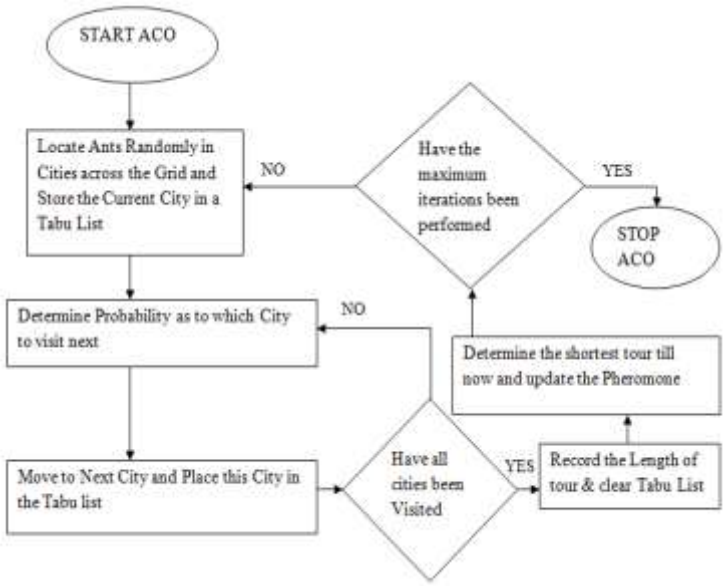

Fig. 4: A Flow-Chart of Ant Colony Optimization Algorithm [14]

b) Bee's Algorithm- This algorithm is based on the behavior of bees to obtain food. In this algorithm, the meta-scheduler works with minimal memory, input, and processor requirements. This work acts as a search bee to find a suitable location. Project scouts are sent to locations that currently require resources. The scout project finds the location using the fitness function. This fitness function performs a task in a given state and evaluates whether it is memory dependent or processor dependent. Fitness is the development of specific activities with designated resources. After determining the resource and location, the search operation returns to the meta plan and performs the carriage function. The Waggle function divides tasks in meta-scheduling based on information such as search, workload, cost, processor, and memory requirements [3].

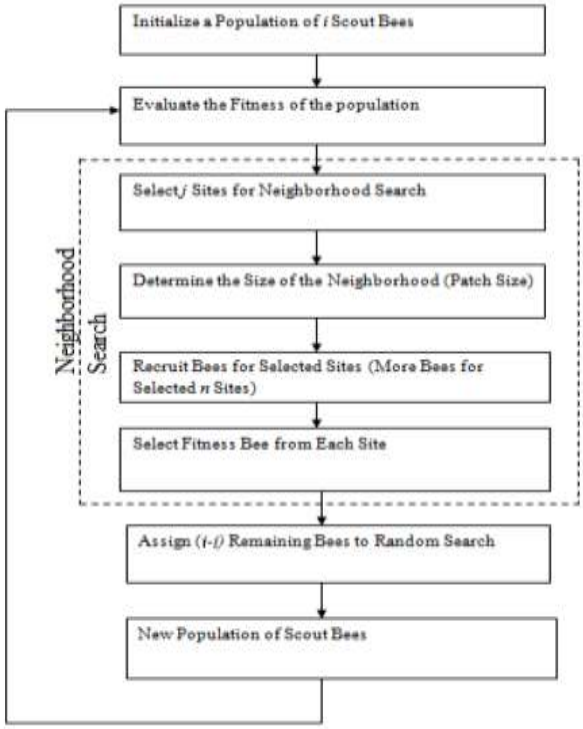

Fig. 5: Flow Chart Diagram of Bee's Algorithm [13]

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].

2. Literature Review

Mohamad Abu et al. proposed that 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].

K. Dinesh et al. describes 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].

Weiwei Lin's et al. describes 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].

3. 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