International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 01
January 2018

A Review on Machine Learning Based Resource Allocation in Cloud Computing

D.Madhavi & Dr.Sateesh kumar nagineni
Research Scholar OPJSU,Churu,Rajasthan,India.
Associate professor, Dep of CSE OPJSU,Churu,Rajasthan,India.

ABSTRACT
Cloud computing is the most advanced technology in the real world environment and provides flexible and convenient possibilities for users to utilize available services. Resource provisioning to the satisfaction of user requirements becomes the most challenging task in the heterogeneous cloud environment. Proper admission control algorithms need to be proposed for better resource provisioning with improved user satisfaction level. The proposed approaches and algorithms have been implemented using CloudSim and the final results have been assessed under the cloud environment. The experimental results indicate that the proposed research methodologies reduce the makespan, cost and result in more efficient satisfaction level for users and service providers when compared to existing algorithms like SLA based model.From the results, the overall findings of the research work can be thus concluded that the proposed research methodologies have the potential to achieve better admission control in cloud environment yielding a better satisfaction of user specified constraints. The comparison of the experimental results show that the proposed research work DLAPUJS provides better results than the previous and existing research methods in terms of profit, number of Virtual Machines(VMs) initiated, execution time and data transfer cost. In future scenario, big data applications can be experimented for the admission control methodologies to adapt to the real world environment. Dynamic switching of resources can be introduced in response to dynamic requests submitted by cloud users. Fault tolerance mechanisms can be integrated with admission control frameworks to avoid the risk factors.

Keywords: SLA, QoS, VMs, Resource allocation

1. INTRODUCTION

Resource allocation is the process of selecting and providing the most suitable resources that satisfy user requirements. Resource allocation is the most complex task in the cloud computing environment where the selection of resources with the satisfaction of Quality of Service (QoS) would be more challenging of the tasks. The resources need to be utilized well for achieving higher profit for the cloud services providers. A lot of research work has been conducted on optimal resource allocation [1].The users can utilize the resources that are allocated by the cloud computing service providers through virtualization environment. Virtualization enables service providers to handle multiple users‟ requests simultaneously. Load level of virtual machines needs to be balanced for better. Resource utilization so that the tasks can be completed in a competent manner. Resource

allocation is divided into three main types based on the allocation methodology [2]. They are • Static Resource Allocation • Dynamic Resource Allocation • Users Self Allocation One of the main goals which need to be satisfied by different Resource allocation techniques are Service Level Agreement (SLA). SLA is defined as the contract made between the cloud service providers and the cloud users regarding some QoS constraints for completing the task execution submitted by the cloud users. These resource allocation techniques should ensure the SLA for the assured delivery of resources and complete execution of tasks submitted by the user. A brief explanation of the static and dynamic resource allocation is given in the following subsection.

1.1 ALLOCATION RESOURCE METHOD

Resource allocation is the process of choosing and providing the most suitable resources that satisfy user specifications. Resource allocation is the most complex task in the cloud computing environment where the selection of resources with the satisfaction of Quality of Service (QoS) would be more challenging of the tasks. The resources need to be utilized well for producing higher profit for the cloud services providers. A lot of research work has been accompanied by optimal resource allocation (Ye Hu et al. 2009).

The users can appropriate the resources that are allocated by the cloud computing service providers through virtualisation environment. Virtualization enables service providers to manage multiple users‟ requests concurrently. Load level of virtual machines requires being adjusted for better resource utilisation so that the tasks can be performed in a competent manner (Suganya et al. 2014).

Resource allocation is classified into three main standards based on the allocation methodology (Bhavani & Guruprasad 2014). They are

Static Resource Allocation

Dynamic Resource Allocation

Users Self Allocation

One of the primary goals which need to be performed by different resource allocation procedures is Service Level Agreement (SLA). SLA is defined as the contract made among the cloud service providers and the cloud users concerning some QoS constraints for completing the task execution offered by the cloud users (Source: https://www.paloaltonetworks.com). These resource allocation techniques should guarantee the SLA for the precise delivery of resources and complete performance of tasks submitted by the user. A brief explanation of the static and dynamic resource allocation is given in the following subsection.

With the transformation of the architecture from mainframes to client-server models, information technology services have been rapidly developed and increased. The cloud computing concept addresses a new paradigm shift in Internet-based services that provides highly scalable distributed computing platforms where computational resources are offered as a service, and hence consumers do not need to understand how it works and can easily access various services via the Internet.

1.1.1 The Static Resource Allocation

Static resource allocation is the way of designating resources to the users before beginning the task execution. It cannot be compared to starting performance of device

use. It is most proper for applications whose workloads are expected apriori and unchangeable until the achievement of execution. In this advance allocation methodology, cloud users would make contact with the cloud service providers and obtain the identical resources that can satisfy their obligations (Bhavani & Guruprasad 2014).

1.1.2 Dynamic Resource Allocation Method

Dynamic resource allocation is defined as the one allowing the users to get their required resources on-demand with the content of QoS constraints. Dynamic allocation is best suited for applications with the variable workload and might vary in runtime. The dynamic resource allocation presents a flexible environment for the users from getting their expected resources on-demand, and they are required to pay only for the resources for which they utilised. Dynamic resource allocation is determined to be better than the static resource allocation methods by scaling down/ scaling up the funds in runtime (Bhavani & Guruprasad 2014).

1.1.3 A User Self Allocation

User self-allocation is also known as cloud self-service wherein the user obtains resources from the service provider by performing an account and giving the resources (Bhavani & Guruprasad 2014).

An Overview of cloud computing

Cloud computing is defined as a model enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [6] . In paper [7] refers cloud computing as:

The applications delivered as services over the Internet

The hardware and systems software in the datacenters

that provide services that is software as a service.

Its characteristics comprise of a broad network access which has the ability to access the network via heterogeneous platforms, on demand self-service, when it provision the computing power automatically. It also has features of the service oriented which is called measures service, with its metering capability, resources can be monitored, controlled and reported and this is transparent to the provider and the customer. Furthermore, it is also equipped with an elastic feature where the scale-out or scale-in provisioning of resources can be implemented rapidly in order to make the resources available, unlimited and can be purchased at any time by the cloud subscriber.

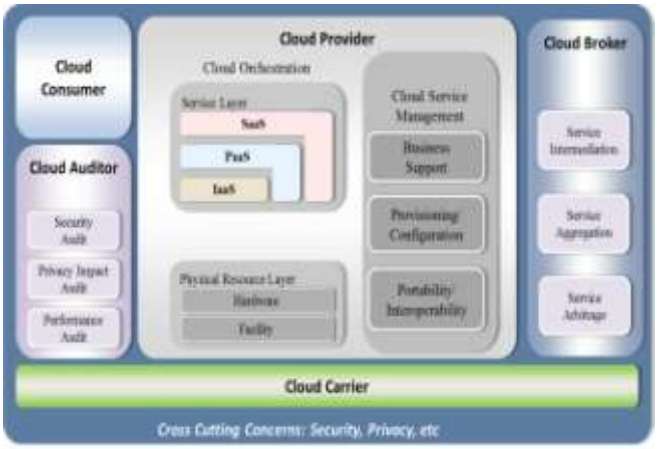
Cloud provider: Person, organization or entity responsible for making a service available to Cloud Consumers

Cloud broker: An entity that manages the use, performance and delivery of cloud services, and negotiates relationships between Cloud Providers and Cloud Consumers.

Cloud consumer: Person or organization that maintains a business relationship with, and uses a service from Cloud Providers.

Cloud carrier: The intermediary that provides connectivity and transport of cloud services from Cloud Providers to Cloud Consumers.

Cloud auditor: A party that can conduct an independent assessment of cloud services, information system operations, performance and security of the cloud implementation.



2. LITERATURE SURVEY

2.1 MACHINE LEARNING BASED RESOURCE ALLOCATION

David Glesser et al. (2015) discussed the role of machine learning algorithm in the cloud computing environment. The authors discussed the impact that can occur by adapting the machine learning algorithm for job scheduling process. Different machine learning algorithms have been adapted for job scheduling which is evaluated in the cloud computing environment in terms of learned knowledge of resource status information. Back filling algorithm is mainly used for finding the impacts that can be caused by machine learning algorithms in terms of finding the suitable resource for the task requests. The back filling algorithm attempts to fill the resource capacity by assigning the maximum tasks. This assignment is done based on the resources capability which is learned using the machine learning approaches. The evaluation results prove that the machine learning algorithms play greater role in the cloud computing environment in terms of their utility.

Kiran Bala & Sahil Vashist (2014) analysed the various resource and task execution status by using the machine learning algorithm to find a better mapping process. This research work generates a secured and effective cloud environment in which the brokers can make decision to map the request to the resources in the proper manner with the consideration of the available resource execution status. This research work makes use of predefined rules that can generate better mapping of resources for the requests which are generated newly. The cloud framework can ensure the successful completion of task execution in terms of better mapping. The overall research of this work proves that the machine learning based mechanism can reduce the computation overhead in a considerable manner.

Venkatesa Kumar & Dinesh (2012) implemented fuzzy neural network algorithm for the successful implementation of job scheduling in cloud computing environment. This approach classifies the user task requests based on QoS constraints such as bandwidth, memory, CPU utilization and size. Each classified task would be given to separate neuron as an input. Fuzzy neural network composed of input layer, hidden layer and the output layer. These requests would be processed in the hidden layer by comparing them with the different resource capacities in terms of their QoS constraints. Matching input neurons would be combined in the hidden layer that can satisfy QoS constraints and finally better combination of neuron would be retrieved as output. This algorithm leads to a better scheduling by selecting the most optimal resource for the task scheduling by adjusting

the weight parameter value in a considerable manner.

Pengcheng Xiong et al. (2011) presented a gain oriented admission control support called Active - SLA for database-as-a-Service (DaaS) environment. As a result of the service delivery nature of DaaS, the system overload impacts the cost of the service provider with payment of penalties to the unsatisfied clients when the performance is below the SLA. The overall performance of the system is monitored in such a way that the SLA parameters are satisfied to avoid unnecessary penalties.

Jaideep Dhok et al. (2010) presented the Opportunistic Admission Control Algorithm based on learning methods for providing Map Reduce as a service. The admission control algorithm allows the user to buffer the Map Reduce jobs in the form of on-demand services. In this approach, a learning algorithm is used to model MapReduce jobs in a ready to use manner which effectively makes the MapReduce as SaaS. The admission control algorithm is based on machine learning algorithm for the prediction of the job admission. Resource sharing and multiplexing the job execution as an alternative of focusing on maximizing user and service provider utility improves the efficiency of the job scheduling approach.

Vipul Mathur et al. (2009) proposed an adaptive admission control scheme in order to provide better scheduling performance for web applications with variable capacity. The approach adopts dynamically by adjusting the load in order to compensate for the variation in the system capacity. The scheme is implemented as a proxy server involving the clients and front-end web servers. The main purpose is to stabilize the changes in response time and throughput, while also avoiding overloads that occur as a result of a reduction in available system capacity.

3. PROPOSED METHODOLOGY

This section discusses the proposed dynamic task scheduling using Particle Swarm Optimization. Table 1 shows an illustrative example where each row represents the particles which correspond to a task assignment that assigns five tasks to three processors, and [Particle 3, T4] =P1 means that, in particle 3, the task 4 is assigned to Processor 1.

The Algorithm used is for the dynamic task scheduling is as follows: The particles are generated based on the number of processors used, number of tasks that have arrived at a particular point of time and the population size specified. Initially the particles are generated at random and the fitness is calculated which decides the goodness of the schedule. The pbest and gbest values are calculated. Then the velocity updation and the position updation are done. The same procedure is repeated for the maximum number of iterations specified. The global solution which is the optimal solution is obtained. When a new task arrives, it is compared with the tasks that are in the waiting queue and a new schedule is obtained. Thus the sequence keeps on changing with time based on the arrival of new tasks. Each particle corresponds to a candidate solution of the underlying problem. Thus, each particle represents a decision for task assignment using a vector of r elements, and each element is an integer value between 1 to n. The algorithm terminates when the maximum number of iterations is reached. The near optimal solution is thus obtained using Particle Swarm Optimization.

3.1 Problem definition

Assume the cloud customer has different jobs and each job is splitter into number task and each task has divided into sub tasks. These tasks are allocated to resources (memory, network, CPU) as virtual machines. The concept of scheduling and resource allocation has different aim. We need to find a schedule to execute a DAG workflow on Infrastructure as a Service computing resource to minimize the execution cost and time. Input: The schedule is defined in the format of S=(R, A, TC, TT) in the form of set of resources (R), task to resource allocation (A), total cost (TC) and total time (TT). For each resource R= (r1, r2... rn) different virtual machines are allocated, and each resource has its start time STri and end time ETri. Here A represents allocation and the allocation consist of set of tuples in the form of at a r t= (ti, rj, STri, ETri). A task ti is scheduled to run on the resourcerj. The total cost TC and total time TT are calculated as follows:
Output: Assignment of tasks to resources in minimized time and cost. The n number of tasks and n number of resources are allocated as {(t1, r1), (t2, r2)... (tn, rn)}. Constraints: Each task must be completed within short time and without interruption. One virtual machine can complete one task at a time. The processing time is depends on the virtual machine allocated. Objective: The aim is to assign each task to matching virtual machine resources and sequence the tasks to minimize the time, cost and throughput. The challenge of job scheduling and resource allocation was optimized by combing GABFO with minimized time and cost, finally reliability is achieved. 4. The standard bacterial foraging optimization BFO algorithm was used to solve optimization problem [18]. Bacterial Foraging Optimization is an evolutionary method based on E.coli bacteria. The area having high level nutrients are searched by bacteria. This task is used for optimization process. By sending signals individual bacterium communicates with others. During foraging locomotion is achieved by a set of tensile flagella. E.coli bacteria tumble or swim using flagella. These are the two basic operation of bacteria performed at the time of foraging. After considering two previous factors foraging decision is taken by bacteria.

4. OBJECTIVES

The present research work focuses on optimal resource allocation for the cloud consumer jobs with consideration of the multiple QoS constraints with reduced execution time. The major objectives of the proposed research work are:

1. To predict the near future resource information accurately by learning historical information about resources using machine learning approaches (Support Vector Machine and Artificial Neural Network) to achieve proper resource allocation with improved user satisfaction level.

To optimize the resource allocation process by utilizing the optimization algorithm instead of machine learning algorithms to support huge volume of tasks and to avoid incorrect indelicate parameters value which may lead to wrong predictions.

To eliminate the computation overheads that occur due to the generation of large volume of rules for more number of arrived tasks by introducing the NMF based clustering approach.

2. To provide priority for the users based on their category and to reduce the data transfer cost by selecting the shortest path from which data can be retrieved for the execution.

4.1 CONTRIBUTION OF THE WORK

The first contribution is to propose a machine learning based approach for SLA

aware resource allocation with admission control. This research considers two machine learning approaches for SLA aware resource allocation, namely Support vector machine (SVM) classification approach, and the Artificial Neural Network (ANN) classification approach. In this approach, QoS constraints like profit and the make span are considered for improved user satisfaction level.

The second contribution of this paper is the Position Balanced Parallel Particle Swarm Optimization (PB-PPSO) method utilized for optimizing the resource allocation process. This approach extracts rules from the training set and then finds the optimal rule based on the best fitness values. By using these rules, optimal resource allocation with the satisfaction of profit and make span parameters is ensured for the current allocation.

The third contribution of this paper is NMF-based optimal clustering and PB-PPSO based resource Prediction method (NMF-PB-PPSO). This approach is used to efficiently handle large volumes of tasks with reduced time by clustering similar kind of tasks and resources together. Once clustering, optimal resource allocation for those clusters is done by using the Position Balanced Parallel Particle Swarm Optimization approach(PPSO).

5. PROPOSED FRAMEWORK

Machine Learning Approaches in Refining Service Level Agreement-Based Admission Control On behalf of a SaaS Provider in Cloud

Software-as-a-Service provider is popular in cloud computing environment which offers a set of reliable services to the cloud consumers regarding their request. To do so, SaaS service providers lease the required resources from the IaaS service providers to achieve their client requirements. Renting resources from the external service providers such as IaaS service providers might increase the administration and maintenance costs of the internal service providers. The rental system of the cloud collects the required resources from the external service provider which would also cause the service quality degradation. To overcome these issues, this proposed research methodology seeks to use, machine learning approaches for ensuring the SLA based admission control. Two machine learning approaches that are proposed in this research work for reaching the SLA based admission control are Support Vector Machine (SVM) approach and Artificial Neural Network (ANN) approach. Both of the proposed approaches attempt to control the admission control process by allocating the tasks to such resources that can provide more Return on Investment (ROI). This is done by learning the profit and the non-profit parameters of the cloud resources and the tasks in an enhanced manner.

5.1 ADMISSION CONTROL IN CLOUD COMPUTING

SLA violation is the critical issue which might occur while renting the resources from the external service providers for internal deployment. This problem might create more impact on the software services that are allocated to cloud users regarding increased administration and maintenance cost. This problem needs to be resolved effectively for the better resource allocation to the cloud users. Hence the better admission control mechanism needs to be implemented between the SaaS providers, and the IaaS providers which can The proposed research methodology of this work makes use of two machine learning algorithms for efficient admission control with the satisfaction of various QoS requirements by learning the machine status information. The proposed research attempts to improve the satisfaction level of both

cloud users and cloud service providers by considering SLA parameters of both. The admission control algorithm proposed in this work will analyze the various user QoS requirement and learned knowledge of machine status and then will find the suitable resources for better allocation. The scheduling would be performed in view of the admission control algorithms choice.

5.2 MACHINE LEARNING BASED ADMISSION CONTROL

Machine learning techniques are most often used in the real world applications and are more suitable for the dynamic cloud environment. They can adapt to the dynamic nature of the resources and tasks flexibly. In this research work, two machines learning algorithms are proposed for better admission control by learning the resource status information effectively. Those algorithms are

1. Support Vector Machine
2. Artificial Neural Network
3.

Both of the algorithms are evaluated in this work for admission control process, and outcomes of these algorithms are finally compared with each other to predict a better mechanism that can produce improved results regarding increased ROI. The working procedures of these methodologies are discussed in detail in the following sub sections.

Support Vector Machine Based Admission Control Support vector machine is an administered learning technique which is utilized to investigate information and perceive patterns.It is also used for regression analysis and classification. Generally, SVM employs two possible classes for each of the input that takes a set of input data and predicts the admission. The conceivable classes frame the contribution by making the SVM a non-probabilistic paired direct classifier.

Artificial Neural Network Based Admission Control

Artificial Neural Networks (ANN) comprises of frameworks that are purposely worked to make utilization of some authoritative esteems looking like those of the human mind. They speak to the promising new making of data preparing frameworks. Neural Networks are great at assignments, for example, pattern coordinating and arrangement, optimization and information bunching. They have countless interconnected preparing components called neurons, which more often than not work in parallel and are sorted out in standard structures. The aggregate conduct of a NN, similar to a human cerebrum, uncovers the capacity to learn, review and sum up from preparing patterns or information. NNs are characterized by

1. Pattern of interconnection between neurons
2. Learning algorithm
3. Activation function.

In an NN, each neuron is connected to the other neuron using directed connected link and with an associated weight. Every ha an interior gaze called as its movement level. In light of the flag stream course, they are divided as bolster forward networks and criticism networks.

5.3 POSITION BALANCED PARALLEL PARTICLE SWARM OPTIMIZATION METHOD FOR RESOURCE ALLOCATION

Scheduling and allocation of resources play a critical role in the cloud environment which might violate the satisfaction level of both cloud users and the cloud service providers. Scheduling would be the most critical task in case of arrival of more number of tasks dynamically. This needs to be optimized to get a better solution of task

resource scheduling. This research work attempts to handle number of incoming tasks that are arriving dynamically by scheduling them in an optimal manner. In this proposed inquire about, Position balanced Parallel Particle Swarm Optimization (PB-PPSO) conspire is produced for proficient designation of cloud assets. In this method, Particle Swarm Optimization (PSO) an unsupervised learning algorithm is used for finding the optimized resources for the group of tasks. In the optimization algorithm, every particle has the gathering of errands and gathering of assets that starts with a self-assertive statement of particle's area and speed. Each particle in the swarm conduct has two particulars: an area which speaks to the proposed area and a speed which implies the speed of moving particle. The particle in the swarm consults over the whole pursuit space and remembers the best area found. The correspondence happens between the particles with the goal that they alter their areas and speeds in light of arrangements found by others. The area of the particle is scored by the wellness. The wellness is figured into the goal capacities. The principle target work is to give errand assignments that will achieve least makespan and least cost for the clients.

ALGORITHM 1: POSITION BALANCED PARALLEL PARTICLE SWARM OPTIMIZATION (PB-PPSO) ALGORITHM

Input: Training samples and class labels
Output: Resource allocation

1. Initialize N number of particles with set of tasks and allocate the resources randomly, a position of particle is denoted by Xi and velocity is denoted as Vi .

2. ***pbest*** represents the best well-known position of particle i and gbest signifies the best position of the entire swarm
3. Particle position is initialized as Xi

4. For every particle i=1, 2… N

5. Compute the fitness value for each particle
6. // Fitness computation
7. Fitness Min = (Makespan, total cost)
8. If the fitness value is higher than the pbest
9. Set the present value as the new pBest
10. Until a termination criterion is met
11. Select the particle with best fitness value of all particles as the gbest
12. Update particle position and velocity
13. (t 1) x (t) v (t 1) i ii x += + +
14. until some stopping condition is met
15. Generate the rules and assign class labels
16. Particles which has less fitness are assigned to "loss" at every iteration
17. Particles which has high fitness are assigned to "profit" at every iteration
18. // Testing Process
19. If a new task is submitted, learn the rules
20. Assign the resources for the tasks

6. EXPERIMENTAL RESULTS

In this section, the resource allocation procedures are implemented by using the CloudSim that is a Cloud environment simulator. The appearance of the proposed NMF-PB-PPSO method is compared to the previous method like PB-PPSO from the user and the resource provider perspectives. On the user side, the total number of requests are accepted, and the user request will be processing. In the experimental results, three performance metrics such as profit in $, Average Response Time in seconds are compared. The actual
values that are obtained during evaluation of proposed and existing methodologies as denoted

6.1 Profit Comparison

Profit is displayed as the difference between the total amount that is invested and the total amount which is retrieved as earnings. The gain of the proposed research methodology should be high for its better performance. Profit is calculated using the procedure as given in Equation 3.1.
The comparison chart of profit of the proposed research methodologies and the existing research method is provided in Figure 5.2.

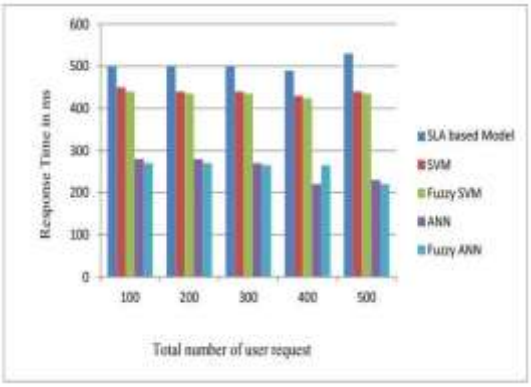


Figure 5.3 shows that the NMF-PB-PPSO method produces less average response time and receiving the least number of VMs when arrival rate is increasing from 100 to 500. When the user request number is multiplied, the average response time is decreased in the proposed method as contrasted to the existing system. NMF-PB-PPSO shows 23% less acknowledgement time than the PB-PPSO methodology.

CONCLUSION

In this paper, we have discussed about resource management in general, the existing resource allocation and monitoring strategies from the current research works. This paper has summarized different method (algorithms technique) and theory which being used to formulate framework and model, derived to provide a better resource allocation and monitoring process in terms of a better performance, competitive and efficiency to meet the required SLA, improved the resource performance and lowered the power consumption. We hope this paper will motivate researchers to explore and formulate a new mechanism to solve issues in allocating and monitoring resources in cloud computing.
The overall findings of this research work conclude that the proposed research methodologies can accomplish the admission control process in the cloud computing environment to the satisfaction of user specified constraints. The experimental tests conducted prove that the proposed research namely DLAPUJS provides better results than the existing and previous research works in terms of increased profit and decreased number of VMs initialized as well as execution time with reduced data transfer cost.

REFERENCE

[1] Bo An, Chunyan Miao & Zhiqi Shen 2007, „Market Based Resource Allocation with Incomplete Information", Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India.

[2] ChakradharaRao Ch, Mogasala Leelarani & Ramesh Kumar, Y 2013, „Cloud: Computing Services And Deployment Models", International Journal Of Engineering And Computer Science, ISSN:2319-7242, vol. 2, Issue 12, pp. 3389-3392.

[3] Chitra, S & Kalpana, B 2014, „Optimum session interval based on particle swarm optimization for generating personalized ontology". Indian Journal of Science and Technology, vol. 7, no. 8, pp. 1137-43.

[4] Chonho Lee, Ping Wang & Dusit Niyato 2015, „A Real-Time Group Auction System for Efficient Allocation of Cloud Internet Applications", IEEE Transactions on Services Computing, vol. 8, no. 2.

[5] David Glesser & Yiannis Georgiou BULL, Denis Trystram LIG 2015, „Improving Job Scheduling by using Machine Learning Yet an another SLURM simulator", Bull atos technologies.

[6] Devendra Dahiphale, Rutvik Karve, Athanasios V. Vasilakos, Huan Liu, Zhiwei Yu, AmitChhajer, Jianmin Wang & Chaokun Wang 2014, „An Advanced MapReduce: Cloud MapReduce, Enhancements and Applications", IEEE transactions on network and service management, vol. 11, no. 1.

[7] Dhingra, A & Paul, S 2014, „Green Cloud: heuristic based BFO tech-nique to optimize resource allocation", Indian Journal of Science and Technology; vol. 7, no. 5, pp. 685-91.

[8] Eun-KyuByun, Yang-Suk Kee, Jin-Soo Kim & Seungryoul Maeng 2011, „Cost optimized provisioning of elastic resources for application workflows", Elsevier Journal of Future Generation Computer systems, vol. 27, Issue 8, pp. 1011-1026.

[9] Garg, SK, Buyya, R & Siegel, HJ 2009, „Time and cost trade-off man-agement for scheduling parallel applications on utility grids". Future Generation Computer System. vol. 26, no. 8, pp. 1344-55.

[10] Gaurav Pandey, Himanshu Gupta & Pabitra Mitra 2005, „Stochastic Scheduling of Active Support Vector Learning Algorithms", Proceedings of the 2005 ACM symposium on Appliedcomputing, pp. 38-42.