ISSN (Online): 2230-7893
www.IJCEM.org

Improved SAH-DB Algorithm for Task Scheduling in Cloud Computing

Johnson Rai[1], Navdeep Kaur[2]

[1,2]Deptt of Computer Science,
SGGSWU, India

Abstract

The cloud computing is the development of distributed computing, parallel computing and grid computing, or defined as the commercial implementation of these computer science concepts. Task scheduling plays a key role in cloud computing systems. Scheduling of tasks cannot be done on the basis of single criteria but under a lot of rules and regulations that we can term as an agreement between users and providers of cloud. This agreement is nothing but the quality of service that the user wants from the providers. Providing good quality of services to the users according to the agreement is a decisive task for the providers as at the same time there are a large number of tasks running at the provider's side. .

The task scheduling problem can be viewed as the finding or searching an optimal mapping/assignment of set of subtasks of different tasks over the available set of resources (processors/computer machines) so that we can achieve the desired goals for tasks. One of the fundamental issues in this environment is related to task scheduling. Cloud task scheduling is an NP-hard optimization problem, and many meta-heuristic algorithms have been proposed to solve it. A good task scheduler should adapt its scheduling strategy to the changing environment and the types of tasks.In this paper we mainly categorize various types of task scheduling approaches and divide the algorithms based on each category.

Keywords:*Cloud Computing,Execution Cost,Meta-heuristicsNP-hard,Optimal,Scheduling*

I. Introduction

Cloud computing is 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. Cloud computing is The word cloud (also phrased as "the cloud") is utilized as an analogy for "the Internet," so the term cloud computing characterizes "a style of Internet-based computing," where totally different services—like servers,storage associated applications are conveyed to an organization's computers and devices through the internet.

The objective of cloud computing is to use ancient supercomputing, or superior computing power, unremarkably employed by military and analysis facilities, to perform vast computations per second, in consumer-oriented applications like monetary portfolios, to deliver customized info, to produce information storage or to power massive, immersive on-line pc games.

Cloud computing uses networks of huge group of servers usually running minimal effort PC technology with specialised connections to unfold data-processing chores across them.

Generally, virtualization techniques are utilized to boost the force of cloud computing..

1. Infrastructure as a Service (IaaS): This layer delivers hardware components (like server and storage) and software as services.

2. Platform as a Service (PaaS): Cloud application developers are the users of this layer. Automatic scaling, load balancing and integration with other services (e.g. email services) are the major benefits to cloud application developer.

3. Software as a Service (SaaS): This layer hosts the software and provide to the customer through Internet. It reduces the purchase and maintenance cost of the customer.

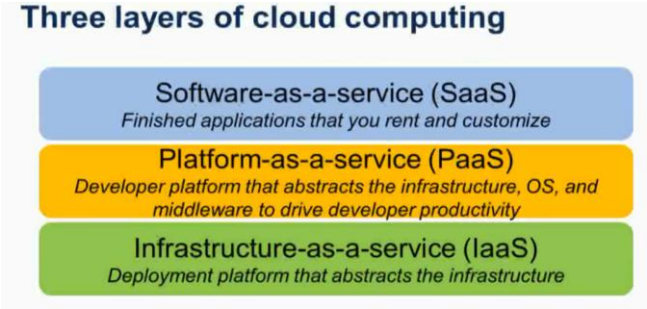

Fig1:Layers in Cloud Computing

II Overview:Task Scheduling in Cloud

Cloud consists of a number of resources that are different with one other via some means and cost of performing tasks in cloud using resources of cloud is different so scheduling of tasks in cloud is different from the traditional

ISSN (Online): 2230-7893
www.IJCEM.org

methods of scheduling and so scheduling of tasks in cloud need better attention to be paid because services of cloud depends on them. Task scheduling plays a key role to improve flexibility and reliability of systems in cloud. The main reason behind scheduling tasks to the resources in accordance with the given time bound, which involves finding out a complete and best sequence in which various tasks can be executed to give the best and satisfactory result to the user. Job Scheduling in cloud computing refers to dispatch the computing tasks to resource pooling between different resource users according to certain rules of resource use under a given cloud circumstances.

In cloud computing, resources in any form i.e. cups, firewall, network are always dynamically allocated according to the sequence and requirements of the task, subtasks. So, this leads task scheduling in cloud to be a dynamic problem means no earlier defined sequence may be useful during processing of task. The reason behind the scheduling to be dynamic is that because flow of task0 is uncertain, execution paths are also uncertain and at the same time resources available are also uncertain because there is a number of tasks are present that are sharing them simultaneously at the same time.. In scheduling algorithms list of tasks is created by giving priority to each and every tasks where setting of priority to different tasks can be based on various parameters. Tasks are then chooses according to their priorities and assigned to available processors and computer machines which satisfy a predefined objective function

Currently, in cloud computing, the issue about how to get the higher QoS (quality of service) by the lower service cost is increasingly attracting the attention. Cloud computing can provide the dynamic and elastic virtual resources for the users to execute the large-scale computing tasks.The tasks scheduling plays an important role in the Cloud computing. It should adopt the scheme to dispatch the computing tasks to the appropriate resources considering some QoS constraints, e.g., task execution time, task completion time, resource utilization, and cost. At present, the scheduling schemes should consider the associate tasks scheduling problem with some constraints in the real applications.

Scheduling in cloud computing can be categorized into three stages.

1.Resource discovering and filtering – Datacenter Broker discovers the resources present in the network system and collects status information related to them.

2.Selecting a target resource (Decision stage) – Target resource is selected basedon certain parameters of task and resource. This is deciding stage.

3. Submission of a particular task to a target resource - Task is submitted to resource selected.

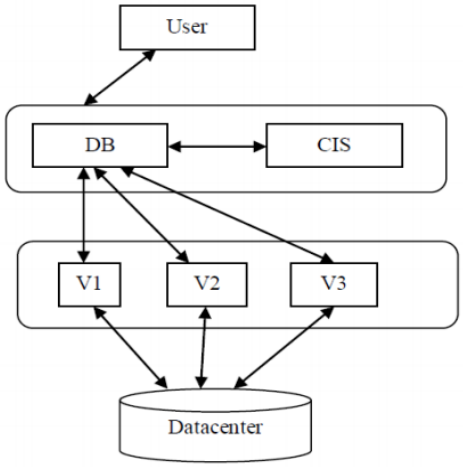

Fig 2: Stages of Scheduling

DB: - Datacenter Broker
CIS: - Cloud Information services
V1,V2,V3 are the virtual machines

Scheduling Types

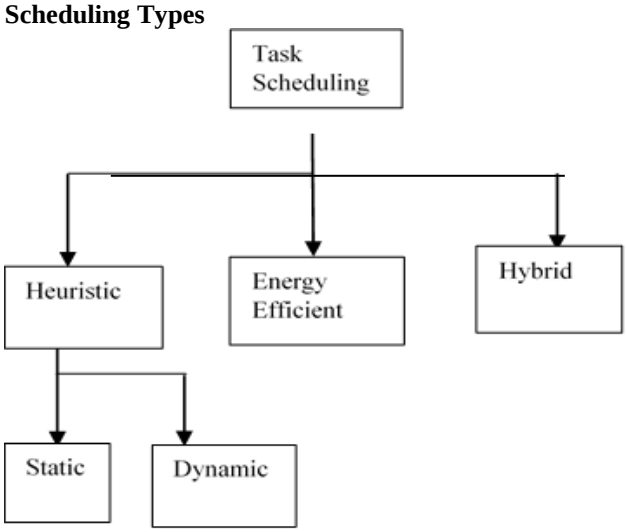

Fig3: Types of Scheduling

1.Heuristic Task Scheduling Approaches

Heuristics planning [1], [2], [3]; gives an ideal arrangement in which it utilizes the knowledge bases for taking the scheduling decisions. Heuristic methodologies can be either static or dynamic. To start with lets look at the static scheduling algorithms.

Static Scheduling Methods: Static scheduling algorithms consider that all tasks come at the same moment of time and they are free of the system resource's states and their availability. The static heuristics incorporate the fundamental straightforward planning procedures like First Come First Serve and Round Robin techniques. FCFS strategies gathers the tasks and queues them until resources

are available and once they get to be available the tasks are assigned to them taking into account their arrival time. It is less complex in nature however does not consider some other criteria for scheduling the tasks to machines.

Dynamic Scheduling Methods: In dynamic scheduling techniques [4],[5],[14] tasks are dynamic in nature. Here tasks land at various purposes of time and it is reliant on the system machine's state. Dynamic scheduling algorithms are grouped into two classes:online mode and batch mode. In online mode tasks are allocated out in a split second once they reach in the system like most-fit task scheduling algorithm where as in batch mode tasks are gathered as a group and planned at predefined times. Min-min, max-min, round robin are a few case for batch mode. MCT, MET, OLB has a place with online mode, and works like static algorithms.

2.Energy Efficient Task Scheduling Approaches

The management of power of a data center depends up on different elements and task scheduling is a huge one among them.The different task scheduling algorithms that basically concentrates on the reducing power consumption, increasing energy efficiency, performance improvement and cost reduction are included in this category[6][7][8][9][10].

3.Hybrid Scheduling Algorithms

Many of these algorithms are novel or are produced on the top of some current strategies joining more planning parameters to enhance the performance[11].

III Study of Existing Task Scheduling Algorithms

Min-Min and Max-Min [1] are two other heuristic techniques utilized for task scheduling. Min-min heuristic chooses the task with lowest time first from all the available task and allots it to a machine which gives the lowest completion time (quickest machine) for that assignment. It expands the aggregate completion time of all the tasks and consequently affects the makespan. Yet, it doesn't consider load of the machines before scheduling as basically allocating smaller tasks on faster machines. Here the expected finish time and execution time for a task are thought to be practically same values or close values. The long tasks need to wait to complete the execution of smaller ones. In any case, the strategy enhances the system's general throughput.

Max-Min is like min-min aside from that it chooses the longest task(with maximum completion time) first to schedule on the best machine accessible on the basis of minimum completion time of that specific task on all available machines. Here the smaller tasks need to starve and load balancing is likewise not considered. Anyways it expands the makespan and system throughput than the min-min procedure since the longest task decides the makespan of all the available tasks in the system. Subsequently in max-min the longer tasks can be executed first in speedier machines and smaller tasks can be executed in parallel on other conceivable machines which results in preferable makespan and adjusted burden over the previous technique.

Genetic Algorithm and Simulated Annealing are two other general techniques in heuristic methodology which is utilized to perform near ideal scheduling.In Genetic Algorithm approach [2] we perform four unique operations, evaluation, selection, cross over and mutation.The underlying population speaks to the possible mappings of the given tasklist on the available machines. Every task is entitled as a vector in which every position of that vector represents to a task in the task list. The worth in every position represents the machine to which the task is mapped. Every task represents a chromosome. Each chromosome has a fitness value demonstrating the general execution time of all the tasks (makespan) which is framed from the mapping of tasks to resources constituting that chromosome and it is chosen such that it decreases makespan.This technique utilizes past results with present results to show signs of improvement mappings and survival of the fittest happens.

Simulated Annealing[3] is an iterative technique which can be represented like genetic algorithm in which it begins with a single solution(mapping) chosen from an arbitrary distribution. The underlying version of SA is assessed to show signs of improvement. After mutation the new makespan is analyzed.On the off chance that it is lower (better) than the previous one then place the old one with the new makespan. Simulated Annealing finds poorer arrangements than Genetic Algorithm. The features of genetic algorithm and simulated annealing can be joined to show improvement scheduling solution.

At present, the scheduling plans ought to consider the associate tasks scheduling issue with a few constraints in the real applications. In this paper, concerning the postponement of the associated tasks scheduling in cloud processing, a structured based hierarchical task models is discussed about and the associated task scheduling algorithms in view of deay bound limitation (SAH-DB) was proposed[4]. The scheduling plan in view of the tasks model can enhance the task execution concurrency. The tasks in the parallel structure can be assembled into one task set belonging to same task layer. Through the estimation of the aggregate tasks execution timebound in every task layer, the associated task are dispatched to the resources with the base execution time. Broad exploratory results showed that the proposed SAH-DB calculations can accomplish preferred execution over CPM and TSSim algorithm in the terms of the total execution expense and resource utilization.

In [5] three possible methodologies proposed for dynamic task scheduling in cloud computing are explored. The three methodologies are having a place with the field of swarm intelligence that is utilized to discover answers for troublesome or inconceivable combinatorial issues. These methodologies are inspired by ant colony nature, the behaviour of particle swarm and honeybee foraging behavior. The principle objective is to give an assessment and near investigation of these methodologies that are utilized to minimize the makespan of a given tasks set. Performance of the algorithms is simulated using toolbox of CloudSim. Algorithms have been compared with each other and with the existing algorithms for dynamic task scheduling issue. The output of the analyses are introduced and the qualities of every algorithm is researched. Experimental results demonstrate that the proposed approaches fulfill expectation, additionally demonstrated that ABC calculation is the prevalent than other algorithms.
In [6];three algorithms are given which mainly concentrates on the most proficient method to handle a solicitation from the clients in heterogeneous frameworks.The first is an benefit driven algorithm in which the tasks are allocated on the best server machines in view of benefit value calculated.This works for heterogeneous systems. For homogeneous frameworks here they are proposing two strategies: control best fit calculation in which they consider the machine with least power consumption increment as its decision for scheduling the task.Also, the other one is load balancing approach in which load balancing is done in view of the power frequency ratio of every resource. Power frequency ratio shows the processing limit of the server.
In [7];energy efficient job scheduling considering the traffic load balancing in cloud datacenters is concentrated. They look on the traffic requirements of the cloud applications. Thus it reduces congestion and communication delays in the network.
In [8];scheduling of tasks is finished by consolidating network awareness and energy efficiency. It fulfills QoS prerequisites and enhances job performance. It diminishes the number of figuring servers and keeps away from hotspots. Network awareness is achieved by utilizing feedback channels from the fundamental network switches. This technique has less computational and memory overhead.
In [9];an improved planning technique is actualized to decrease power utilization alongside fulfilling undertaking response time constraints during scheduling. It is a greedy approach which chooses least number of most proficient servers for scheduling. The tasks are heterogeneous in nature with the goal that they constitute diverse energy consumption levels and have different task response times. Ideal task depends on minimum energy consumption and minimum finish time of a task on a specific machine.
In [10]; an environment friendly strategy for scheduling utilizing the DVFS procedure is proposed. Utilizing Dynamic Voltage Frequency Scaling strategy it lessens the power utilization of infrastructure. Minimizing number of processing servers and time decreases energy usage and can enhance resource utilization. The servers are kept running at various mixes of frequencies and voltages. This technique proficiently plans the tasks to resources without compromising the performance of the system meeting the SLA prerequisites and saving energy.
In [11];task's priority is ascertained for planning them on different resources. In view of the distinctive qualities of the tasks,priorities are calculated for the tasks and they are sorted taking into account that.At that point they are alloted on the machine which delivers the best completion time.Consequently this algorithm enhances execution by having better completion time.
In [12];tasks are scheduled in view of their cost to various resources.The cost of services shifts for various tasks based on their complexity.Thealgoritms considers resource cost and processing capacity of resources.They group tasks in view of the processing limit and chooses some best resources to schedule them in such an approach to lessen cost.This algorithm minimizes the makespan and the processing cost when contrasted with other scheduling technique.
In [13];an algorithm based on conventional minmin algorithm which incorporates scheduling on the basis of load of the servers and considering the user need is build. The users are grouped into two classes as VIP and normal users.Load is adjusted taking into account the maximum loaded resource and the makespan of the system The strategy demonstrates a decent pick up in user satisfaction, makespan and resource usage ratio.
With a specific goal to take care of the issue of task scheduling in cloud computing, this paper proposes a novel dynamic task scheduling algorithm in view of improved genetic algorithm[14].Based on Genetic algorithm,the proposed algorithm gives full control to the dynamic characteristics of the cloud computing environment. The CloudSim simulation platform is chosen for simulation; experimental results demonstrate that the proposed algorithm can adequately enhance the throughput of cloud computing system, and can significantly minimize the execution time of task scheduling.

Conclusion

With the emerging of cloud computing, cloud workflow systems are designed to facilitate the cloud infrastructure to support large scale distributed collaborative e-business and e-science applications. The management and scheduling of resources in Cloud environment is complex, and therefore demands sophisticated tools for analysis the algorithm before applying them to the real system. In this

paper, we have surveyed the various existing task scheduling algorithms in cloud computing and tabulated their various parameters along with tools and so on.

References

[1] Tracy D. Braun, Howard Jay Siegel, Noah Beck, "A Comparison of Eleven Static Heuristics or Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems", Journal of Parallel and Distributed Computing 61, pp. 810-837, 2001.

[2] Sung Ho Jang, Tae Young Kim, Jae Kwon Kim , Jong Sik Lee School, "The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing", International Journal of Control and Automation,pp 157-162,2012.

[3] M. Coli, P. Palazzari, "Real Time Pipelined System Design through Simulated Annealing," Journal of Systems Architecture, vol.42, no. 6-7, 1996, pp. 465-475.

[4] Mao Yingchi, Xu Ziyang, Ping, Wang Longbao," Delay-Aware Associate Tasks Scheduling in the Cloud Computing ",IEEE Fifth International Conference on Big Data and Cloud Computing (BDCloud),pp**:**104 – 109,2015

[5] Gamal F. Elhady and Medhat A. Tawfeek," A Comparative Study into Swarm Intelligence Algorithms for Dynamic Tasks Scheduling in Cloud Computing", IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS),pp 362 – 369,2015

[6] WeichengHuai, Zhuzhong Qian, Xin Li, Gangyi Luo, and Sanglu Lu, "Energy Aware Task Scheduling in Data Centers, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications", volume: 4, number: 2, pp. 18-38, 2013.

[7] Dzmitry Kliazovich1, Sisay T. Arzo, FabrizioGranelli, Pascal Bouvry and Samee Ullah Khan, "e-STAB: Energy-Efficient Scheduling for Cloud Computing Applications with Traffic Load Balancing", IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing e-STAB,pp**:**7 – 13, 2013.

[8] DzmitryKliazovich, Pascal Bouvry, Samee Ullah Khan, "DENS: Data Center Energy-Efficient Network-Aware Scheduling, Cluster Computing", special issue on Green Networks, vol. 16, no. 1, pp. 65-75, 2013.

[9] Ning Liu, Ziqian Dong, Roberto Rojas-Cessa, "Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers", IEEE 33rd International Conference on Distributed Computing Systems Workshops Task, 2013.

[10] Chia-Ming Wu, Ruay-Shiung Chang, Hsin-Yu Chan received, "A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters", Elsevier,pp:141–147, 2014.

[11] Xiaonian Wu, Mengqing Deng, Runlian Zhang, Bing Zeng, Shengyuan Zhou, "A Task Scheduling Algorithm based on QoS driven in Cloud Computing", Information Technology and Quantitative management, 2013.

[12] Mrs. S. Selvarani, Dr. G. Sudha Sadhasivam, "Improved Cost-Based Algorithm For Task Scheduling in Cloud Computing", IEEE,pp:5, 2010.

[13] Huankai Chen, Frank Wang, Na Helian, GbolaAkanmu, "User-Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing", IEEE, pp**:**1 – 8, 2013.

[14] H. Izakian, A. Abraham, V. Snasel, "Comparison of Heuristics for Scheduling Independent Tasks on Heterogeneous Distributed Environments, "International Joint Conference on Computational Sciences and Optimization, IEEE, pp. 8-12,2009