LITERATURE REVIEW: VARIOUS PRIORITY BASED TASK SCHEDULING ALGORITHMS IN CLOUD COMPUTING

[1]MS. POOJA .P.VASANI, [2]MR. NISHANT .S. SANGHANI
[1]M.Tech. [Software Systems] Student, Patel College of Science and Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, M.P., India

[2]M.E. [Computer Engineering] Student, Computer Engineering Department, V.V.P. Engineering College, Rajkot, Gujarat, India

poojavasani@ymail.com, nishantssn@yahoo.co.in,

ABSTRACT: *Cloud computing is a latest new computing paradigm where applications, data and IT services are provided over the Internet. Cloud computing serves different types of the resources in virtualized form which can be utilised dynamically. There are different number of issues which can be researched out for the proper allocation and better utilization of the virtualized resources using scheduling. In the process of scheduling some intensive data or computing an intensive application, it is acknowledged that optimizing the transferring and processing time is crucial to an application program. The Task management is the key role in cloud computing systems. Task scheduling problems are premier which relate to the efficiency of the whole cloud computing facilities. The main aim of this work is to study various problems, issues and types of task scheduling algorithms for cloud computing.*

Keywords — Cloud Computing, Task Scheduling.

1. INTRODUCTION

The cloud computing is a large group of interconnected computers and cloud symbol represents a group of systems or complicated networks. Cloud computing is one way of communication among the various system in the network with the help of internet. [9]

Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application, data and file storage. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. [8]

Cloud computing is a marketing term for technologies that provide computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. [1]

Cloud computing is "on demand resources provisioning" which means to provide the available resources based on the requirement of the resources. Cloud computing is a "subscription based". Cloud computing is a pay as per usage and reliable leads to an efficient network. Cloud Computing is an emerging technique and it's very successful because of the following features like reliable, secure, fast, fault tolerance and efficient communication etc., among different network.[9]

Scheduling is used to allocate particular resources for a certain tasks in particular time. Task scheduling problem is a core and challenging issue in cloud computing. The Task execution time cannot be predicted in cloud computing. Hence the scheduler must be dynamic. The purpose of scheduling is to increase the utilization of resources. [9]

Scheduling process in cloud can be generalized into three stages namely–

Resource discovering and filtering- Datacenter broker discovers the resources present in the network system and collects status information related to them.
Resource selection- Target resource is selected based on certain parameters of task and resource. This is a deciding stage.
Task submission- Task is submitted to resource selected. [10]

CLOUD SERVICES

Cloud computing can be thought as different layers or models which provide different services. Cloud contains three types of services as follows.

1. Infrastructure-as-a-Service(IaaS)- This type of cloud computing distributes a full computer infrastructure via the Internet. Most popular IaaS provider like Amazon Web Services offers virtual server instances with unique IP addresses and block of storage on demand. Here customers usually use the service provider's application program interface to start, stop, access, modify and configure their virtual servers and storage as is needed. In the enterprise, cloud computing allocates services to a company to pay for only as much facility as is required, and bring more flexible tools online as soon as required. [11]

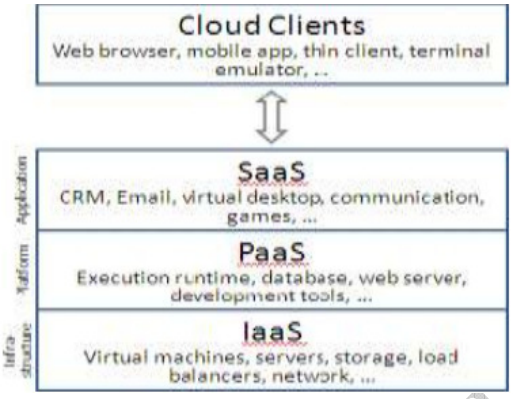

Fig1. Cloud Services [11]

2. Platform-as-a-Service(PaaS)- This type of cloud computing offers a product development tool or environment that users can access and utilize online, even in collaboration with others and hosted on the provider's infrastructure. In PaaS, developers create applications on the service provider's platform over the Internet. PaaS service providers may use Application Program Interfaces (APIs), gateway software or website portals installed on the customer's premises. [11]

3. Software-as-a-Service (SaaS) - This type of cloud computing model offers users the hardware infrastructure, the software product and interrelates with the users through a front-end gateway or portal. Here a provider authorizes an application to clients either as a service on demand in a "pay-as-you-go" model or at no charge by a subscription. Like we can say that we have a sky drive in Hotmail. There we can save our word, PowerPoint document and when we need to edit them we can easily do it. [11]

CLOUD DEPLOYMENT MODELS

Cloud Computing technology and services can be implemented in different ways according to their purpose and characteristics. These different types of deployment of Cloud are categorized in four ways as follows. [1-7]

1. Public Cloud: A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Public cloud services may be free or offered on a pay-per-usage model.

2. Community cloud: Community cloud shares infrastructure between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so only some of the benefits of cloud computing are realized.

3. Hybrid cloud: Hybrid cloud is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. It can also be defined as multiple cloud systems that are connected in a way that allows programs and data to be moved easily from one deployment system to another.

4. Private cloud: Private cloud is infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally.

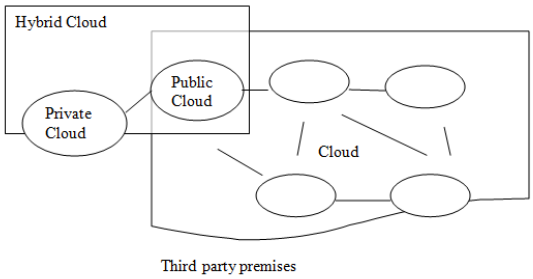

Fig2. Cloud deployment models [1-7]

2. ALGORITHMS AND METRICS

A. Tasks Scheduling optimization for the Cloud Computing Systems

Task scheduling algorithm is an NP-completeness problem which play key role in cloud computing. In Hadoop, the open-source implementation of MapReduce, many scheduling policies such as FIFO scheduling is used by the master node to distribute waiting tasks to computing nodes. So it can still achieve more improvement in the task scheduling process. Therefore, in this paper, we purpose an optimized algorithm based on the Fuzzy-GA optimization which makes a scheduling decision by evaluating the entire group of task in the job queue. [12]

Here, the model works on the following aspects

- Task Model and Related Properties

Suppose that a cloud computing system consist of heterogeneous process unit possess of m (m>1) units. The tasks in this system features are as follows:

a) Tasks are aperiodic; i.e., task arrival times are not known a priori. Every task has the attributes arrival time (ai), worst case computation time (ci), and deadline (di). The ready time of a task is equal to its arrival time.

b) Tasks are nonpre-emptive; each of them is independent.

c) Each task has two ways of access to a process unit:(1)exclusive access, in which case no other task can use the resource with it, or (2) shared access, in which case it can share the resource with another task (the other task also should be willing to share the resource).

- Predicted Execution Time Mode

The statistical modeling approach - Kernel Canonical Correlation Analysis (KCCA), which representing each Hadoop job as a feature vector of job characteristics and a corresponding vector of performance metrics. From the Hadoop job logs, performance feature vectors, such as map time, reduce time, total execution time, map output bytes, HDFS bytes written, and locally written bytes, could be extracted. These results convincingly demonstrate the effectiveness of this approach which can also be used for our mechanism of MapReduce scheduling optimization.

- System Model

The system model describes the data store and computing cluster that jobs could be assigned to the cluster includes machines arranged in a general tree-shaped switched network.

- Reschedule

When a task cannot be completed due to a disk failure, processor failure, or other problems, the job may never get complete. A time out Tout could be set for this situation and the non-completed tasks could be rescheduled in the next calculation. Other than failures, when new jobs arrive, the new jobs also need to be scheduled with uncompleted tasks.

- Objective Function

The objective function for our algorithm is the latest completion time of the task schedule, referred as Makespan. The Makespan is calculated in objective function. Where t represents the time that processor i will have finished the previously assigned jobs and E[t][i] is the predicted execution time that task t is processed on processor i.

- Scheduling Optimizer

A master processor unit in cloud, collecting all tasks, will take charge of dispatching them to other process units. Each process unit has its own dispatch queue (DQ). The master unit communicates with other process units through these dispatch queues. This organization ensures that processor units always find some tasks in the dispatch queue when it finishes the execution of their current task. The master unit works in parallel with other units, scheduling the newly arrived tasks, and periodically updating the dispatch queues. Tasks are sorted ascending by the value of deadlines. Reasons to choose GA as an optimization algorithm is simplicity of operation and power of effect. It is suitable to some NP-hard problems.

  - Fuzzification

The Fuzzification comprises the process of transforming values into grades of membership for linguistic terms of fuzzy sets. The membership function is used to associate a grade to each linguistic term.

  - Genome representation

We use direct representation. For job scheduling problem, a direct representation is obtained as follows. Feasible solutions are encoded in a vector, called schedule, of size n tasks, where the numbers indicates the slot where task i is assigned by the schedule. Thus, the values of this vector are natural numbers included in the range [1; m slots].

  - Genetic Operators

In the initialization process, the slot number (node number) which has the data for the task is chosen as the initial schedule. If there are several choices for one task, the slot number is randomly selected among these slots.

Algorithm

I. Get new tasks to be scheduled. The tasks to be scheduled include the uncompleted task and new jobs. But if jobs arrive in dynamically and make too many jobs waiting to be assigned at one time, the sliding-window technique could be used as an option. The window size is fixed. Tasks fall into the sliding window will get scheduled.
II. Generating E matrix for the job Using KCCA technique to predict the execution time of any individual task assigned to every node.
III. Get the current state of the system.
IV. Fuzzification of all above parameter to get optimized task schedule.

V. The Fuzzify parameter Map in GA to get optimized.
   a. Generate an initial population of chromosomes randomly.
   b. Evaluate the fitness of each chromosome in the population. Evaluate P according to information in E
   c. Create a new population by repeating the following steps until the new population is complete, Selection: Select two parent chromosomes from a population according to their fitness. (The better the fitness, the higher is the chance for getting selected). Crossover With a crossover probability, do cross over operations on the parents to form new offspring. If no crossover is performed, offspring is the exact copy of the parents. Mutation with a mutation probability, mutate new offspring at each locus (Position in chromosome) Acceptance Place the new offspring in the new population.
   d. Using the newly generated population for a further sum of the algorithm.
   e. If the test condition is satisfied, stop and return the best solution in the current population.
   f. Repeat Step c until the target is met.

VI. Finally obtain the optimal solution.

B. Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm

In this paper in order to minimize the cost of the processing a model for task scheduling and propose a particle swarm optimization (PSO) algorithm which is based on small position value rule is formulated. By virtue of comparing PSO algorithm with the PSO algorithm embedded in crossover and mutation and in the local research, the experiment results show the PSO algorithm not only converges faster but also runs faster than the other two algorithms in a large scale. The experiment results prove that the PSO algorithm is more suitable to cloud computing. [14]

The PSO Algorithm

The algorithm begins with k random particle vector and each particle is in n dimensions. Every particle vector is a candidate solution of the underlying problem. The particles are the task to be assigned and the dimensions of the particle are the number of the special tasks in a workflow. Then, each particle moves by the direction on the pbest and gbest until the maximal number of iterations. When the algorithm executes over, the gbest and fitness value are the corresponding task scheduling and the minimal cost of the optimal strategy.

PSO algorithm

1. Initialize particle position vector and velocity vector randomly. The vector's dimension equal to the size of the special tasks.
2. Convert the continuous position vector ($x_k^i = [x_1^i, x_2^i, \ldots, x_n^i]$) to discrete vector ($s_k^i = [s_1^i, s_2^i, \ldots, s_n^i]$) in light of SPV rule. Then, transform the ($s_k^i = [s_1^i, s_2^i, \ldots, s_n^i]$) to processor's vector ($p_k^i = [p_1^i, p_2^i, \ldots, p_n^i]$).
3. If one particle's fitness value is better than current, setting current value replace previous pbest and as the new pbest.
4. Selecting the best particle from the entire particle as the gbest.
5. For all particles update their position and velocity.
6. If reaching to the maximum iteration or getting the ideal result stops, otherwise repeating from Step 2.

C. Comparison of load balancing algorithms in a Cloud

Here the jobs are submitted by the clients to the computing system. As the submitted jobs arrive to the cloud they are queued in the stack. The cloud manager estimates the job size and checks for the availability of the virtual machine and also the capacity of the virtual machine. Once the job size and the available resource (virtual machine) size match, the job scheduler immediately allocates the identified resource to the job in queue. The impact of the ESCE algorithm is that there is an improvement in response time and the processing time. The jobs are equally spread, the complete computing system is load balanced and no virtual machines are underutilized. Due to this advantage, there is reduction in the virtual machine cost and the data transfer cost. [13]

ESCE LOAD ALGORITHM

ACTIVE VM LOAD BALANCER
[START]
Step1:- find the next available VM
Step2:-check for all current allocation count is less than max length of VM list allocate the VM
Step3:- if available VM is not allocated create a new one in
Step 4:- count the active load on each VM
Step5:- return the id of those VM which is having least load
[END]

D. A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment

This paper proposes a scheduling algorithm which addresses these major challenges of task

scheduling in cloud. The incoming tasks are grouped on the basis of task requirement like minimum execution time or minimum cost and prioritized. Resource selection is done on the basis of task constraints using a greedy approach. [10]

Algorithm

An optimum scheduling algorithm is proposed and Implemented in this section. The proposed algorithm works as follows

1. Incoming tasks to the broker are grouped on the basis of their type– deadline constrained or low cost requirement.
2. After initial grouping they are prioritized according to deadline or profit. This is required because the tasks with shorter deadline need to be scheduled first and similarly the tasks resulting in more profit should be scheduled on lost cost machines. Thus, the prioritizing parameter is different based on the nature or type of task.
3.

A. For each prioritized task in deadline constrained group –

i) Turnaround time at each resource is calculated taking following parameters into account.

- Waiting time
- Task length
- Processing Power of virtual machine

ii) The virtual machine with minimum turnaround time that is capable to execute the task is selected and task is scheduled for execution on that machine.

iii) Waiting time and resource capacity of selected machines are updated accordingly.

B. For cost based group

i) Virtual Machines are selected on the basis of processing power of machine and its cost

ii) For each virtual machine cloudlets from the group are scheduled till the resource capacity is permitted.

iii) Resource capacity and waiting time are updated accordingly.

3. CONCLUSION

Cloud task scheduling algorithms are designed to support large scale applications and to support cloud infrastructure. Various algorithms are proposed in order to achieve proper scheduling in cloud computing environment through the better resource allocation and optimization. Better performance and efficiency can be achieved by applying various different proposed techniques discussed above.

In this paper, analysis and overview of various different task scheduling algorithms is carried out which uses different parameters like time, cost, and make span, speed, scalability, resource allocation, scheduling and sophisticated tools.

4. REFERENCES

[1] "Cloud computing resources". http://www.cloud9s.net/cloudcomputingresources.html

[2] http://www.thecloudcomputing.org/2011 http://salsahpc.indiana.edu/CloudCom2010/papers.html

[3] http://www.itworld.com/internet/69141/5-cool-cloud-computing-research-projects

[4] http://en.wikipedia.org/wiki/Cloud_computing

[5] http://www.thecloudcomputing.org/ 2012/history.html

[6] http://www.whereisdoc.co

[7] Torry haris "CLOUD COMPUTING – An Overview"

[8] V. Venkatesa Kumar and K. Dinesh "Job Scheduling Using Fuzzy Neural Network Algorithm in Cloud Environment" Bonfring International Journal of Man Machine Interface, Vol. 2, No. 1, March 2012

[9] Monika Choudhary and Sateesh Kumar Peddoju "A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment" International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, May-Jun 2012, pp.2564-2568

[10] Sudhir Singh "Performance Optimization in Gang Scheduling In Cloud Computing" IOSR Journal of Computer Engineering (IOSRJCE), ISSN: 2278-0661 Vol. 2, Issue 4, July-Aug. 2012, pp. 49-52

[11] Sandeep Tayal "Tasks Scheduling optimization for the Cloud Computing Systems" International Journal Of Advanced Engineering Sciences And Technologies, Vol No. 5, Issue No. 2, pp. 111 – 115

[12] Jaspreet kaur "Comparison of load balancing algorithms in a Cloud" International Journal of Engineering Research and Applications, Vol. 2, Issue 3, May-Jun 2012, pp.1169-1173

[13] Lizheng Guo, Shuguang Zhao, Shigen Shen, Changyuan Jiang "Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm", JOURNAL OF NETWORKS, VOL. 7, NO. 3, MARCH 2012