Analysis of Job Scheduling Algorithms in Cloud Computing

Rajveer Kaur[1], Supriya Kinger[2]

[1]*Research Fellow, Department of Computer Science and Engineering, SGGSWU, Fatehgarh Sahib, India, Punjab (140406)*

[2]*Asst.Professor, Department of Computer Science and Engineering, SGGSWU, Fatehgarh Sahib, India, Punjab (140406)*

Abstract – *Cloud computing is flourishing day by day and it will continue in developing phase until computers and internet era is in existence. While dealing with cloud computing, a number of issues are confronted like heavy load or traffic while computation. Job scheduling is one of the answers to these issues. It is the process of mapping task to available resource. In section (1) discuss about cloud computing and scheduling. In section (2) explain about job scheduling in cloud computing. In section (3) existing algorithms for job scheduling are discussed, section (4) existing algorithms are compared and lastly section (5) conclusion and future work are discussed.*

Keywords- *Cloud Computing, Job Scheduling, FCFS, SJF, Round Robin, Priority, GA.*

1. Introduction

Cloud computing is an Internet-based development and use of computer technology. Cloud computing is one of the upcoming latest new computing paradigm where applications and data services are provided over the Internet. At this time most of the business organizations and educational institutions use cloud environment. Cloud computing is service focused to provide high quality and low-cost information services by pay-per-use model in which guarantees are offered by the cloud service providers [1]. Cloud Computing is the use of computing resources such as hardware and software that are delivered as a service over the internet to the customers. It is a great idea to make many normal computers together to get a super computer and this computer can do a lot of things. It is known to be the concept of cloud computing.

Cloud Computing is a platform which aims to provide shared data to its clients at the same time. It is amongst the buzzwords in today's era. Whether we open an IT magazine or open any website, cloud computing concept is everywhere. As the count of clients for the access of same data increases, catastrophe may occur. Cloud Computing offers various service models. It can be software as a service model, offering software on a single platform. It can be platform as a service model which offers a platform from where the software and data can be accessed. Or it can be infrastructure as a service which provides the security and backup services.

User accesses the cloud services through internet by using Mobile, PC and PDA as shown in fig. 1. Service provider provides the service to user. These services are [2] Infrastructure as a service (IaaS) refers to the sharing of hardware resources for executing services, typically using virtualization technology**.** The others are Platform as a Service (PaaS) approach where offering includes a software execution environment, such as an application server. In the Software as a Service approach (SaaS), complete applications are hosted on the Internet so

that e.g. your word processing software isn't installed locally on your PC anymore but runs on a server in the network and is accessed through a web browser.

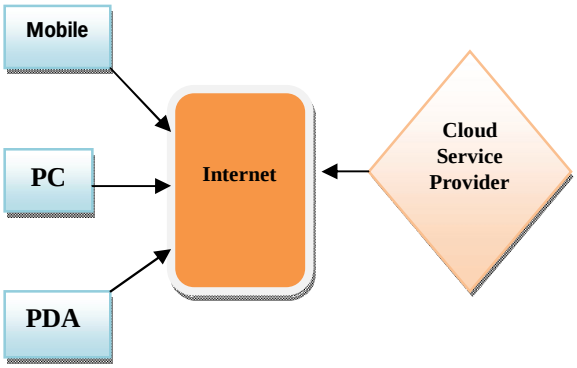

Fig 1: Cloud Architecture

Cloud computing is recently a booming area and has been emerging as a commercial reality in the information technology domain. However the technology is still not fully developed. There are still some areas that are needed to be focused on.

- Resource Management
- Task Scheduling

Task scheduling and provision of resources are main problem areas in both Grid as well as in cloud computing. Cloud computing is emerging technology in IT domain. The scheduling of the cloud services to the consumers by service providers influences the cost benefit of these computing paradigms.

Scheduling is the process of allocating tasks to available resources on the basis of tasks' qualities and need [3]. The main goal of scheduling is increased utilization of the resources without affecting the services provided by cloud.

There are two types of scheduling i.e. resource scheduling and job scheduling. Following are some needs of scheduling in cloud computing .

- Fair resource allocation – Scheduling is done so that allocation of resources is done in fair manners.
- QOS – Resources and jobs are scheduled in such a way so that quality of services is achieved.
- Resource utilization - It is the degree to which the resources of the system are utilized. A good scheduling algorithm provides maximum resource utilization.
- Energy consumption – It is the degree to which the resources of system are consumed. A good scheduling algorithm saves energy consumption.

Scheduling process in cloud is divided into three stages [4] namely:

- Resource discovering and filtering- Datacenter Broker discovers the resources present in the network system and collects status information about the resources.
- Resource selection - Target resource is selected based on certain requirements of task and resource. This is deciding stage.
- Task allocation -Task is allocated to resource selected.

2. Job Scheduling

The Job management is the fundamental concept of cloud computing systems task scheduling problems are main which relates to the efficiency of the whole cloud computing system. Job scheduling is a mapping mechanism from users' tasks to the appropriate selection of resources and its execution. Job scheduling is flexible and convenient. Jobs and job streams can be scheduled to run whenever

required, based on business functions, needs, and priorities. Job streams and processes can set up daily, weekly, monthly, and yearly in advance, and run on-demand jobs without need for assistance from support staff [5].

2.1 Characteristics of the Job Scheduling:

- Job scheduling is global centralized - As cloud computing is a computing model which supply the centralized resource by the mirror service to multiple distributed applications, and this mirroring deployment can make heterogeneous procedures' executing of interoperate become easier, which used to be difficult to deal with. Therefore, virtualized technology and mirroring services make the task scheduling of cloud computing achieve a global centralized scheduling.
- Each node in the cloud is independent - In cloud computing, the internal scheduling of every cloud node is autonomous, and the schedulers in the cloud will not interfere with the scheduling policy of these nodes.
- The scalability of job scheduling - The scale of resources supply from cloud provider may be limited in early stages. With the addition of a variety of computing resources, the size of the abstract virtual resources may become large, and the application demand continues increasing. In the cloud, task scheduling must meet the scalability features, so that the throughput of the task scheduling in the cloud may not be too low.
- Job scheduling can be dynamically self-adaptive - Expanding and shrinking applications in the cloud may be necessary depend on the requirement. The virtual computing resources in cloud system may also expand or shrink at the same time. The resources are constantly changing, some resources may fails, new resources may join in the clouds or restart.
- The set of job scheduling - Task scheduling is divided into two parts: one is used as a unified resource pool scheduling, and primarily responsible for the scheduling of applications and cloud API; the other is for the unified port resource scheduling in the cloud, for example, MapReduce task scheduling. However, each scheduling consists of two two-way processes that are scheduler leases resource from cloud and scheduler callbacks the requested resources after use. The former process is scheduling strategy and the latter one is callback strategy. The combination of the scheduling and callback resource strategy is the set of task scheduling.

2.2 Need of Job Scheduling

The task scheduling goals of Cloud computing is provide optimal tasks scheduling for users, and provide the entire cloud system throughput and QoS at the same time [5]. Following are the needs of job scheduling in cloud computing:

- Load Balance - Load balancing and task scheduling has closely related with each other in the cloud environment, task scheduling mechanism responsible for the optimal matching of tasks and resources. Task scheduling algorithm can maintain load balancing. So load balancing become another important measure in the cloud.

- Quality of Service - The cloud is mainly to provide users with computing and cloud storage services, resource demand for users and resources supplied by provider to the users in such a way so that quality of service can be achieved. When job scheduling management comes to job allocation, it is necessary to guarantees about QoS of resources.
- Economic Principles - Cloud computing resources are widely distributed throughout the world. These resources may belong to different organizations. They have their own management policies. As a business model, cloud computing according to the different requirements, provide relevant services. So the demand charges are reasonable.
- The best running time - jobs can be divided into different categories according to the needs of users, and then set the best running time on the basis of different goals for each task. It will improve the QoS of task scheduling indirectly in a cloud environment.
- The throughput of the system - Mainly for cloud computing systems, throughput is a measure of system task scheduling optimizing performance, and it is also a target which has to be considered in business model development. Increase throughput for users and cloud providers would be benefit for them both

3. Existing Job Scheduling Algorithms

Job Scheduling Algorithms are as follows:

3.1 First Come First Serve Scheduling Algorithm – It is also known as First in First out [7]**.** Shortest job next is advantageous because of its simplicity and because it minimizes the average amount of time each process has to wait until its execution is complete. It is one of the simplest Scheduling algorithms we have it allocate the CPU in the order in which the process arrive. It assumed that ready queue is managed as first in first out which means that the first job will be processed first without other preferences.

Algorithm FCFS:

- Initialize Tasks.
- First task assigned to the queue and add tasks up to n numbers.
- Add next task 'I' at last position in the main queue.

3.2 Shortest Job First Scheduling Algorithm - Shortest job First (SJF) also known as Shortest Job Next (SJN) or Shortest Process Next (SPN) is a scheduling technique that selects the job with the smallest execution time [8]. The jobs are queued with the smallest execution time placed first and the job with the longest execution time placed last and given the lowest priority. This Scheduling algorithm is deal with different approach in this algorithm CPU is allocated to the process with least burst time.

Algorithm SJF:

- for i = 0 to i < main queue-size
  - if task $_{i+1}$ length < task $_{i}$ length then
    - add task $_{i+1}$ in front of task $_{i}$ in the queue
  - end if
  - if main queue-size = 0 then
    - task $_{i}$ last in the main queue
  - end if
- end for

3.3 Round-Robin Scheduling Algorithm - It is one of the oldest, simplest, fairest and most widely used scheduling algorithms, designed especially for time-sharing systems [3]. A small unit of time, called time slices or quantum is defined. All run able processes are kept in a circular queue. The CPU scheduler goes around this queue, allocating the CPU to each process for a time interval of one quantum. New processes are added to the tail of the queue. The CPU scheduler picks the first process from the queue, sets a timer to interrupt after one quantum, and dispatches the process .If the process is still running at the end of the quantum, the CPU is preempted and the process is added to the tail of the queue. If the process finishes before the end of the quantum, the process itself releases the CPU voluntarily.

Algorithm RRS:

- Keep the ready queue as a FIFO queue of processes.
- New processes are added to the tail of the ready queue.
- The CPU scheduler picks the first process from the ready queue, sets a timer to interrupt after 1 time slot, and dispatches the process.
- The process may have a CPU burst of less than 1 time quantum.
    - In this case, the process itself will release the CPU voluntarily.
    - The scheduler will then proceed to the next process in the ready queue.
- Otherwise, if the CPU burst of the currently running process is longer than 1 time quantum,
    - the timer will go off and will cause an interrupt to the OS.
    - A context switch will be executed, and the process will be put at the tail of the ready queue.
    - The CPU scheduler will then select the next process in the ready queue.

3.4 Priority Scheduling Algorithm - This Scheduling algorithm is preemptive in which all things are based on the priority in this scheduling algorithm each process in the system is based on the priority whereas highest priority job can run first whereas lower priority job can be made to wait, the biggest problem of this algorithm is starvation of a process [7].

Algorithm PSA

- for i = 0 to i <main queue-size
    - if priority (task $_{i+1}$) > priority (task $_{i}$) then
        - add task $_{i+1}$ in front of task $_{i}$ in the queue
    - end if
- end for

3.5 Genetic Algorithm (GA) – Genetic algorithm is a problem solving method that uses genetics as its model of problem solving. It is a search technique to find optimizated solution. GA handles a population of possible solution. Each solution is represented through a chromosome. Genetic algorithm is a method of scheduling in which the tasks are assigned resources according to individual solutions (which are called schedules in context of scheduling), which tells about which resource is to be assigned to which task. Genetic Algorithm is based on the biological concept of population generation. In Genetic Algorithm the initial population is generated randomly. Genetic algorithm is a random searching method [9].

Algorithm GA

- Initialization: Great initial random population
- Evaluate
- Keep the best
- While termination not true do
- Selection
- Crossover
- Mutation
- Evaluate
- Elitist
- Check exit
- End while
- Return mapping result.

In genetic algorithm Fitness Function, Selection, Crossover, Mutation are main terms –

- Fitness Function - The fitness function creates a fitness value of each chromosome. Chromosome is a set of parameters which define a proposed solution to the problem that GA is trying to solve.
- Selection Function – Pick two parent individuals from the population according to the fitness value. Selection operator is used to pick superior and eliminate the inferior.
- Crossover Function - The crossover operators are the most important component of any evolutionary-like algorithm. Crossover function, generate the new offspring by changing to the better state to parents. It also randomly pick two parents chromosome.
- Mutation Function – This function is designed to reduce ideal time of processor that are waiting for the data from other processors. There are several mutation operators like Move, Swap, Move&Swap and Rebalancing.
- Evaluation Function - Evaluation is depended on the execution time and execution cost. Those schedules will be selected for next generation whose execution time and execution cost is less.

4. Comparison of Exiting Algorithms

In this section we compare the existing algorithms on the basis of different parameters as shown in table 1.

Algorithms	Complexity	Allocation	Waiting Time	Type of system
FCFS Algorithm	Simplest Scheduling Algorithm	CPU is allocated in the order in which the processes arrive	More	Suitable for Batch system.
SJF Algorithm	Difficult to understand and code	CPU is allocated to the process with least CPU burst time	Lesser than FCFS	Suitable for Batch system
Priority Algorithm	Difficult to understand	Based on priority, So the higher priority job can run first.	Lesser	Suitable for both Batch and time sharing systems
Round Robin Algorithm	Performance heavily depends upon the size of time quantum	The preemption take place after a fixed interval of time	More than all	Suitable for time sharing system
Genetic Algorithm	Complexity depends on the task to be scheduled	This is a greedy algorithm and pick the best job to allocate the CPU	Waiting time is less	It deals with problem where the search space is large

The above table presents the number of task scheduling algorithms and comparison between them on the basis of complexity, allocation, waiting time and type of system. Complexity defines which type

of algorithm is simple or easy to use in processing. Allocation defines how the jobs are assigned to the resources. Waiting Time defines which of the algorithm takes more time for processing. Type of System defines which algorithm is suitable for which type of system. The FCFS algorithm is a simplest algorithm for scheduling but waiting time to process the tasks is much more in this. This algorithm is work on the basis of first come first serve, it means which process or task arrives first will processed first and it is suitable for batch type of systems. SJF algorithm is not a simple algorithm but their waiting time is less than FCFS. This algorithms process the task first having least CPU burst time and it is also suitable for batch systems. Priority algorithm works on the basis of priority. It is difficult to understand because how priority can be assigned to the task is a difficult task. Here waiting time is less because task with higher priority processed first. It is suitable for both batch and time sharing systems. Round robin algorithm works accordingly its fixed intervals of time. Here waiting time is more than all because after a fixed time interval the next task will execute. So problem faced when one task is very heavy and other one is with very simple and small calculations. Genetic algorithm is a bio inspired artificial intelligent scheme. Its complexity depends on the task. The best selected task executes first so waiting time is less here. This algorithm deals with large problems.

5. Conclusion & Future Work

In this paper, a number of existing algorithms for job scheduling are discussed and also compared with each other. First Come First Serve algorithm has some disadvantages like processing time of each job must be known in advance and it is suitable only for batch process. One of the major drawbacks of this scheme is that the average time is often quite more. In Shorts Job First long jobs may wait longer because it has to wait not only for jobs that are in the system at the time of its arrival, but also for all short jobs that are in the system at the time of its arrival. In Priority Algorithm only higher priority jobs get chance to execute. In Round-robin scheduling, like other first-come, first-served methods, doesn't give special priority to more important tasks. This means an urgent request doesn't get handled any faster than other requests in queue. This describes that there is a need to propose a new scheme which achieves all the objectives and as well as provide better performance. So, our future work will explore existing algorithms and merge them with Genetic Algorithm which is bio inspired artificial intelligence scheme and may achieve the best performance.

References

[1] Savitha. P, J Geetha Reddy ,"A Review Work On Task Scheduling In Cloud Computing Using Genetic Algorithm", International Journal of Scientific & Technology Research Vol. 2, Issue 8, August 2013.

[2] Börje Ohlman, René Rembarz," What Networking of Information Can Do for Cloud Computing", 2009 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises.

[3] Dr Ajay jangra, Tushar Saini," Scheduling Optimization in Cloud Computing," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

[4] Ram Kumar Sharma and Nagesh Sharma, "A Dynamic Optimization Algorithm for Task Scheduling in Cloud Computing With Resource Utilization", International Journal of Scientific Engineering and Technology, Volume No.2, Issue No.10, pp. 1062-1068.

[5] Hong Sun, Shi-ping Chen, Chen Jin, and Kai Guo1,"Research and Simulation of Task Scheduling Algorithm in Cloud Computing", Telkomnika, Vol.11, No.11, November 2013, pp. 6664-6672.

[6] Tejinder Sharma, and Vijay Kumar Banga, "Efficient and Enhanced Algorithm in Cloud Computing", International Journal

of Soft Computing and Engineering (IJSCE), Volume-3, Issue-1, March 2013, pp. 385-390.

[7] Lu Huang, Hai-shan Chen and Ting-ting Hu, "Survey on Resource Allocation Policy and Job Scheduling Algorithms of Cloud Computing" ,Journal of Software, Vol. 8, No. 2, February 2013, pp. 480-487.

[8] Van den Bossche, R., Vanmechelen, K., Broeckhove, J, "Cost Optimal Scheduling in Hybrid IaaS Clouds for Deadline Constrained Workloads", in 3rd IEEE International Conference on Cloud Computing, Miami (July 2010)

[9] Savitha. P, J Geetha Reddy ,"A Review Work On Task Scheduling In Cloud Computing Using Genetic Algorithm", International Journal of Sciences & Technology Research, vol 2, Issue 8, August (2013)

[10] Ankur Bhardwaj " Comparative Study of Scheduling Algorithms in Operating System", International Journal of Computers and Distributed Systems, Vol. No.3, Issue I, April-May 2013

[11] V. Krishna Reddy, L.S.S.Reddy, "A Survey of Various task Scheduling Algorithms in Cloud Environment", Global Journal Engineering and Applied Sciences (GJEAS),Volume 2. No 1 , January 2012.