SCHEDULING IN CLOUD COMPUTING

Lipsa Tripathy, Rasmi Ranjan Patra

CSA,CPGS,OUAT,Bhubaneswar,Odisha

Abstract

Cloud computing is an emerging technology. It process huge amount of data so scheduling mechanism works as a vital role in the cloud computing. Thus my protocol is designed to minimize the switching time, improve the resource utilization and also improve the server performance and throughput. This method or protocol is based on scheduling the jobs in the cloud and to solve the drawbacks in the existing protocols. Here we assign the priority to the job which gives better performance to the computer and try my best to minimize the waiting time and switching time. Best effort has been made to manage the scheduling of jobs for solving drawbacks of existing protocols and also improvise the efficiency and throughput of the server.

Keywords

1. INTRODUCTION

Cloud Computing is an emerging technique. Recently it is found that researchers are interested in using cloud for performing scientific applications and even the big organizations are on the verse of switching over to hybrid cloud. Many complex applications require parallel processing to execute the jobs effectively. Due to the communication and synchronization among parallel processes there is a decrease in utilization of CPU resources. It is necessary for a data center to achieve the utilization of nodes while maintaining the level of responsiveness of parallel jobs.

The cloud computing is attracting an increased number of applications to run in the remote data centers. Many complex applications require parallel processing capabilities. Some of the parallel applications show a decrease in utilization of CPU resources whenever there is an increase in parallelism if the jobs are not schedule correctly then it reduces the computer performance.

Several algorithms & protocols are proposed regarding the scheduling mechanism of the cloud computing. But very few algorithms are proposed to detect the scheduling mechanism in cloud computing. Most of the authors consider a regular monitoring region in their protocol, which is not a real life scenario. Practically the monitoring region is always irregular as the clouds are randomly deployed. So we propose an algorithm to schedule the jobs in cloud computing.

Most of the authors consider the FCFS scheduling for processing the jobs. In this condition it decreases the resources utilization and utilization of server. So I take the consideration to improve the utilization of servers allocated to the jobs and to improve the resource utilization by using Backfilling and by assigning the shortest distance resources to the job to minimizing the

makespan. Some authors do not assign priority to the process. Processors process the jobs by assigning same priority in FCFS scheduling. So it decreases the performance of the computer. So I take the consideration of priority to schedule the job. Some authors does not consider the waiting time. For that reason the makespan of the job increases. For that reason performance of the computer decreases. Some authors give the idea to minimize the makespan by decreasing the waiting time but however it doesn't consider the switching time of the resources. So I think there is a better way to minimize the switching time which also minimize the makespan of the job.

So the main goal of my proposed protocol is to

- Improve the utilization of servers allocated to the jobs.
- To process the job having higher priority.
- Improve the resource utilization.
- Minimizes the completion time (makespan) of MapReduce jobs
- Minimizing the waiting time
- Minimizing the switching time

The rest of the paper is organized as follows: Section II discusses about the related work in this field. Section III describes proposed model. Section IV brings the conclusion and future scope of the paper.

2. RELATED WORK

As cloud computing holds different types and huge amount of data so it is called as heterogeneous system. Now days Cloud Computing is an emerging technology. So to improve the utilization of resource in cloud, minimizing the processing cost, increase the performance of the server, minimizing the processing time and completion time it is very necessary schedule the tasks in the cloud. So our main objective is to schedule mechanism of the tasks in cloud. This schedule mechanism was studied by several authors who have proposed various algorithms in order to solve the various problems. The paper [1],[2],[3],[4] discuss a complete survey of cloud computing. The authors of these papers basically discuss basic fundamentals & various applications of Cloud computing.

The author of the paper [5] discuss about the data processing in cloud computing. They proposed a programming model i.e. MapReduce. MapReduce is a programming model and an associated implementation for processing and generating large data sets. The MapReduce programming is widely used at Google for many different purposes. The author attributes this success to several reasons. First, the model is easy to use, even new users who do not have any experience in parallel and distributed systems, since it hides the details of parallelization, fault-tolerance, locality optimization and load balancing. Second, a large variety of problems can be easily expressed as MapReduce computations. Third, the authors [5] have developed an implementation of MapReduce that scales to large clusters of machines comprising thousands of machines. We found several things from [5]. Restricting the programming model which makes it easy to parallelize and distribute computations, make computations fault-tolerant, redundant execution used to reduce the impact of slow machines, and to handle machine failures and data loss.

So in paper [6] the author describes to optimizing the transferring and processing time is very crucial to an application program in the cloud though cloud computing holds lots of data and

process those data to give the services to the user. In order to minimize the cost of the processing the author[] proposed a model for task scheduling and proposed a particle swarm optimization (PSO) algorithm which is based on small position value rule. In order to improve the efficiency the optimizing task scheduling is necessary. In cloud computing resources distribute all over the world, and the data usually is bigger and the bandwidth often is narrower, these problems are more important. In this paper, the author presented the task scheduling optimizing method in cloud computing, and formulates a model for task scheduling to minimize the cost of the problem and solved it by a PSO algorithm. Experimental result manifests that the PSO algorithm both gains optimal solution and converges faster in large tasks than the other two. Moreover, running time is shorter than the other two too and it is obvious that PSO is more suitable to cloud computing.

As an increasing number of complex applications leverage the computing power of the cloud for parallel computing, it becomes important to efficiently manage computing resources for these applications. Many parallel applications show a pattern of decreasing resource utilization along with the increase of parallelism In this research paper [7] the author presented a Generalized Priority algorithm for efficient execution of task and comparison with FCFS and Round Robin Scheduling. The author proposed Workload consolidation method supported by virtualization technologies which is commonly used for improving utilization of resources in data centers. In this paper, the author gave a priority-based workload consolidation method to schedule parallel jobs in data centers to make use of underutilized node computing capacity to improve responsiveness. In the proposed method [7] there is a partition of node's computing capacity into the foreground VM (with high CPU priority) tier and the background VM (with low CPU priority) tier. The performance of the jobs running in the foreground VMs is closer to jobs running in dedicated nodes. The authors gave integrated backfilling and migration techniques to make effective use of the two types of VMs. The author's simulation showed that the consolidation based algorithm Aggressive Migration Supported BackFilling (AMCBF), even without knowing the job execution time, significantly outperforms the commonly used EASY algorithm. In addition, AMCBF is robust in the sense that it allows inaccurate CPU usage estimation of parallel processes.

The author of paper [9] discuss about the problems that can be arise during the schedule which minimizes the overall completion time of a given set of independent MapReduce jobs. The author of the paper [9] designed a novel framework called Balanced- Pools that efficiently utilizes characteristics and properties of MapReduce jobs in a given workload for constructing the optimized job schedule. The authors have evaluated this heuristic with a variety of different MapReduce workloads to measure achievable performance. Data analysis tasks often specified with higher-level SQL-type abstractions like Pig and Hive may result in MapReduce jobs with dependencies.

3. PROPOSED PROTOCOL

Efficiency of scheduling mechanism in cloud computing depends on how efficient it is in managing the processes and increase the performance of the server as well as resources. As we have discussed earlier there are various problems in previous scheduling mechanism, so it needs to be minimized in all possible ways, in order to increase the efficiency.

In this section, we propose a scheduling mechanism that schedules the jobs in an efficient manner to improve the resource utilization. The entire protocol or method has 4 steps.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Step-1

If there is no dependency among the jobs and resources then we take the switching time to consideration, as it is more flexible and more reliable. So the jobs may be processed in this manner. Else check the dependency and sort them in a queue then move to step-2.
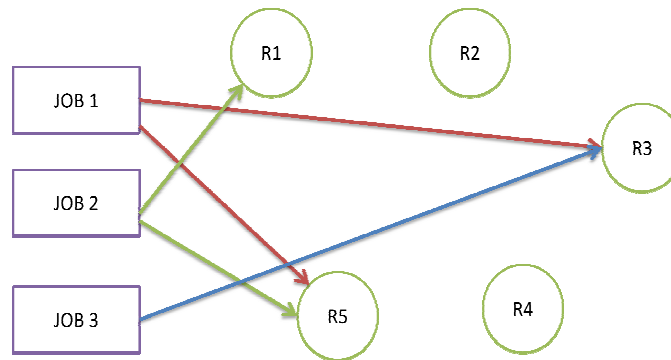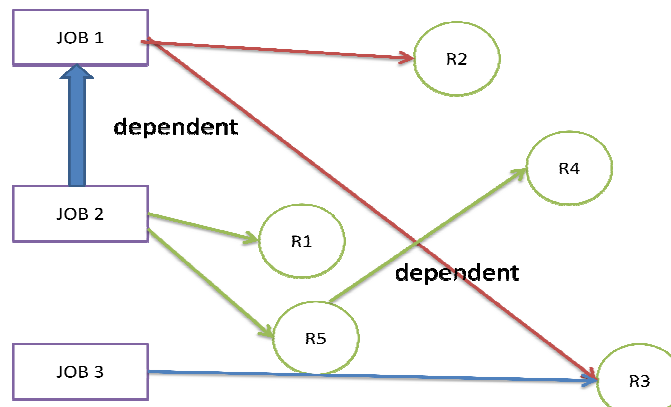

[Describing the jobs and resources having no dependencies]

In the above figure there are no dependencies among the jobs. So we take the consideration of switching time. So we store the jobs in the following manner to schedule the jobs.

JOB2 → JOB1→ JOB3

If there are any dependencies among the jobs or resources then directly jump to step 2. For example in the below figure there are dependency among the resources and jobs also. So there can be chance of deadlock and critical sections. So to avoid this situation we move to step-2.

[Describing the jobs and resources having dependencies]

B.Step-2

As we see in the above problem the server does not assign any priority to the jobs. So to avoid the problem we take priority as another parameter to check which job needs to be processed first. Let us take the Maximum priority 1 and the Minimum priority 5 then we assign the jobs in priority manner to improve the server performance along with resource utilization. As we take the Backfilling technique to improve the resource utilization.

Example:

| JOB1(2.5,**1**)		
	Job3(2,5,**4**)	JOB2(6, 10,**5**)
JOB4(2,5,**2**)		

[Examples to give the Priorities to the jobs]

JOB i(j,k,l)

Where

i= Job numberj= resource required
k= time duration required to complete the jo
l=Priority of the process

If there is any dependency among the resources or jobs then the higher priority always assign to the independent job then to dependent job. So the deadlock situation will be minimized.

C. Step-3

After priority assign to the jobs one question is arise that which job assign to which processor. So to know which job assign to which processor, I take an average time of execution time of all the jobs. And then check the processing speed of the processor. Then assign the jobs to the processor or CPU or server. Let us take an example:
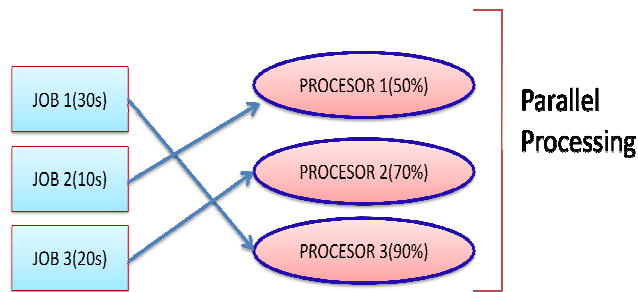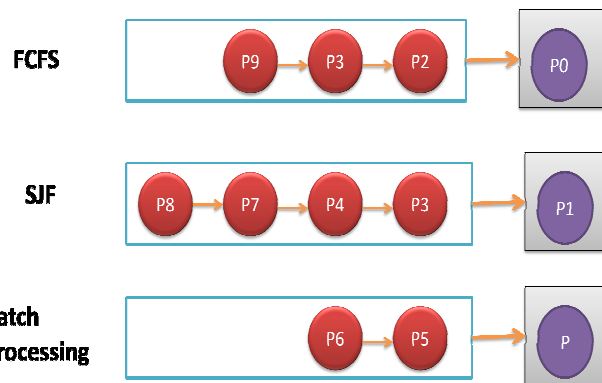
[Describing the job allocation to the processor]

In the above figure there are 3 jobs i.e. job 1, job 2, job 3. So the average time period is 20 sec. So all the jobs must be complete their task in 20sec. So the higher completion time job assign to higher processor. Then the next job having less completion time will assign to the next processor. Like this fashion all the jobs are assign to the processor. So then all the processor starts the execution parallelly.

D. Step-4

Then each processor creates a multi level queue to store the jobs and process the jobs parallelly.

[Describing the parallel Processing]

In the above figure one processor creates 3 queue which having own scheduling mechanism algorithm to process the jobs. Then as the job arrives to the processor, the processor stores the jobs in the appropriate queue then process the jobs parallelly.

4. CONCLUSION AND FUTURE SCOPE

Scheduling mechanism is an important issue in case of cloud computing. Scheduling mechanism is very much necessary to improve the server and resource utilization also to increases the performance of the computer. So in this work I proposed a scheduling mechanism or method to schedule the jobs in the cloud. This method is a very simple & novel method but very efficient method to schedule the jobs in cloud computing. This scheduling mechanism or method is more beneficial as compared to other proposed protocol or algorithm or method because it can be able

to schedule the jobs in very good manner as priority assign to the jobs with minimizing the makespan. And also it increases resource utilization and server utilization.

The proposed model is very simple and easy to understand. This paper is proposed after the relentless effort made by me by taking into consideration various aspects of scheduling mechanism and also detection. However, time complexity and switching and completion time of the job are yet to be done in order to prove the efficiency of the protocol along with the simulation in order to prove the efficiency of the protocol.

References

[1] B. Ramamurthy & K. Madurai : Cloud Computing: Concepts, Technologies and Business Implications
[2] Dan C. Marinescu : Cloud Computing Theory and Practice. Pp. 1-50 .
[3] School of Software, Sun Yat-sen University : Introduction to Cloud Computing .
[4] Introduction to Cloud Computing .Fact Sheet , Fiche d'information. Page 1- 6..
[5] Jeffrey Dean and Sanjay Ghemawat : MapReduce: Simplified Data Processing   on Large Clusters
[6] Lizheng Guo,Shuguang Zhao, Shigen Shen, Changyuan Jiang : Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm. In IEEE Journal of Networks, Vol. 7, No. 3, March 2012
[7] Xiaocheng Liu, Albert Y. Zomaya, Fellow IEEE, Chen Wang, Bing Bing Zhou, Junliang Chen, Ting Yang, : Priority-Based Consolidation of Parallel Workloads in the Cloud. IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 9, September 2013
[8] Sharad Mehrotra ,Sai Wu, Feng Li , Beng Chin Ooi : Query Optimization for Massively Parallel Data Processing.
[9] Ludmila Cherkasova , Roy H. Campbell ,Abhishek Verma : Two Sides of a Coin: Optimizing the Schedule of MapReduce Jobs to Minimize Their Makespan and Improve Cluster Performance.
[10] Anish Das Sarma, Christopher Olston,Xiaodan Wang, Randal Burns : CoScan: Cooperative Scan Sharing in the Cloud.

AUTHORS PROFILE

Rasmi Ranjan Patra received Master in Computer Application with 1st Class with distinction from O.U.A.T, Odisha, India in 2001, M.Tech in Computer Science and Technology from C.E.T , Bhubaneswar ,India in 2010 and completed Ph.D Degree in Utkal University, India in 2012 .He is working as Assistant Professor in Department of Computer Science and Application in O.U.A.T.He has published many papers at national /international Journals and Conferences in the areas of Sensor Network. Mr.Patra is the author of one book in Computer Science area.

Lipsa Tripathy received Master In Computer Application With 1st Class With distinction from Orissa University of Agriculture and Technology(O.U.A.T)., Odisha, India in 2014 and currently pursuing M Tech Degree in Trident Academy of Creative Technology, Bhubaneswar,Odisha,India.