Performance Optimization in Gang Scheduling In Cloud Computing

Sudhir Singh

Research Scholar, Department of Computer Science Engineering Lovely Professional University Phagwara, India

Abstract: *Cloud computing is a latest new computing paradigm where applications, data and IT services are provided over the Internet. The Job Scheduling is the key role in cloud computing systems. One technique is to use gang scheduling where a set of tasks is scheduled to execute simultaneously on a set of processors. Usually tasks are scheduled by user requirements. So, taking existing model new scheduling strategy proposed to overcome the problems of the performance unpredictability with the help of scheduling technique between user and resources.*

Keywords: Cloud Computing, Gang Scheduling, Virtual Machine, Cloud models

I. Introduction

Cloud refers to both the applications delivered as services over the Internet and the hardware and systems software in the Datacentres that provide those services. Cloud Computing is offering the services the applications delivered as services with the help of Internet, hardware and systems software in the DC that provide those services. The services are referred to as SaaS, so we use that term. The datacentre is called a Cloud. [9]

- When a Cloud is made available in a pay-as-you-go manner to the public environment is termed as Public Cloud.
- When taking about Private Cloud it means that to the internal data centres of a business or other organization that are not provided to the public [5].

II. Cloud Models

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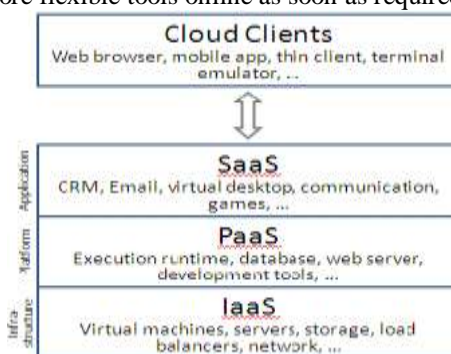
Platform-as-a-Service (PaaS) - This type of cloud computing offers a product development tool or environment that users can access and utilize online, even in collaboration with others and hosted on the provider's infrastructure. In PaaS developers create applications on the service provider's platform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

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. It's a best example of platform as a service providing by the Hotmail as a free of cost **[15]**.

III. Deployment models

Public cloud - Public cloud dynamically provisioned on a fine-grained, self-service basis over the Internet, via web services, from an off-site third-party provider who shares resources and bills on a fine-grained utility computing basis. The cloud infrastructure is owned by a cloud vendor, and is accessible to the general public or a large industry group.
Examples of Public Cloud:
- Google App Engine
- Microsoft Windows Azure
- IBM Smart Cloud

Private Cloud - Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally **[16]**.
Examples of Private Cloud:
- Eucalyptus
- Ubuntu Enterprise Cloud - UEC (powered by Eucalyptus)
- Amazon VPC (Virtual Private Cloud)
- VMware Cloud Infrastructure Suite
- Microsoft ECI data center.

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 **[16]**.
Examples of Hybrid Cloud:
- Windows Azure (capable of Hybrid Cloud)
- VMware vCloud (Hybrid Cloud Services)

IV. Literature Survey

Shinpei Kato in her paper describes about the pre-emptive real-time scheduling of sporadic parallel task systems is studied. Also present an algorithm, called Gang EDF, which applies the Earliest Deadline First (EDF) policy to the traditional Gang scheduling scheme it's also providing schedulability analysis of Gang EDF **[1]**.

Helen d. karatza also gives idea on distributed systems are considered in this paper. Two different cases of job parallelism (LGFS, AFCFS) are examined. The LGFS method provides the best overall performance. Furthermore, LGFS provides fairer service to individual job classes. This is achieved by preventing small gangs to arbitrarily overtake large gangs **[2]**.

Karatza also studied the performance of Adaptive First Come First Serve (AFCFS) and Largest Job First Served (LJFS) gang scheduling policies in cloud computing with providing the scheduling structure. Both algorithms provide similar performance for medium workloads LJFS outperforms AFCFS when workloads get heavier. That is even more apparent in weighted metrics which take into account job sizes. Cost-wise LJFS provides superior cost-performance efficiency than AFCFS and is far-superior in situations of heavy workload **[3]**

Zafeirios C. Papazachos and Helen D. Karatza in their paper studies the performance of scheduling strategies in the case of gangs which are dynamically generated based on precedence constraints with formerly executed gangs. The formation of a secondary gang upon the execution of a preceding gang can be necessary for the further processing of affinity information which resides on the caches of the previously seized processors **[4]**.

V. Problem Formulation

Existing Gang Scheduling Structure- The simulation model consists of a single cluster of Virtual Machines connected with a Dispatcher Virtual Machine (DVM). In this scheduling model, they integrate both a migration mechanism and a starvation handling system into the model [4].

The problems occur in this scheduling structure:

* The starvation problem occurs. That if a large gang job is migrated then the small jobs with earliest deadline have to wait long.
* Jobs requiring a smaller number of processors and thus may increase fragmentation of the system.

VI. Objective

By taking existing model new scheduling strategies is proposed to overcome the problems of the performance unpredictability with the help of scheduling technique between user and resources. New scheduling strategies use some of the conventional scheduling concepts to merge them together with some performance efficient aware strategies to provide solutions for better and more efficient job scheduling. The two job scheduling algorithms AFCFS and LJFS are used in it. AFCFS schedules smaller jobs whose tasks are behind the tasks of the large job and LGFS tasks are placed in increasing job size order in processor queues (tasks that belong to larger gangs are placed at the head of queues). All tasks in queues are searched in order, and the first jobs whose assigned processors are available begin execution. The entry point of the system is the Job Queue. Jobs with degrees of parallelism less than or equal to the available VMs are dispatched immediately. Job Routing will give the job ID to all the jobs arrive in the Queue. The job ID contains the detail of job size, job type, job arrival, job deadline. Then According to size the job ID is forwarded to the LGFS and AFCFS.

The propose scheduling structure consist of 4 modules:

Job Queue - Job Routing will give the job ID to all the jobs arrive in the Queue. The job ID contains the detail of job size, job type, job arrival, job deadline. Then According to size the job ID is forwarded to the LGFS and AFCFS.
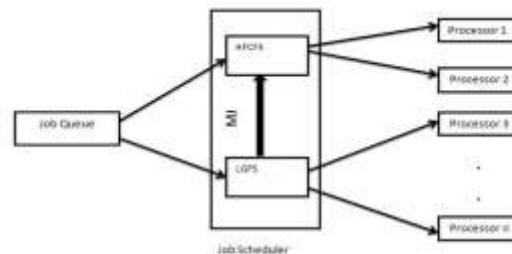
MI- Gang Migration with EDF and precedence constraint
JOB QUEUE- Job routing to the scheduler

Largest Gang First Serve (LGFS) – It will handle all the jobs with Job ID contains detail of the larger job. Then the job need to execute is called from the Job Queue.

Adaptive First Come First Serve (AFCFS) – It will handle the jobs with Job ID contains detail of smaller job. Then the need to execute is called from the Job Queue.

Job Migration with EDF- If the AFCFS is having less job ID's. Then Job ID from LGFS to AFCFS is migrated with considering the EDF scheduling and precedence constraint of the Job ID need to migrated. Then the job is called from the Job Queue which needs to execute.

VII. Conclusion

This study, has integrated two important features, job scheduling with Job ID and Jobs Migration into the previous working model. In case of the migration the generation of a secondary gang was depended on the previous Jobs information. This technique easily improves with providing the Job ID's in the entry point of the Job which will reduce the overburden in the Job Scheduler.

VIII. Future Scope

In the future looking forward to examine new workload models better suited for Cloud computing. Also the application of it in the systems with heterogeneous performance should be studied in depth. Furthermore, it may be possible to examine the case when there is more than one secondary gang with Job ID's are involved in the job migration process.

References

[1]   Yutaka Ishikawa and Shinpei Kato (2009), "EDF Scheduling with gang approach" 30th IEEE Real-Time Systems Symposium
[2]   Helen d. karatza, (2006) "Scheduling gangs in a distributed system" I.J. of SIMULATION, Vol. 7, No. 1
[3]   Helen D. Karatza, Ioannis A. Moschakis, (2011) "Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System with Job Migrations and Starvation Handling" 978-1-4577-0681-3/11/2011 IEEE
[4]   Zafeirios C. Papazachos and Helen D. Karatza, (2007) "Gang Scheduling with Precedence Constraints"
[5]   Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia (2009) "Above the Clouds: A Berkeley View of Cloud Computing" February 10, 2009
[7]   Y. Wiseman and D. G. Feitelson, "Paired gang scheduling," IEEE Trans. Parallel Distrib. Syst., vol. 14, no. 6, pp. 581–592, 2003
[8]   L. Nie and Z. Xu, "An adaptive scheduling mechanism for elastic grid computing," in Proc. of the 2009 Fifth Int. Conf. on Semant, Knowl. and Grid, SKG '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 184–191.
[9]   http://www.technopulse.com/2011/06/cloud-service-models-saas-paas-iaas.html
[10]   Amazon elastic compute cloud (amazon ec2). [Online]. Available: http://aws.amazon.com/ec2/
[11]   Hori A, Tezuka H, Ishikawa Y (1998) Overhead analysis of preemptive gang scheduling. In: Proc of the workshop on job sched strateg for parallel process. Lect notes in comp sci, vol 1459. Springer, Berlin, pp 217–230
[12]   Karatza H (2009) Performance of gang scheduling strategies in a parallel system. Simul Model Pract Theory17:430–441. doi:10.1016/j.simpat.2008.10.001
[13]   Papazachos Z, Karatza H (2009) Performance evaluation of gang scheduling in a two-cluster system with migrations. In: 8th int workshop perform model, evaluation optim of ubiquitous comp and netwsyst, Rome, Italy 2009. doi:10.1109/IPDPS.2009.5161172
[14]   Z. C. Papazachos and H. D. Karatza, "The impact of task service time variability on gang scheduling performance in a two-cluster system," Simul. Modell. Pract. Theory, vol. 17, no. 7, pp. 1276–1289, 2009.
[15]   http://www.cloudtweaks.com/2012/04/cloud-computing-cloud-service-models-part-3/
[16]   http://thecloudtutorial.com/cloudtypes.html

AUTHORS

Sudhir Singh passed his M.Tech degree in Computer Science from Lovely Professional University, India His current research interest includes Performance optimization in cloud computing and Green IT.