International Journal of Academic Research and Development
ISSN: 2455-4197
Impact Factor: RJIF 5.22
www.academicsjournal.com
Volume 3; Issue 1; January 2018; Page No. 884-886

A review on various approaches of job scheduling in cloud computing

Deep Kumar
Software Engineer, Igniva Solutions Private Limited, Mohali, Punjab, India

Abstract
Cloud computing is the emerging technology that has been used in recent applications. Data has been stored on the cloud servers so that this can be managed from remote locations. In the process of cloud computing various users transmit their requests for access to cloud data that has been processed at different servers. In this process a number of requests have been issued on cloud server. Various job scheduling algorithms have been proposed so that request can be respond in minimum time span. In this paper various approaches of cloud computing has been discussed so that approaches can be used for data management and response by the cloud server.

Keywords: Iaas, Paas, Saas, Ga and Pso

1. Introduction

1.1 Cloud Computing

Cloud computing is the long dreamed vision of computing as a utility, where data owners can remotely store their data in the cloud to enjoy on-demand high-quality applications and services from a shared pool of configurable computing resources. Cloud is a new business model wrapped around new technologies such as server virtualization that take advantage of economies of scale and multi-tenancy to reduce the cost of using information technology resources. It also brings new and challenging security threats to the outsourced data. Since cloud service providers (CSP) are separate administrative entities, data outsourcing actually relinquishes the owner's ultimate control over the fate of their data.
Frameworks provide mechanisms for:

- Self-healing
- Self monitoring
- Resource registration and discovery
- Service level agreement definitions

1.2 Various Cloud Models

1.2.1 Private Cloud

The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

1.2.2 Community Cloud

The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

1.2.3 Public Cloud

The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

1.2.4 Hybrid Cloud

The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

1.3 Job Scheduling

Job Scheduling is used to allocate certain jobs to particular resources in particular time. In cloud computing, job scheduling problem is a biggest and challenging issue. Hence the job scheduler should be dynamic. Job scheduling in cloud computing is mainly focuses to improve the efficient utilization of resource that is bandwidth, memory and reduction in completion time. An efficient job scheduling strategy must aim to yield less response time so that the execution of submitted jobs takes place within a possible minimum time and there will be an occurrence of in-time where resources are reallocated. Because of this, less rejection of jobs takes place and more number of jobs can be submitted to the cloud by the clients which ultimately show increasing results in accelerating the business performance of the cloud. There are different types of scheduling based on different criteria, such as static vs. Dynamic, centralized vs. Distributed, offline vs. Online etc. are defined below:

1. Static Scheduling: Pre-Schedule jobs, all information are known about available resources and tasks and a task is assigned once to a resource, so it's easier to adapt based on scheduler's perspective.

2. Dynamic Scheduling: Jobs are dynamically available for scheduling over time by the scheduler. It is more flexible than static scheduling, to be able of determining run time in advance. It is more critical to include load balance as a main factor to obtain stable, accurate and efficient scheduler algorithm.
3. Centralized Scheduling: As mentioned in dynamic scheduling, it's a responsibility of centralized / distributed scheduler to make global decision. The main benefits of centralized scheduling are ease of implementation; efficiency and more control and monitoring on resources. On the other hand; such scheduler lacks scalability, fault tolerance and efficient performance. Because of this disadvantage it's not recommended for large-scale grids.
4. Distributed / Decentralized Scheduling: This type of scheduling is more realistic for real cloud despite of its weak efficiency compared to centralize scheduling. There is no central control entity, so local schedulers' requests to manage and maintain state of jobs' queue.
5. Pre-Emptive Scheduling: This type of scheduling allows each job to be interrupted during execution and a job can be migrated to another resource leaving its originally allocated resource, available for other jobs. If constraints such as priority are considered, this type of scheduling is more helpful.
6. Non Pre-Emptive Scheduling: It is a scheduling process, in which resources are not being allowed to be re-allocated until the running and scheduled job finished its execution.
7. Co-operative scheduling: In this type of scheduling, system have already many schedulers, each one is responsible for performing certain activity in scheduling process towards common system wide range based on the cooperation of procedures, given rules and current system users.
8. Immediate / Online Mode: In this type of scheduling, scheduler schedules any recently arriving job as soon as it arrives with no waiting for next time interval on available resources at that moment.
9. Batch / Offline Mode: The scheduler stores arriving jobs as group of problems to be solved over successive time intervals, so that it is better to map a job for suitable resources depending on its characteristics.

2. Review of Literature

Kunal Kishor *et al* [1] "an efficient service broker policy for cloud computing environment" As almost all the applications are being provided over the internet, the need for computing resources is shifting from the user's location to the service provider. The concept of services has gained popularity with the widespread use of the term "cloud computing", which is a new paradigm that has been defined to address user requests on a pay-per-use basis. With the greatest benefit being elasticity in terms of increase or decrease of computing resources like computation power, storage and bandwidth, cloud is providing better computing solutions to the users of its services. But the success of these solutions lies in the use of efficient policies and algorithms that govern the underlying concept of cloud computing. These policies involve service brokerage, load balancing, virtual machine management and service level agreements. The broker is an intermediary between the client and the cloud service provider and hence the role of a broker is quite significant.

Amir Nahir *et al.* [2] Author proposes a novel scheme that incurs no communication overhead between the users and the servers upon job arrivals, thus removing any scheduling overhead from the job execution's critical path. Furthermore, our scheme is oblivious, that is, it does not use any state information. Our approach is based on creating, in addition to the regular job requests that are assigned to randomly chosen servers, also replicas that are sent to different servers; these replicas are served in low priority, such that they do not add any real burden on the servers. Through analysis and simulations we show that the expected system performance improves up to a factor of 2 (even under high load conditions), if job lengths are exponentially distributed, and over a factor of 5, when job lengths adhere to heavy-tailed distributions. We implemented a load balancing system based on our approach and deployed it on the Amazon Elastic Compute Cloud (EC2). Realistic load tests on that system indicate that the actual performance is as predicted.

Hong Tao *et al.* [3] Author proposed that with the growing demand of data and the increase of the user scale, data allocation technology has become a key technology for improving scalability and flexibility in current mass storage system such as cloud storage system. This paper proposed an efficient dynamic data allocation strategy with data partitioning and load balancing. Based on the basic idea of consistent hashing algorithm, the strategy introduced the concept of virtualization technology and improved the load-balance with employing virtual node. Moreover, the strategy adopted a novel available-storage-capacity-aware and storage-capacity-utilization-aware method to enhance the performance of the cloud storage system. The simulation results demonstrate that the proposed data allocation strategy improves system performance in both homogeneous and heterogeneous distributed storage architectures.

Yuqi Zhang *et al.* [4] Author described that in clouds, many applications need to distribute large data sets from the cloud's storage facility to all compute nodes as fast as possible, especially data-intensive parallel applications. Many multicast algorithms have been used for clusters and grid environments. In order to maximize available bandwidth and avoid bottleneck links, a common approach is to construct one or more spanning trees based on the network monitoring data and network topology. However, in clouds the available bandwidth changes dynamically, so delivering optimal performance becomes difficult. In this paper, we focus on Eucalyptus (an open-source cloud-computing platform) and propose a high performance multicast algorithms `steal-and-p2p' based on `non-steal' and `steal' algorithm mentioned. We evaluate our algorithm on Eucalyptus, and show that the algorithm can achieve high throughput and perform much better having each node downloading all data directly from storage facility.

Er. Amandeep Kaur1 *et al.* [5] Author proposed that as the cloud computing is a new style of computing over internet. It has many advantages along with some crucial issues to be resolved in order to improve reliability of cloud environment. These issues are related with the load management, fault tolerance and different security issues in cloud environment.

In this paper the main concern is load balancing and security issues in cloud computing. The load can be CPU load, memory capacity, completion time of each job and security issues to prevent the data from unauthorized user. From decades well known algorithms like FCFS, Priority has been seen into action to reduce the server load. But with the increase in the complexity of the server needs, they have failed to cope up with the current scenario. In our approach, we are developing a technique named Cross Breed Job scheduling technique which would be a combination of FCFS, Priority and would be monitored by RBAC (Role based access control). RBAC is a system which checks that whether the user of the system has the access to particular content or not. If the user doesn't have the access to the content, he will be denied and the server's load would be minimized.

3. Approaches Used

Vgreedy-Based Algorithm

For a set of jobs and the virtual machines, Greedy-Based Algorithm depends on the local optimal method toallocate resources. That is the reason why we called it Greedy-Based Algorithm based on the Greedy algorithm.

Max-Min Algorithm

Max-Min is almost same as the min-min algorithm except the following: in this after finding out the completion time, theminimum execution times are found out for each and every task. Then among these minimum times the maximum value is selected which is the maximum time among all the tasks on any resources. Then that task is scheduled on the resource on which it takes the minimum time and the available time of that resource is updated for all the other tasks. The updating is done in the same manner as for the Min-Min. All the tasks are assigned resources by this procedure.

Particle Swarm Optimization (PSO) Algorithm

Particle Swarm Optimization (PSO) as a meta-heuristics method is a self-adaptive global search based optimization technique introduced by Kennedy and Eberhart. The PSO algorithm is alike to other population-based algorithms like Genetic algorithms (GA) but, there is no direct recombination of individuals of the population. The PSO algorithm focuses on minimizing the total cost of computation of an application workflow. As a measure of performance, Authors used cost forcomplete execution of application as a metric.

Genetic Algorithm

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.

4. Conclusion

Cloud computing has been used in various applications of business enterprises and data storage services. In the process of cloud computing users does not aware about the hardware that has been conducted with applications. In this process various approaches have been used for cloud computing process so that services that have been provided by the cloud server must be done in minimum response time. In the process of cloud computing various approaches has been reviewed in this paper that can be used for processing of the job scheduling process. Greedy based, Min-Max based, round robin and earliest Job first are the basis approaches that has been used for job scheduling process. Various AI approaches have been discussed that can be used for optimization process of cloud computing network.

5. References

1. Kunal Kishor "an efficient service broker policy for cloud computing environment", InternationalJournal of Computer Science Trends and Technology (IJCST), 2014, pp 56-62.
2. Amir Nahir, "Distributed Oblivious Load Balancing Using Prioritized Job Replication", ISSN 978-3-901882-48-7, IEEE, 2012.
3. Hong Tao "A dynamic data allocation method with improved load-balancing for cloud storage system", ISSN 978-1-84919-707-6, PP 220 – 225, IEEE, 2013
4. Yuqi Zhang, "Dynamic load-balanced multicast based on the Eucalyptus open-source cloud-computing system", ISSN 978-1-61284-158-8, pp. 456 – 460, IEEE, 2011.
5. Magade, Krishnanjali A, "Techniques for load balancing in Wireless LAN's", ISSN978-1-4799-3357-0, PP 1831 – 1836, IEEE, 2014.
6. Yean-Fu Wen "Load balancing job assignment for cluster-based cloud computing", ISSN 14517061, PP 199 – 204, IEEE, 2014.
7. De Mello, MOMC "Load balancing routing for path length and overhead controlling in Wireless Mesh Networks", ISSN 14630778, PP 1-6, IEEE, 2014.
8. Angel Preethima R, Margret Johnson, "Survey on Optimization Techniques for Task Scheduling in Cloud Environment", IJARCSSE,Volume 3, Issue 12, December 2013.
9. Suraj Pandey, Linlin Wu, Siddeswara Mayura Guru, Rajkumar Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments" in 24th IEEE International Conference on Advanced Information Networking and Applications, DOI 10.1109/AINA.2010.31.
10. Uma Somani, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing," 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010).
11. RuWei Huang, Si Yu, Wei Zhuang and XiaoLinGui, "Design of Privacy-Preserving Cloud Storage Framework" 2010 Ninth International Conference on Grid and Cloud Computing.
12. Jianfeng Yang, Zhibin Chen "Cloud Computing Research and Security Issues" Vol. 978-1-4244-5392-4/10/$26.00 ©2010 IEEE.