A Review of Different Resource Scheduling Algorithms in Cloud computing

Er. Amritpal Singh

Abstract— **As our reliance on cloud computing increases day by day so does our expectation of getting efficient service from cloud service providers a cloud can provide optimal service to users only if it manages its resources in an efficient and sagacious manner for this purpose there are different resource scheduling algorithms which aid the cloud service provider in scheduling of the resources across different nodes in a cloud environment each algorithm gives optimal performance in a particular scenario so choosing a right algorithm is important. In this paper we will analyze different resource scheduling algorithms their features, benefits and limitations depending on the requirement, a cloud service provider chooses a suitable algorithm an unsuitable algorithm can hamper the process of service and resource delivery so the key to a streamlined service and optimal functioning of a cloud depends on a right algorithm there are different algorithms like energy efficient algorithm, cost based algorithm, algorithm based on trust degree etc.**

Index Terms— **resource scheduling, algorithms, virtual resource algorithm, optimized resource scheduling**

I. INTRODUCTION

The main purpose of a cloud is to provide services, infrastructure and applications to users in an efficient manner there are numerous nodes in a cloud environment each vying for resources and services so there is a dire need of efficient resource scheduling algorithms as the success of a cloud service provider depends on streamlined resource scheduling and allocation, depending on demand of users or objectives of cloud service provider a suitable algorithm is chosen. Resource scheduling can be defined as request of cloud users and the resources available at the Data Centre some resource allocation methods take into account different parameters like execution time, policy, SLA's, utility function etc.

II. RESOURCE SCHEDULING ALGORITHMS

.

a) Optimized Resource Scheduling: In this method the optimization is done on the basis of SLA as resource scheduling algorithms are NP hard problems Grobner bases theory is applied MGBA is extended to solve stochastic integer programming with two stage recourse Geometric Buchberger Algorithm is applied addressing SLA aware resource composition problem this optimal solution is based on reasonable short time [1].

b) Cost Based Resource Scheduling paradigm: In this technique the resources are allotted as a leveraging market theory to make maximum use of available resources the set of computing resources with lowest price are assigned to the users according to current supplier, resource availability and price this algorithm is depicted in pure JAVA based private cloud platform JavaCloudware this algorithm and protocol is designed for IaaS [2].

c) Double Auction-based Scheduling of scientific applications in Distributed Grid and Cloud environments: In this method a negotiation protocol is introduced between scheduler and resource manager using market based Continuous Double Auction model (CDA) different scheduling strategies are analyzed that can be applied and identify general strategic patterns that can lead to a fast and cheap work flow execution The self-limitation based scheduling brings small improvements. Improvement in execution is achieved by aggressive strategy the simulation of this method is achieved by GridSim [3].

d) **Heavy Traffic Optimal Algorithm**: The join-the-shortest-queue routing and power-of-two-choices routing with MaxWeight scheduling is optimal in throughput and they are queue length optimal in high traffic loads. Calculating the exact queue length is quite difficult so the system in heavy traffic regime (exogenous arrival rate is almost same as boundary of capacity region) was studied. Use of state space collapse (multi-dimensional state reduces to single dimension) was there. The algorithm is applied on multiple models supported by multiple servers. Above models assume system is work conserving. Then the result converges to regulated Brownian notation and simple path optimality in scaled time.

The method in the heavy traffic optimality is simpler and also in unscaled time consists of three steps lower bound (weighted sum of expected queue length by comparing with a single server queue), state-space collapse (state of system collapsing to single dimension, queue length in particular direction increases and in perpendicular direction it is bounded) and upper bound (obtained by natural Lyapunov function). Heavy traffic is obtained when upper and lower bound coincides. The solution contains one routing and one scheduling algorithm. This is the stochastic model for load balancing and scheduling in clusters. The JSQ and MaxWeight is throughput optimal and traffic optimal when all servers identical. And also the power-of-two-choices is

ISSN: 2278 – 1323

also heavy traffic optimal [4][5]

e) **Hybrid multidimensional algorithm for network aware scheduling**: This algorithm handles multiple resource requests for jobs/tasks that arrive on the computing environment of the e- infrastructure it finds a trade- off between execution time and economic cost for processing data intensive applications by taking into account performance parameters at system and network levels as well as the economic cost of computational resources It has an advantage of knowledge of grid infrastructure [6].

f) **Genetic algorithm with Multiple Fitness**: This strategy is a pre migration strategy it takes into consideration three parameters disk I/O, network throughput and CPU utilization for optimal solution a hybrid of genetic algorithm and knapsack problem is considered. This algorithm raises resource utilization and lower energy consumption cost by runtime resource scheduling under cloud environment the algorithm achieves the goal of improving resource utilization and saving energy cost by runtime resource scheduling [7].

g) **Pricing algorithm**: Cloud bank as resource agency provides analysis and guidance for participants and a price update iterative algorithm analysis the historical utilization ratio and iteration current prices. It also gets the availability of resources and final price to consumers. This algorithm is designed to safeguard the interests of the participants in cloud. With this resources achieve macro control. It is not adaptive as it cannot adapt rapid changes of demand and supply. It reduces the cost of providers, maximizes revenue and is more conducive to keep the providers interests [5] [8].

h) Load adaptive model based on ant colony algorithm: It is a load adaptive cloud resource scheduling model based on ant colony algorithm it monitors the virtual machine of performance parameters in real time and if overload is detected it schedules fast cloud resources using ant colony algorithm to bear some load on load free node, this algorithm the nearest idle node and allows it to bear some load meeting the performance and resource requirements of load thus achieving the goal of load balancing [9]

i**) Smart Dynamic resource scheduling algorithm**: It is a two-step dynamic resource scheduling strategy called smart-DRS it fits cloud data centers and strikes a balance between efficiency, cost and instantaneity In this management prototype resource scheduling is just a module which has high forecast accuracy and can deal with load balancing and load consolidation this paper presents a dynamic scheduling strategy which employs Single Exponential Smoothing (SES) algorithm for prediction of resource utilization and Vector Projection (VP) for second step it doesn't consider spending of migration experiment results show that Smart-DRS has a high forecast accuracy and also can deal well with load balancing and load consolidation [10].

j) Dynamic priority scheduling algorithm (Service request scheduling): This algorithm is applied on three tier containing service providers, resource providers and consumers. This algorithm gives more optimal then First Come First Serve (FCFS) and Static Priority Scheduling Algorithm (SPSA). The consumer response time for services has been tried to reduce in this algorithm as running instance is charged as it runs per unit time. The delays in provider side happens but are not counted under the cost charged to the customer so they need to be reduced. In three tiers there needs to be two scheduling: service request scheduling and resource scheduling.

The FCFS concentrates on fairness to task units but it may result in low priority task units perform before than high priority tasks and SPSA makes task units prioritized before the process of scheduling. The DPSA evaluates task unit scheduled and recalculates and set task unit's priority thus optimizing the scheduling process. Though tasks has their initial priorities but the new priorities being set include SLA between user and cloud, task's features, task's source and operations in cloud. This algorithm considers three queues having highest priority, middle priority and lowest priority. Every queue has a threshold i.e. time a task unit will wait in particular queue. When the some task unit crossed that threshold value then the task unit automatically is moved to higher queue. When task reaches the highest queue it is send to the required component. Finally by comparing the average values and variance of priorities by processing time the DPSA comes out to be more efficient than FCFS and SPSA[11][5].

III Conclusion

Cloud computing has redefined the way resources and services are provided the hindrance posed by distance has become obsolete millions of people all over the world avail the services of cloud service providers so providing efficient service is a challenge in this paper reviews different resource scheduling algorithms and their mode of functioning each algorithm is suited for a special situation and each algorithm has its own limitations so choosing the right algorithm is one of the vital step for efficient resource scheduling. There is need of more energy efficient algorithms in the future as cloud computing applicability is increasing more energy will be needed for the further increasing load in use and more of the services will provided in future.

REFERENCES

[1] Qiang Li, Yike Guo "Optimization of Resource Scheduling in Cloud Computing" 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing

[2] Zhi Yang, Changqin Yin, Yan Liu "A Cost-based Resource Scheduling Paradigm in Cloud Computing" 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies

ISSN: 2278 – 1323

[3] Radu Prodan, MarekWieczorek, Hamid Mohammad Fard, 'Double Auction-based Scheduling of Scientific Applications in Distributed Grid and Cloud Environments', J Grid Computing (2011) 9, PP:531–548

[4] Siva Theja Maguluri and R. Srikant, Lei Ying, 'Heavy Traffic Optimal Resource Allocation Algorithms for Cloud Computing Clusters', ARO MURI W911NF-08-1-0233 and NSF grant CNS-0963807

[5] Nimisha Singla, Seema Bawa "Review of Efficient Resource Scheduling Algorithms in Cloud Computing" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013

[6] D. Adami, C. Callegari, S. Giordano, M. Pagano, 'A Hybrid Multidimensional Algorithm for Network-aware resource Scheduling in Clouds and Grids', IEEE ICC 2012-Communication QoS, Reliability and Modeling Symposium, IEEE 978-1-4577-2053-6/12 (2012), PP: 1297-1301

[7] Shi Chen, Jie Wu, Zhihui Lu, 'A Cloud Computing Resource Scheduling Policy Based on Genetic Algorithm with Multiple Fitness', IEEE 12th International Conference on Computer and Information Technology, IEEE 978-0-7695-4858-6/12 (2012), PP: 177-184.

[8] Hao Li, Jianhui Liu,Guo Tang, 'A Pricing Algorithm for Cloud Computing Resources', International Conference on Network Computing and Information Security, IEEE 978-0-7695-4355-0/11 (2011), PP: 69-73.

[9] Xin Lu, Zilong Gu, 'A LOAD-ADAPATIVE CLOUD RESOURCE SCHEDULING MODEL BASED ON ANT COLONY ALGORITHM', Proceedings of IEEE CCIS2011, IEEE 978-1-61284-204-2/11 (2011), PP: 296-300.

[10] Lei Xu, Wenzhi Chen, Zonghui Wang, Shuangquan Yang, 'Smart-DRS: A Strategy of Dynamic Resource Scheduling in Cloud Data Center', IEEE International Conference on Cluster Computing Workshops, IEEE 978-0-7695-4844-9/12 (2012), PP: 120-127.

[11] Zhongyuan Lee, Ying Wang, Wen Zhou, 'A dynamic priority scheduling algorithm on service request scheduling in cloud computing', 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, IEEE 978-1-61284-088-8, Vol. 11 (2011), PP: 4665-4669

ISSN: 2278 – 1323