A Review of Prioritize Task Scheduling in Cloud Computing

Pankaj Kumar Chauhan [1*], Poonam Dabas[2]

[1]University Institute of Engineering & Technology, Kurukshetra, Haryana, INDIA

[2]University Institute of Engineering & Technology, Kurukshetra, Haryana, INDIA

Abstract: Task scheduling plays the key role systems in cloud computing. Scheduling of tasks cannot be done on the basis of single criteria but under a lot of the regulations and rules that we can term as an agreement between users and providers of cloud. This agreement is nothing else the quality of service that the user wants from the providers. Providing good quality of services to the users according to the agreement is a decisive task of the providers at the same time there are a large number of tasks running at the provider's side. The task scheduling problem can be viewed as the finding or searching an optimal assignment or mapping set of the subtasks of different tasks over the available set of the resources (computers or processor machines) so that we can achieve the desired goals for tasks.

Keywords: Cloud Architecture, Cloud Computing, Credit, Deadline, Task Scheduling

I. INTRODUCTION

Cloud computing is a new computing mode. It is similar to utility computing which involves a large number of computers connected through communication n/w. Cloud computing is trend to provide service as resources including hardware, software, network etc. Every service is provided over network that require high speed of network and persistence connection where its services are distributed over the network according to architecture and geo-location. It is based on pay as you go model, means it depends on matrices like usability, durability, cost, load etc. So that consumer does not need to buy any hardware, software etc. The main goal of cloud computing is to achieve higher throughput, availability, scalability, consistency guarantees, and usability, fault tolerance etc. used distributed resources [1]. Cloud computing resources should able to solve large scale of computation problems. Cloud computing uses characteristics of Client–server model, Grid computing, Peer-to-peer, Mainframe computer, Utility computing to provide better services like gaming, tons of computation, message passing, network etc. Cloud computing has an advantage of delivered a flexible, very high performance, pay-as-you-go, on-demand service. Operators should guarantee to the subscribers and stick to the Service Level Agreement. Google adopts Map-Reduce scheduling mechanism scheduling algorithms are relatively simple (First fit etc.). FIFO, default algorithm performs not so well for short jobs. Besides, Facebook proposes fair share scheduler; Yahoo raises computation ability scheduler. However, these scheduling algorithms cannot work out a better scheduling scheme. In fact, tasks scheduling in cloud is a NP complement problem with time limit. That is to say, it is seldom impossible to search out a reasonable solution in polynomial time. To improve performance of cloud computing, efficient task scheduling and resource management is required.

A. CLOUD COMPUTING

According to the R. Buyya " Collection of the inter-connected and virtualized computers in parallel and distributed computing system" are provisioned dynamically and based on service-level agreement (SLA) and presented as unified computing resources established by negotiation between the service providers of cloud and users is known as Cloud[2].

A.1 Deployment Models of Cloud Computing:

It represents a specific type of the cloud environment, firstly distinguished by ownership, size and access. The basic types of cloud deployment models are defined below:

a) Private Cloud: This type of cloud provides its service to department of large organization (single organization) which is managed by either third party or by same organization. It is hosted either internally or externally to consumer organization. Right evaluation can improve business but there can be lot of security issues.

b) Public Cloud: This type of cloud is organized by a cloud service selling organization over network to provide service as per pay-as-go model. For example Amazon AWS, Microsoft Azure and Google play are clouds and accessed via internet.

c) Hybrid Cloud: This type of cloud is composition of Private community or Public cloud. This cloud is used as per type of service requirements, means to access services a private setup is connected to the public setup.

B. *CLOUD COMPUTING ARCHITECTURE*

Front and back end are the most significant components of Cloud computing architecture. End which is visible to the user of cloud is front end. It has applications and computer that user uses to access the cloud. Storage devices and computers are the back end of the cloud computing. Figure 1 shows the graphical view of Cloud Computing architecture.

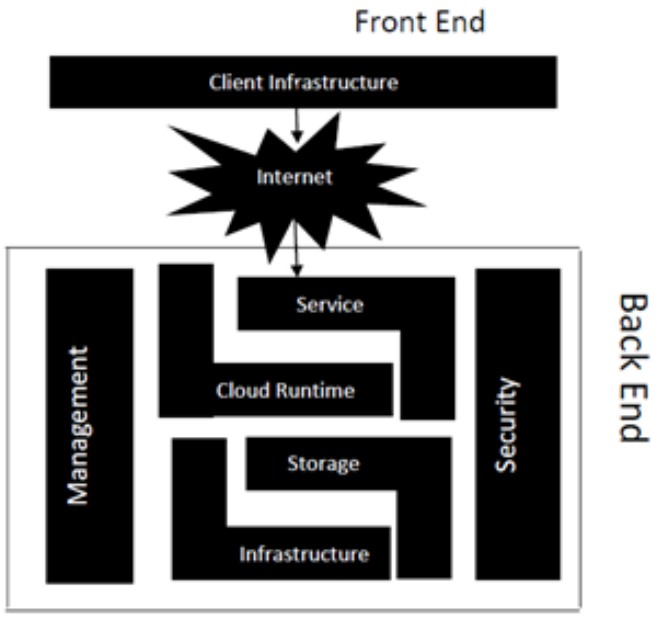

Figure 1: Cloud Computing Architecture

Cloud Architectures address key difficulties related to processing of large amount of data. It is really a hard work to allocate and co-ordinate a large scale job on different-different machines, run procedures on them, and supplies another machine to recuperate if one machine flops during the operation. Applications that built above Cloud Architectures run in-the-cloud where the real physical location of the substructure is determined by the cloud providers and basically take advantage of modest APIs of

Internet-accessible services that are able to scale on-demand, where the difficult logic of scalability and consistency of the services remains executed and hidden inside the huge cloud.

II. TASK SCHEDULING

Task Scheduling plays a key role in the Cloud Computing system. Scheduling of the task cannot be done on basis of single criteria. It is regulations and rules that can term as an agreement between users and providers of cloud. This agreement is nothing else simply the quality of service that the user or client requires. Providing good quality of the services to the users or clients according to the agreement it is a decisive task for the providers at the same time there is large number of tasks running at the side of provider's. The task scheduling problem can be viewed or seen as the finding or searching an optimal mapping of set of subtasks of different tasks over the available set of the resources (e.g. processors/computer machines) hence can be achieved or attained the desired goals for tasks. Scheduling is a method or procedure by which threads, processes and data flows are given access to system resources. Scheduling is the fundamental operating system function, almost all computer resources are scheduled before use. The idea of multiprogramming is relatively simple, if a job is waiting for an I/O request, then the CPU switches from that job to another job, so that it always busy in multiprogramming[3].

A. *Task Scheduling Types*

Task Scheduling is the method by which threads, processes or data flows are given access to system resources. Scheduling is a essential operating system function, nearby all computer resources are scheduled before use. There are various types of task scheduling which are discussed below [4].

a) *Preemptive Scheduling*: New process selected to run also when an interrupt occur when new processes become ready. Tasks are usually allotted with priorities. At periods it is compulsory to run a definite task that has a advanced priority before another task even though it is running. Consequently, the running task is intermittent for some time and continued later when the priority task has done its execution or implementation.

b) *Non-preemptive Scheduling:* New process is selected to run either a process terminates and explicit system request causes a wait state. In sort in non-preemptive scheduling, a task which are running is executed till completion.

c) *First Come-First Serve Scheduling(FCFS):* By far the simplest and easiest CPU-scheduling algorithm is the first come- first serve scheduling algorithm, By using help of this algorithm, processes are assigned the CPU in the order request it. Basically, there is a single queue of ready or prepared processes. Relative importance of jobs measured or calculated just by arrival time (poor choice). The implementation or execution of the FCFS policy is easily managed with a FIFO queue.

d) *Round Robin Scheduling(RR):* The round robin scheduling algorithm is planned or designed specifically for time-sharing systems or structures. It is related to FCFS scheduling, but pre-emption is additional to switch in the middle of processes.

e) *Priority-Based Scheduling:* A priority is allocated for each process. The prepared list contains an entry for the each processes ordered by its priority. The process at the creation of the list (highest priority) is picked first.

III. RELATED WORK

Cloud computing is an emerging technology, which changes things like resource allocation, data processing, monitoring etc. Scalability, availability and consistency are its keen concepts. Its challenges are like internet speed, data portability and interoperability, heterogeneous environment, security etc. A number of researches are provided on its challenges to provide efficient way for handling resources. In the review of literature, let's discuss some review papers of task scheduling in Cloud Computing environment.

Chopra and Singh [5], proposed the Deadline based scheduling process, the tasks using available resources. . This presents an algorithm that decides which resources should be taken on lease from the public cloud to complete the execution workflow with minimum monetary cost for user within the deadline. A new concept of hybrid scheduling algorithm has been proposed which uses a sub-deadline for rescheduling and allocation of resources in public cloud. This algorithm helps in finding the best resources on the public cloud for cost saving and complete workflow execution within deadlines

Han et al. [6], presented the management of resources and scheduling in cloud computing environment, complex undertaking mainly due to resource's geographic distribution, heterogeneity and dynamic. A Qos guided task scheduling model, being composed of the some scheduling strategies and a QoS guided scheduling sufferage-min heuristic algorithm, presented. In this shows that the make span value, the key performance, successfully being shorten. QoS an extensive concept and varies from novel scheduling heuristic by considering QoS factor in scheduling and have proposed some modifications which using the existing Sufferage heuristic and Min-min heuristic.

Chopra and Singh [7], proposed the Cloud computing demand resources for compute and storage requirements. Private cloud a good option for cost saving for executing workflow applications but when the resources in private cloud are not enough to meet the storage and to compute requirements of an application then public clouds are the option left. While public clouds charge the users on basis of pay-per-use, private clouds are owned by users and can be utilized with no charge. Deadline based scheduling the main focus in many of the workflow applications. Proposed algorithm does cost optimization by deciding which of the resources should be taken on the lease from public cloud to complete the workflow execution within deadline.

Santhosh and Ravichandran [8], presented the Pre-emptive Scheduling of On-line Real Time Services with task migration for CC. This presents a new scheduling approach to focus on providing a solution for online scheduling problem of the real-time tasks using "Infrastructure as a Service" model offered by CC. The currently executing task pre-empted as soon as the task with the higher priority enters into the ready queue. The pre-empted task re-enters into the ready queue and sorted according to their priority. A task will be migrated to new virtual machine, when it misses its deadline. Migration algorithm executes the tasks completely and improves the total utility and efficiency.

Dubey et al. [9], introduces an improvement in Task duplication based scheduling Algorithm for Network of Heterogeneous systems by applying the scheduling method of cluster based in order to cluster various submitted tasks and assigned them to appropriate processors. A cloud system, where various tasks are executed on different processors in a network, cost of communication becomes important parameter to be considered. The performance of the improved algorithm show by the comparing Cluster Completion Time of the cluster based scheduling algorithm with the earliest completion time (ECT) of the original TANH algorithm.

Abdullaha and Othmanb [10], investigated the Cost-Based Multi-QoS Job Scheduling using the Divisible Load Theory in cloud computing, attempt to investigate the use of a Divisible Load Theory (DLT) to design efficient strategies. The use and applicability of the DLT paradigm in a compute cloud environment demonstrated by developing a distribution strategy that shown to minimize the overall total cost. In this designed and analyzed a closed form the solution for jobs scheduling problems and validate all our findings via rigorous simulation experiments.

Gao et al. [11], proposed the Cloud computing has attracted significant attention due to the increasing demand for low-cost and energy-efficient computing and high performance. Profit maximization for the cloud service provider a key objective in the large-scale, multi-user and heterogeneous environment of a cloud system. In the envisioned cloud environment, users can constructs their own application and services based on the available set of virtual machines, but are relieved from burden of the resource provisioning and task scheduling. Cloud service provider will then exploit the data parallelism in user workloads to creates an energy platform and deadline aware cloud platform.

Aminizadeh and Yousefi [12], proposed the two categories of reasons have led to the emergence of vehicular cloud computing, vehicular networks and the new generation of well-equipped smart cars and on another, the advent of the cloud computing and a short time-span maturity over it. The goal of this study to offer an application scheduling model for determining optimum response needed for the management of dynamic vehicular cloud resources in a way that tasks are completed with their minimum cost, before deadlines and within the lifetime of the cloud. To solve the mentioned problem, a binary integer program model formulated here, and the impact of changes in various parameters such as applications deadlines, different tasks costs, types and lifetime of the created clouds are analyzed and evaluated.

Lakra and Yadav [13], proposed Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization. A multi-objective task scheduling algorithm form mapping tasks to a Vms in order to improve the throughput of the datacenter and reduced cost without violating the Service Level Agreement for an application in cloud SaaS environment. This proposed algorithm provides an optimal scheduling method. Most of the algorithms schedule tasks based on the single criteria means execution time but in cloud environment it required to consider various criteria like cost, execution time, bandwidth of user etc. This algorithm simulated using Cloud Sim simulator and the result shows better performance and improved throughput.

Thomas et al. [14], proposed that Credit Based Scheduling Algorithm in CC Environment. In order to achieve good services from a cloud, the need for a number of resources arose but cloud providers are limited by the amount of resources have and are thus compelled to strive to maximum utilization. Min-Min algorithm used to reduce the make span of tasks by considering the task length.

Xue et al. [15], main goal was to put forward a task scheduling algorithm in cloud computing with the goal of the maximum load balancing degree, minimum completion time and minimum energy consumption using improved differential evolution algorithm. To improve global search ability in the earlier stage and the local search ability in the later stage, have adopted the adaptive zooming factor mutation strategy and adaptive crossover factor increasing strategy. . In the process of simulation performed the functional verification of the algorithm and compared with the other representative algorithms.

Bochenina [16], proposed a comparative study of scheduling algorithms for the multiple deadline-constrained workflows in the heterogeneous computing systems with time windows. Scheduling tasks

with precedence constraints on a set of resources with different performances a well -known NP-complete problem and a number of effective heuristics has been proposed to solve it. A staged scheme for a multiple deadline-constrained workflow scheduling and run a series of experiments based on the synthetic test examples in order to analyzed and compare three proposed scheduling strategies in accordance with four metrics.

IV. CONCLUSION AND FUTURE SCOPE

Task scheduling is one of the most well-known problems in cloud computing so; there is always a chance of amendment of previously completed work in this specific field. The researchers at their own time performed their work according to their knowledge space and after some interval their work had been carried out some other people. During scheduling they had well thought-out various techniques and applied constraints but as the cloud computing is too vast that they had not been able to capture the all aspects at same time but they mentioned these facts that there is a chance of amendment of algorithms and which part has to be modified.

REFERENCES

[1] Yashpal Singh Jadeja and Kirit Modi, "Cloud Computing - Concepts, Architecture and Challenges", in the proceedings of 4th International Conference on Computing, Electronics and Electrical Technologies, 2012.

[2] R.Buyya, C.S.Yeo, S.Venugopal, J.Broberg, and I.Brandic, Cloud Computing and emerging IT Platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems, 2009, pp. 599-616.

[3] Tayal, Sandeep. "Tasks scheduling optimization for the Cloud Computing systems." International Journal of Advanced Engineering Sciences and Technologies, Vol. 2, 2011, pp. 111-115.

[4] Kamal Kc, and Kemafor Anyanwu. "Scheduling hadoop jobs to meet deadlines." in the proceedings of IEEE 2nd International Conference on Cloud Computing, 2010, pp. 388-392.

[5] Nitish Chopra and Sarbjeet Singh "HEFT based Workflow Scheduling Algorithm for Cost Optimization within Deadline in Hybrid Clouds" in the proceedings of 4th IEEE, ICCCNT, pp. 31661-31666, 2013.

[6] Haiwen Han, Qi Deyu, Weiping Zheng and Feng Bin "A Qos Guided task Scheduling Model in cloud computing environment" in the proceedings of 4th IEEE International Conference on Emerging Intelligent Data and Web Technologies, pp. 72-76, 2013.

[7] Nitish Chopra and Sarbjeet Singh "Deadline and Cost based Workflow Scheduling in Hybrid Cloud" in the proceedings of IEEE International Conference on Advances in Computing Communications and Informatics (ICACCI), pp.840-846, 2013.

[8] R.Santhosh and T.Ravichandran "Pre-emptive Scheduling of On-line Real Time Services With Task Migration for Cloud Computing" in the proceedings of International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp. 271–276, 2013.

[9] Shivani Dubey, Vismay Jain and Shailendra Shrivastava " An Innovative Approach for Scheduling of Tasks in Cloud Environment" in the proceedings of 4th IEEE, ICCCNT, pp. 31661-31668, 2013.

[10] Monir-Abdullaha and Mohamed-Othmanb "Cost-Based Multi-QoS Job Scheduling using Divisible Load Theory in Cloud Computing" in the proceedings of ICCS International Conference on Computational Science, pp. 928–935, 2013.

[11] LizhYue Gao, Yanzhi wang, Sandeep K. Gupta and Massoud Pedram "An Energy and Deadline Aware Resource Provisioning, Scheduling and Optimization Framework for Cloud Systems" in the proceedings of IEEE, pp. 1417-1426, 2013.

[12] Leila Aminizadeh and Saleh Yousefi "Cost Minimization Scheduling for Deadline Constrained Applications on Vehicular Cloud Infrastructure" in the proceedings of 4th IEEE International Conference on Computer and Knowledge Engineering(ICCKE), pp. 358-363, 2014.

[13] Atul Vikas Lakra and Dharmendra Kumar Yadav "Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization" in the proceedings of IEEE International Conference on Intelligent Computing, Communication & Convergence, pp. 321-327, 2014.

[14] Antony Thomas, Krishnalal and Jagathy Raj "Credit Based Scheduling Algorithm in Cloud Computing Environment" in the proceedings of International Conference on Information and Communication Technologies, pp. 913-920, 2014.

[15] Jing Xue, Liutao Li, SaiSai Zhao and Litao Jiao "A Study of Task Scheduling Based On Differential Evolution Algorithm in Cloud Computing" in the proceedings of 6th IEEE International Conference on Computational Intelligence and Communication Networks, pp. 637-640, 2014.

[16] Klavdiya and Bochenina "A comparative study of scheduling algorithms for the multiple deadline-constrained workflows in heterogeneous computing systems with time windows" in the proceedings of 14th International Conference on Computational Science, pp. 509–522, 2014.

