Literature on the Hybrid Approaches and Load Balancing in the Cloud

Y Shyam Sundar, Research Scholar, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)
Dr. Prateek Mishra, Professor, Department of Computer Science, SunRise University, Alwar, Rajasthan (India)

ABSTRACT:

Recommender systems represent a high economic, social, and technological impact at international level due to the most relevant technological companies have been used them as their main services considering that user experience and companies sales have been improved. For this reason, these systems are a principal research area, and the companies optimize their algorithms with hybrid approaches that combine two or more recommendation strategies. A systematic literature review on the hybrid approaches for recommender systems is generated by this work, the objectives are to analyze research line progress and to identify opportunity areas for future investigations. Further, the recent trends about challenges, methodologies, datasets, application domains and evaluation metrics on hybrid approach are identified. An art state from 2016 to 2020 is developed with information systems guide than unlike others works that use less recent guide and software engineering guide. This research will benefit recommender systems community.

Keywords: Load, Balancing, Cloud, Literature

INTRODUCTION:

Nowadays, recommender systems represent a high economic, social and technological impact at international level due to the main technological companies as Google, Facebook, Twitter, LinkedIn, Netflix, Amazon, Microsoft, Yahoo!, eBay, Pandora, Spotify and others more have been used these systems in their leading services [1]. Further, these systems have been contributed on information overload problem, user experience, user decision making, and companies sales [2]. A recommender system is a software tool that suggests items as products, films, jobs, friends, web sites, songs and other more items based on the tastes users [3]. Besides, their performance can be optimized with hybrid approaches that combined two or more recommender algorithms to complement their disadvantages with advantages of other algorithms. For this reason, big money amounts are inverted to optimize algorithms and to develop new research on hybrid approach. This research presents a systematic literature review on the hybrid approaches for recommender systems that is designed with an Okoli and Schram's information systems guide [4]. Systematic literature review is a systematic, explicit, and reproducible method for identifying, evaluating, and synthesizing the completed and recorded research produced by researchers, scholars, and practitioners. Additionally, other related works used a software engineering guide, however, information systems are more related to recommender systems than software engineering. First, load in load balancing refers not only to website traffic but also of memory capacity, CPU load and network on server. Its main function is to ensure every network of system has the same amount of work. It means neither that the system is under used or overloaded. It makes equal distribution of data based on how busy the server is. Without this the client has to wait long for processing this data thiscould be a frustrating for them. During this process, data like CPU process and job arrival rate in the processors are modified. Failures in application of this head have severe consequences like data loss. Different companies utilize various load balancers with numerous load balancing techniques. The most commonly used model or techniques is the "Round Robin" load balancing.

Load indicates not only a traffic website but includes network load, memory capacity and CPU load of any server. This method promises that every network in system has similar number of work at a time. Any of them is highly under or over-loaded use. It gives data based on how busy every node or server is.

Cloud computing is an emerging technology and has attracted a lot of attention in both commercial and academic spheres. Cloud computing has moved computing and data away from desktop and portable PCs into large data centers (Sharma & Banga, 2013). In cloud computing, users do not know where the infrastructures are located. The users only use the services through the cloud infrastructure paradigm and pay for the requested services

(Armbrust et al., 2010). Virtualization is a key enabling technology for cloud computing environments, which makes it possible to run multiple operating systems and multiple applications on the same hardware at the same time, so as to provide services by a virtual unit (Armbrust et al., 2010). The National Institute of Standards and Technology's (NIST) define cloud computing as " a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell & Grance, 2011). In cloud computing, the biggest challenge is how to handle and service the millions of requests that are arriving very frequently from end users efficiently and correctly. Thereby, the need for load balancing. Load balancing is one of the most important issues in cloud computing to improve the performance. Load balancing is a process of distributing load. The load is distributed on individual nodes to maximize throughput, and to minimize the response time. It also removes a condition in which some of the nodes are heavily loaded while some others are light (Ramana, Subramanyam &AnandaRao, 2011) In general load balancing algorithms are classified as static or dynamic and centralized or distributed (Abubakar et al.,2004). A comprehensive study was carried out, where the comparisons are conducted among static, dynamic, and hybrid algorithms. The results show that the hybrid algorithms give efficient results since they inherit the benefits of other algorithms, and avoid drawbacks (Siham, 2017).The objective of this paper review is to unveil the existing hybrid load balancing techniques in cloud computing which are more efficient and will take in to account the overall network loads, energy efficiency with better quality of service satisfaction.

LITERATURE REVIEWED

Literature, shows the presence of many load balancing procedures, resource designation methods, advancement systems and calculations for building throughput and productivity and improve the reaction time in cloud condition. Every methodology has its own advantages.

The Çano y Morisio's quantitative research [3] is a principal related work because they developed an art state from 2005 to 2015 on hybrid recommender systems. Furthermore, they used a software engineering guide to analyze challenges and solutions with data mining and machine learning techniques. The most important challenges identified were cold-start and data sparsity. The new problems identified was recommendations on multi domains, context variation and evolution of tastes users. The foremost application domain was the movies. Accuracy metrics for evaluation were the more used. Finality, the new opportunities identified were context recommendations, parallel hybrid algorithms and processing of biggest datasets.

Resource scheduling and designation assume an indispensable job in CC generally to create performance implementation and use of resources, vitality sparing, clients QoS necessities fulfillment as well as expanding the benefit of cloud suppliers. Moreover, its calculation as well as strategy legitimately impact cloud cost and execution.

Qingxian et al. (2019), have investigated a case wherein the basic data sources (or shared resources) of all stages are known. Building a game that sees each phase as a player, helps incorporation of a game hypothesis with network data envelopment analysis (DEA) for the investigation of the result distribution issue in a three-organize framework. Network DEA models are made for computing the ideal benefits of the framework during resource sharing (i.e, pre-and post- joint effort ideal benefits), and the Shapley esteem technique is applied for the allotment of the expanded benefits of the framework to its stages.Results show the game among stages in a three-organize framework as an added substance. A numerical model is given to show our the technique.

XingChen et al. (2019) have proposed a self-adaptive resource allocation strategy that is really a structure made out of criticism circles, every one of which experiences a planned iterative QoS forecast model and PSO-based runtime decision algorithm. As opposed to past Qo forecast models which foresee a QoS esteem for the last time, thisstrategy improves the anticipated QoS esteem towards the best one. In the forecast, the remaining task at hand is

first utilized along with, the designated resource, the genuine QoS esteem and an activity of resource allocation to deliver a QoS esteem. At that point PSO-based runtime decision algorithm is utilized together with the anticipated QoS incentive to decide on future resource allocation activities. The circles are repeated until the PSO-based algorithm proposes no further improvement over the present resource allocation. The methodology is assessed on RUBiS benchmark. Representing this and based on theequivalent verifiable data, the strategy can accomplish a superior QoS expectation exactness that is 15% higher than the present cutting edge. Besides, a gained 5-6% improvement of the viability of cloud application resource allocation has been demonstrated.

Manasrah, A.M et al. (2018) Cloud computing condition permits resource sharing as well as on-request benefits for customers. Business forms are kept up by work process advancements that feature the difficulties seen in utilization of resources in a productive manner due to conditions between errands. Hybrid GA_PSO's technique reducesexpense and parity heap on the required undertakings. This technique is utilized for the productive use of resources. Exhibition of this technique is contrasted with different techniques like PSO, GA, MTCT,WSGA and HSGA. What's more, it improves the heap adjusting of the work process application over the accessible resources.

Also, the multifaceted nature of the work process scheduling issue, makes building up an enhanced work process scheduling algorithm for work process errands dissemination to the accessible resources inside a sensible overhead, that is, CPU time an exceptionally testing process.

Wei et al. (2018) have indicated the dependence of resource allocation of cloud method on not perfect data Stackelberg game (CSAM-IISG) utilizing HMM in CC condition. CSAM-IISG has appeared to expand the benefit of both resource provider and candidate.

Initially, HMM was utilized to anticipate specialist organization's present offer utilizing the chronicled resources dependant request. Progressive foreseeing of offer helped building up the not perfect data Stackelberg game (IISG). IISG spurs specialistorganizations the pick the ideal offering procedure as per the general utility, accomplishing the most extreme benefits. Dependence on unit costs of various sorts of resources, resource allocation method is suggested to ensure ideal increases for foundation provider. Suggested resource allocation method bolster synchronous allocation for both multiservice suppliers as well as different resources.

Moon, Y.et al. (2017) Computing resources from CC is charged depending on climate. Similarly scheduling resource technique has a complex effect on clients. A novel cloud scheduling technique is utilized. It is like for deciding on this ACO technique that makes a profitable assignment of assets to VM. Techniques of diversification as well as fortification are taken up utilizing slave ants. These ants handle the issue of enhancement. New ACO technique known as SACO with slave ants staying away from overheads provide good execution. This technique deals with the NP-difficult problem even more beneficially.

Thanasias et al. (2016) CC has picked up significance in numerous fields conveying different changes to data advancement. IaaS gives flexible providing and denial of computing assets. It is a compelling method for workloads and a temporary, test or vary startlingly. The prerequisite for assets shifts from time to time and oughtto be kept inside the given budget while keeping up efficiency. The need for resource variation after some resource provisioning such of obtained budget is productively utilized for maintain obtaining job done efficiency is an important challenge. This resource exhibits the task scheduling problem and provisioning resource for numerous tasks in IAAS cloud. It gives novel scheduling and provisioning techniques used for task implementation inside the budget for limit log jam due to budget constraint. Choudhary et al. (2018) had proposed the application of workflow scheduling in CC to provide solution for NP complete issue. Ithas numerous problems, vitality issues, time span and cost. Numerous meta and heuristic techniques are suggested. The technique satisfies the method to some degree showing integration of meta and heuristic technique such as GSA and HEFT. Significant characters are MCR and SLR. Different trial causes are represented in this technique outputs like HGA, HEFT and GSA. Outcomes are produced by ANOVA test.

Future findings are in numerous cloud condition. Madni et al. (2016) had dealt with the allocation of resources using meta-heuristic techniques for IaaS CC condition. Different problems have been noticed in resource allocation meta heuristics techniques, comparative specification as well as experimental instruments used for various method validation. Classification and survey are the bases for additional researches in IaaS CC.

Ma et al. (2014) have presented 5 significant problems in scheduling resource and CC allotment including locality aware,reliability aware and vitality aware resource of scheduling and allocation. SAAS includes scheduling and allocation as well as scheduled work process. They have made thorough analysis and conversion of different current resource allocation as well as the sequence of scheduling and techniques of current issues as away as various specifications. Zhang and Su (2014) have done research on basic cloud information target resource scheduling network with its problems. They have seen cloud data focus thought structure and scheduling resource. They have portrayed a method for scheduling resources for cloud data target with a condition to powerful scheduling of cloud resources and less power task scheduling. This research work shows problems in current field of data target compared to different scheduling resources. Use and profit of resources are low for cloud suppliers and vitality use in data community is high. There is need to improve resource scheduling for data centersfor future work.

The basic technology of CC is resource scheduling in resource management. It deals with techniques to enhance efficiency with dynamic scheduling relying on threshold, improved genetic technique with double available and increased ACO for scheduling as suggested by Huang et al. (2013). Areas taken up have been committed with Map decline scheduling research include graph models, dynamic requirement, temporary weight modify, adaptive scheduling, utility based optimization, customization, forecast, equality numerous clients, audit of map reduce entomb reliance and enhance lessening stage. A major task was to increase overall efficiency, reaction time and increase output producing fairness and locality. Open area of work for newapplications had increased the makespan and improved fairness in various clients (Elghoneimy et al., 2012).

Load balancing is a computer networking method to distribute workload across multiple computers or a computer cluster, network links, central processing units, disk drives, or other resources, to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overload (Kaur &Bansal,2013) Load balancing appears to be the major challenge in cloud computing, due to heterogeneous nature of cloud environment where resource pool is on increase (Gabi, Ismail & Zainal, 2015). The load balancer accepts multiple requests from the client and distributing each of them across multiple computers or network devices based on how busy the computer or network device is. Load balancing helps to prevent a server or network device from getting overwhelmed with requests and helps to distribute the work (Gabi, Ismail & Zainal (2015). Providing an efficient load balancing in cloud computing enables efficient resource utilization, achieved higher user satisfaction and also prioritizes users by applying appropriate scheduling criteria (Katyal & Mishra,2012).

Wu et al. (2013) have suggested the QoS scheduling task with the aim of implementing huge required tasks on resources which has numerous least time. Requirements are concluded to satisy special Qos parameters. The technique is in contrast to Min-Min technique andBerger method as well as the makespan of suggested method had been found to be superior to the other two. Amit Nathani et al. (2011) have suggested a technique in scheduler named Haizea for resource allocation such as best exertion, deadline touchy, adverted reservation and immediate. Haizea is resource lease manager which uses resource leases as abstracts of resource allocation and actualizes leases by VMs allot. The important aim of the authors is to limit the resource dismissal rate and the shuffle cost to give above said resource allocation sequence for the IAAS cloud. It uses 2thoughts, namely, backfilling and swapping for deadline touchy resource allocation criteria. The main idea is leasing 4 specifications for trails, namely, duration, start time, number of hubs and deadline. Kejiang Ye et al. (2011) have suggested resource reservation dependant on live shift structure of different VMs. Focused machine in structure has 4 VMs, namely, migration decision maker and controller,

Resource monitor and reservation control.

The authors have targeted migration performance improvement by live shifting of VMs and suggested 3 optimization methods, namely, source machine optimization, numerous machine parallel migration and workload aware shifting criteria. The authors have used specifications like total migration time, workload efficiency and downtime for enhancing shifting performance. He claims resource reservation method is need at source machine as well as focus machine. Congfeng Jiang et al. (2011) have presented a compelling resource allocation issue that depends on the real time data on workload as well as efficiency request for processing services. They have suggested the stochastic method of resources in virtual condition and scheduling heuristics techniques and resource allocation with service level constraints. Targeted machine efficiency has been taken as efficiency feedback to source for enhancing the viability of things to approach a dynamic workload. This improves the resource allocation method suggested by authors. Linlin Wu et al. (2011) have suggested the resource allocation technique for SAAS suppliers which limits framework cost as well as SLA violation for SAAS purchasers to ensure service satisfaction. The authors have considered buyers QoS specifications, as for instance, framework specifications and reaction time in server start time. They have presented 3 cost driven techniques from 2 shoppers as well as SAAS suppliers view. The first technique was one which increasesprofit by adjusting the number of SLA violations. Consecutivetechniques increased profit by cost reuse of VMs limit, with a huge space. The third technique increased profit by cost reuse VMs limit with less space. The second and the third suggested by authors were simulated on cloud sim condition.

Query structure is based on six keywords and synonyms set for each keyword. The first three keywords are gotten of Çano y Morisio's related work [3], other three keywords are proposed by this work to include challenges, datasets and evaluations on recommender hybrid approaches. The Table 1 describes the keywords with their synonyms to generate the query terms. AND/OR Boolean operators are used to define the query with the previous terms. Two subqueries are joined, the first subquery represents the general searching on hybrid recommender systems, and the second subquery represents the specific searching about challenges, datasets, and evaluations on recommender hybrid approaches. The Table 2 shows the query developed with the terms previously defined. Searching for the literature is realized with five scientific databases previously selected and a query previously defined. The general results are 383,937 publications retrieved of which, Spring Link retrieves 319,124 publications, ACM Digital Library retrieves 37,349 publications, Scopus retrieves 12,921 publications, Web of Science retrieves 12,717 publications and IEEE Xplore retrieves 1,826 publications. Springer Link retrieved the most publications amount, and IEEE Xplore retrieved the least publications amount. However, practical screen and quality appraisal steps will select the valid publications to generate the art state.

Different types of resource allocation techniques have been suggested in cloud. Gunho Lee et al. (2011) have suggested a structure for optimized resource allocation in IAAS based cloud structure. CurrentIAAS structure is unaware of facilitated apps significant. This pathallotted resources free of requirements with a major effect in the efficiency for disseminated data serious apps. Structure that has "what if" method to manage allocate decision taken by IAAS has been suggested for locating this resource allocation problem. The structure used an expectation motor with lightweight simulator for calculation of the efficiency of the given resource allocation and GA to find optimized sequence in large search space.

Javanmardi et al. (2014) proposed a hybrid job scheduling algorithm using genetic algorithm and fuzzy theory. The proposed algorithm assigns jobs to resources without considering the job length and resources capacities. Genetic algorithm was used as the basis of the approach. This was modified with the aid of fuzzy theory to reduce the iteration of producing the population. Two types of chromosomes with different QOS parameters were defined; then the fitness value of all the chromosomes for the mentioned two types were obtained. The new approach with the use of fuzzy theory modified the standard genetic algorithm and improved system performance in terms of execution cost to 45% and total execution time to 50%.This

algorithm was implemented on CloudSim and the results showed that this hybrid approach outperformed other algorithms. Bhowmik et al. (2016) proposed an efficient load balancing approach in a cloud computing platform.This work proposes a priority based virtual machine (VM) arrangement and load balancing using round robin scheduling. The VM is arranged by comparing the vital parameters which involve bandwidth, RAM, and MIPS (million instructions per second). The new algorithm is based on round robin for load balancing and ensures that jobs are not scheduled according to the VM priority and hence no VM is inundate or remains inactive for a longer period of time. This new method eliminates the process of excess burden on any individual VM and enhances the resource utilization. Cloudsim was used as a simulation tool to simulate the algorithms. The experiment results clearly indicated that the proposed method task scheduling consumes lesser time than conventional Round Robin Algorithm. Wang et al., (2011) proposed a two-phase scheduling load balancing algorithm (OLB+LBMM). This algorithm combines Opportunistic Load Balancing (OLB) and Load Balance MinMin (LBMM) scheduling algorithms. OLB scheduling algorithm keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to minimize the execution of time of each task on the node thereby minimizing the overall completion time. This algorithm works to enhance the utilization of resources and enhances the work efficiency. Tasquia et al.(2012) proposed a modified task scheduling algorithm based on the concept of beelife algorithm (BLA) and greedy algorithm to get an optimal service in hybrid cloud. Bee life algorithm are residential for task scheduling while the greedy method will select randomly forone data center. In modified job scheduling the jobs are processed in the queue using non-preemptive priority queue. First the tasks enter to the BLA algorithm and to greedy method to get optimal solution. The make span can be reduced by this hybrid algorithm. Liu, Luo, Zhang, Zhang and Li (2013) proposed a multiobjective genetic algorithm (MOGA)-based scheduling algorithm. This algorithm combines both random and greedy initialization methods. In genetic algorithm the fitness is calculated by energy consumption and profits of the service providers. The best fitness is selected and stored in pareto. Then selection operation can be done by two strategies: elitism and crowding. After the selection process crossover operation can be done by two individuals. Mutation swaps their position to generate new individuals This MO-GA is used to minimize the energy consumption and to maximize the profit of service providers. Parsa and Entezari-Maleki (2009) proposed a new algorithm RASA (Resource Aware Scheduling Algorithm): A new task scheduling algorithm in grid environment. The algorithm was built through a comprehensive study and analysis of two wellknown task scheduling algorithms, Min-min and Max-min. RASA uses the advantages of both algorithms and covers their disadvantages. According to this algorithm at first the scheduler allocates the resource to tasks according to number of available resources. The available resource is added then it

REFERENCES:

[1] Abdulhamid .S. M, Latiff .M. S. A and Idris .I (2015), "Tasks Scheduling Technique using League Championship Algorithm forMakespan Minimization in IaaS Cloud", ARPN Journal of Engineering and Applied Sciences, Vol. 9, pp. 2528-2533.

[2] Alexander .S (2014), "Efficient Cloud Storage Confidentiality to Ensure Data Security", IEEE CCSW 2014 International Conference on Computer Communication and Informatics, Vol. 03, No. 05.

[3] Amanpreet Kaur and Bikrampal Kaur (2019), "Load balancing optimization based on hybrid Heuristic-Metaheuristic techniques in cloud environment", Journal of King Saud University - Computer and Information Sciences.

[4] Amit Nathani, Sanjay Chaudhary and Gaurav Somani (2011),"Policy based resource allocation in IaaS Cloud" in ELSEVIER - Future Generation Computer Systems, Vol. 28, pp. 94-103.

[5] An Qingxian, WenYao, Ding Tao and Li Yongli(2019), "Resource sharing and payoff allocation in a three-stage system: Integrating network DEA with the Shapley value method", Elsevier, Journal of omega, Vol. 85, pp. 16-25.

[6] Anuradha .V.P and Sumathi .D (2014), "A Survey on Resource Allocation Strategies in Cloud Computing", ICICES - S.A.Engineering College, Chennai, Tamil Nadu, India.

[7] Armougum .A, Orriols .E, Gaston-Bellegarde .A, Marle .C. J.-L and Piolino .P (2019), "Virtual reality: A new method to investigate cognitive load during navigation", Journal of Environmental Psychology,101338. doi:10.1016/j.jenvp.2019.101338