Volume 4, Issue 11, November 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Survey of Recent Task Scheduling Strategies in Cloud Computing

Mahesh S. Shinde*
Computer Department, AISSMS COE
Pune University, India

Anilkumar Kadam
Computer Department, AISSMS COE
Pune University, India

Abstract— ***Cloud computing is a recent technology that concern with online distribution of computing resources and services on pay- per- use basis. These dynamically scalable resources within a cloud are managed by cloud service provider and distributed among the number of users according to the contract known as Service Level Agreement (SLA). After realization of benefits of cloud computing, number of users using cloud services are increasing tremendously. Therefore task scheduling plays crucial role in allocating (scheduling) cloud resources among the users efficiently. An efficient task scheduling policy provides proper resource utilization, load balancing and optimization of execution cost and time. In this review paper, we have given an overview of research work done by several researchers in the area of cloud task scheduling.***

Keywords— cloud computing, cloud services, Service level agreement, task scheduling, load balancing, resource utilization.

I. INTRODUCTION

Cloud computing is model that allows you to use dynamically scalable and shared pool of resources over the internet on pay-per-use basis. There are two actors involved in the cloud computing: cloud service provider and cloud service user. Cloud service provider owns the computing resources which are used by cloud consumers. In cloud computing, end user does not need the knowledge about the configuration of service provider because client just uses services on pay-per-use model. Cloud service provider handles all system configuration and resource management. Fig. 1 shows general view of cloud computing environment.

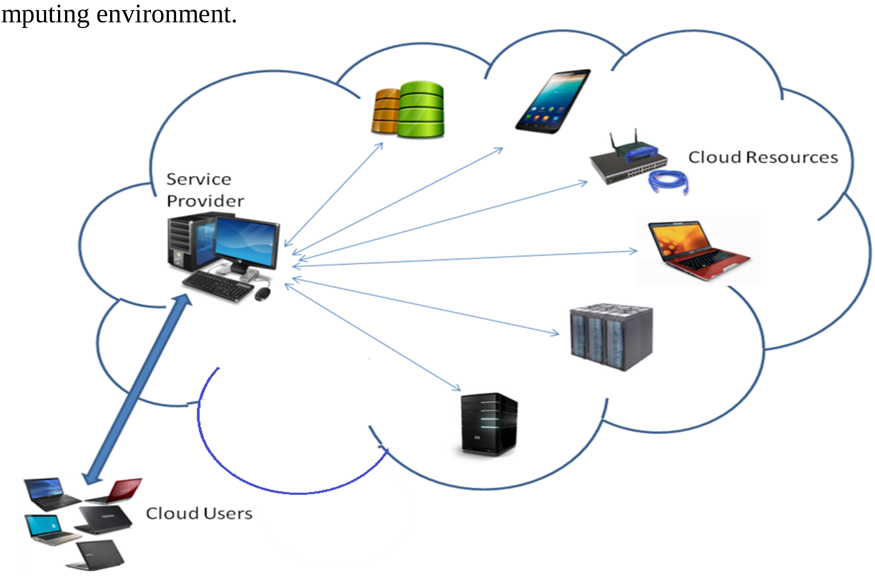

Fig. 1 General view of cloud environment

The services provided by the cloud are categorized into the following three cloud service models [12]:

Software as a Service (SaaS): It provides ability to cloud users to access and use the applications of cloud provider on pay-per-use basis. User can access these applications simply through browser while cloud provider manages the underlying infrastructure required for running such applications.

Platform as a Service (PaaS): In this service model, cloud service provider distributes computing platform so that users can develop their own applications using programming languages without having any overhead of managing underlying hardware and software layers. The provided computing platform may consist of operating system, environment to support program execution, database management systems etc.

Infrastructure as a Service (IaaS): It gives ability to the users for using the infrastructure (Physical resources) such as processors, storage disks, RAMs, routers etc provided by the service provider on pay-per-use basis. Using this model, small organizations can avoid the huge cost of buying such infrastructure. These physical resources are virtualized in order to share them among multiple cloud users.

Also there are following four cloud deployment models which show the ways through which cloud services are used by its users [13].

Private Cloud: Private cloud is build for the exclusive use by single organization. That means all the resources provided by the private cloud are accessed and used only by users of the organization who owns that cloud. Main benefit of using private cloud is its security since its resources are shared within different users of same organization. Another advantage of private cloud is its ability to provide customization which allows organization to mold it according to requirement. But the problem with private cloud is that it provides less scalability.

Community Cloud: Community Cloud allows for sharing its resources among the users of multiple organizations which are having same requirements and objectives. This cloud divides initial establishment cost among several organizations. These clouds provide somewhat more scalability of resources than private cloud.

Public Cloud: Public cloud provides unlimited storage, services and computing environment to the users all over world through web on pay-per-use basis. Public clouds are built and managed by third party agencies. Public clouds provide more scalability, availability and flexibility than private clouds. But insufficient security is a major problem with the public cloud since the resources provided by public cloud are shared among large number of worldwide users from different organizations.

Hybrid Cloud: Hybrid clouds are built by combining the private and public clouds. Hybrid cloud thus aggregates the properties of both private and public clouds such as scalability, flexibility and security. In this model, users of private clouds use the resources of public cloud when its own resources become insufficient. The extra required resources are taken from public cloud on pay-per-use basis.

II. TASK SCHEDULING

Scheduling is the group of strategies that manage the order of execution of multiple tasks on the processors in order to decrease the time and cost required to execute all these tasks. In the cloud environment, task scheduler plays vital role of allocating cloud provider’s resources among the large number of users. Task scheduling deals with distribution of the tasks among the cloud servers which process or execute these tasks for user (or client). An efficient task scheduling policy provides proper utilization of resources, load balancing and optimization of execution cost and time. Therefore today task scheduling is main research topic in the area of cloud computing. There are various types of scheduling such as static, dynamic, pre-emptive, non pre-emptive, centralized and distributed scheduling.

The main aim of this paper is to give the survey of recent strategies related with task scheduling in cloud computing. In the next section, we will take an overview of these strategies.

III. LITERATURE SURVEY

A] A Tri Queue Scheduling (TQS) algorithm : AV.Karthick, Dr.E.Ramaraj and R.Kannan in [1] proposed a Tri Queue Scheduling (TQS) algorithm which groups the jobs into the three separate queues – long queue, medium queue and small queue – depending on time and number of processors required to execute them. The fragmentation of the jobs in case of the existing algorithms like First-Come First-Serve, Shortest Job First, EASY and Combinational Backfill is avoided by using dynamic time slice based Round Robin mechanism of scheduling. Thus this algorithm provides equal opportunity to all jobs which removes the problem of starvation of jobs which was observed in above mentioned existing algorithms. This algorithm thus allows the proper utilization of resources and improves the performance.

B] Cluster Based TANH algorithm: Shivani Dubey, Vismay Jain and Shailendra Shrivastava in [2] presented cluster based TANH algorithm (Task duplication based scheduling Algorithm for Network of Heterogeneous systems) as an improvement over original TANH algorithm to achieve the goal of minimization of overall task execution time in heterogeneous environment. The inter-dependant tasks that are executing on the heterogeneous processors require considerable amount communication cost in terms of time. This modified algorithm groups the tasks which are dependent on each other for execution into a single cluster and assign them to the same processor so that communication cost between the tasks is eliminated. Performance is evaluated by comparing the Cluster Completion Time (CCT) of the cluster based TANH algorithm with the earliest completion time (ECT) of the existing TANH algorithm.

C] A Cost-based Resource Scheduling Paradigm: Zhi Yang , Changqin Yin and Yan Liu in [3] designed algorithm which use the market theory to achieve cost based resource allocation. This proposed algorithm satisfies user’s requirement of getting compute resource at minimum price by considering resource availability and price of the resource provider. Authors have built a three-tier hierarchical architecture (JavaCloudware) which consists of multiple clusters, each for one service provider which has resources for lease along with price policies. As the resource availability changes dynamically due to the usage and release by consumers, this scheduling algorithm works in two phases – a) retrieving resources information from clusters and applying scheduling algorithm, b) committing of the resources. The algorithm also ensures the atomicity of the scheduling that means either all resources reserved during scheduling are released in case of scheduling failure or all resources are committed in case of successful scheduling.

D] Deadline and Cost based Workflow Scheduling in Hybrid Cloud: Nitish Chopra and Sarbjeet Singh in [4] provided the approach of using hybrid cloud in order to minimize the cost of execution workflow applications under the constraint of deadline specified by user for execution. This is level based approach that means all the tasks which are independent on each other are considered at same level and are executed simultaneously. In this approach, firstly resources of private cloud are used for executing the applications since using the private cloud's resources does not charge any cost. If the resources of private cloud are insufficient then the resources of public cloud are used for execution of remaining applications by paying charges according to the usage of resources. Therefore key idea used is that the larger and complex tasks are executed on the private cloud and smaller tasks are executed on public cloud to reduce the overall cost along with considering the deadline of entire application as a constraint. A Sub-deadline for each task is calculated by using the deadline given by user to entire application. This Sub-deadline for each task is compared with the execution time required for that task on the particular virtual machine (VM). If execution time is within that sub-deadline then that task is allowed to execute on that VM otherwise it is given to another VM. If all the VMs within that private cloud failed to complete that task within sub-deadline then task is given to the public cloud. This approach is evaluated by comparing with min-min approach in terms of cost and time.

E] Most-efficient-server-first task-scheduling scheme: Ning Liu, Ziqian Dong and Roberto Rojas-Cessa in [5] presents most-efficient-server-first task-scheduling scheme in order to reduce energy consumption by data center in cloud environment under the task response time constraint. The key idea used for achieving this goal to schedule the tasks among the cloud servers such that it reduces number of active servers used for executing the tasks. For this cloud scheduler arranges the servers in the sequence on the basis of energy efficiency so that it can allocate the tasks to the most efficient server first. All the subsequent tasks are allocated to the same server until that server reach to its saturation point or server's queue get fully filled. After this upcoming tasks are allocated to the next efficient server and so on. The performance of this proposed scheme is compared with random-based task-scheduling scheme by using Matlab simulation. The results from this simulation proved that proposed scheme consumes energy 70 times less than the random-based task-scheduling scheme.

F] A novel approach for task scheduling: R. Vijayalakshmi and Mrs. Soma Prathibha in [6] presented a novel approach for task scheduling based on priority of tasks. The higher priority task is given to the virtual machine which has huge computing capability. Thus this approach tries to manage available resources efficiently which results in lower executing time. Authors have proposed this approach by using the simulation tool- CloudSim, which supports dynamic creation and management of multiple virtual entities (VMs). When cloudlets (tasks) are submitted by user, these cloudlets are prioritized. Also the available VMs are sorted according to the processing power. Cloudlets are mapped on the VMs on the basis of priority of cloudlets and processing power of VMs i.e. higher priority task is mapped to VM with highest processing capability. Executing time required by using this approach is compared with available FCFS (First Come First Served) approach by using different number of cloudlets.

G] Load Balancing with Optimal Cost Scheduling Algorithm: Mrs. Nagamani H. Shahapure and Dr. Jayarekha P in [7] introduced Load Balancing with Optimal Cost Scheduling Algorithm for dynamic workload balancing by considering current state of the resources on cloud. This proposed algorithm fairly allocates the workload to all the servers (VMs) on the cloud which avoids problem of overloading and idling of the cloud resources. Thus it provides the profit to service providers by reducing execution cost and also satisfies user by reducing the execution time. Unlike existing algorithms where separate VM was required for each resource, this proposed algorithm allows to group resources or services into package within a single VM which reduces the execution cost at service provider. The results of this algorithm are compared with existing algorithms (Round Robin and Honey Bee Foraging) with respect to execution time and execution cost at the provider.

H] Optimized Resource Scheduling using Task Grouping: Jignesh Lakhani and Hitesh A. Bheda in [8] proposed task grouping strategy in order to minimize communication overhead between the tasks. When user submits the tasks, Scheduler groups these tasks based on the information provided by Cloud Information Service (CIS) about the available resources. This information consists of availability, processing power, cost of processing of that resource etc. After completion of grouping of all submitted task, dispatcher gives these grouped tasks to corresponding resources as per decided in scheduling. Hence proposed strategy reduces the transition time since group of tasks (or cloudlets) are sent towards particular resource instead of each task is separately sent to that resource. As the groups are made by considering the processing power of the resources, full utilization of a resource capability is achieved.

I] Optimal Cloud Resource Provisioning (OCRP) algorithm: Sivadon Chaisiri, Bu-Sung Lee and Dusit Niyato in [9] provides OCRP algorithm in order to adjust the trade off between two resource provision strategies- resource reservation and on-demand resource allocation. The main goal of the proposed algorithm is to reduce the overall cost of resource provisioning. Resource reservation strategy requires less cost as compared to on-demand resource allocation strategy but is difficult to implement since it requires the prediction about future customer requirement as well as resource cost of the provider. Proposed algorithm is developed by considering the future demand and cost. This proposed Optimal Cloud Resource Provisioning (OCRP) algorithm uses three approaches namely deterministic equivalent formulation, sample-average approximation, and Benders decomposition in order to get the solution. The evaluation of this algorithm is done by two case studies- Two provisioning stage problem (2-PSP) and 12 provisioning stage problem (12-PSP).

J] ThinkAir: Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier and Xinwen Zhang in [10] introduced the framework –ThinkAir, which allows the migration of the applications available on the smart phones to the cloud in order to mitigate the problem of insufficient processing power and energy of the smart phones. This framework achieves the parallelism by creating, using and destroying the VMs when needed. This framework consists of execution environment, application server and profilers. Execution controller within the execution environment decides whether a particular application (or method) should migrate to the cloud or should it be executed on the same smart phone. Application server which is present at the cloud side performs the activities like managing connection with client, receiving and executing the offloaded code from smart phones, returning the results etc. Profilers are used for monitoring and analysis of hardware, software and network parameters. Evaluation of this proposed strategy is done by using micro benchmarks and application benchmarks.

K] Hybrid Haizea-Condor Scheduler (HHCS): Heba Kurdi and Ebtehal T. Alotaibi in [11] proposed Hybrid Haizea-Condor Scheduler which gives the advantages and removes limitations of existing Haizea and Condor scheduler. Haizea scheduler was designed to obtain high performance in terms of execution time reduction by assuming the unlimited pool of resources is available. But this causes the problem of under utilization of resources which is unaffordable for cloud provider. While Condor scheduler removes this problem by formulating and evaluating a logical expression based on user's and cloud service provider's requirements. This results in proper utilization of resources but requires more execution time. So by combining above two, authors proposed the scheduler to enhance the resource utilization without affecting execution time. This is achieved by adapting the matchmaking policy of condor such that modified Haizea satisfies the requirements of both user and provider of cloud service. Evaluation of this HHCS is done by comparing it with Haizea scheduler in terms of average CPU utilization, average turnaround time and average job completion time.

IV. CONCLUSION

Task scheduling plays vital role in managing and sharing cloud resources among the different cloud users. Therefore today task scheduling is main research topic in the area of cloud computing. In this paper, we have highlighted on the various strategies proposed by the researchers regarding to task scheduling in cloud computing. All of the authors have done king-size work related with task scheduling to achieve improvement in performance, load balancing, resource utilization, cost and time optimization, energy conservation and scalability. Above mentioned parameters can be improved by adopting new strategies in task scheduling. For cost and time optimization, we think that the use of optimization techniques like linear programming will give surpassing results.

REFERENCES

[1] AV.Karthick, Dr.E.Ramaraj and R.Kannan, 'An Efficient Tri Queue Job Scheduling using Dynamic Quantum Time for Cloud Environment', International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), IEEE 978-1-4673-6126-2/13, 2013, PP: 871-876.

[2] Shivani Dubey, Vismay Jain and Shailendra Shrivastava, 'An Innovative Approach for Scheduling of Tasks in Cloud Environment', 4th ICCCNT, IEEE-31661, July 4-6, 2013.

[3] Zhi Yang , Changqin Yin and Yan Liu, 'A Cost-based Resource Scheduling Paradigm in Cloud Computing', 12th International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE 978-0-7695-4564-6/11, 2011, PP: 417-422.

[4] Nitish Chopra and Sarbjeet Singh, 'Deadline and Cost based Workflow Scheduling in Hybrid Cloud', International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE 978-1-4673-6217-7/13, 2013, PP: 840-846.

[5] Ning Liu, Ziqian Dong and Roberto Rojas-Cessa, 'Task Scheduling and Server Provisioning for Energy-Efficient Cloud-Computing Data Centers', IEEE 33rd International Conference on Distributed Computing Systems Workshops, IEEE 978-0-7695-5023-7/13, 2013, PP: 226-231.

[6] R.Vijayalakshmi and Mrs. Soma Prathibha, 'A novel approach for task scheduling in cloud', 4th ICCCNT, IEEE-31661, July 4-6, 2013.

[7] Mrs.Nagamani H. Shahapure and Dr. Jayarekha P, 'Load Balancing with Optimal Cost Scheduling Algorithm', INTERNATIONAL CONFERENCE ON COMPUTATION OF POWER, ENERGY, INFORMATION AND COMMUNICATION (ICCPEIC), IEEE 978-1-4 799-3826-1/14, 2014, PP: 24-31.

[8] Jignesh Lakhani and Hitesh A. Bheda, 'An Approach to Optimized Resource Scheduling using Task Grouping in Cloud', International Journal of Advanced Research in Computer Science and Software Engineering, ISSN:2277 128X, Volume 3, Issue 9, September 2013, PP: 594-599.

[9] Sivadon Chaisiri, Bu-Sung Lee and Dusit Niyato, 'Optimization of Resource Provisioning Cost in Cloud Computing', IEEE TRANSACTIONS ON SERVICES COMPUTING, IEEE 1939-1374/12, 2012, PP: 164-177.

[10] Sokol Kosta, Andrius Aucinas, Pan Hui, Richard Mortier and Xinwen Zhang, 'ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading', Proceedings IEEE INFOCOM, IEEE 978-1-4673-0775-8/12, 2012, PP: 945-953.

[11] Heba Kurdi and Ebtehal T. Alotaibi, 'A Hybrid Approach for Scheduling Virtual Machines in Private Clouds', The 9th International Conference on Future Networks and Communications (FNC-2014), ELSEVIER, 2014, PP: 249-256.

[12] Rajesh Piplode and Umesh Kumar Singh, 'An Overview and Study of Security Issues & Challenges in Cloud Computing', International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 2, Issue 9, September 2012, PP: 115-120.

[13] Kalpana Parsi and M. Laharika, 'A Comparative Study of Different Deployment Models in a Cloud', International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 3, Issue 5, May 2013, PP: 512-515.