iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VIII **Month of publication:** August 2015

DOI:

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Level Based Optimized Workflow Scheduling In Cloud Environment

Divya Kataria[1], Dr. Sudesh Kumar[2]
[1]PG Scholar, [2]Head of Department , CSE
BRCM College of Computer Science & Tech., Bahal, India

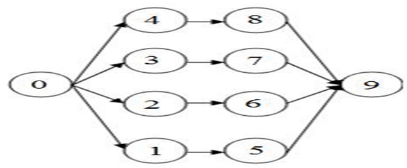
Abstract- Cloud computing is a rapidly growing area offering utility-oriented IT services to the users worldwide over the internet. In cloud, service providers managed and provided resources to users. Software or hardware can be used on rental basis; there is no need to buy them. Most of the cloud applications are modeled as a workflow. In workflows to complete the whole task applications require various sub-tasks to be executed in a particular fashion. Key role in cloud computing systems is managing different tasks. Workflow scheduling is the most important part of cloud computing, because based on the different criteria it decides cost, executiontimeandotherperformances.This research paper describes my proposed algorithm i.e. Level Based Optimized Workflow Scheduling Algorithm and also the comparison of already implemented Algorithm HEFT with the proposed one in terms of Makespan and Cost.
Keywords - Cloud Computing, Workflow Scheduling in Cloud Environment, HEFT

I. INTRODUCTION

The field of distributed and parallel computing has seen technologies rapidly growing from desktop computing, through grid computing, and now to cloud computing delivering. All these technologies focus on delivering computing power to a large number of end-users in a reliable, efficient and scalable manner. More and less, the trend has been to deliver the computing power as a utility, much like how water and electricity is delivered to households these days. Cloud computing delivers infrastructure, platform and software as a service, which are made available as services in a pay-as-you-go model to consumer. These services are referred to as infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS). In [3] Buyya et al. define a cloud as a "type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as a one or more unified computing resources based on service-level agreements". Clouds try to make opportunity to the users all over the world to be able access the services on demand, according to their desired quality of service requirements. So it offers lots of benefits for companies by decreasing management and maintenance costs from leasing IT infrastructure from cloud providers. Besides many applications, cloud computing environment can be used for workflow execution also. Execution of a workflow involves workflow scheduling. Workflow scheduling involves mapping of workflow tasks with available resources in such a way that some predefined criteria is met. Workflow scheduling is well known NP-complete problem [4] and key issue in workflow management system. Moving workflows to Cloud computing enable us to exploit the benefits of cloud for workflow execution. Scheduling can be multi objective also. The multi objective nature of scheduling is more difficult to solve. Many heuristic and meta-heuristic approaches have been proposed by different researchers for workflow scheduling.At present, workflow scheduling algorithms for cloud systems focus on two major parameters viz. Cost and Time. Users may require earlier reliable completion of their workflow tasks within manageable cost .These kinds of requirements make workflow scheduling on clouds more important and complex. In this paper, a level based optimized workflow scheduling algorithm has been proposed. The objective of the proposed scheduling algorithm is to develop workflow schedules which can execute the workflow within manageable cost and makespan as compared to HEFT. The remainder of the paper is organized as follows: Section II describes workflows in Cloud. Section III presents related work. Section IV describes Problem Definition; Section V discusses the proposed approach, Problem evaluation of the algorithm. Experiment and Results are presented in Section VI, Finally, paper is concluded in Section VII.

II. WORKFLOWS IN CLOUD

Executing workflows in clouds is more promising as clouds offer less complex environment than grids. Cloud services like storage, compute and bandwidth are available at much lower costs. Scalability is the prime benefit which is achieved if workflows are moved to cloud. Scalability allow real-time provisioning of resources to meet workflow requirements.Workflows are represented by a Directed Acyclic Graph (DAG) in which each node represents a task and the edge between corresponding nodes represents data dependency between tasks.



Workflow scheduling is the problem of mapping of workflow tasks on suitable resources while satisfying the constraints imposed by the user. Proper workflow scheduling can have significant impact on the performance of the workflow application [5].

III. RELATED WORK

Following are the workflow scheduling algorithms that are important for cloud environment.

PSO-based Heuristic for Scheduling Workflow [7]

This paper proposed a particle swarm optimization based algorithm. In it scheduling of applications considering execution and data transfer cost both. Paper compared the cost savings with existing 'Best Resource Selection' (BRS) algorithm. PSO achieved better distribution of workload on resources with three times cost savings

Workflow Scheduling for SaaS / PaaS [8]

This paper presented an integer linear program formulation. ILP is formulated to schedule SaaS customer's workflows into multiple IaaS providers. It was able to find low-cost solutions, when deadlines were larger the proposed heuristics are effective. Also considered multiple workflows scheduling in the same group of resources and for future work considered fault tolerance mechanisms.

Scheduling Scientific Workflows Elastically [9]

This paper proposed the SHEFT (Scalable HEFT) scheduling algorithm that helps increasing and decreasing the number of resources at runtime. Itprovides facility to resources to scale at runtime outperforms in optimizing workflow execution time.

Optimized Resource Scheduling Algorithm [10]

This paper tells about the optimal use of resources by using virtual machines. It used Improved Genetic Algorithm (IGA). IGA selects optimal VMs byintroducing dividend policy. As compared to traditional GA scheduling method speed of IGA was almost twice and utilization of resources is also larger.

Multiple QOS Constrained Scheduling algorithms [11]

Multiple QOS constrained scheduling is introduced in this paper. It scheduled multiple workflows which were started at different instants. This strategy increased the scheduling success rate considerably and dynamically schedule with minimized execution time and cost.

Deadline and Budget Distribution based Cost-Time Optimization Algorithm [12]

It considered two constraints: deadline and budget. This paper proposed (DBD-CTO) workflow scheduling algorithm. It minimized computation cost before the required deadline for achieving target.

Revised Discrete PSO Algorithm [13]

It scheduled applications that considered data transfer and execution cost both. It compared with the standard PSO and BRS algorithm on makespan, cost savings and cost optimization ratio and achieved better performance and large cost savings on cost optimization and makespan.

Improved cost-based algorithm [14]

In this paper author proposed the approach that is improved cost-based scheduling algorithm. It measured computation performance and resourcecost. It also increased execution/data transfer ratio by combining the tasks. Combining of task is done by analyzing the capability of resource's processing.

HEFT algorithm [15]

This algorithm first calculates average execution time for each task and average communication time between resources of two successive tasks. Then tasks in the workflow are ordered (non-increasing) on a rank function. The task with higher rank value is given higher priority. In the resource selection phase tasks are scheduled in the order of their priorities and each task is assigned to the resource that can complete the task at the earliest time.

IV. PROBLEM DEFINITION

A scheduling system model consists of an application, a target cloud computing environment, and performance criteria for scheduling. We denote an application workflow as a directed acyclic graph (DAG) represented by G=(V,E), where V={T_1, … , T_n} is the set of tasks, and E represents the data dependencies between these tasks. The number of tasks are considered n. G is a n*n matrix of Directed acyclic graph that gained by scientific application. if the amount of $G_{i,j}$ is one, represent task T_i and task $T_j$are dependable and T_i is the parent of T_j, else if the amount of $G_{i,j}$ is zero, shows they aren't dependable. if tasks were dependable, the data of parent task T_i should be transmitted to a child task T_j. a task without any parent is called an entry task and a task without any child is called an exit task. The size of output data of tasks are given in array D as D_i represent size of output data task T_i. We assumed that we want to use some of resources in cloud for scheduling. We have a set of resources R = {1,… , m}. here, we consider the number of resources is m. Each resource has its feature like cost of running tasks or Cost_exe(R_i), cost of incoming data by resources or Cost in(R_i), cost of sending data from resources or Cost_out(R_i), resource availability or Availability(R_i) and resource reliability or Reliability(R_i). The estimated times for computing each task on each resource are given. W is the n*m matrix that represent the estimated times for execute each task on each resource. W_{ij} shows the estimated execution timefor task T_i on resource R_j .The objective function of our workflow scheduling problem is to determine the schedule plan to assign tasks of a given scientific applicationand to run those tasks on cloud resources such that it is done in user's desired time and cost.

For this we take already implemented Algorithm ie. HEFT .

The HEFT algorithm first calculates average execution time for each task and average communication time between resources of two successive tasks. Let time (Ti, r) be the execution time of task Ti on resource r and let Ri be the set of all available resources for processing Ti. The average execution time of a task Ti is defined as :

$$\overline{\omega} = \frac{\sum_{r \in R_i} time(T_i, r)}{|R_i|} \qquad (3.1)$$

Let time (eij , ri, rj)be the data transfer time between resources ri and rj which process the tasks Ti and Tj respectively. Let Ri and Rj be the set of all available resources for processing Ti and Tj respectively. The average transmission time Ti to Tj is defined by:

$$\overline{c_{ij}} = \frac{\sum_{r_i \in R_i, r_j \in R_j} time(e_{ij}, r_i, r_j)}{|R_i|\,|R_j|} \qquad (3.2)$$

Then tasks in the workflow are ordered in HEFT based on a rank function. For a exit task Ti, the rank value is:

$$rank\,(T_i) = \overline{\omega}_i \qquad (3.3)$$

The rank values of other tasks are computed as:

$$rank\,(T_i) = \overline{\omega_i} + \max_{T_j \in succ(T_i)}(\overline{c_{ij}} + rank\,(T_j)) \qquad (3.4)$$

where, succ(Ti) is the set of immediate successors of task Ti. The algorithm then sorts the task by decreasing order of their rank values. The task with higher rank value is given higher priority. In the resource selection phase, tasks are scheduled according to their priorities and each task is assigned to the resource that can finish the task at the earliest time .The makespan and the cost coming by HEFT was efficient but to make it more efficient in terms of make span and cost, We proposed an algorithm that is Level Based Optimized

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Scheduling algorithm.

V. PROPOSED APPROACH

The key idea of our approach is based on HEFT algorithm [25]. We named the proposed approach as a Level Based Optimized Workflow Scheduling Algorithm (LBOWSA). In LBOWSA we try to preserve benefits of HEFT and also apply the data-intensive scheduling features in cloud environment and add few more parameters for calculating the rank of the tasks. LBOWSA has three Phases ,In the level sorting phase, tasks are categorized in levels so that in each level the tasks are independent.In the task prioritization phase, priority is computed and assigned to each task using the attributes Data Received Cost (DRC) from parent, Average Computation Cost (ACCN) of node, Data Transfer Cost (DTC) to child and Average Computation Cost (ACCC) of child node.The rank is computed for each task ni based on DRC, ACCN, DTC and ACCC values.Before presenting the proposed approach in 5-1section , we formulate the parameters and attributes that are used in proposed protocol. And in the section 5-2 we present our protocol.

A. *Problem Formulation*

The algorithm has the three phases.In the level sorting phase, tasks are categorized in levels so that in each level the tasks are independent. In the task prioritization phase, priority is computed and assigned to each task using the attributes Data Received Cost (DRC) from parent, Average Computation Cost (ACCN) of node, Data Transfer Cost (DTC) to child and Average Computation Cost (ACCC) of child node.The rank is computed for each task ni based on its DRC, ACCN, DTC and ACCC values and is given by :

$$\mathbf{rank(ni) = max(DRC(ni)) + ACCN(ni) + max(DTC(ni) + ACCC(ni))}$$

whereDRC of a task ni is the amount of communication costs incurred to transfer the data from immediate parent task to task ni; for an entry node DRC(nentry) = 0.

The ACCN of a task is the average computation cost on all the p processors.

The DTC of a task ni is the amount of communication costs incurred to transfer the data from task ni to its immediate successor tasks; for an exit node DTC(nexit) = 0.

The ACCC of a task ni is the average computation cost of child on all the p processors. At each level, the task with highest rank value receives the highest priority followed by the task with next highest rank value and so on. A tie is broken by selecting the task with a higher ACCN value. The make-span and Cost received by this algorithm is less than or equal to HEFT.

VI. EXPERIMENT AND RESULTS

The conversion of HEFT Algorithm and proposed algorithm from pseudo code to coding in java using Net beans 7.0 is done to conduct our simulation experiment. We have done ten independent executions for all scenarios ie. to gain better results in terms of make span. In this work, we presented a scientific workflow scheduling algorithm on cloud resources that is LBOWSA (Level Based Optimizes Workflow Scheduling Algorithm) is compared with already existing Algorithm ie. HEFT in terms of make span when different DAG models (Montage , Ligo , Cybershake , Epigenomics etc) of 25,100 ,1000 nodes are taken in XML format shown in below table providing better results.

DAG MODEL	HEFT	LBOWSA
EPIGENOMICS1000 Nodes	19049512	19026163
MONTAGE 100 Nodes	53138	53125
INSPIRAL 30 Nodes	318693	313165
MONTAGE 1000 Nodes	541792	541792

VII. CONCLUSION AND FUTURE WORK

In this work, we presented a scientific workflow scheduling algorithm on cloud resourcesnamed as LBOWSA, (Level Based Optimized Workflow Scheduling Algorithm)based on HEFT algorithm. LBOWSA contains three phases : Level sorting , Ranking tasks and selecting resources. We used LBOWSA algorithm to minimize the total cost and Make span of execution of application workflows on cloud environment, based onuser's preference. In evaluation we consider the user is more interested to minimize cost and execution time. We compared the results obtained by our algorithm against HEFT algorithm. We found that LBOWSA achieve improved results in context of make span and Cost.. In addition, LBOWSA considers reliability and availability factor of resources while scheduling. Future scope of this work is to achieve more efficient results in terms of Cost and Make span and also to reflect improved results in all the models of DAG as currently this result can be seen on Montage ,Ligo , Inspiral and Epigenomics.

REFERENCES

[1] Kleinrock L (2003) An Internet Vision: The Invisible Global Infrastructure. Ad-Hoc Networks vol 1, no 1, pp 3-11

[2] Vaquero LM, Rodero-Merino l, Caceres J, Lindner M (2008) A break in the clouds: towards a cloud definition. SIGCOMM Computer Communication Review, vol 39, pp 50-55

[3] Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the5th utility. Future Generation Computer Systems, vol 25, no 6pp 599-616

[4] J. Yu, R. Buyya, (2008) "Workflow scheduling Algorithms for Grid computing", metaheuristics for scheduling in distributing computing environment, Springer, Berlin.

[5] Y. Zhang, A. Mandal, C Koelbel, K. Cooper, (2009) "Combined Fault Tolerance and Scheduling

[6] J.Yu R. Buyya, (2006) "Scheduling Scientific Workflow Applications with Deadline and Budget Constraints using Genetic Algorithms", Scientific Programming Journal, pp. 217-230

[7] S. Pandey, L. Wu, S. Mayura Guru, R. Buyya, "A particleswarm optimization-based heuristic for scheduling workflow applications in cloud computing environments," 24th IEEE international conference on advanced information networking and applications, PP 400-407, 2010.

[8] T. A. L. Genez, L. F. Bittencourt, E. R. M. Madeira, "Workflow scheduling for saas / paas cloud providers considering two SLA levels," IEEE network operations and management symposium (NOMS): mini-conference, pp.906-912, 2012.

[9] C. Lin, S. Lu, "Scheduling scientific workflows elastically for cloud computing," IEEE 4th international conferenceon cloud computing, pp. 246-247, 2011.

[10] H. Zhong, K. Tao, X. Zhang, "An approach to optimized resource scheduling algorithm for open-source cloud systems," Fifth annual china grid conference (IEEE), pp. 124-129, 2010.

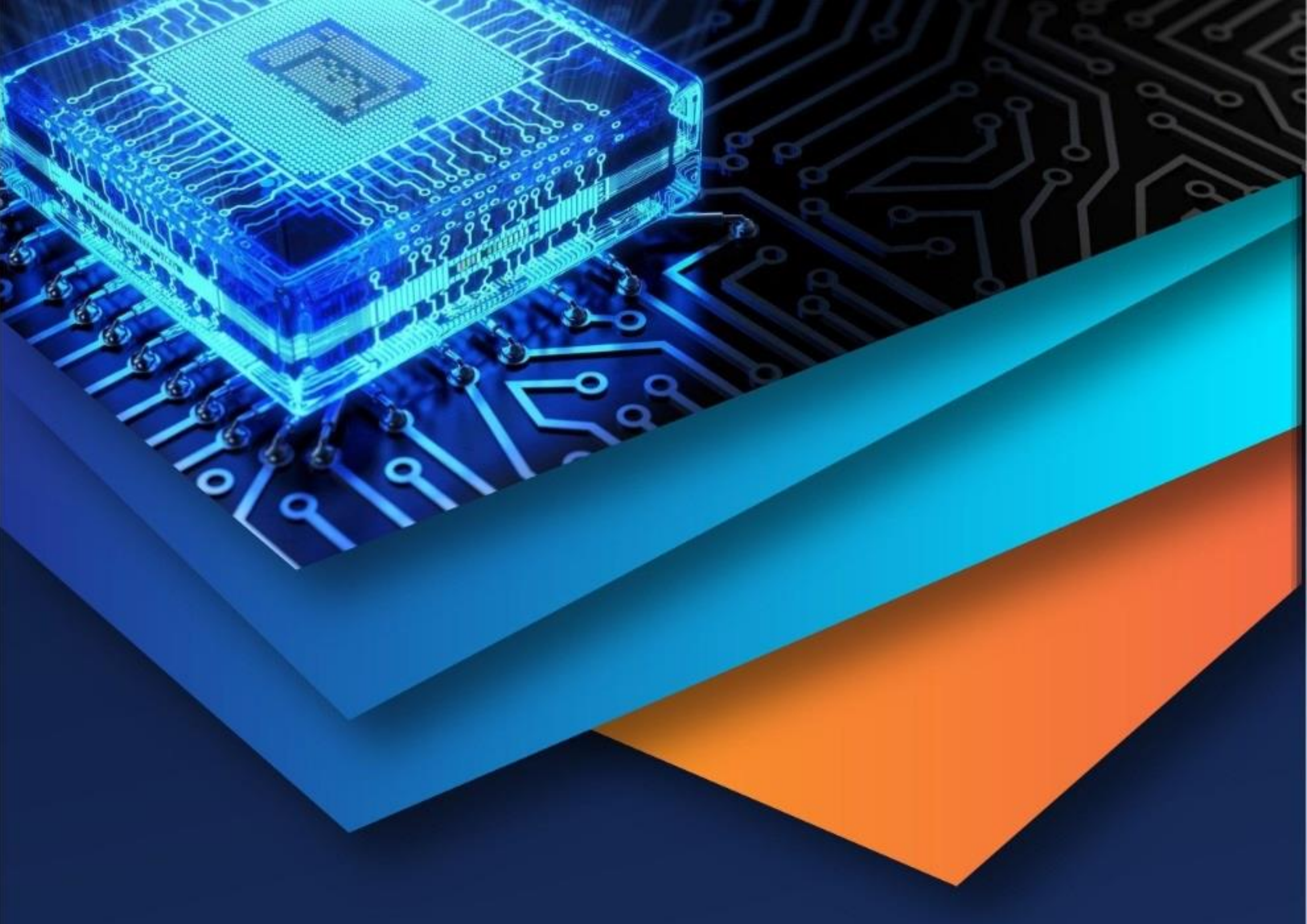
[11] M. Xu, L. Cui, H. Wang, Y. Bi, "A multipleQoSconstrained scheduling strategy of multiple workflows for cloud computing," IEEE international symposium on parallel and distributed processing with applications, pp. 629-634, 2009.

[12] A. Verma, S. Kaushal, "Deadline and budget distribution based cost- time optimization workflow scheduling algorithm for cloud," International conference on recent advances and future trends in information technology (iRAFIT 2012).

[13] Z. Wu, Z. Ni, L. Gu, "A revised discrete particle swarm optimization for cloud workflow scheduling," International conference on computational intelligence and security (CIS), pp. 184-188, 2010.

[14] S. Selvarani, G.S. Sadhasivam, "Improved cost-basedalgorithm for task scheduling in cloud computing,"computational intelligence and computing research, pp.1-5, 2010.

[15] Topcuoglu H, Hariri S, Wu M. Performance effective and low-complexity task scheduling for heterogeneous computing IEEE Transactions on Parallel and Distributed Systems 13(3):260-274



INDEX COPERNICUS

IJSRA JIF

TOGETHER WE REACH THE GOAL

10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY