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:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

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

A Study on Workflow Scheduling Algorithms in Cloud

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

Abstract*—Cloud computing is one of the budding domains that has gained popularity in the recent years. It offers utility-oriented IT services to the users worldwide over the internet. In cloud, service providers manage and provide 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 completion of 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's it decides cost, execution time and other performances. This review paper describes cloud computing, basics of workflows and scheduling, some scheduling algorithms used in workflow management, factors considered by these algorithms, type of algorithm and tool used.
Keywords*— scheduling algorithm; Workflow; Cloud computing; scheduling; DAG

I. INTRODUCTION

Cloud computing provides services, shared resources and common infrastructure on demand through internet. Specific service provider provides these facilities and charge for what a customer used called pay per use [1]. Customer can use storage space, processing capabilities, servers, operating system and application development environments. User can scale up and down the resources in an instant (timely) and on-demand manner in cloud [2]. It also provides flexibility of accessing the resources from different devices. On the cloud, users can manage their applications, develop and deploy with the help of virtualization of resources. Several types of cloud computing environments are there but mainly they classified as Private , Public and Hybrid cloud [8]. Some examples of cloud are as follows

Amazon's Elastic Computing Cloud (EC2) provides resizable compute capacity (CPU cycles) to users [3].

Amazon's Simple Storage Service (S3) provide facility for retrieve and manage large quantity of data, at any instant and from anywhere by internet. This service is provided on rental basis [4].

CRM services provided by salesforce.com which can manage customer information without installing any specialized software [5].

Figure 1 shows 3 -layer architecture or three types of services provided by cloud computing [6].

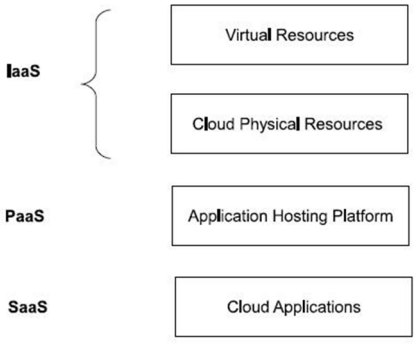


Fig 1: Cloud Computing Services [6]

Infrastructure as a Service (IaaS) - In IaaS the resources are provided by service provider to users without revealing details like location and hardware to users.

Platform as a Service (PaaS) - service provider offer several environments to user for development of applications. User can develop

applications according to the requirements.

Software as a Service (SaaS) – service provider provides software or application on internet and customer used these, with no knowledge of development or maintenance. Example CRM

With efficient scheduling of workflow we can achieve high performance in cloud. It is assumed that each task will be performed on one processor at one time and it cannot be preempted [7]. So, in this paper we try to explain about cloud, workflow and some scheduling algorithms with their parameters. Other sections of the paper are as follows: workflows and its basics discussed in section-II. Section -III is about scheduling. Description and metrics of some scheduling algorithms is in section-IV. Section-V concluded the paper.

II. WORKFLOWS

Besides Cloud, workflows can be executed in Grids also. But due to the complexity of the environment in Grids, 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 a key concern in workflow management systems. 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

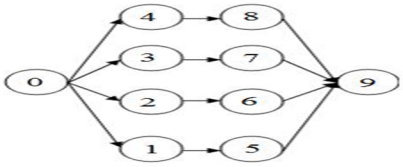


Fig 2: workflow representation in graph

Figure 2 shows the dependencies among different tasks in a workflow graph G. The child tasks 1, 2, 3 and 4 are executed after parent task 0. The child node takes input from the output of parent node. The task 0 acts as entry node and task 9 act as an exit node. After the completion of tasks 5, 6, 7and 8 the task 9 is executed.

In a task graph, an entry task is a task that has no parent and an exit task has no child. Exactly one entry and exactly one exit task is the requirement of task scheduling algorithms. Makespan tells the performance of workflow and it is calculated as ending time minus starting time of a workflow.

Each node and edge in a graph is represented by a particular weight. It is noticed that there is a considerable effect on the schedule when separate techniques are applied for calculating the values of the edges and nodes of the DAG [8].

III. SCHEDULING CONCEPT

The scheduling in terms of cloud means choosing the best suitable resource for task execution and to allocate tasks to machines in order to minimize completion time (makespan). Generally, in scheduling algorithms lists of tasks are constructed by giving priority to every task. Tasks are chosen according to priorities and assigned to a processor which fulfills a predefined objective function [9]. There are two types of scheduling algorithms. First one is static which already has the information of estimation of job execution time, complete structure and mapping of resources before execution. Second is dynamic algorithms which estimate information at the ready state of job before execution [10].

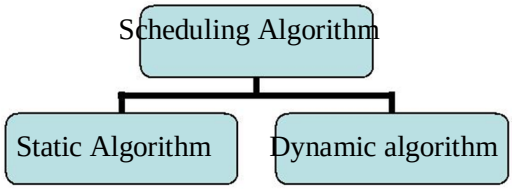


Fig 3: types of scheduling algorithms

New list scheduling algorithm for heterogeneous environment has been proposed [11]. Most works in scheduling were limited to a single workflow application. There are some cases when we required multiple workflow scheduling [12].Multiple workflows are managed online that are submitted at different time by different users. [13]

IV. EXISTING WORKFLOW SCHEDULING ALGORITHMS

The workflow scheduling algorithms that are important for cloud environment are as follows and summarized in table I.

A. A PSO-Based Heuristic for Scheduling Workflow [14]

This paper proposes 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. Better distribution of workload on resources with three times cost savings is achieved by PSO.

B. Workflow Scheduling For SAAS / PAAS [15]

This paper presents 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.

C. Scheduling Scientific Workflows Elastically [16]

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

D. Optimized Resource Scheduling Algorithm [17]

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

E. Multiple Qos Constrained Scheduling Algorithm [18]

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

F. Deadline and Budget Distribution Based Cost-Time Optimization Algorithm [19]

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.

G. Revised Discrete PSO Algorithm [20]

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.

H. Improved Cost-Based Algorithm [21]

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

I. Deadline Constraint Heuristic Based Genetic Algorithm [22]

This paper proposed Heuristic based Genetic Algorithms (HGAs). It scheduled applications in a way to lower the computation cost. Tasks are completed within the timeline. This algorithm had a good performance as compared with Standard Genetic Algorithm (SGA).

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653

Table I: *Summarization of algorithms*

Methods	Factor Considered	Tool Used
Particle Swarm Optimization [14]	Time, Cost Optimization, Resource utilization	Amazon EC2
integer linear program (ILP) formulation [15]	Makespan, Cost, Time	Java Environment
SHEFT workflow scheduling algorithm [16]	Scalability, Execution time	CloudSim
Improved Genetic Algorithm (IGA) [17]	Execution time, Resource Utilization, Speed, CPU utilization,	Eucalyptus
Multiple QoS Constrained Scheduling Algorithm[18]	Time, Cost, Make span, Scheduling success rate	CloudSim
DBD-CTO	Cost,	Java

algorithm [19]	Time	Environment
Particle Swarm Optimization [20]	Makespan and Cost Optimization	Amazon Elastic Compute Cloud
Improved cost based scheduling algorithm [21]	Cost, performance	Cloud Sim
Heuristic based Genetic Algorithms (HGAs) [22]	Execution cost, Execution time, Data transfer cost	Java Environment

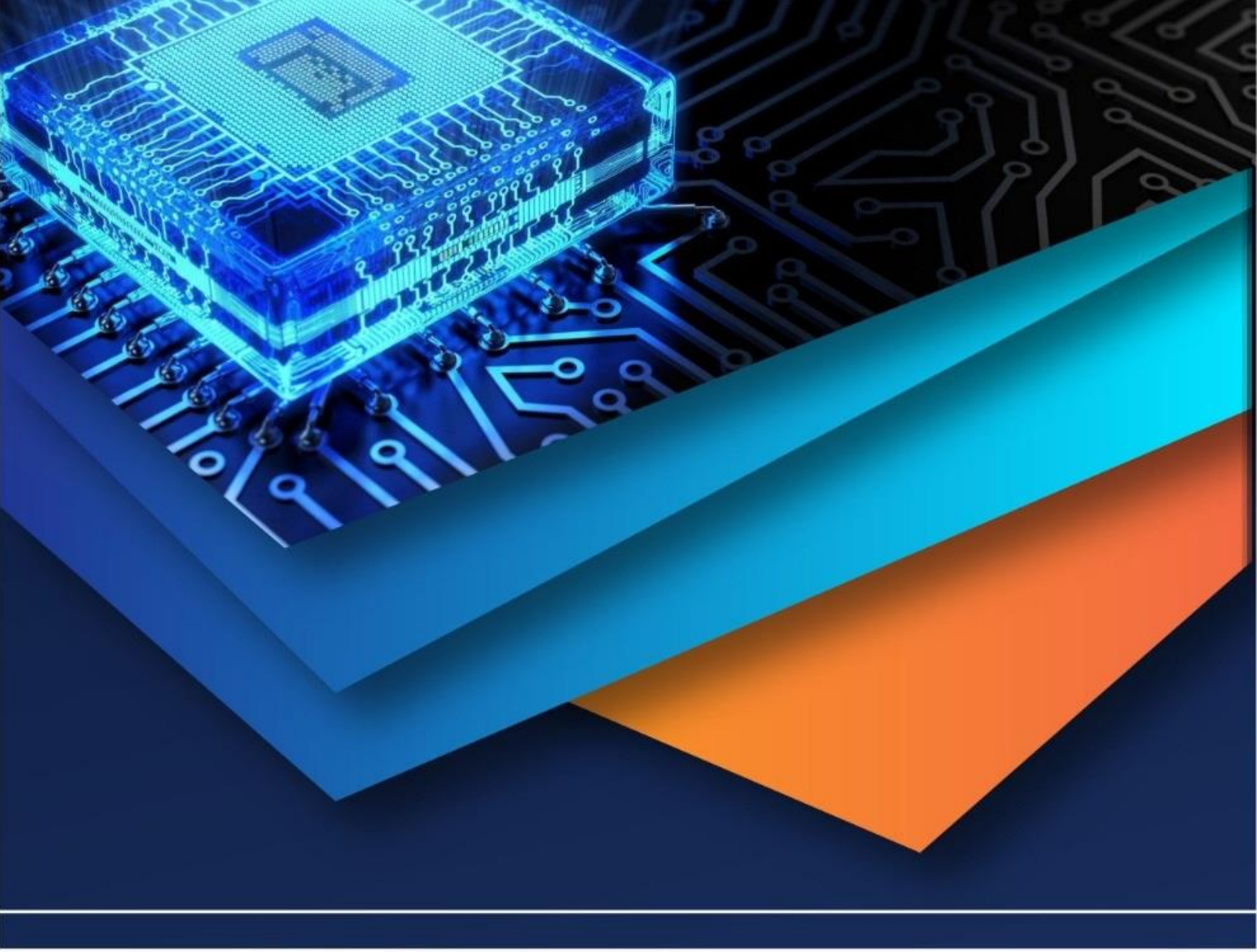
V. CONCLUSION

With the emergence of cloud technology, Workflow systems are designed to support large scale business applications. As resource scheduling and management is complex in cloud, it requires sophisticated tools for analysis of various scheduling algorithm so that it can be applied in real environment. In this paper, we analyzed workflow scheduling algorithms by constructing a table having algorithm used, parameter considered, and development environment. According to analysis the optimized areas in work flow scheduling are resource utilization, cost optimization, make span etc. but areas like reliability, load balancing, reservation conflicts, backup, and fault tolerance require further research.

REFERENCES

[1] S.M. Hashemi, A.Kh. Bardsiri, “Cloud computing vs. grid computing,” ARPN journal of systems and software, vol. 2, No 5, pp. 188-194, May 2012.
[2] H. Alhakami, H. Aldabbas, T. Alwada, "Comparison between cloud and grid computing : review paper," International journal on cloud computing: services and architecture (IJCCSA), vol. 2, No. 4, pp. 1-21, August 2012.
[3] http://aws.amazon.com/ec2.
[4] http://aws.amazon.com/s3.
[5] http://www.salesforce.com/in/crm/what-is-crm.jsp.
[6] M. Shiraz, A. Gani, R. H. Khokhar, R. Buyya, "A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing," IEEE communications surveys & tutorials, vol. 15, no. 3, pp. 1294-1313, 2013.
[7] H. Topcuoglu, S. Hariri, M.Y. Wu, “Performance-effective and low-complexity task scheduling for heterogeneous computing,” IEEE Transactions on parallel and distributed systems, vol. 13, no. 3, pp. 260-274, March 2002.
[8] R. Sakellariou, H. Zhao, “A hybrid heuristic for DAG scheduling on heterogeneous systems,” Proceedings of the
[9] 18th international parallel and distributed processing symposium, 2004.
[10] A. Radulescu, A. Gemund, “Fast and effective task scheduling in heterogeneous systems,” Proceedings of the 9th heterogeneous computing workshop (HCW 2000), pp. 229-238, 2000.
[11] Y. K. Kwok, I. Ahmad, “Dynamic critical-path scheduling: an effective technique for allocating task graphs to multiprocessors,” IEEE transactions on parallel and distributed systems, vol. 7, no. 5, pp. 506-521, May 1996.

[12] G.C. Sih, E.A. Lee, "A compile-time scheduling heuristic for interconnection-constrained heterogeneous processor architectures," IEEE transactions on parallel and distributed systems, vol. 4, no. 2, pp. 175-187, February 1993.
[13] H. Zhao, R. Sakellarious, "Scheduling multiple DAGs onto heterogeneous systems," IEEE 20th international parallel and distributed processing symposium,2006.
[14] Z. Yu, W. Shi, "A planner-guided scheduling strategy for multiple workflow applications," international conference on parallel processing - IEEE workshop, pp. 1-8, 2008.
[15] S. Pandey, L. Wu, S. Mayura Guru, R. Buyya, "A particle swarm 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.
[16] 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.
[17] C. Lin, S. Lu, "Scheduling scientific workflows elastically for cloud computing," IEEE 4th international conference on cloud computing, pp. 246-247, 2011.
[18] 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.
[19] M. Xu, L. Cui, H. Wang, Y. Bi, "A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing," IEEE international symposium on parallel and distributed processing with applications, pp. 629-634, 2009.
[20] 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).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY