A Journey towards Workflow Scheduling of Cloud Computing

Anil Kumar Gupta*, Shashank Shukla**
Research Scholar
Department of CSE*, IT**
GCET Greater Noida UP (INDIA)

Sandeep Saxena
Research Scholar
Department of CSE
NIT Durgapur INDIA

Sanjay Khakhil
Asstt. Prof.
Department of IT
GCET Greater Noida UP (INDIA)

ABSTRACT
Cloud computing is a type of grid computing which is a form of distributed computing and distributed computing is a special type of parallel computing. Presently a lot of services are growing under the single umbrella that is known as *cloud computing.* Cloud computing gain popularity in the several area due its property of everything-as -a-service(XaaS), includes SaaS, PaaS and IaaS. Many problems have been arising when we go for implementation development. Workflow scheduling and appropriate allocation of resources is one of among problems that will decrease the Quality of Service (QoS) of cloud computing. There are many algorithms to automate the workflows in a way to satisfy the Quality of service (QoS) of the user. This paper is the survey of some workflow scheduling algorithms that have been proposed for cloud computing.

Keywords
Cloud computing, grid computing, distributed computing, parallel computing, workflow scheduling Virtualization, and QoS.

1. INTRODUCTION
Cloud computing has grown-up prominence in recent years, with is unbelievable on demand capacity management model. Many cloud providers are now active in the market, providing a rich offering several kind of services to the user. The cloud technology has become mainstream in enterprise data centers, where private, public and hybrid cloud architectures [1] are progressively more adopted. These number of cloud services are manage, develop and deliver to the user through virtualization technology. It will payable by cloud resource users only when user's requirement is fulfilled that mean user pay per use basis. It is referred as QoS that denotes the levels of reliability, performance, security and availability offered by the platform or infrastructure of the cloud. QoS is a major factor for cloud users, who expect from providers to deliver the advertised worth characteristics[2].

The rest of paper organizes as follows: In section-2 we discuss the workflow scheduling follows by its nature and parameters that are considered during the improvement of quality of service of cloud computing. Section-3 is the literature survey on various workflow scheduling algorithms. In section-4, we present the analytical table of the various survey workflow algorithm and finally we draw our conclusion and future work in section-5.

2. WORKFLOW SCHEDULING
Workflow Scheduling compacts with the credentials of suitable resources and allocation of the tasks on those resources. In workflow scheduling, bigger task are divided into different-different small co-tasks these co-task or sub-tasks are allocated to resources in such a way that achieve some pre-define objective. There is various problems in bioinformatics, astronomy and business enterprise in which set of sub-tasks are executed in particular sequence and ordered to carry out the bigger task[6]. In general, a workflow application requires series of steps to be executed in a particular fashion. These steps have the parent-child relationship[7]. The parent task is associated to child-task according to some rules and must be executed before its child-task.

A workflow is represented by Directed Acyclic Graph(DAG). The number of tasks is denoted by vertex(V) and data dependencies among tasks are denoted by an edge(E).

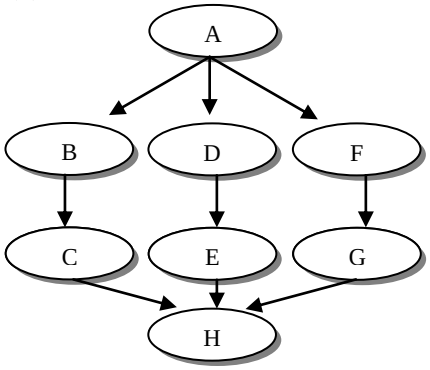

Figure-1 Workflow represented in the form of a graph

2.1 Nature of Workflow Algorithm: The workflow scheduling algorithms can be two type:

2.1.1 Heuristic in Nature: The heuristic algorithms are priority based and mainly problem centric. The maker can use his individual knowledge to allocate priority for workflow applications and cloud resources.

2.1.2 Meta Heuristic in Nature: Meta-heuristic scheduling algorithms are the ones that do not need the human interface and provide the over-all solution to workflow applications.

2.2 Workflow Scheduling Parameters: Workflow scheduling parameters which may consider in different algorithms as per developer requirements.

•**Execution Time(ET):** It refers to the time taken by CPU to execute the task or job.

• **Response Time(RT):** The response time is the time at which system start responding for the submitted task.

RT= Service Time + Wait Time

• **Makespan (M-s):** Time difference between the submission of request and completion time of the last task of workflow is known as makespan time.

M-s= Completion Time - Start

• **Energy Consumption:** Energy Consumption is referred as consumption of power or energy by resources during the service.

• **Throughput:** The rate of production or how much task are completed in the particular unit of time is known as the throughput of the system.

• **Scalability:** The ability of system grow itself according to increasing the demand or increasing the data.

• **Resource Utilization (RU):** Keep the resources as busy as possible. It covers the ranges from 0 to 100% but in practice 100% is not possible.

• **Load Balancing:** It is a process that keep the resources or servers in balance by pooling or pushing the load from one resource to another.

• **Fault Tolerance:** To give the surety of availability and reliability of critical services is known as fault tolerance.

• **Reliability:** Reliability is define as the trust value of the resources, services and Cloud service provider (CSP). It defines user must be received continue service without failures and quality must be meet the SLA agreement.

3. LITERATURE SURVEY ON WORKFLOW SCHEDULING ALGORITHMS

In the current scenario, many scientist/ researcher are working for improving the proper utilization of cloud resources. This problem is similar with some other environment like the grid or distributed. In the cloud computing environment, we can improve QoS with the help of workflow scheduling algorithms. We studied various algorithm and some of them will be discussed here.

3.1 Cost-Based Scheduling Of Scientific Workflow Applications On Utility Grids[8] is based on Nimrog-G[9] and A Market-Based Workflow Management System[10] which schedules independent tasks and assign deadline of individual task respectively. In this algorithm author solved the task independency problem by considering the tasks with certain dependencies. Through this, author minimize the execution cost while meeting the deadline constrain and solve optimally the scheduling problem for sequential tasks by modeling the branch partition as a Markov Decision Process(MDP)[3] that cut down the execution cost and improve the QoS in Utility Grids environment.

3.2 A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing(MQMW)[11] optimize multiple parameters(make-span, the cost of workflow, the success rate of workflow scheduling) of QoS. It removes the problem of previous algorithms such as cost-based scheduling algorithm introduce by Jia Yu which was used for multiple workflows but the relationship between workflows are not consider[12]. In other paper, proposed by Ke Liu et al. for intensive workflow (multiple instances of single workflow) but in multiple workflows means multiple structure[13].

3.3 Scheduling Service Workflows for Cost Optimization in Hybrid Clouds[14] This algorithm work in Hybrid cloud scenario. When cloud user demands new resources from the public cloud to fulfill their need at that time public cloud resources should be schedule in a proper way. The author using PCH[4] strategy that separate the public cloud resources and private cloud resources are requested by cloud user in order to minimize the cost and meet the deadline. After the comparison of proposed work give the better result in compare to the individual only private cloud and only public cloud environment.

3.4 Scheduling Scientific Workflows Elastically For Cloud Computing [15] in this paper Cui Lin et al, optimizing the execution time as well as resources to scale elastically at run time. This was done by formulize the computing environment and scientific workflow.

3.5 Scheduling Technique of Data-Intensive Application Workflows In Cloud Computing[16] The main goal of this paper is the selection of suitable hosts for accessing resources and creating a virtual machine to execute applications for more effective execution. The selection of suitable data centers that have minimum transfer time to access replica among other data centers and create VM for this data center. This procedure undersized the access cost that minimize the broadcast time of input and output data and make more efficient execution.

3.6 Trust-Based And QoS Demand Clustering Analysis Customizable Cloud Workflow Scheduling Strategies[17]. This proposed algorithm is based on 'demand at affordable price' in which all kinds of resources provided to end user on their demand at affordable price [18]. In this paper, defining the key region and key region reliability for the cloud resources, considering the cost, completion time and security as well as set up the reliability level as TOP LEVEL(TL), BOTTOM LEVEL(BL) through this whole process, author able to improve success rate of workflow in restricted period of time.

3.7 A Truthful Dynamic Workflow Scheduling Mechanism For Commercial Multicloud Environments[19]. It is multicloud environment workflow scheduling algorithm which has objected to minimize the completion time and monetary cost while meeting the deadline for delivering the result. Algorithm based on game theory and mathematical approach to analyze the decision of agents in problem for trust accuracy of the cloud resource[20,21].

3.8 Energy Efficient Workflow Job Scheduling for Green Cloud [22] was presenting a realistic model that consume less energy and minimize the CO2 emission. Each task of scientific workflow run without affecting the performance based on DVFS strategy. Resource utilization is maximized in this algorithm, through VM allocation strategy that enhance the system performance. When this algorithm was simulated then its save the energy consumption up to 30% and maximize the resource utilization 25% and reduce emission of CO2.

3.9Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds[23]. This paper proposed a new type of algorithm which is source provisioning and schedule the scientific workflow on infrastructure as-a-service (IaaS). Cloud computing through presenting a meta-heuristic optimization technique, particle swarm optimization(PSO) and aims to reduce the overall workflow execution cost while meeting deadline constrain

3.10 Cluster based Scheduling of Workflow Applications in Cloud[24] in this paper author propose a new type of workflow algorithm that is based on the cluster of the task. In this algorithm reducing the overall makespan is the main objective of the author. Cluster are calculated by obtaining the execution time, based on it author put the task in some cluster.

3.11 An Approach to Workflow Scheduling using Priority in Cloud Computing Environment[5] workflow scheduling algorithm consider the multiple criteria (cost, priority) to optimize the service utilization. Proposed algorithm gives better result over GA algorithm.

4. ANALYSIS OF WORKFLOW ALGORITHMS

As we discussed in previous section each scheduling algorithms consider one or multiple parameters and on the basis of this, scheduling algorithms are improved. Some algorithms are focus on minimizing the execution costs or time whereas some improve the success rate of workflow scheduling, reducing the CO2 emission . At the same time workflow scheduling algorithms try to achieve the Service Level Agreement and maintaining the quality of service (QoS). Each workflow has the different structure that fit in types of cloud architecture and demand scheduling strategies according to it. In this table we compare survey algorithm on the basis of different attributes.

Table 1: A brief description and comparison among various workflow scheduling algorithms

Name of Algorithm	Year	Nature of Algorithm	Objective	Environment	Criteria	Simulation tool
Cost-Based Scheduling of Scientific Workflow Applications On Utility Grids[8]	2005	Heuristic	Execution costs	Cloud Computing	Deadline, Task Dependencies	GridSim
A Multiple QoS Constrained Schedulin -g Strategy of Multiple Workflows for Cloud Computing (MQMW) [11]	2009	Meta-Heuristic	Multi workflows with multiple parameter (make-span, the cost of workflow, the success rate of workflow)	Cloud Computing	Task with minimum surplus time/cost & minimum covariance	Real World
Scheduling Service Workflows for Cost Optimization in Hybrid Clouds[14]	2010	Heuristic	Costs	Hybrid Cloud	Deadline	Hybrid System
Scheduling Scientific Workflows Elastically For Cloud Computing [15]	2011	Heuristic	Execution Time, Elasticity	Cloud Computing	Priority	General Cloud Environment
Scheduling Technique of Data-Intensive Application Workflows In Cloud Computing[16]	2012	Meta-Heuristic	Broadcast time, execution cost, access cost	Cloud Computing	Minimum Network Delay	NA
Trust-Based And QoS Demand Clustering Analysis Customizable Cloud Workflow Scheduling Strategies [17]	2012	Heuristic	Makespan time, Execution success Rate, User Satisfaction	Cloud computing	Trust level	CouldSim
A Truthful Dynamic Workflow Scheduling Mechanism For commercial multicloud Environments[19]	2013	Heuristic	Completion time, Monetary cost,	Cloud Computing	Reliability, Trust value	Real world
Energy Efficient Workflow Job Scheduling for Green Cloud[22]	2013	Heuristic	Energy Consumptions, CO2 emission Minimize VM overhead	Cloud Computing	Dynamic Voltage & Frequency Scaling (DVFS)	CloudSim
Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds[23]	2014	Meta-Heuristic	Execution cost	Cloud computing	Deadline	CloudSim
Cluster-based Schedul -ing of Workflow Applications in Cloud [24]	2014	Heuristic	MakeSpan	Cloud Computing	Clusters	CloudSim
An Approach to Workflow Scheduling using Priority in Cloud Computing Environment [5]	2015	Heuristic	Service Utilization, cost	Cloud Computing	Priority	Real World

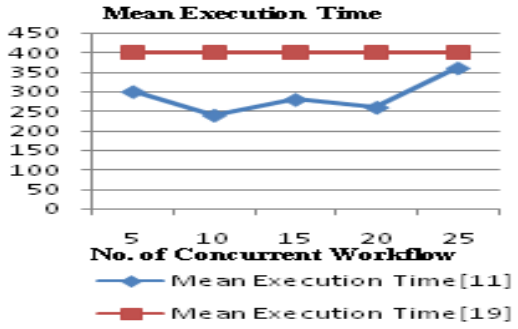

Figure-2 Mean Execution Time Graph

In mean execution time graph when number of concurrent workflows was scheduled by MQMW[11] with multiple QoS parameters and A Truthful Dynamic Workflow Scheduling Mechanism [19] for multiple workflow or multicloud environment than MQMW[11] give better result.

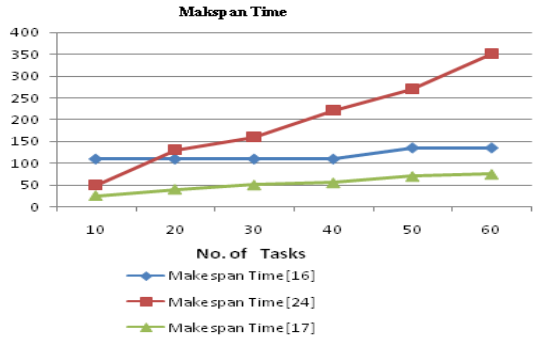

Figure-3 Makespan Time Graph

In makespan time graph, Trust-Based [17] algorithm give better performance by reducing the makespan time over the Intensive-Based[16] and Cluster-Based[24] workflow algorithm.

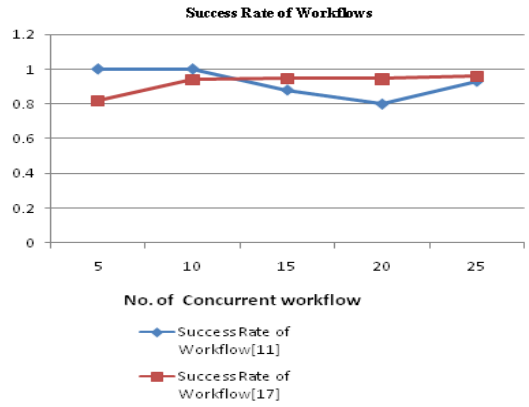

Figure-4 Success Rate of Workflows

In success rate of workflow graph MQMW[11] give the Avg. success rate 92% whereas Trust Based [17] give 92.3% that is better success rate in compare to MQMW.

Note: All graphical data are taken from respective paper and scaled.

5. CONCLUSION AND FUTURE WORK

On the basis of above analysis we had seen that algorithms are change according to user priority parameters. Parameters, which involves to enhance the Quality of Service (QoS) in the area of workflow scheduling such as execution cost, time, elasticity, access time, make-span time, the success rate of work flow, reducing the emission of CO_2, energy efficiency etc.

But still cloud computing is not widely used due to lack of reliability, availability and fault tolerance etc. because user data are stored and used from third party and servers are installed on worldwide locations. Sometimes user data did not available due to large amount of data request /network traffic with effect of that user trust has been decrease rapidly. Sometimes resources are allocate but user task did not complete and reliability had affect. Sometimes reliability did affect due to fault in cloud environment.

6. REFERENCES

[1] Quality-of-service in cloud computing: modeling techniques and their applications Danilo Ardagna, Giuliano Casale, Michele Ciavotta, JuaJuann F Pérezand Weikun Wang in 2014.

[2] Ardagna D, Panicucci B, Trubian M, Zhang L (2012) Energy-aware autonomic resource allocation in multitier virtualized environments. IEEE Trans Serv Comput 5(1):2-19

[3] A. Sulistio and R. Buyya, "A Grid Simulation Infrastructure Supporting Advance Reservation", In *16*th International Conference on Parallel and Distributed Computing and Systems (PDCS 2004), November 911, 2004, MIT Cambridge, Boston, USA.

[4] L. F. Bittencourt and E. R. M. Madeira, "A performance-oriented adaptive scheduler for dependent tasks on grids," Concurrency and Computation: Practice and Experience, vol. 20, no. 9, pp. 1029–1049, 2008.

[5] An Approach to Workflow Scheduling using Priority in Cloud Computing Environment P. Kumar, V. Anandarangan and A. Reshma in International Journal of Computer Applications (0975 – 8887) Volume 109 – No. 11, January 2015

[6] Advance Reservation of Resources in Workflow System, Lalit et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.12, December- 2014, pg. 140-145

[7] A Survey of Workflow Scheduling Algorithms and Research Issues in International Journal of Computer Applications (0975 – 8887) Volume 74– No.15, July 2013.

[8] Jia Yu, Rajkumar Buyyaand Chen Khong Tham in Proceedings of the First International Conference on e-Science and Grid Computing (e-Science'05) 0-7695-2448-6/05 $20.00 © 2005 IEEE.

[9] R. Buyya, J. Giddy, and D. Abramson, "An Evaluation of Economy based Resource Trading and Scheduling on Computational Power Grids for Parameter Sweep Applications", In 2nd Workshop on

Active Middleware Services (AMS 2000), Kluwer Academic Press, August 1, 2000, Pittsburgh, USA.

[10] A. Geppert, M. Kradolfer, and D. Tombros, "Market-based Workflow Management", International Journal of Cooperative Information Systems, World Scientific Publishing Co., NJ, USA, 1998.

[11] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for CloudComputing in 2009 IEEE International Symposium on Parallel and Distributed Processing with Applications

[12] Jia Yu, Rajkumar Buyya and Chen Khong Tham, "Cost-basedScheduling of Scientific Workflow Applications on Utility Grids", In1st IEEE International Conference on e-Science and GridComputing, Melbourne, Australia, Dec. 5-8, 2005.

[13] Ke Liu, Jinjun Chen, Yun Yang and Hai Jin, "A throughputmaximization strategy for scheduling transaction-intensiveworkflows on SwinDeW-G", Concurrency and Computation:Practice and Experience, Wiley, 20(15):1807-1820, Oct. 2008.

[14] Scheduling Service Workflows for Cost Optimization in Hybrid Clouds by Luiz F. Bittencourt, Carlos R. Senna and Edmundo R. M. Madeira in 978-1-4244-8909-1/$26.00 2010 IEEE.

[15] Scheduling Scientific Workflows Elastically for Cloud Computing by Cui Lin and Shiyong Lu in 2011 IEEE 4th International Conference on Cloud Computing.

[16] Jignesh Lakhani and Hitesh Bheda,"Scheduling Technique of Data Intensive Application Workflows In Cloud Computing"Nirma University International Conference On Engineering, Nuicone-2012, 06-08december, 2012.

[17] Trust-based and QoS Demand Clustering Analysis Customizable Cloud Workflow Scheduling Strategies Wenjuan Li Qifei Zhang Jiyi Wu and Jing Li in 978-0-7695-4844-9/12 $26.00 © 2012 IEEE DOI 10.1109/ClusterW.2012.21 in 2012 IEEE International Conference on Cluster Computing Workshops.

[18] Peng Liu.: Cloud Computing, Second Editon [M]. Beijing: Publishing House of Electronics Industry, 2011.

[19] Hamid MohammadiFard, RaduProdan, and Thomas Fahringer, "A Truthful Dynamic Workflow Scheduling Mechanism for Commercial Multicloud Environments," in IEEE Transactions on Parallel and Distributed Systems, VOL. 24, NO. 6, JUNE 2013.

[20] A. Archer and E. Tardos, "Truthful Mechanisms for One-Parameter Agents," Proc. 42nd IEEE Symp. Foundations of Computer Science, pp. 482-491, 2001.

[21] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani," AlgorithmicGame Theory,"Cambridge Univ. Press, 2007.

[22] Fei Cao and Michelle M. Zhu ,"Energy Efficient Workflow Job Scheduling for Green Cloud" in IEEE 27th International Symposium on Parallel & Distributed Processing Workshops and PhD Forum, 2013.

[23] Maria Alejandra Rodriguez and Rajkumar Buyya, "Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds" in IEEE Transactions On Cloud Computing, Vol.2, No.2, April-June 2014 2168-7161 2

[24] Cluster based Scheduling of Workflow Applications in Cloud Ardra V A and Sindhu S in International Journal of Computer Applications (0975 – 8887) Advanced Computing and Communication Techniques for High Performance Applications (ICACCTHPA-2014)