A Survey of Various Workflow Scheduling Algorithms in Cloud Environment

Anju Bala, Assistant Professor
Computer Sci. & Engg.Deptt.
Thapar University
Patiala-147004

Dr.Inderveer Chana, Associate Professor
Computer Sci. & Engg.Deptt.
Thapar University
Patiala-147004

ABSTRACT
Cloud computing is a new benchmark towards enterprise application development that can effectively facilitate the execution of workflows in business process management system. The workflow technology can manage the business processes efficiently satisfying the requirements of modern enterprises. Moving workflows to a cloud computing environment enables the utilization of various cloud services to facilitate workflow execution. Workflow scheduling is one of the key issues in the management of workflow execution. In this paper we have surveyed different types of workflow scheduling algorithms and tabulated their various parameters along with tools, scheduling factors and so on. Existing workflow scheduling algorithms does not consider reliability and availability. Therefore there is a need to implement a workflow scheduling algorithm that can improve the availability and reliability in cloud environment.

General Terms
Cloud Computing, Workflow Scheduling.

Keywords
Workflow Management System, Workflow Scheduling, Cloud Computing.

1. INTRODUCTION
With the promotion of the world's leading companies, cloud computing is attracting more and more attention for providing a flexible, on demand computing infrastructure for a number of applications [13]. The concept of cloud computing continues to spread widely, as it has been accepted recently. Cloud computing has many unique advantages which can be utilized to facilitate workflow execution.

Workflow is concerned with the automation of procedures where documents, information or tasks are passed between participants according to a defined set of rules. A workflow enables the structuring of applications in a directed acyclic graph form [15], where each node represents the constituent task and edges represent inter task dependencies of the applications [16].A single workflow generally consists of a set of tasks each of which may communicate with (depends on or is depended upon) another task in the workflow. Multiple workflows have multiple instances of workflow. Workflows are supported by Workflow Management Systems.

3. WORKFLOWS IN CLOUD COMPUTING
Moving workflows to a cloud computing environment enables the utilization of various cloud services to facilitate workflow execution. Typical examples of workflows include online banking, insurance claim processing and many other e-business and e-government scenarios. Here we take the online banking example to illustrate the characteristics of workflows. Online banking is a service where people can perform financial transactions such as account to account transfers, deposit of cash and withdrawal of cash and so on. In our example, the security is provided by giving a Personal Identification Number (PIN) to every customer to access

The main purpose of a workflow management system (WfMS) is to support the definition, execution, registration and control of business processes. At run time, a workflow enactment engine manages the execution of the workflow by utilizing middleware. Three major components in a workflow enactment engine are the workflow scheduling, data movement and fault management. Workflow scheduling discovers resources and allocates tasks on suitable resources. Data movement is used for communication of data resources and fault management used for failure handling during execution. Workflow scheduling plays a vital role in the workflow management. Proper scheduling can have significant impact on the performance of the system [17].For proper scheduling in workflows various scheduling algorithms are needs to be discussed.

The objective of this paper is to be focus on various workflow scheduling algorithms. The rest of the paper is organized as follows. Section II presents the need for implementing workflows in cloud. Section III presents workflows in cloud computing. Section IV presents various existing workflow scheduling algorithms along with table and section V concludes the paper with a summary of our contributions.

2. NEED FOR IMPLEMENTING WORKFLOWS IN CLOUD
Initially, workflows were being implemented in grids. Due to the reduced performance faced in grids, now there is a need to implement workflows in cloud. The primary benefit of moving to Clouds is application scalability. Unlike Grids, scalability of Cloud resources allows real-time provisioning of resources to meet application requirements. This enables workflow management systems to readily meet Quality of-Service (QoS) requirements of applications, as opposed to the traditional approach that required advance reservation of resources in global multi-user Grid environments. Cloud services like compute, storage and bandwidth resources are available at substantially lower costs. Workflow applications often require very complex execution environments .These environments are difficult to create on grid resources. In addition, each grid site has a different configuration, which results in extra effort each time an application needs to be ported to a new site. Virtual machines allow the application developer to create a fully customized, portable execution environment configured specifically for their application.

his/her account. The system checks whether the PIN entered by the user is valid or not. It also takes care of the remaining balance in the customer's account. If the amount entered by the user to withdraw the cash exceeds the current balance, the system gives a warning message else it allows the user to withdraw the cash. Figure 1 shows a graphical overview of the functionality for an online banking application in terms of actors, their goals (represented as use cases) and dependencies between the use cases.

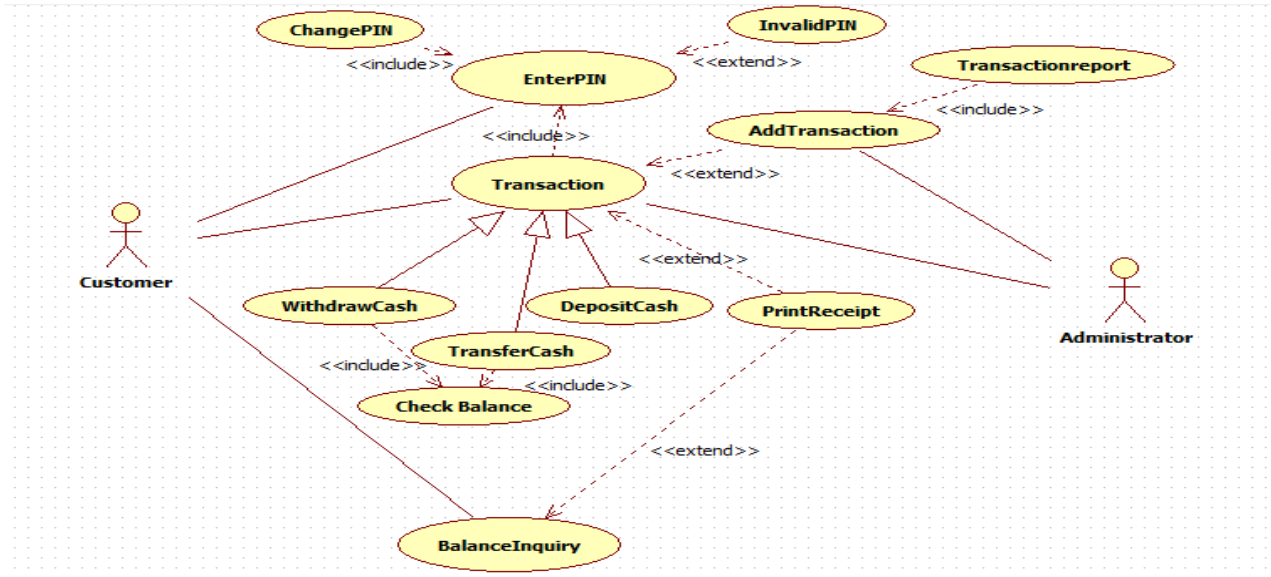

Figure 1: Use Case Diagram for online banking system.

4. EXISTING WORKFLOW SCHEDULING ALGORITHMS

The Following workflow scheduling algorithms are currently prevalent in clouds and these algorithms have been summarized in table1.

4.1 Optimized Resource Scheduling algorithm: Hai Zhong1, Kun Tao1, Xuejie Zhang [14] proposed an optimized scheduling algorithm to achieve the optimization or sub-optimization for cloud scheduling. In this algorithm an Improved Genetic Algorithm (IGA) is used for the automated scheduling policy. It is used to increase the utilization rate of resources and speed.

4.2 Improved Cost-Based Algorithm for Task Scheduling: Mrs.S.Selvarani, Dr.G.Sudha Sadhasivam [2] proposed an improved cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud. This scheduling algorithm measures both resource cost and computation performance, it also Improves the computation/communication ratio.

4.3 Innovative transaction intensive cost-constraint scheduling algorithm: Yun Yang, Ke Liu, Jinjun Chen [3] proposed a scheduling algorithm which takes cost and time. The simulation has demonstrated that this algorithm can achieve lower cost than others while meeting the user designated deadline.

4.4ACompromised-Time-Cost Scheduling Algorithm:Ke Liu, Hai Jin, Jinjun Chen, Xiao Liu, Dong Yuan, Yun Yang [4] presented a novel compromised-time-cost scheduling algorithm which considers the characteristics of cloud computing to accommodate instance-intensive cost-constrained workflows by compromising execution time and cost with user input enabled on the fly.

4.5 A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications: Suraj Pandey, LinlinWu, Siddeswara Mayura Guru, Rajkumar Buyya [5] presented a particle swarm optimization (PSO) based heuristic to schedule applications to cloud resources that takes into account both computation cost and data transmission cost. It is used for workflow application by varying its computation and communication costs. The experimental results show s that PSO can achieve cost savings and good distribution of workload onto resources.

4.6Scalable-Heterogeneous-Earliest-Finish-Time Algorithm (SHEFT): Cui Lin, Shiyong Lu [6] proposed an SHEFT workflow scheduling algorithm to schedule a workflow elastically on a Cloud computing environment. The experimental results show that SHEFT not only outperforms several representative workflow scheduling algorithms in optimizing workflow execution time, but also enables resources to scale elastically at runtime.

4.7Market-Oriented-Hierarchical Scheduling:

Zhangjun Wu, Xiao Liu, Zhiwei Ni, Dong Yuan, Yun Yang [7] proposed a market-oriented hierarchical scheduling strategy which consists of a service-level scheduling and a task-level scheduling. The service-level scheduling deals with the Task-to-Service assignment and the task-level scheduling deals with the optimization of the Task-to-VM assignment in local cloud data centers.

4.8 Multiple QoS Constrained Scheduling Strategy of Multi-Workflows (MQMW):

Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi [8] worked on multiple workflows and multiple QoS.They has a strategy implemented for multiple workflow management system with multiple QoS. The scheduling access rate is increased by using this strategy. This strategy minimizes the make span and cost of workflows for cloud computing platform.

4.9 Optimal Workflow based Scheduling (OWS)algorithm:P.Varalakshmi, Aravindh Ramaswamy

[9] proposed OWS algorithm for scheduling workflows in a cloud environment. The scheduling algorithm finds a solution that meets all user preferred QoS constraints. With this algorithm, a significant improvement in CPU utilization is achived.

4.10Resource-Aware-Scheduling algorithm (RASA):

Saeed Parsa and Reza Entezari-Maleki [10] proposed a new task scheduling algorithm RASA. It is composed of two traditional scheduling algorithms; Max-min and Min-min. RASA uses the advantages of Max-min and Min-min algorithms and covers their disadvantages. The experimental results show that RASA is outperforms the existing scheduling algorithms in large scale distributed systems [14].

4.11Heterogeneous-Earliest-Finish-Time algorithm (HEFT):

Topcuoglu et. al, [11] presented the HEFT algorithm. 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.

Table 1.Existing Workflow Scheduling Algorithms

Scheduling Algorithm	Scheduling Method	Scheduling Parameters	Scheduling factors	Findings	Environment	Tools
Optimized-Resource Scheduling Algorithm[14]	Multiple instances	Speed, Resource Utilization	Request allocation problem	1. Speed of the IGA is almost twice the traditional GA 2.The utilization rate of resources is high	Cloud Environment	Eucalyptus
Improved cost-based algorithm for task scheduling [2]	Batch Mode	Cost, performance	Unscheduled task groups	1.Measures both resource cost and computation performance 2. Improves the computation/communication ratio	Cloud Environment	Cloud Sim
Innovative transaction intensive cost-constraint scheduling algorithm [3]	Batch Mode	Execution cost and time	Workflow with large number of instances	1.To minimize the cost under certain user-designated Deadlines. 2. Enables the compromises of execution cost and time.	Cloud Environment	SwinDeW-C
A compromised-Time-Cost Scheduling Algorithm [4]	Batch mode	Cost and time	An array of workflow instances	1. It is used to reduce cost and cost	Cloud Environment	SwinDeW-C
A Particle Swarm Optimization-based Heuristic for Scheduling [5]	Dependency mode	Resource utilization, time	Group of tasks	1. it is used for three times cost savings as compared to BRS 2.It is used for good distribution of workload onto resources	Cloud Environment	Amazon EC2
SHEFT workflow scheduling algorithm [6]	Dependency Mode	Execution time, scalability	Group of tasks	1. It is used for optimizing workflow execution time. 2. It also enables resources to scale elastically during workflow execution.	Cloud Environment	CloudSim
market-oriented hierarchical scheduling strategy [7]	Virtual clusters	Make span, cost, CPU time	Service level scheduling ,task level scheduling	1.The overall running cost of cloud workflow systems will be minimized 2. It can be used to optimize both make span and cost simultaneously.	Cloud environment	SwinDeW-C

Multiple QoS Constrained Scheduling Strategy of Multi-Workflows [8]	Batch/dependency mode	Scheduling success rate,cost,time,make span	Multiple Workflows	1. It is used to schedule the workflow dynamically. 2. It is used to minimize the execution time and cost	Cloud Environment	CloudSim
Optimal Workflow based Scheduling (OWS) algorithm[9]	Virtual clusters	CPU utilization, Execution time	Multiple Workflows	1. It is used to find a solution that meets all user preferred QoS constraints. 2. It is used to improve CPU utilization.	Cloud Environment	Open nebula
RASA Workflow scheduling [10]	Batch mode	make span	Grouped tasks	1.It is used to reduce make span	Grid Environment	GridSim
HEFT workflow scheduling algorithm [11]	Dependency mode	make span	Highest Upward rank	1.It is used to reduce make span of tasks in a DAG is unbalanced.	Grid Environment	GridSim

5. METRICS FOR WORKFLOW SCHEDULING IN CLOUDS

The existing workflow scheduling algorithms consider various parameters like time, cost, make span, speed, scalability, throughput, resource utilization, scheduling success rate and so on. But, for a multiple workflows, metrics like Reliability and Availability should also be considered.

TABLE 2: Metrics considered by existing Workflow Scheduling Algorithms

Techniques	Time	Cost	Scalability	Scheduling Success Rate	Make span	Speed	Resource Utilization	Reliability	Availability
T1[14]	×	×	×	×	×	√	√	×	×
T2[2]	×	√	×	×	×	×	×	×	×
T3[3]	√	√	×	×	×	×	×	×	×
T4[4]	√	√	×	×	×	×	×	×	×
T5[5]	√	×	×	×	×	×	√	×	×
T6[6]	√	×	√	×	×	×	×	×	×
T7[7]	√	√	×	×	√	×	×	×	×
T8[8]	√	√	×	√	√	×	×	×	×
T9[9]	√	×	×	×	×	×	√	×	×
T10[10]	×	×	×	×	√	×	×	×	×
T11[11]	×	×	×	×	√	×	×	×	×

6. CONCLUSION

Moving workflows to a cloud computing environment enables the utilization of various cloud services to facilitate workflow execution. Workflow scheduling is one of the key issues in the management of workflow execution in cloud environment. In this paper, we have surveyed the various existing workflow scheduling algorithms in cloud computing and tabulated their various parameters along with tools and so on. Existing workflow scheduling algorithms does not consider reliability and availability. Therefore there is a need to implement a workflow scheduling algorithm that can improve the availability and reliability in cloud environment.

7. REFERENCES

[1] Mrs.S.Selvarani1, Dr.G.Sudha Sadhasivam," improved cost-based algorithm for task scheduling in Cloud computing "in IEEE 2010.

[2] Y. Yang, K. Liu, J. Chen, X. Liu, D. Yuan and H. Jin, An Algorithm in SwinDeW-C for Scheduling Transaction-Intensive Cost-Constrained Cloud Workflows, Proc. of 4th IEEE International Conference on e-Science, 374-375, Indianapolis, USA, December 2008.

[3] K. Liu, Y. Yang, J. Chen, X. Liu, D. Yuan and H. Jin, "A Compromised-Time- Cost Scheduling Algorithm in SwinDeW-C for Instance-intensive Cost-Constrained Workflows on Cloud Computing Platform", International Journal of High Performance Computing Applications, vol.24 no.4 445-456,May,2010.

[4] Suraj Pandey1, LinlinWu1, Siddeswara Mayura Guru2, Rajkumar Buyya1," A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments".

[5] Cui Lin, Shiyong Lu," Scheduling ScientificWorkflows Elastically for Cloud Computing" in IEEE 4th International Conference on Cloud Computing, 2011.

[6] Zhangjun Wu1, 2, Xiao Liu2, Zhiwei Ni1, Dong Yuan2, Yun Yang," A Market-Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems" in JSC2010.

[7] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing", in 2009 IEEE International Symposium on Parallel and Distributed Processing.

[8] P.Varalakshmi, Aravindh Ramaswamy, Aswath Balasubramanian, and Palaniappan Vijaykumar," An Optimal Workflow Based Scheduling and Resource Allocation in Cloud", Department of Information Technology, Anna University Chennai, India, pp. 411–420, 2011.

[9] Saeed Parsa and Reza Entezari-Maleki," RASA: A New Task Scheduling Algorithm in Grid Environment" in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.

[10] R. Sakellariou and H. Zhao, "A Hybrid Heuristic for DAG Scheduling on Heterogenerous Systems", The 13th Heterogeneous Computing Workshop (HCW 2004), Santa Fe, New Mexico, USA, April 26, 2004.

[11] W. Chen, J. Zhang, "An Ant Colony Optimization Approach to a Grid Workflow Scheduling Problem With Various QoS Requirements", IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, Vol. 39, No. 1, January 2009.

[12] C.Hoffa, G.Mehta, T.Freeman, et al., "On the Use of Cloud computing for Scientific Workflows", 4th IEEE Int.Conf.on eScience, December 2008.

[13] E.Deelman, "Mapping abstract complex workflows onto Grid environments", in Journal of Grid Computing 2003; 1:25–39.

[14] Hai Zhong1, 2, Kun Tao1, Xuejie Zhang1, 2," An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems ", in Fifth Annual China Grid Conference,2010.

[15] Zhifeng Yu and Weisong Shi, "A Planner-Guided Scheduling Strategy for Multiple Workflow Applications," icppw, pp.1-8, International Conference on Parallel Processing - Workshops, 2008.

[16] J. Yu and R. Buyya, "Workflow Scheduling Algorithms for Grid Computing", Technical Report, GRIDS-TR-2007-10, Grid Computing and Distributed Systems Laboratory, The University of Melbourne, Australia, May 2007.

[17] Yang Zhang1, Anirban Mandal2, Charles Koelbel1 and Keith Cooper," Combined Fault Tolerance and Scheduling Techniques for Workflow Applications on Computational Grids "in 9th IEEE/ACM international symposium on clustering and grid.