A Survey of Various Workflow Scheduling Algorithms in Cloud Environment

Simsy Xavier*, S.P.Jeno Lovesum**

* Post Graduate student, Department of Computer Science and Engineering, Karunya University, India
** Assistant professor (SG), Department of Computer Science and Engineering, Karunya University, India

Abstract- Cloud computing is a technology that uses the internet and the central servers to maintain data and resources.Cloud computing allows consumers and businesses to use the applications without installing and access their personal files at any computer with the help of internet. The users request for available services according to their desired Quality of Service, and they are charged on a pay-per-use basis. One of the most challenging problems in Cloud computing is the workflow scheduling the problem of satisfying the Quality of Service of the users as well as minimizing the cost of workflows executions. Workflow scheduling is one of the major issue in cloud computing environment . This paper surveyed different types of scheduling algorithms and compare their various parameters. Existing workflow algorithms does not consider the execution time. Therefore there is a need to implement a new scheduling algorithm that can minimize the execution time in cloud environment.

Index Terms- Cloud computing, Workflow scheduling, SaaS, Workflow management, Qos scheduling.

I. INTRODUCTION

A Cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and represented as one or more unified computing resources based on service level agreements established through the negotiation between the service providers and consumers. Cloud computing is a term that involves delivering hosted services over the Internet.

By using the virtualization concept, cloud computing can also support heterogeneous resources and flexibility is achieved. Another important advantage of cloud computing is its scalability. Cloud computing has been under growing spotlight as a possible solution for providing a flexible on demand computing infrastructure for a number of applications. All these factors increased the popularity of cloud computing The services are broadly divided into three categories Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS).

Workflows constitute a common model for describing a wide range of scientific applications in distributed systems. Workflow is described by a Directed Acyclic Graph (DAG) in which each computational task is represented by a node, and control dependency between tasks is represented by a directed edge between the corresponding nodes. Workflow scheduling is the problem of mapping each task to appropriate resource and allowing the tasks to satisfy some performance criterion. Workflow is processes that consist of a series of steps which simplifies the complexity of executions and management of applications.

The objective of this paper is to focus mainly on various workflow scheduling algorithms. The remainder of the paper is organized like this. Section II presents the key concepts of this paper. Section III presents existing workflow scheduling algorithms and section V concludes the paper with summary.

I. KEY CONCEPTS

The main concepts dealing in this paper are cloud computing and workflow scheduling. Cloud computing is a term that involves delivering hosted services over the Internet. Workflow scheduling is the problem of mapping each task to appropriate resource allowing the tasks to satisfy some performance criterion. Workflows are supported by Workflow Management Systems. Workflow scheduling discovers resources and allocates tasks on suitable resources. Software as a service (SaaS), sometimes referred to as on-demand software, is a software delivery model in which software and associated data are centrally hosted on the cloud. SaaS is typically accessed by end users, via a web browser .Qos is associated with end users and providers.

A. Cloud computing

A cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and represented as one or more unified computing resources based on service level agreements established through negotiation between the service providers and consumers. According to the type of services provided cloud computing is classified into three. They are called as service models. Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as aService (IaaS) are the service models. IaaS Clouds, like Amazon, provide virtualized hardware and storage on top of which the users can deploy their own applications and services. PaaS Clouds, like Microsoft Azure, provide an application development environment in which the users can implement and run applications on the Cloud. According SaaS cloud there are two types of Cloud, which deliver software applications to the users. The first group offers an entire application as a service to the end users, which can be used without any changes or customization. Examples of these Clouds are Google office automation service like Google Document or Google Calendar. The second group provides rudimentary web services to the users (known as on-demand web services), which can be used to build more complex applications.

Clouds are deployed in four ways. They are community cloud, public cloud, private cloud and hybrid cloud .In community cloud type the resources are shared between organizations. Community clouds are mainly benefited by Government, banks etc.Public cloud provide applications, storage, and resources available to the general public by using a service provider. These services are free or on a pay-per-use model. Generally, public cloud service providers like Microsoft and Google owns and operates the infrastructure and offer access only via Internet.Hybrid cloud is a composition of two or more clouds i.e (private, community or public) Hybrid clouds lack the flexibility, security in applications. Private cloud is mainly operated for single organization,and managed by internally or by a third party that hosted either internally or externally.

B. Workflow scheduling

Workflow scheduling is the problem of mapping each task to appropriate resource and allowing the tasks to satisfy some performance criterion. A workflow consists of a <u>sequence</u> of <u>concatenated</u> (connected) steps. Workflow mainly focused with the automation of procedures and also in order to achieve a overall goal thereby files and data are passed between participants according to a defined set of rules. A workflow enables the structuring of applications in a directed acyclic graph form where each node represents

the task and edges represent the dependencies between the nodes of the applications .A single workflow consists of a set of tasks and each task communicate with another task in the workflow. Workflows are supported by Workflow Management Systems. Workflow scheduling discovers resources and allocates tasks on suitable resources. Workflow scheduling plays a vital role in the workflow management. Proper scheduling of workflow can have an efficient impact on the performance of the system .For proper scheduling in workflows various scheduling algorithms are used.

II. EXISTING WORKFLOW SCHEDULING ALGORITHM

The following workflow scheduling algorithms are currently present in clouds and these algorithms are summarized .

1. A Particle Swarm Optimization based Heuristic for Scheduling Workflow Applications: Pandey, LinlinWu, Guru, Rajkumar Buyya presented a particle swarm optimization (PSO) based heuristic to schedule the applications to cloud resources that takes both computation and data transmission cost. It is used for workflow applications by varying its computation and communication costs. The evaluation results shows that PSO can minimize the cost and good distribution of workload onto resources.

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

3. Heterogeneous Earliest Finish Time algorithm (HEFT): Topcuoglu et. al, presented the HEFT algorithm. This algorithm finds the average execution time of each task and also the average communication time between the resources of two tasks. Then tasks in the workflow are ordered on a rank function. Then the task with higher rank value is given higher priority. In the resource selection phase tasks are scheduled in priorities and each task is assigned to the resource that complete the task at the earliest time.

4. Market-Oriented-Hierarchical Scheduling: Salehi, M.A. and Buyya, R. proposed a market oriented hierarchical scheduling strategy which consists of both service level scheduling and tasklevel 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 Virtual Machine assignment in local cloud data centers.

5. Cost based scheduling of scientific workflow applications on utility grids: Yu, J., Buyya, R. and Tham, C.K. proposed a cost based workflow scheduling algorithm minimizes the execution cost while meeting the deadline for delivering results.It can also adapt to the dealys of service executions by rescheduling unexecuted tasks.

6. Scheduling workflows with budget constraints: Sakellariou, R., Zhao, H., Tsiakkouri, E. and Dikaiakos, M.D proposed a basic model for workflow applications that modelled as directed acyclic graph(DAGs) and that allow to schedule the nodes of DAG onto resources in a way that satisfies a budget constraint and is optimized for overall time.

Table 1
Existing Workflow Scheduling Algorithms

Scheduling algorithm	Scheduling parameters	Scheduling factors	Findings	Environment	Tools
A particle swarm optimization-based heuristic[1]	Resource utilization time	Group of tasks	Good distribution of workload onto resources. Cost savings.	Cloud environment	Amazon EC2
HEFT workflow	Make span	Highest upward rank	Reduce make span in a DAG .	Grid environment	Grid sim

scheduling algorithm[3]					
Multiple Qos constrained scheduling strategy for multi-workflows[2]	Successrate,cost,time,makespan	Multiple workflows	Schedule the workflow dynamically. Minimize the execution time and cost.	Cloud environment	Cloud sim
Market oriented hierarchial strategy[4]	Cost,cpu time	Service and task level scheduling	Running cost is minimized. Optimize both cost and cpu time	Cloud environment	Amazon EC2
Workflow with budget constrained[9]	Makespan,budget	DAG scheduling	Minimize the execution time Makespan is minimized	Cloud environment	Amazon EC2
Cost based scheduling on utility grids[11]	Cost	Task scheduling	Reschedule the unexecuted tasks Minimize the cost	Grid environment	Grid sim

II. CONCLUSION

Workflow scheduling is one of the major issue in cloud computing environment. This paper surveyed the various existing workflow scheduling algorithms in cloud computing and compare their various parameters. Existing workflow algorithms does not consider the execution time. Therefore there is a need to implement a new scheduling algorithm that can minimize the execution time in cloud environment. Moving workflows to a cloud computing environment enables the use of various cloud services to facilitate workflow execution.

REFERENCES

[1] Pandey, S., Wu, L., Guru, S. and Buyya, R. ''A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing environments'', *24th IEEE Int'l Conference on Advanced Information Networking and Applications (AINA), Perth, Australia,* pp. 400–407 (2010).

[2] Xu, M., Cui, L., Wang, H. and Bi, Y. ''A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing'', *IEEE 11th Int'l Symposium on Parallel and Distributed Processing with Applications, Chengdu, China,* pp. 629–634 (2009).

[3] Wieczorek, M., Prodan, R. and Fahringer, T. ''Scheduling of scientific workflows in the ASKALON grid environment'', *SIGMOD Rec., 34(3),* pp. 56–62 (2005).

[4] Salehi, M.A. and Buyya, R. ''Adapting market-oriented scheduling policies for cloud computing'', Proceedings of the 10th Int'l Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010), Busan, Korea,pp. 351–362 (2010).

[5] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J. and Brandic, I. ''Cloud computing and emerging IT platforms: vision, hype and reality for delivering computing as the 5th utility'', *Future Gener. Comput. Syst., 25(6),* pp. 599–616 (2009).

[6] Hoffa, C., Mehta, G., Freeman, T., Deelman, E., Keahey, K., Berriman, B. and Good, J. ''On the use of cloud computing for scientific workflows'', *Fourth IEEE Int'l Conference on e-Science (e-Science 2008), Indiana, USA,* pp. 640–645 (2008).

[7] Juve, G., Deelman, E., Vahi, K., Mehta, G., Berriman, B., Berman, B.P. and Maechling, P. ''Scientific workflow applications on Amazon EC2'', *5th IEEE International Conference on e-Science, Oxford, UK* (2009).

[8] Bharathi, S., Chervenak, A., Deelman, E., Mehta, G., Su, M.H. and Vahi, K.''Characterization of scientific workflows'', *The 3rd Workshop on Workflows in Support of Large Scale Science, Austin, TX, USA,* pp. 1–10 (2008).

[9] Sakellariou, R., Zhao, H., Tsiakkouri, E. and Dikaiakos, M.D. ''Scheduling workflows with budget constraints'', *In Integrated Research in GRID Computing, S. Gorlatch and M. Danelutto, Eds Springer- Verlag.,* pp. 189–202, (2007).

[10] Amazon simple storage service (amazon s3). [Online]. Available: http://aws.amazon.com/s3/.

[11] Yu, J., Buyya, R. and Tham, C.K. ''Cost-based scheduling of scientific workflow applications on utility grids'', *First Int'l Conference on e-Science and Grid Computing, Melbourne, Australia,* pp. 140–147 (2005).

AUTHORS

First Author – Simsy Xavier, Post Graduate student, Department of Computer Science and Engineering, Karunya University, India

Second Author – S.P.Jeno Lovesum, Assistant professor (SG), Department of Computer Science and Engineering, Karunya University, India