Review Paper on- Workflow Scheduling Approaches in Cloud Environment

Harsimerdeep Kaur	**Monika Bharti**
CSE & IT &	*Assistant Professor CSE & IT &*
Chandigarh University, India	*Chandigarh University, India*

Abstract— *Cloud Computing gave a new direction for the betterment of IT industry. Cloud Computing provide services over Internet according to pay per services use. Processing of Workflow in parallel reduces the execution time, execution Cost and Cost computed by data transmission .Workflow is divided into number of stages to compute and each stage is interdependent further mapping is done to process these schedules on parallel machines. This paper focuses on the study of scheduling algorithms to compute schedule for user's workflow.*

Cloud computing environments facilitate applications by providing virtualized resources that can be provisioned dynamically. The advent of Cloud computing as a new model of service provisioning in distributed systems, encourages researchers to investigate its benefits and drawbacks in executing scientific applications such as workflows. There are a mass of researches on the issue of scheduling in cloud computing, most of them, however, are bout workflow and job scheduling. A cloud workflow system is a type of platform service which facilitates the automation of distributed applications based on the novel cloud infrastructure. In this paper, the authors have reviewed different types of workflow scheduling algorithms. The main aim of this work is to study various problems, issues and types of scheduling algorithms for cloud workflows.

Keywords— Cloud Computing, Scheduling Algorithms, Cloud broker, Workflow, Scheduling.

I. INTRODUCTION

As Internet is one of the most influential technology, that provides a platform for organizations and individuals to achieve their goal in a better way. New concept of Cloud Computing further exploits features of internet and make services and information available conveniently regardless of location and time. Cloud Computing is a term that provides services over the Internet and uses central remote servers to organize data. Blend of Cloud Computing and Internet is a boon to small scale as well as to large scale industry. The main target behind Cloud Computing is to reduce the cost of infrastructure, maintenance and users can access services what they want and pay for the services that they use. In Cloud, instead of storing data and running applications at individual desktop one can perform all these operations on Cloud. Anyone can access all the services and data any time regardless of location using internet.

A Cloud data centre consists of physical machines. Cloud use the concept of virtualization, as virtual machines are created over the physical machines. Communication between Cloud users and Cloud data centres is handle by broker. In Fig. 1, role of broker in scheduling is shown. The broker is responsible for allocating Cloud resources to users workflow applications. It uses scheduling algorithms to assign virtual machines to user's workflow applications and setup a written agreement between user and service providers called and SLA (Service Level Agreements). The term parallel processing is helpful in execution of workflows in as must time as possible. To execute the workflow parallel need to plan schedule, for schedule first of all have to compute mapping. There exist numerous of scheduling algorithms. Objective for scheduling algorithms is either to minimize execution cost or to minimize the data transmission cost or to minimize latency and many more.

II. CLOUD COMPUTING

Cloud computing is one of the fastest growing module of IT industry in the last few years. Cloud Computing can be defined as an internet based computing where virtual shared servers provide software, platform, infrastructure, devices and resources [1]. As Cloud computing provides many resources and depending on those different layers can be classified. Main concern of these layers is to reduce the capital expenditure and to provide the best operational opportunities. Instead of buying own software's or license, increasing storage space, one can take all these services on rent to accomplish their goal from Cloud service providers. Application as a Service Cloud application is the top most layer of the cloud which eliminates the need of installing and running the application on customer's computer which further eliminates the burden of software maintenance, ongoing operation and support services [1]. This service is also called as Application as a service. Cloud platform is middle layer well known by platform as a service, provides computing platform as a service. At the bottom of Cloud, infrastructure layer works. Infrastructure as a service provide IT infrastructure through virtualization. Thus, all these layers results in reduction of cost for deployment and maintenance of resources required by any organization.
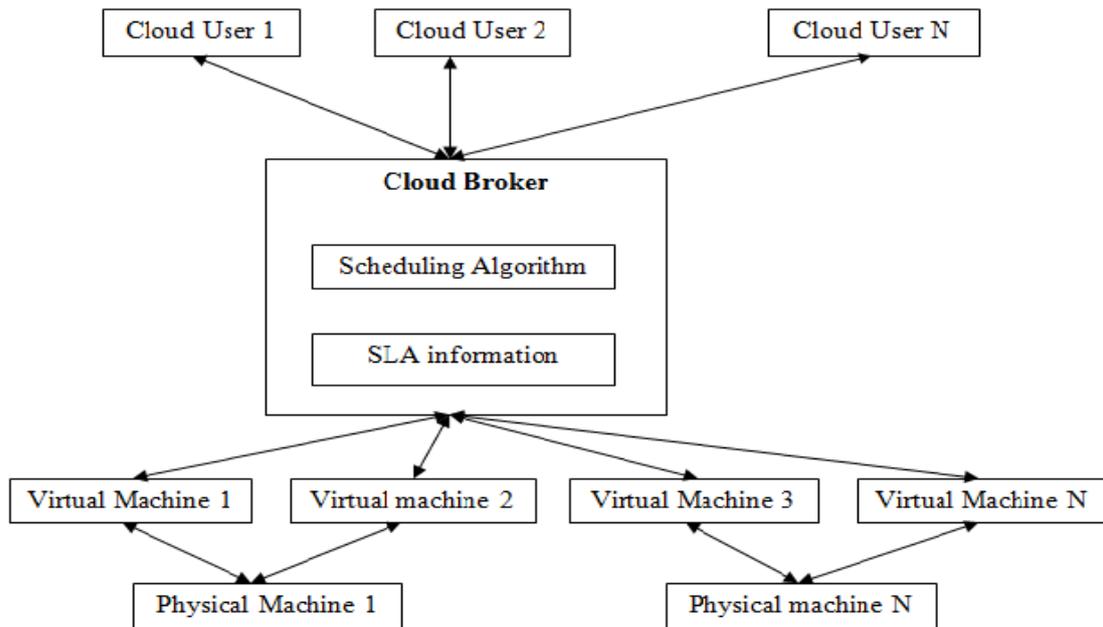
Fig. 1 Role of broker in Scheduling.

III. BENEFITS OF CLOUD COMPUTING

Cloud Computing contributes various benefits to world. Some of them mentioned below are:

A. *Cost Reduction:* Organizations and companies have better opportunity of opting Cloud Computing either investing heavy expenditure in set up of new hardware or buying software's and include benefit of better operational activities that further reduce operational cost. Customer pay for what services they use and for what time they use.

B. *Eliminate Maintenance Cost:* In Cloud Computing, Customer needs not to install any software on its own PC as maintenance of software is done by API's. Further maintenance of Cloud is done by the Cloud service providers, these results to reduction of maintenance cost.

C. *Scalable and Flexible:* A Company can be member of Cloud or can access the Cloud services within Short span of time and can start business from small deployment model to large scale model and according to requirement can again shift back to large scale to small scale deployment model.

D. *Mobility:* Cloud allows its users to access the services and information any time regardless of location just by having internet connectivity. Therefore, mobile users can also take benefit to increase the productivity using Cloud.

IV. CHALLENGES TO CLOUD COMPUTING

Although Cloud Computing has many advantages but it has severe challenges too. Some of them are discussed below:

A. *Security and Privacy:* Security is the most considerable challenge to Cloud. Organizations are not willing in handing over the organization's data to Cloud providers. However, data is stored on services in use by customer. Service provider even claim that there severs are free from risks like virus and malware but even though it is an issue as number of people access the servers. Privacy is another concern. Cloud service providers has to ensure customers that there confidentiality is maintained as no other users can access others data.

B. Service Quality: Service Quality is a factor any company thing about, before shifting to any service. Cloud Computing provide services on Internet. Therefore, required bandwidth should be adequate and continues connectivity is necessary. Second factor which is must, servers should not be down at any time as all the operations are performed on them.

C. Loss of Control: Once an organization stops paying for the services they are using, they will loss control over data as data is only storing at the Cloud no backup is formed.

V. RELATED WORK

A. K. Agrawal et. al [2] in 2010 presented the concept of scheduling and mapping of workflows in Cloud. Author arise the problem of scheduling for linear workflows if mapping is given. In this paper, scheduling algorithm is used either to minimize the period (inverse of throughput) or latency (bounded communication capacities and Communication/computation overlap) or both. Two models are used in scheduling, one is one-port model and another is multi-port model to minimize the period or latency. To minimize latency using these models is easy after computing schedule but to compute schedule to minimize period is NP hard using one-port model.
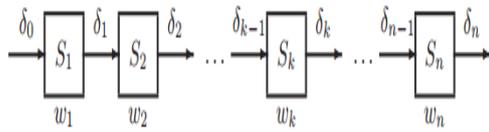
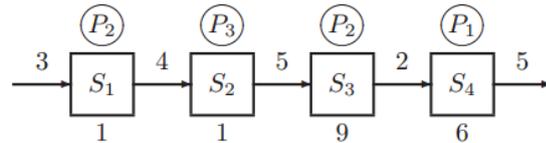Fig. 2 Representation of a linear workflow [2].   Fig. 3 Representation of a linear graph and a mapping [2].

B. Lovjit Singh et. al [3] in 2013 discussed workflow scheduling algorithms required according to the scenario and issues during research in workflow algorithm. Give the factors according to which scheduling algorithms are to be selected. The objective of algorithms is to optimize cost, time, reliability, load balancing or combination of all. At the end author mention the algorithms that can be used for workflow scheduling.
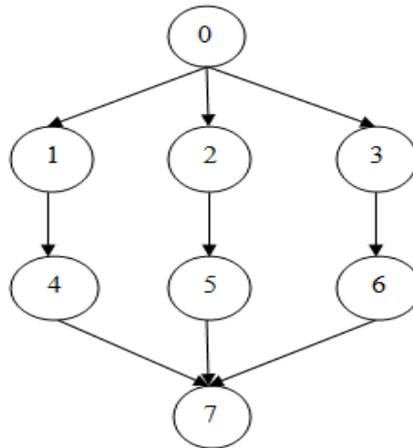

Fig. 4 Graphical representation of Workflow

A workflow application is generally represented as a Directed Acyclic Graph (DAG) such as G (V, E) where V is the number of tasks and E is the information regarding data dependencies among tasks. A task which does not have any parent task is called entry task and a task which does not have any child task is called an exit task. Figure 4 shows the dependencies among different tasks in a workflow graph G. The parent task 0 is executed before child tasks 1, 2 and 3.The output of parent node acts as an input to child node. The task 0 acts as entry node and task 7 act as an exit node. Task 9 is executed after the completion of tasks 4, 5 and 6.

C. Suraj Pandey et. al [4] in 2010 suggested the second factor need to be optimize that is execution cost and cost arising during transmission of data between sources along with execution time. Particular swarm optimization (PSO) is used to schedule applications taking into account to reduce the execution cost and data transmission cost. Author compares the results of PSO with Best Resource Selection (BRS) algorithm. Using PSO, cost reduces to 3 times by varying the communication cost between resources and the execution cost of compute resources as compare to BRS and workload distribution to resources is achieved.

D. Luiz Fernando Bittencourt et. al [5] in 2011 proposed a solution for the problem "Which are the best resources to request from a public cloud based on the current demand and on resources costs? " in hybrid Cloud as hybrid Cloud is aggregation of public Cloud and private Cloud. The algorithm named as HCOC (The Hybrid Cloud Optimized Cost scheduling) algorithm. Work of this algorithm is to decide which resources need to be taken at lease from public cloud and combine with the private cloud to provide enough processing power to execute a workflow within a given execution time. The goal achieved is reduction of cost while keeping second factor desired execution time in parallel.

E. Christian Vecchiola et. al [6] in 2009 give an overview about Cloud Computing and Aneka, an enterprise Cloud computing solution. Author presented a case study on using Aneka for the classification of gene expression data and the execution of fMRI brain imaging workflow. in Fig. 7, Architecture for Aneka is given. Aneka Clouds can be built on top of different physical infrastructures and integrated with other Cloud computing solutions.

F. Rajkumar Buyya et. al [7] discuss the term Cloud Computing, benefits given by Cloud Computing and Services provided by Cloud. In next section author explained the layer architecture of Cloud and functionality at each layer. A challenge is to quantifying the performance of resource allocation policies and application scheduling algorithms at finer details in Cloud computing environments for different application and service models under varying load, energy performance (power consumption, heat dissipation), and system size. Author proposed CloudSim to solve the problem. an extensible simulation toolkit that enables modelling and simulation of Cloud computing environments. The CloudSim toolkit supports modelling and creation of one or more virtual machines (VMs) on a simulated node of

a Data Centre, jobs, and their mapping to suitable VMs. It also allows simulation of multiple Data Centres to enable a study on federation and associated policies for migration of VMs for reliability and automatic scaling of applications.
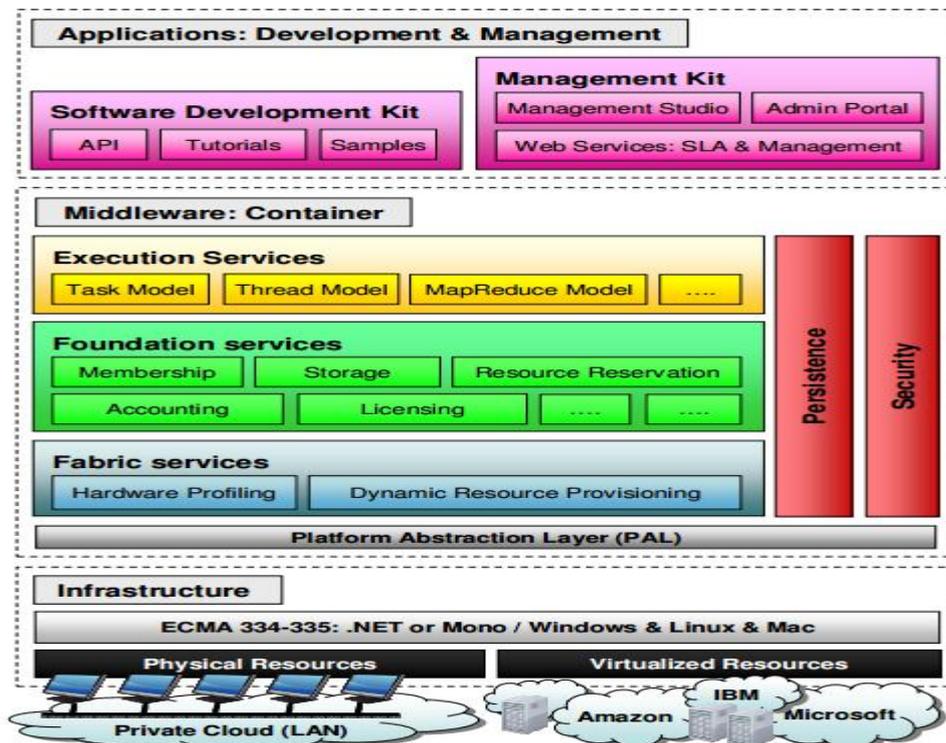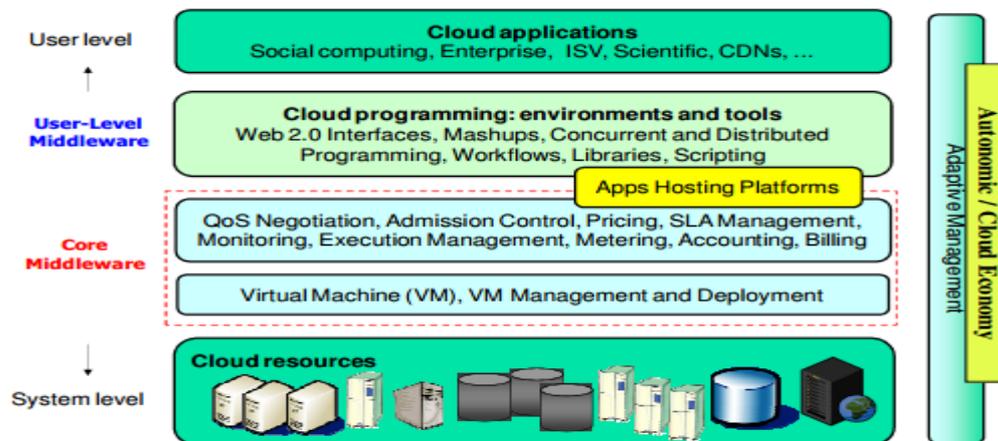

Fig. 5 Aneka Architecture [6]

Fig. 6 Layered Cloud Computing Architecture [7].

REFERENCE

[1]   Naresh Kumar and Shalini Sharma, "Study of Intrusion Detection System for DDoS Attacks in Cloud Computing" IEEE, 2013.

[2]   K. Agrawal, A. Benoit, L. Magnan and Y. Robert, "Scheduling Algorithms for Linear Workflow optimization" IEEE, 2010.

[3]   Lovejit Singh and Sarbjeet Singh, "A Survey of Workflow Scheduling Algorithms and Research Issues" Volume 74, July 2013.

[4]   Suraj Pandey, Linlin Wu, Siddeswara Mayura Guru and Rajkumar Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments" April 2010.

[5]   Luiz Fernando Bittencourt and Edmundo Roberto Mauro Madeira, "HCOC: a cost optimization algorithm for workflow scheduling in hybrid clouds" Volume 2, Issue 3, December 2011.

[6]   Christian Vecchiola, Suraj Pandey and Rajkumar Buyya, "High-Performance Cloud Computing: A view of Scientific Applications" 2009.

[7]   Rajkumar Buyya, Rajiv Ranjan and Rodrigo N. Calheiros, "Modelling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities".

[8]   Nils Gruschka and Meiko Jensen, "Attack Surfaces: A Taxonomy for Attacks on Cloud Services" IEEE, 2010.

[9]   Meiko Jensen, Jorg Schwenk, Nils Gruschka and Luigi Lo Iacono, "On Technical Security issues in Cloud Comptuing" IEEE, 2009.