Review on Workflow Scheduling by Optimization Approach in Cloud Computing

Astha gautam, Meenakshi sharma
Computer Science Engineering
LR Group of Institute, Solan, India

Abstract- 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 review on of different approaches of workflow scheduling using optimization approaches like PSO, ga etc.

Keywords- Cloud computing, Software as a Service, Platform as a Service, Infrastructure as a Service.

I. INTRODUCTION

a. Cloud Computing Evolution

Cloud computing is a distributed design that brings together server resources on an acceptable stage in order to provide on request figuring resources and administrative data [1]. A cloud specialist organization (CSP's) provides the different stages to their customers to use the services and make the web administrative control. This service is similar to a broadband band connection offered by the service provider for the internet connection. Cloud computing provides the services through the internet, these service belongs to hardware and software both. Cloud computing concept is growing high day by day due to its service pay per usage concept. When cloud provides the service in the form of platform it is called as Platform as a service model [1] [4]. When cloud provides the hardware to the consumer it is called as Infrastructure as a service model. When cloud provided the software services it is also called as Software as service [5].

b. Cloud Service Models

Into three service models cloud computing can be circulated: IaaS, PaaS, SaaS. Depending upon their particular needs an organization may acquire any gathering of these service model [3] [12].In Figure 1 these services are appeared.

(a) Software as a Service (SaaS): Over the web the SaaS (Software as a Service) delineates any cloud organization where purchasers can get to programming applications. For both individuals and affiliations the applications are encouraged in "the cloud" and can be used for an expansive assortment of assignments. By techniques for any web engaged contraption the Twitter, Facebook and Flickr are all examples of SaaS, with customers prepared to get to the organizations [4] [8] [17]. Rather than securing it the programming as Service clients, subscribes, and regardless to the things, if all else fails on a month to start. Rather than on singular PCs the applications are acquired and used online with records saved as a bit of the cloud.

(b) Platform as a Service (PaaS): Using instruments given by the provider the Platform as Service licenses clients to make programming applications. Customers can subscribe to the PaaS organizations can contain preconfigured fragments; while discarding those that don't they can combine the parts that meet their necessities.

(c) Infrastructure as a Service (IaaS): In the IaaS clouds the cloud customers clearly use IT bases (dealing with, structures, stockpiling, and other essential planning assets) is given. Virtualization is comprehensively utilized as a touch of IaaS cloud reviewing a legitimate focus to sort out/break down physical resources in an extraordinarily assigned way to deal with meet making or contracting resource request from cloud customers. The essential course of action of virtualization is to setup free VM that are disconnected from both the secured mechanical assembly and specific VMs [7] [10]. This framework is not totally the same as the multi-residency show up, which intends to change the design of application programming so that different cases can keep running on a particular application. A case of IaaS is Amazon's EC2.

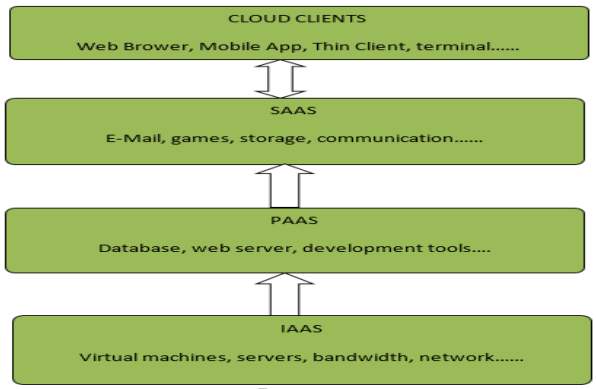

Fig.1: Cloud service Model

c. Cloud Deployment Model

Four sending models have been perceived, each with particular qualities that reinforce the prerequisites of the organizations and customers of the clouds specifically ways [3] [11].

(a) Private cloud: The cloud foundation is worked in a particular union, and facilitated by the association or an untouchable notwithstanding whether it is found prelude or off reason. The motivation to build a private cloud inside an affiliation has two or three perspectives. Regardless, to strengthen and streamline the utilization of existing in-house assets. The second, security concerns includes information protection and trust in like way make PC a likelihood for a couple of affiliations. Third, information exchange passed on from near to IT foundation to a PC is still rather stunning. The fourth, affiliations reliably require full control over mission-basic exercises that stay back of the firewalls [12].

*(b) Community cloud:*A few affiliations usually make and have a relative cloud base and in like way diagrams, necessities, qualities, and concerns. The cloud bunch shapes into a level of gainful and free change.

(c) Public cloud: this is the key sort of current Cloud managing sending model. People considering all things cloud is utilized by the general masses cloud customers and the cloud association supplier has the full duty as to open cloud with its own particular system, respect, and great position, costing, and charging model [9]. Particular comprehended cloud affiliations are open mists including Amazon EC2, S3,and Force.com.

*(d) Hybrid cloud:*The cloud base is a mix of no under two hazes (private, storing up, or open) that stay striking parts however are bound together by systematized or select development that pulls in information and application transportability (e.g., cloud influencing for weight changing between hazes). Affiliations utilize the cream cloud appear with a specific phenomenal concentration to push their advantages for build up their inside points of confinement by margining out edges business limits onto the cloud while controlling spotlight rehearses on-premises through private cloud [11] [16].

d. Cloud Applications

1. Development and Testing: Cloud plays an effective role as it is used for test and development. It saves the cost of setting up environment by setting up physically which includes the manpower and time. The installation and configuration of the software also take more time and this problem is also solved by using clod resources.

*2. Big Data Analytics:*Cloud is using the concept of big data and provides theeffective data extraction of the business value. It provides the effective data to the retailers and suppliers by extracting the buying patterns of the consumers. The buying patterns of the consumers show their liking and disliking of the consumers to the product.

3. File storage: Cloud offers the facility of data storage, retrieval and access from any web- enabled interface. The user can access data anytime, anywhere with high speed and availability. The large organization stores their data on cloud and only pay for the storage of data and they do not worry about the daily maintenance of the storage system.

*4. Disaster Recovery:*Cloud provides the effective data recovery in case of disaster at very effective cost. Data recovery by traditional method is very expensive and slow.

5. Backup: Backing up data is always a complex and time consuming process. The backup includes the tapes and drives to collect the data manually and then dispatching them for backup. Cloud provides the data backup automatically and no need to worry if the data is deleted. By using cloud it is easy to recover the data.

II. RESEARCH MOTIVATION

The distributed cloud computing condition gives numerous administrations to the vast scale logical work processes. Right off the bat, assets can be flexibly provisioned and de-provisioned on-request. It gives the capacity to the work process framework to utilize the assets as per the assignment and its compose artfully and should be prepared at a given point in time. Along these lines, the accessible asset pool can be scaled out and in as the work process prerequisites change. This is a helpful element for logical work processes as normal topological structures, for example, information appropriation and conglomeration [30] prompt noteworthy changes in the parallelism of the work process after some time. This prompts circumstances in which altering the quantity of assets being utilized is very attractive with a specific end goal to build execution and guarantee the accessible assets are proficiently used. Another advantage of sending work processes in cloud computing situations gets from the way that they are for the most part inheritance applications that contain heterogeneous programming segments. Virtualization takes into consideration the execution condition of these parts to be effectively redone. The working framework, programming bundles, registry structures, and info information records, among others, would all be able to be customized for a particular segment and put away as a VM picture. This picture would then be able to be effectively used to send VMs equipped for executing the product segment they were intended for. Another favorable position of utilizing VM pictures for the sending of work process errands is the way that they empower logical approval by supporting analysis reproducibility. Pictures can be put away and redeployed at whatever point an investigation should be imitated as they empower the production of the same correct condition utilized as a part of past analyses.

Benefits of Cloud Computing

- Cloud computing have some essential or unique characteristics as shown in Fig.1.2. is to provide qualitative services. These characteristics are as follows [2]
- **On-demand self-service** This self-advantage notification to the organization given by appropriated registering merchants that enables the course of action of cloud assets on ask for at whatever point they are required. In on-ask for self-advantage, the customer finds the opportunity to cloud benefits through an online control board.
- **Broad network access** Cloud computing isolates computing abilities from their consumers, with the goal that they don't need to keep up the capacities themselves. A consequence of this is the computing abilities are found somewhere else, and must be accessed over a network.
- **Resource pooling** Resource pooling is an Information Technology term used as a piece of distributed computing conditions to depict a situation in which suppliers serve diverse clients, clients or "inhabitants" with impermanent and flexible organizations. These organizations can be usual to suit every client's needs with no developments being clear to the client or end client. Occurrences of advantages combine stockpiling, arranging, memory, and framework data trade restrain.
- **Fast elasticity** It is described as the ability to modify resources both all over as required. To the buyer, the cloud has every one of the reserves of being immense, and the purchaser can purchase to such an extent or as pitiful enlisting power as they need.
- **Measured service** Cloud systems therefore control also, redesign asset use by utilizing a metering limit at some level of direction sensible to the sort of affiliation (e.g., dealing with, stockpiling, information transmission, and dynamic customer accounts).
- Stockpiling, information transmission, and dynamic customer accounts).

III. RELATED WORK

Qi Zhang et al. [1] have discussed the various cloud computing technologies and commercial products in detail. The commercial products have been compared on parameters like cloud provider, computing classes, target application, computation, auto scaling and storage. The research challenges presented in this work are service provisioning automation, server consolidation, virtual machine migration, energy management and data security.

Christian Vecchiola et al. [2] have given a comparison of computing solutions such as Amazon EC2, Google App Engine, and Microsoft Azure. The comparison is done on the basis of parameters like type of service, value added provider, if PaaS, ability to deploy on third party IaaS, platform, Virtualization, deployment Model and interface for user access. The aneka architecture, deployment model and application model are discussed in detail. Programming models like task model, map reduce model, thread model, parameter sweep model, workflow have been compared on the basis of execution services, applications and execution unit.

Yong Zhao et al. [3]have given many opportunities that the cloud has brought in, such as better utilization of resources, improved responsiveness thereby improving user experience, enabling a generation of collaborative scientific workflows and reducing the cost in challenges and opportunities in running scientific workflows on cloud. The challenges faced by the applications are architectural challenges, service challenges for integration tools, high-end computing support language-conversion challenge, challenge in compute intensive applications, challenge for data management, service management challenge.

Jens Sonke et al. [4] a scientific application is executed on FutureGrid, Amazon EC2 cloud and NERSC's Magellan in this paper. The result of this paper have been compared and analyzed to comprehend various challenges that came across during the process. In this work Pegasus workflow management system has been used to execute a scientific application which was used to process data from the Kepler project by NASA to find out planets similar to the earth.

Suraj Pandey et al. [5]has worked on reducing the computing cost of the application by using particle swarm optimization algorithm which is basically a meta-heuristic algorithm used for scheduling. PSO is used for calculating the fitness function. The total cost calculated is the cost of execution and the transfer cost of data. This algorithm ensures the cost of the highest task is reduced by heuristic scheduling. This algorithm helps to schedule the resources and mapping.

AlexandruIosup et al. [6] have explained the differences between the actual scope field of cloud and the requirements of scientific applications. It evaluates the cloud and check the capability of a cloud to run the applications efficiently. The evaluation is done by quantifying the number of users that require scientific computing services followed by evaluating four cloud services mostly used for scientific applications.

Simon Ostermann et al. [7] have discussed about various features of cloud computing which help ease the execution of scientific applications. It evaluates these features by different workloads like SJSI, MJSI and SJMI on Amazon EC2 cloud platform. Different types of benchmarks like Lmbench, Bonnie, and HPC are used to evaluate the performance of EC2 cloud for scientific applications.

EwaDeelman et al. [8] have discussed dependency of cost on execution models in this work. In this work, the cost is

calculated as a function of number of processors. The cost of executing montage workflow has been estimated by running simulation using GridSim tool. Three montage workflows have been Remote I/O, regular and dynamic cleanup. Three montage workflow that are executed are following montage degree 1, degree 2 and degree 4. The cost of running each of the data management models have been compared graphically. To maintain the trade-off between the number of processors and reduction in execution time.

Christina Hoffa et al. [9]worked on four different workflows and four different environments in order to compare the performances. The tools like WMS, DAGman. GridFTP, condoe and GRAM are explained in detail. The different montage degree workflows are executed on local machines and cluster and multiple virtual clusters.

Scott Callaghan et al. [10]have discussed and tackled the problems relating to managing a workflow in this work. In this work, the probabilistic seismic hazard analysis is performed for calculation estimates. It requires ground motions caused by past earthquakes as input. The ground motions are calculated by CyberShake 3D ground motion simulations with analytic wave propagation model.

Table.1 Existing Scheduling Model

Author's Name	Year	Methodology Used	Proposed Work
Netjinda*et al.*	2014	Particle Swarm Optimization	Focused on optimizing the value of buying infrastructure-as-a-service cloud competencies to attain clinical work goes with the flow execution in the unique closing dates
Nancharaiah*et al.*	2013	Ant Colony Optimization Algorithm and Particle Swarm Optimization	Displayed hybrid routing algorithm, Ant Colony Optimization algorithm and Particle Swarm Optimization (PSO) is utilized to enhance the different measurements in MANET routing.
AlexandruIosup*et al.*	2011	Scientific Computing Services	Explained the differences between the actual scope field of cloud and the requirements of scientific applications.
Suraj Pandey *et al.*	2010	Particle Swarm Optimization	Worked on reducing the computing cost of the application by using particle swarm optimization algorithm which is basically a meta-heuristic algorithm used for scheduling.
Christina Hoffa *et al.*	2008	WMS, DAGman. GridFTP, Condoe and GRAM	Worked on four different workflows and four different environments in order to compare the performances.

Christina Hoffa et al. [16] worked on four different workflows and four different environments in order to compare the performances. The tools like WMS, DAGman. GridFTP, condoe and GRAM are explained in detail. The different montage degree workflows are executed on local machines and cluster and multiple virtual clusters. Verma et al. [17] recommended that the users put up their workflows alongside a few QoS constraints like closing date, budget, and consider, reliability and so on. For computation, Authors considered the two constraints: closing date and finances and recommend cut-off date and finances Due date and Budget Distribution based cost-Time Optimization (DBD-CTO) work process scheduling set of rules that points of confinement execution regard while get together time diagram for giving over outcomes and separate the direct of the estimation

IV. CONCLUSION

Workflow scheduling in scientific computing systems is one of the most challenging problems that focuses on satisfying user-defined quality of service requirements while minimizing the workflow execution cost. So, to reduce the cost we use cloud environment. In cloud environment, resources will increase but its utilization is another challenge while using cloud environment. In this thesis, to maintain and utilization of the resources on the cloud computing scheduling mechanism is needed. Many algorithms and protocols are used to manage the parallel jobs and resources which are used

to enhance the performance of the CPU on the cloud environment. In the proposed work PSO (Particles swarm Optimization) and WCA (Water cycle optimization) for effective scheduling. Reason of using PSO because it will optimize global and locally to task and WCA also improve local optimization for VM migration. This work is based on the optimization of Total execution Time and Total Execution Cost

V. REFERENCES

[1]. Alkhanak, EhabNabiel, and Sai Peck Lee. "A hyper-heuristic cost optimisation approach for Scientific Workflow Scheduling in cloud computing." *Future Generation Computer Systems* (2018).

[2]. Netjinda, Nuttapong, BooncharoenSirinaovakul, and TiraneeAchalakul. "Cost optimal scheduling in IaaS for dependent workload with particle swarm optimization." *The Journal of Supercomputing* 68, no. 3 (2014): 1579-1603.

[3]. Abrishami, Saeid, Mahmoud Naghibzadeh, and Dick HJ Epema. "Deadline-constrained workflow scheduling algorithms for infrastructure as a service clouds." *Future Generation Computer Systems* 29, no. 1 (2013): 158-169.

[4]. Nancharaiah, B., and B. Chandra Mohan. "MANET link performance using ant colony optimization and particle swarm optimization algorithms." In *2013 International Conference on Communication and Signal Processing*, pp. 767-770. IEEE, 2013.

[5]. Rahman, Mustafizur, Xiaorong Li, and Henry Palit. "Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment." In *2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*, pp. 966-974. IEEE, 2011.

[6]. Mao, Ming, and M. Humphrey, "Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows" in *International Conference High Performance Computing, Networking, Storage and Analysis (SC)*, pp. 1-12, Nov. 2011.

[7]. Zhao, Yong, et al. "Opportunities and challenges in running scientific workflows on the cloud." *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2011 International Conference on*. IEEE, 2011.

[8]. Vöckler, Jens-Sönke, et al. "Experiences using cloud computing for a scientific workflow application." *Proceedings of the 2nd international workshop on Scientific cloud computing*. ACM, 2011.

[9]. Iosup, Alexandru, et al. "Performance analysis of cloud computing services for many-tasks scientific computing." *IEEE Transactions on Parallel and Distributed systems* 22.6 (2011): 931-945.

[10]. Zhang, Qi, Lu Cheng, and RaoufBoutaba. "Cloud computing: state-of-the-art and research challenges." *Journal of internet services and applications* 1.1 (2010): 7-18.

[11]. Callaghan, Scott, et al. "Scaling up workflow-based applications." *Journal of Computer and System Sciences* 76.6 (2010): 428-446.

[12]. Pandey, Suraj, et al. "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments." *Advanced information networking and applications (AINA), 2010 24th IEEE international conference on*. IEEE, 2010.

[13]. Vecchiola, Christian, SurajPandey, and RajkumarBuyya. "High-performance cloud computing: A view of scientific applications." *Pervasive Systems, Algorithms, and Networks (ISPAN), 2009 10th International Symposium on*. IEEE, 2009.

[14]. Ostermann, Simon, et al. "A performance analysis of EC2 cloud computing services for scientific computing." *International Conference on Cloud Computing*. Springer, Berlin, Heidelberg, 2009.

[15]. Deelman, Ewa, et al. "The cost of doing science on the cloud: the montage example." *High Performance Computing, Networking, Storage and Analysis, 2008. SC 2008. International Conference for*. Ieee, 2008.

[16]. Deelman, Ewa, et al. "The cost of doing science on the cloud: the montage example." *High Performance Computing, Networking, Storage and Analysis, 2008. SC 2008. International Conference for*. IEEE, 2008.

[17]. Aggarwal, Rohit, KunalVerma, John Miller, and William Milnor. "Constraint driven web service composition in METEOR-S." In *IEEE International Conference onServices Computing, 2004.(SCC 2004). Proceedings. 2004*, pp. 23-30. IEEE, 2004.