Volume 5, Issue 9, September 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Review of Workflow Scheduling Methods in Cloud Computing

Diksha Mahajan, Kapil Dewan
CSE, PCET, PTU
Punjab, India

Abstract: Provides services over Internet according to pay per services use 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 like Montage, Sipht and Cybershake. . Most of the algorithms that are currently in use, like First Come First Serve, Roundrobin etc., are ignoring the consideration of dependent and independent tasks that directly influence the overall execution time. We propose an approach based on Max-Min algorithm that will consider dependent and independent tasks separately and process the independent tasks simultaneously. It directly gives profit in minimizing computation time.

***Keywords*: Cloud Computing, Scheduling Algorithms, Cloud broker, Workflow, Scheduling**

I. INTRODUCTION

With the uprise of fourth paradigm that is discovery of science over a prolonged period of time, scientific workflows commence to amend their status amongst innumerous science subject areas including physics, astronomy, biology, chemistry, earthquake science and many more. In Scientific workflows, heavy volumes of data processing is required and workflows with up to a few million tasks are not unusual. Amongst these large-scale and loosely-coupled applications, the similar jobs within these requests are often fairly minor. However, critical amount of calculation and data is characterized by accumulation. The very best example of workflow is the CyberShake workflow that is employed by the Southern California Earthquake Center (SCEC) for machine parallel environment like Grid or the Cloud. An overhead is discrete as time taken to execute various works other than performing the user's computational activities is too much. Overheads can smashingly disturb runtime performance. To avoid and minimize the effects of overheads, concept of task clustering is applied. Task clustering has been employed to merge small tasks into larger jobs in order to diminish the amount of computational activities so that computational granularity can be increased by reducing the system overheads.

1.1 Demonstration of Workflow

A workflow is conventionally demonstrated as a Directed Acyclic Graph (DAG).

I. Node in the DAG signifies a workflow task
II. Edges signify dependency between the tasks (t) that pressure the order in which the tasks are performed.

Every job includes a program and a set of factors that need to be carried out to fulfill the job. A job (j) itself is a single execution unit and it comprises of one or multiple tasks. The dependencies typically means data flow dependencies in the application, because of result files yielded by one task are acquired as inputs for another task.

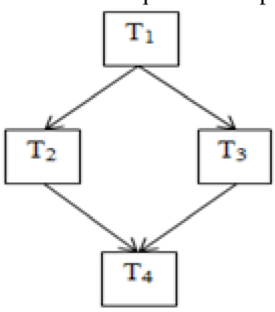

Fig 2.1 Directed Acyclic Graph

A. IMPROVEMENT METHODS IN WORKFLOWS

1.2.1 Using overhead aware (o-DAG) : DAG prototype is extended to overhead aware (o-DAG) because system overheads show a prominent role in workflow execution. With the poorly clustered tasks, overheads adversly affect the overall runtime. Enhancement in a DAG to be an o-DAG provides the ability to signify scheduling overheads such as workflow engine interruption, queue interruption, and postscript delay is shown in Figure 2.2.

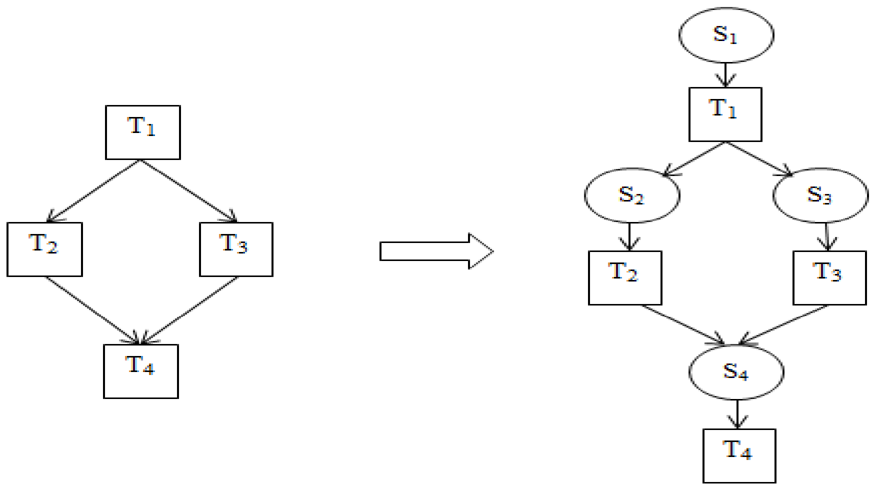

Figure 2.2 Extending DAG to o-DAG

Using o-DAG, one can clearly represent the producer of task clustering. For instance, in Figure 2.2, two tasks are taken without data dependency that is t_1 and t_2 which are merged to form a cluster job j_1. With this step scheduling overheads (s_2) are eliminated but clustering gap (c_1) is added.

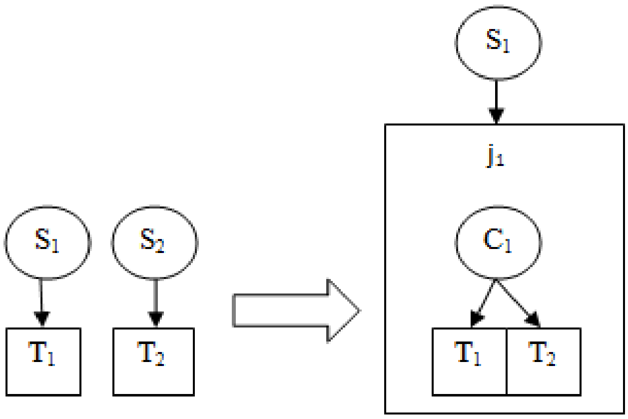

Figure 2.3 Task Clustering

1.2.2Using Workflow partitioning is optional approach that amends workflow organizations by separating a huge workflow into numerous sub-workflows that results in stepping down overall number of tasks and resource requirements that are required by these sub-tasks.
1.2.3 Using Sub-workflow is a workflow and a job of high-level workflow. Sub-workflows are executed on number of different sites and scheduled according to it. The performance of at runtime is improved.

1.3 CHALLENGES IN WORKFLOW APPLICATIONS

- **Imbalance of Computation** :Amongst the common challenge confronted because of merging workflow tasks is the imbalance of computation. Load imbalance is caused because of diversity of tasks at runtime. Numerous tactics have adopted to resolve this issue. A tactic called, Bag-of-Tasks resolve this issue by grouping tasks together on the behalf of related properties of tasks but assumption is taken that tasks are independent that limits the usage of them in scientific workflows.
- **Data Organization within a Workflow**: Scientific applications are often data modifiers for which a team of scientists is pre-requisite who belongs to various organizations, that result to distribution of requested data in scientific workflows. Primarily, with the blend of grids and clouds, scientists can upload their data and introduce their applications on scientific cloud workflow systems regardless of location through the Internet. But we cannot ignore the fact, having input data from or having output data to diverse data hubs is time-consuming and inefficient in case of merged tasks.
- **Resource management:** It is the third big challenge that comes forward with the combination of cloud computing and resource provisioning methods. Demand for different types of resources for execution has been increased rapidly with the increase of the amount of workflows. The potential to deliver resources on-demand is based on pay-per-use model and alter resource volume according to the varying demands of the application is offered by Infrastructure-as-a-Service (IaaS) clouds. Task clustering can still be implemented to this cloud situation. There is need to put attention and take decisions for cloud setup towards resource utilization factor along with the performance-related metrics because resources from commercial clouds generally have economic costs related with them.

- **Fault Tolerance** : Current clustering approaches overlook or undervalue the influence of the incidence of failures on system behavior, in spite of the increasing influence of failures in large-scale distributed systems. The value of fault acceptance design is highlighted by many researchers and they show that the failure rates are important in modern distributed systems. Temporary failures need more attention than the stable failures because they are predictable to be more widespread. There are generally three approaches for organizing workflow failures in failure situation.
  i. Retry the entire job when its computation is unsuccessful as in the Pegasus Workflow Management System. But there is wastage of resources and time in recompiling all the tasks because certain of the tasks within the job may be finished.
  ii. Check-points are used so that when a failure occurs from there only the tasks are recompiled so as to reduce the amount of work to re-execute. However, the overheads of check pointing can bound its profits.
  iii. Avoid location specific failures tasks can be replicated to different nodes. However, unsuitable clustering parameters may origin serious performance degradation if they generate long-running clustered jobs.

II. LITERATURE REVIEW

K. Agrawal et. al [1] 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. Scheduling algorithm is used either to minimize the retro (opposite of output) or potential (restricted message capabilities and Message/calculation overlap) or both. Two models are used in arrangement, one is one-port model and another is multi-port model. In one port model all processes are sequential and restricted communication capacities and communication/ computation overlap comes in multi-port model. To lessen potential using these models is easy after figuring list but to compute schedule to reduce retro is NP hard using one-port model. However period can be lessen in multi-port in polynomial time. Author presents an estimate procedure to lessen the retro in the one- port model and used Directed Acyclic graph to characterize workflow. Further, numerous heuristics to calculate a list of small period are: First is greedy algorithm but the problematic come from transportations: in the case where there are solitary computations, each processor can just compute its operations in any order, and the period is the cycle-time of the busiest processor. This hints to proj ect a new process, which first calculates communication between all processors (excluding Pin and Pout) in the schedule, and then includes greedily all computations. Further, longest first algorithm is used if there is a hole in the agenda a short process can be simply injected than a lengthy one.

Shabnam Khan [2] explained about varied resource allocation methods and their challenges are mentioned thoroughly. In cloud computing multiple cloud users will request range of cloud services at the same time. Thus there should be a provision that every one resources should be out there to requesting user in economical manner. During this paper a review of varied policies like linear scheduling strategy for resource allocation, dynamic resource allocation for parallel processing, their benefits, significance and disadvantages are studied.

Lovjit et. al [3] in 2013 discussed workflow scheduling algorithms vital according to the situation and problems during research in workflow algorithm. Give the factors according to which scheduling algorithms are to be selected. The algorithms are used in order to improve some factors such as cost, time, reliability, load balancing or combination of all. Author mentions the algorithm that can be used for workflow scheduling such as intelligent water drops. A workflow application is generally represented as a Directed Acyclic Graph (DAG) and role of broker in scheduling. At last, Author reference the parts in which effort in not yet given for instance it is vital to discover energy efficient genetic algorithm.

Suraj et. al [4] in 2010 suggested factors required to be optimize that is finishing cost and cost arising through communication of data between foundations along with finishing time. Particular swarm optimization (PSO) is used to list requests taking into account to lessen the completing cost and data transmission cost. PSO is a self-adaptive global search based optimization technique and similar to population-based algorithms like genetic algorithms. There is no direct re-combination of individuals of population in PSO but it relies on the social behaviour of particles. According to local best and global best position particles of every generation amends its path among the population. Author relates the outcomes of PSO with Best Resource Selection (BRS) algorithm. By PSO, cost lessens to 3 times by changing the communication cost between resources and the execution cost of compute resources as compare to BRS and workload supply to resources is attained and in future one can mix PSO into workflow management system to plan workflows of actual requests such as brain imaging analysis, EMO and others.

Luiz et. al [5] in 2011 familiarize with the term hybrid cloud and proposed a solution for the problem “According to existing demand and cost associated with the resources, to request for the best resources from private cloud “ 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 procedure is to resolve which properties need to be taken at contract from public cloud and combine with the private cloud to offer ample processing power to achieve a workflow within a given execution time and which task should be execute at public cloud or private cloud. The goal achieved is reduction of cost; minimize schedule length while keeping third factor to achieve is desired execution time in parallel and use number of workflows to compute. Authors get the best results in reduction of cost as compare to greedy algorithm. For future work, important issue to keep in mind is how to estimate the available bandwidth between two public clouds. Such estimative is important for the scheduling algorithm to resolve when needy tasks could be put in different public clouds.

Christian Vecchiolaet. 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. Fig. , represents architecture of Aneka. Aneka is used for developing scattered applications on the Cloud, as it is a software platform and a framework. An annoying set of APIs are provided by Aneka to makers for visibly exploiting such resources and conveying the soundness of applications by using a selection of programming abstractions. Aneka Clouds can be built on top of different physical infrastructures and integrated with other Cloud computing solutions.

RajkumarBuyyaet. al [7] discussed 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 measuring the performance of resource allocation policies and application scheduling algorithms. Author proposed CloudSim to solve the problem. An extensible imitation toolkit that allows 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.

RajkumarBuyyaet. al [8] in 2009 presents the 21st period image of calculating and detects several IT models hopeful to convey computing as a utility. Secondly, describes the construction for making market-oriented Clouds and computing atmosphere by leveraging technologies such as simulated machines. Further, delivers opinions on market-based resource organization plans that include both customer-driven facility organization and computational risk controlling to sustain SLA-oriented resource allocation. At last Author presents the work carried out as portion of new Cloud Computing creativity, called Cloudbus. Give detail about Cloud, Challenges facing by industry due to the adoption of Cloud e.g. security, privacy of data, reference model of Cloud, services like Iaas, PaaS, SaaS provided by it and components of cloudbus toolkit.

Michael Armbrustet. al [9] in 2009 describes Cloud computing as a utility computing, further discuss the services provided and deployment model of cloud according to pay-as-you-go model. Author presents three aspects of Cloud that are: infinite computing resources are accessible on demand; second is scalability which means a company can increase or decrease resources when there is requirement, instead of up-front commitment and last is user pay only for what they have used and for what time they used. Then, presents obstacles in adoption of cloud by companies e.g. data confidentiality and auditability, data lock-in, data transfer bottlenecks, scalable storage and other.

Luis M. Vaqueroet. al [10] in 2009 this paper elaborate the idea of Cloud to clear the basis that include what actually a cloud and model on which it works. In paper more than twenty definitions for cloud is given based on different papers. Comparison between cloud and grid computing is reveled because both terms are taken as having same features and designate the interactions and differences among the Grid and Cloud tactics. Last, discuss the areas in which research work is still needed to be carried out and features comes in Cloud, like virtualization, Quality of service etc.

Christina Hoffaet. al [11] in researchers focus on the usage of cloud computing for scientific workflows, concentrating on a extensively used astronomy application that is Montage workflow. A new approach is defined to compute the Montage workflow running in a local environment and on virtual environment. Results show that time taken to compute the workflow is efficient but in case of resource scheduling and wide area communication some obstacles arise. Problem of overheads has been seen while calculating Montage in virtual system and a technique could be used called clustering to overcome the problem of overheads in which small jobs are aggregated to form large jobs to reduce execution time instead of executing small jobs individually. Second problem researchers observed is disk space management.

Rodrigo N. Calheiroset. al [12] author define the problems related to Cloud infrastructure for diversity in application and service models under varying load, energy performance and system size. To resolve these issues Cloudsim is proposed that is a novel widespread and extensible simulation framework that enables seamless modeling, simulation, and experimentation of emerging Cloud computing infrastructures and management services. Further layered architecture, design and implementation of CloudSim are elaborated. In future work, planned to incorporate new pricing and provisioning policies to CloudSim, in order to offer an integral support to simulate the presently available Clouds.

III. CONCLUSION

optimize workflow scheduling , Time Shared, Round-Robin, Data aware algorithms are deployed and results obtained are compared with each other. Parameters included are Optimization Time Delay. Workflows used during experiments are CyberShake, SPIHT. give a good performance in Cyber shake but average performance sipht workflow ,If we analysis this result we conclude that those workflow which have longest process greater than low time process and low time process time covered in long time process like cyber shake there Max-min algorithm work good because execute longer process first and meanwhile process low time delay process, if they are not dependent to each other but if they are highly dependent there not work Properly like Sipht work flow . In sipht workflow perform Average .

REFERENCES

[1] Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." *Speech communication* 41.4 (2003): 603-623.

[2] Pan, Yixiong, Peipei Shen, and Liping Shen. "Feature Extraction and selection in speech emotion recognition." *Department of Computer Technology, Shanghai JiaoTong University, Shanghai, China* (2012).

[3] Rathina, X. Arputha, K. M. Mehata, and M. Ponnavaikko. "Basic analysis on prosodic features in emotional speech." *nternational Journal of Computer Science, Engineering and Applications (IJCSEA)* 2.4 (2012): 99-107.

[4] Pierre-Yves, Oudeyer. "The production and recognition of emotions in speech: features and algorithms." *International Journal of Human-Computer Studies*59.1 (2003): 157-183.

[5] Bozkurt, Elif, et al. "Improving automatic emotion recognition from speech signals." *INTERSPEECH*. 2009.

[6] Zheng, Wenming, et al. "A novel speech emotion recognition method via incomplete sparse least square regression." (2014): 1-1.

[7] Tato, Raquel, et al. "Emotional space improves emotion recognition."*INTERSPEECH*. 2002.

[8] Schuller, Björn, Gerhard Rigoll, and Manfred Lang. "Hidden Markov model-based speech emotion recognition." *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. Vol. 2. IEEE, 2003.

[9] Schuller, Björn, Manfred Lang, and Gerhard Rigoll. "Automatic emotion recognition by the speech signal." *Institute for Human-Machine-Communication, Technical University of Munich* 80290 (2002).

[10] Panda Bhoomika , Padhi Debananda, Dash Kshamamayee, Prof. Sanghamitra Mohanty. " Use of SVM Classifier & MFCC in Speech Emotion Recognition System." *international Journal of Advanced Research in Computer Science and Software Engineering* 1.5 (2012).

[11] Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." *Speech communication* 41.4 (2003): 603-623.

[12] Pan, Yixiong, Peipei Shen, and Liping Shen. "Feature Extraction and selection in speech emotion recognition." *Department of Computer Technology, Shanghai JiaoTong University, Shanghai, China* (2012).

[13] Rathina, X. Arputha, K. M. Mehata, and M. Ponnavaikko. "Basic analysis on prosodic features in emotional speech." *nternational Journal of Computer Science, Engineering and Applications (IJCSEA)* 2.4 (2012): 99-107.

[14] Pierre-Yves, Oudeyer. "The production and recognition of emotions in speech: features and algorithms." *International Journal of Human-Computer Studies*59.1 (2003): 157-183.

