

Review on Workflow Scheduling In Cloud Environment: A Comprehensive Study

Pratiksha¹, and Afsaruddin²

¹ PG Scholar, Department of Computer Science & Engineering, Integral University, Lucknow, India

² Assistant Professor, Department of Computer Science & Engineering, Integral University, Lucknow, India

Correspondence should be addressed to Pratiksha; Pratiksha.vermag8@gmail.com

Copyright © 2021 Made Pratiksha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- For enterprise application development, cloud computing is going to new-fangled paradigm that meets the needs of today's enterprises and helps them efficiently executing workflows in business and scientific applications. Workflow applications often require a very intricate runtime environment that is hard to create and maintain. Thus, cloud infrastructure is valued as an implementation platform for positioning technical workflows with features such as on-demand provisioning, scalability, reproducibility and resilience. To schedule multiple workflows in the cloud, we need to design the workflow before scheduling it. Therefore, in this paper we review the diverse work done by researchers for the workflow design as well as workflow scheduling.

KEYWORDS- Cloud Computing, Distributed Computing, Work Flow Scheduling, Benefit Of Clouds, Load Balancing

I. INTRODUCTION

Information is a term which is so huge and enormous content taking in this internet era [1]. Data and information are two important aspect of cloud computing, in these days data are rapidly generated from the various field, like data scientist generated data, artificial intelligence system is generating data, now data are generated from various sources, so a lot of complication is going to be created. cloud computing paradigm give us a lot of solution in this regard [2] . Cloud computing environment ensures the accessibility of computer resources (particularly data storage and computing processing power) on demand, without direct vigorous control from user. Cloud computing provides the storage service, drive of calculation elements and software delivery It delivers to the upcoming generations of computing infrastructures hosted by large companies. Elasticity, pay-as-you-go model and on-demand provisioning of resources are the three main characteristics of cloud computing. These enabling features of cloud computing are very useful for business people. The large infrastructure providers built Elastic Compute Cloud (EC2), Google Cloud and Azure cloud platforms to support the applications around the world.

II. BENEFITS OF CLOUD

A. Cost Effectiveness

The main motive for the move to cloud computing is that it costs significantly less than on-premises technology. The cloud provides massive storage space, eliminating the need to store data on disk, saving corporate money and resources.

B. Elevated Pace

Cloud computing enables users to quickly deploy services with fewer clicks. With this high-speed deployment, one can get the resources in minutes.

C. Exceptional Accessibility

Information stored in cloud can be accessed from anywhere and anytime.

D. Data Back-up and Restore

Once the data are stored in cloud, it is easier to get the back-up and recovery of that, which is quite a time taking process on-premise.

E. Manageability

The requirement upgrading IT infrastructure is eradicated since the service provider ensures well-timed, assured and flawless delivery of the user services and also deals with the entire maintenance of the IT enabled services in accordance with the Facility Equal Contract (SLA).

F. Load balancing

In cloud computing, the number of computing nodes rapidly generates the data, in this situation load of the computing server are increased, but with the concept of load balancing approach all the computing server load can be balanced and further the throughput and efficiency of the computing nodes will be increased [3]

III. DRAWBACKS OF CLOUD COMPUTING

A. Openness to Attacks

The Challenge of Information theft is always there when data is stored on cloud. Although highly developed security measures are deployed then also storing private on cloud is perilous.

B. Network Connectivity Enslavement

To obtain the benefits of cloud services we need steadfast

internet service with superior connection speed as well as bandwidth.

C. Downtime

Due to power interruption or connectivity issue, internet services or data centers may go out of services leading a momentary downtime.

D. Vendor Lock-in

When users intend to change cloud, platforms from different vendors can pose major challenges for the company. Hosting and running applications on other platforms can introduce support issues, configuration complexity, and additional costs. In addition, potential disruptions during migration can leave the business data vulnerable to security attacks.

E. Restricted Control

Cloud clients may have restricted control over their deployment. Cloud services run on isolated servers that are wholly owned and controlled by the service provider, making it complicated for enterprises to get necessary control over their internal infrastructure.

F. Strategic Edge

In accessing the latest mission-critical applications whenever required, cloud computing leads companies to be more competitive than their competitors without investing time and money in installation.

G. Security breach

In cloud computing security is one of the important aspects, several time clients fail to get the actual information, some time data can be loss, this stage is concern to modify the security system every times. Client need to modify the system password and also modify the security protocol every times .

IV. WORKFLOW SCHEDULING

The increasing demand and conglomeration of cloud computing is gaining recognition among the scientific commune to leverage cloud services to implement large-scale electronic applications as workflows (set of tasks). Workflows are regarded as computations performed considering task dependencies. It allows users to straightforwardly elucidate multifaceted multi-step calculation tasks. Workflow tasks are associated to the mechanization of procedures where information, documents or tasks are delivered between partakers in accordance with a specific set of policies and allow the formation of different applications in a Directed Acyclic graph (DAG). In the DAG every node symbolizes the component task and the edges symbolize the dependencies [3][4]. Workflows can be categorized as single and multiple. Single workflow consists of one or more instances using the same architecture Multiple workflows can take into account manifold cases of diverse structures of workflow. A mechanism to manage various workflows on cloud, is known as Workflow Management System (WMS).

The requirement of workflow scheduler is to create the execution plan for workflow task on cloud resources that are utilized to implement the workflow. The components that must be delineated to schedule a workflow are revealed below in Figure 1.

V. WORKFLOW SCHEDULING: STATE OF THE ART

Cloud computing is evolving towards a service-oriented paradigm that makes it easier for scientific applications to take advantage of cloud resources over the Internet. Scientific applications necessitate roadmap processing in which workflow tasks are performed according to task-to-data dependencies. Diverse scientific workflow applications can be structured differently, allowing different cloud resources to manage multiple workflows at the same time. To schedule multiple workflows in the cloud, one needs to design the workflow before scheduling it. Consequently, this section describes a workflow design approach and existing workflow planning algorithms .

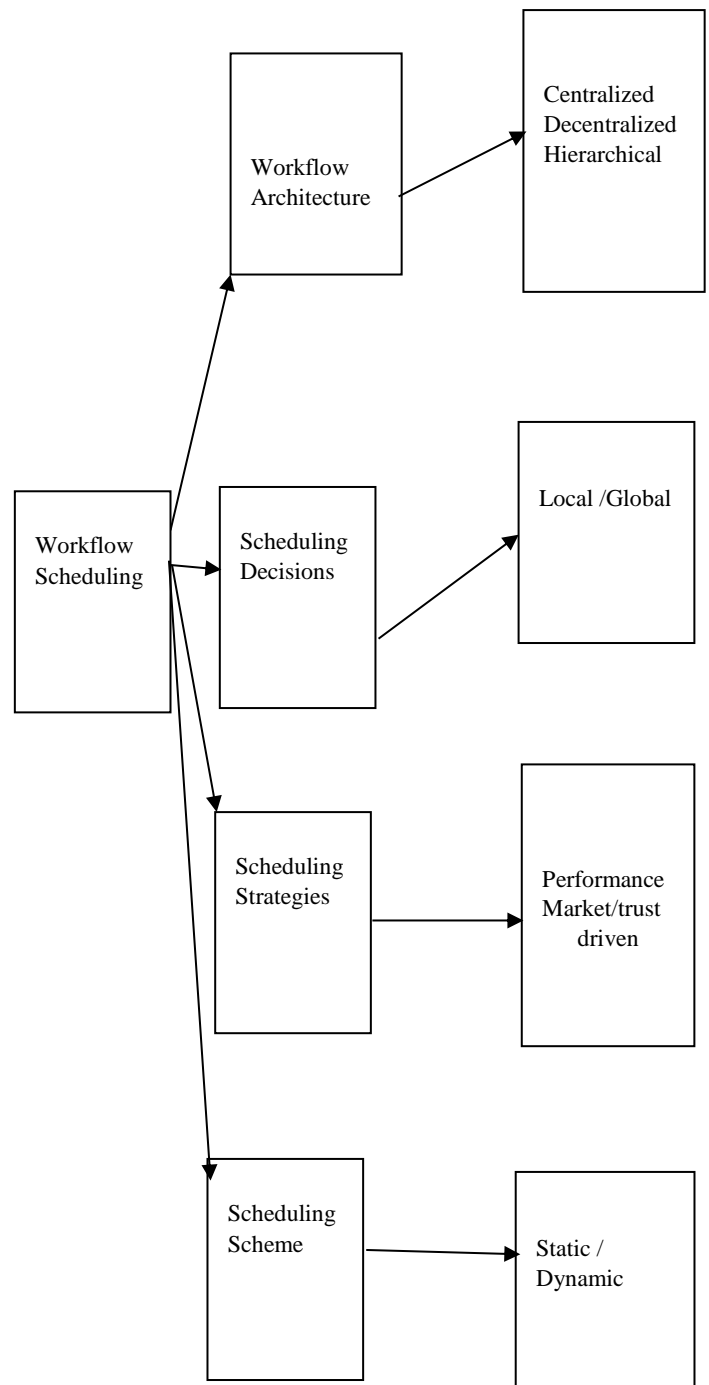


Fig. 1: Components of Workflow Scheduling

Table 1: Various Approaches for Designing Workflows

Workflow Design Approach	Characteristics	Language used	Suggested Future Work	Implementation Setting
Approach based on Top-Down Petri Net [5]	Boosts the dependability of consequential workflow nets	Petri Nets Based JAVA	Risk of Deadlock can be diminished	Disseminated
Product Based Approach [6]	Improves liveness and dynamicity	Petri nets	Optimal strategies can be used	Distributed
Model-Driven Approach [7]	Good for designing user interface	XML	Utilization of more advanced Workflow Analysis Tools	Cloud
Knowledge Based Approach [8]	Manages the exceptions	-	Implementation required	Distributed
ECA Rule Based Design Approach [9]	Solves the complicated business logic or scientific application problems	WSDL	AI techniques need to be merged	Grid Environment
Service based Approach [10]	Designs the monitoring process	Prom	Should be mingled with a different approach	Distributed
Agent Based Approach [11]	knowledge-exhaustive workflows	JAVA	Should be combined With a different approach	Distributed

Table 2: Various Approaches for Scheduling Workflows

Algorithm	Methods	Parameters	Factors	Conclusion	Environment	Tools
Optimized-Resource Scheduling	Multiple instances	Speed, Resource Utilization	Request allocation problem	enhances pace of IGA [12]	Cloud	Eucalyptus
Scheduling based on Transaction Intensive Cost Constraint [13]	Batch Mode	Implementation expenditure and period	Workflow with hefty instances	Minimizes expenditure, Improves the computation/communication ratio	Cloud	SwinDeW-C, CloudSim
A Particle Swarm Optimization based Heuristic [14]	Dependency mode -	Resource utilization, time	Group of tasks	Utilizes for three times cost savings as compared to BRS	Cloud	Amazon EC2, CloudSim

Market-oriented Hierarchical Scheduling Strategy [15]	Virtual clusters	Make span, Expenditure, CPU time	Service level scheduling, Task level scheduling	Optimizes both makes pan and cost at the same time	Cloud	SwinDeW-C
SHEFT algorithm [16]	dependence Mode	Execution time, scalability	collection of tasks	Optimizes work-flow implementation period, enables resources to scale elastically during workflow carrying out	Cloud	CloudSim.
Manifold QoS Constrained Scheduling approach of Multi Workflows [17]	Batch or dependency	Scheduling success rate, Cost, Time, Make span	Multiple Work-flows	Makes Workflow execution plan dynamically, minimizes completing period as well as cost	Cloud	CloudSim.
GRASP[18]	Batch or dependency	pace, outlay	Data concentrated Workflows	Evaluate performances of Min-Min heuristics for workflow applications	Grid	GridSim, NS2
GA[18]	Dependency mode	Reliability, Makespan	n Single workflows	Schedules and maps on heterogenous environment	Grid and Cloud	GridSim, CloudSim.
Optimal Work-flow based Scheduling (OWS) algorithm [19]	Virtual clusters	CPU exploitation, finishing period	manifold Work-flows	perk up CPU exploitation	Cloud	Open nebula.
RASA Workflow scheduling [20]	Batch mode	make span	clustered tasks	Used to reduce make span	Grid	GridSim.
SA[21]	Virtual clusters	Execution Time	Single	Minimizes execution time	Grid, Cloud	GridSim, CloudSim”

Table 2 analyzes the metrics for the workflow planning approach described above. Existing workflow scheduling algorithms take a variety of parameters such as time, makes pan, cost, speed, throughput, scalability, resource utilization, and likelihood of successful scheduling into account. Reliability as well as accessibility has not been reviewed by any of the authors of the reviewed articles

VI. CONCLUSION

To complete workflow tasks in the cloud, scientists require proficient scheduling algorithms to plan synchronized workflows for cloud resources. This paper has provided an extensive analysis of workflow planning algorithms and a design approach. We found that current scheduling heuristics do not prioritize numerous workflows across

cloud resources.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] S. Srivastava, M. Haroon, and A. Bajaj, “Web document information extraction using class attribute approach,” Proc. - 4th IEEE Int. Conf. Comput. Commun. Technol. ICCCT 2013, pp. 17–22, 2013, doi: 10.1109/ICCCT.2013.6749596.
- [2] H. S. Kharkwal, “Automated Task Allotment in Unmanned Submarines by Smart Searching Algorithm,” vol. 13, no. 2, 2017.
- [3] R. Khan, M. Haroon, and M. S. Husain, “Different technique of load balancing in distributed system: A review paper,”

- Glob. Conf. Commun. Technol. GCCT 2015, no. Gcct, pp. 371–375, 2015, doi: 10.1109/GCCT.2015.7342686
- [4] X. Zhou, G. Zhang, J. Sun, J. Zhou, T. Wei, and S. Hu, “Minimizing cost and makespan for workflow scheduling in cloud using fuzzy dominance sort based HEFT,” *Futur. Gener. Comput. Syst.*, vol. 93, pp. 278–289, 2019, doi: 10.1016/j.future.2018.10.046.
- [5] M. S. Husain and D. M. Haroon, “an Enriched Information Security Framework From Various Attacks in the IoT,” *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 8, no. 4, 2020, doi: 10.21276/ijircst.2020.8.4.3.
- [6] P. Wang, Y. Lei, P. R. Agbedanu, and Z. Zhang, “Makespan-Driven Workflow Scheduling in Clouds Using Immune-Based PSO Algorithm,” *IEEE Access*, vol. 8, pp. 29281–29290, 2020, doi: 10.1109/ACCESS.2020.2972963.
- [7] Zhang, G., Sun, J., Zhou, J., Wei, T. and Hu, S., “Minimizing cost and makespan for workflow scheduling in cloud using fuzzy dominance sort-based HEFT”, *Future Generation Computer Systems*, 93, pages.278-289,2019
- [8] Piotr Chrz, Astowski-Wachtel, Boualem Benatallah, Rachid Hamadi, Milton ODell, and Adi Susanto. “ A top-down petri net-based approach for dynamic workow modelling” *LNCS*, 2678:336-353, 2003.
- [9] Irene Vanderfeesten, Hajo A. Reijers, Wil M.P., and van der Aalst, “Product based workow support: A recommendation service for dynamic workow execution” *LNCS*, 5074:571-574, 2008.
- [10] Dr. Mark Klein and Prof. Chrysanthos Dellarocas, “A knowledge-based approach to handling exceptions in workow systems” *Journal of Computer Supported Collaborative Work*, Special Issue on Adaptive Workow Systems, pages 1-11, 2000.
- [11] Lin Chen, Minglu Li Jian, and Cao YiWang, “An eca rule-based workflow design tool for shanghai grid”, *Services Computing (SCC05)*, 2005.
- [12] Macro Comuzzi and Irene T.P. Vanderfeesten, “Product based workflow design for monitoring of collaborating buisness processes” *CAiSE, Bibliography 132*, pages 154-168, 2011.
- [13] Therani Madhusudan, “An agent-based approach for coordinating product design workflows”. *International journal of science direct Computers in Industry*, pages 235-239, 2005.
- [14] P.Varalakshmi, Aravindh Ramaswamy, Aswath Balasubramanian, and Palaniappan Vijaykumar, “An optimal workow based scheduling and resource allocation in cloud”, *International Conference*, pages 411-420, 2011
- [15] Xiaofeng Wang, Chee Shin Ye, Rajkumar Buyya, and Jinshu Su, “Optimizing the makespan and reliability for workow applications with reputation and a look-ahead genetic algorithm” *Future Generation Computer Systems*, (27): 11241134, 2011.
- [16] Saeed Parsa and Reza Entezari-Maleki, “Rasa: A new task scheduling algorithm in grid environment”, *World Applied Sciences Journal 7 Special Issue of Computer, IT*, pages 152-160, 2009.
- [17] Piotr Chrz, Astowski-Wachtel, Boualem Benatallah, Rachid Hamadi, Milton ODell, and Adi Susanto. “ A top-down petri net-based approach for dynamic workow modelling” *LNCS*, 2678:336-353, 2003.
- [18] Meng Xu, Lizhen Cui, Haiyang Wang, and Yanbing Bi, “A multiple qos constrained scheduling strategy of multiple workflows for cloud computing”, *IEEE International Symposium on Parallel and Distributed Processing*, 2009., pages 629-634, 2009.
- [19] Cui Lin and Shiyong Lu, “Scheduling scientific workflows elastically for cloud computing”, *IEEE 4th International Conference on Cloud Computing*, 2011.
- [20] Maria Alejandra Rodriguez and Rajkumar Buyya, “Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds” *IEEE Transactions On Cloud Computing*, (2):222-235, May 2014
- [21] Kumar, M., Sharma, S.C., Goel, A. and Singh, S.P., “A comprehensive survey for scheduling techniques in cloud computing”. *Journal of Network and Computer Applications*, 143, pp.1-33, 2019

ABOUT THE AUTHORS



Pratiksha, M.Tech, Research Scholar, Department of Computer Science & Engineering at Integral University, Lucknow (U.P), India



Afsaruddin, Working as an Assistant Professor, Department of Computer Science & Engineering at Integral University, Lucknow (U.P), India. Research Interest: Cloud Computing, IoT, Artificial Intelligence.