UNIVERSITI PUTRA MALAYSIA

Deadline Guarantee for Scientific Workflow using Dynamic Scheduling Algorithms on IaaS Clouds

Sarah Abdulrahman Shukur Alqaisy

FSKTM 2018 46

Deadline Guarantee for Scientific Workflow using Dynamic Scheduling Algorithms on IaaS Clouds

By

Sarah Abdulrahman Shukur Alqaisy

GS47274

Thesis Submitted to Universiti Putra Malaysia, in Fulfillment of the Degree of Master Computer Science

July 2018

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for Master of Computer Science

Deadline Guarantee for Scientific Workflow using Dynamic Scheduling Algorithms on IaaS Clouds

By

Sarah Abdulrahman Shukur Alqaisy

July 2018

ABSTRACT

Many important scientific applications can be expressed as workflows, which describe the relationship between individual computational tasks and their input and output data in a declarative way. This enables workflows to be automatically adapted to run across different environments. For complex workflows, abstraction also helps scientists to express their workflows at a higher level without being concerned about the details of how individual jobs are invoked or how data is transferred between jobs. Also, large-scale applications expressed as the scientific workflows that are often grouped into ensembles of interrelated workflows (J. Vöckler, G. Juve & G. B. Berriman, 2011),(M. Malawski, G. Juve, E. Deelman & J. Nabrzyski ,2015). Normally, commercial Cloud computing is rapidly becoming the target platform on which to preform scientific computation. Typically, the commercial Cloud services charge on the basis of the number of hours the resources (such as CPU, network bandwidth and amount of storage) are used. This charging model is referred to as pay-per use. The flexibility inherent in the elastic Cloud model, while powerful computing results in inefficient

usage and high costs when inadequate scheduling and provisioning decisions are made. The problem of scientific workflow scheduling in Clouds requires an alternative scheduling approach in mapping tasks to resources while fulfilling the deadline in the workflow of Cloud computing. In this project, our objective achieved better performance in term of success rate when compared to existing scheduling algorithms. In terms of solving the problems we efficiently designed the workflows scheduling on dynamically provisioned Cloud resources, while reducing the computation complexity. Specifically, we enhance two scheduling algorithms Proportional Deadline Constrained (PDC) and Deadline Constrained Critical Path (DCCP) that address the workflow scheduling problem in Infrastructure as a Service (IaaS) Cloud. We will conduct a simulation experiment of scientific workflow algorithms with both of the algorithms mentioned above. In addition, the simulation will be managed under deadline constraints on Infrastructure as a Service (IaaS) Clouds. The performance of the scientific workflow will be based on measuring Success Rate (SR) and Throughput. Finally, we expected the scheduling algorithms PDC and DCCP to be improve the Cloud resource usage high efficiency in the IaaS Cloud with efficient scheduling for scientific workflow.

Abstrak tesis yang dikemukakan kepada Universiti Putra Malaysia sebagai memenuhi keperluan Ijazah Sarjana Sains Komputer

July 2018

ABSTRAK

Banyak aplikasi saintifik yang penting boleh dinyatakan sebagai aliran kerja yang menerangkan hubungan antara tugas-tugas pengiraan individu dengan input serta output data mereka secara deklaratif. Ini membolehkan aliran kerja disesuaikan secara automatik untuk dilaksanakan dalam persekitaran yang berbeza-beza. Bagi aliran kerja yang kompleks, pengabstrakan turut membantu para saintis untuk menyatakan aliran kerja mereka pada tahap yang lebih tinggi tanpa menghiraukan perincian tentang bagaimana setiap kerja dinyatakan atau bagaimana data dipindahkan antara pekerjaan. Selain itu, aplikasi berskala besar dinyatakan sebagai aliran kerja saintifik yang sering dikelompokkan menjadi ensembel aliran kerja yang saling berkait (J. Vöckler, G. Juve & GB Berriman, 2011), (M. Malawski, G. Juve, E. Deelman & J. Nabrzyski, 2015). Biasanya, pengkomputeran Awan komersil dengan cepat menjadi platform sasaran untuk membuat pengiraan saintifik. Biasanya, caj perkhidmatan Awan komersil berdasarkan jumlah jam sumber (seperti CPU, jalur lebar rangkaian dan jumlah penstoran) yang digunakan. Model pembayaran ini dirujuk sebagai bayaran setiap penggunaan. Kefleksibelan wujud dalam model Awan anjal, sementara pengkomputeran yang berkuasa mengakibatkan penggunaan yang tidak cekap dan kos yang tinggi apabila penjadualan dan keputusan peruntukan tidak mencukupi dibuat. Masalah penjadualan aliran kerja saintifik pada Awan memerlukan pendekatan penjadualan alternatif dalam pemetaan tugas serta sumber di samping perlu menepati tarikh akhir dalam aliran kerja pengkomputeran Awan. Dalam projek ini, objektif kami mencapai prestasi yang lebih baik dari segi kadar kejayaan apabila dibandingkan dengan algoritma penjadualan

sedia ada. Dari segi penyelesaian masalah, kami secara cekap telah mereka bentuk penjadualan aliran kerja bagi sumber Awan yang diperuntukkan secara dinamik, di samping mengurangkan kerumitan pengiraan. Secara khusus, kami meningkatkan dua algoritma penjadualan iaitu Kekangan Tarikh Akhir Berkadaran (PDC) dan Laluan Kritikal Terkekang (DCCP) yang menangani masalah penjadualan aliran kerja dalam Infrastruktur sebagai Awan Perkhidmatan (IaaS). Kami akan menjalankan eksperimen simulasi algoritma aliran kerja saintifik terhadap kedua-dua algoritma yang disebutkan di atas. Selain itu, simulasi ini akan diuruskan di bawah kekangan tarikh akhir pada Infrastruktur sebagai Awan Perkhidmatan (IaaS). Prestasi aliran kerja saintifik akan didasarkan pada pengukuran Kadar Kejayaan (SR) dan Pemprosesan. Akhir sekali, kami menjangkakan algoritma penjadualan PDC dan DCCP dapat meningkatkan kecekapan tinggi penggunaan sumber Awan pada Awan IaaS dengan penjadualan yang efisien untuk aliran kerja saintifik.

ACKNOWLEDGEMENT

First, all thanks, praises, and gratitude to the omnipotent Allah, who has favored me with the valuable bounties during my life and has given me the physical and mental power that empowered me to be what I am, I would like to express my deep gratitude after God almighty in the accomplishment of this thesis to my supervisor **Dr. Masnida Hj Hussin** for her time, encouragement, exceptional support, guidance, and fruitful discussion. Big thanks to my thesis committee for their effort to review my work and provide me with their comments. I would also like to thank all the of my faculty members and colleagues for supporting , Especially my friend Haqi kalid who always support, inspire, and encourage me to complete this project, staff members of the Department of Computer Science and the School of Graduate Studies at university Putra Malaysia. Finally, I would eternally thankful to my parents for their encouragement, help and for their psychological and material support during these two years. I would like to dedicate this success to my dear mother who provided me moral and material support during this period without them, I would not be here today. My God offers wellbeing and bliss to all of them.

SUPERVISORY COMMITTEE

Thesis submitted to the Senate of University Putra Malaysia and has been accepted as fulfillment of the requirement for Master of Computer Science (distributed computing)

Masnida Hj Hussin, Ph.D

Senior lecturer

Department of Communication Technology & Network,

Faculty of Computer Science and Information Technology,

Universiti Putra Malaysia,

(Supervisor)

Date: ______________________________ Signature:_________________________

DECLARATION

Declaration by graduate student I hereby confirm that:

- This thesis is my original work;
- Quotations, illustrations and citations have been duly referenced;
- This thesis has not been submitted previously or concurrently for any other
- Degree at any other institutions;
- Intellectual property from the thesis and copyright of thesis are fully-owned by University Putra Malaysia, as according to the University Putra Malaysia.(Research) Rules 2012;
- Written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the University Putra Malaysia (Research) Rules 2012;
- There is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the University Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the University Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

LIST OF ABBREVIATIONS

PDC	Proportional Deadline Constrained
DCCP	Deadline Constrained Critical Path
IaaS	Infrastructure as a Service
SR	Success Rate
QoS	Quality of Service
NIST	National Institute of Standards and Technology
SaaS	Software as a Service
PaaS	Platform as a Service
EC2	Elastic Compute Cloud
S3	Simple Storage Service
SQS	Simple Queue Service
RDBMS	Relational Database Management Systems
WfMC	Workflow Management Coalition 1
Stdout	Standard Output
Stderr	Standard Error
SCEC	Southern California Earthquake Centre
PSHA	Probabilistic Seismic Hazard Analysis

SGTs	Strain Green Tensors
NCBI	National Centre for Biotechnology Information
BLAST	Basic Local Alignment Search Tools
BoTs	Bags of Tasks
VM	Virtual Machine
DET	Deadline Early Tree
MOHEFT	Multi-Objective Heterogeneous Earliest Finish Time
DDR	Deadline Distribution Ratio
DDS	Delay-Based Dynamic Scheduling
CEDA	A Cost-Effective Deadline Aware
DAG	Directed Acyclic Graph
EST	Earliest Start Time
DBL	Deadline Bottom Level
DTL	Deadline Top Level
TLS	Task Level Sets
ECT	Earliest Completion Time
EDF	Earliest Deadline First
FF	First Fit

BF	Best Fit
WF	Worst Fit
CIS	Cloud Information Service
GUI	Graphical User Interface
MFLOPS	Million Floating Point Operations per Second

TABLE OF CONTENTS

Page

ABSTRACT II
ABSTRAK IV
ACKNOWLEDGEMENT VI
SUPERVISORY COMMITTEE VII
DECLARATION VIII
LIST OF ABBREVIATIONS IX
TABLE OF CONTENTS XII
LIST OF TABLES XV
LIST OF FIGURES XVI

CHAPTER 1 INTRODUCTION 1
1.1 Overview 1
1.2 Problem Statement 3
1.3 Research Objectives 4
1.4 Research Scope 4
1.5 Thesis Organization 4

CHAPTER 2 LITERATURE REVIEW 5
2.1 Introduction 5
2.1.1 Cloud Computing 5
2.1.2 Definition 5
2.1.3 Service Models 6
2.1.4 Deployment Models 8

2.1.5 Cloud Vendors 11
2.2 Scientific Workflows 13
2.2.1 Workflow Traces 15
2.2.2 Scientific workflows Generator 16
2.2.3 Workflow Scheduling 21
2.2.4 Workflow Issues: 23
2.2.5 Task Scheduling 26
2.4 Related Work 28
2.5 Summary 36

CHAPTER 3 RESEARCH METHODOLOGY 37
3.1 Application and System Model 38
3.2 PDC & DCCP Algorithms 42
3.2.1 Pre-processing Step 43
3.2.2 Task Prioritization 45
3.2.3 Instance Selection in PDC 48
3.2.4 Instance Selection in DCCP 50
3.2.5 Time Complexity 52
3.3 Simulation Environment 52
3.3.1 CloudSim 53
3.4Simulation Design 55
3.6 Summary 61

CHAPTER 4 RESULT AND DISCUSSION 62
4.1 Evaluation 62
4.2 Performance Analysis 64

4.3 Configuration Scenario 65
4.4 Discussion 71
4.5 Summary 72

CHAPTER 5 CONCLUSION AND FUTURE WORK 73
5.1 Conclusion 73
5.2 Future Work 74

REFERENCES 75

LIST OF TABLES

Page

Table 2.1 : Different Types of Cloud Services 8

Table 3. 1: Ranks Values 49

Table 4. 1 Parameters of Data Centre Creation...67

Table 4. 2 Parameters of Virtual Machine. 67

Table 4. 3 Parameters of Cloudlet Creation 67

LIST OF FIGURES

Page

Figure2. 1: Cloud Services Model. 6
Figure2. 2: Private Cloud. 9
Figure2. 3: Community Cloud 10
Figure2. 4: Public Cloud 10
Figure2. 5: Hybrid Cloud. 11
Figure2. 6: Sample Workflow with Nine Tasks. 14
Figure2. 7: Sample Montage Workflow. 17
Figure2. 8: Sample Cyber Shake Workflow 18
Figure2. 9: Sample Epigenomics Workflow. 19
Figure2. 10: Sample LIGO Workflow. 20
Figure2. 11: Sample SIPHT Workflow. 21
Figure2. 12: Elements of Workflow Scheduling. 23
Figure2. 13: General View of Task Scheduling 28

Figure 3. 1 System Model 39
Figure 3. 2 Scheduling Workflow with PDC And DCCP. 43
Figure 3. 3 CloudSim Architecture. 55
Figure 3. 4 Screenshot of Data Centre loading. 56
Figure 3. 5 Screenshot of Virtual Machine loading. 57
Figure 3. 6 Screenshot of Cloudlet/Tasks loading 58
Figure 3. 7 Screenshot of Completed Cloud Environment. 58
Figure 3. 8 Screenshot of The Scheduling Algorithms Execution 59
Figure 3. 9 Framework of Our Simulation. 60

Figure4. 1 Success Rate for Five Different Datasets 66
Figure4. 2 Throughput of Cybershake and Montage 67
Figure4. 3Throughput of Epignomics, Ligo, and Sipht 68
Figure4. 4 Tasks Execution Time. 70
Figure4. 5 LIGO Systems Utilization. 71

CHAPTER 1

INTRODUCTION

1.1 Overview

Internet-based computing that uses shared software or resources as well as computing; the electricity grid being one such example. Cloud computing usually involves the provision of often virtualized and dynamically scalable resources over the Internet. It also introduces an entirely new consumption, supplement, and delivery model for Internet-based Information Technology (IT) services. The recent advent of large-scale scientific experiments and the power of big data have driven the rapid growth of cloud computing, which strives to directly deliver computing power to the end-user at anytime and anywhere; thus making it a highly attractive computing paradigm. Cloud computing is based on Quality of Service (QoS) user requirements. It enables users to deploy and access applications at competitive costs from anywhere in the world. The main objective of cloud computing is to design and power next-generation datacenters as a virtual services network consisting of application logic, hardware, database, user interface, etc. (Buyya. et, 2009). The most up-to-the-minute developments in parallel computing, grid computing, and distributed computing have resulted in the birth of cloud computing, the progression of which, is to deliver computing much like water and electricity are delivered to households nowadays: as a utility. Cloud computing—with the help of its application development and the constant expansion into cloud computing research—is set to become the computing paradigm and mainstream service of the future. More and more computing power—beyond that of the capability of a single machine—is required in a number of applications that consist of cooperative tasks. Some examples of these applications are Google's Map Reduce (Dean & Ghemawat, 2008) and Microsoft's Dryad (Isard, Budiu, Yu, Birrell, & Fetterly, 2007), which are big data processing workflows, as well as scientific workflows (Juve et al.,

2013) and multi-tier web service workflows (Mao & Humphrey, 2011). On the flip side are highly demanded common infrastructure or shared resources, and cloud computing services on the Internet. Additionally, users can access these facilities from specific service providers through a pay-per-use mechanism (Bardsiri & Hashemi, 2012). This way, the users will be able to access the massive processing capablities, operating systems, application development environments, servers, and storage space that would be otherwise unavailable elsewhere. The cloud enables users to access resources and scale them up or down on demand (AlHakami, Aldabbas, & Alwada'n, 2012).

Currently, workflows are being used to describe important scientific applications that declaratively relate input and output data with individual computational tasks. For the sake of convenience, workflow application development should employ a workflow definition that is abstract, and does not include details that are resource specific. This ensures the automatic adaption of these workflows to various environments. Since there is no need to worry about how data is transferred between jobs or how individual jobs are invoked, scientists working on complex workflows find that abstraction really helps them express higher-level workflows (Vöckler, Juve, & Berriman, 2011). Execution and composition of analysis on distributed resources are now more complex thanks to scientific workflow systems, which are now considered to be necessary tools of the trade. Nowadays, workflow system functionality are found to overlap due their flooding of the market. Therefore, an important requirement is that potential users are able to compare between various tool capabilities available (Deelman, Gannon, Shields, & Taylor, 2009). Besides that, ensembles, which are a group of interrelated workflows, often arise due to larger computational problems. Even though they share a similar structure, workflows in an ensemble vary when it comes to individual task sizes, input data, and number of tasks. Scientific workflow ensembles are required by numerous applications. One example is CyberShake, which generates seismic hazard maps using ensembles. CyberShake creates a hazard map by combining several hazard curves, with each hazard curve for a particular geographic location being first generated from an ensemble. A previous study in 2013 over 286 sites employed CyberShake to generate a set of hazard maps, and ended up requiring an ensemble of 2288 workflows. Montage also works similar to CyberShake; it generates a set of image mosaics from several workflows with different required parameters. These mosaics can be combined to make a single larger mosaic. In addition, the IaaS (Infrastructure-as-a-Service) cloud, a pay-per-use model, enables users to access on-demand resources. The scope of this work is an extension of our previous work involving the DCPP (Deadline Constrained Critical Path) and PDC (Proportional Deadline Constrained) algorithms. The DCPP and PDC algorithms are related and consist of a task assignment phase and a task prioritization phase, which classifies them as list-based scheduling algorithms.

1.2 Problem Statement

- Workflow scheduling in clouds determines the most suitable resources for allocation to certain tasks in order to achieve better task completion time (Tilak & Patil, 2012; Arabnejad, Bubendorfer & Ng, 2017). However, it suffers from heavy execution time and this leads to reduced throughput. This is because the workflow applications are dynamically arriving that explicitly effects execution time. Meanwhile, determining the type and amount of resource is a critical issue in resource provisioning (Arabnejad, Bubendorfer & Ng, 2017; Wu et al., 2015) .
- Hence, scientific workflow applications require an alternative scheduling approach for task-to-resource mapping, and at the same time, meet the deadlines for cloud computing workflows.

1.3 Research Objectives

- To develop effective scheduling for better performance in term of success rates while dynamically provisioning cloud resources for maximizing throughput.

1.4 Research Scope

- Re-simulate a previous paper (Arabnejad, Bubendorfer& Ng, 2017) that recommends the use of two algorithms with one or two workflows to schedule deadline-constrained scientific workflows on dynamically provisioned cloud resources.
- Focus on IaaS provider.
- Aims to improve workflow deadline.

1.5 Thesis Organization

This thesis is divided into five chapters. The following paragraphs provide a brief description of the remaining chapters of this thesis: Chapter 2 presents a comprehensive taxonomy and survey on scheduling algorithms for scientific workflows in IaaS clouds .Chapter 3 introduces different phases of the research methodology of this research. Chapter 4 describes the design, implementation and proves the concepts of the proposed workflow system. Chapters 5 present the conclusion together with thesis limitation and identify some areas for future work.

REFERENCES

- Abrishami, S., Naghibzadeh, M., & Epema, D. H. J. (2013). Deadline-constrained workflow scheduling algorithms for Infrastructure as a Service Clouds. *Future Generation Computer Systems*, *29*(1), 158–169. https://doi.org/10.1016/j.future.2012.05.004
- AlHakami, H., Aldabbas, H., & Alwada'n, T. (2012). Comparison Between Cloud and Grid Computing: Review Paper. *International Journal on Cloud Computing: Services and Architecture*, *2*(4), 1–21. https://doi.org/10.5121/ijccsa.2012.2401
- Arabnejad, V., Bubendorfer, K., & Ng, B. (2016). Budget distribution strategies for scientific workflow scheduling in commercial clouds. *Proceedings of the 2016 IEEE 12th International Conference on E-Science,EScience2016*,137146.https://doi.org/10.1109/eScience.2016.7870894
- Bardsiri, A., & Hashemi, S. (2012). A Review of Workflow Scheduling in Cloud Computing Environment. *International Journal of Computer Science and Management Research*, *1*(3), 348–351. Retrieved from http://goo.gl/0eK1Z
- Buyya, R. (2009). Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility. *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, CCGRID 2009*, *25*(6), 1. https://doi.org/10.1109/CCGRID.2009.97
- Cai, Z., Li, X., Ruiz, R., & Li, Q. (2017). A delay-based dynamic scheduling algorithm for bag-of-task workflows with stochastic task execution times in clouds. *Future Generation Computer Systems, 71,* 57–72. https://doi.org/10.1016/j.future.2017.01.020
- Calheiros, R. N., & Buyya, R. (2014). Meeting deadlines of scientific workflows in public clouds with tasks replication. *IEEE Transactions on Parallel and Distributed Systems*, *25*(7), 1787–1796. https://doi.org/10.1109/TPDS.2013.238

- De Coninck, E., Verbelen, T., Vankeirsbilck, B., Bohez, S., Simoens, P., & Dhoedt, B. (2016). Dynamic auto-scaling and scheduling of deadline constrained service workloads on IaaS clouds. *Journal of Systems and Software*, *118*, 101–114. https://doi.org/10.1016/j.jss.2016.05.011
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, *51*(1), 107. https://doi.org/10.1145/1327452.1327492
- Deelman, E., Gannon, D., Shields, M., & Taylor, I. (2009). Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems*, *25*(5), 528–540. https://doi.org/10.1016/j.future.2008.06.012
- Deldari, A., Naghibzadeh, M., & Abrishami, S. (2017). CCA: a deadline-constrained workflow scheduling algorithm for multicore resources on the cloud. *Journal of Supercomputing*, *73*(2), 756–781. https://doi.org/10.1007/s11227-016-1789-5
- Durillo, J. J., & Prodan, R. (2014). Multi-objective workflow scheduling in Amazon EC2. *Cluster Computing*, *17*(2), 169–189. https://doi.org/10.1007/s10586-013-0325-0
- Genez, T. A. L., Bittencourt, L. F., & Madeira, E. R. M. (2013). Using time discretization to schedule scientific workflows in multiple cloud providers. *IEEE International Conference on Cloud Computing, CLOUD*, 123–130. https://doi.org/10.1109/CLOUD.2013.141
- Haidri, R. A., Katti, C. P., & Saxena, P. C. (2017). Cost effective deadline aware scheduling strategy for workflow applications on virtual machines in cloud computing. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2017.10.009

- Isard, M., Budiu, M., Yu, Y., Birrell, A., & Fetterly, D. (2007). Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks. *ACM SIGOPS Operating Systems Review*, 59–72. https://doi.org/10.1145/1272998.1273005
- Juve, G., Chervenak, A., Deelman, E., Bharathi, S., Mehta, G., & Vahi, K. (2013). Characterizing and profiling scientific workflows. *Future Generation Computer Systems*, *29*(3), 682–692. https://doi.org/10.1016/j.future.2012.08.015
- Mao, M., & Humphrey, M. (2011). Auto-scaling to minimize cost and meet application deadlines in cloud workflows. *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '11*, 1. https://doi.org/10.1145/2063384.2063449
- Poola, D., Garg, S. K., Buyya, R., Yang, Y., & Ramamohanarao, K. (2014). Robust scheduling of scientific workflows with deadline and budget constraints in clouds. *Proceedings - International Conference on Advanced Information Networking and Applications, AINA*, 858–865. https://doi.org/10.1109/AINA.2014.105
- Sahni, J., & Vidyarthi, D. (2015). A Cost-Effective Deadline-Constrained Dynamic Scheduling Algorithm for Scientific Workflows in a Cloud Environment. *IEEE Transactions on Cloud Computing*, *7161*(c), 1–1. https://doi.org/10.1109/TCC.2015.2451649
- Singh, V., Gupta, I., & Jana, P. K. (2018). A novel cost-efficient approach for deadline-constrained workflow scheduling by dynamic provisioning of resources. *Future Generation Computer Systems*, *79*, 95–110. https://doi.org/10.1016/j.future.2017.09.054
- Topcuoglu, H., Hariri, S., & Wu, M. (2002). Performance-effective and low-complexity task scheduling for heterogeneous computing. *Parallel and Distributed Systems, …*, *13*(3), 260–274. https://doi.org/10.1109/71.993206
- Van Den Bossche, R., Vanmechelen, K., & Broeckhove, J. (2010). Cost-optimal

scheduling in hybrid IaaS clouds for deadline constrained workloads. *Proceedings - 2010 IEEE 3rd International Conference on Cloud Computing, CLOUD 2010*, 228–235. https://doi.org/10.1109/CLOUD.2010.58

- Vöckler, J., Juve, G., & Berriman, G. B. (2011). Experiences Using Cloud Computing for a Scientific Workflow Application. *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, SC'08, IEEE Press, Piscataway, NJ, USA, 2008*, *50*(12), 15–24.
- Yuan, Y., Li, X., Wang, Q., & Zhu, X. (2009). Deadline division-based heuristic for cost optimization in workflow scheduling. *Information Sciences*, *179*(15), 2562–2575. https://doi.org/10.1016/j.ins.2009.01.035
- Zheng, W., & Sakellariou, R. (2013). Budget-Deadline Constrained Workflow Planning for Admission Control. *Journal of Grid Computing*, *11*(4), 633–651. https://doi.org/10.1007/s10723-013-9257-4
- Arabnejad, Vahid, Kris Bubendorfer, and Bryan Ng. 2017. "Scheduling Deadline Constrained Scientific Workflows on Dynamically Provisioned Cloud Resources." *Future Generation Computer Systems* 75: 348–64. http://dx.doi.org/10.1016/j.future.2017.01.002.
- S.M. Hashemi, A.Kh. Bardsiri, "Cloud computing vs. grid computing," ARPN journal of systems and software, vol. 2, No 5, pp. 188-194, May 2012.
- H. Alhakami, H. Aldabbas, T. Alwada, "Comparison between cloud and grid computing: review paper," International journal on cloud computing: services and architecture (IJCCSA), vol. 2, No. 4, pp. 1-21, August 2012.
- Lokesh Kumar Arya, Amandeep Verma, " Workflow Scheduling Algorithms in Cloud Environment - A Survey," Institute of Electrical and Electronics Engineers (IEEE),978-1-4799-2291-8/14/$31.00,2014.
- Shun-Ren Yang and Yi-Bing Lin, "Modeling UMTS Discontinuous Re-

ception Mechanism",IEEE Transactions on Wireless Communications, vol. 4, no. 1, January 2005.

- Z. Zhu, G. Cao, R. Keralapura and A. Nucci, "Characterizing Data Services in a 3G Network: Usage, Mobility and Access Issues",IEEE International Conference on Communications, pp.1-6, June 2011.
- Larmo,"Usability of Instant Messaging over WCDMA", M.S.Thesis, Helsinki University of Technology,Finland,2005.
- W. Featherstone, J.P. Wallington and F. Schneider, ”TCP Performance Over UMTS Dedicated Channels with Finite RLC Buffer Size”,6th IEE International Conference on 3G and Beyond,pp.1-5,November 2005.
- L. Wang, A. Ukhanova and E. Belyaev, ”Power consumption analysis of constant bit rate data transmission over 3G mobile wireless networks”,11th International Conference on ITS Telecommunications,pp.217 -223, August 2011.
- D. Hilley, “Cloud computing: A taxonomy of platform and infrastructure-level offerings,” Georgia Institute of Technology, Tech. Rep. GIT-CERCS-09-13, April 2009.
- U. Yildiz, A. Guabtni, and A. H. Ngu, “Business versus scientific workflows: A comparative study,” in Proceedings of the Fourth World Conference on Services (SERVICES). IEEE, 2009, pp. 340–343.
- R. Andonov and S. Rajopadhye, “A sparse knapsack algo-tech-cuit and its synthesis,” in Proceedings of the IEEE International Conference on Application Specific Array Processors, 1994.
- G. Berriman, A. Laity, J. Good, J. Jacob, D. Katz, E. Deelman, G. Singh, M. Su, and T. Prince, “Montage: The architecture and scientific applications of a national-virtual observatory service for computing astronomical imagemosaics,”inProceedings of Earth Sciences Technology Conference, 2006.

- S. Callaghan, P. Maechling, E. Deelman, K. Vahi, G. Mehta, G. Juve, K. Milner, R. Graves, E. Field, D. Okaya, and T. Jordan. Reducing Time-to-Solution Using Distributed High-Throughput Mega-Workflows: Experiences from SCEC CyberShake. In Proceedings of the 4thIEEE International Conference one-Science (e-Science'08),2008
- R. Graves, T. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, D. Okaya, P. Small, and K. Vahi. CyberShake: A Physics-Based Seismic Hazard Model for Southern California. Pure and Applied Geophysics, 168(3-4):367–381, May 2010.
- E. Deelman, S. Callaghan, E. Field, H. Francoeur, R. Graves, N. Gupta, V. Gupta, T. H. Jordan, C. Kesselman, P. Maechling, J. Mehringer, G. Mehta, D. Okaya,K.Vahi and L.Zhao. Managing Large-Scale workflow execution from resource provisioning to provenance tracking: The CyberShake example. In Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing (e-Science 06), Dec. 2006.
- P. Maechling, E. Deelman, L. Zhao, R. Graves, G. Mehta, N. Gupta, J. Mehringer, C. Kesselman, S. Callaghan, D. Okaya, H. Francoeur, V. Gupta, Y. Cui,K.Vahi, T.Jordan,andE.Field. SCEC CyberShake Workflows Automating probabilisticseis michazard analysis calculations. InI.Taylor,E.Deelman,D.Gannon,andM.Shields,editors,Workflows for e-Science, pages 143–163. Springer, 2007.
- Southern California Earthquake Center (SCEC). http://www.scec.org.
- The TeraGrid Project. http://www.teragrid.org.
- Illumina. http://www.illumina.com.

- H.Li,J.Ruan,andR.Durbin. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research, 18(11):1851–1858, 2008.
- MAQ: Mapping and assembly with qualities. http://www.maq.sourceforge. net.
- A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Grsel, S. Kawamura, F. J. Raab, D. Shoemaker, L. Sievers, R. E. Spero, K. S. Thorne, R. E. Vogt, R. Weiss, S. E. Whitcomb, and M. E. Zucker. LIGO: The Laser Interferometer Gravitational-Wave Observatory. Science, 256(5055):325 –333, Apr. 1992.
- Laser Interferometer Gravitational Wave Observatory (LIGO). http://www.ligo. caltech.edu.
- D. A. Brown, P. R. Brady, A. Dietz, J. Cao, B. Johnson, and J. McNabb. A case study onthe use of workflow technologies for scientific analysis: Gravitational wave data analysis.In Workflows for e-Science, pages 39–59. Springer-Verlag, 2007.
- C. D. Capano. Searching for Gravitational Waves from Compact Binary CoalescenceUsing LIGO and Virgo Data. PhD thesis, Syracuse University, 2011.
- J. Livny, H. Teonadi, M. Livny, and M. K. Waldor. High-Throughput, Kingdom-Wide
  prediction and annotation of bacterial Non-Coding RNAs. PLoS ONE, 3(9):e3197, 2008
- DAGMan:DirectedAcyclicGraphManager.http://cs.wisc.edu/condor/
  dagman.
- B. Santhosh, P. K. A, and D. H. Manjaiah, “Comparative Study of Workflow Scheduling Algorithms in Cloud Computing,” pp. 31–37, 2014.
- J. Y. and R. Buyya, “A Taxonomy of Workflow Management Systems for Grid Computing,” J. Grid Comput., vol. 3, no. 3–4, pp. 171–200, 2005.

- J. O. Gutierrez-Garcia and K. M. Sim, "A family of heuristics for agent-based elastic cloud bag-of-tasks concurrent scheduling," Future Generation Computer Systems, vol. 29, no. 7, pp. 1682–1699, 2013.
- M. E. Frincu, S. Genaud, J. Gossa et al., "Comparing provisioning and scheduling strategies for workflows on clouds," in Proceedings of the Third International Workshop on Workflow Models, Systems, Services and Applications in the Cloud (CloudFlow). IEEE, 2013.
- E. Michon, J. Gossa, S. Genaud et al., "Free elasticity and free CPU power for scientific workloads on IaaS clouds," in Proceedings of the Eighteen IEEE International Conference on Parallel and Distributed Systems (ICPADS). IEEE, 2012, pp. 85–92.
- D. Villegas, A. Antoniou, S. M. Sadjadi, and A. Iosup, "An analysis of provisioning and allocation policies for infrastructure-as-a-service clouds," in Proceedings of the Twelfth IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). IEEE, 2012, pp. 612–619.
- "Elastic Compute Cloud EC2," http://aws.amazon.com/ec2, accessed: October 2015.
- "AmazonEC2SpotInstances,"https://aws.amazon.com/ec2/purchasingoptions/spot-instances/, accessed: October 2015.
- R. M. Singh, S. Paul, and A. Kumar, "Task Scheduling in Cloud Computing: Review," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 6, pp. 7940–7944, 2014
- a Radulescu and A. Van Gemund, "Fast and effective task scheduling in heterogeneous systems," Heterog. Comput. Work. 2000.(HCW 2000) Proceedings. 9th, pp. 229–238, 2000.

- J. Vöckler, G. Juve, and G. B. Berriman, "Experiences Using Cloud Computing for a Scientific Workflow Application," *Proc. 2008 ACM/IEEE Conf. Supercomput. SC'08, IEEE Press. Piscataway, NJ, USA, 2008*, vol. 50, no. 12, pp. 15–24, 2011.
- M. Malawski, G. Juve, E. Deelman, and J. Nabrzyski, "Algorithms for cost- and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds," *Futur. Gener. Comput. Syst.*, vol. 48, pp. 1–18, 2015.
- Mohapatra, S., Rekha, K.S., and Mohanty, S., A Comparison of Four Popular Heuristics for LoadBalancing of Virtual Machines in Cloud Computing. International Journal of Computer Applications, 68,(2013).
- Pakize, S.R., Khademi, S.M., and Gandomi, A., Comparison Of CloudSim, CloudAnalyst And CloudReports Simulator in Cloud Computing. International Journal of Computer Science And Network Solutions, 2: pp. 19-27,(2014).
- Wickremasinghe Bhathiya, N., C.R., and Rajkumar, B. CloudAnalyst: A CloudSim-based Visual Modeller for Analysing Cloud Computing Environments and Applications. . in International Conference on Advanced Information Networking and Applications (AINA). IEEE Computer Society, pp. 446-452, (2010).
- Wickremasinghe, B., CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments, (2009).
- S. Bharathi, A. Chervenak, E. Deelman, G. Mehta, M.-H. Su, K. Vahi, Characterization of scientific workflows, in: Third Workshop on Workflows in Support of Large-Scale Science, WORKS 2008, pp. 1–10.
- Y. Yuan, X. Li, Q. Wang, Y. Zhang, Bottom level based heuristic for workflow scheduling in grids, Chinese J. Comput.-Chinese Ed. 31 (2) (2008) 282–290.

- J. Yu, R. Buyya, C.K. Tham, Cost-based scheduling of scientific workflow applications on utility grids, in: First International Conference on e-Science and Grid Computing, 2005, pp. 140–147.
- Y.-K. Kwok, L. Ahmad, Dynamic critical-path scheduling: An effective technique for allocating task graphs to multiprocessors, IEEE Trans. Parallel Distrib. Syst. 7 (5) (1996) 506–521.
- G. Sih, E. Lee, A compile-time scheduling heuristic for interconnectionconstrained heterogeneous processor architectures, IEEE Trans. Parallel Distrib. Syst. 4 (2) (1993) 175–187.
- M.A. Khan, Scheduling for heterogeneous systems using constrained critical paths, Parallel Comput. 38 (4) (2012) 175–193.
- R. Chard, K. Bubendorfer, B. Ng, Network health and e-science in commercial clouds, Future Gener. Comput. Syst. 56 (2016) 595–604.
- M.R. Palankar, A. Iamnitchi, M. Ripeanu, S. Garfinkel, Amazon S3 for science grids: a viable solution? in: Proceedings of the 2008 International Workshop on Data-Aware Distributed Computing, ACM, 2008, pp. 55–64.
- S. Ostermann, A. Iosup, N. Yigitbasi, R. Prodan, T. Fahringer, D. Epema, A Performance Analysis of ec2 Cloud Computing Services for Scientific Computing, in: Lecture Notes of the Institute for Computer Sciences, vol. 34, Social-Informatics and Telecommunications Engineering, Springer, Berlin, Heidelberg, 2010, pp. 115–131.
- M. Mao, M. Humphrey, A performance study on the vm startup time in the cloud, in: Proceedings of the IEEE Fifth International Conference on Cloud Computing, CLOUD’12, IEEE Computer Society, Washington, DC, USA, 2012, pp. 423–430.

- G. Juve, A. Chervenak, E. Deelman, S. Bharathi, G. Mehta, K. Vahi, Characterizing and profiling scientific workflows, Future Gener. Comput. Syst. 29 (3) (2013) 682–692. special Section: Recent Developments in High Performance Computing and Security.
- I.J. Taylor, E. Deelman, D.B. Gannon, M. Shields, Workflows for e-Science: Scientific Workflows for Grids, Springer Publishing Company, Incorporated, 2014.
- H. Topcuoglu, S. Hariri, M.-Y. Wu, Performance-effective and low-complexity task scheduling for heterogeneous computing, IEEE Trans. Parallel Distrib. Syst.13 (3) (2002) 260–274.