UNIVERSITI PUTRA MALAYSIA

IMPATIENT JOB SCHEDULING UNDER CLOUD COMPUTING

NAWFAL A. MAHDI

FSKTM 2012 20

IMPATIENT JOB SCHEDULING UNDER CLOUD COMPUTING

NAWFAL A. MAHDI

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2012

IMPATIENT JOB SCHEDULING UNDER CLOUD COMPUTING

By

NAWFAL A. MAHDI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2012

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

IMPATIENT JOB SCHEDULING UNDER CLOUD COMPUTING

By

NAWFAL A. MAHDI

May 2012

Chairman: Associate Prof. Ali Mamat, PhD

Faculty: Computer Science and Information Technology

Cloud computing, a new concept, refers to a hosted computational environment that can provide elastic computation and storage services for users per demand. This paradigm arises with the huge growth in applications and data sizes. Many agencies, organizations and departments are responsible for time critical tasks and these tasks need to be completed as soon as possible. At the same time, these agencies also face IT problems because of the huge growth of applications, data and solution sizes. Many experts proposed that cloud computing is a solution to these problems such that each agency can execute its tasks via the cloud and expand their requirements based on the situation.

In this thesis, a study on the scheduling of impatient jobs in a cloud environment is presented. The study can be divided into three parts. The first part focuses on reviewing the previous immediate mode scheduling and adopting them on cloud paradigm. The limitations of those algorithms were addressed and this leads to the proposition of an algorithm that has the ability to map the impatient jobs to virtual machines near its input, output, application, or forth party. The challenge was how to consider the file sharing in on-the-fly way. This algorithm has been improved to take into account the input file sharing after analyzing the effect of this case on system performance. The

proposed algorithm is tested via simulation and real datasets. A mathematical model has also been drawn for this problem. It models the cloud computing infrastructure with the ability to inter-operate among the clouds. It assumes the virtual machine as the smallest computational unit in cloud computing. The results have shown better job mapping to resources from the perspective of throughput that is improved by 28%, while the execution time is improved by 29% and the amount of data transfer by 99%.

The second part concerns the bandwidth allocation in a virtualized environment for impatient jobs. In this part, we address the problem of immediate jobs that have huge amount of data and the ability to improve the resource allocation to meet the job deadlines. We modeled the problem as a set of events and proposed an algorithm that finds a proper virtual machine that can donate its bandwidth amount with full compliance to virtual machines deadlines and Quality of Service (QoS) constraints. The proposed algorithm was transplanted in an adopted scheduling algorithm and tested using simulation with a synthetic dataset. The simulation results showed better throughput with dynamic BW allocation by 21.1% than static allocation due to better resource allocation. Furthermore, the algorithm showed 10.07% better bandwidth utilization in a virtualized environment.

The third part looks at the negotiation process of the Service Level Agreement (SLA) under cloud computing. Previously proposed models in literature have many steps for confirmation which consume precious time for impatient jobs. We proposed a model for SLA negotiation which has the ability of offer-bid counter and rapid assigning in an immediate mode. System finite automata and control flow have been drawn. The proposed system is evaluated via simulation using synthetic data. From the results, the proposed algorithm improved the jobs throughput by reducing jobs waiting time by 81.5%, allowing more jobs to meet their deadlines.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia untuk memenuhi keperluan Ijazah Doktor Falsafah

PENJADUALAN KERJA SEGERA DALAM PENGKOMPUTERAN AWAN

Oleh

NAWFAL A. MAHDI

Mei 2012

Pengerusi: Profesor Madya Ali Mamat, PhD

Fakulti: Fakulti Sains Komputer dan Teknologi Maklumat

Pengkomputeran awan, satu konsep baharu, merujuk kepada persekitaran perkomputan berhos yang boleh menyediakan perhitungan anjal dan perkhidmatan penyimpanan untuk pengguna mengikut permintaan mereka. Paradigma ini muncul dengan pertumbuhan yang besar dalam aplikasi dan saiz data. Banyak agensi, jabatan dan organisasi yang bertanggungjawab untuk melaksanakan tugas-tugas kritikal dari segi masa dan tugas-tugas ini perlu disiapkan secepat mungkin. Pada masa yang sama, agensi-agensi ini juga mengalami masalah teknologi maklumat disebabkan pertumbuhan besar dalam saiz aplikasi, data dan penyelesaian. Ramai pakar telah mencadangkan pengkomputeran awan sebagai penyelesaian kepada masalah ini demikian hingga setiap agensi boleh melaksanakan tugas-tugasnya melalui awan dan memperluaskan keperluan mereka berdasarkan kepada situasi.

Dalam tesis ini, kajian tentang penjadualan kerja segera dalam persekitaran awan dibentangkan. Kajian ini dibahagikan kepada tiga bahagian. Bahagian pertama memfokuskan kepada mod penjadualan mod serta-merta yang terdahulu dan menerima mereka dalam paradigma awan. Keterbatasan algoritma tersebut diutarakan dan ini membawa kepada cadangan satu algoritma yang mempunyai kebolehan untuk memetakan kerja segera kepada mesin maya berdekatan dengan input, output, aplikasi

atau pun pihak keempatnya. Cabaran yang dialami adalah bagaimana perkongsian fail boleh boleh dilakukan secara langsung. Algoritma ini telah dipertingkatkan dengan mengambil kira perkongsian fail input setelah menganalisis kesannya ke atas prestasi sistem. Algoritma yang dicadangkan diuji menerusi simulasi dan set data sebenar. Satu model matematik telah juga dibina untuk masalah ini. Ia memodelkan infrastruktur komputer awan dengan kebolehan saling beroperasi di kalangan awan. Mesin maya dianggap sebagai unit perkomputan terkecil dalam pengkomputeran awan. Keputusan menunjukkan pemetaan kerja kepada sumber yang lebih baik daripada perspektif hasil kerja, masa perlaksanaan dan jumlah perpindahan data, masing-masing menunjukkan 28%, 29% dan 99% penambahbaikan.

Bahagian kedua menitikberatkan peruntukan jalur lebar dalam persekitaran maya untuk kerja segera. Dalam bahagian ini, kami mengutarakan masalah kerja segera yang mempunyai jumlah data yang besar dan mempunyai kebolehan menambahbaik peruntukan sumber untuk memenuhi tarikh akhir kerja. Kami memodelkan masalah ini sebagai satu set peristiwa dan mencadangkan satu algoritma yang dapat mencari mesin maya yang sesuai yang boleh menderma sejumlah jalur lebar yang patuh sepenuhnya kepada tarikh akhir mesin maya dan kekangan Kualiti Perkhidmatan (QoS). Masa konfigurasi mesin maya adalah dipertimbangkan untuk mencerminkan sistem sebenar. Algoritma yang dicadangkan telah digunakan dalam satu algoritma penjadualan dan diuji menggunakan simulasi dengan satu set data sintetik. Kepu- tusan simulasi menunjukkan peningkatan jumlah kerja sebanyak 21.1% hasil daripada peruntukan dinamik jalur lebar berbanding peruntukan statik yang menyebabkan peruntukan sumber yang lebih baik. Selanjutnya, algoritma ini memberikan peningkatan penggunaan jalur lebar sebanyak 10.07% dalam persekitaran maya.

Bahagian ketiga melihat kepada proses perundingan bagi Perjanjian Aras Khidmat (PAK) dalam pengkomputeran awan. Model yang dicadangkan sebelum ini dalam ke-

susateraan mempunyai banyak langkah untuk pengesahan yang memakan masa untuk kerja segera. Kami mencadangkan model perundingan PAK yang mempunyai kebolehan untuk menawarkan tawaran balas dan penugasan cepat dalam mod serta merta. Sistem automata terhingga dan aliran kawalan telah dibina. Sistem yang dicadangkan telah dinilai melalui simulasi menggunakan data sintetik. Keputusan menunjukkan bahawa algoritma yang dicadangkan itu menambahbaik jumlah kerja (throughput) dengan mengurangkan masa menunggu kerja segera sebanyak 81.5%, membenarkan lebih banyak kerja untuk memenuhi tarikh akhir mereka.

ACKNOWLEDGEMENTS

This thesis arose in part out of years of research that has been done since I came to the University Putra Malaysia. By that time, I have worked with a great number of people whose contribution in assorted ways to the research and the making of the thesis deserved special mention. It is a pleasure to convey my gratitude to them all in my humble acknowledgment.

In the first place I would like to record my gratitude to Associate Professor Dr. Ali Mamat for his supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences through out the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His truly scientist intuition has made him as a constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student, a researcher and a scientist want to be. I am indebted to him more than he knows.

I gratefully acknowledge Associate Professor Dr. Hamidah Ibrahim and Associate Professor Dr. Shamala K. Subramaniam for their advices, supervision, and crucial contribution, which made the backbone of this research and so to this thesis.

Where would I be without my family? My parents deserve special mention for their inseparable support and prayers. My Father, Professor Dr. Addul Sattar, in the first place is the person who put the fundament my learning character, showing me the joy of intellectual pursuit ever since I was a child. My Mother, is the one who sincerely raised me with her caring and gently love. Aseel, Zena, and Osama thanks for being supportive and caring siblings.

Words fail me to express my appreciation to my wonderful wife Reem whose dedication, love and persistent confidence in me, has taken the load off my shoulder. Her support and encouragement was in the end what made this dissertation possible.

APPROVAL

I certify that a thesis Examination Committee has met on 29^{th} May 2012 to conduct the final examination of Nawfal A. Mahdi on his thesis entitled **Impatient Jobs Scheduling Under Cloud Computing** in accordance with the Universities and University Collage Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The committee recommends that the student be awarded the Doctor of Philosophy. The members of the Examination Committee are as follows:

Abdul Azim Abd Ghani, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Mohamed Othman, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Rusli Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Putchong Uthayopas, PhD
Professor
High Performance Computing and Networking Center
Faculty of Engineering Kasetsart University
Kasetsart University
Thailand
(External)

SEOW HENG FONG, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Ali Bin Mamat, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hamidah Ibrahim, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Shamala A/P K Subramaniam, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

NAWFAL A. MAHDI

Date: 29 May 2012

TABLE OF CONTENTS