



UNIVERSITY OF SCIENCE AND TECHNOLOGY
COLLEGE OF GRADUATE STUDIES AND ACADEMIC
ADVANCEMENT

Faculty of Computer Science and Information Technology

Scheduling Jobs on Cloud Computing Using Cat Swarm

Optimization Algorithm

by

Monika Shohdy Wilson Abdelshahed

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements
in Partial Fulfillment of the Requirement for the Degree of Master of Science in
Computer Science

Supervisor

Dr. Adil yousif

November 2015

Abstract

Cloud computing is an emerging technology that provides all functions of traditional computing as services via the internet. Cloud computing can provide strong processing capacity to tackle huge amounts of requests from many users. Job scheduling problem is the keystone to cloud computing because there are many jobs have to be implemented efficiently. Several job scheduling mechanisms are available. However, still there is a need for new job scheduling methods to minimize the execution time and to optimize the scheduling process.

The goal of this research is to find the best allocation of a set of jobs on a set of resources. This study propose a new job scheduling mechanism based on Cat Swarm Optimization (CSO), using information of jobs (cloudlets) and resources (virtual machines) such as length of job, speed of resource and identifier for both. The proposed mechanism aims to find the best allocation of jobs to resources by imitates the behavior of cats and models into two modes. The scheduling function in the proposed job scheduling mechanism firstly creates a set of jobs and resources to generate the population by assigning the jobs to resources randomly and evaluates the population by using the fitness value which represents the execution time of jobs. Secondly the function used iterations to regenerate populations based on cats behavior to produce the best job schedule that gives the minimum execution time of jobs.

Two phases are considered in the research methodology process, implementing initial program using Java Language and conducting a complete simulation using CloudSim simulator. A number of experimentation scenarios using different numbers of jobs and resources are taken in the evaluation process. The proposed mechanism is compared with First Come First Servers (FCFS) algorithm and experimental results showed that the proposed mechanism has better performance than FCFS for reducing the execution time of jobs.

المستخلص

الحوسبة السحابية هي تكنولوجيا ناشئة التي توفر جميع وظائف الحوسبة التقليدية كخدمات عبر الإنترنت.

الحوسبة السحابية يمكن أن توفر القدرة على معالجة قوية لمعالجة كميات هائلة من الطلبات من العديد من المستخدمين. مشكلة جدولة الوظائف هي حجر الزاوية في الحوسبة السحابية لأن هناك العديد من الوظائف يجب أن تنفذ بكفاءة. العديد من آليات جدولة الوظائف متاحه ومع ذلك، لا تزال هناك حاجة إلى آليات جدولة وظائف جديدة لتقليل وقت التنفيذ وتحسين عملية الجدولة.

الهدف من هذا البحث هو العثور على أفضل توزيع لمجموعة من الوظائف على مجموعة من الموارد. إقترحت هذه الدراسة آلية جدولة وظائف جديدة على أساس خوارزمية Cat Swarm Optimization (CSO)، وذلك باستخدام معلومات من الوظائف (cloudlets) والموارد (الأجهزة الظاهرية) مثل طول الوظيفة، وسرعة المورد ومعرف لكلاهما. وتهدف الآلية المقترحة لجدولة الوظائف إلى إيجاد أفضل توزيع للوظائف من قبل تقليد سلوك القطط والنماذج في الوضعين. دالة الجدولة في آلية جدولة الوظائف المقترحة يخلق أولاً مجموعة من الوظائف والموارد لتوليد السكان عن طريق تعيين الوظائف إلى الموارد بشكل عشوائي وتقييم السكان باستخدام قيمة اللياقة البدنية التي تمثل وقت التنفيذ للوظائف. ثانياً تكرر إعادة توليد السكان على أساس سلوك القطط لإنتاج أفضل جدول الوظيفة الذي يعطي وقت تنفيذ أقل للوظائف.

مرحلتين مأخوذة في عين الاعتبار في عملية منهجية البحث ، وتنفيذ البرنامج الأولي باستخدام لغة جافا وإجراء محاكاة كاملة باستخدام جهاز محاكاة CloudSim. عدد من سيناريوهات التجارب تستخدم فيها أرقام مختلفة من الوظائف والموارد أُنخذت في عملية التقييم . تتم مقارنة الآلية المقترحة مع خوارزمية First Come First Servers (FCFS) وأظهرت النتائج التجريبية أن الآلية المقترحة لديها أداء أفضل من FCFS لتقليل وقت تنفيذ الوظائف.

1.1 Overview

This chapter introduces the research works, state the problem, research objectives, research significant and describe the thesis structure.

1.2 Problem Background

As the IT technologies are growing day by day, the need of computing and storage are rapidly increasing. Cloud Computing has become a widely accepted paradigm for high performance computing, because in Cloud Computing all type of IT facilities are provided to the users as a service. The services of the cloud are provided through the Internet. In Cloud Computing the term Cloud is used for the service provider, which holds all types of resources for storage, computing etc. Mainly three types of services are provided by the cloud. First is Infrastructure as a Service (IaaS), which provides cloud users the infrastructure for various purposes like the storage system and computation resources. Second is Platform as a Service (PaaS), which provides the platform to the clients so that they can make their applications on this platform. Third is Software as a Service (SaaS), which provides the software to the users and hence the user don't need to install the software on their own machines and they can use the software directly from the cloud. Cloud Computing provides many benefits: it results in cost savings because there is no need of initial installation of much resource; it provides scalability and flexibility, the users can increase or decrease the number of services as per requirement; maintenance cost is very less because all the resources are managed by the Cloud providers[1].

Job scheduling is one of the major activities performed in all the computing environments. Cloud computing is one the upcoming latest technology which is developing drastically. To efficiently increase the working of cloud computing environments, job scheduling is one the tasks performed in order to gain maximum profit.

The goal of scheduling algorithms in distributed systems is spreading the load on processors and maximizing their utilization while minimizing the total job execution time Job scheduling, one of the most famous optimization problems[2].

Job scheduling has been considered as one of crucial problems in cloud computing. An optimized scheduler would improve many factors in scheduling of jobs in a cloud system such as throughput and performance. Different Approaches have tried to solve this problem like Genetic algorithm, Ant colony optimization, Particle swarm optimization and etc.

1.3 Problem Statement

There are many jobs need to be distributed over resources in order to be completed efficiently to minimize the execution time. Suppose that $R = \{r_1, r_2, r_3 \dots r_c\}$ are c cloud resources and $J = \{j_1, j_2, j_3 \dots j_n\}$ are n independent jobs. The speed of each resource is expressed in form of MIPS (Million Instructions Per Second), and the length of each job is expressed in the form of number of instructions. So the problem is how to distribute these jobs provided on the resources available, such that an optimum execution time is achieved.

1.4 Research Objectives

The objective of this research is to minimize the execution time by proposing new job scheduling mechanism based on Cat Swarm Optimization (CSO) algorithm and to evaluate the proposed mechanism using CloudSim simulator.

1.5 Thesis Structure

This thesis contains six chapters, Chapter 2 Gives an overall idea of cloud computing, and scheduling algorithm on cloud computing. Chapter 3 Describes the Research Methodology. Chapter 4 Describes the Proposed Scheduling Mechanism. Chapter 5 Presents the Results and Discussion. Chapter 6 Provides the Conclusion and Recommendations.