ENERGY-EFFICIENT AND PERFORMANCE-AWARE SCHEDULING IN CLOUD ENVIRONMENT

SUMMARY REPORT

SUBMITTED FOR THE DEGREE OF PH.D.

Faculty of Engineering and Technology

M. M. University, Sadopur, Ambala

November 2016
Summary

1. Abstract

Cloud computing is a dynamic platform and has gained unquestionable reputation among users and developers. Cloud has attracted a large number of users, both Non-IT and IT professionals. Ever-increasing usage and dependence on cloud platforms has opened new thinking paradigm both in terms of opportunities and challenges. Enterprises have shifted their infrastructure from in-house to cloud platforms. Normal users keep their valuable data on clouds. Governments rely heavenly on cloud services for e-governance. E-commerce has seen a large growth after cloud platforms appeared on the horizon. As the promises are huge in cloud platforms so are the challenges. Clouds are an enterprise product but still in its infancy. Huge servers hosting millions and thousands of applications consume a lot of energy. It requires huge investments in power and infrastructure. Applications are hosted and running on clouds require to be scheduled optimally to harness most out of available resources and maximize the profits by minimizing the power consumptions. Resource management and power management while maintaining the QoS as specified in SLA documents are tradeoffs. It is difficult to obtain desired level of services under such a competitive environment.

Research community has understood the importance of rising challenges and opportunities and has devised several solutions of relevance and stature. Present research is bully on reducing the carbon emissions from huge servers and maximizing the utilization of resources and profits from investments.
In this research work, several proposals have been made to schedule the user’s jobs and application on the hosts such that it minimizes the makespan and maximize the utilization. VM migration has been explored and improved with selection of best VM to migrate from overloaded to lightly loaded hosts and switching of under-utilized hosts. A novel heuristic was proposed to harness the available resources for a set of independent tasks. PSO was modified using proposed heuristic to two-objective functionality by considering the cost-effectiveness as secondary objective. An effort has been made for efficient workflow scheduling by extending the existing dynamic scheduling algorithms. Further, a robust planning algorithm was proposed which first sequenced the workflow tasks into groups of independent tasks. Set of independent tasks were subjected to available resources by proposed heuristic. The results and improvements have been subjected to error analysis after a fair number of repetitions. All the proposals have been simulated in CloudSim toolkit and compared with existing best representative of their category.

2. Introduction
Cloud computing is a development of parallel, distributed and grid computing which provides computing power or potential as a service rather than a product. The user can avail resources, software, information and other devices as a metered service over a network through cloud computing.

Cloud computing utilizes services which runs on Internet that move computations from self-managed individual resources to on-demand pool of resources in centralized infrastructures. Cloud computing has removed the limitations that once existed in
traditional computing such as platform, hardware, software, architecture and geographical location etc. and emerged as a leading trend in high performance computing. In cloud computing, the end users contract with cloud vendors for customized Virtual Machines (VM) and interact with the VMs using only a console/browser through the Internet, with all the data and applications maintained on the remote servers accessible to end users from any device, anywhere and at any time. Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at distant locations. The Cloud services examples are web mail, online business applications, online file storage and social networking sites.

Cloud computing delivers infrastructure, platform, and software (applications) as services, which are provided to customers as subscription-based services under the pay-as-you-go model (customers pay for services on pay-per-use basis). In particular, Cloud resources need to be allocated not only to satisfy Quality of Service (QoS) requirements specified by users via Service Level Agreements (SLAs), but also to reduce energy usage.

Cloud computing is offering utility-oriented IT services to customers worldwide. Based on a pay-as-you-go model, it provides hosting of pervasive applications from consumer, scientific and business domains. But the data centers hosting Cloud applications consume lot of energy, adding to high operational costs and carbon footprints to the environment. So, Green Cloud computing solutions are required that can not only save energy for the environment but also reduce operational costs.
3. Objectives

1. Study and performance evaluation of various existing scheduling algorithm in terms of parameters relevant to cloud environment

2. Develop a heuristic approach to schedule the service requests on cloud resources without violating the negotiated Service Level Agreements.

3. Design an adaptive resource scheduling algorithm which makes a trade-off between economic effectiveness and performance.

4. To propose an energy-aware dynamic scheduling that adapts to variation in task types and load.

4. Overview of Contributions

- A review on task scheduling taxonomies was proposed. This work identifies a new taxonomy based on Goal Oriented Task Scheduling (GOTS) and Constrained Oriented Task Scheduling (COTS). A thorough review was presented with features, limitation and scope of latest work in task scheduling.

- A VM migration scheme based on minimizing variance of remaining loads on hosts after VM migration was proposed. Scheme converges on best VM Selection (Var_Sel) for migration. Var_Sel was compared and proved betters than several existing VM migration policies under various parameters besides minimizing VM migration and reduced instances of SLA violations. Also an extended version of this work was carried by roping in more performance
parameters. Var_Sel proved to be an energy efficient procedure in cloud computing.

- A variance aware scheduling scheme for workflows was proposed called Nearest Neighbor (NN). NN was then used to obtain hybrid variants of several exiting independent tasks scheduling schemes. This work proposed hybrid variants of Minimum Completion Time (MCT), NN_MaxMin, NN_MinMin. The hybrid versions of legacy scheduling schemes were improved in performance than their original variants.

- A cost aware hybrid meta-heuristic based scheduling was proposed. Considering PSO and Cost Aware PSO as base variants, NNCA_PSO was proposed. Comparing PSO, CA_PSO and NNCA_PSO on makespan, energy efficiency and utilization etc. NNCA_PSO proved better than existing hybrid PSO variant.

- A planning algorithm called Robust HEFT (RHEFT) was proposed for workflows by using an Interior Scheduling (IS) approach. Tasks in workflow were ranked and grouped into sets of independent tasks. Set of independent tasks were then subjected to scheduling using IS approach. The comparison of RHEFT with HEFT and DHEFT was performed and proved better than both in terms of energy, makespan and utilization etc.

5. Organization of Thesis

Research work of this thesis has been organized into 8 chapters. Chapter 1 introduces cloud computing and its challenges and applications. A brief background on research work in performance tuning of clouds is also presented. Chapter 2 presents a detailed
literature Review on scheduling, VM Migration and Energy Efficiency. Chapter 3 presents a new taxonomy on Task Scheduling. Chapter 4 presents VM migration approach proposed during this research work. Chapter 5 presents a Hybrid metaheuristic based scheduling schemes for set of independent tasks. Chapter 6 describes workflow scheduling implementations. Chapter 7 explores in Novel Robust Planning algorithm carried out during this research work. Finally Chapter 8 presents the conclusion and future aspects of this research work.