

Chapter 89

QoS Oriented Enhancement based on the Analysis of Dynamic Job Scheduling in HPC

Reshmi Raveendran

SBM College of Engineering and Technology, India

D. Shanthi Saravanan

PSNA College of Engineering and Technology, India

ABSTRACT

With the advent of High Performance Computing (HPC) in the large-scale parallel computational environment, better job scheduling and resource allocation techniques are required to deliver Quality of Service (QoS). Therefore, job scheduling on a large-scale parallel system has been studied to minimize the queue time, response time, and to maximize the overall system utilization. The objective of this paper is to touch upon the recent methods used for dynamic resource allocation across multiple computing nodes and the impact of scheduling algorithms. In addition, a quantitative approach which explains a trend line analysis on dynamic allocation for batch processors is depicted. Throughout the survey, the trends in research on dynamic allocation and parallel computing is identified, besides, highlights the potential areas for future research and development. This study proposes the design for an efficient dynamic scheduling algorithm based on the Quality-of-Service. The analysis provides a compelling research platform to optimize dynamic scheduling of jobs in HPC.

1. INTRODUCTION

High Performance Computing (HPC) refers to the practice of aggregating computing power to a single problem in ways that would deliver much higher performance than that derived through a typical desktop computer or workstation. Usually such problems pose significant challenges in sci-

ence, engineering and business applications. The purpose of having a high performance computing is to address these challenges by grouping of individual nodes that can work together to solve a larger problem much efficiently than any single computer can do.

A change is eminent in nature, and this seems certainly to be true, in the market of HPC. This

DOI: 10.4018/978-1-4666-9562-7.ch089

market has progressively undergone rapid change of vendors, architectures, technologies and the usage of systems. The performance of micro-processors has been increasing in an exponential way for the last four decades and so the scope of computer's processing ability does not have a fixed definition. A computer is considered to be high performance, if it uses multiple processors (tens, hundreds or even thousands) connected together by some network to exceed the performance of a single processor. Use of multiple processors to enhance computing performance is known as parallel computing. Parallel computing includes several computations carried out simultaneously. In parallel computation, the jobs must be scheduled in a better way to achieve better results.

The first systematic approach to scheduling problem was undertaken in the mid of 1950s. In the second half of the seventies the introduction of vector computer systems marked the beginning of modern supercomputing. And in the eighties, the integration of vector computation in conventional computing environment became more important. Thereafter Massive Parallel Processing (MPP), which utilizes a large number of processors to perform a set of coordinated computations in parallel, became successful. The price/performance ratios contributed to the success. Later it was replaced by microprocessor based Symmetric Multiprocessor Systems (SMS) which are tightly

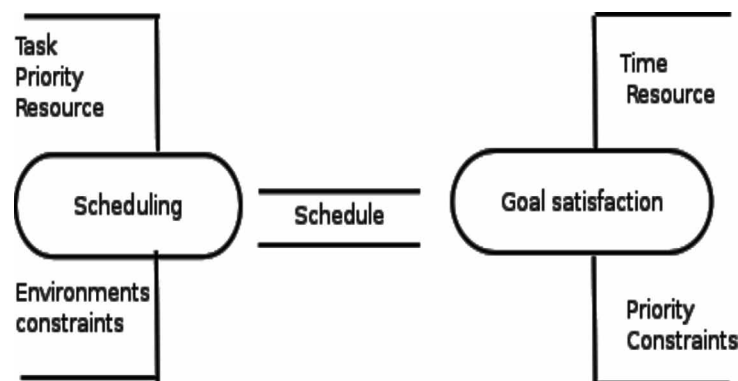
coupled multiprocessor systems with a pool of homogenous processors running independently on different data and with the capability of sharing resources.

High performance computation accelerates calculations much faster than a conventional processor, thereby speeding up many time consuming computations. The objective of High Performance Center is not only to minimize the make span or waiting time but also to maximize the user's satisfaction regarding the system behavior. Better job allocation gives better performance which leads to user satisfaction.

Scheduling the execution of parallel algorithms on parallel computer is an important and challenging area in current research. It is the job of a scheduler to determine when, where and how the given task should run and upon which the resource managers can be directed inappropriately. Parallel computers can be used to solve scheduling problems very fast. Although the importance of parallel scheduling algorithms has been widely recognized, only few results are obtained so far. Scheduling problem can be graphically depicted as in Figure 1.

Programming parallel applications is difficult and not worth the effort unless large performance gains can be realized. Scheduling is a key part of the workload management software which usually performs queuing, scheduling, monitor-

Figure 1. A scheduler



16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/qos-oriented-enhancement-based-on-the-analysis-of-dynamic-job-scheduling-in-hpc/142702

Related Content

Modeling-Centered Data Warehousing Learning: Methods, Concepts and Resources

Nenad Jukic and Boris Jukic (2012). *International Journal of Business Intelligence Research* (pp. 74-95).

www.irma-international.org/article/modeling-centered-data-warehousing-learning/74735

Artificial Intelligence Effects on Contracts and Contracting

Fatmah Baothman (2021). *Innovative and Agile Contracting for Digital Transformation and Industry 4.0* (pp. 149-160).

www.irma-international.org/chapter/artificial-intelligence-effects-on-contracts-and-contracting/272638

SCAX: Measuring Organizational Readiness to Embrace Supply Chain Analytics

Hamid Nemati and Antara Udiavar (2013). *International Journal of Business Intelligence Research* (pp. 19-38).

www.irma-international.org/article/scax-measuring-organizational-readiness-embrace/78274

Integrated Data Mining and Business Intelligence

S M Monzurur Rahman, Md Faisal Kabir and Muhammad Mushfiqur Rahman (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 1234-1253).

www.irma-international.org/chapter/integrated-data-mining-and-business-intelligence/107322

Price Discounts and Consumer Load-Shifting Behavior in the Smart Grid

Eeyad Al-Ahmadi and Murat Erkoc (2018). *International Journal of Business Analytics* (pp. 33-54).

www.irma-international.org/article/price-discounts-and-consumer-load-shifting-behavior-in-the-smart-grid/192167