Chapter 13
Resource Scheduling for Big Data on Cloud: Scheduling Resources

K. Indira Suthakar
Thiagarajar College of Engineering, India

M. K. Kavitha Devi
Thiagarajar College of Engineering, India

ABSTRACT

Cloud computing is based on the concepts of distributed computing, grid computing, utility computing and virtualization. It is a virtual pool of resources which are provided to users via Internet. It gives users virtually unlimited pay-per-use computing resources without the burden of managing the underlying infrastructure. Cloud computing service providers’ one of the goals is to use the resources efficiently and gain maximum profit. This leads to task scheduling as a core and challenging issue in cloud computing. This paper gives different scheduling strategies and algorithms in cloud computing.

INTRODUCTION

Cloud computing dates back to the 1960’s when John McCarthy opined that “computation may someday be organized as a public utility”. Amazon played a key role in cloud computing development by launching Amazon web service on utility basis in 2006. Before scheduling tasks on cloud computing, the characteristics of the cloud should be taken into account. Some of the characteristics of cloud include:

1. On-demand self service
2. Ubiquitous network access
3. Location independent resource pooling
4. Rapid elasticity
5. Pay per use

DOI: 10.4018/978-1-4666-9767-6.ch013
Millions of users share cloud resources by submitting their computing tasks to the cloud system. Scheduling these millions of tasks is a challenge for cloud computing environments. Different scheduling strategies are proposed in the cloud resource scheduling environment. These strategies consider different factors like cost matrix generated by using credit of tasks to be assigned to a particular resource, quality of Service (QoS) based meta-scheduler and Backfill strategy based lightweight virtual machine scheduler for dispatching jobs, QoS requirement heterogeneity of the cloud environment and workloads. Optimal resource allocation or task scheduling in the cloud should decide the optimal number of systems required in the cloud so that the total cost is minimized and the SLA is upheld. Cloud computing is highly dynamic, and hence, resource allocation problems have to be continuously addressed, as servers become available/non-available while at the same time the customer demand fluctuates. Thus this study focuses on scheduling algorithms in cloud environment considering above mentioned characteristics, challenges, and strategies.

**TASK SCHEDULING TYPES**

**Cloud Service Scheduling**

Cloud service scheduling is categorized at user level and system level. At user level scheduling deals with problems raised by service provision between providers and customers. The system level scheduling handles resource management within the datacenter. Datacenter consists of many physical machines. Millions of tasks from users are received; assignment of these tasks to physical machines is done at datacenter. This assignment or scheduling significantly impacts the performance of datacenter. In addition to system utilization, other requirements like QoS, SLA, resource sharing, fault tolerance, reliability, real-time satisfaction, etc. should be taken into consideration.

**User Level Scheduling**

Market-based and auction-based schedulers are suitable for regulating the supply and demand of cloud resources. Market-based resource allocation is effective in cloud computing environments where resources are virtualized and delivered to users as a service. Development of a pricing model using processor-sharing for clouds, the application of this pricing model to composite services with dependency consideration and the development of two sets of profit-driven scheduling algorithms are proposed in. Service provisioning in Clouds is based on Service Level Agreements (SLA). SLA represents a contract signed between the customer and the service provider stating the terms of the agreement including non-functional requirements of the service specified as Quality of Service (QoS), obligations, and penalties in case of agreement violations. Thus there is a need of scheduling strategies considering multiple SLA parameters and efficient allocation of resources. A novel scheduling heuristic considering multiple SLA parameters for deploying applications in cloud. The scheduler algorithm that allows re-provisioning of resources on the cloud in the event of failures. The focus of model is to provide fair deal to the users and consumers, enhanced quality of service as well as generation of optimal revenue. A novel cloud scheduling scheme uses SLA along with trust monitor to provide a faster scheduling of the over flooding user request with secure processing of the request. A novel approach of heuristic-based request scheduling at each server, in each of the geographically distributed data centers, to globally minimize the penalty charged to the cloud.
Related Content

Resource Allocation Policies in Cloud Computing Environment
[www.igi-global.com/chapter/resource-allocation-policies-in-cloud-computing-environment/174757?camid=4v1a](www.igi-global.com/chapter/resource-allocation-policies-in-cloud-computing-environment/174757?camid=4v1a)

IBM’s Watson Analytics for Health Care: A Miracle Made True
[www.igi-global.com/chapter/ibms-watson-analytics-for-health-care/164580?camid=4v1a](www.igi-global.com/chapter/ibms-watson-analytics-for-health-care/164580?camid=4v1a)

Forecasting the Trends in Cloud Computing and its Impact on Future IT Business

Challenges in Using Cloud Technology for Promoting Learner Autonomy in a Spanish Language Course: Reshaping Pedagogical Design
Jaya Kannan and Pilar Munday (2017). *Integration of Cloud Technologies in Digitally Networked Classrooms and Learning Communities* (pp. 261-290).
[www.igi-global.com/chapter/challenges-in-using-cloud-technology-for-promoting-learner-autonomy-in-a-spanish-language-course/172275?camid=4v1a](www.igi-global.com/chapter/challenges-in-using-cloud-technology-for-promoting-learner-autonomy-in-a-spanish-language-course/172275?camid=4v1a)