

Optimizing Cloud-Service Performance: Efficient Resource Provisioning via Optimal Workload Allocation

ABSTRACT:

Cloud computing is being widely accepted and utilized in the business world. From the perspective of businesses utilizing the cloud, it is critical to meet their customers' requirements by achieving service-level-objectives. Hence, the ability to accurately characterize and optimize cloud-service performance is of great importance. In this paper a stochastic multi-tenant framework is proposed to model the service of customer requests in a cloud infrastructure composed of heterogeneous virtual machines. Two cloudservice performance metrics are mathematically characterized, namely the percentile and the mean of the stochastic response time of a customer request, in closed form. Based upon the proposed multi-tenant framework, a workload allocation algorithm, termed maxmin-cloud algorithm, is then devised to optimize the performance of the cloud service. A rigorous optimality proof of the max-min-cloud algorithm is also given. Furthermore, the resource-provisioning problem in the cloud is also studied in light of the max-min-cloud algorithm. In particular, an efficient resource-provisioning strategy is proposed for serving dynamically arriving customer requests. These findings can be used by businesses to build a better understanding of how much virtual resource in the cloud they may need to meet customers' expectations subject to cost constraints.

SHIELD TECHNOLOGIES,
2232, 3RD FLOOR, 16TH B CROSS, YELAHANKA NEW TOWN, BANGALORE-64
Mail us: shieldtechnobl@gmail.com / manager@shieldtechno.com
Contact: 9972364704 / 8073744810