

Load balancing and server consolidation in cloud computing environments: a meta-study

ABSTRACT

The data-center is considered the heart of cloud computing. Recently, the growing demand for cloud computing services has caused a growing load on data centers. In terms of system behavior and workload, patterns of cloud computing are very dynamic; and that might serve to imbalance the load among data center resources. Eventually, some data-center resources could come to be over-loaded/under-loaded, which leads to an increase in energy consumption in addition to decreased functioning and wastage of resources. Just considering energy-efficiency (that can be attained efficiently by consolidate the servers) may not be enough for real applications because it may cause problems such as unbalanced load for each Physical Machine (PM). Therefore, this paper surveys published load balancing algorithms that achieved by server consolidation via a meta-analysis. Load balancing with server consolidation enriches the exploitation of resource utilization and can enhance Quality of Service (QoS) metrics, since data-centers and their applications are increasing exponentially. This meta-study, reviews the literature on load balancing and server consolidation and presents a ready reference taxonomy on the most efficient algorithms that achieve load balancing and server consolidation. This work attempts to present a taxonomy with a new classification for load balancing and server consolidation, such as migration overhead, hardware threshold, network traffic, and reliability.

Keyword: Cloud computing; Load balancing; Server consolidation; Energy efficiency; VM live migration