Statistical techniques for characterizing cloud workloads: a survey

Derdus Kenga, Vincent Omwenga and Patrick Ogao

Abstract

Cloud computing infrastructure is becoming indispensable in modern IT. Understanding the behavior and resource demands of cloud application workloads is key in data center capacity planning, cloud infrastructure testing, performance tuning and cloud computing research. Additionally, cloud providers want to ensure Quality of Service (QoS), reduce Service Level Agreement (SLA) violations and minimize energy consumption. To achieve this, cloud workload analysis is critical. However, scanty information is known about the characteristics of these workloads because cloud providers are not willing to share such information for confidentiality and business reasons. Besides, there is lack of documented techniques for workload characterization. To alleviate this situation, in this paper we perform the first meticulous study on statistical techniques that can be used to characterize cloud workloads. Through this review, we identify a statistical technique and its role in understanding cloud workload characteristics and its weakness. Throughout the review, we point out relevant examples where and how (particularly workload prediction), such techniques have been applied by pointing out to Google Cluster Trace (GCT) and Bitbrain's Grid Workload Archive Trace (GWA-T-12).