

Jordan University of Science and Technology

Autonomic Cloud Computing Resource Scaling

Authors: Ahmad Al-Dahoud, Ziad A. Al-Sharif, Luay Alawneh and Yaser Jararweh

Abstract: Cloud computing provides computing services on demand with an acceptable QoS. These services are provided and maintained by the Cloud Service Provider (CSP). Therefore, a Cloud User (CU) can focus more on business advancement and less on system maintenance, resource provisioning and service continuity as these became the duties of CSPs. A Cloud computing model gives CSPs the ability to retain multiple workloads on a single physical system. However, efficient resource provisioning and possible system fault management in the cloud can be a challenge to CSPs, which are bounded by SLAs and a very competitive market. Early fault detection provides room to recover from potential faults before impacting QoS. Moreover, current static techniques of fault management in computing systems are not satisfactory to safeguard the QoS required by CUs. Thus, new smart techniques are needed. This paper presents the ACCRS framework for cloud computing infrastructures to advance system's utilization level, reduce cost and power consumption and fulfil SLAs. The ACCRS framework employs state monitoring, planning, decision making, fault predication, fault detection, root cause analysis, and recovery actions to improve system's reliability, availability, and utilization level by scaling resources in response to changes in the cloud system state.