Abstract

In recent years, Infrastructure-as-a-Service (IaaS) clouds have become increasingly popular as a flexible and inexpensive platform for ad-hoc parallel data processing. Major players in cloud computing have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. However, currently used processing frameworks have been designed for static, homogeneous cluster systems and do not support the new features which distinguish the cloud platform. In this paper discussion is being done on the research project Nephele. Nephele is the first data processing framework to explicitly exploit the dynamic resource allocation offered by today’s IaaS clouds for both, task scheduling and execution. First performance results of Nephele are presented and its efficiency is compared with one of the well-known software, MapReduce. MapReduce is chosen for comparison since it is open source software and currently enjoys high popularity in the data processing community.
Keywords
Cloud computing  parallel data processing  Nephele  MapReduce