Università degli Studi di Salerno

Dipartimento di Scienze Aziendali – Management and Innovation Systems

---

Dottorato di Ricerca

Management and Information Technology

Curriculum in Informatica, Sistemi Informatici e Tecnologie del Software

XV Ciclo

Tesi di Dottorato

Parallel Genetic Algorithms in the Cloud

Tutor

Prof.ssa Filomena Ferrucci

Coordinatore

Prof. Andrea De Lucia

Candidato

Pasquale Salza

Matr. 8887600010

Anno Accademico 2015/2016

Abstract

Genetic Algorithms (GAs) are a metaheuristic search technique belonging to the class of Evolutionary Algorithms (EAs). They have been proven to be effective in addressing several problems in many fields but also suffer from scalability issues that may not let them find a valid application for real world problems. Thus, the aim of providing highly scalable GA-based solutions, together with the reduced costs of parallel architectures, motivate the research on Parallel Genetic Algorithms (PGAs).

Cloud computing may be a valid option for parallelisation, since there is no need of owning the physical hardware, which can be purchased from cloud providers, for the desired time, quantity and quality. There are different employable cloud technologies and approaches for this purpose, but they all introduce communication overhead. Thus, one might wonder if, and possibly when, specific approaches, environments and models show better performance than sequential versions in terms of execution time and resource usage.

This thesis investigates if and when GAs can scale in the cloud using specific approaches. Firstly, Hadoop MapReduce is exploited designing and developing an open source framework, i.e., elephant56, that reduces the effort in developing and speed up GAs using three parallel models. The performance of the framework is then evaluated through an empirical study. Secondly, software containers and message queues are employed to develop, deploy and execute PGAs in the cloud and the devised system is evaluated with an empirical study on a commercial cloud provider. Finally, cloud technologies are also explored for the parallelisation of other EAs, designing and developing cCube, a collaborative microservices architecture for machine learning problems.