

Lecturer:

Ioannis Konstantinou

Title

Automatic application elasticity in the clouds.

Abstract

Elasticity, i.e., the ability to scale up or down resources according to observed demand, is one of the most important requirements in today's cloud applications. Elasticity allows the commitment of just the right amount of resources based on application demand, performance and requirements resulting in optimal use of infrastructure and significant reductions in costs.

Both IaaS cloud providers and cloud application developers can benefit from elasticity. The former can provide the elasticity feature as a service through their infrastructure and optimize their resource utilization. The latter can utilize the elasticity mechanisms to adapt their applications to quickly respond to unpredictable workload variations while being charged only for the resources they need.

In this talk we present a system to provide automatic, multi-grained elasticity for generic applications in the cloud. We also present a decision making mechanism that can be used to achieve this in an adaptive, real time and automatic manner. Finally, we showcase the architectural approach of the CELAR project towards automated cloud elasticity depicting its main components and modules.