

# Machine Learning for Online Dynamic Cloud Configuration Optimization

The ever-increasing penetration of cloud-based services across the web and mobile platforms has resulted in highly demanding performance requirements from prevalent cloud computing architectures. Today's cloud platforms must be scalable, but also highly agile in order to accommodate rapidly changing usage patterns. They must be energy and cost efficient and must maintain the hardware health. At the same time, computation performance of the cluster must be sufficiently maintained. These conflicting requirements are often challenging to balance, but critical towards building and maintaining the next generation of smart clouds. This project will leverage state-of-the-art machine learning and optimization approaches for development of smart real-time deployable cloud configuration algorithms.

## Critical Performance Analysis of Cloud Computing Platforms

Cloud computing powers millions of mobile and web applications delivered to billions of consumers worldwide. Each of these apps places a unique set of performance requirements from the cloud computing platform. The architecture of a cloud-based platform is developed keeping generality in mind, in the interest of flexibility to host a wide variety of applications, scalability to handle user-base growth, practicality of operational costs and of course, sufficient computational performance. Although prevalent cloud platforms have demonstrated remarkable generality, there is scope to tailor the cloud configuration to be application-specific, thereby optimizing computational performance, operational costs, cluster health, energy efficiency (and in turn environmental impact) and availability. This project will investigate the relationship between such cloud configuration parameters and application-specific behavior and usage patterns. The goal will be to develop a principled algorithmic approach towards inferring the effect of application and user behavior on cloud parameters. This can potentially help in automated application-specific cloud configuration tailoring.

Contact Persons:

[Salman.Toor@it.uu.se](mailto:Salman.Toor@it.uu.se) [Prashant.Singh@it.uu.se](mailto:Prashant.Singh@it.uu.se)