Abstract

On-Demand-Computing technologies have reached a high level of development, but adopters underline core shortcomings related to QoS-guarantees, trustiness, and dependability for commercial applications and services. This is a significant problem for outsourcing mission critical applications in cloud infrastructures despite negotiation and availability of service level agreements (SLA).

In this talk we will discuss methods for SLA-aware resource management in order to provide QoS guarantees for high-performance applications. Resource outages threaten the guaranteed resource reservations, therefore methods for application transparent fault tolerance are developed. These methods are based on checkpointing and migration, if necessary over multiple administrative domains. Alternative compute resources are discovered using a peer-to-peer based search engine. Finally, methods for risk assessment and management help to reduce the risk of SLA violation by providing indicators for self-organising fault tolerance. Applications such as information management on cloud infrastructures benefit from the developed methods. The presented results are achieved in a number of national and international projects such as HPC4U, AssessGrid and BisGrid.