Dreamcloud

Dreamcloud

The essence of High Performance Computing (HPC) lies in sharing the large-scale hardware resources among the software applications using them. Efficient application allocation on the available compute resources is a key aspect of any HPC infrastructure functionality. The well-established HPC schedulers, such as a Portable Batch System (PBS), offer effective in terms of the offered scheduling features algorithms and techniques to manage the execution of computational tasks, i.e., in the HPC terminology – batch jobs, on distributed compute nodes. However, with the emergence of high-level e-Infrastructures, such as Grid and Cloud, the traditional cluster scheduling techniques have proved useful to a limited extent only. The main reason for this is that applications running on those infrastructures require a job scheduler to offer a much more extensive set of features in terms of scalability, fault tolerance, and usability, which the traditional, static (with regard to the application) scheduling techniques are not able to meet. The execution frameworks of new-generation parallel applications, such as Hadoop/MapReduce, require the underlying infrastructure scheduler to be more interactive with regard to the applications, in order to enable more intelligent allocation of resources within and also beyond a batch job, i.e., the property of dynamism.

The DreamCloud (Dynamic Resource Allocation in Embedded and High-Performance Computing) project started in September 2013, partially funded by the European Commission. The project aims to develop novel load balancing mechanisms that can be applied during runtime in a wide range of parallel and high performance computing systems, allowing for a fine-tuning of the trade-off between performance guarantees and system efficiency according to the application needs.

A number of techniques will be explored as the underlying allocation heuristics, including bio-inspired and market-inspired techniques and control-theoretic closed loop mechanisms that rely on the monitoring capabilities of the different kinds of systems. Such mechanisms will be organised in distinct types of cloud-like system software infrastructure that will manage the workload on different kinds of systems. Embedded Clouds will be used in systems with time-critical behaviour (such as the flight control in an aircraft), allowing for restricted load balancing and privileging strict performance guarantees. Micro Clouds will rely on novel extensions to operating systems and virtual machines, allowing for the dynamic migration of threads or full virtual machines from one core to another. Finally, High Performance Clouds will balance highly dynamic workloads, aiming for full utilisation of the underlying platform but at the same time providing performance guarantees to selected applications.

The total duration of the project is 3 years.

Project Partners:
- X/Open Company Ltd. (United Kingdom)
- Aicas GmbH (Germany)
- University of York (United Kingdom)
- Centre National de la Recherche Scientifique (France)
- Robert Bobsch GMBH (Germany)
- Rheon Media Ltd. (United Kingdom)
- HLRS (Germany)

Project website: www.dreamcloud-project.org

Contact: Dr. Alexey Cheptsov
Höchstleistungsrechenzentrum Universität Stuttgart
Nobelstraße 19, 70569 Stuttgart, Germany
Phone: +49-711-68560470
E-Mail: cheptsov@hlrs.de