Understanding and mastering dynamics in computing grids: processing moldable tasks with user-level overlay

Mościcki, J.T.

Citation for published version (APA):
Jakub T. Mościcki is a researcher and software engineer at CERN, Geneva, Switzerland. He obtained the MSc in Computer Science from the AGH University of Science and Technology in Kraków, Poland. He is a lead developer of the Ganga project and creator of the DIANE framework, which are used to support very large LHC user communities as well as users of multidisciplinary applications in theoretical physics, medical and radiation studies, bio-informatics, drug design, telecommunications. His research interests focus on scheduling and management of distributed and parallel applications, large-scale computing infrastructures such as grids, and various forms of High Throughput and High Performance Computing.

Thousands of scientific users witness every day inherent instabilities and bottlenecks of large-scale task processing systems: lost or incomplete jobs and hard-to-predict completion times. They are struggling to resubmit failed jobs and get consistent results. And it is always difficult to catch up with latest deployed software environments or system configurations. In addition, the users have often more than one system to deal with: they continue to use locally available computing power (a desktop PC, a nearby computing center, a small cluster next door) while exploiting global resources such as grids. On top of this, grids use a large variety of middleware stacks, which are customized in different ways by user communities. Quality of Service and usability are the two keywords probably most frequently echoed in the corridors of many “grid-enabled” research labs.

This PhD dissertation presents scientific research from the problem statement, system analysis, modeling and simulation, to validation through experimental results. It captures and characterizes complexity and dynamics of global task processing systems using as an example the largest scientific grid to date - the EGEE/EGI Grid. A task processing model developed in this work allows to rigorously explain why the late-binding method is superior to traditional task scheduling based on early binding. A study of statistical properties of task processing times is complemented by Monte Carlo simulation.

This book is also addressed to grid practitioners: developers and users. Presenting several successful application examples from diverse domains, it explains how heterogeneity and dynamics of global task processing systems may be addressed and mastered in a cost-effective way directly by the users. It describes a User-level Overlay, based on two software packages, Ganga and DIANE, which are ready to use with little or no customization for your application. Advanced resource selection strategies and scheduling approaches developed in this book may be reused in your environment.