

SORS: Optimization Issues in Data-Intensive Flows

Speaker: Anastasios Gounaris, an Assistant Professor at the Dept. of Informatics of the Aristotle University of Thessaloniki, Greece.

Abstract: Data-intensive flows are increasingly encountered in various settings, including business intelligence and scientific scenarios. As the data flows become more and more complex and operate in a highly dynamic environment, we argue that we need to resort to automated cost-based optimization solutions rather than relying on efficient designs by human experts. In this talk, we are going to discuss three complementary aspects of dataflow optimization. First, motivated by the fact that current approaches tend to employ multiple execution engines, we discuss state-of-art solutions to the problem of allocating flow activities to specific heterogeneous and interdependent execution engines while minimizing the flow execution cost. Second, we are going to discuss novel approaches to automatically define the execution order of the constituent tasks in a flow, thus relieving the designer from the burden of manually deciding the exact execution plan in full detail. Third, we narrow our focus on MapReduce-like systems and their descendants like Spark, and we discuss trade-offs between individual executor load and data transmission over the network during shuffling.

Short Bio: Anastasios Gounaris is an Assistant Professor at the Dept. of Informatics of the Aristotle University of Thessaloniki, Greece. Prior to that, he was a visiting lecturer with the University of Cyprus, and a researcher with the School of Computer Science of the University of Manchester and the Centre of Research and Technology Hellas CERTH. A. Gounaris received his PhD from the University of Manchester (UK) in 2005. He was also awarded an MPhil in Computation in 2002 by UMIST (UK) and a BSc in Electrical and Computer Engineering in 1999 by the Aristotle University of Thessaloniki. His research interests are in the area of autonomic, adaptive and wide-area data management, massive parallelism, flow and query optimization, data mining and resource scheduling. He has served as a program committee member in several international conferences and workshops. He has more than 50 publications in international journals and conferences. His work has received over 1000 citations. He is a member of ACM and IEEE. More details can be found at <http://delab.csd.auth.gr/~gounaris/>

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 15 jul 2024 - 08:46): <https://www.bsc.es/ca/research-and-development/research-seminars/sors-optimization-issues-data-intensive-flows>