Parallel and Distributed Data Management

Rizos Sakellariou¹, Salvatore Orlando¹, Josep Lluís Larriba-Pey²,
Srinivasan Parthasarathy², and Demetrios Zeinalipour-Yazti²

¹ Topic Chairs

² Members

The manipulation and handling of an ever increasing volume of data by current data-intensive applications requires novel techniques for efficient data management. Despite recent advances in every aspect of data management (storage, access, querying, analysis, mining), future applications are expected to scale to even higher degrees, not only in terms of volumes of data handled but also in terms of users and resources, often making use of multiple, pre-existing autonomous, distributed or heterogeneous resources. The notion of parallelism and concurrent execution at all levels remains a key element in achieving scalability and managing efficiently such data-intensive applications, but the changing nature of the underlying environments requires new solutions to cope with such changes. In this context, this topic sought papers in all aspects of data management (including databases and data-intensive applications) whose focus relates to some form of parallelism and concurrency.

Twelve papers discussing some of the aforementioned issues were submitted to this topic. Each paper was reviewed by at least three reviewers and a total of 49 reviews were obtained. Finally, after a severe debate, we decided to select two regular papers.

The two accepted papers discuss very interesting issues concerning distributed data mining and workflow enactment in dynamic environments.

In particular, the paper entitled “A Distributed Approach to Detect Outliers in Very Large Data Sets”, by F. Angiulli, S. Basta, S. Lodi, and C. Sartori, proposes a distributed approach addressing the problem of distance-based outlier detection in very large data sets. The authors discuss a solution that aims at reducing the execution time with respect to the centralized version, and is able to deal with distributed data sets. The second paper, entitled “Federated enactment of workflow patterns”, by G. Yaikhom, C. Sun Liew, L. Han, J. van Hemert, M. Atkinson, and A. Krause, considers a data mining application from the life sciences as a use case for discussing techniques that are applicable to large-scale workflow compositions, where the resources could change dynamically during the life-time of an application. In particular, they propose techniques for abstracting and cataloguing recurring workflow patterns, and for optimizing the mapping from workflow patterns to actual resources at runtime.

We would like to take the opportunity to thank the authors who submitted a contribution, as well as the Euro-Par Organizing Committee, and the referees who, with their highly useful comments and efforts, have made this conference and this topic possible.