

Topic 5

Parallel and Distributed Databases, Data Mining, and Knowledge Discovery

Bernhard Mitschang, David Skillicorn, Philippe Bonnet, and Domenico Talia
Topic Chairs

Database systems and knowledge discovery tools are two key technologies for storing, querying, and mining large volumes of data available today. The number of people and organizations that use data analysis techniques in their daily activities is increasing significantly and in several application domains the need of processing and mining large data sets is becoming a standard task. However, these intensive data consuming applications suffer from performance problems and single database sources. Introducing data distribution and parallel processing help to overcome resource bottlenecks and to achieve guaranteed throughput, quality of service, and system scalability. High-performance computers supported by high speed networks and intelligent data management middleware offer parallel and distributed databases and knowledge discovery systems a great opportunity to support cost-effective every day applications.

Data processing and knowledge discovery on large amounts of data can benefit from the use of parallel computers both to improve performance and quality of data selection. Implementation of data mining tools on high-performance parallel computers allows for analyzing massive databases in a reasonable time. Faster processing also means that users can experiment with more models to understand complex data. Furthermore, high performance makes it practical for users to analyze greater quantities of data. Distribution of data sources and data mining tasks is another paramount issue that the increasing decentralization of human activities and large availability of connection facilities are making more and more critical.

This year, 14 papers discussing some of those issues were submitted to this topic and this was more than in previous years. Each paper was reviewed by at least three reviewers and, finally, we were able to select 3 regular papers and 3 short ones. The accepted papers discuss very interesting issues such as database replication on PC clusters, pipelined queries in large databases, parallel clustering, join site selection methods, parallel suffix arrays, and sensor database systems.

We would like to take the opportunity of thanking the authors who submitted a contribution, as well as the Euro-Par Organizing Committee, and the referees with their highly useful comments, whose efforts have made this conference, and Topic 5 possible.