

Topic 5

Parallel and Distributed Databases, Data Mining and Knowledge Discovery

Domenico Talia, Hillol Kargupta, Patrick Valduriez, and Rui Camacho

Topic Chairs

To manage the very large amount of data available today, computer scientists are working on efficient systems, algorithms and applications that can handle and analyze very large databases. Intensive data consuming applications are running on very large databases (on data warehouses, on multimedia databases) with the task to extract information diamonds. Data mining is one of the key applications here. 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. Distributed architectures, cluster systems and P2P systems, supported by high performance networks and intelligent middleware offer parallel and distributed databases a great opportunity to support cost-effective everyday applications.

Data processing and knowledge discovery on large data sources can benefit from parallel and distributed computing both to improve performance and quality of results. Development 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 key issue that the increasing decentralization of human activities and large availability of connection facilities are making more and more critical.

This year, 9 papers discussing some of those issues were submitted to this topic. Each paper was reviewed by at least three reviewers and, finally, we were able to select 3 regular papers. The accepted papers discuss very interesting issues such as middleware for database replication, mining global association rules on Grids, and hierarchical aggregation in networked data management.

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.