Topic 5
Parallel and Distributed Databases, Data Mining and Knowledge Discovery

David Skillicorn, Abdelkader Hameurlain, Paul Watson, and Salvatore Orlando

Topic Chairs

This topic addresses the parallel storage and processing of data, ranging from traditional database systems, through OLTP and OLAP systems, to data mining. The primary motivation for parallelism in this setting is performance. Parallel database systems have been in use for several decades, but the use of parallelism for data mining is more recent. Interestingly, a new motivation for parallelism in handling data is becoming important – the need to compute results from multiple datasets without necessarily having full access to each of them. This idea of preserving some level of privacy for the contents of each part of a dataset generates some interesting problems and we look forward to papers in this area in years to come.

We received 15 submissions this year, and were able to accept 6, an acceptance rate of 40%.

The accepted papers cover a wide range of topics. Fagni, Perego and Silvestri address how to improve the performance of a concurrent web search cache, showing that a cache region dedicated to high-frequency queries improves overall performance. Lima, Mattoso and Valduriez use a new method for the virtual partitioning of OLAP queries on clusters to improve performance. Dash, Petrutiu and Scheuermann show how hierarchical clustering can exploit the property that clusters are not well-separated until the final stages of a clustering. Glimscher and Agrawal find a way to express both the E and M stages of the EM clustering technique as generalized reductions, which can then be implemented in parallel using a middleware system they have developed. Coppola, Pesciullesi, Ravazzolo and Zoccolo describe the design of a parallel customer relationship management system. Finally, di Giacomo, Martinez and Scott describe SearchPlus, a distributed system for bibliographic records.

We thank all of the referees whose work helped to select an interesting program.