

Topic 5

Parallel and Distributed Databases, Data Mining and Knowledge Discovery

Patrick Valduriez, Wolfgang Lehner, Domenico Talia, and Paul Watson

Topic Chairs

Managing and efficiently analysing the vast amounts of data produced by a huge variety of data sources is one of the big challenges in computer science. The development and implementation of algorithms and applications that can extract information diamonds from these ultra-large, and often distributed, databases is a key challenge for the design of future data management infrastructures. Today's data-intensive applications often suffer from performance problems and an inability to scale to high numbers of distributed data sources. Therefore, distributed and parallel databases have a key part to play in overcoming resource bottlenecks, achieving guaranteed quality of service and providing system scalability. The increased availability of distributed architectures, clusters, Grids and P2P systems, supported by high performance networks and intelligent middleware provides parallel and distributed databases and digital repositories with a great opportunity to cost-effectively support key everyday applications. Further, there is the prospect of data mining and knowledge discovery tools adding value to these vast new data resources by automatically extracting useful information from them.

We solicited submissions in either Experience and Application or System and Research in distributed and parallel data management. We received 15 paper submissions. We thank all authors for their submissions. All papers were reviewed by 4 reviewers. We selected the 4 following papers to be presented at EuroPar 2006, in one session: "Dynamic and Distributed Reconciliation in P2P-DHT Networks", "HyParSVM - A New Hybrid Parallel Software for Support Vector Machine Learning on SMP Clusters", "Supporting a Real-Time Distributed Intrusion Detection Application on GATES", and "A Semantic Web Service Based Approach to Supporting Augmented Provenance on the Grid".