

Topic 5: Parallel and Distributed Databases

Domenico Talia*, Josep Lluís Larriba-Pey*, Hillol Kargupta*,
and Esther Pacitti*

Advances in data management, including store, access, query, retrieval, and analysis, are inherent to current and future information systems. Today, accessing very large volumes of information is a reality. Tomorrow data intensive management systems will enable huge user communities to transparently access multiple pre-existing autonomous, distributed and heterogeneous resources (data, documents, images, and services). Existing data management solutions do not provide efficient techniques for exploiting and mining Tera-datasets available in clusters, P2P and Grid architectures. Parallel and distributed file systems, databases, datawarehouses, and digital libraries are a key element for achieving scalable, efficient systems that will both cost-effectively manage and extract knowledge from huge amounts of highly distributed and heterogeneous digital data repositories.

Intensive data oriented applications are running on very large databases (data warehouses, multimedia databases). However, these intensive data consuming applications suffer from performance problems and centralized 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. Distribution of data sources and data analysis tasks are key issues are becoming more and more critical due to the increasing decentralization of human activities and the large availability of network facilities.

This year, 13 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 6 regular papers. The accepted papers discuss very interesting issues about transaction processing and the application of parallelism to web and multimedia content search engines.

In particular, paper “Reducing Transaction Abort Rates with Prioritized Atomic Multicast Protocols” by E. Miedes, F. Muñoz and H. Decker, compares a set of classic total order message delivery protocols with their prioritized counterparts. Paper “Fault-Tolerant Partial Replication in Large-Scale Database Systems” by P. Sutra and M. Shapiro, investigates the use of a decentralized approach to committing transactions in replicated databases. Paper “A Search Engine Index for Multimedia Content”, by M. Marín, G. V. Gil-Costa and C. Bonacic, proposes a distributed index structure to support parallel query processing of multimedia content in search engines. Paper “Complex Queries

* Topic Chairs.

for Moving Object Databases in DHT-based Systems” by C. Hernández, A. Herández and M. Marín, proposes a distributed indexing method to efficiently support complex spatio/temporal queries. Paper “Exploiting Hybrid Parallelism in Web Search Engines” by C. Bonacic, M. Marín, C. García, M. Prieto and F. Tirado, proposes a hybrid technique devised to take advantage of the multithreading facilities provided by multicore nodes for search engines. Finally, paper “Scheduling Intersection Queries in Term Partitioned Inverted Files”, by M. Marín, C. Gómez-Panto, S. González and G. V. Gil-Costa, proposes and compares different scheduling algorithms for load balancing query traffic on distributed inverted files.

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 this topic possible.