Workshop on

Parallel and Distributed Spatial Data Structures (P-SPADS)(*)

Enrico Nardelli[1,2]

(1) Dip. di Matematica Pura ed Applicata, Univ. di L'Aquila, Via Vetoio, Località Coppito, 67100 L'Aquila, Italy, e-mail: nardelli@vxscaq.aquila.infn.it.
(2) Istituto di Analisi dei Sistemi ed Informatica, C.N.R., Viale Manzoni 30, 00185 Roma, Italy.

Abstract

The main objective of this workshop is to investigate parallel and distributed architectures, algorithms and data structures for the processing of spatial data with the aim of identifying advantages and disadvantages of the different combinations of hard and soft technologies for different classes of applications.

1. Rationale

Processing of spatial data is characterized by computations which usually involve data which are local (that is processing does not require that in a single point it is necessary to access information distributed all over the space under consideration). Therefore spatial data processing can be efficiently parallelized, but it is not at all clear which is the best combination of "hard" technologies (i.e. kind of processing units, memory architecture and topology of communication network) and "soft" technologies (i.e. data structures for representing data and algorithms for their manipulation), and which is the best way of solving trade-off issues in the space vs time context.

This strongly depends on the specific nature of processing, which for spatial data may be carried out mainly in two ways: (i) with simple image-processing-like computations (as it is done when two different maps of the same piece of land are intersected and all zones which are homogeneous with respect to both maps are identified); (ii) with complex mathematical models which carry out many computations on the spatial data (as it happens in the case of environmental planning). In the first case, infact, a distributed memory architecture, based on a message-passing approach, with many simple processing units and a simple communication networks, may work very well, while in the second one it seems that a shared memory architecture, with a few of powerful processing elements may provide better performances.

(*) Partially supported by project "Multidata" of CNR's research program "Progetto Finalizzato Sistemi Informatici e Calcolo Parallelo" and by projects "Amusing" and "Itu-land" of ESPRIT's research program "Basic Research Action".

An equally important aspect, which deserve consideration, is the kind of parallelism which is best suited to computing needs: sometimes it is important to parallelize main memory computation, in other cases it may be more efficient to provide parallel access to secondary storage; sometimes it is better to keep data in a single physical memory whose access is shared among processors, in still other cases better performances can be obtained by a suitably defined distribution and/or replication of data.

In the very last years the technological advances in the field of parallel and distributed computing have made it available computing architectures with high computational power at a cost between 50 K$ and 200 K$. But the identification of a cost effective solution for the end-users does not only depend on the availability of powerful systems. The full satisfaction of end-user needs requires the identification of an architecture able to support efficiently software requirements and able to provide good performances at a reasonable overall cost.

Therefore it is of the utmost importance, in the field of spatial data management based on a parallel/distributed computation approach, to provide different kinds of approach to parallel computing, with different degrees of mixing depending on the different user needs. This means to be able to define global solutions where kind of parallelism (processing units, secondary memory, ...), hardware architecture (shared memory, message passing, ...), system software (centralized/distributed system services, ...), and application software (replication/ distribution of data/ functions, ...) are carefully chosen and tuned according to given requirements.

2. Main issues

There are four main research areas which are central to the field of parallel and distributed spatial data structures:

(1) Theoretical Issues:
The main points here are (i) to define computation models suited for theoretical characterization of spatial data, (ii) to investigate both time- and space- complexity issues of spatial data processing, and (iii) to define performance evaluation models which are both formally sound and usable in practice and which allow to deal with trade-off

0-8186-6322-7/94 $04.00 © 1994 IEEE

issues in the space vs time context.

(2) Secondary Memory Representations:
In this area the issue at stake is to define and analyze performant secondary memory data structures and algorithms for parallel processing of spatial data, considering issues as distribution and replication policies of data on different physical devices, static and dynamic processing requirements, and so on. A key point is the identification of criteria for evaluating the relative merit of different proposals, based also an experiments with real life data and queries, but having a strong and sound theoretical basis.

(3) Architectures:
Here it is important (i) to identify and characterize through experiments those classes of parallel and/or distributed architectures which are best suited for processing spatial data in a parallel way and (ii) to identify critical issues affecting performances of parallel and distributed architectures in spatial data processing.

(4) Application Projects
Here the point is to be able to identify, by means of reference case studies, how to define globally effective solutions for classes of applications, where hardware architecture, system software, application software are carefully chosen and tuned according to requirements of the considered application class.

The objective of the P-SPADS workshop is to exchange ideas among researchers actively working or interested in the field, to discuss initiatives able to make it easier the starting of an eventual cooperation among research groups already working on the subject, and to possibly enrich and refine research directions through discussions with end-users.