

Data Management over the GRID: Heaven or Hell? (Panel Description)

Yannis Ioannidis
Department of Informatics and Telecommunications
University of Athens, Hellas (Greece)
yannis@di.uoa.gr

The last decade has seen unprecedented advances in network and distributed-system technologies, which have opened up the way for the construction of global-scale systems based on completely new conceptions of computation and sharing of resources. The dream of integrating unlimited levels of processing power, unlimited amounts of information, and an unlimited variety of services, and offering the entire package in a reliable and seamless fashion to widely distributed users is quickly becoming reality. Scientific applications will be among the first to take advantage of such environments, as the demands of current and future experimental studies for intensive computation and processing of very large amounts of information are pressing.

GRID technologies are at the forefront of these developments. While much has been written about computation in the GRID environment, data management has received less attention in the literature. Nevertheless, the GRID offers tremendous opportunities for large-scale distributed data management and at the same time poses major technical challenges in the area as well. The goal of this panel discussion is to identify these opportunities and challenges and examine whether the positive aspects of the GRID outweigh the negative ones or vice versa. In this direction, the panelists are called to answer some of the following questions:

- Is there new research to be done on data management over the GRID? Are there any new problems that arise from managing data over the GRID? For example, are there new problems with respect to security or heterogeneous information integration?
- Do classical problems require new solutions or do conventional approaches work well in the GRID environment? For example, how does one address issues of concurrency control, recovery, query processing and optimization?
- How does the GRID compare with other architectures, e.g., peer-to-peer, with respect to data management?
- Is the existing distributed computing infrastructure developed for the GRID, e.g., Condor, Globus, or Unicore, adequate for supporting the required data management functionality?
- Are there any particular difficulties when dealing with management of scientific data over the GRID compared to other kinds of data?
- How does workflow management interact with data management over the GRID?