Developing Scientific Database Applications in a Grid Environment

— Panel Discussion —

Abstract

Whether for astrophysics, bioinformatics or any other discipline, developing a scientific database application in a distributed computing environment remains a challenging task. Part of this challenge lies in having to make critical choices with respect to the underlying computing technologies and the roles they play in such projects. Emerging Grid technologies are adding further to the choices available to developers. Although the Grid promises significant advances in support of e-Science applications, much of the envisaged infrastructure, including data management, is still under development. The panel will discuss the role of data management in the e-Science process and the integration of database services into a Grid necessary to enable the realisation of this role.

1 Introduction

Grid computing is aimed at supporting dynamic integration of services in a distributed heterogeneous computing environment, allowing the formation of a "virtual organisation" which may include one or more organisations collaborating to achieve a particular objective. The Grid plays a key role in enabling the vision of e-Science, where scientists around the world can join large scale collaborative scientific enterprises through the Internet.

The Grid is generally described in terms of three layers. The data/computation layer “deals with the way that computational resources are allocated”, the information layer “deals with the way that information is represented, stored, accessed, shared and maintained”, the knowledge layer “is concerned with the way that knowledge is acquired, used, retrieved, published and maintained”, where data refers to uninterpreted bits and bytes, information is data and its meaning, and knowledge describes information applied to achieve a goal or solve a problem [1]. Of course, this is only a conceptual model and implementations of the Grid will vary in the degree to which they adhere to this model.

Foster et al. [2] propose the Open Grid Services Architecture (OGSA) for the realisation of the Grid. All required functionalities are abstracted in terms of Grid services which are described using the Web Services Description Language (WSDL). Hierarchies of such services are envisaged to support various levels of abstractions in the design of distributed systems. Each service specification may be supported by one or more service implementation in a Grid.

Grid applications initially considered were mostly using files for data storage. More recent work ([3],[5]), however, has also been looking at the integration of databases and the Grid. Paton et al. [3] have proposed an initial draft of a collection of database services that articulate with the service-based approach of OGSA. A comprehensive evaluation of data requirements for the Grid can be found in [4].

2 Discussion

The panel will discuss the role and current state of data management in different e-Science applications, the limitations and problems of software that is in use today to integrate databases into a Grid environment, and possible future developments, such as alternative software architectures, to address the existing problems.


References