Grid Data Management: Open Problems and New Issues

Esther Pacitti

Atlas Team Member, INRIA&LINA University of Nantes

Abstract. Initially developed for the scientific community, Grid computing is now gaining much interest in important areas such as enterprise information systems. This makes data management critical since the techniques must scale up while addressing the autonomy, dynamicity and heterogeneity of the data sources. In this paper, we discuss the main open problems and new issues related to Grid data management. We first recall the main principles behind data management in distributed systems and the basic techniques. Next we make precise the requirements for Grid data management and introduce the main techniques needed to address these requirements mainly by using using P2P techniques. To illustrate the use of these techniques that may be implemented in Grid infrastructures, we present a peer to peer multi-master replication approach for collaborative applications, and continous stream processing algorithms built over distributed hash tables using gossiping.