Towards a scalable scientific data grid model and services

ABSTRACT

Scientific Data Grid mostly deals with large computational problems. It provides geographically distributed resources for large-scale data-intensive applications that generate large scientific data sets. This required the scientist in modern scientific computing communities involved in managing massive amounts of a very large data collections that are geographically distributed. Research in the area of grid has given various ideas and solutions to address these requirements. However, nowadays the number of participants (scientists and institutions) that are involved in this kind of environment is increasing tremendously. This situation has lead to a problem of scalability. In order to overcome this problem we need a data grid model that can scale well with the increasing number of users. Peer-to-peer (P2P) is one of the architectures that is a promising scale and dynamism environment. In this paper, we present a P2P model for Scientific Data Grid that utilizes the P2P services to address the scalability problem. By using this model, we study and propose various decentralized discovery strategies that intend to address the problem of scalability. We also investigate the impact of data replication that addresses the data distribution and reliability problem for our Scientific Data Grid model on the propose discovery strategies. For the purpose of this study, we have developed and used our own data grid simulation written using PARSEC. We illustrate our P2P Scientific Data Grid model and our data grid simulation used in this study. We then analyze the performance of the discovery strategies with and without the existence of replication strategies relative to their success rates, bandwidth consumption and average number of hop.