International Journal of Engineering, Applied and Management Sciences Paradigms, Vol. 06, Issue 01, August 2013
ISSN (Online): 2320-6608
www.ijeam.com

Abstract Details

Title: Fault Tolerance Approach for Data Replication in Data Intensive Scientific Applications

Author: Sinu Nambiar, Prof. Rakesh Pandit, Prof. Sachin Patel

Abstract: Data Intensive scientific computing involves organizing, moving, visualizing, and analyzing massive amounts of data from around the world, as well as employing large-scale computation. There are systems that try to unify job and data management, but these are two different tasks to face in a Grid environment. Data replication and job scheduling are two different but complementary functions in Data Grids: one to minimize the total file access cost (thus total job execution time of all sites), and the other to minimize the MakeSpan (the maximum job completion time among all sites). The two main challenges: first, how to formulate a problem that incorporates not only data replication but also job scheduling, and which addresses both total access cost and maximum access cost; and second, how to find an efficient algorithm that, if it cannot find optimal solutions of minimizing total/maximum access cost, gives near-optimal solution for both objectives in the proposed system. The prime motive of data replication in intensive application is that to reduce the data file transfer time and bandwidth consumption. In order to minimize the MakeSpan we use Optimal MakeSpan algorithm and for polynomial time Nominal Distribution strategy has been used and it reduces the total data file access delay by at least half of that reduced is proposed in this paper. Both the algorithms are adaptive to the dynamic change of file access patterns in Data Grids. Since grid applications run in a very heterogeneous computing environment, fault tolerance is important in order to ensure their correct behaviour. A novel method to achieve maximum fault tolerance in the Grid environment system by using a fault recovery pool is proposed. Using GridSim, a popular distributed Grid simulator, it can be demonstrated that the technique significantly outperforms an existing popular file caching technique in Data Grids.

Keywords: MakeSpan, Optimal MakeSpan, Nominal Distribution, Data Grid, GridSim.