Name: Catherine Lynn Ruby
E-mail address: clruby@cse.buffalo.edu

Completed Thesis Title:

Three Data Grid Initiatives for the Advanced Computational Data Center

Thesis Committee:
          Chair: Dr. Russ Miller
          Member: Dr. Bina Ramamurthy

Abstract:

Advances in parallel computing have resulted in the growth of grid computing, a model that enables the integration of highly distributed and heterogeneous compute elements into a single, ubiquitous environment for solving computational problems. Often applied to large-scale efforts in fields such as structural biology, computational chemistry and high-energy physics, grid applications require infrastructure that supports the sharing and storage of increasing quantities of data. Grid initiatives are challenged to serve the growing demands of these data-intensive grid applications.

A data grid is a grid solution that integrates storage devices to provide a data service network for grid applications. Data grids house and serve data to the grid, and must be a robust, scalable, and seamless environment for both users and their computational jobs. The distribution of files across a network of storage devices can represent significant challenges to data management and delivery in a grid environment. A data grid is an essential component of a generic grid.

We will discuss three initiatives for the Advanced Computational Data Center Grid (ACDC Grid) that we have designed to address these issues and to provide a reliable data grid infrastructure. First, an examination of usage patterns and bottlenecks has resulted in the development of the Scenario Builder, a tool for studying potential storage device load scenarios that may result from data grid use. Next, a simulation tool, the Intelligent Migrator, has been designed for exploring methods of using historical data usage information to estimate future file utilization and replicate data before it is required for computational jobs. Finally, we present the current production ACDC Data Grid and the ways in which the Scenario Builder and Intelligent Migrator enable simulations of its design and performance. These three systems illustrate the ongoing efforts to deliver a robust and scalable data service to the ACDC Grid.