A Dataflow-Based Software Integration Model in Parallel and Distributed Computing and Applications

by Gang Cheng

Abstract of Dissertation

As an important component in parallel software engineering, software integration plays an indispensable role in building large-scale, multi-disciplinary parallel and distributed applications on high-performance computing systems. In this study we propose a software integration framework that is based on a dataflow programming paradigm and execution model for general-purpose HPDC applications. Advantages of this dataflow-based model are investigated and demonstrated. We study applications of this model in several specific, yet typical, areas that have general needs for parallel software integration and are well suited to this dataflow model. Drawing from extensive software and system integration case studies in those areas, we present the general requirements for a software integration framework in an HPDC application. These areas include:

1. integration of multiple parallel programming paradigms in a single unified environment;

2. integration of interactive data visualization and parallel computations on a heterogeneous distributed environment;

3. integration of networking navigation and parallel relational database management system (RDBMS); and
4. integration of World Wide Web computing and information processing.

By viewing different parallel programming paradigms as an essentially heterogeneous approach in mapping "real-world" problems to parallel systems, we discuss methodologies in integrating multiple programming models on a massively parallel system. Three applications in earth science, financial modeling, and computational electromagnetics are studied to demonstrate the integration of data-parallel and message-passing modules in the proposed multi-paradigm programming environment.

We develop interactive visualization systems on heterogeneous parallel and distributed computers. Using the dataflow model of an AVS testbed in two case studies in financial modeling and computational electromagnetics applications, we demonstrate a modular approach to couple parallel simulation modules into an interactive remote visualization environment.

Viewing the client-server model as an essentially restricted dataflow model, we describe the concept of integrating distributed computing in a World Wide Web (WWW) environment for data-intensive Internet applications. We examine approaches in integrating parallel relational databases and a Web-based networking navigation interface. Realization of the proposed dataflow integration model in Java and WWW is also described.