Performance Analysis and Modeling of a Computational Biology Code on CMP Clusters

Timothy Campbell
Texas A&M University College Station Texas 77840
Distributed Mentor Project
tcampbell106@neo.tamu.edu

(Research Advisor: Dr. Valerie Taylor and Dr. Xingfu Wu)

Abstract

The current trend in parallel computing systems is shifting towards cluster systems with CMPs (chip multiprocessors). Further, the CMPs are usually configured hierarchically (e.g., multiple CMPs compose a multi-chip module and multiple multi-chip modules compose a node) to compose a node of the parallel system. A major challenge to be addressed is efficient use of such cluster systems for large-scale scientific applications. In this research, we analyze the performance of a computational biology code MrBayes on two supercomputers: DataStar p655 at San Diego Supercomputer Center (SDSC) and Hydra at Texas A&M Supercomputing Facilities, and quantify the performance gap resulting from using different number of processors per node. We use PAIDE to instrument the source code of MrBayes to collect the performance, upload the performance data to Prophesy database, then use Prophesy system to model the performance online.

Basic Tasks:

1. Parallel Environments and two supercomputers (SDSC p655 and TAMU Hydra)
2. Application: a computational biology code MrBayes
3. Using PAIDE to instrument the source code to collect the performance
4. Scalability Analysis (with increasing number of processors and problem sizes)
5. Processor Partitioning Impacts (with different number of processors per node)
6. Automatically upload performance data to Prophesy Database
7. Online performance modeling using Prophesy system