Slade Scholar Program: Performance Testing and Software Additions to EnsembleMD Toolkit

Nikhil Shenoy
Computer Engineering Class of 2016
RUID: 145003806

Spring 2016

Abstract

Much of today's scientific experiments have become computation driven, and the amount of data that needs to be processed can be very large. Traditional data processing techniques relied on scheduling tasks directly through the job scheduler on a cluster, and then running the job using a Message Passing library to achieve the desired parallelism. However, many drawbacks were discovered with this approach, with the most prominent being extended wait times for jobs in the scheduling queue. This problem, coupled with scientists' increasing demand for fast, responsive simulations, has led to inefficient systems and a lack of progress.

This research seeks to eliminate the drawbacks of traditional processing methods by utilizing an existing Pilot-Job system, RADICAL-Pilot, to process problems in molecular dynamics. The resultant software, the EnsembleMD Toolkit, is designed to to efficiently manage and assign tasks to distributed resources using Pilots, and to abstract the details of resource management away from the user. This abstraction is designed for the molecular dynamics community, which works on simulating the N-body interactions of atoms and molecules.To run these simulations, the scientists must submit their tasks to distributed computing resources (DCRs) using software specifically designed for their experiments. There are many such softwares throughout the community, but since there is no standard that integrates all the functionalities into a single framework, progress is hindered due to a lack of portability and extensibility. The absence of this standard prevents scientists from adapting to and creating new simulation patterns, leading to the larger problem of unscalable software. EnsembleMD Toolkit allows the user to focus on the science of his simulation rather than the complexities of running the simulation. Specifically, this work analyzes the performance of EnsembleMD modules and details the development of additional modules for the API.

Research Team

- Nikhil Shenoy (nrs76@scarletmail.rutgers.edu)

- Prof. Shantenu Jha (shantenu.jha@rutgers.edu)

Project Plan

- Technical
    - Continue performance testing on EnsemblMD Toolkit
        * Current testing is pattern-based; may change to other components in the future
        * Identify bottlenecks and help fix them.
    - Develop new modules towards the Toolkit
        * Work on new version of Pipeline pattern (bag of pipelines)
    - Participate in release testing
    - Maintain functionality of example code in the documentation
    - Complete Slade Scholar program by writing a cohesive technical paper summarizing all work done over the two semesters of the program.

- Non-Technical
    - Devote at least 10 hours per week towards research
    - Maintain weekly journals documenting activity for the week.
    - 1-1 meetings to discuss progress