

Spare CPU Cycles as a Service – A New Cloud Computing Paradigm

Kireet Kokala, David Andersen, AOS CSc 8320, 8-24-2009
Dr. Yanqing Zhang, Georgia State University

ABSTRACT

We propose the merging of two technologies: the Folding@Home (FAH) distributed application and cloud computing. By relocating the utilization of spare CPU cycles from disparate and often unreliable distributed volunteer nodes, which now perform the protein folding calculations, to the large, secure, and dedicated server farms of large cloud computing vendors, we can demonstrate concrete increases in efficiency and throughput. As our proposed algorithm would still leverage only the currently unused spare computing power available from the many computers in the cloud, the costs for this model would still be incredibly reasonable and fit into the cloud's existing Software as a Service (SaS) business model.

1. INTRODUCTION

With the increasing demand for reusable and easily accessible data in domains like health-care, military fields, the commercial industry, and scientific research, the need for efficient means of computation is at an all time high. While many technologies, solutions, and implementations are available for such tasks, this paper explores cloud-based and distributed computing. Specifically, it draws on the reliability of both methods and theorizes on a hybrid approach for a real-time problem: protein folding [1-3].

Cloud computing has garnered a great deal of attention in recent years. The attraction of this computing resource model for many technology departments, large and small, is

threefold. First, by embracing the cloud as their information technology backbone, an IT department can eliminate the large initial capital expenditures required to set up a computing infrastructure and reduce the significant overhead costs of the department by offloading the software upgrades and hardware maintenance to a third party. Secondly, and perhaps more importantly, the department can optimize its budget by eliminating the need to overbuy technology to allow for future needs—the cloud provides instant scalability with no up-front costs. And finally, when the potential for the additional security and monitoring provided by the dedicated staff of the cloud vendor are added to the equation, the advantages of the cloud model are quite compelling. Simply put, the cloud offers the potential for not only a large capacity of spare computing power, but importantly a reliable source.

Contrast this model with the current implementation of the FAH distributed application, which not only suffers from a continually varying processing capacity, but loses efficiency because several calculations must be restarted or verified and the application must recover from errors because of the inherent unreliability of data being returned by the volunteer nodes.

The main focus of the paper is to analyze the search complexity of a cloud-based computing method vs. a distributed computing method and directly compare and contrast it with that of the hybrid simulation. An L-Kynureninase enzyme mutation conducted during 2003-2005 in the

University of Georgia biochemistry department will be used as the baseline for a real-time computation problem featuring multiple mutation sequences with appropriate depth searches, best, and worst case scenarios [3-4].

2. FUTURE POTENTIAL

The current proposal is a specific implementation or in a strict sense, a hybrid porting of functionality from a closed system to an open one. A future direction of such an ambitious project could be to learn from the architecture and implementation complexity of the system, and to establish an open framework that can be built cross-platforms, which can have several user interface, back-end functionality, and even decentralized management, which can in turn enable control and repository maintenance of FAH-type applications on cloud-distributed system hybrids [6,7].

3. REFERENCES

- [1] E. Stoneman, "Cloud-Distributed Computing," <http://geekswithblogs.net/EltonStoneman/archive/2008/12/02/cloud-distributed-computing.aspx>, 12-2-2008.
- [2] http://en.wikipedia.org/wiki/Cloud_computing, "Cloud Computing".
- [3] Sriganesh V. Koushik et Robert S. Phillips, "Cloning, Sequence, and Expression of Kynureninase from Pseudomonas fluorescens," *Archives of Biochemistry and Biophysics*, 1997, 344 301-308.
- [4] Gawandi, V., Liskey, D., Lima, S., and S. Phillips, "Reaction of beta-Benzoyl-

L-alanine with Kynureninase from Pseudomonas fluorescens: Detection of a New Reaction Intermediate and a Change in Rate-determining Step," *Biochemistry*, 2004, 43, 3230-7.

- [5] Hiroshi Takemiya, Yoshio Tanaka, Satoshi Sekiguchi, Shuji Ogata, Rajiv K. Kalia, Aiichiro Nakano, Priya Vashishta, "Sustainable adaptive grid supercomputing: multiscale simulation of semiconductor processing across the pacific," *Conference on High Performance Networking and Computing archive Proceedings of the ACM*, No. 106, 2006.
- [6] Chuck Pheatt, "An easy to use distributed computing framework," *Technical Symposium on Computer Science Education archive Proceedings of the 38th SIGCSE technical symposium on Computer science education*, Pages 571-575, 2007.
- [7] Jerome Verbeke, Neelakanth Nadgir, Greg Ruetsch, Ilya Sharapov, "Framework for Peer-to-Peer Distributed Computing in a Heterogeneous, Decentralized Environment," Lecture Notes In Computer Science; Vol. 2536. *Proceedings of the Third International Workshop on Grid Computing*, Pages 1-12, 2002.