

IMPLEMENTATION METHODS FOR ESTIMATING HAPLOTYPES WITH GRID COMPUTING TECHNOLOGY

Zafri M. , Sanjaya K.C. , Phua K.K.

Institute for Research in Molecular Medicine (INFORMM) , Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia.

ABSTRACT

Introduction: PHASE is a software for haplotype reconstruction and recombination rate estimation from population data and implements the methods for estimating haplotypes from population genotype data. This software require enormous computing power due to the complexity algorithms used. Recent advancement in grid computing and multi-core processing technology has resulted in the creating of virtual supercomputers (that are capable of massive parallel computations).

Method: PHASE version 2.0 was downloaded from PHASE's homepage (<http://stephenslab.uchicago.edu/software.html>), and installed on 10 Apple iMac computers (20 CPUs) running Mac OS 10.5 using Apple™ Xserve as a controller. A PHP web-based application was developed to enable easy job submission and result retrieval. Haplotype estimation jobs for PHASE were submitted using a web browser running either from Windows®, Linux® or Macintosh™ OS 10.5 operating systems. Jobs consisting of 20 input files of genomic loci (SNPs) were submitted to the Xgrid via the controller and the job completion time for each run were recorded and compared.

Results: The job completion time for estimating 20 input files and returning the results to the controller using XGrid (20 CPUs) took 4,980 seconds compared with 58,800 seconds using a single computer (2 CPUs), representing 91.5% reduction in computing time.

INTRODUCTION

- Grid computing is applying the resources of many computers in a network to a single problem at the same time. This procedure can enhance computing performance by connecting a large number of parallel computers or PC clusters with high performance networks. Therefore, grid system can help speed up the experimentation time.
- PHASE can analyze bi-allelic (SNPs), microsatellites, or multi-allelic locus (tri-allelic SNPs, or HLA allele). PHASE provides two main models for haplotype reconstruction.

OBJECTIVES

The main objective of this study was to investigate the use of grid technology to accelerate the multiple halotypes estimation using PHASE software.

METHODOLOGY

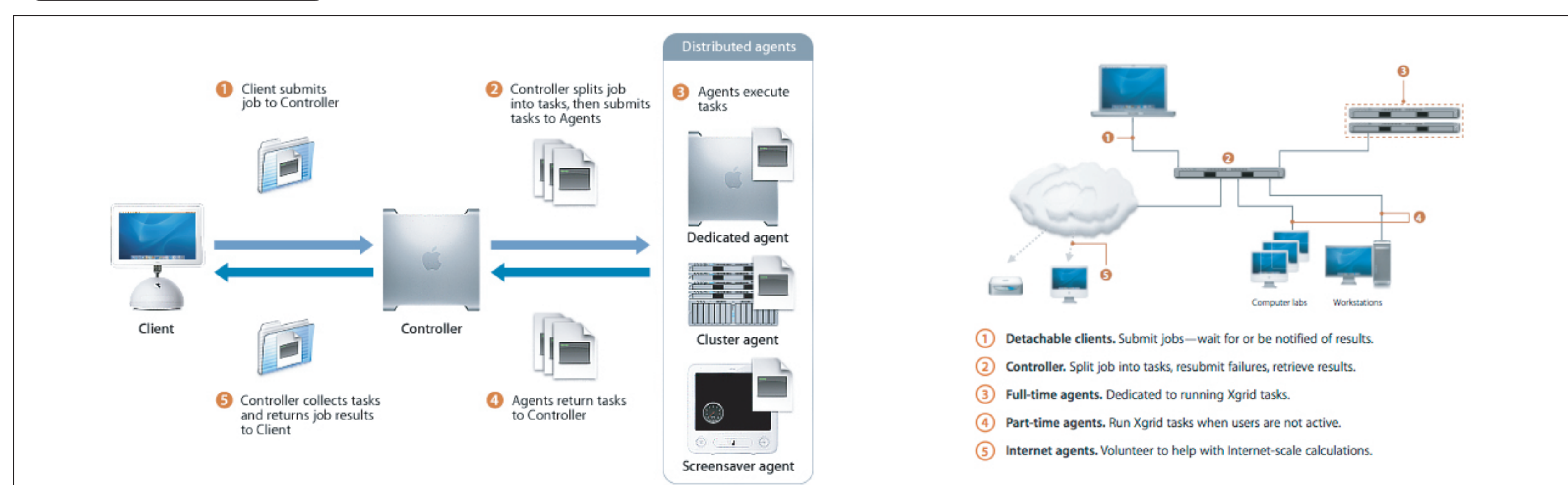
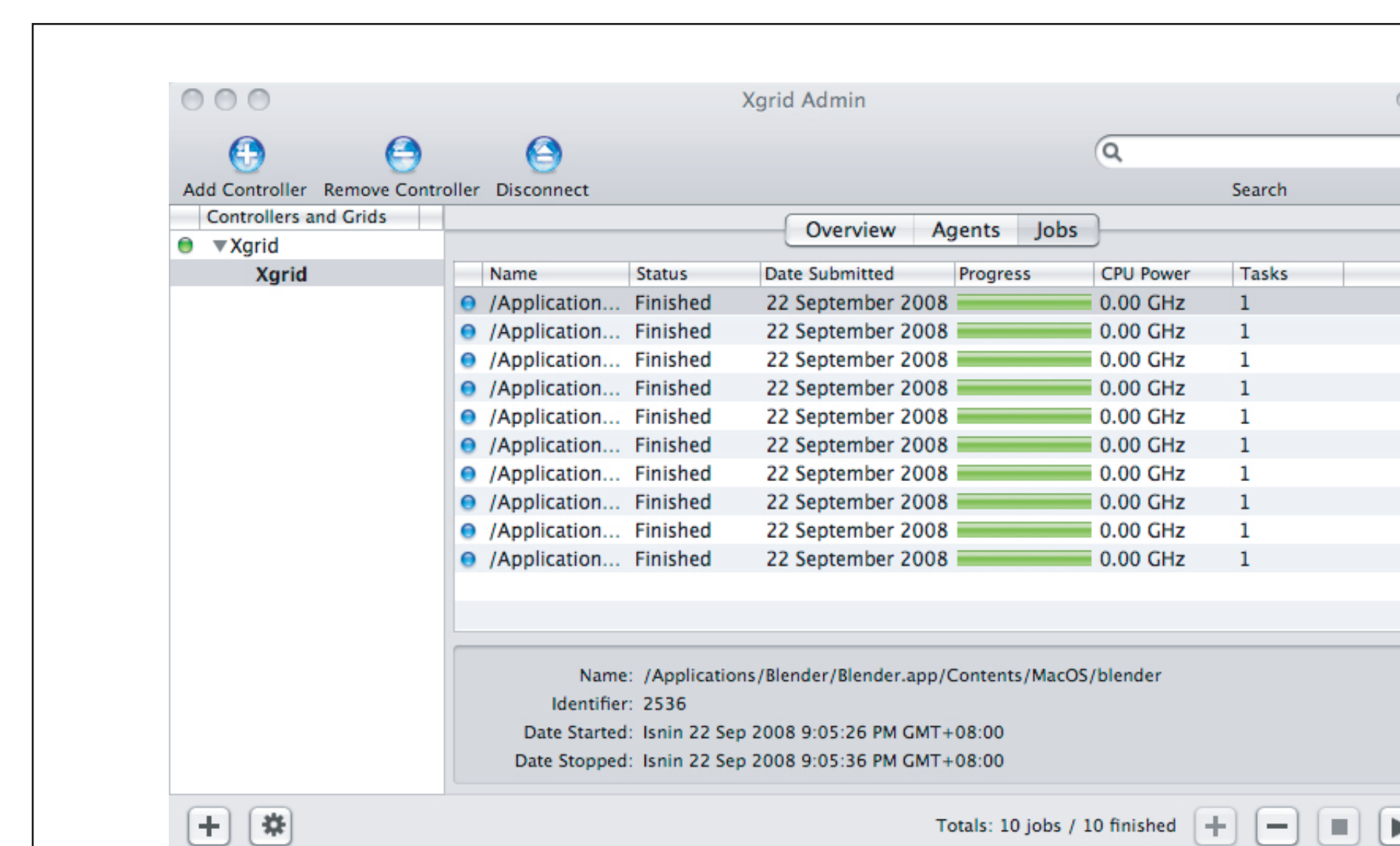


Figure 1: Schematic diagram showing the set up of the Xgrid.

RESULTS



Xgrid admin panel showing all submitted haplotyping estimation jobs were completed.

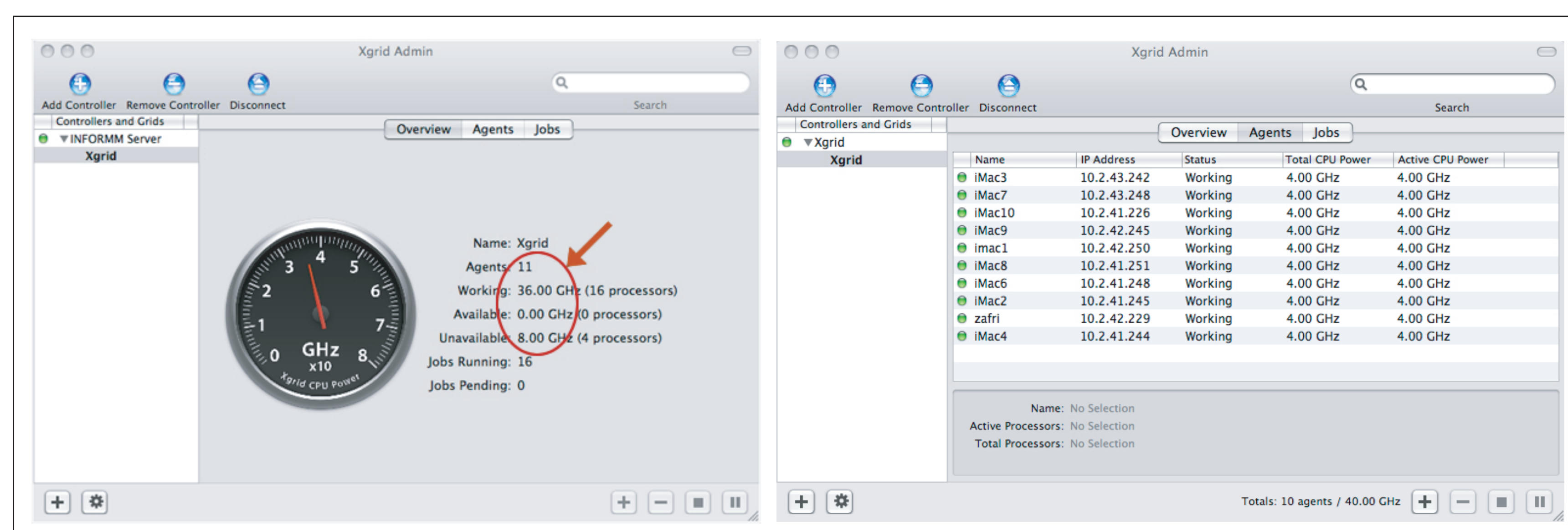


Figure 2: Xgrid admin panel indicating the working CPU speed of the Grid. Instead of 1 CPU (2 GHz), the grid above showed an increase in CPU speed (36 GHz) depending on the number of agents.

Haplotypes Estimation (20 SNPs) With Xgrid

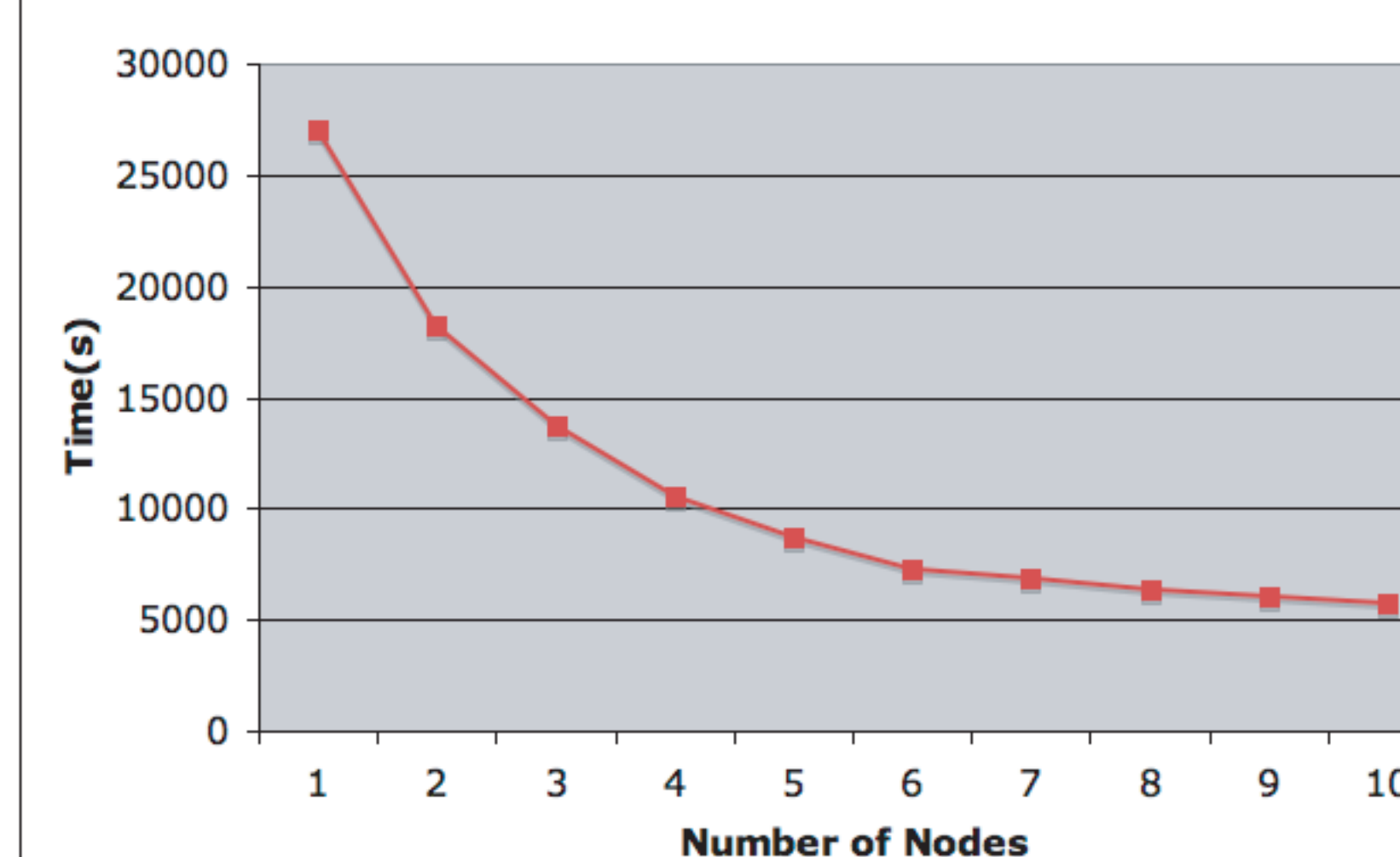


Chart showing decreasing time taken by PHASE to complete haplotyping estimation of 20 SNPs using Xgrid with 10 nodes.

CONCLUSIONS

- Xgrid technology can be used to create a low-cost supercomputer to accelerate CPU intensive task, save time and cost. Job submission and retrieval of results can be simplified using our PHP web application.

REFERENCE

- <http://stephenslab.uchicago.edu/software.html>, The PHASE website.
- <http://www.apple.com/acg/>, Apple Computer Inc. website
- <http://www.ibm.com/grid/>, The IBM Grid website. Grid Computing: Past, Present and Future, An Innovation Perspective.

ACKNOWLEDGEMENTS

- USM Grid Computing Cluster Grant (RU Grant No: 1001/PPTM/817006)
- Institute for Research In Molecular Medicine, USM.
- Fabio P. Valente, School of Bioinformatics, University of Paris Diderot, Paris, France.
- Constance S.N. Chew, School of Pathology and Laboratory Medicine, University of Western Australia.