

PROJECT INTRODUCTION

Objectives

To create an application that would enable the running of multiple BLAST runs utilizing distributed resources on the Grid.

Project Investigator / Manager

Dr. Arun Krishnan
Bioinformatics Institute
arun@bii.a-star.edu.sg

Period of Project

Mar 2002 – Dec 2003

Website

<http://gridblast.bii.a-star.edu.sg>

Abstract

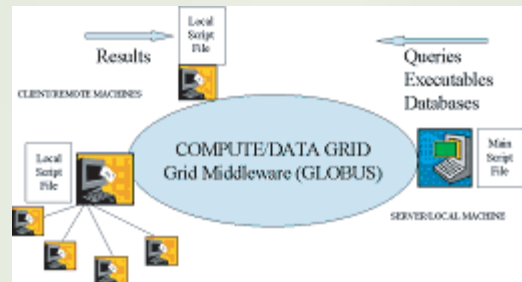
Deals with creating a grid-enabled, high-throughput, standalone version of a bioinformatics application, BLAST, using Globus as the Grid middleware. BLAST is a sequence alignment and search technique that is embarrassingly parallel in nature and thus amenable to adaptation to a Grid environment.

PROJECT DETAILS

Description

The development of improved DNA (deoxyribonucleic acid) sequencing technologies in the 1980s and 1990s heralded the start of a new era in biology. A vast amount of data is also being generated from DNA microarrays, mass spectrometry, DNA sequencing and structure analysis. In order to be useful, the data acquired with these technologies needs to be processed and interpreted. The need for processing this vast and exponentially growing information requires large parallel computing resources.

The large resource needs of bioinformatics allied to the large number of data-parallel applications in this field and the availability of a powerful, high-performance computing Grid environment lead naturally to opportunities for developing grid-enabled bioinformatics applications.



Our work focuses on creating a stand-alone, high-throughput implementation of BLAST in a Grid computing environment. Moreover, this work undertakes a closer study of SPMD type Grid applications, by investigating the effect of problem size and scheduling methodology on the feasibility or infeasibility of grid-enabling applications. In addition, this work also provides a detailed methodology for building a high-throughput Grid application. This can, in turn, be used as a template for developing similar applications on the Grid.

To the best of our knowledge, GridBLAST was the FIRST grid-enabled application, developed in Singapore.