

On the Migration of the Scientific Code Dyana from SMPs to Clusters of PCs and on to the Grid

Michela Taufer¹, Thomas Stricker¹, Gérard Roos¹, Peter Güntert²

¹ Department of Computer Science ² Institute for Molecular Biology and
Biophysics
ETH Zentrum ETH Hoenggerberg
CH-8092 Zurich, Switzerland CH-8092 Zurich, Switzerland
stricker,taufer@inf.ethz.ch guentert@mol.biol.ethz.ch

Abstract

Dyana is a molecular biology code used in the study of infectious prion proteins. Like many other scientific codes, Dyana was migrated successfully from vector supercomputers to a more cost-effective cluster of commodity PCs. A further migration to a widely distributed Grid computing platform looks very tempting because many of these platforms promise the use of nearly free compute-cycles on the Internet.

Not all codes are equally suited for all platforms. Even embarrassingly parallel codes might require a significant re-engineering effort for a migration from one platform to another. A better understanding of the performance characteristics of a code is required before a migration is attempted.

To address this problem, we present a systematic method to study the viability of a code migration from one platform to another. We construct an analytic performance model of the application. We use the previous migration from SMPs to commodity clusters of PCs to validate and calibrate the model. Finally we extrapolate the performance of Dyana to new platforms including widely distributed computing on the Grid and we suggest optimizations in the process of migration.

We demonstrate our method with the molecular biology code Dyana. In particular our general model predicts that Dyana can efficiently use up to 42000 processors with its current workload and is therefore well suited for grid computing on the Internet.

Keywords: *performance evaluation and modelling, computation-intensive applications, widely distributed supercomputing, migration of scientific codes, software engineering, grid architectures.*