CompBioGrid - Projected Timeframe to Full-Production Mode

Jeffrey Dutton and Ion I. Moraru

University of Connecticut Health Center

Richard D. Berlin Center for Cell Analysis and Modeling

263 Farmington Avenue

Farmington, CT 06030-1507

11 June 2009

INTRODUCTION

The current status of our Virtual Organization (VO), CompBioGrid, is documented as inactive with the caveat that we are ramping up in our ability towards full-production capabilities. What has led us to be assigned this status is the fact that CompBioGrid has existed as a VO for some time but has not utilized Open Science Grid (OSG) resources in a substantial way towards the VO's scientific goals – and nor have we been able to offer computational resources to the OSG community to assist the broader user base in their scientific computing goals. This is not to say that work has not been progressing towards full-production capabilities. There have been a number of advancements to date that were, and continue to be, necessary conditions towards making it possible for CompBioGrid to begin "production-level" scientific work utilizing OSG resources.

INFRASTRUCTURE SUPPORT

1. A separate network infrastructure had to be purchased and implemented by our institutions' IT department. This separate network infrastructure was necessary to prevent OSG network traffic from potentially flooding the institutional network and to provide the site UCHC_CBG an independent pipeline to Internet2. An important consideration in developing this independent network was to isolate network traffic related to open access OSG activities from the institutional network, as our campus location is an academic medical center with a university hospital. This ensures that external network connections to UCHC_CBG resources can never reach HIPAA-protected data on the institutions network and are thus not subject to considerably complex additional security and monitoring requirements.

2. A number of major infrastructure upgrade projects were undertaken in our Datacenter to enhance cluster computing capabilities and facilitate the deployment of an additional cluster dedicated to the UCHC_CBG CE site. A clustered storage system was purchased to allow scalable high-performance access to large capacity shared storage for multiple compute clusters. Additionally, a major upgrade of

the core switching infrastructure was performed. Two enterprise-class switches were purchased to support Terabit core switching and multiple switches were upgraded for10gigE connectivity and link aggregation to allow aggregate non-blocking traffic to all clusters' compute nodes and storage nodes on all public and private LANs.

RECENT ADVANCEMENTS

In recent months, progress has been made on building the Compute Element (CE) UCHC_CBG. This is a critical component towards reaching full-production for the science CompBioGrid will initially undertake. It is very desirably, and necessary, to be able to do the development work on two major science goals using an onsite CE for the purposes of optimization of the software and workflow as well as debugging capabilities inherent when seeing an application run from the client and server side.

TIMELINE

The following is a best-case scenario for ramping up to full-production for two different science projects:

By June 30th – Complete CE validation of UCHC_CBG.

By July 31st – Initial development completed to run Virtual FRAP software on OSG resources (see below for description of the application).

By September 15th – An initial version of the Virtual Cell software that will run on OSG resources (see below for description of the application).

ADDITIONAL INFORMATION PER PROJECT

1. The Virtual FRAP (VFRAP) is a software tool part of the Virtual Microscopy suite that is being recently developed at our center. These tools use VCell (see second project below) simulation technology in customized applications to analyze experimental data of the dynamic behavior of

fluorescently labeled molecules in live cells. VFRAP allows the extraction of quantitative parameters from Fluorescence Redistribution After Photobleaching (FRAP) experiments that cannot be obtained by analytical analysis of a selected subset of the data. Unlike VCell (see below), VFRAP is being made available to biologist users as a standalone software to be installed and run on the users' computer (currently beta preview version prior to the first official public release). It uses an extension of our existing linear transport parameter estimation algorithm to more general nonlinear reaction/diffusion models and requires multiple compute-intensive simulation runs per analysis that are currently run as child processes on the same computer. We plan to grid-enable VFRAP to allow these simulations to be dispatched as asynchronous batch jobs to remote resources.

2. The Virtual Cell (VCell) is a modeling and simulation framework that has been developed and deployed for more than ten years by our center. It is currently the most widely used software platform for building quantitative spatial models of cellular biochemistry, with more than 2,000 worldwide users who currently store over 30,000 models and over 160,000 simulation results in the VCell database. The target audience comprises mostly cell biologists with limited computational knowledge and resources. As such, VCell has been designed and is being deployed as a web-based distributed application; all users use a client application to create and store models and run simulations by connecting to dedicated servers and clusters at our HPC facility. We plan to grid-enable VCell in order to keep up with the rapidly expanding compute and storage needs of a continuously growing user base that utilize continuously expanding simulation features of VCell.

LINKS

Information about VFRAP can be found at http://vcell.org/vfrap

Information about VCell can be found at http:// vcell.org/