

Award Abstract # 0849980

Collaborative: CSR-DMSS, TM: Live Updates of Scientific Workflows

NSF Org:	CNS Division Of Computer and Network Systems
Recipient:	ARIZONA STATE UNIVERSITY
Initial Amendment Date:	August 25, 2008
Latest Amendment Date:	July 28, 2011
Award Number:	0849980
Award Instrument:	Continuing Grant
Program Manager:	Marilyn McClure mmclure@nsf.gov (703)292-5197 CNS Division Of Computer and Network Systems CSE Directorate for Computer and Information Science and Engineering
Start Date:	September 1, 2008
End Date:	August 31, 2013 (Estimated)
Total Intended Award Amount:	\$700,000.00
Total Awarded Amount to Date:	\$715,600.00
Funds Obligated to Date:	FY 2008 = \$190,000.00 FY 2009 = \$350,000.00 FY 2010 = \$160,000.00 FY 2011 = \$15,600.00
History of Investigator:	Rida A Bazzi (Principal Investigator) bazzi@asu.edu
Recipient Sponsored Research Office:	Arizona State University 660 S MILL AVENUE STE 204 TEMPE AZ US 85281-3670 (480)965-5479
Sponsor Congressional District:	04
Primary Place of Performance:	Arizona State University 660 S MILL AVENUE STE 204 TEMPE AZ US 85281-3670
Primary Place of Performance Congressional District:	04
Unique Entity Identifier (UEI):	NTLHJXM55KZ6
Parent UEI:	
NSF Program(s):	Information Technology Researc, Special Projects - CNS, CSR-Computer Systems Research, DATA-INTENSIVE COMPUTING
Primary Program Source:	01000809DB NSF RESEARCH & RELATED ACTIVIT 01000910DB NSF RESEARCH & RELATED ACTIVIT 01001011DB NSF RESEARCH & RELATED ACTIVIT 01001112DB NSF RESEARCH & RELATED ACTIVIT
Program Reference Code(s):	7354, 9178, 9216, 9251, HPCC
Program Element Code(s):	164000, 171400, 735400, 779300
Award Agency Code:	4900
Fund Agency Code:	4900
Assistance Listing Number(s):	47.070

ABSTRACT

Live Updates of Scientific Workflows

ABSTRACT

Computer systems are being applied to increasingly demanding cross disciplinary applications. The environments in which they function and the resources they manage are gradually more diverse, distributed and dynamic. This is particularly observed for

https://www.nsf.gov/awardsearch/show-award?AWD_ID=0849980

Feedback

biomedical applications where the need for flexible, dynamic, data intensive computing functionalities are substantially critical. This revolution of modern medicine requires novel and efficient approaches to support resource integration and scientific workflows (SWF) that capture, analyze, retrieve, integrate, and produce molecular information including biological specimens derived from tissue, cells, or blood combined with medical records or clinical trial data. Such workflows are increasingly used in e-science discovery and are characterized by frequent changes that are part of the scientific discovery process. The research develops a uniform approach to support live updates of scientific workflows accessing distributed resources. The goal of live updates is to allow continuous execution in the presence of changes. The research supports updates of both the structure and nature of SWFs processes and the data they access and create (produce and consume). The research divides the live update problem for SWFs into four main technical aims: software update, persistent storage update, workflow update, and evaluation and testing. It develops the mechanisms for updates as well as automatic checks for the safety of a planned update. The research impacts information-based drug discovery and development, and healthcare decision support systems.

PROJECT OUTCOMES REPORT

Disclaimer

This Project Outcomes Report for the General Public is displayed verbatim as submitted by the Principal Investigator (PI) for this award. Any opinions, findings, and conclusions or recommendations expressed in this Report are those of the PI and do not necessarily reflect the views of the National Science Foundation; NSF has not approved or endorsed its content.

Overview of the Project

The project aimed at developing techniques and tools for *dynamic software update*. Unlike the software update that many are familiar with and typically requires restarting the software application or even the device for the update to take effect, dynamic software update updates a software application while it is running. Our study was not confined to traditional software application, but also looked at updates of *scientific workflows*. A scientific workflow typically consists of many software components that can be developed by different developers and that are put together in a workflow to achieve a scientific analysis. The execution of scientific workflows is time consuming and there is a need to support experimentation with scientific workflows in which modified versions of the same workflow are executed in order to find the best way to analyze scientific data.

An important goal of the work was to expose students in computer science to work with scientists from the general biotechnology field through internships at the Translational Genomics Institute.

Outcomes

We have been successful in achieving many of the project goals.

1. **Publications.** Ten publications in conferences and workshops resulted from this project with some additional publications in preparation.
2. **Development of a general mechanism for dynamic software update.** We have developed a general approach and software tool for updating programs while they are executing. We call the tool UpStare. It is a general dynamic software update tool and it is available together with a user manual for upload and use by other researchers.
3. **Identification of the inherent overhead introduced by dynamic software update.** We have shown that any general dynamic software update tool that does not place limitation on when an update can occur must introduce overhead to the execution of a program even when the program is not being updated. We have proposed, but not implemented, a mechanism whose overhead would be the lowest possible.
4. **Improvement of the Structural Prediction for pProtein folding UTILITY System (SPROUTS).** SPROUTS was supported by a previous NSF grant and is available to scientists to compare and integrate various analyses for protein folding. SPROUTS was made much more useable and significant improvements were made to the ability to compare multiple analyses as part of effort under this project. As the results of the improvements, SPROUTS' collection grew to 900 entries for various proteins.
5. **Involvement of undergraduate students in research.** Five undergraduate students worked on research related to the project. The work led to one student being a co-author on a workshop paper. Another student who has started working on the project before the expiration of the grant is continuing to work on the topic of dynamic database update. A high school student was also hosted at ASU as part of the PI's outreach effort to local high schools.
6. **Training of workforce.** Three computer science and engineering students working on the grant got the chance to work with life scientists and help in making it easier for the scientists to solve their scientific problems. One PhD student and one Masters student already graduated and one PhD and one Masters student are expected to graduate in 2014. One computer science student spent part of a summer in France working with French scientists.
7. **Identification of conditions under which a dynamic software update can b...**

Please report errors in award information by writing to: awardsearch@nsf.gov.