



[e-Infrastructure colloquium 22 June 2012](#)

Towards Scalable Distributed Applications and Systems: The P* Model of Pilot-Abstractions

by **Shantenu Jha**, Project leader SAGA, Rutgers University School of Engineering, Department of Electrical and Computer Engineering, USA

Friday 22 June 2012, 15h30 - 16h30 drinks afterwards

Science Park Amsterdam, UvA FNWI, room 1.110, Science Park 904, 1098 XH Amsterdam

Science Park Amsterdam is easily accessible by public transport, train station Amsterdam Science Park. Paid parking for cars.

www.scienceparkamsterdam.nl/en/contact/how-to-find-us

Admission: free, no registration

Abstract

Towards Scalable Distributed Applications and Systems: The P* Model of Pilot-Abstractions

Many scientifically important questions require the efficient use of high-performance and distributed computing in order to provide answers with the accuracy needed at large-scales. We begin by analyzing how and why it has been necessary to develop "effective abstractions" in order to successfully utilize production high-performance distributed cyberinfrastructure, such as NSF's TeraGrid/XSEDE. For example, pilot-jobs are arguably one of the most widely-used distributed computing abstractions, and have been shown to support scalable and dynamic utilization of distributed resource. However, there does not exist a well defined, unifying conceptual model of pilot-jobs which can be used to define, compare and reason across different implementations of pilot-jobs; this presents a barrier to extensibility and interoperability. We introduce the P* Model the first known conceptual model of pilot-jobs, validate its implementation via the Pilot-API -- by concurrently using multiple distinct pilot-job frameworks on distinct production distributed cyberinfrastructures, and propose extensions of the P* Model to data. We will discuss the application of the pilot-abstraction to support the infrastructural and algorithmic requirements of several Grand Challenge problems facing the Computational Biology community, for example data-analytics for next-generation gene sequencing, enhanced sampling molecular algorithms, and in-silico personalized and predictive health-care.

Shantenu Jha

<http://www.ece.rutgers.edu/faculty/jha>

<http://www.ece.rutgers.edu/node/339>

Shantenu is an Assistant Professor at Rutgers University, a member of the Graduate Faculty in the School of Informatics at the University of Edinburgh (UK), and a Visiting Scientist at University College London. He is also the Associate Director for Advanced Research Cyberinfrastructure at the nascent Rutgers Discovery Informatics Institute. Before moving to Rutgers, he was the lead for Cyberinfrastructure Research and Development at the CCT at Louisiana State University. His research interests lie at the triple point of Applied Computing, Cyberinfrastructure R&D and Computational Science. Shantenu is the lead investigator of the SAGA project (<http://www.saga-project.org>), which is a community standard and is part of the official middleware/software stack of most major Production Distributed Cyberinfrastructure -- such as US NSF's XSEDE and the European Grid Infrastructure. His research has been funded by multiple NSF awards, US National Institute for Health (NIH) as well as the UK EPSRC (OMII-UK project and Research theme at the e-Science Institute). Jha has won several prestigious awards at ACM/IEEE Supercomputing and the International Supercomputing Series. Jha is writing a book on "Abstractions for Distributed Applications and Systems: A Computational Science Perspective". Jha seeks fearless and revolutionary young minds to join the RADICAL (thinking) group! Away from work, Jha tries middle-distance running and biking, tends to be an economics-junky, enjoys reading and writing random musings and tries to use his copious amounts of free time with a conscience.

e-Infrastructure colloquia are organized by [BiG Grid](#), [e-BioGrid](#), [NBIC](#), [SARA](#), [Nikhef](#), [EGI](#), [NLeSC](#), [UvA](#).