Tu Mai Anh Do

Contact Information

Full Birth Name:	Tu Mai Anh Do	Email:	tudo@isi.edu
Date of Birth:	21 May 1993	Address:	Room 1220, 4676 Admiralty Way, Suite 1001
Nationality:	Vietnamese		Marina del Rey, California 90292, USA

Research Interests

In situ Processing, Big Data Analytics, Scientific Workflows, Workflow Managment System, High Performance Computing, Distributed Systems

Education

Aug. 2017 – Now | University of Southern California (USC), Los Angeles, California, USA
Viterbi School of Engineering
Ph.D. Candidate in Computer Science
Advisor: Ewa Deelman

Sep. 2011 – April. 2016 | Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh, Vietnam
Faculty of Computer Science and Engineering
B.E. in Computer Engineering, Honors Program
GPA: 8.55/10.00

Experience

Aug. 2017 – Now | Information Sciences Institute, Marina Del Rey, California
Graduate Research Assistant
Advisor: Ewa Deelman
Science Automation Technologies Group
Working on In Situ Data Analytics for Next Generation Molecular Dynamics Workflows project (Analytics4MD is funded by NSF (National Science Foundation))
Tools: DataSpaces, ADIOS, Decaf, Pegasus, Cmake, Shippable, Gromacs, Plumed2, A4MD, TAU

May. 2018 – Aug. 2018 | Lawrence Livermore National Laboratory, Livermore, California
Student Intern
Advisor: Ming Jiang
Computation Directorate/Institute for Scientific Computing Research
Enabled data analytics workflows that couple high-performance computing simulations with Big Data analytics using node-local storage
Tools: Spark, Ascent, Pegasus

Oct. 2014 – Jul. 2017 | High Performance Computing Center, Ho Chi Minh City University of Technology
Research Assistant
Advisor: Nam Thoai
Studied and developed abnormal behavior detection techniques for large-scale parallel applications in message-passing programming model
Tools: C, C++, MPI, Makefile, Shell Script

Mar. 2017 – Jul. 2017 | Novobi Vietnam, Ho Chi Minh City
Part-time Software Engineer
Built automated system for deploying, testing and delivering package of health care application
Tools: Batch Script, Powershell Script, Atlassian Bamboo, Visual Studio, SQL Server Management Studio

May. 2015 – August. 2015 | DEK Technology Vietnam, Ho Chi Minh City
Summer Intern
Built and automate deploy small high availability cluster
Tools: Shell Script, Git

Teaching Experience

Fall 2020	University of Southern California (USC), Los Angeles, California, USA
⋮	*Teaching Assistant*
	Courses: *CSCI 585 - DATABASE SYSTEMS*
Spring 2021	*Instructor*: Saty Raghavachary

Sep. 2016	Ho Chi Minh City University of Technology, Ho Chi Minh City
⋮	*Teaching Assistant*
	Courses: *PARALLEL PROGRAMMING AND DISTRIBUTED SYSTEMS, FUNDAMENTALS OF PROGRAMMING*
May. 2017	*Instructors*: Nam Thoai, Sach Le

Notable Projects

- **HCMUT B.E Graduation Thesis**

  Title: Developing Methods To Help Large-Scale Parallel Applications More Reliable
  Advisor: Nam Thoai
  Description: This thesis explored the idea that scalability is the main cause of suffering high probability of error in large-scale parallel applications. To help the applications more reliable for better performance goal, the technique was proposed by detecting abnormal behaviors, which are considered as highly vulnerable from errors.
  Score: 9.96/10.00 – Highest score among concurrent defensed theses

- **Analytics4MD**

  Name: In Situ Data Analytics for Next Generation Molecular Dynamics Workflows
  Advisor: Ewa Deelman
  Description: Molecular dynamics simulations studying the classical time evolution of a molecular system at atomic resolution are widely recognized in the fields of scientific computing. Moving to exascale requires the simulations to analyze data as it is generated and store only necessary data. The analysis need to perform in-situ. We propose a paradigm for transforming the centralized, off-line nature of the molecular dynamics analysis into performing in-situ processing via in-memory staging area or in-transit processing via optimized parallel file systems or emerging burst buffer. This effort provides the ability to interleave simulations and analytics to improve data analyzing performance and processing more data.

Activities

- *Attended* Student Volunteers program of 2019 Supercomputing (SC19), November 16th, 2019 - November 22nd, 2019
- *Attended* 2019 eScience Conference, September 24th, 2019 - September 27th, 2019
- *Presented* a poster titled "Enabling Data Analytics Workflows using Node-Local Storage" at the USC ISI Graduate Student Symposium 2019, Marina Del Rey, California, March 26, 2019
- *Attended* 2018 Supercomputing (SC18) as a Student Volunteer and *presented* an acceptted poster titled "Enabling Data Analytics Workflows using Node-Local Storage", Nov 10, 2018 - Nov 16, 2018
- *Visited* Rutgers Discovery Informatics Institute (RDI^2), New Brunswick, New Jersey, May 13th, 2018 - May 17th, 2018
- *Presented* a poster titled "In Situ Data Analytics for Next Generation Molecular Dynamics Workflows" at the USC ISI Graduate Student Symposium 2018, Marina Del Rey, California, April 5th, 2018

Awards and Honors

2017 | ISI Distinguished Top-Off Fellowship

2016 | Recommended Candidate of 2017 VEFSTA Fellowship Program

2016 | 18th Eureka Scientific Research Student Award Finalist

2015 | 7th HCMC Information and Communication Technology Award for Student

2014 | DATALOGIC Vietnam's Scholarship, CSC Vietnam's Scholarship

Publications

- **T. M. A. Do**, L. Pottier, S. Caíno-Lores, R. Ferreira da Silva, M. A. Cuendet, H. Weinstein, T. Estrada, M. Taufer, and E. Deelman, "A Lightweight Method for Evaluating In Situ Workflow Efficiency," in Journal of Computational Science, 48, 101259, 2020
- **T. M. A. Do**, L. Pottier, S. Thomas, R. Ferreira da Silva, M. A. Cuendet, H. Weinstein, T. Estrada, M. Taufer, and E. Deelman, "A Novel Metric to Evaluate In Situ Workflows," in International Conference on Computational Science (ICCS), 2020.
- S. Thomas, M. Wyatt, **T. M. A. Do**, L. Pottier, R. Ferreira da Silva, H. Weinstein, M. A. Cuendet, T. Estrada, E. Deelman, and M. Taufer, "Characterization of In Situ and In Transit Analytics of Molecular Dynamics Simulations for Next-generation Supercomputers," in 15th eScience Conference, 2019.
- R. Ferreira da Silva, S. Callaghan, **T. M. A. Do**, G. Papadimitriou, and E. Deelman, "Measuring the Impact of Burst Buffers on Data-Intensive Scientific Workflows," Future Generation Computer Systems, vol. 101, p. 208–220, 2019.
- **T. M. A. Do**, M. Jiang, B. Gallagher, A. Chu, C. Harrison, K. Vahi, and E. Deelman, "Enabling Data Analytics Workflows using Node-Local Storage," in The International Conference for High Performance Computing, Networking, Storage, and Analysis (SC18), Poster, 2018.
- **T. M. A. Do**, D. Diep, and N. Thoai, "Race Condition and Deadlock Detection for Large-Scale Applications," 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016.
- **T. M. A. Do**, D. Diep, and N. Thoai, "Message Leak Detection in Debugging Large-Scale Parallel Applications," 2015 International Conference on Advanced Computing and Applications (ACOMP), 2015