Ilkay Altintas

Dr. Ilkay Altintas is the Chief Data Science Officer at the San Diego Supercomputer Center (SDSC), UC San Diego, where she is also the Founder and Director for the Workflows for Data Science Center of Excellence and a Founding Faculty and Fellow of the Halicioglu Data Science Institute (HDSI). In her various roles and projects, she leads collaborative multi-disciplinary with a research objective to deliver impactful results through making computational data science work more reusable, programmable, scalable and reproducible. Since joining SDSC in 2001, she has been a principal investigator and a technical leader in a wide range of cross-disciplinary projects. Her work has been applied to many scientific and societal domains including bioinformatics, geoinformatics, high-energy physics, multi-scale biomedical science, smart cities, and smart manufacturing. She is a co-initiator of the popular open-source Kepler Scientific Workflow System, and the co-author of publications related to computational data science at the intersection of workflows, provenance, distributed computing, big data, reproducibility, and software modeling in many different application areas. Among the awards she has received are the 2015 IEEE TCSC Award for Excellence in Scalable Computing for Early Career Researchers and the 2017 ACM SIGHPC Emerging Woman Leader in Technical Computing Award.

Abstract

Over the last two decades, workflows made change. They went from experimental tools to commodity environments with many capabilities to support the end-to-end process for scientific applications. A challenge for today’s workflows is the ability to respond to such heterogeneous needs and lowering the barriers to computing for long tail researchers as well as supporting the most cutting edge computing applications. Such needs require software systems that can coordinate applications across different scales of computing, data and networking needs. A number of software innovations like cluster virtualization, containers and gateways have enabled system software be more portable and lowered the barriers for many more to take advantage of computing. However, there is still a need at the converged application level to enable communications with data and computing middleware, while optimizing resources and dynamically adapting to the changes in application workflows. Even though workflow systems have been a part of the HPC and HTC ecosystem for task coordination and management, we are just discovering their potential for managing dynamic-data driven applications and decision support using advances in big data platforms and on-demand computing systems. They provide an ideal programming model for deployment of computational and data science applications on all scales of computing and provide a platform for system integration of data, modeling tools and computing while making the applications reusable and reproducible. This talk will overview the plethora of applications supported by workflows over the last two decades and introduce some of our recent work on the development of methodologies and tools for effective workflow driven collaborations, namely the PPoDS methodology and family of SmartFlows tools for the practice and smart utilization of workflows.