Seminar

Big Data & Data Science: Fourth Paradigm Framework

Abstract. The emergence of massive data collections (i.e., "Big Data") has ushered in a paradigm shift in the way scientific research is conducted and new knowledge is discovered. The traditional observe-hypothesize-test (OHT) model of scientific endeavor is increasingly augmented with, and in some cases supplanted by, collaborative scientific research involving multi-disciplinary teams from distributed organizations coming together to solve a common problem. Research is increasingly data-intensive, with individual investigators and collaborative teams generating and harnessing significant amounts of data that require complex data mining, integration and analysis to discover and test new models and hypotheses. This paradigm shift, often referred to as the "fourth paradigm", is central to data-intensive research and data-enabled scientific discovery. This talk will examine a fourth paradigm framework for defining the research life-cycle needs in the evolving cyber-data environment and the Data Intensive Science and Engineering (DISE) system needed for performing data-intensive research. We will discuss emerging solutions to these issues such as the integrated Rule Oriented Data Systems and the DataBridge.

Dr. Arcot Rajasekar

Professor, School of Information and Library Science, Chief Domain Scientist, Data Renaissance Computing Institute University of North Carolina, Chapel Hill, NC
rajasekar@unc.edu
https://sils.unc.edu/people/faculty/profiles/Arcot-Rajasekar

Friday October 19, 2018
Time: 2:00 – 2:50pm
Room: Bate 1021

Biography. Arcot Rajasekar is a Professor in the School of Library and Information Sciences at the University of North Carolina at Chapel Hill, a Chief Scientist at the Renaissance Computing Institute (RENCI) and co-Director of Data Intensive Cyber Environments (DICE) Center at the University of North Carolina at Chapel Hill. Previously he was at the San Diego Supercomputer Center at the University of California, San Diego, leading the Data Grids Technology Group. He has been involved in research and development of data grid middleware systems for over a decade and is a lead originator behind the concepts in the Storage Resource Broker (SRB) and the integrated Rule Oriented Data Systems (iRODS), two premier data grid middleware developed by the Data Intensive Cyber Environments Group. A leading proponent of policy-oriented large-scale data management, Rajasekar has several research projects funded by the National Science Foundation, the National Archives, National Institute of Health and other federal agencies. Rajasekar has a PhD in Computer Science from the University of Maryland at College Park and has more than 200 publications in the areas of data grids, digital library, persistent archives, artificial intelligence and smart cities. His latest projects include the Datanet Federation Consortium, the Data Bridge and Smart and Connected Communities Initiative