The Need for Integrated Analytic Platforms and Multithreaded Runtime Systems

The availability of data is changing science, commerce, society, and defense. Unfortunately, data comes in many different varieties and formats. Data is both historic and real-time creating enormous alignment, collection, processing, and error correcting challenges. Analytic workflows are complex graphs of software packages comprising a plethora of algorithms, data structures, data format, and programming models. Implementing the workflows by linking methods from subject-specific vertical libraries introduces significant data and parallel inefficiencies that cripple performance and scalability. I will review several application exemplars that epitomize the complexity of emerging workflows and argue for horizontal platforms that support a wide spectrum of data types, data structures, and methods without prejudice. We need platforms that support equally well linear algebra, relational operations, graph methods, and machine learning. We need platforms that can view data as matrices, tensors, tables, and graphs. We need platforms that can implement data parallelism, synchronous and asynchronous tasks, map-reduce, and client-server protocols. We need platforms that can tolerate small messages and fine-grain synchronization events (dynamic task scheduling and atomic memory operations).

I will conclude by describe a scalable, high-performance mixed analytic platform that we have developed at PNNL that meets the requirements of emerging workflows. The platform supports a variety of analytic methods, data structures, and parallel programming models. It processes effectively at-rest and streaming data. The platform relies on aggressive multithreading and data aggregation, rather than data partitioning, to tolerate fine-grain irregular memory accesses and atomic memory operations. I will provide both application examples and performance results.

Bio: Dr. John Feo is a senior manager at Pacific Northwest National Laboratory where he manages a variety of DOD and DOE research project in data analytics, scalable software platforms, and parallel programming models. The projects are investigating cyber security, critical infrastructures, and transportation systems. Dr. Feo received his Ph.D. in Computer Science from The University of Texas at Austin. He began his career at Lawrence Livermore National Laboratory where he managed the Computer Science Group and was the principal investigator of the Sisal Language Project. Dr. Feo then joined Tera Computer Company (now Cray Inc) where he was a principal engineer and product manager for the MTA-1 and MTA-2, the first two generations of the Cray's multithreaded architecture. He has taken short sabbaticals to work at Sun Microsystem, Microsoft, and Context Relevant. Dr. Feo's has held academic positions at UC Davis and Washington State University.