International Collaboration Framework for Extreme Scale Experiments (ICEE)

Scientific Collaborations at Extreme-Scale
DOE ASCR Lab 12-695

Principal Investigator

Name(s)	John Wu
Division	CRD
Institution	LBNL
Address	One Cyclotron Road MS50B3238 Berkeley, CA 94720
Phone number (Voice)	510-486-6609
Phone number (Fax)	510-486-4004
Email	kwu@lbl.gov

Co-Principal Investigators:

Name	CS Chang	Scott Klasky	Yeong-Kook Oh
Organization	PPPL	ORNL	KSTAR
Email	cschang@cims.nyu.edu	klasky@ornl.gov	ykoh@kbsi.re.kr
Phone	609-243-2127	865-241-9980	82-42-865-3494
Fax	609-243-2665	865-241-4811	82-42-865-3459

Abstract

Large-scale scientific exploration in domains such as high-energy physics, fusion, and climate are based on international collaborations. As these collaborations produce more and more data, the existing workflow management systems are hard pressed to keep pace. A necessary solution is to process, analyze, summarize and reduce the data before it reaches the relatively slow disk storage system, a process known as in transit processing (or in-flight analysis). We propose to dramatically increase the data handling capability of collaborative workflow systems by leveraging the popular in transit processing system known as ADIOS, and integrating this with FastBit to provide selective data accesses. These new features will contribute to a new collaborative system named ICEE that aims at significantly improving the data flow management for distributed workflows. The improved data processing capability will enable large international projects to make near real-time collaborative decisions. As scientific teams tackle increasingly complex problems, many data analysis workflows have to dynamically adjust to the experimental conditions and sensor output; thus, workflows need to also be modified dynamically for evolving user requirements. The ICEE framework will not only allow users to modify parameters of a workflow, but also dynamically modify its processing elements and alter its structure. Additionally, we plan to incorporate data mining features to provide feedback and recommendations while the user is constructing or modifying a workflow. Overall, the ICEE framework will allow researchers to conduct distributed analyses on extreme scale data efficiently and easily. It will enable collaborative decisions in near real-time for geographically distributed teams, reduce the turn-around time on large instruments, and improve scientific productivity.