Session TBD

James Ahrens - KEYNOTE

Senior Scientist at Los Alamos National Laboratory

“Supercharging the Scientific Process Via Data Science at Scale”

Abstract

Supercharging the Scientific Process Via Data Science at Scale

Historically, the scientific process is used to explain phenomena by iteratively formulating a theory and running real-world experiments to test and improve the theory. Advances in the field of computer engineering, driven by Moore¹s law has fundamentally changed the scientific process in two ways:

- The **first change** is the availability of inexpensive but highly accurate sensors composed of integrated circuits. The sensors, such as extremely high resolution cameras and signal recorders, enable the collection of scientific data and are used in all scientific disciplines.

- The **second change** is the addition of highly detailed scientific simulations that run on high performance computing (HPC) platforms. The performance of these HPC platforms has increased by approximately six orders of magnitude over the past two decades from terascale (10 ^12 Floating Points Operations Per Second (FLOPS)) to petascale (10 ^15 FLOPS). This performance increase has enabled the creation of extremely detail scientific simulations. These simulations augment the scientific process by providing a proving ground for theories and an environment for virtual experimentation.

Both changes produce massive data streams that need to be effectively processed, transformed, analyzed, visualized and understood through data science. In this talk, I will present new developments in data science, highlighting how the data analysis and visualization process needs to change for exascale supercomputers (10 ^18 FLOPS). Exascale supercomputers are bounded by power and storage constraints. These constraints require us to transition from standard, storage-based, post-processing, data science approaches to intelligent, automated, streaming, in situ ones. I will present a new approach that focuses on automatically identifying and tracking areas of interest, and then on selecting and visually presenting these areas to scientists. The work will be presented in the context of solving real-world data science problems for the climate, cosmological, asteroid impact and experimental material science communities.

Bio

Dr. James Ahrens is a senior research scientist at the Los Alamos National Laboratory (LANL). He is the founder and design lead of ParaView, a widely adopted visualization and data analysis package for large-scale scientific simulation data (http://paraview.org). ParaView has had an extremely positive impact on the large-scale data analytic capabilities available to simulation

scientists around the world. Dr. Ahrens graduated in 1989 with a B.S. in computer science from the University of Massachusetts and in 1996 with a Ph.D. in computer science from the University of Washington. At LANL, he is part of a data science team of twenty staff, postdocs and students. He is also a national leader of programmatic initiatives important to the United States Department of Energy's National Nuclear Security Administration and Office of Science. Dr. Ahrens is the Data Analysis and Visualization lead for the U.S. Exascale Computing Project and the general chair for this year's IEEE Scientific Visualization conference to be held in Phoenix, AZ in early October.