Exceptional service in the national interest

Sandia National Laboratories

Enabling Production-Quality Scientific-Discovery Tools with Data and Execution Models

Kenneth Moreland *Sandia National Laboratories*
Jeremy Meredith *Oak Ridge National Laboratories*
Berk Geveci *Kitware, Inc.*

Successful deployment of production scientific visualization tools like VisIt and ParaView have enabled prolific scientific discovery through advanced computing. As scientists pursue new discoveries using more powerful computers, it is vital that we maintain thoroughly functional visualization tools for the ever-increasing data sizes. The challenges in updating our high-performance tools as we move to extreme-scale computing are greater than ever as we observe fundamental changes in computer hardware, programming models, compiler technology, and system behavior.

Research, funded mostly by DOE initiatives, is well underway to understand new, effective algorithms for scientific visualization at extreme scale. However, it is a long proverbial road from a sparse collection of algorithms to production-ready tools. Our current visualization tools take advantage of a common framework that allows multiple algorithms and components to coexist and interact. To continue our success, we must provide similar mechanisms that work well at extreme scale. In particular, it is essential that we provide the appropriate execution and data models that provide a common convention for disparate visualization components.

An execution model provides the structure for the operation of an algorithm. It provides the internal and external behavior, defines input and output data, and designates the scope of influence. An execution model simplifies the design, debugging, and support of an algorithm by encapsulating important details such as scheduling, memory management, synchronization, and communication. The execution model also provides a common interface between algorithms that facilitates the amalgamation of independently designed components. Studies show that the execution models used by our current scientific visualization tools are inadequate to fully realize the high level of concurrency required by emerging computer design.

A data model, in addition to simply describing scientific data, also establishes storage mechanisms and access patterns. In scientific discovery tools, the data model must generally encompass those of the applications and experiments generating the data. However, data models have not evolved as necessary to meet the requirements of expanding and maturing scientific data sources. Furthermore, a data model provides opportunities for the execution model to gain efficiency and performance portability. For example, studies show a more flexible array layout can drastically reduce memory movement on many-core devices, and that new mesh representations can reduce the memory requirements of common analysis cases by an order of magnitude. Opportunities to grow support for new scientific data and evolving hardware architectures will be missed without advances in scientific data models.

In conclusion, the design of extreme-scale visualization algorithms is necessary but insufficient to address our future scientific discovery needs. In addition, it is critical to also provide appropriate data and execution models that form the basis for a common framework in which components can be brought together. Only then can we hope to provide complete, production-quality solutions to support analysis and visualization for scientific discovery.

SAND 2012-10796P

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Research sponsored by the Laboratory Directed Research and Development (LDRD) program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

This work was supported in part by the Director, Office of Advanced Scientific Computing Research, Office of Science, of the U.S. Department of Energy under Contract No. 12-015215, through the Scientific Discovery through Advanced Computing (SciDAC) Institute of Scalable Data Management, Analysis and Visualization.

This work was supported in part by the DOE Office of Science, Advanced Scientific Computing Research, under award number 10-014707, program manager Lucy Nowell.