SciDAC Scalable Data Management, Analysis, and Visualization Institute

The SciDAC SDAV Institute will actively work with application teams to assist them in achieving breakthrough science and will provide technical solutions in the data management, analysis, and visualization regimes that are broadly applicable in the computational science community.

As the scale of computation has exploded, the data produced by these simulations has increased in size, complexity, and richness by orders of magnitude, and this trend will continue. Users of scientific computing systems are faced with the daunting task of managing and analyzing their datasets for knowledge discovery, frequently using antiquated tools more appropriate for the teraflop era. While new techniques and tools are available that address these challenges, often application scientists are not aware of these tools, aren't familiar with the tools' use, or the tools are not installed at the appropriate facilities.

SDAV will deploy, and assist scientists in using, technical solutions addressing challenges in three areas:

- **Data Management** – infrastructure that captures the data models used in science codes, efficiently moves, indexes, and compresses this data, enables query of scientific datasets, and provides the underpinnings of in situ data analysis
- **Data Analysis** – application-driven, architecture-aware techniques for performing in situ data analysis, filtering, and reduction to optimize downstream I/O and prepare for in-depth post-processing analysis and visualization
- **Data Visualization** – exploratory visualization techniques that support understanding ensembles of results, methods of quantifying uncertainty, and identifying and understanding features in multi-scale, multi-physics datasets

The team will work directly with application scientists to assist them and in the process will learn from the scientists where SDAV tools fall short. Technical solutions to any shortcomings will be developed to ensure that our tools address and overcome mission-critical challenges in the scientific discovery process. State-of-the-art techniques in software development and quality assurance will be applied so that the software developed and deployed meets the high standards needed to ensure the correctness and performance of science codes.

In addition to connecting with application teams, close ties to leading compute facilities are important for successful deployment and adoption of SDAV tools. The Institute includes facility partners from NERSC, ANL, and ORNL who are responsible for software installation at their respective site. These partners will also inform SDAV team members of upcoming system architectures, guiding development of SDAV tools to ensure that they will be ready as new systems come online.

Institute Director:
Arie Shoshani, LBNL

Deputy Director:
Robert Ross, ANL

Executive Council:
Arie Shoshani, LBNL (chair)
James Ahrens, LANL
Wes Bethel, LBNL
Hank Childs, LBNL
Scott Klasky, ORNL
Kwan-Liu Ma, UC Davis
Valerio Pascucci, U Utah
Robert Ross, ANL

Key Personnel:
Sean Ahern, ORNL
James Ahrens, LANL
Wes Bethel, LBNL
Peer-Timo Bremer, LLNL
Eric Brugger, LLNL
Hank Childs, LBNL
Alok Choudhary, Northwestern
Berk Geveci, Kitware
Charles Hansen, U Utah
Chris Johnson, U Utah
Kenneth Joy, UC Davis
Scott Klasky, ORNL
Robert Latham, ANL
Kwan-Liu Ma, UC Davis
Anatoli Melechko, UT Knoxville
Kenneth Moreland, SNL
Michael Papka, ANL
Manish Parashar, Rutgers
Valerio Pascucci, U Utah
Tom Peterka, ANL
Norbert Podhorszki, ORNL
David Pugmire, ORNL
Nagiza Samatova, NC State
William Schroeder, Kitware
Karsten Schwan, GA Tech
Han-Wei Shen, OSU
Venkat Vishwanath, ANL
Matthew Wolf, GA Tech
Jonathan Woodring, LANL
John Wu, LBNL

SciDAC Scalable Data Management, Analysis, and Visualization Institute

In addition to one-on-one collaborations between SDAV and science teams, SDAV team members will organize tutorials and workshops that will help inform the larger community about the tools the Institute makes available, train potential users, and provide opportunities to gather information from other researchers and potential customers. These activities will be coordinated with leading conferences (e.g., ACM/IEEE Supercomputing) and DOE computing facility activities (e.g., the ALCF *Getting Started Workshop* series).

SDAV is a collaboration tapping the expertise of researchers at six laboratories: Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Sandia national laboratories and in seven universities: Georgia Tech, North Carolina State, Northwestern, Ohio State, Rutgers, the University of California at Davis, and the University of Utah. Kitware, a company that develops and supports specialized visualization software, is also a partner in the project. The team will build on their successes from the SciDAC Scientific Data Management (SDM) Center for Enabling Technologies, the Visualization and Analytics Center for Enabling Technologies (VACET), and the Institute for Ultra-Scale Visualization (UltraVis) and provide the tools and knowledge required to achieve breakthrough science in this data rich era.

Contact Information for Executive Council:
Arie Shoshani, shoshani@lbl.gov, 510-486-5171
Robert Ross, rross@mcs.anl.gov, 630-252-4588
James Ahrens, ahrens@lanl.gov, 505-667-5797
Wes Bethel, ewbethel@lbl.gov, 510-486-7353
Hank Childs, hchilds@lbl.gov, 510-486-4154
Scott Klasky, klasky@ornl.gov, 865-241-9980
Kwan-Liu MA, ma@cs.ucdavis.edu, 530-752-6958
Valerio Pascucci, pascucci@sci.utah.edu, 801-585-6513

The SDAV Toolkit

Software tools are the vehicles through which our expertise can be applied to address application needs. This list captures the current set of tools provided by the SDAV team.

I/O Frameworks
ADIOS
Darshan
Parallel netCDF
ROMIO
ViSUS/IDX

In Situ Processing
ActiveSpaces
DataSpaces, DART
DIY
FFS, EvPath
GLEAN

Indexing and Compression
FastBit
ISABELA

Statistics and Data Mining
NU-Minebench
STPMiner
Importance-Driven Analysis

Topological Methods
Topologika

Visualization Frameworks
IceT
ParaView
Ultravis-V
VisIt
VTK

Flow Visualization
Ultravis-P