Overview

SDAV provides comprehensive expertise in scientific data management, analysis, and visualization aimed at transferring state of the art techniques into operational use by application scientists on leadership-class computing facilities over its five year period. A primary goal of SDAV is to provide technical solutions that are broadly used by the computational science community. These technical solutions will come in the form of software. This poster describes the SDAV Software Plan and also gives an overview of the tools supported by SDAV.

SDAV Software Plan

Our strategy for customer impact is to work directly with application scientists to assist them by applying the best tools and technologies at our disposal. Further, we will learn from the scientists where our tools fall short. Technical solutions to any disposal. Further, we will learn from the scientists applying the best tools and technologies at our with application scientists and to assist them by learning from the scientists applying the best tools and technologies at our

SDAV’s Data Management Tools

- ADIOS: ADIOS is a data management end-user tool that deals with parallel I/O for large-scale scientific applications. The purpose of this tool is to provide fast, adaptable and scalable I/O interfaces so that scientific codes can run efficiently across all computing platforms. ADIOS provides a simple, flexible interface for scientists to describe the data (either files or streams) in their code that may need to be written, read, or processed outside of the running simulation. ADIOS received a 2013 R&I 100 Award.
- ALACRITY Compressed Data Indexing: This tool offers storage-lightweight indexing over compressed scientific data, for the purpose of enabling accelerated response times on exploratory query-driven analysis and visualization. It is designed to function either during post processing or in situ. The ultimate goal is to facilitate scientific discovery by reducing time to results.
- Darshan: Darshan is an I/O instrumentation library that collects I/O access pattern information from applications running on leadership-class hardware. This information is collected with minimal memory overhead by application scientists, programmers, and I/O researchers to better understand and tune I/O behavior. Darshan is deployed in production on a variety of systems including major platforms at NERSC and ANL.

DataSpaces: DataSpaces is a data management framework that enables dynamic and asynchronous interactions between applications. It provides support for coupled applications workflows, including loose and tight coupling, data analysis workflows with in-situ and in-transit processing.

FastBit: FastBit is a tool for querying large multi-dimensional data sets with a large number of numerical variables. It supports subsets and filtering with predicate conditions over the variable, as well as histogramming and similar aggregate operations. Its primary strength is the extensive set of efficient bitmap indexing methods.

FastQuery: FastQuery intends to provide querying capability on top of data files in scientific data formats including ADIOS, HDF5, netCDF and PnetCDF. This tool gives scientists advanced searching capability without requiring them to load their data into a data management system. FastQuery was optimized to run on a large number of cores.

FlexPath: FlexPath is a component of Adios designed to provide memory-to-memory data transfers and transport-level data reduction through ‘data conditioning plugins’. It is designed to aid such things as data management, analysis and visualization.

GLEAN: GLEAN is an I/O acceleration and simulation-time analysis framework.

HDF5: Production quality, high performance parallel I/O library with self-describing data format, and accompanying tool suite.

ISOBAR: ISOBAR compression is effective in situ lossless compression to improve end-to-end I/O throughput when writing (or reading) scientific data, including other types of “hard-to-compress” data. ISOBAR compression is compliant with the JPEG2000 standard. It is an advanced version of an algorithm that encodes block data and can be used to compress data sets with additional information (e.g. timestamps, provenance, or units).

MLOC: MLOC offers multi-level storage layout optimization for scientific datasets to speed up queries with heterogeneous access patterns. It reorganizes datasets to a hybrid multi-level layout scheme to improve the locality for space-constrained, value-constrained scenarios. It supports a byte-level precision-based multi-resolution data storage and access.

ParalalNetCDF: ParallelNetCDF (PnetCDF) is a library providing high-performance parallel I/O while still maintaining file-format compatibility with Unidata’s NetCDF. The library provides a data model based on multi-dimensional arrays of typed data, which allows for annotation of datasets and variables with additional information (e.g. timestamps, provenance, or units).

PIDX: PIDX is a high performance parallel I/O library that writes data in hierarchical multi-resolution Z order (tiled) format suitable for analysis of large data sets. The library integrates with the visualization elements also aids in real-time monitoring of large scale simulations.

TALASS: The TALASS toolkit (which is part of the Topologyka suite) incorporates both a highly effective streaming computation of feature-based segmentations as well as an interactive environment to explore the temporal evolution of features and their properties.

SDAV’s Visualization & Analysis Tools

- EAVL: EAVL is a visualization library that targets the approaching hardware/software ecosystem by (1) updating the traditional data model to handle modern simulation codes and a wider range of data, (2) leveraging an updated data and execution model to achieve necessary computational, I/O, and memory efficiency, and (3) enabling developers to realize these efficiency gains through high productivity programming constructs.
- ICet: ICet is the Image Composition Engine for Tiles (ICet) is a high-performance sort-last parallel rendering library. ICet is principally used in large-scale scientific visualization tools to provide scalable and interactive rendering.
- ImageVis3D: ImageVis3D and the ‘tool companion’ ‘Ivo’ library is a visualization tool meant to transparently solve the problem of large data visualization. The tool can easily visualize terascale data sets on commodity hardware.
- MSECER: MSECER (Morse-Smale Complex Extraction and Reasoning) is a set of libraries and a visualization tool for data analysis and visualization using Morse-Smale complexes, which is part of the Topologyka suite. The software provides command-line tools for pre-computing and storing an MS complex, simplifying and extracting features, and interactive exploration and visualization.
- OSUPFlow: OSUPFlow is a parallel flow analysis and particle tracing library that supports analysis of large scale flow data sets. The library allows for scalable computation of streamlines, pathlines, stream surfaces, and stream surfaces, and can be deployed on DOE’s leadership computing facility.
- ParaView: ParaView is a general-purpose scientific data analysis and visualization tool. It was designed originally to work with large data. It supports an MPI-enabled data-parallel back-end (server) and client-server computing.
- ParaView-Catalyst: Catalyst is an open-source, scalable data analysis and visualization library designed to be tightly coupled with simulation codes. It can be directly embedded into parallel simulation codes to perform in situ analysis at run time. Furthermore, it can be coupled with the ParaView In Situ Analysis Framework to perform run-time visualization of data extracts and steering of the data analysis pipeline.
- PISTON: PISTON is a framework for creating such operators that employs the data-parallel programming model, which enables portable performance over many architectures. PISTON consists of a user-extensible set of operators (such as isosurface, Threshold, and cut surfaces) implemented using this model.
- Ultravis-P: Ultravis-P supports the visualization and statistical analysis of large-scale particle data that is produced by complex simulations. It is designed to work with massive volumes using very modest computing resources.

To learn more visit: http://www.sdav-scidac.org