Red Storm Data Analysis and Visualization Environment

Constantine Pavlakos and David White
Sandia National Laboratories

Abstract

This paper describes the data analysis and visualization environment that is being implemented for the Red Storm platform at Sandia National Laboratories, in conjunction with the Department of Energy's Advanced Simulation and Computing (ASC) program. The environment is based on an integrated hardware infrastructure that spans from the supercomputer to the desktop that includes: data/visualization servers in the form of graphics clusters; large storage in the form of a parallel file system; an integrated archive (tertiary storage); 3D graphics desktops, and high performance networking. The high performance graphics capabilities in the environment leverage the use of commodity PC-graphics cards, both at the desktop and within the parallel servers. A unique capability still under consideration for the Red Storm platform includes the deployment of graphics nodes on the platform itself.

A set of scalable software tools have been developed that take advantage of the high performance hardware, delivering demonstrated 3D-graphics rendering performance as high as 1.5 billion polygons per second, while also offering a choice of distributed high performance visualization usage models. Scientists, engineers, and application developers will have access to the EnSight-Gold commercial visualization tool; the open source ParaView tool, which is providing a framework for advanced custom development (ParaView is based on the popular Visualization Tool Kit, or VTK); and a suite of data manipulation tools. Importantly, the full power of Red Storm's data/visualization environment will be accessible from the office/desktop.

1 Introduction

The Red Storm platform will provide Sandia National Laboratories with a substantial increase in local supercomputing capability. In order to support real computational science and engineering applications, Red Storm's supercomputing capability must be complemented with commensurate hardware infrastructure and high performance software tools that enable effective data analysis and visualization[1,2], for comprehension of supercomputing results.

An integrated end-to-end environment is being implemented that emphasizes the following strategies:

- Scalability, in both hardware and software.
- Leveraging the use of commodity-based graphics hardware, both as parallel hardware components in cluster visualization servers and to provide local graphics power in the office.
- Accessibility of the complete environment and full set of services from the desktop/office.
- Ease of use.

The data analysis and visualization process is depicted in Figure 1. Note that the process is quite complex. A common misconception is to associate visualization purely with the process of rendering, which translates data objects such as surfaces and polygons into an image. In actuality, the process involves accessing data from potentially disparate sources, manipulating data in a diversity of ways, attaching visual attributes to the data, rendering of the data to produce images, and delivery to one or more of a multitude of possible display environments, all while under the control of whatever interfaces are provided to the end-user. Keep in mind that this process is often spread out over a highly distributed geographic and/or hardware environment. An effective data analysis and visualization environment must support a breadth of such functionality.

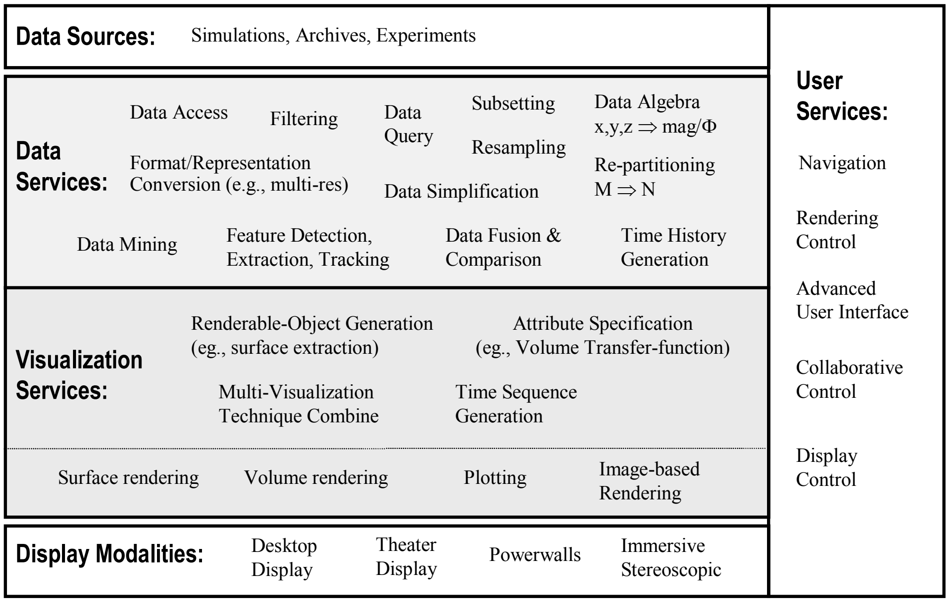


Figure 1: The data analysis / visualization process.

The data/visualization environment being deployed at Sandia is, in many respects, the culmination of a cooperative research and development effort that has been enabled through the ASC program. Over the course of a half-dozen years or so, Sandia's ASC data/visualization partners have included: the other two ASC labs, Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory (LLNL); Computational Engineering International (CEI)[3], providers of the EnSight suite of visualization tools, for development of high performance and other custom features; Kitware[4], developers and supporters of the open-source Visualization Tool Kit (VTK), for parallel VTK and ParaView development; Stanford University, for development of the Chromium[5] OpenGL-based parallel rendering system; Princeton University, for development of scalable displays[6]; University of Utah[7], for large data visualization algorithms; RedHat and Tungsten Graphics, for development of Chromium and the Distributed X Server (DMX)[8]; NVIDIA, for development and optimization of graphics card drivers for the Linux operating system; IBM, for development of ultra-resolution flat-panel displays; and the National Center for Supercomputing Applications, for development and support of the HDF5[9] scientific data format.

2 Overall Architecture

An overall end-to-end architecture being deployed in various forms by the three ASC labs is shown in Figure 2. Nominally, it includes: the ASC supercomputing platform(s); a data/visualization partition ("VIEWS Partition") directly on the platform that has direct access to data residing on the platform file system; high performance data/visualization servers that are powerful enough to support the post-processing of data sets of a magnitude anticipated in accordance with the platform(s); an archive, or tertiary storage system; display environments, including advanced facilities, such as powerwalls and/or immersive displays, and offices; and the underlying networking and communications infrastructure that supports the transmission of data and/or images. Transmission of image data has historically made use of analog hardware, but the trend is toward increasing deployment of digital-based approaches, not to mention transmission over conventional switched IP networks. An important feature of the VIEWS Partition on the platform is that simulation results can be accessed on the platform without any prior movement of the data – this is particularly useful for large data sets.

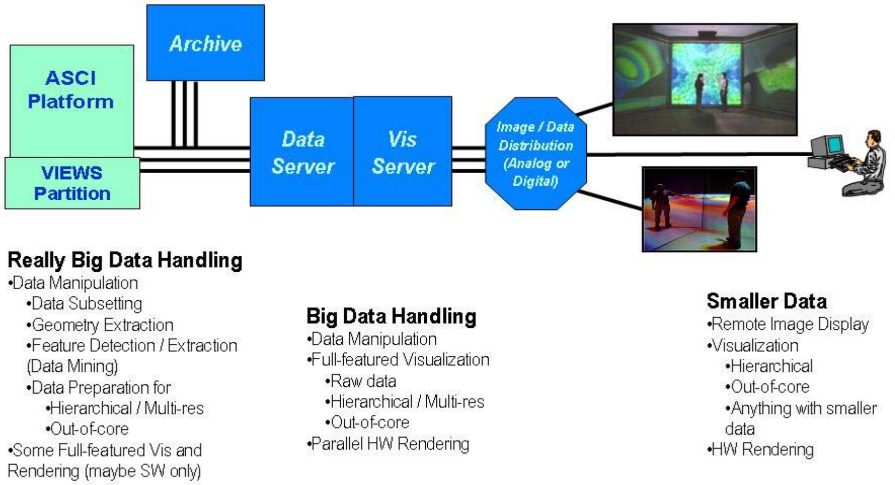


Figure 2: An ASC Tri-lab end-to-end architecture for data analysis & exploration.

A key to delivering a rich data analysis and visualization capability for such a supercomputing environment is to enable flexible access to the appropriate high performance parts of the end-to-end system, all the way from the display environments, shown on the right hand side in the figure above. Since the supercomputing platform itself, shown on the left hand side above, provides the most raw processing power in the distributed environment, and since simulation results generally originate there, it is important to enable at least partial processing of very large data in the proximity of the platform. Recent instantiations of ASC platforms (ASC White at LLNL and ASC Q at LANL) have included general purpose computing resources on the platform that can be used to support the back end of the data/visualization process. Since they have not included graphics rendering hardware, such resources are generally used to do various forms of data manipulation. While some tools have enabled full-featured visualization directly on the platform, including rendering, such rendering would generally be performed in software.

Data/visualization servers have been deployed to serve a couple of purposes: (1) to provide shareable high performance data processing and hardware-assisted graphics capability; and (2) to enable more distributed use with high-end visualization facilities, such as powerwalls. Note that these can be very substantial systems in their own right that are capable of handling large data sets and/or data objects – it becomes an issue of how to get the right data to the server. Recall that there is a preference to not move entire data sets if at all possible. With the emergence of shareable high performance file systems, it will become increasingly possible for supercomputing platforms and data/visualization servers to have equitable access to the same data storage.

Today's office graphics workstation is a vital part of the end-to-end. Aside from providing the end-users everyday interface to the full system environment, it has processing and graphics power which should be leveraged by the overall system. When possible, the system should accommodate the delivery of data-of-interest in forms and at scales (e.g., through multi-resolution techniques) that enable processing by the desktop. It is important to remember that the desktop, unlike other shareable system resources, is dedicated to the end user who possesses it, providing convenience and maximal opportunity for on-demand interactive use.

3 Hardware and Associated Infrastructure

The Red Storm platform is conceptualized in Figure 3. The platform is designed to support both classified ("red") and unclassified ("black") processing. Center compute-node portions (which run a light-weight operating system) are designed to be switchable from one side to the other. Each end of the platform incorporates service nodes that provide machine access, I/O nodes that provide access to a parallel file system, and I/O nodes that provide external high performance connectivity. The system is designed to deliver sustained I/O to the parallel file system at 50 GB/sec and sustained external I/O at 25 GB/sec (50 10-Gb-Ethernet links on each end).

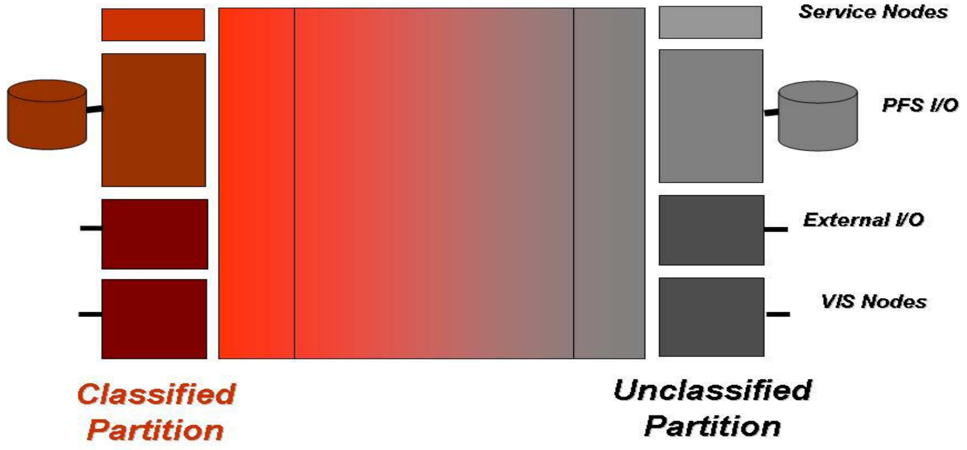


Figure 3: A conceptual depiction of Sandia's Red Storm platform, which has been designed to accommodate visualization ("VIS") nodes in each of the "Red" and "Black" partitions.

The platform has also been designed to accommodate the additional deployment of dedicated visualization nodes ("VIS nodes"). The plan has been to procure and deploy 256 graphics nodes, with integrated PCI-Express graphics cards, on each end of the machine. These nodes, like other service and I/O nodes, would run a full-service operating system to support the straightforward migration of data/visualization tools. These nodes would make the system somewhat unique, in that hardware-assisted graphics nodes would reside directly on the platform. These nodes would provide dedicated data manipulation resources, as on other recent ASC platforms, but also fast 3D rendering, at a negligible additional cost for the graphics cards themselves. The availability of rendering services directly on the platform negates the need for any transmission of data off of the machine, other than the resulting images. Among other possibilities, having such nodes may be more conducive to run-time visualization, in which visualization is performed in concert with the running simulation. The ultimate procurement and deployment of these nodes is still uncertain at this time.

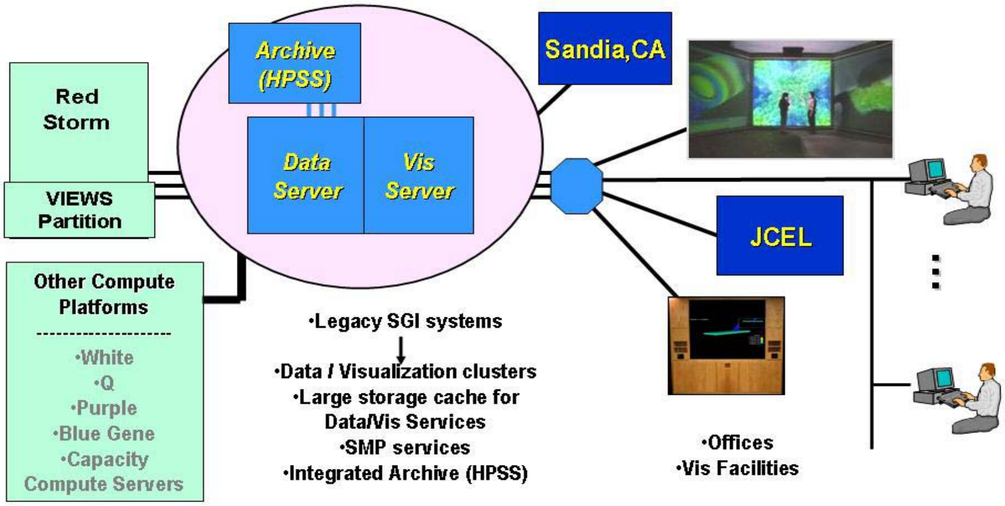


Figure 4: Building a data/visualization "center".

In addition to any data/visualization components on the Red Storm platform itself (i.e., the "VIEWS Partition"), Sandia is building a data / visualization "center" that provides a breadth of services for data-intensive processing associated with high performance computing. The notion is to provide a center for pre and post processing, data manipulation services, archival services, etc. In order to accommodate processing of large amounts of data, this center will incorporate large server(s) and a very large storage system, to be used as a data cache of sorts, where applications can deposit data for a few hours to a few weeks or months, as needed to support current intensive data processing activities. The center as a whole will be well integrated with other components in the environment, which include a more complete set of compute servers available to Sandia computational analysts. In fact, mechanisms are being explored that would allow the center's storage system to be shared directly by many of the local Sandia platforms (using the CFS Lustre[10] parallel file system).

Where the large data/visualization servers of old were once SGI Infinite-Reality-based systems, they are being replaced by data/visualization clusters. Sandia has been building and using graphics clusters for some time -- about 5 years now. Current data/visualization cluster deployments at Sandia include:

> "Feynman" – A 3 year old unclassified development platform for scalable visualization and data services. Includes 128 graphics nodes and 80 compute nodes, 15 TB of RAID storage, and .2 PB of HPSS (tertiary) tape storage.
>
> "DaVinci" – A production classified system that stands up the earliest parts of what are to evolve into "Red RoSE" (see Red Storm Environment clusters below). Currently compute processors only, no graphics. 8 TB of RAID storage and .2 PB of HPSS tape storage.
>
> "Wilson" – The oldest of our current clusters at about 5 years old. A 64-node graphics cluster that is largely dedicated to Sandia's "VIEWS Corridor"; the cluster is used to drive a 48-projector, 60 million-pixel tiled display.
>
> "RoSEbud" – A relatively new 12-node graphics cluster deployed in a visualization laboratory in Sandia's Joint Computational & Engineering Laboratory (JCEL) building. Each node supports dual-ported display outputs to enable driving a 6x4 tiled display. (A similar 24-node cluster has also been procured that is expected to be deployed in the same facility for classified use.)

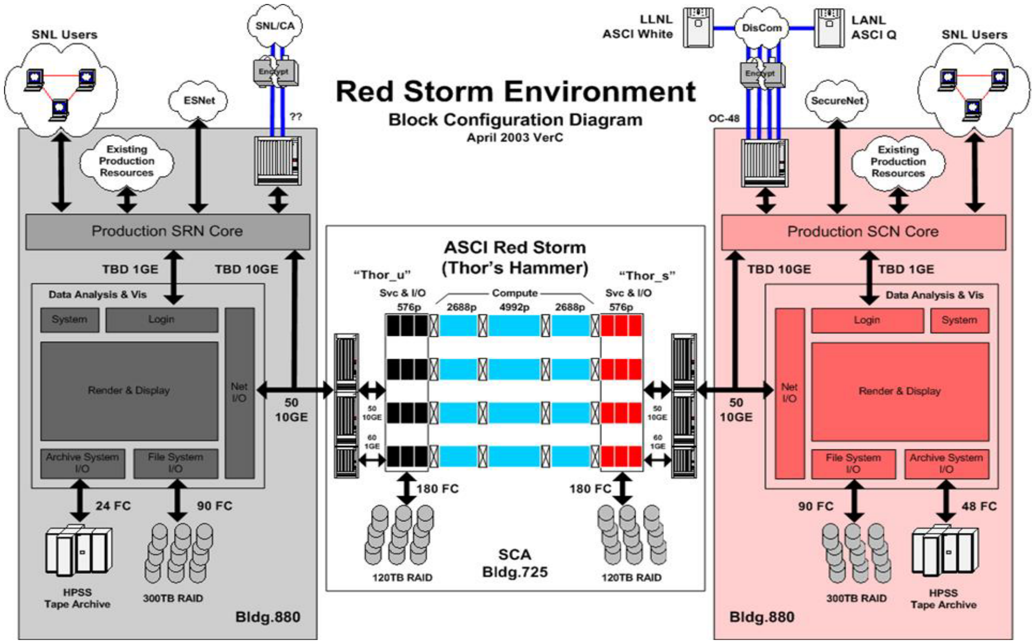
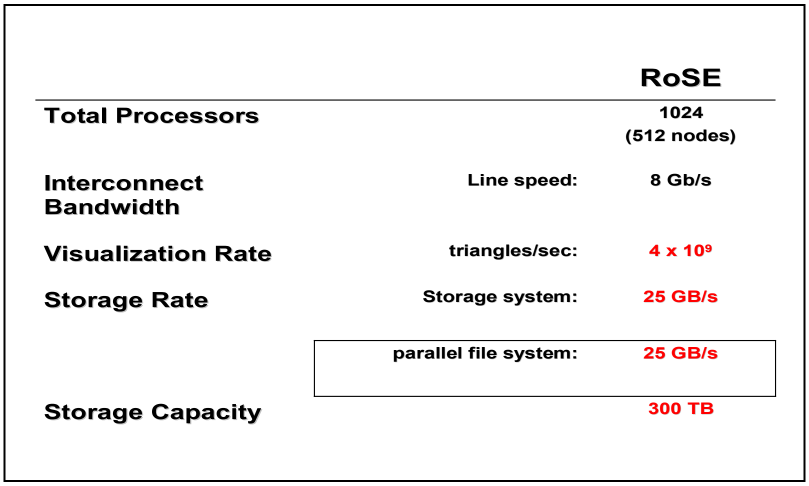


Figure 3: The Red Storm platform (center) and associated "RoSE" clusters, which are at the heart of Sandia's data / visualization center.

Efforts are currently underway to build and deploy Red Storm Environment ("RoSE) clusters that would serve at the heart of the data/visualization center described above (see Figure 5). Current RoSE hardware on hand includes:

- 136 dual Xeon nodes (3.06 GHz, 4 GB RAM) that are meant to serve as storage and external I/O nodes.
- 264 dual Xeon graphics nodes (3.6 GHz EM64T, 4 GB RAM), NVIDIA Quadro FX 3400 graphics cards, PCI-Express-based.
- A 4X Infiniband interconnect, providing ~800 MB/sec measured bandwidth per port.
- 480 TB SATA RAID storage, supporting ~20 GB/sec aggregate bandwidth.

The targeted configurations for the RoSE clusters (see also Table 1) are designed to be well matched with the characteristics and capabilities of the Red Storm platform. Fully deployed configurations would support 25 GB/sec external I/O, and 25 GB/sec to the RoSE-attached parallel file system – this is meant to allow data to be deposited to RoSE storage as fast as it can be taken off of the Red Storm platform. Ultimately realized configurations for "Red RoSE" and "Black RoSE" will depend on available funding.



		RoSE
Total Processors		1024 (512 nodes)
Interconnect Bandwidth	Line speed:	8 Gb/s
Visualization Rate	triangles/sec:	4 x 10^9
Storage Rate	Storage system:	25 GB/s
	parallel file system:	25 GB/s
Storage Capacity		300 TB

Table 1: Targeted design specifications for the RoSE clusters.

The RoSE configurations targeted for delivery in spring/summer of 2005 are as follows:

"Red RoSE" – the classified system.

- 8-12 GB/sec file system bandwidth, ~200 TB capacity
- 8 GB/sec network bandwidth to Red Storm
- 264 visualization nodes
- Infiniband interconnect
- 20 tape drives for archival storage; the tape silo has ~2 PB available capacity

"Black RoSE" (to be fulfilled using aging Feynman parts for now) – the unclassified system.

- 4-6 GB/s file system bandwidth, ~100 TB capacity
- 8 GB/s network bandwidth to Red Storm
- 128 visualization nodes, 80 compute nodes
- Myrinet interconnect
- 16 tape drives for archival; the tape silo has ~2 PB available capacity

Note that the high performance connectivity between the Red Storm platform and the RoSE clusters, together with the large disk storage capacity in the RoSE data center, will enable fast migration of data off of the Red Storm platform that is targeted for analysis or tape archival.

Completing the end-to-end infrastructure, commodity-based graphics workstations are in common use as engineering-science office systems at the laboratory. Gig-Ethernet connectivity is also becoming increasingly available into technical office environments. The challenge, then, is to provide software tools and seamless access to high performance services that deliver the full power of the infrastructure to the desktop.

4 Software Tools and Results

Sandia, together with its ASC partners, has been working to develop software tools for data analysis and visualization that support high performance computing applications. The tools can be broadly characterized as visualization tools and/or data management/manipulation tools.

The commercial EnSight-Gold product (whose development has been supported by ASC to address high performance and other ASC requirements), provided by CEI, is in use as a production visualization tool at Sandia, and is available at all of the ASC tri-labs. An important feature is the EnSight Server-of-Server capability, which has been in use for some time in our ASC tri-lab environments to support distributed visualization of large data sets, without forcing movement of the complete data set. This feature enables use of parallel servers on a supercomputer (where simulation results originate), for example, to extract 3D surfaces from 3D volumetric data, which can be passed to a distributed client (even a thousand miles away) for interactive graphical-rendering. This model for using EnSight will be supported on the RoSE clusters – it will also be supported directly on the Red Storm platform if VIS-nodes are ever procured and deployed, or if other service nodes on the platform can be used instead.

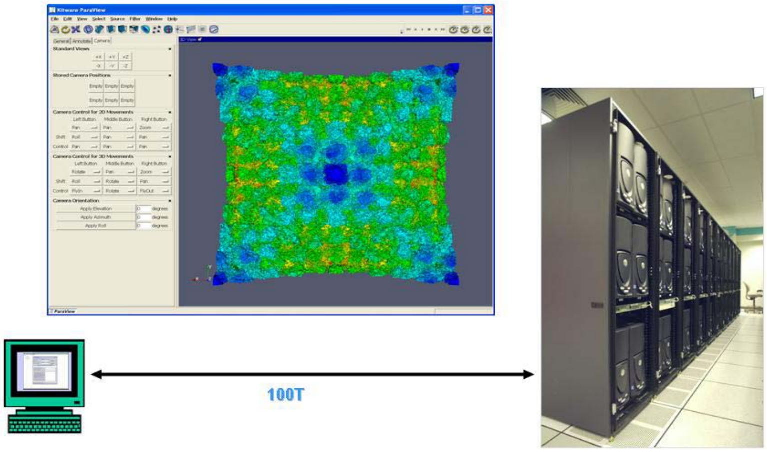
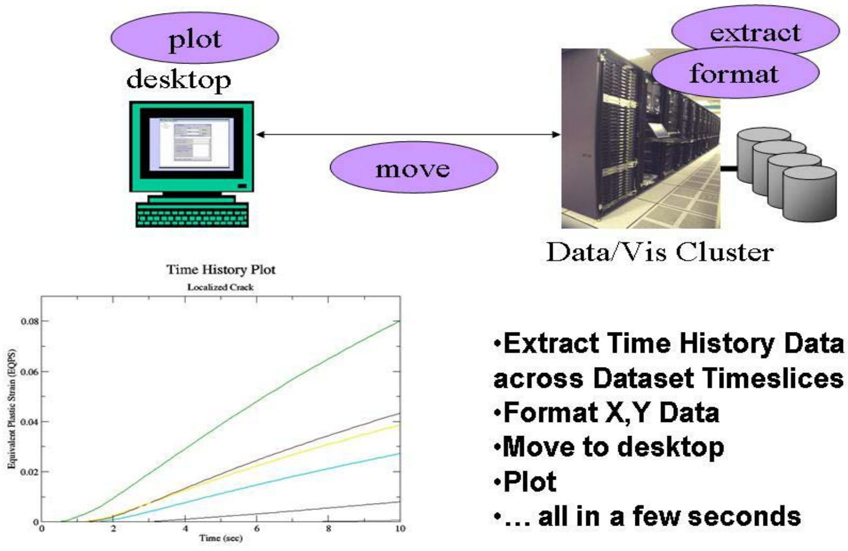


Figure 4: The ParaView visualization application has been used in conjunction with Sandia's visualization clusters to demonstrate scalable interactive visualization of large datasets, such as the 470-million triangle turbulence-data surface shown above. This figure depicts the distributed workstation ↔ cluster configuration used to demonstrate a

rendering performance of 1.5 billion polygons per second, and up to as many as 15 frames per second for some datasets, even when constrained by a 100T connection.

The open-source VTK and its associated ParaView application are being used at Sandia to support advanced in-house development and the rapid development and delivery of custom features that address in-house application requirements. Sandia has become a substantial contributor to VTK and ParaView, introducing features that enable scalability (including "ICE-T" parallel rendering[11]), desktop-image compression and delivery software ("Squirt"), and certain advanced techniques (e.g., the correct rendering of higher order unstructured elements). The resulting stock ParaView application has been used, together with 128 new RoSE graphics-nodes, to demonstrate rendering performance, delivered to the desktop, at 1.5 billion triangles per second (see Figure 6; see also joint Sandia-NVIDIA press release[12]). The image delivery software has produced seemingly impressive frame rate results also – in experiments limited by a 100 Base T connection, frame rates as high as 15 frames per second have been observed. While precise measurements have not been taken, the software-based compositing and image delivery software, which must read back image data from the graphics card(s) that do the rendering, are apparently benefiting from the faster data transfer rates provided by PCI-Express.

To give a sense of how much progress has been made in recent years, the turbulence data shown in Figure 6, when rendered using a single SGI Infinite Reality pipe about 5 years ago, took between 6 and 10 minutes to render for a single frame. The same data is now being rendered and delivered to the desktop at 3 frames per second.



- •Extract Time History Data across Dataset Timeslices
- •Format X,Y Data
- •Move to desktop
- •Plot
- •... all in a few seconds

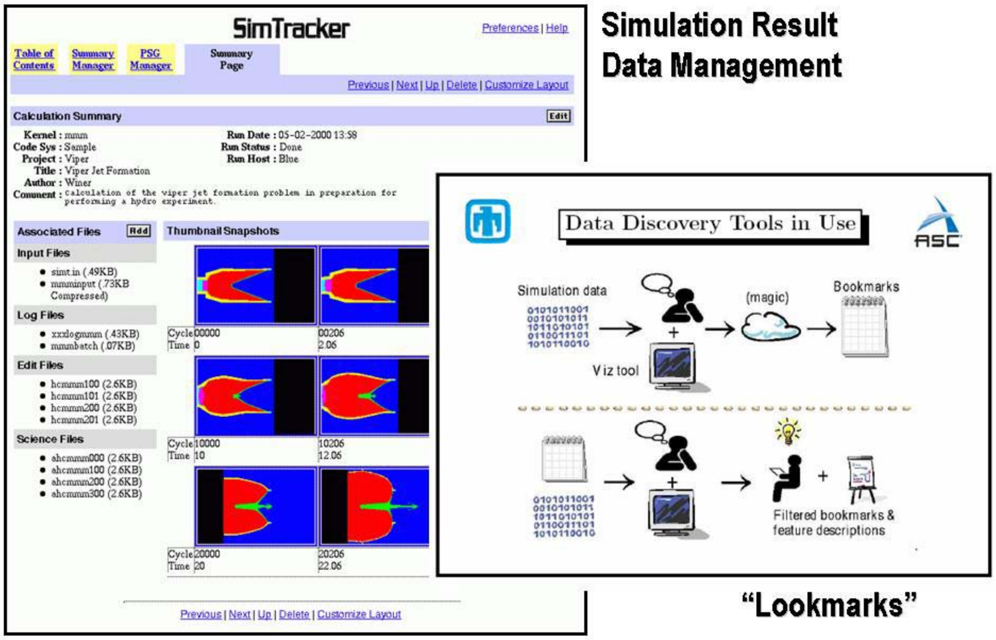
Figure 5: A data service example illustrating the use of high performance data/visualization cluster services to manipulate large data, in this case extracting a relatively small amount of data which can then be successfully studied on the local workstation.

As noted earlier, data analysis and visualization is about more than just rendering. While full-featured visualization tools normally include certain data manipulation capabilities, Sandia is working to complement such tools with data service tools that deliver functionality such as data management, data query, data reduction, data extraction, data transformation, data mining, and data derivation. A usefulness of such tools is captured in the example shown in Figure 7.

A beta-release product, the “Data Services Tool Kit (DSTK)”, is now available at Sandia. Initial features include:

- A Python-based scriptable interface
- Data extraction, output, math, filtering, selection, mesh modification, query, and subsetting functions
- Integrated plotting
- Matlab output
- Support for Sandia’s Exodus 2 and SAF data formats

The product internals are designed to support parallelism, but parallelism is currently enabled only through the use of parallel Python.



SimTracker

Preferences | Help

Table of Contents | Summary Manager | PSG Manager | Summary Page

Previous | Next | Up | Delete | Customize Layout

Calculation Summary Edit

Kernel : mmm
Code Sys : Sample
Project : Viper
Title : Viper Jet Formation
Author : Winer
Comment : calculation of the viper jet formation problem in preparation for performing a hydro experiment.

Run Date : 05-02-2000 13:58
Run Status : Done
Run Host : Blue

Associated Files Add

Input Files

- simt.in (.49KB)
- mmminput (.73KB Compressed)

Log Files

- xxxlogmmm (.43KB)
- mmmbatch (.07KB)

Edit Files

- hcmmm100 (2.6KB)
- hcmmm101 (2.6KB)
- hcmmm200 (2.6KB)
- hcmmm201 (2.6KB)

Science Files

- ahcmmm000 (2.6KB)
- ahcmmm100 (2.6KB)
- ahcmmm200 (2.6KB)
- ahcmmm300 (2.6KB)

Thumbnail Snapshots

Cycle Time	00000 0	00206 2.06
Cycle Time	10000 10	10206 12.06
Cycle Time	20000 20	20206 22.06

Previous | Next | Up | Delete | Customize Layout

Figure 6: Tools to help manage simulation results data and discover/mark features of interest in simulation data are also being developed and integrated into Sandia’s suite of post-processing tools.

A couple of other examples of data management tools that are influencing Sandia’s data/visualization tool suite are captured in Figure 8. “SimTracker” is a tool that originated at LLNL and has been adopted somewhat by the ASC tri-labs. It uses metadata to capture and present information about simulation runs, allowing the analyst to manage simulation data at a higher level than just file names. Approaches like this are under consideration for simplifying the modeling and simulation workflow that computational analysts use to complete their work.

"Lookmarks" borrows from the "Bookmarks" feature of internet browsers. The notion is to be able to mark specific locations in data that are of special interest while visually browsing the data. Data characteristics at these locations can then be used to generate saliency attributes which might then, in turn, be used to automate the location of similar features in subsequently explored data. A manual Lookmark feature has been implemented for use with VTK/ParaView.

5 Acknowledgements

The data/visualization team at Sandia is centered in three Departments managed by the two authors and Jerry Friesen – the environment presented herein is the result of the collective work of those organizations and a few others. At the risk of naming individuals whose work is represented, they include: Brian Wylie, Ken Moreland, Lee Ann Fisk, Lisa Ice, David Karelitz, Jon Goldman, John Greenfield, David Thompson, Jeff Jortner, Gary Templet , Rena Haynes, Brett Clark, Jake Jones, Daniel Sands, Philip Kegelmeyer, Wendy Koegler, Eric Stanton, Vic Holmes, Han Lin, Milt Clauser, Steve Monk, Chris Maestas, Randy Scott, David Logsted and Carl Diegert.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

6 References

[1] Constantine Pavlakos and Philip Heermann, "Issues and Architectures in Large-Scale Data Visualization", The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, Elsevier Inc. 2005, chapter 28, pages 551-568.

[2] Philip Heermann and Constantine Pavlakos, "Desktop Delivery: Access to Large Datsets", The Visualization Handbook, chapter 25, pages 493-510.

[3] http://www.ceintl.com/

[4] http://www.kitware.com/

[5] http://chromium.sourceforge.net/

[6] K. Li et al, "Building and Using a Scalable Display Wall System", IEEE Computer Graphics and Applications, pages 29-37, IEEE Computer Society, July/August 2000.

[7] http://www.cs.utah.edu/research/areas/graphics/

[8] http://dmx.sourceforge.net/

[9] http://hdf.ncsa.uiuc.edu/HDF5/

[10] http://www.clusterfs.com/

[11] K. Moreland, B. Wylie and C. Pavlakos, "Sort-Last Parallel Rendering for Viewing Extremely Large Data Sets on Tile Displays", 2001 Symposium on Parallel and Large-Data Visualization and Graphics, Proceedings, IEEE, October 2001.

[12] http://www.nvidia.com/object/IO_19962.html