Distributed data and immersive collaboration.

by Daniel A. Reed, Roscoe C. Giles and Charles E. Catlett

The National Computational Alliance is developing technologies that will provide distributed data archives and high performance collaborative environments for the members of virtual research teams. The alliance has thus assembled application technology, enabling technology, scalable input/output, data and collaboration, and desktop access teams tasked to design and test the technologies and formats that will enable these teams to focus on their science and engineering tasks instead of on the mechanics of data management.

© COPYRIGHT 1997 Association for Computing Machinery Inc.

Eliminating the barriers of time and space separating potential collaborators means virtual research teams can assemble quickly via distributed data archives and high-resolution, high-modality collaborative environments.

TODAY'S EXPERIMENTAL SCIENTIST OR ENGINEER IS A CONSTANT INTERNATIONAL TRAVELER, MEETING COLLEAGUES AND ANALYZING DATA AT WIDELY DISPERSED SITES. oBSERVATIONAL ASTRONOMERS SPEND TIME AT MAUNA KEA, KITT PEAK, THE HUBBLE SPACE TELESCOPE SCIENCE INSTITUTE, AND THE NATIONAL RADIO ASTRONOMY OBSERVATORY, EXPLOITING THE SPECIAL CAPABILITIES OF INSTRUMENTS AT EACH SITE AND CORRELATING INCREASINGLY LARGE VOLUMES OF OBSERVATIONAL DATA WITH THAT OBTAINED BY OTHER COLLABORATORS.

Although much briefer, the history of computational science follows the same path as that of experimental science, evolving from single independent investigators to encompass distributed, collaborative, multidisciplinary groups and massive computational science datasets. Today's computational science and engineering research teams consist of multi-institutional and cross-disciplinary groups requiring advanced capabilities for multiple investigators to continually interact while using multiple remote computer data repositories, scientific instruments, virtual environments, and teraflop compute engines.

In the early 1980s, computational science meant a researcher, his or her code, and a vector supercomputer. Computing cycles were scarce, and individual researchers traveled to national laboratories or foreign supercomputer centers for local access. By the late 1980s, NSFnet (begun as a 56kbps backbone connecting the National Science Foundation) made possible remote access to high-performance computing systems, and tools like NCSA Image and NCSA Dataslice brought scientific visualization to the desktop.

In the 1990s, high-performance scalable parallel systems, fueled by the microprocessor performance revolution, enabled Grand Challenge teams of computational and computer scientists to attack previously intractable problems (e.g., computational cosmology models of the large-scale structure and evolution of the universe, as discussed by Ostriker and Norman in this issue). To glean insights from the multiple-gigabyte datasets produced by these multiscale simulations, researchers turned to high-resolution volumetric visualizations and virtual environments like the CAVE Automatic Virtual Environment (CAVE) for interactive data navigation and correlation with experimental data [1].

Today, cutting-edge computational science and engineering require distributed access to very large data archives, sophisticated information mining and visualization techniques, and collaborative exploration and data analysis. For example, the National Computational Science Alliance's multidisciplinary Environmental Hydrology Application Technologies (AT) team is expanding current models of the fluid flow, chemistry, and biology of the Chesapeake Bay and adjacent Atlantic Ocean coastal region. The goal is to incorporate important processes influencing shelf circulation, such as wind, river inflow, and offshore currents. Important practical applications of this research include predicting crab larva dispersal and determining how the ocean currents affect the trajectories of oil spills. These developments will be pursued in partnership with other NCSA-affiliated programs, like the U.S. Defense Department's Modernization Program [ILLUSTRATION FOR FIGURE 1 OMITTED].

Creating such a comprehensive model requires not only linking multiple scientific models (e.g., atmospheric precipitation, surface water flow, and watershed ecosystems) but integration of parallel software tools, processing of multiple-terabyte datasets, and interactive, immersive visualization and collaboration tools for coupling computational and computer scientists with scientific data. Equally important, though often overlooked, is the need for ubiquitous desktop collaboration tools for unobtrusive, daily information sharing and discussion.

Each of the other Alliance AT teams has similar distributed data access and multidisciplinary interaction requirements. Taken together, the needs of the AT teams have three

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity

Distributed data and immersive collaboration.

profound implications for future data management and collaboration:

* High-performance, distributed storage and information management. Distributed access to massive datasets necessitates interoperable, high-performance secondary and tertiary storage systems supporting intelligent data mining and retrieval.

* High-modality, collaborative visualization and virtual environments. Data access and collaboration mechanisms need to be ubiquitous, with universal desktop access as well as high-resolution displays and virtual environments.

* Ubiquitous desktop data access and collaboration software. The spatial and temporal separation of research groups, coupled with massive datasets and multidisciplinary applications, requires high bandwidth, multimedia support for collaboration, data exploration, computational steering, and replay.

The sustained I/O performance of secondary and tertiary storage systems today is substantially less than that achievable for computing systems, with little interoperability and limited support for intelligent data mining. Similarly, distributed collaboration and exploration of massive datasets via shared desktop and immersive virtual environments are possible with only a handful of expensive research prototypes. The limitations of extant data management and collaboration tools constrain the range of computational approaches by limiting access to (local and remote) multiple-terabyte datasets and preclude natural modes of interactive, collaborative data exploration and display.

The goal of the Alliance's Data and Collaboration Enabling Technologies (ET) team is to make ubiquitous what is now available only as point demonstrations and research prototypes. Put simply, the goal is to enable a new model of distributed, multidisciplinary computational science for the 21st century.(1)

High-Performance Distributed Storage and Information Management

Despite the clear importance of high-capacity, high-performance I/O software and hardware for scalable computing systems, scalable I/O remains the poor stepchild of the computation, communication, and storage triumvirate, with sustained I/O rates measured in MB/sec on systems with near-teraflop processing rates. To enable efficient creation and processing of and distributed access to massive datasets, the Data and Collaboration team is working with the staff of the NCSA Leading Edge site to integrate three tiers of I/O services:

* Advanced file systems for application-directed secondary storage I/O

* Interoperable hierarchical storage management (HSM) for unified tertiary storage access

* Parallel object/relational databases for structured data manipulation and mining

High-performance, parallel I/O. Performance increases for secondary storage devices have not kept pace with processor and network performance increases, creating a hardware I/O bottleneck on high-performance computing systems. More perniciously, current commercial parallel file systems deliver only a small fraction of hardware I/O bandwidth to parallel applications, and their performance is extremely sensitive to even small changes in application I/O request patterns [9]. The widening disparity between computation and I/O rates is now a major impediment to achieving high performance for a range of Grand Challenge applications, especially for multidisciplinary applications with dynamic, irregular behavior and time-varying I/O demands, such as those being developed by Alliance AT partners.

An extensive analysis of application and physical I/O patterns by multiple research groups has shown that no single parallel file policy is likely to yield high performance for every application code. Instead, next-generation file systems must aggressively exploit I/O access pattern knowledge in caching and prefetching systems to obtain a substantial fraction of peak hardware I/O performance.

Few vendors today support parallel file systems, and standard APIs for parallel I/O are only now emerging. Consequently, most scientific application developers eschew parallel I/O systems in favor of more portable, though very-low-performance, application-specific solutions. In a vicious circle, parallel file system and storage hierarchy designers have little empirical data on parallel I/O access patterns to guide design of parallel I/O APIs and file systems. The Message Passing Interface I/O is an example of a promising new API (http://lovelace.nas.nasa.gov/MPI-IO).

To redress the limitations of current parallel file systems, the Alliance will build on the results of the national Scalable I/O (SIO) Initiative. This broad attack on the I/O problem exploits application I/O characterization to drive development of APIs encapsulating common I/O access patterns, creates parallel network and file-system policies adaptable to changing access patterns, and explores compilation techniques for optimizing low-level I/O accesses from high-level specifications. Because the SIO initiative includes leaders from the Alliance and from the

- Reprinted with permission. Additional copying is prohibited. -
GALE GROUP
Information Integrity

Distributed data and immersive collaboration.

NSF's partnership programs, it targets all the major hardware platforms in the PACI program, in both the Alliance and the National Partnership for Advanced Computational Infrastructure. In addition, SIO members are working closely with the Department of Energy's Accelerated Strategic Computing Initiative, an effort in nuclear weapons stockpile stewardship based on integrated computational simulation rather than on continued testing via underground detonations. (For more on SIO, go to http://www.cacr.caltech.edu/SIO.)

As part of SIO, the Data and Collaboration team will work with the AT teams to optimize I/O performance on the Alliance's new Hewlett-Packard/Convex SPP-2000 and Silicon Graphics/Cray Origin 2000 distributed shared-memory systems as well as large-scale Windows/NT clusters. The team will measure, analyze, and tune application I/O behavior. In addition, it will implement and test new parallel file system APIs and an associated set of flexible file system policies and compilation techniques. Collectively, these efforts will create the toolkit needed to achieve high-performance file I/O on all Alliance hardware platforms.

Distributed data archives and data mining.
High-performance access to secondary storage is necessary but not sufficient to enable effective distributed collaboration. Because parametric computational science simulations produce far more data than can be retained on secondary storage and because these simulations execute at multiple partner sites, computational scientists generally interact with a number of popular archive systems (for example, UniTtee, HPSS, and Cray Data Migration Facility DMF). For the most part, data is stored and accessed as entire files, and ftp is the common-denominator user interface among the systems.

Consequently, analyzing datasets from multiple archives currently requires that users collect the relevant files at a single site. This requirement is unwieldy and impractical for large numbers of multiple-gigabyte files. Further, the model does not scale well to multiple researchers collaborating from multiple sites.

Thus, the need to analyze and correlate very large, multidisciplinary datasets (for example, geospatial data, such as that produced by NASA's Mission to Planet Earth and analyzed by the Alliance's Environmental Hydrology team) is poorly matched to the current model of whole-file access via hierarchical storage management systems. Instead, such analysis involves a variety of nonsequential access patterns. Consequently, new data archives must decouple data access from the storage format and enable access at a finer granularity than entire files.

To separate data access and storage for scientific applications, NCSA developed the hierarchical data format (HDF) API and library. HDF was created specifically for scientists to store and exchange data, including images, multidimensional arrays, tables, and data attributes. With HDF, scientists can easily define and assemble data collections suited to their specific needs that are also portable across architectures and computing platforms. Reflecting its flexibility and generality, HDF was selected by NASA as the primary file format for its Mission to Planet Earth, a project that will eventually process more than 2TB of new data per day from dozens of instruments. (For more on HDF, go to http://hdf.ncsa.uiuc.edu.)

Building on the success of HDF in the sequential computing domain, the Alliance is extending it to support both data-parallel and message-passing environments, and a new implementation of the HDF library is being designed to support very large objects and files.

Self-describing data metaformats and libraries like HDF can simplify data management by collaborative groups but are still based on an underlying model of files and secondary storage. Alliance team members from the University of Wisconsin are leading an effort to apply Paradise, an advanced, parallel object-relational database management system (ORDBMS), for scalable access to multiple-terabyte data archives. As part of this effort, the Paradise group will continue implementing an HDF compatibility library, enabling existing HDF-based applications to gain seamless access to Paradise's advanced features. (For more on Paradise, go to http://www.cs.wisc.edu/paradise.)

Like other ORDBMSs, Paradise supports nontraditional applications via an extensible type system. In addition to the normal scalar types, Paradise's type system includes multidimensional arrays, geolocated raster images (for storing satellite imagery), text, video, and a full set of spatial data types, including points, polylines, and polygons. Paradise is unique in its integrated support for tertiary storage. Individual columns of a table or whole tables can be stored on tape. By integrating support for tertiary storage and the database system (instead of using a separate hierarchical storage manager), the Paradise query optimizer and execution engine can optimize execution of queries to tape-resident data.

Finally, to create a seamless environment for network computing and distributed collaboration, the Data and Collaboration team will exploit ORDBMS support for objects to create a database that can act as a common name service for multiple HSM systems, allowing multiple Alliance and PACI-wide HSMs to be distributed and tightly integrated across computational facilities.

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity

Distributed data and immersive collaboration.

To support the Alliance's development of next-generation parallel file systems and distributed-data-archive software, NCSA is dramatically upgrading its secondary and tertiary storage systems. Currently based on UniTtee and running on a Hewlett-Packard/Convex C3820 system, the archive today contains roughly 30TB of data, with monthly traffic of 2.5TB (input) and 750GB (output). Secondary storage consists of a 300GB disk pool, with tertiary storage using robotic tape libraries with IBM 3590 (10GB) tapes.

With introduction of the Silicon Graphics/Cray Origin2000, typical secondary storage available for computations has increased from 20-30GB to 200-300GB, resulting in larger files to be archived. As an example of the anticipated increase in storage requirements, a recent project with an Alliance partner, the University of Minnesota's Laboratory for Computational Science and Engineering (LCSE), produced 2TB of data during an eight-day simulation on 128 processors on a Silicon Graphics/Cray Origin2000. This computation resulted in a 544 x 544 x 544-resolution model of a star the size of the Sun.

Due to the extraordinary size of this computation, all of the data was stored directly on 165GB Ampex D2 tapes. To provide very large data handling capacity as a general service, NCSA is increasing the capabilities of the current UniTree and working with Silicon Graphics/Cray to integrate Cray's DMF directly with the Origin. The UniTree archive will be run on a Hewlett-Packard/Convex Exemplar SPP-2000 with 1TB of disk cache and use IBM 3590 tape drives. UniTree will provide center-wide backup facilities and general-purpose archiving for anyone in the Alliance. The DMF archive will be tuned for high performance for very large datasets. The initial system will use an eight-processor Origin, 1TB of disk cache, and high-performance tape drives.

Enhancements to DMF will allow data to be transferred directly from the Origin disk to tertiary storage, avoiding having to make an extra disk copy to DMF disk cache. These efforts will tightly integrate processors and memory with the archive, reducing latencies and data copying.

A variety of other archive systems are available across the Alliance, including HPSS at the Maul High Performance Computing Center, ADSM at Argonne National Laboratory, and DMF at the Ohio Supercomputer Center. The Alliance plans to create a seamless environment for networked computing and distributed collaboration by allowing uniform, secure access to each site's archive.

High-Modality, Collaborative Visualization, and Virtual Environments

To enable intuitive, immersive data exploration and distributed collaboration, the Data and Collaboration [TABULAR DATA FOR TABLE 1 OMITTED] team will integrate three virtual environment software technologies for distributed collaboration: multimedia interaction; capture/replay of interaction experiences; and real-time software manipulation and optimization. This work builds on a technology development and transfer effort derived from early demonstrations of the CAVE at Siggraph '92, Siggraph '94, and the pioneering I-WAY [2] at Supercomputing '95.

Technology deployment. The Alliance is deploying CAVEs, ImmersaDesks, and PowerWalls at partner sites to support high-modality exploration of complex data and real-time computational steering. The goal is to enable researchers to do problem solving by visualizing and interacting with the results of their simulation codes in real time using the CAVE and Immersadesk projection-based virtual-reality displays developed by the University of Illinois at Chicago's Electronic Visualization Laboratory (EVL), the PowerWall developed by Minnesota's LCSE, and the large-screen stereo Infinity Wall, jointly developed by EVL, LCSE, and NCSA. Table 1 lists all sites that currently own projection-based visualization and virtual reality displays.

The CAVE is a room-sized virtual environment based on rear-projection displays and high-performance Silicon Graphics Infinite Reality Engines that allow users to interact with virtual objects using a wand and simple LCD shutter glasses [ILLUSTRATION FOR FIGURE 2 OMITTED]. To support CAVE-like collaboration and computational steering throughout the Alliance, the ImmersaDesk, a lower-cost version of the CAVE supporting a single display - one wall of the CAVE - is being deployed for use by Alliance members.

EVEs CAVE library has all the functions necessary to create a CAVE program, whether to run in the CAVE, on the ImmersaDesk, or on the Infinity Wall; programs written for one device automatically port to the others. The library deals with the synchronization of all the devices, the synchronization of the walls, the calculations of the stereo transformation, and many other virtual reality-specific tasks. LCSE's PowerWall is a large-screen, high-resolution, monoscopic projection display capable of 7.6-million pixel resolution for the visualization and analysis of very-high-resolution data. The large size and extraordinary detail of these displays bring fidelity and scale to other computer displays and virtual environments comparable to that previously experienced only in IMAX theaters.

LCSE is working with Seagate Technology to develop RAID systems based on Seagate's new 9GB 3.5-inch disks with 100MB/sec Fibre Channel interfaces. While the

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity

ultimate goal of visual supercomputing (see Kennedy et al. in this issue) is to do data simulation and rendering in real time, LCSE's current approach is to compute, then render, then store a hierarchy of data (either spatial, like volume data, or temporal, like animation). The images are then "played back" off disk onto tiled screens to form images at higher resolutions. LCSE intends to perform this rendering in parallel on multiple rendering engines installed in multiple machines and driving displays on multiple panels of a large, very-high-resolution next-generation PowerWall, the Great Wall of Power [ILLUSTRATION FOR FIGURE 3 OMITTED]. This system can scale upward and downward in resolution, size, and cost.

By attaching the visual supercomputer's video output to ATM networks, EVL, LCSE, NCSA, Argonne, and the University of Utah will develop remote steering capabilities for users with ATM access but no access to advanced graphics hardware. This concept, based on current work by LCSE, EVL, and NCSA, uses lower-resolution graphics as a "window into the oven" during a computation, allowing the remote user to select a visual path through the data, then render it at higher resolution after the computation has been completed.

Broad access to CAVEs, ImmersaDesks, and PowerWalls will enable individual researchers as well as groups of dispersed collaborators to exploit high-resolution displays and virtual reality environments for interaction with massive datasets. Future Silicon Graphics computer rendering speeds and large memory could make the beautiful high-resolution images produced for the recent Cosmic Voyage IMAX sequences (requiring rendering a 165GB dataset created from thousands of hours of computer runs at NCSA and SDSC) available interactively to scientists using the Leading Edge facility.

Multimedia interactions. For physically distributed research teams, virtual environments are necessary but still inadequate; many of the most important interactions among collaborating team members are spontaneous, centering on discussions of shared computational and experimental data. To be effective vehicles for group interaction, virtual environments must also support collaborative manipulations with integrated audio/video for, group discussions. In such a shared virtual environment, manipulation of scientific visualizations, application steering, or software performance optimization must be easily passed among the participants. Moreover, the locations of objects within the space as well as discussions and manipulations of the environment's objects must be available to all participants. That is why the Data and Collaboration team is building on the CAVE software toolkit.

First, the Distributed Computing ET team (see Stevens et al. in this issue) will exploit EVL's prototype CAVERN (CAVE Research Network) software [3] and extend the CAVE and PowerWall toolkits to support linked visualization and virtual environments with massive image stores. CAVERN currently supports two-way interaction between a pair of sites; the extended version will support multiple interactions, the ability to enter and leave sessions anytime, and active shared manipulation of an environment's objects. (A recent EVL collaborative virtual reality educational application that uses CAVERN, called NICE (Narrative, Immersive, Constructionist/Collaborative Environments), is 16-way collaborative [6-8].)

A second effort will map live digital audio and video from collaborating sites onto objects within virtual environment applications. Leveraging public-domain protocols and standards like those used in the successful multicast backbone (called MBONE), the toolkit will include real-time transport clients and servers sending and receiving audio/video streams.

Collaborative interaction, capture, replay. Many studies have shown that recordings of results and histories are key components in collaborative data analysis and user information internalization. These annotations consist of records preserving contextual information, descriptions, and conclusions. If a collaboration environment is to provide a basis for contextual sharing among physically and temporally separated research group members, it must allow users to record interactions and annotate virtual representations and behavior for subsequent playback and review.

Computational scientists and engineers using 3D virtual reality displays have an interface for data exploration far more powerful than the 2D workstation screen; however, without appropriate tools, navigating through large datasets and producing recorded motion sequences are unwieldy tasks. Virtual Director [10] is just such a powerful tool, interfacing real-time data navigation with animation recording and simplifying the steering, editing, and recording of large datasets for the CAVE and ImmersaDesk. Scientists control a "virtual camera" to record visualizations using voice and gesture commands with the ease of operating a Handycam.

Features of an existing prototype version of Virtual Director enable real-time recording, editing, and playback of camera motion through time-varying data. Alliance team members will generalize and extend Virtual Director's functionality to support a variety of data classes (particle, grid, and polygonal models) and rendering techniques (volumetric, isosurface). The resulting API will also drive commercially available renderers (e.g., Alias and

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity

Wavefront) for the highest-quality recording on videotape or video servers. Virtual Director will also enable the recording of CAVE-to-CAVE remote collaborations - the key to supporting asynchronous communication.

Building on Virtual Director and collaboration technology for the capture and intelligent recall of research group activities, team members will develop a virtual reality/video server (called SerVR) for the capture and indexing of virtual experiences, as well as associated navigation software for following spatial, temporal, and logical patterns in replayed experiences. Digital audio/video technology provides the mechanism for recording multimedia interactions; capture/recall technology supports collaboration capture and analysis; and Virtual Director supports the recording and structured editing and annotation of virtual reality sessions.

Direct software manipulation and optimization. Although high-modality visualization and interaction software is routinely applied to scientific data, application developers study and modify complex software via a small workstation window on the complex world of large-scale, heterogeneous distributed computing. To provide high-modality support for configuring, monitoring, and optimizing software across clustered parallel systems, the Alliance will build on its successful design of virtual environments for performance analysis and visualization, as in Figure 4.

This effort, led by the University of Illinois at Urbana-Champaign, will extend existing software to create shared virtual environments to allow scientific research team members to directly manipulate software components and their behavior while immersed in scalable, hierarchical representations of software structure and real-time performance data [5]. The software will support a hierarchy of parallel software components and their interactions, ranging from geographically distributed computations on the Alliance’s networked parallel systems to module interactions within a single task. In this way, it can serve as a distributed monitoring system for the Alliance’s computing and network infrastructure and as an analysis environment for individual applications.

Ubiquitous Desktop Data Access and Collaboration

During the Alliance’s first five years, virtual environment technology and collaboration software will continually migrate from high-end systems to the desktop. This evolution, fueled by declining hardware costs and increasing hardware capabilities, will be accelerated by the Data and Collaboration team’s integration of desktop and immersive collaboration tools. Given the explosion of commercial interest in desktop information sharing and collaboration software (Lotus Notes, Microsoft Net Meeting, and Web-based intranets), the focus will be on desktop collaboration extensions addressing the specific needs of the Alliance AT teams. Concurrently, commercial desktop collaboration tools will be deployed throughout the Alliance.

The first of three major efforts will integrate commercial collaboration tools with the HDF and ORDMBS data access interfaces described earlier, allowing groups to retrieve, discuss, and display scientific data. The second effort will create interaction software for sharing CAVE and ImmersaDesk sessions via Virtual Reality Modeling Language (VRML) windows. Desktop users will be able to participate in immersive discussions and push integration of the underlying software frameworks for collaboration between the high end and many-user, ubiquitous systems. Desktop versions of the audio/video interaction software will also be integrated with their virtual environment equivalents to support the transmission, receipt, capture, and replay of audio/video collaborations via the same intelligent capture-and-recall technology.

Commercial desktop collaboration tools. To provide Alliance-wide access to shared documents, distributed meeting scheduling, and archived communications and plans, as well as support for Alliance team collaboration and management processes, NCSA will deploy an intranet throughout the Alliance based on current Web and Java client technologies as well as on commercial server collaboration tools like Lotus Domino. The Data and Collaboration team will enhance these capabilities with Alliance-developed software to provide shared document repositories, asynchronous and synchronous collaboration and meeting spaces, and a variety of database services and reporting functions.

To support synchronous interactions within and among Alliance teams, all distributed across multiple institutions, the Data and Collaboration team is exploring commercial technology supporting virtual meetings, such as Microsoft’s Net Meeting. An early Alliance effort will seek to augment these meeting tools with the emerging desktop audio/video teleconferencing standards, such as H.323/320.

Application-sharing infrastructures. The Alliance is focusing on collaboration needs specific to science research and science education, anticipating the simultaneous rapid deployment of commercial tools for office-oriented applications and for interpersonal communication. Design requirements include dealing with very large datasets (for example, through use of Globus/Nexus routing, described by Stevens et al. in this issue), interacting within complex graphical analytical

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP
Information Integrity

Distributed data and immersive collaboration.

tools, leveraging existing domain-specific applications, meeting the universal need for robustness in collaborative systems, and supporting a heterogeneous hardware inventory.

Much of this work will capitalize on the NCSA Habanero collaboration infrastructure, a Java-based object framework for synchronous interaction. Alliance members have created viewers for HDF and Flexible Image Transport System data formats and have linked an extant chemical reactor model (the Kansas Engineering and Science Interface system) with Habanero (see McRae in this issue) for discussing results and selecting parameter values. A project with the U.S. Nuclear Regulatory Commission is linking the Logos graphics package, written in Fortran, to Habanero/Java via CORBA IDL. (For more on Habanero, go to http://www.ncsa.uiuc.edu/SDG/Software/Habanero.)

Habanero and two other Alliance collaboration projects - Syracuse University's Tango and the University of Michigan's DistView - function by sharing state information encapsulated in serialized Java objects. A central server may host part of the application, and clients may present differing views of the information. Architectural dissimilarities arise from the fact that Habanero's design focus is on computational scientists, Tango's on education needs, and DistView's on real-time science observations by large teams of the Upper Atmosphere Research Collaboratory community, a University of Michigan experimental testbed for wide-area collaboration among geographically dispersed space scientists.

Tango uses a Web-based architecture, employing server-resident code, Java applets, and plug-ins to the Netscape Web browser. Computational science's need for significant investment in legacy code and to prepositioned large datasets resulted in Habanero's being a Java application, allowing it to go beyond the limitations previously placed on applets by Web browsers. The need for reliable support and software version control for many users led to a DistView architecture based on Web server distribution of Java applets, without plug-ins, and accessible by the major commercial Web browsers [4]. These projects have or are developing community-specific facilities for session capture/replay, privacy, object persistence, and support for large numbers of users. (For more on Tango, go to http://trurl.npac.syr.edu/tango/papers/tangowp.html.)

Conclusions

The Alliance aims to eliminate the barriers of time and space separating potential collaborators, making it possible to quickly assemble and reconfigure virtual research teams. Linking team members via distributed data archives and high-resolution, high-modality collaborative environments will enable them to focus on the essentials of computational science and engineering rather than on the mechanics of data management and information sharing.

1 This work is supported by the NSF's PACI program, effective October l, 1997.

REFERENCES

1. Cruz-Neira, C., Sandin, D., and DeFanti, T. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In Proceedings of Siggraph '93 Computer Graphics Conference (Los Angeles, Aug. 1-6, 1993). ACM Press, New York, 1993, pp. 135-142.

2. DeFanti, T., Foster, I, Papka, M., Stevens, R., and Kuhfuss, T. Overview of the I-WAY: Wide area visual supercomputing. Int. J. Supercomput. Appl. 10, 2/3 (1996), 123-130.

3. Leigh, J., Johnson, A., and DeFanti, T. Issues in the design of a flexible distributed architecture for supporting persistence and interoperability in collaborative virtual environments. In Proceedings of SC'97 (San Jose, Calif., Nov. 15-21, 1997). ACM/IEEE, New York, 1997.

4. Prakash, A., and Shim, H. DistView: Support for building efficient collaborative applications using replicated objects. In Proceedings of the 1994 ACM Conference on Computer-Supported Cooperative Work (Oct. 1994). ACM Press, New York, 1994, pp. 153-164.

5. Reed, D., Elford, C., Madhyastha, T., Smirni, E, and Lamm, S. L. The next frontier: Closed loop and interactive performance steering. In Proceedings of the International Conference on Parallel Processing Workshop (St. Charles, Ill., Aug. 1996). Penn State Press, State College, Penn., 1996, pp. 20-31.

6. Roussos, A., Johnson, J., Leigh, C., Barnes, C., Vasilakis, C., and Moher, T. The NICE project: Narrative, Immersive, Constructionist/Collaborative Environments for learning in virtual reality. In Proceedings of the ED-MEDIA/ED-TELECOM '97, (Calgary, Alberta, Canada, June 14-19, 1997). AACE, Charlottesville, Va., 1997, pp. 917-922.

7. Roussos, A., Johnson, J., Leigh, C., Barnes, C., Vasilakis, C., and Moher, T. NICE: Combining constructionism, narrative, and collaboration in a virtual learning environment. Comp. Gr. 31, 3 (Aug. 1997), 62-63.

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity

Distributed data and immersive collaboration.

8. Roussos, A., Johnson, J., Leigh, C., Vasilakis, C, and Moher, T. Constructing collaborative stories within virtual learning landscapes. In Proceedings of the European Conference on Al in Education (Lisbon, Portugal, Sept. 30 - Oct. 2, 1996). Colibri Editions, Lisbon, Portugal, 1996, pp. 129-135.

9. Smirni, E., and Reed, D. Workload characterization of input/output intensive parallel applications. In Proceedings of the 9th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation (June 1997).

10. Thiebaux, M. Virtual Director, Master's Thesis, Dept. of Electrical Engineering and Computer Science, Univ. of Illinois at Chicago, 1997.

DANIEL A. REED (reed@cs.uiuc.edu) is a professor in and chair of the Department of Computer Science at the University of Illinois at Urbana-Champaign, a member of the Alliance Executive Committee, and co-leader of the Data and Collaboration team.

ROSCOE C. GILES (roscoe@bu.edu) is an associate professor in the Department of Electrical, Computer, and Systems Engineering at Boston University, a member of the Alliance Executive Committee, co-leader of the Data and Collaboration team, and chair of the Alliance Education, Outreach, and Training Committee.

CHARLES E. CATLETT (catlett@ncsa.uiuc.edu) is senior associate director of the Science and Technology Division of the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign.

- Reprinted with permission. Additional copying is prohibited. -

GALE GROUP

Information Integrity