GA–A23741

DATA MANAGEMENT, CODE DEPLOYMENT, AND SCIENTIFIC VISUALIZATION TO ENHANCE SCIENTIFIC DISCOVERY THROUGH ADVANCED COMPUTING

by
D.P. SCHISSEL, A. FINKELSTEIN, I.T. FOSTER, T.W. FREDIAN,
M.J. GREENWALD, C.D. HANSEN, C.R. JOHNSON, S.A. KLASKY,
K. LI, D.C. McCUNE, Q. PENG, R. STEVENS, and M.R. THOMPSON

AUGUST 2001

DISCLAIMER

GA–A23741

DATA MANAGEMENT, CODE DEPLOYMENT, AND SCIENTIFIC VISUALIZATION TO ENHANCE SCIENTIFIC DISCOVERY THROUGH ADVANCED COMPUTING

by
D.P. SCHISSEL, A. FINKELSTEIN,† I.T. FOSTER,‡ T.W. FREDIAN,◊ M.J. GREENWALD,◊ C.D. HANSEN,∆ C.R. JOHNSON,∆ S.A. KLASKY,† K. LI,† D.C. McCUNE,† Q. PENG, R. STEVENS,‡ and M.R. THOMPSON£

This is a preprint of a paper to be presented at the 3rd IAEA Technical Committee Meeting on Control, Data Acquisition, and Remote Participation for Fusion Research, Padova, Italy, July 16–19, 2001 and to be published in *Fusion Engineering and Design.*

†Princeton Plasma Physics Laboratory
‡Argonne National Laboratory
◊Massachusetts Institute of Technology
∆University of Utah
£Lawrence Berkeley National Laboratory

Work supported by
the U.S. Department of Energy under
Contracts DE-AC03-99ER54463, DE-AC02-76CH03073,
W-31-109-ENG-38, and General Atomics IR&D Funds

GENERAL ATOMICS PROJECT 30033
AUGUST 2001

ABSTRACT

The long–term vision of the Fusion Collaboratory described in this paper is to transform fusion research and accelerate scientific understanding and innovation so as to revolutionize the design of a fusion energy source. The Collaboratory will create and deploy collaborative software tools that will enable more efficient utilization of existing experimental facilities and more effective integration of experiment, theory, and modeling. The computer science research necessary to create the Collaboratory is centered on three activities: security, remote and distributed computing, and scientific visualization. It is anticipated that the presently envisioned Fusion Collaboratory software tools will require three years to complete.

1. INTRODUCTION

In 1999, the United States Department of Energy (USDOE) Office of Fusion Energy Sciences (OFES) established the Plasma Science Advanced Computing Initiative (PSACI) to revolutionize fusion research by greatly enhancing simulation and modeling capabilities made accessible by terascale computing. These advanced computational resources aim to improve fusion research by improving scientific understanding of experimental data, by stimulating new theories, and by providing better designs for future facilities. The power of advanced computing to solve critical plasma science problems can be fully exploited only if a capable infrastructure is established and effective software tools are made available. This infrastructure includes establishing standardized data structures and access methods, synthetic diagnostics, standard analysis and visualization utilities, and common code interfaces. Work to date has included support of two PSACI pilot programs: Macroscopic Modeling and Microturbulence Simulation.

For 2001, the PSACI has been folded into the larger SciDAC initiative, Scientific Discovery through Advanced Computing Initiative, created by the USDOE Office of Science. Leveraging the lessons learned from the two PSACI pilot projects, a proposal was submitted to the SciDAC initiative for the design and implementation of a Fusion Collaboratory. This proposal describes a three-year collaboration between experts in the magnetic fusion and the computer science and enabling technology communities to create a Collaboratory. Preliminary notification from the Office of Science indicates that the proposal will be funded beginning in the second half of calendar year 2001. This paper describes the design of the Collaboratory.

Goals of the Collaboratory. The Fusion Collaboratory will advance scientific understanding and innovation in magnetic fusion research by enabling more efficient use of existing experimental facilities and more effective integration of experiment, theory, and modeling. Specifically, this project will create and deploy collaborative software tools throughout the magnetic fusion research community comprised of over one thousand researchers from over forty institutions. Built on a foundation of established computer science toolkits, successful deployment of the Collaboratory will nevertheless require significant computer science research to extend the toolkits beyond their present capabilities. The Fusion Collaboratory will enable networked real–time data analysis and instantaneous communication amongst geographically dispersed teams of experimentalists and theoreticians. This represents a fundamental paradigm shift for the fusion community where data, analysis and simulation codes, and visualization tools will be thought of as network services. In this new paradigm, access to resources (data, codes, visualization tools) is separated from their implementation, freeing the researcher from needing to know about software implementation details and allowing a sharper focus on the physics (Fig. 1). The aim of the Collaboratory is to (1) create transparent and secure access to local/remote computation, visualization, and data servers, (2) develop collaborative visualization that allows interactive sharing of graphical images among control room display devices, meeting room displays, and with offices over a wide area network, and (3) Enable real–time access to high–powered remote computational services allowing such capabilities as between pulse analysis of experimental data and advanced scientific simulations.

Benefits to Fusion. The Fusion Collaboratory will increase the productivity of both the physicists and the code and tool developers. Physics productivity will be increased by (1) enabling more efficient utilization of experimental time on the three large facilities through more powerful between pulse data analysis resulting in a greater number of experiments at less cost, (2) allowing more transparent access to analysis and simulation codes, data, and visualization tools resulting in more researchers having access to more resources, (3) creating a standard tool set for remote data access, security, and visualization allowing more researchers to build these services into their own tools, (4) facilitating the comparison of theory and experiment, and (5) facilitating multi–institution collaborations. The Collaboratory will also increase the productivity of code and tool developers by (1) supporting more users with fewer installations at reduced cost, (2) facilitating shared code development projects resulting in more rapid code creation, and (3) creating a standard tool set for remote data access, security, and visualization allowing these services to be easily built into new tools.

Anticipated New Research Environment. The following usage scenarios illustrate types of collaboration that are presently not possible but that have been requested by the fusion community.

- Enhanced Experimental Operations: In the first minute after a DIII–D plasma pulse, a scientist based at Wisconsin dispatches a detailed physics analysis of the DIII–D data to a Princeton Plasma Physics Laboratory (PPPL) compute farm. This analysis code automatically retrieves data from the DIII–D MDSplus server and saves its results back to the server upon completion. The physicist at Wisconsin generates a visualization comparing the results to other analysis run locally at DIII–D. The visualization is shared and discussed with the DIII–D control room and with collaborators at Lawrence Livermore National Laboratory (LLNL) and PPPL. Based on this information, the experimental team is able to make a decision that affects the configuration of subsequent plasma pulses.
- Remote Data Access and Display: An investigator at Columbia logs onto her workstation and authenticates herself. She then runs an analysis code that compares data from General Atomics (GA) and PPPL, displaying the results in a visualization that can be seen

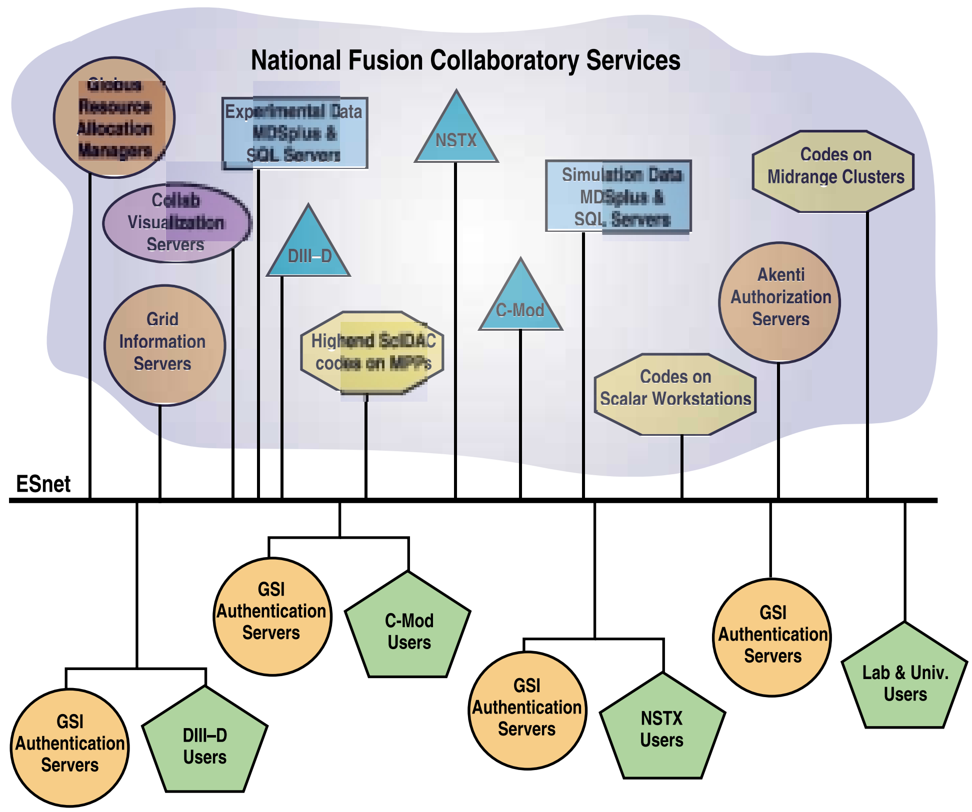

Fig. 1. The Fusion Collaboratory will provide a unified framework in which data, does, and visualization tools are available securely and transparently over the Internet.

simultaneously by researchers at all three sites. Spotting a particular regime of interest, she has the data processed by a powerful rendering engine at GA and an interactive 3D display is made available to all meeting participants.

- Remote/Distributed Analysis: A researcher at MIT acquires data from a local experiment, pre-processes the data on local computers, creates an entry for a code run in a local relational database then sends a message to a queue server at PPPL to run a large analysis job. The queue server verifies his identity, finds that he is authorized to use the code and compute facilities and queues the job on a compute cluster. The job reads data over the Internet from MIT, passing along the credential that identifies him as an authorized user, does its computation and writes the results back into the MIT data archive. It also reaches over the Internet to update the relational database entry to reflect the successful completion of the job.
- Detailed Comparisons Between Simulation Data and Experiments: A principal investigator (PI) from LLNL runs a very large simulation on an MPP at NERSC, writing the results onto a data server at GA. He runs a synthetic diagnostic on the data on his local workstation that accesses the necessary portions of the simulation run over the net.

He then pulls up fluctuation data from the experiment at MIT that has been taken with a physical diagnostic with the same characteristics as the one run on the simulation. The two data sets along with the experimental uncertainties are compared and displayed.

Computer Science Research. To accomplish these goals, fusion scientists with expertise in large experiments and simulation code development will join computer scientists with expertise in security, distributed computing, and visualization to form a closely coordinated team. This team, leveraging existing computer science technology where possible, will deploy a collaboratory prototype. For requirements not met by current capabilities, new technologies will be developed. The variety of users, resources, applications, and policies encountered will serve as an excellent proving ground for new technologies that prepares the way for their use in other scientific disciplines. The computer science research necessary to create the Collaboratory is centered on three main activities: security, remote and distributed computing, and scientific visualization.

Security. The sharing of data, codes, and visualization tools as network services requires a system for protecting these valuable resources against unauthorized use. The Collaboratory will exploit state-of–the–art authentication, authorization, and encryption technologies provided by the Globus Security Infrastructure [1] and the Akenti authorization service [2]. Existing fusion community codes will be modified to use this infrastructure for remote execution and data access. To meet the needs of the Collaboratory, the current version of these middleware tools will be extended. In particular, it will be necessary to incorporate rules for fair use of shared resources into the policy enforced by the security model and to enhance tools that enable valid credentials to propagate automatically from resource to resource.

Distributed Computing. The remote and distributed computing requirements of the Collaboratory will utilize the Globus facilities including remote job scheduling, monitoring, exception handling, and accounting. This will enable researchers and their institutions to share the community's computational resources. The components of the Globus toolkit that can be immediately deployed to create the foundation of the Collaboratory are Grid Information Services, Grid Security Infrastructure, and Globus Resource Allocation Manager. Research components that are required to fully meet the needs of the Collaboratory and that will create new functionality for Globus include managing batch versus preemptive job priorities, providing status display and accountability to users, monitoring the adherence of resources to policies, and providing advance reservations. Fusion community codes to be adapted to the Collaboratory include serial and parallel MHD stability codes and serial and parallel transport codes.

Scientific Visualization. The demand placed on visualization tools by the Collaboratory is intense due to both the highly collaborative nature of fusion research and the dramatic increase in data resulting from the enhanced computing capabilities. The visualization component of the

Collaboratory will focus on the development of a collaborative control room, collaborative meeting room, and enhanced visualization tools. The collaborative visualization requirements will utilize the Access Grid [3] that enables distributed meetings and collaborative teamwork sessions. Extensions to the Access Grid software include a more closely integrated shared experience with the researcher's current work environment and the support of large tiled displays that will provide collaborative capabilities to large-format remote visualizations. New software will be tested on display walls [4] that already exist within the proposal team. Extensions to display wall software include the ability to have visualizations not tied to an individual projector allowing the size of the visualization to vary depending on the researcher's need. Extensions to the visualizaton toolkit will be the ability to quantitatively compare theory to experiment with uncertainties.

Benefits to Computer Science Toolkits. The Collaboratory project will enhance the existing computer science toolkits by (1) enabling automatic propagation of security credentials in multi–server contexts, (2) Allowing complex use policy implementation and remote monitoring of computational resources, (3) extending security architecture to encompass commercial databases, (4) creating pre–emptive scheduling capability and advance reservation of computational resources, (5) extending the Access Grid to large tiled display walls, and (6) developing a quantitative visualization capability allowing data comparison with uncertainties.

Summary. The team designing the Collaboratory is comprised of scientists and software engineers from the three large experimental magnetic fusion facilities and a group of computer scientists from universities and national laboratories with the expertise to achieve the research and development advances required to make the Collaboratory a success. The long–term vision of the Fusion Collaboratory described above is to transform fusion research and accelerate scientific understanding and innovation so as to revolutionize the design of a fusion energy source. The project will create and deploy collaborative software tools that will enable more efficient utilization of existing experimental facilities and more effective integration of experiment, theory, and modeling. These tools will initial link together the large magnetic fusion research community that is spread over more than 40 sites throughout the United States. However, the tools can be easily extended so as to include our international partners within the Collaboratory framework. The large scale and unique characteristics of the Fusion Collaboratory make it an important testbed for these toolkits. The proposed collaborative project would span three years and represents a dramatic shift for the fusion community where access to resources (data, codes, visualiation tools) would be separated from details of their implementation and eliminate barriers to their widespread use by the research community.

REFERENCES

[1] I. Foster, et al., “A Security Architecture for Computational Grids,” Proc. of the 5th ACM Conference on Computer and Communications Security (1998), http://www.globus .org.

[2] M. Thompson, et al., “Certificate-Based Access Control for Widely Distributed Resources," Proceedings of the Eight Usenix Security Symposium (1999).

[3] L. Childers, et al., “Access Grid: Immersive Group-to-Group Collaborative Visualization,” Proc. of the 4th International Immersive Projection Technology Workshop (2000), http://www.accessgrid.org.

[4] K. Li, et al., “Early Experiences and Challenges in Building and Using a Scalable Display Wall System,” IEEE Computer Graphics and Applications **20** (2000) 671-680.

ACKNOWLEDGMENT

Work supported by U.S. Department of Energy under Contracts DE-AC03-99ER54463, DE-AC02-76CH03073 and W-31-109-ENG-38. Proposal prepared under General Atomics internal funds.