ITR/AP+IM: High Performance with Petaobjects: The Hall D Virtual Experiment Environment

Project Summary

This project conducts the research necessary to integrate all of the information resources and software tools required by the planned search for exotic mesons in Hall D at Thomas Jefferson National Accelerator Facility. The resulting Hall D Virtual Experiment Environment (Hall D VEE) is an end-user problem solving environment that includes the high performance object system required to manage the needed collaboration, data management, and simulation tools used by geographically distributed experimental teams in conducting scientific research. To accomplish this, the project addresses many critical information technology research challenges to provide the foundation for VEE's. These research efforts establish the formal specifications and procedures necessary to integrate archived and derived information (distributed objects) into an efficient, common access framework.

This project builds upon existing ITR efforts and integrates object-oriented technology into scientific projects. The ITR program is already funding the GriPhyN Project to build and deploy the software middleware to manage virtual data grids for large scale physics projects similar to the one envisioned for Hall D. That effort can be leveraged to build the network and computing infrastructure within which the Hall D VEE will operate. However, extensive research remains in order to build a scalable, secure framework for managing the information, collaboration, process workflow, and scientific results. This research must meet the challenges inherent both in creating the collaborative, problem-solving environment and in managing the number, diversity, and complexity of distributed real and virtual objects.

This project explores a completely object-oriented approach for scientific data that differs significantly from those employed by commercial object-oriented databases. These differences are used to help enable efficient data mining and remote object access. The project investigates ways to merge existing repositories of materialized objects and virtual repositories of objects (objects that can be generated by object factories) into a single information resource and determines optimum strategies by which queries on this resource may be satisfied. The project also incorporates collaborative object technology into a Grid-based, problem-solving framework. This framework provides groups of experimenters with the ability to manage complex, distributed, large-scale simulation, and data analysis problems involving quadrillions of individual objects.

Significant quality, and efficiency gains are a direct result of the Hall D VEE. More scientists can be involved in the experiment because the learning curve for computing is reduced significantly. Implementation of the VEE also allows for better quality control over the applications and information associated with the experiment. Personnel and capital resources can be used more efficiently.

The research group assembled for this project consists of computer scientists and physicists from Florida State University and Indiana University who are primary participants in the Hall D project and have significant experience in distributed and collaborative computing and in large-scale physics experiments. This group is also in an excellent position to leverage other resources and to utilize complementary research efforts to help meet the goals of this project.

ITR:AP+IM: High Performance with Petaobjects: The Hall D Virtual Experiment Environment

Introduction

The search for exotic mesons being planned for Hall D[1,2] at the Thomas Jefferson National Accelerator Facility (JLab), like many modern scientific endeavors, will produce large volumes of complex, high-quality data from many sources. These data include experimental measurements, information on the state of the experimental conditions, information describing the status and results of data analysis, and simulations of the detector required to understand its response during the experiment. The exploration of physical phenomena with Hall D depends on our ability to efficiently extract information from the large volume of distributed data using a complex set of interrelated computing activities.

The goal of this proposal is to create an environment in which everything associated with the Hall D experiment—the entire complex, evolving network including the detector, experimental measurements, subsequent analyses, computer systems, technicians, and experimenters—will be integrated into a simple, collaborative fabric of information resources and tools. We plan to create a Virtual Experiment Environment (VEE), which will make it very easy for groups of distributed collaborators to conduct, analyze, and publish Hall D experiments based on the composition and analysis of these resource objects.

To facilitate this integration, the experiment has already committed to making all information within the Hall D project available as distributed objects with corresponding metadata descriptions. Within the VEE, every application will accept only properly described distributed objects as input and produce corresponding objects as output. We estimate that the Hall D activities will create about 10^{15} distinct objects for each year of its operation including event records, which are macro-objects containing micro-objects like detector signals and processed information like track segments and particle identifications. There will also be objects describing reports and presentations and the information needed to specify input and output of simulations initiated by any of the approximately 30 institutions involved in the experiment.

Our main research foci will be to establish the formal specifications and procedures necessary to integrate all archived and derived information (distributed objects) into a common access framework and to apply them to create the Hall D Virtual Experiment Environment (Hall D VEE). We believe that significant quality and efficiency gains will result from the creation of the Hall D VEE and that those gains will translate directly into higher-quality physics results.

The ITR program is already funding the GriPhyN Project[3] to build and deploy the software middleware to manage virtual data grids for large scale physics projects similar to the one envisioned for Hall D. Two members of our team are unfunded senior investigators on the GryPhyN project. We will leverage the GriPhyN virtual data grid middleware in the construction of the network and computing infrastructure within which the Hall D VEE will operate (see below). The VEE will focus on the end-user problem solving environment, with its high performance object system, that is required to manage the needed collaboration, data access, and simulation tools that experimental teams will use to conduct their scientific research.

Extensive research is required to create the Hall D VEE. Critical research challenges that must be met include:

- Developing a test-bed to support the research activities of this project.
- Researching strategies for specifying all information within a VEE in a way that facilitates searching, assembling, creating, and organizing it. We must ensure that these specifications allow high performance implementations.
- Finding ways to describe the organization of information within a VEE. This must include structured information associated with the raw and processed events as well as the more scattered and heterogeneous objects that describe the scientific analysis and collaborative discussions associated with planning and the certification of results.
- Adapting procedures for high performance storage and access from the GriPhyN Project for the Hall D objects. A primary research challenge is in achieving levels of performance for storing and accessing objects that are competitive with traditional approaches.

- Researching techniques for supporting distributed queries on the objects of the VEE that provide the enhanced search capabilities inherent in the object-oriented approach.
- Discovering procedures for including real-time measurements, derived information, and archived data into a VEE. This includes incorporating the necessary security and access restrictions for real-time measurements and distributed computing activities.
- Investigating specifications and procedures for defining and managing versioning of virtual information in the VEE.
- Finding a way to integrate Hall D objects into collaborative frameworks that can share the metadata stored in the Hall D object properties. We must ensure that these properties are designed to support collaborative functions needed for the VEE.
- Developing tools for applying these specifications and procedures to real-world problems. The tools must be simple enough so that people who are Hall D experts are able to use them reliably.
- Ensuring that the VEE architecture created for Hall D will scale to other experiments that could involve larger number of more heterogeneous objects managed in the context of much bigger collaborations.

Experimental Context of Hall D

The approximately 30 geographically distributed institutions involved in the Hall D experimental program will collaborate to build the detector and develop data acquisition and analysis software, run the experiments, analyze the measurements and publish the results. When the detector is completed, physicists from these remote institutions will travel to JLab (Newport News, VA) to conduct experiments year-round (whenever the accelerator can deliver beam). It is reasonable to assume that these experiments will continue for approximately 10 years.

Funding for these activities comes primarily from the National Science Foundation, the Department of Energy, and various universities. We will be able to leverage such support to provide the computer infrastructure required by Hall D, to develop the required application software, and to support this effort. Some Hall D collaborators are from outside the United States. Normally their respective governments and universities provide support for their involvement.

As participants in the Hall D experiment, scientists are required to jointly develop plans for detector construction, software, and experiments. Analysis of the experimental data and the production and publication of final results require participation of the entire collaboration. Such activities place a premium on close collaboration and as a result have historically been conducted primarily by research groups that reside near the experiment or that are able to place scientists in residence at the experiment.

We will integrate this research project into the Hall D management plan to help ensure the successful completion of both this and the Hall D project. The integration helps us leverage the resources necessary to define and create the required Hall D objects and to insure that the real-world needs of the Hall D collaboration are met.

The primary Hall D computing activities are quite diverse and must be supported by a variety of high-performance computer systems. Experimental data acquisition collects measurements from the detector components, eliminates unnecessary data items, organizes the resulting information as objects, and stores the results. First-pass offline analysis converts the experimental information (each event is about 1000 smaller raw-data micro-objects) into more meaningful physical information (processed objects), such as particles, momenta, and energies Detector calibrations provide information for converting measured electrical pulses to times, energies, and positions Event simulations model the detector's response to particles of known masses, energies, and momenta Physics analyses use the measured particles, momenta, and energies together with simulated detection efficiencies to reconstruct the reactions taking place.

The current plan for the operation of Hall D indicates that it should generate approximately three Pbytes/year from three main sources: experimental data (~0.75 Pbytes/year), offline analysis (~1.5 Pbytes/year) and event simulations (~0.75 Pbytes/year). This implies that each year the experiment will generate some 10^{15} basic objects (nuggets or micro-objects of information) organized into 10^{11} macro-objects such as events or analysis instances. Even though the Hall D detector will generate experimental information at a very high rate (up to 100 Mbytes/s), computer programs will generate the majority of the information. In particular, the offline analysis and event simulations require significant computing resources.

The distributed computing resources required for the Hall D experiment will be created as a set of computer systems each designed to address a specific computing task. The computing activities are described here to illustrate the scope and diversity of the information involved in the Hall D project and to emphasize the natural organization of Hall D information as distributed objects. This proposal does not include any funds for creating these computing resources or for developing the specific computer software needed to carry out these steps.

Experimental data acquisition takes places on computers dedicated to the task. These computers acquire the event information from the detector and distribute it to a computer cluster for filtering in real time. The filtered data are stored on a tape archive at Jefferson Lab. During data acquisition, one of the critical real-time activities is the automated collection and recording of information about the state of the 50,000 active components that comprise the Hall D detector. Data calibration requires repeated analysis of small subsets of the experimental data. The acquisition system will archive approximately 1% of the Hall D data for calibrations and carry out the detector calibrations.

For the experiments currently running at Jefferson Lab, data location, efficiency, and quality control issues have forced most first-pass analysis activities to be performed at Jefferson Lab. Jefferson Lab computing facilities will be expanded to meet the archival and computing needs for the first-pass analysis of the Hall D data.

Event simulations are required to quantify the systematic errors in the Hall D experiment. We anticipate that the quality of our results will be limited by systematic errors and therefore highly dependent on the quality and quantity of simulations. Our plan is to distribute the simulation effort among several of the institutions within the Hall D collaboration.

Traditionally, large distributed teams of graduate students and researchers develop specialized computer programs and software systems to conduct an experiment's computing activities. Such systems generally have an unacceptably high dependence on the original authors, require large ramp-up times to use, and are very labor intensive to produce and maintain. Furthermore, the entire process does not scale well with the data size and complexity because the programs and computing environments are frequently lax about the way they keep track of what has been done. As a result a significantly large fraction of each scientist's time in conducting such research must be devoted to learning how to use the programs and to routine computing activities such as code management, installation, and data management.

Technology Approach

To explain our proposed approach, we first summarize the nature of the experimental data and its associated metadata. At one level are the events. These come in highly structured streams of simple objects that represent the response of detector components to physical events. At another level are the metadata that describe each class of event object, the types of processing functions that are applied to them, the descriptions of the instruments that generated them, and the scientific documentation and annotations related to each observation derived from the analysis. The metadata form a deeply nested hierarchy of class descriptions that changes from one experiment to the next. The event objects form highly structured collections that are archived in fast mass storage systems.

While it has been well understood for some time that object-oriented technology was an important organizing conceptual framework for managing and interacting with the data[4], commercial object database systems do not provide the right framework for writing efficient analysis and mining tools. The problem is one of separating the metadata from the data. A typical object model, such as that used in Java, is very inefficient for storing scientific data[5,6]. This is because each Java object is a complex beast that is capable of answering questions about itself (reflection) and is very costly to serialize. This is because the metadata about the class of the object is closely bound to the object itself. Scientific experiment data is not stored that way and the application filters that process this data must exploit this fact or the processing will be too slow to be useful.

In this proposal we will explore an alternative approach that draws upon our experience with compilers for object-oriented systems[7,8] for parallel and distributed computing. Our plan is to allow the user to describe data mining operations that search the event streams as transformations on the metadata associated with the objects. A compiler can generate analysis codes that would operate directly on the densely packed event streams. This same technology would allow any object to have a virtual object interface that would allow it to be accessed over the network as if it were a true remote object (in the Java or CORBA sense.)

Thus, our approach to events and the objects contained therein is to treat them as hierarchically labeled macro-objects. Similarly, physics analyses produce large objects that contain a multitude of smaller ones organized into particular histogram bins or partial wave components.

There will also be a multitude of less structured information. At the higher levels, the macro objects themselves will not naturally be organized in a structured fashion; they will be labeled by the independent ideas of the distributed science team members who conduct the analysis. Here we must conduct the research necessary to join our careful compiler-based methodology to traditional distributed database technology.

We will build on the virtual data ideas being developed in GriPhyN to tackle the distributed data produced by simulation and analysis farms around the world. We expect this to enhance the efficiency of our system as the inner loops and small objects will all be tackled by the compiler-based approach. However, an important research area will be to find the best distributed organization that will optimize data-locality and hence efficient macro-object use.

An experimental environment makes extensive use of object factories—programs that can be used to generate objects of a specific type and with specific initial conditions. The Hall D event simulators are object factories that are used to create massive numbers of immutable objects. These objects, once created, are added to the Hall D object repositories and used by scientists is various analysis activities. Scientists may issue queries that are satisfied by all of the objects in a single simulation (one macro object) or by a filtered subset of the objects in one or more macro objects.

We have an opportunity in this project to investigate ways to merge existing repositories of materialized objects and virtual repositories of objects that can be generated by object factories into a single information resource. A query on this resource may be satisfied by accessing existing objects or by using an object factory to create new ones.

Obvious trade-offs exist between the cost of accessing and filtering objects in repositories and the cost of generating a new set of objects that exactly match the query. A merger of repositories and factories will allow users to create queries without having to consider whether sufficient objects have been materialized to satisfy the query. The compiler-based optimization research mentioned above will require additional effort to find ways to automatically and efficiently process queries on these merged information resources.

Implementation

We need to build a scalable, secure framework for managing the information, collaboration, process workflow, and scientific results that make up a HALL–D experiments. This framework must interoperate with and leverage existing technology as well as results from other Grid related ITR research efforts.

We will build our system as a set of uniform objects so that a single system can manage them. The Hall D objects have well-defined properties and methods defined in XML. We will allow implementation in any object model and will provide wrappers for data that comes from non-object-oriented sources. We plan to use custom Java, CORBA, or XML middleware, and not an object database, so we can build a flexible research system to test essential ideas such as scalable performance. We will of course build on standard component technology like Enterprise Javabeans[9]. We have experience at both Indiana University[10,11,12,13] and Florida State University[14,15,16] with all these approaches and their inter-operation. We will achieve uniformity for legacy objects by wrapping them.

We need to support both synchronous and asynchronous collaboration. We will first install Access Grid[17] technology at every major institution to provide high-end audio-video conferencing. This will be supplemented to the desktop by using the rapidly evolving voice-over-IP commercial offerings. HearMe, Lipstream, and Net2Phone are the leaders[18,19]. These multi-point systems can support any mix of Internet and traditional telephone traffic. Florida State University is installing a HearMe server and working with this company to support full recording of all sessions. This will produce SMIL[20] (the appropriate W3C XML standard) sessions with digital multi-media interspersed with changes in shared objects being viewed in collaborative sessions. The technology is quite well understood from early work such as TangoInteractive[21,22,23,24] from Fox's group (at Syracuse) and now the excellent commercial support from Centra, WebeX, Placeware, and Microsoft NetMeeting.

We will support the full event sharing of all properties exposed in the Hall D object API. For example, suppose Hall D documents include properties (metadata) for file name, author, and so on. By wrapping a Microsoft word document and defining the Hall D properties, we can immediately share it at the level of opening it on collaborating clients. However, the sharing of non-Hall D properties (such as clicking buttons on the Word taskbar) will only be supported in shared event mode if the particular collaboration system has support for this application. As in the commercial systems mentioned above, we will offer the limited shared-display collaborative mode for all applications. VNC[25] is a well-known open source shared-display tool. We expect developments such as the Gecko[26] layout engine from mozilla.org to be very important for building custom collaborative portals which will use the W3C DOM standard for which we can build universal shared event support much more easily than for proprietary interfaces such as Word and PowerPoint. Our long range goal is to incorporate this collaborative object technology into a Grid-based, problem solving framework that gives groups of experimenters the ability to manage complex, distributed, large-scale simulation and data analysis problems that involve quadrillions of individual objects.

Education and Outreach

We will build on the membership of both NCSA and Indiana University in the NCSA Alliance where we work on both the EOT (Education, Outreach and Training) and Enabling Technologies thrusts. This will provide us an appropriate audience, in association with leading nationwide organizations, and appropriate methodology for our work. The PACI EOT also gives us contacts with the Minority Serving Institutions. The collaborative framework sketched in this proposal will build on ongoing work with EOT to provide education and community VEE. Further our proposed object API will be a superset of the proposed standards from IEEE[27], IMS, [28] and DoD ADL[29] for learning objects. Thus we anticipate that we will be able to broadly disseminate both the CS and physics results of our proposal through any standards compliant web-based tools.

Benefits of the Hall D Virtual Experiment Environment

Creating an environment that enables scientists to efficiently access all the computing and information resources for an experiment will produce a significant reduction in the time they must devote to computing activities. We will enable more scientists to participate in science rather than computing by concentrating on improving the efficiency of the VEE and will include more people in the process by reducing the learning curve for computing.

In the process we will develop general tools to facilitate data description and interpretation and develop a customizable view of the integrated computational and data management services. The result will be similar to a personalized electronic logbook that organizes existing information and the resources needed to create additional information.

Adopting this approach has several distinct advantages for Hall D and similar data-intensive scientific computing research projects:

- It produces a single point of contact and a common interface for all computing activities. This is expected to dramatically reduce the learning curve for those joining the effort.
- It greatly simplifies the software that a user must install to fully participate in computing activities—only a Web browser is required. This should be contrasted with some of the existing systems that require significant, repeated software installations and extensive coordination efforts.
- It allows us to design efficient computing systems to deal with specific computing problems and to greatly reduce scheduling problems found in general purpose systems.
- It helps ensure that the services underlying the VEE are of high quality. Everyone uses the same version in the same environment so problems can be found quickly and resources directed to their solution.
- It forces all information to be presented in a standardized self-describing format. This allows broader access to information and the development of standard tools for accessing and organizing the information and services.
- It provides universal access to the data and services, subject to reasonable security considerations. This allows more efficient remote participation in data taking, data analysis, and problem solving.
- The systematic use of objects allows use of modern electronic collaboration systems based on sharing objects. This will cut the need for travel to a central location and help the distributed collaboration.

The net results will be a substantial reduction in the overall cost of computing and an improvement in the quality and timeliness of scientific enterprises.

Notes and References

- Alex Dzierba and Curtis Meyer, "The Search for QCD Exotics," *American Scientist Magazine*, September-October 2000, Volume 88, No. 5, and http://www.sigmaxi.org/amsci/articles/00articles/Dzierba.html.
- Hall D Collaboration, "The Hall D Conceptual Design Report," http://dustbunny.physics.indiana.edu/HallD/.
- Paul Avery and Ian Foster, "The GriPhyN Project," http://www.griphyn.org/.
- See, for example, systems like ROOT developed at CERN in 1995 (http://root.cern.ch/root/Mission.html) or the AIPS++ tools used for radio astronomy data (http://aips2.nrao.edu/docs/aips++.html).
- F. Breg, D. Gannon, A Customizable Implementation of RMI for High Performance Computing", *Proceedings of the International Workshop on Java for Parallel and Distributed Computing*. IPPS/SPDP April, 1999.
- Madhusudhan Govindaraju, Aleksander Slominski,Venkatesh Choppella, Randall Bramley, Dennis Gannon, Requirements for and Evaluation of RMI Protocols for Scientific Computing,' *Proceedings SC2000.*
- Elizabeth Johnson and D. Gannon. "Programming with the HPC++ Parallel Standard Template Library," *Proceedings of ACM ICS*, July 1997, Vienna.
- A. J. C. Bik and D. Gannon, Automatically Exploiting Implicit Parallelism in Java, *Concurrency: Practice and Experience*, Vo. 9. no. 6, June 1997, pp. 579-621.
- *Enterprise Java Beans Technology*, http://java.sun.com/products/ejb/.
- R. Bramley, K. Chiu, S. Diwan, D. Gannon, M. Govindaraju, N. Mukhi, B. Temko, M. Yechuri, A Component Based Services Architecture for Building Distributed Applications, *Proceedings HPDC 2000*.
- Rob Armstrong, Dennis Gannon, Al Geist, Katarzyna Keahey, Scott Kohn, Lois McInnes, Steve Parker, Brent Smolinski, "Toward a Common Component Architecture for High-Performance Scientific Computing," *Proceedings, High Performance Distributed Computing Conference 1999.*
- William Johnston, Dennis Gannon, Bill Nitzberg, Alex Woo, Bill Thigpen, Leigh Ann Tanner, Computing and Data Grids for Science and Engineering, *Proceedings SC2000.*
- Juan Villacis, Madhusudhan Govindaraju, David Stern, Andrew Whitaker, Fabian Breg, Prafulla Deuskar, Benjamin Temko, Dennis Gannon, Randall Bramley, "CAT: A High Performance, Distributed Component Architecture Toolkit for the Grid," *High Performance Distributed Computing Conference 1999*.
- Erol Akarsu, Geoffrey Fox, Tomasz Haupt, Alexey, Kalinichenko, Kang-Seok Kim, PraveenSheethaalnath, and Choon-Han Youn, "Using Gateway System to Provide a Desktop Access to High Performance Computational Resources," *Proceedings of HPDC-8 Conference*, Redondo Beach Ca., Aug 3-6, 1999, IEEE Press.
- Fox, G.C., K. Hurst, A. Donnellan, J. Parker , "Introducing a New Paradigm for Computational Earth Science: A web-object-based approach to Earthquake Simulations", in *Geocomplexity and the Physics*

of Earthquakes, edited by J.B. Rundle, D.L. Turcotte, and W. Klein, *Geophysical Monograph 120*, Am. Geophys. Un., Wash. D.C., and published by AGU, 219–244, 2000.

- Gateway website http://www.gatewayportal.org/.
- Argonne National Laboratory NCSA Alliance, *Access Grid Digital Conferencing System*, http://www-fp.mcs.anl.gov/fl/accessgrid/default.htm.
- http://www.webex.com and http://www.centra.com are commercial collaboration systems to be used in VEE.
- http://www.hearme.com is a commercial collaboration system under evaluation.
- *Synchronized Multimedia Integration Language (SMIL) 1.0 Specification*. http://www.w3.org/TR/REC-smil.
- Fox, G., and Podgorny, M., “Real Time Training and Integration of Simulation and Planning using the TANGO Interactive Collaborative System,” in *Proceedings of the International Test and Evaluation Workshop on High Performance Computing*, June 10, 1998.
- Fox, G., Scavo, T., Bernholdt, D., Markowski, R., McCracken, N., Podgorny, M., Mitra, D., and Malluhi, Q., “Synchronous Learning at a Distance: Experiences with TANGO Interative.” *Supercomputing 98 Conference*, November 1998.
- Geoffrey C. Fox and Jianxiang Jin, *Overview of Collaborative Computing and Some NPAC Experience,* ERDC Technical Report May 2000, http://www.new-npac.org/users/fox/documents/collabcompmay00/ collabviz.html, unpublished.
- Fox, G.C. “Portals and Frameworks for Web Based Education and Computational Science”, *Proceedings of the Second International Conference on the Practical Application of Java*, Editor Omer Rana, Manchester England April 12-14 2000, http://www.practical-applications.co.uk/PAJAVA2000/index.html and http://www.new-npac.org/users/fox/documents/pajavaapril00/
- *VNC* or *Virtual Network Computing* at http://www.uk.research.att.com/vnc/. *Virtual Network Computing*, Tristan Richardson, Quentin Stafford-Fraser, Kenneth R. Wood & Andy Hopper, IEEE Internet Computing, Vol.2 No.1, Jan/Feb 1998 pp33-38.
- Gecko, also known as NGLayout is found at http://www.mozilla.org/newlayout
- *IEEE Learning Technology Standards Committee* (LTSC), http://ltsc.ieee.org/
- *IMS (Instructional Management System) Project,* established by Educause, http://www.imsproject.org/
- *ADL SCORM: Sharable Courseware Object Reference Model*, http://www.adlnet.org/ADL-TWG/documents.htm