Incorporating High Energy Physics Data Capabilities into Large-Scale DoD Simulations

Julian J. Bunn and Thomas D. Gottschalk
Center for Advanced Computing Research, Caltech
Pasadena, California
julian@cacr.caltech.edu, tdg@cacr.caltech.edu

ABSTRACT

The data management and data exploitation issues for large-scale, distributed DoD simulations have striking parallels within a number of existing large-scale High Energy Physics (HEP) projects, in particular, the experiments associated with the Large Hadron Collider (LHC) in Geneva, Switzerland. The significant commonalities include: data rates of 10-100 GBytes/day, data distribution and database operations over very large scale, high-speed networks, and sophisticated data exploitation objectives. In this regard, the lessons learned over the past decade of preparations for LHC operations have obvious significance and relevance for operational (fielded) DoD information exploitation systems. The requirements for persistent, scheduled, secure data access and data mining within the HEP environment are similar to many aspects of future large-scale DoD simulation environments, such as the Sentient World Simulation (SWS).

This paper explores three particular areas of DoD data exploitation needs with significant parallels within existing HEP/LHC work. The first involves robust, scalable database design and management, such as the distributed simulation and data system within the Joint SemiAutomated Forces project now under development within the US Joint Forces Command. Important aspects here include operational transparency and efficiency from the perspective of a single user/analyst at a workstation. The second general area involves support for "user toolkits" - significant additional computational subsystems such as data-mining/knowledge-discovery procedures and "what if" Monte Carlo excursions that go well beyond straightforward queries of a distributed database. The final area has to do with "real-time" considerations, where this term is to be understood in the more general sense of legitimate, possibly urgent user needs that exceed available computational resources. Strategies are discussed for leveraging the demonstrated HEP expertise toward DoD data management and exploitation problems, using SWS, on occasion, as a template for some specific DoD requirements.

ABOUT THE AUTHORS

Julian Bunn is a Member of the Professional Staff at Caltech, where he works in the Center for Advanced Computing Research. He held a Physicist position at CERN, the European Centre for Particle Physics for 15 years before joining Caltech. His research interests include high performance network and computing systems and Grid architectures for solving the data distribution and analysis challenges posed by CERN's Large Hadron Collider experiments. He gained his B.Sc (Hons) in Physics from the University of Manchester in 1980, and his Ph.D. in Experimental Particle Physics from the University of Sheffield in 1983.

Thomas D. Gottschalk is a Member of the Professional Staff, a Senior Research Scientist at the Center for Advanced Computing Research (CACR) and Lecturer in Physics at the California Institute of Technology. He has been with CACR for nearly a decade. Much of his research has been on the use of parallel computers to simulate various physical phenomena. His instructional duties include his upper division course on Statistics for Physics Graduate students. He also consults for the AeroSpace Corporation and other scientific, financial and technical organizations. He received a B.S. in Physics from Michigan State University and a Ph.D. in Theoretical Physics from the University of Wisconsin.