Grand Challenge Application on HENP Data *

D. Olson, G. Odyniec, R. Porter, F. Wang, and HENP GC Collabration

Research Objective

The purpose of this project is to develop techniques and tools that will enable efficient access to the massive datasets of modern High-Energy and Nuclear Physics (HENP) experiments, particularly for experiments at the Relativistic Heavy ion Collider (RHIC) beginning in late 1999, in their search for the Quark-Gluon-Plasma (QGP). Access to these 100TB+ datasets by hundreds of scientists, in conjunction with carrying out the large-scale computations necessary to refine and reduce them to the essential physical properties hidden within, is one of the forefront problems of high-performance computing today. By capitalizing on recent advances in storage systems (HPSS), object database technology (Objectivity Inc.), high-performance scientific computing (NERSC) and expertise from several institutions across the U.S. we are able to address this important problem affecting fundamental science.

Computational Approach

The principle approach to this problem is cast it in the form of an extremely large hierarchical collection of objects (some persistent and some transient) that experience has shown is very well matched to this event-based experimental data. Modern object-database technology allows us to address the issue of physical storage layout separately from the logical relationships between the objects. The primary issue for physical storage is to organize the data so that it is stored according to how it is accessed rather than by how it is generated. The NERSC T3E is used to simulate the relativistic heavy ion collisions that will occur at RHIC and produce a dataset with these characteristics for which expected access patterns can be studied.

Coupling large robotic tape storage systems to object databases is a very recent development and there are a number of issues that need to be addressed to enable applying the required amount of high-performance computing cycles to these massive object databases. Addressing these issues is a principle area of R&D in this project that is enabled by use of the newly installed HPSS system at NERSC.

Accomplishments

The NERSC T3E is used to generate simulated data according to theoretical predictions of what may occur in the relativistic nuclear matter collisions at RHIC. Since these collisions will produce matter under conditions that have never before been measured in the laboratory, there is considerable uncertainty about what exactly will occur, and just what signatures will show up in the data measured by the detectors. One method being employed to address this uncertainty is to calculate this matter under various initial conditions and then propagate it in time (and space) in order to see the effects in the detector data. See the contribution by F. Wang for a more detailed description of this work.

Significance

The direct significance of this work is to enhance the capabilities of large-scale experimental high-energy and nuclear physics to achieve their full scientific potential in exploring new regions of the physical world.

Related benefits are the advancement of computational and data management technology in high-performance access to very large databases. As the digital world becomes more connected and pervasive the issues of dealing with massive quantities of data have a direct impact not only on science but on the commercial and financial worlds as well.

Footnotes and References

*http://www-rnc.lbl.gov/GC/
http://www.nersc.gov/
http://www.rhic.bnl.gov/