

Data Management in High Energy and Nuclear Physics

God, it appears, does play dice: fundamental physical truth cannot be determined by single measurements. To probe deeper into fundamental physics we must record and analyze ever greater numbers of collisions between particles. Like the science of particle accelerators, the science of data analysis is integral to progress in experimental fundamental physics.

Data management in high-energy and nuclear physics (HENP) is unique in its combination of scale and complexity. Each HENP experiment requires the intellect and labor of hundreds or thousands of physicists at universities and laboratories all over the nation or the world. Our data analysis capability limits the number of collisions that can be studied. Today's limit for the complex and granular HENP data is around one petabyte. At this limit, it can take many months for a student to try a simple new analysis idea.

We propose a revolutionary advance in the science of distributed data management and analysis in high-energy and nuclear physics. The goal is to give data-management intelligence to networked computers and storage such that queries

that might have taken months or years could be completed in minutes or hours. We also propose an increasingly rigorous approach to the design and exploitation of HENP data-management systems. Systems are built from mass storage (tape), caches (disk), wide-area networks, computers and HENP-specific or commercial software components. The rigorous approach requires that existing systems be instrumented and that models be developed that can predict the behavior of new or improved systems. The models will

facilitate rapid evaluation and improvement of new approaches in the science of HENP data management. Dedicated testbeds will be created so that the more adventurous new ideas can be explored. The potential benefits of this HENP-motivated data-management science for other data-intensive sciences and for industry will be vigorously pursued.

We will be able to deal with the number of col-

lisions in HENP experiments by factors of tens or hundreds. The need to make dangerous guesses about which collisions are worth recording will be reduced. The precision of physics measurements will increase and the sensitivity to the unexpected will be vastly improved.

