

Researchers tackle problem of data storage for next-generation supercomputers

September 7 2006

The U.S. Department of Energy has awarded a five-year, \$11 million grant to researchers at three universities and five national laboratories to find new ways of managing the torrent of data that will be produced by the coming generation of supercomputers.

The innovations developed by the new Petascale Data Storage Institute will enable U.S. scientists to fully exploit the power of these new computing systems, which will be capable of performing millions of billions of calculations each second.

The institute combines the talents of computer scientists at Carnegie Mellon University, the University of California at Santa Cruz and the University of Michigan with those of researchers at the DOE's Los Alamos, Sandia, Oak Ridge, Lawrence Berkeley and Pacific Northwest national laboratories.

Increased computational power is necessary because scientists depend on computer modeling to simulate extremely complicated phenomena, such as global warming, earthquake motions, the design of fuel-efficient engines, nuclear fusion and the global spread of disease. Computer simulations provide scientific insights into these processes that are often impossible through conventional observation or experimentation. This capability is critical to U.S. economic competitiveness, scientific leadership and national security, the President's Information Technology Advisory Committee concluded last year.

But simply building computers with faster processing speeds -- the new target threshold is a quadrillion (a million billion) calculations per second, or a "petaflop" -- will not be sufficient to achieve those goals. Garth Gibson, a Carnegie Mellon computer scientist who will lead the data storage institute, said new methods will be needed to handle the huge amounts of data that computer simulations both use and produce.

"Petaflop computers will achieve their high speeds by adding processors -- hundreds of thousands to millions of processors," said Gibson, an associate professor of computer science. "And they likely will require up to hundreds of thousands of magnetic hard disks to handle the data required to run simulations, provide checkpoint/restart fault tolerance and store the output of these modeling experiments.

"With such a large number of components, it is a given that some component will be failing at all times," he said.

Today's supercomputers, which perform trillions of calculations each second, suffer failures once or twice a day, said Gary Grider, a co-principal investigator at the Los Alamos National Laboratory. Once supercomputers are built out to the scale of multiple petaflops, he said, the failure rate could jump to once every few minutes. Petascale data storage systems will thus require robust designs that can tolerate many failures, mask the effects of those failures and continue to operate reliably.

"It's beyond daunting," Grider said of the challenge facing the new institute. "Imagine failures every minute or two in your PC and you'll have an idea of how a high-performance computer might be crippled. For simulations of phenomena such as global weather or nuclear stockpile safety, we're talking about running for months and months and months to get meaningful results," he explained.

Collaborating members in the Petascale Data Storage Institute represent a breadth of experience and expertise in data storage. "We felt we needed to bring the best and brightest together to address these problems that we don't yet know how to solve," said Grider, leader of Los Alamos' High Performance Computing Systems Integration Group.

Source: Carnegie Mellon University

Citation: Researchers tackle problem of data storage for next-generation supercomputers (2006, September 7) retrieved 14 July 2024 from <https://phys.org/news/2006-09-tackle-problem-storage-next-generation-supercomputers.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.