SDSC, NCAR, LLNL, IBM Team Sets U.S. Records in Weather Simulation

Collaborative Effort Results in New U.S. Records for Speed, Scale, Detail and Parallelism

October 10, 2007

Paul Tooby and Jan Zverina

A team of researchers from the National Center for Atmospheric Research (NCAR), the San Diego Supercomputer Center (SDSC) at UC San Diego, Lawrence Livermore National Laboratory (LLNL), and IBM Watson Research Center has set U.S. records for size, performance, and fidelity of computer weather simulations, modeling the kind of "virtual weather" that society depends on for accurate weather forecasts.

For the highly detailed weather simulations, the researchers used the sophisticated Weather Research and Forecast (WRF) model, widely used for continuous weather forecasting by government, military, and commercial forecasters as well as for weather and climate research in hundreds of universities and institutions worldwide.

The team's efforts open the way for simulations of greatly enhanced resolution and size, which will serve as a key benchmark for improving both operational forecasts and basic understanding of weather and climate prediction.

The scientific value of the research goes hand-in-hand with the computational achievements. The "non-hydrostatic" WRF weather code is designed for greater realism by including more of the physics of weather and capturing much finer detail than simpler models traditionally used for global scale weather prediction. Running this realistic model using an unprecedented number of computer processors and simulation size enabled researchers to capture key features of the atmosphere never before represented in simulations covering such a large part of the Earth's atmosphere. This is an important step towards understanding weather predictability at high resolution.

"The scientific challenge we're addressing is the question in numerical weather prediction of how to take advantage of coming petascale computing power," said weather scientist Josh Hacker of NCAR. "There are surprisingly complex questions about how to harness the higher resolution offered by petascale systems to best improve the final quality of weather predictions." Petascale computing refers to next generation supercomputers able to compute at a petaflop (10^15 calculations per second), equivalent to around 200,000 typical laptops.

The researchers set a speed performance record for a U.S. weather model running on the Cray XT4 "Franklin" supercomputer at the Department of Energy's National Energy Research Computing Center (NERSC) at Lawrence Berkeley National Laboratory. Running on 12,090 processors of this 100 peak teraflops system, they achieved the important milestone of 8.8 teraflops - the fastest performance of a weather or climate-related application on a U.S. supercomputer. One teraflops is one trillion, or a thousand billion, calculations per second. It would take a person operating a hand-held calculator more than 30,000 years to complete one trillion calculations.

The team also set a record for "parallelism," or harnessing many computer processors to work together to solve a large scientific problem, running on 15,360 processors of the 103 peak teraflops IBM Blue Gene/L supercomputer at Brookhaven National Laboratory, jointly operated by Brookhaven and Stony Brook University.

"We ran this important weather model at unprecedented computational scale," added NCAR's Hacker. "By collaborating with SDSC computer scientists to introduce efficiencies into the code, we were able to scale the model to run in parallel on more than 15,000 processors, which hasn't been done with this size problem before, achieving a sustained 3.4 teraflops."

Added John Michalakes, lead architect of the WRF code, "To solve a problem of this size, we also had to work through issues of parallel input and output of the enormous amount of data required to produce a scientifically meaningful result. The input data to initialize the run was more than 200 gigabytes, and the code generates 40 gigabytes each time it writes output data."

With this power the researchers were able to create "virtual weather" on a detailed 5 kilometer horizontal grid covering one hemisphere of the globe, with 100 vertical levels, for a total of some two billion cells - 32 times larger and requiring 80 times more computational power than previous simulation models using the WRF code.

"The calculation, which is limited by memory bandwidth and interprocessor communication, is representative of many other scientific computations," said Allan Snavely, director of the Performance Modeling and Characterization (PMaC) lab at SDSC, whose group helped tune the model to run at these unprecedented scales. "This means that what we learn in these large simulations will not only improve weather forecasts, but help a number of other applications as they enter the petascale realm."

The work was presented in November at SC07, the international conference for high performance computing, networking, storage, and analysis, where it was a finalist in the prestigious Gordon Bell Prize competition in high performance computing.

"Modeling weather systems is an enormously challenging endeavor, and forecast accuracy depends on the ability to represent many components of the environment and their complex interactions," said Fran Berman, director of SDSC. "The WRF team used sophisticated optimizations to create a new breakthrough in resolution which will lead the way to better predictions, and lays the groundwork for runs on next generation 'petascale' supercomputers. We congratulate them on these exciting results."

In preparing for the groundbreaking runs on the Stony Brook-Brookhaven and NERSC systems, the extensive problem-solving required to achieve these results was made possible by running the WRF code on the Blue Gene system at the Department of Energy's Livermore lab, the fastest supercomputer on the Top500 list, and the large Blue Gene system at the IBM Watson Research Center.

Tuning and testing were also carried out at the National Center for Computational Sciences at Oak Ridge National laboratory and on SDSC's Blue Gene system, a resource in the National Science Foundation-supported TeraGrid, an open scientific discovery infrastructure combining leadership class resources at nine partner sites. In these ongoing collaborations the team anticipates further record-setting results.

Team members include John Michalakes, Josh Hacker, and Rich Loft of NCAR; Michael McCracken, Allan Snavely, and Nick Wright of SDSC; Tom Spelce and Brent Gorda of Lawrence Livermore; and Robert Walkup of IBM.

This research has been supported by the National Science Foundation and the Department of Energy.

Media Contacts:

National Science Foundation: Lisa-Joy Zgorski, 703-292-8311
San Diego Supercomputer Center: Warren Froelich, 858-822-3622 or Jan Zverina, Zverina 858 534-5111
National Center for Atmospheric Research: David Hosansky, 303-497-8611

Comment:

Allan Snavely, 858-534-5158
Josh Hacker, 303-497-8188
John Michalakes, 303-497-8199