COMPUTERWORLD
Storage

IT to Help Avoid Astronomical Armageddon

Predicting asteroid impacts will require supercomputers and petabytes of data

Patrick Thibodeau

September 6, 2004 (Computerworld) The computers at the Minor Planet Center recently sorted through some 10,000 observations of astronomic phenomena and deemed a handful as deserving follow-up because they were either newly discovered or on a path in Earth's general vicinity.

The preliminary trajectories plotted by the MPC were based on too few observations to be accurate, but unbeknownst to center officials, the computer had found one object on a collision path with Earth. The discovery was posted on the Web late in the day on Jan. 13, 2004.

Alert amateur astronomers in Europe first noticed the problematic path of the near-Earth object (NEO), and a Yahoo message board used by astronomers began to fill with activity about the bogy.

Mapping the Cosmos

In some ways, scientists were lucky to have found this potential threat, because finding NEOs today is literally a hit-or-miss activity.

The first warning of an asteroid or comet impact may come from calculations performed by some old workstations clustered together at the MPC at the Smithsonian Astrophysical Observatory at Harvard University.

These systems aren't large enough to map the sky and provide scientists with a comprehensive view of everything that could do serious damage. But that's changing. Some big projects are under way to attempt to find as many threatening objects as possible. These sky-mapping efforts will create extraordinarily large data sets and require new searching and processing approaches.

An asteroid hitting Earth, as depicted by an artist, is the scenario data-intensive programs like LSST and Pan-STARRS are aimed at preventing.

One ambitious project is the Large Synoptic Survey Telescope (LSST), a public/private partnership based in Tucson, Ariz., intended to provide 3-D maps of the universe, covering a very wide area.

When it's in operation in 2011 at a site still to be determined, the telescope being built for the LSST project will collect data at a rate of about 6GB (equivalent to one DVD) per 10 seconds, generating many petabytes of data over time. One petabyte equals roughly 100 times the printed contents of the Library of Congress.

The LSST project "pushes forward database technology dramatically," says Philip Pinto, a physics professor at Steward Observatory at the University of Arizona and a member of the LSST project's board of directors. "The LSST database will probably be the largest known nonproprietary database in the world."

Because operation of the telescope lies some seven to eight years out, the scientists working on it are hopeful that processing capabilities and storage densities will increase enough to handle this data, but they can still imagine a supercomputer system of 1,000 or so systems networked together.

Work on the software has already started, however. "It's not entirely clear how you do it," says Jeff Kantor, a computer scientist in charge of data management on the project. He's developing algorithms for handling the data, such as determining from multiple observations whether a particular dot in the sky is the same object moving from frame to frame.

A second effort, called Pan-STARRS (Panoramic Survey Telescope and Rapid Response System), is being developed by the University of Hawaii's Institute for Astronomy.

Instead of a large telescope, the Pan-STARRS project will use four smaller ones that combine image data and generate data on a scale similar to that of the LSST. These multiple systems, further delineated by 64 detectors in each of the four cameras, will be well suited to parallel processing.

But instead of the CPU parallelization common in high-performance technical computing, researchers in Hawaii are working on data parallelization in which every processor works on different data but executes the same instruction at the same time. "A lot of techniques have been explored in that particular realm," says Eugene Magnier, image-processing pipeline technical lead. Pan-STARRS is expected to be operational by 2008.

Scientists hope that LSST and Pan-STARRS will help them identify most of the NEOs that may threaten Earth and locate those on a dangerous path long before they strike. They hope that information will give them time to develop ways to deflect the NEO.

These system developments will happen, of course, only if the Earth isn't first destroyed by a comet or asteroid.

Predictive Processes

Among those following the activity on the Yahoo message board back in January was Alan Harris, a senior research scientist at the Space Science Institute in Boulder, Colo. Harris ran some calculations and found that the object was "heading straight for us at around 18km/sec." and would hit in 26 hours.

The NEO was estimated to be about 30 meters in diameter. Depending on its composition, it could have disintegrated in the atmosphere or hit the Earth. (The mile-wide, 570-foot Barringer Meteorite Crater in Arizona was created by a 45-meter iron object.)

Harris was nonetheless skeptical about the object's preliminary path. The MPC trajectory was based on only a few observations—enough to help astronomers find it, but not accurate enough to determine its actual course.

At 8:30 p.m., Brian Marsden, director of the MPC, got a call from NASA's Jet Propulsion Laboratory. The NASA official "sort of was wondering" about the path of the object, Marsden recalls.

Marsden ordered further observations, and the bogy was determined not to be on a threatening path after all.

In the aftermath of this incident, the MPC changed the computer program that provides the preliminary path of an NEO. Next time the computer forecasts an "impacting solution," as astronomers call it, the program will alert staffers before the data is posted on the Web.

The Large Synoptic Survey Telescope Project

Objective: Map the entire sky to track near-Earth objects using time-lapse digital imaging

Camera size: 3 billion pixels

Data acquisition rate: 6GB every 10 seconds, or 3 petabytes per year

Total database size over project life: 30 petabytes

Source: LSST Corp., Tucson, Ariz.