Sandia Labs Details Update to Graph500 Benchmark

June 26, 2012

ALBUQUERQUE, N.M., June 25 -- Supercomputing performance is getting a new measurement with the Graph500 executive committee's announcement of specifications for a more representative way to rate the large-scale data analytics at the heart of high-performance computing.

An international team that includes Sandia National Laboratories announced the single-source shortest-path specification to assess computing performance on Tuesday at the International Supercomputing Conference in Hamburg, Germany.

The latest benchmark "highlights the importance of new systems that can find the proverbial needle in the haystack of data," said Graph500 executive committee member David A. Bader, a professor in the School of Computational Science and Engineering and executive director of High-Performance Computing at the Georgia Institute of Technology.

The new specification will measure the closest distance between two things, said Sandia National Laboratories researcher Richard Murphy, who heads the executive committee. For example, it would seek the smallest number of people between two people chosen randomly in the professional network LinkedIn, finding the fewest friend of a friend of a friend links between them, he said.

Graph500 already gauges two computational techniques, called kernels: a large graph that links huge numbers of participants and a parallel search of that graph. The first two kernels were relatively easy problems; this third one is harder, Murphy said. Once it's been tested, the next kernel will be harder still, he said.

The rankings are oriented toward enormous graph-based data problems, a core part of most analytics workloads. Graph500 rates machines on their ability to solve complex problems that have seemingly infinite numbers of components, rather than ranking machines on how fast they solve those problems.

Big data problems represent a $270 billion market and are increasingly important for businesses such as Google, Facebook and LexisNexis, Murphy said.

Large data problems are especially important in cybersecurity, medical informatics, data enrichment, social networks and symbolic networks. Last year, the Obama administration announced a push to develop better big data systems.

Problems that require enormously complex graphs include correlating medical records of millions of patients, analyzing ever-growing numbers of electronically related participants in social media and dealing with symbolic networks, such as tracking tens of thousands of shipping containers of goods roaming the world's oceans.

Medical-related data alone could potentially overwhelm all of today's high-performance computing, Murphy said.

Graph500's steering committee is made up of more than 30 international experts in high-performance computing who work on what benchmarks supercomputers should meet in the future. The executive committee, which implements changes in the benchmark, includes Sandia, Argonne National Laboratory, Georgia Institute of Technology and Indiana University.

Bader said emerging applications in healthcare informatics, social network analysis, web science and detecting anomalies in financial transactions "require a new breed of data-intensive supercomputers that can make sense of massive amounts of information."

But performance can't be improved without a meaningful benchmark, Murphy said.

"The whole goal is to spur industry to do something harder" as they jockey for top rankings, he said.

"If there's a change in the list over time — and there should be — it's a big deal," he added.

Murphy sees Graph500 as a complementary performance yardstick to the well-known Top 500 rankings of supercomputer performance, based on speed processing the Linpack code. Nine computers made the first Graph500 list in November 2010; by last November, the number had grown to 50. Its fourth list, released at the conference in Germany, ranked 88. Rankings are released twice a year at the Supercomputing Conference in November and the International Supercomputing Conference in June.

"A machine on the top of this list may analyze huge quantities of data to provide better and more personalized health care decisions, improve weather and climate prediction, improve our cybersecurity and better integrate our online social networks with our personal lives," Bader said.

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration. With main facilities in Albuquerque, N.M., and Livermore, Calif., Sandia has major R&D responsibilities in national security, energy and environmental technologies and economic competitiveness.

Source: Sandia Labs
