

username

LOGIN

Register

Forgot Password

English | en Español | auf Deutsch | in Italiano | 中文 | in India



Published by MIT

Advertisement



For a life go

BLOGS

- HOME
- COMPUTING
- WEB
- COMMUNICATIONS
- ENERGY
- MATERIALS
- BIOMEDICINE
- BUSINESS
- BLO



Mims's Bits

B I C

Christo
and sc

New Benchmark Ranks Supercomputers More Thoughtfully

The Graph500 test measures "deep analysis" rather than ability to perform basic arithmetic.

CHRISTOPHER MIMS 12/14/2010

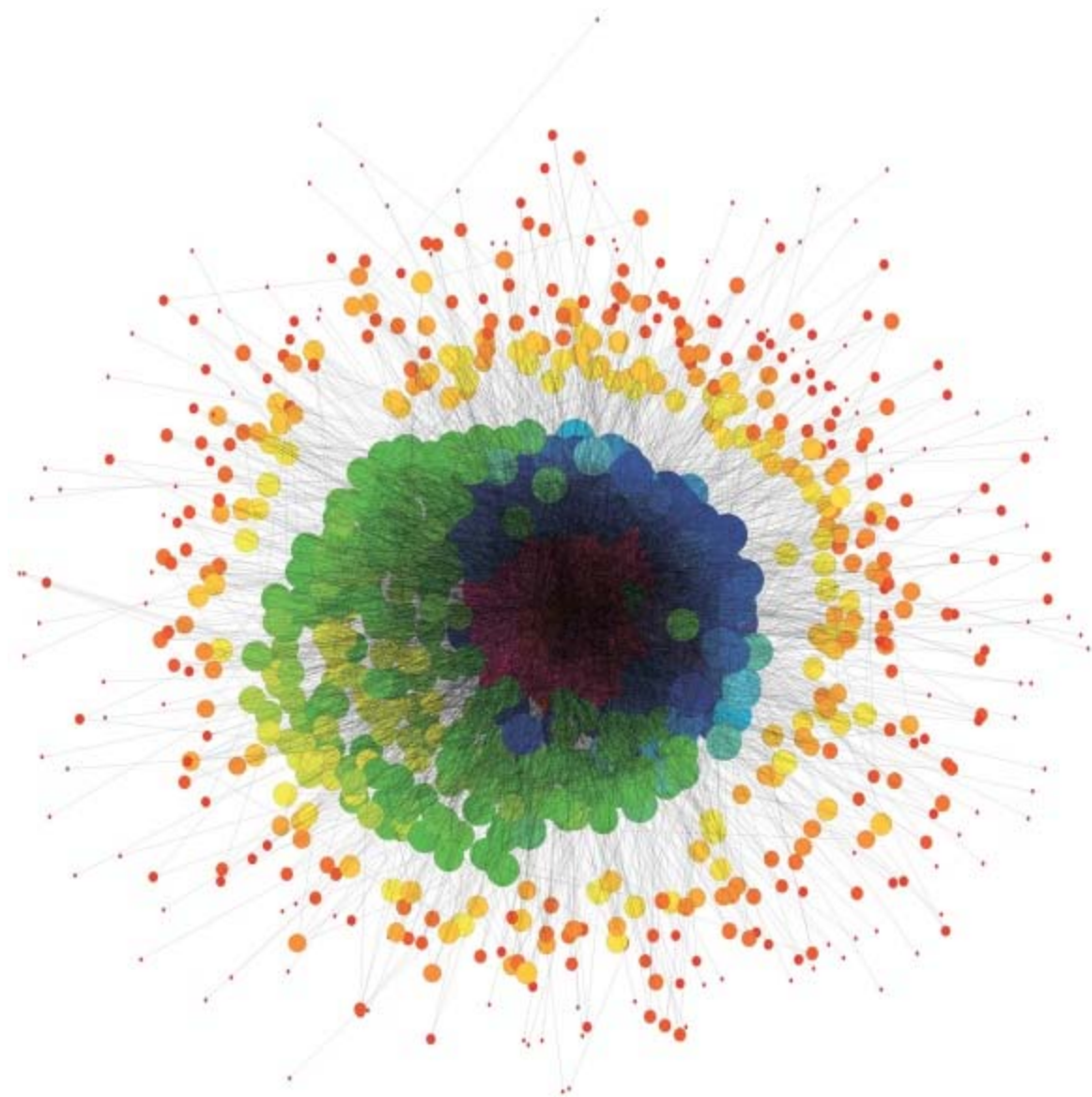


New

How

A Le

Cars



blekl
pass

In the Graph500 benchmark, supercomputers race each other to process every "edge" (connection) in an artificially-generated synthetic graph such as this one.

Credit: [Jeremiah Willcock](#), Indiana University

There are dozens of ways to usefully measure the performance of the world's fastest supercomputers -- everything from the speed of a system's internal communications network and its ability to randomly access memory, to the traditional measure of a high performance system's speed: its ranking on the [Linpack benchmark](#), which is used to "officially" rate systems on the international Top500 list.

Now a new ranking, proposed by a steering committee including 30 of the world's top high performance computing (HPC) experts, has proposed: the Graph500 list, which measures a supercomputer's ability to chew through "graph" data.

Graph data turns out to be the lingua franca of countless areas, from biomedicine (drug discovery, protein interactions, gene networks, public health) to homeland security, fraud

detection, and the social graphs that represent users on networks like Facebook. All these fields, and countless others, could benefit from faster parsing of graphs.

The first-ever Graph500 list was unveiled in November, at [Supercomputing 2010](#) in New Orleans. The list is so new that results from only 9 systems were submitted, with the Department of Energy's IBM BlueGene/P powered Intrepid system, using 8192 of its 40960 nodes, coming out on top.

The Intrepid managed an impressive 6.6 Giga-edges per second (GE/s), which means it managed to leap from one node (or vertex) on an artificially generated graph to another (along the connections between them, known as edges) 6.6 billion times per second.

This level of speed is required for "traversing", or stepping through, very large graphs, such as all the members (or vertices) of Facebook and all the connections (or edges) between them. Imagine, for example, that you want to know what is the shortest number of connections between any two members of Facebook -- depending on the sophistication of the algorithm doing the searching, a computer might have to search many potential paths through the social graph to determine which members comprise the chain of degrees of separation between two nodes.

Interestingly, the third and fourth fastest systems on the Graph500 list -- two Cray XMT supercomputers housed at Pacific Northwest National Laboratory and Sandia National Laboratories -- managed their high ranking despite having relatively few processors: 128 apiece. At an average performance for each machine of 1.2 Giga edges per second, these Cray machine were more than ten times faster, as measured per node, than the IBM Blue Gene/P-powered Intrepid system.

That should be no surprise, when you consider that the [Cray XMT is explicitly designed to tackle graph data](#). This specialization extends to the machine's development environment.

"John Thale, who has the number three system [on the Graph500 list,] said it took him an hour or two to write the code [required to run the Graph500 benchmark,] whereas on a conventional system it takes days," says Shoaib Mufti, director of knowledge management at Cray.

It's this kind of specialization that will be required for tackling the "data-intensive computing" that will characterize the uses to which supercomputers will be put in the future. As Richard Murphy of Sandia National Laboratories, who led the founding of the Graph500 list, argued in a talk delivered in September, no amount of acceleration of conventional systems, including the increasingly popular GPGPU-supercomputers, can conquer the applications to which high performance systems will be put in the coming years: "deep analysis" of data designed to extract complicated relationships between data points.

This kind of computing, and not the kind measured by the traditional Linpack benchmark, will determine the true power of systems applied to graph data, Murphy argued in his talk, the slides of which included this interpretation of the bottom line for high performance systems:

This is a survival of the fittest question... the person with more complete data or better algorithms will win by:

- Being more profitable
- Beating you to the nobel prize
- Etc.

How to Prevent a Gawker-Style Hack From Endangering You

To comment, please sign in or [register](#)

username

LOGIN

[Forgot my password](#)

Featuring a chat with speaker **Jose Bravo**,
Chief Scientist, Physical Separations, Shell



© 2010 MIT
Technology Review

ABOUT US
ADVERTISE
PRESS
PERMISSIONS
REPRINTS
STAFF

HELP
CUSTOMER SERVICE
FEEDBACK
CONTACT US
PRIVACY
TERMS OF USE
SITMAP

SUBSCRIPTIONS
SUBSCRIBE
RENEW
GIFT SUBSCRIPTIONS
BACK ISSUES
CUSTOMER SERVICE
NEWSLETTERS
IPAD APP
IPHONE APP