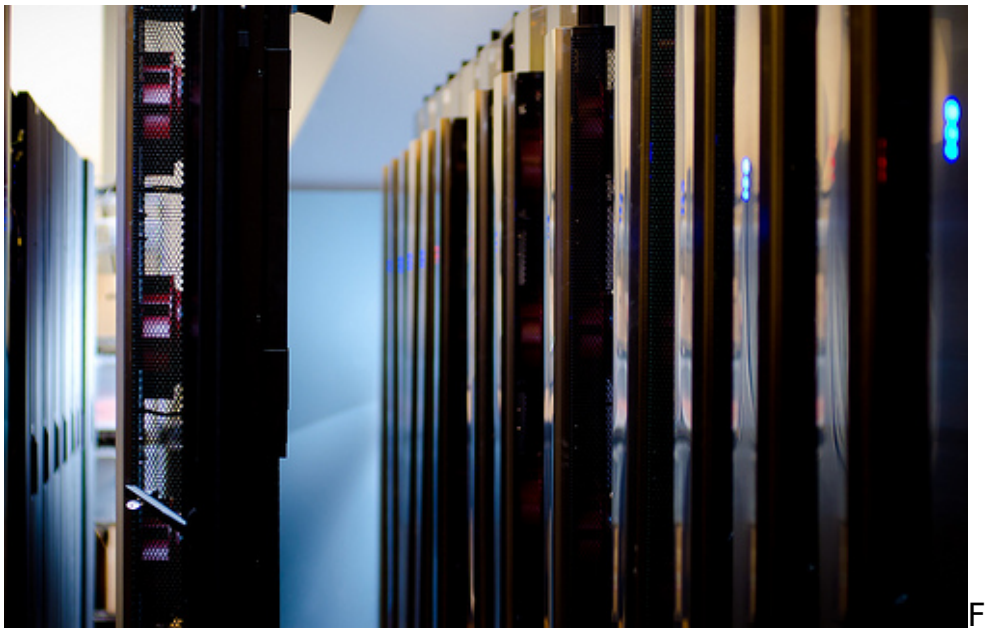
PNNL expertise highlighted at Supercomputing

Pacific Northwest National Laboratory

Algorithm to identify matching patterns, MPI code translator to be discussed Wednesday



From identifying common patterns in data to speeding up computers, researchers from the Department of Energy's Pacific Northwest National Laboratory will share their computational expertise at this year's Supercomputing conference.

Also referred to as SC12, the annual gathering is the international conference for high-performance computing, networking, storage and analysis. It runs Nov. 10-16 at the Salt Palace Convention Center in Salt Lake City. Two noteworthy talks featuring PNNL research are described below.

New algorithm pin-points similar data in seconds
Data is everywhere these days. Biologists sift through vast amounts of error-prone data to understand how our cells work. Even librarians slog through mountains of information to better understand the materials they catalog. The key to comprehending today's information explosion is finding meaningful patterns buried in the data — and then finding comparable data patterns in other, related sources. This technique is called network alignment. Computational scientists at PNNL and Purdue University have developed new methods to identify similar patterns in any type of data. Their procedures help find proteins that act the same in humans and mice, and help find ideas that act the same for librarians and Wikipedia editors.

The existing methods used to solve these kinds of problems have been too slow to cope with the growing amount of data, prompting the PNNL and Purdue team to make them faster. To do this, they developed a new algorithm that uses an approach called approximate matching, which saves time by matching nearly

identical patterns instead of exactly identical ones. They also developed new computer implementations that enabled the algorithm to use all a computer's processors in parallel to quickly identify relationships between two different networks. Tests using both of these improvements showed that the algorithm found similar interactions between thousands of proteins in two species in just seconds and found comparable ideas between hundreds of thousands of topics in library systems and Wikipedia entries in less than a minute.

PNNL's Mahantesh Halappanavar led the research on how to quickly find approximate matchings with help from Purdue's Arif Khan and Alex Pothen. And, Purdue's David Gleich led the work on how to use approximate matchings to align networks. Gleich will present a paper describing this research Wednesday.

4:30-5 p.m., Wed., Nov. 14: A multithreaded algorithm for network alignment via approximate matching, Arif Khan, David Gleich, Mahantesh Halappanavar & Alex Pothen, Room 355-EF.

Software translates code, speeds up data-crunching

Large and complex networks in parallel computers can lead to inefficient communications between processors that also slows down computation. This makes it difficult to achieve exascale computing, which is one thousand times faster than today's fastest petascale supercomputers. Scientists are developing strategies to reduce the time it takes to compute data and communicate those results between parallel processors. A team of researchers from PNNL, University of California, San Diego, and Lawrence Livermore National Laboratory have developed new software called Bamboo to help do just that.

Traditionally, scientists have broken up a complex algorithm to speed things up. Different processors calculate bits of the algorithm and then each processor communicates its results to the others. Such division of labor is quicker than one processor doing all the work by itself. But communicating bunches of data between multiple processors can cause information bottlenecks that slow down the whole process. One solution is to initially calculate a portion of a processor's data and communicate those results while the other portion is still being calculated. Called overlapping communications and calculations, this approach can reduce the overall time it takes to complete a job, but it requires extremely complex codes. That's where Bamboo comes in. Bamboo automatically translates standard MPI parallel codes into a format that can easily overlap communication with available computation. Without Bamboo, scientists have the onerous task of manually developing overlapping MPI code. Tests showed Bamboo-generated code was as good as or better than human-developed codes.

PNNL's Eric Bylaska drew on his experience developing complex code for NWChem, DOE's premier molecular modeling software package, to help develop realistic test programs for the Bamboo framework. The University of California, San Diego's Scott Baden, who led the project, will present a paper describing the team's results Wednesday.

Source: http://www.pnnl.gov/news/release.aspx?id=953 [1]

Source URL (retrieved on *08/31/2014 - 5:27am*):
http://www.ecnmag.com/news/2012/11/pnnl-expertise-highlighted-supercomputing

Links:
[1] http://www.pnnl.gov/news/release.aspx?id=953