SDSC to venture capitalists: Data-intensive supercomputing is here

April 22 2011

The exponentially increasing amount of digital information, along with new challenges in storing valuable data and massive datasets, are changing the architecture of today's newest supercomputers as well as how researchers will use them to accelerate scientific discovery, said Michael Norman, director of the San Diego Supercomputer Center (SDSC) at the University of California, San Diego (UCSD).

In a presentation during the 3rd Annual La Jolla Research & Innovation Summit this week, Norman said that the amount of digital data generated just by instruments such as DNA sequencers, cameras, telescopes, and MRIs is now doubling every 18 months.

"Digital data is advancing at least as fast, and probably faster, than Moore's Law," said Norman, referring to the computing hardware belief that the number of transistors which can be placed inexpensively on an integrated circuit doubles approximately every 18 months. "But I/O (input/output) transfer rates are not keeping pace – that is what SDSC's supercomputers are designed to solve."

SDSC, a key resource for UCSD researchers as well as the UC system and nationally, will later this year deploy a new data-intensive supercomputer system named Gordon, which will be the first high-performance supercomputer to use large amounts of flash-based SSD (solid state drive) memory. Flash memory is more common in smaller devices such as mobile phones and laptop computers, but unique for supercomputers, which generally use slower spinning disk technology.

1/3

The result of a five-year, $20 million grant from the National Science Foundation, Gordon will have 250 trillion bytes of flash memory and 64 I/O nodes, and be capable of handling massive data bases while providing up to 100 times faster speeds when compared to hard drive disk systems for some queries.

"We are re-engineering the entire data infrastructure in SDSC to support the capabilities offered by Gordon," Norman said.

This makes Gordon ideal for data mining and data exploration, where researchers have to churn through tremendous amounts of data just to find a small amount of valuable information, not unlike a web search.

"Gordon is a supercomputer that will do for scientific data analysis what Google does for web search," Norman told the summit, adding that SDSC likes to call the new system "the largest thumbdrive in the world."

SDSC researchers are already doing preliminary tests on several potential applications using 16 I/O nodes of the Gordon system now in operation. Such data mining applications include 'de novo,' or 'from the beginning' genome assembly from sequencer reads, or classification of objects found in massive astronomical surveys.

"The future of personalized genomic medicine will require technologies like those prototyped in Gordon," Norman said.

The new supercomputer also is expected to aid researchers in conducting interaction network analysis for new drug discovery. Other data-intensive computational science that will benefit from Gordon's unique configuration include the solution of inverse problems – or converting observed measurements into information about a physical object or system – in oceanography, atmospheric science, and oil exploration, as well as using the system's large shared memory system to research

modestly scalable codes in quantum chemistry, structural engineering, and computer-aided design/computer-aided manufacturing (CAD/CAM) applications.

Earlier this year, SDSC deployed a new high-performance computer called Trestles, the result of a $2.8 million award from the NSF. Trestles is appropriately named because it will serve as a bridge between SDSC's unique, data-intensive resources available to a wide community of users both now and into the future.

"These new systems were designed with one goal in mind, and that is to enable as much productive science as possible as we enter a data-intensive era of computing," said Norman.

The annual La Jolla Research & Innovation Summit is organized by CONNECT, a San Diego-based group that links inventors and entrepreneurs with resources needed for commercialization by promoting collaborations between industry, venture capital sources, and research organizations including academic institutions such as UCSD.

Provided by University of California - San Diego