MIT, Intel unveil new initiatives addressing ‘big data’

Massachusetts Institute of Technology
MIT has been selected from among 55 institutions that submitted 157 proposals to host a new Intel research center that will concentrate on what’s come to be called “big data” — new techniques for organizing and making sense of the huge amounts of information generated by Web users and new networked sensors.

That was just one of three related announcements made Wednesday [1] at an event at MIT’s Stata Center — home of the Computer Science and Artificial Intelligence Laboratory (CSAIL) — that featured the president of MIT, the governor of Massachusetts, and the chief technology officer of Intel.

The Intel research center will be the cornerstone of a new CSAIL initiative known as “bigdata@CSAIL [2],” led by Associate Professor Sam Madden and Adjunct Professor Michael Stonebraker, both of MIT’s Department of Electrical Engineering and Computer Science (EECS). In addition to Intel, the initiative’s sponsors include AIG, EMC, SAP and Thomson Reuters.

Mass. Gov. Deval Patrick was also on hand to announce the Massachusetts Big Data Initiative, which will sponsor several programs, including a grant-matching program, an internship program and a project to investigate how big-data technologies can improve government.

As part of its new Intel Science and Technology Center at CSAIL, Intel will hire a handful of new researchers who will be based in Cambridge and will work closely with MIT faculty on technologies related to big data. Researchers at the University of California at Santa Barbara, Portland State University, Brown University, the University of Washington and Stanford University will also be affiliated with the center. Intel has committed $2.5 million a year to the center for at least the next three years, with an additional two-year commitment possible if the center passes a three-year review.

Data detonation

“We are witnessing an unprecedented period of growth in unstructured digital data on the Web and in the cloud. And this will only further accelerate through the rapid growth of mobile devices such as smartphones and connected cars,” Intel’s CTO, Justin Rattner, said at the event. “While this amount of data is already staggering — with numbers so big they defy our imagination — they will really pale in comparison to the amount of data that will be generated in real time from the ‘Internet of things,’” an envisioned network that connects computing devices embedded in ordinary household items. “So if you think this is a lot of data, you ain’t seen nothing yet.”

“‘Big data’ has become one of the hottest new phrases,” added MIT President Susan Hockfield. “You know that a phrase is really hot when you get on the [subway] and the ads include that phrase. We’re there.”

After the event, Rattner pointed out that while existing data-retrieval methods may have to contend with huge amounts of data, the operations they perform are not very computationally intensive. The techniques that the Intel and CSAIL centers will develop, he said, will be much more complex.

Madden agreed. “I think a part of big data is about doing more sophisticated statistical analysis that is a lot more computationally intensive than conventional database workloads,” he said.

Research agenda

At the event, Madden described four main research topics that bigdata@CSAIL will address: first, scalable platforms — computing infrastructures that can expand indefinitely to accommodate increasing data loads; second, the data-analysis tools that will run on those platforms; third, scalable techniques for handling security and privacy; and fourth, new techniques for visualizing data.

Madden offered a few examples of the types of problems that big-data research will address. One is the analysis of biological data — ferreting out the connections between genetic variations or concentrations of biological molecules in the cell and the processes underlying disease or physiological development. He also pointed to recent work [3] by Andrew Lo — a professor in MIT’s Sloan School of Management and a principal investigator at CSAIL — who has used techniques borrowed from computer science to mine credit-bureau data and data about the transactions conducted by customers of financial institutions to more accurately predict the risk of default or delinquency.

Madden himself is involved in a host of big-data projects, including the CarTel initiative [4], which he co-leads with Hari Balakrishnan, the Fujitsu Professor in EECS, and Daniela Rus, the newly appointed director [5] of CSAIL, who also spoke at the event. Madden explained that the CarTel project has equipped a fleet of Boston-area cabs with sensors and developed innovative algorithms to process the data they collect. One application he described was an algorithm for computing the most reliable route between any two points; another was a map of potholes.

“I just want to point out the examples you were using were city roads and not state roads,” Patrick joked.

Source URL (retrieved on *07/12/2014 - 10:41am*):
http://www.ecnmag.com/news/2012/05/mit-intel-unveil-new-initiatives-addressing-%E2%80%98big-data%E2%80%99

Links:

[1] http://www.csail.mit.edu/node/1750
[2] http://bigdata.csail.mit.edu/
[3] http://www.argentumlux.org/documents/CRisk_final.pdf
[4] http://cartel.csail.mit.edu/doku.php
[5] http://web.mit.edu/newsoffice/2012/rus-named-csail-director-0522.html