

# IBM Researchers Develop Energy Efficient Method to Analyze the Quality of Data at Record Speeds

February 25 2010

---



IBM scientists in Zurich, Dr. Costas Bekas (left) and Dr. Alessandro Curioni, stand next to a Blue Gene/P system similar to the one where they conducted a record-breaking experiment to validate nine terabytes of data (nine million million or a number with 12 zeros) in less than 20 minutes, without compromising accuracy. Photo credit: Michael Lowry

(PhysOrg.com) -- IBM Research today unveiled a breakthrough method based on a mathematical algorithm that reduces the computational complexity, costs, and energy usage for analyzing the quality of massive amounts of data by two orders of magnitude. This new method will greatly help enterprises extract and use the data more quickly and

efficiently to develop more accurate and predictive models.

In a record-breaking experiment, IBM researchers used the fourth most powerful supercomputer in the world -- a Blue Gene/P system at the Forschungszentrum Julich in Germany -- to validate nine terabytes of data (nine million million or a number with 12 zeros) in less than 20 minutes, without compromising accuracy. Ordinarily, using the same system, this would take more than a day. Additionally, the process used just one percent of the energy that would typically be required.

The breakthrough will be presented today at the Society for Industrial and [Applied Mathematics](#) conference in Seattle.

"In a world with already one billion transistors per human and growing daily, data is exploding at an unprecedented pace," said Dr. Alessandro Curioni, manager of the [Computational Sciences](#) team at IBM Research - Zurich. "Analyzing these vast volumes of continuously accumulating data is a huge computational challenge in numerous applications of science, engineering and business. This breakthrough greatly extends the ability to analyze the quality of large volumes of data at rapid speeds."

One of the most computation-intensive, yet critical factors in analytics is the measurement of the quality of the data, which shows how reliable the data is that is being used and also generated by the model. In areas ranging from traffic management, financial management and water management this method could pave the way to create more powerful, complex and accurate models with greater predictability.

For example:

- A water authority could analyze real time, map-based information and geo-analytics to develop predictive models showing problems before they occur across the sprawling infrastructure of pipes, valves, public

fire hydrants, collection pipes, man holes and water meters. This can be done by analyzing an enormous amount of data and uncovering patterns related to weather conditions, water use, and hundreds of other variables

- Supply chains face many challenges when it comes to logistics, such as road construction, traffic or poor weather that may get in the way of delivering the final product on time. With multiple suppliers to source parts from, along with a variety of transportation modes and tight deadlines the variables and challenges are endless. Using GPS-data, traffic sensors, a database of suppliers and demand forecasting, analytics can aid in making realtime decisions when these types of unforeseen obstacles arise

The amount of digital data is increasing at enormous rates - due also to the ever more ubiquitous presence of sensors, actuators, RFID-tags or GPS-tracking-devices. These miniature computers measure everything from the degree of pollution of ocean water to traffic patterns to food supply chains.

With all of this data come new challenges as organizations are now struggling to not only extract the relevant information out of it, but to also make sure it's accurate. IBM researchers are pursuing leading edge research and actively engaging in client projects to extend the ability for analytics to predict outcomes and improve the speed and quality of business decisions.

"Determining how typical or how statistically relevant the data is, helps us to measure the quality of the overall analysis and reveals flaws in the model or hidden relations in the data," explains Dr. Costas Bekas of IBM Research - Zurich. "Efficient analysis of huge data sets requires the development of a new generation of mathematical techniques that target at both reducing computational complexity and at the same time allow for their efficient deployment on modern massively parallel resources."

The new method demonstrated by the IBM scientists brings down computational complexity and has very good scaling characteristics that reach to the full scale of the JuGene Supercomputer at the Forschungszentrum Julich with its 72 racks of IBM's Blue Gene/P system, 294,912 processors and a peak performance of one petaflop.

"In the next years supercomputing will provide us with unique insights and will help to create added value with new technologies," says Prof. Dr. Thomas Lippert, Director of the Julich Supercomputing Centre. "A cornerstone for the future will be innovative tools and algorithms helping us to analyze the huge amount of data provided by simulations on the most powerful computers."

Source: IBM

Citation: IBM Researchers Develop Energy Efficient Method to Analyze the Quality of Data at Record Speeds (2010, February 25) retrieved 23 April 2024 from <https://phys.org/news/2010-02-ibm-energy-efficient-method-quality.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.