En direct des Labs : la Recherche d'IBM développe une méthode économe en énergie pour analyser un nombre de données sans précédent à une vitesse record.

Cette méthode permet de valider 9 térabits de données en moins de 20 minutes ainsi que d'en améliorer l'analyse

****

Made in IBM Labs: IBM Researchers Develop Energy Efficient Method to Analyze Unprecedented Amounts of Data at Records Speeds

Nine Terabytes of Data Validated in Less than 20 Minutes; Will Further Analysis of Information Produced by Sensors

Seattle, USA, and Zurich, Switzerland, February 25, 2010 – At the Society for Industrial and Applied Mathematics conference, IBM (NYSE: IBM) Research today will unveil a novel method-based on a mathematical algorithm that reduces the computational complexity, costs, and energy usage for analyzing massive amounts of data by two orders of magnitude. The new method ensures that the data produces more accurate and predictable models.

In a record-breaking experiment, the IBM researchers used the fourth most powerful supercomputer in the world - a Blue Gene/P system at the Forschungszentrum Jülich, in Julich, Germany -- to validate nine terabytes of data (nine million million or a number with 12 zeros) in less than 20 minutes without compromising accuracy. Extrapolated, this would be like analyzing the entire online catalogue of the US Library of Congress in less than four hours**. Ordinarily, using the same system, this would have taken more than a day. Additionally, the process used just one percent of the energy than would typically be required**.

One of the most computation-intense, 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 economics, finance and portfolio management, climate modeling, geology to astrophysics, the new method by IBM could pave the way to create more powerful, complex and accurate models with greater predictability.

"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."*

"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 Jülich Supercomputing Center**.*" 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."*

Through the emergence of new smarter systems, such as smart grids and traffic monitoring, which connect to sensors, actuators, RFID-tags, GPS-tracking-devices, the amount of digital data is increasing at enormous rates. These miniature computers measure everything from the degree of pollution of ocean water to traffic patterns to food supply chains. The information these devices produce need to be analyzed rapidly to help people make decisions.

Further complicating matters, the rate at which data is stored greatly exceeds the computing capacity of standard computational   analytics   techniques.

"*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 amount of computations needed for measuring the quality of massive data sets with current techniques requires exaflops of computations - far beyond the current computing capabilities.

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 Jülich. With its 72 racks of IBM's Blue Gene/P system, 294,912 processors and a peak performance of one petaflop, JuGene is the fourth most powerful supercomputer in the world.

Data Analytics For Better Decision Making
This year the amount of digital information has reached 988 exabytes -- equivalent to a stack of books from the sun to Pluto and back. Despite the massive amounts of accumulated and available data, organizations are struggling to extract the relevant information out of it. In the 2009 IBM Global CIO Study, 83 percent of respondents identified business intelligence and analytics -- the ability to see patterns in vast amounts of data and extract actionable insights in short time -- as the way they will enhance their organizations' competitiveness.

At IBM Research labs around the world, scientists are pursuing leading edge research to extend current analytics capabilities and are working with clients to help them both improve the speed and quality of business decisions while better understanding the consequences and business outcomes of those decisions.

*based on 100 TB May 2009, 3.7 h, the estimation is based on using the same hardware
**JuGene requites about 52800 kWh for one day of operation on the full machine, the IBM demonstration required an estimated 700 kWh

About Forschungszentrum Jülich
Forschungszentrum Jülich pursues cutting-edge interdisciplinary research on solving the grand challenges facing society in the fields of health, energy and the environment, and also information technologies. In combination with its two key competencies – physics and supercomputing – work at Jülich focuses on both long-term, fundamental and multidisciplinary contributions to science and technology as well as on specific technological applications. With a staff of about 4,400, Jülich – a member of the Helmholtz Association – is one of the largest research centres in Europe. The Jülich Supercomputing Centre hosts regularly world leading Supercomputers and supports a user community of over 200 science and research groups by developing algorithms, models, tools, and methods in a variety of fields of computational science and engineering.
http://www.fz-juelich.de/jsc/

About IBM and Analytics
For more information, visit
http://www.ibm.com/press/us/en/presskit/27163.wss