CWI

Centrum Wiskunde & Informatica

Invitation

CWI Lectures in Mathematics and Computer Science 2010

### Data Intensive Research

Friday 25 June 2010

A Fourth Paradigm (J. Gray) is emerging. The data-centric setting requires a new look at computing architectures, algorithms and strategies to exploit the accumulation of data in social networks, Internet archives, physical sciences and genomics. We are soon dealing with Petabytes if not Exabytes of data to harvest information from.

The opportunities for large scale text analysis is discussed by Hector Garcia-Molina, who brings experience in the analysis of large web archives and the challenges scalability poses for domain specific information extraction. He will survey the research issues on entity-resolution at a world scale.

Data intensive research in numerical data is discussed by Alex Szalay. He brings experience in the development of the pivotal Sloane Digital Skyserver, which enables data intensive research in astronomy. He will explore the consequences and lessons learned for eScience research at large, including the fusion with computational science simulations.

Stratos Idreos and Daan Crommelin complete the grand visions with concrete data intensive research in database architectures and modelling at CWI.

www.cwi.nl/lectures2010
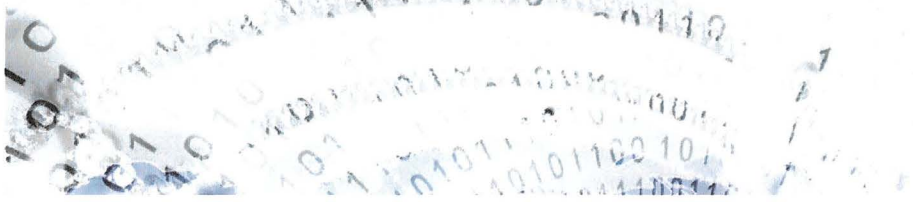
Speakers:

Hector Garcia-Molina is professor at the Department of Computer Science and Electrical Engineering at Stanford University (USA). He is author of over 370 papers on distributed databases, digital libraries and social networks. His current interest is to push the technology for knowledge extraction at Internet scale.

Alex Szalay is the Alumni Centennial Professor of Astronomy and Computer Science at the Johns Hopkins University. He is a cosmologist, working on the statistical measures of the spatial distribution of galaxies and galaxy formation. He has written over 450 papers, covering areas from theoretical cosmology to observational astronomy, spatial statistics and computer science.

Stratos Idreos is a tenure track researcher in the Information Systems Department at CWI. His interests are algorithms for adaptive database architectures.

Daan Crommelin is a researcher in the Dynamical Systems and Numerical Analysis group at CWI. His research focuses on stochastic methods for multiscale systems estimation of stochastic-dynamical models from data and applications in atmosphere-ocean science.

CWI LECTURES

Program

13.25	**Welcome by Martin Kersten**
13.30	**Hector Garcia-Molina**
	Entity Resolution: Identifying Real-World Entities in Large Data Sets
14.15	**Stratos Idreos**
	Database Cracking: Towards Auto-tuning Database Kernels
14.40	**Break**
15.00	**Alex Szalay**
	Amdahl's Laws and Extreme Data Intensive Computing
15.45	**Daan Crommelin**
	Blending Data with Models in Climate Science
16.10	**Wrapup by Jan Karel Lenstra**
16.15	**Drinks**

You are cordially invited!

Please register by filling out the online registration form at
http://www.cwi.nl/lectures2010

CWI

Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, The Netherlands
www.cwi.nl
+31205929333

CWI

Centrum Wiskunde & Informatica

Founded in 1946, CWI is the national research institute for mathematics and computer science. It is located at Science Park Amsterdam and is supported by the Netherlands Organisation for Scientific Research (NWO). The institute holds a strong international position. More than 150 researchers conduct pioneering research in about 70 projects, and share their acquired knowledge with society. Some 30 researchers are professor at a university. The institute generated 16 spin-off companies.

www.cwi.nl/lectures2010