Monitoring the Data Tsunami
(Johannes Gehrke)

By 2020, we will live in a world where small-scale sensors will be ubiquitous and pervasive, and where they will have the computational power to sense and monitor every corner of the world. Monitoring and understanding this flood of data is a grand research challenge. I will discuss two approaches for monitoring event streams. First, I will discuss how to summarize data streams through small synopsis data structures that enable answering of complex queries over the stream with little resource usage. Second, I will describe Cayuga, a system for large-scale notifications based on monitoring complex events. I will discuss the Cayuga query algebra, its compilation into finite state automata, and the performance of the system. I will conclude with some open problems in the field.

Johannes Gehrke is a Professor in the Department of Computer Science at Cornell University. Johannes' research interests are in the areas of database systems, data mining, and data privacy. Johannes received an NSF CAREER Award, an Arthur P. Sloan Fellowship, an IBM Faculty Award, and an Humboldt Prize. He co-authored the undergraduate textbook Database Management Systems (McGrawHill (2002), currently in its third edition), used at universities all over the world. Johannes is also an Adjunct Faculty Member at the University of Tromsø in Norway. From 2007 to 2008, he was Chief Scientist at FAST, A Microsoft Subsidiary. He is currently visiting the Max Planck Institute for Software Systems in Kaiserslautern and Saarbrucken.