

Keynote Talk

Information Management in the Cloud - Parallel Dataflow Programming Beyond Map/Reduce

Prof. Volker Markl

Technical University Berlin (TU-Berlin)

Abstract

We have been researching a massively parallel data processor in the Stratosphere Research Unit, a DFG funded project among TU Berlin, FU Berlin, and HPI Potsdam. The research of the data management group at TU Berlin in this area focuses on a new flavor of data processors that goes beyond the popular map/reduce paradigm. We propose a programming model based on second order functions that describe what we call parallelization contracts (PACTs). PACTs are a generalization of the map/reduce programming model, extending it with additional higher order functions and output contracts that give guarantees about the behavior of a function. A PACT program is transformed into a data flow for a massively parallel execution engine, which executes its sequential building blocks in parallel and provides communication, synchronization and fault tolerance. The concept of PACTs allows the system to abstract parallelization from the specification of the data flow and thus enables several types of optimizations on the data flow. The system as a whole is as generic as map/reduce systems, but can provide higher performance through optimization and adaptation of the system to changes in the execution environment. Moreover, it enables the execution of tasks that traditional map/reduce systems cannot execute without mixing data flow program specification and parallelization, like joins, time-series analysis or data mining operations. We will present our research vision and preliminary research results that we have achieved during the last year. We will also highlight our research agenda for the upcoming year.

Bio: Volker Markl is a Full Professor and Chair of the Database Systems and Information Management (DIMA) group at the Technische Universität Berlin (TU-Berlin). Prior to joining TU-Berlin, Dr. Markl led a research group at FORWISS, the Bavarian Research Center for Knowledge-based Systems in Munich, Germany, and was a research staff member and project leader at the IBM Almaden Research Center in San Jose, California, USA. His research interests include: information as a service, new hardware architectures for information management, information integration, autonomic computing, query processing, query optimization, data warehousing, electronic commerce, and pervasive computing.

Volker has presented over 100 invited talks in numerous industrial settings and at major conferences and research institutions worldwide. He has authored and published more than 50 research papers at world-class scientific venues. Volker regularly serves as member and chair for program committees of major international database conferences. He also is a member of the Board of Trustees of the VLDB Endowment. Volker has 5 patent awards, and he has submitted over 20 invention disclosures to date. Over the course of his career, he has garnered many prestigious awards, including the European Information Society and Technology Prize, an IBM Outstanding Technological Achievement Award, an IBM Shared University Research Grant, an HP Open Innovation Award, and the Pat Goldberg Memorial Best Paper Award.

Website: <http://www.dima.tu-berlin.de/menue/dimaparameteren>