Tim Kaldewey

Resume

505 Governors Ct.
Philadelphia, PA 19146
http://www.kaldewey.com

PHONE: (650) 799-1189
E-MAIL: tim@kaldewey.com

EDUCATION

Ph. D. Computer Science, University of California, Santa Cruz, March 2010
Thesis: *Predictable High-Performance Data Management*, Advisor: Prof. Scott Brandt

M. S. Computer Science, University of California, Santa Cruz, March 2008

M. S. Network Engineering, Institute Eurecom, Sophia Antipolis, FRANCE & University of Technology Darmstadt, GERMANY, September 2004

HONORS

IBM awards	2015, 2014, 2013
Best paper awards	SIGMOD 2010, RTAS 2008
Fellowships	Oracle 2010, 2009, 2008, University of California Regent's 2006, 2005, Eurecom 2003

SELECTED PUBLICATIONS

M. Gschwind, T. Kaldewey, D. K. Tam, "Optimizing Efficiency of Deep Learning Through Accelerator Virtualization", **IBM Journal of Research and Development**, Volume 61, Issue 4/5, 2017.

T. Kaldewey, D. Wendt, "Accelerating Document Retrieval and Ranking for Cognitive Applications", **GPU Technology Conference 2017**,

E. Sitaridi, R. Mueller, T. Kaldewey, G. Lohman, K. Ross "Massively-Parallel Lossless Data Decompression", **ICPP'16**,

T. Kaldewey, G. Lohman, R. Mueller, P. Volk., "GPU Join Processing Revisited", **DaMoN**'12

T. Kaldewey, S. Tata, E. Shekita., "Clydesdale: Structured Data Processing on MapReduce", **EDBT**'12.

T. Kaldewey, A. Di Blas, "Large-Scale GPU Search", **GPU Computing Gems**, Morgan Kaufmann, 2011.

PROFESSIONAL EXPERIENCE

May 2014–present	**IBM** Watson Group (R&D), New York NY, *Performance Architect*
November 2010–April 2014	**IBM** Almaden Research, San Jose CA, *Research Staff Member*
October 2011–present	**University of Pennsylvania**, Philadelphia PA, *Assistant Adjunct Professor*
Previously	**Oracle** Special Projects, **SAP** Research, **Lufthansa** Technik, **Software AG**

EXPERTISE

Performance/efficiency analysis and optimization at an architect level – leading technical teams

Design and development of parallel algorithms for low-latency and/or high-throughput computing

Parallel programming in CUDA (GPU), Assembly (SSE), C; Linux Kernel, real-time scheduling

INTERESTS

Performance/efficiency optimization of AI workloads

Parallel algorithms optimized for parallel architectures

System resource management to guarantee application performance

High-performance and large-scale data management on parallel systems