

Summary

Key strength are very strong practical skills combined with deep technical and theoretical abilities.
Background in databases, distributed systems, and programming languages; Expert in C++, Linux administration & networking.

Education

PH.D. COMPUTER SCIENCE from University of California, Davis, CA. **Received A+ in all graded courses** 2010
M.SC. COMPUTER SCIENCE from Technical U Ilmenau, Germany. **Graduated with best achievable grade. Top 1.5%** 2005

Professional Experience

SOFTWARE ENGINEER — **Google, Inc.**, Seattle, WA, USA Since Oct. 2015

- Work on Google Compute Engine: Memory-overhead reductions for virtual machines; harmonized customer and infrastructure view of CPU utilization; various other contributions (non-public).

COMPUTER SCIENTIST — **LogicBlox, Inc.**, Seattle, WA, USA Oct. 2011 – Sept. 2015

- Research & development work on LogicBlox Database engine. I contributed several core components including the multi-threaded server component, a decimal built-in datatype, hierarchical import/export, etc.
- Research & development into query compilation techniques. Prototype improves query time by 8x-15x.
- Research & development into using GPUs to accelerate relational queries. Prototype improves query times up to 100x.
- Foundational research into distributed Datalog evaluation. Work resulted in Best Paper at PODS.

INTERN — **Google, Inc.**, Mountain View, CA, USA — MapReduce Team July – Sept. 2008

- Designed, implemented, and benchmarked grouping strategies for MapReduce's shuffling for performance gains.

INTERN — **Google, Inc.**, Mountain View, CA, USA — Platforms Team July – Sept. 2007

- Modified Linux Kernel to gather kernel- and user-space stack traces during context switches to analyze performance of applications; integrated stack tracing using Google's trace gathering mechanism as well as the Linux Tool Tracekit (LTT).

INTERN — **MLDesign Technologies, Inc.**, Palo Alto, CA, USA Oct. 2002 – March 2003

- Built a model library for simulating network stacks based on the TCP/IP protocol suite.

ACTIVE MEMBER — **FeM Non-profit** Association, Ilmenau, Germany 1999 – 2005

- Designed and wrote an Oracle database application that automatically configures all of the approximately 2,000 student computers via DHCP, manages switch port configuration, security options, and VLAN configuration for the computers.
- Maintained several Linux servers and services, including SSH, DHCP, NTP, and SAMBA.

Selected Awards, Scholarships, Talks, and Presentations

Invited Talk: General-Purpose Join Algorithms for Large-Scale Graph Analysis. SUNY, Buffalo 2015

PODS 2014 **Best Paper Award**. Joint work with T. Ameloot, B. Ketsman, and F. Neven 2014

Above work selected to the **19th annual list of notable items published in computing** by ACM Computing Reviews 2014

Invited Talk: MapReduce, Hadoop, and Beyond. Amazon Cloud Computing for Workshop. UC Davis Genome Center 2010

ACM **Travel Award** (one of 7 awardees), ACM Conference on Computer and Communications Security 2007

\$12,000 Block Grant **Fellowship**, CS Department, UC Davis 2007

\$15,000 **Tuition Scholarship**, UC Davis 2005 – 2006

EUR 12,000 **Fellowship** from SD&M AG for superior career performance 2002 – 2003

Research Accomplishments

I (co-)authored 29 peer-reviewed articles. My work was published at Top-Tier CS venues: PODS, TODS, TOCL, 2xICDT, 3xICDE, CCGRID, VLDB, and CCS. Work was featured by BBC, WSJ Business Technology Blog, and others, including the German magazine Spiegel, which interviewed me for the article. Please see CV for more details.

I am an EU citizen and a permanent resident of the USA.

Daniel Zinn

(Detailed CV)

Skill Summary

Extensive experience in research, development, and engineering.

My key strength are very strong practical skills combined with deep technical and theoretical abilities.

My expertise is in databases, distributed systems, and programming languages.

I am expert in C++, Perl, C; Linux administration & networking.

Education

PH.D. COMPUTER SCIENCE — University of California, Davis, CA 2010

Advisor: Bertram Ludäscher and Todd J. Green

GPA 4.00 on a 4.00 scale, straight A+

(M.Sc.) DIPLOM COMPUTER SCIENCE (MATHEMATICS MINOR) — Technical University of Ilmenau, Germany 2005

Graduated with distinction. GPA 1.0 (Best achievable grade on German scale), among top 1.5%

Professional Experience

SOFTWARE ENGINEER — *Google, Inc., Seattle, WA, USA* Since Oct. 2015

- Work on Google Compute Engine in the VMX team.
- Memory-overhead reductions for virtual machines; harmonized customer and infrastructure view of CPU utilization; various other contributions (non-public).

COMPUTER SCIENTIST — *LogicBlox, Inc., Seattle, WA, USA* Oct. 2011 – Sept. 2015

- Research & development work on cloud-based distributed Database engine. I contributed several core components including the multi-threaded server component, a decimal built-in datatype, hierarchical import/export, etc.
- Research & development into query compilation techniques. Prototype improves query time by 8x-15x.
- Research & development into using GPUs to accelerate relational queries. Prototype improves query times up to 100x.
- Foundational research into distributed Datalog evaluation. Work resulted in Best Paper at PODS.

INTERN — *Google, Inc., Mountain View, CA, USA* July – Sept. 2008

- Designed, implemented, and benchmarked grouping strategies for MapReduce's shuffling for performance gains.
- Prototyped a proof-of-concept implementation of a fuse-based solution to mounting the Google-filesystem GFS.
- Worked closely with the MapReduce team at Google.

INTERN — *Google, Inc., Mountain View, CA, USA* July – Sept. 2007

- Modified Linux Kernel to gather kernel- and user-space stack traces during context switches to analyze performance of applications; integrated stack tracing using Google's trace gathering mechanism as well as the Linux Tool Tracekit (LTT).
- Proposed visualization for system-level process performance; co-designed and implemented prototype for processing trace data to generate the proposed graphical output images.

INTERN — *MLDesign Technologies, Inc., Palo Alto, CA, USA* Oct. 2002 – March 2003

- Built a model library for simulating network stacks based on the TCP/IP protocol suite.

ACTIVE MEMBER — *FeM Non-profit Association, Ilmenau, Germany* 1999 – 2005

- Designed and wrote an Oracle database application that automatically configures all of the approximately 2,000 student computers via DHCP, manages switch port configuration, security options, and VLAN configuration for the computers.
- Acted as system administrator for my dormitory. Work included projection and installation of optical fiber backbone.
- Maintained several Linux servers and services, including SSH, DHCP, NTP, and SAMBA.

COLLEGIATE EMPLOYEE — *Mission Level Design GmbH, Ilmenau, Germany* 2003 – 2005

- Extended the tool MLDesigner by proposing and developing a new domain that links it to the open source network simulator NS2 to co-simulate network applications and protocols. Efforts led to a new revision of the MLDesigner software.
- Built abstract model of the IEEE 802.11 standard for performance simulation of multiple 802.11b and 802.11g nodes.

Awards and Scholarships

PODS 2014 Best Paper Award . Joint work with T. Ameloot, B. Ketsman, and F. Neven	2014
Above work selected to the 19th annual list of notable items published in computing by ACM Computing Reviews	2014
ACM Travel Award (one of 7 awardees), ACM Conference on Computer and Communications Security	2007
\$12,000 Block Grant Fellowship , CS Department, UC Davis	2007
\$15,000 Tuition Scholarship , UC Davis	2005 – 2006
EUR 12,000 Fellowship from SD&M AG for superior career performance	2002 – 2003
3rd prize in state-wide mathematical contest (Thuringia, Germany)	1997
Winner of county-wide mathematical contest (Sonneberg/Neuhaus, Germany)	1996

Presentations

Invited Talk: General-Purpose Join Algorithms for Large-Scale Graph Analysis. SUNY, Buffalo.	Feb. 2015
Invited Talk: MapReduce, Hadoop, and Beyond. Amazon Cloud Computing for Workshop. UC Davis Genome Center.	2010
Research paper presentations: ICDDT'12, CCGrid'11, WORKS'10, IPAW'10, ICDE'10, WORKS'09, PTMini'09, ICDE'09, EDBT PhD'08, CCS'07, VLDB'07, PTMini'07. <i>See Publication section for details.</i>	

Research Experience

INDUSTRIAL RESEARCH — <i>LogicBlox, Inc., Seattle, USA</i>	since 2011
Research into distributed execution of relational queries and Datalog. <ul style="list-style-type: none">• Foundational work on coordination-complexity won best paper award at PODS 2014.• Research into query compilation. Prototype shows 8x-15x speed improvement over current LogicBlox engine.• Adapted Leapfrog-Tree-Join for NVidia GPUs with up to 100x speed improvement over current engine.	
POSTDOCTORAL RESEARCH — <i>Genome Center at University of California at Davis, CA, USA</i>	Feb. 2010 – Sept. 2011
Research into innovative programming models and parallel execution strategies for distributed applications. <ul style="list-style-type: none">• Research is motivated by Bioinformatics use-cases from the UCD Genome Center as well as business planning applications.• Theoretical research investigating the coordination-complexity of recursive Datalog with negation. Work resulted in ICDDT 2012 publication <i>Win-Move is Coordination-Free (Sometimes)</i>.	
DOCTORAL RESEARCH — <i>University of California at Davis, CA, USA</i>	2006 – 2010
Dissertation: <i>Design and Optimization of Scientific Workflows</i> . <ul style="list-style-type: none">• Formalization of VDAL (Virtual Data Assembly Lines), a dataflow-oriented paradigm for scientific workflows with data organized in nested collections [EDBT PhD'08, FGCS'08, WORKS'09].• Optimization of VDAL execution by exploiting data parallelism via a compilation to MapReduce/hadoop [JCSS'10].• Optimization of VDAL execution via static analysis of component configurations to reduce data shipping costs [ICDE'09].• Supporting workflow design via static analysis of component configurations: predicting workflow's output schema and identifying "idle" workflow components [ICDE'10].• Supporting workflow design by anticipating summarized provenance graphs using static analysis [IPAW'10].• Implementation of a process-network engine for a Linux cluster with Kepler as front-end [PTMini'08].	
Other Projects: <ul style="list-style-type: none">• Investigation of the Great Firewall of China: Used latent semantic analysis to efficiently probe for blocked words [CCS'07]. News coverage by BBC, Schneider on Security, EWeek, Slashdot, Epoch Times, Ars Technica, WSJ Business Technology Blog, etc., and foreign news sources of 10 different languages. I was interviewed by the German magazine Spiegel. Details under http://www.conceptdoppler.org/.• Outreach: Article in German journal on Internet censorship [digma 03/08].• Representation of fuzzy objects and probability distributions via a series decomposition over orthogonal, smooth basis functions. Implemented the data type with querying and index functionality in PostgreSQL [VLDB'07].	
MASTERS RESEARCH — <i>Technical University of Ilmenau, Germany</i>	2005
Thesis: <i>Skyline Queries in P2P Systems</i> [DBISP2P'05, CoopIS'07]. <ul style="list-style-type: none">• Design and analysis of a novel data structure called "QTree", and demonstrated its effectiveness as a routing index for rank-aware queries in P2P systems.• Definition of "approximated skyline" with proposed query strategies to process top-N, skyline queries, and approximated skyline queries in P2P systems.	

Technical Skills

- Programming Languages** Expert level in C++ [STL, Boost, MPL], Perl [CPAN], C.
Good working knowledge with Python, Java, OCaml, Haskell, i86-Assembler.
- Programming Experience** Writing production code for cloud-based smart database engine. (at LogicBlox)
Writing production code inside the Google code-base. (at Google)
Improving the Google MapReduce Implementation. (at Google)
Distributed applications using MapReduce, PVM, plain sockets on Linux Cluster. (at UC Davis)
Relational databases: PostgreSQL, MySQL. (at UC Davis)
Linux kernel programming. (at Google)
Embedded systems development based on ATMEL micro controlers. (personal interest)
System level modeling. (at MLDesigner Inc.)
- Operating Systems** Expert-level at Linux system administration & networking. (at FeM, personal interest)
General experiences with Windows.
- Technical Skills** Network design, config' and management with HP, Intel, Cisco, and 3COM devices. (at FeM)
VOIP solutions. (personal interest)

Selected Side Projects

- VOIP PHONE SYSTEM since 2006
- Asterisk VOIP server (first on SOEKRIS hardware, now on OpenWRT router).
 - Provides my US and German landline; free calls to German landlines.
- HOME AUTOMATION SYSTEM 2002
- Designed and implemented a distributed system for controlling lighting and music in my apartment.
- Wrote controlling server application, and firmware for the ATMEL microprocessor.
 - Designed and soldered circuit board for microprocessor; soldered AC-power-switching devices.
 - Wrote client/server programs for controlling the system over the Internet via custom web-interface.

Teaching Experience

Taught the following laboratory sessions as a teaching assistant at UC Davis and TU Ilmenau. Responsibilities: developing and leading discussion sessions, drafting quizzes, grading homework, quizzes and exams.

Introduction to Theory of Computation	UC Davis	Winter Quarter 2005
Practical Computer Science (Non-CS Majors)	TU Ilmenau	Winter Semester 2001/2002
Theoretical Computer Science	TU Ilmenau	Summer Semester 2001
Algorithms and Data Structures	TU Ilmenau	Winter Semester 2000/2001

Community Service

Program Committee member: 3rd Workshop on Scientific Cloud Computing (ScienceCloud 2012). 4rd USENIX Workshop on the Theory and Practice of Provenance (TaPP 2012), Intl. Conf. on Scientific and Statistical Database Management (SSDBM'11), 2nd ACM Workshop on Scientific Cloud Computing. Intl. Conf. on Scientific and Statistical Database Management (SSDBM'10).

Journal reviewer: The VLDBJ Journal 2012, ACM Transactions on Database Systems (TODS) 2011, Scientific Programming Journal, Special Issue on Science-driven Cloud Computing (SPJ ScienceCloud 2011), Transactions on Computational Biology and Bioinformatics (TCBB) 2010, IEEE Transactions on Automation Science and Engineering (T-ASE) 2008, IEEE Transactions on Knowledge and Data Engineering (TKDE) 2008.

Conference reviewer: PODS'12, SIGMOD'11, ICDT'11, EDBT'10, eScience'09, VLDB'08 demo, ESCW'07, VLDB'06.

Workshop reviewer: WORKS'10, WORKS'09, IDEAS'09, WebDB'09, DILS'09, DILS'07.

Personal

- Work permissions: EU Citizen, US permanent resident.
- Languages: native German, fluent English.

Publications – Major Conferences

- ICDT'15 Datalog Queries Distributing over Components. Tom J. Ameloot, Bas Ketsman, Frank Neven, Daniel Zinn. *18th International Conference on Database Theory (ICDT 2015)*, Brussels, Belgium. March 2015.
- PODS'14 Weaker Forms of Monotonicity for Declarative Networking: a More Fine-grained Answer to the CALM-conjecture. Tom J. Ameloot, Bas Ketsman, Frank Neven, Daniel Zinn. *Symposium on Principles of Database Systems (PODS 2014)*, Snowbird, UT, USA. June 2014.
PODS'14 Best Paper Award.
Selected to the 19th annual list of notable items published in computing by ACM Computing Reviews.
- ICDT'12 Win-Move is Coordination-Free (Sometimes). Daniel Zinn, Todd J. Green, Bertram Ludäscher. *15th International Conference on Database Theory (ICDT 2012)*, Berlin, Germany. March 2012.
- CCGrid'11 Towards Reliable, Performant Workflows for Streaming-Applications on Cloud Platforms. Daniel Zinn, Yogesh Simmhan, Michail Giakkoupis, Quinn Hart, Timothy McPhillips, Viktor K. Prasanna, Bertram Ludäscher. *11th IEEE Int'l Symposium on Cluster Computing and the Grid*, Newport Beach, CA, USA, May 2011.
- ICDE'11 Scientific Workflow Design 2.0: Demonstrating Streaming Data Collections in Kepler. Lei Dou, Daniel Zinn, Timothy McPhillips, Sven Köhler, Sean Riddle, Shawn Bowers, and Bertram Ludäscher. *27th IEEE International Conference on Data Engineering (ICDE 2011). Demo Track*, Hannover, Germany, April, 2011.
- ICDE'10 XML-Based Computation for Scientific Workflows. Daniel Zinn, Shawn Bowers, Bertram Ludäscher. *26th IEEE Int'l Conference on Data Engineering (ICDE 2010)*, Long Beach, CA, USA, March, 2010.
- ICDE'09 X-CSR: Dataflow Optimization for Distributed XML Process Pipelines. Daniel Zinn, Shawn Bowers, Timothy McPhillips, Bertram Ludäscher. *25th IEEE Int'l Conference on Data Engineering (ICDE 2009)*, Shanghai, China, March 2009.
- CCS'07 ConceptDoppler: A Weather Tracker for Internet Censorship. Jedidiah R. Crandall, Daniel Zinn, Earl Barr, Michael Byrd, and Rich East. *14th ACM Conference on Computer and Communications Security (CCS 2007)*, Alexandria, VA, USA, October 2007.
- CoopIS'07 A Relaxed but not Necessarily Constrained Way from the Top to the Sky. Katja Hose, Christian Lemke, Kai-Uwe Sattler, Daniel Zinn. *15th International Conference on Cooperative Information Systems (CoopIS 2007)*, Vilamoura, Algarve, November 2007.
- VLDB'07 Modeling and Querying Vague Spatial Objects Using Shapelets. Daniel Zinn, Jim Bosch, and Michael Gertz. *33rd Int'l Conference on Very Large Data Bases (VLDB 2007)*, Vienna, Austria, September 2007.
- MASCOTS'03 System-Level Simulation Modeling with MLDesigner. Gunar Schorcht, Ian A. Troxel, Keyvan Farhangian, Peter Unger, Daniel Zinn, Colin K. Mick, Alan D. George, and Horst Salzwedel. *11th International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2003)*, Orlando, FL, USA, October 2003.

Publications – Journals and Book Chapters

- TOCL'17 Datalog Queries Distributing over Components. Tom J. Ameloot, Bas Ketsman, Frank Neven, Daniel Zinn. *ACM Transactions on Computational Logic (TOCL)*, Volume 18 Issue 1, March 2017.
- TODS'16 Weaker Forms of Monotonicity for Declarative Networking: a More Fine-grained Answer to the CALM-Conjecture. Tom J. Ameloot, Bas Ketsman, Frank Neven, Daniel Zinn. *ACM Transactions on Database Systems*, Volume 40 Issue 4, February 2016.
- LNCS'13 First-Order Provenance Games. Sven Köhler, Bertram Ludäscher, Daniel Zinn. *In Search of Elegance in the Theory and Practice of Computation*, pages 382–399. 2013.
- JCSS'10 Parallelizing XML Processing Pipelines via MapReduce. Daniel Zinn, Shawn Bowers, Sven Köhler, Bertram Ludäscher. *Journal of Computer and System Sciences (JCSS 2010)*, doi:10.1016/j.jcss.2009.11.006.

- digma 03/08 Wer sucht, der findet (nicht immer). Daniel Zinn. *digma – Zeitschrift für Datenrecht und Informationssicherheit*, Heft 3, September 2008. Who searches will (not always) find. Article on Internet censorship.
- FGCS'08 Scientific Workflow Design for Mere Mortals. Timothy McPhillips, Shawn Bowers, Daniel Zinn, Bertram Ludäscher. *Future Generation Computer Systems (FGCS 2008)*, doi:10.1016/j.future.2008.06.013.

Publications – Smaller Conferences and Workshops

- GPGPU @PPoPP'16 General-purpose join algorithms for large graph triangle listing on heterogeneous systems. Daniel Zinn, Haicheng Wu, Jin Wang, Molham Aref, Sudhakar Yalamanchili. *General-Purpose GPU @ Principles and Practice of Parallel Programming (GPGPU@PPoPP 2016)*, Barcelona, Spain. March 2016.
- ADMS'14 Multipredicate Join Algorithms for Accelerating Relational Graph Processing on GPUs. Haicheng Wu, Daniel Zinn, Molham Aref, and Sudhakar Yalamanchili. *Fifth Int'l Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS 2014)*, Hangzhou, China. September 2014.
- SSDBM'11 ProPub: A Declarative Approach for Publishing Customized, Policy-Aware Provenance. Saumen C Dey, Daniel Zinn, Bertram Ludäscher. *23rd International Conference on Scientific and Statistical Database Management (SSDBM 2011)*, Portland, OR, USA. July 2011.
- SSDBM'11 Improving Workflow Fault Tolerance through Provenance-based Recovery. Sven Köhler, Sean Riddle, Daniel Zinn, Timothy McPhillips, Bertram Ludäscher. *23rd International Conference on Scientific and Statistical Database Management (SSDBM 2011)*, Portland, OR, USA. July 2011.
- WORKS'10 Streaming Satellite Data to Cloud Workflows for On-Demand Computing of Environmental Products. Daniel Zinn, Quinn Hart, Bertram Ludäscher, Yogesh Simmhan. *5th Workshop on Workflows in Support of Large-Scale Science (WORKS 2010)*, In conjunction with SC 2010, New Orleans, LA, November 2010.
- IPAW'10 Abstract Provenance Graphs: Anticipating and Exploiting Schema-Level Data Provenance. Daniel Zinn, Bertram Ludäscher. *Third Int'l Provenance and Annotation Workshop (IPAW 2010)*, Troy, NY, USA, June, 2010.
- WORKS'09 Scientific Workflow Design with Data Assembly Lines. Daniel Zinn, Shawn Bowers, Timothy McPhillips, Bertram Ludäscher. *4th Workshop on Workflows in Support of Large-Scale Science (WORKS 2009)*, In conjunction with SC 2009, Portland, OR, November 2009.
- PTMini'09 Parallel Virtual Machines in Kepler. Daniel Zinn, Xuan Li, Bertram Ludäscher. *8th Biennial Ptolemy Miniconference*, Berkeley, CA, USA. April 2009.
- EDBT PhD'08 Modeling and Optimization of Scientific Workflows. Daniel Zinn. *EDBT 2008 Ph.D. Workshop (EDBT Ph.D. 2008)*, Nantes, France, March 2008.
- PTMini'07 Scheduling Data-Intensive Workflows in Kepler. Daniel Zinn, Tim H. Wong, Bertram Ludäscher. *7th Biennial Ptolemy Miniconference*, Berkeley, CA, USA. Februar 2007.
- DBISP2P'05 Processing Rank-Aware Queries in P2P Systems. Marcel Karnstedt, Katja Hose, Anke Koch, Kai-Uwe Sattler, and Daniel Zinn. *Third International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P 2005)*, Trondheim, Norway, August 2005.
- WebDB'05 Processing Top-N Queries in P2P-based Web Integration Systems with Probabilistic Guarantees. Katja Hose, Marcel Karnstedt, Kai-Uwe Sattler, and Daniel Zinn. *Proceedings of the Eight International Workshop on the Web & Databases (WebDB 2005)*, Baltimore, Maryland, USA, June 2005.

Publications – Misc

- arxiv'15 General-Purpose Join Algorithms for Listing Triangles in Large Graphs. Daniel Zinn.