

JOY ARULRAJ

BARRY DICKMAN EARLY CAREER
ASSISTANT PROFESSOR

SCHOOL OF COMPUTER SCIENCE
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332

arulraj@gatech.edu
<http://www.cc.gatech.edu/~jarulraj>

Contents

I	EARNED DEGREES	2
II	EMPLOYMENT HISTORY	2
III	HONORS AND AWARDS	2
IV	RESEARCH	2
A	REFERRED PUBLICATIONS AND SUBMITTED ARTICLES	2
A.1	Thesis	2
A.2	Tutorials	2
A.3	Conference Publications	3
A.4	Workshop Publications	3
A.5	Invited Articles	4
A.6	Preprints	4
B	SOFTWARE ARTIFACTS	4
C	PRESENTATIONS (SELECTED)	5
C.1	Academic Talks	5
C.2	Industry Talks	6
D	SOCIETAL AND POLICY IMPACTS	6
D.1	Research Coverage in the News	6
V	TEACHING	7
A	COURSES TAUGHT	7
B	INDIVIDUAL STUDENT GUIDANCE	7
B.1	Ph.D. Students (Currently Advising or Co-advising)	7
B.2	PhD/MS Students (Informal or Past advising)	7
VI	SERVICE	8
A	EXTERNAL SERVICE ACTIVITIES	8
A.1	Program Committee Member	8
A.2	Journal Reviewer	8
A.3	Workshop Program Committee Member	8
A.4	Tutorials	8
B	INTERNAL SERVICE ACTIVITIES	8

I EARNED DEGREES

Ph.D.	Computer Science, Carnegie Mellon University	2018
M.S.	Computer Sciences, University of Wisconsin, Madison	2013
B.E.	Computer Science and Engineering, College of Engineering, Guindy	2011

II EMPLOYMENT HISTORY

Assistant Professor	School of Computer Science, Georgia Institute of Technology, Atlanta, GA	2018–present
Research Intern	Database Group, Microsoft Research, Redmond, WA	Fall 2016
Engineering Intern	Amazon, India	Summer 2011

III HONORS AND AWARDS

1. [ACM SIGMOD Jim Gray Doctoral Dissertation Award](#), 2019
2. [Class of 1969 Teaching Fellow](#), 2019
3. [Barry Dickman Early Career Professorship](#), 2018
4. [Carnegie Mellon Presidential Fellowship](#), 2017
5. [ACM Heidelberg Laureate](#), 2017
6. [Samsung Ph.D. Fellowship](#), 2016
7. [UW-Madison Alumni Graduate Fellowship](#), 2011

IV RESEARCH

A REFERRED PUBLICATIONS AND SUBMITTED ARTICLES

A.1 Thesis

1. **Ph.D. Thesis**
Title: The Design and Implementation of a Non-Volatile Memory DBMS
Advisor: Andrew Pavlo
Carnegie Mellon University

A.2 Tutorials

1. Joy Arulraj and Andrew Pavlo.
[How to Build a Non-Volatile Memory Database System.](#)
SIGMOD'17: 44th ACM SIGMOD Int'l Conf. on the Management of Data, Chicago, IL, 2017.

A.3 Conference Publications

1. Jinho Jung, Hong Hu, Joy Arulraj, Taesoo Kim, and Woon-Hak Kang.
[APOLLO: automatic detection and diagnosis of performance regressions in database systems.](#)
VLDB'20: 45th Int'l Conf on Very Large Data Bases, Tokyo, Japan, 2020.
2. Qi Zhou, Joy Arulraj, Shamkant B. Navathe, William Harris, and Dong Xu.
[Automated verification of query equivalence using satisfiability modulo theories.](#)
VLDB'19: 45th Int'l Conf on Very Large Data Bases, Los Angeles, CA, 2019.
3. Joy Arulraj, Justin Levandoski, Umar Farooq Minhas, and Per-Ake Larson.
[BzTree: A High-Performance Latch-free Range Index for Non-Volatile Memory.](#)
VLDB'18: 44th Int'l Conf on Very Large Data Bases, Rio de Janeiro, Brazil, 2018.
4. Kai Ren, Qing Zheng, Joy Arulraj, and Garth Gibson.
[SlimDB: A Space-Efficient Key-Value Storage Engine For Semi-Sorted Data.](#)
VLDB'18: 44th Int'l Conf on Very Large Data Bases, Rio de Janeiro, Brazil, 2018.
5. Joy Arulraj, Matthew Perron, and Andrew Pavlo.
[Write-Behind Logging.](#)
VLDB'17: 43rd Int'l Conf on Very Large Data Bases, Munich, Germany, 2017.
6. Yingjun Wu, Joy Arulraj, Jiexi Lin, Ran Xian, and Andrew Pavlo.
[An Empirical Evaluation of In-Memory Multi-Version Concurrency Control.](#)
VLDB'17: 43rd Int'l Conf on Very Large Data Bases, Munich, Germany, 2017.
7. Andrew Pavlo, Gustavo Angulo, Joy Arulraj, Haibin Lin, Jiexi Lin, Lin Ma, Prashanth Menon, Todd C. Mowry, Matthew Perron, Ian Quah, Siddharth Santurkar, Anthony Tomasic, Skye Toor, Dana Van Aken, Ziqi Wang, Yingjun Wu, Ran Xian, and Tieying Zhang .
[Self-Driving Database Management Systems.](#)
CIDR'17: 8th Conf. on Innovative Data Management, Chaminade, CA, 2017.
8. Joy Arulraj, Andrew Pavlo, and Prashanth Menon.
[Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads.](#)
SIGMOD'16: 43rd ACM SIGMOD Int'l Conf. on the Management of Data, San Francisco, CA, 2016.
9. Joy Arulraj, Andrew Pavlo, Subramanya Dullloor.
[Let's Talk About Storage & Recovery Methods for Non-Volatile Memory Database Systems.](#)
SIGMOD'15: 42nd ACM SIGMOD Int'l Conf. on the Management of Data, Melbourne, Australia, 2015.
10. Joy Arulraj, Guoliang Jin, Shan Lu.
[Leveraging the Short-term Memory of Hardware to Diagnose Software Failures.](#)
ASPLOS'14: 19th ACM Int'l Conf. on Architectural Support for Programming Languages and Operating Systems, Salt Lake City, UT, 2014.
11. Joy Arulraj, Po-Chun Chang, Guoliang Jin, Shan Lu.
[Production-run Software Failure Diagnosis via Hardware Performance Counters.](#)
ASPLOS'13: 18th ACM Int'l Conf. on Architectural Support for Programming Languages and Operating Systems, Houston, TX, 2013.

A.4 Workshop Publications

1. Lin Ma, Joy Arulraj, Sam Zhao, Andrew Pavlo, Subramanya R. Dullloor, Michael J. Giardino, Jeff Parkhurst, Jason L. Gardner, Kshitij Doshi, and Col. Stanley Zdonik.
[Larger-than-Memory Data Management on Modern Storage Hardware for In-Memory OLTP Database Systems.](#)
DAMON @ SIGMOD'16: 12th Int'l Workshop on Data Management on New Hardware, San Francisco, CA, 2016.
2. Justin DeBrabant, Joy Arulraj, Andrew Pavlo, Michael Stonebraker, Stanley B. Zdonik, Subramanya Dullloor.
[A Prolegomenon on OLTP Database Systems for Non-Volatile Memory.](#)
ADMS @ VLDB'14: 5th Int'l Workshop on Accelerating Data Management Systems Using Modern Processor and

A.5 Invited Articles

1. Joy Arulraj
[Data management on non-volatile memory](#)
SIGMOD'19: 46th ACM SIGMOD Int'l Conf. on the Management of Data, Amsterdam, The Netherlands, 2019.
2. Joy Arulraj and Spyros Blanas.
[Data management on non-volatile memory.](#)
ACM SIGARCH Computer Architecture Today Blog, September 2018.
3. Peter Bailis, Camille Fournier, Joy Arulraj, and Andrew Pavlo.
[Research for Practice: Distributed Consensus and Implications of NVM on Database Management Systems.](#)
ACM Queue, July 2016.

A.6 Preprints

1. Joy Arulraj, Ran Xian, Lin Ma, and Andy Pavlo.
[Predictive Indexing.](#)
2. Joy Arulraj, Andy Pavlo, and Krishna Teja Malladi.
[Multi-Tier Buffer Management and Storage System Design for Non-Volatile Memory.](#)

B SOFTWARE ARTIFACTS

1. **EQUITAS** is a tool for automated verification of equivalence of SQL queries. It is based on a novel approach to determining query equivalence based on symbolic representation. The key idea is to effectively transform a wide range of SQL queries into first order logic formulae and then use an SMT solver to efficiently verify their equivalence.
2. **APOLLO** is a toolchain for automatically detecting, reporting, and diagnosing performance regressions in DBMSs. It automates the generation of regression-triggering queries, simplifies the bug reporting process for users, and enables developers to quickly pinpoint the root cause of performance regressions.
3. **PELTON** is a new database management system (DBMS) designed for emergent non-volatile memory (NVM) technologies. As part of my doctoral dissertation, I have worked on several research projects based on this system.
 - **WRITE-BEHIND LOGGING** is a novel logging and recovery algorithm that improves the availability of the system by 100× compared to the ubiquitous write-ahead logging algorithm. It leverages the durability and byte-addressability properties of NVM to avoid unnecessary data duplication.
 - **BZTREE** is a high-performance latch-free range index tailored for NVM that supports near-instantaneous recovery without requiring special-purpose recovery code. Its design reduces software development and maintenance complexity by enabling the same implementation to run on both volatile DRAM and on NVM without any code changes.
 - **LOGICAL TILE ALGEBRA** allows the DBMS to execute queries over data stored under different hybrid storage layouts without requiring separate execution engines and with minimal overhead.
 - **FLEXIBLE STORAGE MODEL** generalizes the canonical n-ary and decomposition storage models. It enables continuous incremental refinement of the database's physical design using a machine learning model on evolving hybrid transaction/analytical processing workloads.
 - The **SELF-DRIVING ARCHITECTURE** of Peloton leverages advancements in deep neural networks, improved hardware, and high-performance database architectures to remove the human capital impediments of deploying DBMSs and allow organizations to more easily derive the benefits of data-driven decision making applications.
 - **PREDICTIVE INDEXING** continuously improves a database's index configuration using lightweight physical design changes. It uses a reinforcement learning model to forecast the utility of these changes, and continuously

refines the index configuration of the database to handle evolving workloads.

4. **N-STORE** is a full-featured DBMS testbed that is specifically designed for exploring storage engine architectures that efficiently support on-line transaction processing on NVM. The NVM-aware storage engine architectures enable higher performance compared to their traditional counterparts, support instantaneous recovery from system failures, and extend the lifetime of the NVM device itself. N-STORE is included as a part of Wisconsin's **WHISPER benchmark suite** for NVM technologies.
5. **POSTGRESQL-CPP** is a port of the PostgreSQL DBMS to the C++ language. It spawned a subsequent official porting effort by PostgreSQL developers.
6. **SQLCHECK** is a dialect-agnostic static-analysis tool for automatically detecting common anti-patterns in SQL queries. It presents domain-specific knowledge in the form of helpful hints.

C PRESENTATIONS (SELECTED)

C.1 Academic Talks

- **Debugging Your Database System Using Apollo**
HPTS, November 2019, Pacific Grove, CA
- **ACM SIGMOD Jim Gray Doctoral Dissertation Award Talk**
SIGMOD, June 2019, Amsterdam, The Netherlands
- **BzTree: A High-Performance Latch-free Range Index for Non-Volatile Memory**
VLDB, August 2018, Rio de Janeiro, Brazil
- **SlimDB: A Space-Efficient Key-Value Storage Engine For Semi-Sorted Data**
VLDB, August 2018, Rio de Janeiro, Brazil
- **Data Management on Non-Volatile Memory**
University of California, October 2018, Santa Cruz, CA
University of California, May 2018, Berkeley, CA
Princeton University, April 2018, Princeton, NJ
University of Texas, April 2018, Austin, TX
University of California, April 2018, Santa Barbara, CA
Yale University, April 2018, New Haven, CT
Microsoft Research, March 2018, Redmond, WA
Swiss Federal Institute of Technology, March 2018, Zurich, Switzerland
University of Illinois Urbana-Champaign, March 2018, Champaign, IL
Northwestern University, March 2018, Evanston, IL
Georgia Institute of Technology, March 2018, Atlanta, GA
Cornell University, March 2018, Ithaca, NY
University of Utah, March 2018, Salt Lake City, UT
Duke University, February 2018, Durham, NC
IBM Research, February 2018, Almaden, CA
University of California, February 2018, Los Angeles, CA
VoltDB, January 2018, Bedford, MA
- **What Non-Volatile Memory Means for Database Systems**
MIT Database Group, November 2017, Cambridge, MA
Duke Database Group, November 2017, Durham, NC
- **OLTP on Non-Volatile Memory**
HPTS, October 2017, Pacific Grove, CA
- **Non-Volatile Memory Database Systems**
Stanford InfoLab, Sep 2016, Stanford, CA
- **Write-Behind Logging**

PDL Retreat, October 2017, Bedford, PA

VLDB, Aug 2017, Munich, Germany

- **How to Build a Non-Volatile Memory Database System**
NEDB @ MIT, Jan 2017, Cambridge, MA
- **Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads**
ACM SIGMOD, June 2016, San Francisco, CA
PDL Retreat, October 2015, Bedford, PA
- **Let's Talk About Non-Volatile Memory Database Systems**
ACM SIGMOD, June 2015, Melbourne, Australia
MIT Database Group, September 2014, Cambridge, MA
Brown Data Management Research Group, September 2014, Providence, RI
PDL Retreat, October 2014, Bedford, PA
- **So You Want to Fork PostgreSQL**
HPTS, September 2015, Pacific Grove, CA
- **Leveraging the Short-term Memory of Hardware to Diagnose Software Failures**
ACM ASPLOS, March 2014, Salt Lake City, UT
- **OLTP Database Systems for Next Generation Non-Volatile Memory**
PDL Retreat, October 2013, Bedford, PA
- **Production-run Software Failure Diagnosis via Hardware Performance Counters**
ACM ASPLOS, March 2013, Houston, TX

C.2 Industry Talks

1. **Write-Behind Logging**
SAP Labs, Sep 2017, Walldorf, Germany
2. **BzTree: An NVM-Aware Latch-Free B+Tree**
Microsoft Research, Dec 2016, Redmond, WA
3. **What Non-Volatile Memory Means for the Future Of Database Systems**
VMware Research, Sep 2016, Palo Alto, CA
4. **Thus Spoke Arulraj: Non-Volatile Memory for Databases**
Intel ISTC Big Data Retreat, Aug 2016, Hillsboro, OR
5. **Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads**
Oracle Labs, April 2016, Redwood City, CA
6. **The Case for Flexible Storage Model**
Huawei, April 2016, Santa Clara, CA
7. **Rethinking Database Systems for Next-Generation Memory Technologies**
Samsung Research, March 2016, Mountain View, CA
8. **Let's Talk About Non-Volatile Memory Database Systems**
IBM Research, September 2015, San Jose, CA
EMC Flash Forum, September 2014, Hopkinton MA

D SOCIETAL AND POLICY IMPACTS

D.1 Research Coverage in the News

- Squashed Bugs, Served Hot and Fresh with Failure Rate Heatmaps
[PingCAP Blog](#), December 2019
- N-Store is a part of Wisconsin-HPL Suite for Persistence
[WHISPER](#), March 2017

- Self-driving database management systems
[The Morning Paper Blog](#), [HackerNews](#), January 2017
- Let's talk about storage and recovery methods for non-volatile memory database systems
[The Morning Paper Blog](#), [HackerNews](#), September 2016
- C++ port of Postgres
[PostgreSQL Hackers Mailing List](#), August 2016
- Research for Practice: Distributed Consensus and Implications of NVM on Database Management Systems
[ACM Queue](#), July 2016

V TEACHING

A COURSES TAUGHT

Semester	Course Number	Course Title	Enrollment
Fall 2019	CS 8803 DDL	Data Analytics using Deep Learning	42
Spring 2019	CS 4420/6422	Database System Implementation	50
Fall 2018	CS 8803 DDL	Data Analytics using Deep Learning	12

B INDIVIDUAL STUDENT GUIDANCE

B.1 Ph.D. Students (Currently Advising or Co-advising)

- **Jaeho Bang**
Fall 2018-
Status: Pre-Qualifier
Topic: *Exploratory Video Analytics*
- **Pramod Chunduri**
Fall 2018-
Status: Pre-Qualifier
Topic: *Exploratory Video Analytics*
- **Jinho Jung**
Fall 2018-
Status: Pre-Qualifier (co-advised with Taesoo Kim)
Topic: *Debugging Database Management Systems*
- **Qi Zhou**
Fall 2018-
Status: Pre-Qualifier (co-advised with Bill Harris)
Topic: *Automated SQL Solver*

B.2 PhD/MS Students (Informal or Past advising)

- CMU: Matthew Perron (First employment: PhD student at MIT), Ran Xian (First employment: Facebook), Qian Li (First employment: PhD student at Stanford), Lin Ma (First employment: PhD student at CMU), Ziqiang Feng (First employment: PhD student at CMU), Edward Yeo (First employment: Two Sigma)

VI SERVICE

A EXTERNAL SERVICE ACTIVITIES

A.1 Program Committee Member

- SIGMOD 2020
- VLDB 2020-21
- CIDR 2019

A.2 Journal Reviewer

VLDB J., 2019; DAPD, 2016-18; JPDC, 2016; IEEE TKDE, 2017

A.3 Workshop Program Committee Member

SIGMOD SRC 2019

A.4 Tutorials

- Invited tutorial at SIGMOD'17: 44th ACM SIGMOD Int'l Conf. on the Management of Data, Chicago, IL, 2017, titled *How to Build a Non-Volatile Memory Database System*, with Andy Pavlo, 2017.

B INTERNAL SERVICE ACTIVITIES

- SCS School Advisory Committee, 2020-21
- SCS Ph.D. Visit Day Committee, 2019