

JOY ARULRAJ

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	3
II	EMPLOYMENT HISTORY	3
III	HONORS AND AWARDS	3
IV	RESEARCH, SCHOLARSHIP, AND CREATIVE ACTIVITIES	3
A	PUBLISHED BOOKS, BOOK CHAPTERS, AND EDITED VOLUMES	3
A.1	Published Books	3
B	REFERRED PUBLICATIONS AND SUBMITTED ARTICLES	4
B.1	Thesis	4
B.2	Tutorials	4
B.3	Published and Accepted Journal Articles	4
B.4	Conference Presentation with Proceedings (Refereed)	5
B.5	Workshop Publications	6
B.6	Invited Articles	7
B.7	Demonstrations	7
B.8	Preprints	7
C	OTHER PUBLICATIONS AND CREATIVE PRODUCTS	8
C.1	Software Artifacts	8
D	PRESENTATIONS (SELECTED)	10
D.1	Academic Talks	10
D.2	Industry Talks	12
E	GRANTS	12
E.1	As Principal Investigator	12
E.2	As Co-Principal Investigator	14
E.3	As Senior Personnel or Contributor	14
E.4	Pending Proposals	14
E.5	Proposals Submitted but Not Funded (Last Two Years)	14
E.6	Equipment Grants	14
F	OTHER SCHOLARLY AND CREATIVE ACCOMPLISHMENTS	14
G	SOCIETAL AND POLICY IMPACTS	14

G.1	Research Coverage in the News	14
V	TEACHING	15
A	COURSES TAUGHT	15
B	INDIVIDUAL STUDENT GUIDANCE	15
B.1	Ph.D. Students (Currently Advising or Co-advising)	15
B.2	Ph.D. Students (Graduated)	16
B.3	M.S. Students (Special Problems, Project, and Thesis)	16
B.4	Undergraduate Students	16
B.5	Service on Thesis or Dissertation Committees	16
B.6	PhD/MS Students (Informal or Past advising)	16
C	EDUCATIONAL INNOVATIONS AND OTHER CONTRIBUTIONS	17
VI	SERVICE	17
A	PROFESSIONAL CONTRIBUTIONS	17
A.1	Program Committee Member	17
A.2	Journal Reviewer	17
A.3	Workshop Program Committee Member	18
A.4	Tutorials	18
B	PUBLIC AND COMMUNITY SERVICE	18
C	INSTITUTE CONTRIBUTIONS	18

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. NSF CAREER Award, 2023
2. IEEE TCDE Rising Star Award, 2022
3. ACM SIGMOD Jim Gray Doctoral Dissertation Award, 2019
4. Class of 1969 Teaching Fellowship, 2019
5. Barry Dickman Early Career Professorship, 2018
6. Carnegie Mellon Presidential Fellowship, 2017
7. ACM Heidelberg Laureate, 2017
8. Samsung Ph.D. Fellowship, 2016
9. UW-Madison Alumni Graduate Fellowship, 2011

IV RESEARCH, SCHOLARSHIP, AND CREATIVE ACTIVITIES

A PUBLISHED BOOKS, BOOK CHAPTERS, AND EDITED VOLUMES

A.1 Published Books

1. Joy Arulraj and Andrew Pavlo.
Non-Volatile Memory Database Management Systems.

Synthesis Lectures on Data Management, 14:1, 191 pages
Morgan & Claypool Publishers, 2019.

B REFERRED PUBLICATIONS AND SUBMITTED ARTICLES

B.1 Thesis

1. Ph.D. Thesis

Title: The Design and Implementation of a Non-Volatile Memory DBMS

Date: June 2018

Advisor: Andrew Pavlo

Carnegie Mellon University

B.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.

B.3 Published and Accepted Journal Articles

Legend: Student co-authors are written in boldface.

1. Jaeho Bang, Gaurav Kakkar, Subrata Mitra, and Joy Arulraj.

SEIDEN: Revisiting Query Processing in Video Database Systems.

VLDB'23: 48th Int'l Conf on Very Large Data Bases, Vancouver, Canada, 2023.

Proceedings of the VLDB Endowment, Vol. 16, No. 11.

2. John A. Christian and Joy Arulraj.

Review of the k-Vector and Its Relation to Classical Data Structures

Journal of Guidance, Control, and Dynamics 2022.

3. Abhijit Suprem, Joy Arulraj, Calton Pu, and Joao Ferreira.

ODIN: Automated Drift Detection and Recovery in Video Analytics.

VLDB'20: 45th Int'l Conf on Very Large Data Bases, Tokyo, Japan, 2020.

Proceedings of the VLDB Endowment, Vol. 13, No. 11.

4. 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.

Proceedings of the VLDB Endowment, Vol. 13, No. 1.

5. 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.

Proceedings of the VLDB Endowment, Vol. 12, No. 11.

6. 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. Proceedings of the VLDB Endowment, Vol. 11, No. 5.
7. 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. Proceedings of the VLDB Endowment, Vol. 10, No. 13.
 8. Joy Arulraj, Matthew Perron, and Andrew Pavlo.
Write-Behind Logging.
VLDB'17: 43rd Int'l Conf on Very Large Data Bases, Munich, Germany, 2017. Proceedings of the VLDB Endowment, Vol. 10, No. 4.
 9. 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. Proceedings of the VLDB Endowment, Vol. 10, No. 7.

B.4 Conference Presentation with Proceedings (Refereed)

1. **Pramod Chundhuri, Jaeho Bang, Yao Lu, and Joy Arulraj.**
Zeus: Efficiently Localizing Actions in Videos using Reinforcement Learning.
SIGMOD'22: 49th ACM SIGMOD Int'l Conf. on the Management of Data, Philadelphia, PA, 2022.
2. **Zhuangdi Xu, Gaurav Kakkar, Joy Arulraj, and Umakishore Ramachandran.**
EVA: A Symbolic Approach to Accelerating Exploratory Video Analytics with Materialized Views.
SIGMOD'22: 49th ACM SIGMOD Int'l Conf. on the Management of Data, Philadelphia, PA, 2022.
3. **Jiashen Cao, Ramyad Hadidi, Joy Arulraj, and Hyesoon Kim.**
FiGO: Fine-Grained Query Optimization in Video Analytics.
SIGMOD'22: 49th ACM SIGMOD Int'l Conf. on the Management of Data, Philadelphia, PA, 2022.
4. **Qi Zhou, Joy Arulraj, Shamkant Navathe, William Harris, and Jinpeng Wu.**
SPES: A Symbolic Approach to Proving Query Equivalence Under Bag Semantics.
ICDE '22: 38th IEEE Int'l Conf. on Data Engineering, Virtual, 2022.
5. **Xinyu Liu, Qi Zhou, Joy Arulraj, and Alessandro Orso.**
Automatic Detection of Performance Bugs in Database Systems using Equivalent Queries.
ICSE'22: 44th ACM SIGSOFT Int'l Conf. on Software Engineering, Pittsburgh, PA, 2022.
6. **Qi Zhou, Joy Arulraj, Shamkant Navathe, William Harris, and Qiupeng Wu.**
Sia: Optimizing Queries using Learned Predicates.
SIGMOD'21: 48th ACM SIGMOD Int'l Conf. on the Management of Data, Xi'an, Shaanxi, China, 2021.
7. **Xinjing Zhou, Joy Arulraj, Andy Pavlo, and Dave Cohen.**
SPITFIRE: A Three-Tier Buffer Manager for Volatile and Non-Volatile Memory.

- SIGMOD'21: 48th ACM SIGMOD Int'l Conf. on the Management of Data, Xi'an, Shaanxi, China, 2021.
8. Prashanth Dintyala, Arpit Narechania, and Joy Arulraj.
[SQLCheck: Automated Detection and Diagnosis of SQL Anti-Patterns.](#)
SIGMOD'20: 47th ACM SIGMOD Int'l Conf. on the Management of Data, Portland, OR, 2020.
 9. Andrew Pavlo, Gustavo Angulo, Joy Arulraj, Haibin Lin, Jiexi Lin, Lin Ma, Prashanth Menon, Todd C. Mowry, Matthew Perron, Ian Quah, Siddharth Santurkar, Anthony Tomic, 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.
 10. 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.
 11. 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.
 12. 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.
 13. 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.

B.5 Workshop Publications

1. Gaurav Tarlok Kakkar, Jiashen Cao, Pramod Chunduri, Zhuangdi Xu, Suryatej Reddy Vyalla, Anirudh Prabakaran, Jaeho Bang, Kaushik Ravichandran, Ishwarya Sivakumar, Aryan Rajoria, Ashmita Raju, Tushar Aggarwal, Shashank Suman, Myna Prasanna Kalluraya, Subrata Mitra, Ali Payani, Yao Lu, Umakishore Ramachandran, Joy Arulraj.
[EVA: An End-to-End Exploratory Video Analytics System.](#)
DEEM @ SIGMOD'23: 7th Workshop on Data Management for End-to-End Machine Learning, Seattle, WA, 2023.
2. Jiashen Cao, Ramyad Hadidi, Joy Arulraj, Hyesoon Kim.
[Video analytics from edge to server: work-in-progress.](#)
CODES+ISSS: Int'l Conference on Hardware/Software Codesign and System Synthesis Companion, 2019.

3. Lin Ma, Joy Arulraj, Sam Zhao, Andrew Pavlo, Subramanya R. Dulloor, 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.
4. Justin DeBrabant, Joy Arulraj, Andrew Pavlo, Michael Stonebraker, Stanley B. Zdonik, Subramanya Dulloor.
[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 Storage Architectures, Hangzhou, China, 2014.

B.6 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.

B.7 Demonstrations

1. Arthita Ghosh, Arpit Narechania, Visweswara Sai Prashanth Dintyala, Su Timurturkan, Joy Arulraj, and Deven Bansod.
[Interactive Demonstration of SQLCHECK.](#)
Proc. VLDB Endow., 2021.

B.8 Preprints

1. Jiashen Cao, Rathijit Sen, Matteo Interlandi, Joy Arulraj, and Hyesoon Kim.
[Revisiting Query Performance in GPU Database Systems.](#)
CoRR 2023
2. Jaeho Bang, Pramod Chunduri, and Joy Arulraj.
[EKO: Adaptive Sampling of Compressed Video Data.](#)
CoRR 2021

3. Joy Arulraj, Ran Xian, Lin Ma, and Andy Pavlo.
Predictive Indexing.
CoRR 2019

C OTHER PUBLICATIONS AND CREATIVE PRODUCTS

C.1 Software Artifacts

1. **EVA** is a new AI-Relational Database Management System (DBMS) for efficiently and effectively analysing videos and other types of unstructured data at scale. There are several research projects based on this system:
 - EVA automatically materializes and reuses the results of expensive user-defined functions (UDFs) to facilitate faster exploratory data analysis. It differs from the state-of-the-art reuse algorithms in traditional DBMSs in three ways. First, it focuses on reusing the results of UDFs as opposed to those of sub-plans. Second, it takes a symbolic approach to analyze predicates and identify the degree of overlap between queries. Third, it factors reuse into UDF evaluation cost and uses the updated cost function in critical query optimization decisions like predicate reordering and model selection (results presented in SIGMOD'22).
 - EVA enables automated analysis of videos at scale using computationally-intensive deep learning models. To reduce the computational overhead of these models, it uses an ensemble of models that support a range of throughput-accuracy tradeoffs. Second, it adopts a fine-grained approach to query optimization. It processes different chunks of the video using different models in the given ensemble to meet the user's accuracy requirement. Lastly, it uses a lightweight technique to prune the model ensemble to lower the query optimization time. We empirically show that these techniques enable EVA to outperform the state-of-the-art video analytics systems (results presented in SIGMOD'22).
 - A significant challenge that video database systems encounter lies in the drift in real-world visual data. For instance, a model for self-driving vehicles that is not trained on images containing snow does not work well when it encounters them in practice. This drift phenomenon limits the accuracy of models employed for visual data analytics. EVA automatically detects and recovers from drift. It uses adversarial autoencoders to learn the distribution of high-dimensional images. It leverages unsupervised algorithm for detecting drift by comparing the distributions of the given data against that of previously seen data. When it detects drift, it invokes a drift recovery algorithm to deploy specialized models tailored towards the novel data points. These specialized models outperform their non-specialized counterpart on accuracy, performance, and memory footprint (results presented in VLDB'21).
 - Detection and localization of actions in videos is an important problem in practice. State-of-the-art video analytics systems are unable to efficiently and effectively answer action queries, because actions are often rare events, spread across a sequence of frames; detecting and localizing them often require computationally expensive deep neural networks. It

is also important to consider the entire sequence of frames to effectively answer the query. EVA uses a novel technique for efficiently answering these queries using deep reinforcement learning. It trains a reinforcement learning agent that learns to adaptively modify the input video segments that are subsequently sent to an action classification network. The agent alters the input segments along three dimensions - sampling rate, segment length, and resolution. To meet the accuracy target, the query optimizer trains the agent based on an accuracy-aware, aggregate reward function (results presented in SIGMOD'22).

2. **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 satisfiability modulo theory (SMT) solver to efficiently verify their equivalence.
3. **SQLCHECK** is a tool for assisting database application developers by automatically detecting common anti-patterns in SQL queries. It leverages query and data analysis algorithms to detect anti-patterns with high precision and recall. It suggests fixes for high-impact anti-patterns using rule-based query refactoring techniques.
4. **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 assists database developers in quickly pinpointing the root cause of performance regressions.
5. **PELTON** is a new DBMS designed for emergent non-volatile memory (NVM) technologies. As part of my doctoral dissertation, I 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 $100\times$ 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.
6. **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.
 7. **POSTGRESQL-CPP** is a port of the PostgreSQL DBMS to the C++ language. It spawned a subsequent official porting effort by PostgreSQL developers.

D PRESENTATIONS (SELECTED)

D.1 Academic Talks

- **EVA: An End-to-End Data System for Querying Videos At Scale**
Workshop on Video Analytics, May 2022, Stanford, CA
ICDE, May 2022, Virtual
DaMoN, June 2022, Philadelphia, PA
CIDR, January 2022, Virtual
- **Buffer Management on Non-Volatile Memory**
Imperial College, November 2020, Remote
- **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

D.2 Industry Talks

1. **EVA: An End-to-End Data System for Querying Videos At Scale**
DARPA, January 2023, Virtual
Cisco, March 2022, Virtual
Adobe Research, July 2021, Virtual
2. **Buffer Management on Non-Volatile Memory**
Samsung Research, January 2021, Remote
Oracle Labs, October 2020, Remote
3. **Write-Behind Logging**
SAP Labs, Sep 2017, Walldorf, Germany
4. **BzTree: An NVM-Aware Latch-Free B+Tree**
Microsoft Research, Dec 2016, Redmond, WA
5. **What Non-Volatile Memory Means for the Future Of Database Systems**
VMware Research, Sep 2016, Palo Alto, CA
6. **Thus Spoke Arulraj: Non-Volatile Memory for Databases**
Intel ISTC Big Data Retreat, Aug 2016, Hillsboro, OR
7. **Bridging the Archipelago between Row-Stores and Column-Stores for Hybrid Workloads**
Oracle Labs, April 2016, Redwood City, CA
8. **The Case for Flexible Storage Model**
Huawei, April 2016, Santa Clara, CA
9. **Rethinking Database Systems for Next-Generation Memory Technologies**
Samsung Research, March 2016, Mountain View, CA
10. **Let's Talk About Non-Volatile Memory Database Systems**
IBM Research, September 2015, San Jose, CA
EMC Flash Forum, September 2014, Hopkinton MA

E OTHER SCHOLARLY AND CREATIVE ACCOMPLISHMENTS

F SOCIETAL AND POLICY IMPACTS

F.1 Research Coverage in the News

- EVA AI-Relational Database System
[Hacker News](#), May 2023
- 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: Implications of NVM on Database Management Systems
[ACM Queue](#), July 2016

V TEACHING

A COURSES TAUGHT

Semester	Course #	Course Title
Spring 2023	8803 ASI	Database System Implementation (Part II)
Fall 2022	4420/6422	Database System Implementation (Part I)
Spring 2022	8803 ASI	Database System Implementation (Part II)
Fall 2021	4420/6422	Database System Implementation (Part I)
Spring 2021	8803 ASI	Database System Implementation (Part II)
Fall 2020	4420/6422	Database System Implementation (Part I)
Fall 2019	8803 DDL	Data Analytics using Deep Learning
Spring 2019	4420/6422	Database System Implementation
Fall 2018	8803 DDL	Data Analytics using Deep Learning

B INDIVIDUAL STUDENT GUIDANCE

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

- **Pramod Chunduri**
Fall 2018-
Status: Post-Qualifier
Topic: *Enabling Action and Object Reidentification Queries in Video Database Systems*
- **Xinyu Liu**
Fall 2019-
Status: Post-Qualifier (co-advised with Alex Orso)
Topic: *Inferring Properties of User-Defined Functions in Database Applications*
- **Jiashen Cao**
Fall 2020-
Status: Post-Qualifier (co-advised with Hyesoon Kim)
Topic: *Adaptive Query Processing in Video Database Systems*

- **Gaurav Kakkar**
Fall 2021-
Status: Post-Qualifier
Topic: *Model-Aware Storage Management in Video Database Systems*
- **Sayan Sinha**
Fall 2023-
Status: Pre-Qualifier (co-advised with Aditya Prakash)
Topic: *Improving Accuracy of Time-Series Forecasting*

B.2 Ph.D. Students (Graduated)

- **Jinho Jung**
Fall 2018-Spring 2021
Status: Graduated (co-advised with Taesoo Kim)
Topic: **Practical Systems for Strengthening and Weakening Binary Analysis Frameworks**
- **Qi Zhou**
Fall 2019-Fall 2020
Status: Graduated (co-advised with Bill Harris)
Topic: **Symbolic Reasoning for Query Verification and Optimization**

B.3 M.S. Students (Special Problems, Project, and Thesis)

- 2022: Karan Sarkar (First employment: AWS Redshift), Divyanshu Goyal (First employment: Adobe), Jaeho Bang (First employment: Rebellion), Sylesh Suresh (First employment: Meta), Varsha Ramesh (First employment: Google), Devshree Bharatia (First employment: Walmart), Aubhro Sengupta (First employment: LineaPy)
- 2020: Prashanth Dintyala (First employment: Nvidia), Xi Cheng, Ramesh Ravi, Akhilesh Siddhanti, Swati Ghosh, Jiashen Cao (First employment: PhD student at Georgia Tech), Arthita Ghosh (First employment: Salesforce), Gaurav Kakkar (Project Option), Nitin Chauhan (Project Option), Vinod Kumar (Project Option), Eric Gastineau (Thesis Option)
- 2019: Sonia Matthew (First employment: Amazon), Arpit Narechania (First employment: PhD student at Georgia Tech), Apaar Shanker, Sarmistha Dutta, Sanjana Garg, Dean Zhu, Jacob Logas (First employment: PhD student at Georgia Tech), Shreya Varshini, Arindum Roy

B.4 Undergraduate Students

- 2020: Albert Chen, Ashwin Dubey, Jeremy Hua, Luke Kim, Sylesh Suresh, Vivian De Sa Thiebaut, Abdullah Shah, Yukun Zheng, Neel Raja

B.5 Service on Thesis or Dissertation Committees

- 2020: Pradeep Fernando (Advisor: Ada Gavrilovska), Mehmet Emre Gursoy (Advisor: Ling Liu), Sanidhya Kashyap (Advisor: Taesoo Kim)

B.6 PhD/MS Students (Informal or Past advising)

- 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)

C EDUCATIONAL INNOVATIONS AND OTHER CONTRIBUTIONS

Developing a new two-part series of courses on implementing database systems:

- **Database System Implementation (CS 4420/6422)**
- **Advanced Database System Implementation (CS 8803)**

These courses contain significantly updated course materials: lectures, programming assignments, exercise sheets, and exams. Received the **Class of 1969 Teaching Fellowship** for further strengthening this course.

All of the programming assignments are based on the **BuzzDB academic DBMS**. The goal of these assignments is to familiarize the students with the internals of database management systems and to give them immediate feedback on their solutions.

The courses are based on similar programming-centric courses offered at CMU, Technische Universität München, and MIT. The key difference is that as the topics are covered across two semesters (instead of one semester), the students implement the entire database system stack in the programming assignments. There is a significant emphasis placed on correctness and performance of the solutions. BuzzDB leverages the Google Testing Library and the Valgrind memory error detector to accomplish these goals.

- **Course 1: Database System Implementation**
 - Labs: C++ Tutorial, Buffer Manager, B+-Tree, Query Execution
- **Course 2: Advanced Database System Implementation**
 - Labs: External Sort, Logging And Recovery, Concurrency Control, Query Optimizer

VI SERVICE

A PROFESSIONAL CONTRIBUTIONS

A.1 Program Committee Member

- SIGMOD 2023 (Publicity Chair)
- SIGMOD 2022 (Remote Participation Chair)
- SIGMOD 2020-21
- VLDB 2020-21
- CIDR 2019

- OSDI 2021 (External reviewer)

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 PUBLIC AND COMMUNITY SERVICE

C INSTITUTE CONTRIBUTIONS

- SCS Faculty Recruitment Committee, 2021, 2023
- SCS School Advisory Committee, 2020-21
- SCS Ph.D. Visit Day Committee, 2019