Immanuel Trummer

itrummer@cornell.edu
www.itrummer.org

Education

École Polytechnique Fédérale de Lausanne (EPFL) 2010-2016
PhD in Computer Science; Advisor: Christoph Koch.
EPFL Thesis Special Distinction.

University of Stuttgart & École Centrale de Nantes 2003-2010
Double Diploma in Computer Science and Engineering.
Graduated with Distinction.

Academic Appointments

Cornell University 2024-Present
Associate Professor, Cornell Bowers CIS.

Cornell University 2016-2024
Assistant Professor, Cornell Bowers CIS.

École Polytechnique Fédérale de Lausanne (EPFL) 2010-2016
Graduate Research Assistant, LIA/DATA Lab.

Publications

Books

- Immanuel Trummer. Data Analysis with LLMs: Text, Tables, Images and Sound. Manning Publications, 2025, pp. 1–232.

Refereed Conference and Journal Publications[1]

- Immanuel Trummer. Generating Highly Customizable Python Code for Data Processing with Large Language Models. In: *VLDBJ* 34.21 (2025), pp. 1–19.
- Victor Giannakouris and Immanuel Trummer. λ-Tune: Harnessing Large Language Models for Automated Database System Tuning. In: *SIGMOD* 3.1 (2025), pp. 1–26.
- Saehan Jo and Immanuel Trummer. SpareLLM: Automatically Selecting Task-Specific Minimum-Cost Large Language Models under Equivalence Constraint. In: *SIGMOD* (2025).
- Immanuel Trummer. Customizing Operator Implementations for SQL Processing via Large Language Models. In: *Data Engineering Bulletin* 49.1 (2025), pp. 45–56.
- Tharushi Jayasekara and Immanuel Trummer. Demonstrating CEDAR: A System for Cost-Efficient Data-Driven Claim Verification. In: *SIGMOD*. 2025, pp. 1–4.
- Jiale Lao and Immanuel Trummer. Demonstrating SQLBarber: Leveraging Large Language Models to Generate Customized and Realistic SQL Workloads. In: *SIGMOD*. 2025, pp. 1–4.
- Victor Giannakouris and Immanuel Trummer. SwellDB: Dynamic Query-Driven Table Generation with Large Language Models. In: *SIGMOD*. 2025, pp. 1–4.
- Immanuel Trummer. DB-BERT: Making Database Tuning Tools "Read" the Manual. In: *VLDB Journal* 33 (2024), pp. 1085–1104.
- Saehan Jo and Immanuel Trummer. ThalamusDB: Approximate Query Processing on Multi-Modal Data. In: *SIGMOD* 2.3 (2024), pp. 1–26.

[1] PVLDB publications are presented yearly at the VLDB conference.

- Ziyun Wei and Immanuel Trummer. ROME: Robust Query Optimization via Parallel Multi-Plan Execution. In: *SIGMOD* 2.3 (2024), pp. 1–25.
- Junxiong Wang, Kaiwen Wang, Yueying Li, Nathan Kallus, Immanuel Trummer, and Wen Sun. JoinGym: An Efficient Join Order Selection Environment. In: *RLJ* 1 (2024), pp. 64–91.
- Victor Giannakouris and Immanuel Trummer. DBG-PT: A Large Language Model Assisted Query Performance Regression Debugger. In: *PVLDB* 17.12 (2024), pp. 4337–4340.
- Immanuel Trummer. Generating Succinct Descriptions of Database Schemata for Cost-Efficient Prompting of Large Language Models. In: *PVLDB* 17.11 (2024), pp. 3511–3523.
- Immanuel Trummer. Large Language Models: Principles and Practice. In: *ICDE*. 2024, pp. 5354–5357.
- Victor Giannakouris and Immanuel Trummer. Demonstrating λ-Tune: Exploiting Large Language Models for Workload-Adaptive Database System Tuning. In: *SIGMOD*. 2024, pp. 508–511.
- Junxiong Wang, Immanuel Trummer, Ahmet Kara, and Dan Olteanu. ADOPT: Adaptively Optimizing Attribute Orders for Worst-Case Optimal Join Algorithms via Reinforcement Learning. In: *PVLDB* 16.11 (2023), pp. 2805–2817.
- Immanuel Trummer. Demonstrating GPT-DB: Generating Query-Specific and Customizable Code for SQL Processing with GPT-4. In: *PVLDB* 16.12 (2023), pp. 4098–4101.
- Junxiong Wang, Immanuel Trummer, Ahmet Kara, and Dan Olteanu. Demonstrating ADOPT: Adaptively Optimizing Attribute Orders for Worst-Case Optimal Joins via Reinforcement Learning. In: *PVLDB* 16.12 (2023), pp. 4094–4097.
- Simran Arora, Brandon Yang, Sabri Eyuboglu, Avanika Narayan, Andrew Hojel, Immanuel Trummer, and Christopher Re. Language Models Enable Simple Systems for Generating Structured Views of Heterogeneous Data Lakes. In: *PVLDB* 17.2 (2023), pp. 92–105.
- Immanuel Trummer. Can Large Language Models Predict Data Correlations from Column Names? In: *PVLDB* 16.13 (2023), pp. 4310–4323.
- Manuel Schoenberger, Immanuel Trummer, and Wolfgang Mauerer. Quantum-Inspired Digital Annealing for Join Ordering. In: *PVLDB* 17.3 (2023), pp. 511–524.
- Immanuel Trummer. Demonstrating NaturalMiner: Searching Large Data Sets for Abstract Patterns Described in Natural Language. In: *SIGMOD*. 2023, pp. 139–142.
- Immanuel Trummer. BABOONS: Black-Box Optimization of Data Summaries in Natural Language. In: *PVLDB* 15.11 (2022), pp. 2980–2993.
- Ziyun Wei and Immanuel Trummer. SkinnerMT: Parallelizing for Efficiency and Robustness in Adaptive Query Processing on Multicore Platforms. In: *PVLDB* 16.4 (2022), pp. 905–917.
- Immanuel Trummer. CodexDB: Synthesizing Code for Query Processing from Natural Language Instructions using GPT-3 Codex. In: *PVLDB* 15.11 (2022), pp. 2921–2928.
- Immanuel Trummer. From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management. In: *PVLDB* 15.12 (2022), pp. 3770–3773.
- Junxiong Wang, Debabrota Basu, and Immanuel Trummer. Procrastinated Tree Search: Black-box Optimization with Delayed, Noisy, and Multi-fidelity Feedback. In: *AAAI* 36.9 (2022), pp. 10381–10390.
- Immanuel Trummer. Towards NLP-Enhanced Data Profiling Tools. In: *CIDR*. 2022, pp. 1–1.
- Immanuel Trummer. DB-BERT: a Database Tuning Tool that "Reads the Manual". In: *SIGMOD*. 2022, pp. 190–203.
- Immanuel Trummer. Demonstrating DB-BERT: A Database Tuning Tool that "Reads" the Manual. In: *SIGMOD*. 2022, pp. 2437–2440.
- Immanuel Trummer. Database Tuning using Natural Language Processing. In: *SIGMOD Record* 50.3 (2021), pp. 27–28.
- Junxiong Wang, Immanuel Trummer, and Debabrota Basu. UDO: Universal Database Optimization using Reinforcement Learning. In: *PVLDB* 14.13 (2021), pp. 3402–3414.
- Ziyun Wei, Immanuel Trummer, and Connor Anderson. Robust Voice Querying with MUVE: Optimally Visualizing Results of Phonetically Similar Queries. In: *PVLDB* 14.11 (2021), pp. 2397–2409.
- Immanuel Trummer. The Case for NLP-Enhanced Database Tuning: Towards Tuning Tools that "Read the Manual". In: *PVLDB* 14.7 (2021), pp. 1159–1165.
- Immanuel Trummer. WebChecker: Towards an Infrastructure for Efficient Misinformation Detection at Web Scale. In: *IEEE Data Eng. Bull.* 44.3 (2021), pp. 66–77.

- Immanuel Trummer, Junxiong Wang, Ziyun Wei, Deepak Maram, Samuel Moseley, Saehan Jo, Joseph Antonakakis, and Ankush Rayabhari. SkinnerDB: Regret-bounded Query Evaluation via Reinforcement Learning. In: *ACM Transactions on Database Systems* 46.3 (2021), pp. 1–45.
- Immanuel Trummer and Anderson Connor. Optimally Summarizing Data by Small Fact Sets for Concise Answers to Voice Queries. In: *ICDE*. 2021, pp. 1715–1726.
- Junxiong Wang, Immanuel Trummer, and Debabrota Basu. Demonstrating UDO: A Unified Approach for Optimizing Transaction Code, Physical Design, and System Parameters via Reinforcement Learning. In: *SIGMOD*. 2021, pp. 2794–2797.
- Ziyun Wei, Immanuel Trummer, and Anderson Connor. Demonstrating Robust Voice Querying with MUVE: Optimally Visualizing Results of Phonetically Similar Queries. In: *SIGMOD*. 2021, pp. 2798–2802.
- Georgios Karagiannis, Mohammed Saeed, Paolo Papotti, and Immanuel Trummer. Scrutinizer: Fact Checking Statistical Claims. In: *PVLDB* 13.12 (2020), pp. 2965–2968.
- Georgios Karagiannis, Immanuel Trummer, Saehan Jo, Shubham Khandelwal, Xuezhi Wang, and Cong Yu. Mining an "Anti-Knowledge Base" from Wikipedia Updates with Applications to Fact Checking and Beyond. In: *PVLDB* 13.4 (2020), pp. 561–573.
- Immanuel Trummer. Demonstrating the Voice-Based Exploration of Large Data Sets with CiceroDB-Zero. In: *PVLDB* 13.12 (2020), pp. 2869–2872.
- Georgios Karagiannis, Mohammed Saeed, Paolo Papotti, and Immanuel Trummer. Scrutinizer: A Mixed-Initiative Approach to Large-Scale, Data-Driven Claim Verification. In: *PVLDB* 13.12 (2020), pp. 2508–2521.
- Saehan Jo, Jialing Pei, and Immanuel Trummer. Demonstration of ScroogeDB: Getting More Bang for the Buck with Deterministic Approximation in the Cloud. In: *PVLDB* 13.12 (2020), pp. 2961–2964.
- Georgios Karagiannis, Mohammed Saeed, Paolo Papotti, and Immanuel Trummer. Scrutinizer: a System for Checking Statistical Claims. In: *BDA*. 2020, pp. 72–73.
- Saehan Jo and Immanuel Trummer. BitGourmet: Deterministic Approximation via Optimized Bit Selections. In: *CIDR*. 2020, pp. 1–7.
- Saehan Jo and Immanuel Trummer. Demonstration of BitGourmet: Data Analysis via Deterministic Approximation. In: *SIGMOD*. 2020, pp. 2801–2804.
- Saehan Jo, Immanuel Trummer, Weicheng Yu, Xuezhi Wang, Cong Yu, Daniel Liu, and Niyati Mehta. AggChecker: a Fact-Checking System for Text Summaries of Relational Data Sets. In: *PVLDB* 12.12 (2019), pp. 1938–1941.
- Immanuel Trummer. Data Vocalization with CiceroDB. In: *CIDR*. 2019, pp. 1–8.
- Saehan Jo, Immanuel Trummer, Weicheng Yu, Xuezhi Wang, Cong Yu, Daniel Liu, and Niyati Mehta. Verifying Text Summaries of Relational Data Sets. In: *SIGMOD*. 2019, pp. 299–316.
- Immanuel Trummer, Yicheng Wang, and Saketh Mahankali. A Holistic Approach for Query Evaluation and Result Vocalization in Voice-Based OLAP. In: *SIGMOD*. 2019, pp. 936–953.
- Immanuel Trummer, Junxiong Wang, Deepak Maram, Samuel Moseley, Saehan Jo, and Joseph Antonakakis. SkinnerDB: Regret-Bounded Query Evaluation via Reinforcement Learning. In: *SIGMOD*. 2019, pp. 1039–1050.
- Immanuel Trummer. Exact Cardinality Query Optimization with Bounded Execution Cost. In: *SIGMOD*. 2019, pp. 2–17.
- Immanuel Trummer, Samuel J Mosley, Joseph Antonakakis, and Saehan Jo. SkinnerDB: Regret-Bounded Query Evaluation via Reinforcement Learning. In: *PVLDB* 11.12 (2018), pp. 2074–2077.
- Immanuel Trummer, Mark Bryan, and Ramya Narasimha. Vocalizing Large Time Series Efficiently. In: *PVLDB* 11.11 (2018), pp. 1563–1575.
- Immanuel Trummer and Christoph Koch. Multi-Objective Parametric Query Optimization. In: *VLDB Journal* 26.1 (2017), pp. 107–124.
- Immanuel Trummer, Jiancheng Zhu, and Mark Bryan. Data Vocalization: Optimizing Voice Output of Relational Data. In: *PVLDB* 10.11 (2017), pp. 1574–1585.
- Immanuel Trummer and Christoph Koch. Multi-Objective Parametric Query Optimization. In: *Communications of the ACM* 60.10 (2017), pp. 81–89.
- Immanuel Trummer and Christoph Koch. Solving the Join Ordering Problem via Mixed Integer Linear Programming. In: *SIGMOD*. 2017, pp. 1025–1040.
- Immanuel Trummer and Christoph Koch. Parallelizing Query Optimization on Shared-Nothing Architectures. In: *PVLDB* 9.9 (2016), pp. 660–671.

- Immanuel Trummer and Christoph Koch. Multiple Query Optimization on the D-Wave 2X Adiabatic Quantum Computer. In: *PVLDB* 9.9 (2016), pp. 648–659.
- Immanuel Trummer and Christoph Koch. Multi-Objective Parametric Query Optimization. In: *ACM SIGMOD Research Highlights* 45.1 (2016), pp. 24–31.
- Immanuel Trummer and Christoph Koch. A Fast Randomized Algorithm for Multi-Objective Query Optimization. In: *SIGMOD*. 2016, pp. 1737–1752.
- Immanuel Trummer, Alon Halevy, Hongrae Lee, Sunita Sarawagi, and Rahul Gupta. Mining Subjective Properties on the Web. In: *SIGMOD*. 2015, pp. 1745–1760.
- Immanuel Trummer and Christoph Koch. An Incremental Anytime Algorithm for Multi-Objective Query Optimization. In: *SIGMOD*. 2015, pp. 1941–1953.
- Immanuel Trummer and Christoph Koch. Multi-Objective Parametric Query Optimization. In: *PVLDB* 8.3 (2014), pp. 221–232.
- Immanuel Trummer and Christoph Koch. Approximation Schemes for Many-Objective Query Optimization. In: *SIGMOD*. 2014, pp. 1299–1310.
- Immanuel Trummer, Boi Faltings, and Walter Binder. Multi-Objective Quality-Driven Service Selection - A Fully Polynomial Time Approximation Scheme. In: *IEEE Transactions on Software Engineering* 40.2 (2013), pp. 167–191.
- Mehdi Riahi, Thanasis Papaioannou, Immanuel Trummer, and Karl Aberer. Utility-Driven Data Acquisition in Participatory Sensing. In: *EDBT*. 2013, pp. 251–262.
- Immanuel Trummer and Boi Faltings. Dynamically Selecting Composition Algorithms for Economical Composition as a Service. In: *ICSOC*. 2011, pp. 513–522.
- Immanuel Trummer and Boi Faltings. Optimizing the Tradeoff between Discovery, Composition, and Execution Costs in Service Composition. In: *ICWS*. 2011, pp. 476–483.
- Walter Binder, Daniele Bonetta, Cesare Pautasso, Achille Peternier, Diego Milano, Heiko Schuldt, Nenad Stojnic, Boi Faltings, and Immanuel Trummer. Towards Self-Organizing Service-Oriented Architectures. In: *IEEE World Congress on Services*. 2011, pp. 115–121.
- Immanuel Trummer, Frank Leymann, Ralph Mietzner, and Walter Binder. Cost-Optimal Outsourcing of Applications into the Clouds. In: *CloudCom*. 2010, pp. 135–142.

Refereed Workshop Publications

- Immanuel Trummer and Davide Venturelli. Leveraging Quantum Computing for Database Index Selection. In: *Q-Data*. 2024, pp. 14–26.
- Immanuel Trummer. Towards Out-of-Core Simulators for Quantum Computing. In: *Q-Data*. 2024, pp. 1–1.
- Ibrahim Sabek, Immanuel Trummer, and Stefan Prestel. First Workshop on Quantum Computing and Quantum-Inspired Technology for Data-Intensive Systems and Applications (Q-Data). In: *SIGMOD*. 2024, pp. 663–664.
- Manuel Schoenberger, Immanuel Trummer, and Wolfgang Mauerer. Quantum Optimisation of General Join Trees. In: *QDSM*. 2023, pp. 1–12.
- Victor Giannakouris and Immanuel Trummer. Building Learned Federated Query Optimizers. In: *VLDB PhD Workshop*. 2022, pp. 1–5.

Patents

- Immanuel Trummer. Verifying Text Summaries of Relational Data Sets (US Patent Number 11113275). 2021.
- Davide Venturelli and Immanuel Trummer. System and Method to Hardcode Integer Linear Optimization Problems on Physical Implementations of the Ising Model (US Patent Number 10691771). 2020.

Theses

- Immanuel Trummer. From Massive Parallelization to Quantum Computing: Seven Novel Approaches to Query Optimization. PhD thesis. EPFL, 2016, pp. 1–253.
- Immanuel Trummer. Cost-Optimal Provisioning of Cloud Applications. Diploma Thesis. University of Stuttgart, 2010, pp. 1–98.

Awards and Honors

- Best of VLDB Paper, 2024.
- SIGMOD Best Demo Runner Up Award, 2023.
- NSF CAREER Award, 2022.
- BDA Best Demo Award, 2020.
- Best of SIGMOD Paper (TODS), 2019.
- Google Faculty Research Award, 2018.
- Jim Gray Award, Honorable Mention, 2017.
- EPFL PhD Thesis Special Distinction, 2017.
- Google Faculty Research Award, 2016.
- ACM SIGMOD Research Highlight, 2015.
- Best of VLDB Paper, 2015.
- Google PhD Fellowship, 2015.
- EPFL Teaching Assistant Award, 2015.
- First Graduation Prize, 2010.
- German National Academic Foundation, 2008.
- Scholarship for Academic Excellence, 2007.
- Scholarship for TIME Double Degree, 2005.

Funded Proposals

Internal

- CIDA RIF Seed Grant: "Co-Creating a Human-Machine Interface Better Adapted for On-Farm Data Recording, Curation, Management, and Use" (2021-2023). PI: Louis Longchamp. Co-PIs: Wendy Ju, Immanuel Trummer, Diane Bailey, Mike Stanyard, Erik Smith, Michael Hunter. Total Amount: $150,000.
- Huawei Research Initiative: "Deterministic Approximation via BitGourmet" (2019). PI: Immanuel Trummer. Total Amount: $116,831.

External

- NSF 2239326: "CAREER: Mining Hints from Text Documents to Guide Automated Database Performance Tuning" (2023-2028). PI: Immanuel Trummer. Total Amount: $594,874.
- NSF 1910830: "III: Small: Regret-Bounded Query Evaluation via Reinforcement Learning" (2019-2022). PI: Immanuel Trummer. Total Amount: $499,999.
- Google AI-Powered Immediate Response to Pandemics Initiative: "Data-driven Automatic Fact Checking of Coronavirus Claims" (2020). PIs: Immanuel Trummer, Paolo Papotti (Co-PI). Total Amount: $120,000.
- Fact-Checking Development Grant (IFCN & YouTube): "Automatic Fact-Checking of Coronavirus Claims" (2020). PIs: Paolo Papotti, Immanuel Trummer (Co-PI). Total Amount: $49,800.
- Google Faculty Research Award: "Mining an 'Anti-Knowledge Base' for Fact Checking from Wikipedia Updates" (2018). PI: Immanuel Trummer. Total Amount: $65,235.
- Google Faculty Research Award: "Optimizing Voice-Based Output of Relational Data" (2016). PI: Immanuel Trummer. Total Amount: $57,407.

Talks

Tutorials

- 5/14/24: Tutorial on Large Language Models: Principles and Practice (90 Minutes), ICDE 2024.

- 3/8/23: From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management (180 Minutes), BTW 2023.
- 9/8/22: From BERT to GPT-3 Codex: Harnessing the Potential of Very Large Language Models for Data Management (90 Minutes), VLDB 2022.

Panels

- 6/14/2024: Foundation Models for and in Databases,
  SIGMOD 2024 Workshop on Exploiting Artificial Intelligence Techniques for Data Management.
- 5/13/2024: Knowledge Graphs and LLMs: A Relationship under Investigation,
  ICDE 2024 Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores Graph Edition.
- 6/18/23: Data Management Challenges in LLM-Powered Solutions,
  SIGMOD 2023 Workshop on Data Management for End-to-End Machine Learning.
- 4/11/22: Futuristic Data Interfaces - Voice Interfaces for Data Access,
  Conference on Database Systems for Advanced Applications 2022.
- 10/3/19: Anticipating Societal Impact, Cornell CIS Celebration 2019.

Invited Talks

- 3/4/2025: CheaPT: Using Language Models Without Breaking the Bank, Invited talk at ML4Sys and Sys4ML.
- 11/12/2024: Novel Applications for Large Language Models in Data Management, **Keynote** at KBC-LM.
- 6/14/2024: Novel Applications for Large Language Models in Data Management, **Keynote** at aiDM.
- 5/23/24: Novel Applications for Large Language Models in Data Management, **Keynote** at NEDB.
- 4/4/24: Novel Use Cases for Large Language Models in Data Management, UIUC (Virtual).
- 12/15/23: Next-Generation Data Management with Large Language Models, Invited talk at TRL.
- 9/1/23: Applications for Large Language Models in Data Management, **Keynote** at AIDB.
- 9/1/23: Playing Dice with the Universe: a Brief History of Quantum Computing and Potential Use Cases in Data Management, **Keynote** at QDSM.
- 7/14/23: Novel Use Cases for Large Language Models in Data Management, TUM.
- 7/13/23: Novel Use Cases for Large Language Models in Data Management, MPI.
- 7/11/23: Towards Highly Customizable Database Systems via Generative AI, University of Stuttgart.
- 7/7/23: Novel Use Cases for Large Language Models in Data Management, TU Berlin.
- 7/5/23: Next-Generation Data Management via Large Language Models, **Keynote** at ScaDS.AI.
- 4/18/23: Towards Highly Customizable Database Systems via Generative AI, Boston University.
- 4/19/23: Towards Highly Customizable Database Systems via Generative AI, MIT.
- 4/18/23: Towards Highly Customizable Database Systems via Generative AI, Harvard.
- 4/7/23: Towards Tuning Tools that "Read" the Manual, University of Buffalo.
- 4/3/23: Towards AI-Generated Database Management Systems, **Keynote** at DBML.
- 3/7/23: Towards AI-Generated Database Management Systems,
  BTW 2023 Tutorial Workshop on ML for Systems and Systems for ML.
- 2/25/23: Leveraging Generative AI for Data Processing, CWI (Virtual).
- 12/14/22: Towards Tuning Tools that "Read" the Manual, TU Darmstadt (Virtual).
- 12/1/22: Towards Tuning Tools that "Read" the Manual, NYU.
- 11/3/22: Towards Tuning Tools that "Read" the Manual, OSU (Virtual).
- 8/15/22: Novel Applications for Language Models in Data Management, Stanford DB Group.
- 8/3/22: Intra-Query Learning for Worst-Case Optimal Join Algorithms, FDB Workshop (Virtual).
- 3/11/22: Mining Hints for Database Performance Tuning from the Web, Aarhus University (Virtual).
- 7/21/22: Towards Tuning Tools that "Read" the Manual, RelationalAI (Virtual).
- 11/17/21: Towards Tuning Tools that "Read" the Manual, UC Berkeley (Virtual).
- 2/24/21: SkinnerDB and the Case for Intra-Query Learning, UCSD (Virtual).
- 8/27/20: SkinnerDB and the Case for Intra-Query Learning, Snowflake (Virtual).

- 8/20/20: SkinnerDB and the Case for Intra-Query Learning, Google (Virtual).
- 8/11/20: SkinnerDB and the Case for Intra-Query Learning, Oracle (Virtual).
- 7/31/20: SkinnerDB and the Case for Intra-Query Learning, MSR (Virtual).
- 6/30/20: Recent Advances in Automated Fact-Checking, Global Fact 7 (Virtual).
- 3/5/20: SkinnerDB: the Case for Intra-Query Learning,
  German Research Foundation, Special Interest Group on Databases.
- 3/3/20: Exploring Data Using Your Voice, NYC Open Data Week.
- 10/14/16: Query Optimization for Data Analysis, UW.
- 4/14/16: Query Optimization for Data Science, UIUC.
- 3/30/16: Query Optimization for Data Science, UCSD.
- 3/22/16: Query Optimization for Data Science, Rice University.
- 3/14/16: Query Optimization for Data Science, ETH Zürich.
- 3/8/16: Query Optimization for Data Science, Cornell.
- 3/3/16: Query Optimization for Data Science, UMD.
- 2/25/16: Query Optimization for Data Science, University of Edinburgh.
- 2/22/16: Query Optimization for Data Science, University of Chicago.

Teaching

At Cornell

- Fall'24: Co-taught Introduction to Database Systems (CS 4320, 5320). 417 Students Enrolled.
- Fall'24: Co-taught Practicum in Database Systems (CS 4321, 5321). 49 Students Enrolled.
- Fall'24: Co-taught Seminar in Database Systems (CS 7390). 19 Students Enrolled.
- Spring'24: Advanced Database Systems (CS 6320). 33 Students Enrolled.
- Spring'24: Seminar in Database Systems (CS 7390). 6 Students Enrolled.
- Fall'23: Introduction to Database Systems (CS 4320, 5320). 368 Students Enrolled.
- Fall'23: Practicum in Database Systems (CS 4321, 5321). 25 Students Enrolled.
- Fall'23: Seminar in Database Systems (CS 7390). 7 Students Enrolled.
- Spring'23: Advanced Database Systems (CS 6320). 24 Students Enrolled.
- Spring'23: Seminar in Database Systems (CS 7390). 14 Students Enrolled.
- Fall'22: Introduction to Database Systems (CS 4320, 5320). 346 Students Enrolled.
- Fall'22: Practicum in Database Systems (CS 4321, 5321). 51 Students Enrolled.
- Fall'22: Seminar in Database Systems (CS 7390). 11 Students Enrolled.
- Spring'22: Seminar in Database Systems (CS 7390). 5 Students Enrolled.
- Fall'21: Introduction to Database Systems (CS 4320, 5320). 258 Students Enrolled.
- Fall'21: Seminar in Database Systems (CS 7390): 3 Students Enrolled.
- Spring'21: Advanced Database Systems (CS 6320). 13 Students Enrolled.
- Fall'20: Introduction to Database Systems (CS 4320, 5320). 279 Students Enrolled.
- Fall'20: Practicum in Database Systems (CS 4321, 5321). 33 Students Enrolled.
- Fall'20: Seminar in Database Systems (CS 7390). 7 Students Enrolled.
- Spring'20: Advanced Database Systems (CS 6320). 11 Students Enrolled.
- Fall'19: Introduction to Database Systems (CS 4320, 5320). 276 Students Enrolled.
- Fall'19: Practicum in Database Systems (CS 4321, 5321). 32 Students Enrolled.
- Fall'19: Seminar in Database Systems (CS 7390). 6 Students Enrolled.
- Spring'19: Advanced Database Systems (CS 6320). 14 Students Enrolled.
- Fall'18: Introduction to Database Systems (CS 4320, 5320). 266 Students Enrolled.
- Fall'18: Practicum in Database Systems (CS 4321, 5321). 49 Students Enrolled.
- Fall'18: Seminar in Database Systems (CS 7390). 6 Students Enrolled.
- Spring'18: Advanced Database Systems (CS 6320). 10 Students Enrolled.

- Fall'17: Introduction to Database Systems (CS 4320, 5320). 184 Students Enrolled.
- Fall'17: Practicum in Database Systems (CS 4321, 5321). 26 Students Enrolled.
- Fall'17: Seminar in Database Systems (CS 7390). 5 Students Enrolled.
- Spring'17: Introduction to Database Systems (CS 4320, 5320). 168 Students Enrolled.
- Spring'17: Practicum in Database Systems (CS 4321, 5321). 26 Students Enrolled.
- Fall'16: Advanced Database Systems (CS 6320). 5 Students Enrolled.

External

- Introduction to Database Systems. Online lecture introducing students to topics in database systems with a total duration of 25 hours. The lecture received over **one million** views.
- The Cornell, Maryland, Max Planck Pre-doctoral Research School 2020. I gave a virtual lecture on "Learning to Process Data Faster via Reinforcement Learning".

Mentoring

Current PhD Students

- Jiale Lao (2024-Present).
- Tharushi Jayasekara (2024-Present).
- Victor Giannakouris (2021-Present). Passed A-exam on 12/6/24.

Graduated PhD Students

- Junxiong Wang (2019-2024). First position: Together AI.
- Saehan Jo (2018-2025). First position: Amazon.
- Ziyun Wei (2018-2024). First position: ByteDance.

Awards by Mentored Students

- CRA Outstanding Undergraduate Research Award 2018, Honorable Mention. Mark Bryan.
- NSF VLDB 2018 Travel Grant. Samuel Moseley.
- JP Morgan Award 2018. Mark Bryan, Ramya Narasimha.
- Lockheed Martin Award 2017. Mark Bryan, Jiancheng Zhu.

Master of Science Theses

- Ankush Rayabhari. "Adaptive Join Execution in Compilation-Based Execution Engines of Databases". 2022.

Student Organizations

- Cornell Data Engineering Team, Cornell Data Science (2018-Present).

Master of Engineering Projects

- Ashutosh Agarwal (Spring 2017).
- Niveditha Shenoy Badiadka (Spring 2018).
- Janice Chan (Spring 2020).
- Elva Gao (Fall 2022).
- Shubham Khandelwal (Spring 2019).
- Haram Kim (Fall 2019).
- Daniel Liu (Spring 2017).
- Rahul Madanahalli (Spring 2018).
- Jenny Mallette (Fall 2016).

- Niyati Mehta (Spring 2017).
- Samuel Moseley (Fall 2018).
- Ramya Narasimha (Fall 2017).
- Maarij Tahir (Fall 2016).
- Rong Tan (Spring 2020).
- Sanjana Thirumalai (Fall 2016).
- Zhaoyu Wang (Spring 2023).
- Weicheng Yu (Fall 2016).
- Jiancheng Zhu (Fall 2016).

Service

Internal

- Reviewer for Cisco Strategic Initiative (2023).
- Reviewer for CIDA Research Innovation Fund (2022).
- PhD Admissions Committee (2016, 2017, 2020, 2021, 2022).
- Recruitment Committee, Ad-Hoc or Full Member (2016-Present).

External

- Co-Founder and Co-Chair of Q-DATA Workshop (2024, 2025).
- Associate Editor for SIGMOD (2024).
- Co-Chair of VLDB Demo Track (2022).
- Associate Editor of SIGMOD Record (2019-Present).
- Associate Editor of VLDB (2026).
- Reviewer for IEEE TKDE (2022).
- Panelist for the National Science Foundation (2021, 2023).
- Jim Gray Dissertation Award Committee (2021).
- Reviewer for ACM TODS (2020).
- PC Member for VLDB (2018-Present).
- PC Member for SIGMOD (2018, 2019, 2020, 2026).
- PC Member for SoCC (2017).
- Juror for SIGMOD Undergrad Competition (2016).