

Zheguang Zhao

Senior Software Engineer
NetApp Inc.
Level 1, 550 Bourke Street
Melbourne, VIC 3000
Australia

Phone: +61-433-440-582
Email: zheguang.zhao@alumni.brown.edu
Homepage: zheguang.github.io
LinkedIn: www.linkedin.com/in/zheguang
Google Scholar: [goo.gl/DR8pSa](https://scholar.google.com/citations?user=go0gl/DR8pSa)

Education

Ph.D. in Computer Science, Brown University, Rhode Island, 2022

Advisors: Stan Zdonik, Seny Kamara

Readers: George Kollios, Moti Yung

Thesis: *Building a Structurally-Encrypted Relational Database*

Software: [KafeDB](#)

M.Sc. in Computer Science, Brown University, Rhode Island, 2016

Advisor: Stan Zdonik

Readers: Tim Kraska, Ugur Cetintemel

Thesis: *Approximate Data Structures for Visualization*

B.Sc. in Computer Science, University of Wisconsin at Madison, Wisconsin, 2012

Advisor: Jignesh Patel, Ben Liblit

Experiences

NetApp Inc., Cloud Storage, *Senior Software Engineer* (Current)

Full-time Apache Kafka Contributor; Kafka Connector for NetApp ONTAP storage

Developed multi-cloud management platform; automated OS upgrade for 1000+ nodes

The University of Melbourne, *Research Fellow in Secure Systems* (2023)

Security in adversarial ML and human-AI collaboration. Hosts: Prof. Olya Ohrimenko and Prof. Ben Rubinstein

Technical University of Darmstadt, *Postdoc Researcher* (2021-2022)

Secure federated learning and databases with Intel SGX. Host: Prof. Carsten Binnig

Aroki Systems (now part of MongoDB), *Database Scientist* (2017-2021)

KafeDB, an end-to-end encrypted relational database based on SparkSQL

Los Alamos National Laboratory, *Research Intern* (2019)

ML model reconstruction of mixed dynamics in cyber-physical systems, with application to network verification and security

Microsoft AI & Research, *Research Intern* (2017)

Constraint learning. Host: DMX Group

Intel Labs, *Research Intern* (2015)

Efficiency of machine learning algorithms in Apache Spark
In-memory transactional processing using non-volatile memory

Hadapt (now part of Teradata), *Software Engineer* (2013-2014)

Enterprise SQL-on-Hadoop system including query execution, storage engine, high availability and analytics

Kosmix (now part of @WalmartLabs), *Software Engineer* (2012-2013)

In-memory distributed queue system for in-house stream processing

Great Lakes Bioenergy Research Center, *Software Engineer* (2009-2012)

Scientific database for biological enzyme research

Open Source

Apache Kafka Contributor

KafeDB: End-to-End Structurally-Encrypted Relational Database [\[link\]](#)

Approximate Computation using GPGPU [\[simulator\]](#)[\[benchmark\]](#)

ML framework for Cyber-physical Systems [\[link\]](#)

Encrypted Searchable Signal [\[link\]](#)

Macau: statistical hypothesis testing based on resampling [\[link\]](#)

Kafka Connector for AWS, NetApp S3 Storage [\[link\]](#)

Machine learning algorithms in Spark [\[link\]](#)

Consistency control for machine learning algorithms [\[link\]](#)

R-tree in Rust [\[link\]](#)

Spark performance analysis tool [\[link\]](#)

VoltDB on non-volatile memory [\[link\]](#)

Publications

Benchmarking the Second Generation of Intel SGX Hardware
Proceedings of the 18th International Workshop on Data Management on New Hardware

Towards Decentralized Parameter Servers for Secure Federated Learning DATA, 2022

ACID-V: towards a New Class of DBMSs for Data Sharing
Polystores Workshop at VLDB, 2021

Encrypted Databases: from Theories to Systems
CIDR, January 2021

Dynamic Query Refinement for Interactive Data Exploration
EDBT/ICDT Joint Conference, March 2020

Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis
CHI Conference, April 2018

Controlling False Discoveries During Interactive Data Exploration
SIGMOD Conference, May 2017

Safe Visual Data Exploration
SIGMOD Conference, Demo, May 2017

Bridging the Gap between HPC and Big Data frameworks
VLDB Journal, 2017

Towards Sustainable Insights
CIDR Conference, January 2017

Towards a Benchmark for Interactive Data Exploration
IEEE Data Engineering Bulletin, 2016.

Larger-than-memory Data Management on Modern Storage Hardware for In-memory OLTP Database Systems
SIGMOD DaMoN Workshop, June 2016

VisTrees: Fast Indexes for Interactive Data Exploration
SIGMOD HILDA Workshop, June 2016

Data Tiering in Heterogeneous Memory Systems
EuroSys Conference, April 2016

Certification & Technologies

ML/Stats: Tensorflow, Julia, Matlab, Deep Learning Specialization [\[link\]](#)

Systems: Kafka, Spark, Linux kernel, Relational databases, GPGPU

PL: C/++, Assembly, Go, Haskell, Java, Python, Prolog, Scala, SQL, TLA+

Infra: AWS, Jenkins, Git, Jira, Docker, Kubernetes

Teaching

Security Analytics, The University of Melbourne, 2023.

Data Management Labs, TU Darmstadt, 2021.

Introduction to Database Systems, Brown University 2015, 2018

Advanced Topics in Database Systems, Brown University 2015

Services

Reviewer: OSDI'16, TVCG'21, TODS'21, IEEE S&P (Journal)'24, IEEE S&P (Oakland)'24

Program committee: VLDB'22 (Demo Track)

Member: IACR, ACM

Honors

Graduate Fellowship, Brown University, 2014-2019

Dean's Honor List, University of Wisconsin at Madison, 2007, 2008, 2009

Eta Kappa Nu, 2009

Upsilon Pi Epsilon, 2010

Golden Key International Honour Society, 2010

Dissertation Fellowship, Brown University, 2018-2019

References

Available upon request.