

Zeyuan Shang

Cambridge, MA, USA

☎ 4014104368 | ✉ zeyuans@mit.edu | 🏠 www.shangzeyuan.com | 📱 zeyuanxy | 🌐 zeyuanxy

Research Interest

Systems for ML, ML for Systems, Large-scale Interactive Data Analytics

Education

Massachusetts Institute of Technology

Cambridge, MA, USA

PH.D. STUDENT, ELECTRICAL ENGINEERING & COMPUTER SCIENCE

2018 - Now

- Advisor: Prof. Tim Kraska

Tsinghua University

Beijing, China

B.ENG IN COMPUTER SCIENCE

2011 - 2015

- Advisor: Prof. Guoliang Li
- Major GPA: 91/100
- Undergraduate Thesis: *K-Join: Knowledge-Aware Similarity Join*(Outstanding Undergraduate Thesis)
- Graduated with Distinction

Experience

Database Group, Brown University

Providence, RI, USA

RESEARCH ASSISTANT

2017 - 2018

- Automated machine learning problems
- Finding efficient execution plans for DBMSs
- Building powerful ETL systems

Hulu

Beijing, China

FULLTIME SOFTWARE DEVELOPER

2015 - 2017

- Designed and implemented a streaming ingest framework using Spark Streaming, Kafka and HBase, which has been deployed into product
- Designed and implemented a realtime platform for monitoring the status of cluster, usage of HDFS and data statistics in HBase
- Designed and implemented a data dumping pipeline on Spark supporting UDF, task scheduling along with a fancy UI based on React
- Maintained and improved a real-time distributed OLAP query engine for Parquet-like nested data (e.g., supporting more aggregation operations, filter pushdown and code generation)
- Maintained and improved a data management platform for marketing team and research team (e.g., writing MapReduce code for batch pipeline, programming ETLs to load and dump data, improving web portal's UI)
- Rewarded several promotions for hard work, dedication and excellent performance

Publications

Towards Interactive Curation & Automatic Tuning of ML Pipelines

CARSTEN BINNIG, BENEDETTO BURATTI, YEOUNOH CHUNG, CYRUS COUSINS, TIM KRASKA, ZEYUAN SHANG,

ISABELLA TROMBA, ELI UPFAL, ROBERT ZELEZNIK, EMANUEL ZGRAGGEN

- DEEM 2018 (2nd Workshop on Data Management for End-to-End Machine Learning)
- SysML 2018

DITA: Distributed In-Memory Trajectory Analytics

ZEYUAN SHANG, GUOLIANG LI, ZHIFENG BAO

- SIGMOD 2018
- SIGMOD 2018 (Demo)

DIMA: A Distributed In-Memory Similarity-Based Query Processing System

Ji SUN, ZEYUAN SHANG, GUOLIANG LI, DONG DENG, ZHIFENG BAO

- VLDB 2017 (Demo)

K-Join: Knowledge-Aware Similarity Join

ZHEYUAN SHANG, YAXIAO LIU, GUOLIANG LI, JIANHUA FENG

- IEEE Transactions on Knowledge and Data Engineering (TKDE), 2016
- ICDE 2017 (Extended Abstract)

Projects

Alpine Meadow: Interactive and Automatic Machine Learning

DATABASE GROUP, MIT

- Built several components of the project: infrastructure for accessing and processing data, pipeline for training and testing data, back-end RPC service for interacting with front-end
- Helped design and implement the rule-based optimizer, a component which automatically finds the optimal solution to prediction problems
- Adopted the agile development technique based on my industry experiences to help coordinate the rapid progress of the project for the whole team.

Fast ETL

DATABASE GROUP, BROWN UNIVERSITY

- Implemented a CSV reader that inferred the data schema automatically by sampling data, and generated native C code to read CSV data using its schema to improve performance.
- Implemented some benchmarks in Spark for comparison with our implementation.
- Set up the continuous integration of our system, using Docker and Travis.

Honors & Awards

2018 **Ashar Aziz Presidential Fellowship**, MIT EECS

2018 **Student Travel Award**, SIGMOD 2018

2015 **Outstanding Graduate(10/124)**, Computer Science Department, Tsinghua University

2015 **Outstanding Undergraduate Thesis Award(6/124)**, Tsinghua University

2013, 2014 **Academic Excellence Scholarship(10/124)**, Tsinghua University

2014 **First Prize, Challenge Cup(Top Prize for Innovation)**, Tsinghua University