

Jason (Jia) Teoh

jiateoh at gmail | <https://jiateoh.github.io> | [linkedin.com/in/jiateoh](https://www.linkedin.com/in/jiateoh)
Los Angeles, CA

Education

-
- Ph.D., Computer Science** Sept '16 - Present
Advisor: Dr. Miryung Kim
Research Interests: Debugging in Big Data Systems
University of California, Los Angeles | GPA: 3.940 (ongoing)
- M.S., Computer Science** Sept '16 - March '19
University of California, Los Angeles | GPA: 3.933
- B.A., Computer Science** Aug '10 – May '13
University of California, Berkeley | GPA: 3.868 (High Distinction)

Research Experience

-
- Software Engineering and Analysis Laboratory (SEAL)** March '18 - Present
- Investigate debugging precision of big data systems with a focus on root cause analysis.
 - Combine latency instrumentation with data provenance to compute record-level latency and identify input subsets responsible for causing computation skew.
 - Improve data provenance trace precision of suspicious or faulty output records in dataflow applications by merging taint analysis with influence-based aggregation provenance.
 - Generate workloads which reproduce desired performance symptoms by leveraging targeted user-defined function fuzzing and skew-inspired mutation operators.
- Scalable Analytics Institute (ScAi)** Sept '16 - March '18
- Built a mobile dataflow computing system that integrates with cloud computing for distribution of data and computation.
 - Developed Android framework integrated with Spark Catalyst Optimizer to automatically export sub-plans and coordinate with Spark jobs through Apache Kafka.

Industry Project Experience

-
- Cloud Spanner** (<https://cloud.google.com/spanner/>) June '19 - Sept '19
Software Engineering Intern - Google
- Designed and implemented an automated rule-based diagnostics framework to analyze query executions and identify performance problems or anomalies.
 - Integrated the rule-based analysis framework with the existing diagnostics infrastructure to visually bind detected issues with corresponding query execution information.
- Thirdeye** (github.com/linkedin/pinot/tree/master/thirdeye) Oct '15 - Sept '16
Senior Software Engineer - LinkedIn
- Open source anomaly detection framework for monitoring business metrics.
 - Designed and implemented multifaceted dashboard for efficient anomaly investigation.

- Investigated and designed solutions to improve anomaly detection scalability.
- Onboarded new use cases from data bootstrapping to server setup.

LinkedIn Segmentation and Targeting Tool (LISTT)

June '13 - Sept '15

Senior Software Engineer (5/15- 9/15), Software Engineer (6/13 - 5/15) - LinkedIn

- Suite of self-service applications for targeting audiences via custom member attributes.
- Enhanced computational flow to improve data quality and application robustness for over 500 metrics per each of LinkedIn's 400 million members at up to daily refresh cycle.
- Heterogenous data integration for data definitions defined via Teradata, Hive, Pig.
- Leveraged Apache Datafu Hourglass to reduce redundant computation and optimize up data delivery.
- Independently developed flow-monitoring CLI and graph UI for time and stability analysis of application

RaINier

May '12 - Aug '12

Software Engineer Intern - LinkedIn

- Developed self-service business analytics dashboard tool for key metrics and splits.
- Implemented multidimensional (OLAP) cube caching layer with background refresh.
- Multiple project presentations for project fair, business users, and prospective developers.

Publications

[SoCC 2020] Jason Teoh, Muhammad Ali Gulzar, and Miryung Kim. 2020. Influence-Based Provenance for Dataflow Applications with Taint Propagation. *ACM Symposium on Cloud Computing (SoCC '20)*. 24.4% acceptance rate

[SoCC 2019] Jason Teoh, Muhammad Ali Gulzar, Guoqing Harry Xu, and Miryung Kim. 2019. PerfDebug: Performance Debugging of Computation Skew in Dataflow Systems. *ACM Symposium on Cloud Computing (SoCC '19)*. 24.8% acceptance rate

Honors/Awards

Member of Upsilon Pi Epsilon (Berkeley chapter)

Spring '12

Dean's Honors List (UC Berkeley)

Fall '11 to Spring '13

Skills

Programming: Scala, Java, TypeScript, Python, SQL, Bash, Apache Pig, Apache Hive, Javascript, Ruby, R, C, LaTeX

Frameworks/Libraries: Spark, Hadoop, Kafka, Spring, Lucene, AngularJS, jQuery, Sigma (js), Datafu Hourglass, Ruby on Rails, Maven, Gradle

Teaching

Software Engineering (CS130), Teaching Assistant - UCLA [remote]

Fall '20

Database Systems (CS143), Teaching Assistant - UCLA

Fall '17