Jia (Jason) Teoh

jiateoh@gmail.com | linkedin.com/in/jiateoh | https://jiateoh.github.io
Los Angeles, CA

Education

Ph.D., Computer Science 9/16 - Present
UC Los Angeles, GPA: 3.940 (ongoing)

M.S., Computer Science 9/16 - 3/19
UC Los Angeles, GPA: 3.933

B.A., Computer Science 8/10 – 5/13
UC Berkeley, GPA: 3.868 (High Distinction)

Research Experience

Software Engineering and Analysis Laboratory (SEAL) 3/18 - Present

- Investigate performance debugging of big data systems such as Apache Spark with a focus on data impact
- Combine latency instrumentation with data provenance to produce record-level performance provenance which associates application performance with input data and program definitions.

Scalable Analytics Institute (ScAi) 9/16 - 3/18

- Investigated mobile dataflow computing and bridging the gap between mobile and cloud computing
- Developed Android framework integrated with Spark Catalyst Optimizer to automatically export sub-plans and integrate with Spark jobs through Apache Kafka.

Publications

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

Project Experience

Cloud Spanner (https://cloud.google.com/spanner/) 6/19-9/19
Software Engineering Intern - Google

- Migrated an internal diagnostics infrastructure from JavaScript to TypeScript
- Designed and implemented an automated rule-based diagnostics framework to analyze query executions and identify performance problems or anomalies. Detected issues are presented to users of the tool alongside additional documentation and potential fixes.
- Collaborated with Spanner developers to identify dozens of query execution characteristics that typically indicate poor performance. Implemented and refined a small subset of these rules for demonstration.
- 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) 10/15 - 9/16
Senior Software Engineer - LinkedIn

- Open source anomaly detection framework for monitoring business metrics.
  - Utilizes Pinot (realtime distributed OLAP datastore with horizontal scaling and low latency).
- Designed and implemented multifaceted dashboard for straightforward anomaly investigation.
- Investigated and designed solutions to improve anomaly detection scalability.
- Onboarded new use cases from data bootstrapping to server setup.
- Start[in]: Coached new college graduates in workplace topics such as communication and owning one's role.

LinkedIn Segmentation and Targeting Tool (LISTT) 6/13 - 9/15

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

- Suite of self-service applications for targeting audiences and computing user-defined member-level attributes.
- Enhanced computational flow to improve data quality and application robustness for over 500 metrics per each of Linkedin's 400 million members, on a daily or weekly refresh cycle.
- Heterogenous data integration for hundreds of data definitions defined by SQL (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
- Mentored intern and onboarded new hires.
- Attended Hadoop Summit (2014-2015) to learn about big data technologies such as Spark, Mesos, YARN.

RaINier 5/12 - 8/12

Software Engineer Intern - LinkedIn

- Developed self-service business analytics dashboard tool for viewing key metrics and dimensional splits
- Implemented multidimensional (OLAP) cube caching layer with background refresh daemons.
- Presented project to a variety of audiences: project fair, business users, and prospective developers.

Teaching

Database Systems

Teaching Assistant (CS 143) - UCLA

- Lead discussions in class and online (Piazza), held office hours, administered grades, managed class projects, conducted extra study reviews.

Selected Coursework

Distributed Systems, Databases, Machine Learning, Software Engineering, Operating Systems, Artificial Intelligence, Statistical Programming, Probability, Efficient Algorithms and Intractable Problems, Compilers

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