

Vicky Papavasileiou

San Francisco, CA || (805) 886-5282 || vicky.papavas@gmail.com || linkedin.com/in/vicky-papavasileiou

Education

- PhD in Computer Science — University of California, San Diego** 09/2010 - 10/2018
Thesis: A Declarative Framework for Big Graph Analytics and their Provenance
Advisors: Alin Deutsch, Ken Yocum
- MSc in Computer Science — University of Crete, Greece** 10/2007 - 12/2009
Thesis: Detecting Deterministically High-level Changes for RDF/S Knowledge Bases
Advisor: Vassilis Christophides
- BS in Computer Science — University of Crete, Greece** 09/2002 - 09/2007

Experience

- Human Longevity Inc.**, San Diego, CA, USA — Software Engineer Intern 06/2016 - 08/2016
- Developed a Python library for tracking provenance in a pipeline of ML analytics running on AWS. The library detects outliers in the output, faulty runs and erroneous DNA samples in the input and reports provenance information used by data scientists to improve performance and correctness of their analytics.
- NEC Laboratories Inc.**, Cupertino, CA, USA — Research Assistant Intern 06/2015 - 08/2015
- Extended Apache Giraph to support asynchronous execution. I implemented non-blocking messaging with high-parallelism read/write operations and removed unnecessary synchronization barriers. Experiments revealed that running time decreased by 1.7x for PageRank.

Research

- Declarative large-scale graph processing** 09/2013 - 10/2018
- Developed a distributed query processor for Big Graph analytics executing on Vertex-Centric (VC) large-scale graph processing engines. I implemented a compiler that translates Datalog queries to optimized VC graph analytics. The prototype is built on Apache Giraph. Experiments on common graph analytics reveal a speedup of 2x compared to their imperative counterparts.
 - Designed a provenance query language (PQL) based on Datalog and implemented a provenance management system for Big Graph analytics on VC engines. I implemented two novel provenance query evaluation methods that incur 1.3x overhead over the baseline compared to the 8x overhead of the traditional approach.
- Middleware query processor** 09/2010 - 09/2013
- Participated in the development of a main-memory query processor with federated database capabilities. I designed algebraic operators for SQL++, a semi-structured query language and implemented logical plan optimizations that transform tuple-at-a-time nested queries to set-at-a-time evaluation plans offering 7x speedup.

Technical Skills

Languages: Java, Python, C, SQL.

Technologies: Apache Giraph, Apache Hadoop, GraphLab, PostgreSQL, Apache Spark, HBase.

Publications

- V. Papavasileiou, K. Yocum, A. Deutsch. Ariadne: Online Provenance for Big Graph Analytics. Sigmod, 2019.
- W. Moustafa, V. Papavasileiou, K. Yocum, A. Deutsch. Datalography: Scaling Datalog Graph Analytics on Graph Processing Systems. Big Data, 2016.
- Y. Fu, K. W. Ong, Y. Papakonstantinou, V. Papavasileiou, R. Vernoux. Holistic Data Access Optimization for Analytics Reports. Technical Report, 2013.
- V. Papavasileiou, G. Flouris, I. Fundulaki, D. Kotzinos, V. Christophides. High-Level Change Detection in RDF(S) KBs. Transactions on Database Systems, 38(1). 2013.
- V. Papavasileiou, G. Flouris, I. Fundulaki, D. Kotzinos, V. Christophides. On Detecting High-Level Changes in RDF/S KBs. ISWC, 2009. Best Student Paper Award.