

Jia Yu

PH.D. CANDIDATE

☎ (+1) 480-297-8386 | ✉ jjayu2@asu.edu | 🏠 [jjayuasu.github.io](https://github.com/jjayuas) | 📱 [jjayuasu](https://www.linkedin.com/in/jjayuas) | 📺 [jia-yu-27632182](https://www.youtube.com/channel/UCjia-yu-27632182)

Summary

Jia Yu is a PhD candidate in the Department of Computer Science at Arizona State University working with Assistant Professor [Mohamed Sarwat](#). He is the recipient of 2019 ASU Fulton Engineering Graduate Fellowship, 2017 ACM SIGSPATIAL Student Research Competition bronze medal, and 2019 SSTD Best Demo Paper Runner-up. Jia's research focuses on large-scale database systems and geospatial data management. In particular, he worked on distributed geospatial data management systems, database indexing, and geospatial data visualization. His research outcomes have appeared in the most prestigious database / GIS conferences and journals, including SIGMOD, VLDB, ICDE, SSTD and Geoinformatica Journal. He is the main architect of several open-source systems which include GeoSpark project that receives 10,000 downloads per month and has users / contributors from major companies (e.g., Facebook, Uber, AT&T, and MoBike).

Education

Arizona State University

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

Tempe, Arizona, U.S.A

Sept. 2013 - May 2020 (expected)

- Advisor: Assistant Professor [Mohamed Sarwat](#) (started in January 2015)
- Thesis: System Support for Large-scale Geospatial Data Analytics
- Thesis committee: Mohamed Sarwat, Kasim Selcuk Candan, Ming Zhao, Wenwen Li (from ASU geography department)

Northwest Agriculture and Forestry University

BACHELOR OF ENGINEERING IN SOFTWARE ENGINEERING

Yangling, Shaanxi, China

Sept. 2009 - Jul. 2013

- Outstanding Graduate (200 / 5600)

Employment History

My internship experience matches my expertise in database systems and geospatial data management: (1) Microsoft is the birthplace of SQL Server. (2) IBM -Almaden is the birthplace of relational model, SQL, and DB2. (3) Apple is the birthplace of Apple Map.

Microsoft Research

RESEARCH INTERN, DATABASE GROUP

Redmond, Washington, U.S.A

Jun. 2019 - Aug. 2019

- Mentor / Collaborators: Umar Farooq Minhas, David Lomet, Jaeyoung Do, Yinan Li, Chi Wang, Badrish Chandramouli, Johannes Gehrke, Donald Kossmann
- I participated in the ALEX project to explore a realistic design of updatable learned indices.

IBM Almaden Research Center

RESEARCH INTERN, EXPLORATORY DATABASES GROUP

San Jose, California, U.S.A

May. 2018 - Aug. 2018

- Mentor / Collaborators: Vijayshankar Raman, Yingjun Wu, Yuanyuan Tian, Ronald Barber, and Richard Sidle
- I participated in the Hermit project to design a succinct secondary index. I also explored the code generation issues on compressed database tables using the LLVM JIT engine.

Apple

SOFTWARE DEVELOPMENT INTERN, APPLE MAP TEAM

Cupertino, California, U.S.A

Jun. 2016 - Aug. 2016

- Mentor / Collaborators: Huang-Hsiang Cheng, Alex Radeski
- I worked on cluster computing frameworks and resource management systems such as Apache Spark and Apache Mesos. I developed internal evaluation tools to assist in large-scale spatial analysis.

Research Experience

My school projects: (1) have 1000+ stars and forks on GitHub (2) have 10,000+ monthly downloads (3) are used by many companies in production (4) are implemented in the kernels of widely used data systems, Apache Spark and PostgreSQL.

Research Assistant - Arizona State University, Tempe, Arizona, USA

January, 2015 - Present

Tabula: turbocharge geospatial visualization dashboards via a materialized sampling cube

MAIN DEVELOPER

- Designed Tabula, a middleware that sits between the data system and the geospatial visualization dashboard to accelerate the interactive visual analysis. It adopts a materialized sampling cube approach, which pre-materializes sampled answers for a set of potential queries. Tabula allows user-defined sample accuracy loss and guarantees bounded deterministic accuracy loss.
- Results of this work are published in ICDE 2020 (P3). A demo paper about Tabula is under review by SIGMOD 2020 (??).
- Implemented Tabula in SparkSQL, open-sourced the system on [GitHub](#) and released a [video](#) to demonstrate how Tabula works.

Hippo: a fast yet lightweight database indexing scheme

MAIN DEVELOPER

- Designed Hippo, a fast yet lightweight database indexing scheme. It significantly shrinks the index storage and mitigates maintenance overhead without compromising much on the query execution performance. Hippo stores disk page ranges instead of tuple pointers in the indexed table to reduce the storage space occupied by the index. It maintains simplified histograms that represent the data distribution. When a query is issued, Hippo leverages the page ranges and histogram-based page to prune irrelevant pages.
- Results of Hippo are published in VLDB 2016 (P17) and demonstrated in ICDE 2017 (P16). Hippo for spatial data is published in SSTD 2017 (P14).
- Hippo is a PostgreSQL 9.6 built-in index. I open-sourced Hippo on [GitHub](#) and released a [video](#) to demonstrate how Hippo works.

GeoSpark: a cluster computing framework for processing big spatial data

MAIN DEVELOPER

- I am leading the GeoSpark project, a cluster computing framework for processing big spatial data, and collaborating with 30 contributors from GeoSpark community. In particular, I contributed to the following components in GeoSpark ecosystem.
 1. Designed [GeoSpark](#), a cluster computing system which extends Apache Spark and SparkSQL to support spatial data types, indexes, and queries at scale. Users interact with the system using Spatial SQL, Scala, Java, Python and R.
 2. Designed [GeoSparkViz](#), a large-scale geospatial map visualization framework which extends GeoSpark to provide native support for general cartographic design. It encapsulates the main steps of the geospatial map visualization process, e.g., rasterize spatial objects, aggregate pixels, into a set of massively parallelized RDD transformations.
 3. Mentored [GeoSparkSim](#), a scalable traffic simulator which extends Apache Spark to generate large-scale road network traffic datasets. GeoSparkSim converts road networks to Spark graphs and simulated vehicles to VehicleRDDs. It takes into account microscopic traffic models such as traffic lights, lane changing, and car following.
- I have published several papers featuring different components of GeoSpark ecosystem.
 1. Results of [GeoSpark](#) are published in Geoinformatica Journal (P12) and demonstrated in ICDE 2016 (P18). I also gave GeoSpark tutorials in ICDE 2019 (P9) and SIGSPATIAL 2019 Workshop (P4). Preliminary results are published in SIGSPATIAL 2015 (P19).
 2. Results of [GeoSparkViz](#) are published in SSDBM 2018 (P13) and demonstrated in ICDE 2019 (P11). Some preliminary results are published in SIGSPATIAL 2017 Student Research Competition (P15).
 3. Results of [GeoSparkSim](#) are published in MDM 2019 (P5) and demonstrated in SSTD 2019 (P6, Best Demo Paper Runner-Up).
- I maintain [GeoSpark GitHub source code](#) and [GeoSpark website](#). GeoSpark ecosystem receives 10K downloads per month (see [statistics from Maven Central](#)). Users and contributors are from Facebook, Uber, MoBike and numerous startups. [Databricks](#) (the tech unicorn behind Apache Spark) featured GeoSpark in its [blog post](#) and provides an [interactive GeoSpark notebook](#) for its Spark Cloud. A British company, [Gyana](#), uses GeoSpark as the backend engine of its Location Intelligence dashboard ([demo video](#)).

Research Intern - Microsoft Research, Redmond, Washington, USA

June, 2019 - August, 2019

ALEX: an updatable adaptive learned index

TEAM MEMBER

- Collaborated with researchers in Microsoft Research Database group to design a new set of model-based learned index structures called ALEX, that work beyond static data. ALEX reorganizes storage structures for index nodes that support efficient data insertion. As a side effect, the reorganized storage structures also improve the search performance. Moreover, ALEX equips an adaptive index structure to gain model robustness when data distribution shifts.
- Results of this work are published in SIGMOD 2020 (P1).

Research Intern - IBM Almaden Research Center, San Jose, California, USA

May, 2018 - August, 2018

Hermit: design a succinct secondary index by exploiting column correlations

TEAM MEMBER

- Collaborated with researchers in IBM Almaden Research Center Database group to design Hermit, a succinct secondary indexing mechanism for modern RDBMSs. Hermit judiciously leverages the rich soft functional dependencies hidden among columns to prune out redundant structures for indexed key access. Instead of building a complete index that stores every single entry in the key columns, Hermit uses a succinct, ML-enhanced data structure to navigate any incoming key access queries to an existing index built on the correlated columns.
- Results of this work are published in SIGMOD 2019 (P7) and demonstrated in VLDB 2019 (P8).

Peer-Reviewed Publications

- Research papers published on Database 1st tier conferences (SIGMOD, VLDB and ICDE) and GIS 1st tier journal / conference (Geoinformatica and SSTD)
- 19 publications: 13 first-author, 5 second-author, 1 third-author
- 2 papers are under review
- The first GeoSpark paper (P19, Nov. 2015) is [the most cited paper](#) among all 633 papers from 2014 - 2019 and [the 7th most cited paper](#) among all 935 papers from 2011 to 2019 in ACM SIGSPATIAL (one of the most prestigious conferences on spatial data management)
- Google Scholar citation: 330+

P1 Jialin Ding, Umar Farooq Minhas, **Jia Yu**, Chi Wang, Hantian Zhang, Yinan Li, Jaeyoung Do, Donald Kossmann, Johannes Gehrke, David Lomet, Badrish Chandramouli, and Tim Kraska. ALEX: An Updatable Adaptive Learned Index. In *Proceedings of the ACM International Conference on Management of Data, SIGMOD*, page to appear, 2020 [**Summer 2019 intern work at Microsoft Research**]

P2 **Jia Yu** and Mohamed Sarwat. Big Geospatial Data Processing Made Easy: A Working Guide to GeoSpark. In *Handbook of Big Geospatial Data*, page to appear. 2020 (Book chapter)

P3 **Jia Yu** and Mohamed Sarwat. Turbocharging Geospatial Visualization Dashboards via a Materialized Sampling Cube Approach. In *Proceedings of the International Conference on Data Engineering, ICDE*, page to appear, 2020 ([pdf](#)) [**Accepted directly after 1st round of review, 3% direct acceptance rate**]

P4 **Jia Yu** and Mohamed Sarwat. Spatial Data Wrangling with GeoSpark - A Step by Step Tutorial. In *International Conference on Advances in Geographic Information Systems Spatial API Workshop, ACM SIGSPATIAL Spatial API Workshop*, page to appear, 2019 (Tutorial) ([pdf](#))

P5 Zishan Fu, **Jia Yu**, and Mohamed Sarwat. Building Microscopic Road Network Traffic Simulators in Apache Spark. In *Proceedings of the International Conference on Mobile Data Management, MDM*, pages 320--328, 2019 ([pdf](#))

P6 Zishan Fu, **Jia Yu**, and Mohamed Sarwat. Demonstrating GeoSparkSim: A Scalable Microscopic Road Network Traffic Simulator Based on Apache Spark. In *Proceedings of the International Symposium on Advances in Spatial and Temporal Databases, SSTD*, pages 186--189, 2019 (Demo paper) ([pdf](#)) [**Best Demo Paper Runner-Up**]

P7 Yingjun Wu, **Jia Yu**, Yuanyuan Tian, Ronald Barber, and Richard Sidle. Designing Succinct Secondary Indexing Mechanism by Exploiting Column Correlations. In *Proceedings of the ACM International Conference on Management of Data, SIGMOD*, pages 1223--1240, 2019 ([pdf](#)) [**Summer 2018 intern work at IBM Almaden Research Center**]

P8 Yingjun Wu, **Jia Yu**, Yuanyuan Tian, Richard Sidle, and Ronald Barber. HERMIT in action: Succinct secondary indexing mechanism via correlation exploration. *Proceedings of the VLDB Endowment, PVLDB*, 12(12):1882--1885, 2019 (Demo paper) ([pdf](#)) [**Summer 2018 intern work at IBM Almaden Research Center**]

- P9 **Jia Yu** and Mohamed Sarwat. Geospatial Data Management in Apache Spark: A Tutorial. In *Proceedings of the International Conference on Data Engineering, ICDE*, pages 2060--2063, 2019 (Tutorial) ([pdf](#), [website](#))
- P10 Yuhan Sun, **Jia Yu**, and Mohamed Sarwat. Demonstrating Spindra: A Geographic Knowledge Graph Management System. In *Proceedings of the International Conference on Data Engineering, ICDE*, pages 2044--2047, 2019 (Demo paper) ([pdf](#))
- P11 **Jia Yu**, Anique Tahir, and Mohamed Sarwat. GeoSparkViz in Action: A Data System with built-in support for Geospatial Visualization. In *Proceedings of the International Conference on Data Engineering, ICDE*, pages 1992--1995, 2019 (Demo paper) ([pdf](#))
- P12 **Jia Yu**, Zongsi Zhang, and Mohamed Sarwat. Spatial Data Management in Apache Spark: the GeoSpark Perspective and Beyond. *Geoinformatica Journal*, 23(1):37--78, 2019 ([pdf](#))
- P13 **Jia Yu**, Zongsi Zhang, and Mohamed Sarwat. GeoSparkViz: a scalable geospatial data visualization framework in the Apache Spark ecosystem. In *Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM*, pages 15:1--15:12, 2018 ([pdf](#))
- P14 **Jia Yu** and Mohamed Sarwat. Indexing the Pickup and Drop-Off Locations of NYC Taxi Trips in PostgreSQL - Lessons from the Road. In *Proceedings of the International Symposium on Advances in Spatial and Temporal Databases, SSTD*, pages 145--162, 2017 ([pdf](#))
- P15 **Jia Yu**. SRC: Geospatial Visual Analytics Belongs to Database Systems: the BABYLON approach. *International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL*, 9(3):2--3, 2017 (Extended Abstract) ([pdf](#)) [**Third Place of ACM SIGSPATIAL Student Research Competition**]
- P16 **Jia Yu**, Raha Moraffah, and Mohamed Sarwat. Hippo in Action: Scalable Indexing of a Billion New York City Taxi Trips and Beyond. In *Proceedings of the International Conference on Data Engineering, ICDE*, pages 1413--1414, 2017 (Demo paper) ([pdf](#))
- P17 **Jia Yu** and Mohamed Sarwat. Two Birds, One Stone: A Fast, yet Lightweight, Indexing Scheme for Modern Database Systems. *Proceedings of the VLDB Endowment, PVLDB*, 10(4):385--396, 2016 ([pdf](#))
- P18 **Jia Yu**, Jinxuan Wu, and Mohamed Sarwat. A demonstration of GeoSpark: A cluster computing framework for processing big spatial data. In *Proceedings of the International Conference on Data Engineering, ICDE*, pages 1410--1413, 2016 (Demo paper) ([pdf](#))
- P19 **Jia Yu**, Jinxuan Wu, and Mohamed Sarwat. GeoSpark: a cluster computing framework for processing large-scale spatial data. In *International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL*, pages 70:1--70:4, 2015 (Short paper) ([pdf](#))

Under Review

- R1 **Jia Yu**, Zishan Fu, and Mohamed Sarwat. Dissecting GeoSparkSim: A Scalable Microscopic Road Network Traffic Simulator in Apache Spark. *Distributed and Parallel Databases Journal*, 2019 (Under review)
- R2 **Jia Yu** and Mohamed Sarwat. GeoSparkViz: A Cluster Computing System for Visualizing Massive-Scale Geospatial Data. *The VLDB Journal*, 2020 (Under review)

Before 2015

- Zijiang Yang, Bei-Bei Yin, Junpeng Lv, Kai-Yuan Cai, Stephen S. Yau, and **Jia Yu**. Dynamic Random Testing with Parameter Adjustment. In *IEEE Annual Computer Software and Applications Conference, COMPSAC Workshops*, pages 37--42, 2014
- Lei Zhang, Bei-Bei Yin, Junpeng Lv, Kai-Yuan Cai, Stephen S. Yau, and **Jia Yu**. A History-Based Dynamic Random Software Testing. In *IEEE Annual Computer Software and Applications Conference, COMPSAC Workshops*, pages 31--36, 2014

Teaching Experience

2019	Instructor , CSE511 Data Processing at Scale (graduate, introduction)	<i>ASU, U.S.A</i>
2019	Teaching Assistant , CSE412 Database Management (undergraduate)	<i>ASU, U.S.A</i>
2018	Coursera course designer , ASU MS Degree of CS: Data Systems, over 10K learners (statistics)	<i>ASU, U.S.A</i>
2018	Teaching Assistant , CSE412 Database Management (undergraduate)	<i>ASU, U.S.A</i>
2016	Teaching Assistant , CSE512 Distributed Database Systems (graduate)	<i>ASU, U.S.A</i>
2015	Teaching Assistant , CSE512 Distributed Database Systems (graduate)	<i>ASU, U.S.A</i>
2014	Teaching Assistant , CSE543 Information Assurance (graduate)	<i>ASU, U.S.A</i>
2014	Teaching Assistant , CSE240 Introduction to Programming Languages (undergraduate)	<i>ASU, U.S.A</i>

Research Funding

III: Small: Towards Data Systems Support for Geospatial Visualization	<i>U.S.A</i>
SUBMITTED TO NATIONAL SCIENCE FOUNDATION CISE IIS (INFORMATION AND INTELIGENT SYSTEMS)	<i>2019</i>
• Collaborated with my advisor Mohamed Sarwat (PI) to write the proposal.	

Patent

Systems and Methods for an End-To-End Visual Analytics System for Massive-Scale Geospatial Data	<i>U.S.A</i>
FILED AS UNITED STATES NON-PROVISIONAL PATENT 16355261 (LINK)	<i>2019</i>

Mentoring Experience

As a senior Ph.D. student in my lab, I had the opportunity to mentor the following junior students:

PhD Students

Ankita Sharma	<i>Arizona State University</i>
PHD STUDENT IN COMPUTER SCIENCE	<i>Fall 2019 - Spring 2020</i>

- We collaborated on the unreliable computation issue of large-scale Internet-of-Things streaming data.

Kanchan Chowdhury	<i>Arizona State University</i>
PHD STUDENT IN COMPUTER SCIENCE	<i>Fall 2019 - Spring 2020</i>

- We collaborated on the front-end interface of Tabula.

Master Students

Zishan Fu	<i>Arizona State University</i>
MASTER STUDENT IN COMPUTER SCIENCE, NOW SDE AT CERNER, USA	<i>Spring 2019</i>

- Thesis: Scalable Microscopic Traffic Simulator
- We collaborated on GeoSparkSim. Results are published in MDM 2019 (P5) and demonstrated in SSTD 2019 (P6).

Anique Tahir	<i>Arizona State University</i>
MASTER STUDENT IN COMPUTER SCIENCE, NOW PHD STUDENT AT ASU, USA	<i>Spring 2018</i>

- We collaborated on GeoSparkViz. Results are demonstrated in ICDE 2019 (P11).

Zongsi Zhang	<i>Arizona State University</i>
MASTER STUDENT IN COMPUTER SCIENCE, NOW SDE AT GRABTAXI, SINGAPORE	<i>Fall 2017 - Spring 2018</i>

- We collaborated on GeoSpark and GeoSparkViz. Results are published in Geoinformatica Journal (P12) and SSDBM 2018 (P13).

Jinxuan Wu	<i>Arizona State University</i>
MASTER STUDENT IN COMPUTER SCIENCE, NOW SENIOR SDE AT BLOOMBERG, USA	<i>Fall 2015 - Spring 2016</i>

- We collaborated on GeoSpark. Preliminary results are published in SIGSPATIAL 2015 (P19) and demonstrated in ICDE 2016 (P18).

Academic Service

2020	Invited Journal Reviewer , IEEE Transactions on Parallel and Distributed Systems	<i>U.S.A</i>
2020	Invited Journal Reviewer , Computers and Geosciences	<i>U.S.A</i>
2018 - 19	Invited Journal Reviewer , IEEE Transactions on Cloud Computing	<i>U.S.A</i>
2018 - 19	Invited Journal Reviewer , Geoinformatica Journal	<i>U.S.A</i>
2018 - 19	Invited Journal Reviewer , International Journal of Geographical Information Science (IJGIS)	<i>U.S.A</i>
2018 - 19	Invited Journal Reviewer , ACM Transactions on Spatial Algorithms and Systems (ACM TSAS)	<i>U.S.A</i>
2018 - 19	Invited Journal Reviewer , VLDB Journal	<i>U.S.A</i>
2018	Publicity Chair , GeoRich Workshop @ SIGMOD	<i>U.S.A</i>
2016 - 19	External Reviewer , SIGMOD, PVLDB, ICDE, TODS, SIGSPATIAL, SSTD	<i>U.S.A</i>

Honors & Awards

2019	Best Demo Paper Runner-Up , SSTD	<i>Vienna, Austria</i>
2019	Engineering Graduate Fellowship , Ira A. Fulton Schools of Engineering	<i>ASU, U.S.A</i>
2016 - 19	NSF Student Travel Grant (3 times) , IEEE ICDE	
2015 - 19	NSF Student Travel Grant (4 times) , Microsoft Student Travel Grant , ACM SIGSPATIAL	
2017	Third Place of Student Research Competition , ACM SIGSPATIAL	<i>Los Angeles, U.S.A</i>
2013	Outstanding Graduate , Northwest A & F University	<i>Yangling, China</i>
2011 - 12	First-class Scholarship (2 times) , Northwest A & F University	<i>Yangling, China</i>
2011 - 12	Merit Student (2 times) , Northwest A & F University	<i>Yangling, China</i>
2010 - 12	Outstanding Student Leader (3 times) , Northwest A & F University	<i>Yangling, China</i>
2011	President of the Student Association of College of Info. Engr. , Northwest A & F University	<i>Yangling, China</i>

Presentation

- Conference talks (8 times): VLDB, ICDE, SIGSPATIAL, SSTD, MDM, ApacheCon (Apache Software Foundation annual conference)
- Company talks (7 times): Microsoft Research, IBM Almaden Research Center, Apple, NVidia, StateFarm, Vocareum
- The slides of my 12 talks are available. Other talks are confidential due to NDA.

Spatial Data Wrangling With GeoSpark: A Step-by-Step Tutorial

ACM SIGSPATIAL SPATIALAPI WORKSHOP

Chicago, Illinois, U.S.A

Nov. 2019

- [Talk slides](#) and [coding examples](#)

GeoSpark and Geospatial Data Management in Apache Spark

APACHECON 2019 NORTH AMERICA

Las Vegas, Nevada, U.S.A

Sept. 2019

- [Talk slides](#)

ALEX: An Updatable Learned Index

MICROSOFT RESEARCH

Redmond, Washington, U.S.A

Aug. 2019

- Slides not available due to NDA

Designing Succinct Secondary Indexes by Exploiting Column Correlations

MICROSOFT RESEARCH

Redmond, Washington, U.S.A

Jul. 2019

- [Presentation video](#)

Building a Large-Scale Microscopic Road Network Traffic Simulator in Apache Spark

MDM 2019

Hong Kong, China

Jun. 2019

- [Conference presentation slides](#)

Geospatial Data Management in Apache Spark: A Tutorial

ICDE 2019

Macau, China

Apr. 2019

- [Tutorial website](#)

- Spatial Data Management in Apache Spark - The GeoSpark Perspective and Beyond** *Arizona, U.S.A*
 NVIDIA *Aug. 2018*
 • [Talk slides](#)
- Code-generation for Fast Queries on Compressed Data** *San Jose, California, U.S.A*
 IBM ALMADEN RESEARCH CENTER *Aug. 2018*
 • Slides not available due to NDA
- Deploy Distributed Database Course Project on Vocareum** *Arizona, U.S.A*
 VOCAREUM *Feb. 2018*
 • [Talk slides](#)
- Geospatial Visual Analytics belongs to Database Systems** *Redondo Beach, California, U.S.A*
 ACM SIGSPATIAL 2017 STUDENT RESEARCH COMPETITION *Nov. 2017*
 • [Conference presentation slides](#)
- Interactive and Scalable Exploration of Geospatial Data** *Arizona, U.S.A*
 STATE FARM *Sept. 2017*
 • [Talk slides](#)
- Two Birds, One Stone: A Fast, yet Lightweight, Indexing Scheme for Modern Database Systems** *Munich, Germany*
 VLDB 2017 *Aug. 2017*
 • [Conference presentation slides](#)
- Indexing the Pickup and Drop-off Locations of NYC Taxi Trips in PostgreSQL – Lessons from the Road** *Washington D.C., U.S.A*
 SSTD 2017 *Aug. 2017*
 • [Conference presentation slides](#)
- Data affinity for computation on Spark and Mesos** *Cupertino, California, U.S.A*
 APPLE *Aug. 2016*
 • Slides not available due to NDA
- GeoSpark: A Cluster Computing Framework for Processing Large-Scale Spatial Data** *Seattle, Washington, U.S.A*
 ACM SIGSPATIAL 2015 FAST FORWARD SESSION *Nov. 2015*
 • [Conference presentation slides](#)

References

Mohamed Sarwat

Assistant Professor, Computer Science Department, Arizona State University

faculty.engineering.asu.edu/sarwat

msarwat@asu.edu (**send letter requests to: send.Sarwat.018A1FCF78@interfolio.com**)

699 South Mill Avenue, Tempe, Arizona 85281

Yingjun Wu

Software Engineer, Redshift Team, Amazon Web Services

(former researcher, Database Group, IBM Almaden Research Center)

yingjunwu.github.io

yjwu@amazon.com (**send letter requests to: send.Wu.E08E714A6A@interfolio.com**)

2100 University Avenue, East Palo Alto, California 94303

David Lomet

Principal Researcher, Database Group, Microsoft Research

Fellow of ACM, IEEE, and AAAS

Member of the National Academy of Engineering

www.microsoft.com/en-us/research/people/lomet

lomet@microsoft.com (**send letter requests to: send.Lomet.E8E3605F0D@interfolio.com**)

One Microsoft Way, Redmond, Washington 98052

Umar Farooq Minhas

Principal Researcher, Database Group, Microsoft Research

www.microsoft.com/en-us/research/people/ufminhas

ufminhas@microsoft.com (**send letter requests to: send.Minhas.B0B972C660@interfolio.com**)

One Microsoft Way, Redmond, Washington 98052