Research Statement

The vision of my research is to build systems that improve developer productivity through automated debugging and testing of big data analytics. Broadly, I am interested in designing novel tool support for data-centric software development. My approach combines insights from software engineering, distributed systems, and databases. I have also been awarded the Google Ph.D Fellowship in the sub-field of software engineering.

Education

- **PhD in Computer Science**  
  University of California, Los Angeles  
  Advised by Miryung Kim  
  Thesis: Interactive and Automated Debugging for Big Data Analytics  
  2014 – Present

- **B.S in Computer Science**  
  Lahore University of Management Sciences  
  2010 – 2014

Research and Work Experience

- **Software Engineer, Intern**  
  Google Inc. Mountain View  
  Jun ’19 – Sept 19

- **Software Engineer Tools and Infrastructure Intern**  
  Google Inc. Mountain View  
  Jun ’18 – Sept 18

- **Summer Research Assistant**  
  NEC Labs America, Princeton NJ  
  Jun ’16 – Sep ’16

- **Research Intern**  
  Koç University, Turkey  
  May ’12 – Aug ’12

Awards and Grants

- Google Ph.D Fellowship 2017-20 (with 3rd year extension)
- NSF I-Corps Grant (Entrepreneurial Lead) 2018
- Gold medal at ACM Student Research Competition at ICSE 2018
- SIGMOD Student Travel Award 2017
- SoCC Student Travel Award 2017 and 2019
- Graduation with Distinction, Dean’s Honor List Award, National Mathematics Olympiad Finalist

Research Track Publications

- **[P1] White-Box Testing of Big Data Analytics with Complex User-Defined Functions**  
  Muhammad Ali Gulzar, Shaghayegh Mardani, Madan Musuvathi, Miryung Kim  
  ESEC/FSE’19

  Jason Teoh, Muhammad Ali Gulzar, Harry Xu, Miryung Kim.  
  SoCC’19

- **[P3] LogLens: A Real-Time Log Analysis System**  
  Biplob Debnath, Mohiuddin Solaimani, Muhammad Ali Gulzar, Nipun Arora, Cristian Lumezanu, Jianwu Xu, Bo Zong, Hui Zhang, Guofei Jiang, Latifur Khan.  
  IEEE 38th International Conference on Distributed Computing Systems. 11 Pages.  
  ICDCS’18

- **[P4] Adding data provenance support to Apache Spark**  
  Matteo Interlandi, Ari Ekmejki, Kshitij Shah, Muhammad Ali Gulzar, Sai Tetali, Miryung Kim, Todd Millstein, Tyson Condie.  
  The VLDB Journal (Special Issue on “The Best Papers of” VLDB 2018). 21 Pages.  
  VLDB Journal’18
Automated Debugging in Data-Intensive Scalable Computing
Muhammad Ali Gulzar, Matteo Interlandi, Xueyuan Han, Mingda Li, Tyson Condie, Miryung Kim.
Acceptance Rate: 23.6%

BigDebug: Debugging Primitives for Interactive Big Data Processing in Spark
Muhammad Ali Gulzar, Matteo Interlandi, Seunghyun Yoo, Sai Tetali, Tyson Condie, Todd Millstein, Miryung Kim.
Acceptance Rate: 19%
The proceedings of the 38th International Conference on Software Engineering. 12 Pages.

Optimizing Interactive Development of Data-Intensive Applications
Matteo Interlandi, Sai Tetali, Muhammad Ali Gulzar, Joseph Noor, Tyson Condie, Miryung Kim, Todd Millstein.
Acceptance Rate: 25.1%

Titian: Data Provenance Support in Spark
Matteo Interlandi, Kshitij Shah, Sai Tetali, Muhammad Ali Gulzar, Seunghyun Yoo, Miryung Kim, Todd Millstein, Tyson Condie.
Acceptance Rate: 21.2%

Demonstration Track Publications

BigSift: Automated Debugging of Big Data Analytics in Data-intensive Scalable Computing
Muhammad Ali Gulzar, Siman Wang, Miryung Kim.
Acceptance Rate: 38.8%
The 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, Research Demonstration Track. 4 Pages.

Debugging Big Data Analytics in Spark with BigDebug
Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim.
Acceptance Rate: 34%
The Proceedings of The 2017 ACM SIGMOD/PODS Conference Demonstration Track. 4 Pages.

BigDebug: Interactive Debugger for Big Data Analytics in Apache Spark
Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim.
Acceptance Rate: 40%
The proceedings of the 24th Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Research Demonstration Track. 5 Pages.

Short/Workshop/Industry Track Publications

Perception and Practices of Differential Testing
Muhammad Ali Gulzar, Yongkong Zhu, Xiaofeng Han.
Acceptance Rate: 22.2%
The proceedings of the 41th International Conference on Software Engineering Software Engineering in Practice Track. 10 Pages.

Interactive Debugging for Big Data Analytics
Muhammad Ali Gulzar, Xueyuan Han, Matteo Interlandi, Shaghayegh Mardani, Sai Tetali, Tyson Condie, Todd Millstein, Miryung Kim.
Acceptance Rate: 30.8%
The 8th USENIX Workshop on Hot Topics in Cloud Computing. 7 Pages.

Teaching Experience

CS130 Software Engineering
University of California, Los Angeles
Spring ‘17 and Fall ‘15

CS282 Network Centric Computing
Lahore University of Management Sciences
Spring ‘14

CS310 Algorithms
Lahore University of Management Sciences
Spring ‘13

CS340 Databases
Lahore University of Management Sciences
Fall ‘13
Mentoring Experience

Xueyuan Han  *(Undergraduate at UCLA. Currently a PhD student at Harvard University)*  
Simon Wang  *(Research Intern (CSST) at UCLA. Currently a Masters student at UCSD)*  
Jason Teoh  *(PhD student at UCLA)*  
Fabrice Harel-Canada  *(PhD student at UCLA)*  
Usama Hameed  *(PhD student at UCLA)*  

Jan ’16 - May ’17

Service

Journal Reviewer:  Transaction on Software Engineering (TSE 2018)
Student Volunteer:  Co-lead Student Volunteer ICSE 2016
Mentor:  Incoming M.S. Computer Science Class in 2015
Panelist:  UCLA Computer Science PhD Open-house year 2017 and 2018
Research Host:  UCLA Cross-disciplinary Scholars in Science and Technology 2017

Invited Talks

- White-Box Testing of Big Data Analytics  
  Google PhD Fellowship Summit ’19
- Automated Debugging and Testing of Big Data Analytics  
  ICSE’18
- Automated Debugging in Data-Intensive Scalable Computing  
  FSE’18, SoCC’17
- Debugging Big Data Analytics in Spark  
  Spark Summit’17
- Debugging Big Data Analytics with BigDebug  
  SIGMOD’17
- Interactive Debugger for Big Data Analytics  
  FSE’16
- BigDebug: Debugging Primitives for Interactive Big Data Processing in Spark  
  ICSE’16
- Interactive Debugging for Big Data Analytics  
  NEC Labs America, 2016
- Towards Big Data Debugging in Apache Spark  
  Databricks Inc., 2015

Academic References

- **Miryung Kim**  
  Professor of Computer Science  
  University of California, Los Angeles  
  miryung@cs.ucla.edu
- **Harry Xu**  
  Associate Professor of Computer Science  
  University of California, Los Angeles  
  harryxu@cs.ucla.edu
- **Todd Millstein**  
  Professor of Computer Science  
  University of California, Los Angeles  
  todd@cs.ucla.edu
- **Madan Musuvathi**  
  Partner Research Manager  
  Microsoft Research  
  madanm@microsoft.com