

# MUHAMMAD ALI GULZAR

Suite 2224, KWII  
Blacksburg, VA, 24060

Assistant Professor  
Department of Computer Science  
Virginia Tech

[gulzar@cs.vt.edu](mailto:gulzar@cs.vt.edu)  
(540) 231 0851

## 1 Research Interests

My research vision is to build systems that improve developer productivity through automated debugging and testing for emerging applications. My research delivers integral insights to transfer traditional software engineering techniques to the next generation software in emerging domains such as ML/AI, computational/data science, web, and big data.

## 2 Education

<b>Ph.D.</b>	University of California, Los Angeles Computer Science <i>Interactive and Automated Debugging for Big Data Analytics</i>	2014 - 2020
<b>B.S.</b>	Lahore University of Management Sciences Computer Science	2010 - 2014

## 3 Employment History

08/2020 - Current	<b>Assistant Professor</b> , Computer Science, Virginia Tech
06/2019 - 09/2019	<b>Software Engineering Intern</b> , Google, Inc.
06/2018 - 09/2018	<b>Software Engineering Tools and Infrastructure Intern</b> , Google, Inc.
06/2019 - 09/2016	<b>Summer Research Assistant</b> , NEC Labs America, Princeton, NJ.

## 4 Awards and Grants

1. Rising Start Faculty Award, Computer Science Department, Virginia Tech	2022
2. NSF Medium:Reinventing Fuzz Testing for Data and Compute Intensive Systems Total. \$900. Sole PI at VT. VT's Share: \$323,982	2021 - 2025
3. Outstanding Computer Science Graduating Ph.D. Student	2020
4. Northrop Grumman Outstanding Computer Science Graduate Student Research Award	2020
5. NSF i-Corps, Amount: \$50K, Role: Entrepreneurial Lead	2018
6. Google Ph.D. Fellowship 2017-20 (with 3rd-year extension)	2017 - 2020
7. Gold medal at ACM Student Research Competition at ICSE'18	2018

## 5 Professional Activity

### Journal Reviewer

IEEE Transaction on Software Engineering, TSE (since 2018)  
ACM Transactions on Software Engineering and Methodology, TOSEM (since 2019)  
IEEE Transactions on Dependable and Secure Computing, TDSC (since 2021)

### Program Committee

ACM European Software Engineering and Foundations of Software Engineering (ESEC/FSE '22) 2022  
ACM European Software Engineering and Foundations of Software Engineering Demonstration Track (ESEC/FSE '22) 2022  
International Workshop on Testing Database Systems (DBTest '22) co-located with SIGMOD 2022

International Conference on AI Engineering – SE for AI (CAIN'22) co-located with ICSE 2022  
ACM Symposium of Cloud Computing (SoCC '22 ) 2022  
The Theory and Practice of Provenance (TaPP'21) 2021  
ACM Symposium of Cloud Computing (SoCC '21 ) 2021

### Grant Proposal Panels

NSF 2022  
The Department of Energy (DOE) (2021)

## 6 Teaching Activity

CS 3304 **Comparative Languages**, Computer Science, Virginia Tech Spring '22.  
CS 6704 **Advanced Topic on Software Engineering**, Computer Science, Virginia Tech Fall '21.  
CS 5614 **Big Data Engineering**, Computer Science, Virginia Tech Spring '21.  
CS 130 **Software Engineering, TA**, Computer Science, UCLA Spring '17 and Fall '15.

## 7 Selected Publications

H-index 12 , Citations 493 (as of 6/14/2022)

### 7.1 Research Track Publications

1. Detecting Build Conflicts in Software Merge for Java Programs via Static Analysis. Sheikh Towqir, Bowen Shen, Muhammad Ali Gulzar, and Na Meng. The 37th IEEE/ACM International Conference on Automated Software Engineering. **ASE 2022**.
2. A Characterization Study of Merge Conflicts in Java Projects. Bowen Shen, Muhammad Ali Gulzar, Fei He, Na Meng. Transactions on Software Engineering and Methodology. **TOSEM 2022**.
3. Sibylvariant Transformations for Robust Text Classification. Fabrice Harel-Canada, Muhammad Ali Gulzar, Nanyun Peng, Miryung Kim. The Proceedings of the Conference of the 60th Annual Meeting of the Association for Computational Linguistics **ACL Findings 2022**.
4. TrackerSift: Untangling Mixed Tracking and Functional Web Resources. Abdul Haddi Amjad, Danial Saleem, Muhammad Ali Gulzar, Zubair Shafiq, Fareed Zaffar. The proceedings of the 21st ACM Internet Measurement Conference. **ACM IMC 2022**.
5. OptDebug: Fault-Inducing Operation Isolation for Dataflow Applications. Muhammad Ali Gulzar, Miryung Kim. The proceedings of the 12th ACM Symposium on Cloud Computing. **SoCC 2021**.
6. Is neuron coverage a meaningful measure for testing deep neural networks?. Fabrice Harel-Canada, Lingxiao Wang, Muhammad Ali Gulzar, Quanquan Gu, Miryung Kim. The proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. **ESEC/FSE 2020**
7. BigFuzz: efficient fuzz testing for data analytics using framework abstraction. Qian Zhang, Jiyuan Wang, Muhammad Ali Gulzar, Rohan Padhye, Miryung Kim. The proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. **ASE 2020**
8. Influence-based provenance for dataflow applications with taint propagation. Jason Teoh, Muhammad Ali Gulzar, and Miryung Kim. The proceedings of the 11th ACM Symposium on Cloud Computing. **SoCC 2019**
9. HeteroRefactor: Refactoring for Heterogeneous Computing with FPGA. Jason Lau\*, Aishwarya Sivaraman\*, Qian Zhang\*, Muhammad Ali Gulzar Jason Cong, Miryung Kim. The proceedings of the 42nd International Conference on Software Engineering. **ICSE 2020**
10. White-Box Testing of Big Data Analytics with Complex User-Defined Functions. Muhammad Ali Gulzar, Shaghayegh Mardani, Madan Musuvathi, Miryung Kim. The proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. **ESEC/FSE 2019**.
11. PerfDebug: Performance Debugging of Computation Skew in Dataflow Systems. Jason Teoh, Muhammad Ali Gulzar, Harry Xu, Miryung Kim. The proceedings of the ACM Symposium on

Cloud Computing. **SoCC 2019**

12. 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. 2018 IEEE 38th International Conference on Distributed Computing Systems. **ICDCS 2018**.
13. Adding data provenance support to Apache Spark. Matteo Interlandi, Ari Ekmekji, Kshitij Shah, Muhammad Ali Gulzar, Sai Tetali, Miryung Kim, Todd Millstein, Tyson Condie. The VLDB Journal — The International Journal on Very Large Data Bases. **VLDB Journal 2018**.
14. Automated Debugging in Data-Intensive Scalable Computing. Muhammad Ali Gulzar, Matteo Interlandi, Xueyuan Han, Mingda Li, Tyson Condie, Miryung Kim. The proceedings of the 2017 Symposium on Cloud Computing. **SoCC 2017**.
15. 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. The proceedings of the 38th International Conference on Software Engineering. **ICSE 2016**.
16. Optimizing Interactive Development of Data-Intensive Applications. Matteo Interlandi, Sai Tetali, Muhammad Ali Gulzar, Joseph Noor, Tyson Condie, Miryung Kim, Todd Millstein. The proceedings of the Seventh ACM Symposium on Cloud Computing. **SoCC 2016**.
17. Titian: Data Provenance Support in Spark. Matteo Interlandi, Kshitij Shah, Sai Tetali, Muhammad Ali Gulzar, Seunghyun Yoo, Miryung Kim, Todd Millstein, Tyson Condie. The proceedings of the VLDB Endowment, Volume 9, Issue 3. **VLDB 2016**.

## 7.2 Workshops

1. Towards a Serverless Bioinformatics Cyberinfrastructure Pipeline. Shunyu David Yao, Muhammad Ali Gulzar, Liqing Zhang, and Ali R. Butt. In Proceedings of the 1st Workshop on High Performance Serverless Computing. **HiPS 2021**.
2. Perception and Practices of Differential Testing. Muhammad Ali Gulzar, Yongkang Zhu, and Xiaofeng Han. In Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice. **ICSE SEIP 2019**.
3. Interactive Debugging for Big Data Analytics. Muhammad Ali Gulzar, Xueyuan Han, Matteo Interlandi, Shaghayegh Mardani, Sai Deep Tetali, Todd Millstein, and Miryung Kim. In 8th USENIX Workshop on Hot Topics in Cloud Computing. **HotCloud 16**.

## 7.3 Demonstration Track

1. BigTest: Symbolic Execution Based Systematic Test Generation for Apache Spark. Muhammad Ali Gulzar, Madan Musuvathi, Miryung Kim. The 42nd International Conference on Software Engineering. Research Demonstration Track. **ICSE Demo 2020**.
2. BigSift: Automated Debugging of Big Data Analytics in Data-intensive Scalable Computing. Muhammad Ali Gulzar, Siman Wang, Miryung Kim. The 26th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Research Demonstration Track. **ESEC/FSE Demo 2018**.
3. Debugging Big Data Analytics in Spark with BigDebug Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim. The Proceedings of The 2017 ACM SIGMOD/PODS Conference. Demonstration Track. **SIGMOD Demo 2017**.
4. BigDebug: Interactive Debugger for Big Data Analytics in Apache Spark Muhammad Ali Gulzar, Matteo Interlandi, Tyson Condie, Miryung Kim. The proceedings of the 24th Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. Research Demonstration Track. **ESEC/FSE Demo 2016**.

## 7.4 Patent

1. Systems and Methods with a Realtime Log Analysis Framework. Biplob Debnath, Nipun Arora, Hui Zhang, Guofei Jiang, Mohiuddin Solaimani, Muhammad Ali Gulzar. US Patent App. 15/784,393. 2018.

## 8 Student Advising

### Current Students

1. Abdul Hadi Amjad, Ph.D. Student at VT. Joined Fall 2021
2. Waris Gill, Ph.D. Student at VT. Joined Fall 2021
3. Ahmad Humayun, Ph.D. Student at VT. Joined Fall 2021
4. Ming Chen, M.S Student at VT. Joined Spring 2022
5. Sabaat Haroon, Ph.D. Student at VT. Joined Fall 2022

### Graduated Students

1. Satish Venkatesan, VT. MS Thesis: Automatic Restoration and Management of Computational Notebooks, First Job: Salesforce. Graduated March 2021