Wei Zhang

2415 Rosy Billed Dr. APT 208
Charlotte, NC, USA, 28262
Mobile: +1 (806) 786 - 9088
Email:zhangwei217245@gmail.com
Website:https://zhangwei217245.wixsite.com/curriculumvitae

Final-year Ph.D Student
Top-tier publications in CS
7+ Years industry experience
High-Performance Computing
Distributed Computing,
Big Data

Links

Github:// **zhangwei217245**
LinkedIn:// **zhangwei217245**
ORCID:// **0000-0002-6921-4926**

Skills

OS
Linux, Unix, Windows

Languages
C, Java, Python, Scala, NodeJS, Bash, Rust, C++

Server-side Development
Spring, JPA, RESTful web service

Databases
RDBMS: MySQL, Oracle, SQL Server,
NoSQL: Memcache, Redis,
HBase, MongoDB, Neo4J

Data Mining
Spark, Hadoop, MPI, Numpy,
SciPy, Matplotlib

Software Engineering
Design Patterns, UML,
Continuous Integration

Others
AWS & GCP experience, K8s,
HTML, JavaScript, CSS, XML,
XSTL, AJAX

Coursework

Analysis of Algorithm
Parallel Computing
Distributed Computing
Advanced Operating System
Big Data Management
Information Retrieval
Pattern Recognition
Machine Learning

Education

2016-NOW
Ph.D in Computer Science
Texas Tech University, USA

2003-2007
BSc in Computer Science
Hebei University of Science and Technology, China

Experience

2017-NOW **DISCL @ Texas Tech University** — **Research Assistant**

- Major Research Focus: metadata indexing and querying over massive data collections in a distributed setting featuring networked computing nodes, interconnected CPU cores, and heterogeneous storage devices.
- Related Research Topics: in-memory data indexing, distributed data indexing, graph-based metadata management, graph partitioning, performance optimization, load balancing, etc.
- Research Outcomes: Papers published in HPDC'17, CCGrid '18, PACT '18, SC '19; Software release: https://bitbucket.org/berkeleylab/miqs/src/master/
- Collaborator and Research Fund Support: Lawrence Berkeley National Laboratory.

C, Numpy, SciPy, Rust, MongoDB

2016-2017 **STAR Lab @ Texas Tech University** — **Research Assistant**

- Project Goal: Geo-spatial data mining over geo-tagged data mining.
- Major Contribution: Setting up computing platform (HDFS + Spark), conducting geo-spatial data mining using sentimental analysis and geo-spatial clustering.
- Outcomes: Papers published as a co-author in JAG, TGRS, ISPRS. Code base: https://github.com/zhangwei217245/GeoAnalysis.

Spark, Scala, HDFS, Node.js, Python, GDAL, Redis

2014-2016 **Beijing Serious Technology Co., Ltd.** — **Senior System R&D Engineer**

- Major Contribution: Designed and developed the backbone of the back-end data service architecture featuring performance, availability and scalability for Enjoy!
- Code Base: https://github.com/zhangwei217245/Lego.

Memcache, Redis, MySQL, RabbitMQ, Java, Spring

2010-2013 **Sina.com Technology (China) Co.,Ltd.** — **System R&D Engineer**

- Weibo User Data Service: Designed and developed the second generation of the user data service of Weibo.com, serving 2.2 trillion active users each day (back in 2013).
- T.cn URL Shorten Service: Designed and developed the URL shorten service T.cn serving millions of transactions per minute.

Memcache, Redis, MySQL, Java, Spring, Jersey (JAX-RS)

Side Projects

2019 **ActiveDR** — **Data Analysis**

An activeness-based data retention solution for HPC storage. Paper is under review. See ActiveDR for details.

2016 **SqeezeItBoard** — **AI, Board Game**

An interesting board game developed using JavaFX and Java for the AI course project. Try SqeezeItBoard for detail!

Selected Publications

2019 W.Zhang, S.Byna, H.Tang, B. Williams, Y.Chen. MIQS: Metadata Indexing and Querying Service for Self-describing File Formats. Accepted to appear in The Proceedings of The 31st ACM/IEEE Supercomputing Conference (SC'19), Denver, CO, 2019. (first-around acceptance rate: 72/344=21%, another 15 papers being asked for major revisions per SC'19) [Paper Link]

2018 W.Zhang, H.Tang, S.Byna, Y.Chen. DART: Distributed Adaptive Radix Tree for Efficient Affix-based Keyword Search on HPC Systems. In The Proceedings of The 27th International Conference on Parallel Architectures and Compilation Techniques (PACT'18), 2018. (acceptance rate: 36/126=28.6%) [Paper Link]

2018 W.Zhang, Y.Chen, D.Dai. AKIN: A Streaming Graph Partitioning Algorithm for Distributed Graph Storage Systems. In The Proceedings of The 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '18). (acceptance rate: 20.8%). [Paper Link]

Please refer to my Google Scholar profile for my other publications