KEWEN WANG

Email: wangkewen001@gmail.com Website: http://wangkewen.github.io
Address: Redmond, WA

EDUCATION

University of Connecticut *2014 -2020*
Ph.D. in Computer Science, GPA: 4.0
Beihang University *2010 -2013*
M.S. in Computer Science, GPA: 3.3
Beijing Information Science and Technology University *2005 -2009*
B.S. in Computer Science, GPA: 3.5

TECHNICAL SKILLS

Computer Languages	Java, Scala, Python, Linux Shell, C, Go
Open Source	Apache Spark, Apache Hadoop, Apache Mesos, Aurora, MySQL, Postgres
Web Development	ReactJS, Node.js, JavaScript, Apache Tomcat

CODING COMPETITION

Google Code Jam 2022	Qualification Round Rank #1203/32k, Round 1B Rank #2971/11k
Topcoder Open 2020	Advance to Round 3, Rank #109/238, T-shirt Award
Google Kickstart 2018	Round B Rank #122/753
Google Code Jam 2017	Qualification Round Rank#1483/25k, Round 1C Rank#1664/3775

RESEARCH PROJECTS

Performance Prediction and Improvement for Apache Spark Jobs Aug 2014 - May 2019
Research Assistant *University of Connecticut*

· Developed a Spark analytics system in Java to parse JSON logs of Apache Spark event, and predict time, I/O overhead, memory consumption using analytical approaches.
· Developed a dynamical job predictor in Java to predict the execution time of multiple Spark jobs in Xen, and implemented a job scheduler in Java and Bash to reduce the total execution time.
· Implemented a Spark optimizer in Java to predict and mitigate potential task stragglers and skewed task distribution problems for Apache Spark platform to improve job performance.
· Designed and implemented a middleware to dynamically allocate computing resources for Apache Spark applications to improve resource utilization.

Optimizing Hadoop MapReduce Nov 2011 - Dec 2012
Research Assistant *Beihang University*

· Applied BTrace to trace MapReduce job functions, and monitor resource consumption using Ganglia.
· Implemented a MapReduce optimizer in Java through constructing Hadoop performance model for execution time prediction and designing heuristic search algorithm to find near optimal configurations for MapReduce jobs.

WORK EXPERIENCE

Senior Software Engineer Feb 2022 - Jan 2025
Salesforce. Seattle, WA

· Distributed systems. Spark Structured Streaming.

Software Engineer II July 2019 - Feb 2022
Oscar Health. New York, NY

· Working in Engineering Effectiveness team.
· Building platform services.

Research Intern May 2018 - Aug 2018
HashiCorp. San Francisco, CA

· Developed a system performance predictor in Python for Consul cluster workload prediction using Machine Learning algorithms such as SVM, Random Forest, Gradient Boosting Tree.
· Implemented server buffer in Go to improve cluster stability and reduce response latency and failure.

Software Engineer Intern Mar 2010 - May 2010
NDtech Inc. Beijing, China

· Analyzed ANTLR (an open source parser generator) to learn C# parser and Script#.
· Applied Script# to write JavaScript using C#.

AWARDS

Salesforce Astro Award *2024*
Salesforce, Inc
FGCS Outstanding Reviewer *2023*
The Future Generation Computer Systems Journal
ISPASS Student Travel Grants *2019*
IEEE International Symposium on Performance Analysis of Systems and Software
Predoctoral Fellowship *2017*
Computer Science and Engineering department at University of Connecticut
Third Class Scholarship *2011*
Beihang University
Academic Scholarship *2008*
Beijing Information Science and Technology University
Municipal 2nd Prize of 21st National Middle School Students Physics Competition *2004*
City of Xianning, China

PUBLICATIONS

1. An Adaptive Graph Sampling Framework for Graph Analytics. Wang, Kewen. Social Network Analysis and Mining (2023).
2. A Dynamic Resource Allocation Framework for Apache Spark Applications. Wang, Kewen, Mohammad Maifi Hasan Khan, and Nhan Nguyen. 2020 IEEE 44th Annual Computer Software and Applications Conference (COMPSAC).
3. A Model Driven Approach towards Improving the Performance of Apache Spark Applications. Wang, Kewen, Mohammad Maifi Hasan Khan, Nhan Nguyen, and Swapna Gokhale. 2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS).
4. Modeling Interference for Apache Spark Jobs. Wang, Kewen, Mohammad Maifi Hasan Khan, Nhan Nguyen, and Swapna Gokhale. IEEE 9th International Conference on Cloud Computing (CLOUD), 2016.

5. Performance Prediction for Apache Spark Platform. Wang, Kewen, Mohammad Maifi Hasan Khan. IEEE 17th International Conference on High Performance and Communications (HPCC), 2015.

6. Predator - An experience guided configuration optimizer for Hadoop MapReduce. Wang, Kewen, Xuelian Lin, and Wenzhong Tang. IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), 2012.