

KEWEN WANG

Email: wangkewen001@gmail.com Website: <http://wangkewen.github.io>

Address: Storrs Mansfield, CT 06269

EDUCATION

University of Connecticut 2014 -2019

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, Python, Linux Shell, C, Go

Open Source Apache Spark, Apache Hadoop, MySQL, Xen, BTrace, Ganglia

Web Development JSP, Ajax, CSS, JavaScript, jQuery, Apache Tomcat

CODING COMPETITION

Google Code Jam 2017 Qualification Round Rank#1483/25k, Round 1C Rank#1664/3775

Google Kickstart 2018 Round B Rank #122/753

ACADEMIC SERVICE

Reviewer of IEEE Transactions on Parallel and Distributed Systems (TPDS).

Reviewer of ACM Transactions on Architecture and Code Optimization (TACO).

Reviewer of Journal of Experimental and Theoretical Artificial Intelligence (JETAI).

RESEARCH PROJECTS

Performance Prediction and Improvement for Apache Spark Jobs Oct 2014 - May 2018

Research Assistant (Java, Spark, ML, Bash) *University of Connecticut*

- Developed a tool in Java to parse json logs of Apache Spark event, predict time, I/O overhead and memory consumption using analytical and Machine Learning approaches.
- Developed a tool 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 by up to 71 %.

Learning Environment for Smart Grid Security

Aug 2013 - Feb 2014

Research Assistant (JSP, MySQL, jQuery, Bash)

Georgia State University

- Implemented an online tool using JSP, jQuery and MySQL to schedule Smart Grid emulator.
- Implemented course information display and scheduling calendar using JSP and jQuery, developed Java Servlets and Bash scripts to launch and control Smart Grid emulator, loaded user profile and course schedule information into MySQL database.

Optimizing Hadoop MapReduce

Nov 2011 - Dec 2012

Research Assistant (Java, Hadoop, Bash)

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

Research Intern (Python, ML, Bash, Go)
HashiCorp. San Francisco, CA

May 2018 - Aug 2018

- Developed a performance model in Python for Consul cluster workloads 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.

Full Stack Developer (JSP, JavaScript, MySQL, Bash)
Institute of Science and Technology at Beihang University

Nov 2011 - Jan 2012

- Designed and implemented a website on Struts+Spring+Hibernate framework.
- Implemented information retrieval and display using JSP, JavaScript and Ajax, and loaded project archives into MySQL database
- Implemented Java Servlets and filters for service actions (such as information update, remove) and security validation (such as user privilege), applied Apache Tomcat as the web server in Linux.

Software Engineer Intern (JavaScript, C#)
NDtech Inc. Beijing, China

Mar 2010 - May 2010

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

AWARDS

Predoctoral Fellowship Computer Science and Engineering department at University of Connecticut	<i>2017</i>
Third Class Scholarship Beihang University	<i>2011</i>
Academic Scholarship Beijing Information Science and Technology University	<i>2008</i>
Municipal 2nd Prize of 21st National Middle School Students Physics Competition City of Xianning, China	<i>2004</i>

PUBLICATIONS

1. 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).
2. Design and implementation of an analytical framework for interference aware job scheduling on Apache Spark platform. Wang, Kewen, Mohammad Maifi Hasan Khan, Nhan Nguyen, and Swapna Gokhale. Cluster Computing (2017).
3. 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.
4. Performance Prediction for Apache Spark Platform. Wang, Kewen, Mohammad Maifi Hasan Khan. IEEE 17th International Conference on High Performance and Communications (HPCC), 2015.
5. 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.