# 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 |
| **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

**Improving Performance of Apache Spark Jobs** Aug 2015 - May 2018
*Research Assistant* *University of Connecticut*

· Developed an interference model to predict the execution time of multiple Spark jobs, and implemented an interference aware job scheduler to reduce the total execution time.
· Implemented a framework to predict and mitigate potential task stragglers and skewed task distribution problems for Apache Spark platform.

**Performance Prediction for Apache Spark Jobs** Oct 2014 - May 2015
*Research Assistant* *University of Connecticut*

· Parsed JSON format event logs of Apache Spark jobs, and analyzed task execution pattern.
· Developed an analytical performance model to predict time, I/O overhead and memory consumption.

**Learning Environment for Smart Grid Security** Aug 2013 - Feb 2014
*Research Assistant* *Georgia State University*

· Implemented an online tool using JSP, jQuery and MySQL to schedule Smart Grid emulator.

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

· Applied BTrace to trace MapReduce job functions, and monitor resource consumption using Ganglia.
· Constructed Hadoop performance model for execution time prediction.
· Designed heuristic search algorithm to find near optimal configurations for MapReduce jobs.

## WORK EXPERIENCE

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

· Developed a performance model for Consul cluster workloads using Machine Learning algorithms.
· Implemented server buffer to improve cluster stability and performance.

**Full Stack Developer** Nov 2011 - Jan 2012
**Institute of Science and Technology at Beihang University**

· Designed and implemented a website on Struts+Spring+Hibernate framework.
· Implemented information retrieval and display framework using JSP, MySQL, JavaScript and Ajax.

**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

**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. 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. CSMiner: An Automated Tool for Analyzing Changes in Configuration Settings across Multiple Versions of Large Scale Cloud Software. Nguyen, Nhan, Mohammad Maifi Hasan Khan, and Kewen Wang. 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.