Wang, Daoyuan

Cellphone:*+86-152-1686-1267*
E-mail:*me at daoyuan.wang*

EDUCATION

Bachelor of Engineering, Zhejiang University, Hangzhou, China Oct 2009 to Jun 2013
Enrolled in HE Zhijun Honored Class and graduated with a HE Zhijun Certification.
Won *Student of Excellence* in 2011-2012.
Involved in GIVE Lab during undergradute. Major: Computer Science
Overall GPA $3.83/4$

WORK EXPERIENCE

Senior Software Engineer at Intel Jul 2013 to Now

After Daoyuan's graduation, he joined Big Data Technologies organization of Intel SSG. He works on optimization of Hadoop/Spark eco-system software, developing new features and improving performance for typical workloads.

- *Project Panthera ASE* Jul 2013 to Feb 2014
  An approach to support PL/SQL on Apache HIVE, to provide support for legacy queries with Intel Distribution for Hadoop(*IDH). The project was developed by three engineers including Daoyuan, source code is available at Github. Daoyuan redesigned the basic processing structure of co-related subquery un-nesting, improving the passing rate of queries a lot. He committed a lot of code to this project involving bug fixing and new features such as error tracing. It was part of Intel Distribution for Hadoop(IDH).*
- *HiBench March 2014 to Dec 2014*
  Originated in Big Data Technologies, HiBench is a widely adopted big data benchmark suite in the industry. Before Daoyuan's contributions, HiBench could only work on MapReduce v1 of Apache Hadoop. Daoyuan fixed a lot of compatibility issues for different Hadoop versions, and finally made it support MapReduce v2 of Apache Hadoop (Hadoop Yarn), with the official release of HiBench 3.0. He then maintained this project till the end of 2014, answering question from community. This project is open-source, available at Github.
- *Apache Spark Apr 2014 to Now*
  Apache Spark is a fast and general engine for large-scale data processing. Apache Spark has become a widely used platform for computing in both clusters and clouds. Daoyuan contributed a lot of code to Apache Spark, mainly focused on Spark SQL.
- *Intel OAP (codename: Spinach) Feb 2017 to Now*
  OAP is the efforts from Intel BDT for ad-hoc queries support on Spark SQL. A lot of companies have adopted Apache Spark as their default analysis engine for large volume of data, and data scientist may require the analysis platform as a real-time query engine. While Spark is not designed for ad-hoc queries, OAP is the spark package to unlock the power of hardware, accelerating query execution using mechanism like indexing and caching. Baidu, one of the largest internet companies in the world, has been using OAP in their real-time analysis engine in production for advertisement strategies. And they reported a 1.5x-5x performance gain using OAP. This project is open-sourced from Jun 2017. Daoyuan is a core developer of this project.

Software Engineer Intern at Intel Sept 2012 to Dec 2012

Daoyuan joined Intel IT in 2012 as software engineer intern. He worked on the project of TAS(Transcode as a service) on Hadoop, providing transcoding service using Hadoop platform and Intel hardware transcoding technologies. He successfully traced down a BSOD bug, which was later identified as a bug in the driver program of

Intel Graphics. He also developed a thorough test framework for the project, to enable automatic regression tests.

- Won 2nd place on Intel® SWPC China 2012.

COMPUTER SKILLS

◇ *Languages & Software:* JAVA, Scala, Python, C/C++, C#, PHP, SQL, R, Ruby, BASIC, Assembly, JavaScript, Linux Shell, Windows Shell, Verilog HDL, Matlab, OpenCV, OpenGL, OpenMP, LaTeX, MS Office.

STANDARD TESTS

TOEFL-iBT	105	Nov 2017
GRE	$152 + 170 + 3$	Nov 2017

AWARDS

- Third prize(the 19th place) in the 11th Zhejiang University Programming Contests
- Third prize(the 10th place) in the 12th Zhejiang University Programming Contests
- Intel SSG 2017 Q3 Group Recognition Award

TRANSLATION

Learning Spark: Lightning-fast Data Analysis (Chinese version) O'Reilly

PUBLIC TALKS

Tuning Garbage Collection for Spark Applications
QCon Shanghai 2015, Shanghai, Oct 2015

Spinach: Run Ad-hoc Queries on top of Spark SQL
DTCC 2017, Beijing, May 2017

OAP: Optimized Analytics Package for Spark Platform
Spark Summit 2017, San Francisco, Jun 2017

OAP: Optimized Analytics Package for Spark Platform
Strata Beijing 2017, Beijing, Jul 2017

MISCS

Github: https://github.com/adrian-wang