

## 王道远

联系电话: +86-152-1686-1267

E-mail: me at daoyuan.wang

教育经历	何志均班 计算机科学与技术, 浙江大学, 本科 总 GPA 86/100 获浙江大学 2011-2012 学年三好学生 获学士学位, 何志均荣誉证书	2009/08 至 2013/06
工作经历	资深软件工程师 @ 英特尔亚太研发有限公司 毕业后加入英特尔软件与服务事业部 (SSG) 大数据部门, 从事 Hadoop/Spark 大数据生态系统软件优化工作, 结合实际用例为平台开发新功能并进行性能调优。	2013 年七月至今
	<ul style="list-style-type: none"><li>• <i>Panthera ASE</i> 2013 年 7 月至 2014 年 2 月 该项目尝试基于 Apache Hive 支持 PL/SQL 语法, 并被集成到 IDH (Intel 的 Hadoop 发行版) 中, 支持客户一些固有查询无需修改即可迁移到 Hive 平台。项目主要共由三人进行开发, 代码托管在 Github 上。负责重新设计了关联子查询的处理逻辑, 大大提高了对复杂查询语句的支持。另外, 还修复了很多错误, 开发了很多新功能, 提高了查询支持度, 还实现了一套针对错误查询语句报错的机制。</li><li>• <i>HiBench</i> 2014 年 3 月至 2014 年 12 月 HiBench 是由英特尔大数据部门发起的开源大数据基准测试工具, 在业界有比较广泛的应用。在本人参与项目前, HiBench 仅支持 Hadoop 0.x 和 1.x 的提供的 MRv1。本人修复了许多关于不同 Hadoop 版本的兼容性问题, 并确保 HiBench 3.0 可以同时支持 MRv1 和 MRv2, 并负责发布了 HiBench 3.0 版本。该项目托管于 Github。</li><li>• <i>Apache Spark</i> 2014 年 4 月至 2016 年 9 月 Apache Spark 是一个快速的分布式通用大规模数据计算平台。Apache Spark 在集群和公有云/私有云中都有广泛的部署和应用。本人贡献了很多代码到 Spark 开源社区, 主要集中在 Spark SQL 模块。结合用户的需求与反馈以及对程序的测试分析, 所贡献的代码包括新功能实现、错误修复、性能优化等各方面, 提高了 Spark 的易用性与功能, 优化了性能。期间翻译出版了《Spark 快速大数据分析》一书。</li><li>• <i>Intel OAP (曾用名 Spinach)</i> 2017 年 2 月至今 OAP 是英特尔大数据团队基于 Spark SQL 实现的即席查询加速库。许多公司在生产环境中尝试了部署 Spark 作为默认的大规模数据分析引擎, 让数据科学家通过平台使用 SQL 进行一些数据分析, 然而原生的 Spark SQL 性能可能无法满足即席查询的需求, 毕竟 Spark 并不是为这种场景专门优化的。OAP 是我们提供的 Spark SQL 优化包, 充分利用集群的硬件条件, 使用索引与缓存等机制加速查询的执行, 同时对新硬件提供支持。 该项目于 2017 年 6 月开源, 本人为项目核心开发者, 主导了 0.2 与 0.3 版本的开发工作。百度在生产环境的广告策略实时分析引擎中使用了 OAP 包, 并在真实的查询中观察到了 1.5 倍至 5 倍的性能提升。</li></ul>	
	实习软件工程师 @ 英特尔亚太研发有限公司 以实习生身份工作于英特尔 IT FLEX 部门, 参与基于 Hadoop 实现的 TAS (Transcode as a service) 项目。该项目使用 Hadoop 平台和英特尔集成显卡硬件解码功能提供视频	2012 年 9 月至 2012 年 12 月

转码服务。在测试中发现了系统隔一段时间会随机蓝屏的错误，后确定为显卡驱动的错误。负责使用 C# 开发了项目的自动化测试框架，进行日常回归测试。

计算机技能	◇ 编程语言与软件: Spark, Hive, Hadoop, Scala, JAVA, Python, SQL, C/C++, C#, PHP, R, Ruby, BASIC, Assembly, JavaScript, Linux Shell, Windows Shell, Matlab, OpenCV, OpenGL, L <sup>A</sup> T <sub>E</sub> X, MS Office.	
获奖	第 11 届浙江大学大学生程序设计竞赛三等奖（总第 19 名）. 第 12 届浙江大学大学生程序设计竞赛三等奖（总第 10 名）.	
翻译著作	<i>Spark 快速大数据分析</i> (Learning Spark: Lightning-fast Data Analysis 中文版) 图灵程序设计丛书，人民邮电出版社	
演讲	<i>Spark 应用 GC 调优</i> QCon 上海 2015, 2015 年 10 月 <i>Spinach: 使用 Spark SQL 执行 Ad-hoc 查询</i> 中国数据库技术大会 2017, 2017 年 5 月, 北京 <i>OAP: Optimized Analytics Package for Spark Platform</i> Spark Summit 2017, 2017 年 6 月, 旧金山 <i>OAP: Optimized Analytics Package for Spark Platform</i> Strata Beijing 2017, 2017 年 7 月, 北京	
相关链接	Github 个人主页	<a href="https://github.com/adrian-wang">https://github.com/adrian-wang</a> <a href="http://daoyuan.wang">http://daoyuan.wang</a>