

翟士丹 (mail@stanzhai.site)

自我介绍	6年+工作经验，现居北京 专注大数据领域，Apache Spark Contributor，有丰富的BI系统构建经验；熟悉SQL、Web技术栈及Linux；工具控；热爱编码，喜欢探索未知领域。
教育经历	曲阜师范大学 软件工程，2008-2012，本科
职业相关	方向： 大数据平台、分布式存储/计算；后端开发；基础平台 技术栈： Spark、Hive、Hadoop、Nodejs、Docker、OpenResty 编程语言： Scala、Python、Go ... 编辑器： Vim
工作经历	海致： 架构师，2014.12，SaaS BI(BDP)大数据平台负责人，核心研发 搜狐·畅游： 研发组长，2012.07，游戏数据分析平台；爬虫、舆情分析系统；信息安全审计系统 浪潮： 实习，2011.12，空军装备研究院，作战指挥系统（C++）
项目经验	BDP商业数据分析平台 Spark, BI 2014-至今 https://www.bdp.cn <ul style="list-style-type: none">- 基于Spark实现通用查询及建模服务，为BDP提供基础计算能力- Spark性能优化及Bug修复（Spark 0.9 -> 2.3）- 实现Spark任务过载保护及自愈机制- 实现多级查询缓存模块- 实现支持Update、Delete的数据接入服务- 基于分区表实现物化视图执行机制，改进超大表建模速度- 基于Go实现基于优先级及Quota机制的查询调度模块- 参与实现并行DAG数据建模任务调度系统- 基于Node.js实现运维管理平台，用于BDP服务监控及问题排查- WIP: 基于Structured Streaming实现通用流式计算平台 MOPlatform数据分析监控平台 ASP.NET MVC4, RESTful, MVVM 2013/10-2014/12 <ul style="list-style-type: none">- 建立高度自定义的报表模块，用于数据可视化，通过灵活的报表配置，自定义报表内容- 基于正则，实现通用的日志解析工具，将日志文件处理为带Schema的结构化数据- 实现专用ETL模块，抽象日志元数据生成可以直接分析入库的数据集，统一数据源管理- 构建数据分析模块，实现常用的阈值算法（极值，四分位等）对异常数据进行分析报警- 实现基于Phantomjs的服务端图表生成模块，保证最佳的图表生成质量 舆情监控系统-天兵 C#, Node.js, Phantomjs 2012/07-2014/12 <ul style="list-style-type: none">- 实现通用爬虫系统（队列，排重，正文提取，存储，分布式，任务管理及任务指派等）- 设计并实现舆论分析系统（贝叶斯分类，实时图表，舆情数据可视化，语音及邮件报警等）- 开发Web信息采集器，带领团队完成特定信息的采集和分析，为竞品分析提供基础数据源- 实现通用网页正文提取算法：https://github.com/stanzhai/Html2Article 被一些公司采用 Android五棍棋游戏 2010/10-2010/11 五棍棋是一款深具中国传统文化特色的棋类游戏。我负责Android版的五棍棋项目开发，采用AndEngine游戏引擎实现，优化了AI核心算法，实现了蓝牙对战、人机对战。 参加移动MM百万青年创业大赛，全国入围
其他	GitHub: https://github.com/stanzhai 专栏: https://zhuanlan.zhihu.com/bigdata-spark Blog: https://stanzhai.site 折腾: https://www.zhihu.com/question/51314788/answer/144324074