

## 教育背景

- 硕士 | 中国科学院软件研究所 (保送) | 计算机软件与理论 | 工学硕士 2013.9 - 2016.6
- 学士 | 东北大学 | 软件工程 | 工学学士 | 名次: 8/304 | 通过 CET-6, 口语良好, 阅读良好 2009.9 - 2013.6

## 专业技能

- 熟悉大数据技术, 能够使用 Hadoop、Spark 等相关技术处理海量数据, 熟悉 Hive、Spark SQL
- 熟悉 Java 与 Java 企业级开发, 了解 C/C++、Scala、SQL 等常用语言
- 熟练使用 Linux 系统, 熟悉 Linux 常用指令及 Shell 脚本, 能够在 Linux 环境下开发应用
- 熟悉设计模式, 数据结构及算法, 熟悉软件工程开发流程, 懂得团队合作, 良好的编码风格

## 项目实践

**海量数据交互式查询、统计分析及预处理** | 主力开发人员 | 中国科学院软件研究所 2015.7 - 至今

- 毕业设计课题, 使用 Spark SQL 替代 Hive 实现海量数据交互式查询, 基于 Spark 实现 HDFS 只读文件系统数据的预处理功能, 并封装成 UDF 为用户提供 Web 接口的大数据预处理及分析平台
- 负责 Spark 源码研究, Spark SQL 查询优化, UDF 封装, HDFS 数据预处理等
- 关键技术: Spark, Spark SQL, UDF, Tachyon, HBase, ORCFile

**Haflow: 大数据分析平台** | 主力开发人员 | 中国科学院软件研究所 2014.6 - 2015.5

- 863 项目, Haflow 是一个灵活可扩展的大数据分析平台, 集成多种大数据分析工具与算法, 如 R, Hive, Mahout 等。通过 Web 接口提供海量数据资源管理、统计查询及可视化分析、流程化数据挖掘等功能。现已应用在中国卒中数据中心提供国家脑卒中大数据分析平台
- 负责基于 Sqoop 实现数据集成模块, 使用 Hive 实现海量数据查询与统计分析, 并通过 Echarts 实现可视化, 使用 Linux Shell 实现集群自动化部署等
- 关键技术: Hadoop, Hive, Oozie, Mahout, R, Echarts, SSH

**基于 ETL 的数据集成技术** | 研发实习生 | 锐易特软件有限公司 2012.12 - 2013.5

- 本科毕业设计课题, 东软医疗卫生实际项目。数据来源于辽宁省各大医院, 实现将这些形式各异、结构各异的数据源通过基于 ETL 的数据集成技术装载到东软卫生医疗标准的数据仓库中
- 负责研究基于 ETL 的数据集成技术, Kettle 源码研究, 使用缓存库设计业务逻辑, 业务组件化等
- 关键技术: ETL, Kettle, Java 多线程, 业务组件化

**Java 企业级开发实习** | 关键技术: SSH, Ajax, Maven, JUnit, JSP 等

- 人力资源管理系统 (一期) | 研发实习生 | 联想研究院 2015.3 - 2015.5
- 酒店客房管理系统、物流管理系统 | 评审负责人 | 东软实训中心 2012.5 - 2012.10

## 所获奖励

- 硕士 | 中国科学院大学优秀学生、2014 年 OW2 国际开源代码比赛二等奖
- 本科 | 国家奖学金(10/304)、优秀学生标兵(8/304)、东北大学优秀毕业生(10/304)、曾宪梓优秀奖学金(3/304)、东软集团奖学金(15/304)、优秀团员标兵(8/304)、优秀学生一等奖学金(10/304)等