

王佳希

北京市海淀区 170-9014-5202

wangjiaxi.cv@qq.com

教育背景

2013.9-2017.6	天津商业大学	信息与计算科学
<ul style="list-style-type: none">主修课程：数学分析、高等代数、概率论与数理统计、数学建模、数值分析、离散数学、数据结构、数据库系统、算法导论、Java 程序设计、神经网络等。获得毕业生奖学金、全国大学生数学建模竞赛（CUMCM）国家二等奖、天津市二等奖主持并完成国家级“大学生创新创业训练计划项目”——在校大学生就业信息分享平台		

工作经历

2017.7-2018.8	天津精准互动科技有限公司	技术部-数据开发工程师
2018.12-2019.3	北京中软国际科技服务有限公司	数据服务业务线-软件开发工程师

知识技能

- 熟练掌握 Java、Scala 开发
- 熟悉 Hadoop 生态系统的技术框架，如：MapReduce、Yarn、Zookeeper、Flume、Spark、Kafka、HBase 等
- 熟悉分布式文件系统 HDFS 读写及工作机制，熟练掌握分布式计算框架 MapReduce 及运行机制，熟悉分布式资源调度平台 Yarn 的资源调度机制
- 熟悉分布式协调系统 Zookeeper 集群搭建，了解 Zookeeper 内部投票选举工作机制
- 理解 Spark 运行机制，具有一定的 Spark 调优经验
- 掌握 Spark Streaming 实时框架，Spark Core 常用算子，Spark SQL 数据统计
- 熟悉 Flume 工作机制，通过自定义 Source 实现日志数据采集
- 熟悉 Linux 系统，掌握常用的 Shell 命令，能编写简单脚本
- 熟悉 Redis，MySQL 数据库，了解 ES
- 熟悉 IDEA 等开发工具，熟悉 Maven 等项目管理工具

项目经验

2015.5-2017.5	在校大学生就业信息分享平台	天津商业大学
<ul style="list-style-type: none">类型：在校科研项目简介：学习论坛类网站搭建技术框架，使用 bootstrap 前端框架和腾讯云 wamp 集成云服务器，并用 MVC 架构搭建简单网站。职责：负责论坛主要模块开发，并负责人向学院范围师生（在校和已毕业）推广。技术：bootstrap + wamp		
2017.7-2018.8	广告智能推送数据管理平台	天津精准互动科技有限公司
<ul style="list-style-type: none">类型：商业项目简介：不同的人群，网络行为爱好差异大，产生了不同类型的广告投放需求。该项目通过分析用户的上网行为、爱好习惯等数据，抽象出一个用户的商业全貌，为用户进行画像，这样即可根据广告主的需求，确定广告的受众群体，进行精准投放。		

- 职责：
 - (1) 数据采集：利用 Flume 组件自定义 Source 采集 Redis 溢写到磁盘上的原始日志，并通过 File Channel 和 HDFS Sink 写入 HDFS。
 - (2) 统计报表：离线数据分析：地域、媒体、终端设备、渠道来源、原始请求数、有效请求数、广告请求数、参与竞价数、成功竞价数、广告点击数、广告展示数、点击量等等，并将结果存入 MySQL，供 WEB 系统查询，展示。
 - (3) 用户画像：将用户数据标签化，根据原始日志和知识库给用户打上关键字标签，商圈标签等，并不断完善公司知识库。对上下文标签进行合并，并进行统一用户识别，完成用户标签的最终合并。
 - ① 关键字标签：把一些没有意义的关键字收入停用词库，对数据进行分析时，使用 BroadCast 广播停用词库，操作关键字标签时将关键字与停用词库进行匹配，如果停用词库包含这个关键字，则不需要打这个关键字标签。
 - ② 商圈标签：对原始数据携带的经纬度进行 GeoHash 编码，对接百度的逆地理位置编码服务，获得经纬度所对应的商圈信息建立商圈知识库，基于我们的商圈知识库，对用户打上商圈标签。
 - ③ 统一用户识别：使用 IMEI、MAC 等编号标记用户，但是在日志中相同的用户的这些字段不一定同时有值，所以使用 Graph 进行统一用户识别。
 - ④ 用户标签合并：将 Graph 识别后的同一用户的标签进行合并，标签权重值*0.8 的衰减系数得到最终的用户画像。

● 技术：Scala + Spark Core + Spark SQL + Spark Graphx + Flume + Redis + MySQL + Hadoop + Zookeeper

2018.12-2019.3

苏州公安在线数据攻坚

北京中软国际科技服务有限公司

- 类型：政企行业项目
- 简介：苏州市公安局数据平台建设，该项目属于苏州市公安局“警务大脑”项目的一部分
- 职责：
 - (1) 实时流数据开发。使用华为 Fusion Insight 管理集群，将接收到的各分局和派出所警务数据通过 Kafka 消费并存入到 MPPDB 数据库，并在 Yarn 上运行 Spark Streaming 任务，配合苏州市各派出所开发人员创建 topic 和数据表，每天处理 40G 左右数据量。
 - (2) 集群监控与维护。通过 FI 的管理界面，不定期查看集群各个组件的运行情况，检查是否存在宕机、任务出错、数据为空等情况，并及时重启任务、修补数据。
- 技术：Java + Scala + Kafka + Yarn + Spark Streaming + Redis

个人能力

- 良好的英语口语表达及文档阅读能力，获得 CET-6 证书
- 扎实的数学基础，优秀的逻辑思维能力、推理判断能力和主动学习能力
- 具有良好的沟通技巧，高度的团队精神，追求极致
- 喜欢钻研技术，善于积累和分享，个人博客：<https://wangjiayi.github.io>