

罗旭耿

✉ luoxugeng12581@163.com · ☎ (+86) 185-180-71569 ·

🎓 教育背景

广东工业大学, 广州
学士 计算机科学与技术

2011 年 9 月 – 2015 年 6 月

💼 工作经历

快手北京

2018 年 6 月 – 今

数据开发工程师

2018 年 2020 年 7 月, 参与海外数仓, LOGSDK 公共基础数仓建设, 2020 年 7 月至今, 在风控团队负责基础数据和小组资源治理工作

- 负责海外项目 Zynn 数据仓库从 0 到 1 建设
- 2020 年 K3 战役中负责春节预热活动数据全链路服务, 包括数据埋点设计, 验收, 活动数据指标建设

关键结果 & 技能点:

- 搭建风控团队的基础数据, 如审核, 举报数据, 构建惯犯的基础特征和数据召回评估
- 深圳团队资源治理交付结果: 减少存储 300TB, 优化资源计算量 4TB* 天
- 使用 Spark 引擎重构新业务日活表逻辑, 相同的资源任务耗时缩短 50%;
- 2019 年参与海内外 DB 隔离工作, 结果: 在不影响业务前提下, DB 机器减少 80%

华兴资本北京

2017 年 4 月 – 2018 年 6 月

数据开发负责人

负责华兴资本数据从 0 到 1 的搭建. 设计了从数据抓取到数据仓库建设整个框架并实现. 在完成主流机构融资数据建设后, 自发对融资内容进行文本分析, 输出机构的偏好给业务同学做数据支持.

- 挑战点: 没有现成数据, 需要自己使用 Scrapy 对外部数据进行抓取. 为了效率的提升, 采用分词和后期人工介入来清洗异构数据.

腾讯-微信支付研发部门 – 深圳

2015 年 6 月 – 2017 年 4 月

后台开发/数据开发

作为数据组 4 名成员之一, 主要负责基础数据清洗和数据主题域建设. 满足分析师日常取数需求. 在 2016 年 88 无现金日, 独立完成产品需求沟通, 按时输出亿级用户过去一年支付数据统计指标.

- 活动期间, 根据设备/登陆地/ip 等方式, 支持风控同学识别疑似薅羊毛用户
- 活动期间, 发现对账系统滞后严重, 对其进行重构. 结果: 使对账时间缩短 40%

⚙️ 技能

- 熟悉大数据仓库整个建设的理论和实践, 擅长自顶向下分析
- 了解整个 Hadoop 的整个生态, 熟练 Hql 以及优化方式;
- 了解调度集成工具 Airflow, 分布式消息系统 Kafka, 并行计算框架 Spark
- 熟悉 mysql, 了解其底层存储引擎底层实现, 了解基本的性能调优以及死锁分析
- 熟悉常用数据结构与算法, 设计模式, 了解多线程编程模型, 如 Libevent
- 编程语言: hql > C == Python > C++