



谢祁峰



男 | 32岁 ☎ 17640410805 📩 xieqifeng2018@gmail.com

9年工作经验 | 求职意向：数据开发 | 期望城市：北京

个人优势

- 喜欢钻研技术并解决工作中遇到的疑难杂症，通过技术博客打造从输入到输出的闭环知识体系，有自己专属的个人技术公众号，公众号：junzibuqi_x
- 熟练使用AI智能体搭建工作流，包括dify, claude code等工具
- 有强烈的求知欲，喜欢读书和写作，曾在1年内读过100本书籍
- 喜欢哲学和心理学，对工作和生活具备良好的心态
- 喜欢做技术分享给身边的人，乐于助人帮从中找到意义感和快乐

工作经历

北京百易图信息科技有限公司 数据开发 2023.12-至今

负责短视频数据离线推荐数据的开发和建设，包括基础数据、上层离线特征、推荐核心指标看板及看板指标的监控和告警。

搭建基于ClickHouse的实时数据和离线数据的BI支撑，保证clickhouse的平台优化，包括表设计、表管理及实时数据的稳定性，提升数据处理效率。

参与Flink实时直播域推荐数据相关任务的开发，包括直播实时特征、用户归因及实时指标看板。

主导公司大数据调度平台的调研和建设，涉及dolphinscheduler 2.0和3.2.1以及DataWorks on EMR沙特站的调研、跟进及部署，同时监控整个任务的执行情况。完成后期迁移所有任务到dataworks，完成对dolphinscheduler 2.0和3.2.1版本的下线。

元气森林（北京）食品科技集团有限公司 数据仓库 2020.12-2023.11

1. 数仓一期模型建设

终端域：负责网点域/资产域/业务员域，包括网点订单/外勤/资产投放，期间上游业务系统重构了2次，数仓ods数据重新接入并整合历史数据以及模型重构

主数据建设：包括人员/部门/网点/冰柜，一期主数据由数仓独自开发完成，后期由单独主数据团队负责，数仓从主数据集成数据

指标设计：主导一期指标设计和开发原子指标/派生指标/复合指标，一期指标包括派生指标，复合指标后台统计个数600+

指标系统：主导调研并基于阿里dataworks数据建模系统设计和开发模型及指标并团队内宣讲二

2. hologres 建设

前期调研和开通hologres实例，对其使用场景和使用方式进行组内宣讲

结合业务指标需求探索并应用roaringbitmap结构用于支持产出去重型指标类型

业务系统实时数据集成方案切换为hologres

纷享销客数据数据集成方案切换为hologres

3. 数据应用

销售一张表一二期

销售周报

销售月报

诊断模型

销售工作台

4. 数据治理系统开发

费用统计 包括 dataworks/maxcompute/hologres/flink

作业费用统计，分项目空间/分用户/top 费用作业统计并告警

无下游数仓表治理

5. 跨店宝(电商数据)

财务数据需求：对 rpa 抓取的各大电商后台账单入仓并计算加工分摊后结果表

业务数据需求：销售业绩看板/供应链看板/即时库存/在途库存

火币科技控股有限公司

数据开发

2018.06-2020.11

1. 安全实时数仓

设计并搭建安全数仓，主要包括集团日志(waf/堡垒机/auditbeat/k8s/防火墙)实时采集，入仓及后续统计分析及告警需求

部署kafka 2.5集群(三节点)/elk 7.8.1/ open-falcon 部署监控（11台机器及进程）/配置logstash数据源(包括mysql日志数据源/kafka数据源),以及输出源es集群

使用structed streaming实时采集kafka写S3，用于后续运营告警看板

部署clickhouse集群19.16版本(3shard2replica)，开发kafka connect sink ck组件实时数据写入ck

搭建superset bi用于clickhouse即席查询

2. azkaban部署及开发

部署集群

根据梳理好的依赖关系开发任务流并逐渐迁移到 azkaban

规范化shell调度脚本模版

合理规划任务流，将属于相同业务的任务部署到同一个工作流中

根据业务要求合理分配任务到相应的队列中

合理规划任务流的执行时间，保证集群任务的平稳运行，资源充足，尽可能减少任务在资源上的争抢

导致的 oom以及运行缓慢

统计集群各个业务数据量占比

开发 azkaban 报警功能，修改 azkaban 在阿里云被禁用端口号导致报警邮件无法发送

3. 离线数仓

合约insight二期（用户-用户分析模块）

新增交易用户（交割合约/永续合约/全部）（下钻维度：合约币种，交易方式，认证国家，费率等级，合约类型，交易倍数，订单来源）（日报/周报/月报）

新开户人数（交割合约/永续合约/全部）（下钻维度：认证国家，费率等级）（日报/周报/月报） - 重点交易用户

top100交易金额（按本周taker+maker交易总量排序）（周报） -

海外用户分析（国内，海外，其他交易金额占比）（日报/周报/月报） - 用户交易币种数量分析（日报/周报/月报） -

用户访问分析（浏览量和访问数）（日报）

Global用户行为二期（留存率模型模块）

维度：客户端/国籍/具体国籍（可多选）/渠道数据（可多选）

日留存（次1日至次7日/次15日/次30日）/周留存（次周至此七周）/月留存（次月至次七月） -

新注册用户留存率（N日登录数/某日注册数） -

登录留存率（N日登录数/某日登录数）

访问留存率 (N日访问数/某日访问数)

交易留存率 (N日交易数/某日交易数)

去哪儿网 数据开发

2017.06-2018.04

1. 实时广告监测系统

所用技术：java,spring,mybatis,mysql,kafka stream,qmq,hive,redis,hbase,spark streaming

项目描述：去哪儿作为广告主和各个渠道广告商合作投放广告，该项目用于实时接受点击数据和激活数据进行归因处理，系统数据库数据用于实时统计点击和激活数，日志数据每小时离线同步到数仓小时分区用于离线计算

使用springmvc开发对接各渠道开发广告点击接口

使用sparkstreaming实时接收客户端安卓激活数据过滤和处理

将上游安卓激活数据写入qmq供下游使用

使用spring和qmq开发安卓激活数据消费接口

开发归因模块，将各渠道点击数据和激活数据根据优先级和first click进行归因处理

开发激活回调模块将归因后的激活数据回传给该激活数据所属渠道

教育经历

吉林大学

本科

计算机软件

2011-2015