

姓名: 李健  
出生年月: 1995.09

邮箱: lj\_cs@foxmail.com  
电话: 13120362233

## 教育经历

- |            |            |                   |
|------------|------------|-------------------|
| ● 北京航空航天大学 | 计算机科学与技术硕士 | 2015.09 - 2018.04 |
| ● 长沙理工大学   | 网络工程学士     | 2011.09 - 2015.06 |

## 工作经历

- |   |                  |                   |
|---|------------------|-------------------|
| ● 快手  | 音视频技术部数据研发组研发工程师 | 2019.10 - 至今      |
| <ul style="list-style-type: none"><li>○ 实时数据开发和数仓建设</li><li>○ 音视频异常检测和归因系统设计与开发</li><li>○ 指标管理以及相应的查询订阅服务、质量管理能力建设</li><li>○ 数据春节等活动重保</li></ul>  |                  |                   |
| ● 百度  | 知识图谱部离线架构组研发工程师  | 2018.04 - 2019.09 |
| <ul style="list-style-type: none"><li>○ 百度知识库数据存储系统Restful批处理接口, 流计算场景下的RPC服务开发</li><li>○ 知识计算框架实体检索模块优化以及算子网关开发</li><li>○ 百度知识图谱SPO收录机制的开发</li><li>○ Hadoop、HBase、ES、MongoDB集群和服务的维护工作</li></ul> |                  |                   |

## 专业技能

- 熟悉Java栈、Python工程开发; 熟悉sql, shell, Javascript等脚本开发
- 熟悉流处理框架Flink, 以及Hadoop生态Yarn、Hive、Spark
- 熟悉OLAP引擎Clickhouse, Druid, NoSQL存储引擎HBase, Elasticsearch
- 熟悉分布式数据系统的基本原理, 并且有相应的应用调优经验
- 熟悉微服务架构, 具有Spring Boot, gRPC应用开发经验
- 熟悉常用的数据库、消息队列和缓存技术, 例如MySQL、Redis、Kafka、ZooKeeper等
- 具有一定的数仓建设和数据治理的经验, 包括有模型设计以及流程规范等

## 项目经验

- 快手-音视频体验优化实时数仓
  - 项目描述: 面向快手主站和海外产品下音视频全链路上报数据, 基于音视频指标体系构建实时数仓, 支持CDN质量管理、排障、实时A/B、OLAP数据分析、策略算法模型及异常检测归因等数据应用
  - 主要工作: 1. 快手直播全链路质量实时数仓建设, 从主播推流、源站、CDN分发、拉流 2. 生产侧实时数仓建设 3. 流批一体能力建设 4. 活动时数据重保
  - 难点: 1. GB级别下的实时数据开发和参数调优 2. 音视频业务场景复杂带来的维度多、链路长、数据不完全、数据不统一等难点 3. 需要高效的流批一体以及实时数据回放机制 4. 快手活动给数据任务保障带来的挑战
  - 相关技术: Java、Flink、Clickhouse、Druid、ES、数仓建模、Hadoop
- 快手-音视频异常检测与归因系统
  - 项目描述: 人工故障发现、定位、止损恢复耗时久不及时, 需要智能的异常检测、归因、自愈的自动化能力提高故障发现和止损效率、保障音视频用户体验。系统目前共接入共计160+个指标, 覆盖了快手主站、海外Kwai以及其他音视频中台支持产品下点播、生产、直播、网络库、静态资源下载、客户端等主题域下10多个业务过程; 快手主站直播全链路监控建设发现90%直播故障, 近三个月21起有效反馈; 主站核心指标准确率达92%; 支持CDN自动故障调度, 故障恢复时长从日均7h可以降低到5min, 并且为为CDN云厂商服务质量优化提供了长效的数据支持和评价依据
  - 主要工作: 系统历经了两个大阶段, 第一阶段基于人工阈值类策略, 第二阶段升级为无监督智能算法策略, 负责从0到1的系统工程构建以及相应的维护工作
  - 难点: 1. 高效的算法策略开发、上线和排障框架设计 2. 工程和存储架构的设计支持系统可扩展 3. 指标体系设计支持算法和工程优化 4. 通用能力的复用
  - 相关技术: Python、服务编排Conductor、Hadoop、Spring Boot
- 快手-音视频指标管理服务
  - 项目描述: 音视频业务下存在指标取数效率低、指标口径难以统一、生产消费链路耦合、消费场景管控缺失的痛点, 因此构建音视频指标管理服务, 提供统一的指标接入以及查询服务, 同时建立相应的数据质量管理和保障机制

- 主要工作: 1. 指标查询和订阅服务开发 2、数据质量SLA体系建设
- 难点: 1、指标、维度统一管理、与数据表解耦 2、统一的查询用户视图 3、数据质量保障
- 相关技术: 指标维度建模、gRPC、Spring Boot

#### ● 百度-知识库数据存储系统

- 项目描述: 基于HBase(300个节点)开发的一套PB级别知识库数据存储系统, 支撑百度知识图谱PB级全量数据和每天约200TB的增量更新, 主要实现以下功能, 1)提供大规模结构化数据存储服务 2)封装了常见的HBaseAPI操作: 包括数据批量和单条的CRUD操作, 简化了用户使用Hbase的成本 3)在Hbase基础上增加了一套权限控制方案 4)提供了数据备份、容灾机制 5)为数据提供了干预、回灌等功能, 为数据生产提供了遍历
- 主要工作: 1. RESTful和RPC接口开发 2. 权限系统升级和维护 3. 运维工具开发
- 相关技术: 1. Java多线程, 网络编程 2. HBase客户端开发 3. RPC服务

#### ● 百度-知识计算框架KB检索模块优化

- 项目描述: 知识计算框架对知识的计算, 存储功能进行了封装, 实现了DAG任务流。检索模块为其中的一个工作任务(组件), 支持SPO消歧等任务, 在Elasticsearch(48台机器)索引(10亿文档)中对数据进行查询; 从搜索策略、ES索引调整、集群部署优化三个方面进行了优化, 服务qps由100提升到700
- 主要工作: KB检索模块优化方案设计和代码开发
- 相关技术: 1. elasticsearch, mongodb 2. python开发