

# 赵澍源

(+86) 185-1860-8351 · shuyuan.zhao@hotmail.com

## 个人总结

作为大数据工程师，对数据充满热情，有较强的业务理解和问题抽象能力。具备丰富的数据产品研发及项目管理经验。从事多年 JAVA、Python 语言开发，熟悉分布式系统的设计和应用，熟悉分布式、缓存、存储等机制。熟练掌握 Spark、Hadoop、ClickHouse 等大数据核心组件，具备海量数据处理和实时数据计算工程经验。熟悉广告及市场营销业务，擅长处理复杂的业务场景，挖掘数据价值，用数据驱动业务。具备优秀的沟通能力和团队合作精神。

## 技术能力

- 编程语言: Java, Python, Scala, SQL
- 大数据: Spark, ClickHouse, Yarn, Hive, OLTP/OLAP
- 中间件: Redis, Elasticsearch, Zookeeper, Thrift, MongoDB, 消息队列

## 工作经历

### 字节跳动 | 大数据工程师, 北京

2019.12 至今

- 负责巨量云图数据营销产品的后端及数据开发，助力客户的营销决策和长效经营，提供人群、内容资产运营，广告投放价值度量，小店生意追踪能力，为营销场景提供科学有效的解决方案；
- 作为项目 owner 负责项目的排期管理和技术实现，完成包括需求分析、方案设计、技术选型、开发上线、线上质量保障等研发流程；
- 负责搭建基于 Yarn 集群的 Spark 分布式离线计算服务，参与 ClickHouse 在线计算服务开发，提供 PB 级数据底层计算能力；
- 负责基于 RPC 框架的后端接口开发，涉及微服务架构、缓存、ES、MySQL 等后端核心技术；
- 参与数据仓库及离线数据流搭建，以品牌为视角实现用户数据资产沉淀。

### Capital One | 数据分析师, 美国

2018.2-2019.11

- 负责 Capital One 信用卡市场营销和客户管理数据的数据分析及可视化。为商业决策及客户增长提供依据；
- 搭建基于 Snowflake、Redshift 的 Spark 数据管道实现用户分层及用户数据分析。

## 项目经历

### 广告投后度量——营销活动的一站式复盘工具

字节跳动

- 项目介绍：为广告主提供营销活动广告投放度量分析、优化复盘的能力，涉及触达覆盖、人群流转、品牌力提升、长效转化等分析视角，支持人群、创意、频次和触点等关键元素的下钻分析。
- 技术实现：1) 基于 Spring 框架搭建 Spark 计算服务，以离线任务提交的方式产出结案报告，核心解决数据倾斜、Spark 性能调优等问题；2) 利用 Clickhouse 实现排序、TopN、下钻分析等场景的实时查询；3) 基于数据仓库对 hive 数据进行离线预聚合，将结果写入 MySQL/ES/Redis 供后端服务查询；4) 基于 Thrift 框架开发后端 RPC 接口，查询计算结果并处理返回。

### 人群分析——人群资产运营解决方案

字节跳动

- 项目介绍：为品牌人群资产提供人群分析和行业洞察能力，并提供用户画像、用户分层及目标人群的圈选能力，以实现广告的精准投放。核心解决人群分层、用户画像、人群预估、粉丝运营、破圈拉新等人群运营需求。
- 技术实现：利用 ClickHouse RocksMergeTree 引擎构建 bitmap 解决人群流转、实时画像、人群规模预估等场景的实时查询，实现百亿级数据 uv 计算秒级返回。

## 教育背景

乔治华盛顿大学, 数据科学, 理学硕士

2015.8 - 2017.5

中南大学, 统计学, 理学学士

2011.9 - 2015.6