

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。具备丰富的数据平台开发及项目管理经验。精通 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, Doris
- 中间件: Redis, Elasticsearch, Zookeeper, Thrift

工作经历

字节跳动 | 大数据工程师, 北京 2019.12 至今

- 负责巨量云图数据平台的数据及后端服务开发，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 框架的后端服务开发，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。

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

营销分析平台开发 字节跳动

- 负责数据平台开发，为广告主提供广告投放度量分析能力，支持人群、创意、频次和触点等多视角下钻分析。
- 基于 Spring 框架搭建 Spark 离线计算服务，核心解决数据倾斜、性能调优等问题；
- 搭建数据仓库实现数据分层，开发 Hive 离线数据管道；
- 负责后端微服务开发，涉及业务功能实现，RPC 接口开发，ES、Mysql、Doris 存储查询。

数据管理平台开发 字节跳动

- 基于 ClickHouse bitmap 实现人群圈选、人群预估、统计分析。
- 实现千亿级别人群包交、并、补计算结果秒级产出。
- 搭建 Spark 离线服务，产出用户标签。

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