

易斌鑫

个人信息

- 易斌鑫/男/1993年/28
- 西安电子科技大学/本科/数学系/2010.9-2014.7
- 工作年限: 7年
- 期望职位: 大数据开发工程师
- Tel: 18710849538
- Email: ybxlife@foxmail.com

工作经历

- 迅捷联动（北京）科技有限公司（2018年1月 ~ 2021年5月）
- 西安优云软件有限公司（2017年6月 ~ 2017年12月）
- 西安导学教育科技有限公司（2014年7月 ~ 2017年5月）

工作项目

- **大数据平台**（2018/01 - 2021/05）
 - 负责大数据平台技术方案调研和实施，数据仓库整体架构设计、ETL开发、任务调度开发、元数据管理；
 - 完善业务应用端、服务端数据埋点采集方案，实现业务多源数据库（Mysql、Redis）数据同步集成；
 - 开发数据仓库上层应用，离线统计监控、实时统计监控、SQL自助查询引擎；
 - 挖掘业务数据，用户画像分群、数据分析模型（ROI、LTV）；
 - 搭建K8s、Docker容器环境Kafka集群，开发Kafka Stream流式数据处理服务；
 - 编写项目文档、用户内容行为反垃圾数据模型。

为业务产品提供数据支撑，熟练掌握Scala、Hadoop、Spark、Docker、Zeppelin、Fluentd、Kafka，Team Leader角色

- **AB测试实验平台**（2020/03 - 2020/09）
 - 书写并分享学习笔记，AB实验分流系统架构设计；
 - 跟进实验CheckList管理与流量分配方法原则；
 - 实现实验分层管理、流量分域分配、白名单规则、用户受众规则、事件指标配置和收集；
 - 应用多实验分层，互斥实验分组和简单的流量百分比自动调整。并与产品共同推广AB实验优化的基本工作流程。

减少业务开发的工作量，仅一次API调用。使产品业务优化探索效率提升一倍

- **内容推荐系统**（2019/05 - 2020/03）
 - 开发特征处理逻辑，训练CTR预测模型（FM），提升AUC指标，实现多路召回、重排序及冷启动策略，提高内容分发效率；
 - 使用Spark、Angle框架实现分布式和增量模型训练，通过Faiss、Redis提高Embedding向量检索速度；
 - 使用BERT模型实现word2vec的内容文本向量化特征处理，增加内容文本I2I召回，提升内容主题相似度；
 - 实现模型服务每小时更新，开发内容用户多维度CTR指标数据统计。

为公司省掉第三方推荐接入成本，提升产品首页内容分发效率；了解TensorFlow、Keras等机器学习框架工具。

- **日志收集监控服务**（2016/07 - 2017/05）
 - 研究和实践ELK技术栈，使用轻量级Filebeat工具进行日志收集，减轻服务器压力；
 - 编写Nginx、Tomcat的Logstash日志处理器，完成日志格式预处理；
 - 添加Kibana数据监控规则，监控服务允许状态和服务压力；
 - 搭建CDH数据集群，实现业务数据集成处理，搭建Solr分布式索引服务。
- **K12题库搜索系统**（2014/09 - 2016/06）
 - 研究Lucene索引排序和OCR模糊搜索算法，优化图像处理效果和上传速度；
 - 搭建Solr分布式搜索引擎，完善索引词典建设，编写LaTeX分词算法，提升10倍搜索请求QPS；
 - 实现HTML+LaTeX文本内容图片转码服务和PDF、DOCX文档格式转换服务，提供公共API，通过多任务阻塞队列实现并发快速处理千万任务；
 - 参与翻译Solr使用指南，为其它项目提供搜索服务。

技术能力

- 熟练：Java、Scala、Python、Shell、Linux、Docker、Spark SQL、Hive
- 熟悉：Redis、MySQL、Hadoop、Kafka、K8s、Spring Boot
- 其它：英语六级、Git、Maven

教育经历

- 西安电子科技大学 本科 / 数学与应用数学 2010.09 - 2014.07

自我评价

- 工程开发经验丰富，具有较强独立面对难题的能力，工作态度认真，敢于探索钻研。