何宗谨 求职岗位: 大数据开发工程师 | 北京 | 随时到岗

懂 年 龄: 27岁 ♂ 性 别: 男

🛅 工作年限: 3年经验 💮 📞 电 话: 18738991359

◯ 教育背景

☆ 开发技能

- 熟练使用Java,Scala进行代码开发
- 熟练使用Spark相关的技术组件,熟练使用SparkCore,SparkSql进行离线开发,熟练使用SparkStreaming进行实时开发,能够对Spark 进行优化,解决数据倾斜,OOM等问题
- 熟悉Hive组件作用,熟练使用HQL语句进行离线开发,能够编写UDF扩展Hive功能,对HQL进行调优
- 熟练使用Flink DataStream,Flink SQL进行实时开发,熟悉Checkpoint过程
- 熟悉Hbase/Redis/ES/Clickhouse等数据库,熟悉Hbase架构原理,能够对Hbases数据存取,做出合理的表设计
- 熟悉Kafka 消息中间件及其原理

▲ 工作经验

2020-05 ~ 2023-04

中国数码信息(北京无一科技有限公司)

大数据开发工程师

- 参与制定数仓规范,基于大数据技术的开发,完成公司其他部门的多种数据要求。
- 负责分析所有数据的来源,以及数据的抽取、转换、加载工作。根据业务定位做数据转化。
- 完成产品的迭代升级以及底层架构的升级研发,构建强大的OLAP系统,提供实时、离线的分析数据
- 具有较强需求分析和设计能力,保障迭代内容按质按量交付,解决在工作过程中前台开发人员遇到的数据相关问题,并寻找恰当方法,着手解决。
- 根据需求要实现的功能进行表结构设计。完成数据文档,并根据测试提交的BUG进行定位、 解决问题。

🗂 项目经验

OMO数据分析平台

项目架构:datax+hive+spark+redis+kafka

背景:

业务数据,日志数据分散难以管理,下游业务使用多次关联,重复建设.卖家在产品研究和选择上遇到困难,缺乏可靠的分析工具来监测和评估自己的销售绩效,并与行业标准进行比较,以便做出针对性的改进和优化。缺乏能够提供可视化的数据指标和图表,帮助卖家快速了解市场动态并做出决策,业务部门需要准确、及时的订单指标报表以支持决策和业务优化。

实现:

与业务团队深入沟通,了解他们的需求和报表使用场景,明确关键指标和分析要求。

通过ETL工具将来自多个数据源的数据抽取、清洗和整合,确保数据的准确性和完整性。

基于业务需求和数据模型设计,开发灵活的报表查询和分析逻辑,并优化报表性能。

实施自动化数据质量检查和异常监控,及时发现和解决数据质量问题。

职责:

数据划分:负责用户域,交易域模型建设

主题建设:构建用户域用户信息,用户活跃,历史订单等主题,构建交易域订单,商品,维权,评价等dwd层主题建设代码编写:编写代码计算dwd层明细数据和dws聚合数据

数据缓存:负责代码开发,将ads层数据写入redis加速读取数据监控:配置dqc保障数据质量,数据及时产出

结果:

成功构建了用户域,交易域等10+张报表,提供准确、可靠的关键业务指标和洞察,业务团队的数据需求得到满足,报表被广泛使用,为业务部门提供了数据驱动的决策支持。通过报表的分析和洞察,成功降低了订单处理时间、提高了客户满意度,并实现了20%的销售增长。

一号线索池

项目架构: hive+hbase+phoenix+spark+es+kafka+flink

背景:

企业数据繁多,庞杂无法准确定位目标客户,销售人员疲于奔命却成交率低下,亟需有效的方法来整理、分析和提取有用的企业洞察。 将市场数据和其他相关数据进行整合,以获得全面的市场视图和准确的企业画像。需要基于企业画像来进行目标客户圈选细分、销售 渠道选择、营销活动定制等战略决策,以提升销售效果和市场竞争力。

实现:

数据预处理:将初始数据进行ETL数据清洗,去除空值,重复值和不符合规则的值,并将时间,货币等值进行统一

标签计算:使用spark按照规则计算标签并写入hbase

写入宽表:将企业基本数据和标签数据组成宽表写入es

接口查询:提供接口供业务方使用,圈选人群

更新状态:使用flink同步保护状态至es

结果:

计算60+标签精确识别不同类型的客户群体,了解其偏好和购买决策过程。

通过企业画像的应用,能够更加精准地选择适合的销售渠道和推广方式,最终成交率提升30%+

难点:

全量去读hbase数据较为缓慢解决办法:初始存储数据将hbase改为hive

大量写入es数据es节点会宕机解决办法:调整写入并法度,修改es更新时间,写入量参数等进行参数调优

门户云

项目架构:kafka+flink+clickhouse+spark+redis

背景:

公司对于网站的访问情况和用户行为缺乏清晰的了解。需要了解网站的流量来源、用户行为和转化率等关键指标,以便进行网站优化和改进。

面对大量的访问数据和日志文件,缺乏有效的方法来提取、整理和分析这些数据。要对不同渠道、页面和用户进行细分分析,以了解不同群体的行为特征和偏好。需要监测网站的性能指标和用户体验,以便及时发现问题并采取措施进行修复。

实现:

数据收集:使用flink读取神策的埋点数据,将数据展开写入clickhouse的ods层

数据清洗:对收集到的数据进行清洗和预处理,去除重复、无效或错误的数据,并进行数据格式化和标准化

数据处理:构建dwd层流量域流量分析,来源,受访页面,访客分析等主题建设,进行dws层数据聚合以及ads层报表生成

加速查询:使用spark将ads层报表写入redis缓存加速查询

结果:

通过页面的访问量,对页面进行优化,提升页面的可见性和吸引力。

针对加载速度慢或页面响应时间长的情况。改善网站性能能够提升用户体验,用户流失率减少20%,转化率增加25%

通过网站访问数据分析,提供有价值的商业洞察和数据驱动的决策支持。

用户pv和访问时长增加30%

OMO驾驶舱-实时数据链路建设

项目架构:cdc+kafka+flink+redis

背景:

传统的报表和分析工具无法及时获取最新的业务数据和关键指标,导致缺乏对业务实时状态和趋势的洞察。这可能使卖家在做出

决策时依赖过时的数据或不完整的信息。难以

及时发现异常和问题:缺乏实时数据大屏可能导致延迟或错过对关键业务指标的异常和问题的发现。没有实时的数据监控和告警机制,无法及时识别业务过程中的异常情况,从而影响对问题的迅速响应和解决。

实现:

数据拉取:使用flink cdc将下单 支付 退单等信息拉取到kafka,同时将神策埋点数据拉取到本地kafka

数据计算:使用flink进行维度关联计算出买家信息,以及下单支付退单等关键指标信息

数据写入:将计算后的指标写入redis供接口查询

任务监控:编写监控程序,每5分钟触发监控,任务挂掉自动重启

结果:

计算30+标签毫秒级响应数据更新,提供可视化的数据展示,通过图表、图形和仪表盘等方式直观地呈现数据,有效支撑卖家促销活动,总体GMV提升20%+

难点:

数据背压,数据倾斜:增大并发,优化聚合

□ 其他

语言: 英语 CET-6

兴趣爱好: 游泳, 跑步, 读书, 研究新技术