

技能清单

- 熟悉Java、Python。
- 熟练使用SQL，熟悉常用窗口分析函数，有SQL的优化经验。
- 熟悉Hive、Spark，了解基本运行原理。
- 熟悉Airflow、DataX的使用。
- 熟悉数据建模理论，熟悉数据分层和维度建模，能够根据业务场景设计数据模型。
- 了解元数据管理、数据质量管理，有一定的实践经验。
- 熟练使用Linux、Git、Maven。
- 了解Spring Boot、Mybatis的使用，熟悉RESTful API设计。

工作经历

2018.04-2020.01 上海致维网络科技有限公司

- 负责数据模型的设计和数据库的ETL
- 负责数据任务的日常维护
- 负责数据服务接口开发

项目经验

商户数据采集系统

项目描述：之前对商户数据的采集主要是通过Airflow自定义Operator调用DataX命令进行抽取，每个商户有多个DAG任务，Airflow的压力比较大，同时也不方便管理。商户数据采集系统可以更方便的采集数据以及管理数据采集任务。

职责：在该项目中负责采集系统的开发，ODS、DWD、DIM层的模型设计和ETL。

成果：

- 使用Spring Boot将DataX包装成服务，可以非常灵活的对外提供数据抽取的服务。
- 记录每次抽取任务的执行信息到数据库，任务的执行状态由系统和数据库进行维护，Airflow上商户数据采集的任务从几百个减少到2个，大大减小了Airflow的压力。
- 记录任务执行状态、时长等信息并进行分析，便于及时发现任务和对任务进行优化。

超级导购重构

项目描述：超级导购各级面板数据经常出现不一致的情况，影响商户日常使用，重构主要是解决数据不准确和不一致问题。

职责：在该项目中负责数据模型的设计和ETL。

成果：

- 梳理并确定数据口径后，重新设计数据模型，划分数据层级，解决了面板数据不准确和不一致的问题，同时减少30%计算任务。
- 重新设计后数据结构更加清晰，更方便复用，避免烟囱式开发。

商户自定义指标

项目描述： 商户在做促销活动或统计业绩等场景下，不只是想看到基本的统计指标，也会有更为复杂的自定义指标的需求。

职责： 在该项目中主要负责ADS层表模型设计， Spark计算任务开发， 数据服务接口开发以及接口文档的编写。

成果：

- 目前有会员升级、会员PLUS以及品类、品牌、SKU自由组合等指标，在基础经营指标外，让商户从更多的角度看到了经营情况。项目上线两个月，已有过半商户使用此功能，评价良好。
- 自定义指标涉及多个类型，而且可能持续扩展，指标的计算根据指标类型和商品规则有不同的算法，计算采用策略模式而非硬编码，让算法和计算逻辑解耦。
- 项目有导出Excel的需求，使用EasyExcel和阿里云OSS编写了通用的Excel导出接口，满足了本项目的需求，其他项目也可以复用。

自我评价

- 喜欢思考，喜欢总结，能从多个角度看待事物。
- 热爱学习且学习能力强，工作中遇到问题大部分可以通过自己搜索查阅资料快速解决，业余时间也会主动学习。
- 对数据领域有着浓厚的兴趣，相信数据能产生价值。