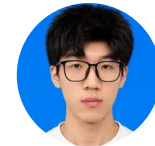


王佳伟



男 | 20岁 | 18692174538 | 2924539178@qq.com

求职意向: 数据治理 | 期望薪资: 6-8K

工作经历

神州医疗科技股份有限公司 数据治理 2026.02-2026.06

内容:

参与神州医疗与西京医院合作项目, 负责医院多源异构数据同步、调度运维、电子病历解析及入库相关工作, 涉及Oracle、SQL Server、MySQL、PostgreSQL 等数据库, 以及DolphinScheduler、DataX、FlinkCDC、Kettle等工具。

业绩:

- 根据业务需求调整原有增量抽取逻辑, 参与6年的全量历史数据重抽方案落地, 处理Oracle中文字段抽取至PostgreSQL 过程中的乱码问题。
- 维护DolphinScheduler 调度任务, 针对失败任务进行日志排查、问题定位与重跑处理, 保障同步链路稳定运行。
- 编写并部署DataX同步脚本至DolphinScheduler, 实现批量任务的定时调度与自动化执行。
- 基于Dinky编写FlinkSQL, 结合FlinkCDC实现医院Oracle到PostgreSQL 的实时数据同步。
- 编写CDR脚本, 围绕业务需求完成表结构创建及ODS宽表向更细粒度明细表的拆分, 支持不同科室的数据使用场景。
- 基于平台API 参与OCR手写病历识别流程, 处理非结构化病历图像的文本抽取结果, 辅助后续结构化解析。
- 参与OCR返回结果的JSON解析, 对识别结果进行字段提取、清洗与映射, 支撑后续数据库入库流程。
- 参与电子病历解析结果入库, 围绕结构化字段抽取结果完成表字段适配、数据写入及异常数据处理, 提升非结构化病历数据的可用性。
- 参与了解医院大数据同步与清洗整体架构, 覆盖全量/增量抽取、目标端数据清洗、分表及非结构化数据处理等环节。
- 参与编写各类电子病历的增量任务 对电子病历进行t+1的增量设计 支持业务上对昨日新增病历的及时使用。

项目经历

基于多模态模型的图片文字识别与结构化解析工作流 ai平台搭建 2026.03-2026.04

- 基于可视化 workflow 平台BISHENG搭建文件识别流程, 完成文件上传、文本提取、识别结果处理与结构化输出的全链路设计。
- 使用Python编写预处理代码, 对.docx文件进行解析, 提取文档中的文本内容, 并按固定长度切分为chunks, 提升长文本场景下的模型处理稳定性。
- 调用Qwen3-VL-8B进行图片文字识别, 结合文件内容与分块文本完成多模态信息抽取, 并通过二次代码处理拼接完整识别结果。

教育经历

西安石油大学 本科 数据科学与大数据技术 2023-2027

专业技能

- 个人信息介绍网站：www.vpymhk.cn
- 熟悉SQL 编写，能够完成多表关联、窗口函数、CTE、条件聚合等复杂查询
- 熟悉Python，能够使用pandas 进行数据处理、脚本开发及JSON结果解析
- 熟悉DataX、DolphinScheduler、Flink CDC、Linux、Kettle
- 具备非结构化病历OCR识别结果处理经验，了解平台API 调用、JSON解析及结构化入库流程
- 熟悉MySQL、Oracle、PostgreSQL、SQL Server、Hive
- 熟悉多模态大模型的应用prompt工程以及ai-agent的搭建大语言模型的使用
- 熟悉linux shell脚本的编写 以及linux中常用命令的使用