

陈磊简历



18727131430



znzchenlei@163.com



大数据开发



基本资料

姓名：陈磊

年龄：24 岁

籍贯：湖北襄阳

意向薪资：15K

工作年限：3 年

专业：计算机应用技术



个人技能

- 掌握 python 相关内置函数，面向对象思想编程
- 熟悉 python 结合其他相关大数据组件开展业务
- 熟练运用关系型数据库 mysql 的事务操作机制
- 熟练使用 linux 操作系统，熟悉 shell 编程
- 能够基于 sqoop 对关系型数据库和非关系型数据库之间的数据导入
- 熟悉 hadoop 生态圈，能够熟练操作 hadoop 生态圈各个产品
- 能够基于 hive 进行数仓分层,使用 hql 进行开发，熟练使用开窗函数
- 掌握 sparkcore、sparksql 编程，熟练使用 rdd 相关算子
- 熟悉 kafka 消息队列框架，理解 ack 和生产者和消费者，以及消费组的概念
- 熟悉 spark structured streaming 和 kafka 的整合流计算
- 熟悉 flume 采集工具、oozie 调度工具的使用



工作经历

上海携宁计算机科技股份有限公司深圳分公司 2020.3-至今 大数据开发工程师

宁波中茂科技有限公司深圳分公司 2019.6-2020.1 运维工程师



项目经验

项目一

SwanComesHome 用户画像项目

2021.10-2022-4

项目描述：

中国城镇化进程推进、提升家庭小型化、人口老龄化以及全面二孩政策等社会变迁因素为家政服务业提供潜在的消费需求，加上居民购买力有所提升，中国家政行业市场规模持续扩大。通过对潜在客户打标签的形式，准确的找到目标受众，以便在后期的服务中，更好的技术更好的满足客户的需求。

技术架构：

Sqoop+Hive+Hadoop+Spark+Kafka+Dolphinscheduler+Es+Mysql

项目职责：

- 实现 hive 数仓分层，ods 存储原始数据，dwd 明细数据，dwb 存储预聚合临时层数据。
- 使用 es-hadoop 方式，在 hive 数据库中建立 es 的外部表，使用 hql 语句导入数据到 es 中。
- 实现年龄标签计算程序，实现根据 4 级标签查询 es 数据，根据 5 级标签匹配到标签 id。
- 使用 spark dsl 和 sql 格式代码，实现根据 4 级标签查询 es 数据，根据 5 级标签匹配到标签 id。
- 其他匹配类打标签流程程序开发，使用 spark dsl 风格的代码，给用户打上规则匹配类标签。
- 实现配置文件加载、kafka 生产者、消费者、文件读取等工具类
- 实现基于离线标签匹配和统计，以及挖掘标签的基类标签流程。

项目二

T3 出行车联网大数据平台

2020.12-2021.7

项目描述：

随个全国出行打车的需求越来越高，打车人数的增加，使得打车相关数据的锐增，基于平台实时产生的各种日志和业务数据，统计当日活跃用户、订单量、交易额、区域等情况进行分析，以便更好的为乘客提供安全、便捷、专业的出行服务，建立统一的数据中心，做好针对性的数据分析，为决策层提供有效的决策支持

技术架构：

Mysql+Sqoop+hadoop+Presto+Spark+Python+Kafka+Flume

项目职责：

- 通过 flume 监听每天的日志目录，将日志文件上传致 hdfs 的目录下
- 数据 ods、dwd、dws、app 的数仓分层
- 使用 mapreduce 对数据进行清洗，生成规范数据文件上传到 hdfs
- 负责每日的地区司机活跃情况、司机接单情况、司机订单评论情况等司机相关主题指标的统计
- 进行 hql 的调优，以及相关参数调优
- 使用 spark 结合自定义 UDF 函数对部分数据进行数据清洗
- Dolphischeduler 进行任务调度

项目三

智慧餐饮管理数仓系统

2020.5-2020.10

项目描述：

美味不用等智慧餐饮服务平台，针对解决餐厅效率、客流、资金、数管理等问题，全国合作餐品品牌上百家，覆盖全国 300 多个城市，累计注册客户 2 亿，由于传统的数据库已无法满足平台所产生的庞大数据量，根据对门店相关数据信息进行维度分析，年月日活跃用户数量，年月日销售量等做出统计，相应的数据给予决策者决策支持。

技术架构：

Mysql+Hadoop+Sqoop+Hive+Spark+Zookeeper+Finbi

项目职责：

- 使用 Sqoop 将存储于 Mysql 的业务数据导入到 HDFS，并将日志数据与业务数据的分析结果从 Hive 导出到 Mysql
- 将每日新增的日志数据和业务数据经过压缩后加载数据仓库 ODS 层
- 对数据仓库的 ODS 层数据进行判空清洗，行式存储改为列式存储，并更改压缩方式，然后存入 DWD 层
- 在数据仓库 DWD 层建立并维护拉链表，记录从开始到当前所有的变化信息
- 在数据仓库 DWS 层，整合经常查询的字段组建宽表，以便之后进行统计分析，如订单主题宽表中订单相关的字段。
- 负责日活、周活、月活、年活、订单量、退单量、差评率、等订单主题相关的指标统计
- 在 SparkSql 中自定义 UDAF，以满足特殊需求



教育背景

学校

武昌首义学院

专业

计算机应用专业



自我评价

- 具有良好的沟通协作能力学习能力，乐于分享
- 有实际的大数据分布式项目经验，了解各个组件的使用场景，有离线、实时的处理经验
- 积极主动，对新技术、新事务有很好的探索欲和求知欲
- 休息时间喜欢运动和烹饪