

# 赵坤

意向岗位：数据开发  
手机号：15866867491  
邮箱：2953457349@qq.com



性别：男 年龄：26  
学历：本科 年限：4

## 个人优势

- 数据仓库架构：精通Kimball/Inmon模型，熟悉ODS/DWD/DWS/ADS分层设计。
- ETL开发：熟练使用Informatica、DataX、Kettle等工具，实现高效数据抽取与转换。
- 实时/离线数仓：掌握Lambda/Kappa架构，有Flink实时计算和Hive离线批处理经验。
- SQL优化：擅长复杂SQL编写（窗口函数、索引优化），熟悉执行计划分析。
- 关系型数据库：精通MySQL、Oracle、PostgreSQL，具备性能调优经验，底层原理。
- 大数据生态：熟练使用Hadoop、Hive、Spark、Flink、HBase、Kafka处理海量数据，编写自定义函数
- 云数仓：有阿里云MaxCompute,华为dgc,mrs等云平台实战经验。
- 编程语言：Java/Python/Scala开发数据处理脚本及自动化工具。
- 掌握机器学习基础，熟悉特征工程和模型部署流程。
- 数据治理：掌握数据质量监控、元数据管理及数据血缘追踪。
- BI工具：熟悉帆软，支持业务报表开发与可视化分析。
- 数据湖：了解Delta Lake、hudi等湖仓一体技术。
- Linux/Shell：能编写Shell脚本完成服务器日志分析及任务调度。

## 亚信科技有限公司

数据开发工程师 数据仓库开发工程师 2023.01-2025.09

## 济南银泉科技有限公司

数据开发工程师 数据仓库开发工程师 2021.09-2022.12

## 项目经历

首发集团智慧高速数据治理平台建设项目 数据开发 2024.04 - 2025.06

项目架构:Flink, Kafka, Spark, 华为DGC, Hive, GaussDB, HDFS, Python, Java, FineReport

项目描述:

项目基于华为DGC数据治理中心构建高速公路智能数据中台，整合ETC收费系统、路网监测系统、设备运维系统等8大核心业务系统数据，构建"采、存、算、管、用"一体化的数据治理体系。平台日均处理ETC交易记录1.2亿条、物联网设备数据45TB，通过实时数据分析和离线批处理相结合的方式，为路网调度指挥、收费稽核管理、设备预测性维护等业务场景提供数据支撑，显著提升高速公路运营管理效率和智能化水平

个人职责:

- 实时数据处理:开发Flink实时处理作业实现ETC交易秒级处理，通过优化checkpoint机制将延迟控制在800ms以内，支撑路网实时运营决策;
- DGC平台开发:基于华为DGC平台开发Python数据加工脚本，构建完整的数据处理流程，实现100%自动化调度，日均稳定处理1.2亿+交易记录;
- 批处理优化:设计DGC批处理作业完成每日数据清洗和汇总，通过参数调优使加工效率提升60%，关键数据服务响应时间控制在300ms以内;
- 数据质量管理:编写数据质量检查规则，建立完善的质量监控体系，确保数据准确率达99.9%，异常处理时效提升80%;
- 系统运维保障:监控作业运行状态及时处理异常，维护数据血缘关系保障可追溯性，通过资源分配策略优化降低整体运行成本30%;

光大银行数字金融BI系统2023年升级改造项目 数据开发

2023.07-2024.04

项目架构：Impala，Hive，Kudu，GaussDB，teradata，MapReduce，HDFS，python，finereport

项目描述：

光大银行数字金融数仓项目致力于从整体提高中国光大银行电子渠道为客户服务的质量，提高客户对光大银行电子渠道的忠诚度，提高电子渠道的竞争能力，最终提高电子渠道的收益水平，需要通过对客户，渠道和产品等方面进行细分和观察，即开展电子渠道数据挖掘(BI)项目，为电子渠道产品开发提供参考。业务方面：统计业务指标包括支付类的（云闪付，微信支付，支付宝，京东白条等不同渠道的支付笔数金额统计）查看支付类的业务开展情况理财存款产品购买类的包含（公募基金理财，私募理财，保险，存款产品，代销理财等笔数金额同环比统计）查看产品购买情况权益类活动的指标包括（权益客户领取活动，领取的数量，兑换权益的数量，库存等指标）查看活动情况

个人职责：

- 1，需求分析，对业务提出的需求进行分析，包括需求的内容，背景和目的，首次开发还是修改，报表权限，风险点等
- 2，需求设计，对分析完的需求进行设计，包括取数据来源，数据处理逻辑捋顺，保留策略，跑批策略，表及字段的设计
- 3，需求开发，主要对所做需求进行代码开发封装成 python 脚本，在 Linux 服务器进行脚本测试，以及写 mapping 规则

4，对一些脚本进行优化运行速度通过优化 SQL 代码如：大小表 join，创建多次使用的临时表，优化建表语句索引等，提升了脚本运行速度从10分钟优化到5分钟

以内。

- 5，脚本迁移，从 frq 迁移到帆软，需要固话成 python 跑批脚本，历史数据保留在高斯，再帆软从高斯取数据，进行数据核对。
- 6，修改由于金额越来越大导致数据类型不够用的情况，修改帆软报错情况，如取错了数据集等导致报错。

光大银行1104 数据开发

项目架构：Oracle kettle shell sqlldr 帆软 2022.08-2023.06

项目描述：

我们主要做的就是贷款业务。早期国家由于对银行监管不严格所以会造成很多非法贷款，为了避免损失所以要求银行向中国银监会做业务报送监管。提取大量 SQL 公共逻辑，封装存储过程增加逻辑复用的同时减少计算时常。

个人职责：

- 1.设计 Kettle 任务，将处理好的数据获取 MD 5校验，以保证数据的完整性。在将数据采取压缩在发送到 Oracle 服务器，提升数据的准确性以及数据传输时间。
- 2.创建临时表与正式表，避免加载数据时构建索引时间过长。
- 3.编写存储过程实现动态创建分区、刷新索引。提升工作效率。
- 4.对贷款业务划分主题、例如贷款质量迁移、各项贷款分析、各合作金融公司贷款情况。
- 5.针对于各主题的 SQL 提取公共逻辑，提升数据计算时间70%以上。
- 6.编写 Shell 脚本、设置定时调度任务

中国农业银行贷款离线数仓 数据开发

项目架构：hadoop，spark，impala，kudu，hive，hbase，cdh，elasticsearch 2021.08-2022.07

项目描述：

本项目是从银行的核心系统，用户系统，借贷系统，销售系统，财务系统，日志流等各个子系统中提取数据，分析用户借贷情况，还款和逾期情况，每个银行分店的借贷流程，贷款审批状态等，将不同的数据抽取到 hive 数据库。并对数据进行清洗面临的主要问题是集成过程中保护敏感数据同时保持数据完整性和准确性。通过数据集成过程符合相关的法律法规要求，进行合规性审计和风险评估。提高了数据传输效率和提高数据安全能力，有效防止了数据泄露和跨部门使用权限等一系列问题。

个人职责：

- 1.参与数仓分层设计，参与维度建模维度抽取
- 2.开发贷款任务，实现小额贷款多维度指标统计
- 3.优化和实现 hive 任务，解决数据倾斜问题和 join 性能过慢问题
- 4.对任务依赖关系进行设定，保证任务执行逻辑正确性
- 5.梳理项目文档和任务 mapping 逻辑，保证数据粒度唯一性
- 6.使用 kettle 实现数据 ETL，开发对数脚本保证数据质量

教育经历

