姜植蔡

求职意向：数据架构师

北京 | jiangzhicong2516@163.com | 17600192516

个人简介

5 年以上数据开发与数据架构经验，精通 Python / SQL 数据管道开发，独立设计并落地企业级数据平台（ODS / DWD / DWS 分层），具备日均百万级数据处理实战经验。数学专业背景，熟悉维度建模与数据仓库设计方法，擅长将业务需求转化为可扩展的数据架构方案。

专业技能

- 数据开发：Python（pandas / numpy）、SQL、Java，5 年 ETL 管道设计与开发经验
- 数据架构：数据分层建模（ODS / DWD / DWS / ADS）、维度建模、数据仓库设计
- 数据采集：akshare、爬虫、API 对接，定时增量采集方案设计
- 数据存储：MySQL、DuckDB、华为云 OBS；了解 Hive、ClickHouse
- 数据计算：离线 ETL、实时数据处理、A/B 实验数据支撑
- 可视化与推送：Flask 看板开发、ECharts、钉钉机器人推送
- 算法基础：Word2Vec、TF-IDF、LR，特征工程与模型评估（AUC）

项目经历

A 股量化数据平台 | 数据架构 / 全栈开发（进行中） | 2026.10 – 至今

- 独立设计平台整体数据架构：采集层（akshare 定时拉取 A 股日 K 线、资金流向）→ 数仓层（DuckDB，ODS / DWD / DWS 分层建模）→ 应用层（策略信号、Flask 可视化看板）
- 实现交易日历感知的增量采集调度，每日收盘后自动拉取，当日数据 T+1 进入 DWD 清洗层
- DWS 层沉淀技术指标（MA5/10/20/60、涨跌幅、换手率、量比），支撑均线突破、动量等多策略信号计算
- 策略信号通过钉钉机器人每日推送；看板展示策略组合收益率曲线与沪深 300 对比，直观呈现是否跑赢大盘
- 规划中：成交截图 OCR 识别（PaddleOCR），自动计算实盘收益率并与策略收益联动分析

企业级数据运营平台 | 数据架构 / 后端开发 | 2023.2 – 2023.12

- 从零独立完成平台数据架构设计：定义 ODS 贴源层、DWD 明细层、DWS 汇总层三层模型，支撑运营指标实时查询
- 设计并实现海量日志数据的全链路 ETL：采集 → 清洗 → 入库 → 聚合 → 可视化，日均处理百万级日志数据
- 基于 SQL 与 Python（pandas / numpy）构建数据清洗规则库，处理缺失值、异常值、重复数据，数据准确率达 99.5%
- 通过精心设计的数据分析模型，从日志中提取关键业务指标，为管理层决策提供数据支撑，获优秀治理团体奖

数据质量治理专项 | 项目主导 | 2024.1 – 2024.12

- 建立数据质量监控指标体系：完整性、准确性、时效性三维度，覆盖 600+ 数据源
- 设计异常数据识别规则与自动化工单流转，将异常数据单从月均 200+ 降至 70+，降幅 65%
- 推动数据标准化：统一账号、权限、证书等多维数据的口径与命名规范，沉淀数据字典

推荐系统数据体系 | 算法工程师（数据方向） | 2018.7 – 2023.1

- 负责推荐系统数据底座：用户行为日志的采集、存储、清洗与管理，制定多路召回的数据策略

- 采用 Word2Vec 计算物品 Embedding，基于用户点击序列构建兴趣模型，召回率从 2% 提升至 5%（提升 150%）
- 基于 TF-IDF 对点击序列加权重排序，排序模型 AUC 达 75%，显著提升推荐转化效果

工作经历

- 北京外企德科人力资源有限公司 | 数据开发工程师 | 2023.2 – 2025.4
- 北京盛创恒达房地产开发有限公司 | 算法工程师（数据方向） | 2018.7 – 2023.1

教育背景

- 中国政法大学 | 国际贸易学硕士 | 2015.9 – 2018.6
- 中国地质大学 | 数学与应用数学学士 | 2009.9 – 2013.6

JIANG Zhicong

Target Role: Data Architect

Beijing, China | jiangzhicong2516@163.com | +86 17600192516

Professional Summary

Data engineer with 5+ years of experience in data pipeline development and data architecture. Designed and delivered enterprise data platforms end-to-end (ODS / DWD / DWS layered modeling) with proven million-records-per-day processing at scale. Mathematics background with solid dimensional modeling and data warehouse design skills; adept at translating business needs into scalable data architecture.

Technical Skills

- Data Engineering: Python (pandas / numpy), SQL, Java; 5 years of ETL pipeline design and development
- Data Architecture: Layered modeling (ODS / DWD / DWS / ADS), dimensional modeling, data warehouse design
- Data Ingestion: akshare, web crawlers, API integration; scheduled incremental collection design
- Data Storage: MySQL, DuckDB, Huawei Cloud OBS; familiar with Hive and ClickHouse
- Data Computing: Offline ETL, real-time data processing, A/B testing data support
- Visualization & Alerting: Flask dashboards, ECharts, DingTalk bot push notifications
- Algorithms: Word2Vec, TF-IDF, logistic regression; feature engineering and model evaluation (AUC)

Project Experience

A-Share Quantitative Data Platform | Data Architecture / Full-Stack (In Progress) | 2026.10 – Present

- Independently designed the platform data architecture: ingestion (akshare scheduled pull of daily K-lines and capital flow) → warehouse (DuckDB, ODS / DWD / DWS layered modeling) → application (strategy signals, Flask visualization dashboard)
- Built trading-calendar-aware incremental scheduling: auto-pull after market close daily, landing in DWD cleansing layer on T+1
- DWS layer materializes technical indicators (MA5/10/20/60, price change %, turnover rate, volume ratio) powering multi-strategy signal computation (moving-average breakout, momentum)
- Daily strategy signals pushed via DingTalk bot; dashboard compares strategy portfolio returns against CSI 300 to visualize outperformance
- Planned: trade-screenshot OCR (PaddleOCR) to auto-compute live returns and correlate with strategy performance

Enterprise Data Operations Platform | Data Architecture / Backend | 2023.2 – 2023.12

- Designed the platform data architecture from scratch: three-layer model (ODS source / DWD detail / DWS summary) supporting real-time operational metrics
- Built the full log-data ETL chain: collection → cleansing → warehousing → aggregation → visualization, processing millions of log records daily

- Developed data-cleansing rule library with SQL and Python (pandas / numpy) handling missing values, outliers, and duplicates at 99.5% data accuracy
- Extracted key business indicators from massive logs through analytical models, grounding executive decisions in data; won team governance award

Data Quality Governance Initiative | **Project Lead** | **2024.1 – 2024.12**

- Established data-quality monitoring across completeness, accuracy, and timeliness dimensions covering 600+ data sources
- Designed anomaly-detection rules with automated ticket routing, cutting anomaly tickets from 200+/month to 70+ (65% reduction)
- Drove data standardization: unified metric definitions and naming conventions across account, permission, and certificate data; published data dictionary

Recommendation System Data Infrastructure | **Algorithm Engineer (Data)** | **2018.7 – 2023.1**

- Owned the recommendation data foundation: collection, storage, and cleansing of user behavior logs; defined multi-channel recall data strategies
- Computed item embeddings with Word2Vec and built user-interest models on click sequences, lifting recall from 2% to 5% (+150%)
- Applied TF-IDF weighted re-ranking on click sequences; ranking model reached 75% AUC, significantly improving conversion

Work Experience

- FESCO Adecco, Beijing | Data Development Engineer | 2023.2 – 2025.4
- Beijing Shengchuang Hengda Real Estate | Algorithm Engineer (Data) | 2018.7 – 2023.1

Education

- China University of Political Science and Law | Master of International Trade | 2015.9 – 2018.6
- China University of Geosciences | B.S. in Mathematics and Applied Mathematics | 2009.9 – 2013.6