

# 张勇

## 资深数据架构师

北京 · 13911083192 · zyjarge@163.com · blog.zhang-yong.site · github.com/zyjarge

### 个人简介

- 15 年数据架构与数仓建设经验，主导过 PB 级数据平台、实时数仓与公司级数据中台建设
- 擅长数据治理、数据建模与数据团队管理，曾管理 15 人数据技术团队
- 深度实践 LLM 应用与 Agent 架构设计，完成数据治理 Agent PoC，并将日常建模、数据纠错、质检等流程沉淀为 Skill + 工作流的 AI 数据开发模式

### 工作经历

#### 华耐家装

2023.04 – 2026.07

资深数据架构开发

- 负责公司各业务线的数据开发工作，并作为 AI 开发布道者推动 AI 在数据团队落地
- 从 0 到 1 搭建公司数据开发工作流，覆盖数据建模、代码开发、工作流编排、数据质量监控、数据治理等各个环节
- 将 AI Agent 应用到数据开发工作流各环节：设计并实现数据治理 Agent，覆盖元数据自动标注、血缘/影响分析、智能问数（Text-to-SQL）等场景
- 完成 PoC 验证：覆盖 200+ 核心表的元数据自动标注，标注准确率 85%+，智能问数日均为 20+ 名数据团队成员提供服务
- 将日常建模、数据纠错、数据质检等操作抽象为可复用 Skill，通过工作流编排纳入日常数据开发流程
- 技术栈：Airflow, dbt, OpenMetadata, DataX, Grafana + Prometheus; LLM (Codex/DeepSeek 等) + Prompt Engineering + Function/Tool Calling

#### 去哪儿旅行

2021.04 – 2023.03

资深数据架构师

- 主导设计并实现公司级数据中台建设，服务 100+ 业务线，日均处理数据量 10TB+，支持实时数据分析和决策
- 设计实现实时数据处理平台，基于 Flink + Kafka 构建，处理延迟 <5s，支撑订单、用户行为等高并发场景
- 推进数据治理项目，基于 Atlas + OpenMeta 构建数据资产管理平台，数据使用效率提升 200%，运维成本降低 40%
- 主导数据架构升级，引入 Data Vault 2.0 模型，提升数据加工效率 50%，数据一致性达到 99.9%

#### 北京风行视频有限公司

2013.01 – 2021.03

数据中心高级经理

- 搭建公司统一数据平台，整合 7+ 业务系统数据，日均处理 PB 级数据，支持千万级用户数据分析
- 设计并实现实时用户画像系统，基于 Spark Streaming 构建，支持秒级标签更新，提升广告转化率 35%
- 主导数据仓库重构，采用 Lambda 架构，优化查询性能提升 8 倍，支持亿级数据秒级查询
- 管理 15 人技术团队，建立数据开发规范和评审机制，团队效能提升 80%

#### 亚信（中国）技术有限公司

2010.03 – 2012.12

高级工程师

- 负责中国移动省级数据仓库建设，日均处理话单数据 3 亿+，查询性能提升 5 倍
- 设计实现用户行为分析平台，支持亿级用户精准营销，营销转化率提升 40%

- 主导 ETL 框架优化，处理效率提升 3 倍，支持增量更新和实时计算

项目经历

风行大数据平台 & 数据中台2019.01 – 2021.03

- 主导大数据平台向数据中台转型，集成多维分析、元数据管理、数据质量和数据安全等核心功能
- 数据需求响应时间从小时级优化至分钟级，服务效率提升 80%；通过数据复用和自助分析降低数据开发成本 60%；数据质量达标率提升至 99%
- 技术栈：ClickHouse, Presto, Kylin, Atlas, DataHub, Griffin

风行商业经营分析系统2018.03 – 2020.08

- 构建企业级广告投放数据分析平台，采用星型模型构建多维分析体系，支持亿级数据秒级查询
- 平台日均处理广告数据 500GB+，支持 100+ 位商业化人员的日常运营决策；帮助广告运营团队提升投放效率 40%，广告转化率提升 35%
- 技术栈：Kafka, Storm, Hive, ClickHouse, Redis, Grafana

风行数据大屏2019.06 – 2020.03

- 设计并实现企业级实时数据可视化平台，基于 Flink + Kafka 构建实时计算链路
- 系统支持同时展示 50+ 核心业务指标，数据更新延迟 <1s，稳定性达 99.99%，支持 7×24 小时持续运行
- 技术栈：Vue3, ECharts, SpringBoot, WebSocket, Flink, Kafka

技能专长

数据架构与平台

- 精通 Hadoop、Spark、Flink 等大数据计算框架
- 精通数据仓库模型设计（Data Vault、维度建模）
- 熟悉 dbt、DataX 等数据开发与集成工具，OpenMetadata 元数据平台
- 熟悉数据质量管理体系建设、元数据、数据血缘管理
- 熟悉 Grafana + Prometheus 监控体系、数据安全和隐私保护策略

开发技能

- 精通 Java、Python 等后端开发语言
- 掌握 Docker、K8s 等容器化技术
- 熟悉 TypeScript、Next.js 等前端技术
- 熟练使用 Git、Airflow 等开发工具

AI 与 Agent 开发

- 掌握 LLM 应用开发：Prompt Engineering、RAG、Function/Tool Calling
- 掌握 Agent 架构设计，实践基于 Skill + 工作流编排的数据开发 Agent
- 完成数据治理 Agent PoC：元数据自动标注、血缘/影响分析、智能问数（Text-to-SQL）
- 熟悉 Dify、Coze 等 AI 应用平台

团队管理

- 多年技术团队管理经验，带领 15 人数据团队完成多个重大项目
- 建立完善的技术评审和知识分享机制
- 具备数据业务价值分析能力，善于数据价值挖掘与变现

教育背景

北京航空航天大学 · 经济管理学院2019.05 – 2021.09  
工商管理硕士（MBA）

北京信息科技大学 · 计算机与自动化系2001.09 – 2005.07  
工学学士