

# 曹文栋

求职意向：数据开发/数据分析 | 随时到岗

年 龄 27 岁 性 别 男  
电 话 18906724954 微 信 C8651402  
籍 贯 湖南常德 邮 箱 18948191209@163.com



## 教育背景

2016-09 ~ 2020-06

河海大学

(本科)

## 个人技能

### • I 关系型数据库与 SQL 应用

熟悉 Oracle、MySQL 等关系型数据库，能独立使用 SQL 语言完成业务需求的数据查询、更新、统计等操作，且具备 SQL 语句性能调优能力，保障数据操作效率。

### • I Hadoop 生态核心组件掌握

熟悉 HDFS、YARN、MapReduce 的架构原理与工作机制，了解 ZooKeeper 分布式协调服务的原理与架构，了解 Hive 数据仓库的工作原理。

### • I 实时计算引擎与湖仓一体工具

熟悉 Doris、Analytic DB 等实时性湖仓一体数据平台的特性，了解 Spark、Flink 实时计算引擎的核心逻辑，可使用阿里云 Flink 实现数据实时采集、处理与分析，满足实时业务需求。

### • I ETL 工具

了解 Sqoop、Flume、Kettle、DataX 等主流 ETL 工具的使用方法，能根据业务场景选择合适工具完成数据抽取、转换、加载全流程。

### • I 任务调度工具应用

熟悉 Azkaban 调度工具的核心功能与操作流程，能使用该工具进行数据处理任务的创建、配置、调度与监控，保障数据任务按计划稳定执行。

### • I Linux 系统与 Shell 脚本开发

熟悉 Linux 系统的基本操作与常用命令，能编写并封装 Shell 脚本，实现数据处理、任务执行等操作的自动化。

### • I 数据可视化工具应用

了解 Tableau、FineBI 等可视化工具的基本操作流程，具备使用帆软 Report 进行报表开发的实战经验，支撑数据成果展示。

### • I 分布式系统理论与数据治理

掌握分布式系统核心理论，包括 CAP 定理和 BASE 理论的思维逻辑，了解数据治理的相关思想。

### • I 数据仓库维度建模

具备维度建模能力，熟悉星型模型、雪花模型和星座模型的设计逻辑与建构方法，曾实际运用多模型完成数据仓库维度层的搭建，熟悉电商漏斗分析模型，支撑数据分层管理与分析。

## 工作经验

2022-09 ~ 2025-09	任拓数据科技有限公司	大数据开发工程师
2020-07 ~ 2022-08	上海璞恒新能源科技有限公司	数据开发工程师

## 项目经历

2025-01 ~ 2025-09	清美鲜食实时数仓选型与应用
-------------------	---------------

### 项目背景

在离线数仓稳定运行后，公司不满足于离线数仓对核心指标的T+1展示，希望实现实时订单量、实时销售额、库存告警商品数等关键数据的实时可视化，构建支撑业务实时监控的技术体系。

### 工作内容

**【调研与选型】**市面成熟开源实时数仓方案多采用 Flume+Kafka+Flink 架构实现，但该方案需投入更多硬件资源，且后期运维成本较高；综合团队人力配置、成本控制及业务落地效率，最终确定采用阿里云一体化服务，降低技术复杂度与运维压力。

**【数据传输与存储】**通过阿里云 DTS 工具实现数据实时传输，先将业务数据导入 RDS 数据库构建 ODS 层（原始数据层）；再借助 Flink 计算引擎将数据清洗后导入 DataHub，建立 DWD 层（明细数据层）；后续抽取离线数仓 DIM 层（维度数据层）与 DWD 层数据，联合导入 Analytic DB 数据库，形成支撑报表分析的 ADS 层（应用数据层）。

**【报表展示】**采用阿里云 DataV 作为可视化报表工具，直接以 ADS 层数据为数据源，生成地区销售排名、商品实时销量、销售额趋势等图表，满足业务实时监控需求。

2023-09 ~ 2024-12	清美鲜食数仓搭建
-------------------	----------

### 项目背景

依托线上业务，应对业务增长带来的MySQL 业务库和用户行为日志整合难、数据分散产生数据孤岛、历史数据分析滞后、运营日报 / 财务溯源 / 商品销量排名等报表需求频繁的问题，构建一套稳定、可扩展、分层清晰的Hadoop离线数据仓库，实现对清美鲜食全产业链（供应链、生产、物流、门店、会员、财务）历史数据的统一整合与分析，支撑T+1的经营分析、报表生成与决策支持。

### 工作内容

**【分层架构落地】**按 ODS-DWD-DWS-DWT-ADS 规范搭建，参与维度建模，实现维度退化，设计事务型、周期型、累积型等多张事实表；DWS 层构建用户/设备主题日活宽表，DWT 层整合全量历史数据，ADS 层开发活跃设备数、会员行为漏斗分析等报表。

**【核心模型设计】**开发用户拉链表解决缓慢变化维度问题，设计商品维度表关联 SPU / 分类 / 品牌信息，构建订单累积型事实表追踪从创建到退款的全生命周期，保障数据一致性与业务追溯。

**【数据调度】**通过 Azkaban 编排全链路调度（从 MySQL 抽数到 ADS 导出），用 Sqoop 将 ADS 层数据导出至 MySQL 供报表使用。

**【优化情况】**后续根据实际业务情况不断进行更新，让取数逻辑更贴近业务实际情况，为业务发展不断赋能。

2022-09 ~ 2023-08

电商数据迁移及数仓搭建

## 项目背景

电商平台为解决 Oracle 数据库数据存储压力大、数据与业务分析需求适配性不足的问题，同时搭建统一数据处理与查询体系，开展 Oracle 数据库到 Hadoop 生态的迁移及 Hive 数仓搭建工作，实现原始数据安全存储、高效处理与业务指标计算。

## 工作内容

**【数据迁移】** 参与 Oracle 数据库全量与增量数据迁移至 Hadoop 生态的 HDFS 存储过程，通过数据校验机制保障原始数据完整性，最大限度减少迁移过程中的数据丢失情况。

**【数仓搭建】** 参与 Hive 离线数仓搭建，结合业务分析需求开展维度建模工作，根据不同业务场景（如用户分析、商品分析）选择适配的建模方案，确保数仓结构贴合业务使用场景。

**【数据清洗】** 基于 Hadoop 大数据处理框架，对迁移至 HDFS 的原始数据进行清洗、去重、格式转换等预处理操作，提升数据质量，为后续分析奠定基础。

**【数据查询】** 通过 HQL 语言完成各业务主题表的数据导入与查询操作，结合用户、订单、商品等核心业务主题，完成活跃用户数、商品周转率等关键指标的计算。

**【可视化情况】** 配合业务团队开展可视化报表开发工作，提供清洗后的标准数据与计算指标，为业务团队打造直观的数据展示工具，支撑业务决策。

2020-07 ~ 2022-08

新能源锂电池租赁项目

## 项目背景

随着新能源汽车市场快速发展，电池成本高、充电不便等问题催生了电池租赁新模式。为提升电池资产利用率、优化用户换电体验，企业需对海量电池使用、租赁、运维数据进行统一管理。当前数据分散在各系统，缺乏实时监控与深度分析能力，难以支撑精细化运营。亟需构建一个集电池状态追踪、用户行为分析、故障预警于一体的新能源电池租赁大数据平台。该平台将实现电池全生命周期管理，助力企业降本增效，提升服务竞争力。

## 工作内容

**【数据与报表开发】** 整合订单、押金、租金、保险、服务单数据构建订单宽表；制作专项报表，同步梳理财务数据，为收入溯源查询奠定基础。

**【数仓分层搭建】** 针对重复开发问题，设计并落地数仓 DWD 层，构建核心维度表（租户、城市、BD、门店）与事实表（订单、电池库存、电柜归属、调度记录），统一数据口径以减少重复开发。

**【模型与代码优化】** 迭代优化中期不合理的 DWD 层模型，同步更新电池租赁、资产调度相关指标计算逻辑，确保数据准确性。

**【平台支撑与数据保障】** 提供接口或 SQL 支持公司业务大屏、BI 驾驶舱的指标需求。

## 自我评价

1. 工作积极认真，细心负责，善于在工作中提出问题、发现问题、解决问题，有较强的分析力。
2. 勤奋好学，踏实肯干，坚毅不拔，吃苦耐劳，喜欢迎接新挑战。
3. 善于沟通，理解能力强，能很好的理解业务需求。
4. 接受公司外派和商务应酬安排。