

龚文东

✉ gwd163nom@163.com · 📞 +86 18810373739 · 🗣️ gwdweixinom · 🌐 gwdgithubnom

🏠 (出生) 1992 · 🏠 (户口) 北京 · 🏢 (职称) 工程师

工作经历

滴滴

2023/07 – 至今

高级数据研发工程师，主要业绩：(1) 拉齐各业务方向的指标口径，梳理与完善指标体系的建设；梳理与优化网约车特区数据资产 (2) 构建网约车全局应用模型，实现特区数据任务、数据资产的成本优化，丰富特区数仓在自选车业务线上的数据建设；建设特区实验效果评估的数据体系；(3) 特区架构与模型优化，完成数据生产任务生产与回溯成本的 20%~60% 的成本优化，10%~80% 的性能优化

小米

2018/04 – 2023/07

高级软件研发工程师，主要业绩：(1) 从 0 到 1 建设 Zili 数据的采集、仓库、应用服务与平台，通过搭建数据分析平台，完成了对业务运营分析与活动策划的支持 (2) 再对数据系统流程、架构的迭代优化，满足了业务整体对数据的需求 (3) 以及平台组件进行 Bug 修复与优化，保障了业务数据的生产需要与稳定 (4) 最后完成短视频业务的公有云的平滑改造与迁移，提升了团队人才的技术能力

教育背景

中央民族大学

2015/09 – 2018/06

计算机科学与技术 (硕士研究生) GPA: 3.85/4

长春工业大学

2012/09 – 2015/06

计算机科学与技术 (本科) / (统招)

相关技能

- 编程语言: (熟练) SQL/Java/Scala, (熟悉) Python/Shell, (了解) Go/JS
- 组件工具: Spark/ Flink/ Hbase/ Kudu/ Pegasus/ Iceberg 等
- 工程管理: PlantUML/ Jira/ MS Project/ Git/ Gradle/ Maven

项目经历

滴滴-网约车数据的数据研发与治理

2023/07 – 2024/07

该项目是数仓方面的项目，主要围绕业务，开展数据建设、数据治理与数据赋能落地工作。主要角色：数据研发

- 数据建设：参与滴滴核心业务的数据建设，优化重点业务的数据架构，实现打造适应滴滴网约车业务需要的数仓体系效果 (负责围绕网约车全局应用层 & 特区数仓)；
- 数据赋能：承担业务需求的设计与实现，梳理与抽象核心业务流程，通过构建数据架构与模型与开发数仓中间层和数据应用产品，实现数据的高效生产、加工计算，并沉淀出业务通用分析与决策的数据资产；
- 数据治理：优化特区耗时、耗成本的任务链路，解决特区回溯成本高的痛点问题，在不断完善数据体系与流程，持续提升了现有数仓建设的质量和效率；

小米-Zili 数据中心建设与开发

2020/08 – 2023/02

Zili 是一款印度的千万日活短视频软件，一款小米成功海外出圈短视频 APP。Zili 数据中心是为其提供内容挖掘与运营、达人分析、内容监控等功能的数据项目。本人在该项目中主要承担了数据建设与

数据赋能的架构设计与实施工作。主要角色：负责人

- 总体设计：负责调研与制定数据标准与规范，规划与设计 Zili 的数据架构，并通过搭建数据产品、数据服务等方式，实现数据驱动业务增长的目标
- 详细实现：建设 Zili 离线实时数据仓库，并从 0 到 1 完成搭建数据体系，组织与安排人员参与完成业务场景需求多元化的数据支持
- 组件维护：对 Kudu/Hbase/Spark/Flink 等底层基础 Client Api 组件的二次开发与维护
- 构建业务的数据指标体系，并对数据架构进行迭代与优化，完成数据统计服务、数据报表生产服务的构建支持，对数据赋能进行多元化输出
- 数据治理：维护业务的数据资产，梳理与完善业务核心指标体系与主题建设
- 技术攻关：解决数据中心系统建设中遇到的阻碍，对组件进行封装，支持服务端、数据服务场景中 Api 使用，保障 OLAP 查询服务的构建方案落地，对 Kudu/Hbase/Spark/Flink/DolphinScheduler 等集群进行数据开发的脚本维护

小米-业务数据仓库建设与治理

2021/04 – 2023/07

该项目是数仓方面的项目，主要围绕业务，开展数据建设、数据治理与数据赋能落地工作。主要角色：负责人/参与者

- 数据架构：负责阅读数据仓库、收音机数仓、经营分析、小米之家运营分析数仓的架构设计与验证，以及项目实施落地，项目中后期维护与治理
- 数据建设：离线实时的数据 ETL 工作，为业务提供定制化的数据支持，并优化任务的性能，实现业务个性化的数据爬取 & 参与埋点规范建设与实施 & 基于 Flink 完成定制化的数据接入 Hdfs /Iceberg /Hbase /Pegasus /Doris 等
- 数据赋能：指标体系搭建与实施 & 基于 Hive/Spark/PySpark 完成日常的业务数据跑数与分析统计 & SQL 模型开发 & 看板建设与报表 & 报告输出 & 数据服务搭建与维护

小米-Zili 短视频数据分析平台

2020/08 – 2023/02

基于 SpringMVC+Redis+MongoDB+ES+Hbase 主要角色：负责人

- 负责 Zili 数据分析平台建设，面向产品、运营提供交互式/可视化分析报告、报表的平台
- 负责项目系统的需求分析、设计，并主导完成详细设计和编码任务，确保项目的进度和质量
- 负责项目系统以及底层 Dao/Service 等模块的设计与开发，维护与优化 Spark/Flink 的数据报表处理模块
- 针对服务故障的技术攻关，以及平台服务模块的架构维护与改造

小米-Zili 数据中心的迁移与改造

2021/10 – 2022/03

该项目是对 Zili 进行公有云的迁移的项目 主要角色：负责人

- 数据中心的组件调研与改造，确保数仓与服务的平滑迁移，完成公有云的组件对内部服务的兼容，保证服务的稳定性
- 制定迁移的方案，推进与实施，解决中途遇到的技术难题，迁移数据中心到 Aws/Gcp，推动 Zili 整体的项目顺利迁移

基于 Dhash 的短视频去重服务

2020/11 – 2023/02

Python 项目，一种基于 MongoDB 的 Dhash 短视频去重算法服务 角色：负责人

- 负责视频去重服务的维护与开发
- 负责去重服务的私有云迁移公有云 (Aws cloud/Google Cloud) 的改造

海外 MIUI 应用的数据建设与服务系统开发

2018/03 – 2020/08

该项目是面向 MIUI 应用进行数据采集、数据集成、数据分析工作。主要角色：数据研发

- 负责构建部门数据采集系统，包含不限于流程设计、模块规划、逻辑架构设计与物理架构选型与实施；
- 拓展数据采集系统部门贡献，保障数据采集系统的应用方需求实现，支持海外浏览器 & 游戏 & 画报等业务的数据爬取，以及日常运营活动主题的数据推荐工作；
- 支持产品运营个性化专项洞察分析与主题的挖掘

🏆 个人荣誉

- 曾获 **2024 集团奖 数据治理优秀团队奖**
- 曾获 **2022 部门奖 数据治理优秀团队奖**
- 曾获 **2021 集团奖 小米杯 CTF**
- 曾获 **2021 集团奖 青年工程师**
- 曾获 **2020 部门 优秀员工**

📖 其他方面

- 技术博客: <http://gwdgithubnom.github.io/>
- 热爱社区，拥抱变化，相信相信的力量