

徐鹏飞

17772743579 | sherlock.xpf@gmail.com | 杭州 |
大数据工程师 | www.xufei.com | xuf-95.github.io/xufei-wiki/

Summary

- 拥有超过3年的数据开发领域工作经验，具备大型数据工程经验和企业级数据治理（DCMM）经验，有效提升数据处理效率达30%以上
- 熟悉Java、Python和Shell编程语言，具备在Scala语言上的基础理解，能够灵活运用多种编程语言解决复杂数据处理问题
- 熟练掌握离线数仓开发及数据建模技术，具有丰富的报表开发实践，能高效实现数据洞察与业务决策支持
- 深入了解并实操MapReduce、Spark、Flink和Yarn技术框架，结合Hive、DataFrame、Pandas等进行高效数据查询与分析，优化数据处理流程
- 对分布式系统架构、批处理及流处理架构有深刻理解，包括存储、计算和消息队列等核心组件，确保数据开发项目在高并发环境下的稳定运行
- 擅长使用阿里大数据生态圈内的RDS、MaxComputer、Dataphin、Hologres、EMR及QuickBI等大数据开发工具，创建完整的项目工程及数据治理和数据分析，快速进行POC项目开发和交付

Experience

移通科技（杭州）有限公司

2021.11 - 2023.09

数据架构师

杭州市

- 作为核心架构师，主导设计企业级数据中台解决方案，实现跨部门数据整合与共享，打造统一的数据服务能力，支撑政务数字化转型
- 领导团队完成DCMM数据管理能力成熟度认证项目，建立数据治理框架和数据标准规范，提升企业数据资产管理水平，实现数据价值最大化
- 设计并实施数据质量管理平台，构建包含数据探查、质量监控、问题追踪、评估报告的全流程质量管理体系，数据准确率提升40%
- 优化数据开发全生命周期管理，开发自动化工具提升团队效率，设计标准化开发规范和流程，开发数据建模和开发工具，使团队整体效率提升20%+
- 深度参与数据价值挖掘，通过机器学习算法和高级分析模型，为政务决策提供数据支撑，构建预测分析模型，设计可视化分析报表，直接支撑多个重要决策项目

软通动力信息技术（集团）股份有限公司

2019.10 - 2022.11

大数据开发工程师（海外POC项目交付）

杭州市

- 作为核心技术专家，负责阿里云海外客户的数据迁移及架构设计，提供端到端的技术支持和最佳实践方案，确保客户业务平稳上云
- 主导制定客户数据中台建设方案，基于业务场景设计数据架构，完成包括数据采集、存储、计算、治理、可视化在内的全链路技术选型和容量规划
- 独立完成多个海外大客户的POC交付，通过深入分析业务痛点，设计创新解决方案并进行现场演示与技术答疑，实现POC到商业化转化率90%+

浩鲸云计算科技股份有限公司

2018.08 - 2019.10

大数据开发工程师（海外POC项目交付）

杭州市

- 主导大规模数据架构设计与技术选型，负责需求分析、数据探查及上云方案设计，成功完成TB级数据平台规划与落地，平台稳定性99.9%
- 设计并实现企业级数据仓库分层架构(ODS/DWD/DWS/ADS)，构建标准化数据模型，提供可复用的数据服务，支撑跨部门数据协作需求
- 开发高性能ETL处理框架，实现海量数据的采集、清洗、转换和加载，优化数据处理流程，数据处理效率提升50%，保障数据及时性达99%

Projects

- 项目背景：响应数据化政务改革，构建改市政府区域性数字政府平台大数据底座，整合多方异构数据源数据，涉及15+政府部门及100+的省市回流数据，包含经济、民生、医疗、就业、助学、瓜农、工坊、乡贤等多维度分析的区域数据模型，打造服务300万+市民的智慧政务决策系统，项目获评市级数字政府建设示范案例
- 技术架构：基于Apache Dolphin构建数据集成框架，采用Dataworks+MaxCompute进行数据开发，通过SpringBoot+FastAPI提供微服务接口，结合DataX实现异构数据源同步，应用DataV进行可视化展现，实现数据全链路开发与治理
- 项目职责：负责平台整体数据架构设计，构建跨部门数据融合方案和统一数据标准；设计实现面向共同富裕的主题数据模型和500+核心指标体系；搭建数据质量管控平台，确保TB级数据准确性；开发共富指数评估模型和区域经济画像系统，为政府决策提供数据支撑，决策效率提升40%

- 项目描述：项目基于“e 家富”数据仓，获取住房数据、新生儿数据、生育保险数据、就诊结算数据、教育学籍数据等，进行二次加工处理，集成共富背景下各种类型家庭的算法识别和研判模型，智能识别不同类型共富家庭，辅助并推动以家庭为单位的财税政策体系落地
- 项目职责：构建数字化家庭图谱和家庭共富画像，配合合作团队进行数据进行数据集成与清洗，数据逻辑模型构建及多维数据指标分析，其中重点涉及贫困代际传递风险预测模型的构建与分析，多维指标的底层模型的构建等；协助前后端团队构建家庭画像经济图谱

- 项目描述：为美西SEARS构建线上系统，实现商品、订单、库存、采购、会员等多维度数据的统一管理和智能分析，支持日均百万级订单处理，为业务决策提供实时数据支持，平台整体性能提升60%，数据分析效率提升90%
- 技术架构：采用Dataphin进行智能建模和数据开发，基于MaxCompute构建分布式计算引擎，使用RDS提供在线数据服务，通过QuickBI实现数据可视化分析，打造一站式数据开发与分析平台
- 项目职责：主导数据中台架构设计与业务域划分，开发自动化工具提升数据同步和维护效率；设计实现智能建模体系，包含维度建模、业务过程、原子指标和派生指标；负责CDM层建模和ETL开发，确保数据准确性和及时性；提供QuickBI报表开发支持，优化数据可视化呈现，解决性能瓶颈，最终使数据加载性能提升80%

- 项目描述：为联合国世界粮食计划署构建全球饥饿监测数据平台，实现多源数据自动采集、清洗、分析和可视化，支持全球200+国家地区的饥饿风险预警，平台日处理数据量达TB级，为人道主义援助决策提供数据决策支持
- 技术架构：基于阿里云ECS部署MySQL数据库，通过Java API实现数据采集存储至OSS，使用MaxCompute(ODPS)+DataWorks构建数据处理平台，采用DataX实现数据同步，通过DataV进行全球地图可视化展示
- 项目职责：主导设计跨系统数据流处理方案，实现从数据采集、存储到分析的全链路自动化；开发Python脚本实现数据清洗和预处理，构建预测模型评估饥饿风险；负责多地域RDS数据分发和同步机制优化，确保全球数据实时更新；配合联合国团队进行敏捷开发，数据处理效率提升150%，预测准确率达85%

Education