唐国鑫

基本信息

31岁|男|湖北襄阳|未婚| `18986377139` **(电话微信同号)|** `tang2472854207@outlook.com`

- **主业**：ABC（AI+BigData大数据+Cloud云原生）｜解决方案架构师｜离职随时到岗
- **副业**：自媒体"道语星航"（全网统一账号名）｜商业合作请加微信私聊。

Github： `github.com/Hermesfuxi` | **个人网站**： `https://blog.star-sea.site/`

专业技能

- **数学基础（线性代数、概率论、数值分析、微积分、统计学等）和计算机基础（计组、操作系统、网络、数据结构与算法）**
- **编程语言**：Python、C++、Java、Scala、JavaScript、TypeScript、Golang、Linux/Shell；
- **JavaWeb技术栈**：
  - **数据库**：MySQL（MyCat分库分表）、PostgreSQL、TiDB、OceanBase、Rocksdb、Redis、mongo、ElasticSearch、Neo4j、ArangoDB；
  - **后端**：MyBatis、Spring全家桶(Springboot、SpringCloud、SpringCloud Alibaba等)、dobble；
  - **前端**：jQuery、Bootstrap、Vue、React、AngularJS、Antd、ECharts 、BizCharts、G2、qiankun；
- **大数据技术栈**：
  - **大数据平台**：有完整的大数据平台项目设计、开发及部署经验，曾调研过开源社区和各公司的数据平台产品，且有数据治理经验；
  - **大数据处理**：熟悉整个大数据的完整处理流程（数据的采集、清洗、预处理、存储、分析挖掘、机器学习和数据可视化等），熟悉实时/离线数仓、数据湖等架构，有丰富的数据开发经验，熟悉数仓领域的业务梳理、主题划分、模型设计、架构分层等；
  - **熟悉大数据相关组件**：Zookeeper、Kafka、Hadoop、Spark、Flink、Hive、Presto/Trino、Hbase、Impala、Presto、Alluxio、ClickHouse、Doris、Kylin、Azkaban、Atlas、Flume、Sqoop、DolphinScheduler、xxl-job、iceberg、hudi等；
  - **熟悉大数据相关的开源项目和相关技术工具引进**：StreamPark、dlink、Flink-CDC、ChunJun（前FlinkX）、datax、waterdrop、DataHub、superset等；
  - **熟悉阿里云大数据产品**（DataHub、MaxCompute、DataWorks、RDS、QuickBI、ECS）；
- **云原生技术栈**：
  - **云平台**：国内云（阿里云、腾讯云、华为云、七牛云）、国外云（AWS, Azure, GCP, IBMCloud）、私有云、混合云
  - **容器和容器编排**：docker、podman、containerd、K8s、K3s、OpenShift；
  - **云原生平台**：Rainbond、Kubesphere、KubeVela；
  - **DevOps**：熟悉软件工程全流程，包括规划-编码-构建-测试-部署-操作-监测、 IaC（Terraform、-CloudFormation）、CI/CD（Jenkins, GitLab）、自动化运维（Ansible、Puppet）、监控日志（Prometheus、Grafana）、性能监控（Datadog、SigNoz）、安全合规（OpenSSL）；
- **AI 技术栈**：
  - **数据分析**：NumPy、Pandas、SciPy、Matplotlib、
  - **机器学习**：scikit-learn、HanLP、Alink；
  - **深度学习**：Tensorflow、PyTorch、PaddlePaddle、CNNs、RNNs/LSTMs、Transformers (如BERT, GPT系列)、- Autoencoders、GANs、NLP（NLTK、Hugging Face Transformers）、CV（OpenCV）；
  - **AI大模型**：Alagent

- **AI应用**：
- **AIGC**：ChatGPT、midjourney、stable-diffusion
- **软技能**：良好的团队合作与协调沟通能力，小规模的团队建设管理能力，分析复杂问题和提高解决方案的能力。

工作经历

大数据技术负责人（暂定）- 杏达银通（上海）企业发展有限公司 2022-07 - 2023-07

职责：主要负责智慧停车项目的行业市场调研分析、相关技术咨询和商业解决方案探讨

1、对当前智慧停车行业的市场现状进行梳理和研究，帮助团队确定公司在产业链中的定位为“车位信息平台”，经营模式为轻资产模式（信息资源整合模式）。

2、洞察行业趋势和消费者行为，研究行业竞争对手的商业模式和解决方案，确定其中可吸收可借鉴的部分，从而优化产品设计和服务策略，并结合公司的商业计划，制定该项目的商业模式、产品定位、解决方案。

3、配合团队，参与拆分决策课题，列出课题任务清单，设计验证步骤，模拟真实用户场景，验证关键性条件假设，以支撑下一步决策。

4、利用大数据相关技术，设计并搭建“智慧停车信息综合管理服务平台”，提供实时监控、预测分析和智能优化等功能，大幅提升了停车运营效率和用户体验，以供决策参考。

成就：给团队提供行业市场信息报告，参与探讨项目方向、商业模式、团队建设、产品体系、品牌战略、运营资源、增长路径等议题，“从0到1”地设计并验证解决方案，推动技术项目落地，快速适应市场需求变化。

大数据研发工程师/大数据产品需求解决方案- 云智慧（北京）科技有限公司 2021-05 - 2022-06

中国银联数据智能运营平台项目DODP：一个项目，两个角色

DODP产品已经下架，部门被解散，前身产品DOCP (点击官网)还在

前期角色：数据组leader（上海-研发五部）

职责：前期作为数据组leader，主要负责建设和带领团队，设计并实施项目迭代流程规范，对接并挖掘客户的数据需求，设计-开发-测试-维护相关的数据任务，设计并架构实时数仓，参与平台产品的功能调研-设计开发-保障运维，完成数据任务的完整交付和数据平台产品的阶段交付；

技术：

1. 设计多引擎OLAP系统高可用架构，以“ClickHouse为主，Doris为辅”的策略，形成高低搭配的OLAP技术选型。技术参考方案：京东OLAP亿级查询高可用实践 / 京东ClickHouse高可用实践
2. 基于“湖仓一体架构”设计架构准实时数仓，技术选型为：HDFS + Iceberg + Alluxio + Presto。 业内的同类方案：Presto+Alluxio 加速 Iceberg 数据湖访问
3. 开源工具的选型引进和定制化改造集成：Clickhouse运维工具ckman、kafka运维管控平台KnowStreaming（替代KafkaManager）、调度工具DolphinScheduler（替代XXL-Job）；
4. 参考腾讯Oceanus、 开源的Dlink和StreamPark，设计基于Flink的实时计算平台，实现Flink作业可视化管理，应用构建支持FlinkSQL与Flink-Jar，支持任务的多版本管理和全局UDF算子管理；
5. 经多次实践，总结出Flink与Clickhouse的最佳交互模式『读本地表-写本地表』，并开发连接器 flink-clickhouse-connector实现，同时添加自动故障转移功能，解决大批量数据下clickhouse读写压力过大导致Clickhouse 连不上、写不进、读超时等问题；
6. flink任务优化和流程设计：
   a. 针对多种实时/离线数据源场景，使用FlinkSQL流批混搭设计；
   b. 针对『无事件』触发 场景，使用人造『时间序列流』Time interval join 方案；
   c. 在部分Flink任务中添加数据辅助字段（如标记字段、版本字段、时序字段、kafka 元数据字段等）；
   d. 优化性能（UDF重用、对象重用、shuffle/join 方式选择、参数调优等）；
   e. 开发数据mock功能，并实施数据单元测试（如：实时与离线对数、明细与汇总对数）；
7. Flink + Alink 实现智能实时异常监测：批任务导出训练样本CSV至hdfs上，在Flink任务中使用Alink孤立森林算法加载训练样本，同时接入生产kafka数据，将结果输出至另一个kafka，接入指标体系中，设定阈值，识别异常数据；
8. FlinkSQL数据治理解决方案：重点解决 FlinkSQL字段级的血缘关系、面向用户级别的行级数据访问控制。

业绩：

1. OLAP引擎能支撑数据分析场景数十种以上，支持千万级、亿级大表关联查询，可以实现秒级响应；
2. 数据湖仓（HDFS + Iceberg + Alluxio + Presto)被项目定位为Hive离线数仓加速版，一般存放历史数据或当作数据中间层，在Alluxio缓存加速下，查询数据湖可达到分钟级响应；
3. 推进数据平台支持多种SQL查询语法，并默认使用Presto/Trino SQL，以解决多种SQL查询引擎语法不统一的问题，实现查询引擎与计算引擎解耦；
4. 通过引入开源运维工具，集成至平台产品中，实现手工运维→自动化运维→平台化运维的升级，并制定规范，系统性解决存在性能不稳定、达不到预期、使用不正确、运维艰难等诸多问题，同时拓展了产品功能；
5. 监控并维护日常任务和平台运行，设计、实现、跟踪、优化相关数据计算指标的任务方案，涉及用户域、交易域、商户域等近百张表、几十个关键指标和数据链路，将链路过长且难维护的Kappa架构，拆分替换为Lambda架构，将离线链路与实时链路任务分开，并缩短数据链路，保证任务互不影响，避免任务链路中关键环节出现问题所导致的全局性“重启”，以业务的角度，对数据开发做整体规划，保持整个数据链路的业务语义一致，增强相关任务可维护性和可扩展性。

6. 设计

后期角色：产品需求解决方案（上海-交付部）

职责：后期转为项目产品需求解决方案，主要负责 数据本身的治理运营 和 数据平台的治理运营，详情如下：

1. 反复与客户沟通，了解业务流程和企业现状，挖掘客户深层需求，分析需求场景，形成用户故事，输出需求文档和排期计划；
2. 针对产品现状和项目需求，选择部分国内数据厂商产品作竞品分析，提出产品规划和设计建议；
3. 针对数据乱象，规划设计数据治理体系，包括不限于组织分工、角色职责、产品落地、评估实施等工作；
4. 针对平台本身的治理问题，规划平台治理方案，设计功能落地；

业绩：

1、数据治理方面，参考DAMA框架和《数据资产管理实践5.0》，系统的梳理所有平台数据，按“数据域-数据主题-数据实体-设计数据模型”的逐层分解，补充相关的数据标准和元数据信息、目录标签信息，并配置对应的标准规范、质量监控、安全监控，通过相关指标。具体实施包括数据架构管理、数据标准管理、主数据管理、数据质量管理、数据安全管理、元数据管理、数据资产管理、数据生命周期管理等多个环节，并落地至产品中，开发相关的约束性管理功能。

重点解决数据定义不正确、信息不完整、碎片化、元数据陈旧、质量低下、使用随意、查找定位困难、分类不清、职责不明、管理混乱、缺乏规范、堆积杂乱、业务不可用、只汇不治、形聚神散、安全难保障等诸多问题，实现数据在平台运营过程中可见、可用、可管，覆盖数据流程的全过程。

2、数据平台运营方面，针对正在运行的平台功能，需要监控平台相关服务和组件，实时动态更新监控数据，并对其进行多维度画像分析，根据分析结果匹配治理规则，调用自动化运维脚本或输出服务分析报告，实现平台本身的自治和自调整；而针对一些新需求，通过建立生产级『持续集成-交付-部署』流水线，快速添加或升级自研/开源组件，并自动纳入平台治理功能中。

重点解决数据平台在扩展的同时保障性能和质量的问题，借鉴DataOps理论，结合项目实际情况，探索符合团队的平台运营方式，形成敏捷自动化的数据供应链。

java/大数据开发工程师- 杭州博彦信息技术有限公司 2018-09 - 2021-04

阿里云外包团队TL，负责管理团队，并对接阿里云相关的项目需求，完成独立交付，涉及技术范围比较广（web前后端、大数据、运维都做过），完整全程参与的有两个大型项目，详细如下：

一、某全域动态配置化大数据仓库中台项目

该项目是阿里内部的中台项目，对接其数据技术产业部，项目采用Lambda 架构（离线/实时数仓分层设计架构）。

职责：本人参与的是存储部门的业务业务数仓DaaS部分（部分也涉及PaaS）模块的搭建，离线/实时统计相关的指标（如：PV/UV/DAU/MAU/ARPU等），并对用户行为数据进行分析（如：用户活跃分析/新用户留存分析/归因分析/漏斗分析），建立部分用户画像标签，前台DA（数据应用层）的运营-风控-推荐相关功能的规则决策引擎系统、查询缓存模块的设计开发、动态SQL引擎开发。项目重点在于掌握数据仓库的相关领域知识、模型设计方法和管理技能（元数据管理、数据质量、主数据管理、性能调优等），尤其是决策引擎，集活动创建、执行、管理、控制、反馈、迭代为一体，能够通过用户行为、属性、标签等数据筛选受众，实现目标人群的精准触达，提升关键指标和运营效率。

技术：用BitMap/HyperLogLog/布隆过滤器等技术实现去重类计算和层级聚合（自定义UDF/UDAF/UDTF），实现数据分析的各种报表和大屏展示，实现读写数据流的精确一致性(旧版sparkStreaming→新版Flink)，实现实时动态分组、动态配置加载、动态SQL，实现离线实时的数据拆分与整合，Flink源码修改和重新编译部署，解决双流Join的数据延迟问题（维表关联），解决数据热点、数据倾斜等问题。

业绩：解决了部门表逻辑不清晰问题（逻辑分层方案），数据孤岛问题（维度建模方案-主题域划分，沉淀中间结果），找表用表难不敢用的问题（数据地图方案-查看元数据追溯数据血缘依赖关系），指标定义混乱/重复计算/数据不一致问题（指标字典方案-命名规范管理，统一口径规则）等等，项目实施后，经调研，业务看板报表下线50多张，表分区支持定时清理30多张，HDFS小文件减少到原来的30%，元数据信息减少至原来的60%，存储空间释放：rawdata(4PB)/warehouse(0.4PB)，风险控制延时达到毫秒级（可动态调整）、相关运营效率提升到原来的150%。

二、阿里云某存储部门的内部资源管理系统和任务调度平台

简单来说，该项目是为了内部资源管理调度业务的自动化、配置化、白屏化、可视化、闭环化、服务化。

职责：负责带领外包团队，在主管的指挥下，参与设计、开发、测试、维护整个内部资源管理调度系统，对相关业务流程进行调研和探索，经过抽象定义与流程建模，形成新系统框架，有计划地对已有旧系统逐步升级改造、重构，对新模块进行设计开发（包括需求的调研汇总、设计评估、开发测试、优化维护），同时负责整理项目文档、参与制定标准规范、引入权限审批系统与监控报警系统、构建实时反馈Dashboard页面，部门内部的培训和宣贯等等。

技术：前端：React前端开发（Antd Pro框架），业务系统：SpringBoot+Mybatis+Mysql+工作流引擎Activiti。项目难点在于业务流程的汇总梳理、逻辑分层、抽象建模，需求的沟通挖掘、设计的评审确认、风险把控以及平台的人性化体验等等。

业绩：项目实施后效果十分显著，机器资源的下线-克隆-上线-扩容-维修-归还等完整环节全部流水线自动化，耗费时间缩短至原来的10%，部分环节缩短至5%以下，业务调度错误率降低至0.01%以下，业务流程周转速度至少是原来的5倍以上，许多模块都是前所未有的，有效解决了指标混乱，资源抢占，烟囱式利用资源等工作难题，极大地提高了部门内部的工作效率。

我将上述的大数据中台项目中的部分功能拆开，并用开源软件重新实现，现放在github上，欢迎各位业内大牛多多指教和斧正：

1、datayi 数易数据运营系统 - https://github.com/Hermesfuxi/datayi

一套离线数据运营平台（模拟阿里云数仓中台），包含数据采集汇聚、数据仓库、用户画像、OLAP平台等模块，实现数仓搭建的完整流程（数据产生-采集-ODS-DWD-DWS-DWT-ADS），统计相关的指标信息，以及用户画像功能（基于机器学习，只是Demo入门级别），个性化推荐功能等等

2、eagle鹰眼实时智能运营系统 - https://github.com/Hermesfuxi/eagle

一套实时风控+实时分析系统（仿阿里的实时风控，简化demo版本），简单来说，就是一个基于事件驱动且可进行动态规则计算的实时系统；在技术上，它是通用的；本套架构及系统内核，不仅可以用于“实时运营”，也可以用于“实时风控”，“实时推荐”，“实时交通监控”等场景。技术重点是要能在作业运行的时候去添加和删除规则，而不会因停止和重新启动作业从而造成高昂的代价。项目本身是基于Flink+ClickHouse的lambda架构，使用drools规则引擎，基于Spring boot+Vue构建规则的管理系统（还在构建中，支持规则、模板、策略、黑白名单等的增删改查），并能基于模板引擎Beetl生成动态SQL，并存储到Mysql中，由canal 监听到Mysql的binlog 后加载到Kakfa，再由Kafka流入Flink和ClickHouse，Flink做用户行为的实时计算，ClickHouse做离线计算，支持动态数据分区与规则配置（Flink广播流），支持类与Jar文件的动态编译与动态加载，利用ProcessFunction复杂的自定义逻辑来“模拟”窗口，redis做缓存，HBase存储用户画像数据（模拟生成，后续会建立实时画像模块），后期打算接入机器学习-专家系统等模块，项目现阶段仍处于构建阶段，文档也在补充当中。

java开发工程师- 成都晟峰软件有限公司 2015-08 - 2018-08

一、互联网金融后台管理项目

该项目主要包括数据接口系统、后台管理系统、支付系统、第三方接口对接系统、定时任务系统、营销活动系统，红包系统。

技术：Nginx（负载均衡），Redis（分布式缓存处理），Dubbo（远程调用），Mycat（读写分离和分库分表），FastDFS（分布式文件系统存储图片），Zookeeper（协调选举提供高可用的消息队列集群），ActiveMQ（采用消息队列实现异步处理）。

职责：

1、负责 Dubbo 服务的开发，根据前端业务需求提供底层 Dubbo 服务并与前端业务系统整合联调，并搭建 Zookeeper 集群环境，提供 Dubbo 服务；

2、搭建 MySQL 数据库主从复制集群，采用 Mycat 实现数据库读写分离，搭建 Nginx 服务器实现平台的负载均衡与静态分离；

3、负责首页轮播图、产品列表、产品详情、专题活动的开发；

4、实现会员 用户的注册、验证、开户、投资、充值、个人金库、账户统计等功能，其中账户统计包括可用余额、所获利息、所借到的金额、投标中的金额、已逾期的项目；

5、搭建 Redis 高可用集群环境，并对注册用户数、利率、总投资、投资排行榜等统计数据进行缓存；6、搭建 ActiveMQ 消息队列集群，并在用户抢标、生日短信、账户资金变动等功能模块中采用 ActiveMQ 消息队列实现异步处理；

7、参与项目的测试部署，并编写项目文档。

二、重工业行业的客户关系管理系统CRM

该项目为重工业行业的客户关系管理系统，主要用于对国内外治金、石油化工、造船、压力容器、装潢、金属结构及机械制造等诸多行业客户的开发和管理。

技术：主要是CRUD（前后端），SSM 架构，MySQL 数据库，Maven，BootStrap 框架，Apache Common，RBAC 权限管理

职责：工作模块包括业务报告、审批流程、客户公海、客户组、联系人、市场活动、线索公海、交易情况、售后回访、统计图

表、账务报表、销售订单、发货单等等，并与销售、营销、推广、策划、人事等多部门业务对接。