

单梦凡

电话: 18017661124

Email: dd.famous@gmail.com

工作经历

Block

2021 年 5 月 ~ 至今

开发工程师 - MarTech 数据团队

从头开始设计、搭建、更新并维护 MarTech 数据平台，包括数据集成，数据仓库，作业调度，数据质量和数据服务，满足营销和广告等在线业务对数据的需求。开发多个数据驱动的线上服务，如数据洞察、用户数据平台、资格服务等。同时也负责新技术研究、疑难问题攻关、开源组件接入，提升平台的运作效率和稳定性。

上海数禾信息科技有限公司

2019 年 4 月 ~2021 年 4 月

大数据架构师 - 大数据部

牵头成立数据平台组，负责公司大数据平台的架构设计、开发部署、监控管理、优化升级等，并负责大数据平台的架构演进。期间主要负责数据平台架构升级，从自建小规模数据平台，主导负责数据上云和云平台迁移，并最终完成数据中台的设计和落地。与数仓团队和数据应用团队合作，负责公司核心数据产出的 SLA，保证数据质量、数据平台和数据服务的稳定性，并平衡平台运营成本。

Works Applications (万革始应用软件有限公司)

2016 年 8 月 ~2019 年 4 月

团队负责人-企业数据处理平台

负责设计、开发大数据开发及运行平台，为业务开发人员提供简便的大大数据开发接口，提高业务数据开发效率。平台主要依托 Hadoop/Spark 提供高效低成本的并发方案，并处理不同需求场景中的数据处理请求。

项目经历

资格服务

项目负责人

2022 年 9 月 ~2023 年 6 月

设计并开发了资格服务。资格服务是为下游广告、促销等活动服务提供用户触达，并为市场营销人员量身定制的营销平台。它提供基于规则的精准营销，动态细分，并集成的 A/B 测试等功能。同时，资格服务与数仓集成，保证线上离线逻辑一致，方便分析师对营销活动在上架前后分析、改进。资格服务将营销活动时效性从天级提高到秒级，并极大提高了营销活动的准确性。

实现: Java/Kotlin, Snowflake, Antlr4, UDF

用户洞察

项目负责人

2022 年 6 月 ~2023 年 3 月

设计并开发了用户洞察服务，一个通用的用户数据平台。用户洞察集成了所有需要的数据源，项目包含数据服务，数据处理，跨平台数据集成等，在为各种场景下的客户属性和特征创建一个跨平台的系统，为营销、广告等线上服务提供统一的用户数据服务。数据平台集成了数据科学团队和离线模型团队的离线特征数据和强实效的交易、埋点等数据。通过去耦数据管道和消费者服务，实现独立添加属性/特点，在不依赖于其他团队的情况下，优化市场营销的效果。

实现: Kotlin/Java, Kafka, DynamoDB, Snowflake, Airflow, Spark/Hive/Snowpark, EMR, S3, SQS, Glue

数据洞察	数据负责人	2021 年 5 月 ~2023 年 3 月
------	-------	------------------------

设计、部署并落地了多个数据洞察服务，为接入 Cash 和 Afterpay 的商户提供自助的数据分析平台，是公司数据闭环的关键。项目包括的数据平台基础设施，数据处理，数据质量、数据服务以及数据展示。数据洞察会将商户和用户各种纬度的数据和各种对比趋势提供给销售、运营人员，让商户清楚了解运营情况。同时，集成深度学习模型和大模型，给出运营建议。目前服务超过 6000 个各地商户，并帮助商家做出数据驱动的决策。

实现：Clickhouse, Kylin, Spark, Hive, OpenAI, AWS EMR, Airflow, Ansible, Spring, Java

数据中台	平台负责人	2019 年 6 月 ~2021 年 3 月
------	-------	------------------------

作为大数据平台负责人，设计并主导了大数据平台的整体迭代。大数据平台从最初的小规模 CDH 自建，随着数据量增长逐步数据上云，又因云供应商切换进行云迁移、混合云架构，最终在完成数据中台。保证在数据迁移和平台重构的过程中对业务无影响，并减少大数据平台整体的运行成本。数据平台从二十余台机器，十多 TB 数据，发展到数百台机器，PB 级数据，很好支持了业务的增长。

实现：Cloudera, AWS, AliCloud, HBase, Hive, S3, OSS, Dataphin, Kafka, Flume

特征平台	平台负责人	2020 年 10 月 ~2021 年 3 月
------	-------	-------------------------

特征平台用于为模型训练及模型预测提供所需特征。传统的特征平台主要聚焦于数据清洗、归一化、分箱等数据处理阶段，但对于数据获取并没有很好的解决方案。而在实际的模型训练到模型部署上线预测的过程中，最困难的问题就在于数据获取。因此，本系统直接聚焦于这一点，所有设计将为同一点服务：离线训练特征与实时预测特征统一。

实现：HBase, Flink, Hive

核心优势

-
1. 长期大数据工作，熟悉主流的数据架构，由 0 到 1 数据平台设计、搭建经验。有数据中台、数据湖、DataMesh 等数据平台设计、实施经验。
 2. 熟悉公有云（AWS, AliCloud）上主要数据产品，基于各种云产品的整合、开发数据平台经验，并基于云有 IaC 经验。
 3. 较强的问题处理能力。有基于源码二次开发（数据血缘等），修复发现开源组件的问题（HDFS-9406, HIVE-19994 等）经验。
 4. 可以根据不同的业务需求设计开发高可用、高吞吐、低延时服务，满足业务需求。
 5. 有跨时区、跨团队项目设计、沟通、开发经验。

教育背景

南京大学	2010 年 9 月 ~2016 年 6 月	博士，计算机科学与技术系
明尼苏达大学双城分校	2014 年 10 月 ~2015 年 10 月	访学，计算机科学与工程系
南京大学	2006 年 9 月 ~2010 年 6 月	学士，计算机科学与技术系