单梦凡

地址: 虹口区天宝路 466 弄, 200086　　　**电话:** 18017661124　　　**Email:** dd.famous@gmail.com

工作经历

Block, 上海　　　　　　　　　　　　　　　　　　　　　　　　　　　　　2021 年 5 月至今

开发工程师 - MarTech 数据团队　　我 2021 年加入 Afterpay，在 2022 年初加入 Block。期间主要工作是从头开始设计、搭建和维护 MarTech 数据平台，包括数据平台基础设施搭建维护，跨平台数据集成，数据任务，强数据驱动的在线服务，如多种数据洞察服务，用户数据平台（CDP）等。同时也负责核心技术研究、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

Hadoop 架构师 - 大数据部　　　成立并负责大数据平台组，牵头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，负责大数据平台的架构演进，任职期间从自建小规模大数据平台，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，负责公司的数据服务。

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（Eligibility Service），Block 我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筛选，并为市场营销人员量身定制的数据平台。它提供基于规则的定位，动态细分，实时响应，以及集成的 A/B 测试，以提高市场营销的效率，并提高可扩展性和准确性。同时数据服务于数据分析、机器学习以支持数据驱动的决策和优化。Kotlin, Armeria, DynamoDB, antlr4，Snowflake

用户洞察（CustomerIQ），Block 我设计并开发了用户洞察，一个通用的用户数据平台，包含数据服务，数据管道，夸平台数据集成等功能，旨在为各种场景下的客户属性预测创建一个多功能的系统，重点为激励/广告观众定位和广告排名/个性化。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。Kotlin, DynamoDB, Snowflake, Airflow

数据洞察（Insight Service），Block 设计、部署并落地了多个数据洞察服务，包括数据平台基础设施，数据流，数据质量系统和数据服务。数据洞察是一个自助的数据服务平台，是我建数据闭环归的关键，将商户数据整合后提供个各地商户，并提供各种对比趋势，集成 AI 分析提供数据分析和营销建议。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。Clickhouse, Kylin, Spark, Hive, DynamoDB，OpenAI

数据中台，数禾科技 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，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，同时修复发现开源组件的问题（HDFS-9406, HIVE-19994），并减少大数据平台整体的运行成本。Cloudera, AWS EMR, S3, Ali EMR, OSS, Dataphin

特征平台，数禾科技 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 HBase, Flink, Hive

常用技能

语言：　Java，Kotlin，Python，Ansible，shell

组件：　Clickhouse, Kylin, Hive, HBase, Hadoop, Spark, Flume, Flink, Zookeeper, Kafka, neo4j, HugeGraph

技能：　AWS，阿里云，Snowflake

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

代码库

GitHub　　　　　　　　　　　　　　　　　　　　　　　　　https://github.com/ddfamou

Mengfan (Fox) SHAN

Address: 466 Tianbao Road, Shanghai, 200086   **Tel:** 18017661124   **Email:** dd.famous@gmail.com

WORKING EXPERIENCE

Block, Shanghai   Since May, 2021

I joined Ads & Growth Team, Afterpay in 2021 and transitioned to Cash, Block in 2022. My responsibilities were designing and implementing Martech Data Platform, including data infrastructures & integrations, data pipelines & data-driven services. During my time at work, I took the initiative to consolidate internal/external data and construct the Martech Data Platform. This platform encompassed various essential components, including data integration, data processing, and the provision of diverse data services, catering to the specific requirements of downstream operations. These services included data insights, a user data platform, and user engagement tools, ensuring seamless access and utilization of valuable information for our team's success.

Shuhe Group, Shanghai   April, 2019   April, 2021

Leading the Data Platform Team at Shanghai's Shuhe Group. My role included designing, developing and optimizing BigData Platform, collaborating with both the Data Warehouse and Data Application Teams to ensure high availability. The architecture of the data platform has undergone a transformative journey, starting from an on-premise setup and gradually transitioning to cloud-based infrastructure. This evolution involved the migration of data to the cloud, leading to the ultimate achievement of a robust and comprehensive data middle platform.

Works Applications, Shanghai   August, 2016   April, 2019

As Team Leader of EDP2 (Base Technologies Div) at Works Applications. I oversaw the development of a platform for enterprise applications on the cloud. This included managing a server for job assignments and a Hadoop/Spark cluster for job execution.

PROJECTS

Eligibility Service, Block

I designed and implemented Eligibility Service, a new, dynamic audience targeting service to significantly scale Boost to all types of incentive, to highly reduce operational overhead for Boost set-up and to improve reporting processes of Boost. Kotlin, Armeria, DynamoDB, antlr4, Snowflake

CustomerIQ, Block

I designed and implemented the service and related data pipeline jobs of CustomerIQ, a generic Customer Data Platform. CustomerIQ integrated data from various Clouds/Platforms and projecting customer attributes across multiple scenarios optimised for boost targeting and ad personalisation. Its design promoted independent addition of attributes/traits enhancing marketing effectiveness. Kotlin, DynamoDB, Snowflake, Airflow

Insight Service, Block

Design and implement data infrastructures, pipelines, data quality check system and the data service for MarketplaceIQ & AfterpayIQ, a self-serving insight product for over 6000 merchants. Our integrated insight service leverages data from various dimensions and utilizes AI to analyze reports for 6000 merchants. This project steps towards building a closed-loop attribution system, unlocking the potential of our first-party data, and enables data-driven merchant decisions. Clickhouse, Kylin, Spark, Hive, DynamoDB, OpenAI

Data Mid-Platform, Shuhe

Design the architecture and the migration plan of Data on Cloud . The BigData infra was migrated from On-Promise to AWS EMR, from AWS to AliCloud and finally Data Mid-Platform. To minimize migration impact and downtime to business, I needed to be compatible with changes, to fix bugs on opensource components (HDFS-9406, HIVE-19994, etc) and to provide all kinds of automation tools for migration and further management. Cloudera, AWS EMR, S3, Ali EMR, OSS, Dataphin

FeatureHub, Shuhe

FeatureHub is a platform managing and providing features, for training models and online-prediction. Unlike traditional feature platforms, which focus on data processing, FeatureHub provides processed data for both online and offline services with the unified calibre. It provides real time features for online services with low latency, as well as historical data for training. HBase, Flink, Hive

SKILLS

Languages: Java, kotlin, python, Ansible, shell
Middleware: Clickhouse, Kylin, Hive, HBase, Hadoop, Spark, Flume, Flink, Zookeeper, Kafka, neo4j, HugeGraph
Skills: AWS, Aliyun, Snowflake

EDUCATIONS

Nanjing University, Nanjing   09/2010 - 06/2016
Ph.D. in Computer Software and Theory, Department of Computer Science and Technology

University of Minnesota, Twin Cities, MN   10/2014 - 10/2015
Visiting Scholar supervised by Prof. Tian He, Department of Computer Science and Engineering

Nanjing University, Nanjing   09/2006 - 06/2010
Bachelor, Department of Computer Science and Technology

REPOSITORIES

GitHub   https://github.com/ddfamou