

『基金行业金融科技获奖成果宣传活动』银  
华基金：基于大模型的基金公司新一代实时  
数据资产服务平台

## 摘要

行业普遍存在外购核心系统与自研管理系统并存、数据孤岛严重、数据质量不高的问题。为解决数据贯通难题，企业多采用数据仓库、中台等“体外循环”方式实现局部数据整合，但存在处理延迟大、及时性不足等局限。随着市场竞争加剧与精细化、实时化需求提升（如实时交易、风控和实时头寸），传统数据架构已显乏力。尽管AI与大模型技术迅速发展，但因成本高、周期长、落地难等因素，其在业务中的深度融合仍较为有限。为突破困局，我们转向以“AI+数据”双驱动为核心，推动数据从“依附应用”向“自主赋能”转变，优化数据价值链。重点围绕三方面展开：构建AI驱动的新数据资产服务模式、打造支持全域实时集成的新一代智慧数据底座、建立可快速迭代的业务创新敏捷工厂，以重塑数据资产服务体系，提升实时服务与持续创新能力。

**关键词：**优化数据价值链；AI驱动；AI+数据；实时数据服务；自主赋能

## 一、背景情况

当前，基金行业在数字化建设过程中，普遍面临系统架构复杂、数据价值释放不充分的挑战。长期以来，行业核心业务系统多依赖外购，为满足多元化管理需求，各机构又围绕核心系统自主研发了各类配套管理系统，形成了外购与自研系统并存但彼此割裂的格局。系统间数据无法顺畅贯通，导致“数据孤岛”现象突出。早期建设以支撑业务流程为首

要目标，对数据本身的规范性、一致性与完整性重视不足，进一步加剧了数据质量低、投资、运营、合规、市场等关键领域的数据难以融合，跨部门、跨场景的数据协同如同“跨组织协调”，成本高、效率低。

与此同时，人工智能技术正经历革命性演进。自 2022 年底 ChatGPT 引爆全球 AI 热潮以来，大模型技术在多模态理解、复杂推理等方面能力飞速提升，为金融行业的深度智能化带来了前所未有的机遇。然而，受限于技术门槛高、投入成本大、建设周期长以及在垂直业务领域落地见效慢等实际困难，行业对大模型技术的应用仍多处于探索与试点阶段，尚未实现与核心业务流程的大规模、深度融合。整体智能化水平仍较多停留在传统的统计分析与经营看板层面，未能充分释放 AI 的核心价值。

政策层面，国家相继推出《推动数字金融高质量发展行动方案》等指导文件，明确鼓励金融机构深化数字化转型，利用科技赋能业务创新。在此背景下，行业亟待突破数据服务时效滞后、智能化能力不足、数据价值转化链条冗长等核心痛点。

因此，变革传统数据建设模式，构建以“AI+数据”双轮驱动的新型数据资产服务体系，推动数据从被动、依附于应用的资源，转变为能够主动、实时赋能业务的战略资产，并促使 AI 从辅助工具升级为核心业务引擎，已成为行业顺应技术趋势、响应政策号召、提升核心竞争力的关键方向。这不仅是技术架构的升级，更是服务模式与价值释放路径的根本性重构，旨在为写好“数字金融”大文章奠定坚实的智

能数据基础。

## 二、解决方案

### (一) 设计思路

在建设目标上，我们设定了三个核心方向：

1. 是构建 AI 驱动的数据资产服务新模式，优化价值释放链条。
2. 是打造新一代智慧数据底座，逐步实现全域数据实时集成，打造自下而上实时服务能力中心。
3. 是搭建业务创新敏捷工厂，保障体系可快速持续迭代。

### (二) 系统架构

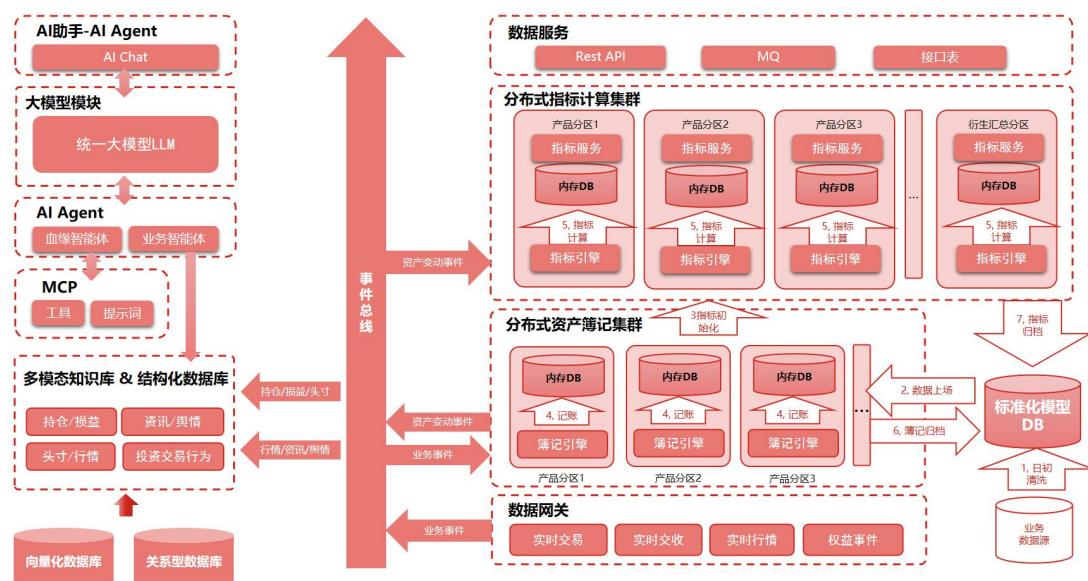


图 1 系统架构图

首先，系统实现异构数据统一接入。通过对接各个交易市场、结算机构的数据以及对接内部系统数据和外部系统数据，构成了统一的数据源，打破数据孤岛，实现全域实时数据底座。

其次，通过事件驱动、流式计算的技术手段，替代传统

轮询模式，满足应用场景对低延时、高时效的核心需求。

再次，以元数据、配置数据、业务数据及标签数据作为输入，借助大模型的推理能力，输出精细的数据标签，同时结合基础血缘数据及代码逻辑后，输出字段级血缘图谱，以实现基于AI技术的智能数据管理。同时基于相关数据作为知识库，促进智能数据消费效果。而消费的异常结果，又为数据治理的持续优化提供了素材，构建了智能数据治理+智能数据消费的双向闭环。

数据网关（异构整合）+实时计算（EDA驱动）+智能治理（大模型）+指标管理+智能数据消费构成了平台的核心。平台以自下而上的构建思路，全链路开放敏捷的研发模式，使得平台构建的不仅是一个系统，更是一种极速实时赋能各类业务场景的能力与生态，结合智能数据消费，将极大程度缩短数据的价值释放链条，提升数据应用价值。

### （三）业务架构



图 2 系统业务图

通过构建实时资产数据，实现重数据轻应用新模式，以全域实时数据资产为基础，构建轻量化应用体系。基于大模

型，完成智能元数据管理、血缘管理、指标定义相关模块，以大模型驱动构建消费链路，提供多元化消费，达到善用数据柔性的目的。

#### (四) 技术实现

1. 从功能层面上看：平台具备从异构整合、AI 驱动的数据治理、EDA 驱动的极速计算、分布式调度、灵活的指标配置研发、快速数据服务发布、智能消费的数据生产全链能力。

2. 从底层技术架构上看：应用平台是基于 Spring-boot 技术栈进行管理平台开发，以 Kafka 为消息中间件，前端则使用 React+Antd+AntV+UMI 开发前端工程；

3. 数据架构则采用 Spark、自研 EDA 技术作为主要数据加工框架、自主优化后的 Apache Ignite 做极速内存计算查询支撑，通过平台进行统一管理，进而实现流批一体的计算框架。

4. 大模型相关技术则以千问、DS 为模型，以 Dify 为框架，以 Milvus 为向量库实现框架，构建核心能力。

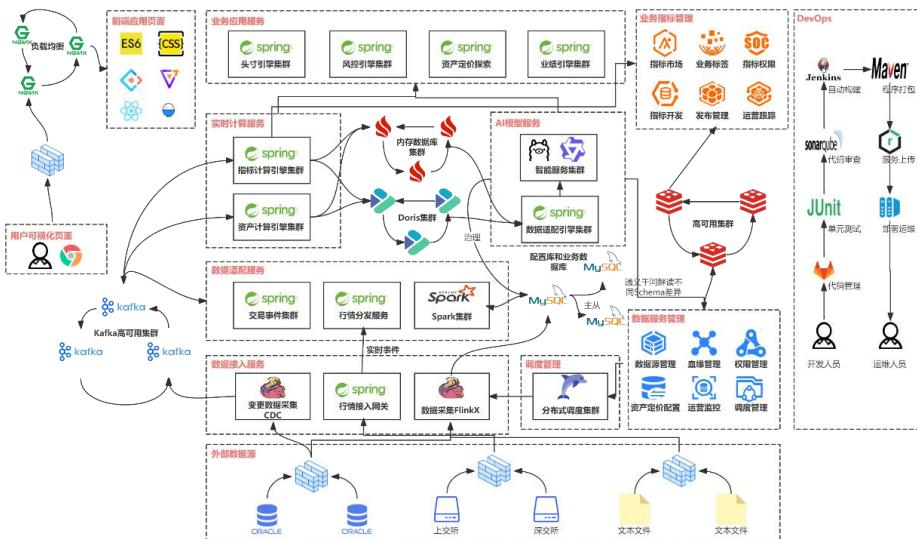


图 3 技术实现图

## **(五) 关键措施**

### **1. 业务稽核**

在关键数据处理链条上，对前后环节的关键指标进行定时比对。例如，交易明细、持仓明细、行情价格等进行核对，若偏差超过预设阈值，立即触发高级别预警。内置多种业务规则进行实时判断。例如，校验“基金单位净值不应为负”、“单日净值波动率通常应在特定范围内”（如±10%等）。一旦系统发现计算结果违反这些基本规则，将触发警报。监控相邻周期计算结果的突变情况。例如，本期基金净值相较于上期通常不应出现断崖式下跌或飙升，一旦检测到异常波动，系统会自动预警，由人工介入核实是否为市场真实情况或计算错误。

### **2. 数据稽核**

针对数据稽核场景，项目实现全流程数据异常场景，明确界定 4 类核心异常：

- (1) 数据接入类：离线/实时数据漏接、接入超时、数据源连接失败。
- (2) 数据质量类：关键字段缺失、字段格式错误、数据值超出业务值域。
- (3) 数据逻辑类：关联数据不匹配、业务规则冲突（如回购利率异常）、跨系统数据不一致。
- (4) 计算执行类：指标计算失败、SQL 语句执行报错、批量处理中断。

### **3. AI 应用优化**

基于 MCP 架构，我们构建了一套提示词优化体系，旨在

将传统静态、固化的数据查询指令，升级为能够驱动智能体协同工作的动态会话核心。该体系以三层提示词结构为基础：底层是封装单一数据源或分析任务的原子能力提示词，确保基础操作的精准性；中层是负责理解复杂业务意图、进行任务拆解与智能体路由的协调层提示词，实现跨系统调度；上层则是维持多轮对话、管理上下文与探索过程的会话层提示词，保障交互的连续与智能。

## （六）平台工具

本次平台工具主要使用包括：

1. 通义千问大语言模型
2. Dify AI 应用开发平台
3. 顶点 LiveData 数据管理平台
4. 顶点 IBLive 事件网关
5. 顶点 IBLive 簿记中心
6. 顶点 IBLive 指标中心

## 三、实践成效

### （一）技术成效

异构数据接入成效：接入了行情（上交所、深交所、彭博）、交易（指令、委托、成交）、估值、申赎等 70+类业务事件，数百张数据表。

事件中心管理成效：基于 Apache Kafka 构建，负责“事件分发”。抽象设计三类事件：基础指标事件（如“成交事件”）、原子指标事件（如“持仓变动事件”）、衍生指标事件（如“持仓占比变动事件”）。抽象了 6 类业务事件（交

易类、交收类、权益类、人工操作、指标类、其他），通过时序技术实现了业务会话一致性管理。通过 3 层监控，确保事件不丢失。当前日均，日均接入事件数数十万个，事件接入准确运行稳定，事件延迟达到毫秒级。

指标中心计算成效：记录基金的实时持仓、资金流水等核心数据，数据不落地，计算在内存中完成，绝大部分全链路指标计算耗时 5 秒内。

服务中心能力成效：负责“指标服务发布”，把计算好的实时指标通过 API、消息订阅等方式提供给业务系统。目前已发布数十个实时 API 接口，支撑组合管理、实时风控等核心系统，接口平均响应时间毫秒级。

AI 治理能力成效：智能数据治理能力，通过大模型形成元数据管理数万条，形成基础标签数十类，形成行为特征标签数百个。

AI 数据消费成效：业务同事基于 AI 数据消费，实现了个性化的实时头寸、交易管理等场景，具体场景下单人效率提升数十倍。

## （二）业务效能

一是组合管理系统全面激活：通过实时数据资产的构建，组合管理系统的数据服务效率提升了近 200 倍，投中估值精度最低偏差可控制在百万分位。

二是合规场景的敏捷处置：24 年场内合规业务场景风险紧急处置，基于事件驱动框架快速配置了“实时合规校验”程序，效能提升 20 余倍，成功化解了公司的合规风险，并直接帮助公司节约实时类需求研发成本。

三是货币基金投资赋能：平台全面支撑了公司货币基金和短债基金的流动性管理，实现秒级各类“可用头寸”“申赎预测”“质押式回购余额”等精细化管控，帮助基金经理精准把握资金流向，显著提升资金利用率。

## 四、经验总结与展望

回顾整个项目建设，我们最深的体会是：数据价值的核心，在于持续放大其资产属性，降低其负债属性。

### （一）项目总结

通过这一轮建设，我们实现了几个突破：

#### 1. 资产属性提升

（1）服务能力提升，平台直接赋能盘中投资、合规、运营等核心场景；

（2）价值释放链条优化，构建了“数据即服务、服务即价值”的模式，跳过业务系统构建的“中间商”环节；

（3）联通价值提升，基于AI数据治理，数据的可连通性更强，可用范围更广；

（4）时间价值提升，实时服务让数据“保鲜度”更高；

#### 2. 成本属性降低

（1）风险成本降低，通过提升合规场景的技术应对能力，有效降低潜在风险成本。

（2）AI驱动的数据治理及自下而上的建设模式，有效提升了数据质量及数据复用度，减少了重复建设成本；

（3）敏捷研发模式提升了研发效率，缩短了响应业务的时间；

(4) 全栈信创信创也在长期降低了技术依赖成本。

## 五、未来展望

### (一) 以业务场景为纲的持续扩展

基于全链路开放的敏捷研发能力，快速孵化更多应用场景，比如多策略组合的实时管理、实时风控等，让数据价值渗透到基金运作的每一个环节。

### (二) 以人为本的场景持续增强

在 AI 赋能的基础上，持续放大 DaaS、SaaS 模式的杠杆效应，更大范围内优化数据价值释放链条。继续推进基于 AI Agent 构建模块化场景化数据消费场景，持续拓展数据即服务的应用半径。

### (三) 联通范围的持续拓展

在 AI 驱动的数据治理模式下，持续扩大平台数据范围，物理上实现离线体系与实时体系的一体化归并，不再以技术形式人为划定边界，让数据资产的流通性更强。