

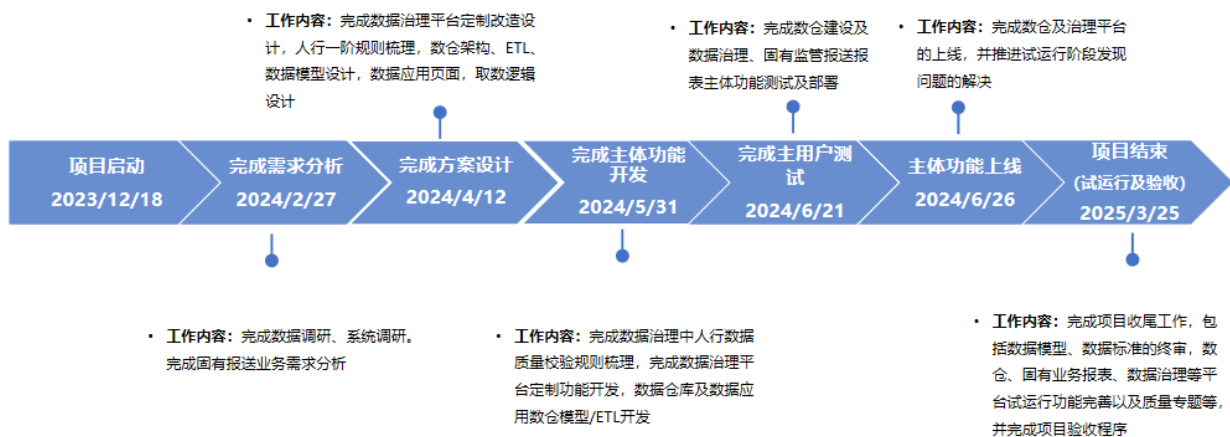
苏州信托-数据仓库及数据治理一期实践分享（上）

建设历程与核心成果

苏州信托数据仓库及数据治理一期项目于2023年12月正式启动，以“构建数据治理体系、夯实数据仓库基座、试点监管报送应用”为核心目标，历经需求分析、系统设计、开发测试、试运行优化等阶段，于2025年3月通过多部门联合评审并正式验收。项目围绕数据治理能力搭建、数据仓库基础建设、固有监管报送试点三大主线，构建了覆盖“标准 - 架构 - 应用”的数据管理闭环，形成了数据汇聚、治理到价值输出的基础能力，为公司数字化转型奠定了核心数据底座。

一、项目建设总览

项目遵循“现状调研 - 需求分析 - 系统设计 - 开发测试 - 试运行优化”的科学路径，先后通过十余次内外部评审，确保各阶段成果符合业务与监管要求：2024年4月完成需求分析与系统设计评审，2024年6月上线主体功能，2025年2月完成试运行优化并结项，最终于2025年3月25日正式验收。目前数仓系统运行稳定、性能优良，为后续数据应用拓展奠定了坚实基础。



二、整体规划及本期核心目标

（一）整体规划

在项目开始前，基于信托业数字化转型背景，公司面临的数据管理挑战等内外部因素，为全面提升数字化能力，已初步设计了数据仓库与数据治理建设三年规划方案，提出了“1 体系+2 平台+1 中心”的建设思路，即以数据治理体系为总纲，建设数据仓库平台、数据质量与资产平台的技术支持底座，以数据应用能力为中心发挥数据价值。



（二）本期核心目标

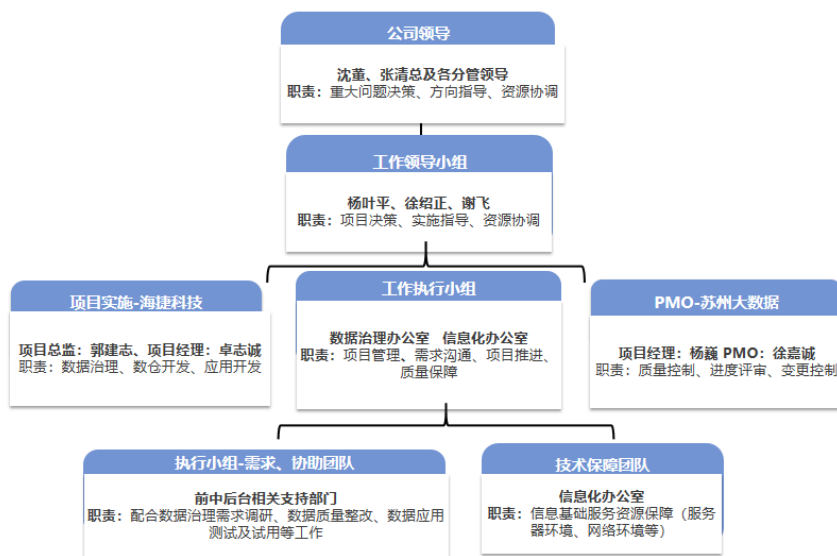
本期项目主要目标是建设数据仓库聚合企业核心的业务数据，加强治理体系建设,提升数据质量,试点支持监管报表应用，具体目标如下：



三、项目启动及需求分析

（一）项目启动及准备

为保障项目的顺利进行，在项目组织架构方面，建立三级组织架构：决策层由公司领导牵头，负责战略方向与资源协调；管理层设工作领导小组，统筹项目决策与实施指导；执行层由数据治理办公室、信息化办公室及实施团队组成，承担需求落地与技术实现。



在工作职责方面，项目组明确了需求涉及的具体部门及职责。由数据治理办公室全程统筹协调相关部门进行需求调研、设计、评审、测试等工作。信息化办公室负责项目管理、沟通事项，并提供基础技术资源保障。运营管理总部配合指导监管报送问题和质检规则梳理。固有业务部门配合梳理和测试固有 1104、固有 EAST4.0 相关报表取数规则和数据准确性。计划财务部配合梳理和测试利率报备、固有 1104、固有 EAST4.0 相关报表取数规则和数据准确性。信息化办公室负责项目管理、沟通事项，并提供基础技术资源保障。

在沟通机制方面，通过项目周例会、月度会、阶段性总结报告等方面进行沟通，保障信息同步。

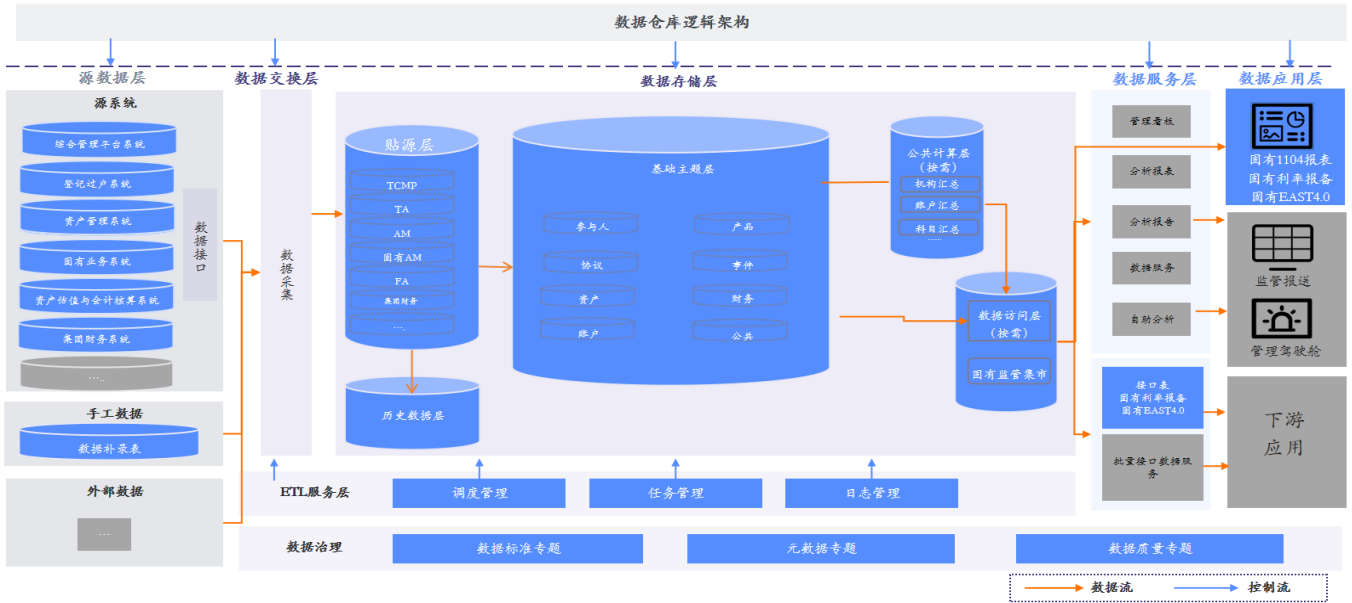
（二）需求调研与分析

数据仓库梳理 6 个核心系统数据资源，明确相关规范与主题域规划；数据治理及质量方面，明确数据治理体系各项制度，确定平台优化需求点，开展专题质量规则库建设；试点固有监管报表梳理 59 张报表取数口径及实现方式。

需求 分类	需求细项
数据 仓库	1. 数据采集（6 个核心系统全量接入） 2. 数据模型建设（贴源层、基础主题层等分层设计） 3. 调度监控（ETL 及任务调度）
数据 治理	1. 数据治理体系规划（含制度、流程制定） 2. 元数据专题（梳理与监控） 3. 数据标准专题（制定与落地） 4. 数据质量专题（规则库建设、专题治理） 5. 数据质量管理平台定制功能（10 项功能优化）
数据 应用	1. 固有 1104 非现场监管报表（51 张报表自动化） 2. 利率报备固有业务数据接口（3 张报表） 3. EAST 4.0 固有业务数据接口（5 张报表）

四、设计与开发

基于业务需求、系统情况及未来规划等综合考虑，公司数据仓库采用传统 Oracle 离线数仓形式进行设计开发。整体架构如下：



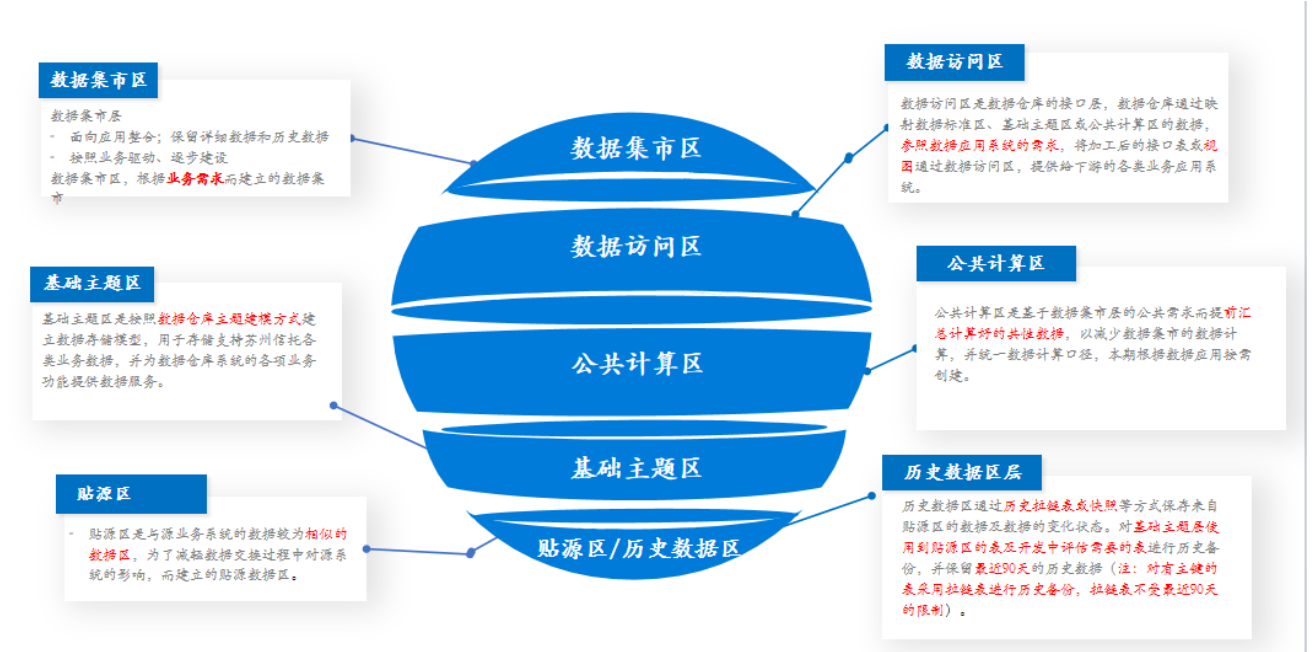
（一）源数据

是数据仓库系统的数据来源，为数据仓库系统提供数据。本次接入 6 个核心系统：综合管理平台（TCMP）、登记过户系统（TA）、资产管理系统（AM）、资产估值与会计核算系统（FA）、固有业务系统（GYAM）、集团财务系统（金蝶）。

（二）数据交换层

从各个数据源系统获取数据，经过一系列的 ETL 转换，将数据装载到数据仓库系统。本层采用 DataX 工具实现 ETL 流程，将源系统数据按照数仓要求的形式全量 / 增量抽取。ETL 过程包括数据抽取、数据转换、数据加载等。另外，数据交换层还要完成 ETL 任务间依赖关系的定义，任务调度的自动化，错误处理机制，例外情况处理，ETL 运行监控等工作。

（三）数据存储层



是整个数据仓库的核心，按照用途划分为贴源区、基础主题区、公共计算区、数据集市区、数据访问区等区域。源系统数据通过数据交换进入贴源区，其数据结构和数据基本与源系统贴近，贴源区的数据再经过数据清洗、数据转换、数据标准化等 ETL 处理，进入基础主题区。基础主题区按照数据仓库主题模型存储数据，保存最详细的全量历史数据。同时，为了统一加工计算，降低数据处理成本，建设公共计算区，按业务主题和数据应用需求轻度汇总数据；为了支持不同的业务应用需求，建设数据集市区及数据访问区，其中数据集市区用于支撑公司内部各类查询、分析、报表等 BI 应用；数据访问区通过视图、表等方式对下游业务应用系统提供数据接口服务。数据集市区及数据访问区的数据可来自贴源区、基础主题区、公共计算区、数据集市区等，按照应用的要求进行汇总加工，为业务应用服务。

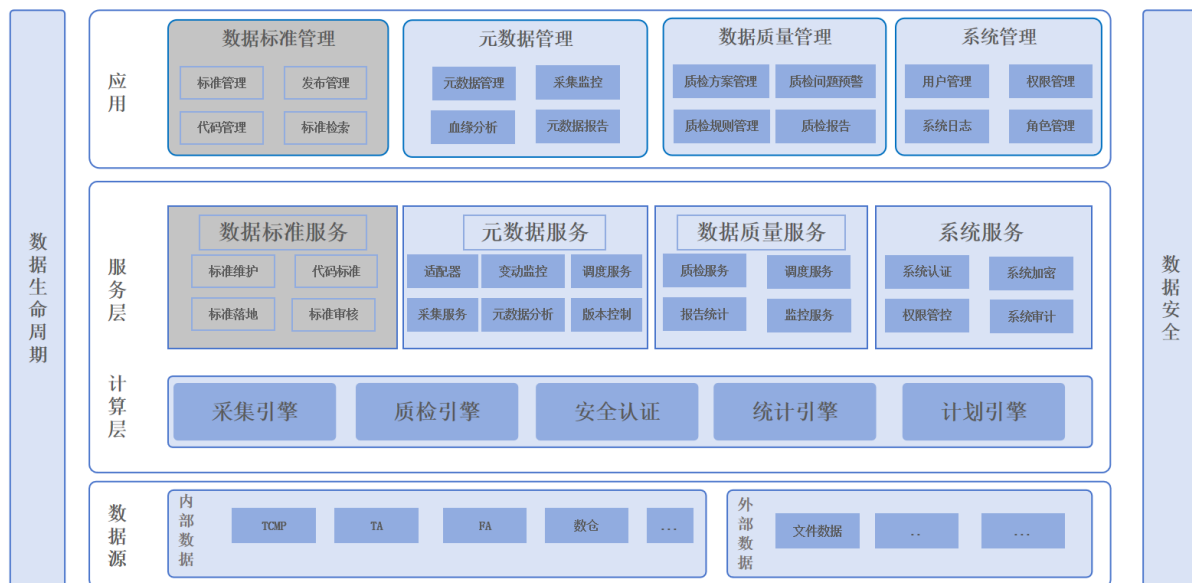
历史数据区用于保存来自贴源区的数据，并通过历史拉链等方式，保存历史数据的变化情况，可用于查询、回溯历史时点的数据状态。

（四）数据应用层



聚焦固有业务监管场景，实现 EAST 4.0、1104 报表等自动化生成，提升报送效率与准确性。。

(五) 数据治理层



对数据行使统一管理的集合，包含覆盖数据架构、数据标准、数据模型、元数据、数据质量、数据安全等多个专题领域。本期项目将实施数据治理体系规划及元数据专题、数据标准专题和数据质量专题等数据治理专题。

五、测试与上线试运行

（一）测试内容

1. 单元测试

覆盖范围：完成 798 个用例 的单功能验证，覆盖 4 大模块，通过率 100%。

数据仓库功能测试：44 个用例，含 6 个核心系统数据抽取、9 大主题层转换验证。

数据治理平台：10 个用例，聚焦定制功能（如登录界面、规则编码自定义等）。

数据应用：140 个用例，覆盖 1104 报表 51 张、EAST 4.0 报表 5 张、利率报备报表 3 张及补录功能。

其他模块：剩余用例未明确模块，推测为基础组件或接口测试。

2. 集成测试

测试内容：跨模块联动测试，验证数据链路完整性（如数据抽取→主题层转换→报表生成）。

执行强度：每日耗时 2-3 小时，确保模块间交互逻辑准确、数据流转无误。

3. 功能测试

参与方：业务部门深度参与，通过人工核对报表数据，验证业务逻辑一致性与准确性。

核心目标：确保系统功能满足实际业务需求，数据输出符合监管报送标准。

（二）试运行及优化

1. 上线进度

主体功能上线：2024 年 6 月完成开发测试后，核心功能正式投入使用。

试运行周期：2024 年 6 月至 2025 年 2 月，持续约 8 个月，期间逐步优化系统性能与功能。

结项时间：2025 年 2 月完成试运行及功能优化，并进行项目结项。

2. 试运行成果

系统稳定性：试运行期间整体运行稳定，并优化了上游数据变动对下游采集失败的解决方案，经内部、外部单位评审，确认具备验收条件。

监管报送验证：对业务部门使用固有监管报送功能的新需求点进行优化，并通过实际场景验证，满足合规要求。

六、项目核心成果

整体来看，本次项目以“数据治理筑基、数据仓库构体、数据应用赋能”为主线系统推进，为公司数据应用能力奠定核心基础。项目核心成果如下：

01**数据治理**

- ◆ **数据标准**：制定了符合苏信业务的企业级数据标准，共计**242**项基础数据标准，在数仓落标率**100%**
- ◆ **元数据管理**：全面盘点及梳理了元数据，覆盖核心业务系统，共计**205**张表和**10766**个字段并补充了中文说明、数仓内部数据血缘全覆盖**1639**项
- ◆ **数据质量专题**：结合人行资管的问题反馈及业务规则，编制了**120**条质量规则，发现**49**条问题，正在推进整改
- ◆ **数据治理平台**：上线数据质量监控、元数据管理模块

02**数据仓库**

- ◆ **数据架构**：完成了贴源层、历史数据层、主题模型层、公共计算层和数据访问层等5层的建设
- ◆ **数据采集**：贴源层共采集**2238**张表，TCMP：455张表、FA：389张表、TA：509张表、信托AM：475张表、固有AM：323张表、集团财务：87张表
- ◆ **数据模型及开发**：结合业务及同业实践构建9大主题，**78**个逻辑实体、**1439**个属性项、落地数据标准242项
- ◆ **ETL/调度监控**：完成了4大工作流开发，覆盖数仓数据加工、历史数据切片、报表重采提数等模块

03**数据应用**

- ◆ **EAST4.0报表（固有）**：完成了**5**张报表，**186**个指标的梳理与开发
- ◆ **利率报备报表（固有）**：完成了**3**张报表，**26**个指标的梳理与开发
- ◆ **1104报表（固有）**：完成了**50**张报表，**11103**个指标及数据项的梳理与开发
- ◆ **固有监管报送数据补录**：完成了**11**张业务补录表，**115**个数据项的梳理与开发

数仓及数据治理一期项目通过“治理体系化、仓库标准化、应用场景化”的实施路径，初步构建了符合信托业务特色的数据管理体系，实现了从数据无序到有序、从分散到集中、从被动合规到主动赋能的跨越。本期项目不仅为固有业务监管报送提供了高效工具，更通过数据治理机制的建立，为数据资产开发奠定了基础。