

## 数据仓库漫谈：信托业数据仓库架构设计与技术

数据仓库不仅是信托公司内部数据整合、分析和应用的关键基础设施，更是推动信托业务创新、提高风险管理和决策效率的重要工具。在深入推进数字化转型过程中，通过数字化技术记录大量的、形式多样的金融业务全流程数据，并且需要对这些数据进行挖掘分析，因此促进了金融数据仓库技术的快速发展和不断创新。承接《信托业数据仓库基础与发展现状》一文，本文将进一步探讨信托业数据仓库的架构设计、数据模型构建以及建设所需的关键技术。

### 一、数据仓库架构设计

#### （一）数据仓库架构设计原则

数据仓库建设作为企业层面的关键工程，其架构构建必须紧贴信托公司业务拓展与转型战略。信托公司数据仓库架构设计准则如下：



#### • 贴合业务实际需求

以数据分析和决策支撑为导向，建设目的精准对接信托业务，切实服务核心业务板块并攻克业务难题。

#### • 着眼企业整体目标

数据仓库架构的规划要全面考量信托公司的整体业务运作流程以及各类应用系统的设计

布局，进而塑造出企业级的数据架构体系。

- **架构清晰、数据流畅**

呈现出清晰的层次架构，各层级划分合理，职能界定明晰，数据在各层级间的流转有条不紊。

- **全流程标准化**

设计规范、流程规范、质量管控规范以及安全规范均需做到清晰明确、细致入微且具备切实的可操作性。

- **融合数据治理**

数据治理工作对高质量数据仓库建设起到强有力的推动作用，同时数据仓库的设计能够有效梳理数据流转脉络，促进数据治理工作的顺利开展。

## （二）数据仓库典型设计架构

信托公司数据仓库的结构设计通常分为五个层级，它们依次为：贴源数据层、整合模型层、汇总模型层、数据集市层以及数据应用层。然而，也有实践将之简化为三个层级，即：贴源层、整合层和应用层。具体采用何种分层模型，应基于信托业务需求和技术架构的具体情况来确定。

在五层架构的数据仓库中，各层的定义和作用如下：



数据仓库分层架构示例

**贴源数据层：**此层负责采集和存储原始数据，是信托业务数据的原始记录层，它直接从信托管理系统、交易系统等源头采集数据，不做任何改动，确保数据的真实性和完整性。

**整合模型层：**作为数据仓库的核心，该层根据信托业务特性构建抽象的数据主题模型，对所有业务数据进行有效的重组、整合和汇聚，形成稳定的数据基础。

**汇总模型层：**通过统一加工各类应用系统的共性指标和标签，比如信托总资产、平均收益率等，这些数据用于快速查询和报告，减少重复计算，提高效率。

**数据集市层：**针对不同的业务应用需求，进行数据的个性化处理，并提供对应的数据服务。如风险分析、客户服务、产品管理等，提供定制化的数据集。

**数据应用层：**该层负责将加工后的数据转化为具体的业务信息，比如通过数据分析来指导信托产品的设计、优化客户投资策略等，支持决策制定和业务流程的优化。

## 二、数据仓库模型

### （一）数据模型设计原则

根据信托行业的客户管理、运营和绩效管理、财务管理、风险管理、信息管理等领域对数据使用的要求，数据仓库模型设计应遵循如下原则。

01 兼顾泛化性和实用性

03 核心模型与扩展模型分离

05 成本与性能平衡



07 持续优化数据模型

02 高内聚低耦合

04 公共处理逻辑下沉

06 重视数据治理

#### • 兼顾泛化性和实用性

信托业务涉及多种产品和复杂的交易流程。数据模型设计时，要对业务进行深入分析，创

建的模型需要既能概括多种业务场景又能具体到操作细节。

- **高内聚低耦合**

不同类型的数据（如客户信息、交易记录、财务报表等）需要有不同的处理方式。相似或高度相关的数据应该整合在一起，形成内聚性高的数据模型，而业务差异大的数据则应保持独立，减少耦合，以便于管理和维护。

- **核心模型与扩展模型分离**

有些数据是日常运营的关键，如信托账户信息、交易数据等，这些应构成核心模型。而一些特定场景下的数据，如市场分析报告、客户个性化服务等，可以作为扩展模型。这种分离可确保核心模型的稳定性和扩展模型的可定制性。

- **公共处理逻辑下沉**

在数据模型处理中，公共的逻辑如日期计算、货币转换等应该在数据仓库的底层统一处理，避免在多个地方重复相同的逻辑，提高数据处理的一致性和效率。

- **成本与性能平衡**

信托业务产生的数据量可能非常大，因此数据模型设计时需要考虑存储成本和查询性能的平衡。适当的数据逆规范化可以提高查询效率，同时需要制定合理的数据更新策略以控制成本。

- **重视数据治理**

信托业务对数据的准确性、完整性和一致性有极高的要求。数据模型设计必须与数据治理紧密结合，确保数据质量符合监管要求，符合业务要求且模型定义标准化，便于不同系统间的数据交换和合规报送。

- **持续优化数据模型**

信托业务是不断发展的，数据模型也需要随之进化。通过建立评价体系，持续监控业务整合能力、数据模型质量和使用情况，及时对模型进行优化，确保其始终满足业务需求。

## （二）数据模型层次

数据仓库模型按照抽象程度由高到低依次是概念模型、逻辑模型、物理模型。



**概念模型：**关注的是业务的核心和关键部分。就像绘制信托业务的蓝图，只包括最基础和关键的业务实体，以及这些实体之间的关系，如委托人与受托人之间的信托关系。概念模型由业务需求驱动，帮助我们理解信托业务的主要流程，而不涉及具体的操作细节。

**逻辑模型：**逻辑模型是对概念模型的进一步细化，它详细描述了数据需求。相当于将信托业务的蓝图转化为具体的建筑计划，添加了更多的细节，如委托人的个人信息、信托财产的详细信息、受益权的分配规则等。但它不受特定技术或实施条件的限制，也就是说，它不关心数据最终如何在数据库中存储，只关心数据应该如何被组织和描述。

**物理模型：**是逻辑模型的实现，它将逻辑模型中的数据实体、属性、约束和关系转换为数据库中实际存储的结构。这就像将建筑计划转化为实际的建筑，物理模型设计确定了如何在数据库软件中创建表、字段、索引和其他数据库对象。例如，它定义了如何将信托合同、资产信息、交易记录等逻辑实体转换为数据库中的表和字段，并考虑了数据库的性能和存储效率。

### （三）数据仓库建模方式

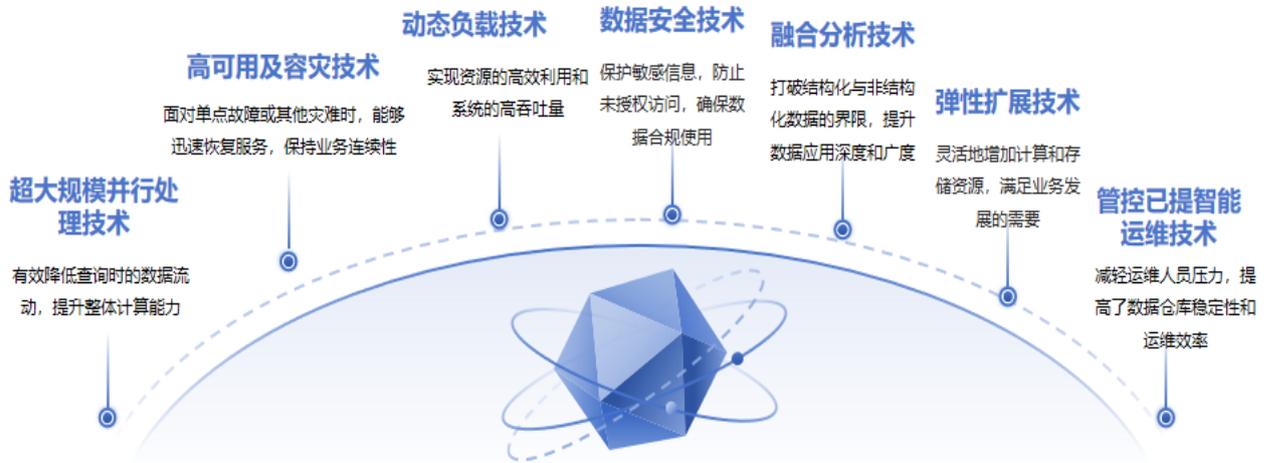
数据仓库建模方式主要包括范式建模和维度建模，这两种方法各有优缺点：

建模方式	优点	缺点
范式建模	<ol style="list-style-type: none"> <li>1. 范式化设计，数据冗余度低；</li> <li>2. 模型稳定，泛化能力强；</li> <li>3. 维护成本低。</li> </ol>	<ol style="list-style-type: none"> <li>1. 开发周期长，设计难度高；</li> <li>2. 大量表关联影响查询性能；</li> <li>3. 模型变动对数据应用的影响较大。</li> </ol>
维度建模	<ol style="list-style-type: none"> <li>1. 面向分析设计，模型简单，紧密贴合业务目标；</li> <li>2. 反范式化设计，开发周期短；</li> <li>3. 查询性能高、使用门槛低。</li> </ol>	<ol style="list-style-type: none"> <li>1. 数据大量冗余；</li> <li>2. 预处理阶段开销大，维护成本高；</li> <li>3. 受业务变动影响大。</li> </ol>

通过分析不同建模方式优缺点，贴源数据层直接采集业务系统数据，不需要建模；整合模型层、汇总模型层抽象信托业务形成统一视图，模型稳定性很重要，一般采取范式建模。数据集市层和数据应用层贴合业务需求，需求变化快，一般以维度建模为主。

### 三、信托业数据仓库的关键技术

数据仓库运用的关键技术是平台能力强大的支撑，以应对海量数据的处理、保障系统高可用性、管理多样化负载、确保数据安全、促进数据分析融合、实现弹性扩展等目标。



### （一）超大规模并行处理技术

超大规模并行处理技术是核心，通过分布式架构和云原生架构，实现数据的分布式存储和并行处理。这不仅提高了查询效率，还确保了 PB 级数据的高性能计算和存储需求得到满足。通过节点间和节点内的并行技术，能够有效降低查询时的数据流动，提升整体计算能力。

### （二）高可用及容灾技术

高可用及容灾技术对于信托业务至关重要。采用数据多副本、集群内高可用性设计以及双集群容灾部署，可以确保数据仓库在面对单点故障或其他灾难时，能够迅速恢复服务，保持业务连续性。

### （三）动态负载管理技术

动态负载管理技术可以根据不同的业务需求，智能分配系统资源。包括静态和动态资源管理、优先级控制、异常处理和快慢查询管理，从而实现资源的高效利用和系统的高吞吐量。

### （四）数据安全技术

在数据安全方面，数据安全技术通过透明数据加密、数据脱敏和权限分离等措施，保护敏感信息，防止未授权访问，确保数据合规使用。

### （五）融合分析技术

融合分析技术打破了结构化与非结构化数据的界限，使得核心仓库与大数据区数据能够高效关联，提升了数据分析的深度和广度。

### （六）弹性扩展技术

为了适应不断增长的数据量，弹性扩展技术提供了在线按需扩展的能力。无论是物理部署还是云化部署，都能够灵活地增加计算和存储资源，满足业务发展的需要。

### （七）管控一体的智能运维技术

管控一体的智能运维技术通过全面的监控、在线运维和智能诊断，减轻了运维人员的压力，提高了系统稳定性和运维效率。包括资源监控、系统监控、SQL 诊断和智能运维，确保了数据仓库的长期稳定运行。

在信托业数据仓库的架构设计与技术发展过程中，技术的创新与业务的融合是推动行业进步的双引擎。通过对架构设计的精细打磨、数据模型的精准构建，以及关键技术的深入应用，数据仓库已成为信托行业内支撑业务决策、提升运营效率、强化风险管理的重要基石。