



大数据技术标准推进委员会

Big Data Technology and Standard Committee

企业数据架构实践指南

(1.0版)

CCSA TC601 大数据技术标准推进委员会

2024年12月

版权声明

本报告版权属于 CCSA TC601 大数据技术标准推进委员会，并受法律保护。转载、摘编或利用其它方式使用本报告文字或者观点的，应注明“来源：CCSA TC601 大数据技术标准推进委员会”。违反上述声明者，将追究其相关法律责任。

编制说明

本指南的撰写得到了企业数据治理领域多家企业与专家的支持和帮助,主要参与单位与人员如下。

参编单位

中国信息通信研究院云计算与大数据研究所、中国平安人寿保险股份有限公司、中国东方航空股份有限公司、中国建设银行股份有限公司、中国移动通信集团有限公司、中国移动通信集团浙江有限公司、中泰证券股份有限公司、阿里云计算有限公司、上海星顺信息技术有限公司、上海汉得信息技术股份有限公司、中体彩彩票运营管理有限公司、杭州比智科技有限公司、中国南方航空股份有限公司、海亮教育科技服务集团、人保信息科技有限公司、湖北数据集团有限公司、杭州衡泰技术股份有限公司、深圳市城市交通规划设计研究中心股份有限公司、中汽数据(天津)有限公司、交通银行股份有限公司、联通数字科技有限公司、中国联通北京分公司、中国联通集团、振华智造(西安)科技有限公司、中电信人工智能科技(北京)有限公司、华夏银行股份有限公司、中兴通讯股份有限公司、中移(苏州)软件技术有限公司、平安银行股份有限公司、国际商业机器公司、中信证券股份有限公司、海洋石油工程股份有限公司、湖北中烟工业有限责任公司、国泰君安证券股份有限公司、国家石油天然气管网集团有限公司、北京智网数科技术有限公司、海尔集团公司、中车信息技术有限公司、中车南京浦镇车辆有限公司、一汽-大众汽车有限公司

参编人员

阚鑫禹、张一鸣、王妙琼、姜春宇、王宇龙、刘思达、尹正、高倩倩、周京晶、张娇婷、周圣文、田明慧、林木森、高海暘、季文彬、周鹏、刘静莉、查毓灵、车春雷、王冀彬、江勇、陈永刚、宋秀红、郝玉刚、何强、蔡枫桦、曹正炎、童文斌、李锦、王鹏、李晓燕、项灵刚、史赞、张东飞、薛良、史凯、楚思思、蒋正东、潘任虹、卓锦坤、周强、唐晶晶、李金夏、朱瑞、王远、王项男、文雪颖、袁雪梅、薄冉、刘洋、谭立何、马亚东、张荣楷、刘继荣、范超、王斌、李乌英嘎、王灏、王桂强、刘殿兴、陈辉华、王会峰、张良轩、王子申、陈永刚、胡业伟、吴靖宁、王烨、张玲岩、于涛、宋一纯、张建军、徐加兴、李明旭、孟庆亮、谢葳、武艳军、徐聪颖

前言

随着新一轮科技革命和产业变革深入发展，数据作为关键生产要素的价值日益凸显，标志着以数据为关键要素的数字经济进入新时代。数据要素的规模报酬递增、低成本复用、融合创新等特点，正在催生新业态新模式，培育经济发展新动能。

数据架构是对数据资源基本特性的描述和管理，企业数据架构立足企业全局视角对全量数据资源进行统一管理和规范，为数据要素驱动的新业态、新模式提供最基础支撑。一方面企业数字化改革转型的持续深入，应用系统之间和数据应用场景之间的关联关系愈发复杂，数据架构已经成为深度刻画企业数字化环境的核心手段，支持更高水平的应用集成和数据集成。另一方面以数据质量和数据安全为代表的企业数据治理工作要求持续提高，亟需数据架构管理作为核心枢纽进行数据管理一体化能力的贯通。

本指南对企业数据架构的概念、框架和管理活动进行了阐述，提出企业数据架构实践的意义和误区，结合典型企业实践案例进行分析，并展望了未来企业数据架构的发展趋势。

目 录

一、企业数据架构概述

（一）数据架构典型框架.....	1
（二）企业数据架构管理意义.....	3
（三）企业数据架构管理挑战.....	4

二、企业数据架构框架

（一）企业数据架构管理定义.....	5
（二）企业数据架构管理定位.....	5
（三）企业数据架构参考框架.....	7

三、企业数据架构管理活动

（一）数据架构规划.....	9
（二）数据标准管理.....	10
（三）数据分布管理.....	13
（四）数据模型管理.....	16
（五）主数据管理.....	17

四、企业数据架构误区

（一）过度强调技术的先进性.....	19
（二）过度依赖 IT 架构能力.....	19
（三）数据架构与项目开发脱节.....	19
（四）将湖仓等同于数据架构.....	20
（五）数据架构局限在 IT 部门.....	20
（六）数据架构不是必选项.....	20

五、企业数据架构案例

（一）东方航空.....	21
（二）平安人寿.....	26
（三）交通银行.....	30
（四）中车浦镇.....	33

六、企业数据架构展望

图目录

图 1 数据架构与其他架构的关系.....	6
图 2 数据架构与数据管理活动的关系	7
图 3 企业数据架构框架	8
图 4 企业数据架构管理活动	9
图 5 数据标准间关联关系	10
图 6 指标数据构成.....	11
图 7 数据标准与数据模型的关系.....	12
图 8 业务架构下的数据流关系.....	13
图 9 业务流程 CU 矩阵.....	13
图 10 业务流程 CU 矩阵	14
图 11 零售企业数据仓库总线矩阵	15
图 12 数据采集接口清单.....	15
图 13 企业级数据模型分层框架	16
图 14 企业级数据模型应用和管控路径.....	17
图 15 东航数据架构体系.....	22
图 16 东航数据模型主题域划分	23
图 17 东航主题概念数据模型.....	24
图 18 东航数据模型管控流程.....	25
图 19 东航概念模型 - 航班运行领域	25
图 20 东航航班飞行领域逻辑模型	25
图 21 平安人寿数据底座全景.....	27
图 22 平安人寿数据底座 ADS 建设.....	27
图 23 平安人寿企业架构元模型	28
图 24 交通银行数据架构领域.....	31
图 25 双“T”数据模型	31
图 26 数据标准管控流程.....	32
图 27 浦镇公司数据质量稽核示例	35
图 28 浦镇公司概念模型 - 售后管理域	36
图 29 浦镇公司逻辑模型示例 - 售后管理域	37
图 30 浦镇公司数据分布视图示例	37
图 31 浦镇公司数据应用	38

一、企业数据架构概述

本章聚焦企业数据架构在实践过程中面临的机遇和挑战，梳理了典型数据架构理论框架的特点和使用情况，阐明企业数据架构管理的重要意义，为理解企业数据架构提供宏观概念。

（一）数据架构典型框架

在企业架构实践和数据管理实践的推动下，企业数据架构已经形成了多种不同的框架理论，为企业数据架构的规划、建设和管理工作提供了理论和实践指导。下文中列举了国内外典型的数据架构（包含信息架构，在本文中不作区分）理论框架，这些框架都是在企业实践中广为传播和使用的，是当下业界开展数据架构相关工作的理论基础。

Zachman 1987 年 Zachman 在《信息系统架构框架》中提出了首个企业架构框架，被业界视作企业架构的源头。该框架主要表现为一个 6 行 6 列的矩阵，6 行是基于 5W1H 的分类，即 Why（动机）、What（数据）、Who（角色）、When（时间）、Where（分布）、How（功能），6 列是基于不同人员的视角（高层、业务管理者、架构师、工程师、技术员、企业用户），完成从抽象概念到实例的转换（识别、定义、表示、详述、配置、实例化）。Zachman 架构框架给出了企业架构内容的描述和分类，确保每个干系人的每个关注点都被照顾到、并有效地串联成一个完整的体系，提供了一种有效分解复杂企业系统的方法。

IFW IFW（Information Frame Work）是由 IBM 为金融服务行业量身定做的信息架构。IFW 框架的核心内容涵盖了企业模型、数据模型、功能模型、数据仓库模型等多个方面，通过整合不同的技术和业务模型，支持银行与金融机构的数字化转型和业务升级。通过核心系统建设、数据仓库建设等方式，IFW 在国内银行业机构中得到了广泛应用。虽然 IFW 起初是为金融服务业开发的产品，但是其核心理论方法在各行业中同样适用。

TOGAF TOGAF 标准是 The Open Group 的企业架构标准，目前最新版本为第 10 版。TOGAF 是目前全球市场认可度最高、使用最广泛的企业架构框架及实践指南。在 TOGAF 框架中，数据架构是企业架构的关键组成部分之一，定义了数据收集、存储、整合、访问和使用的策略、模型和标准。TOGAF 框架中数据架构部分由数据模型、数据流、数据字典、数据标准、数据治理等组成，旨在确保数据一致性、支持业务需求、提高数据质量、促进数据共享利用。

DMBOK DMBOK 是国际数据管理协会（DAMA）的数据管理知识体系，是当前业界开展数据管理工作的重要参考理论，也是数据治理从业人员的学习教材。DMBOK 强调数据架构是数据管理的基础，数据架构由企业数据模型和数据流设计组成，并指出简化数据架构这一复杂问题的两种方式，一是面向质量，专注于业务和 IT 开发周期内的数据架构改进；二是面向创新，专注于业务和 IT 转换。

FEAF 在《克林格 - 科恩法案》法案的推动下，美国联邦政府在 1999 年发布了 FEAF（Federal Enterprise Architecture Framework）框架，用于指导联邦政府各部门创建企业架构。FEAF 框架通过绩效参考模型、业务参考模型、数据参考模型、应用参考模型、基础设施参考模型和安全参考模型描述战略、业务、数据、应用、基础设施和安全六个子架构。FEAF 框架主要包括 8 个部分，分别是体系架构驱动、战略方向、当前体系架构、目标体系架构、变化过程、体系架构模型和标准。

DoDAF DoDAF（美国国防部体系结构框架）是支持国防部首席信息官履行《克林格 - 科恩法案》所要求的架构开发和维护职责的主要支柱之一。DoDAF 起源于 Zachman 框架，由 C4ISR 体系架构框架演化而来，使美国国防部各级管理者能够打破国防部、联合能力域（JCA）、部门或项目等层次界限，实现有序的信息共享，提高关键决策能力。最新的 DoDAF 2.0 以数据为中心，以支持核心决策过程、开发符合用户需要与目的为根本出发点，是一个真正实现以数据为中心的全新架构框架。DoDAF 2.0 定义了数据与信息、能力、作战等 8 个视图，可以全方位、全过程的对复杂军事系统的分析设计提供路线图和建模框架。

DCMM GB/T-36073《数据管理能力成熟度评估模型》（简称“DCMM”）是我国首个数据管理领域正式发布的国家标准。在 DCMM 中数据架构是关于企业如何组织和设计数据资源的整体规划，旨在确保数据在企业内部的流通、集成、共享，以及管理的一致性和有效性。DCMM 的数据架构域包括数据模型、数据分布、数据集成与共享和元数据管理。随着 DCMM 贯标评估工作在全国范围内的持续、深入开展，DCMM 中对数据架构管理的能力要求得到了企业的广泛重视和采纳，并以 DCMM 框架作为企业自身开展数据架构管理的重要指导理论。

华为数据之道《华为数据之道》中介绍了华为的企业级信息架构，以结构化的方式描述在业务运作和管理决策中所需要的各类信息及其关系的一套整体组件规范，包括数据资产目录、数据标准、企业级数据模型和数据分布四个组件。该框架是华为在实践过程中构建的一套对业务运作数据进行有效管理的信息架构方法论，用于指导企业内部各部门的信息架构建设工作，让管理者、专家和员工之间有共同语言。华为企业级信息架构在业界得到了广泛传播和应用，特别是数据资产目录 L1-L5 的 5 层架构体系，已经成为诸多企业描述自身数据架构的核心框架。

（二）企业数据架构管理意义

数据架构在企业的整体架构中起到了承上启下的关键作用，作为连接纽带向上支持业务架构，向下驱动应用架构和技术架构。数据架构作为数据管理工作的核心指引手册，扮演着至关重要的角色。它不仅是数据治理的基石，更是企业数据战略实施的蓝图，为数据管理工作提供了明确的指导和方向。

指导数字化规划 数据架构站在企业级视角，建立数据模型和数据流向，为业务系统建设、重构、优化提供输入，包括数据模型及集成关系的落地等，为企业的数字化规划提供了清晰的数据蓝图，能够帮助后续信息化项目简化、省去调研和梳理业务及数据的过程，有助于应用系统的快速迭代和优化，避免重复建设。

促进数据互联互通 关注数据应用的程度越深，则对数据架构的管控力度则越强。企业数据架构对数据互联互通起着决定性作用，直接决定了数据应用的程度和水平。企业数据架构通过定义统一的数据标准、构建数据集成平台、优化数据流通过程以及强化数据安全机制，确保了企业内外数据资源的无缝对接与高效共享，促进了不同系统、不同部门之间的数据流通与协作。

提升数据一致性 企业数据架构通过对数据描述、数据分布、数据流动的约束和管理，有利于统一数据定义、确认权威数据源、消除冗余数据、消除数据混乱，确保在数据生命周期全过程的一致性和标准化，为数据需求的理解、部门间沟通协调、数据质量主动设计提供根本保障。

提升数据处理效率 在大数据技术性能提升和资源追加投入的基础上，通过企业数据架构提供结构化的效率提升方式，从根本上减少不必要的数据处理环节、优化数据流转链路、科学追加存储和计算资源，从而在全局层面提高数据查询、处理、传输等工作效率，减少延迟、开销和冗余。同时还可以显著降低数据维护成本，提高自动化水平并减少人工干预和错误。

降低企业管理风险 数据和 IT 系统是企业风险管理的基础，构建与业务模式、经营结构和风险配置相适应的数据架构，有利于企业应对海量的内、外部风险数据，为各类风险的识别、计量、监测和控制提供数据基础。良好的企业数据架构为风险数据加总、风险暴露、风险报告、信息披露等提供支持，提高内控合规水平并降低企业管理风险。

（三）企业数据架构管理挑战

企业数据架构管理得到了众多企业的重视，并将架构作为优先管理内容，但是受数据结构、管理协同、存量系统等现实因素的影响，数据架构在执行层面面临严重挑战。

多模态数据管理难 在企业数字化转型的浪潮中，多模态数据管理难题愈发凸显，涉及结构化、半结构化和非结构化数据的并存与复杂交互。多源数据接入、多模态数据检索、多模态数据融合分析、多模态数据存储等问题给企业数据架构带来重大挑战，对企业数据架构的兼容性、灵活性、拓展性提出了较高要求。

企业架构协同难 企业架构通常包括业务架构、数据架构、应用架构和技术架构（4A），由于 4A 架构各有特点和侧重，表现形式差异大，而相互之间又紧密联系并相互依赖，在实际的架构建设和架构管理过程中企业经常会面临 4A 协同困难的局面。

架构遵从执行难 企业架构和数据架构的落地实践是企业面临的首要挑战，背后有几方面的原因。一是业务需求和业务形态快速变化，架构的响应和支持不及时；二是 4A 架构分属不同团队管理，且业务部门具有较强的自主性，企业架构和数据架构难以统一推进；三是多数企业的架构建设落后于 IT 系统建设，架构规划和架构建设工作长期处于滞后状态，企业架构处于被动管理局面。

存量系统改造难 存量系统是数据资源积累的核心载体，绕开存量系统的数据架构将沦为空谈。企业数据架构的实践过程中需要协调不同 IT 系统之间的矛盾和冲突，处理好数据质量的一致性、完整性和实时性问题。同时应谨慎论证和规划新系统的建设和引进，如果架构遵从不到位则增量部分在系统上线后便会成为新的“孤岛”。

二、企业数据架构框架

本章给出了企业数据架构管理定义，阐述了企业数据架构在企业架构和企业数据治理体系中的管理定位，并给出了企业数据架构的参考框架，供企业在实践过程中参考和裁剪，以便开展业务实例化。

（一）企业数据架构管理定义

企业数据架构管理以企业内所有数据为管理对象，对数据的静态属性和动态特性进行管理，对数据的定义、分布、集成、整合等进行规划、管控和应用，确保数据与业务需求保持一致。

（二）企业数据架构管理定位

企业数据架构框架是一套指导企业开展数据架构建设和管理的参考框架，企业根据自身实际情况进行裁剪和挑选使用。数据架构作为一个独立整体时，既是企业架构的重要组成部分，又是数据管理体系的重要基础。

企业架构通常包含业务架构、数据架构、技术架构和应用架构，数据架构承担着从业务架构向应用架构、技术架构落地的桥梁作用。数据架构负责描述业务架构中的数据实体和数据需求，并将其转换为应用架构中的物理数据对象。数据架构与业务架构完全对齐，即便业务活动尚未得到应用系统的支撑，或尚未实现业务在线化、数字化，业务架构中的数据实体也应包含在数据架构中。

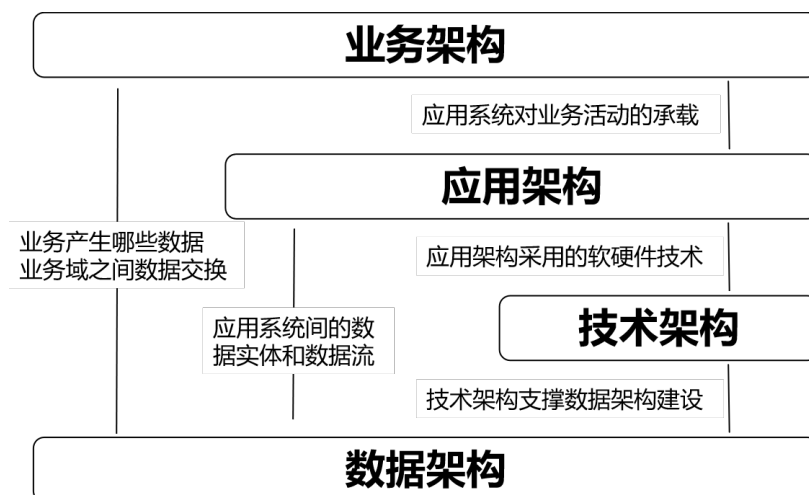


图 1 数据架构与其他架构的关系

当下数据治理的理念已在企业数字化转型过程中达成了广泛共识和重视，部分企业将数据架构的建设和管理作为数据治理工作的核心内容，已不单通过企业架构建设工作构建数据架构。数据架构与数据治理组织体系、数据共享、数据质量、数据安全、数据开发等诸多领域有紧密的关联关系。企业在管理数据架构的过程中常面临“管理孤岛”的挑战，将数据架构的建设和维护与其他工作割裂，难以形成联动局面。下图简要说明了数据架构在数据治理体系中的特殊作用，具体如下：

- ① 数据治理的组织体系和管理体系本质上由数据架构体系决定，以数据架构指导数据认责或归口管理工作开展；
- ② 数据质量管理可以借助数据架构进行问题原因排查，推动数据质量在源头得到改进和治理；
- ③ 数据开发工作在数据架构的支持下，可以大幅降低需求理解门槛，减少数据开发工作量，同时避免任务的重复开发；
- ④ 数据安全治理借助数据架构可以快速构建全生命周期的整体视图，提高数据分类分级的覆盖程度，准确识别关键数据的关键安全环节；

⑤ 数据共享工作可以通过数据架构快速高效定位数据资源，并有利于将应用系统内外的数据资源有序归集，形成高质量的数据集。

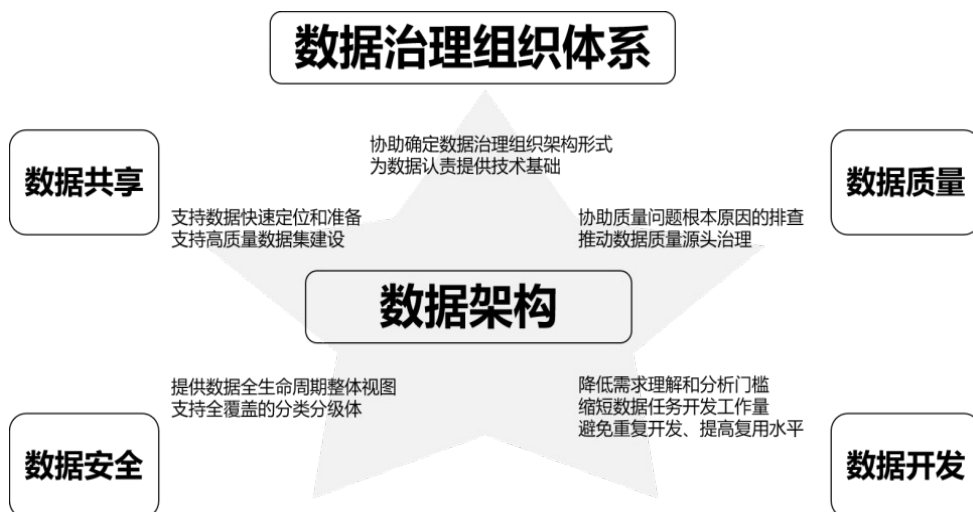


图 2 数据架构与数据管理活动的关系

（三）企业数据架构参考框架

当前数据架构的参考框架繁多，本节以企业实践现状为基础，围绕企业中联机分析处理（OLAP）和联机事务处理（OLTP）两类主要数据处理形式，从数据静态描述和动态描述视角提出数据架构的参考框架，主要内容如下：

① OLTP 侧以应用系统集成为核心，OLAP 侧以数据集成整合为核心，分别支持业务交易活动和业务分析活动的顺利开展；

② OLTP 侧以范式数据模型为核心，确保应用系统满足业务交易的数据需求；OLAP 侧以维度模型为主，承接 OLTP 侧的物理数据实体后进行转换整合，满足跨业务域的数据分析挖掘等需求；

③ OLTP 和 OLAP 侧均需要数据标准的全面规范和支持，前者以数据项标准为主，后者以指标数据标准为主；

④ OLTP 侧以数据分布、数据流管理为主，核心在于构建数据实体与业务流程的关联关系；OLAP 侧以数据采集接入、总线矩阵管理为主，确保数据资源的准确采集和高效汇聚，保证数据分析结果回流到业务活动中；

⑤主数据是介于应用集成和数据集成之间的一种特殊状态，其本身是业务数据的一种特别状态，但其既可以有力推动应用集成，又可为数据集成和数据分析挖掘提供高质量的核心业务数据资源；

⑥企业数据架构参考框架需要结合企业实际的数字化现状进行重构，并进行业务的实例化才能发挥作用。

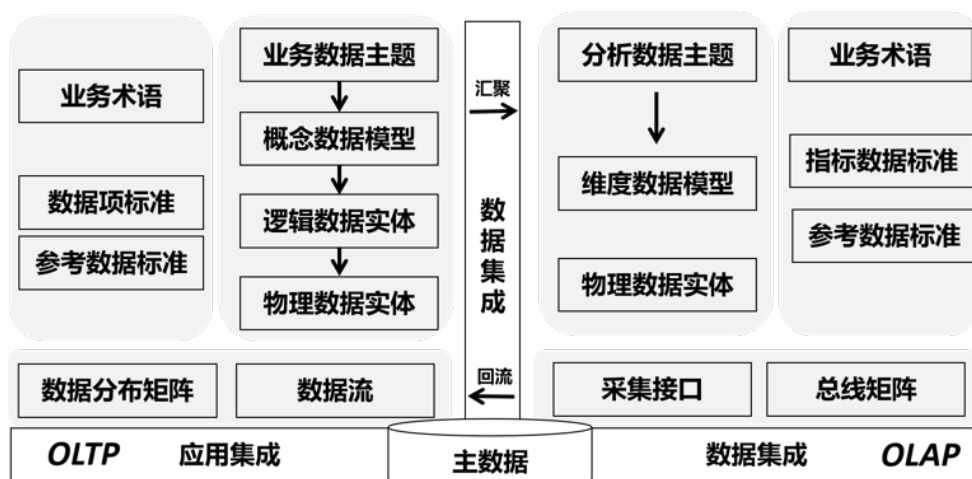


图 3 企业数据架构框架

三、企业数据架构管理活动

本章主要阐述企业数据架构的主要管理活动，将架构规划、数据标准、数据模型和数据分布作为企业数据架构管理活动的核心内容，用以指导企业数据架构的建设和管理。由于主数据管理是在企业全局视角下围绕核心业务对象的数据管理工作，保证核心业务对象数据的全局一致性，故将主数据作为数据架构落地的一种典型模式。

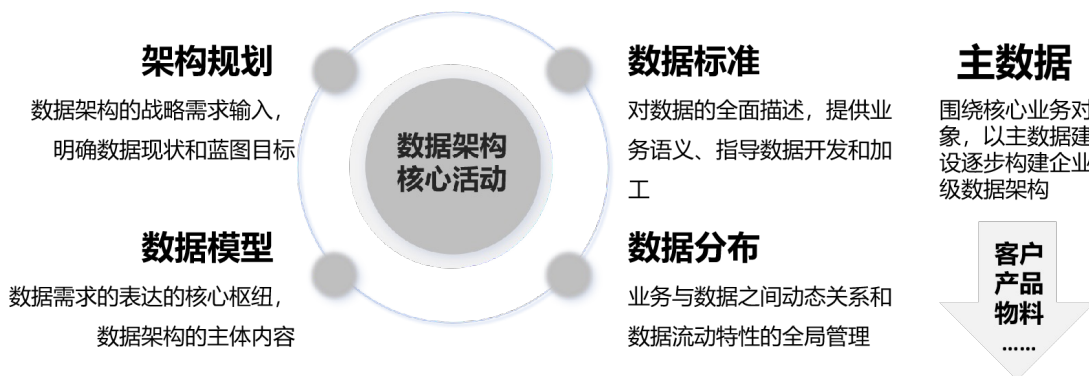


图 4 企业数据架构管理活动

（一）数据架构规划

数据需求分析 数据需求分析是数据架构规划的第一步，宏观、微观的数据需求是开展数据架构规划、建设和管理的源头动力。结合业务需求，分析数据的来源、类型、结构和质量等要求，为数据架构的设计提供支持，主要任务包括业务需求收集和分析挖掘、确定数据量规模和类型、评估数据处理性能需求、分析数据安全需求等。数据需求分析的本质是对业务活动的分析，收集和整理业务活动的信息需求、数据需求，要求分析人员具有业务背景和业务知识，是一个常常被忽视的重要环节。

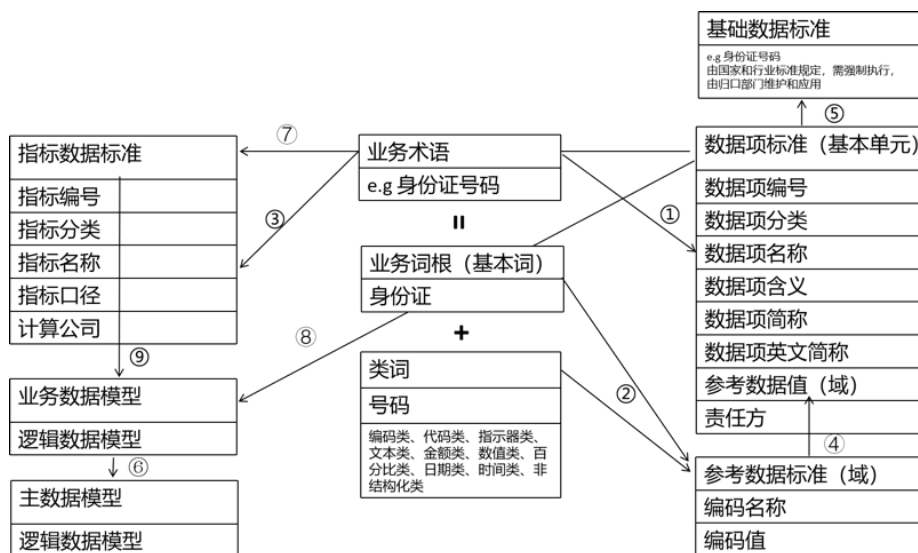
数据资源盘点 数据资源盘点是对企业数据资源现状的深入分析和全局情况归集，为数据管理的整体工作提供数据基础，很多企业都将数据资源盘点作为数据管理工作的起点，并形成了工作方法和技术要求。数据资源盘点涉及企业内所有数据资源，包括数据库、文件系统、云存储和第三方数据源等，除了对数据库、表、字段的结构和含义进行盘点外，还会对关联关系、使用频率、敏感等级、管理部门等情况进行分析。

构建数据全景视图 企业开展数据资源盘点的现实目标是构建企业数据全景视图或数据地图，配合资源目录、数据门户等工作，以快速、精准、自助式地找到数据。通常会将盘点作为年度的重点工作任务，通过几期项目完成全域数据盘点工作，完成数据架构管理的初始化。在实际盘点工作中，还会涉及到业务架构和应用架构的盘点工作，盘点工作大多是从 IT 系统的维度展开，部分企业为梳理清楚数据与业务的动态关系，还会从业务流程入手梳理数据。

数据架构蓝图设计 数据架构蓝图设计的主要内容包括定义数据标准、设计数据模型、设计数据分布和集成关系、制定质量和安全策略、设计数据访问策略等。企业数据架构的蓝图是数据架构管理各项工作的目标图，为各项管理工作提供清晰的目标和路径，并纳入企业中长期的数据战略或数字化战略规划中。蓝图的设计应以数据需求分析结果和数据资源盘点结果为基础，兼顾企业当前现状并对未来的发展做前瞻性规划，但是企业目前面临的挑战在于数据架构蓝图与业务规划、应用规划等紧密关联，经常出现数据架构蓝图缺失的情况。

（二）数据标准管理

数据标准管理已经作为企业数据管理体系一个独立领域，部分企业会设置独立的数据标准管理处进行管理。数据标准制定活动包括定义业务属性标准、技术属性标准、管理属性标准等，其中技术属性标准是数据架构重点管理的范畴，主要包括数据命名规则、格式、类型和质量标准，确保数据在整个组织中的一致性和可比较性，促进跨部门和跨系统的数据整合。由于数据标准是描述数据架构的基础，数据标准之间的关联关系直接决定和影响数据架构内容，本指南不对数据标准的具体管理过程进行阐述，重点分析不同数据标准对象之间的关联关系。



- ① 业务术语是数据项定义的基础，基于业务术语可以做业务含义的演化和派生，基于业务词根和数据元组合的方式可以实现；
- ② 业务词根和数据元指导参考数据的定义，形成字段含义的内容取值清单；
- ③ 业务术语支持指标口径和指标名称的定义，基于业务词根和数据元组合的方式可以实现；
- ④ 参考数据可以视作数据项标准的一部分，由于很多参考数据本身具有重要的业务和技术内涵，需单独进行管理；
- ⑤ 越是重要的参考数据、数据项，国家和行业的管控力度越强，例如币种、地区等，所以需要单独进行维护，也是对接国家和行业，支撑跨企业数据协同的基础；
- ⑥ 主数据是业务数据的子集，主要突出跨系统和跨流程共享；
- ⑦ 指标数据在构建过程中需要以数据项标准为基础；
- ⑧ 业务数据在进行数据建模过程中，需要以数据项标准为基础，也就常说的贯标。需要指出的是，在数据标准的管理范畴中，数据模型一般指逻辑模型，并非直接在系统中产生的物理数据记录；
- ⑨ 在进行业务数据建模的过程中，需要统筹考虑指标数据标准和指标应用需求，内置部分加工计算逻辑。

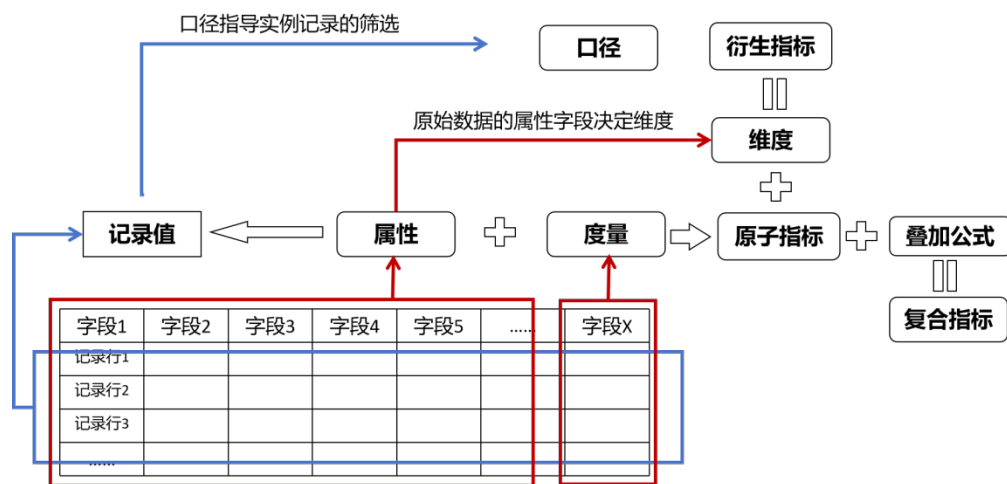


图 6 指标数据构成

指标数据是企业数据管理和应用的核心内容，指标数据标准与字段、记录值、计算公式等内容存在联动关系，对实际的数据资源归集和数据加工处理起着主导作用。由于指标口径指导原始数据记录的筛选，属性字段决定指标构成和指标维度，指标数据标准既需要一套完整的指标体系框架，又需要和实际数据资源进行联动考虑，指标的加工来自于经营管理需求，但本质上受到数据架构的限定。

数据标准是数据架构的核心组件，标准对数据模型和数据库开发有着重要影响，无论是交易侧的业务系统数据库开发还是分析侧的数据整合应用，都需要企业级统一的数据标准作为支撑。在系统开发和数据开发过程中对字段标准和参考数据标准进行审查，在分析应用的数据开发过程中对指标数据标准的审查，是保证数据标准落地的重要措施。企业开展的数据标准贯标检查工作，也是在提高数据架构的一致性。

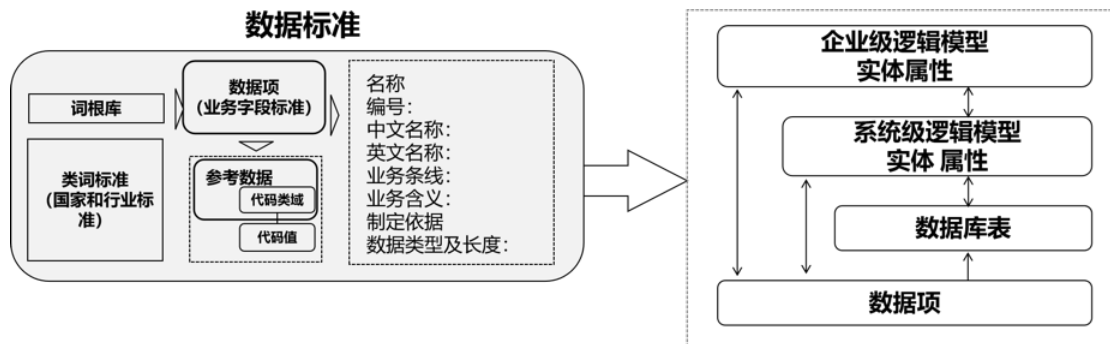


图7 数据标准与数据模型的关系

在数据架构管理的视角下，企业数据标准管理面临的重要挑战一是没有数据标准意识，在数据汇聚整合后出现大量矛盾和分歧，难以快速开展加工应用；二是数据标准与数据架构缺少关联关系，数据标准的贯彻应用缺少具体路径，数据标准与应用架构脱节；三是数据标准规模失控，关联关系的缺失导致数据标准数量快速增长，“同名不同义、同义不同名”的情况频繁出现。

(三) 数据分布管理

企业数据分布管理是极具挑战的一项管理内容，在实践过程中首先面临的挑战是数据分布的管理范围和表现形式是什么，因为这直接决定了数据分布管理的范围、对象、颗粒度、任务量。数据分布管理应是对企业整体范围内所有数据的管理，主要聚焦在数据源头、数据分布、数据流、数据集成关系等动态关系的管理，管理数据实体与业务流程、IT 系统的关联关系。常见的数据分布管理手段有以下几种。

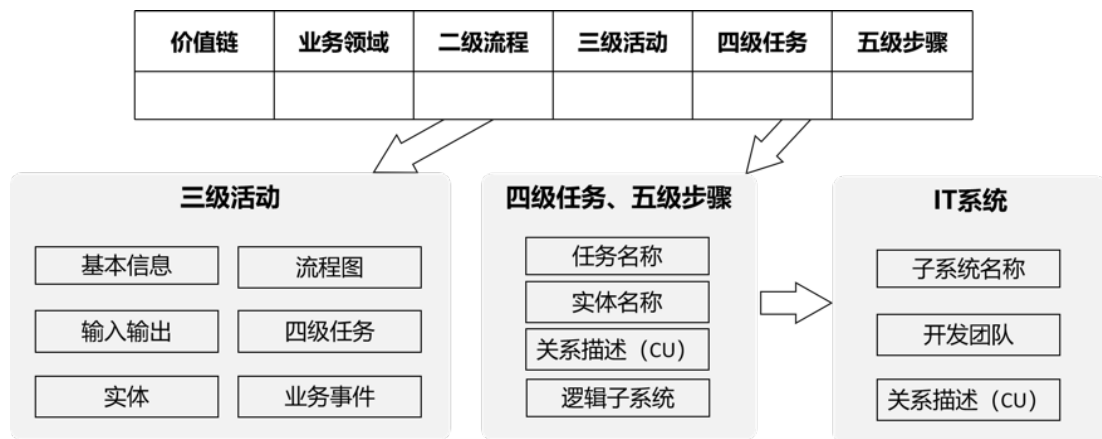


图 8 业务架构下的数据流关系

数据流 从业务架构视角下数据分布管理以数据流的形式展现，在业务领域、流程、活动、任务、步骤的五级业务架构中，数据流用于描述某一业务对象在信息系统中的引用关系。完善的数据流结构依赖于完整的业务流程体系，数据流程不仅可以指导 IT 系统的开发，还可以为数据资源定位、数据质量和安全问题溯源提供支持。

数据项/业务流程	文件保管	问题意见	维护手册	交易指令	服务合约	项目方案	
合约帐户	☐ C ☐ R ☒ U	☐ C ☐ R ☒ U	☐ C ☐ R ☒ U	☐ C ☒ R ☐ U	☐ C ☒ R ☐ U	☒ C ☐ R ☐ U	☐ C ☐ R ☐ U

图 9 业务流程 CU 矩阵

CU 矩阵 为了更加直观地对数据流进行管理，可以通过 CU 矩阵对数据分布关系进行管理，CU 矩阵以二维表格形式将数据项和业务流程进行关联，记录业务对象在每一个流程步骤中被创建（Create）、读取（Read）、更新（Update）、删除（Delete）的情况，明确业务对象在业务流程中的 CURD 操作。CU 矩阵既可以管理数据对象的分布情况，还可以了解各业务环节的数据操作内容，是在企业实践过程中经常使用的管理方式。在主数据的规划和建设过程中，也可以将属性字段与 IT 系统进行关联，从而准确得到不同业务流程和 IT 系统间的共享属性字段，为主数据标准建设和主数据记录分发提供支持。

数据项/IT系统	CRM	PDM	MES	CWM	ERP	
客户	☒ C ☐ R ☐ U	☐ C ☐ R ☐ U	☐ C ☒ R ☐ U	☐ C ☒ R ☐ U	☐ C ☐ R ☐ U	☐ C ☐ R ☐ U

图 10 业务流程 CU 矩阵

权威数据源管理 权威数据源是以信息链和数据流为基础，分别分析数据在业务流程和信息系统中首次被创建的流程步骤或系统功能，并经企业数据管理组织认证，确定为数据在真实的业务或系统中的源头。权威数据源应在企业范围内统一发布，作为唯一数据源头被周边业务或信息系统调用。权威数据源通常也会作为企业数据认责工作的重要内容，明确各部门归口管理的数据对象。

数据分布应尽量消除数据冗余，确保数据权威、一致，但是考虑到数据使用过程的效率、时效和性能等各种因素，为了满足不同业务之间对重要数据的共享需要，主辅数据源管理至关重要。一般由产生数据、频繁使用、业务关联紧密的系统作为主数据源，为了缩短业务操作过程、提高数据使用效率、提高系统平台性能等，允许系统间进行数据同步或复制。但是数据同步需要遵循基本的架构管控原则，例如所有辅数据源应由主数据源提供数据、辅助数据源的数据更新不应早于主数据源等。

总线矩阵 总线矩阵是数据仓库建设的重要内容和成果，基于业务过程构建一致性维度和一致性事实，指导数据仓库的建设。总线矩阵可以帮助各方人员直观了解数据架构体系，提高数据一致性、避免重复建设、关联主数据资源，是需要架构师进行长期维护和全局共享的重要内容，下图所示为 Kimball 提出的零售企业数据仓库总线矩阵。总线矩阵是数据仓库建设的体系性指导，其独立于具体技术，用于规划和指导数据仓库、BI 等系统的长期数据内容需求，既保证各团队相对独立地开展工作，又确保各项成果可以整合成一个完整的数据仓库。

业务过程	公共维度						
	日期	产品	仓库	商店	促销	客户	雇员
提出购买订单	×	×	×				
接受仓库存货	×	×	×				×
仓库库存	×	×	×				
接收商店存货	×	×	×	×			×
商店库存	×	×		×			
零售	×	×		×	×	×	×
零售预测	×	×		×			
零售促销跟踪	×	×		×	×		
客户退货	×	×		×	×	×	×
退货至供应商	×	×		×			×
常客注册	×			×		×	×

图 11 零售企业数据仓库总线矩阵

数据采集接口清单 企业为提高数据资源的汇聚整合程度，在大数据体系建设和数据挖掘应用等工作中，对数据采集接入十分重视。由于数据来源的持续增多，企业需要持续维护和更新数据采集规范，除采集周期、接口技术方式、采集对象等基本内容外，还会制定采集策略、原则和职责。另外，数据采集工作需要符合安全合规要求，并指定明确的数据提供方和数据接收方，以保证数据采集汇聚工作安全有序。当系统建设变更时，需要同步进行接口的规划和维护，确保数据采集工作稳定连续，保障数据质量。

编号	接口名称	详细说明	源数据周期	源数据对象	采集方式	周期内数据量	可采集频率	及时性要求	平台保存周期	源端联系人	数据提供方负责人	数据接收方负责人
1	帐单	XXX	实时	文件	增量	1G	实时	延后1小时	7天	联系人A	联系人B	联系人C

图 12 数据采集接口清单

（四）数据模型管理

数据模型是企业数据架构的核心内容，数据模型的设计、实施、管理和优化是企业数据架构主要活动内容，以确保数据需求得到有效表达、数据资源得到充分共享和利用。在实践中企业将数据模型划分为企业级数据模型和系统级数据模型，企业级数据模型对 OLTP 侧的 IT 系统开发和 OLAP 侧的数据开发工作具有指导作用。企业级数据模型一般由主题域、企业级概念模型和企业级逻辑模型构成，其中企业级逻辑模型需要对业务数据需求进行充分识别、定义和表达，关联业务流程建设和数据标准建设等工作，建设和维护成本高昂。企业级数据模型可以指导系统级模型的构建，实体和属性可以继承引用，降低系统开发成本的同时提高数据模型的规范性。

数据资源目录的建设和管理有两个重要路径，一是采用 L1-L5 的 5 层结构作为数据架构体系的基础，目录的建设和管理需要从盘点工作尽快转为企业级数据模型的建设和管理；二是将目录作为数据资源、数据资产、数据产品和服务等成果的展现方式，目录的建设和管理尽可能与数据门户、数据运营推广等工作配合推进。

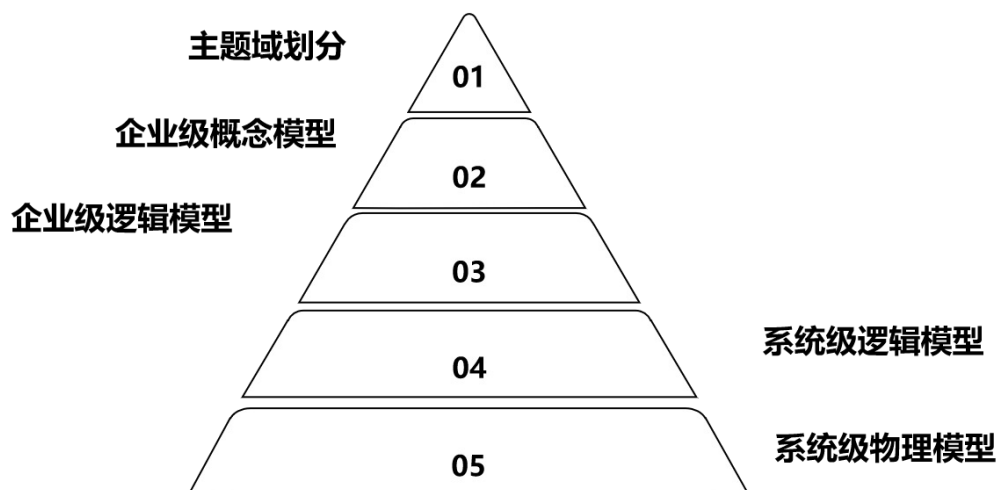


图 13 企业级数据模型分层框架

目前企业普遍意识到数据模型建设和管理的重要性，引入相关工具进行管理，但是数据模型管理的挑战依然巨大，主要挑战包括：

一是缺少专业人才，不具备自顶向下的正向建模能力，以物理数据库表逆向建模完成模型初始化工作；

二是缺少大型项目契机，无法构建全面的企业级数据模型，数据模型建设和管理工作在各开发团队中难以全面推广；

三是企业级数据模型的维护成本较高，数据模型难以得到持续的更新和维护；

四是与企业级数据模型配套的业务流程管理、数据标准管理、精细化开发管理等工作，尚未形成联动建设和管理的局面。

企业级数据模型的正向管理和应用是构筑企业核心数据资产结构、掌握数据资产主权的关键，下图展现了企业级数据模型的贯彻应用和管控路径。在应用系统或分析项目开发过程中，目标系统承接企业级数据模型的实体和属性，形成系统级的逻辑数据模型后指导物理数据模型的开发工作。企业级、领域级和系统级数据模型之间相互关联、互为支撑，为了保证整个技术流程得到有效贯彻，系统级的开发项目通常需要得到领域架构师的评审，如果是企业级的项目则需要在架构委员会进行评审。企业级数据模型管控路径的执行需要足够规模和足够能力的架构师团队，同时要设置架构例外原则以保证数据架构的灵活性和兼容性。

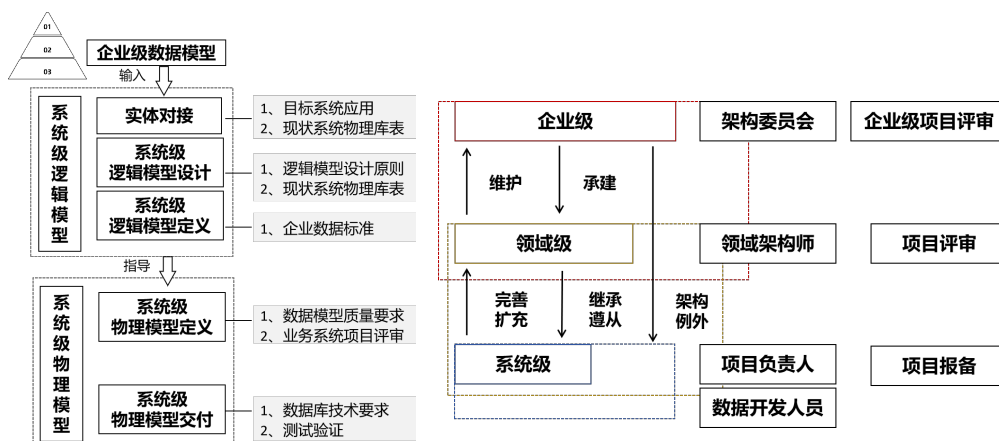


图 14 企业级数据模型应用和管控路径

（五）主数据管理

主数据管理是一系列规则、应用和技术，用以协调和管理与组织核心业务实体相关的系统记录数据。主数据管理负责企业核心业务数据记录的统一管理和分发，与业务流程和 IT 系统紧密相关，主数据具备打通全局数据架构的基本功能，可以作为数据架构逐步落地的一种务实方法。在《主数据管理实践白皮书（2.0）》中总结了主数据管理的三种模式，可以作为以核心业务对象为中心的数据架构落地模式。

在集中式应用模式中，所有主数据都通过主数据管理系统进行创建、查询和维护操作，再由主数据系统同步给业务系统。适用于组织结构相对简单，或决策权力高度集中，或对数据一致性和集中控制有较高要求时。联邦式应用模式一般由多个源头系统作为主数据权威源头，各类应用系统既可以将主数据记录传递给主数据管理系统进行统一分发，也能够直接在源头业务系统间落实主数据记录的同步和分发。适用于组织中业务分布广泛，但仍需要一定程度的数据集中管理。各业务单元保留一定程度的自治权，但通过联邦机制进行协同。分析式应用模式是在数据分析侧对主数据对象进行分析提炼，进入数据仓库和数据湖的数据中包含了主数据对象，经统一清洗、转换、集成，形成准确、一致的主数据集合，为分析型应用提供数据支撑。适用于组织结构较为分散，各业务单元有较大独立性，需要灵活管理主数据的情况。每个业务单元的数据管理自治，跨业务协同。

四、企业数据架构误区

本章聚焦企业数据架构实践过程中容易出现的误区，阐述其主要现象及不良影响，为企业提供实践策略和风险提示，提高架构相关工作成效、避免无效工作投入。

（一）过度强调技术的先进性

数据架构的规划与设计往往由技术团队主导，过度关注于采用最新、最先进的技术，而忽视了业务需求的核心驱动力。这种做法可能导致技术选型与实际应用场景脱节，收集、存储和处理的数据无法有效支撑业务决策，最终无法满足用户期望。数据架构的规划应紧密围绕业务需求展开，确保技术选型能够切实解决业务痛点，从而提升业务效率。在引入新技术时，应进行充分的业务影响分析和成本效益评估。虽然高可用性和高性能对于任何数据系统都至关重要，但将其作为数据架构设计的唯一目标是片面的。过分强调技术性能可能导致忽视其他关键因素，如数据的准确性、一致性、安全性和灵活性。这种过度追求技术指标的做法，可能在短期内提升系统性能，但从长期来看，会限制数据架构的适应性和扩展性。

（二）过度依赖 IT 架构能力

很多企业在规划和设计数据架构时，往往依赖于现有的信息化系统。企业通常会认为数据架构的设计应当与当前的信息系统保持一致，以便降低实施的复杂性和成本。然而，这种思维方式忽略了数据架构的独立性和长期性。数据架构不应只是信息系统的附属品，而应作为企业整体架构的一部分，具有更广泛的视角和战略意义。

（三）数据架构与项目开发脱节

企业的项目开发实施与数据架构活动间的脱节是一个常见的现象。在项目开发过程中，企业往往专注于实现短期的业务目标，而忽略了与数据架构的协调。这种情况下，项目开发可能在没有充分考虑数据架构设计的前提下，直接进入实施阶段，导致架构设计与实际开发的不一致。这

种脱节不仅会导致重复工作，还可能引发数据不一致、功能缺失等问题。有效的数据架构设计必须与企业在项目的各个阶段保持高度的一致性，确保数据架构与业务需求、技术实施同步推进，避免因缺乏协同而导致的效率低下或数据资源的浪费。

(四) 将湖仓等同于数据架构

企业将数据架构的建设与管理简单等同于对数据湖、数据仓库的管控工作，认为只要管理好湖仓数据，就完成了企业数据架构的工作，忽略了数据架构的全局性。实际上，数据架构不仅仅是存在于湖仓的数据存储、整合和管理任务，也需要覆盖到源头业务系统，它涵盖了数据的生命周期管理。企业需要从全局角度出发，建立覆盖数据生命周期的管理机制，确保数据架构不仅支持数据存储，还能为业务发展和战略决策提供坚实基础。

(五) 数据架构局限在 IT 部门

传统组织架构的固化思维是导致数据架构工作局限于 IT 部门的重要原因之一。在这种架构下，IT 部门往往被视为技术支持的提供者，而非企业战略决策的参与者。技术与业务之间的鸿沟也是不可忽视的因素。随着技术的快速发展，IT 部门逐渐形成了自己的专业壁垒，与业务部门之间的沟通变得困难，导致数据架构的设计与实施难以充分融入业务需求。同时，企业内部缺乏有效的跨部门协作机制，使得 IT 部门在数据架构工作中难以获得其他部门的支持和配合。

(六) 数据架构不是必选项

很多企业错误地认为数据架构是一个可选项，尤其是在数据管理需求尚未显现时。这种错误的观点显然低估了数据架构在支持企业战略决策和业务扩展中的重要作用。数据架构不仅仅是技术实施的一部分，它直接影响企业的业务全流程和数据治理能力。忽视数据架构的战略地位的问题可能在初期不明显，但随着企业数据量的增长和业务的复杂化，缺乏有效数据架构的支持将导致数据孤立、数据质量下降、决策不准确等问题的产生。

五、企业数据架构案例

本章对民航业、银行业、保险业、装备制造业的企业数据架构实践案例进行了汇编，通过企业案例更加生动、具象地展现企业数据架构全貌，为企业开展数据架构工作提供参考。

（一）东方航空

1. 信息化“硬仗”推动架构落地

自 2010 年起，中国东方航空股份有限公司（以下简称“东航”）每年的战略解码都有一场信息化硬仗，树立了“IT 引领”的发展理念。在 2014 年，东航在信息部成立的企业架构团队，引入基于 TOGAF 的企业架构建模方法，自上而下进行顶层设计。2017 年，东航设立数据职能部门，构建数据管理组织体系、管理制度及流程，有力促进了跨领域信息的交互和共享。经过持续迭代优化，企业架构逐步成为连接东航业务发展战略与信息化项目的关键指导工具，并形成了以企业运营流程为中心，运用信息技术固化和优化业务流程的工作模式。

2. 企业级数据模型的体系架构

东航数据架构管理以企业信息化战略为核心，遵循“业务驱动，统一目录、统一模型、规范分布”的原则，促进数据与业务的深度融合，在跨业务、跨部门、跨系统数据拉通建立基础。根据东航信息化战略顶层设计及推进工作，参照 TOGAF 企业架构框架，以业务架构为输入，持续开展数据架构相关活动。东航数据治理方法以数据架构和数据模型为核心，从自上而下、自下而上、由外而内、由内而外四个视角开展企业级数据模型设计。

一是自上而下构建企业级分层数据架构。东航采用行业企业架构方法，在业务流程的基础上，通过识别各流程节点的输入和输出信息，形成数据实体。与业务流程分级相同，数据架构也分层进行组织，包含主题模型、概念模型、详细模型和系统模型。其中，前三层为企业级数据模型。企业级数据模型作为系统模型的设计参考，系统模型应遵从企业模型。

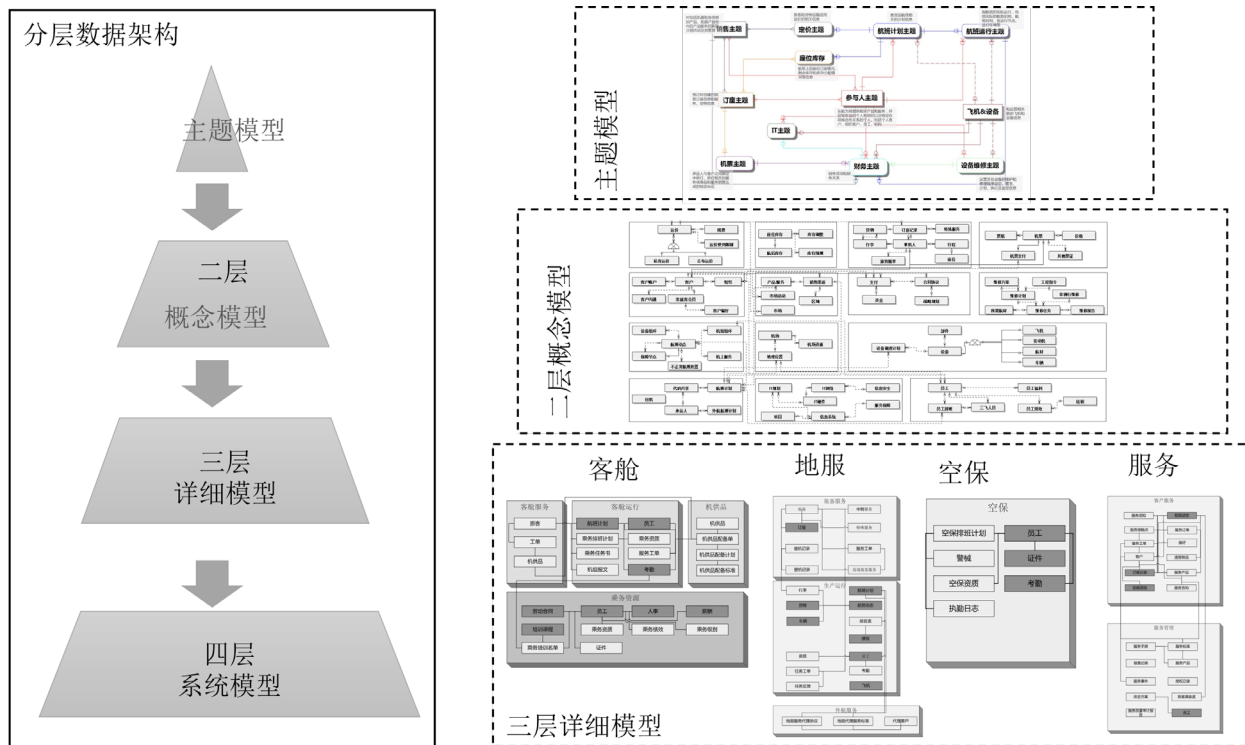


图 15 东航数据架构体系

二是自下而上统筹数据架构接纳信息化建设成果。东航信息化建设历时十多年，建设了 400+ 个应用系统。企业级数据模型设计必须通过现状调研，在充分掌握各应用系统承载的业务流程、运行中的数据库模型和接口定义等信息之后，提炼总结成为跨越系统和业务领域的企业级数据模型。

三是由外向内遵从民航业高标准数据架构要求。航空运输对技术要求高、设备操作复杂，需要机场、空管、航司等多家单位互相协同。为此，IATA（国际航空运输协会）和 ICAO（国际民用航空组织）等国际组织制定了业务程序和标准，帮助各国管理机构更好地协同。例如报文规范、NDC 标准等。国内民航局也发布了《中国民航航班运行数据开放资源目录》和《智慧机场数据规范与交互技术指南》，定义了各单位间的数据传输格式，对各单位制定数据标准、设计数据模型有很好的借鉴作用。东航企业数据模型还借鉴了 TOGAF 发布的《商业航空公司参考模型》、Oracle 航空数据模型参考、Teradata 旅游行业数据模型参考等材料。

四是由内向外保障产业链条数据兼容一致。为了更好地保障航班、服务旅客，东航要与民航局、机场、空管等多家单位协同，通过数据共享，建立完整的航班保障链和旅客服务链。受限于各系统建模差异，外部报文数据缺失航班 OneID 和旅客 OneID，东航设计了用于公司内外部共享的独立数据模型，通过数据标准从根本上保证系统间数据的兼容性和一致性。

3. 企业数据架构建设历程

数据架构作为业务与应用系统建设的桥梁，数据架构基于业务架构（业务模式、流程、规则等）识别出业务数据需求，统一数据语言及操作手段，提供应用架构（系统功能、组件、接口等）和技术架构（技术指标、技术选型等）设计和开发的依据。东航数据模型发展可概括为三个阶段：

（1）系统应用实现阶段（2010—2016 年）

在此阶段，数据模型处于“无意识”管理状态。针对公司信息系统数据模型呈现“各自为战”阶段，数据分散存储于各类信息系统中，原始系统中的数据格式、数据模型、数据标准均尚未统一。

（2）企业模型实践阶段（2016—2018 年）

依托业务模型建立企业级数据模型，指导作用信息系统建设。东航的企业级数据模型参考国内、外行业级组织发布的航空运输行业数据模型，并结合东航业务系统的业务特性进行构建。根据业务属性，东航将业务数据规划为 14 个主题数据域模型和主题域概念模型，分别是定价、座位库存、订座、销售、航班计划、客户和参与人、维修、飞机和设备、航班运行、机票、财务、IT、员工和位置。

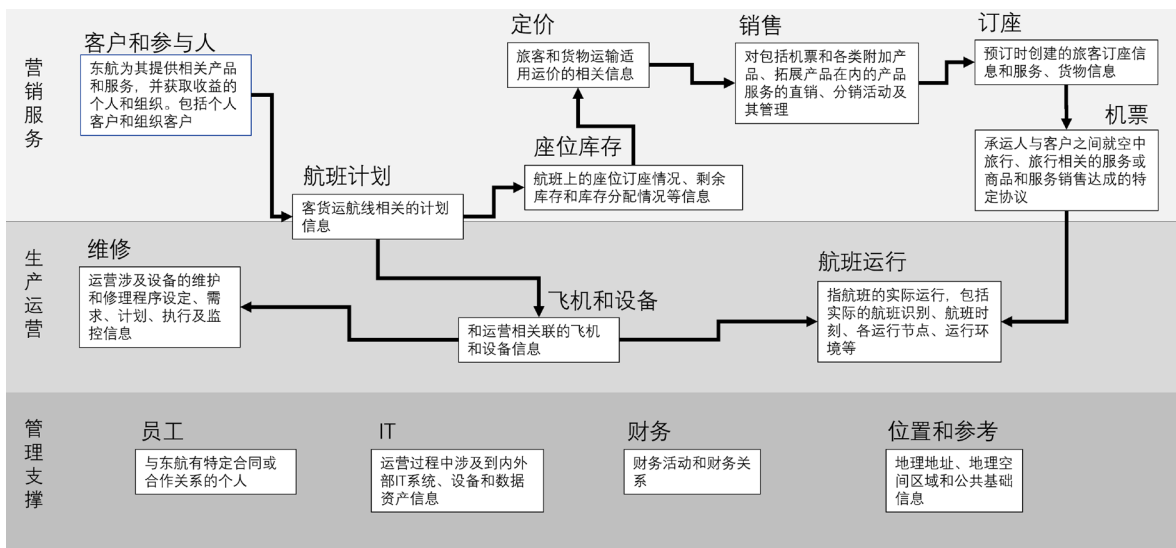


图 16 东航数据模型主题域划分

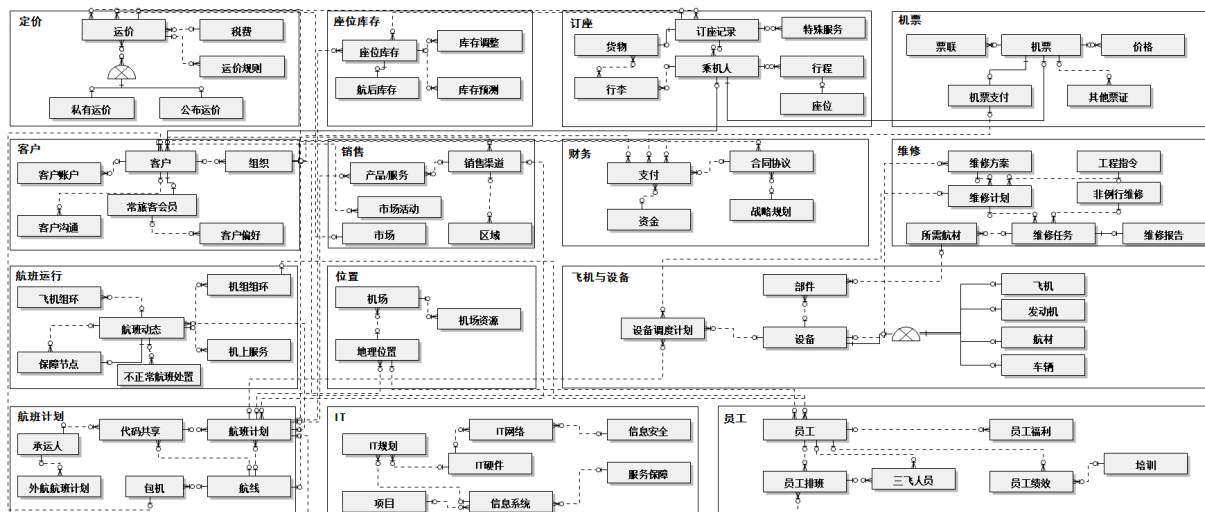


图 17 东航主题概念数据模型

（3）模型集中管理阶段（2018 年至今）

通过引入统一建模工具，实现数据资产管理闭环，形成完整数据地图及多套数据目录体系。

东航已实现系统数据模型事前统一、事中评审、事后监控的全流程管理。事前统一即通过统一的建模工具，进行模型设计。事中评审即增加模型评审环节，架构师负责评审，模型工具配合模型规范性检查。事后监控即数据模型部署生产环境后，通过模型工具捕捉模型差异，监控模型的一致性、完整性。在系统模型的管控过程中，了解业务发展，及时分析数据目录差异，同步更新企业数据模型。系统模型应遵从企业级数据模型。

东航通过项目立项环节识别主数据、审查一致性，保证主数据建模源头一致；通过系统模型事中评审和事后监控，验证主数据建模的一致性，检查数据建模的遵从。同时，东航持续开展相关主题数据域的业务梳理，组织编制相关数据的数据标准，形成相关领域的概念模型和子领域的逻辑模型。

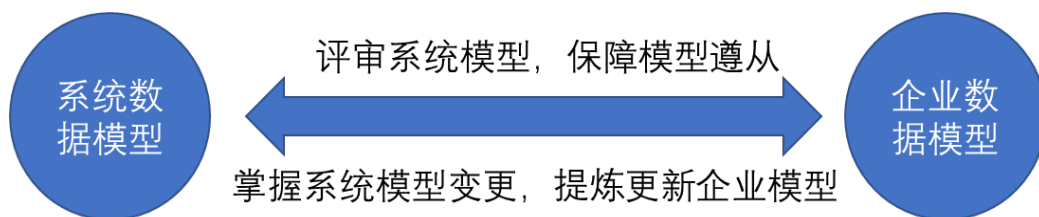


图 18 东航数据模型管控流程

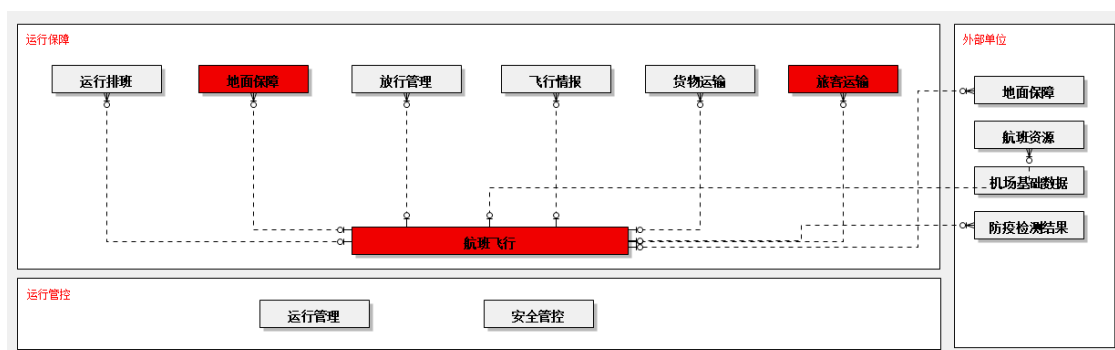


图 19 东航概念模型—航班运行领域

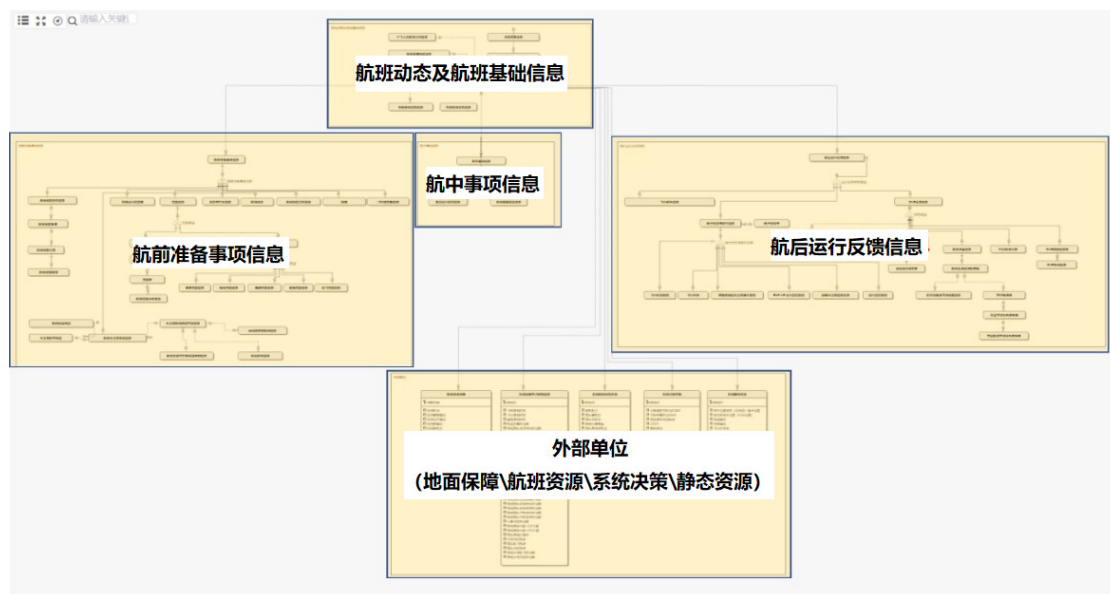


图 20 东航航班飞行领域逻辑模型

4. 企业架构促进业务创新发展

东航数据架构和发展蓝图日趋完善，已形成统一的数据资产管理体系与框架，统筹对接数据类建设、共享需求，打破数据壁垒，减少数据理解成本，避免重复投资，优化数据治理，推动业务与 IT 深度融合，以快速响应业务需求，促进业务发展和持续创新变革。

(二) 平安人寿

1. 以数据架构探索数字化改革

中国平安人寿保险股份有限公司（以下简称“平安人寿”）作为保险行业的引领者和开拓者，随着内外部环境的变化，面临着客户需求升级、人口红利变化、老龄化趋势加速的机遇与挑战，需要通过数字化改革构建新的增长曲线，而数据是数字化改革的重要引擎。平安人寿作为数字原生企业高度重视数字化对企业发展的赋能作用，从 2019 年起在行业内率先探索改革转型，打造数字寿险新模式，并以数据架构全面升级支持公司数字化转型战略。

2. 企业数据架构建设思路

平安人寿的数据架构建设是在新老系统并存、系统间数据架构交叉的背景下，面对日益丰富的数据使用场景和快速增长的数据量，亟需升级企业数据架构摆脱烟囱式的建设模式。以下从数据分析侧和应用系统侧两个方面介绍下平安人寿数据架构的建设思路。

（一）数据分析侧数据架构建设思路

平安人寿在数据分析侧摒弃了烟囱式的数据集市模式，构建统一的数据底座，建设了全新的平安人寿数据仓库。数据底座中形成了统一的公共维度模型层，并构建了覆盖从设计评审到效果评价全过程的流程、规范，包括主题域管理、总线矩阵设计、维度表设计、词根管理、模型发布等。



图 21 平安人寿数据底座全景

在统一的公共维度模型之上，为了快速高效产出各类指标，平安人寿建设原子和派生的指标设计模型和ADS层，然后基于设计的原子指标，在不同的维度进行抽象生成派生指标并进行输出，提升数据一致性。

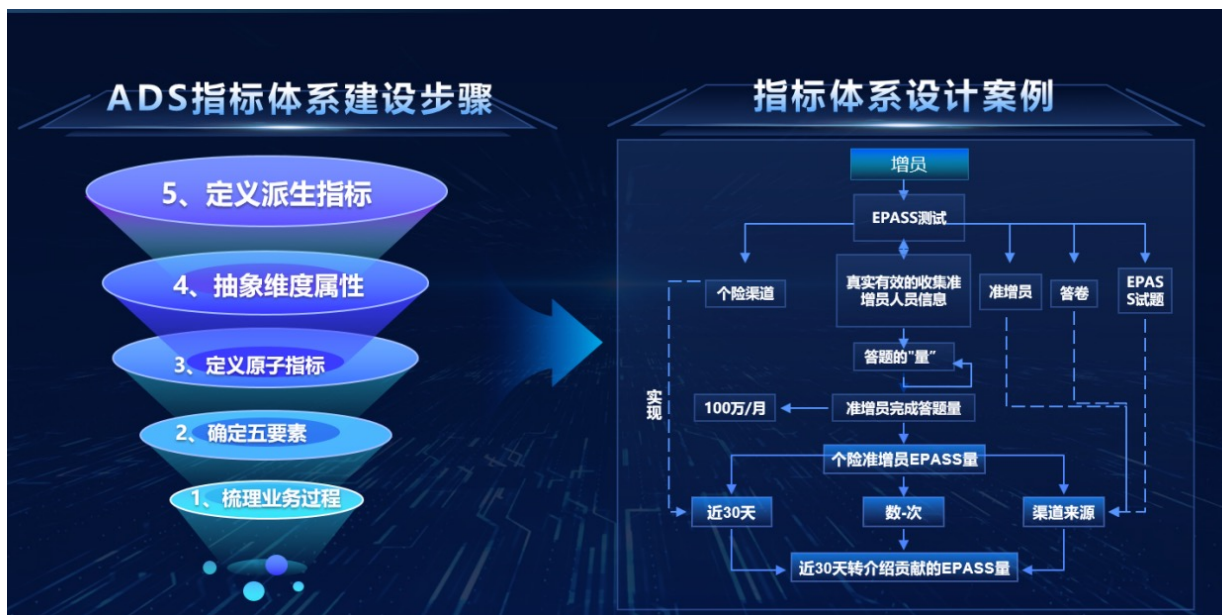


图 22 平安人寿数据底座 ADS 建设

（二）应用系统侧数据架构建设思路

平安人寿以 TOGAF 为指导构建完善的企业架构管理体系，数据架构作为其中的核心环节。通过建立架构基本法，明确数据架构的交付物和标准，对存量系统按架构基本法要求分阶段进行架构资产盘点梳理，对增量系统要求按照架构基本法规范进行架构设计，统一通过架构设计平台进行线上化管理。

通过设计企业架构元模型，明确各元素定义和关系，指导设计建模，确保数据架构与应用架构、技术架构、业务架构对齐一致，快速响应业务变化，高效支持业务目标达成。以企业架构元模型为基础，构建企业级数据模型，根据业务架构全景划分主题域，依据业务流程设计，抽象识别业务对象并形成概念模型。在概念模型基础上，根据业务场景添加关键属性，对业务活动、业务规则进行更加明确的定义和表示，形成逻辑模型。

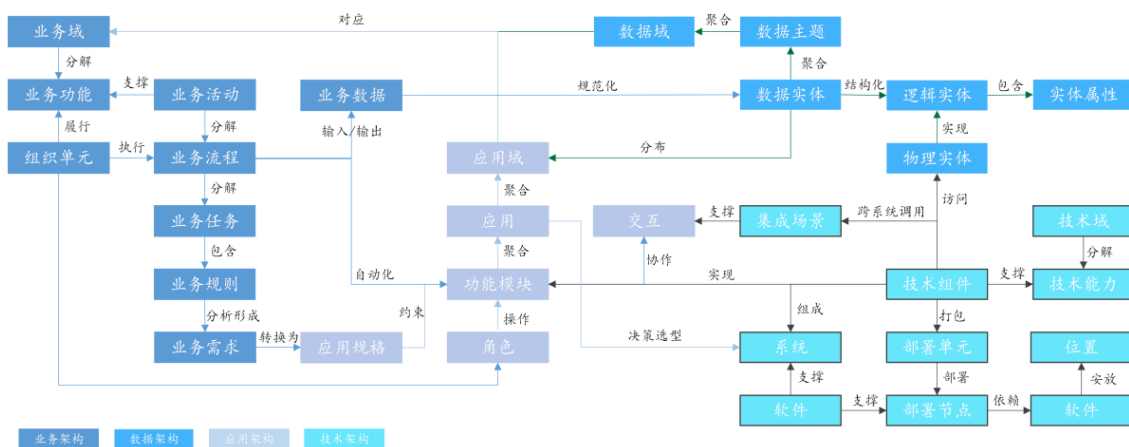


图 23 平安人寿企业架构元模型

3. 企业数据架构建设历程

平安人寿的数据体系在过去的 20 多年时间里，经历了从传统企业数据仓库到大数据平台再到数据主题集市的发展升级。从 08 年开始，采用 IBM Cognos 数仓系统，基于 Oracle 和小型机，构建企业级数据仓库；从 16 年开始，基于业务需求驱动，基于 Hadoop 平台搭建了面向不同渠道、不同业务的独立主题集市；从 20 年开始，公司通过三年的时间，实现数据中台化；最后从数据架构升级到企业架构，全面支持数字化转型。平安人寿数据架构的发展历程总体分为五个阶段。

第一阶段以报表驱动，按需开发数据报表对业务系统进行集中的数据处理。

第二阶段是在公司积累了大量报表之后，通过对这些报表按照业务主题驱动，来响应不同业务团队的需求，分领域建设平安人寿的“数据集市”。

第三阶段则是随着平安人寿数据体量越来越大，从 2016 年开始，公司根据行业的开放体系引入了大数据平台，这个阶段主要是利用大数据的存储和计算能力来处理很多比较大的数据表，提升报表的实效，但仍是报表驱动的模式。

第四阶段是从 2020 年平安人寿设计数据中台开始，公司在数据中台的建设上引入了 DAMA 的数据支持能力体系，并把数据中台按照数据治理、数据底座、数据产品三个层面进行建设，同时在技术架构上也引入了湖仓一体的体系架构。

第五阶段则是由数据架构升级到企业架构，在企业架构框架指导下建设统一架构管理体系，提升架构成熟度，全面支持业务战略需求。采用自上而下设计，建立企业级数据模型，实现从主题域、概念模型、逻辑模型、物理模型不同阶段数据架构资产的全流程管理。

4. 企业数据架构价值成效

平安人寿通过构建企业数据架构，全面支持数字化转型，在数据管理、业务支持、风险控制、经济效益方面都取得了显著成绩。在数据管理方面，平安人寿提升了数据管理能力水平，并在 2023 年获得了 DCMM 量化管理级认证；在业务支持方面，帮助识别业务流程中的瓶颈和低效环节，从而推动流程优化；通过数据模型优化，数据需求处理时效提升 50%；在风险控制方面，基于建立的数据架构，实现统一的数据安全和合规管理，降低数据风险；在经济效益方面，通过整合数据资源，推动架构资产复用，避免系统重复建设，降低 IT 运维成本，实现降本增效。

(三) 交通银行

1. 数据架构升级助力数据治理

近年来，商业银行在数据治理与应用领域不断探索前行，在制度体系构建、组织架构保障、系统平台建设等方面均取得了阶段性成果，初步具备了数据管理基础能力。然而，随着数字化转型的深入和业务发展的需求，对数据的灵活性和时效性提出了更高要求，监管科技的升级也使得数字化监管成为新常态，这既带来了新机遇，也对商业银行提出了新挑战。一是架构遵从执行难，虽然已建立数据治理组织架构，但缺乏各层级间的共识，业务人员对数据治理的重要性认识不足，导致其推进效果有限。二是业数融合难，面对快速发展的业务需求，数据开发仍依赖传统模式，缺乏敏捷性和自助分析能力。三是存量系统改造难，银行存量系统庞大且复杂，跨部门流程繁琐，使得数据治理改造困难。为进一步深化数据治理，发挥数据价值，交通银行通过升级数据架构体系，推进数据与业务紧密融合，加强数据治理制度体系、组织架构和系统平台建设，实现数据规范化管理、自助化分析、智慧化服务和数据数字化等能力提升，引入大模型、人工智能等先进工具进一步提升数据架构管控水平，相关工作已取得阶段性成果。

2. 全面推进企业数据架构建设

交通银行建立了董监高层级的数据治理决策领导体系，制定了全行数据治理规划，并将其作为全行数据工作的战略性指导文件，提出“规范化治理、智慧化服务、平台化支撑、协同化组织”的体系建设目标，明确业务、数据、技术三方职责分工，深化全行数字化发展理念共识，总分协同、全面推进企业级数据架构建设。数据架构包含企业级数据模型、数据标准、数据分布和数据流转四个领域。

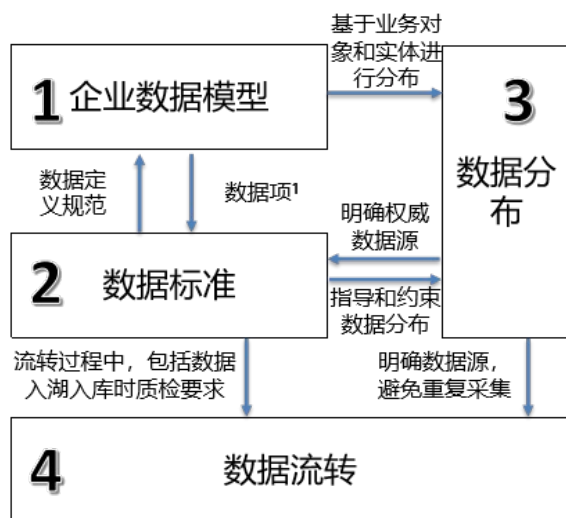


图 24 交通银行数据架构领域

(1) 企业级数据模型

交通银行建立了由各业务部门专家参与的专项团队，全面梳理应用架构现状和产品分类，结合各业务领域现状，同时分析借鉴业界已有的 IBM 的 FSDM 及 Teradata 的 FSLDM 等参考架构和方法、证券期货行业数据模型，采用双“T”架构构建企业级数据模型，覆盖组织业务经营管理和决策数据需求，通过“自顶向下、逐步求精、弱范式化、松散化”的设计原则，制定了面向全行的架构管控体系，指导企业级数据模型的开发和管理，依托数据建模工具，实现数据模型标准化管控从数据架构到应用架构领域的拓展和落地实践。

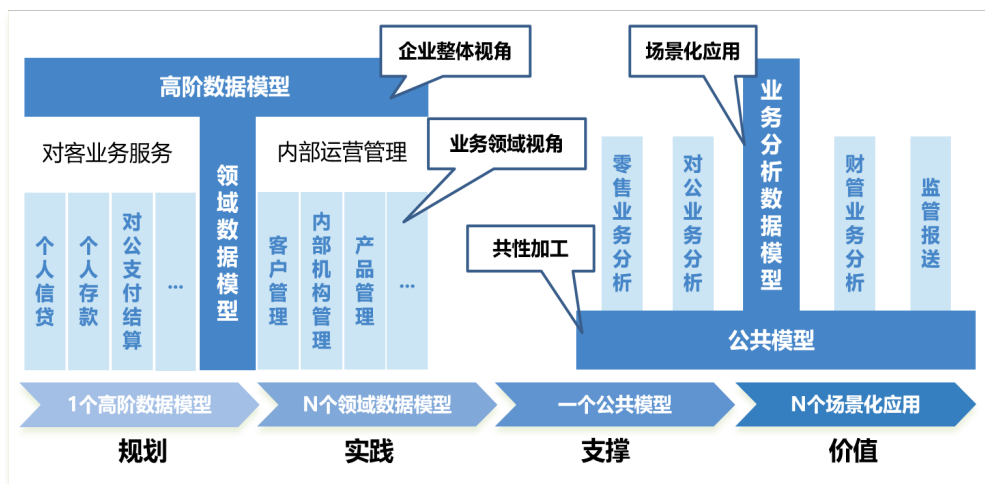


图 25 双“T”数据模型

（2）数据标准

数据标准是实现企业数据一致性的关键，要求企业各部门使用统一化、标准化的语言描述数据。交通银行基于已编制的覆盖零售、公司等各业务领域的企业级数据字典，一方面，推进“治理前移”，将数据标准的管理工作前置到架构评审、需求评审等环节，以制度和流程保证“基础数据跨系统一致、指标数据跨报表一致”，沉淀全行统一的指标库与标签库；另一方面，加强“工具管控”，搭建数据建模管控工具，在开发流程中实现数据落标的“强管控”，确保跨系统之间数据开发的一致性。

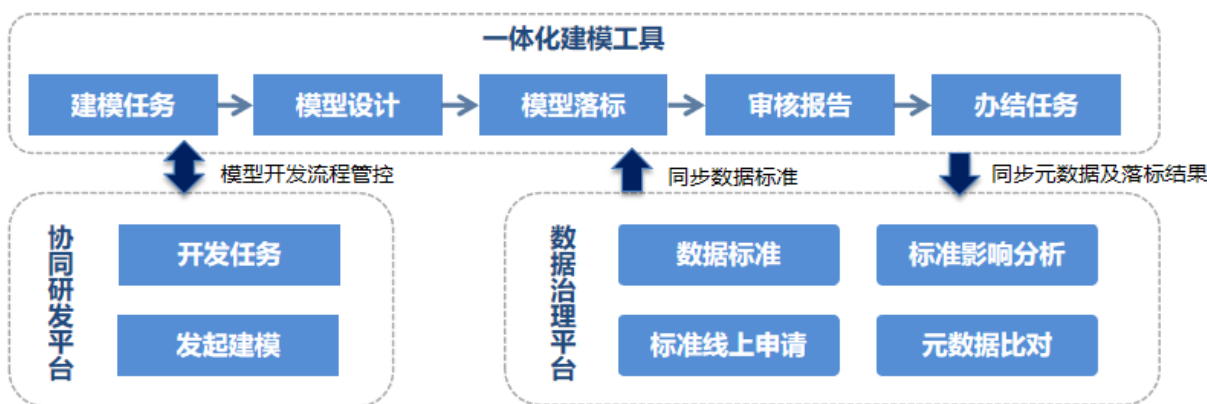


图 26 数据标准管控流程

（3）数据分布

数据分布是针对企业级数据模型中数据的定义，明确数据在系统、组织和流程等方面的分布关系，定义数据类型，明确权威数据源，为数据相关工作提供参考和规范。交通银行通过结合企业级数据模型，以主数据为抓手进行数据分布的管理，形成主数据流向及分布清单，同时将数据资产的业务场景与业务部门的实际业务诉求进行关联，从而实现了业务价值导向的数据分布管理方案。通过搭建治理平台和智能应用运维平台，集中管理数据的发布、使用及审批流程，实现对数据分布关系的线上化管理和监控，逐步提升数据分布的自动维护能力和智能化管理水平。

（4）数据流转

数据流转是建立起组织内各应用系统、各部门之间的集成共享机制，通过组织内部数据集成共享相关制度、标准、技术等方面的管理，促进组织内部数据的互联互通。交通银行通过建立企

业级的数据集成共享规范，明确全行数据资产主人是全行数据资产的所有者和最高决策者，明确了数据集成和共享平台的管理机制和使用步骤，并通过企业级数据中台支持企业内的数据集成和共享，实现组织内多种类型数据的整合，形成一个统一的数据视图，为业务分析和决策提供支持。

3. 轻量智能开放的企业数据架构

交通银行数据架构体系日趋完善，数据管理的规范化水平持续提升，数据治理要求在全行得到有效衔接和落实。交通银行自主研发的双“T”数据模型架构，具备“轻量化”“智能化”“开放化”等特性，提升了建模工作的效率和模型成果的准确度；编制了覆盖各业务领域的企业级字典规范，将标准化流程嵌入开发过程，推进业务系统全面贯标，实现从需求到投产的全生命周期的数据管控，提升了数据质量；自动维护数据分布关系，量化数据相关业务价值；企业级数据中台基本建成并有力支持前台业务发展，内外部数据融合应用不断深化并持续赋能业务经营和风险防控。

（四）中车浦镇

1. 以数据架构支撑智能工厂

中车南京浦镇车辆有限公司（以下简称“浦镇公司”）是从事轨道交通装备研究和制造的专业化生产企业，是中国铁路装备制造业大型一档企业、中国铁路空调双层客车研制基地、中国城市轨道交通车辆生产定点企业。2020年，浦镇公司启动了以精益数字管理体系和“三大驾驶舱”为指导的“城市轨道交通装备智能工厂”建设项目，支撑“城市轨道交通装备智能工厂”建设。

随着数字化转型的不断深入，浦镇公司建设的信息系统日益增多，为顺利推进“城市轨道交通装备智能工厂”建设工作，浦镇公司以数据架构建设推动数据的生产者、使用者、管理者及数据标准的制定者之间的有效协同，对齐数据管理维度和颗粒度。

2. 问题导向的数据架构建设

浦镇公司数据架构建设工作以解决数据管理问题为导向，确立了数据资产目录、数据标准、数据模型和数据分布为大数据架构组件。在“产品全生命周期数据驾驶舱”建设过程中，碰到找不到数、取不了数等问题，选择“数据资产目录”作为数据架构组件进行支撑；在企业主价值链贯通过程中碰到的数据孤岛问题，选择“数据标准”作为数据架构组件进行支撑；在信息化项

目规划与设计阶段，为实现业务贯通和数据贯通开展的数据集成规划，选择“数据模型”作为数据架构组件进行支撑；在“经营管理驾驶舱”“生产指挥驾驶舱”建设过程中，碰到的数据源头难以确认等问题，选择“数据分布”作为数据架构组件进行支撑。

（1）数据资产目录

数据资产目录作为数据架构的基础，必须能够承接浦镇公司各业务环节的使用需求和报告分析最小粒度的要求，并支撑数据标准、数据模型和数据分布。每个数据资产都应定义唯一的数据责任人，并与业务责任人保持一致，负责数据资产的定义、管理和共享授权。

（2）数据标准

数据标准是浦镇公司业务和管理活动中所涉及数据的规范化定义和统一解释。数据标准的制定必须全面考虑跨业务、跨流程的应用场景，确保对所有关键数据进行准确识别、科学分类、明确定义及标准化处理，以有效支撑应用系统开发、数据管理及数据分析。数据的定义和规则在公司范围内必须具有唯一性。

（3）数据模型

数据模型是浦镇公司数据架构的核心，是数据贯通规划的重要技术手段，通过数据模型，连接业务架构与应用架构，用于对公司业务模式和业务规则的数据需求进行分析和重组，以支撑应用系统设计和开发。

（4）数据分布

在浦镇公司，数据分布主要用于数据溯源工作。通过对核心业务流程中的业务对象及信息系统中的关键数据进行深入分析与梳理，明确关键数据的源头部门、工作步骤及所属信息系统的模块与功能，为数据集成与数据应用提供了便捷的取数路径。

3. 企业数据架构建设历程

随着数据应用和数据质量之间的矛盾不断凸显，浦镇公司于 2022 年开展了以数据质量全面提升为目标的数据架构建设工作，整个数据架构建设工作主要分为三个阶段。

（1）夯实基础，全面提升数据质量（2021—2022 年）

随着浦镇公司投资建设了大量的业务系统，数据集成、数据贯通问题日益突出，同时，业务和管理都衍生出大规模数据应用的需求。故而，在本阶段，浦镇公司主要开展了三项工作，分别是盘点数据、制定标准和提升质量。

盘点数据是指在公司范围内，全面开展线上数据库和线下表单的盘点，并以此为基础，编制数据资产目录。

制定标准是指依据数据资产目录提供的数据清单，对其中的主数据、指标数据、字典数据以及关键的业务数据制定数据标准。

提升质量是指基于数据资产目录和数据标准，对相关数据进行质量稽核，对质量问题较大的数据开展数据质量提升行动。数据质量提升行动包括数据清洗、对系统录入增加数据标准校验、统一数据维度、颗粒度等。

质量规则类型	质量维度	作用规则	字段	数据类型	质量规则说明	检查时间	样本数量	问题数量	问题占比率	问题说明
格式规范	有效性	单列	行驶里程	数值型	数值型数据是表示数量、可以进行数值运算的数据类型；数值型数据由数字、小数点、正负号和表示乘幂的字母 E 组成	2022-07-31	30845	1283	4.16%	1、顿号：88项； 2、“.”122项； 3、“/”301项； 4、“0”开头并含有有效数字3项； 5、含“.”2项； 6、含中文字符和英文字母651项； 7、“\”116项
格式规范	有效性	单列	故障响应时间	数值型	数值型数据是表示数量、可以进行数值运算的数据类型；数值型数据由数字、小数点、正负号和表示乘幂的字母 E 组成	2022-07-31	30845	140	0.45%	1、空白136项； 2、含字母和汉字3项； 3、含“.”1项
格式规范	有效性	单列	发生日期	日期时间型	通过 YYYYMMDD hh:mm:ss.SSS 的形式表达的类型 GB/T 7408-2005 中的日期完全表示法，“YYYY”表示年份，“MM”表示月份，“DD”表示日期，“hh”表示小时，“mm”表示分钟，“ss”表示秒，“SSS”表示毫秒。 本字段要求为精确到秒，为YYYY-MM-DDThh:mm:ss	2022-08-01	30845	19350	62.73%	1、YYYY-MM-DDThh:mm:ss有11494项； 2、YYYY-MM-DDThh:mm有19350项；
格式规范	有效性	单列	报告日期	日期型	通过 YYYYMMDD 的形式表达的值的类型 GB/T 7408-2005 中的日期完全表示法，“YYYY”表示年份，“MM”表示月份，“DD”表示日期，本字段为：YYYY-MM-DD	2022-08-01	30845	0	0.00%	
长度约束	有效性	单列	故障描述	字符型	以事实为依据设置输入长度限制200个字节。	2022-08-01	30845	1229	3.98%	当前长度： <100,2988项； 100~200,26661项； >200,1229项
长度约束	有效性	单列	初步原因分析	字符型	以事实为依据设置输入长度限制200个字节。	2022-08-01	30845	226	0.73%	当前长度： <100,29924项； 100~200,695项； >200,226项

图 27 浦镇公司数据质量稽核示例

（2）联接共享，数据全价值链贯通（2022—2023 年）

在数据质量得到初步保障并持续提升的基础上，本阶段以数据全价值链贯通为目标，开展了四项工作，分别是固化业务流程、设计业务对象、确定数据模型和梳理数据分布。

固化业务流程是指浦镇公司针对现有流程不完整、流程颗粒度不细、流程说明缺失、流程与业务现状不符等现象，以原有流程文件为基础，会同相关责任部门，重新梳理了核心价值链，完善并细化了流程体系，使其能够支撑数据架构建设工作的开展。

设计业务对象是指浦镇公司对于表达业务关系，承载业务运作和管理涉及的重要信息的人、事、物的对象进行设计，用于统一业务领域的重要业务概念。浦镇公司明确了通过业务流程识别业务对象的方法与步骤，并根据有代表性的业务场景编制了相关案例，利用确定性的方法与步骤，辅以滚动培训，让业务对象的识别成为标准化的行动。

确定数据模型是指基于业务流程、业务对象建设成果，对业务流程及流程涉及的表、证、单、书开展分析，梳理业务对象涉及的业务活动和业务交互，并以此为基础，开展了关键业务领域的概念模型、逻辑模型设计。

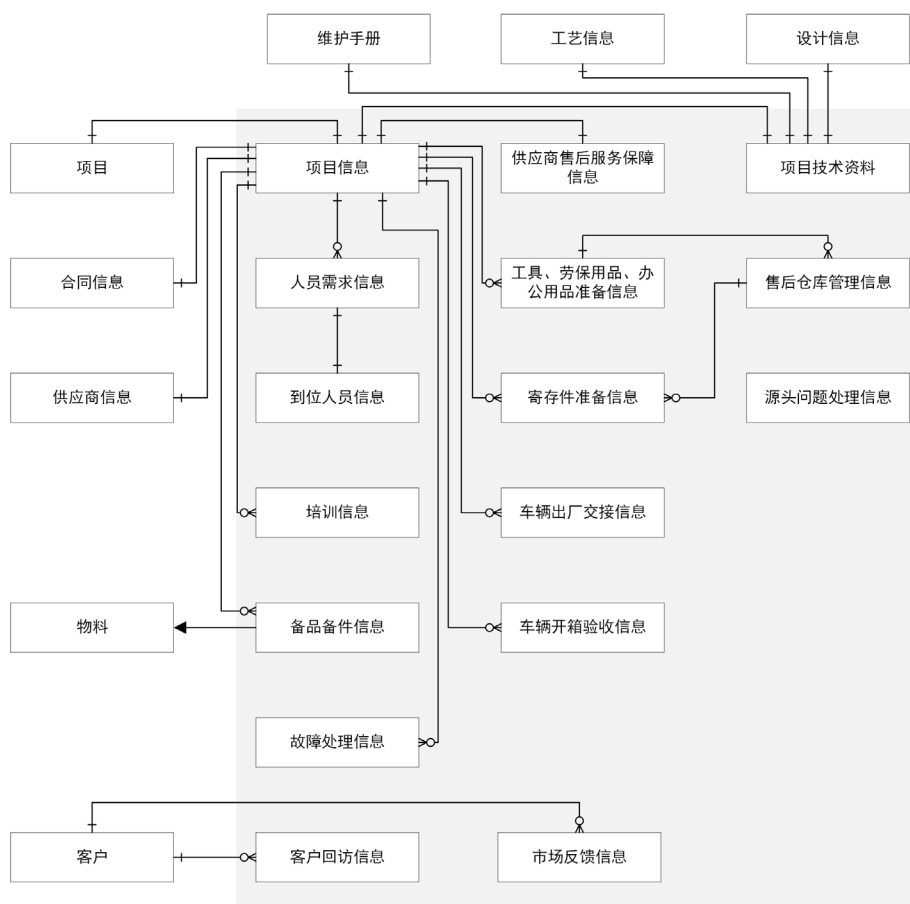


图 28 浦镇公司概念模型 - 售后管理域

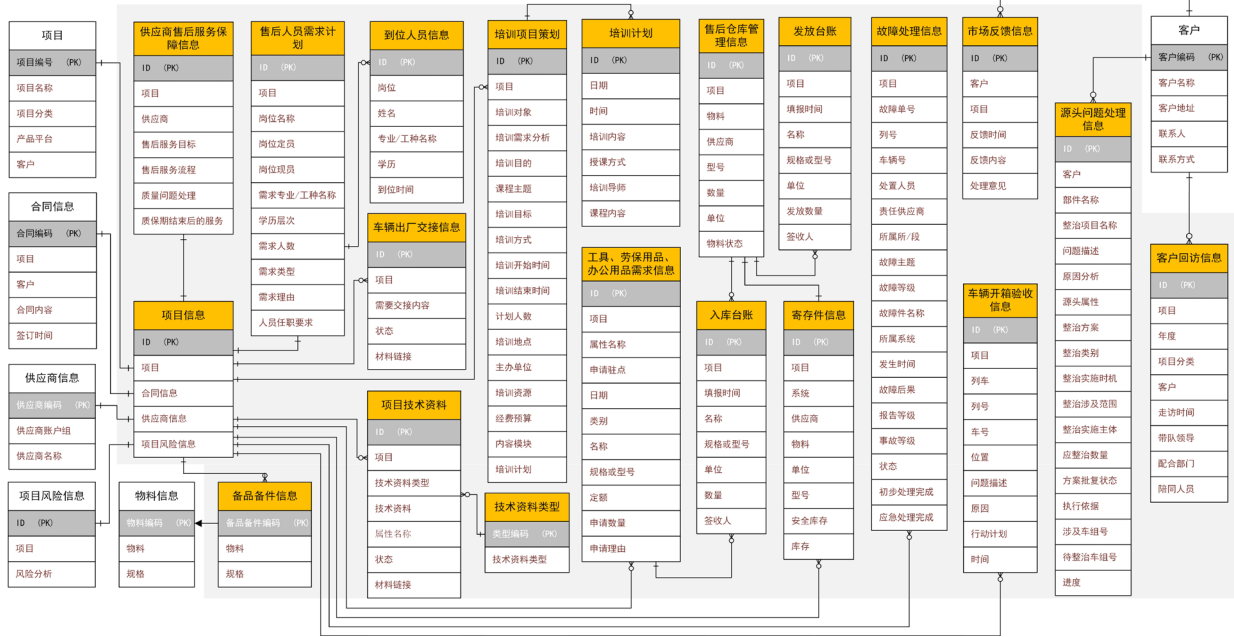


图 29 浦镇公司逻辑模型示例 - 售后管理域

梳理数据分布是指对业务流程涉及的表、证、单、书以及信息系统中的关键数据进行分析与梳理，确定产生关键数据的源头部门、工作步骤以及相关信息系统的模块和功能，梳理关键数据在各业务流程及信息系统间的流动情况，形成公司级的数据分布视图。

四级流程	流程步骤	项目编码	项目名称	人员
1.1.1.1 市场开拓	收集市场信息	R	R	
1.1.1.1 市场开拓	了解竞争态势	R	R	
1.1.1.2 投标启动	招标文件发布	R	R	R 投标人员 编制人 审核人 批准人
1.1.1.2 投标启动	投标启动会议	R	R	R 部门分管领导 部门项目负责人
1.1.1.2 投标启动	投标启动会议	R	R	R 负责人
1.1.1.2 投标启动	投标启动	R	R	
1.1.1.2 投标启动	投标启动	R	R	
1.1.1.2 投标启动	投标启动	R	R	R 编写人
1.1.1.2 投标启动	投标启动	R	R	
1.1.1.2 投标启动	投标启动	R	R	

图 30 浦镇公司数据分布视图示例

（3）数据驱动，无障碍的数据应用（2023—2024 年）

本阶段是基于前两个阶段的工作成果，全面开展数据驱动型企业建设。数据驱动型企业建设包括业务系统的全面贯通和自助式、服务化的大规模数据应用。

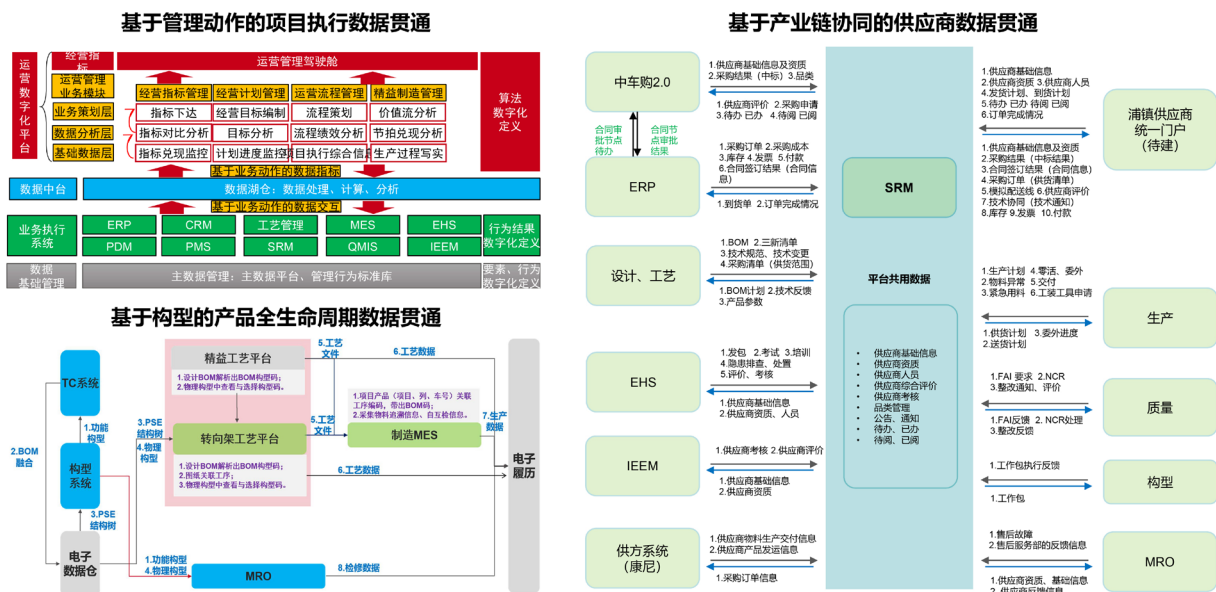


图 31 浦镇公司数据应用

4. 数据架构推动自助式服务

通过构建数据架构，浦镇公司已建立起较为完善的数据管控技术体系，该体系有效支撑了业务架构和应用架构的建设。浦镇公司的数据架构不仅有助于打破业务孤岛和数据孤岛，实现业务数据的全面互联互通，还为增强数据安全、建设数据湖仓及数据治理平台提供了坚实基础。此外，数据架构还为浦镇公司将现有的“保姆式”数据服务能力升级为“自助式”数据分发与共享能力提供了有力的技术保障。

六、企业数据架构展望

未来，企业数据架构将得到更多关注和投入，常态化、一体化和多样化的发展趋势将推动企业数据架构在各行业实践落地，每个企业都可以找到适合自身发展的数据架构管理模式。

常态化数据架构管理。企业数据资源规模和应用场景不断扩大，企业数据架构扮演更为关键的角色，为确保企业数据架构体系得到持续建设和维护，数据架构管理工作将作为一项常态化工作开展，与数据架构配套的管控评审机制、架构师团队、架构工具平台等工作将日趋完善。

一体化数据架构建设。企业数据架构的建设将摆脱孤立局面，一方面将与业务流程、系统迭代、数字化创新等工作协同建设推进，将数据架构融入到各领域中；另一方面企业数据架构将与数据安全、数据质量、数据开发等领域融合，形成一体化的数据治理体系。

多样化数据架构体系。企业数据架构受到业务架构和公司治理体系的影响，企业将根据组织架构、业务模式和 IT 系统等因素形成多样化的数据架构体系，制定适合自身发展情况的数据架构管理组件和数据架构管理模式。



地 址：北京市海淀区花园北路52号

邮 编：100191

邮 箱：kanxinyu@caict.ac.cn

网 址：www.tc601.com