



西安电子科技大学
XIDIAN UNIVERSITY

第四章 主数据管理

马 晶

经济与管理学院 信息管理系

Email: majing@xidian.edu.cn

系统孤立带来的苦恼

- 政务系统中，很多部门或者子系统都存有本地居民的个人信息，这些个人信息对于相关业务的政策运转是非常重要的。
- 理想情况下，这些个人信息在不同的子系统应该都是完整、一致且不过时的。



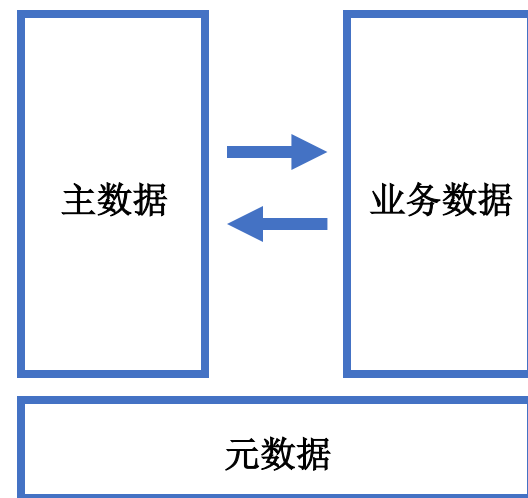
- 主数据与上一章的元数据是不同的两个概念。
- 元数据是“描述数据的数据”，而主数据则指实体数据。
- 同一个系统可能在不同的企业或组织中被多次部署，在不考虑个性化定制的情况下，每套系统的元数据都基本是一样的，但不同企业或组织中维护的主数据可能各不相同。
- 主数据管理的重点在于如何将不同职能部门中**共享**的那些重要的**实体数据**统一地管理起来，进而提高数据质量并发现重要的业务洞察。

课前案例

以小组为单位阅读并讨论案例“某电线电缆集团的主数据管理实践”，回答以下问题：

- 该公司目前在主数据管理上存在什么问题，导致这些问题的原因是什么？
- 该公司主数据管理项目的大致实现流程是怎样的？
- 该公司的主数据类型主要有哪些？
- 项目实践过程中的主要任务有哪些，是如何实践的？
- 项目采用了哪些主数据管理方案？
- 你认为通过实施该项目，对该公司的意义是什么，可以为公司创造哪些价值？

- 主数据是整个企业范围内各个系统（包括操作或事务型应用系统及分析型系统）间**要共享的业务实体的数据**，如客户、产品、厂商、账户等信息。
- 主数据指的**不是企业内所有的业务数据**，只有那些有必要在各个系统间共享的重要数据才是主数据，如大部分的交易数据、账单数据等都不是主数据。
- 与记录业务活动，波动较大的业务数据相比，**主数据变化缓慢**。在正规的关系数据模型中，交易记录（例如，订单行项）可通过关键字（例如，订单头或发票编号和产品代码）调出主数据，类似于会员信息表、商品资源表、供应商信息表等。
- 主数据管理和治理工作重点是要尽可能共享和复用。**
- 相比较主数据，业务数据体现的是业务实时变化的过程，例如单纯的会员信息表是主数据，而会员浏览的内容明细日志，或者交易订单明细则是业务数据。



主数据（Master Data）是企业内部数据的一个子集，是跨越多个部门或多种服务被重复使用的高价值核心数据，并具有一致、准确、完整、可控的性质。

- **一致性**管理主要是为了保证多个主数据实体或数据表之间的一致性。
- **准确性和完整性**是为了保证数据如实地反映真实世界的情况。
- **可控性**是为了保证主数据不会发生非预期的变化和使用，即其更新、读/写、运算都是受监管、可追溯的。

主数据的特征

- **高价值**：主数据是所有业务处理都离不开的实体数据，其数据质量的好坏直接影响到数据集成、数据分析和数据挖掘的结果。
- **高共享**：主数据是跨部门、跨系统高度共享的数据。
- **相对稳定**：与业务数据相比，主数据是相对稳定的，变化频率较低。但并不意味着一成不变。



主数据管理的基本思路

- 主数据管理是指一整套规范、技术和方案，用于生成和维护企业主数据，并保证主数据的4个重要性质。
- 一致性**对应主数据的标准化管理，**准确性和完整性**对应主数据的质量管理，**可控性**对应主数据的安全性管理。

标准化管理

- 管理标准：**是指对企业标准化领域中需要协调统一的管理事项所制定的标准。针对主数据来说，管理标准应当包括主数据获取和组织的方法、程序和规程，以及涉及主数据获取、维护、使用的各个部门的工作内容、指责范围、工作程序、工作方法、工作质量、考核奖惩办法等。
- 业务标准：**是对主数据业务含义的统一解释及要求，包括来源、管理级次、业务环境相关的描述及含义解释、主数据相关的业务规则等。编码规则、分类规则、描述规则都属于主数据的业务标准。
- 模型标准：**包括逻辑模型和物理模型。逻辑模型主要完成从业务场景、业务概念、业务规则到主数据实体、模式、属性、规则的逻辑层映射；物理模型负责制定主数据的实际数据结构、存储结构、存储方式、读/写模式、底层数据库选取等物理层设计。

主数据管理的基本思路

质量管理

- 采用狭义的数据质量定义，即**准确性**和**完整性**。
- 为了保证主数据质量，首先要遵循严格的标准化管理。除此之外，还需要专门针对主数据质量控制进行统一的规划和监管，保证可实时跟踪和及时反馈。

安全性管理

- 数据安全性管理包括制度安全、存储安全、运算安全、传输安全、服务安全、隐私保护等多个方面。对于主数据来说，数据安全性管理是保证主数据可控性的一个重要手段。

主数据管理系统

主数据管理系统为企业数据创建单一的事实来源，帮助打破数据孤岛，改善大数据管理。构造精良、适合应用场景的主数据管理系统是实现高效主数据管理的基础。

主数据管理系统大致可以分为三类：

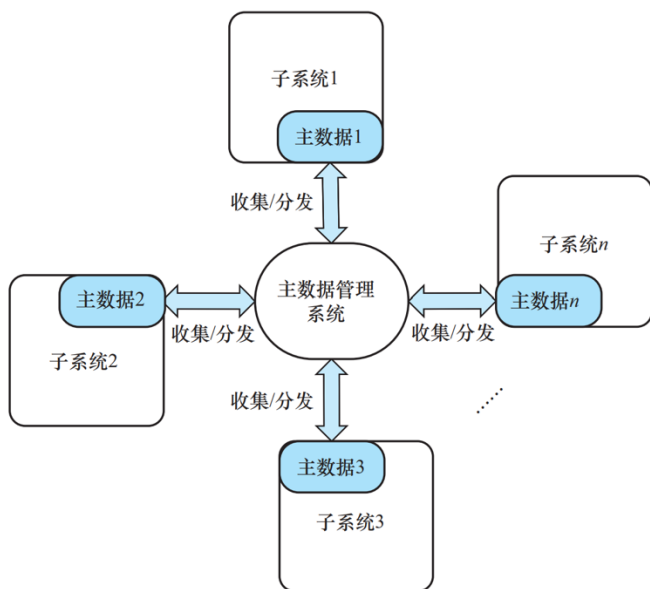
(1) 主数据管理作为某个套装软件的一部分，为套装软件的其他模块提供服务。由于主数据管理作为附属于整个软件的一个组件，其功能的完备性往往取决于套装软件本身所面向的需求。

(2) 主数据管理主要面向分析应用。其整体的完整性受制于所面向的应用。

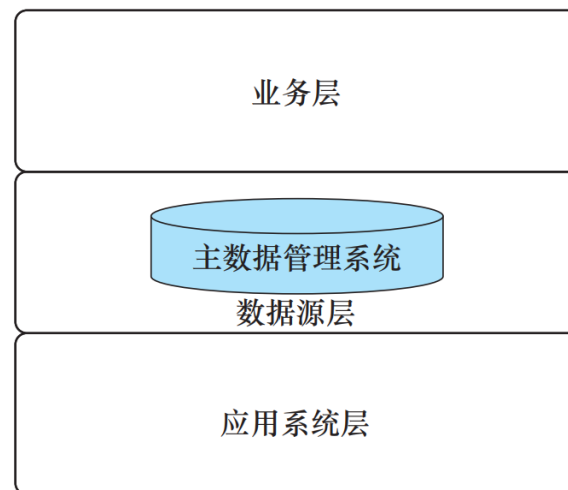
(3) 主数据管理本身自成一套独立、完整的解决方案。这种方案突出主数据本身的重要性，因此从主数据管理的角度，其整体架构和功能方面的完整性和先进性都优于前两种方案。

主数据管理系统的架构设计

◆ 在逻辑功能上，主数据在整个系统生态中处于**核心地位**，其功能是将正确的、完整的、一致的主数据及时同步分发到各子系统中。

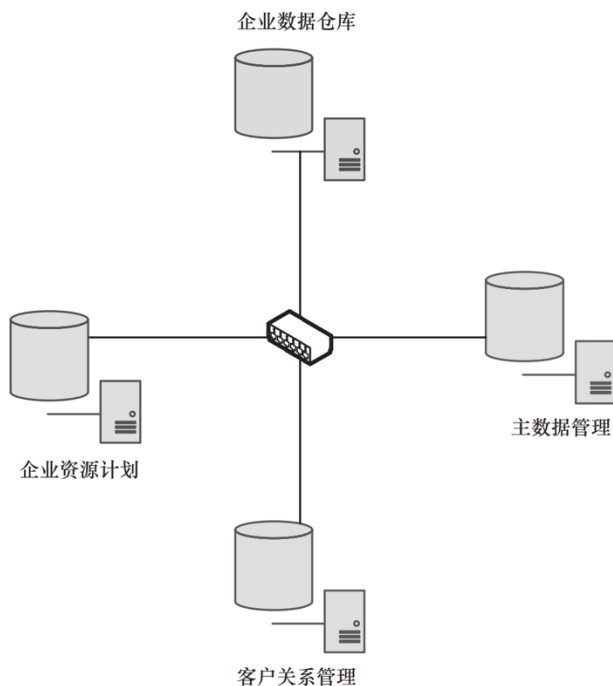


◆ 从系统的分层（layer）视图上来看主数据管理，可以认为主数据管理位于应用系统层和业务层之间，起**协调承接作用**。



主数据管理系统的架构设计

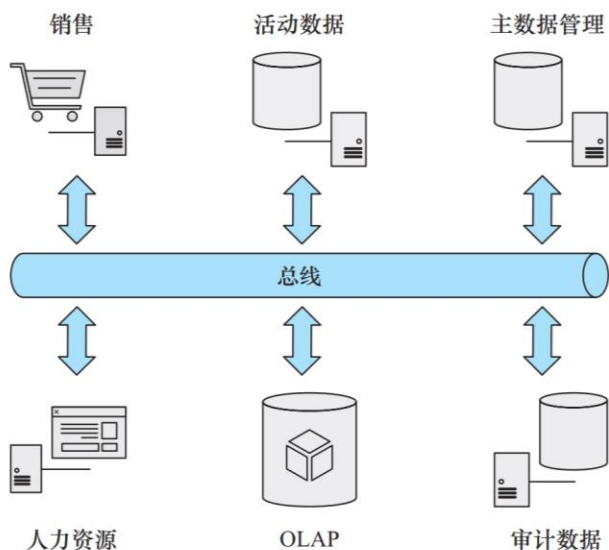
在物理上，主数据管理既可以采用**集中式管理**，也可以采用**分布式管理**。集中式数据管理使用全局唯一的主数据存储，分布式主数据管理则使用超过一个主数据存储。



1. 本地模式

- 本地模式是三种主数据管理系统架构模式中最简单的模式。
- 企业规模较小、系统组成较为简单的情况下可以直接使用本地模式，其只需要单独维护一份主数据存储，并施加相应的管理即可。
- 在非常简单的情况下，企业可能只需要维护一张或数张主数据表。

主数据管理系统的架构设计

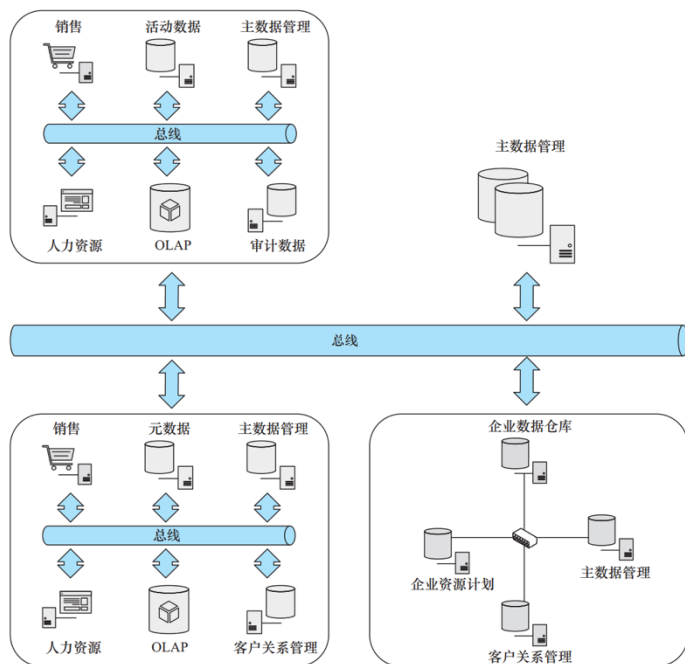


2. 联邦模式

- 联邦模式仍然设计了全局唯一的主数据管理，但在整体结构上允许不同的部门和系统使用异构的数据管理系统。
- 联邦模式适合于稍大规模的企业。
- 各个部门的系统之间可能存在着固有的异构性，消除这种异构性代价很高且在很多情况下并无必要。更理性的方案是保留这种异构性，将不同的异构系统当成不同的“联邦”个体，主数据管理也仅作为“联邦”个体之一，维护全局一致的主数据，并提供适当的主数据服务。

主数据管理系统的架构设计

3. 混合模式



- 在混合模式中，前两种模式可以共存。这种模式并不假设存在全局唯一的主数据管理，而允许存在多个主数据管理组件或子系统。
- 混合模式被设计出来以应付更复杂的情况。
- 如果企业大到了一定程度或不同的企业之间存在共享主数据的需求，那么就可以选择这种模式。
- 这种模式提供更高的自由性，不再强制全局唯一的主数据管理，允许按照具体需求定制层次化的主数据管理。

主数据管理系统的核心功能

1. 模型管理

该模块负责定义实体和属性，并维护实体、属性之间的关系和层次，包括实体模型、属性模型、业务规则、存储模型等关于主数据逻辑模型和物理模型方面的设计、管理和更新。

2. 数据整合和维护

该模块负责主数据相关的增、删、改、查的业务逻辑，包括主数据收集、清洗、融合、审核、校验等功能。在这个模块中，数据质量管理是非常重要的一个环节。

3. 数据服务

该模块负责向所有相关业务系统提供其需要的主数据服务，包括主数据查询、分发、订阅等。

4. 系统管理

该模块提供基本的安全管理、系统检测、报表分析等功能，包括用户权限管理、参数设置、系统日志管理、系统状态监控等。

主数据管理系统的实现风格

1. 合并风格

- 合并风格是将不同来源的主数据经过转换、清理、匹配和集成后，合并为一致、准确、完整的唯一视图，得到的唯一视图中的记录被称为**黄金记录**（golden record）。
- 适用于**分析型主数据管理**，目的是提供值得信赖的报告和分析数据源。
- **优点**：允许企业或组织从现有的系统中提取主数据，并进行中央集成式管理。

2. 登记风格

- 主要通过对来自各种源系统的数据运行清洗和匹配算法来发现匹配的重复记录，通过为匹配的重复记录分配唯一的**全局标识符（identity）**，该风格的主数据管理可以识别主数据的唯一版本。这种风格提供了低成本、高效、一定程度实时的主数据集成，有助于解决复杂环境中的应用。
- 登记风格和合并风格的区别在于，合并风格真正地存储黄金记录，但登记风格只存储标识符。

主数据管理系统的实现风格

3. 共存风格

- 通过与合并风格相同的方式构造黄金记录，但是主数据存储在中**央主数据管理系统**中，并在其源系统中进行更新。共存风格**允许在多处创建和存储主数据**。
- **优点：**数据可以在和谐共存的同时仍然提供真实的单一版本；主数据的质量提高，且访问速度更快；主数据单一存储，形成分析报告更加容易。

4. 集中/事务风格

- 使用链接、清洗、匹配和丰富算法来存储和维护主数据属性，以管理主数据。可以将丰富后的主数据发布回其各自的源系统。主数据 hub支持主数据记录的合并，并且源系统可以订阅由中央主数据管理系统发布的主数据更新，通过从服务接口获取的数据来更新主数据，保证完全的一致性。适用于**协同型**、**操作型**和**分析型主数据管理**。
- **优点：**维护的主数据始终准确且完整；中心主数据管理系统可以支持数据属性级别的安全性和可见性策略。

主数据管理成熟度

对企业来说，有必要提供主数据全生命周期管理，涉及如下几个关键步骤：

(1) 主数据建模

从业务、物理、UI、模板角度创建主数据模型，为主数据制定标准的编码规则。

(2) 主数据建立及共享

设置加载、分发的数据权限，对对象、属性、记录等三种级别配置数据权限。

(3) 主数据维护

基础维护、版本管理、审批流程定义执行、查重及合并、主数据日志、数据溯源等。

(4) 副本监管

副本模型的建立和维护，包括数据转换、消息处理、副本日志管理、副本分发、副本监控等。

(5) 系统管理

支持系统服务注册、业务系统注册、系统参数设置、多数据库等。

主数据管理成熟度

根据上述几个步骤实施的复杂程度，参照Jill Dyche和Evan Levy的观点大体可以把主数据管理可以分为六个层次，从低到高反映主数据管理的不同成熟度。





本章要点

- 掌握主数据概念、主数据管理的基本思路
- 了解主数据管理系统的架构设计、核心功能及实现风格
- 了解元数据管理的成熟度

课后任务

- 完成学在西电平台作业1

案例讨论

关于编码？

某企业正在讨论制定物料编码的管理方案，来自生产、财务、信息等部门的相关负责人参与了讨论，大家就“同一种物料，供应商不同，应该使用相同编码还是不同编码？”持不同意见。

生产部门的负责人表示：“在设计和生产环节，我们对不同供应商的同款物料不加以区分，因此应该使用相同编码。”

财务部门的负责人表示：“不同供应商的物料在财务核算时可能需要分开进行，因此建议使用不同编码。”

你作为信息部门的负责人，你会给出什么意见？

接下来大家又遇到了新的问题，“同一物料，型号、规格都形同，但存在不同颜色，该如何处理呢？”



小组案例作业

数据治理案例调研分析

请结合身边的现实观察（例如：学校、政务、企业/平台、社会问题、规范标准制定等），选择一个案例，从数据治理的角度以小组为单位展开调研分析，并撰写案例分析报告。报告结构可参照主数据管理章节的案例，大致包括案例背景或问题、解决方案、成效或结论等几个部分。

要求：

1. 案例调研尽可能详尽，分析案例的痛点、难点，有独立的思考和见解，避免抄袭。
2. 组内成员分工合作，需在报告中明确成员分工或贡献度。
3. 报告格式规范，数据引用有据，字数不限。
4. 各小组确定案例主题后，及时在群内发布，避免选题重复。
5. 报告提交时间为第15周（12月15日前），请将电子版报告发送至 majing@xidian.edu.cn，不用提交纸质版。
6. 课程将根据进度择优安排汇报研讨。